THE ROLE OF FEEDBACK
IN MULTI-USER INFORMATION THEORY

by
JOY A. THOMAS
STANFORD UNIVERSITY

TECHNICAL REPORT NO. 73
JUNE 1990

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION
GRANT NCR-89-14538

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
THE ROLE OF FEEDBACK
IN MULTI-USER INFORMATION THEORY

by
JOY A. THOMAS
STANFORD UNIVERSITY

TECHNICAL REPORT NO. 73
JUNE 1990

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION
GRANT NCR-89-14538

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
Abstract

Feedback does not increase the capacity of single user memoryless channels. But for memoryless multi-user channels with many senders and receivers, feedback helps the senders and receivers to cooperate more effectively and enables an increase in the total rate of communication. In this report, we will explore the various roles of feedback in communication networks.

For a multiple access channel (many senders sending information to a single receiver), we show how feedback helps the senders cooperate and show that for an \( m \)-user multiple access channel, feedback increases the total capacity by at most a factor of \( m \). For the Gaussian multiple access channel, feedback at most doubles the total capacity.

For simple multiple access channels, feedback does not increase capacity. We will prove a source channel separation theorem for such channels and show how we can analyze cascades of such channels with deterministic broadcast channels.

For a broadcast channel (a single sender sending information to many receivers), we illustrate how feedback helps to increase the capacity by enabling the sender to send common information to the receivers. We also show how feedback increases the capacity of a Gaussian interference channel (two senders and two receivers with crosstalk) and derive inner and outer bounds on its capacity region.

We describe the role of feedback in a general network of senders and receivers and describe an outer bound on the rate of flow of information in such a network. We study some of the properties of these bounds, which are stated in terms of the mutual information between subsets of random variables. We derive some inequalities for entropy rates of subsets and show how they can be used to derive new inequalities for determinants of positive matrices.

The flow of information bounds for a general network allow us to unify the earlier results on the role of feedback in multi-user communication channels. For simple multi-user networks, feedback does not help very much, but in general, feedback could open up new communication paths that are more efficient than the paths available without feedback and could increase capacity by an arbitrarily large factor.
Acknowledgements

Working with Prof. Tom Cover has been the most valuable educational experience of my life. His ideas have shaped my thinking, my tastes, my philosophy. His insight and intuition underlie many of the results in this thesis. We have spent many pleasant hours working together on the thesis and the book, *Elements of Information Theory*. I will be forever grateful for his enthusiasm and his faith in me.

I would like to thank Prof. Bob Gray and Prof. Giovanni De Micheli for their reading of this thesis and to Prof. Abbas El Gamal, Prof. John Gill and Prof. Thomas Kailath for their support and advice all through my research.

The office has always been a lively place, with plenty of technical and non-technical discussions. Thanks to William Equitz, Drew Nobel, Mitchell Trott and Jim Roche for making the office such a pleasant place to work. I had the privilege of working with other students and visitors at ISL including Sandeep Pombra, Amir Dembo and Bixio Rimoldi. I would particularly like to thank Prof. Hirosuke Yamamoto for his careful reading of the thesis and very helpful comments.

Education was only part of the Stanford experience. My friends here made all the difference. Shyam, Rita and Sandhya Pillalamarri accepted me as a member of their family. I learnt about life over countless cups of coffee with Ramana Venkata. My friends, Prasad Akella, Nita Goyal, Tamali Sengupta and others in and around Stanford were constant sources of support and encouragement. I would particularly like to thank my housemate, Ajit Paranjpe, for his patience and enthusiasm over five very enjoyable years together. My family, particularly my brother, Tony Thomas, has always been a source of love and inspiration. It is to them that I dedicate this thesis.

The research described here was supported by an NSF contract number ECS85-20136, by an IBM Graduate Fellowship and by a grant from Bell Communications Research. I would like to express my gratitude to Dr. F. Jelinek and Dr. D. Nahamoo (IBM) and Dr. B. Gopinath (Bellcore) for their assistance in funding this research.
Contents

1 Introduction .......................................................... 7
   1.1 Feedback and communication capacity ....................... 8
   1.2 Other uses of feedback ......................................... 9

2 Feedback in Single User Channels ............................. 11
   2.1 Discrete Memoryless Channels with Feedback ............... 12
   2.2 Single user Gaussian channels with feedback .............. 16
   2.3 Counterexample to factor of two for discrete channels .... 18
   2.4 Summary ...................................................... 21

3 Feedback and Multiple Access Channels ....................... 22
   3.1 Introduction .................................................. 22
   3.2 General converse for multiple access channels ............ 26
   3.3 Gaussian multiple access channels .......................... 30
      3.3.1 Case of equal powers ..................................... 33
      3.3.2 Case of unequal powers .................................. 38
   3.4 Feedback in discrete multiple access channels ............ 39
   3.5 Simple multiple access channels ............................. 41
   3.6 The source channel separation theorem ....................... 44
   3.7 Implications for networks of channels ....................... 50
   3.8 Summary ...................................................... 53

4 Feedback and Broadcast Channels .............................. 54
   4.1 Introduction .................................................. 54
   4.2 Broadcast channels with feedback ............................ 57
   4.3 Bounds on broadcast channel capacity ....................... 61
   4.4 Gaussian broadcast channels with feedback .................. 64
5 Feedback and other multiple user channels
5.1 The relay channel with feedback .......................... 72
5.2 Interference Channels ...................................... 74
  5.2.1 Interference channels without feedback ............ 75
  5.2.2 Interference channels with feedback ............... 77
5.3 The two way channel ....................................... 86
5.4 Summary ..................................................... 87

6 Feedback in a communication network .................. 88
6.1 A model for a communication network .................. 88
6.2 Properties of capacity regions without feedback .... 95
6.3 Networks with no feedback ............................... 99
  6.3.1 Networks with no common information ............... 101
6.4 Networks with complete feedback ....................... 103
6.5 Networks with single node feedback .................... 104
  6.5.1 Networks with no common information ............... 109
6.6 Properties of capacity regions with feedback ....... 109
6.7 Summary ..................................................... 111

7 Subset inequalities ....................................... 112
7.1 Introduction ............................................... 112
7.2 Subset inequalities for entropy ........................ 112
7.3 Subset inequalities for mutual information ............ 116
7.4 Subset inequalities and information flow ............. 119
  7.4.1 Multiple access channels .......................... 120
  7.4.2 General networks of channels ...................... 121
7.5 Consequences for Determinant Inequalities ............ 123
7.6 Summary ..................................................... 126

8 Other roles of feedback .................................. 127
8.1 Feedback simplifies encoding and decoding .......... 127
8.2 Finds properties of channel ............................ 129
8.3 Allows simple protocols ................................ 130
8.4 Increases zero-error capacity ......................... 131
8.5 Feedback in discrete communication..............133
8.6 Summary ........................................134

9 Summary ..........................................135

Bibliography ........................................137
List of Figures

2.1 A communication system ........................................ 12
2.2 A communication system with feedback ..................... 12
2.3 The Gaussian Channel ........................................... 16
2.4 Counterexample for discrete channels with memory ........ 19
3.1 A multiple access channel ...................................... 23
3.2 Capacity region for binary adder multiple access channel .. 24
3.3 Capacity region for a two user multiple access channel .... 29
3.4 Gaussian Multiple Access Channel ............................. 30
3.5 Bounds on the capacity of the Gaussian multiple access chan-
  nel with feedback for m = 8 senders and signal to noise ratio
  \( P/N = 1 \) ...................................................... 36
3.6 Slepian-Wolf encoding of correlated sources ................ 45
3.7 Transmission of correlated sources over a multiple access channel 46
3.8 Asymmetric multiple access channel with asymmetric inputs 47
3.9 Cascade of noiseless channels with a multiple access channel 50
3.10 Cascade network as a general 4 node network ................ 52
4.1 A Broadcast Channel ............................................ 55
4.2 Broadcast channel for Dueck example ........................ 59
4.3 Capacity regions in Dueck's example ........................ 60
4.4 Cascade of deterministic broadcast channel with a multiple
  access channel ................................................ 67
4.5 Possible configurations for the capacity regions of the deter-
  ministic broadcast channel and the multiple access channel .... 69
4.6 Cascade network as a general 4 node network ............... 71
5.1 The relay channel ............................................... 73
Chapter 1

Introduction

Feedback is a part of many communication systems. For example, during a telephone conversation, we receive feedback from the person on the other end of the line, which we can use to improve the communication. Most communication is over an inherently two-way channel, even in cases when the information transfer is only in one direction.

Even though feedback is present in many common communication situations, it is not well understood. Feedback is used in communication systems, but the theory behind the use has not been developed fully. Throughout this report, we will consider various communication systems with feedback and derive bounds on their performance.

Let us begin by defining what we mean by feedback. Intuitively, feedback in a communication channel is any mechanism that sends information back from the receiver to the sender so the sender can use the information to decide what to transmit next. In practical systems, there is normally noise and delay in the feedback link. But for the purposes of this report, we will assume that the feedback is ideal, so that all the received symbols are sent back immediately and noiselessly to all the transmitters.

There are many reasons for considering such an idealized system. The first is that most of the results that we derive are upper bounds on the performance of systems with feedback, and since any non-idealities would only degrade the performance, the upper bounds remain valid. Also, the ideal system is easier to analyze. And lastly, there are practical systems where we have near-ideal feedback. For example, in a computer network like Ethernet, all the computers listen to the same cable, so that they all receive
the same signal. Thus the senders know what the receivers have received—which corresponds to ideal feedback at no extra cost. Another example is a passive satellite that reflects everything which it receives back to earth. Again all the earth stations receive the feedback signal.

We assume that the feedback link is available at no cost. We do not directly consider the question of whether it is worthwhile to divert some of the resources to build a feedback link when none is available; but the general tenor of the results on feedback indicate that this cannot be justified on the grounds of increasing capacity. But there are a number of other advantages of feedback which might make the feedback link worthwhile. It is not possible to make a general statement about the usefulness of feedback.

1.1 Feedback and communication capacity

Throughout this report, the main focus of attention is the role of feedback in increasing capacity. We will begin by considering single user channels and describe previous results that show that feedback does not increase the capacity for a discrete memoryless channel. For channels with memory, on the other hand, feedback enables the sender to predict what will happen with the noise and thus combat the noise more effectively. We will briefly describe some results on the capacity of Gaussian colored noise channels with feedback. Then we will give an example to illustrate why the results for the Gaussian channel do not extend to a general discrete channel with memory.

The major contributions of this report are concerned with the role of feedback in increasing capacity in multi-user channels, i.e., networks with many senders and receivers. In this case, feedback helps for a very different reason, namely, feedback helps the senders and receivers to cooperate more effectively. To isolate this effect from the effect of memory in the noise, we will consider only memoryless multi-user networks. We will describe the various kinds of cooperation possible in a network and show how feedback helps convert independent information into common information, thus allowing for more effective cooperation. We will first consider the two simplest channels, the multiple access channel (many senders talking to one receiver) and the broadcast channel (one sender talking to many receivers), and prove bounds on the increase in capacity due to feedback. We will also consider the interference channel with feedback and prove an achievable rate region that is
tight at the corner points. We also describe the role of feedback for a relay channel and a two-way channel.

The results for the two-way channel motivate us to consider a general $m$-node network. In such a network, there are many possible communication paths and many different strategies that can be used. Though it is not possible to solve the network in general, we give upper bounds on the rate of flow of information between two subsets of nodes in such a network. These bounds illustrate some simple properties of networks with and without feedback.

The bounds derived for a general network are stated in terms of the mutual information between inputs and outputs corresponding to different subsets of the nodes. That motivates us to study the properties of the entropy rate of subsets of random variables. We will also derive some simple applications to these subset inequalities to prove determinant inequalities for submatrices of a positive definite matrix.

1.2 Other uses of feedback

In addition to increasing capacity, feedback has a number of other uses in communication.

- Feedback simplifies encoding and decoding.
- Feedback enables the sender to determine the characteristics of the channel.
- Feedback is used in simple protocols that allow a large number of users to share a common channel with low delay.
- Feedback enables the senders to monitor the traffic on a communication channel and to control it to ensure that the system is stable.
- For the communication of information with no error (zero-error capacity), feedback helps increase the capacity even for the memoryless channel.
- Feedback helps reduce communication complexity (the total number of bits transmitted) when the receiver has side information.
In this report, we will deal mainly with the increase in capacity of communication systems due to feedback. We will describe the other uses of feedback in greater detail in Chapter 8.
Chapter 2

Feedback in Single User Channels

Communication can be defined as a process by which information is sent from one point to another so as to produce a desired state of knowledge at the receiver. In this report, we will consider the model of a communication system (Figure 2.1) introduced by Shannon[72]. In this model, the channel is described as a system whose output depends probabilistically on the input. The sender transmits a sequence of input symbols that depend on the message that he wants to send, and the receiver looks at the output of the channel and tries to decide which message was sent. Shannon showed that associated with the channel was a capacity which depended only on the probabilistic relationship of the input and output of the channel, such that it was possible to send messages with an arbitrarily low probability of error at any rate below capacity and that at rates above capacity, the probability of error was bounded away from zero. The capacity of the discrete memoryless channel without feedback was shown to be

\[ C = \max_{p(x)} I(X; Y) = \max_{p(x)} \sum_{x,y} p(x)p(y|x) \log \frac{p(x)p(y|x)}{p(x)p(y)}. \]  

In 1956, Shannon[73] considered the discrete memoryless channel with feedback and showed that feedback does not increase capacity. As this is perhaps the most important result on the role of feedback in communication, we will describe it in greater detail in the next section.
2.1 Discrete Memoryless Channels with Feedback

Shannon [73] showed that feedback does not increase the capacity of a discrete memoryless channel. This result is rather surprising since it does not match our intuition about the role of feedback in communication. We expect that feedback would help the sender know what the receiver's uncertainty is and therefore would enable the sender to tailor his transmissions to optimally correct the receiver's uncertainty. But Shannon's result shows that this does not change the maximum rate of transmission of information.

A channel with feedback is illustrated in Figure 2.2. We assume that all the received symbols are sent back instantaneously and noiselessly to the transmitter, who can then use them to decide which symbol to send next.

We will begin with a few definitions of feedback codes and feedback capacity. These definitions can be extended in the obvious way to the multiple user channels considered in the later chapters in the report.
**Definition:** An \((M, n)\) feedback code for the channel \((\mathcal{X}, p(y|x), \mathcal{Y})\) consists of the following

1. An index set \(\{1, 2, \ldots, M\}\),
2. An encoding function \(x_i : \{1, 2, \ldots, M\} \times \mathcal{Y}^{i-1} \to \mathcal{X}\), and
3. A decoding function

\[
g : \mathcal{Y}^n \to \{1, 2, \ldots, M\},
\]

which is a deterministic rule which assigns a guess to each possible received vector.

**Definition:** Let

\[
\lambda_w = \Pr(g(Y^n) \neq w | X^n = x^n(w))
\]

be the conditional probability of error given that index \(i\) was sent.

**Definition:** The (arithmetic) average probability of error \(P_e^{(n)}\) for an \((M, n)\) code is defined as

\[
P_e^{(n)} = \frac{1}{M} \sum_{w=1}^{M} \lambda_w.
\]

Note that

\[
P_e^{(n)} = \Pr(W \neq g(Y^n)),
\]

if the index \(W\) is chosen uniformly on the set \(\{1, 2, \ldots, M\}\).

**Definition:** The rate \(R\) of an \((M, n)\) code is

\[
R = \frac{\log M}{n} \text{ bits/transmission.}
\]

**Definition:** A rate \(R\) is said to be achievable with feedback if there exist a sequence of \((\lfloor 2^{nR} \rfloor, n)\) feedback codes such that the average probability of error \(P_e^{(n)}\) tends to 0 as \(n \to \infty\).
Later, we will write \((2^{nR}, n)\) codes to mean \((2^{nR}, n)\) codes to simplify notation.

**Definition:** The (operational) capacity with feedback, \(C_{FB}\), of a discrete memoryless channel is the supremum of all rates achievable by feedback codes.

**Theorem 2.1.1** The (operational) capacity of a discrete memoryless channel is not increased by feedback. Thus

\[
C_{FB} = C = \max_{p(x)} I(X; Y). \tag{2.7}
\]

**Proof:** Since a non-feedback code is a special case of a feedback code, any rate that can be achieved without feedback can be achieved with feedback, and hence

\[
C_{FB} \geq C. \tag{2.8}
\]

Proving the inequality the other way is slightly more tricky. Let \(W\) be uniformly distributed over \(\{1, 2, \ldots, 2^nR\}\). Then

\[
nR = H(W) = H(W|Y^n) + I(W; Y^n) \leq 1 + P_e^n nR + I(W; Y^n), \tag{2.9}
\]

by Fano’s inequality[36], [23].

Now we can bound \(I(W; Y^n)\) as follows:

\[
I(W; Y^n) = H(Y^n) - H(Y^n|W) \tag{2.11}
\]

\[
= H(Y^n) - \sum_{i=1}^{n} H(Y_i|Y_1, Y_2, \ldots, Y_{i-1}, W) \tag{2.12}
\]

\[
= H(Y^n) - \sum_{i=1}^{n} H(Y_i|Y_1, Y_2, \ldots, Y_{i-1}, W, X(f)) \tag{2.13}
\]

\[
= H(Y^n) - \sum_{i=1}^{n} H(Y_i|X_i), \tag{2.14}
\]

since \(X_i\) is a function of \(Y_1, \ldots, Y_{i-1}\) and \(W\); and, conditioned on \(X_i, Y_i\) is independent of \(W\) and past samples of \(Y\). Then using...
the fact that the entropy of a collection is less than the sum of the individual entropies, we have

\[ I(W; Y^n) = H(Y^n) - \sum_{i=1}^{n} H(Y_i|X_i) \]  
\[ \leq \sum_{i=1}^{n} H(Y_i) - \sum_{i=1}^{n} H(Y_i|X_i) \]  
\[ = \sum_{i=1}^{n} I(X_i; Y_i) \]  
\[ \leq nC \]  

from the definition of capacity for a discrete memoryless channel. Putting these together, we obtain

\[ nR \leq P_e^{(n)} nR + 1 + nC, \]  

and dividing by \( n \) and letting \( n \to \infty \), we conclude

\[ R \leq C. \]  

Thus we cannot achieve any higher rates with feedback than we can without feedback, and

\[ C_{FB} = C. \]  

\[ \square \]

Thus feedback does not increase the capacity of discrete memoryless channels. One intuitive explanation of this is that when the noise samples are independent, feedback does not give any information about the behavior of the noise in the future. It is then as efficient to send new information as to send correction information. In channels with memory, this is not true. In this case, feedback enables the sender to learn something about the future behavior of the noise and to combat it more effectively. For channels with memory, feedback can help increase the capacity, as illustrated by the example in Section 2.3.
2.2 Single user Gaussian channels with feedback

A common model for a continuous alphabet communication channel is the Gaussian channel, where the output of the channel at time $i$,

$$Y_i = X_i + Z_i$$

(2.22)

where the $\{Z_i\}$ forms a Gaussian process. The input is assumed to satisfy a power constraint $P$, which means that all the codewords are assumed to have an average power less than $P$, i.e.,

$$\frac{1}{n} \sum_{i=1}^{n} x_i^2(w) \leq P$$

(2.23)

for all $w \in \{1, \ldots, 2^{nR}\}$.

If $\{Z_i\}$ forms an i.i.d. Gaussian process with variance $N$, this channel is called an Additive White Gaussian Noise (AWGN) channel. The capacity of this channel without feedback was found by Shannon[72] to be

$$C = \frac{1}{2} \log \left(1 + \frac{P}{N}\right) \text{ bits/transmission.}$$

(2.24)
Kadota, Zakai and Ziv[47] showed that feedback does not increase the capacity of continuous memoryless channels, and in particular, the AWGN channel. Kailath and Schalkwijk[49] described a simple feedback coding scheme which achieves the capacity of the AWGN channel. In this scheme, the sender attempts to send a real number to the receiver. Using the feedback link, the sender can calculate the same estimate of the message as the receiver, and it sends a scaled version of the error to the receiver in the next transmission. The receiver uses the noisy version of this transmission to update its estimate of the message. A detailed analysis of this scheme shows that the estimate gets more accurate with each transmission, and that the probability of error decreases doubly exponentially with the block length. Schemes similar to the Kailath Schalkwijk scheme were suggested for other channels with feedback[71], [87].

If the \( Z_i \) are not independent or identically distributed, the channel is called a colored noise channel. The capacity of the stationary colored noise channel without feedback was found by Holsinger[44], who analyzed the problem in the spectral domain, and showed that optimum procedure was to put the signal power in the valleys of the power spectrum of the noise.

Butman[9],[10] analyzed coding schemes for the colored noise channel with feedback and gave examples where feedback increased the capacity of the channel. Ebert[31] and Pinsker(unpublished) showed that feedback at most doubles the capacity of the colored noise Gaussian channel. Various authors have proposed and analyzed coding schemes for Gaussian channels with feedback; a good survey of the early work on this subject can be found in Lucky[56].

Cover and Pombra[22] analyzed the Gaussian channel with feedback and defined the capacity of the channel with and without feedback in terms of the covariance matrices of the input and the output of the channel. They then proved two simple bounds on the capacity of the channel with feedback:

\[ C_{FB} \leq C_{NFB} + \frac{1}{2} \text{bits.} \]  
\[ C_{FB} \leq 2C_{NFB} \]  

**Theorem 2.2.1**

**Proof:** See [22]. \( \Box \)
They also showed the achievable capacity with and without feedback as defined in terms of the covariance matrices. Ihara[45] extended the investigation and gave an example of a channel where the factor of 2 bound is tight. Dembo[26] gave more accurate bounds on the capacity of the channel with feedback. Ozarow[65] has analyzed new coding schemes for the channel with feedback and in [64] proved some bounds on the capacity of the Gaussian channel with feedback.

Although the colored noise Gaussian channel with feedback has been extensively studied, there is still no explicit characterization for the capacity of the channel with feedback. In [22], the capacity is given as a solution to a maximization problem over a set of covariance matrices that satisfy the power constraint. This characterization was used by Dembo[26] and Ihara[46] to derive bounds on the capacity of the channel with feedback. There are still a number of open questions in the area and this remains an important area of study.

2.3 Counterexample to factor of two for discrete channels

The simple bounds on the increase in capacity due to feedback for a Gaussian channel may cause us to believe that feedback only helps by a small factor for all channels. But we can modify a simple example due to Dobrushin[27] to show that feedback can help by an arbitrary amount for a general channel.

Example 2.3.1:

Consider the channel shown in Figure 2.4. This channel has $M$ input symbols \{1, 2, \ldots, M\} and 2 output symbols \{0, 1\}. The channel has $M$ states, which correspond to one of the input symbols (say i) getting mapped onto 1 and all the other inputs getting mapped onto 0. If we knew the state, i.e., which symbol was mapped onto 1, then we can send 1 bit per transmission through this channel. But if we did not know which symbol was mapped to 1, then the best we can do is use all the inputs equally. We
Figure 2.4: Counterexample for discrete channels with memory
calculate the capacity as
\[
C = \max_{p(x)} I(X; Y) = \max_{p(x)} H(Y) = H\left(\frac{M-1}{M}, \frac{1}{M}\right) \approx \frac{\log M}{M}.
\]
(2.27)

This is the capacity of the channel without feedback when we do not know the state of the channel.

With feedback, on the other hand, it is easy to use the feedback to find the state of the channel (for example, by sending a test sequence of length $M$). After finding the state of the channel, we can send 1 bit/transmission. So with feedback, the capacity of the channel is 1 bit/transmission.

Feedback has thus increased the capacity of the channel by a factor of $M/\log M$. We can therefore obtain an arbitrary multiplicative factor for the increase in capacity due to feedback. We can also modify the example to obtain an arbitrary additive factor for the increase in capacity due to feedback (by considering a number of parallel channels, for example).

Thus we see that the fact that feedback does not help too much for Gaussian channels does not extend to all channels.

It may be argued that the example above is somewhat pathological. For example, the channel is not ergodic.

**Definition:** A channel is said to be ergodic if an input ergodic source produces a jointly ergodic input-output process[39].

But we can easily modify the example to make the channel ergodic. Let the state of the channel (which input symbol gets mapped to 1) vary very slowly with time as the state of an irreducible Markov chain. In this case, the sequence of states of the channel forms an ergodic process and we can show that the channel is ergodic. But if the state of the channel varies very slowly, the sender can use the feedback link to learn the state at use it to increase capacity as in the non-ergodic case. Thus we can also construct ergodic channels with an arbitrary increase in capacity due to feedback.
There are practical channels where the state of the channel is unknown and the capacity of the channel depends to a large degree on the knowledge of the state. In such channels, feedback may make a large difference to the capacity. For example, in telephone line modems, the channel characteristics can vary with the switching, crosstalk, etc. Feedback is used by the equalizers in the modem to get rid of some of these effects and thus increase the capacity. Thus the example above is not very unrealistic; feedback can help considerably in increasing capacity.

2.4 Summary

In this chapter, we reviewed earlier results on the increase in capacity for a single user channel. We have shown that feedback does not increase the capacity of a memoryless channel, but can help for channels with memory. For the Gaussian channel, there are simple bounds on the increase in capacity due to feedback, but these bounds do not generalize to a general discrete channel with memory.

In the rest of the report, we will consider the capacity of multi-user channels with feedback. To isolate the effect of the dependencies in the noise from the effect of the cooperation in the network, we will only consider memoryless networks in the rest of this report. If there is memory in the network, feedback can be used to exploit the memory and increase capacity in the same way as in a single-user communication channel. We will begin our discussion of networks by describing the simplest and best understood multi-user channel, the multiple access channel.
Chapter 3

Feedback and Multiple Access Channels

3.1 Introduction

In situations like satellite communication, we have many senders trying to communicate to a single receiver - a multiple access channel. The channel is illustrated in Figure 3.1.

The multiple access channel is the simplest and best understood multiple user communication channel. The capacity region of this channel without feedback was found by Ahlswede[3] and Liao[54]. In the two user case, it is given by the following theorem:

Theorem 3.1.1 (Multiple Access Channel Capacity).

The capacity of a multiple access channel \( \mathcal{X}_1 \times \mathcal{X}_2; p(y|x_1, x_2), Y \) is the closure of all \((R_1, R_2)\) satisfying

\[
R_1 < I(X_1;Y|X_2,Q) \tag{3.1}
\]

\[
R_2 < I(X_2;Y|X_1,Q) \tag{3.2}
\]

\[
R_1 + R_2 < I(X_1,X_2;Y|Q) \tag{3.3}
\]

for some product distribution \(p(q)p_1(x_1|q)p_2(x_2|q)\) on \(\mathcal{X}_1 \times \mathcal{X}_2\), where \(Q\) is an auxiliary time-sharing random variable.

