INTERFERENCE CHANNELS

BY

MAX HENRIQUE MACHADO COSTA

TECHNICAL REPORT NO. 47
FEBRUARY 1983

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION
GRANT ECS82-11568

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
INTERFERENCE CHANNELS

by

Max Henrique Machado Costa

TECHNICAL REPORT NO. 47
FEBRUARY 1983

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION
GRANT ECS82-11568

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
Interference Channels

Abstract

The interference channel, first considered by Shannon, models the communication between transmitters and receivers in which each transmitter wants to send information reliably to its intended receiver in the presence of interference from the other transmitters. The performance of this model is characterized by its capacity region, i.e., the set of rate points that can be simultaneously achieved with arbitrarily small probability of error. The capacity region of the general interference channel remains an open problem. This thesis presents results concerning particular cases of this problem. We divide the thesis into the following three parts:

1) Gaussian interference channels:

We describe the model and summarize the cases that have been solved by other researchers. Then we establish the optimality of two extreme points in the achievable region of the general Gaussian interference channel. Also we prove that the class of degraded Gaussian interference channels is equivalent to the class of Z (one-sided) interference channels. Finally, we show a curious example of a Z-interference channel with degraded message sets, in which the effects of interference can be completely eliminated.

2) A strengthened Entropy Power Inequality:

We prove a strengthened version of Shannon's Entropy Power Inequality for the case where one of the random vectors involved is Gaussian. In particular we show that if we add independent Gaussian noise to an arbitrary multivariate random variable, the entropy power of the resulting random variable is a concave function of the variance (power) of the added noise. This fact is precisely what allows us to establish the optimality of the new points in part 1). As a byproduct of this development we present
an analog of the isoperimetric inequality in the domain of entropy powers.

3) Discrete interference channels:

We establish the capacity region of two classes of discrete memoryless interference channels.
Contents

Abstract .............................................................................................................. 1

Contents ......................................................................................................... iii

List of Figures .................................................................................................. v

1. Introduction

1.1. Subject ....................................................................................................... 1
1.2. History ....................................................................................................... 1
1.3. Outline ....................................................................................................... 3

2. The Gaussian Interference Channel

2.1. Introduction ............................................................................................... 5
2.2. The Gaussian Z-Interference Channel ..................................................... 13
2.3. Proof of Theorem 2.1 .............................................................................. 17
2.4. Bounds on the Equivocation .................................................................... 23
2.5. A Model with Degraded Message Sets .................................................... 27

3. Strengthening the Entropy Power Inequality

3.1. Introduction ............................................................................................... 35
3.2. The Concavity of Entropy Power with Added Noise ............................... 37
3.3. Conjecture and Special Cases ................................................................. 50
3.4. Similarity with the Brunn Minkowski Inequality .................................... 54
4. On Discrete Interference Channels

4.1. Introduction ........................................................................................................ 59

4.2. A Class of Deterministic Interference Channels ............................................. 60

4.3. Extension ............................................................................................................ 66

4.4. Discussion ......................................................................................................... 69

Appendix 1: Equivalent channels ........................................................................... 70

Appendix 2: From almost Gaussian to Gaussian .................................................... 73

Appendix 3: Bound on the derivative of entropy power ........................................ 78

Appendix 4: Vanishing surface integrals ................................................................ 79

References ................................................................................................................ 85
List of Figures

2.1. The Gaussian interference channel ........................................ 4
2.2. The complete model ................................................................... 7
2.3. Satellite communication ............................................................... 8
2.4. Best known achievable rate sum .................................................. 11
2.5. Typical Han and Kobayashi region ............................................... 12
2.6. The Gaussian Z-interference channel .......................................... 14
2.7. Equivalent channels ................................................................. 16
2.8. Model used in the proof .............................................................. 17
2.9. Relations between variables ....................................................... 19
2.10. Equivocation as function of added noise power .......................... 28
2.11. Bounds on equivocation ............................................................. 27
2.12. A model with degraded message sets ........................................... 28
2.13. A variation of the Gaussian Shannon channel ............................ 29
3.1. Typical growth of entropy power with added Gaussian noise .......... 38
4.1. Class of discrete interference channels under study ...................... 61
4.2. A class of Z-interference channels .............................................. 67
4.3. Example of interference channel in Fig. 4.2 ................................. 68
4.4. Capacity region of channel in Fig. 4.3 with $\epsilon = .1$ .................... 69
1. Introduction

1.1. Subject

The interference channel models the communication between \( M \) transmitter-receiver pairs in which the \( i \)'th transmitter wishes to send information reliably to the \( i \)'th receiver in the presence of interference from the other senders. It was first considered by Shannon [1] and has subsequently been investigated by Ahlswede, Carleial, Sato, Benzel, Han and Kobayashi among other researchers [2]-[13]. As in the previous studies, this dissertation is primarily concerned with establishing fundamental limits on the rates of information flow in interference channels. We shall restrict our attention to two transmitter-receiver pairs, for simplicity and because this case seems to capture the essence of the problem.

1.2. History

The fundamental bounds on the rates of information flow in a multiple user channel is characterized by its capacity region. The capacity region of the general memoryless interference channel is still unknown. Shannon (1961) was the first to consider the interference channel. In his pioneering paper on the two-way channel, Shannon proposed a variation of the two-way channel in which “the transmitted sequence at each terminal depends only on the message and not on the received sequence at that terminal”, or equivalently, where “the transmission and reception points at each end were at different places with no direct cross communication” [1, page 636]. Shannon proposed that his inner bound to the capacity region of the two-way channel was indeed the capacity region of this model, later termed an interference channel. This region is expressed by

\[
\mathcal{C} = \bigcup_{P_{1},P_{2}} \{(R_{1},R_{2}) : R_{1} \leq I(X_{1}, Y_{1} | X_{2}), R_{2} \leq I(X_{2}, Y_{2} | X_{1})\},
\]

where the union is over
all product input probability distributions, $I(\cdot;\cdot)$ denotes mutual information and $co A$ denotes the convex hull of the set $A$.

More than a decade later, Ahlswede presented the first specific study of interference channels [2]. By means of an instructive example Ahlswede showed that the region given by $co \bigcup_{P_1P_2} \{(R_1, R_2): R_1 \leq I(X_1, Y_1), R_2 \leq I(X_2, Y_2)\}$, where the union is over all product input probability distributions, is not optimal for the interference channel. The same example shows that the region Shannon conjectured to be optimal is not achievable in general.

In 1975, Carleial introduced the Gaussian interference channel with power constraints and showed that very strong interferences may be completely harmless [3]-[5]. Carleial also put forth an achievable region for the general interference channel. Until 1981 this region was the largest known achievable region for the interference channel. Another contribution of Carleial's work is the capacity region of interference channels with statistically equivalent outputs.

Two years later, Sato presented a detailed study of the interference channel and established an outer bound to its capacity region [6]. In the following year, Sato introduced an outer bound for the capacity region of the degraded Gaussian interference channel [8]. This outer bound touched Carleial's inner bound at one point, thus establishing its optimality. Sato and Tanabe, following up on Carleial's work, found the capacity region of discrete interference channels with very strong interference [7], [9].

Another important result was obtained by Benzel in 1979 [10]. He used a clever argument to establish the capacity region of a class of degraded additive interference channels.
More recently, an important result was contributed by Han and Kobayashi [11]. They put forth an achievable region that is in general strictly larger than Carleial’s region. This region is a strong candidate for the capacity region of the general interference channel. Another recent result, obtained by Han and Kobayashi [11] and Sato [12], is the capacity region of Gaussian interference channels with strong interference.

1.3. Outline

In the next chapter we investigate the Gaussian interference channel. We describe the model and review the cases that have been solved. Then we establish the optimality of two new points in the capacity region of the general Gaussian interference channel. Also, we show that degraded Gaussian interference channels are equivalent to Z (one-sided) interference channels in the sense that for every channel in one of these classes, we can find a channel in the other class having identical capacity region. To conclude chapter 2, we present a curious example with degraded message sets in which the effects of interference can be completely eliminated.

The main contribution of this dissertation lies in the new techniques developed to prove the optimality of the new rate points in chapter 2. These new tools are examined in chapter 3. The results of this chapter are of independent interest from the interference channel problem. The idea involves a strengthening of Shannon’s entropy power inequality in which one of the random variables is Gaussian. As a consequence of this stronger inequality, the entropy power is shown to grow as a concave function of added noise power. The result suggest a number of conjectures. In particular, we observe a strong similarity between the entropy power inequality and the Brunn Minkowski inequality in geometry. This similarity allows us to propose an analog of the isoperimetric inequality involving entropy and Fisher Information.
Finally, in chapter 4, we find the capacity region of a class of deterministic interference channels. This solution is extended to yield the capacity region of class of Z-interference channels.
2. The Gaussian Interference Channel

2.1. Introduction

In this chapter, we consider the continuous alphabet interference channel known as the Gaussian interference channel shown in Fig. 2.1.

![Diagram of the Gaussian interference channel]

Fig. 2.1. The Gaussian interference channel.

Its inputs $X_1$ and $X_2$ and its outputs $Y_1$ and $Y_2$ are real numbers related by

$$Y_1 = X_1 + b X_2 + Z_1$$

$$Y_2 = a X_1 + X_2 + Z_2$$

where $a$ and $b$ are given interference coefficients and $Z_1$ and $Z_2$ are Gaussian noise random variables with zero mean and unit variance. Also, the input signals are required to satisfy power constraints given by
The Gaussian interference Channel

\[ \frac{1}{n} \sum_{i=1}^{n} x_{(i)1}^2 \leq P_1 \]

\[ \frac{1}{n} \sum_{i=1}^{n} x_{(i)2}^2 \leq P_2 \]  \hspace{1cm} (2.2)

for all codewords \( x_1 = (x_{(1)1}, x_{(2)1}, \ldots, x_{(n)1}) \) and \( x_2 = (x_{(1)2}, x_{(2)2}, \ldots, x_{(n)2}) \) of any block length \( n \).

The noise terms \( Z_1 \) and \( Z_2 \) are independent of the input signals. Since there is no cooperation between the two receivers, the capacity region of this channel depends on the joint distribution \( p(y_1, y_2 | z_1, z_2) \) only through the two marginals \( p(y_1 | z_1, z_2) \) and \( p(y_2 | z_1, z_2) \). Therefore it is immaterial whether or not the noise terms \( Z_1 \) and \( Z_2 \) are independent of each other.

It may seem that we are losing generality by considering only those channels with transmission coefficients and unit noise powers as shown in Eqs. (2.1), but as Carleial showed [5], we can always apply a scaling transformation to a Gaussian interference channel with arbitrary transmission coefficients and noise powers and reduce it to an equivalent channel in this restricted class. A channel in this class is said to be in its standard form. A Gaussian interference channel in standard form is completely specified by its interference coefficients \( a \) and \( b \) and by its available powers \( P_1 \) and \( P_2 \).

This model is usually studied in its discrete time form. Results obtained for the discrete time version can be readily extended to its continuous time, band limited counterpart in much the same way as we can obtain the capacity of continuous time, band limited Gaussian Shannon channels from the capacity of discrete time Gaussian Shannon channels [14].

The complete model under investigation is shown in Fig 2.2. There are two
independent and uniformly distributed sources, one at sender 1 producing an integer $W_1 \in \mathcal{M}_1 = \{1, 2, \ldots, M_1\}$ and another at sender 2 producing an integer $W_2 \in \mathcal{M}_2 = \{1, 2, \ldots, M_2\}$. Encoder 1 maps $W_1$ into $X_1 = (X_{(1)1}, \ldots, X_{(n)1})$ and encoder 2 maps $W_2$ into $X_2 = (X_{(1)2}, \ldots, X_{(n)2})$, where all $X_1$ and $X_2$ satisfy the constraints set by (2.2). At the outputs, decoder 1 maps $Y_1 = (Y_{(1)1}, \ldots, Y_{(n)1})$ into $\mathcal{M}_1$ and decoder 2 maps $Y_2 = (Y_{(1)2}, \ldots, Y_{(n)2})$ into $\mathcal{M}_2$.

![Fig. 2.2. The complete model.]

Let an $(M_1, M_2, n)$ code for this channel be a set of two encoding functions $e_1 : \mathcal{M}_1 \to \mathbb{R}^n$, $e_2 : \mathcal{M}_2 \to \mathbb{R}^n$, with all image point satisfying the appropriate power constraint expressed in (2.2), and two decoding functions $d_1 : \mathbb{R}^n \to \mathcal{M}_1$, $d_2 : \mathbb{R}^n \to \mathcal{M}_2$. We then define the (average) probabilities of error

$$p_{e_1} = \frac{1}{M_1 M_2} \sum_{w_1, w_2} P \{ d_1(Y_1) \neq w_1 \mid W_1 = w_1, W_2 = w_2 \},$$

$$p_{e_2} = \frac{1}{M_1 M_2} \sum_{w_1, w_2} P \{ d_2(Y_2) \neq w_2 \mid W_1 = w_1, W_2 = w_2 \}$$

(2.3)

and
and

$$\max \{ p_{e,1}^n, p_{e,2}^n \} = p_e^n. \quad (2.4)$$

A rate pair \((R_1, R_2)\) is said to be achievable if there is a sequence of \((e^{nR_1}, e^{nR_2}, n)\) codes with \(p_e^n \to 0\) as \(n \to \infty\). The capacity region of this channel is defined as the closure of the set of all achievable rate pairs.

The Gaussian interference channel is perhaps the most important interference channel because it is a realistic model for a number of practical communication networks. Fig. 2.3 shows an example of these networks, in which two satellites attempt to send two independent messages to two separate Earth stations in the midst of added thermal noise and interference from each other.

![Diagram of satellite communication](image-url)

**Fig. 2.3.** Satellite communication.
From the capacity theorem for the Gaussian Shannon channel, it is clear that if $a = b = 0$, the capacity region of the Gaussian interference channel is given by the rectangular region formed by the rate pairs $(R_1, R_2)$ satisfying $\dagger$.

$$0 \leq R_1 \leq C_1 = \frac{1}{2} \log(1 + P_1)$$

$$0 \leq R_2 \leq C_2 = \frac{1}{2} \log(1 + P_2).$$ (2.5)

When either $a$ or $b$ is nonzero, it is clear that the capacity region is a subset (not necessarily a strict subset) of this rectangular region.

The capacity region of the Gaussian interference channel with arbitrary parameters $a$, $b$, $P_1$ and $P_2$ is yet to be established. The problem has been solved in the two special cases that follow.

In 1975, Carleial [4] demonstrated the striking fact that very strong interference is as innocuous as no interference. Specifically, he showed that if $a^2 \geq 1 + P_2$ and $b^2 \geq 1 + P_1$, the capacity region of the Gaussian interference channel is the full rectangular region described by (2.5). The reason is that the interfering signals are so strong in this case that the receivers may decode them reliably even if they consider their intended signals as noise. Having decoded the interfering signals, the receivers may clean their channels by subtracting out the interference.

In 1981, Han and Kobayashi [11] and Sato [12] found the capacity region for the strong interference case, where $a \geq 1$ and $b \geq 1$. They showed that for interference parameters in this range, both receivers would be able to reliably decode both messages, regardless of the particular coding technique being used. Thus the capacity region is simply the intersection of the capacity regions of the two multiple access channels

$\dagger$ Here and throughout this dissertation $\log(\cdot)$ denotes natural logarithm.
embedded in the interference channel. This region is the subset of rate pairs \((R_1, R_2)\) of the rectangle given by (2.5) for which

\[
R_1 + R_2 \leq \min \{ \frac{1}{2} \log(1 + a^2 P_1 + P_2), \frac{1}{2} \log(1 + P_1 + b^2 P_2) \}. \tag{2.6}
\]

The Gaussian interference channel acquires greater practical importance when the interference parameters are smaller than 1. A look at Fig. 2.3 will convince the reader that if one of these parameters is greater than 1, we should rather improve the alignment of the system antennas than invest in more sophisticated coding techniques.

Unfortunately, the capacity region of the Gaussian interference channel is not known in the cases when \(a\) or \(b\) or both are in the open interval \((0, 1)\). For interference parameters in this interval, we may define weak and moderate interference in the following manner. Consider for a moment the symmetrical interference channel where \(a = b\) and \(P_1 = P_2 = P\). We say the interference is weak if \(0 \leq a = b \leq a^*\) where \(a^*\) is the positive solution of the equation \(2 a^2 (1 + a^2 P) = 1\). If \(a = b\) falls in the range \((a^*, 1)\) the interference is said to be moderate. The meaning of \(a^*\) is illustrated in Fig. 2.4. It is the critical value where the largest known achievable rate sum \(R_1 + R_2\), regarded as a function of the interference parameter \(a\), exhibits a sudden change in slope, with its derivative going from a strictly negative value to zero. Therefore, the largest known achievable rate sum \(R_1 + R_2\) is a strictly decreasing function of weak interference parameters and a constant for parameters in the range of moderate interference. The curve in Fig. 2.4 is given by the following expressions:

\[
R_1 + R_2 = \log(1 + \frac{P}{1 + a^2 P}), \quad \text{if } 0 \leq a < a^*;
\]

\[
R_1 + R_2 = \frac{1}{2} \log(1 + 2 P), \quad \text{if } a^* \leq a < 1;
\]

\[
R_1 + R_2 = \frac{1}{2} \log(1 + P + a^2 P), \quad \text{if } 1 \leq a < \sqrt{1 + P};
\]
\[ R_1 + R_2 = \log(1 + P), \quad \text{if} \quad \sqrt{1 + P} \leq a. \quad (2.7) \]

**Fig. 2.4.** Best known achievable rate sum as a function of the interference parameter.

For weak interference the best known coding scheme simply treats the interference signal as noise. In the moderate interference range, we know of no technique better than time division (or frequency division) multiplexing. For strong and very strong interference, the optimal coding technique is known to employ superposition. Upper bounds for the curve in Fig. 2.4 in the weak and moderate interference ranges have been put forth by Sato [6] and Carleial [13], but optimality is yet to be established.