Proof: See [23]. \(\square\)
We will call the region in the above theorem $R_f$. It is the capacity of the channel without feedback. The case of the multiple access channel with common information was solved by Slepian and Wolf[75] and for some other special source hierarchies by Han[41].

Unlike the simple discrete memoryless channel, feedback in the multiple access channel can increase capacity even when the channel is memoryless. This is because feedback would enable the senders to cooperate with each other to a greater extent than is possible without feedback. This was first demonstrated by Gaarder and Wolf[37]. Cover and Leung[21] proved an achievable rate region for the multiple access channel with feedback. Later, Willems[82] proved that the Cover-Leung region is indeed that capacity region for a certain class of channels including the binary adder channel. Ozarow[63] has found the capacity region for the two user Gaussian multiple access channel, using a modification of the Kailath-Schalkwijk scheme for simple Gaussian channels. However, his method does not generalize easily to three or more users. The problem of determining the capacity region for a multiple access channel with feedback remains open.

We will illustrate why feedback helps increase capacity by using the original example of Gaarder and Wolf[37].
Example 3.1.1: Consider the binary adder multiple access channel, which is a two user channel with binary inputs $X_1$ and $X_2$ and a ternary output $Y$ which is the sum (real addition) of the two inputs defined by

$$ Y = X_1 + X_2. \quad (3.4) $$

We can calculate the capacity region as given in Theorem 3.1.1. The region is illustrated in Figure 3.2.

The maximum achievable sum of rates without feedback is 1.5 bits per transmission.

If the two senders cooperate to send a common message, then they can treat this as a single-user channel. In that case, the capacity is $C = \max_{p(x_1,x_2)} I(X_1, X_2; Y) = \log 3 = 1.58496$ bits per transmission. Thus the senders can send at a higher total rate when they cooperate than when they send independent messages.
What happens when we have feedback? Feedback enables each sender to see what the other sender is sending. For example, Sender 1 can try to determine the message that is being sent by Sender 2. Since Sender 1 knows which codeword $X^*_1$ he is sending, he can “subtract” it out from the received signal $Y^n$ and thus have a better idea of the second sender’s message than the receiver. After the senders have decoded each other’s messages, they can cooperate to resolve the uncertainty at the receiver.

For the binary adder channel, let the two senders send a sequence of $n$ independent equiprobable bits. The receiver $Y$ will receive a sequence of ternary digits, about $1/4$ of which will be 0’s, $1/2$ will be 1’s and $1/4$ will be 2’s. When the output is 0 or 2, the receiver knows immediately which two inputs were sent. When the output of the channel is 1, the receiver will not know whether $(X_1, X_2) = (1, 0)$ or $(0, 1)$. Using the feedback link, each of the senders knows the output and can easily calculate what the other sender has sent.

At the end of $n$ transmissions, each of the senders knows the $n$ bits that were sent by the other sender, but the receiver has about $n/2$ bits of uncertainty corresponding to the times when the output was 1. The two senders now cooperate to resolve this uncertainty at the receiver. Since they can cooperatively send information at $\log 3$ bits/transmission, they can resolve this uncertainty using another $n/(2\log 3)$ transmissions. With this method, it is possible to send $2n$ bits of information in $n + n/(2\log 3)$ transmissions, giving an effective sum of rates of 1.52 bits per transmission, which is higher than the maximum rate achievable without feedback.

For this channel, Willems\[82\] has shown that the Cover-Leung region is optimal and in \[83\] showed that the best achievable sum of rates is 1.58226 bits per transmission.

Gaarder and Wolf\[37\] showed the capacity of a multiple access channel with feedback is bounded by

$$R_1 \leq I(X_1; Y|X_2)$$  \hspace{1cm} (3.5)
\[ R_2 \leq I(X_2; Y|X_1) \]  
\[ R_1 + R_2 \leq I(X_1, X_2; Y) \]

for some joint distribution \( p(x_1, x_2)p(y|x_1, x_2) \). We will call the union of these regions over all joint distributions \( R_O \). This is an outer bound on the capacity of the multiple access channel with feedback.

In this chapter, we prove bounds on the capacity increase with feedback for multiple access channels. As we shall show in the next section, feedback does not increase the rate of individual senders — it only helps them to cooperate and keep out of each other’s way. Hence the only increase will be in the sum of the rates. Throughout this chapter, we will define the capacity of the multiple access channel to be the maximum achievable sum of rates of all the senders.

### 3.2 General converse for multiple access channels

In this section we will extend the converse due to Ozarow[63] and Gaarder and Wolf[37] to the \( m \)-user case and use it to show that capacity cannot be increased by a factor of more than \( m \) by feedback.

The general multiple access channel is characterized by an input alphabet \( (\mathcal{X}_1 \times \mathcal{X}_2 \times \ldots \times \mathcal{X}_m) \), a probability transition matrix \( p(y|x_1, x_2, \ldots, x_m) \) and an output alphabet \( \mathcal{Y} \) (Figure 3.1).

To simplify notation, let \( S \) denote any arbitrary subset of \( \{1, 2, \ldots, m\} \). Let \( X_S \) denote the set \( \{X_j : j \in S\} \). For example, if \( S = \{1, 3\} \), then \( X_S = \{X_1, X_3\} \). Let \( W_1, W_2, \ldots, W_m \) denote the input messages, each \( W_j \) uniformly distributed in \( (1, 2^nR_j) \) and independent of the other messages. Let \( W_S \) denote the set \( \{W_j : j \in S\} \).

Since we have feedback, the input symbol \( X_{ji} \) of sender \( j \) at time \( i \) is a function of the message at that sender \( W_j \) and the past values \( Y_1, Y_2, \ldots, Y_{i-1} \) of the output. Thus \( X_{ji} = X_{ji}(W_j, Y^{i-1}) \).

**Theorem 3.2.1** The rates of any feedback code for a multiple access channel must satisfy

\[ \sum_{j \in S} R_j \leq I(X_S; Y|X_S) \]  

for some joint distribution \( p(x_1, x_2, \ldots, x_m)p(y|x_1, x_2, \ldots, x_m) \), for all subsets \( S \) of the senders.
Proof: By Fano's inequality[36], for any feedback code, we have

\[ H(W_S|Y^n) \leq n\epsilon_n, \]  

(3.9)

with \( \epsilon_n \to 0 \) as \( n \to \infty \). Thus

\[
\sum_{j \in S} nR_j = H(W_S) 
\leq I(W_S; Y^n) + n\epsilon_n 
= H(W_S) - H(W_S|Y^n) + n\epsilon_n 
\leq (a) H(W_S|W_{\overline{S}}) - H(W_S|Y^n, W_{\overline{S}}) + n\epsilon_n 
= I(W_S; Y^n|W_{\overline{S}}) + n\epsilon_n 
\leq (b) \sum I(W_S, X_{S_i}; Y_i|Y^{i-1}, W_{\overline{S}}, X_{\overline{S}_i}) + n\epsilon_n 
\leq (c) \sum I(X_{S_i}; Y_i|X_{\overline{S}_i}) + n\epsilon_n, 
\]  

(3.10)

(3.11)

(3.12)

(3.13)

(3.14)

(3.15)

(3.16)

where (a) follows from the fact that \( W_S \) and \( W_{\overline{S}} \) are independent, (b) from the fact that \( X_{S_i} \) is a function of \( (W_S, Y^{i-1}) \) and \( X_{\overline{S}_i} \) is a function of \( (W_{\overline{S}}, Y^{i-1}) \), and (c) follows since \( (Y^{i-1}, W_S, W_{\overline{S}}) \rightarrow (X_{S_i}, X_{\overline{S}_i}) \rightarrow Y_i \) forms a Markov chain.

Now let \( Q \) be a random variable uniformly distributed on \( \{1, 2, \ldots, n\} \), independent of all the other random variables. Define \( X_j \overset{\text{def}}{=} X_{jQ} \) and \( Y \overset{\text{def}}{=} Y_Q \). Then

\[
\sum_{j \in S} R_j \leq \frac{1}{n} \sum I(X_{S_i}; Y_i|X_{\overline{S}_i}) + \epsilon_n 
\leq I(X_{S_Q}; Y_{X_{\overline{S}_Q}}, Q) + \epsilon_n 
= I(X_S; Y|X_{\overline{S}}) + \epsilon_n 
\leq I(X_S; Y|X_{\overline{S}}) + \epsilon_n 
\]  

(3.17)

(3.18)

(3.19)

(3.20)

by the Markovian relation \( Q \rightarrow (X_S, X_{\overline{S}}) \rightarrow Y \) induced by the channel and the definition of \( Q \). Now letting \( n \to \infty \), we get the desired converse

\[
\sum_{j \in S} R_j \leq I(X_S; Y|X_{\overline{S}}) 
\]  

(3.21)
for some joint distribution \( p(x_1, x_2, \ldots, x_m)p(y|x_1, x_2, \ldots, x_m) \), for all subsets \( S \). □

This converse is a special case of the general converse for an \( m \)-node network that will be proved later (Theorem 6.5.1).

Since we have not placed any restrictions on the joint distribution, the converse is not tight in general. The basic problem in determining multiple access channel capacity with feedback is to find the class of joint distributions achievable using feedback.

The capacity region of a discrete memoryless multiple access channel without feedback was first determined by Ahlswede[3] and Liao[54]. It is given by the convex closure of regions bounded by (3.8) for arbitrary input product distributions \( p(x_1)p(x_2)\ldots p(x_m)p(y|x_1, x_2, \ldots, x_m) \).

We will derive a few simple consequences of this converse.

- Feedback does not increase the individual rates of the senders. We will illustrate it for \( m = 2 \) users.

\[
I(X_1; Y|X_2) = \sum_{x_2} P(X_2 = x_2)I(X_1; Y|X_2 = x_2) \tag{3.22}
\]

\[
\leq \max_{x_2} I(X_1; Y|X_2 = x_2) \tag{3.23}
\]

Hence

\[
\max_{p(x_1,x_2)p(y|x_1,x_2)} I(X_1; Y|X_2) = \max_{p(x_1)p(x_2)p(y|x_1,x_2)} I(X_1; Y|X_2) \overset{\text{def}}{=} R_{1c},
\]

since the maximum is achieved by a degenerate product distribution obtained by setting \( X_2 \) to the value that best opens up the channel between \( X_1 \) and \( Y \). Hence the maximum rate from \( X_1 \) to \( Y \) is not increased by feedback. This can be easily justified by the fact that since \( X_2 \) does not get any better look at \( X_1 \) than does \( Y \), he cannot help any more than by opening up the channel as best as he can. We can similarly define \( R_{2c} \) as the maximum rate of transmission from \( X_2 \) to \( Y \).

- Feedback increases the total capacity (sum of the rates of all the senders) by at most a factor of \( m \) for any \( m \)-user multiple access channel.
Figure 3.3: Capacity region for a two user multiple access channel

We have shown that feedback does not increase the maximum individual rates of any senders. Ahlswede’s results imply that the rate pairs $(R_{1c}, 0)$ and $(0, R_{2c})$ (points A and B on Figure 3.3) are achievable without feedback. By timesharing, one can achieve all points on the line joining A to B.

With feedback, the converse implies that the individual rates are less than $R_{1c}$ and $R_{2c}$ respectively. Therefore the capacity region with feedback lies within the rectangle defined by the points A, B and C. The maximum sum of rates within this rectangle is at C, which is less than twice the sum of the rates at D, the midpoint of the main diagonal. Since D is achievable without feedback, the maximum achievable sum of rates with feedback is less than twice the maximum achievable sum of rates without feedback.

One can easily generalize this argument to $m$-users to obtain a factor of $m$ as the bound on the increase in capacity using feedback.
3.3 Gaussian multiple access channels

\[ Y_i = \sum_{j=1}^{m} X_{ji} + Z_i \]  
(3.25)

where the \( Z_i \) are i.i.d. \( \sim N(0, N) \) (Figure 3.4). There is a power constraint \( P_j \) on each of the senders, i.e. for all senders, messages and output sequences, we must have

\[ \frac{1}{n} \sum_{i=1}^{n} X_{ji}^2(W_j, Y_{i-1}) \leq P_j \]  
(3.26)

The capacity of the Gaussian channel without feedback was first proved by Cover[18] and Wyner[86]. The capacity region can be written as

\[ \sum_{j \in S} R_j \leq C \left( \frac{\sum_{j \in S} P_j}{N} \right), \]  
(3.27)

for every subset \( S \) of \( \{1, 2, \ldots, m\} \), and

\[ C(x) \overset{\text{def}}{=} \frac{1}{2} \log(1 + x) \]  
(3.28)
The dominant bound is the first one and the total capacity without feedback is $C(\sum_j P_j/N)$. Note that the capacity with total cooperation is

$$C \left( \frac{\left( \sum_j \sqrt{P_j} \right)^2}{N} \right),$$

(3.29)

which can be much larger than $C(\sum_j P_j/N)$. The main objective of this section is to show that even with feedback, the Gaussian multiple access channel cannot attain the cooperative bound and in fact is at most a factor of 2 better than the capacity without feedback.

Ozarow[63] has determined the feedback capacity region for the two user case, using a modification of the Kailath-Schalkwijk scheme. His expression for the capacity region is

$$R_1 \leq C \left( \frac{P_1(1 - \rho^2)}{N} \right)$$

(3.30)

$$R_2 \leq C \left( \frac{P_2(1 - \rho^2)}{N} \right)$$

(3.31)

$$R_1 + R_2 \leq C \left( \frac{P_1 + P_2 + 2\sqrt{P_1 P_2 \rho}}{N} \right)$$

(3.32)

for all $0 \leq \rho \leq 1$.

We will now prove that feedback at most doubles the total capacity of the Gaussian multiple access channel. We first extend the converse of the last section to the Gaussian case. We will need a simple lemma.

**Lemma 3.3.1** Let $X_1, X_2, \ldots, X_k$ be an arbitrary set of zero mean random variables with covariance matrix $K$. Let $S$ be any subset of $\{1, 2, \ldots, k\}$. Then

$$h(X_S | X_{\bar{S}}) \leq h(X^*_S | X^*_{\bar{S}})$$

(3.33)

where $(X^*_1, X^*_2, \ldots, X^*_k) \sim N(0, K)$.

**Proof:**

$$0 \leq D(f_{X_S|X_{\bar{S}}}(x_S|x_{\bar{S}}) \| f_{X^*_S|X^*_{\bar{S}}}(x_S|x_{\bar{S}}))$$

(3.34)
\[ = \int f_{X_S, X_{\overline{S}}}(x_S, x_{\overline{S}}) \log \frac{f_{X_S|X_{\overline{S}}}(x_S|x_{\overline{S}})}{f_{X_{\overline{S}}|X_S}(x_S|x_{\overline{S}})} \]  

(3.35)

\[ = -h(X_S|X_{\overline{S}}) - \int f_{X_{\overline{S}}}(x_{\overline{S}}) f_{X_S|X_{\overline{S}}}(x_S|x_{\overline{S}}) \log (f_{X_S|X_{\overline{S}}}(x_S|x_{\overline{S}})) \]

(a) \[ = -h(X_S|X_{\overline{S}}) \]

\[ - \int f_{X_S, X_{\overline{S}}}(x_S, x_{\overline{S}}) \left( -\frac{n}{2} \log 2\pi - \frac{1}{2} \log |\hat{K}| \right) \]

\[ + \frac{1}{2} (x_S - Ax_{\overline{S}})' \hat{K}^{-1} (x_S - Ax_{\overline{S}}) \]

(b) \[ = -h(X_S|X_{\overline{S}}) \]

\[ - \int f_{X_{\overline{S}}, X_S}(x_S, x_{\overline{S}}) \left( -\frac{n}{2} \log 2\pi - \frac{1}{2} \log |\hat{K}| \right) \]

\[ + \frac{1}{2} (x_S - Ax_{\overline{S}})' \hat{K}^{-1} (x_S - Ax_{\overline{S}}) \]

\[ = -h(X_S|X_{\overline{S}}) - \int f_{X_{\overline{S}}, X_S}(x_S, x_{\overline{S}}) \log (f_{X_S|X_{\overline{S}}}(x_S|x_{\overline{S}})) \]

\[ = -h(X_S|X_{\overline{S}}) + h(X_S^*|X_{\overline{S}}^*) \]  

(3.36)

where

(a) follows from the definition of a conditional normal density,

where \( Ax_{\overline{S}} \) is the conditional mean and \( \hat{K} \) is the conditional covariance matrix of \( X_S \), and

(b) is true since \( \log (f_{X_S|X_{\overline{S}}}(x_S|x_{\overline{S}})) \) is a quadratic form in \( K \). The proof of the lemma is complete. \( \square \)

Starting from (3.16) in the general converse

\[ \sum_{j \in S} nR_j \leq \sum I(X_{Si}; Y_i|X_{\overline{Si}}) + n\epsilon_n \]  

(3.37)

\[ = \sum h(Y_i|X_{\overline{Si}}) - h(Y_i|X_{Si}, X_{\overline{Si}}) + n\epsilon_n \]  

(3.38)

\[ \leq \sum h(Y_i|X_{\overline{Si}}) - h(Z_i) + n\epsilon_n \]  

(3.39)

Now using the same method as in the general case, defining the random variable \( Q \), etc., we get

\[ \sum_{j \in S} R_j \leq h(Y_S|X_{\overline{S}Q}, Q) - \frac{1}{2} \log ((2\pi\epsilon)^N) + \epsilon_n \]  

(3.40)
\[
\begin{align*}
\leq & \ h(Y_q | X_{\tilde{S}Q}) - \frac{1}{2} \log((2\pi e)N) + \epsilon_n \quad (3.41) \\
\leq & \ h(Y^* | X^*_S) - \frac{1}{2} \log((2\pi e)N) + \epsilon_n, \quad (3.42)
\end{align*}
\]

by Lemma 3.3.1. Now letting \( n \to \infty \), we can get rid of the \( \epsilon \)'s. The covariance matrix of the \( X^*_S, X^*_\tilde{S} \) is the average of the covariance matrices for each time instant and should therefore have diagonal elements \( \leq P \) to satisfy the power constraint (3.26).

Hence, dropping the * for simplicity, we get

\[
\sum_{j \in S} R_j \leq \ h(Y | X_{\bar{S}}) - \frac{1}{2} \log((2\pi e)N) \quad (3.43)
\]

\[
= \frac{1}{2} \log \left( (2\pi e) \text{var}(Y | X_{\bar{S}}) \right) - \frac{1}{2} \log ((2\pi e)N) \quad (3.44)
\]

\[
= \frac{1}{2} \log \left( \frac{\text{var}(Y | X_{\bar{S}})}{N} \right). \quad (3.45)
\]

Thus we can bound the total capacity in terms of the covariance of the inputs and outputs.

### 3.3.1 Case of equal powers

We will first consider the case when all the powers are equal, i.e., \( P_1 = P_2 = \cdots = P_m = P \), so that all the senders are equivalent. Let \( X = (X_1, X_2, \ldots, X_m)^t \) be the vector of inputs. Let \( K \) be the covariance matrix of \( X \). We shall first show that one of the possible values of \( K \) that maximizes the total capacity is a highly symmetric Toeplitz form -

\[
K = \begin{pmatrix}
P & \rho P & \rho P & \cdots & \rho P \\
\rho P & P & \rho P & \cdots & \rho P \\
\rho P & \rho P & P & \cdots & \rho P \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\rho P & \rho P & \rho P & \cdots & P 
\end{pmatrix} \quad (3.46)
\]

where \( \rho \) is the correlation coefficient and \(-1 \leq \rho \leq 1\)

This is because all the senders are identical. Let us assume that there is some other form of \( K \) that maximizes that sum of the rates. By appropriately
relabeling the rows and columns of $K$, we have a new matrix which has the same total capacity. By timesharing between the two forms of $K$, we can obtain a more symmetric form. Proceeding in this way, by timesharing between all possible relabelings of the rows and columns of $K$ (corresponding to all possible relabelings of the senders), we can obtain the symmetric form above. Hence, we can restrict our attention to this form of $K$, and we will obtain our bounds using it.

Let $e = (1, 1, 1, \ldots, 1)$. Divide the matrix $K$ into submatrices corresponding to $S$ and $\bar{S}$.

$$K = \begin{bmatrix} S & \bar{S} \\ \bar{S}^T & C \end{bmatrix}$$

(3.47)

where $A$ includes the rows and columns corresponding to $S$ and $C$ to $\bar{S}$.

Hence, the variance of $Y$ given $X_{\bar{S}}$ is

$$\text{var}(Y|X_{\bar{S}}) = \text{var}(Z) + \text{var}(\sum_{j \in \mathcal{S}} X_j|X_{\bar{S}})$$

(3.48)

Let $V = \sum_{j \in \mathcal{S}} X_j$ and let $\hat{V}$ be the least mean square estimate of $V$ given $X_{\bar{S}}$. Then

$$\text{var}(V|X_{\bar{S}}) = E_x \text{var}(V|X_{\bar{S}} = x_{\bar{S}})$$

(3.49)

$$\leq E_x \text{var}(V - \hat{V})$$

(3.50)

$$= \text{lms error in estimating $V$ from $X_{\bar{S}}$}$$

(3.51)

$$\overset{(a)}{=} R_{VV} - R_{VX_{\bar{S}}}R_{X_{\bar{S}}X_{\bar{S}}}^{-1}R_{X_{\bar{S}}V}$$

(3.52)

$$= e(A - BC^{-1}B^T)e,$$  

(3.53)

where $(a)$ follows from the standard expressions for lms error of estimation (see for example [[48], pg.11]).

Using this bound for the symmetric form of $K$ and after some algebra, we obtain the following bounds on the rates. (Let $T = mR$ be the total capacity of the channel.)

$$T \leq C \left( m(1 + (m - 1)\rho) \frac{P}{N} \right)$$

(3.54)

$$T \leq \frac{m}{m - 1} C \left( (1 - \rho)(m - 1)(1 + (m - 1)\rho) \frac{P}{N} \right)$$

(3.55)
\[
T \leq \frac{m}{m-2} C \left( (1 - \rho)(m - 2) \frac{1 + (m - 1)\rho P}{1 + \rho} \right) \quad (3.56)
\]
\[
\vdots
\]
\[
T \leq m C \left( (1 - \rho) \frac{1 + (m - 1)\rho P}{1 + (m - 2)\rho N} \right) \quad (3.57)
\]

In general, when \( S \) has \( m - l \) elements
\[
T \leq \frac{m}{m - l} C \left( (1 - \rho)(m - l) \frac{1 + (m - 1)\rho P}{1 + (l - 1)\rho N} \right) \quad (3.58)
\]

The bounds are illustrated for the case \( m = 8 \) and \( P/N = 1 \) in Figure 3.5.

These bounds have various interesting properties -

- The bounds reduce to Ozarow's capacity region for \( m = 2 \).
- Since the first bound is less than \( C(mP/N) \) for negative \( \rho \), we can restrict our attention to positive \( \rho \).
- The first bound is monotonically increasing with \( \rho \), the last one monotonically decreasing with \( \rho \), and all the others first rise to a maximum, then decrease with \( \rho \) falling to 0 when \( \rho = 1 \).

Since each of these functions is a concave function of \( \rho \), the pointwise minimum is also a concave function and has an unique maximum value.

To prove the factor of two bound for \( T \), we need to use only one of the bounds corresponding to \( l \approx m/2 \). We will first illustrate it in the case of even \( m \).

\[
T \leq \frac{m}{m - l} C \left( (1 - \rho)(m - l) \frac{1 + (m - 1)\rho P}{1 + (l - 1)\rho N} \right) \overset{\text{def}}{=} T_l(\rho) \quad (3.59)
\]

On differentiating and setting to 0, we find that the value of \( \rho \) that maximizes \( T_l(\rho) \) is
\[
\rho = \frac{\sqrt{l(m - l)(m - 1) - (m - 1)}}{(m - 1)(l - 1)} \quad (3.60)
\]
Figure 3.5: Bounds on the capacity of the Gaussian multiple access channel with feedback for $m = 8$ senders and signal to noise ratio $P/N = 1$
Substituting this value of $\rho$ and using $m = 2l$, we get

$$T \leq 2C \left( \left( \frac{l}{l-1} \right)^2 \left( 2l - 2\sqrt{2l - 1} \right) \frac{P}{N} \right)$$  \hspace{1cm} (3.61)$$

$$\left( \frac{l}{l-1} \right)^2 \left( 2l - 2\sqrt{2l - 1} \right) = 2 \left( \frac{l^2}{(l-1)^2} \right) \left( \frac{l^2 - (2l - 1)}{l + \sqrt{2l - 1}} \right)$$ \hspace{1cm} (3.62)$$

$$= 2l \left( \frac{l}{l + \sqrt{2l - 1}} \right)$$ \hspace{1cm} (3.63)$$

$$\leq 2l$$ \hspace{1cm} (3.64)

Hence

$$T \leq 2C(2lP/N) = 2C(mP/N) = 2T_{nfb}$$ \hspace{1cm} (3.65)

We proceed in the same manner for odd $m$. In this case we use $m = 2l+1$. Substituting $\rho_m$ and $m = 2l + 1$

$$T \leq \frac{2l+1}{l+1} C \left( \left( \frac{l\sqrt{2} - \sqrt{l+1}}{l-1} \right)^2 \right) \frac{P}{N}$$ \hspace{1cm} (3.66)

Now

$$\left( \frac{l\sqrt{2} - \sqrt{l+1}}{l-1} \right)^2 \left( l + 1 \right) \leq 2l + 1$$ \hspace{1cm} (3.67)

if

$$2l^3 - 3l^2 + 1 \geq 0$$ \hspace{1cm} (3.68)

But this is true for all positive $l$ since the minimum of the left hand side for positive $l$ is 0 occurring at $l = 1$. Hence

$$T \leq \frac{2l+1}{l+1} C(2l + \frac{1}{N}) \leq 2C(mP/N) = 2T_{nfb}$$ \hspace{1cm} (3.69)

We have thus proved that the capacity can at most be doubled using feedback for both odd and even $m$. 

37
3.3.2 Case of unequal powers

So far we have been dealing only with the case when all the transmitters have the same power constraints. Now let us assume that the powers are $P_1, P_2, \ldots, P_m$.

1. Without feedback

The dominating constraint on the sum of the rates is

$$T \leq C \left( \frac{P_1 + P_2 + \cdots + P_m}{N} \right)$$

(3.70)

Defining $P = \frac{1}{m} \sum P_i$, then

$$T \leq C \left( \frac{P}{N} \right)$$

(3.71)

2. With feedback

Let $S(P_1, P_2, \ldots, P_m)$ be the capacity with feedback. By the symmetry of the problem, $S$ is a symmetric function of its arguments. By timesharing, we can easily show that $S(\cdot, \ldots, \cdot)$ is a concave function. Hence by the properties of symmetric concave functions [[57], pg.104], we have

$$S(P_1, P_2, \ldots, P_m) \leq S(\frac{\sum P_1}{m}, \frac{\sum P_2}{m}, \ldots, \frac{\sum P_i}{m})$$

(3.72)

$$= S(P, P, \ldots, P)$$

(3.73)

$$\leq 2C(m \frac{P}{N})$$

(3.74)

$$= 2S_{nfb}(P_1, P_2, \ldots, P_m),$$

(3.75)

where (3.74) follows from the result for equal powers.

Hence, even with different powers at the different transmitters, the capacity with feedback is less than twice the capacity without feedback.

We can summarize the result in the following theorem:

Theorem 3.3.1 For any $m$-user Gaussian multiple access channel, feedback increases the total capacity by at most of factor of 2, i.e.

$$C_{FB} \leq 2C \left( \frac{\sum P_i}{N} \right).$$

(3.76)
3.4 Feedback in discrete multiple access channels

In the previous section, we have proved that feedback at most doubles the capacity of the Gaussian multiple access channel. We have not been able to prove a corresponding result for discrete memoryless channels. Neither have we been able to find a counterexample. It therefore remains open whether feedback at most doubles the capacity of any discrete memoryless multiple access channel.

In this section, we will examine in greater detail the conditions under which feedback does indeed increase the capacity of a multiple access channel. In the next section, we will define a special class of channels for which feedback does not increase the capacity and later prove some interesting properties of this class of channels.

We consider channels for which the inner bound (the capacity of the channel without feedback) is not equal to the outer bound to the capacity of the channel with feedback. We will first define two quantities related to maximum achievable sum of rates.

**Definition:** We will call

\[ R_c = \max_{p(x_1,x_2)} I(X_1,X_2;Y) \]  
(3.77)

the *cooperative capacity* of the multiple access channel and

\[ R_{nc} = \max_{p(x_1)p(x_2)} I(X_1,X_2;Y) \]  
(3.78)

the *non-cooperative capacity* of the multiple access channel.

The non-cooperative capacity \( R_{nc} \) is the maximum achievable sum of rates for the channel without feedback. The cooperative capacity \( R_c \) is an upper bound on the sum of rates with feedback. If \( R_{nc} = R_c \) then feedback does not increase the maximum achievable sum of rates. We can use these ideas to divide multiple access channels into two categories:

1. Channel where \( R_c > R_{nc} \).

   In this case, we will quote a result due to Leung:
Lemma 3.4.1 Assume that $R_c > R_{nc}$ for a multiple access channel $X_1 \times X_2 \rightarrow Y$, and let $p^*(x_1)p^*(x_2)$ be a distribution that achieves the maximum in the definition of $R_{nc}$. If for this product distribution,

$$I(X_1; Y | X_2) + I(X_2; Y | X_1) - I(X_1, X_2; Y) > 0 \quad (3.79)$$

then feedback does increase the total capacity of the multiple access channel.

Proof: The proof is given in Leung[53]. It essentially involves the extension of the Gaarder-Wolf scheme to a general multiple access channel. $\Box$

It is interesting to examine the second condition (3.79) for increase in capacity due to feedback. Expanding the mutual informations, we can rewrite the condition as

$$I(X_1; Y | X_2) + I(X_2; Y | X_1) - I(X_1, X_2; Y) \quad (3.80)$$

$$= H(Y | X_2) - H(Y | X_1, X_2) + H(Y | X_1) - H(Y | X_1, X_2)$$

$$- H(Y) + H(Y | X_1, X_2) \quad (3.81)$$

$$= H(Y, X_2) - H(X_2) + H(Y, X_1) - H(Y, X_1, X_2)$$

$$+ H(X_1, X_2) - H(Y) \quad (3.82)$$

$$= H(X_2 | Y) - H(X_2) + H(X_1 | Y) - H(X_1) - H(X_1, X_2 | Y)$$

$$+ H(X_1, X_2) \quad (3.83)$$

$$= I(X_1; X_2 | Y) - I(X_1; X_2). \quad (3.84)$$

For a product distribution, $I(X_1; X_2) = 0$, and hence the condition reduces to $I(X_1; X_2 | Y) > 0$, which is equivalent to the condition that $X_1$ and $X_2$ are not independent given $Y$. This gives us another interpretation of the role of feedback. Feedback makes independent messages conditionally dependent and since dependent information is transmitted more efficiently, feedback increases the capacity.

If for all product distributions, $I(X_1; X_2 | Y) = 0$, then feedback cannot increase total capacity, as is shown by Willems and Hekstra[84].

2. Channels for which $R_c = R_{nc}$, but for which the outer bound is not equal to the inner bound. In this case, feedback can enlarge the capacity region without increasing the total capacity.
We can use a modified version of the Gaarder Wolf scheme to show that the capacity region is increased with feedback for such a channel, if in addition, all distributions that achieve the non-cooperative capacity also satisfy the inequality

\[ I(X_1; Y|X_2) + I(X_2; Y|X_1) > I(X_1, X_2; Y). \] (3.85)

The arguments are similar to those in Leung[53] and will not be repeated.

The above arguments give a partial answer to the question: When does feedback help to increase the capacity of a discrete multiple access channel? But a complete characterization is still not possible.

### 3.5 Simple multiple access channels

Now we will consider a class of channels for which feedback does not increase the capacity. We will use this to demonstrate a connection between capacity with feedback and the source channel decomposition theorem for multiple access channels.

Consider a multiple access channel \((\mathcal{X}_1 \times \mathcal{X}_2, p(y|x_1, x_2), \mathcal{Y})\). For any input distribution \(p(x_1, x_2)\), we define a mutual information triplet \(A = (a_1, a_2, a_3)\), where

\[
\begin{align*}
    a_1 &= I(X_1; Y|X_2) \quad (3.86) \\
    a_2 &= I(X_2; Y|X_1) \quad (3.87) \\
    a_3 &= I(X_1, X_2; Y). \quad (3.88)
\end{align*}
\]

Let \(\mathcal{A}\) be the set of achievable mutual information triplets for all possible joint distributions on the inputs. Then \(\mathcal{A}\) corresponds to the outer bound on the capacity of a multiple access channel with feedback, since Theorem 3.2.1 shows that for a multiple access channel with feedback, \((R_1, R_2, R_1 + R_2) \in \mathcal{A}\).

Now consider input distributions of the form \(p(q)p(x_1|q)p(x_2|q)\) for some auxiliary random variable \(Q\). Let

\[
\begin{align*}
    b_1 &= I(X_1; Y|X_2, Q) \quad (3.89) \\
    b_2 &= I(X_2; Y|X_1, Q) \quad (3.90) \\
    b_3 &= I(X_1, X_2; Y|Q), \quad (3.91)
\end{align*}
\]
and let $\mathcal{B}$ be the set of triplets $B = (b_1, b_2, b_3)$ that can be achieved for some joint distribution of the above form. Then $\mathcal{B}$ corresponds to the capacity region of the multiple access channel without feedback, i.e., if $(R_1, R_2)$ is achievable for the multiple access channel without feedback, then $(R_1, R_2, R_1 + R_2) \in \mathcal{B}$ (Theorem 3.1.1).

**Definition:** We will call a multiple access channel *simple* if for the channel the sets $\mathcal{A}$ and $\mathcal{B}$ defined above coincide, i.e., the set of achievable mutual information triplets is not changed when we allow arbitrary joint distributions rather than mixtures of product distributions on the input.

As we shall show, simple multiple access channels are very similar to memoryless single user channels in their behavior. In particular, we will show that feedback does not increase the capacity of simple multiple access channels. Also, we will prove a source channel separation theorem for simple multiple access channels. The same results are not true for general multiple access channels, as our examples later will illustrate.

We will first give two examples of simple multiple access channels:

**Example 3.5.1: Noisy binary exor multiple access channel.**

Let $X_1 = X_2 = Y = Z = \{0, 1\}$ and let $Y = X_1 \oplus X_2 \oplus Z$, where $Z \sim \text{Bern}(p)$, i.e.,

$$Z = \begin{cases} 
0, & \text{with probability } 1 - p \\
1, & \text{with probability } p
\end{cases} \quad (3.92)$$

Then for any joint distribution $p(x_1, x_2)$, we have

$$a_1 = H(Y|X_2) - H(Y|X_1, X_2) \leq 1 - H(p) \quad (3.93)$$

$$a_2 = H(Y|X_1) - H(Y|X_1, X_2) \leq 1 - H(p) \quad (3.94)$$

$$a_3 = H(Y) - H(Y|X_1, X_2) \leq 1 - H(p) \quad (3.95)$$

All three limits are attained by an uniform distribution over the input pairs

<table>
<thead>
<tr>
<th>$p(x_1, x_2)$</th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$\frac{1}{4}$</td>
<td>$\frac{1}{4}$</td>
</tr>
<tr>
<td>1</td>
<td>$\frac{1}{4}$</td>
<td>$\frac{1}{4}$</td>
</tr>
</tbody>
</table>
Thus the region $\mathcal{A}$ for this channel is the entire cube $a_1 \leq 1 - H(p), a_2 \leq 1 - H(p), a_3 \leq 1 - H(p)$. Since the uniform distribution is also a product distribution, the region $\mathcal{B}$ for this channel is also the same, namely the unit cube. So this channel is an example of a simple channel.

**Example 3.5.2:** Consider a deterministic two input multiple access channel with $\mathcal{X}_1 = \mathcal{X}_2 = \{0, 1, 2, 3\}$ and $\mathcal{Y} = \{0, 1, 2, 3, 4, 5, 6, 7\}$, given by the following function:

<table>
<thead>
<tr>
<th>$X_1 \backslash X_2$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>5</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>6</td>
<td>7</td>
<td>4</td>
</tr>
</tbody>
</table>

Again, a uniform distribution on the inputs simultaneously achieves the maximum values for $I(X_1; Y|X_2) = H(Y|X_2) \leq 2, I(X_2; Y|X_1) = H(Y|X_1) \leq 2$ and $I(X_1, X_2; Y) = H(Y) \leq 3$. Thus for this channel, the region $\mathcal{A}$ is the region defined by $a_1 \leq 2, a_2 \leq 2, a_3 \leq 3$ and is equal the region $\mathcal{B}$. Thus the channel is simple. The capacity region of this channel without or with feedback is

$$R_1 \leq 2$$  \hspace{1cm} (3.96)

$$R_2 \leq 2$$  \hspace{1cm} (3.97)

$$R_1 + R_2 \leq 3.$$  \hspace{1cm} (3.98)

We now prove some of the properties of simple multiple access channels.

**Theorem 3.5.1** The capacity region of a simple multiple access channel is not changed by feedback.
Proof: With feedback, the capacity region of the multiple access channel is bounded (Theorem 3.2.1) by

\[
\begin{align*}
R_1 & \leq I(X_1; Y|X_2) & (3.99) \\
R_2 & \leq I(X_2; Y|X_1) & (3.100) \\
R_1 + R_2 & \leq I(X_1, X_2; Y) & (3.101)
\end{align*}
\]

for some joint distribution \(p(x_1, x_2)p(y|x_1, x_2)\). Thus \((R_1, R_2, R_1 + R_2) \in \mathcal{A}\). But since the channel is simple, \(\mathcal{A} = \mathcal{B}\) and hence \((R_1, R_2, R_1 + R_2) \in \mathcal{B}\). But by Theorem 3.1.1, \((R_1, R_2, R_1 + R_2) \in \mathcal{B}\) implies that \((R_1, R_2)\) is achievable without feedback. Thus feedback does not increase the capacity region of a simple multiple access channel. □

3.6 The source channel separation theorem

We now consider the source channel separation theorem for a multiple access channel. The source channel separation theorem is one of the most important results for single user channels. In essence, it states that source can be transmitted noiselessly over a channel if and only if its entropy rate is less than the capacity of the channel. Thus we can separate the noiseless source coding problem from the channel coding problem.

Unlike single user channels, the source channel separation theorem does not hold in general for all multiple access channels and all sources. This was first demonstrated by Cover, El Gamal and Salehi[20]. Before we consider the source channel separation theorem, we will first describe the source coding problem for a distributed source.

The problem of efficient noiseless encoding a correlated source was first considered by Slepian and Wolf[76]. The situation they considered is shown in Figure 3.6. They showed that given a joint i.i.d. source \((U, V)\) with distribution \(p(u, v)\), it is possible to encode \(U\) and \(V\) separately at rates \(R_1\) and \(R_2\) and recover \((U, V)\) with a common decoder if

\[
\begin{align*}
R_1 & \geq H(U|V) \\
R_2 & \geq H(V|U) \\
R_1 + R_2 & \geq H(U, V)
\end{align*}
\]
We will call the set of rates \((R_1, R_2)\) satisfying these constraints the Slepian-Wolf region \(R_{SW}(U,V)\).

Consider now the problem of sending a correlated source over a multiple access channel. It is natural to consider whether we could combine the results of Slepian-Wolf coding with the capacity results for the multiple access channel. Is it true that we can transmit a source over the channel if and only if the rate region of the source lies in the capacity region of the channel? The situation is illustrated in Figure 3.7.

It is clear that if \(R_{SW}\) contains a point in the capacity region of the multiple access channel, then there exists a Slepian-Wolf encoding of the source that lies within the capacity region of the multiple access channel and it will be possible to transmit the source over the channel. But is this condition also necessary?

The question was answered in the negative by Cover, El Gamal and Salehi[20], who gave an example of a source whose Slepian-Wolf rate region lay outside the capacity region of the channel and yet it was possible to transmit the source over the channel with no error. Thus the condition for transmission of a joint source over the multiple access channel is not simply that the Slepian-Wolf rate region of the source lie within the capacity region.
Figure 3.7: Transmission of correlated sources over a multiple access channel of the channel. Cover, El Gamal and Salehi also showed that it is possible to transmit the source over the channel if

\[
\begin{align*}
H(U|V) &\leq I(X_1; Y|X_2, V) \\
H(V|U) &\leq I(X_2; Y|X_1, U) \\
H(U, V) &\leq I(X_1, X_2; Y)
\end{align*}
\]

for a joint distribution of the form \( p(u, v, x_1, x_2, y) = p(u, v)p(x_1|u)p(x_2|v)p(y|x_1, x_2) \). Dueck[28] later showed that this region is not optimal.

De Bruyn, Prellov and Van der Meulen[25] have shown that for an asymmetric multiple access channel with correlated sources (this is a multiple access channel with one source available at both inputs, as shown in Figure 3.8) the source channel separation theorem does hold, so that the source can be transmitted over the channel if and only if its Slepian-Wolf rate region lies within the capacity region for the asymmetric multiple access channel. A complete characterization of the conditions necessary for the transmission of a joint source over a multiple access channel is not known at this time.

The reason for the failure of the source channel separation theorem is that the capacity of the multiple access channel depends on the correlation between the senders. If the senders are sending dependent messages, the total
Figure 3.8: Asymmetric multiple access channel with asymmetric inputs

capacity is higher than if they were sending independent messages. Slepian-Wolf encoding on the other hand gets rid of the correlation that is present among the random variables. The achievable region due to Cover, El Gamal and Salehi is obtained by preserving the correlation so that the codewords sent over the channel are also correlated.

We first prove the following theorem:

**Theorem 3.6.1** If an i.i.d. joint source \((U,V)\) can be transmitted over a multiple access channel (Figure 3.7 with deterministic encoders), then the source must satisfy the following inequalities

\[
\begin{align*}
H(U|V) & \leq I(X_1;Y|X_2) \\
H(V|U) & \leq I(X_2;Y|X_1) \\
H(U,V) & \leq I(X_1,X_2;Y)
\end{align*}
\] (3.102)

for some joint distribution \(p(x_1,x_2)p(y|x_1,x_2)\).

**Proof:** Consider a deterministic block code for this source channel combination as shown in Figure 3.7. By Fano's inequality,
since the pair \((U^n, V^n)\) can be recovered from \(Y^n\) with a low probability of error, we have
\[
H(U^n, V^n|Y^n) \leq n\epsilon_n \tag{3.105}
\]
and since adding conditioning reduces entropy, we obtain
\[
H(U^n|Y^n, V^n) \leq n\epsilon_n \tag{3.106}
\]
\[
H(V^n|Y^n, U^n) \leq n\epsilon_n \tag{3.107}
\]

Now since \((U_i, V_i)\) are independent and identically distributed, we have

\[
nH(U|V) = H(U^n|V^n) \tag{3.108}
\]
\[
= I(U^n; Y^n|V^n) + H(U^n|Y^n, V^n) \tag{3.109}
\]
\[
\leq I(U^n; Y^n|V^n) + n\epsilon_n \tag{3.110}
\]
\[
= H(Y^n|V^n) - H(Y^n|U^n, V^n) + n\epsilon_n \tag{3.111}
\]
\[
= H(Y^n|V^n, X_2^n) - H(Y^n|U^n, V^n, X_1^n, X_2^n) + n\epsilon_n \tag{3.112}
\]
\[
\leq H(Y^n|X_2^n) - H(Y^n|U^n, V^n, X_1^n, X_2^n) + n\epsilon_n \tag{3.113}
\]
\[
= I(X_1^n; Y^n|X_2^n) + n\epsilon_n \tag{3.114}
\]
\[
= \sum_{i=1}^{n} I(X_1^n; Y_i|X_2^n, Y^{i-1}) + n\epsilon_n \tag{3.115}
\]
\[
= \sum_{i=1}^{n} H(Y_i|X_2^n, Y^{i-1}) - \sum_{i=1}^{n} H(Y_i|X_{1i}, X_{2i}, Y^{i-1}, X_1^{i-1}, X_2^{i-1}, X_2^{i+1}, X_1_i+1, X_2_{i+1}) + n\epsilon_n \tag{3.116}
\]
\[
\leq \sum_{i=1}^{n} H(Y_i|X_2^n) - H(Y_i|X_{1i}, X_{2i}) + n\epsilon_n \tag{3.117}
\]
\[
= \sum_{i=1}^{n} I(X_{1i}; Y_i|X_{2i}) + n\epsilon_n \tag{3.118}
\]
\[
\leq nI(X_1; Y|X_2) + n\epsilon_n, \tag{3.119}
\]
where
(a) follows from Fano’s inequality,
(b) from the fact that $X_1^n$ and $X_2^n$ were deterministic functions of $U^n$ and $V^n$ respectively,
(c) follows since conditioning reduces entropy,
(d) from the definition of the channel since $Y^n$ depends only on $X_1^n$ and $X_2^n$,
(e) from the fact that conditioning reduces entropy and the memorylessness of the channel, and
(f) from standard techniques of adding a time-sharing random variable, and using the memorylessness of the channel.

Similarly, we get

$$nH(U,V) = H(U^n,V^n)$$
$$= I(U^n,V^n;Y^n) + H(U^n,V^n|Y^n)$$
$$\leq I(X_1^n,X_2^n;Y^n) + n\epsilon_n$$
$$= nI(X_1,X_2;Y) + n\epsilon_n.$$  \hspace{1cm} (3.120)

Dividing by $n$ and taking the limit as $n \to \infty$, we have the required converse. $\Box$

If the multiple access channel is simple, then we can transmit the source over the multiple access channel if and only if the Slepian-Wolf rate region has a non empty intersection with the capacity region of the multiple access channel and hence we have the following result:

**Theorem 3.6.2** If the multiple access channel is simple, then the source $(U,V)$ can be transmitted over the channel if and only if $R_{SW}(U,V) \cup R_I \neq \phi$, i.e., $R_{SW}$ contains at least one point that lies in the capacity region of the multiple access channel.

**Proof:** From Theorem 3.6.1, it follows that a joint source can be transmitted over a multiple access channel only if $(H(U|V),H(V|U),H(U,V)) \in \mathcal{A}$ for the channel. For a simple channel, $\mathcal{A} = B$ and hence the Slepian-Wolf encoding of the source lies in the capacity region of the channel without feedback. Thus the source can be encoded first into rates $R_1$ and $R_2$ required for the Slepian Wolf procedure.
Figure 3.9: Cascade of noiseless channels with a multiple access channel

and transmitted over the channel. Thus we have a source channel separation theorem in this case.  

Thus we see that simple multiple access channels behave in some ways like single user channels. In the last section of this chapter, we will explore some of the consequence of this for a cascade of channels.

3.7 Implications for networks of channels

Assume that we have a network consisting of a pair of noiseless channels followed by a multiple access channel as shown in Figure 3.9.

A network that consists of a noiseless non-interfering channel followed a simple multiple access channel can be analyzed as a cascade of channels. If the multiple access channel is not simple, then the network will have to be analyzed as an unit; it will not be possible to consider it as a cascade of channels.

We will prove the following theorem:
Theorem 3.7.1 For the cascade shown in Figure 3.9, the capacity of the network from A to B is

\[
C = \max_{p(x_1)p(x_2|q)p(x_1|q)} \min \{ C_1 + C_2, C_1 + I(X_2; Y|X_1, Q), C_2 + I(X_1; Y|X_2, Q), I(X_1, X_2; Y|Q) \}
\] (3.124)

if the multiple access channel is simple. Otherwise, the expression in (3.124) is a lower bound on the capacity.

Proof: Achievability. Fix \( p(q)p(x_1|q)p(x_2|q) \). It is easy to see how to achieve the capacity in (3.124) by using the standard results for achievability for a multiple access channel, since rates \( (R_1, R_2) \) satisfying

\[
\begin{align*}
R_1 &\leq I(X_1; Y|X_2, Q) \\
R_2 &\leq I(X_2; Y|X_1, Q) \\
R_1 + R_2 &\leq I(X_1, X_2; Y|Q)
\end{align*}
\] (3.125)

for the joint distribution \( p(q)p(x_1|q)p(x_2|q) \) are achievable for the multiple access channel. Thus any pair \( (R_1, R_2) \) such that \( R_1 \leq C_1 \) and \( R_2 \leq C_2 \) satisfying these equations is achievable for the cascade. Graphically, it implies that any point in the intersection of the rectangle \( R_1 \leq C_1, R_2 \leq C_2 \) with the pentagon defined by (3.125) is achievable. By considering various cases, it is easy to show that the minimum of the four quantities in (3.124) lies within the intersection. A generalization of this argument is explored in greater detail in Section 4.5.

Converse: It is slightly more difficult to prove the converse. We will simply apply the general results of Theorem 6.5.1 to this case. This channel can be considered as a four node network, with A as node 3, and B as node 4. The noiseless channel can be represented by letting \( X_3 = (X_3', X_3'') \) and letting \( Y_1 = f(X_3') \) such that \( \max I(X_3'; Y_1) = \max H(Y_1) = C_1 \), and similarly for \( Y_2 \). The new network is shown in Figure 3.10.

Using various choices of the sets \( S \) in the theorem, we obtain

\[
C \leq I(X_3; Y_1, Y_2, Y|X_1, X_2) = C_1 + C_2
\] (3.126)
Figure 3.10: Cascade network as a general 4 node network

\[
C \leq I(X_3, X_1; Y_2, Y|X_2) = C_2 + I(X_1; Y|X_2) \quad (3.127)
\]
\[
C \leq I(X_3, X_2; Y_1, Y|X_1) = C_1 + I(X_2; Y|X_1) \quad (3.128)
\]
\[
C \leq I(X_3, X_1, X_2; Y) = I(X_1, X_2; Y) \quad (3.129)
\]

for some joint distribution \(p(x_1, x_2, x_3)\). Since \(X_3\) does not affect \(Y\), we need only consider all joint distributions \(p(x_1, x_2)\).

Now comes the crucial part. If the multiple access channel is simple, then for any joint distribution \(p(x_1, x_2)\) on the inputs, there exists a corresponding distribution of the form \(p(q)p(x_1|q)p(x_2|q)\) which achieves the same mutual informations. Thus there exists a distribution of this form such that

\[
C \leq C_1 + C_2 \quad (3.130)
\]
\[
C \leq C_2 + I(X_1; Y|X_2, Q) \quad (3.131)
\]
\[
C \leq C_1 + I(X_2; Y|X_1, Q) \quad (3.132)
\]
\[
C \leq I(X_1, X_2; Y, Q) \quad (3.133)
\]

Thus the expression in (3.124) is the best we can do and is the capacity of this cascade network. □
We cannot apply the same argument to a general multiple access channel. For a general multiple access channel, the "capacity region" depends on the correlation between the inputs, and in such cases, it is not possible to analyze the network as a cascade. We have to analyze the whole network at once. Although the information flow bounds of Theorem 6.5.1 are applicable in this case, we cannot simply characterize the joint distribution of the inputs in the theorem.

In the last section of the next chapter, we will extend this result to a cascade of a deterministic broadcast channels. These results extend the results of Aref[4] and El Gamal[34].

3.8 Summary

In this chapter, we considered the role of feedback in increasing the capacity of multiple access channels. We showed that for a discrete multiple access channel, feedback increases the total capacity by a factor less than the number of senders. For a Gaussian multiple access channel, feedback at most doubles the total capacity.

We defined a special class of multiple access channels called simple multiple access channels for which correlation between the inputs does not increase the capacity region. We showed that feedback does not increase the capacity of simple multiple access channels and that the source channel separation theorem holds for such multiple access channels. A cascade of pair of noiseless channels followed by a simple multiple access channel satisfies the max flow min cut theorem with equality and its capacity can be simply determined from the intersection of the capacity region of the pair of noiseless channels with the capacity region of the multiple access channel.

The results of Sections 3.2 and 3.3 were published earlier in [77].
Chapter 4

Feedback and Broadcast Channels

4.1 Introduction

The broadcast channel is a communication channel in which there is one sender and two or more receivers. It is illustrated in Figure 4.1. Examples of broadcast channels include a TV station transmitting to many receivers, a lecturer talking in a class, etc.

The problem of finding the rates of communication that can be achieved simultaneously to both receivers was introduced by Cover[17]. To date, this problem has not been solved. The special case of stochastically degraded channels has been solved by Bergmans[6] and Gallager[38].

We will begin our discussion of the broadcast channel with a few definitions.

Definition: A (two-user) broadcast channel consists of an input alphabet, $\mathcal{X}$, and two output alphabets, $\mathcal{Y}_1$ and $\mathcal{Y}_2$, and a probability transition function, $p(y_1, y_2|x)$.

We define codes, probability of error, achievability and capacity regions for the broadcast channel as we did for the single user channel. We will extend the definitions to include common information to be sent to both receivers.