To date, the largest achievable rate region for the general Gaussian interference channel is that put forth by Han and Kobayashi [11]. In the present chapter we establish the optimality of two extreme points in the Han and Kobayashi region \(^\dagger\). A typical Han

\(^\dagger\) These points are also extreme points in the achievable region described by Carleial [5], [6].
and Kobayashi region is shown in Fig. 2.5 where the extreme points shown to be optimal are marked with stars. We prove the following theorem.

![Graph showing R1, R2, C1, C2 with points marked for optimal rates]

Fig. 2.5. Typical Han and Kobayashi region.

**Theorem 2.1:**

For rate pairs \((R_1, R_2)\) in the capacity region of a Gaussian interference channel with arbitrary positive interference parameters \(a\) and \(b\) and arbitrary power constraints \(P_1\) and \(P_2\), the following statements hold:

a) Let \(\epsilon > 0\). If \(R_2 \geq C_2 - \epsilon\), then

\[
R_1 \leq \frac{1}{2} \log(1 + \frac{a^2 P_1}{1 + P_2}) + \delta_1(\epsilon),
\]  

(2.8.a)

where \(\delta_1(\epsilon) \to 0\) as \(\epsilon \to 0\).

b) If \(R_1 \geq C_1 - \epsilon\), then
\[ R_2 \leq \frac{1}{2} \log(1 + \frac{b^2 P_2}{1 + P_1}) + \delta_2(\epsilon), \] (2.8.b)

where \( \delta_2(\epsilon) \to 0 \) as \( \epsilon \to 0 \).

Clearly, it suffices to prove one of these statements. We will prove assertion a). Theorem 2.1 simply says that if one of the senders (e.g., encoder 2) is transmitting at its maximal possible rate (the maximum it can achieve in the absence of interference), then the other sender (encoder 1) is forced to decrease its rate to the point where both receivers can reliably decode its message. Before proving Theorem 2.1, we examine a simpler model, that of a Gaussian interference channel with one of the interference parameters (e.g., \( b \)) equal to zero. Then we prove Theorem 2.1 for this class of interference channels. Finally a simple argument extends the proof to arbitrary Gaussian interference channels.

2.2. The Gaussian Z-Interference Channel

Consider a Gaussian interference channel where one (and only one) of the interference coefficients is zero. This model, called a Gaussian Z-interference channel, is illustrated in Fig. 2.6. It is completely specified by the powers \( P_1 \) and \( P_2 \) available to each encoder and by the positive interference coefficient \( a \).

If \( a \geq 1 \), the result of Han and Kobayashi [11] and Sato [12] for channel with strong interference can be easily extended to establish the capacity region of this channel as the subset of rate pairs \((R_1, R_2)\) in the rectangle of (2.6) such that

\[ R_1 + R_2 \leq \frac{1}{2} \log(1 + a^2 P_1 + P_2). \] (2.9)

Moreover, if \( a \) is in the very strong interference range, i.e., if \( a^2 \geq 1 + P_2 \), then (2.9) is dominated by the two bounds in (2.6) and the capacity region is the full
rectangle of rate pairs. Therefore, the unsolved case of this channel occurs when $a \in (0, 1)$.

Now we show that the class of Gaussian Z-interference channels with weak or moderate interference ($a \in (0, 1)$) and the class of degraded Gaussian interference channels are equivalent, i.e., for every Gaussian Z-interference channel with $0 < a < 1$ there is a degraded Gaussian interference channel with identical capacity region. Consider the four channels shown in Fig. 2.7. The first channel is a Gaussian Z-interference channel with parameters $P_1$, $P_2$ and $a < 1$. The second channel is obviously equivalent to the first, because $Y_2$ and $a^2 Y_2'$ are identically distributed and scaling the output of a channel does not affect its capacity.

![Diagram of the Gaussian Z-interference channel](image)

**Fig. 2.8.** The Gaussian Z-interference channel.

The third channel, in turn, is equivalent to the second because they have identical marginal distributions $p(y_1 \mid z_1, z_2')$ and $p(y_2' \mid z_1, z_2)$. Here we need the assumption that $a \leq 1$ to guarantee that $N_2 = \frac{1-a^2}{a^2}$ is non-negative. Finally, the fourth
channel of Fig. 2.7, a degraded Gaussian interference channel, is equivalent to the previous channels. A proof of this statement is presented in Appendix 1. To support it informally we observe that \( Y'_2 \) is a degraded version of \( Y'_1 \). Therefore, if receiver 2 can decode \( X_2 \) from \( Y'_2 \) with arbitrarily small probability of error, receiver 1 will perform at least as well in decoding \( X_2 \) from \( Y'_1 \). This receiver can thus subtract \( X_2 \) from \( Y'_1 \) and obtain the cleaner output \( Y_1 \) which is available to receiver 1 in the three other channels of Fig. 2.7. Therefore, the four interference channels illustrated have identical capacity regions.

We concentrate on the third channel of Fig. 2.7. That model is reproduced in Fig. 2.8 where we simplify notation by dropping the primes from the various variables and define the auxiliary signal \( Y_0 = X_1 + Z_1 + Z_2 \). The channel outputs are given by \( Y_1 = X_1 + Z_1 \) and \( Y_2 = X_1 + Z_1 + Z_2 + X_2 \), where \( X_1 \) and \( X_2 \) must satisfy power constraints \( P_1 \) and \( Q_2 = P_2 / a^2 \) and \( Z_1 \) and \( Z_2 \) are the Gaussian noise terms having variances 1 and \( N_2 = \frac{1-a^2}{a^2} \).

Remark: The notation \( H(\cdot) \) is used in this chapter to denote two different functionals of the probability distribution of the argument. If the argument is a discrete random variable it denotes the Shannon entropy (a functional of the probability mass distribution \( p(\cdot) \)) given by \( H = -\sum p \log p = -E \log p \). When the argument is continuous \( H(\cdot) \) denotes the differential entropy (a functional of the probability density function \( f(\cdot) \)) given by \( H = -\int f \log f = -E \log f \).
Fig. 2.7. Equivalent channels.
2.3. Proof of Theorem 2.1

First we prove the assertion of Theorem 2.1 for the model pictured in Fig. 2.8. Then we extend the proof to arbitrary Gaussian interference channels.

Suppose $R_2$ is very close to maximal, in the channel of Fig. 2.8. Then we show that $R_1$ must not exceed $\frac{1}{2} \log (1 + \frac{P_1}{1 + N_2 + Q_2})$ by very much.

From Fano’s inequality we have

$$H(W_1 \mid Y_1) \leq n R_1 p_{\epsilon,1}^n + h(p_{\epsilon,1}^n) = n \epsilon_{1n}, \quad (2.10)$$

$$H(W_2 \mid Y_2) \leq n R_2 p_{\epsilon,2}^n + h(p_{\epsilon,2}^n) = n \epsilon_{2n}. \quad (2.11)$$

where $h(\cdot)$ is the binary entropy function and $\epsilon_{1n}, \epsilon_{2n} \to 0$ as $p_{\epsilon}^n \to 0$.  

Fig. 2.8. Model used in the proof.
Therefore we have

\[
n R_2 = H(W_2) \leq I(W_2; Y_2) + n\epsilon_{2n}
\]

\[
\leq I(X_2; Y_2) + n\epsilon_{2n}
\]

\[
\leq I(X_2; Y_2 | X_1) + n\epsilon_{2n}
\]

\[
\leq \frac{1}{2} \log(1 + \frac{Q_2}{1 + N_2}) + n\epsilon_{2n},
\]

(2.12)

where \(\epsilon_{2n} \to 0\), as \(p^n \to 0\). In this sequence of expressions we used Fano's inequality (first line), the data processing inequality (second line), the independence between \(X_1\) and \(X_2\) (third line) and the fact that the Gaussian distribution maximizes entropy given a variance constraint (fourth line).

By hypothesis, encoder 2 is transmitting at a rate very close to maximal, i.e.,

\[
R_2 \geq \frac{1}{2} \log(1 + \frac{Q_2}{1 + N_2}) - \epsilon.
\]

(2.13)

Combining (2.12) and (2.13) we find

\[
I(X_2; Y_2 | X_1) - I(X_2; Y_2) \leq n(\epsilon + \epsilon_{2n})
\]

(2.14)

which can be written as

\[
H(X_2 | Y_2) - H(X_2 | X_1, Y_2) \leq n(\epsilon + \epsilon_{2n}).
\]

(2.15)

This expression is equivalent to

\[
H(X_1 | Y_2) - H(X_1 | X_2, Y_2) \leq n(\epsilon + \epsilon_{2n}).
\]

(2.16)

Now we note that the auxiliary signal \(Y_0 = X_1 + Z_1 + Z_2\) equals \(Y_2 - X_2\) and that \(X_1, Y_0\) and \(X_2\) are independent. Therefore we have

\[
H(X_1 | Y_2) - H(X_1 | Y_0) \leq n(\epsilon + \epsilon_{2n}).
\]

(2.17)
From (2.12) and (2.13) it is easy to show that

$$
\frac{n}{2} \log(2\pi e (1 + N_2 + Q_2)) - H(Z_1 + Z_2 + X_2) \leq n(\epsilon + \epsilon_{2n}).
$$

(2.18)

Let $Y_3 = X_1 + Z_1 + Z_2 + Z_3$, where $Z_3$ is a Gaussian multivariate random variable with covariance matrix $Q_2 I$ ($I$ is the identity matrix). Using Lemma A.2.1, proved in Appendix 2, we find that (2.18) implies

$$
H(X_1 | Y_3) - H(X_1 | Y_2) \leq n \epsilon_0
$$

(2.19)

where $\epsilon_0 \to 0$ as $\epsilon$ and $\epsilon_{2n} \to 0$.

This is a consequence of the fact that encoder 2 transmits information at rates very close to maximal. Hence $X_2 + Z_1 + Z_2$ can just barely be differentiated from a Gaussian process of power $1 + N_2 + Q_2$, even though $X_2$ itself, being a random vector drawn from the finite set of codewords, is far from appearing Gaussian.

From (2.10), (2.17) and (2.19), we may write the two following conditions:

$$
H(W_1 | Y_1) \leq n\epsilon_{1n},
$$

(2.20.a)

$$
H(X_1 | Y_3) - H(X_1 | Y_0) \leq n(\epsilon + \epsilon_0 + \epsilon_{2n}.
$$

(2.20.b)

---

**Fig. 2.9.** Relations between variables.
Fig. 2.9 illustrates how these variables relate to each other. To the codewords $X_i$ we sequentially add noise processes of powers $1$, $N_2$ and $Q_2$, resulting in the variables $Y_1$, $Y_0$ and $Y_3$, respectively. Conditions (2.20) merely state that the equivocation of $W_1$ with respect $Y_1$ is very small and that the equivocations of $X_1$ with respect to $Y_0$ and $Y_3$ are approximately equal. As we show in the next section, these two conditions imply that the equivocation of $W_1$ with respect to $Y_2$ is also very small.

To continue the proof of Theorem 2.1 we need the following inequality, which is proved in the next section: For $X_1$, $Y_1$, $Y_0$ and $Y_3$ as defined above we have

$$
\frac{1}{n} H(X_1 | Y_3) \geq \frac{1}{2} \log \left( \frac{N_2 (1 + N_2 + Q_2) e^{n \left( H(X_1 | Y_1) + H(X_1 | Y_0) \right)}}{(1 + N_2) (N_2 + Q_2) e^{n \left( H(X_1 | Y_1) - Q_2 e^{n \left( H(X_1 | Y_1) - H(X_1 | Y_0) \right)} \right)}} \right)
$$

(2.21)

We substitute condition (2.20.b) into (2.21) to obtain

$$
\frac{2^{n H(X_1 | Y_3)}}{e^{n H(X_1 | Y_3)}} \geq \frac{N_2 (1 + N_2 + Q_2) e^{n \left( H(X_1 | Y_1) + H(X_1 | Y_0) - \sigma (c + \epsilon_0 + \epsilon_1) \right)}}{(1 + N_2) (N_2 + Q_2) e^{n \left( H(X_1 | Y_1) - Q_2 e^{n \left( H(X_1 | Y_1) - H(X_1 | Y_0) \right)} \right)}}
$$

(2.22)

Now it takes a simple algebraic manipulation to show

$$
H(X_1 | Y_3) - H(X_1 | Y_1)
$$

\leq \frac{n}{2} \log \left( \frac{(1 + N_2) (N_2 + Q_2) e^{2(c + \epsilon_0 + \epsilon_1)} - N_2 (1 + N_2 + Q_2)}{Q_2} \right)

(2.23)

Defining $\delta$ to equal half the logarithmic expression above, we have

$$
H(X_1 | Y_3) - H(X_1 | Y_1) \leq n \delta,
$$

(2.24)
where $\delta \to 0$ as $\epsilon$ and $\epsilon_2 \to 0$.

Expanding both $H(W_1, X_1 \mid Y_3)$ and $H(W_1, X_1 \mid Y_2)$ in two ways by the chain rule and subtracting, we have

$$H(W_1 \mid Y_3) \leq H(X_1 \mid Y_3) - H(X_1 \mid Y_1) + H(W_1 \mid Y_1)$$

$$+ H(X_1 \mid W_1, Y_3) - H(X_1 \mid W_1, Y_3)$$

$$\leq H(X_1 \mid Y_3) - H(X_1 \mid Y_1) + H(W_1 \mid Y_1),$$  \hspace{1cm} (2.25)

and from (2.20.a) and (2.24) we find

$$H(W_1 \mid Y_3) \leq n(\delta + \epsilon_1).$$  \hspace{1cm} (2.26)

It is easy to show that $H(W_1 \mid Y_2) \leq H(W_1 \mid Y_3)$. Therefore, we obtain the desired inequality

$$H(W_1 \mid Y_2) \leq n(\delta + \epsilon_1).$$  \hspace{1cm} (2.27)

This is a Fano type inequality that can be used together with (2.10), in the traditional way, to establish the bound

$$n(R_1 + R_2) \leq \frac{1}{2} \log (1 + \frac{P_1 + Q_2}{1 + N_2}) + n(\delta + 2\epsilon_1),$$  \hspace{1cm} (2.28)

which, together with (2.13) yields

$$R_1 \leq \frac{1}{2} \log (1 + \frac{P_1}{1 + N_2 + Q_2}) + \epsilon + \delta + 2\epsilon_1.$$  \hspace{1cm} (2.29)

Now, letting the probability of error $p_e^2$ tend to zero, the Fano inequality terms $\epsilon_1$ and $\epsilon_2$ also go to zero. Finally, letting $\epsilon \to 0$ we have $(\epsilon + \delta) \to 0$. This shows the desired result for the channel of Fig. 2.8: if sender 2 transmits at capacity, sender 1 must also communicate reliably to receiver 2.
To extend this result to the two-sided Gaussian interference channels of Theorem 2.1, we note that the capacity region of these channels can only be enlarged if one of the interference links is removed. In other words, if we add an interference link to a Z-interference channel, any outer bound to the capacity region of the Z channel will also hold for the capacity region of the derived two-sided channel. Now suppose we remove the link with gain \( b \) from the two-sided channel of Fig. 2.1. If \( a \geq 1 \), the full capacity region of the remaining Z-interference channel is known [11], [12], and the bounds in Theorem 2.1 indeed hold. Alternatively, if \( 0 < a < 1 \), the remaining Z channel is the one pictured in Fig. 2.7 (i), which we have shown to be equivalent to the channel of Fig. 2.8. Substituting \( Q_2 = P_2/a^2 \) and \( N_2 = \frac{1-a^2}{a^2} \) in (2.29), we establish the bound in part a) of Theorem 2.1 for the channel of Fig. 2.7 (i), and hence for the original two-sided Gaussian interference channel. In a completely analogous way, we can show part b) of Theorem 2.1.

Theorem 2.1 is a converse type theorem. It establishes upper bounds on the performance of Gaussian interference channels, but says nothing about the achievability of these bounds. It may happen, as it does in some cases, that a stronger bound holds for \( R_1 \) when \( R_2 \) is constrained to be maximal, and similarly for \( R_2 \) when \( R_1 \) is maximal. In what follows, we give achievable upper bounds on \( R_1 \) when \( R_2 = C_2 \).

For \( a \in (0,1] \) and \( b \) arbitrary we can show, using standard techniques, that an achievable upper bound on \( R_1 \), when \( R_2 \) is maximal, is given by

\[
R_1 \leq \frac{1}{2} \log \left( 1 + \frac{a^2 P_1}{1 + P_2} \right). \tag{2.30}
\]

When \( b \in [0,1] \), another upper bound holds for \( R_1 \) if \( R_2 \) is maximal. This bound

\[\text{The word achievable is used here in a slightly broader sense. In this context, a point is said to be achievable if it is in the capacity region, i.e., if it is a cluster point of a sequence of achievable (in the canonical sense) points.}\]
follows from the work of Sato [8] on the degraded Gaussian interference channel, the
equivalence among the channels of Fig. 2.7, and the fact that removing an interference
link from a two-sided interference channel can only enlarge its capacity region. It asserts
that \( b \epsilon [0,1] \) and \( R_2 \geq C_2 - \epsilon \) implies

\[
R_1 \leq \frac{1}{2} \log \left( 1 + \frac{P_1}{1 + b^2 P_2} \right) + \delta,
\]

where \( \delta \to 0 \) as \( \epsilon \to 0 \).

Using this result and Theorem 2.1 we can easily show that an achievable upper
bound on \( R_1 \), when \( R_2 \) is constrained to be maximal, is given by

\[
R_1 \leq \min \left\{ \frac{1}{2} \log \left( 1 + \frac{a^2 P_1}{1 + P_2} \right), \frac{1}{2} \log \left( 1 + \frac{P_1}{1 + b^2 P_2} \right) \right\}.
\]

Clearly, analogous results can be obtained when \( R_1 \) is constrained to be maximal.

2.4. Bounds on Equivocation

The crucial step in the proof of Theorem 2.1 is the lower bound on the equivocation
\( H(X_1 \mid Y_3) \) provided by (2.21). In the following theorem, this lower bound and an upper
bound on \( H(X_1 \mid Y_3) \) are established.