A $((2^{nR_0}, 2^{nR_1}, 2^{nR_2}), n)$ code for a broadcast channel with independent
Figure 4.1: A Broadcast Channel
information consists of an encoder:

\[ X : (\{1, 2, \ldots, 2^{nR_0}\} \times \{1, 2, \ldots, 2^{nR_1}\} \times \{1, 2, \ldots, 2^{nR_2}\}) \rightarrow \mathcal{X}^n \]  

(4.1)

and two decoders,

\[ g_1 : \mathcal{Y}_1^n \rightarrow \{1, 2, \ldots, 2^{nR_0}\} \times \{1, 2, \ldots, 2^{nR_1}\} \]  

(4.2)

and

\[ g_2 : \mathcal{Y}_2^n \rightarrow \{1, 2, \ldots, 2^{nR_0}\} \times \{1, 2, \ldots, 2^{nR_2}\}. \]  

(4.3)

We define the probability of error as the probability the decoded message is not equal to the transmitted message, averaged over all possible messages, i.e.,

\[ P_e^{(n)} = P(g_1(Y_1^n) \neq (W_0, W_1) \text{ or } g_2(Y_2^n) \neq (W_0, W_2)) \]  

(4.4)

where \((W_0, W_1, W_2)\) are uniformly distributed over \(2^{nR_0} \times 2^{nR_1} \times 2^{nR_2}\).

**Definition:** A rate pair \((R_0, R_1, R_2)\) is said to be *achievable* for the broadcast channel, if there exists a sequence of \((2^{nR_0}, 2^{nR_1}, 2^{nR_2}), n)\) codes with \(P_e^{(n)} \rightarrow 0\).

**Definition:** The *capacity region* of the broadcast channel is the closure of the set of achievable rates.

**Definition:** A broadcast channel is said to be *physically degraded* if \(p(y_1, y_2|x) = p(y_1|x)p(y_2|y_1)\).

**Definition:** A broadcast channel is said to be *stochastically degraded* if its conditional marginal distributions are the same as that of a physically degraded broadcast channel, i.e., if there exists a distribution \(p'(y_2|y_1)\) such that

\[ p(y_2|x) = \sum_{y_1} p(y_1|x)p'(y_2|y_1). \]  

(4.5)

Since the capacity of broadcast channel depends only on the conditional marginals, the capacity region of the stochastically degraded broadcast
channel is the same as that of the corresponding physically degraded channel. The capacity of a degraded broadcast channel is given in the following theorem.

**Theorem 4.1.1** The capacity region for sending independent information over the degraded broadcast channel $X \rightarrow Y_1 \rightarrow Y_2$ is the convex hull of the closure of all $(R_0, R_1, R_2)$ satisfying

\begin{align*}
R_0 + R_2 &\leq I(U; Y_2) \\
R_1 &\leq I(X; Y_1 | U)
\end{align*}

for some joint distribution $p(u)p(x|u)p(y, z|x)$ and some auxiliary random variable $U$ with cardinality

$$|U| \leq \min\{|\mathcal{X}|, |\mathcal{Y}_1|, |\mathcal{Y}_2|\}$$

The proof of the theorem can be found in the above papers. Cover[16] and Van der Meulen[78] have put forward achievable rate regions for the general broadcast channel.

The capacity region for the broadcast channel for degraded message sets was found by Körner and Marton[52]. The capacity region for the less noisy broadcast channel was found by Körner and Marton[51] and for the more capable case by El Gamal[32]. The largest known achievable region for the general broadcast channel has been put forward by Marton[58], but it is not known whether this is indeed the capacity region. The capacity region for the general broadcast channel remains one of the most important unsolved problems in network information theory.

### 4.2 Broadcast channels with feedback

We will consider a broadcast channel with complete feedback where the received sequences of all the receivers are fed back to the sender, who can use that information to decide what to send next.

**Definition:** A $((2^{nR_1}, 2^{nR_2}), n)$ code for a broadcast channel with feedback with independent information consists of a sequence of encoders:

$$X_i : (\{1, 2, \ldots, 2^{nR_1}\} \times \{1, 2, \ldots, 2^{nR_2}\} \times \mathcal{Y}_1^{i-1} \times \mathcal{Y}_2^{i-1}) \rightarrow \mathcal{X}, \quad i = 1, 2, \ldots, n$$

(4.9)
which map the messages and the past received symbols onto the next transmitted symbol, and two decoders,

\[ g_1 : \mathcal{Y}_1^m \to \{1, 2, \ldots, 2^{nR_1}\} \]  \hspace{1cm} (4.10)

and

\[ g_2 : \mathcal{Y}_2^m \to \{1, 2, \ldots, 2^{nR_2}\}. \]  \hspace{1cm} (4.11)

We define the probability of error, achievability, and the capacity region as in the case of the broadcast channel without feedback.

El Gamal[33] showed that for a physically degraded broadcast channel, feedback does not increase the capacity region of the channel. In this way a physically degraded broadcast channel acts like a single user channel.

Dueck[29] provided an example of a non-degraded channel where feedback increased the capacity region. We will give the example later in this section. Ozarow and Leung[66] showed that for a Gaussian broadcast channel that is not physically degraded, feedback does increase the capacity region of the channel.

In this section, we will present Dueck’s example of a broadcast channel where feedback does increase capacity. This example illustrates the basic reason why feedback helps, namely, it allows the sender to send common correction information to all the receivers, and this information can be sent more efficiently than independent information.

We will illustrate the example for two receivers. The sender \( X \) sends a triplet of binary random variables \((X_0, X_1, X_2)\), where \( X_0 = X_1 = X_2 = \{0, 1\} \). Each of the receivers receives a pair of binary random variables. The receiver \( Y_1 \) receives \((X_0, X_1 \oplus Z)\), where \( Z \in \mathcal{Z} = \{0, 1\} \) is binary noise uniformly distributed on 0 and 1. The receiver \( Y_2 \) receives \((X_0, X_2 \oplus Z)\), where the noise \( Z \) is the same as the noise that effects \( Y_1 \).

The channel is illustrated in Figure 4.2.

- **Without feedback.** Because the noise \( Z \) is uniformly distributed on \( \{0, 1\} \), we have \( X_1 \oplus Z \) independent of \( X_1 \) (it is the output of a binary symmetric channel with input \( X_1 \) and cross over probability 1/2). Hence without feedback, the second component of each receiver's output is independent of the corresponding input. Only the first component \( X_0 \) carries any information. Thus without feedback, the capacity
The capacity regions without and with feedback for the Dueck example are illustrated in Figure 4.3.

The example illustrates the basic reason why feedback helps increase the capacity of memoryless broadcast channels. The main problem with a broadcast channel is that we have to send independent information to the two receivers; if we were sending the same information to both of them, the information could be sent a lot more efficiently. Feedback enables the sender to calculate the correction information and find the common component, which he can then send very efficiently to all the receivers.
Figure 4.3: Capacity regions in Dueck's example
4.3 Bounds on the capacity of a broadcast channel with feedback

In this section, we will derive an outer bound on the capacity of a broadcast channel with feedback and derive a few consequences of this outer bound.

Theorem 4.3.1 The capacity region of the broadcast channel with feedback is a subset of the region $C_o = \{(R_0, R_1, R_2) :$

\[
R_0 + R_1 \leq I(U_1; Y_1) \tag{4.13}
\]

\[
R_0 + R_2 \leq I(U_2; Y_2) \tag{4.14}
\]

\[
R_0 + R_1 + R_2 \leq I(U_1, U_2; Y_1, Y_2) \tag{4.15}
\]

for some joint distribution $p(u_1, u_2)p(x|u_1, u_2)p(y_1, y_2|x)$.

Proof: Assume that we are given a $(2^{nR_1}, 2^{nR_2}, n)$ feedback code for the broadcast channel. Then since the probability of error is low, by Fano's inequality, we have

\[
H(W_1, W_0|Y_1^n) \leq P_e^{(n)} n(R_1 + R_0) + H(P_e^{(n)}) \triangleq n\epsilon_n, \tag{4.16}
\]

where $\epsilon_n \to 0$ as $n \to \infty$. Similarly, we have

\[
H(W_2, W_0|Y_2^n) \leq n\epsilon_n, \tag{4.17}
\]

and

\[
H(W_1, W_2, W_0|Y_1^n, Y_2^n) \leq 2n\epsilon_n. \tag{4.18}
\]

We can now derive a chain of inequalities,

\[
n(R_1 + R_0) = H(W_0, W_1) \tag{4.19}
\]

\[
= I(W_0, W_1; Y_1^n) + H(W_0, W_1|Y_1^n) \tag{4.20}
\]

\[
\leq I(W_0, W_1; Y_1^n) + n\epsilon_n \tag{4.21}
\]

\[
= \sum_{i=1}^n I(W_0, W_1; Y_1^i|Y_1^{i-1}) + n\epsilon_n \tag{4.22}
\]

\[
\leq \sum_{i=1}^n I(W_0, W_1; Y_1^{i-1}; Y_{1i}) + n\epsilon_n \tag{4.23}
\]

\[
= \sum_{i=1}^n I(U_{1i}; Y_{1i}) + n\epsilon_n \tag{4.24}
\]
where $U_{i} \triangleq (W_{0}, W_{1}, Y_{i}^{i-1})$. Similarly, we have

$$n(R_{0} + R_{2}) \leq \sum_{i=1}^{n} I(U_{2i}; Y_{2i}) + n\epsilon_{n},$$

(4.25)

and

$$n(R_{0} + R_{1} + R_{2}) \leq \sum_{i=1}^{n} I(U_{1i}, U_{2i}; Y_{1i}, Y_{2i}) + 2n\epsilon_{n}.$$  (4.26)

Dividing all these equations by $n$ and introducing a timesharing random variable $Q$ such that $Q$ is uniformly distributed on $\{1, 2, \ldots, n\}$, we can rewrite these equations as

$$R_{0} + R_{1} \leq \frac{1}{n} \sum_{i=1}^{n} I(U_{1i}; Y_{1i} | Q = i) + \epsilon_{n} \quad (4.27)$$

$$= I(U_{1Q}; Y_{1Q} | Q) + \epsilon_{n} \quad (4.28)$$

$$\leq I(U_{1Q}, Q; Y_{1Q}) + \epsilon_{n}.$$

(4.29)

Similarly we can show that

$$R_{0} + R_{2} \leq I(U_{2Q}, Q; Y_{2Q}) + \epsilon_{n} \quad (4.30)$$

and

$$R_{0} + R_{1} + R_{2} \leq I(U_{1Q}, U_{2Q}, Q; Y_{1Q}, Y_{2Q}) + 2\epsilon_{n}.$$  (4.31)

Defining $U_{1} \triangleq (U_{1Q}, Q)$, $U_{2} \triangleq (U_{2Q}, Q)$, etc. and taking the limit as $n \to \infty$, we have the desired outer bound. It is clear from the memoryless nature of the channel that $(W_{0}, W_{1}, W_{2}, Y_{i}^{i-1}, Y_{2}^{i-1}) \to X_{i} \to (Y_{i}, Y_{2i})$ forms a Markov chain, giving us the joint distribution of the theorem. $\square$

This outer bound has a few direct consequences:

1. **Feedback does not increase the maximum individual rates to any receiver.** This is because the maximum rate with feedback

$$R_{1,\text{max}} = \max I(U_{1}; Y_{1}) \leq \max I(X; Y_{1}),$$

(4.32)

by the data processing inequality[23] since $U_{1} \to X \to Y_{1}$ forms a Markov chain, and the upper bound can be achieved without feedback (when $R_{0} = R_{2} = 0$).
2. Feedback does not increase the maximum rate for common information.
With feedback, the maximum value of $R_0$ is
\[
R_0 = \max_{p(u_1, u_2)} \min\{I(U_1; Y_1), I(U_2; Y_2)\} \tag{4.33}
\]
\[
= \max_{p(x)} \min\{I(X; Y_1), I(X; Y_2)\} \tag{4.34}
\]
which can be achieved without feedback by the results of Körner and Marton[52].

3. Feedback does not increase the total capacity $T = \sum_{i=1}^{m} R_i$ by more than a factor of $m$, where $m$ is the number of receivers.
We can use the same geometric argument that was used for the multiple access channel. Since feedback does not increase the maximum individual rates, the sum of the rates can be at most increased by a factor of $m$. The extension of the Dueck example to $m$ receivers shows that this bound is tight.

4. Feedback does not increase the capacity of a deterministic broadcast channel.
The capacity of a deterministic broadcast channel without feedback was found by Pinsker[67] and Marton[58]. It is the union of regions of the form
\[
R_1 \leq H(Y_1) \tag{4.35}
\]
\[
R_2 \leq H(Y_2) \tag{4.36}
\]
\[
R_1 + R_2 \leq H(Y_1, Y_2), \tag{4.37}
\]
over all distributions $p(x)$.
Since $I(U_1; Y_1) \leq H(Y_1)$, etc., it is easy to see that the upper bound of the theorem cannot be larger than the capacity region without feedback for a deterministic broadcast channel. Hence, the capacity of a deterministic broadcast channel is not increased by feedback. But this result is not very surprising, since for a deterministic broadcast channel, the sender can calculate what the receivers will receive and hence feedback will not provide it any new information.

63
Cover[16] and Van der Meulen[78] have proved that the region of (4.15) is achievable without feedback if we only allow distributions of the form

\[ p(u_1, u_2, x, y_1, y_2) = p(u_1)p(u_2)p(x|u_1, u_2)p(y_1, y_2|x). \]  

(4.38)

This illustrates an aspect of a general theme: without feedback, we can achieve a rate region corresponding to independent auxiliary random variables. In general, this region is optimal for multiple access channels without feedback, but not for broadcast channel. With feedback, the outer bounds give auxiliary random variables that are not independent in general. We will expand on this theme in Chapter 6 on feedback in networks.

### 4.4 Gaussian broadcast channels with feedback

A common model for a continuous broadcast channel is the Gaussian broadcast channel, where the noise is assumed to be independent and identically distributed Gaussian random variables. Thus the received signals are

\[ Y_1 = X + Z_1 \]  

(4.39)

\[ Y_2 = X + Z_2 \]  

(4.40)

where \((Z_1, Z_2)\) are i.i.d. \(\sim \mathcal{N}(0, K_Z)\). We will let \(N_1\) denote the variance of \(Z_1\) and \(N_2\) denote the variance of \(Z_2\). We have a power constraint on the input, so that all codewords are constrained to have an average power less than \(P\).

The capacity of a broadcast channel without feedback depends only on the marginal distributions \(p(y_1|x)\) and \(p(y_2|x)\)[6]. All Gaussian channels are stochastically degraded and the capacity region without feedback can be calculated from the results for degraded broadcast channels. For the Gaussian channel with \(N_1 < N_2\), they can be written as [17] [6] [38]

\[ R_1 \leq C \left( \frac{\alpha P}{N_1} \right) \]  

(4.41)

\[ R_2 \leq C \left( \frac{(1-\alpha)P}{N_2 + \alpha P} \right) \]  

(4.42)
for all $0 \leq \alpha \leq 1$. $\alpha$ is the fraction of power used to send information to receiver 1.

El Gamal[33] has shown that the capacity of a physically degraded Gaussian broadcast channel is not increased by feedback. It was initially conjectured that this was true for all stochastically degraded broadcast channels as well, but Ozarow and Leung[66] have shown that if the channel is not physically degraded, feedback does help to increase the capacity of the channel. They used a modified Kailath-Schalkwijk scheme, similar to the scheme used by Ozarow[63] for the multiple access channel with feedback. Willems and Van der Meulen[85] suggested some improvements to the Ozarow-Leung scheme that increased the capacity region for the Gaussian broadcast channel with feedback.

For the Gaussian broadcast channel with feedback, we do not have any stronger bounds than the bounds in Theorem 4.3.1. The $m$-receiver Gaussian broadcast channel can be written as

$$Y_i = X + Z_i, \quad i = 1, 2, \ldots, m,$$

(4.43)

where the $Z = (Z_1, Z_2, \ldots, Z_m)$ are drawn i.i.d. $\sim \mathcal{N}(0, K_Z)$. We will assume that the noise powers are $N_1 \leq N_2 \leq \cdots \leq N_m$ and the signal power is $P$.

In the case when the noise in the different channels are independent, we can calculate these upper bounds for a $m$-receiver broadcast channel. In this case, the covariance matrix of the noise is diagonal, i.e.,

$$K_Z = \begin{bmatrix}
N_1 & 0 & \cdots & 0 \\
0 & N_2 & \cdots & 0 \\
\vdots & \vdots & \ddots & 0 \\
0 & 0 & \cdots & N_m
\end{bmatrix}$$

(4.44)

- **Without feedback.** The maximum achievable sum of rates without feedback is the maximum throughput to the best receiver, namely, receiver 1. Thus $T_{nfb} = C(P/N_1)$.

- **With feedback.** The strongest bound on the sum of all the rates in Theorem 4.3.1 is

$$\sum_{i=1}^{m} R_i \leq I(X; Y_1, Y_2, \ldots, Y_m) = h(Y_1, \ldots, Y_m) - h(Y_1, \ldots, Y_m|X).$$

(4.45)
When the noise samples are independent, the covariance matrix of the received signals is

\[
K_Y = \begin{bmatrix}
P + N_1 & P & \cdots & P \\
P & P + N_2 & \cdots & P \\
\vdots & \vdots & \ddots & \vdots \\
P & P & \cdots & P + N_m
\end{bmatrix}.
\] (4.46)

Hence we can calculate

\[
h(Y_1, \ldots, Y_m) = \frac{1}{2} \log(2\pi e)^m |K_Y|
\]

\[
= \frac{1}{2} \log(2\pi e)^m \left(1 + P \left(\frac{1}{N_1} + \cdots + \frac{1}{N_m}\right)\right) N_1 N_2 \ldots N_m
\]

and

\[
h(Y_1, \ldots, Y_m | X) = \frac{1}{2} \log(2\pi e)^m |K_Z|
\]

\[
= \frac{1}{2} \log(2\pi e)^m N_1 N_2 \ldots N_m
\] (4.49)

Thus the upper bound on the total capacity with feedback is

\[
T_{fb} = \sum_{i=1}^{m} R_i \leq \frac{1}{2} \log \left(1 + P \left(\frac{1}{N_1} + \cdots + \frac{1}{N_m}\right)\right).
\] (4.50)

Since \(N_1\) is the smallest noise variance, we have

\[
P \left(\frac{1}{N_1} + \cdots + \frac{1}{N_m}\right) \leq m \frac{P}{N_1},
\] (4.51)

and therefore

\[
T_{fb} \leq C \left(m \frac{P}{N_1}\right) \leq T_{nfb} + \log m.
\] (4.52)

Thus feedback increases the total capacity of set of \(m\) independent Gaussian broadcast channels by at most \(\log m\) bits.

This result shows that for Gaussian broadcast channels with independent noise, feedback does not help by a term that grows as the logarithm of the number of users. This result is not true of discrete broadcast channels or Gaussian broadcast channels with dependent noise.

66
4.5 Cascade of deterministic broadcast channels and simple multiple access channels

In the previous chapter, we introduced the concept of simple multiple access channels, for which the capacity region is unchanged if we allow joint distributions instead of product distributions. Simple multiple access channels have a number of properties of single-user channels. For example, their capacity is not changed by feedback and we have a source channel separation theorem for such channels.

In this section, we will consider a network which consists of a deterministic broadcast channel followed by a simple multiple access channel as shown in Figure 4.4. We will call such a channel a cascade of a broadcast channel with a multiple access channel. We will show that the capacity of the cascade has a nice max flow min cut interpretation and that the bounds derived in Chapter 6 for a general network are tight in this case.

The cascade is illustrated in Figure 4.4. For such a cascade, we can prove
the following theorem:

**Theorem 4.5.1** The capacity of a cascade of a deterministic broadcast channel with a simple multiple access channel is given by

$$C = \max_{p(x)p(q)p(x_1)q(x_2|\theta)} \min \{ H(Y_1, Y_2), H(Y_1) + I(X_2; Y|X_1, Q), H(Y_2) + I(X_1; Y|X_2, Q), I(X_1, X_2; Y|Q, \theta) \}.$$  

If the multiple access channel is not simple, then this expression is a lower bound on the capacity of the cascade.

**Proof:** Achievability: Fix $p(x)p(x_1)p(x_2)$. We will show that a rate

$$R = \min\{H(Y_1, Y_2), H(Y_1) + I(X_2; Y|X_1), H(Y_2) + I(X_1; Y|X_2), I(X_1, X_2; Y)\}.$$  \hfill (4.54)

is achievable for any distribution $p(x)p(x_1)p(x_2)$. We can then achieve the rate given in the theorem by time-sharing.

To prove achievability, we will use the known results for the deterministic broadcast channel and the simple multiple access channel. We know that any rate pair $(R_1, R_2)$ is achievable for the deterministic broadcast channel if

$$R_1 \leq H(Y_1)$$  \hfill (4.55)

$$R_2 \leq H(Y_2)$$  \hfill (4.56)

$$R_1 + R_2 \leq H(Y_1, Y_2)$$  \hfill (4.57)

This region for a fixed $p(x)$ defines a pentagon in the $(R_1, R_2)$ plane. Also, since $H(Y_1, Y_2) \leq H(Y_1) + H(Y_2)$, the last bound is always less than or equal to the sum of the other two bounds.

Also for the multiple access channel, we know that the set of rate pairs $(R_1, R_2)$ satisfying

$$R_1 \leq I(X_1; Y|X_2)$$

$$R_2 \leq I(X_2; Y|X_1)$$

$$R_1 + R_2 \leq I(X_1, X_2; Y)$$  \hfill (4.58)
are achievable for any product distribution \( p(x_1)p(x_2) \). Again, for any product distribution,

\[
I(X_1; Y|X_2) + I(X_2; Y|X_1) - I(X_1, X_2; Y) = I(X_1; X_2|Y) - I(X_1; X_2) \geq 0
\]

by the same arguments as for (3.84). Thus the last bound in (4.58) is always less than the sum of the other two.

For the cascade of the deterministic broadcast channel with the multiple access channel, any point in the intersection of the two pentagons is achievable. By considering the various possible configurations of the two pentagons as shown in Figure 4.5, it can be seen that the minimum of the four quantities in (4.54) always lies in the intersection of the two pentagons.

Thus the rate in (4.54) is achievable for any product distribution \( p(x)p(x_1)p(x_2) \). We can then use time-sharing to achieve the rate in (4.53).
Converse: We will use the general converse for an $m$-node network derived in Chapter 6. We rewrite the channel as a four node network as shown in Figure 4.6, and using various choices of subset $S$ in Theorem 6.5.1, we obtain

\begin{align*}
C & \leq I(X; Y_1, Y_2, Y | X_1, X_2) \quad (4.60) \\
C & \leq I(X, X_1; Y_2, Y | X_2) \quad (4.61) \\
C & \leq I(X, X_2; Y_1, Y | X_1) \quad (4.62) \\
C & \leq I(X, X_1, X_2; Y) \quad (4.63)
\end{align*}

for some joint distribution $p(x, x_1, x_2)$.

Consider the second bound

\begin{align*}
I(X, X_1; Y_2, Y | X_2) &= H(Y_2, Y | X_2) - H(Y, Y | X_1, X_2, X) \quad (4.64) \\
&\leq H(Y_2 | X_2) + H(Y | X_2) - H(Y | X_1, X_2, X) \\
&= H(Y_2 | X_2) + H(Y | X_2) - H(Y | X_1, X_2) \quad (4.65) \\
&\leq H(Y_2) + I(X_1; Y | X_2) \quad (4.66)
\end{align*}

with equality in (4.66) if the joint distribution is $p(x)p(x_1, x_2)$.

By similar arguments, we can show that

\begin{align*}
C & \leq H(Y_1, Y_2) \quad (4.67) \\
C & \leq H(Y_1) + I(X_2; Y | X_1) \quad (4.68) \\
C & \leq H(Y_2) + I(X_1; Y | X_2) \quad (4.69) \\
C & \leq I(X_1, X_2; Y) \quad (4.70)
\end{align*}

for some joint distribution $p(x)p(x_1, x_2)$.

Now comes the crucial part for simple multiple access channels. Since the multiple access channel is simple, we can replace the joint distribution $p(x_1, x_2)$ by a distribution of the form $p(q)p(x_1 | q)p(x_2 | q)$ that achieves the same mutual informations. Thus we have shown that there exists a distribution of the form $p(x)p(q)p(x_1 | q)p(x_2 | q)$ such that

\begin{align*}
C & \leq H(Y_1, Y_2) \quad (4.71) \\
C & \leq H(Y_1) + I(X_2; Y | X_1, Q) \quad (4.72) \\
C & \leq H(Y_2) + I(X_1; Y | X_2, Q) \quad (4.73) \\
C & \leq I(X_1, X_2; Y | Q) \quad (4.74)
\end{align*}
$$(X_1, Y_1)$$

$$(X, \cdot)$$

$$(\cdot, Y)$$

$$(X_2, Y_2)$$

Figure 4.6: Cascade network as a general 4 node network

and hence the rate in (4.53) is the best possible, completing the proof of the converse. □

Simple multiple access are the next level of generalization (after non-interfering channels) for which we can calculate the capacity of cascades. The results extend the results of Aref[4] on cascades of deterministic broadcast channels with non-interfering multiple access channels.

4.6 Summary

Feedback helps increase the capacity of broadcast channels since it converts some of the independent information to be sent to the receivers into common information, which can be sent more efficiently to the receivers. We have derived bounds on the capacity of a broadcast channel with feedback and shown that feedback can increase the capacity of a broadcast channel by at most a factor of $m$. The Dueck example shows that this bound is tight.

We also analyzed a cascade of a deterministic broadcast channel with a simple multiple access channel and showed that in this case, the capacity of the network satisfied the max flow min cut theorem. These results generalize earlier work on relay networks.
Chapter 5

Feedback and other multiple user channels

5.1 The relay channel with feedback

The relay channel was introduced by Van der Meulen\cite{81}; it is a channel with one sender and one receiver and a number of intermediate nodes which act as relays to facilitate the communication from the sender to the receiver. This channel was studied by Cover and El Gamal\cite{19}, who derived inner and outer bounds on the capacity, which coincide for the degraded relay channel.

The simplest case is when there is only one intermediate or relay node. In this case the channel consists of four finite sets, $\mathcal{X}$, $\mathcal{X}_1$, $\mathcal{Y}$ and $\mathcal{Y}_1$, and a collection of probability distributions $p(.,|x, x_1)$ on $\mathcal{Y} \times \mathcal{Y}_1$, one for each $(x, x_1) \in \mathcal{X} \times \mathcal{X}_1$. The interpretation is that $x$ is the input to the channel and $y$ is the output, $y_1$ is the relay's output and $x_1$ is the input symbol chosen by the relay as shown in Figure 5.1. The problem is to find the capacity of the channel between the sender $X$ and the receiver $Y$.

The relay channel combines a broadcast channel ($X$ to $Y$ and $Y_1$) and a multiple access channel ($X$ and $X_1$ to $Y$). As in the case of the broadcast channel, the capacity of the relay channel is not known in general. It is known only for the special case of a degraded relay channel, in which the relay is better than the final receiver.

**Definition:** The relay channel $(\mathcal{X} \times \mathcal{X}_1, p(y, y_1|x, x_1), \mathcal{Y} \times \mathcal{Y}_1)$ is said to be
physically degraded if \( p(y, y_1|x, x_1) \) can be written in the form

\[
p(y, y_1|x, x_1) = p(y_1|x, x_1)p(y|y_1, x_1).
\]

(5.1)

Thus \( Y \) is a random degradation of the relay signal \( Y_1 \).

For the physically degraded relay channel, the capacity is given by the following theorem.

**Theorem 5.1.1** The capacity \( C \) of a physically degraded relay channel is given by

\[
C = \sup_{p(x,x_1)} \min \{ I(X, X_1; Y), I(X; Y_1 | X_1) \},
\]

(5.2)

where the supremum is over all joint distributions on \( X \times X_1 \).

The proof of this theorem can be found in Cover and El Gamal[19].

When we add feedback to a general relay channel, we assume that the received signals \( Y \) and \( Y_1 \) are fed back to both senders. In that case, the relay effectively receives both \( Y \) and \( Y_1 \). Since \( Y \) is a physically degraded version of \( (Y, Y_1) \), feedback converts a general relay channel to a physically degraded channel. This enables us to derive the following theorem[19]:

**Theorem 5.1.2** The capacity of a relay channel with complete feedback from \( Y \) and \( Y_1 \) to both \( X \) and \( X_1 \) is given by

\[
C = \sup_{p(x,x_1)} \min \{ I(X; Y, Y_1 | X_1), I(X, X_1; Y) \}.
\]

(5.3)
The proof of the achievable part of the theorem only uses the feedback from the receiver $Y$ to the relay $X_1$. The outer bound does not depend on the feedback links. So if feedback from $Y$ to $X_1$ is present, the capacity is as given in the above theorem; the other feedback links are not necessary.

The relay channel is rather unusual in that feedback simplifies the calculation of the capacity of the channel. The capacity of a general relay channel is not known, but the capacity of the relay channel with feedback is known.

For the physically degraded relay channel, the feedback link from $Y$ to the relay does not increase the information available to the relay. Hence in this case, we have the following theorem:

**Theorem 5.1.3** Feedback does not increase the capacity of a physically degraded relay channel.

**Proof:** The proof follows from the previous two theorems and the fact that for a physically degraded relay channel, $I(X; Y, Y_1|X_1) = I(X; Y_1|X_1)$. □

5.2 Interference Channels

An interference channel consists of a pair of senders $X_1$ and $X_2$ and a pair of receivers, $Y_1$ and $Y_2$. The outputs $Y_1$ and $Y_2$ depend on both the inputs according a conditional distribution $p(y_1, y_2|x_1, x_2)$. The sender $X_1$ attempts to send information to receiver $Y_1$, and the sender $X_2$ attempts to send information to $Y_2$. The problem is to find the set of simultaneously achievable rates for the two senders.

The interference channel was introduced by Shannon[74]. It was studied by Ahlswede[2], who gave an example to show that the region conjectured by Shannon was not the capacity region of the interference channel. Carleial[12] introduced the Gaussian interference channel with power constraints, and showed that very strong interference is equivalent to no interference at all. Sato and Tanabe[70] extended the work of Carleial to discrete interference channels with strong interference. Sato[69] and Benzel[5] dealt with degraded interference channels. The most general achievable region for the general interference channel is due to Han and Kobayashi[43]. This region gives the capacity for Gaussian interference channels with interference parameters greater than 1, as was shown in Han and Kobayashi[43] and

5.2.1 Interference channels without feedback

In this section, we will summarize some of the earlier results on the capacity region of an interference channel without feedback. We will consider the Gaussian interference channel with two users, since it is probably the most important example and illustrates many of the basic ideas. The Gaussian interference channel is the simplest Gaussian channel for which the capacity region is not known in general. It is a realistic model for many practical communication networks - for example, satellite networks.

The Gaussian interference channel is shown in Figure 5.2.

Though the form of the interference channel shown in the figure (with transmission coefficients as shown) appears to be a special case, Carleial[12] has shown that any Gaussian interference channel can be put into this form. This form is called the standard form for the interference channel and is completely specified by its signal powers $P_1$ and $P_2$ and the interference pa-
rameters $a$ and $b$. The capacity of the interference channel without feedback depends only on the marginal distributions of the noise and not on the joint distribution.

It is clear that if the interference parameters $a$ and $b$ are zero, then we have two independent channels and the capacity region for the network is the rectangle given by

$$R_1 \leq \frac{1}{2} \log \left(1 + \frac{P_1}{N_1}\right)$$  
$$R_2 \leq \frac{1}{2} \log \left(1 + \frac{P_2}{N_2}\right).$$  

(5.4) \hspace{1cm} (5.5)

If the interference parameters are greater than zero, then the capacity region is a subset of the above rectangle.

The interference channel has different modes of operation depending on the value of the interference parameter. In the symmetric case ($a = b$, $P_1 = P_2 = P$, $N_1 = N_2 = 1$), four modes have been identified. Let $a^*$ be the positive solution to

$$2a^2(1 + a^2P) = 1.$$  

(5.6)

1. **Low interference:** $0 < a \leq a^*$. In this case, the best known scheme (for maximum sum of rates) treats the interference as part of the noise. In this case, the maximum achievable sum of rates is

$$R_1 + R_2 = 2C \left(\frac{P}{1 + a^2P}\right).$$  

(5.7)

2. **Moderate interference:** $a^* \leq a < 1$. For moderate interference, the best known scheme is time division multiplexing (or frequency division multiplexing). The best achievable sum of rates in this case is

$$R_1 + R_2 = C(2P).$$  

(5.8)

3. **Strong interference:** $1 < a \leq \sqrt{1 + P}$. Since the interference parameter is greater than 1, the second receiver has a better look at the codeword than the intended receiver. Sato[68] and Han and Kobayashi[43] has used this to show that in the optimal coding scheme, both receivers must decode both messages and thus the capacity region for this channel is the intersection of the capacity regions for the multiple access
channels that correspond to receiver 1 and receiver 2 respectively. The best achievable sum of rates for this case is

\[ R_1 + R_2 = C \left( P + a^2 P \right). \]  
(5.9)

4. **Very strong interference**: \( \sqrt{1 + P} \leq a \). This case was considered by Carleial\[12\], who showed that the receivers could decode the signal meant for the other receiver first, treating the signal destined for itself as part of the noise. The receiver can then subtract out the signal meant for the other receiver, thus obtaining a channel with no interference. Thus, very strong interference is equivalent to no interference. The best achievable sum of rates in this case is

\[ R_1 + R_2 = 2C(P) \]  
(5.10)

In the case of strong and very strong interference, the schemes described above are optimal. In general, the best known coding scheme is due to Han and Kobayashi\[43\]. For weak and moderate interference, it is not known whether this scheme is optimal. Costa\[14\] has proved the optimality of the corner points of the Han-Kobayashi region.

### 5.2.2 Interference channels with feedback

We will assume that we have have complete feedback in the interference channel, so that both the received symbols \( Y_1 \) and \( Y_2 \) are fed back instantaneously and noiselessly to both the senders. The feedback link then acts like a channel between the senders. Feedback increases the capacity of the interference channel since it allows the two senders to co-operate in sending information to the two receivers. Without feedback, the codewords sent by the two senders are independent; with feedback, however, the codewords need not be independent.

With feedback, sender \( X_2 \) can act as a relay to help transmission of information from sender \( X_1 \) to receiver \( Y_1 \). Thus feedback can increase the maximum individual rates as well. This is in contrast to the broadcast channel or the multiple access channel, where feedback does not increase the maximum individual rates.
In particular, when the interference parameter \(a\) is very large, we have an almost noiseless channel between the senders due to the feedback link. After the senders communicate with each other through the feedback link, they can send information to the receivers in a cooperative manner at a much higher rate. More precisely, without feedback, the achievable rates satisfy (5.5). With feedback, on the other hand, we can achieve a rate

\[
R_1 = \frac{1}{2} \log \left( 1 + \frac{(1 + a^2)P}{N} \right), \quad R_2 = 0,
\]

(5.11)

if \(a = b, P_1 = P_2 = P\) and \(N_1 = N_2 = N\). (We will prove this later). If \(a \gg 1\), the individual rates can be increased by a very large amount by feedback and so can the sum of the rates.

We will thus not be able to prove simple factor of 2 bounds on the increase in capacity for the general interference channel with feedback. We will however examine a few characteristics of the interference channel with feedback.

- The capacity region without feedback depends only on the marginals of the conditional distribution. The capacity with feedback on the other hand depends on the joint distribution. In the Gaussian case, the capacity depends on the covariance of the noise terms \(Z_1\) and \(Z_2\) and not just on their individual variances.

- Since the senders receive feedback from both receivers and can subtract out their own codeword, they have a better look at the other senders transmissions than the intended receivers. Thus at the end of the transmission of a pair of codewords, \(X_1\) will be able to decode the message sent by \(X_2\) and vice versa.

Using the general outer bound for information flow in networks (Theorem 6.5.1), we can prove the following bounds for the rates for the interference channel with feedback:

**Theorem 5.2.1** The rates of the interference channel with feedback are bounded by

\[
R_1 \leq I(X_1, X_2; Y_1)
\]

(5.12)
\[ R_1 \leq I(X_1; Y_1, Y_2 | X_2) \]  \hspace{1cm} (5.13)
\[ R_2 \leq I(X_1, X_2; Y_2) \]  \hspace{1cm} (5.14)
\[ R_2 \leq I(X_2; Y_1, Y_2 | X_1) \]  \hspace{1cm} (5.15)
\[ R_1 + R_2 \leq I(X_1, X_2; Y_1, Y_2) \]  \hspace{1cm} (5.16)

for some joint distribution \( p(x_1, x_2)p(y_1, y_2 | x_1, x_2) \).

**Proof:** The proof follows directly from various choices of subsets in Theorem 6.5.1. \(\square\)

For the proof of the lower bound on the capacity region of the interference channel with feedback, we need some standard results\[23\] on joint typicality of sequences of random variables, which we repeat here for reference.

**Definition:** The set \( A_\varepsilon^{(n)} \) of jointly typical sequences \( \{(x^n, y^n)\} \) with respect to the density \( f(x, y) \) is the set of sequences whose empirical entropies are \( \varepsilon \)-close to the true entropies, i.e.,

\[ A_\varepsilon^{(n)} = \{(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n : \]  \hspace{1cm} (5.17)
\[ -\frac{1}{n} \log f(x^n) - h(X) < \varepsilon, \]  \hspace{1cm} (5.18)
\[ -\frac{1}{n} \log f(y^n) - h(Y) < \varepsilon, \]  \hspace{1cm} (5.19)
\[ -\frac{1}{n} \log f(x^n, y^n) - h(X, Y) < \varepsilon, \]  \hspace{1cm} (5.20)

where

\[ f(x^n, y^n) = \prod_{i=1}^n f(x_i, y_i). \]  \hspace{1cm} (5.21)

**Lemma 5.2.1 (Joint AEP.)** Let \( (X^n, Y^n) \) be sequences of length \( n \) drawn i.i.d. according to \( f(x^n, y^n) = \prod_{i=1}^n f(x_i, y_i) \). Then

1. \( P((X^n, Y^n) \in A_\varepsilon^{(n)}) \to 1 \) as \( n \to \infty \).

2. If \( (\tilde{X}^n, \tilde{Y}^n) \sim f(x^n) f(y^n) \), i.e., \( \tilde{X}^n \) and \( \tilde{Y}^n \) are independent with the same marginals as \( f(x^n, y^n) \), then

\[ P \left( (\tilde{X}^n, \tilde{Y}^n) \in A_\varepsilon^{(n)} \right) \leq 2^{-n(I(X;Y) - 3\varepsilon)}. \]  \hspace{1cm} (5.22)
Proof: See [35], [23]. □

We now state a lower bound on the capacity of the Gaussian interference channel with feedback.

Theorem 5.2.2 For the Gaussian interference channel shown in Figure 5.2, the set of rates satisfying

\[ R_1 \leq \min \left\{ C \left( \frac{\alpha_3 P_1 + (\sqrt{\alpha_1 P_1} + b \sqrt{\beta_1 P_2})^2}{b^2 \beta_3 P_2 + (\sqrt{\alpha_2 P_1} + b \sqrt{\beta_2 P_2})^2 + N_1} \right) \right\}, \]

\[ C \left( \frac{\alpha_3 P_1}{N_1} + \frac{a^2 \alpha_3 P_1}{N_2} \right) \}

\[ C \left( \frac{\beta_3 P_2 + (a \sqrt{\alpha_2 P_1} + \sqrt{\beta_2 P_2})^2}{a^2 \alpha_3 P_1 + (a \sqrt{\alpha_1 P_1} + \sqrt{\beta_1 P_2})^2 + N_2} \right), \]

\[ C \left( \frac{b^2 \beta_3 P_2}{N_1} + \frac{\beta_3 P_2}{N_2} \right) \}

are achievable for the channel with feedback, for all \( \alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \beta_3 \) such that

\[ \alpha_1 + \alpha_2 + \alpha_3 = 1 \]

\[ \beta_1 + \beta_2 + \beta_3 = 1 \]

Proof: We first describe the main idea of the proof. The theorem is related to the coding theorem for the relay channel[19]. We use block Markov encoding and superposition to achieve cooperation between the senders.

Each of the senders divides its power into three parts. Sender 1 uses power \( \alpha_3 P_1 \) to send new information, \( \alpha_1 P_1 \) to send information to cooperatively resolve the uncertainty of receiver 1 and \( \alpha_2 P_1 \) to cooperatively resolve the uncertainty of receiver 2. Sender 2 uses power \( \beta_3 P_2 \) to send new information, \( \beta_2 P_2 \) to send information to cooperatively resolve the uncertainty of receiver 2 and \( \beta_1 P_2 \) to cooperatively resolve the uncertainty of receiver 1. The feedback link enables the two senders to decode the information.
from the other sender and to then cooperate in the resolving the uncertainty at the receivers. The block Markov encoding procedure (sending information in one block to resolve the uncertainty in the previous block) and the superposition of information for the two receivers allows the two senders to cooperate with each other more effectively than they could for the interference channel without feedback.

**Generation of codebooks:**

1. Generate $2^{nR_1}$ codewords i.i.d. $\sim \mathcal{N}(0, \alpha_1 P_1 - \epsilon)$. Index these codewords by $U_1(w_1), w_1 \in \{1, \ldots, 2^{nR_1}\}$.

2. Generate $2^{nR_2}$ codewords i.i.d. $\sim \mathcal{N}(0, \beta_2 P_2 - \epsilon)$. Index these codewords by $U_2(w_2), w_2 \in \{1, \ldots, 2^{nR_2}\}$.

3. Generate $2^{nR_{01}}$ codewords i.i.d. $\sim \mathcal{N}(0, \alpha_1 P_1 - \epsilon)$. Index these codewords by $S_1(s_1), s_1 \in \{1, \ldots, 2^{nR_{01}}\}$.

4. Generate $2^{nR_{02}}$ codewords i.i.d. $\sim \mathcal{N}(0, \beta_2 P_2 - \epsilon)$. Index these codewords by $S_2(s_2), s_2 \in \{1, \ldots, 2^{nR_{02}}\}$.

5. Randomly partition the set of indices $\{1, \ldots, 2^{nR_1}\}$ into $2^{nR_{01}}$ bins, choosing the bin independently and uniformly over the set of bins. Let the bin index corresponding to $w_1$ be denoted by $s_1(w_1)$.

6. Randomly partition the set of indices $\{1, \ldots, 2^{nR_2}\}$ into $2^{nR_{02}}$ bins, choosing the bin independently and uniformly over the set of bins. Let the bin index corresponding to $w_2$ be denoted by $s_2(w_2)$.

**Coding:** We assume that we send $B - 1$ message pairs $(w_1, w_2)$ in $B$ blocks of length $n$. We will describe the coding starting at block $i$. We assume that at the end of block $i - 1$, each sender has decoded the message sent by the other sender in block $i - 1$. Let the estimate made by sender 1 of the message of sender 2 in block $i - 1$ be denoted $\hat{w}_{2,i-1}$. Sender 1 calculates the bin indices of $w_{1,i-1}$ and $\hat{w}_{2,i-1}$. Similarly, sender 2 calculates $s_1(\hat{w}_{1,i-1})$ and $s_2(w_{2,i-1})$.  

81
Then the codeword sent by sender 1 in block $i$ is

$$X_1(i) = U_1(w_{1,i}) + S_1(s_1(w_{1,i-1})) + \sqrt{\frac{\alpha_2 P_1}{\beta_2 P_2}} S_2(s_2(w_{2,i-1})),$$  \hspace{1cm} (5.27)

and the codeword sent by sender 2 is

$$X_2(i) = U_2(w_{2,i}) + \sqrt{\frac{\beta_1 P_2}{\alpha_1 P_1}} S_1(s_1(w_{1,i-1})) + S_2(s_2(w_{2,i-1})).$$ \hspace{1cm} (5.28)

The codewords $U_1$ and $U_2$ carry the new information. The codewords $S_1$ and $S_2$ are used to cooperatively resolve the receivers uncertainty.

Let the received sequences at receivers 1 and 2 be denoted by $Y_1(i)$ and $Y_2(i)$.

**Decoding:** We assume that the estimates made by the senders at the end of block $i - 1$ are correct.

1. We declare an error if any of the codewords exceeds the power constraint. By the strong law of large numbers, the probability of this goes to 0 as $n \to \infty$.

2. The senders use the feedback link to decode the other sender's message. Sender 1 first subtracts the known signals $S_1$ and $S_2$ from the received signals and then uses these signals to decode $w_{2,i}$. It is easy to show that after subtraction, we have an effective channel as shown in Figure 5.3.

For this channel, we can decode the message with an arbitrarily low probability of error if

$$R_2 \leq C \left( \frac{\beta_3 P_2}{N_2} + \frac{\beta_3^2 P_2}{N_1} \right).$$ \hspace{1cm} (5.29)

3. Similarly, sender 2 can decode $w_{1,i}$ if

$$R_1 \leq C \left( \frac{\alpha_3 P_1}{N_1} + \frac{\alpha_3^2 P_1}{N_2} \right).$$ \hspace{1cm} (5.30)
Figure 5.3: Effective channel from Sender 2 to sender 1

4. Receiver 1 decodes the bin index of $w_{1,i-1}$ by looking for the unique codeword $S_1(s_1)$ such that $(S_1(s_1), Y_1)$ are jointly typical. We know that $s_1$ can be decoded with an arbitrarily small probability of error if

$$R_{01} \leq C \left( \frac{(\sqrt{\alpha_1 P_1} + b\sqrt{\beta_1 P_2})^2}{\alpha_3 P_1 + b^2 \beta_3 P_2 + (\sqrt{\alpha_2 P_1} + b\sqrt{\beta_2 P_2})^2 + N_1} \right).$$

(5.31)

5. Similarly, receiver 2 can decode the bin index of $w_{2,i-1}$ if

$$R_{02} \leq C \left( \frac{(a\sqrt{\alpha_2 P_1} + \sqrt{\beta_2 P_2})^2}{a^2 \alpha_3 P_1 + \beta_3 P_2 + (a\sqrt{\alpha_1 P_1} + \sqrt{\beta_1 P_2})^2 + N_2} \right).$$

(5.32)

6. Receiver 1 constructs a list of possible messages $w_{1,i-1}$ such that $(U_1(w_1), S_1(s_1(w_{1,i-2})), Y_1(i-1))$ are jointly typical. It then intersects this set with the set of indices in the bin corresponding to $s_{1,i}$. By Lemma 7 of [19], there will be an
unique index in the intersection with high probability if

\[ R_1 \leq I(U_1; Y_1 | S_1) + R_{01} \]

\[ = C \left( \frac{\alpha_3 P_1}{b^2 \beta_3 P_2 + (\sqrt{\alpha_2 P_1} + b \sqrt{\beta_2 P_2})^2 + N_1} \right) + C \left( \frac{(\sqrt{\alpha_1 P_1} + b \sqrt{\beta_1 P_2})^2}{\alpha_3 P_1 + b^2 \beta_3 P_2 + (\sqrt{\alpha_2 P_1} + b \sqrt{\beta_2 P_2})^2 + N_1} \right) \]

\[ = C \left( \frac{\alpha_3 P_1 + (\sqrt{\alpha_1 P_1} + b \sqrt{\beta_1 P_2})^2}{b^2 \beta_3 P_2 + (\sqrt{\alpha_2 P_1} + b \sqrt{\beta_2 P_2})^2 + N_1} \right). \]

Thus receiver 1 can decode \( w_{1,i-1} \) with a low probability of error if this condition is satisfied.

7. Similarly, receiver 2 can decode \( w_{2,i-1} \) with a low probability of error if

\[ R_2 \leq C \left( \frac{\beta_3 P_2 + (a \sqrt{\alpha_2 P_1} + \sqrt{\beta_2 P_2})^2}{a^2 \alpha_3 P_1 + (a \sqrt{\alpha_1 P_1} + \sqrt{\beta_1 P_2})^2 + N_2} \right). \]

Thus at the end of block \( i \), receiver 1 has decoded \( w_{1,i-1} \) and receiver 2 has decoded \( w_{2,i-1} \). Sender 1 has decoded \( w_{2,i} \) and sender 1 has decoded \( w_{1,i} \). By the same arguments as in [19], we can show that after \( B \) blocks, we will have decoded \( B - 1 \) message pairs with an arbitrarily small probability of error.

\[ \square \]

In particular, if we are only sending information from sender 1 to receiver 1, then sender 2 acts as a relay. In this case, we can set \( \alpha_2 = \beta_2 = \beta_3 = 0 \) and \( \beta_1 = 1 \). The rate for sender 1 is then

\[ R_1 \leq \min \left\{ C \left( \frac{\alpha_3 P_1 + (\sqrt{\alpha_1 P_1} + b \sqrt{P_2})^2}{N_1} \right), C \left( \frac{\alpha_3 P_1}{N_1} + \frac{a^2 \alpha_3 P_1}{N_2} \right) \right\}, \]

for all choices of \( \alpha_1, \alpha_3 \) such that \( \alpha_1 + \alpha_3 = 1 \).
This rate is optimal, since when we are interested in only sending information from sender 1, then the channel reduces to a degraded relay channel, for which the capacity is known\(^{19}\) to be given by the expression above.

The above expression for the achievable region is in general not the capacity of the interference channel with feedback. This is because it does treat the interference as noise, which is appropriate for small values of the interference parameter. But for strong and very strong interference, it is better to decode the interfering signal and subtract it out. We could modify the above achievable region to take this into account. In this case, we can prove an achievable region of the form

\[
R_1 \leq \min \left\{ C \left( \frac{\alpha_3 P_1 + (\sqrt{\alpha_1 P_1} + b\sqrt{\beta_1 P_2})^2}{N_1} \right), \frac{\alpha_3 P_1}{N_1} + \frac{a^2 \alpha_3 P_1}{N_2} \right\},
\]

\[
C \left( \frac{a^2 \alpha_3 P_1 + (a\sqrt{\alpha_1 P_1} + \sqrt{\beta_1 P_2})^2}{\beta_3 P_2 + (a\sqrt{\alpha_1 P_1} + \sqrt{\beta_2 P_2})^2 + N_2} \right) \right\}
\]

\[
R_2 \leq \min \left\{ C \left( \frac{\beta_3 P_2 + (a\sqrt{\alpha_2 P_1} + \sqrt{\beta_2 P_2})^2}{N_2} \right), \frac{b^2 \beta_3 P_2}{N_1} + \frac{\beta_3 P_2}{N_2} \right\},
\]

\[
C \left( \frac{b^2 \beta_3 P_2 + (\sqrt{\alpha_2 P_1} + \sqrt{\beta_2 P_2})^2}{\alpha_3 P_1 + (\sqrt{\alpha_1 P_1} + b\sqrt{\beta_1 P_2})^2 + N_1} \right) \right\}. \quad (5.37)
\]

This region is larger than the region given in the theorem for large values of the interference parameters. For these values, this region includes the capacity region of the interference channel without feedback for strong interference\(^{43}\).

In summary, feedback in interference channel enables the senders to act as relays for each other. This can increase the achievable rate region by an arbitrary factor. We have proved an achievable rate region for the interference channel with feedback which is optimal for the cases when one of the rates is 0. The capacity of the interference channel with feedback in general has not been solved.
5.3 The two way channel

The two way channel was introduced by Shannon [74]; it was the first multi-user channel to be studied. The channel is illustrated in Figure 5.4.

There are two senders trying to talk to each other over this channel. They can use the past received symbols to decide what to send next; thus there is feedback already present in the definition of the channel.

Shannon showed that the following region was achievable: the convex hull of rate pairs $(R_1, R_2)$ satisfying

\begin{align*}
R_1 &\leq I(X_1; Y_2 | X_2) \\
R_2 &\leq I(X_2; Y_1 | X_1)
\end{align*}

for a product distribution of the form $p(x_1)p(x_2)p(y_1,y_2|x_1,x_2)$.

Shannon also proved that for any code for the two channel with a low probability of error, the rates must satisfy

\begin{align*}
R_1 &\leq I(X_1; Y_2 | X_2) \\
R_2 &\leq I(X_2; Y_1 | X_1)
\end{align*}

for some joint distribution of the form $p(x_1,x_2)p(y_1,y_2|x_1,x_2)$. 

86
The Shannon inner bound is derived by sending codewords that do not depend on the past received symbols. Thus the Shannon inner bound corresponds to no feedback. The Shannon outer bound on the other hand allows an arbitrary joint distribution. But it may not be always possible to achieve any joint distribution on the inputs using the feedback. Shannon tried to characterize the set of distributions that can be achieved using the feedback to set up dependence. Willems and Hekstra[84] have derived stronger bounds on the capacity of a two way channel using these ideas, but the problem of characterizing the possible joint distributions that can be set up (and thus the characterization of the capacity region of the channel) is still open. For example, the capacity region of the Blackwell multiplying channel, which is a single output two way channel with binary alphabets and \( Y_1 = Y_2 = X_1X_2 \), is still not known.

In the next chapter, we will generalize some of the ideas for two way channels when we consider the general \( m \) node network. We will consider the flow of information with and without feedback and prove bounds analogous to Shannon’s inner and outer bounds.

5.4 Summary

Feedback converts a general relay channel to a degraded relay channel, for which the capacity region is known. In this case, feedback simplifies the capacity calculations.

For the interference channel, feedback enables the senders to cooperate more effectively. In the case of very strong interference, the senders could use the feedback link to send information to each other and then relay it to the appropriate receiver. We prove an achievable region for the interference channel with feedback and show how feedback could help increase the total capacity by an arbitrary factor.