Theorem 2.2:

For \( X_1, Y_1, Y_0 \) and \( Y_3 \) as defined in the previous section the following inequalities
hold:

\[
\frac{1}{n} H(X_1 \mid Y_3) \geq \frac{1}{2} \log \left( \frac{N_2(1 + N_2 + Q_2) e^{\frac{2}{n} H(X_1 \mid Y_1) + H(X_1 \mid Y_0)}}{(1 + N_2)(N_2 + Q_2) e^{\frac{2}{n} H(X_1 \mid Y_1) - Q_2 e^{\frac{2}{n} H(X_1 \mid Y_0)}}} \right)
\]
\[
\frac{1}{n} H(X_1 \mid Y_3) \leq \frac{1}{2} \log \left( \frac{(1 + N_2 + Q_2)}{(1 + N_2)} e^{\frac{1}{n} H(X_1 \mid Y_0)} \right)
\]

(2.34)

Proof:

Let \( X \) be an arbitrarily distributed multivariate random variable and \( Z_1 \) be Gaussian random vector, independent of \( X \), with covariance matrix \( tI \). In chapter 3 we show that \( e^{\frac{1}{n} H(X + Z_1)} \) is a concave function of \( t \). This implies that

\[
\frac{1}{n} H(Y_0) \geq \frac{N_2}{N_2 + Q_2} e^{\frac{1}{n} H(Y_4)} + \frac{Q_2}{N_2 + Q_2} e^{\frac{1}{n} H(Y_2)}
\]

(2.35)

Multiplying this equation through by

\[
\frac{2}{n} (H(X_1) + H(Z_1)) e^{\frac{1}{n} H(Y_4) + H(Y_2) - H(Y_0) - H(Y_1)}
\]

and recognizing that

\[
H(X_1 \mid X_1 + Z_1) = H(X_1) + H(Z_1) - H(X_1 + Z_1)
\]

(2.36)

we have

\[
(N_2 + Q_2) e^{\frac{2}{n} H(Z_1 + Z_4) + \frac{2}{n} H(X_1 \mid Y_1) + H(X_1 \mid Y_3)}
\]

\[
\geq N_2 e^{\frac{2}{n} H(Z_1 + Z_4) + \frac{2}{n} H(X_1 \mid Y_1) + H(X_1 \mid Y_3)}
\]

\[
+ Q_2 e^{\frac{2}{n} H(X_1) + H(X_1 \mid Y_3) + H(X_1 \mid Y_2)}
\]

(2.37)

Now we notice that \( e^{\frac{2}{n} H(Z_1 + Z_4 + Z_3)} = 2\pi e(1 + N_2 + Q_2) \), with analogous expressions holding for the other Gaussian terms. Using these relations in (2.36) and rearranging terms we find
\[
\frac{2^n H(X_1 | Y_0)}{2^n (1 + Q_2) e^n (2^n H(X_1 | Y_1) + H(X_1 | Y_0))} \geq \frac{2^n H(X_1 | Y_1) - 2^n H(X_1 | Y_0)}{(1 + N_2)(N_2 + Q_2) e^n - Q_2 e^n} \tag{2.38}
\]

From this, (2.33) follows immediately.

A proof of (2.34) uses Shannon's entropy power inequality. This inequality is investigated in detail in chapter 3. It states that

\[
\frac{2^n H(Y_1)}{e^n} \geq \frac{2^n H(Y_0)}{e^n} + \frac{2^n H(Y_2)}{e^n} \tag{2.39}
\]

Using Eq. (2.36) and rearranging terms as before we find

\[
(1 + N_2 + Q_2) e^{-2^n H(X_1 | Y_1)} \geq (1 + N_2) e^{-2^n H(X_1 | Y_0)} + Q_2 e^{-2^n H(X_1)}
\]

\[
\geq (1 + N_2) e^{-2^n H(X_1 | Y_0)} \tag{2.40}
\]

This yields the desired proof of (2.34).

There is a simple interpretation for the bounds given in (2.33) and (2.34). To see this, consider the family of equivocation functions of the form shown in Fig. 2.10 and expressed by

\[
\frac{1}{n} H(X_1 | X_1 + Z_t) = 0
\]

for \( t < N_0 \) and by

\[
\frac{1}{n} H(X_1 | X_1 + Z_t) = \frac{1}{2} \log \left( \frac{t e^{2R}}{t + N_0 (e^{2R} - 1)} \right) \tag{2.41}
\]

for \( t \geq N_0 \). Suppose \( e^{nR} \) messages are mapped into their associated codewords \( X_1 \) and reliably decoded by a receiver that observes the sum of \( X_1 \) and a Gaussian noise process of power \( N_0 \). The function above gives the asymptotic value of the equivocation of \( X_1 \).
with respect to $X_1 + Z_i$ when the code that requires the least possible amount of power is utilized.

![Graph](image)

**Fig. 2.10.** Equivocation as function of added noise variance.

Now, suppose the values of $H(X_1 \mid Y_1)$ and $H(X_1 \mid Y_0)$ are fixed. Let $R$ and $N_0$ be adjusted so that the equivocation function expressed in (2.41) satisfies these two constraints. Then inequality (2.33) states that the value of this function at $t = 1 + N_2 + Q_2$ is a lower bound for $H(X_1 \mid Y_3)$. On the other hand, (2.34) says that the fastest growth in equivocation is logarithmic, achieved when $X_1$ is Gaussian with independent, identically distributed components. In Fig. 2.11 these lower and upper bound are shown for constrained values of $H(X_1 \mid Y_1)$ and $H(X_1 \mid Y_0)$. 
2.5. A Model with Degraded Message Sets

In this section we present a new problem, a $Z$-interference channel with degraded message sets, in which the effects of interference can be completely eliminated. The model under study is shown in Fig. 2.12. It differs from the usual Gaussian $Z$-interference channel in that encoder 2 has access to both messages $W_1$ and $W_2$ to be sent over the channel, even though all it wishes to accomplish is to reliably communicate $W_2$ to receiver 2.

We show that the capacity region of this channel equals the full rectangle of rate
pairs \((R_1, R_2)\) satisfying \(R_1 \leq C_1\) and \(R_2 \leq C_2\).

Fig. 2.12. A model with degraded message sets.

In order to establish this result we consider the communication problem illustrated in Fig. 2.13. In this model, an index \(W\), uniformly distributed over the set \(\{1, \ldots, M\}\), is to be sent to the receiver in \(n\) uses of the channel. Here \(M\) is the greatest integer smaller than or equal to \(e^{nR}\) and \(R\) is the rate in nats per transmission. The state \(S\) of the channel for \(n\) transmissions is assumed to be a sequence of independent, identically distributed \(N(0, Q)\) random variables. The state \(S\) is known to the transmitter but not to the receiver. Based on \(W\) and \(S\), the encoder sends a codeword \(X\) which must satisfy the power constraint \(\frac{1}{n} \sum_{i=1}^{n} x_i^2 \leq P\).

The channel output is given by \(Y = X + S + Z\), where the channel noise \(Z\) is distributed according to \(N(0, NI)\). Upon receipt of \(Y\) the decoder creates an estimate \(\hat{W}\) of the index \(W\). The probability of error \(P_e\), is given by

\[
P_e = \frac{1}{M} \sum_{k}^{M} \text{Prob}\{\hat{W} \neq k \mid W = k\}. \tag{2.42}
\]
This is the standard Gaussian channel with input power constraint $P$, where the encoder is informed of part of the Gaussian additive noise sequence that will be added to his signal. Unfortunately, this information is not made available to the decoder, who will have to base his estimate $\hat{W}$ solely on the channel output $Y$.

![Diagram](image)

**Fig. 2.13.** A variation of the Gaussian Shannon channel.

There is an obvious encoding scheme to communicate over the channel of Fig. 2.13. If $P > Q$, the encoder may use part of his available power to cancel $S$. He can then use the remaining power $P - Q$ to send information at rate $\frac{1}{2} \log(1 + \frac{P - Q}{N})$. In general, using a fraction $0 \leq \alpha \leq \min\{1, Q/P\}$ of the transmitter power to partially cancel $S$ yields a rate $\frac{1}{2} \log(1 + \frac{(1-\alpha)P}{N + (\sqrt{Q} - \sqrt{\alpha P})^2})$. This scheme may be justified by one's temptation to reduce the problem to a previously solved one, but as we show, it is not an optimal encoding procedure. In fact the optimal encoding uses codewords in the direction of $S$. It looks at the space surrounding the vector $S$ and chooses codewords that are compatible with the power constraint and far enough apart to be distinguishable when viewed from the channel output. In so doing, the encoder actually adapts its signal to the state $S$ instead of trying to erase it.
We show that all rates \( R < C^* = \frac{1}{2} \log(1 + \frac{P}{N}) \) are achievable. Then, it follows easily that \( C^* \) is indeed the capacity of this model.

First we recall the following result. Gel'fand and Pinsker [15] and El Gamal and Heegard [16] have shown that the capacity of a discrete memoryless channel with random state \( S \) known to the encoder is given by

\[
C = \max_{p(u,x|s)} \{ I(U;Y) - I(U;S) \},
\]

where the maximum is over all joint distributions of the form \( p(s)p(u,x|s)p(y|x,s) \), where \( U \) is a finite alphabet auxiliary random variable.

Their random coding argument, assuming discrete, finite alphabets and unconstrained input, can be outlined as follows. First generate \( \exp\{ n(I(U;Y) - \epsilon) \} \) independent identically distributed sequences \( U \), according to the uniform distribution over the set of typical \( U' s \). Next, distribute these sequences uniformly over \( e^{nR} \) bins. For each sequence \( u \) let \( i(u) \) be the index of the bin containing \( u \). For encoding, given the state vector \( S \) and the message \( W \), look in bin \( W \) for a sequence \( U \) such that \( (U,S) \) is jointly typical. Declare an error if no such \( U \) can be found. If the number of sequences in bin \( W \) is larger than \( \exp\{ n(I(U;S) + \delta) \} \), the probability of finding no such \( U \) decreases to zero exponentially as \( n \) increases. Next, choose \( X \) such that \( (X,U,S) \) is jointly typical and send it through the channel. At the decoder look for the unique sequence \( U \) such that \( (U,Y) \) is jointly typical. Declare an error if more than one or no such sequence exist. Then set the estimate \( \hat{W} \) equal to the index of the bin containing the obtained sequence \( U \). If \( R < I(U;Y) - I(U;S) - \epsilon - \delta \), the probability of error averaged over all codes decreases exponentially to zero as \( n \rightarrow \infty \). This shows the existence of a code that achieves rate \( R \) with arbitrarily small probability of error.

This result can be readily extended to memoryless channels with discrete time and
continuous alphabets [17], by considering the supremum of $I(U_d ; Y_p ) - I(U_d ; S_q )$ over all finite alphabet variables $U_d$ and all partitions $Y_p$ and $S_q$ of the channel output and state.

Thus the problem is reduced to that of finding an appropriate auxiliary variable $U$. We consider $U = X + \alpha S$ where $X$ and $S$ are independent random variables distributed according to $N(0, P)$ and $N(0, Q)$, respectively, and $\alpha$ is a parameter to be determined. Note that there could be a loss of generality in restricting attention to such $U'$s, but as we shall see, the derived answer is clearly optimal. Recalling that $Y = X + S + Z$ with $Z$ distributed according to $N(0, N)$, the relevant mutual informations can be calculated to yield

$$I(U ; Y) = H(X + S + Z) - H(X + S + Z | X + \alpha S)$$

$$= H(X + S + Z) + H(X + \alpha S) - H(X + S + Z ; X + \alpha S)$$

$$= \frac{1}{2} \log((2\pi e)^2 (P + Q + N)(P + \alpha^2 Q))$$

$$- \frac{1}{2} \log((2\pi e)^2 ((P + Q + N)(P + \alpha^2 Q) - (P + \alpha Q)^2))$$

$$= \frac{1}{2} \log \left( \frac{(P + Q + N)(P + \alpha^2 Q)}{PQ(1-\alpha)^2 + N(P + \alpha^2 Q)} \right) \quad (2.44)$$

and

$$I(U ; S) = \frac{1}{2} \log \left( \frac{P + \alpha^2 Q}{P} \right). \quad (2.45)$$

Let

$$R(\alpha) = I(U ; Y) - I(U ; S). \quad (2.46)$$
Then

\[ R(\alpha) = \frac{1}{2} \log \left( \frac{P(P + Q + N)}{PQ(1-\alpha)^2 + N(P + \alpha^2 Q)} \right). \]  \hspace{1cm} (2.47)

Maximizing \( R(\alpha) \) over \( \alpha \), we get

\[ \max_\alpha R(\alpha) = R(\alpha^*) = \frac{1}{2} \log \left( 1 + \frac{P}{N} \right) = C^* \]  \hspace{1cm} (2.48)

obtained for \( \alpha^* = P/(P + N) \).\footnote{It is interesting to note that \( \alpha^* \) as well as \( R(\alpha^*) \) do not depend on \( Q \).}

This is the desired answer. We know that the capacity of the channel cannot exceed \( \max_{p(x \mid s)} I(X ; Y \mid S) \), for this is the capacity when both encoder and decoder know the sequence \( S \). This expression can be easily shown to equal \( C^* \). But from (2.48), we achieve \( C^* \). Thus the optimality of the present scheme is immediately established.

In what follows we reexamine the code generation and the encoding-decoding procedures outlined before to show that we can send information at rate \( C^* \) with arbitrarily small probability of error and still satisfy the power constraint \( P \). First, we generate \( \exp\{ n(I(U^*; Y) - \epsilon) \} \) sequences \( U(\epsilon > 0, \text{arbitrarily small}) \) with components independently drawn according to \( N(0, P + \alpha^2 Q) \). Then, we place these sequences into \( e^{nR} = \exp\{ n(C^* - 2\epsilon) \} \) bins in such a way that each bin will contain the same number of sequences. The code book formed by the generated sequences and their assignments to the \( e^{nR} \) bins is known to both the encoder and the decoder. Given a state vector \( S = S_0 \) and a message \( W = k \), the encoder looks for a jointly typical pair \((U, S_0)\) among the \( U' s \) in bin \( k \). This is equivalent to looking for a sequence \( U \) such that

\[ \left| (U - \alpha^* S_0)' S_0 \right| \leq \delta \]  \hspace{1cm} (2.49)

for some appropriately small \( \delta \). In other words, the encoder searches for the sequence
(U - α^*S_0) which is nearly orthogonal to S_0. The encoder declares an error if no such sequence is found. The number of sequences in each bin is \(\exp\{n(I(U^*; S) + \epsilon)\}\). By arguments similar to those used in [18], [19] it can be shown that the probability of finding no suitable sequence U vanishes exponentially as \(n \to \infty\). Call U_0 the sequence obtained in this way. The encoder calculates \(X_0 = U_0 - \alpha^*S_0\). With high probability \(X_0\) will be typical, which is to say \(X_0\) will satisfy the power constraint \(\frac{1}{n} \sum_{i=1}^{n} (X(i)_0)^2 \leq P\). If not, the encoder declares an error.

Supposing no errors have occurred during the encoding procedure, the encoder sends \(X_0\) over the channel. Note that the set of \(X^i\)'s one might transmit is a continuum. At the other end, the decoder receives \(Y = Y_0\) and then looks for a sequence U such that \((U, Y_0)\) is jointly typical. He declares an error if he can find more than one or no such sequence. With high probability the decoder will find only one such sequence and it will be equal to U_0. Finally, the decoder sets his estimate \(\hat{W}\) as the index of the bin containing this sequence. The probability of error averaged over the random choice of code decreases exponentially to zero as \(n \to \infty\). This demonstrates the existence of a code satisfying the power constraint and achieving \(C^*\) with arbitrarily small probability of error.

Returning our attention to the Z-interference model with degraded message sets shown in Fig. 2.12, we assume sender 1 transmits at a rate \(R_1\) very close to \(C_1\). It may accomplish that by assigning each message \(W_1\) to one codeword \(x_1\) randomly generated according to a Gaussian distribution of variance \(P_1\). Then, as encoder 2 knows both messages to be sent over the channel, it also knows the particular codeword \(X_1\) that will be interfering with its transmission. Therefore, it may employ the scheme just examined to achieve rates arbitrarily close to \(C_2\). This argument shows that the rate pair \((C_1, C_2)\) is in the capacity region of the channel. On the other hand, this rate pair is clearly
optimal. Hence, the capacity region of this example is the full rectangle of rate pairs.
3. Strengthening the Entropy Power Inequality

3.1. Introduction

Let $H(X) = -\int p(x) \log p(x) \, dx$ be the differential entropy of an n-dimensional random variable $X$ with probability density $p(x)$. We extend this definition by defining $H(X) = -\infty$ if the distribution of $X$ assigns positive mass to one or more singletons in $\mathbb{R}^n$. The entropy power of $X$ is defined as the variance of the independent, identically distributed components of an n-dimensional white Gaussian random variable having entropy $H(X)$. This definition was introduced by Shannon in his historical 1948 paper [20]. The entropy power, denoted by $N(X)$, and the entropy $H(X)$ of a random vector $X$ are related by

$$ N(X) = \frac{e^{\frac{2}{n} H(X)}}{2\pi e} \quad (3.1) $$

Also in his 1948 article, Shannon proposed the entropy power inequality:

$$ N(X+Z) \geq N(X) + N(Z) \quad (3.2) $$

or

$$ e^{\frac{2}{n} H(X+Z)} \geq e^{\frac{2}{n} H(X)} + e^{\frac{2}{n} H(Z)} \quad (3.3) $$

for any two independent multivariate random variables $X$ and $Z$. Shannon demonstrated extraordinary insight in proposing this inequality. He proceeded to use it in establishing a lower bound to the capacity of a band limited channel with arbitrarily distributed additive noise. Shannon provided a variational argument to show that the entropy power of the sum of two independent random vectors of given entropy powers has zero variation if the two random vectors are Gaussian with proportional covariance matrices.

---

† We assume throughout this thesis that the random variables considered either have a probability density function or assign positive mass to at least one singleton.
Unfortunately, this argument does not rule out the possibility that the stationary point at the Gaussian distribution is in fact a local minimum or a saddle point. A formal proof of this inequality was given eleven years later by Stam [21]. Stam's proof was later simplified by Blachman [22]. Since then, the entropy power inequality has been used in the proofs of some important theorems of Information Theory. For example, it is instrumental in establishing the capacity region of the Gaussian broadcast channel [23].

In our study of the Gaussian interference channel it is necessary to strengthen this inequality. The entropy power inequality involves a constrained minimization of the entropy of $X + Z$, where $X$ and $Z$ are independent random vectors and the entropies of $X$ and $Z$ are fixed. Now let $Z$ be a Gaussian $n$-vector with independent identically distributed components and consider the related problem of minimizing the entropy of $X + \alpha Z$, $\alpha \in [0,1]$, when the entropy of $X$ and that of the sum $X + Z$ are fixed. We prove the following theorem:

Theorem 3.1:

Let $X$ be an arbitrarily distributed random vector and $Z$ be a Gaussian vector, independent of $X$, with covariance matrix proportional to the identity matrix. Then

\[
\frac{2}{n} H(X + \alpha Z) \geq (1-\alpha^2) e^{\frac{2}{n} H(X)} + \alpha^2 e^{\frac{2}{n} H(X + Z)} \quad \alpha \in [0,1]
\]  

(3.4)

where equality is obtained if and only if $X$ also is a Gaussian random vector with covariance matrix proportional to the identity matrix.

By substituting $\beta = 1/\alpha$ and $Z' = \alpha Z$ in the above inequality, we can show that it also provides an upper bound for $H(X + \beta Z')$ when $\beta \geq 1$.

The inequality above is stronger than the entropy power inequality. To see this we
subtract $e^{-\frac{2}{\alpha}H(X)}$ from both sides of (3.4) and divide through by $\alpha^2$, obtaining

$$
\frac{1}{\alpha^2} \left( e^{-\frac{2}{\alpha}H(X+\alpha Z)} - e^{-\frac{2}{\alpha}H(X)} \right) \geq e^{-\frac{2}{\alpha}H(X+Z)} - e^{-\frac{2}{\alpha}H(X)}
$$

(3.5)

Now, letting $\beta = 1/\alpha$, ($\beta \geq 1$) and using the relation $\beta^2 e^{-\frac{2}{\alpha}H(X)} = e^{-\frac{2}{\alpha}H(\beta X)}$, we get

$$
e^{-\frac{2}{\alpha}H(\beta X + Z)} - e^{-\frac{2}{\alpha}H(\beta X)} \geq e^{-\frac{2}{\alpha}H(X+Z)} - e^{-\frac{2}{\alpha}H(X)}, \quad \beta \geq 1.
$$

(3.6)

In this expression $X$ is an arbitrary random variable. Therefore, (3.6) merely states that the function $f(a)$, defined by

$$
f(a) = e^{-\frac{2}{\alpha}H(aX+Z)} - e^{-\frac{2}{\alpha}H(aX)}, \quad a \geq 0
$$

(3.7)

is a monotonically increasing function of $a$. Noting that $f(0) = e^{-\frac{2}{\alpha}H(Z)}$, we can express the entropy power inequality in terms of $f(\cdot)$ as $f(1) \geq f(0)$, whereas the inequality in Theorem 3.1 is equivalent to $f'(a) \geq 0$. Clearly, the latter is a stronger inequality.

To prove Theorem 3.1 we may assume, without loss of generality, that $Z$ has an identity covariance matrix. Let $t > 0$. Define $X_i = X + Z_i$ where $Z_i$ is independent of $X$ and is normally distributed with covariance matrix $t I$. Using this notation we may write $Z = Z_1$. Then, proving Theorem 3.1 is equivalent to showing that

$$
e^{-\frac{2}{\alpha}H(X+Z_i)} \geq (1-t) e^{-\frac{2}{\alpha}H(X)} + t e^{-\frac{2}{\alpha}H(X+Z_i)}, \quad t \in [0,1].
$$

(3.8)

This in turn is equivalent to showing that $e^{-\frac{2}{\alpha}H(X)}$ is a concave function of $t$, our task in the next section.

### 3.2. The Concavity of Entropy Power with Added Noise

Let $X$, $X_i$ and $Z_i$ be as defined above. We show that the entropy power of $X_i$,
given by \( N(X_t) = \frac{e^n}{2\pi e} \), is a concave function of \( t \).

Before giving a formal proof, let us verify intuitively why it is so. In Fig. 3.1 we have a typical graph of \( \frac{e^n}{2\pi e} \) as a function of \( t \) for some arbitrary \( X \).

![Graph showing the growth of entropy power with added Gaussian noise.](image)

**Fig. 3.1.** Typical growth of entropy power with added Gaussian noise.

When \( t = 0 \) the function equals the entropy power of \( X \). As \( t \) increases, \( N(X_t) \) also increases. In fact, as shown in Appendix 3, it follows from the entropy power inequality that the derivative of \( N(X_t) \) with respect to \( t \) is greater than one for all \( t \in (0, \infty) \) except when the components of \( X \) are independent identically distributed Gaussian random variables, in which case \( N(X_t) \) is a linear function of \( t \). It makes sense to expect that, as \( t \) becomes very large, \( X_t \) will behave more as a Gaussian random vector and the slope of \( N(X_t) \) will approach one. It is also reasonable that this approach to slope 1 occurs in a monotonic way. The monotonicity of this approach to unit slope (or to Gaussian
behavior) implies the concavity of $N(X_t)$, i.e., $\frac{d^2}{dt^2}N(X_t) \leq 0$. The above discussion points out a basic difference between the entropy power inequality and Theorem 3.1. The former relates to the first derivative of $N(X_t)$ whereas the latter deals with the second derivative of $N(X_t)$.

We divide the proof of the concavity of $N(X_t)$ into three parts:

1) First we show that

$$\frac{d}{dt} H(X_t) = \frac{1}{2} \int_{\mathbb{R}^s} \| \nabla p_t(x_t) \|^2 \frac{dx_t}{p_t(x_t)}$$

where

$$p_t(x_t) = \frac{1}{(2\pi t)^{n/2}} \int_{\mathbb{R}^s} p(x) \exp\left(-\frac{\|x_t-x\|^2}{2t}\right) dx$$

is the probability density of $X_t$, $p(x)$ is the probability density of $X$, and $\| \cdot \|$ denotes Euclidean norm. Due to the smoothing caused by the added Gaussian noise, for $t > 0$ $p_t(x_t)$ is well defined even if $X$ does not have a density function. (In this case we define $p_t(x_t)$ in terms of the cumulative distribution function of $X$.) Moreover, $p_t(x_t)$ satisfies a number of desirable properties, among which are, for $t > 0$:

(i) $p_t(x_t) > 0$ for all $x_t$,

(ii) $\lim_{\|x_t\| \to \infty} p_t(x_t) = 0$,

(iii) $p_t(x_t)$ is infinitely differentiable everywhere,

(iv) $p_t(x_t) \leq \frac{1}{(2\pi t)^{n/2}}$ for all $x_t$.

These properties are crucial for the proofs in this section and for those in Appendix 4.

The integral in the right side of Eq. (3.9) is the trace of the Fisher Information matrix of the translation family of the probability density $p_t(x_t)$. According to Stam [21] the scalar analog of this relation is due to de Bruijn.
2) Then we demonstrate the following theorem:

Theorem 3.2:

Let \( X \) be an arbitrary random vector and \( Z_t \) a Gaussian random vector with independent identically distributed components having mean zero and variance \( t \). Set \( X_t = X + Z_t \). Then

\[
\frac{d^2}{dt^2} e^{\frac{1}{2} H(X_t)} = e^{\frac{1}{2} H(X_t)} \left( \left( \frac{E}{n} \sum_{i=1}^{n} \frac{\partial^2 \log p_t(x_i)}{\partial x(i)_t^2} \right)^2 - E \frac{1}{n} \sum_{i=1}^{n} \left( \frac{\partial^2 \log p_t(x_i)}{\partial x(i)_t^2} \right)^2 \right) \]

\[
- E \frac{1}{n} \sum_{i \neq j} \left( \frac{\partial^2 \log p_t(x_i)}{\partial x(i)_t \partial x(j)_t} \right)^2 \right). \tag{3.11}
\]

where \( x_t = [x_{(1)}_t, x_{(2)}_t, \cdots, x_{(n)}_t] \) and \( E \) denotes expectation with respect to \( X_t \).

3) Finally, we use Theorem 3.2 to prove the following.

Theorem 3.3:

Let \( X \) and \( Z_t \) be as above. The entropy power of \( X + Z_t \) is a concave function of \( t \), which is in fact strictly concave unless \( X \) is a Gaussian vector with independent identically distributed components.

3.2.1. Proof of (3.9)

Due to the smoothing properties of the normal distribution we can differentiate (3.9) inside the integral (the integrand is continuous and differentiable) to show that \( p_t(x_t) \) satisfies the diffusion (heat) equation
\[
\frac{d}{dt}p_t(x_t) = \frac{1}{2} \nabla^2 p_t(x_t)
\]  
(3.12)

where

\[
\nabla^2 p_t(x_t) = \sum_{i=1}^{n} \frac{\partial^2}{\partial x_i(x_t)} p_t(x_t)
\]

Interchanging derivative and integral once more, we get

\[
\frac{d}{dt} H(X_t) = - \int_{\mathbb{R}^n} \frac{d}{dt} p_t(x_t) \, dx_t - \int_{\mathbb{R}^n} \left( \frac{d}{dt} p_t(x_t) \right) \log p_t(x_t) \, dx_t
\]

\[
= 0 - \frac{1}{2} \int_{\mathbb{R}^n} (\nabla^2 p_t(x_t)) \log p_t(x_t) \, dx_t.
\]  
(3.13)

We now recall Green's identity [24]: if \( \phi(x) \) and \( \psi(x) \) are twice continuously differentiable functions in \( \mathbb{R}^n \) and \( V \) is any set bounded by a piecewise smooth, closed, oriented surface \( S \) in \( \mathbb{R}^n \), then

\[
\int_V \nabla \cdot \nabla \psi \, dV = \int_S \nabla \psi \cdot \nu \, ds - \int_V \nabla \phi \cdot \nabla \psi \, dV
\]  
(3.14)

where \( \nabla \psi \) denotes the gradient of \( \psi \), \( ds \) denotes the elementary area vector and \( \nabla \phi \cdot ds \) is the inner product of these two vectors. This identity plays the role of integration by parts in \( \mathbb{R}^n \).

To apply Green's identity to (3.13) we let \( V_r \) be the \( n \)-sphere of radius \( r \) centered at the origin and having surface \( S_r \). Then we use Green's identity on \( V_r \) and \( S_r \) with \( \phi(x_t) = \log p_t(x_t) \) and \( \psi(x_t) = p_t(x_t) \) and take the limit as \( r \to \infty \). In Appendix 4 the surface integral over \( S_r \) is shown to vanish in the limit. Hence we obtain

\[
\frac{d}{dt} H(X_t) = - \frac{1}{2} \int_{\mathbb{R}^n} \nabla p_t(x_t) \cdot \nabla \log p_t(x_t) \, dx_t = \frac{1}{2} \int_{\mathbb{R}^n} \left\| \nabla p_t(x_t) \right\|^2 \, dx_t
\]  
(3.15)

as wanted.

Eq. (3.15) can also be written as


\[
\frac{d}{dt} H(X_t) = \frac{1}{2} E \left\| \nabla P_t \right\|^2_{p_t}. \tag{3.16}
\]

3.2.2. Proof of Theorem 3.2:

Using (3.16) we have

\[
\frac{d^2}{dt^2} e^{\frac{2}{n} H(X_t)} = e^{\frac{2}{n} H(X_t)} \left( \left\{ \frac{2}{n} \frac{d}{dt} H(X_t) \right\}^2 + \frac{2}{n} \frac{d^2}{dt^2} H(X_t) \right)
\]

\[
= e^{\frac{2}{n} H(X_t)} \left( \left\{ \frac{1}{n} E \left\| \nabla P_t \right\|^2_{p_t} \right\}^2 + \frac{1}{n} \frac{d}{dt} E \left\| \nabla P_t \right\|^2_{p_t} \right). \tag{3.17}
\]

Define

\[
T_1 = E \left\| \nabla P_t \right\|^2_{p_t} \tag{3.18}
\]

\[
T_2 = \frac{d}{dt} E \left\| \nabla P_t \right\|^2_{p_t} \tag{3.19}
\]

It can be easily verified that

\[
E \nabla^2 \log P_t = E \frac{\nabla^2 P_t}{p_t} - E \left\| \nabla P_t \right\|^2_{p_t} = 0 - T_1 \tag{3.20}
\]

That the first term above equals zero follows from the diffusion equation (3.12) and

\[
\int p_t(x_t) dx_t = 1. \quad \text{This gives the following expression for } T_1:
\]

\[
T_1 = - E \sum_{i=1}^n \frac{\partial^2 \log p_t}{\partial x_{(i)}^2} \tag{3.21}
\]

Now we concentrate on term \(T_2\). Differentiating (3.19) inside the integral we get

\[
T_2 = \int_{\mathbb{R}^n} \frac{d}{dt} \frac{\left\| \nabla p_t(x_t) \right\|^2_{p_t}}{p_t(x_t)} dx_t - \frac{1}{2} \int_{\mathbb{R}^n} \frac{\left\| \nabla p_t(x_t) \right\|^2_{p_t} \nabla^2 p_t(x_t)}{p_t(x_t)} dx_t
\]
\[ T_2 = \int R^* \frac{\nabla^2 p_l(x_i) \cdot \nabla (\nabla^2 p_l(x_i))}{p_l^2(x_i)} \, dx_i + \frac{1}{2} \int R^* \frac{\nabla^2 p_l(x_i) \cdot \nabla (\nabla^2 p_l(x_i))}{p_l^2(x_i)} \, dx_i. \]
Again, the surface integral disappears in the limit as shown in Appendix 4. Using the relations

\[
\nabla \left( \frac{\| \nabla p_t \|^2}{p_t^2} \right) = \nabla \left( \frac{\| \nabla p_t \|^2}{p_t^2} \right) - 2 \frac{\| \nabla p_t \|^2 \nabla p_t}{p_t^3} \tag{3.27}
\]

and \( \nabla p_t \cdot \nabla p_t = \| \nabla p_t \|^2 \) we find

\[
\int_{\mathbb{R}^n} \frac{\nabla^2 p_t(x_t) \| \nabla p_t(x_t) \|^2}{p_t^2(x_t)} \, dx_t = -\int_{\mathbb{R}^n} \frac{\nabla p_t(x_t) \cdot \nabla (\| \nabla p_t(x_t) \|^2)}{p_t^2(x_t)} \, dx_t
\]

\[+ \frac{2}{3} \int_{\mathbb{R}^n} \frac{\| \nabla p_t(x_t) \|^4}{p_t^4(x_t)} \, dx_t. \tag{3.28}\]

Now we consider the following relation:

\[
E(\nabla^2 \log p_t)^2 = E \left( \frac{\nabla^2 p_t}{p_t} - \frac{\| \nabla p_t \|^2}{p_t^2} \right)^2
\]

\[= E \left( \frac{\nabla^2 p_t^2}{p_t^2} - 2 \frac{\nabla^2 p_t \| \nabla p_t \|^2}{p_t^3} + \frac{\| \nabla p_t \|^4}{p_t^4} \right). \tag{3.29}\]

Combining (3.25), (3.28) and (3.29) we find

\[
T_2 = -E(\nabla^2 \log p_t)^2 - E \frac{\nabla^2 p_t \| \nabla p_t \|^2}{p_t^3} + \frac{1}{2} E \frac{\nabla p_t \cdot \nabla (\| \nabla p_t \|^2)}{p_t^3}. \tag{3.30}\]

Expanding the terms in this expression yields

\[
T_2 = -E \sum_{i=1}^{n} \left( \frac{\partial^2 \log p_t}{\partial x_{(i)}^2} \right)^2 - E \sum_{i \neq j} \frac{\partial^2 \log p_t}{\partial x_{(i)}^2} \frac{\partial^2 \log p_t}{\partial x_{(j)}^2} - E \frac{1}{p_t^3} \sum_{i=1}^{n} \frac{\nabla^2 p_t}{\partial x_{(i)}^2} \left( \frac{\partial p_t}{\partial x_{(i)}^2} \right)^2
\]

\[+ \frac{1}{2} E \frac{1}{p_t^3} \sum_{i \neq j} \frac{\partial^2 p_t}{\partial x_{(i)}^2} \left( \frac{\partial p_t}{\partial x_{(i)}^2} \right)^2 + \frac{1}{2} E \frac{1}{p_t^3} \sum_{i=1}^{n} \frac{\partial^2 p_t}{\partial x_{(i)}^2} \left( \frac{\partial p_t}{\partial x_{(i)}^2} \right)^2
\]

\[+ \frac{1}{2} E \frac{1}{p_t^3} \sum_{i \neq j} \frac{\partial^2 p_t}{\partial x_{(i)}^2} \left( \frac{\partial p_t}{\partial x_{(i)}^2} \right)^2. \tag{3.31}\]
The third and fifth terms cancel out. Expanding the second term above with the aid of the relation
\[
\frac{\partial^2 \log p_t}{\partial x_{(i)t}^2} = \frac{1}{p_t^2} \left( p_t \left( \frac{\partial^2 p_t}{\partial x_{(i)t}^2} - \left( \frac{\partial p_t}{\partial x_{(i)t}} \right)^2 \right) \right)
\] (3.32)
and rearranging terms we can write (3.31) as
\[
T_2 = -E \sum_{i=1}^n \left( \frac{\partial^2 \log p_t}{\partial x_{(i)t}^2} \right)^2 - E \frac{1}{p_t^2} \sum_{i \neq j} \frac{\partial^2 p_t}{\partial x_{(i)t}^2} \frac{\partial^2 p_t}{\partial x_{(j)t}^2}
+ E \frac{1}{p_t^2} \sum_{i \neq j} \left( \frac{\partial^2 p_t}{\partial x_{(i)t}^2} \left( \frac{\partial p_t}{\partial x_{(j)t}} \right)^2 + \frac{\partial p_t}{\partial x_{(i)t}} \frac{\partial p_t}{\partial x_{(j)t}} \frac{\partial^2 p_t}{\partial x_{(i)t} \partial x_{(j)t}} \right)
- E \frac{1}{p_t^4} \sum_{i \neq j} \left( \frac{\partial p_t}{\partial x_{(i)t}} \right)^2 \left( \frac{\partial p_t}{\partial x_{(j)t}} \right)^2.
\] (3.33)
The first term in (3.33) is already in the form we need. As for the remaining terms we proceed as follows. We define
\[
T_3(i,j) = E \frac{1}{p_t^2} \frac{\partial^2 p_t}{\partial x_{(i)t}^2} \frac{\partial^2 p_t}{\partial x_{(j)t}^2}
\] (3.34)
\[
T_4(i,j) = E \frac{1}{p_t^2} \frac{\partial^2 p_t}{\partial x_{(i)t}^2} \left( \frac{\partial p_t}{\partial x_{(j)t}} \right)^2
\] (3.35)
\[
T_5(i,j) = E \frac{1}{p_t^2} \frac{\partial p_t}{\partial x_{(i)t}} \frac{\partial p_t}{\partial x_{(j)t}} \frac{\partial^2 p_t}{\partial x_{(i)t} \partial x_{(j)t}}
\] (3.36)
\[
T_6(i,j) = E \frac{1}{p_t^4} \left( \frac{\partial p_t}{\partial x_{(i)t}} \right)^2 \left( \frac{\partial p_t}{\partial x_{(j)t}} \right)^2
\] (3.37)
and the new term
\[
T_7(i,j) = E \frac{1}{p_t^2} \left( \frac{\partial^2 p_t}{\partial x_{(i)t} \partial x_{(j)t}} \right)^2.
\] (3.38)