We also introduce the two-way channel, which is the most basic channel with feedback. We will generalize this channel to an \( m \)-node network in the next chapter.
Chapter 6

Feedback in a communication network

In this chapter, we will consider the role of feedback in a general communication network. We will first define a network of \( m \) nodes and define two kinds of feedback models for such a network. One model reduces the network to the multiple access channel with feedback that we had considered earlier. The second model is more realistic and we will derive an outer bound on the capacity of the network with feedback. We will compare this with a general inner bound on the capacity without feedback. We will end this chapter with a discussion of capacity regions for communication networks and discuss the role of feedback in changing the capacity region.

6.1 A model for a communication network

In this section, we will describe a common model that can be used to describe all discrete memoryless communication networks. This model was briefly introduced by El Gamal and Cover[35], but has not been studied in detail earlier. Surprisingly, there has not been much effort in multi-user information theory to find a common model for all networks. Van der Meulen[81] classified a number of multi-user channels and their capacity regions and in [79] and [80] described the known results for multiple access and broadcast channels respectively. In their book[24], Csiszár and Körner describe a large class of networks which combine distributed sources and multi-user channels. But
their model essentially deals with communication in one direction and it is very difficult to incorporate feedback in such a model. In his thesis, Changjia Chen[13] developed the model for communication for a special case of the network when all the nodes have the same output random variable. He proved an achievable region using finite state encoders at each node. Some of the results described in this chapter on the properties of the capacity region are generalizations of the results derived by Chen.

We will begin with a few definitions. In this chapter, we will use superscripts for node indices and subscripts for time indices.

A general multiple user network consists of $m$ nodes, each with an associated transmitter and receiver. At each instant of time, the $i$-th node sends a symbol $x^{(i)}$ that depends on the messages he wishes to send and (perhaps) the past received $y^{(i)}$ symbols. We assume that the simultaneous transmission of $(x^{(1)}, x^{(2)}, \ldots, x^{(m)})$ results in a random collection of received symbols $(Y^{(1)}, Y^{(2)}, \ldots, Y^{(m)})$ drawn according to the conditional probability distribution $p(y^{(1)}, y^{(2)}, \ldots, y^{(m)}|x^{(1)}, x^{(2)}, \ldots, x^{(m)})$, where $p(\cdot|\cdot)$ expresses all the effects of the noise and interference present in the network. If $p(\cdot|\cdot)$ takes on only the values 0 and 1, the network reduces to a deterministic network with interference and no noise.
The model is very general in that it includes all the channels that we have considered earlier as special cases. We wish to find that set of achievable communication rates for this network. Since many of the special cases of this network are still unsolved, we do not attempt to find the capacity region for the network. Instead, we will find inner and outer bounds, which are very loose, but do illustrate some of the features of such networks.

**Definition:** A $m$-node communication network,

$$\{\mathcal{X}^{(1)}, \mathcal{X}^{(2)}, \ldots, \mathcal{X}^{(m)}, p(y^{(1)}, y^{(2)}, \ldots, y^{(m)}|x^{(1)}, x^{(2)}, \ldots, x^{(m)}), y^{(1)}, y^{(2)}, \ldots, y^{(m)}\},$$

consists of $m$ input alphabets, $m$ output alphabets, and a probability transition matrix which describes the probability distribution on the outputs given the inputs.

Let $\mathcal{M} = \{1, 2, \ldots, m\}$ denote the set of nodes.

In the network defined above there are many possible communication paths. Some of them join a pair of nodes, others join a set of nodes to some other set of nodes.

**Definition:** A communication path $j$ in the network is defined by a non-empty set of input nodes $S_j \subset \mathcal{M}$ and a non-empty set of output nodes $T_j \subset \mathcal{M}$, such that $S_j \cap T_j = \emptyset$. The set of all possible paths will be denoted by $\mathcal{P}$. We will denote a path $j$ by its input and output sets, i.e., $(S_j|T_j)$

The total number of paths in a network is the number of ways of choosing two disjoint subsets of the set of $m$ nodes. Since every node must belong to the set $S_j$ or to the set $T_j$ or to the set of unused nodes, the total number of possible choices of sets is $3^m$. Of these, $2^{m+1} - 1$ have either $S_j$ or $T_j$ empty, and hence the number of paths in the network is

$$|\mathcal{P}| = 3^m - 2^{m+1} + 1. \quad \text{(6.1)}$$

We will have different messages being transmitted along the different paths. A message being sent along a path is available at all the input nodes that belong to the path. We have reliable transmission of the message if all the output nodes belonging to the path are able to decode the message correctly.
Definition: Let the set of paths incident on node $i$ be denoted $F_i$. Thus

$$F_i = \{ j \in \mathcal{P} : i \in S_j \}. \quad (6.2)$$

Let the set of paths that exit at node $i$ be denoted $G_i$. Thus

$$G_i = \{ j \in \mathcal{P} : i \in T_j \}. \quad (6.3)$$

Definition: A $(\{2^{nR_j}, j \in \mathcal{P}\}, n)$ code for the $m$ node network consists of the following:

1. *Message sets*, one for each path in the network. We will denote them by

$$\mathcal{W}_j = \{1, 2, \ldots, 2^{nR_j}\}. \quad (6.4)$$

2. *Encoding functions*, which map all the messages at a particular node onto to a sequence of inputs. There are two different types of encoding functions:

   - Without feedback. In this case, the transmitted symbols at a node depend only on the message to be sent from that node. Thus the encoders map the messages incident on the node to the codeword transmitted at that node, i.e.,

$$X^{(i)} : \{ \mathcal{W}_j : j \in F_i \} \rightarrow \left(\mathcal{X}^{(i)}\right)^n. \quad (6.5)$$

   - With feedback. The transmitted symbols depend not just on the messages but also on the past outputs available at that node. Thus the encoders map the messages incident on the node and the past received symbols to the next transmitted symbol, i.e.,

$$X_k^{(i)} : \{ \mathcal{W}_j : j \in F_i \} \times \left(\mathcal{Y}^{(i)}\right)^{k-1} \rightarrow \mathcal{X}_k^{(i)} \quad (6.6)$$

3. *Decoding functions*, which look at the output sequence at each node and map it into estimates of the messages meant for that node. The decoder $D^{(i)}$ maps the output sequence

$$D^{(i)} : \left(\mathcal{Y}^{(i)}\right)^n \times \{ \mathcal{W}_l : l \in F_i \} \rightarrow \{ \mathcal{W}_j : j \in G_i \}. \quad (6.7)$$
Figure 6.2: The multiple access channel without feedback

**Definition:** The probability of error at node $i$ is the probability that the decoded messages at node $i$ are not equal the transmitted messages, for a uniform distribution on all messages.

**Definition:** A rate vector $R_j, j \in \mathcal{P}$ is said to be achievable if there exists a sequence of $(\{2^{nR_j}, j \in \mathcal{P}\}, n)$ codes with the probability of error at each node going to 0.

**Definition:** The capacity region of the network is the closure of the set of achievable rates.

**Definition:** The capacity $C_j$ of a path $j$ is the maximum rate achievable on path $j$.

In Figures 6.2 to 6.10, we show how we can write all the channels that we have talked about so far in this framework.

The basic theorem in this chapter is an outer bound on the rate of flow of information in a network. This bound is valid for both networks with and without feedback. This theorem will be specialized to the various cases considered in this chapter. But before we talk about the outer bound, we will describe some of the properties of the capacity region of a general network.
\((X^{(1)}, Y^{(0)})\) •  
\((X^{(2)}, Y^{(0)})\) •  
\((X^{(0)}, \cdot)\) •  
\((X^{(0)}, Y^{(1)})\) •  
\((X^{(0)}, (Y^{(1)}, Y^{(2)}))\) •  
\((\cdot, Y^{(1)})\) •  
\((\cdot, Y^{(2)})\) •  
\((\cdot, Y^{(2)})\) •

Figure 6.3: The multiple access channel with feedback

Figure 6.4: The broadcast channel without feedback

Figure 6.5: The broadcast channel with feedback
\( (X^{(0)}, \cdot) \bullet (\cdot, Y^{(2)}) \)

Figure 6.6: The relay channel without feedback

\( (X^{(0)}, (Y^{(1)}, Y^{(2)})) \bullet (\cdot, Y^{(2)}) \)

Figure 6.7: The relay channel with feedback

\( (X^{(1)}, \cdot) \bullet (\cdot, Y^{(3)}) \)

\( (X^{(2)}, \cdot) \bullet (\cdot, Y^{(4)}) \)

Figure 6.8: The interference channel without feedback
\[(X^{(1)}, Y^{(3)}, Y^{(4)}) \bullet (\cdot, Y^{(3)})\]
\[(X^{(2)}, Y^{(3)}, Y^{(4)}) \bullet (\cdot, Y^{(4)})\]

Figure 6.9: The interference channel with feedback

\[(X^{(1)}, Y^{(1)}) \bullet (X^{(2)}, Y^{(2)})\]

Figure 6.10: The two way channel

### 6.2 Properties of capacity regions without feedback

The capacity region for the \(m\) node network was defined in the previous section. In this section, we will consider some of the properties of the capacity region without feedback. Later, we will introduce feedback and indicate some of the changes that can occur due to the feedback.

The capacity region must satisfy three basic properties:

1. **Convexity.** This is easily seen by timesharing. Thus if rate vectors \(R_1\) and \(R_2\) are achievable, then any convex combination \(\lambda R_1 + (1 - \lambda)R_2\) is achievable.

2. **Domination.** If a rate vector \(R\) is achievable, then any rate vector which is component by component less than or equal to \(R\) is also achievable. We need not use the entire rate to send useful information.
3. **Consistency.** This is a slightly more difficult concept, which we will illustrate by an example. Consider the broadcast channel with two receivers as shown in Figure 6.4. There are three possible paths in the network, namely (0|1), (0|2) and (0|1, 2). Let us call the rates along these three paths $R_1$, $R_2$ and $R_3$. If $(R_1, R_2, R_3)$ is achievable, then $(R_1 + \lambda R_3, R_2 + (1 - \lambda)R_3, 0)$ is also achievable, since we can send some of the information $(\lambda R_3)$ meant for receiver 1 along the common path (0|1, 2) and some of the information $((1 - \lambda)R_3)$ meant for receiver 2 along the common path. Both receivers will decode the common information. But each receiver will use only the part of the common information intended for it.

In general, for a $m$-node network described in the previous section, instead of sending information along a path $(S_j|T_j)$, we can send it along a path $(S'_j|T'_j)$ where $S'_j \subseteq S_j$ and $T'_j \supseteq T_j$. The information will be sent from fewer senders and decoded by more receivers, but the additional receivers can just discard the information. Thus any rate achievable along the path $(S'_j|T'_j)$ can be transferred to the path $(S_j|T_j)$. The capacity region of the network must be closed under such transfers. We will call the property consistency of the capacity region. This was first described in detail by Hajek and Pursley[40], who showed that the achievable regions for the broadcast channel proposed by Cover[16] and Van der Meulen[78] can be increased by using the above procedure.

To describe the consistency of the capacity region of a network in general, we will introduce two partial orderings on the set of paths $\mathcal{P}$ in a network.

**Definition:** We will say that a path $j \in \mathcal{P}$ is a sub-path of path $k \in \mathcal{P}$ (denoted by $j < k$) if $S_j \subset S_k$, and $T_j = T_k$. Thus a sub-path starts in fewer nodes and ends in the same nodes as the original path.

The operation $<$ defines a partial order on the set of paths. The partial ordering is illustrated for a 3 sender multiple access channel in Figure 6.11.

We can define another partial order based on the destination sets of a path.

**Definition:** We will say that a path $j \in \mathcal{P}$ is more versatile than path $k \in \mathcal{P}$ (denoted by $j \ll k$) if $S_j = S_k$, and $T_k \subset T_j$. Thus a more versatile path has
Figure 6.11: Partial order induced by \( \prec \) for 3 sender multiple access channel

the same input set and a larger output set.

The operation \( \ll \) defines another partial order on the set of paths. The partial ordering is illustrated for a 3 receiver broadcast channel in Figure 6.12.

We can use these properties to clarify the third property of capacity regions. We can split the requirement of consistency of the capacity region into two basic requirements on the rates for paths in the path hierarchy defined above:

1. If \( j \prec k \), and if a rate vector with \( R_j = a \) and \( R_k = b \) is achievable, then a rate vector with \( R_j = 0, R_k = a + b \) is achievable for any \( a, b \geq 0 \). This is because the nodes in \( S_j \) also have the message to be transmitted on path \( k \). So they can send this message to all the nodes in \( T_k = T_j \) using the codes for path \( j \) and thus transfer the achievable rate on the path \( j \) to the path \( k \).

This can be used to show that the capacity of the path \( k \) is greater than (or equal to) the capacity of path \( j \), i.e., if \( j \prec k \), then \( C_j \leq C_k \). In fact, the capacity of path \( k \) must be at least the maximum achievable sum of rates for all sub-paths of path \( k \). For example, if \( S_k = \{1, 3\} \), then \( C_k \geq R_{1|T_k} + R_{3|T_k} \) for any achievable rate vector.
2. If \( j \ll k \), and if a rate vector with \( R_j = a \) and \( R_k = b \) is achievable, then a rate vector with \( R_j = 0, R_k = a+b \) is achievable for any \( a, b \geq 0 \). This is because any information that is to be sent to \( T_k \) can also be sent to the larger set \( T_j \). The remaining nodes in \( T_j \) will just ignore the information.

An immediate consequence of this property is the fact that if \( j \ll k \), \( C_j \leq C_k \). The capacity of the network to send information to a set of nodes is greater than its capacity to send information to a larger set.

These two properties are implied by and imply the consistency property defined above. The reason for defining two separate orderings on the paths (rather than just one) will be made clear when we consider the role of feedback.

The above properties continue to hold when we add feedback as well. We can show that the regions defined by the outer bounds derived in this chapter also satisfy these properties. In the last section of this chapter, we will examine the role of feedback in modifying the capacity region.
6.3 Networks with no feedback

In this section, we will consider a network with no feedback, i.e., one for which the codewords sent depend only on the messages but not on the past received symbols. To begin, we will define what we mean by no feedback for the $m$ node network defined in the previous section.

**Definition:** A code for the $m$-node network without feedback consists of a set of encoders, one for each node, which map the messages for all paths that start at that node onto a sequence of input symbols for that node.

Note that the encoders do not use the past received symbols at the node.

**Theorem 6.3.1** The total rate of flow of information from a subset of the nodes $S$ to another disjoint subset $T$ is bounded by the mutual information between the inputs in $S$ and the outputs in $T$, conditioned by the inputs in $S^c$, i.e.,

$$\begin{align*}
R(S, T) &= \sum_{S_j \subset S, T_j \cap T \neq \emptyset} R^{(j)} \\
&\leq I(X^{(S)}; Y^{(T)}|X^{(S^c)})
\end{align*} \tag{6.8}$$

for some joint distribution $p(x^{(1)}, x^{(2)}, \ldots, x^{(m)})p(y^{(1)}, y^{(2)}, \ldots, y^{(m)}|x^{(1)}, x^{(2)}, \ldots, x^{(m)})$.

**Proof:** The proof follows the same lines as the proof of the converse for the multiple access channel. Let $P = \{j : S_j \subset S, T_j \cap T \neq \emptyset\}$ be the set of paths that cross from $S$ to $T$, and let $P^c$ be all the other paths in the network. Then

$$\begin{align*}
\sum_{j \in P} R^{(j)} &= 0 \\
&= H(W^{(j)}) \quad \text{(a)} \tag{6.9}
\end{align*}$$

$$\begin{align*}
&= H(W^{(P)}) \quad \text{(b)} \tag{6.10}
\end{align*}$$

$$\begin{align*}
&= H(W^{(P)}|W^{(P^c)}) \quad \text{(c)} \tag{6.11}
\end{align*}$$

$$\begin{align*}
&= I(W^{(P)}; Y_1^{(T)}, \ldots, Y_n^{(T)}|W^{(P^c)}) \quad \text{(d)} \tag{6.12}
\end{align*}$$

$$\begin{align*}
&+ H(W^{(P)}|Y_1^{(T)}, \ldots, Y_n^{(T)}, W^{(P^c)}) \quad \text{or} \tag{6.13}
\end{align*}$$

$$\begin{align*}
&\leq I(W^{(P)}; Y_1^{(T)}, \ldots, Y_n^{(T)}|W^{(P^c)}) + n \epsilon_n \quad \text{or} \tag{6.14}
\end{align*}$$
\begin{align}
\sum_{k=1}^{n} I(W^{(P)}; Y_k^{(T)}|Y_1^{(T)}, \ldots, Y_{k-1}^{(T)}, W^{(P^c)}) + n\epsilon_n \tag{6.15} \\
\sum_{k=1}^{n} H(Y_k^{(T)}|Y_1^{(T)}, \ldots, Y_{k-1}^{(T)}, W^{(P^c)}) \\
- H(Y_k^{(T)}|Y_1^{(T)}, \ldots, Y_{k-1}^{(T)}, W^{(P^c)}, W^{(P)}) + n\epsilon_n \tag{6.16} \\
\sum_{k=1}^{n} H(Y_k^{(T)}|Y_1^{(T)}, \ldots, Y_{k-1}^{(T)}, W^{(P^c)}, X_k^{(S^c)}) \\
- H(Y_k^{(T)}|Y_1^{(T)}, \ldots, Y_{k-1}^{(T)}, W^{(P^c)}, W^{(P)}, X_k^{(S)}, X_k^{(S^c)}) + n\epsilon_n \tag{6.17} \\
\sum_{k=1}^{n} H(Y_k^{(T)}|X_k^{(S^c)}) - H(Y_k^{(T)}|X_k^{(S)}, X_k^{(S^c)}) + n\epsilon_n \tag{6.18} \\
\sum_{k=1}^{n} I(X_k^{(S)}; Y_k^{(T)}|X_k^{(S^c)}) + n\epsilon_n \tag{6.19} \\
\sum_{k=1}^{n} \frac{1}{n} I(X_k^{(S)}; Y_k^{(T)}|X_k^{(S^c)}, Q = k) + n\epsilon_n \tag{6.20} \\
\sum_{k=1}^{n} I(X_k^{(S)}; Y_k^{(T)}|X_k^{(S^c)}, Q = k) + n\epsilon_n \tag{6.21} \\
H(Y_k^{(T)}|X_k^{(S^c)}, Q) - H(Y_k^{(T)}|X_k^{(S)}, X_k^{(S^c)}, Q) + n\epsilon_n \tag{6.22} \\
H(Y_k^{(T)}|X_k^{(S^c)}) - H(Y_k^{(T)}|X_k^{(S)}, X_k^{(S^c)}) + n\epsilon_n \tag{6.23} \\
\end{align}

where
(a) follows from the fact that the messages \( W^{(j)} \) are uniformly distributed on the range \( \{1, 2, \ldots, 2^{nR(j)}\} \),
(b) follows from the definition of \( W^{(P)} = \{ W^{(j)} : j \in P \} \) and the fact that the messages are independent,
(c) follows from the independence of the messages for \( P \) and \( P^c \),
(d) follows from Fano's inequality, since the messages \( W^{(P)} \) can be decoded from \( Y^{(T)} \) and \( W^{(P^c)} \),
(e) is the chain rule for mutual information,
(f) follows from the definition of mutual information,
(g) follows from the fact that \( X_k^{(S^c)} \) is a function of the messages \( W^{(P^c)} \), and the fact that adding conditioning reduces the second term,
(h) from the fact that $Y_k^{(T)}$ depends only on the current input symbols $X_k^{(S)}$ and $X_k^{(S^e)}$, 
(i) follows after we introduce a new time sharing random variable $Q$ uniformly distributed on $\{1, 2, \ldots, n\}$, 
(j) follows from the definition of mutual information, and 
(k) follows from the fact that $Y_Q^{(T)}$ depends only the inputs $X_Q^{(S)}$ and $X_Q^{(S^e)}$ and is conditionally independent of $Q$.

Thus there exist random variables $X^{(S)}$ and $X^{(S^e)}$ with some arbitrary joint distribution which satisfy the inequalities of the theorem. □

For the network without feedback, we can apply Theorem 6.3.1, which gives a weak outer bound on the capacity of the network. The bound is weak since it does not say anything about the joint distribution of $X^{(1)}, X^{(2)}, \ldots, X^{(m)}$. For the no feedback case, we can derive some further conditions on the joint distribution, since the $X$'s depend only on the messages and the messages are independent. The general relationship between the symbols is determined by the relationships of the messages on the various paths, and is very difficult to express in general. So we will consider a special case to illustrate the concepts.

### 6.3.1 Networks with no common information

We will consider networks where the only active paths have a single input node and a single output node. Thus no two nodes have the same message to transmit, which is why we will call these networks with no common information.

**Definition:** A network without common information has $m(m - 1)$ active paths from every node to every other node, and all other paths inactive. We will denote the rate from node $i$ to node $k$ by $R^{(ik)}$.

We can then specialize Theorem 6.3.1 to this case to prove the following theorem:
Theorem 6.3.2 The flow of information in a network with no feedback and no common information can be bounded by

\[ \sum_{i \in S, k \in T} R^{(ik)} \leq I(X^{(S)}; Y^{(T)}|X^{(S^c)}, Q) \]  \hspace{1cm} (6.24)

for some joint distribution \( p(q) \prod_{i=1}^{m} p(x^{(i)}|q)p(y^{(1)}, y^{(2)}, \ldots, y^{(m)}|x^{(1)}, x^{(2)}, \ldots, x^{(m)}) \).

Furthermore, this region is not changed if we restrict the cardinality of \( Q \) to \( 2^m \).

Proof: The chain of inequalities in the proof of Theorem 6.3.1 goes through without much change. We will stop the chain at \( (6.20) \) to obtain

\[ n \sum_{i \in S, k \in T} R^{(ik)} = H(W^{(ik)}) \leq nI(X_Q^{(S)}; Y_Q^{(T)}|X_Q^{(S^c)}, Q) + n\epsilon_n, \]  \hspace{1cm} (6.26)

where \( \epsilon_n \to 0 \) as \( n \to \infty \). Now dividing by \( n \) and taking the limit at \( n \to \infty \), we have

\[ \sum_{i \in S, k \in S_c} R^{(ik)} \leq I(X_Q^{(S)}; Y_Q^{(T)}|X_Q^{(S^c)}, Q) \]  \hspace{1cm} (6.27)

To complete the proof of the theorem, we have to argue that \( \{X_Q^{(i)}\} \) are independent given \( Q \). But the distribution of \( \{X_Q^{(i)}\} \) is determined by the encoding and given \( Q = i \), the distribution depends only the messages being sent. Since all the messages are independent and each \( X_Q^{(i)} \) depends on a different set of messages, the \( X_Q^{(i)} \) are independent. Setting \( X^{(i)} \triangleq X_Q^{(i)} \) and \( Y^{(i)} \triangleq Y_Q^{(i)} \), we have

\[ \sum_{i \in S, k \in S_c} R^{(ik)} \leq I(X^{(S)}; Y^{(T)}|X^{(S^c)}, Q) \]  \hspace{1cm} (6.28)

for a joint distribution \( p(q) \prod_{i=1}^{m} p(x^{(i)}|q)p(x^{(1)}, x^{(2)}, \ldots, x^{(m)})p(y^{(1)}, y^{(2)}, \ldots, y^{(m)}) \).

This theorem gives an upper bound on the rate of information flow between any two subsets in the network. The bound is not tight however; these
bounds correspond to all the receivers in $T$ co-operating with each other to
decode the messages from $S$ to $T$. We can prove an inner bound on the ca-
pacity region generalizing the Cover-Van der Meulen region for the broadcast
channel.

**Theorem 6.3.3** For network with no feedback and no common information,
the following rates are achievable

$$R^{(ik)} \leq I(U^{(ik)}; Y^{(k)})$$

(6.29)

for some auxiliary random variables $U^{(ik)}$ with a joint distribution

$$\prod_{i=1}^{m} \prod_{k=1}^{m} p(u^{(ik)}) \prod_{i=1}^{m} p(x^{(i)}|u^{(i1)}, \ldots, u^{(im)})p(y^{(1)}, y^{(2)}, \ldots, y^{(m)}|x^{(1)}, x^{(2)}, \ldots, x^{(m)})$$

with the understanding that $U^{(ii)} = \phi$ and $R^{(ii)} = 0$.

**Proof:** The proof follows the same lines as the proof in Cover[16].

The achievable region given in the above theorem is not the capacity
region of the network, since it does not give the capacity region for the
broadcast channel. The theorem can be extended to the case of common
information by the introduction for some more auxiliary random variables
respecting the common information. However, even this does not give
the complete capacity region of the broadcast channel. In fact, Hajek and
Pursley[40] have shown that even after the addition of common information,
the region does not satisfy the consistency property of capacity regions (de-
scribed in Section 6.2) and that the region can be extended by using this
property.

### 6.4 Networks with complete feedback

In all the multiple user channels that we have considered so far in this report,
we have assumed that with feedback, the received symbols from all the
receivers are fed back instantaneously and noiselessly to all the senders. For
the $m$-node network considered in this chapter, this would imply that all
the $m$ received symbols at time $i$, i.e., $Y_i^{(1)}, \ldots, Y_i^{(m)}$, are fed back to all
the nodes before time $i+1$. So all the nodes effectively receive a vector of
symbols $Y^{(1)}, Y^{(2)}, \ldots, Y^{(m)}$. The network with feedback therefore is similar
to a multiple access channel with feedback and some of the earlier results for
feedback in multiple access channels can be extended to this case. But we
will not proceed in this direction, since the model seems rather unrealistic.
Instead, we will concentrate on a more realistic model of feedback described
in the next section. The network with complete feedback can be described with
the single-node feedback model by letting all the outputs to be the same.

6.5 Networks with single node feedback

In Section 6.3, we considered a network without feedback where the transm-
ted symbols depended only on the messages that needed to be transmitted
and not on the past received symbols. If we let the transmitted symbol at a
node depend on the past received symbols at that node, we have a network
with single node feedback. This is a much more realistic model than the
complete feedback model of the previous section. But feedback in this case
is slightly different from that discussed in the previous chapters, in that the
feedback is not explicit but implicit in the received symbols. In this sense,
the feedback is similar to the generalized feedback for the multiple access
channel considered by King[50].

In a real network, some of the nodes will have messages to transmit; the
others may act just as relays. The various possible strategies and coding
schemes are very complex to analyze in general. We will derive an outer
bound on the flow of information between a subset of nodes and its comple-
ment, and use that to prove some simple properties of such networks.

We will now use this formulation to derive an upper bound on the flow
of information in any multiterminal network. We will divide the nodes into
two sets, $S$ and the complement $S^c$. We will then bound the rate of flow of
information from nodes in $S$ to nodes in $S^c$. This theorem generalizes the

Theorem 6.5.1 The total rate of flow of information from $S$ to $S^c$ is bound-
ed by the mutual information between the inputs in $S$ and the outputs in $S^c$,  

104

conditioned by the inputs in $S^c$, i.e.,
\[
\sum_{j : S_j \subseteq S, T_j \cap S^c \neq \emptyset} R^{(j)} \leq I(X^{(S)}; Y^{(S^c)}|X^{(S^c)})
\]  
(6.30)

for some joint distribution $p(x^{(1)}, \ldots, x^{(m)})$.