Then, by Fubini's Theorem
\[ T_3(i,j) = \int_{\mathbb{R}^n} \frac{1}{p_l(x_i)} \frac{\partial^2 p_l(x_i)}{\partial x_{(i)t}^2} \frac{\partial^2 p_l(x_i)}{\partial x_{(j)t}^2} \, dx_i \]

\[ = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{1}{p_l(x_i)} \frac{\partial^2 p_l(x_i)}{\partial x_{(i)t}^2} \frac{\partial^2 p_l(x_i)}{\partial x_{(j)t}^2} \, dx_{(i)t} \, dx_{(j)t} \]  

\[ (3.39) \]

where \( dx_{(i)}^t = dx_{(1)t} \, dx_{(2)t} \cdots dx_{(i-1)t} \, dx_{(i+1)t} \cdots dx_{(n)t} \). There are \( n \) integrations in the above expression. We integrate (3.39) by parts with respect to \( x_{(i)t} \). Recalling the formula for integration by parts, \( \int u \, dv = uv - \int v \, du \), and letting \( u = \frac{1}{p_l} \frac{\partial^2 p_l}{\partial x_{(i)t}^2} \) and \( dv = \frac{\partial^2 p_l}{\partial x_{(j)t}^2} \), we get

\[ T_3(i,j) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left( \frac{1}{p_l(x_i)} \frac{\partial p_l(x_i)}{\partial x_{(i)t}} \frac{\partial^2 p_l(x_i)}{\partial x_{(j)t}^2} \right) \bigg|_{x_{(i)t}=-\infty}^{x_{(i)t}=\infty} \, dx_i \]

\[ - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{1}{p_l(x_i)} \frac{\partial p_l(x_i)}{\partial x_{(i)t}} \frac{\partial^2 p_l(x_i)}{\partial x_{(i)t} \partial x_{(j)t}^2} \, dx_i \]

\[ + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{1}{p_l^2(x_i)} \left( \frac{\partial p_l(x_i)}{\partial x_{(i)t}} \right)^2 \frac{\partial^2 p_l(x_i)}{\partial x_{(j)t}^2} \, dx_i . \]

\[ (3.40) \]

The first integral above is over \( n-1 \) dimensions only. Using Fatou's lemma and the technique applied in Appendix 4 (to prove that \( L_2 \) and \( L_2' \) are zero) we can show that this integral equals zero.

We recognize the third term to be \( T_4(i,j) \) defined in Eq. (3.35). Also we define

\[ T_8(i,j) = E \frac{1}{p_l^2} \frac{\partial p_l}{\partial x_{(i)t}} \frac{\partial^2 p_l}{\partial x_{(i)t} \partial x_{(j)t}^2} . \]

(3.41)

Therefore we have

\[ T_3(i,j) = -T_8(i,j) + T_4(i,j) . \]

(3.42)

Similarly, we integrate \( T_7(i,j) \) by parts with respect to \( x_{(j)t} \) with
\[ u = \frac{1}{p_t} \frac{\partial^2 p_t}{\partial x(i)_t \partial x(j)_t} \quad \text{and} \quad dv = \frac{\partial^2 p_t}{\partial x(i)_t \partial x(j)_t} \to \text{obtain} \]

\[ T_\gamma(i,j) = \int \int \ldots \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{p_t(x_i)} \frac{\partial^2 p_t(x_i)}{\partial x(i)_t \partial x(j)_t} \frac{\partial^2 p_t(x_i)}{\partial x(i)_t \partial x(j)_t} \cdot dx_{(j)_t} \cdot dx_{(j)_t} \]

\[ = \int \int \ldots \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left( \frac{1}{p_t(x_i)} \frac{\partial p_t(x_i)}{\partial x(i)_t} \right) \frac{\partial^2 p_t(x_i)}{\partial x(i)_t \partial x(j)_t} \cdot dx_{(j)_t} \]

\[ - \int \int \ldots \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{p_t(x_i)} \frac{\partial p_t(x_i)}{\partial x(i)_t} \frac{\partial^2 p_t(x_i)}{\partial x(i)_t \partial x(j)_t} \cdot dx_{(j)_t} \]

\[ + \int \int \ldots \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{p_t^2(x_i)} \frac{\partial p_t(x_i)}{\partial x(i)_t} \frac{\partial p_t(x_i)}{\partial x(j)_t} \frac{\partial^2 p_t(x_i)}{\partial x(i)_t \partial x(j)_t} \cdot dx_{(j)_t} \quad (3.43) \]

Once more, Fatou's lemma and the technique of Appendix 4 can be used to show that the \((n-1)\)-dimensional integral above vanishes. Hence we find

\[ T_\gamma(i,j) = T_\delta(i,j) + T_\delta(i,j) \quad (3.44) \]

Subtracting (3.44) from (3.42) we have

\[ T_\delta(i,j) - T_\gamma(i,j) = T_\delta(i,j) - T_\delta(i,j) \quad (3.45) \]

or

\[ T_\delta(i,j) - T_\delta(i,j) = T_\gamma(i,j) - T_\delta(i,j) \quad (3.46) \]

Here, we consider the following relation

\[ \mathbb{E} \left( \frac{\partial^2 \log p_t(x_i)}{\partial x(i)_t \partial x(j)_t} \right)^2 = \mathbb{E} \frac{1}{p_t^2} \left( \frac{\partial^2 p_t}{\partial x(i)_t \partial x(j)_t} \right)^2 - 2 \mathbb{E} \frac{1}{p_t^2} \frac{\partial p_t}{\partial x(i)_t} \frac{\partial p_t}{\partial x(j)_t} \frac{\partial^2 p_t}{\partial x(i)_t \partial x(j)_t} \]

\[ + \mathbb{E} \frac{1}{p_t^4} \left( \frac{\partial p_t}{\partial x(i)_t} \right)^2 \left( \frac{\partial p_t}{\partial x(j)_t} \right)^2 \]

\[ = T_\gamma(i,j) - 2 T_\delta(i,j) T_\delta(i,j) \quad (3.47) \]
From Eq. (3.33), by direct substitution, we can write

\[ T_2 = -E \sum_{i=1}^{n} \left( \frac{\partial^2 \log p_t}{\partial x(i)_t} \right)^2 - \sum_{i \neq j} \left[ T_3(i,j) - T_4(i,j) - T_5(i,j) + T_6(i,j) \right]. \quad (3.48) \]

Now, using (3.46) we get

\[ T_2 = -E \sum_{i=1}^{n} \left( \frac{\partial^2 \log p_t}{\partial x(i)_t} \right)^2 - \sum_{i \neq j} \left[ T_3(i,j) - 2T_5(i,j) + T_6(i,j) \right] \quad (3.49) \]

which yields the expression we need

\[ T_2 = -E \frac{1}{n} \sum_{i=1}^{n} \left( \frac{\partial^2 \log p_t(x_i)}{\partial x(i)_t} \right)^2 - E \frac{1}{n} \sum_{i \neq j} \left( \frac{\partial^2 \log p_t(x_i)}{\partial x(i)_t \partial x(j)_t} \right)^2. \quad (3.50) \]

Finally, we plug into Eq. (3.17) the expressions for \( T_1 \) and \( T_2 \) obtained in (3.21) and (3.50). With this we obtain Eq. (3.11) thereby proving Theorem 3.2.

3.2.3. Proof of Theorem 3.3:

It follows from the convexity of the square function and Jensen’s inequality (using

\[ E \frac{1}{n} \sum_{i=1}^{n} (\cdot) \] as the averaging operation) that

\[ \left( E \frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2 \log p_t(x_i)}{\partial x(i)_t} \right)^2 \leq E \frac{1}{n} \sum_{i=1}^{n} \left( \frac{\partial^2 \log p_t(x_i)}{\partial x(i)_t} \right)^2. \quad (3.51) \]

Also it is obvious that

\[ -E \frac{1}{n} \sum_{i \neq j} \left( \frac{\partial^2 \log p_t(x_i)}{\partial x(i)_t \partial x(j)_t} \right)^2 \leq 0. \quad (3.52) \]

Thus it follows from Theorem 3.2 that \( \frac{d^2}{dt^2} e^{-\frac{a}{n}H(x_t)} \leq 0 \), i.e., the entropy power of \( X_t \) is a concave function of \( t \). This proves the inequality in Theorem 3.2.

Now, let us investigate conditions for equality. Suppose the entropy power of \( X_t \) is
a linear function of $t$ so that $\frac{d^2}{dt^2} e^{\frac{2}{d}H(x_t)} = 0$. From Theorem 3.2 and the conditions for equality in Jensen's inequality it follows that

$$a) \quad \frac{\partial^2 \log p_i(x_t)}{\partial x_{(i)t}^2} \text{ is constant for all } x_t \in \mathbb{R}^n \text{ and all } i = 1, \cdots, n.$$ 

Also it is clear that we need

$$b) \quad \frac{\partial^2 \log p_i(x_t)}{\partial x_{(i)t} \partial x_{(j)t}} \equiv 0 \text{ for all } x_t \in \mathbb{R}^n \text{ and all } i, j = 1, \cdots, n \text{ such that } i \neq j.$$ 

Suppose the constant in condition a) above equals $A$. Integrating this condition twice we obtain

$$\log p_i(x_t) = A x_{(i)t}^2 + b_i(x_t^{(i)}) x_{(i)t} + c_i(x_t^{(i)}) \quad (3.53)$$

for each $i = 1, \ldots, n$, where $b_i(x_t^{(i)})$ and $c_i(x_t^{(i)})$ are arbitrary functions of $x_t^{(i)}$, the vector with all the components of $x_t$ except the $i^{th}$ one.

Now we consider condition b). Integrating with respect to $x_{(i)t}$ and $x_{(j)t}$ we have

$$\log p_i(x_t) = x_{(j)t} d_i(x_t^{(i)}) + c_i(x_t^{(j)}) \quad (3.54)$$

for all $i, j = 1, \ldots, n$ such that $i \neq j$, where $d_i(\cdot)$ and $c_i(\cdot)$ are arbitrary functions. Alternatively, we can write $p_i(x_t) = \exp(x_{(j)t} d_i(x_t^{(i)}) \exp(c_i(x_t^{(j)})$. The first exponential above does not depend on $x_{(i)t}$ and the second exponential does not depend on $x_{(j)t}$. Furthermore this relation is valid for all $i$ and $j$ in $\{1, \ldots, n\}$ such that $i \neq j$. Therefore we have $p_i(x_t) = p_1_t(x_{(1)t}) p_2_t(x_{(2)t}) \cdots p_{n_t}(x_{(n)t})$, i.e., $X_t$ has independent components. This observation implies that there are no cross terms in (3.53). By examining the forms that the functions $b_i(\cdot)$ and $c_i(\cdot)$ may take in Eq. (3.53) we can further restrict the class of densities under consideration to those satisfying
\[ \log p_t(x_t) = A \|x_t\|^2 + \sum_{i=1}^{n} B_i x_{(i)t} + B_0 \]  

(3.55)

where \( B_i, i = 0,1,...,n \) are constants. This is precisely the class of \( n \)-variate Gaussian distributions with independent and identically distributed components. Conversely, if \( X \) is Gaussian with covariance matrix proportional to the identity matrix we can easily show that the entropy power of \( X_t \) is a linear function of \( t \). Thus we have completed the proof of Theorem 3.3.

3.3. Conjecture and Special Cases

The strengthening of the entropy power inequality given in Theorem 3.1 seems to be valid for arbitrarily distributed \( Z \). We thus state the following conjecture.

Conjecture 3.1:

For \( X \) and \( Z \) independent random vectors we have

\[ e^{nH(X+\alpha Z)} \geq (1-\alpha^2)e^{nH(X)} + \alpha^2 e^{nH(X+Z)}, \quad \alpha \in [0,1] \]  

(3.56)

where equality is obtained if and only if \( X \) and \( Z \) are Gaussian random vectors with proportional covariance matrices.

We can emulate Shannon's variational argument and show that if the entropies of \( X \) and \( X+Z \) are fixed, the entropy of \( X+\alpha Z \) has a stationary point where \( X \) and \( Z \) are Gaussian vectors with proportional covariance matrices. As before, this argument does not constitute a proof of our conjecture for it does not exclude the possibility that other probability distributions might yield an equal or lower value for the entropy of \( X+\alpha Z \).

In this section we demonstrate Conjecture 3.1 under some special cases. First,
suppose $X$ is a Gaussian vector. For simplicity, set $EX = 0$. Let $X'$, independent of $X$ and $Z$, be distributed according to the same multivariate normal as $X$. Then, for $\alpha \in (0,1]$, the vectors $\frac{X}{\alpha}$ and $\frac{\sqrt{1-\alpha^2}X'}{\alpha} + X$ have identical distributions, since they both have multivariate normal distributions with identical means and covariance matrices. Hence, we can write

$$e^{\frac{1}{n}H(X+\alpha Z)} = \alpha^2 e^{\frac{1}{n}H(X^+Z)}$$

$$= \alpha^2 e^{\frac{1}{n}H(\frac{\sqrt{1-\alpha^2}X'}{\alpha} + X + Z)}$$

$$\leq \alpha^2 e^{\frac{1}{n}H(\frac{\sqrt{1-\alpha^2}X'}{\alpha})} + \alpha^2 e^{\frac{1}{n}H(X+Z)}$$

$$= (1-\alpha^2)e^{\frac{1}{n}H(X)} + \alpha^2 e^{\frac{1}{n}H(X+Z)},$$

(3.57)

thus establishing the conjectured inequality (3.56). The only inequality above is simply the entropy power inequality applied to the independent random variables $\frac{\sqrt{1-\alpha^2}X'}{\alpha}$ and $X+Z$. We have already shown in the statement following Eq. (3.7) that (3.56) implies the entropy power inequality. We have thus proven the following:

**Proposition 3.1:**

If $X$ is normally distributed, Conjecture 3.1 is true and is equivalent to the entropy power inequality.

As a second case, let $X$ and $Z$ be univariate independent random variables satisfying the same symmetric stable law with exponent $\alpha$. Stable laws [25], [26] form a class of probability distributions discovered by Levy circa 1923 having a characteristic
function of the form \( e^{-\gamma|\xi|^\alpha} \), where \( \alpha \in (0,2] \) and \( \gamma \) is a constant. Symmetric stable laws are those where \( \gamma \) is a positive real constant. Typical examples are the normal distribution (\( \alpha=2 \)) and the Cauchy distribution (\( \alpha=1 \)). They have the following property: if \( X \) and \( X' \) are two independent random variables identically distributed according to a symmetric stable law of exponent \( \alpha \), then \( k_1 X + k_2 X' \) has the same distribution as \((k_1^\alpha + k_2^\alpha)^{1/\alpha} X\), for any positive constants \( k_1 \) and \( k_2 \). Now, suppose the distributions of \( X \) and \( Z \) have characteristic functions \( e^{-\gamma_1|\xi|^\alpha} \) and \( e^{-\gamma_2|\xi|^\alpha} \), respectively. Then, for \( \epsilon = e^{(\gamma_1 - \gamma_2)} \) we have that \( \epsilon X \) and \( Z \) are independent identically distributed random variables. Therefore, letting \( X \) and \( X' \) be independent and identically distributed, we find

\[
\begin{align*}
f(a) &= e^{2H(aX+Z)} - e^{2H(aX)} \\
&= e^{2H(aX+\epsilon X')} - e^{2H(aX)} \\
&= e^{2H((a^\alpha + \epsilon^\alpha)^{1/\alpha} X)} - e^{2H(aX)} \\
&= (a^\alpha + \epsilon^\alpha)^{2/\alpha} - a^2 e^{2H(X)} 
\end{align*}
\]  

(3.58)

From this point it is trivial to show that \( f'(a) > 0 \) for \( \alpha \in (0,2) \) and \( f'(a) = 0 \) for \( \alpha = 2 \). Therefore, since \( \alpha = 2 \) corresponds to the normal distribution, \( f(a) \) is a strictly increasing function except when \( X \) and \( Z \) are normally distributed, in which case \( f(a) \) is a constant. As we have seen in the discussion around Eq. (3.7), this suffices to establish our conjecture for symmetric stable laws.

Now we would like to extend this scalar relation to the vector case. Unfortunately, we cannot prove this extension when the vectors components are dependent random variables. Suppose \( X \) and \( Z \) have independent components so that \( H(X) = \sum_{i=1}^n H(X_i) \) and \( H(Z) = \sum_{i=1}^n H(Z_i) \). Also suppose their components satisfy (3.56), i.e.,
\[ e^{2H(X_i+\alpha Z_i)} \geq (1-\alpha^2)e^{2H(X_i)} + \alpha^2 e^{2H(X_i+Z_i)}, \]

(3.59)

for \( \alpha \in [0,1], \ i=1,\ldots,n \). We want to obtain the conjectured vector inequality (3.56) from these assumptions. The cases \( \alpha=0 \) or 1 are trivial. For \( \alpha \in (0,1) \), the function \( \log((1-\alpha^2)e^u + \alpha^2 e^{-u}) \) is a strictly convex function of \( u \) since its second derivative equals \( \frac{4\alpha^2(1-\alpha^2)}{((1-\alpha^2)e^u + \alpha^2 e^{-u})^2} > 0 \). Letting \( v_i = H(X_i) + H(X_i+Z_i) \) and \( u_i = H(X_i) - H(X_i+Z_i) \) and using the assumption (3.59) and Jensen’s Inequality, we have

\[
\frac{2}{n} \sum_{i=1}^{n} H(X_i+\alpha Z_i) \geq \frac{1}{n} \sum_{i=1}^{n} \log((1-\alpha^2)e^{u_i} + \alpha^2 e^{-u_i}) + \frac{1}{n} \sum_{i=1}^{n} v_i
\]

\[
\geq \log((1-\alpha^2)e^{\frac{1}{n} \sum_{i=1}^{n} u_i} + \alpha^2 e^{-\frac{1}{n} \sum_{i=1}^{n} u_i}) + \frac{1}{n} \sum_{i=1}^{n} v_i
\]

\[
= \log((1-\alpha^2)e^{\sum_{i=1}^{n} H(X_i)} + \alpha^2 e^{\sum_{i=1}^{n} H(X_i+Z_i)})
\]

(3.60)

as desired. Equality in (3.60) occurs only if \( X_i, Z_i, i=1,\ldots,n \), are Gaussian random variables and \( u_i = H(X_i) - H(X_i+Z_i) \) is constant for \( i=1,\ldots,n \). This requires that \( X \) and \( Z \) have proportional covariance matrices. Another way to arrive at (3.60) from our assumptions is to use a property of the geometrical mean listed as Theorem 185 in [27].

From the above discussion we can state:

**Proposition 3.2:**

Conjecture 3.1 is true for \( X \) and \( Z \) independent random vectors having independent components such that \( X_i \) and \( Z_i \) are distributed according to the same stable law of exponent \( \alpha_i, i=1,\ldots,n \).
3.4. Similarity with the Brunn Minkowski Inequality

In this section we present an intriguing similarity between the entropy power inequality and the Brunn Minkowski inequality in geometry. We recast both inequalities in a form that enhances this similarity. As a byproduct of the discussion we prove an isoperimetric inequality for entropy, namely the fact that the normal distribution with independent identically distributed components minimizes the trace of the Fisher Information matrix of the translation family of distributions given an entropy constraint; just as a sphere minimizes surface area given a volume constraint.

The entropy power inequality states that the effective variance (entropy power) of the sum of two independent random variables is greater than the sum of their effective variances. In mathematical terms we write

\[
\frac{e^n}{2\pi e} \frac{\mathbb{E} H(X+Z)}{2\pi e} \geq e^n \frac{\mathbb{E} H(X)}{2\pi e} + e^n \frac{\mathbb{E} H(Z)}{2\pi e}.
\]  

(3.61)

To recast this inequality we observe that a normal multivariate random variable \( Y \) with independent identically distributed components of variance \( \sigma^2 \) has entropy

\[
H(Y) = \frac{n}{2} \log(2\pi e\sigma^2).
\]  

(3.62)

By inverting, we see that if \( Y \) is normal with independent and identically distributed components and entropy \( H(Y) \), then the variance of its components is

\[
\sigma^2 = \frac{e^n}{2\pi e} \frac{\mathbb{E} H(Y)}{2\pi e}.
\]  

(3.63)

Thus the entropy power inequality is an inequality between effective variances and we can recast (3.61) in the following equivalent form:

\[
H(X + Z) \geq H(X' + Z')
\]  

(3.64)
where $X'$ and $Z'$ are independent normal variables with covariance matrices proportional to the identity matrix and corresponding entropies

$$H( X' ) = H( X )$$

$$H( Z' ) = H( Z )$$  \hspace{1cm} (3.65)

Verification of this statement follows from the use of (3.63) in (3.61) and the fact that $X' + Z'$ is normal and has covariance matrix proportional to the identity matrix.

Now we turn to the Brunn Minkowski inequality. Let $A$ and $B$ be two measurable sets in $\mathbb{R}^n$. The set sum $C = A + B$ of these sets may be written

$$C = \{ x + z : x \in A, z \in B \}.$$  \hspace{1cm} (3.66)

Let $V(A)$ denote the volume of $A$. The Brunn Minkowski inequality states that the effective radius of the set sum of two sets is greater than the sum of their effective radii, or in mathematical terms,

$$\frac{1}{V^n( A + B )} \geq \frac{1}{V^n( A )} + \frac{1}{V^n( B )}.$$  \hspace{1cm} (3.67)

To recast this inequality we observe that an $n$-sphere $S$ with radius $r$ has volume $V(S) = c_n r^n$. Thus if $S$ is a sphere of volume $V(S)$, its radius is $r = (V(S)/c_n)^{1/n}$. Hence the Brunn Minkowski inequality can be viewed as an inequality between the radii of the spherical equivalents of the sets. Rewriting, we have the following restatement of the Brunn Minkowski inequality:

$$V( A + B ) \geq V( A' + B' )$$  \hspace{1cm} (3.68)

where $A'$ and $B'$ are spheres with volumes $V(A') = V(A)$ and $V(B') = V(B)$.

Inspite of the obvious similarity between the above inequalities, there is no apparent similarity of any of the various known proofs of the Brunn Minkowski
inequality [28], [29] with the Stam and Blachman proofs of the entropy power inequality, nor have we succeeded in finding a new common proof. Nevertheless, the striking similarity of these inequalities suggests that we may find new results relating to entropies from known results in geometry and vice versa. We present two applications of this reasoning.

3.4.1. On the Isoperimetric Inequality

It is known that the sphere minimizes surface area for given volume. A proof follows immediately from the Brunn Minkowski inequality [30] as shown below. For "regular" sets $A$, the surface area $S(A)$ of $A$ is given by

$$S(A) = \lim_{\epsilon \to 0} \frac{V(A + S_\epsilon) - V(A)}{\epsilon}$$

(3.69)

where $S_\epsilon$ is a sphere of radius $\epsilon > 0$. Using the Brunn Minkowski inequality we have

$$S(A) \geq \lim_{\epsilon \to 0} \frac{V(A' + S'_\epsilon) - V(A')}{\epsilon} = S(A'),$$

(3.70)

where $A'$ is a sphere with volume $V(A') = V(A)$. Of course, $S'_\epsilon = S_\epsilon$ and $A' + S'_\epsilon$ is a sphere. Thus the surface area of $A$ is greater than that of a sphere with the same volume.

We now proceed to perform the same steps on the entropy power. Noting that $e^H$, like $V$, is a measure of volume, we define $S(X)$, the "surface area" of a multivariate random variable $X$, by

$$S(X) = \lim_{\epsilon \to 0} \frac{e^{H(X + Z_\epsilon)} - e^{H(X)}}{\epsilon},$$

(3.71)

where $Z_\epsilon$ is Gaussian with covariance matrix $\epsilon I$. Thus $S(X)$ is the rate of change of the "volume" $e^H$ when a small normal random variable is added. Using our notation in the previous sections we may write
\[ S(\mathbf{X}) = \lim_{t_0 \to 0} \left[ \frac{d}{dt} e^{H(\mathbf{X} + Z_t)} \right]_{t = t_0}. \] (3.72)

Assuming \( \mathbf{X} \) has a reasonably smooth density \( p(x) \) and using Eq. (3.16) we have

\[ S(\mathbf{X}) = \frac{1}{2} \mathbb{E} \frac{\| \nabla p \|^2}{p^2} e^{H(\mathbf{X})}. \] (3.73)

Also, by the recasted form of the entropy power inequality we have

\[ e^{H(\mathbf{X} + Z_t)} \geq e^{H(\mathbf{X'}) + Z_t)}, \] (3.74)

where \( \mathbf{X'} \) is a Gaussian \( n \)-vector as before, with entropy \( H(\mathbf{X'}) = H(\mathbf{X}) \). Thus, from (3.71) and (3.74), we obtain the following bound on \( S(\mathbf{X}) \):

\[ S(\mathbf{X}) \geq \lim_{\epsilon \to 0} \frac{e^{H(\mathbf{X} + Z_t)} - e^{H(\mathbf{X'})}}{\epsilon} = S(\mathbf{X'}) \] (3.75)

For the Gaussian vector \( \mathbf{X'} \), the "surface area" is found from (3.73) to be

\[ S(\mathbf{X'}) = \frac{n \pi e}{e^n} \frac{e^{H(\mathbf{X'})}}{e^{H(\mathbf{X'})}}. \] (3.76)

Finally, recalling that \( H(\mathbf{X'}) = H(\mathbf{X}) \), we combine (3.73), (3.75) and (3.76) to obtain the entropy analog of the isoperimetric inequality:

\[ \frac{1}{n} \mathbb{E} \frac{\| \nabla p \|^2}{p^2} \geq \left( \frac{\frac{1}{2} e^{H(\mathbf{X})}}{2 \pi e} \right)^{-1}. \] (3.77)

This inequality establishes that the Gaussian distribution minimizes the trace of the Fisher Information matrix given an entropy constraint. The scalar version of the above relation was proved in [21].
3.4.2. More on Concavity

Here we present an example of how a result involving entropy may yield a conjecture in geometry. The concavity of entropy power with added Gaussian noise suggests the following conjecture.

Conjecture 3.2:

Let \( A \) be an arbitrary measurable set and \( S \) a sphere of unit radius in \( \mathbb{R}^n \). Then

\[
\frac{1}{V^n(A + tS)} \text{ is a concave function of } t \geq 0.
\]

If \( A \) is a convex set, a proof goes as follows. Let \( \theta \in [0,1] \). Then we have

\[
\frac{1}{V^n(A + \theta S)} = \frac{1}{V^n((1-\theta)A + \theta A + \theta S)}
\]

\[
\geq \frac{1}{V^n((1-\theta)A)} + \frac{1}{V^n(\theta A + \theta S)}
\]

\[
= (1-\theta)\frac{1}{V^n(A)} + \theta \frac{1}{V^n(A + S)}, \tag{3.78}
\]

where we used the convexity of the set \( A \) in the first line and the Brunn Minkowski inequality in the second line. As \( A \) is an arbitrary convex set, this completes the proof. As far as we know, for arbitrary measurable sets in \( \mathbb{R}^n \), this problem is still open.
4. On Discrete Interference Channels

4.1. Introduction

In this chapter we establish the capacity region of a class of deterministic discrete memoryless interference channels. We then partially relax the determinism, to obtain the capacity region of a class of Z-interference channels and illustrate it with an example.

Discrete interference channels are those where the input and output alphabets are finite sets given by \( \mathcal{X}_1 = \{1, \cdots, I\} \), \( \mathcal{X}_2 = \{1, \cdots, J\} \), \( \mathcal{Y}_1 = \{1, \cdots, K\} \) and \( \mathcal{Y}_2 = \{1, \cdots, L\} \). These channels are described by two marginal probability distributions \( p(y_1 | x_1, x_2) \) and \( p(y_2 | x_1, x_2) \), where \( x_1 \in \mathcal{X}_1 \), \( x_2 \in \mathcal{X}_2 \), \( y_1 \in \mathcal{Y}_1 \) and \( y_2 \in \mathcal{Y}_2 \).

The assumption of memorylessness means that for vectors 
\[ x_1 = (x_{(1)1}, x_{(2)1}, \cdots, x_{(n)1}), \quad x_2 = (x_{(1)2}, x_{(2)2}, \cdots, x_{(n)2}), \quad y_1 = (y_{(1)1}, y_{(2)1}, \cdots, y_{(n)1}) \]
and 
\[ y_2 = (y_{(1)2}, y_{(2)2}, \cdots, y_{(n)2}), \]
we have

\[
p(y_1 | x_1, x_2) = \sum_{i=1}^{n} p(y_{(i)1} | x_{(i)1}, x_{(i)2})
\]

(4.1)

and

\[
p(y_2 | x_1, x_2) = \sum_{i=1}^{n} p(y_{(i)2} | x_{(i)1}, x_{(i)2}).
\]

(4.2)

The capacity region of the discrete memoryless interference channel has only been found for very few cases. They can be summarized in the following two groups:

1) a class of additive degraded discrete interference channels [10];

2) the class of discrete interference channels with strong interference, i.e., those satisfying
On Discrete Interference Channels

\[ I(X_1; Y_1 | X_2) \leq I(X_1; Y_2 | X_2) \]

and

\[ I(X_2; Y_2 | X_1) \leq I(X_2; Y_1 | X_1) \]

for all product probability distributions on \( X_1 \times X_2 \).\(^\dagger\)

The capacity region of the second class above was conjectured by Sato [12]. A proof of Sato's conjecture follows from Proposition 1 in [14]. The class of discrete interference channels with strong interference includes two classes of channels for which the capacity regions were separately obtained. These are:

a) channels with statistically equivalent outputs [2], [3], [6];

b) the class of channels with very strong interference, i.e., those for which

\[ I(X_1; Y_1 | X_2) \leq I(X_1; Y_2) \quad \text{and} \quad I(X_2; Y_2 | X_1) \leq I(X_2; Y_1) \]

for all product probability distributions on \( X_1 \times X_2 \) [7], [9].

### 4.2. A Class of Deterministic Interference Channels

The model under investigation is shown in Fig. 4.1. The outputs \( Y_1 \) and \( Y_2 \) and the interferences \( V_1 \) and \( V_2 \) are (deterministic) functions of the inputs \( X_1 \) and \( X_2 \)

\[ Y_1 = f_1(X_1, V_2), \quad (4.3) \]

\[ Y_2 = f_2(X_2, V_1), \quad (4.4) \]

\[ V_1 = g_1(X_1), \quad (4.5) \]

\[ V_2 = g_2(X_2), \quad (4.6) \]

\(^\dagger\) The solution of this case can be easily modified to yield the capacity region of discrete Z-interference channels with strong interference.
where \( f_1(\cdot, \cdot) \) and \( f_2(\cdot, \cdot) \) satisfy the conditions

\[
H(Y_1 | X_1) = H(V_2)
\] (4.7)

and

\[
H(Y_2 | X_2) = H(V_1)
\] (4.8)

for all product probability distributions on the inputs \( X_1 \times X_2 \).

These conditions are equivalent to requiring the existence of functions \( h_1(\cdot, \cdot) \) and \( h_2(\cdot, \cdot) \) such that \( V_2 = h_1(X_1, Y_1) \) and \( V_1 = h_2(X_2, Y_2) \).

Fig. 4.1. Class of discrete interference channels under study.

The model includes two independent and uniformly distributed sources, one at sender 1 producing an integer \( W_1 \in M_1 = \{1, \cdots, M_1\} \) and the other at sender 2 producing an integer \( W_2 \in M_2 = \{1, \cdots, M_2\} \). Encoder 1 maps \( W_1 \) into
\( X = (X_{(1)}, \cdots, X_{(n)}) \) and encoder 2 maps \( W_2 \) into \( X_2 = (X_{(1)}^2, \cdots, X_{(n)}^2) \). The channel itself consists of four finite alphabets \( \mathcal{X}_1, \mathcal{X}_2, \mathcal{Y}_1, \mathcal{Y}_2 \) and four deterministic functions in agreement with (4.3) - (4.8).

Remark: We note that if \( g_1(\cdot) \) is one-to-one, it will also be onto and, therefore, invertible. Then \( X_1 \) and \( V_1 \) would have identical entropies and we could, without loss of generality, disregard \( g_1(\cdot) \) or assume it to be the identity function. It is only when \( g_1(\cdot) \) is not invertible that it is of interest to us. The same, of course, is valid for \( g_2(\cdot) \).

An \((M_1, M_2, n)\) code for this channel is a set of two encoding functions

\[
e_1 : M_1 \rightarrow \mathcal{X}_1^n
\]  \hspace{1cm} (4.9)

\[
e_2 : M_2 \rightarrow \mathcal{X}_2^n
\]  \hspace{1cm} (4.10)

and two decoding functions

\[
d_1 : \mathcal{Y}_1^n \rightarrow M_1
\]  \hspace{1cm} (4.11)

\[
d_2 : \mathcal{Y}_2^n \rightarrow M_2.
\]  \hspace{1cm} (4.12)

Define the (average) error probabilities by

\[
p_{e,1}^n = \frac{1}{M_1 M_2} \sum_{x_{1,1}, x_{1,2}} P \{ d_1(Y_1) \neq w_1 | W_1 = w_1, W_2 = w_2 \},
\]  \hspace{1cm} (4.13)

\[
p_{e,2}^n = \frac{1}{M_1 M_2} \sum_{x_{2,1}, x_{2,2}} P \{ d_2(Y_2) \neq w_2 | W_1 = w_1, W_2 = w_2 \}
\]  \hspace{1cm} (4.14)

and

\[
\max \{ p_{e,1}^n, p_{e,2}^n \} = p^*_e.
\]  \hspace{1cm} (4.15)
A rate pair \((R_1, R_2)\) is said to be achievable if there is a sequence of 
\((2^{nR_1}, 2^{nR_2}, n)\) codes with \(p_e^n \to 0\) as \(n \to \infty\). As before, the capacity region is defined as the closure of the set of all achievable rate pairs. The capacity region of the channel under study is given in the following theorem:

**Theorem 4.1:**

Let \(C\) denote the capacity region of the channel of Fig. 4.1 satisfying conditions (4.7) and (4.8). Then \(C\) equals the union of the set of all rate pairs \((R_1, R_2)\) satisfying

\[
R_1 \leq H(Y_1 | V_2), \quad (4.16)
\]

\[
R_2 \leq H(Y_2 | V_1), \quad (4.17)
\]

\[
R_1 + R_2 \leq H(Y_1 | V_1, V_2) + H(Y_2), \quad (4.18.a)
\]

\[
R_1 + R_2 \leq H(Y_1) + H(Y_2 | V_1, V_2), \quad (4.18.b)
\]

\[
R_1 + R_2 \leq H(Y_1 | V_1) + H(Y_2 | V_2), \quad (4.18.c)
\]

\[
2R_1 + R_2 \leq H(Y_1) + H(Y_1 | V_1, V_2) + H(Y_2 | V_2), \quad (4.19)
\]

\[
R_1 + 2R_2 \leq H(Y_1 | V_1) + H(Y_2) + H(Y_2 | V_1, V_2) \quad (4.20)
\]

over all product probability distributions on the inputs.