**Proof:** The proof follows the same lines as the proof of Theorem 6.3.1. Let $P = \{j : S_j \subseteq S, T_j \cap S^c \neq \emptyset\}$ be the set of links that cross from $S$ to $S^c$, and let $P^c$ be all the other links in the network. Then

\[
\begin{align*}
&n \sum_{S_j \subseteq S, T_j \cap S^c \neq \emptyset} R^{(j)} \\
&\overset{(a)}{=} \sum_{S_j \subseteq S, T_j \cap S^c \neq \emptyset} H(W^{(j)}) \\
&\overset{(b)}{=} H(W^{(P)}) \\
&\overset{(c)}{=} H(W^{(P)}|W^{(P^c)}) \\
&= I(W^{(P)}; Y_1^{(S^c)}, \ldots, Y_n^{(S^c)}|W^{(P^c)})
\end{align*}
\]  
(6.31)
\[ + H(W^{(P)}|Y_1^{(S_e)}, \ldots, Y_n^{(S_e)}, W^{(P_c)}) \]  

\[ \leq I(W^{(P)}; Y_1^{(S_e)}, \ldots, Y_n^{(S_e)}|W^{(P_c)}) + n\epsilon_n \]  

\[ \leq \sum_{k=1}^n I(W^{(P)}; Y_k^{(S_e)}|Y_1^{(S_e)}, \ldots, Y_{k-1}^{(S_e)}, W^{(P_c)}) + n\epsilon_n \]  

\[ \leq \sum_{k=1}^n H(Y_k^{(S_e)}|Y_1^{(S_e)}, \ldots, Y_{k-1}^{(S_e)}, W^{(P_c)}) \]  

\[ - H(Y_k^{(S_e)}|Y_1^{(S_e)}, \ldots, Y_{k-1}^{(S_e)}, W^{(P_c)}, W^{(P)}) + n\epsilon_n \]  

\[ \leq \sum_{k=1}^n H(Y_k^{(S_e)}|Y_1^{(S_e)}, \ldots, Y_{k-1}^{(S_e)}, W^{(P_c)}, X_k^{(S_e)}) \]  

\[ - H(Y_k^{(S_e)}|Y_1^{(S_e)}, \ldots, Y_{k-1}^{(S_e)}, W^{(P_c)}, W^{(P)}, X_k^{(S_e)}, X_k^{(S_e)}) + n\epsilon_n \]  

\[ \leq \sum_{k=1}^n H(Y_k^{(S_e)}|X_k^{(S_e)}) - H(Y_k^{(S_e)}|X_k^{(S_e)}, X_k^{(S_e)}) + n\epsilon_n \]  

\[ = \sum_{k=1}^n I(X_k^{(S_e)}; Y_k^{(S_e)}|X_k^{(S_e)}) + n\epsilon_n \]  

\[ \geq \frac{1}{n} \sum_{k=1}^n I(X_Q^{(S_e)}; Y_Q^{(S_e)}|X_Q^{(S_e)}, Q = k) + n\epsilon_n \]  

\[ \geq n I(X_Q^{(S_e)}; Y_Q^{(S_e)}|X_Q^{(S_e)}, Q) + n\epsilon_n \]  

\[ = n \left( H(Y_Q^{(S_e)}|X_Q^{(S_e)}, Q) - H(Y_Q^{(S_e)}|X_Q^{(S_e)}, X_Q^{(S_e)}, Q) \right) + n\epsilon_n \]  

\[ \leq n \left( H(Y_Q^{(S_e)}|X_Q^{(S_e)}) - H(Y_Q^{(S_e)}|X_Q^{(S_e)}, X_Q^{(S_e)}) \right) + n\epsilon_n \]  

\[ = n I(X_Q^{(S_e)}; Y_Q^{(S_e)}|X_Q^{(S_e)}) + n\epsilon_n, \]  

where
(a) follows from the fact that the messages $W^{(i)}$ are uniformly distributed on the range $\{1, 2, \ldots, 2^{nR^{(i)}}\}$,
(b) follows from the definition of $W^{(P)} = \{W^{(i)} : j \in P\}$ and the fact that the messages are independent,
(c) follows from the independence of the messages for $P$ and $P^c$,
(d) follows from Fano's inequality, since the messages $W^{(P)}$ can be decoded from $Y^{(S)}$ and $W^{(P_c)}$,
(e) is the chain rule for mutual information,
(f) follows from the definition of mutual information,
(g) follows from the fact that $X_k^{(S_e)}$ is a function of the past received symbols $Y^{(S_e)}$ and the messages $W^{(P_e)}$, and the fact that adding conditioning reduces the second term,

(h) from the fact that $Y^{(S_e)}$ depends only on the current input symbols $X^{(S)}$ and $X^{(S_e)}$,

(i) follows after we introduce a new time sharing random variable $Q$ uniformly distributed on $\{1, 2, \ldots, n\}$,

(j) follows from the definition of mutual information, and

(k) follows from the fact that $Y_Q^{(S_e)}$ depends only the inputs $X_Q^{(S)}$ and $X^{(S_e)}$ and is conditionally independent of $Q$.

Thus there exist random variables $X^{(S)}$ and $X^{(S_e)}$ with some arbitrary joint distribution which satisfy the inequalities of the theorem. \(\square\)

Unlike Theorem 6.3.1, in this theorem, we only bound the information flow between a set of nodes and its complement, and not between any two sets. It is not possible to extend this bound to flow between any two sets because the nodes outside these two sets could act as relays or helpers in the communication. This was not possible in the network of Theorem 6.3.1, in which there is no feedback.

As with the other bounds that we have derived in this chapter, this bound is not tight. However, we can use it to illustrate some of the basic properties of the feedback capacity region.

1. Feedback does not increase the maximum rate at which information can enter a node.

This is because the maximum rate at which information can be transmitted to a node (say node 1) is when all the other nodes cooperate to send it information. This is the maximum rate for the path whose input set is the set of nodes $S = \{2, 3, \ldots, m\}$ and whose output set is $S^c = \{1\}$. This maximum rate is

$$R(S, S^c) = \max_{p(x^{(1)}, x^{(2)}, \ldots, x^{(m)})} I(X^{(2)}, \ldots, X^{(m)}; Y^{(1)} | X^{(1)})$$

(6.45)

$$= \max_{x^{(1)}} \max_{p(x^{(2)}, \ldots, x^{(m)})} I(X^{(2)}, \ldots, X^{(m)}; Y^{(1)} | X^{(1)} = x^{(1)})$$

(6.46)
which can be achieved without feedback when all the nodes in $S$ are sending the same message.

Thus the maximum rates on all paths with $m - 1$ input nodes and 1 output node are not changed by feedback.

2. The maximum achievable total throughput of the network (the sum of the rates of all paths in the network) is increased by at most a factor of $m$ by feedback.

This is a generalization of the geometric argument that gave a factor of $m$ for the multiple access channel with feedback. All paths in the network must end in some node (or nodes). So all paths must contribute to $\hat{R}(i^c,i)$ for some node $i$. Thus the total throughput in the network is bounded by

$$R_T \leq \sum_{i=1}^{m} \hat{R}(i^c,i).$$  \hfill (6.47)

The maximum value of $\hat{R}(i^c,i)$ can be attained without feedback when all the nodes other than $i$ are sending the same message. Since the sum of $m$ numbers is less than $m$ times the largest number, the maximum throughput with feedback is less than $m$ times the maximum achievable without feedback.

This result is not as impressive as it first appears, since it compares the total throughput with feedback with the total throughput achievable over all paths without feedback. However, in any network, we are generally interested only in a small subset of the possible paths and the total throughput on these paths can increase by an arbitrary factor (as illustrated by the interference channel results).

3. Feedback can increase the maximum rate achievable on a single link.

For a $m$ node network without feedback, the maximum rate of communication from node $i$ to node $k$ is achieved when all the other nodes keep quiet. But when we allow feedback, the other nodes can act as relays to help the communication from node $i$ to node $j$. Thus feedback can increase the rate of communication along a single link.

This is unlike the case of the multiple access channel or the broadcast channel with feedback. There feedback did not increase the maximum rate along any single link.
Feedback does not increase the maximum achievable rate for common information. Feedback essentially enables the conversion of independent information into common information, which can be communicated more efficiently.

6.5.1 Networks with no common information

In this case, all the messages originate at a single node and are destined for a single node. But since we now allow feedback, other nodes could act as relays to help the communication. Thus feedback can increase the maximum achievable rates along any single link.

The upper bound derived earlier on the rate of information flow for the network with feedback can be applied to this case as well. The only active paths are those from one single node to another single node, so we can write the upper bound as

$$R(S, S^c) = \sum_{i \in S, k \in S^c} R^{(ik)} \leq I(X^{(S)}; Y^{(S^c)}|X^{(S^c)}),$$

(6.48)

for some joint distribution $p(x^{(1)}, x^{(2)}, \ldots, x^{(m)})$. There are two differences between the bounds with and without feedback:

1. Without feedback, the bounds allowed only for a product distribution on the input symbols $X^{(1)}, X^{(2)}, \ldots, X^{(m)}$. With feedback, the bounds are for a joint distribution.

2. The bounds without feedback applied to flow between any two sets of nodes. With feedback, we have bounds only for the flow between a set and its complement.

6.6 Properties of capacity regions with feedback

In Section 6.2, we discussed the properties of a capacity region without feedback. These properties continue to hold when we add feedback to the network. In addition, feedback can increase the capacity region in two essentially different ways:
1. *Multiple Access Channels.* Consider two paths $j$ and $k$ such that $T_j = T_k$ and $S_j \cap S_k = \emptyset$, and consider the path $l$ where $T_l = T_j = T_k$ and $S_l = S_j \cup S_k$. The path $l$ has the same destination set and has a source set that is the union of the source sets of $j$ and $k$. Thus $j \prec l$ and $k \prec l$. From the first property on pg. 97, the capacity of the path $l$ is at least the sum of the rates achievable on paths $j$ and $k$.

If we had complete feedback from the destination set $T_l$ to all the senders in $S_l$, then the senders in $S_j$ will be able to find out what the senders in $S_k$ are sending and similarly senders in $S_k$ can find out about senders in $S_j$. Then they can cooperate in sending information along path $l$ to the destination set. Thus feedback converts some of the independent information on paths $j$ and $k$ to common information to be sent along path $l$, and since the capacity of path $l$ is higher, feedback helps to increase the maximum sum of rates.

2. *Broadcast channels.* Consider two paths $j$ and $k$ such that $S_j = S_k$ and $T_j \cap T_k = \emptyset$, and consider the path $l$ where $S_l = S_j = S_k$ and $T_l = T_j \cup T_k$. The path $l$ has the same source set and has a destination set that is the union of the destination sets of $j$ and $k$. Thus $l \ll j$ and $l \ll k$.

If we had complete feedback from the destination set $T_l$ to all the senders in $S_l$, then the senders in $S_l$ will be able to calculate the uncertainty at the receivers. They can then calculate the common part of this uncertainty and send that information along the path $l$. Thus feedback converts some of the independent information on paths $j$ and $k$ to common information to be sent along path $l$. Since the same information sent along path $l$ is playing a double role (correcting the uncertainty at receivers in $T_j$ as well as $T_l$), feedback increases the maximum achievable sum of rates along paths $j$ and $k$.

Thus in both cases, we can explain the role of feedback by considering the fact that feedback converts independent information into common information, which can then be sent more efficiently. The same principle can be used to explain why feedback increases the capacity of an interference channel.
6.7 Summary

In this chapter, we set up a common framework for the description of communication in a network. We gave a definition of feedback within this framework and derived outer bounds on the rate of flow of information in such a network with and without feedback.

Network information theory would have a simple and elegant solution if the outer bounds derived in this chapter were indeed achievable. But they are not achievable, even in some simple cases like the broadcast channel. In Sections 3.7 and 4.5, we have described classes of networks where these outer bounds are tight. But in general the outer bounds are in terms of a joint distribution, whereas the achievable regions are given in terms of a product distribution. In Chapter 9, we will give some further philosophical comments on these outer bounds.
Chapter 7

Subset inequalities

7.1 Introduction

In the previous chapters, we have derived various bounds on the rate of flow of information in a network. Many of the bounds were related to the mutual information between random variables associated with various subsets of the set of nodes in the network. In this chapter, we will study the behavior of the entropy and mutual information of subsets of random variables. We will apply some of these results to make a few simple statements about the flow of information in a network.

The inequalities derived in this chapter are of independent interest since they are interpolations of the fact that the entropy of a collection of random variables is less than the sum of the individual entropies. When we let the set of random variables be jointly normal, then we can derive some determinant inequalities, including Szasz's interpolation of Hadamard's inequality. We will describe some of these inequalities later in this chapter.

7.2 Subset inequalities for entropy

Let $X_1, X_2, \ldots, X_n$ be a set of $n$ random variables with an arbitrary joint distribution. Let $S$ be any subset of the indices $\{1, 2, \ldots, n\}$. We will use $X^{(S)}$ to denote the subset of random variables with indices in $S$ and $S^c$ to denote the complement of $S$ with respect $\{1, 2, \ldots, n\}$. For example, if $S = \{1, 3\}$, then $X^{(S)} = \{X_1, X_3\}$ and $X^{(S^c)} = \{X_2, X_4, X_5, \ldots, X_n\}$. Let
the entropy \( h(X) \) of a random vector \( X \in \mathbb{R}^k \) with density function \( f(x) \) be defined by

\[
h(X) = - \int f(x) \log f(x) \, dx.
\]  \hspace{1cm} (7.1)

If \( S = \{i_1, i_2, \ldots, i_k\} \), let

\[
h(X^{(S)}) = h(X_{i_1}, X_{i_2}, \ldots, X_{i_k}).
\]  \hspace{1cm} (7.2)

Let

\[
h^{(n)}_k = \frac{1}{\binom{n}{k}} \sum_{S:|S|=k} \frac{h(X^{(S)})}{k}
\]  \hspace{1cm} (7.3)

be the entropy rate per element for subsets of size \( k \) as the averaged over all \( k \)-element subsets. Here \( h^{(n)}_k \) is the average entropy in bits per symbol of a randomly drawn \( k \)-element subset of \( \{X_1, X_2, \ldots, X_n\} \). We can then prove the following theorem (this theorem is implicit in the results of Han[42]):

**Theorem 7.2.1**

\[
h^{(n)}_1 \geq h^{(n)}_2 \geq \cdots \geq h^{(n)}_n.
\]  \hspace{1cm} (7.4)

**Proof:** We will first prove the last inequality, i.e., \( h^{(n)}_n \leq h^{(n)}_{n-1} \).

We write

\[
h(X_1, X_2, \ldots, X_n) = h(X_1, X_2, \ldots, X_{n-1}) + h(X_n|X_1, X_2, \ldots, X_{n-1}),
\]  \hspace{1cm} (7.5)

\[
h(X_1, X_2, \ldots, X_n) \leq h(X_1, X_2, \ldots, X_{n-2}, X_n) + h(X_{n-1}|X_1, X_2, \ldots, X_{n-2}, X_n)
\]

\[
\quad \leq h(X_1, X_2, \ldots, X_{n-2}, X_n) + h(X_{n-1}|X_1, X_2, \ldots, X_{n-2})
\]

\[
\vdots
\]

\[
h(X_1, X_2, \ldots, X_n) \leq h(X_2, X_3, \ldots, X_n) + h(X_1),
\]  \hspace{1cm} (7.6)

where the inequalities follow from the fact that conditioning reduces entropy. Adding these \( n \) inequalities and using the chain rule, we obtain

\[
n \cdot h(X_1, X_2, \ldots, X_n) \leq \sum_{i=1}^n h(X_1, X_2, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n) + h(X_1, X_2, \ldots, X_n)
\]  \hspace{1cm} (7.7)

\[113\]
or
\[
\frac{1}{n} h(X_1, X_2, \ldots, X_n) \leq \frac{1}{n} \sum_{i=1}^{n} h(X_1, X_2, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n) \frac{1}{n - 1}
\]
(7.8)

which is the desired result \( h^{(n)}_n \leq h^{(n)}_{n-1} \).

We now prove that \( h^{(n)}_k \leq h^{(n)}_{k-1} \) for all \( k \leq n \), by first conditioning on a \( k \)-element subset, then taking a uniform choice over its \((k-1)\)-element subsets. For each \( k \)-element subset, \( h^{(k)}_k \leq h^{(k)}_{k-1} \), and hence the inequality remains true after taking the expectation over all \( k \)-element subsets chosen uniformly from the \( n \) elements.

\( \Box \)

**Corollary:** Let \( r > 0 \), and define
\[
e^{(n)}_k = \frac{1}{\binom{n}{k}} \sum_{S: |S| = k} e^{r h(X^{(S)})}.
\]
(7.9)

Then
\[
e^{(n)}_1 \geq e^{(n)}_2 \geq \cdots \geq e^{(n)}_n.
\]
(7.10)

**Proof:** Starting from (7.8) in the proof of Theorem 7.2.1, we multiply both sides by \( r \), exponentiate, and then apply the arithmetic mean geometric mean inequality to obtain
\[
e^{1/n} r h(X_1, X_2, \ldots, X_n) \leq \left( \frac{1}{n} \sum_{i=1}^{n} e^{r h(X_1, X_2, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)} \right)^{1/n - 1} \leq \frac{1}{n} \sum_{i=1}^{n} e^{r h(X_1, X_2, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)}, \text{ for all } r \geq 0,
\]

which is equivalent to \( e^{(n)}_n \leq e^{(n)}_{n-1} \). Now we use the same arguments as in Theorem 7.2.1, taking an average over all subsets to prove the result that for all \( k \leq n \), \( e^{(n)}_k \leq e^{(n)}_{k-1} \). \( \Box \)

This theorem implies Szasz’s theorem[61], which is a generalization of Hadamard’s inequality for positive definite matrices, as will be shown later.

We can use similar techniques to prove a conditional version of this subset inequality. Surprisingly, the conditioning causes this sensitive inequality to
go the other way. Perhaps we should call this inequality Zsazs's theorem. We will use the inequality to prove a new matrix inequality similar to Szasz's theorem.

**Definition:** The conditional entropy rate per element for a \( k \) element subset \( S \) is \( h(X^{(S)}|X^{(S^c)})/k \).

The average conditional entropy rate per element for all subsets of size \( k \) is the average of the above quantities for \( k \)-element subsets of \( \{1, 2, \ldots, n\} \); i.e.,

\[
g_k^{(n)} = \frac{1}{\binom{n}{k}} \sum_{S:|S|=k} \frac{h(X^{(S)}|X^{(S^c)})}{k}
\]  

(7.11)

Here \( g_k(S) \) is the entropy per element of the set \( S \) conditional on the elements of the set \( S^c \). When the size of the set \( S \) increases, one could expect a greater dependence between the elements of the set \( S \), and expect a decrease in the entropy per element. This explains Theorem 7.2.1.

In the case of the conditional entropy per element, as \( k \) increases, the size of the conditioning set \( S^c \) decreases and the entropy of the set \( S \) increases since conditioning reduces entropy. In the conditional case, the increase in entropy per element due to the decrease in conditioning dominates the decrease due to additional dependence between the elements and hence we have the following theorem (also implicit in Han[42]):

**Theorem 7.2.2**

\[
g_1^{(n)} \leq g_2^{(n)} \leq \cdots \leq g_n^{(n)}.
\]  

(7.12)

**Proof:** The proof proceeds on lines very similar to the proof of the theorem for the unconditional entropy per element for a random subset. We will first prove that \( g_n^{(n)} \geq g_{n-1}^{(n)} \), and then use this to prove the rest of the inequalities.

By the chain rule, the entropy of a collection of random variables is less than the sum of the entropies, i.e.,

\[
h(X_1, X_2, \ldots, X_n) \leq \sum_{i=1}^{n} h(X_i).
\]  

(7.13)
Subtracting both sides of this inequality from \( nh(X_1, X_2, \ldots, X_n) \), we have

\[
(n - 1)h(X_1, X_2, \ldots, X_n) \geq \sum_{i=1}^{n} \left( h(X_1, X_2, \ldots, X_n) - h(X_i) \right)
\]

\[
= \sum_{i=1}^{n} h(X_1, X_2, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n | X_i).
\]

Dividing this by \( n(n - 1) \), we obtain

\[
\frac{h(X_1, X_2, \ldots, X_n)}{n} \geq \frac{1}{n} \sum_{i=1}^{n} \frac{h(X_1, X_2, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n | X_i)}{n - 1},
\]

which is equivalent to \( g_n^{(n)} \geq g_{n-1}^{(n)} \).

We now prove that \( g_k^{(n)} \geq g_{k-1}^{(n)} \) for all \( k \leq n \), by first conditioning on a \( k \)-element subset, then taking a uniform choice over its \((k - 1)\)-element subsets. For each \( k \)-element subset, \( g_k^{(k)} \geq g_{k-1}^{(k)} \), and hence the inequality remains true after taking the expectation over all \( k \)-element subsets chosen uniformly from the \( n \) elements.

\[ \square \]

### 7.3 Subset inequalities for mutual information

In this section, we will prove some inequalities for mutual information between subsets. The first one is a direct consequence of the previous two theorems:

**Theorem 7.3.1** Let

\[
f_k^{(n)} = \frac{1}{\binom{n}{k}} \sum_{S:|S|=k} \frac{I(X^{(S)}; X^{(S^c)})}{k}.
\]

Then

\[
f_1^{(n)} \geq f_2^{(n)} \geq \cdots \geq f_n^{(n)}.
\]
**Proof:** The theorem follows from the identity \( I(X^{(S)}; X^{(S^c)}) = \h(X^{(S)}) - \h(X^{(S)}|X^{(S^c)}) \) and Theorems 7.2.1 and 7.2.2. □

We now prove an inequality for the average mutual information between a subset and its complement, averaged over all subsets of size \( k \) in a set of random variables, without normalization by \( k \). This inequality will be used to prove yet another determinant inequality along the lines of Szasz's theorem; however, unlike the inequalities in the previous section, there is no normalization by the number of elements in the subset.

Let
\[
i_k^{(n)} = \frac{1}{\binom{n}{k}} \sum_{S:|S|=k} I(X^{(S)}; X^{(S^c)})
\]
(7.18)

be the average mutual information between a subset and its complement averaged over all subsets of size \( k \). By the symmetry of mutual information and the definition of \( i_k^{(n)} \), it is clear that \( i_k^{(n)} = i_{n-k}^{(n)} \).

**Theorem 7.3.2**

\[
i_1^{(n)} \leq i_2^{(n)} \leq \cdots \leq i_{\lfloor \frac{n}{2} \rfloor}^{(n)}
\]
(7.19)

**Proof:** Let \( k \leq \lfloor \frac{n}{2} \rfloor \). Consider a particular subset \( S \) of size \( k \). \( S \) has \( k \) subsets of size \( k - 1 \). Let \( S_j \) denote the subset \( S - j \). Then

\[
kI(X^{(S)}; X^{(S^c)}) - \sum_{j \in S} I(X^{(S_j)}; X^{(S_j^c)})
\]

\[
= \sum_{j \in S} I(X^{(S_j)}, X_j; X^{(S^c)}) - I(X^{(S_j)}; X^{(S^c)}), X_j)
\]

\[
= \sum_{j \in S} I(X^{(S_j)}; X^{(S^c)}) + I(X_j; X^{(S^c)}|X^{(S_j)}) - I(X^{(S_j)}; X^{(S^c)})
\]

\[
- I(X^{(S^c)}; X_j|X_j)
\]

\[
= \sum_{j \in S} h(X_j|X^{(S_j)}) - h(X_j|X^{(S_j)}, X^{(S^c)}) - h(X_j|X^{(S^c)})
\]

\[
+ h(X_j|X^{(S^c)}), X^{(S_j)}
\]

\[
= \sum_{j \in S} h(X_j|X^{(S^c)}) - h(X_j|X^{(S^c)}).
\]
(7.22)
Summing this over all subsets of size \( k \), we obtain
\[
\sum_{S:|S|=k} \left[ kI(X^{(S)}; X^{(S^c)}) - \sum_{j \in S} I(X^{(S_j)}; X^{(S^c)}) \right] = \sum_{S:|S|=k} \sum_{j \in S} h(X_j | X^{(S_j)}) - h(X_j | X^{(S^c)}). \tag{7.23}
\]
Reversing the order of summation, we obtain
\[
\sum_{S:|S|=k} \left[ kI(X^{(S)}; X^{(S^c)}) - \sum_{j \in S} I(X^{(S_j)}; X^{(S^c)}) \right] = \sum_{j=1}^{n} \sum_{S:|S|=k, S \ni j} h(X_j | X^{(S_j)}) - h(X_j | X^{(S^c)}) \tag{7.24}
\]
\[
= \sum_{j=1}^{n} \sum_{S':|S'|=k-1, S' \ni j} h(X_j | X^{(S')}) - h(X_j | X^{(S)^c \cup (S')}) \tag{7.25}
\]
\[
= \sum_{j=1}^{n} \left[ \sum_{S':|S'|=k-1} h(X_j | X^{(S')}) \right. \\
- \left. \sum_{S'':S' \subset \{j\}^c, |S''|=n-k} h(X_j | X^{(S'')}) \right]. \tag{7.26}
\]
Since \( k \leq \lfloor \frac{n}{2} \rfloor \), \( k-1 < n-k \). So we would expect that the second sum in (7.26) to be less than the first sum, since both sums have the same number of terms but the second sum corresponds to entropies with more conditioning. We will prove this by using a simple symmetry argument.

The set \( S'' \) with \( n - k \) elements has \( \binom{n-k}{k-1} \) subsets of size \( k-1 \). For each such subset \( S' \) of size \( k-1 \), we have
\[
h(X_j | X^{(S'')}) \leq h(X_j | X^{(S')}) \tag{7.27}
\]
since conditioning reduces entropy. Since (7.27) is true for each subset \( S' \subset S'' \), it is true of the average over subsets. Hence
\[
h(X_j | X^{(S'')}) \leq \frac{1}{\binom{n-k}{k-1}} \sum_{S':|S'|=k-1, S' \subset S''} h(X_j | X^{(S')}). \tag{7.28}
\]

118
Summing (7.28) over all subsets $S''$ of size $n - k$, we get
\[
\sum_{S'' : |S''| = n - k} h(X_j | X^{(S'')}) \leq \sum_{S'' : |S''| = n - k} \frac{1}{\binom{n-k}{k-1}} \sum_{S' : S' \subseteq S'' , |S'| = k-1} h(X_j | X^{(S')})
\]
\[
= \sum_{S' : |S'| = k-1} h(X_j | X^{(S')}), \tag{7.29}
\]

since by symmetry, each subset $S'$ occurs in $\binom{n-k}{n-2k+1} = \binom{n-k}{k-1}$ sets $S''$.