**Proof:**

i) Achievability: This part follows easily from the inclusion of \(C\) in the region of Han & Kobayashi [11].

ii) Converse: First we note that from the concavity of the entropy function, it
follows that $C$ is a convex region.

Constraints (4.16) and (4.17) are identical to the ones in the usual outer bound for the general interference channel [6], thus requiring no proof.

We proceed to prove constraints (4.18) - (4.20). From Fano’s inequality, we have

\[
H(W_1 \mid Y_1) \leq n R_1 p^n_{e,1} + h(p^n_{e,1}) = n \epsilon_{1n},
\]

(4.21)

\[
H(W_2 \mid Y_2) \leq n R_2 p^n_{e,2} + h(p^n_{e,2}) = n \epsilon_{2n}
\]

(4.22)

where $h(\cdot)$ is the binary entropy function and $\epsilon_{1n}, \epsilon_{2n} \to 0$ as $p^n_{e} \to 0$.

Now consider

\[
n(R_1 + R_2) = H(W_1) + H(W_2)
\]

\[
= I(W_1; Y_1) + I(W_2; Y_2) + H(W_1 \mid Y_1)
\]

\[
+ H(W_2 \mid Y_2).
\]

(4.23)

Substituting from (4.21) and (4.22) we get

\[
n(R_1 + R_2) \leq I(W_1; Y_1) + I(W_2; Y_2) + n(\epsilon_{1n} + \epsilon_{2n})
\]

\[
\leq I(X_1; Y_1) + I(X_2; Y_2) + n(\epsilon_{1n} + \epsilon_{2n})
\]

\[
\leq I(X_1; V_1 Y_1 \mid V_2) + H(Y_2) - H(Y_2 \mid X_2) + n(\epsilon_{1n} + \epsilon_{2n})
\]

\[
= I(X_1; V_1 \mid V_2) + I(X_1; Y_1 \mid V_1 V_2) + H(Y_2) - H(V_1)
\]

\[
+ n(\epsilon_{1n} + \epsilon_{2n})
\]

\[
= H(Y_1 \mid V_1 V_2) + H(Y_2) + n(\epsilon_{1n} + \epsilon_{2n})
\]
\[
\sum_{i=1}^{n} \left[ H( Y_{(i)1} \ | \ V_{(i)1} V_{(i)2} ) + H( Y_{(i)2} ) \right] + n(\epsilon_{1n} + \epsilon_{2n}). \tag{4.24}
\]

In the development above we have made use of assumptions (4.7), (4.8) and of the independence between \((X_1, V_1)\) and \((X_2, V_2)\).

In a completely analogous way, we obtain

\[
n( R_1 + R_2 ) \leq \sum_{i=1}^{n} \left[ H( Y_{(i)1} ) + H( Y_{(i)2} | V_{(i)1} V_{(i)2} ) \right] + n(\epsilon_{1n} + \epsilon_{2n}). \tag{4.25}
\]

Using (4.23) once more, we get

\[
n( R_1 + R_2 ) \leq I( X_1 ; Y_1 ) + I( X_2 ; Y_2 ) + n(\epsilon_{1n} + \epsilon_{2n})
= H( Y_1 ) - H( V_1 ) + H( Y_2 ) - H( V_2 ) + n(\epsilon_{1n} + \epsilon_{2n})
\leq H( Y_1 ) - I( V_1 ; Y_1 ) + H( Y_2 ) - I( V_2 ; Y_2 ) + n(\epsilon_{1n} + \epsilon_{2n})
= H( Y_1 | V_1 ) + H( Y_2 | V_2 ) + n(\epsilon_{1n} + \epsilon_{2n})
\leq \sum_{i=1}^{n} \left[ H( Y_{(i)1} | V_{(i)1} ) + H( Y_{(i)2} | V_{(i)2} ) \right] + n(\epsilon_{1n} + \epsilon_{2n}). \tag{4.28}
\]

Next consider

\[
n( 2R_1 + R_2 ) \leq 2H( W_1 ) + H( W_2 ).
\]

Using (4.21) and (4.22) we obtain

\[
n( 2R_1 + R_2 ) \leq 2I( W_1 ; Y_1 ) + I( W_2 ; Y_2 ) + n(\epsilon_{1n} + \epsilon_{2n})
\leq I( X_1 ; Y_1 ) + I( X_1 ; Y_1 V_1 | V_2 ) + I( X_2 ; Y_2 ) + n(\epsilon_{1n} + \epsilon_{2n})
= H( Y_1 ) - H( V_2 ) + H( Y_1 | V_1 V_2 ) + H( Y_2 ) + n(\epsilon_{1n} + \epsilon_{2n})
\leq H( Y_1 ) + H( Y_1 | V_1 V_2 ) + H( Y_2 | V_2 ) + n(\epsilon_{1n} + \epsilon_{2n})
\]
\[ \leq \sum_{i=1}^{n} \left[ H(Y(i)_{1}) + H(Y(i)_{1} \mid V_{i}) + H(Y(i)_{2} \mid V_{i}) \right] \\
+ n(\epsilon_{1n} + \epsilon_{2n}). \quad (4.27) \]

Analogously

\[ n(R_{1} + 2R_{2}) \leq \sum_{i=1}^{n} \left[ H(Y(i)_{2} \mid V_{i}) + H(Y(i)_{2}) + H(Y(i)_{2} \mid V_{i}) \right] \\
+ n(\epsilon_{1n} + \epsilon_{2n}). \quad (4.28) \]

Now, letting \( n \to \infty \), we combine (4.24) - (4.28) and the convexity of \( C \) to conclude that there is a product probability distribution on \( X_{1} \times X_{2} \) for which (4.16) - (4.20) hold. This completes the proof of Theorem 4.1.

4.3. Extension

If we make \( V_{2} \equiv 0 \) and allow \( Y_{1} \) to depend randomly on \( X_{1} \), according to \( p(y_{1} \mid x_{1}) \), we get the Z-interference channel shown in Fig. 4.2. If condition (4.8) still holds for \( f_{2}(\cdot, \cdot) \) we can extend the result of Theorem 4.1 to obtain the capacity of this channel.

Theorem 4.2:

The capacity region of the channel of Fig. 4.2 satisfying (4.8) is given by the union of all pairs \((R_{1}, R_{2})\) for which

\[ R_{1} \leq I(X_{1}; Y_{1}), \quad (4.29) \]

\[ R_{2} \leq H(Y_{2} \mid V_{1}), \quad (4.30) \]
\[ R_1 + R_2 \leq I(X_1; Y_1 | V_1) + H(Y_2) \]  

(4.31)

over all product probability distributions on \(X_1 \times X_2\).

**Fig. 4.2.** A class of Z-interference channels.

**Proof:**


ii) Converse: Constraints (4.29) and (4.30) follow immediately. Application of Fano's inequality and the data processing inequality yields

\[
n(R_1 + R_2) = H(W_1) + H(W_2) 
\]

\[ \leq I(X_1; Y_1) + I(X_2; Y_2) + n(\epsilon_1 + \epsilon_2) \]

\[ \leq I(X_1; V_1 Y_1) + H(Y_2) - H(V_1) + n(\epsilon_1 + \epsilon_2) \]
\[ I(X_1; Y_1 | V_1) + H(Y_2) + n(\varepsilon_{1n} + \varepsilon_{2n}) \]
\[ \leq \sum_{i=1}^{n} \left[ I(X_{(i)1}; Y_{(i)1} | V_{(i)1}) + H(Y_{(i)2}) \right] + n(\varepsilon_{1n} + \varepsilon_{2n}). \quad (4.32) \]

It is easy to show the convexity of the region in this theorem. From this convexity and (4.32) the converse follows.

Example:

Fig. 4.3 shows an example of this one-sided interference channel. Note that if \( \varepsilon = 0 \), perfect communication (i.e., 1 bit per channel use) can be achieved by both sender-receiver pairs if sender 1 restricts his alphabet to the set \{0, 2\}. On the other hand, if \( \varepsilon = 1 \) the effect of interference becomes devastating and the capacity region is the triangle achieved by time-sharing between rate pairs (0, 1) and (1, 0). The capacity region of this channel when \( \varepsilon = .1 \) is shown in Fig. 4.4.

![Diagram](image-url)

Fig. 4.3. Example of interference channel in Fig. 4.2.
4.4. Discussion

We can interpret conditions (4.7) and (4.8) as a requirement that each receiver, having decoded the message sent by his sender, will be able to know exactly the interference caused by the other sender. This observation follows immediately by noting that conditions (4.7) and (4.8) are equivalent to requiring the functions $f_1(x_1, \cdot)$ in (4.3) and $f_2(x_2, \cdot)$ in (4.4) to be one-to-one mappings for each $x_1 \in X_1$ and each $x_2 \in X_2$.

In general, a receiver will not be able to decode the message addressed to the other receiver. In this respect, the class of channels studied here differs from all the interference channels for which capacity regions have been established. For those channels, at least one of the receivers is sure to decode the message interfering in his communication.
Appendix 1: Equivalence between channels

We say that two interference channels are equivalent if they have identical capacity regions. In this appendix we prove the equivalence between two Gaussian interference channels. These channels are pictured in Figs. 2.7 (iii) and (iv).

Suppose \(e^{nR_1}\) messages \(W_1\) are encoded into a vector \(X_1\) with power not exceeding \(P_1\). Independently, \(e^{nR_2}\) messages \(W_2\) are encoded into a vector \(X_2\) with power no greater than \(P_2\). Let \(Z_1\) and \(Z_2\) be Gaussian random vectors with independent, identically distributed components.

The first interference channel of our consideration has output vectors \(Y_1\) and \(Y_2\) given by

\[
Y_1 = X_1 + Z_1 \\
Y_2 = X_1 + X_2 + Z_1 + Z_2. \tag{A1.1}
\]

Having observed the output sequence \(Y_1\), a decoder produces an estimate of the message \(W_1\). A separate decoder observes \(Y_2\) and produces an estimate of the message \(W_2\).

The second channel we consider is known as the degraded Gaussian interference channel. Its outputs \(Y'_1\) and \(Y'_2\) are given by

\[
Y'_1 = X_1 + X_2 + Z_1 \\
Y'_2 = X_1 + X_2 + Z_1 + Z_2. \tag{A1.2}
\]

From knowledge of \(Y'_1\), a decoder gives an estimate of the message \(W_1\), and from \(Y'_2\), a separate decoder gives an estimate of the message \(W_2\).

First we note that these two channels have identical second outputs. Therefore,
their second decoders may use the same decoding rule and achieve the same error performance.

Next we observe that the capacity region of the first channel must contain the capacity region of the second channel. The reason is that, for the sake of decoding $W_1$, the channel output vector $Y_1$ is cleaner than $Y_1'$.

Therefore, to establish the equivalence between these channels, it suffices to show that if a rate pair is achievable† by the first channel, it is also achievable by the second channel.

Using an argument given in [31] we can show that the capacity region of either channel does not decrease if the encoding rule that maps $W_2$ into $X_2$ is restricted to the class of deterministic rules. Thus we may assume that $H(X_2 \mid W_2) = 0$.

Suppose a code is used with rate pair in the capacity region of the first channel. Then by Fano's inequality,

$$H(W_1 \mid Y_1) \leq n\epsilon_{1n}$$

$$H(W_2 \mid Y_2) \leq n\epsilon_{2n}$$

(A1.3)

where $\epsilon_{1n}$ and $\epsilon_{2n}$ can be made arbitrarily small with increasing $n$. We show that $H(W_1 \mid Y_1') \leq n(\epsilon_{1n} + \epsilon_{2n})$.

It is clear that $H(W_2 \mid Y_2') \leq n\epsilon_{2n}$ and that $H(W_2 \mid Y_1') \leq n\epsilon_{2n}$. Expanding $H(W_1, W_2 \mid Y_1')$ by the chain rule, in two different ways, we find the following relation:

$$H(W_1 \mid Y_1') = H(W_2 \mid Y_1') + H(W_1 \mid W_2, Y_1') - H(W_2 \mid W_1, Y_1')$$

$$\leq n\epsilon_{2n} + H(W_1 \mid W_2, Y_1')$$

(A1.4)

† We use the definitions of rate pairs and achievability given in Chapter 2.
Using the chain rule again to expand $H(W_1, X_2 \mid W_2, Y_1')$, we obtain

$$H(W_1 \mid W_2, Y_1') \leq H(X_2 \mid W_2, Y_1') + H(W_1 \mid X_2, W_2, Y_1')$$

$$\leq H(W_1 \mid Y_1)$$

$$\leq n\epsilon_{1n} \quad (A1.5)$$

In the sequence of inequalities above we made use of the relation $Y_1 = Y_1' - X_2$ and of the fact that $X_2$ is a deterministic function of $W_2$. Combining (A1.4) and (A1.5) we get the desired inequality: $H(W_1 \mid Y_1') \leq n(\epsilon_{1n} + \epsilon_{2n})$. From this it follows that the rate pair in use is also in the capacity region of the second channel, completing the proof.
Appendix 2: From almost Gaussian to Gaussian

Recall the notation introduced in Chapter 2. $Y_2$ and $Y_3$ are $n$-dimensional random vectors given by

$$Y_2 = X_1 + Z_1 + Z_2 + X_2$$
$$Y_3 = X_1 + Z_1 + Z_2 + Z_3,$$ \hspace{0.5cm} (A2.1)

where $X_1$ and $X_2$ are input vectors to the Gaussian interference channel shown in Fig. 2.8 and $Z_1$, $Z_2$ and $Z_3$ are zero-mean Gaussian vectors with covariance matrices equal to $I$, $N_2I$ and $Q_2I$, respectively. In Chapter 2, power constraints were imposed on $X_1$ and $X_2$ such that

$$\frac{1}{n} \sum_{i=1}^{n} x_{(i)1}^2 \leq P_1$$
$$\frac{1}{n} \sum_{i=1}^{n} x_{(i)2}^2 \leq Q_2$$ \hspace{0.5cm} (A2.2)

for all codewords $x_1 = (x_{(1)1}, x_{(2)1}, \ldots, x_{(n)1})$ and $x_2 = (x_{(1)2}, x_{(2)2}, \ldots, x_{(n)2})$. In this appendix, we allow arbitrary probability distributions on $X_1$ and $X_2$ subject to the constraint that (A2.2) holds with probability one.

We prove the following lemma:

**Lemma A2.1:**

In the context defined above, if the entropy of $(Z_1 + Z_2 + X_2)$ is close to its maximal value, i.e., if

$$\frac{n}{2} \log \left( 2\pi e \left( 1 + N_2 + Q_2 \right) \right) - H(Z_1 + Z_2 + X_2) \leq n \epsilon,$$ \hspace{0.5cm} (A2.3)

then
Appendix 2

\[ H(X_1 \mid Y_2) - H(X_1 \mid Y_2) \leq n \delta, \quad (A2.4) \]

where \( \delta \rightarrow 0 \) as \( \epsilon \rightarrow 0 \).

Proof:

Let \( (Z_1 + Z_2 + X_2) \) have density denoted by \( f(\cdot) \). The variable \( (Z_1 + Z_2 + Z_3) \) is normally distributed. Let its density be denoted by \( \phi(\cdot) \). We evaluate the Kullback-Leibler distance between \( f \) and \( \phi \) as follows.

\[
D(f \parallel \phi) = \int_{\mathbb{R}^n} f \log \frac{f}{\phi} \nonumber
\]

\[
= -\int_{\mathbb{R}^n} f \log \phi + \int_{\mathbb{R}^n} f \log f \nonumber
\]

\[
= n \frac{1}{2} \log(2\pi(1 + N_2 + Q_2)) + \frac{E(\|Z_1 + Z_2 + X_2\|^2)}{2(1 + N_2 + Q_2)} - H(Z_1 + Z_2 + X_2) \nonumber
\]

\[
\leq \frac{n}{2} \log(2\pi e (1 + N_2 + Q_2)) - H(Z_1 + Z_2 + X_2) 
\]

\[
\leq n \epsilon. \quad (A2.5)
\]

The first inequality above follows from the power constraint imposed on \( X_2 \). The second inequality is just the hypothesis expressed in (A2.3). This upper bound on \( D(f \parallel \phi) \) indicates that \( f \) and \( \phi \) are close to each other in the Kullback-Leibler sense, i.e., that \( (Z_1 + Z_2 + X_2) \) is "almost" Gaussian.