Combining (7.26) and (7.29), we get
\[
\sum_{S : |S| = k} \left[ k I(X^{(S)} ; X^{(S')}) - \sum_{j \in S} I(X^{(S)} ; X^{(S_j)}) \right] \geq 0. \tag{7.30}
\]

Since each set of size $k-1$ occurs $n - k + 1$ times in the second sum, we have
\[
\sum_{S : |S| = k} k I(X^{(S)} ; X^{(S')}) \geq \sum_{S : |S| = k} \sum_{j \in S} I(X^{(S)} ; X^{(S_j)}) \tag{7.31}
\]
\[
= (n - k + 1) \sum_{S' : |S'| = k-1} I(X^{(S)} ; X^{(S')}).
\]

Dividing this equation by $k \binom{n}{k}$, we have the theorem
\[
i_k^{(n)} = \frac{1}{\binom{n}{k}} \sum_{S : |S| = k} I(X^{(S)} ; X^{(S')}) \geq \frac{1}{\binom{n}{k-1}} \sum_{S' : |S'| = k-1} I(X^{(S)} ; X^{(S')}) = i_{k-1}^{(n)}. \tag{7.32}
\]

\[\Box\]

### 7.4 Subset inequalities and information flow

We now apply some of these subset inequalities to prove some simple properties of information flow in networks.
7.4.1 Multiple access channels

1. Multiple access channels without feedback.

By Theorem 3.1.1, the rates of the senders for a \( m \) user multiple access channel without feedback are bounded by

\[
\sum_{k \in S} R_k \leq I(X^{(S)}; Y | X^{(Sc)}, Q) \tag{7.33}
\]

for a product distribution \( p(q)p(x_1|q)p(x_2|q)\ldots p(x_m|q)p(y|x_1,\ldots,x_m) \) on the inputs. Since by the chain rule for entropy,

\[
I(X^{(S+j)}; Y | X^{(Sc-j)}, Q) = I(X_j; Y | X^{(Sc-j)}, Q) + I(X^{(S)}; Y | X^{(Sc)}, Q), \tag{7.34}
\]

we have

\[
I(X^{(S+j)}; Y | X^{(Sc-j)}, Q) \geq I(X^{(S)}; Y | X^{(Sc)}, Q). \tag{7.35}
\]

If we divide the sum of the rates of senders in a set \( S \) of size \( k \) by the number of elements in the set \( S \) (for a bound on the average rate in bits per user), we obtain

\[
\mathcal{R}(S) \triangleq \frac{1}{k} \sum_{i \in S} R_i \tag{7.36}
\]

\[
\leq \frac{1}{k} I(X^{(S)}; Y | X^{(Sc)}, Q) \tag{7.37}
\]

\[
= \frac{1}{k} h(X^{(S)}|X^{(Sc)}, Q) - \frac{1}{k} h(X^{(S)}|X^{(Sc)}, Y, Q) \tag{7.38}
\]

\[
= \frac{1}{k} h(X^{(S)}|Q) - \frac{1}{k} h(X^{(S)}|X^{(Sc)}, Y, Q), \tag{7.39}
\]

where the last equality follows from the fact that the \( X \)'s are independent.

We will define the average of \( \mathcal{R}(S) \) of all sets of size \( k \) to be \( \mathcal{R}_k \).

By symmetry, each sender appears in the same number of sets of size \( k \), and hence each rate makes an equal contribution to \( \mathcal{R}_k \). Therefore it is not difficult to see that

\[
\mathcal{R}_k = \frac{1}{m} \sum_{i=1}^{m} R_i \triangleq \mathcal{R}. \tag{7.40}
\]

120
Using different set sizes give different bounds on $\bar{R}$. If we take the average over all sets of size $k$ in (7.39), we obtain

$$
\bar{R} \leq \frac{1}{k} \left( \frac{m}{k} \right) \sum_{S: |S|=k} h(X^{(S)}|Q) - \frac{1}{k} \left( \frac{m}{k} \right) \sum_{S: |S|=k} h(X^{(S)}|X^{(S^c)}, Y, Q). \quad (7.41)
$$

Since the $X$'s are conditionally independent, $h(X^{(S)}|Q) = \sum_{i \in S} h(X_i|Q)$, and hence the first term in (7.41) by symmetry is equal to $\frac{1}{m} \sum_{i=1}^{m} h(X_i|Q)$. Thus the first term in (7.41) does not depend on $k$. Applying a conditional version of Theorem 7.2.2 to the second term, we see that the second term is maximized when $k = m$. Thus the strongest bound on $\bar{R}$ is obtained for $k = m$. This bound is

$$
\bar{R} \leq \frac{1}{m} I(X_1, \ldots, X_m; Y|Q). \quad (7.42)
$$

2. Multiple access channel with feedback.

With feedback, the bounds allow for an arbitrary joint distribution on the inputs $X_1, X_2, \ldots, X_m$. In this case, (7.35) still holds. But the bounds on the average rate $\bar{R}_k$ are not ordered with respect to $k$. An example is given in Figure 3.5 for the Gaussian multiple access channel. The bounds here correspond to the bounds for different values of $k$, the size of the subset. As can be seen in the figure, different bounds are strongest for different values of $\rho$. Thus which is the most dominant bound depends on the joint distribution of the inputs, and we cannot identify the strongest bound as in the case of the multiple access channel without feedback.

### 7.4.2 General networks of channels

1. Deterministic networks, no feedback, no common information.

The bounds on the flow of information in a deterministic network without feedback and without common information are given by Theorem 6.3.2 as

$$
\sum_{i \in S, j \in T} R^{(ij)} \leq I(X^{(S)}; Y^{(T)}|X^{(S^c)}, Q) \quad (7.43)
$$

121
for some joint distribution \( p(q) \prod_{i=1}^n p(x^{(i)}|q)p(y^{(1)}, y^{(2)}, \ldots, y^{(m)}|x^{(1)}, x^{(2)}, \ldots, x^{(m)}) \).

For a deterministic network, \( H(Y^{(T)}|X^{(S)}, X^{(S^c)}) = 0 \) and hence the bounds become

\[
\sum_{i \in S, i \in T} R^{(i)} \leq H(Y^{(T)}|X^{(S^c)}, Q). \tag{7.44}
\]

For a given set \( S = \{i\} \), let the set \( T \) be of size \( k \). Then the bound on the average of the rates out of node \( i \) is \( \frac{1}{k} H(Y^{(T)}|X^{(S^c)}) \). This is minimized when \( T \) is of size \( m - 1 \) by Theorem 7.2.1.

Similarly for a fixed \( T = \{l\} \), let the size of \( S \) be \( k \). Then the average of the rates going into node \( l \) is bounded by \( \frac{1}{k} H(Y^{(T)}|X^{(S^c)}) \). Since \( X^{(S)} \) and \( X^{(S^c)} \) are independent,

\[
\frac{1}{k} H(Y^{(T)}|X^{(S^c)}) = \frac{1}{k} H(Y^{(T)}|X^{(S^c)}) - \frac{1}{k} H(X^{(S)}|X^{(S^c)}) + \frac{1}{k} H(X^{(S)}) \tag{7.45}
\]

\[
= \frac{1}{k} H(Y^{(T)}|X^{(S^c)}) - \frac{1}{k} H(X^{(S)}, Y^{(T)}|X^{(S^c)}) + \frac{1}{k} H(X^{(S)}) \tag{7.46}
\]

As in the case of the multiple access channel without feedback, the first term is maximized when the set \( S \) is of maximum size. The average of the second term over all sets \( S \) of size \( k \) is equal to \( \frac{1}{n} \sum_{i=1}^n H(X_i) \), and does not depend on \( k \). Thus the strongest bounds on the average rate of flow into a node is obtained when the size of the set \( S \) is \( m - 1 \).

2. Information flow across a boundary.

By Theorem 6.5.1, the rate of flow of information across any cut-set in the network is bounded by

\[
R(S, S^c) \leq I(X^{(S)}; Y^{(S^c)}|X^{(S^c)}) \tag{7.47}
\]

for some joint distribution on the inputs of the network. As the example of the multiple access channel indicates, it is not possible to prove a general subset inequality for the quantity on the right hand side of (7.47), we can prove a few simple inequalities for special cases.
3. Flow in both directions across boundary.

We can bound the sum of the flows in both directions across a cutset

\[ R(S, S^c) + R(S^c, S) \leq I(X^{(S)}; Y^{(S^c)} | X^{(S^c)}) + I(X^{(S^c)}; Y^{(S)} | X^{(S)}). \]

(7.48)

We can rewrite the right hand side of (7.48) by using the chain rule for mutual information to obtain

\[
I(X^{(S)}; Y^{(S^c)} | X^{(S^c)}) + I(X^{(S^c)}; Y^{(S)} | X^{(S)}) \\
= I(X^{(S)}; Y^{(S)}; X^{(S^c)}; Y^{(S^c)}) - I(X^{(S)}; X^{(S^c)}) - I(Y^{(S)}; Y^{(S^c)} | X^{(S)}, X^{(S^c)}).
\]

For a deterministic network, the last term is zero. If the senders are independent (as in case of the network without feedback and no common information), the penultimate term is also zero. The average of the first term on the right hand side of the equation is maximized when the set size is about half the number of nodes (Theorem 7.3.2). Thus in this case the biggest bounds on the rate of flow in both directions across a boundary occurs when the boundary divides the set into about two equal halves.

### 7.5 Consequences for Determinant Inequalities

The theorems in the previous section result in inequalities for determinants if we let \( X_1, X_2, \ldots, X_n \) be jointly normal with a covariance matrix \( K \). Let \( K(i_1, i_2, \ldots, i_k) \) be the \( k \)-rowed principal submatrix of \( K \) formed by the rows and columns with indices \( i_1, i_2, \ldots, i_k \).

**Theorem 7.5.1 (Szasz)** If \( K \) is a positive definite \( n \times n \) matrix and \( P_k \) denotes the product of all the principal \( k \)-rowed minors of \( K \), i.e.,

\[
P_k = \left( \prod_{1 \leq i_1 < i_2 < \ldots < i_k \leq n} |K(i_1, i_2, \ldots, i_k)| \right)^{\frac{1}{k!}}, \tag{7.49}
\]

then

\[
P_1 \geq P_2 \geq P_3 \geq \cdots \geq P_n \tag{7.50}
\]
Proof: Follows directly from Theorem 7.2.1 with the substitution of $X_1, X_2, \ldots, X_n \sim \mathcal{N}(0, K)$. □

Define the geometric mean of $(|K|/|K(S^c)|)^{1/k}$ over $k$-element subsets by

$$Q_k = \left( \prod_{\mathcal{S}:|\mathcal{S}|=k} \frac{|K|}{|K(S^c)|} \right)^{1/k} \tag{7.51}$$

**Theorem 7.5.2**

$$\left( \prod_{i=1}^{n} \sigma_i^2 \right)^{\frac{1}{n}} = Q_1 \leq Q_2 \leq \cdots \leq Q_{n-1} \leq Q_n = |K|^{1/n}. \tag{7.52}$$

**Proof:** The theorem follows immediately from Theorem 7.2.2 and the identification

$$h(X(S)|X(S^c)) = \frac{1}{2} \log(2\pi e)^k \frac{|K|}{|K(S^c)|}. \tag{7.53}$$

□

The outermost inequality, $Q_1 \leq Q_n$ can be rewritten as

$$|K| \geq \prod_{i=1}^{n} \sigma_i^2, \tag{7.54}$$

where

$$\sigma_i^2 = \frac{|K|}{|K(1, 2, \ldots, i-1, i+1, \ldots, n)|} \tag{7.55}$$

is the minimum mean squared error in the linear prediction of $X_i$ from the remaining $X$'s. It is the conditional variance of $X_i$ given the remaining $X_j$'s if $X_1, X_2, \ldots, X_n$ is jointly normal. Combining this with Hadamard's inequality gives upper and lower bounds on the determinant of a positive definite matrix:

**Corollary:**

$$\prod_i K_{ii} \geq |K| \geq \prod_i \sigma_i^2. \tag{7.56}$$
Hence the determinant of a covariance matrix lies between the product of the unconditional variances $K_{ii}$ of the random variables $X_i$ and the product of the conditional variances $\sigma_i^2$.

Let

$$R_k = \left( \prod_{S:\#S=k} \frac{|K(S)||K(S^c)|}{|K|} \right)^{\frac{1}{(2)}}$$

(7.57)

**Theorem 7.5.3**

$$R_1 \geq R_2 \geq \cdots \geq R_{n-1} \geq R_n.$$  

(7.58)

**Proof:** The theorem follows immediately from Theorem 7.3.1 and the identity

$$I(X^{(S)}; X^{(S^c)}) = \frac{1}{2} \log \frac{|K(S)||K(S^c)|}{|K|}.$$  

(7.59)

\[ \square \]

In particular, the outer inequality $R_1 \geq R_n$ results in

$$\left( \prod_{i=1}^{n} \frac{|K_{ii}||K(i^c)|}{|K_i|} \right)^{\frac{1}{n}} \geq 1.$$  

(7.60)

We can also convert Theorem 7.3.2 into a determinant inequality by the same procedure.

Let

$$T_k = \left( \prod_{S:\#S=k} \frac{|K(S)||K(S^c)|}{|K|} \right)^{\frac{1}{(2)}}$$

(7.61)

**Theorem 7.5.4**

$$T_1 \leq T_2 \leq \cdots \leq T_{\lceil \frac{n}{2} \rceil}.$$  

(7.62)

**Proof:** The theorem follows directly from the previous theorem and the identity

$$I(X^{(S)}; X^{(S^c)}) = \frac{1}{2} \log \frac{|K(S)||K(S^c)|}{|K|}.$$  

(7.63)

\[ \square \]
7.6 Summary

The entropy per unit volume of a $k$-element set chosen at random decreases with $k$ because the dependence among the $\{X_i\}$ becomes unavoidable. The conditional entropy per unit volume of a set $S$ chosen at random, given its complement $S^c$, increases with $|S|$ because the dependence grows more slowly than the entropy increase resulting from the reduction of the amount of information in the conditioning set $S^c$.

It is interesting to note that the above theorems are true only on the average over all subsets, and are not true for fixed subsets. For example, even if $S \subset S'$, it does not follow that

$$\frac{h(X^{(S)})}{|S|} \geq \frac{h(X^{(S')})}{|S'|}. \quad (7.64)$$

However the average of the left hand side over all subsets $S$ of size $k$ is greater than the average of the right hand side over all subsets of size $l$ if $k < l$.

Also since the essential inequality in all the proofs was the fact that conditioning reduces entropy, we have equality in the theorems if and only if conditioning does not change the entropy, i.e., if and only if the random variables are independent, i.e., if and only if the covariance matrices are diagonal.

Some of the subset inequalities in this chapter were published earlier in [15].
Chapter 8

Other roles of feedback

In this chapter, we will describe some of the other uses of feedback in communication. As mentioned in the introduction, feedback helps in many ways even when it does not increase the capacity of the channel. We also describe the role of feedback in communication situations where errors are not allowed (zero-error capacity).

8.1 Feedback simplifies encoding and decoding

Intuitively, it appears that the ability to be able to correct a receiver’s misunderstanding of the message sent should help improve the rate of communication. But as we showed in Chapter 2, the maximum rate of communication is not increased for single user memoryless channels. However, feedback can greatly simplify the encoding and decoding procedures that need to be used to achieve capacity. For example, consider the binary erasure channel shown in Figure 8.1.

If feedback is available for the binary erasure channel, it is very clear what to do: if a bit is lost, retransmit it until it gets through. Since the bits get through with probability of $1 - \alpha$, and each time a bit gets through, we succeed in sending one bit, the effective rate of transmission is $1 - \alpha$. In this way we are easily able to achieve a capacity of $1 - \alpha$ with feedback.

Without feedback, the capacity is also $1 - \alpha$ bits per transmission. But it requires complex codes with long block lengths to achieve this capacity.
Figure 8.1: Binary erasure channel
Thus we see that knowing the mistakes made by the channel simplifies our communication procedure, even though it does not increase the maximum achievable rate.

Another example of coding simplification is the Kailath-Schalkwijk scheme for the Gaussian channel with feedback. In this scheme, the sender attempts to send a real number to the receiver. Due to the noise in the channel, the receiver sees a corrupted version of this number. The receiver estimates the number sent. Because of the feedback link, the sender can also produce the same estimate, and calculate the error. In the next transmission, the sender transmits a scaled version of the error, which the receiver uses to update his estimate of the message. Kailath and Schalkwijk[49] have shown that this scheme not only achieves capacity but also has a doubly exponential decay of the probability of error with block length, so that it is possible to achieve very low probabilities of error with small block lengths.

Although it is possible to design intricate coding schemes for channels with feedback, in most practical channels, feedback is used with just a simple ARQ (automatic request) scheme, in which the sender just retransmits the messages that are received in error. ARQ error control underlies many protocols for computer networks[7]. Even though more complex coding schemes may be more efficient, they are often not worth additional cost in processing at each node.

For networks, feedback sometime allows for particularly simple code constructions that use some of the properties of single user feedback codes. For example, it is possible to use the feedback erasure codes for a binary erasure channel with $\alpha = 1/2$ to decode one of the senders in the binary adder multiple access channel. Such simple code constructions do not in general make use of the possible increase in capacity due to feedback.

### 8.2 Feedback enables the sender to determine the characteristics of the channel.

Some communication channels have parameters that may not be known at the sender. And as the example in Section 2.3 indicates, knowledge of the state of the channel may greatly improve the rate of the communication. Feedback can be used to determine the state of the channel and thus improve the rate
of communication.

A practical example of this is the adaptive equalization schemes used for digital communication over phone lines. The characteristics of the noise and the frequency response of a phone line are not constant. The equalizer uses the feedback to determine these characteristics and correct for them. It is possible to achieve much higher rates with equalization than without equalization.

8.3 Feedback is used in simple protocols that allow a large number of users to share a common channel with low delay.

Feedback underlies communication protocols like ALOHA, which enable a large number of bursty users to share a common channel. We assume that the sender send information in packets, which fit to fixed length time slots. In the ALOHA scheme, each sender transmits whenever it has a message available. If there is more than one user transmitting at a time, there is a collision and the messages are lost. Because of the feedback link, the senders know when this happens. They each choose a random delay and then retransmit after that delay. Abramson[1] showed that it is possible to transmit a maximum of \(1/e = 0.37\) packets/slot with this scheme. The scheme allows an arbitrary number of users to share a common channel with a low delay. Variations of this scheme are used in computer networks like Ethernet.

Massey and Mathys[60] have analyzed this channel without feedback and have shown that for a fixed number of users, it is possible to achieve a throughput of \(1/e\) packets/slot without feedback.

There is a substantial field of work analyzing schemes like the ALOHA channel. There are different factors of interest in such schemes: capacity or throughput, average delay, stability, robustness, etc. There are many different schemes that allow a number of users to share the same channel; some of them are described the book by Bertsekas and Gallager[7]. A survey of recent result on the capacity of random access systems can be found in Massey[59].
8.4 Feedback can help increase the zero error capacity, even for single user memoryless channels

Shannon[73] considered the problem of sending information over a communication channel if no errors were allowed at all (as opposed to an asymptotically low probability of error). He showed that in this case, the capacity was the normalized chromatic number of the n-th extension of the graph induced by the channel. He solved for the capacity of simple channels, but the smallest non-trivial example, the pentagon channel, remained open until it was settled by Lovasz[55]. A computable expression for the zero error capacity of a general single user communication channel is not known.

In [73], Shannon also considered the problem of determining the zero error capacity of the single user channel with feedback. In this case, he proved the following theorem:

**Theorem 8.4.1** In a discrete memoryless channel with feedback, the zero error capacity with feedback with finite length block codes $C_{of}$ is zero if all pairs of input letters are adjacent (they can both cause a common output with positive probability. Else,

$$C_{of} = \log \frac{1}{P_0}$$

(8.1)

where

$$P_0 = \min_{P_i} \max_{j \in S_j} \sum_{i \in S_j} p_i$$

(8.2)

where $S_j$ is the set of possible inputs that could have caused output $j$ with positive probability and $p_i$ is a probability assignment to the input of the channel.

**Proof:** See [73]. □

Blackwell[8] showed that if an infinite coding delay is allowed, then zero error transmission is possible with feedback at any rate up to capacity, provided there exists at least one input-output pair $(x, y)$, such that $p(y|x) = 0$.

The zero-error capacity of a channel depends only on which elements of the probability transition matrix of the channel are non-zero, not on the exact values of the transition probabilities.
Figure 8.2: Example for zero error capacity

The results of Shannon indicate that feedback can help by an arbitrary amount to increase zero-error capacity. For example, consider the channel shown in Figure 8.2. Let the channel transition matrix be

$$p(y|x) = \begin{bmatrix}
\frac{1}{2} & \frac{1}{2} & 0 \\
0 & \frac{1}{2} & \frac{1}{2} \\
\frac{1}{2} & 0 & \frac{1}{2}
\end{bmatrix}$$ \hspace{1cm} (8.3)

The zero-error capacity of this channel without feedback is 0, since there are no two symbols that are not confusable. The zero-error capacity of the channel with feedback is $\log 3/2$ bits, which can be seen from the above theorem using a uniform input distribution. Note that $\log 3/2$ bits is also the capacity ($C = \max I(X;Y)$) of the channel with the above transition matrix. Thus feedback can increase the zero-error capacity from 0 to the capacity $C = \max I(X;Y)$ of the channel.

There has been very little work on the extending these results for a multi-user network. Dueck\cite{30} has determined the zero error capacity region for a class of multiple access channels with feedback but the problem of determining the capacity region for a general multiple access channel remains open.
8.5 Feedback in discrete communication

Recently, Orlitsky[62] considered the problem of sending information to a receiver when the receiver has some correlated side information. He analyzed the worst case performance when no errors are allowed and showed that in this case, allowing the receiver to send some information back to the sender can help, and in fact could reduce the total number of bits communicated by an exponential factor. Each transmission in either direction is considered as a message and the objective is to minimize the total number of bits transmitted. He showed that allowing more than two messages will reduce the total number of bits transmitted by at most a factor of 4 compared to the two message case, which could be as low as the logarithm of the number of bits required when only one message is allowed.

Example 8.5.1: To illustrate some of ideas when we have feedback in a communication channel with side information at the receiver, we will use an example due to Orlitsky. Suppose we have a league of \( n \) teams and two of the teams play each other. Suppose that night, Alice heard on the radio team A played team B and before Alice could hear the result, Bob snatched the radio from Alice and he heard the name of the winning team. The question is “How many bits must Bob send to Alice to tell Alice the identity of the winning team”?

If only Bob is allowed to communicate, we can see that Bob needs to send at least \( \log n \) bits (the identity of the winning team). If on the other hand, Alice is allowed to send Bob some information first, we can reduce the total number of bits needed. A protocol that uses \( \log \log n + 1 \) bits will be described below:

Alice and Bob agree on a numbering scheme so that each team is allotted a number of \( \log n \) bits. Alice sends Bob the index of the first place in which the identities of the two teams A and B differ (this would need \( \log \log n \) bits. Bob then sends the corresponding bit of the winning team.

So we have a protocol for two way communication that reduces the total number of bits sent from \( \log n \) to \( \log \log n + 1 \). The
results of Orlitsky[62] show that this is the maximum reduction in the number of bits transmitted due to feedback.

8.6 Summary

Feedback has many roles to play in communication. In this report, we have concentrated our attention on the role of feedback in increasing the capacity of communication networks. In this chapter, we examined some of the other uses of feedback. We showed how feedback could greatly reduce the complexity of encoding and decoding even when it does not increase the capacity. It also enables the sender to find out the characteristics of the channel and it allows a number of users to share a channel with a simple protocol. It may increase the zero-error capacity of a channel and could reduce the communication complexity (the total number of bits transmitted) by a large factor when the receiver has some side information.

These are only a few of the many uses of feedback in communication. The role of feedback is still not perfectly understood and there are still a number of open issues concerning the use of feedback in communication.
Chapter 9

Summary

In this report, we have considered the role of feedback in increasing the capacity of multi-user networks. We proved upper bounds on the capacity regions of memoryless networks with feedback and showed that in some simple cases, feedback does not help very much. In general, feedback may open up a new communication path that is much more efficient than the communication paths available without feedback, and thus can increase the capacity by an arbitrarily large factor.

The subject of multi-user or network information theory has a number of basic open problems. The capacity regions of some simple networks like the general broadcast channel and the interference channel are still unknown. We have tried to characterize the role of feedback in a network and its relationship to the capacity of the network (the maximum rates of information transfer possible in the network). For a multiple access channel, feedback converts the independent information from the senders to common information, which the sender can then send cooperatively to the receiver. For a broadcast channel, feedback converts independent information destined for the individual receivers to common information, which can be sent efficiently to all the receivers at once. For a general network, both these properties of feedback can be used to increase the capacity region.

For both a multiple access and a broadcast channel, feedback does not increase the maximum individual capacities and can increase the total capacity by at most a factor of \( m \). For the Gaussian multiple access channel, feedback increases the total capacity by at most a factor of two. For the interference channel, feedback could increase capacity by an arbitrary factor.
We set up the problem of communication in a general network, and proved upper bounds on the rate of information flow in such networks. These upper bounds were stated in terms of entropies and mutual informations of subsets. This prompted us to study more carefully the properties of the entropy rates of subsets and prove some new theorems regarding these quantities. These theorems can be translated into theorems about the determinants of positive definite matrices, and we have proved some novel determinant inequalities by this method.

Finally, we surveyed some of the other roles of feedback in communication. We briefly described the use of feedback to simplify encoding, to allow simple protocols, to determine the properties of the channel, and to increase zero-error capacity. We also introduced the field of communication complexity, and described a problem where feedback helped a great deal in reducing the communication complexity.

There are still a number of open problems in this area. Instead of listing these problems, we identify a few basic issues that have to be solved. The relationship between the correlation of the inputs of the multiple access channel and the capacity of a multiple access channel is still poorly understood. This problem is closely linked with the problem of sending correlated information over a multiple access channel. We feel that an understanding of this problem will be crucial in understanding the role of feedback in multiple access channels. For broadcast channels, it is important to understand the difference between sending common information to the receivers to sending independent information. Again, the problem is related to the problem of sending correlated information over the broadcast channel, and this is not fully solved.

The role of feedback in networks is still not completely understood. It may be long time before we have a complete theory of information flow in networks, but there can be no doubt that such a theory would have important practical implications for the design of communication and computer networks.
Bibliography


138


139