Now suppose \( X_1 \) has density \( g(\cdot) \). Then the variables \( Y_2 \) and \( Y_3 \) are distributed according to \( f \ast g \) and \( \phi \ast g \), respectively. The Kullback-Leibler distance between these two densities is bounded by the data processing inequality as follows:

\[
D(f \ast g \parallel \phi \ast g) \leq D(f \parallel \phi). \quad (A2.6)
\]

\[\footnote{In strict terms, \( X_1 \) may not have a density since it may be drawn from a finite set of codewords. We assume the existence of \( g(\cdot) \) only for ease of notation; it is not crucial to the proof, because the densities of \( Y_2 \) and \( Y_3 \) always exist.}\]
Using (A2.5) we have

\[ \int_{\mathbb{R}^n} (f \ast g) \log \frac{f \ast g}{\phi \ast g} \leq n \varepsilon. \quad (A2.7) \]

Hence the Kullback-Leibler distance between \( f \ast g \) and \( \phi \ast g \) is small. Another distance measure of interest to us is the variational distance, defined by

\[ d(f \ast g, \phi \ast g) = \int_{\mathbb{R}^n} |f \ast g - \phi \ast g|. \quad (A2.8) \]

The Kullback-Leibler and the variational distances are related to each other through the following inequality [32]:

\[ \frac{1}{2} \, d^2(f \ast g, \phi \ast g) \leq D(f \ast g \| \phi \ast g). \quad (A2.9) \]

Adding and subtracting \( \int_{\mathbb{R}^n} (\phi \ast g) \log (\phi \ast g) \) to (A2.7) and rearranging terms we find

\[ H(Y_1) - H(Y_2) \leq \int_{\mathbb{R}^n} |f \ast g - \phi \ast g| + |\log(\phi \ast g)| + n \varepsilon. \quad (A2.10) \]

Our next step is to show that \( \int_{\mathbb{R}^n} |f \ast g - \phi \ast g| + |\log(\phi \ast g)| \) is small. Let \( V_r \) be the \( n \)-sphere of radius \( r \) and center at the origin and let \( V_r^c \) be the \( \mathbb{R}^n \)-complement of \( V_r \). We break the integral over \( \mathbb{R}^n \) in the following way:

\[ \int_{\mathbb{R}^n} |f \ast g - \phi \ast g| + |\log(\phi \ast g)| = \int_{V_r} |f \ast g - \phi \ast g| + |\log(\phi \ast g)| \]

\[ + \int_{V_r^c} |f \ast g - \phi \ast g| + |\log(\phi \ast g)|. \quad (A2.11) \]

The power constraint expressed in (A2.2) ensures that \( X_1 \) lies in the sphere \( V_{P_1} \). Therefore, \( \phi \ast g \) goes to zero no faster than \( \beta \exp(-\alpha \| y \|^2) \), for some positive constants \( \alpha \) and \( \beta \). This implies that, for \( r \geq P_1 \), we can upper bound \( |\log(\phi \ast g)| \) over \( V_r^c \) by
\[ |\log \beta| + \alpha \|y\|^2 \]. Then, we can write:

\[
\int_{V_r} |f * g - \phi * g| \|\log(\phi * g)\| \leq \int_{V_r} |f * g - \phi * g| |\log \beta| \]

\[ + \alpha \int_{V_r} |f * g - \phi * g| \|y\|^2 dy \]  \hspace{1cm} (A2.12)

Since \(Y_2\) and \(Y_3\) have finite variances, the integrals above tend to zero as \(r \to \infty\).

Hence we can find \(\delta_1 > 0\) such that \(\int_{V_r} |f * g - \phi * g| \|\log(\phi * g)\| \leq n \delta_1\), where \(\delta_1 \to 0\) as \(r \to \infty\).

By the same argument, we upper bound \(|\log(\phi * g)|\) inside \(V_r\) by \(|\log \beta| + \alpha r^2\).

Therefore we have

\[
\int_{V_r} |f * g - \phi * g| \|\log(\phi * g)\| \leq (|\log \beta| + \alpha r^2) \int_{V_r} |f * g - \phi * g| \]

\[ \leq (|\log \beta| + \alpha r^2) d(f * g, \phi * g) , \]  \hspace{1cm} (A2.13)

so that (A2.7) and (A2.9) yield

\[
\int_{V_r} |f * g - \phi * g| \|\log(\phi * g)\| \leq (|\log \beta| + \alpha r^2) \sqrt{2n\epsilon} . \]  \hspace{1cm} (A2.14)

Now we choose \(r = 1/\epsilon^{1/8}\) so that \(r \to \infty\) as \(\epsilon \to 0\) and

\[
\int_{V_r} |f * g - \phi * g| \|\log(\phi * g)\| \leq n (|\log \beta| \sqrt{2\epsilon} + \alpha \sqrt{2} \epsilon^{1/4}) . \]  \hspace{1cm} (A2.15)

Letting \(\delta_2 = |\log \beta| \sqrt{2\epsilon} + \alpha \sqrt{2} \epsilon^{1/4}\) and \(\delta_0 = \delta_1 + \delta_2\), we find that

\[
\int_{\mathbb{R}^d} |f * g - \phi * g| \|\log(\phi * g)\| \leq n \delta_0\), \]  where \(\delta_0 \to 0\) as \(\epsilon \to 0\). Using this result in (A2.10) we obtain

\[
H(Y_3) - H(Y_2) \leq n (\delta_0 + \epsilon) . \]  \hspace{1cm} (A2.16)
To conclude, we use the chain rule to expand $H(X_1 | Y_3) - H(X_1 | Y_2)$ and arrive at

$$H(X_1 | Y_3) - H(X_1 | Y_2) = H(Z_1 + Z_2 + Z_3) - H(Y_3) - H(Z_1 + Z_2 + X_2) - H(Y_2)$$

$$\leq n (\delta_0 + 2\epsilon)$$

$$= n \delta,$$

(A2.17)

where $\delta \rightarrow 0$ as $\epsilon \rightarrow 0$. This completes the proof of Lemma A2.1.
Appendix 3: A lower bound on the growth of entropy power

Let $X$ and $Z_t$ be as in Chapter 3. We show that, for $t > 0$,

$$\frac{d}{dt} \frac{e^{\frac{1}{2}H(X+Z_t)}}{2\pi e} \geq 1. \quad (A3.1)$$

Let $t_0 > 0$. Using the definition of derivative, we write

$$\frac{d}{dt} \frac{e^{\frac{1}{2}H(X+Z_t)}}{2\pi e} \bigg|_{t=t_0} = \lim_{t \to t_0} \frac{e^{\frac{1}{2}H(X+Z_t)} - e^{\frac{1}{2}H(X+Z_{t_0})}}{2\pi e(t - t_0)} \quad (A3.2)$$

Now we suppose $t > t_0$ and apply the entropy power inequality to $e^{\frac{1}{2}H(X+Z_t)}$ in the right side of (A3.2) to find

$$\frac{d}{dt} \frac{e^{\frac{1}{2}H(X+Z_t)}}{2\pi e} \bigg|_{t=t_0} \geq \lim_{t \to t_0} \frac{e^{\frac{1}{2}H(X+Z_t)} + e^{\frac{1}{2}H(Z_{t_0})} - e^{\frac{1}{2}H(X+Z_{t_0})}}{2\pi e(t - t_0)}$$

$$= \lim_{t \to t_0} \frac{2\pi e(t - t_0)}{2\pi e(t - t_0)}$$

$$= 1. \quad (A3.3)$$

For $t < t_0$ the result follows analogously.
Appendix 4: Vanishing surface integrals

Let \( X_t, Z_t \) and \( X_t \) be as in Chapter 3. Also let \( V_t \) be the \( n \)-sphere of radius \( r \) centered at the origin and having surface denoted by \( S_r \). We prove the following relations:

\[
L_1 = \lim_{r \to \infty} \int_{S_r} \log p_t(x_t) \nabla p_t(x_t) \cdot ds = 0, \tag{A4.1}
\]

\[
L_2 = \lim_{r \to \infty} \int_{S_r} \nabla^2 p_t(x_t) \nabla \log p_t(x_t) \cdot ds = 0, \tag{A4.2}
\]

\[
L_3 = \lim_{r \to \infty} \int_{S_r} \frac{\| \nabla p_t(x_t) \|^2}{p_t^2(x_t)} \nabla p_t(x_t) \cdot ds = 0, \tag{A4.3}
\]

where \( p_t(x_t) = \frac{1}{(2\pi t)^{n/2}} \int_{\mathbb{R}^n} p(x) \exp(-\frac{\|x_t-x\|^2}{2t}) dx \) is the probability density of \( X_t \) and \( p(x) \) is the arbitrary density of \( X \). These limits of surface integrals appear in Chapter 3 when Green’s identity is applied to integrals on \( \mathbb{R}^n \) in Eqs. (3.15), (3.23) and (3.20).

We first examine the scalar case. When \( n = 1 \), application of Green’s identity is the same as integration by parts. Then, the yielded terms corresponding to \( L_1 \), \( L_2 \) and \( L_3 \) are:

\[
\log p_t(x_t) \frac{dp_t(x_t)}{dx_t} \bigg|_{x_t=-\infty}^\infty, \tag{A4.4}
\]

\[
\frac{d^2 p_t(x_t)}{dx_t^2} \frac{dp_t(x_t)}{dx_t} \bigg|_{x_t=-\infty}^\infty, \tag{A4.5}
\]

\[
\left( \frac{dp_t(x_t)}{dx_t} \right)^3 \bigg|_{x_t=-\infty}^\infty, \tag{A4.6}
\]

\( \dagger \) To prove that \( L_1 = 0 \) we use the condition \( H(X_t) < \infty \). The proofs that \( L_2 \) and \( L_3 \) equal zero do not require this condition.
respectively.

In [22], Blachman uses Schwarz’s inequality to show that (A4.4) equals zero. To show that (A4.5) and (A4.6) also equal zero we use L’Hospital’s rule. A simple truncation argument can show that both numerators and denominators in (A4.5) and (A4.6) tend to zero in the limit as \(|x_t| \to \infty\). Therefore, applying L’Hospital’s rule to the ratio in (A4.6) we find

\[
\lim_{|x_t| \to \infty} \frac{(\frac{dp_t(x_t)}{dx_t})^3}{(p_t(x_t))^2} = \frac{3}{2} \lim_{|x_t| \to \infty} \frac{d^2p_t(x_t)}{dx_t^2} \frac{dp_t(x_t)}{p_t(x_t)},
\]  

(A4.7)

where we recognize the expression in (A4.5). Two applications of L’Hospital’s rule to the ratio in (A4.5) yields

\[
\lim_{|x_t| \to \infty} \frac{d^2p_t(x_t)}{dx_t^2} \frac{dp_t(x_t)}{dx_t} = 3 \lim_{|x_t| \to \infty} \frac{d^3p_t(x_t)}{dx_t^3}.
\]  

(A4.8)

The limit on the right side above can be shown to equal zero using the aforementioned truncation argument. This establishes the desired result in the scalar case.

We proceed to show (A4.1)-(A4.3) in higher dimensions. First note that the limits \(L_1, L_2\) and \(L_3\) all exist, because we can write each of these limits as the sum of two converging integrals, by use of Eqs. (3.15), (3.23) and (3.26). Next, we consider the integral from \(r = 0\) to \(r = \infty\) of the surface integral in (A4.2). Noting that

\[
\nabla \log p_t(x_t) = \frac{\nabla p_t(x_t)}{p_t(x_t)}\]

we have

\[
\left| \int_0^\infty \int_{S_r} \frac{\nabla^2 p_t(x_t)}{p_t(x_t)} \nabla p_t(x_t) \cdot ds, dr \right| \leq \int_{R^+} \frac{\|\nabla^2 p_t(x_t)\|}{p_t(x_t)} \|\nabla p_t(x_t)\| \|ds\| \|dr\|
\]
\[ = \int_\mathbb{R}^n p_1(x_t) \frac{\|\nabla p_t(x_t)\|}{p_t(x_t)} \frac{\|\nabla p_t(x_t)\|}{p_t(x_t)} \, dx_t. \quad (A4.9) \]

The Laplacian of \( p_t(x_t) \) is given by

\[ \nabla^2 p_t(x_t) = \sum_{i=1}^n \int_{\mathbb{R}^n} \left( \frac{(x(i)_t-x(i)_t)^2}{t^2} - \frac{1}{t} \right) \frac{p(x)}{(2\pi t)^{n/2}} \exp\left(-\frac{\|x_t-x\|^2}{2t}\right) \, dx. \quad (A4.10) \]

Therefore, we have

\[ \frac{\nabla^2 p_t(x_t)}{p_t(x_t)} = \sum_{i=1}^n \int_{\mathbb{R}^n} \left( \frac{(x(i)_t-x(i)_t)^2}{t^2} - \frac{1}{t} \right) \frac{p(x) \exp\left(-\frac{\|x_t-x\|^2}{2t}\right)}{(2\pi t)^{n/2} p_t(x_t)} \, dx \]

\[ = \sum_{i=1}^n \mathbb{E} \left( \frac{(X(i)_t-X(i)_t)^2}{t^2} - \frac{1}{t} \mid X_t = x_t \right) \]

\[ = \mathbb{E} \left( \frac{\|X_t - X\|^2}{t^2} \mid X_t = x_t \right) - \frac{n}{t}. \quad (A4.11) \]

Also we have

\[ \frac{\|\nabla p_t(x_t)\|}{p_t(x_t)} = \left\{ \sum_{i=1}^n \left[ \int_{\mathbb{R}^n} \frac{(x(i)_t-x(i)_t)}{t} \frac{p(x) \exp\left(-\frac{\|x_t-x\|^2}{2t}\right)}{(2\pi t)^{n/2} p_t(x_t)} \, dx \right]^2 \right\}^{1/2} \]

\[ = \left\{ \sum_{i=1}^n \mathbb{E} \left( \frac{X(i)_t-X(i)_t}{t} \mid X_t = x_t \right) \right\}^{1/2} \]

\[ \leq \left\{ \sum_{i=1}^n \mathbb{E} \left( \frac{(X(i)_t-X(i)_t)^2}{t^2} \mid X_t = x_t \right) \right\}^{1/2} \]

\[ = \mathbb{E} \left( \frac{\|X_t - X\|^2}{t^2} \mid X_t = x_t \right)^{1/2}. \quad (A4.12) \]

Thus, using Jensen’s inequality, we can write the following chain of inequalities:
\[
\int_{\mathbb{R}^*} \frac{p_l(x_t)}{p_l(x_t)} \frac{\|\nabla p_l(x_t)\|}{p_l(x_t)} \, dx_t \leq E \left\{ E \left[ \left( \frac{\|X_t\|^2}{l^2} \right)^{3/2} \mid X_t = x_t \right] \right\}^{1/2} + \frac{n}{t} \left\{ E \left[ \left( \frac{\|X_t\|^2}{l^2} \right)^{1/2} \mid X_t = x_t \right] \right\}^{1/2}
\]

\[
\leq E \left\{ E \left[ \left( \frac{\|X_t\|^2}{l^2} \right)^{1/2} \mid X_t = x_t \right] \right\}^{1/2} + \frac{n}{t} \left\{ E \left[ \left( \frac{\|Z_t\|^2}{l^2} \right)^{1/2} \mid X_t = x_t \right] \right\}^{1/2}
\]

\[
< \infty. \tag{A4.13}
\]

Hence, the integral considered in (A4.9) is finite. This fact and the existence of \( L_2 \) suffice to prove that \( L_2 = 0 \).

To show that \( L_2 = 0 \) we proceed in a similar way. We consider the integral from \( r = 0 \) to \( r = \infty \) of the surface integral in (A4.3), and find the sequence of relations below:

\[
\left| \int_0^\infty \int_{S_r} p_l(x_t) \frac{\|\nabla p_l(x_t)\|}{p_l^2(x_t)} \frac{\nabla p_l(x_t)}{p_l(x_t)} \, dS_r \, dr \right|
\]

\[
\leq \int_{\mathbb{R}^*} p_l(x_t) \frac{\|\nabla p_l(x_t)\|}{p_l(x_t)^3} \, dx_t
\]

\[
= E \left( \frac{\|\nabla p_l(x_t)\|^3}{p_l(x_t)^3} \right)
\]

\[
\leq E \left\{ E \left[ \left( \frac{\|X_t\|^2}{l^2} \right)^{3/2} \mid X_t = x_t \right] \right\}^{1/2}
\]

\[
\leq E \left\{ E \left( \frac{\|X_t\|^2}{l^2} \right)^{3/2} \mid X_t = x_t \right\}
\]
\[ = \mathbb{E}( \frac{\|X_i - X\|^3}{t^3} ) \]
\[ = \mathbb{E}( \frac{\|Z_i\|^3}{t^3} ) \]
\[ < \infty. \quad (A4.14) \]

Here also, the finiteness of this integral and the existence of \( L_3 \) are sufficient to establish that \( L_3 = 0 \).

To prove that \( L_1 = 0 \) we assume \( X_t \) has finite entropy. As seen in Chapter 3, this assumption does affect our application of this result since the concavity of \( e^{-H(X_t)} \) has no meaning if \( H(X_t) = \infty \).

In this proof, the magic of Green's identity comes into play once more. Let \( e_{S_r}(x_i) \) be the unit vector normal to \( S_r \) at the point \( x_i \). Under this notation \( ds_r = \|ds_r\| e_{S_r}(x_i) \). We integrate over \( r \geq 0 \) the surface integral in (A4.1) and apply Green's identity to find the following relations:

\[
\int_0^\infty \int_{S_r} \nabla p_t(x_i) \cdot ds_r \, dr = \int_{\mathbb{R}^*} \nabla p_t(x_i) \cdot (\log p_t(x_i) e_{S_r}(x_i)) \, dx_i
\]
\[ = \lim_{r \to \infty} \int_{S_r} p_t(x_i) \log p_t(x_i) e_{S_r}(x_i) \, ds_r
\]
\[ - \int_{\mathbb{R}^*} p_t(x_i) \nabla \cdot (\log p_t(x_i) e_{S_r}(x_i)) \, dx_i. \quad (A4.15) \]

The integral in the first term above can be written as \( \int_{S_r} p_t(x_i) \log p_t(x_i) \|ds_r\| \) and its limit as \( r \to 0 \) is easily shown to be zero, since \( X_t \) has finite entropy.

Now we note that the absolute value of the divergence in the second term satisfies
the following relation:

\[ | \nabla (\log p_t(x_t) e_S(x_t)) | = \frac{|\nabla p_t(x_t) \cdot e_S(x_t)|}{p_t(x_t)} \leq \frac{\| \nabla p_t(x_t) \|}{p_t(x_t)}. \] 

(A4.16)

Therefore we can write

\[ | \int_{\mathbb{R}^n} p_t(x_t) \nabla (\log p_t(x_t) e_S(x_t)) \, dx_t | \leq \int_{\mathbb{R}^n} p_t(x_t) \frac{\| \nabla p_t(x_t) \|}{p_t(x_t)} \, dx_t \]

\[ = E \left( \frac{\| \nabla p_t(x_t) \|}{p_t(x_t)} \right). \] 

(A4.17)

Using the steps in (A4.13) we can show that this expectation is finite. Hence the integral in (A4.15) is finite. Together with the fact that the limit \( L_1 \) exists, this suffices to prove that \( L_1 = 0 \).
References


