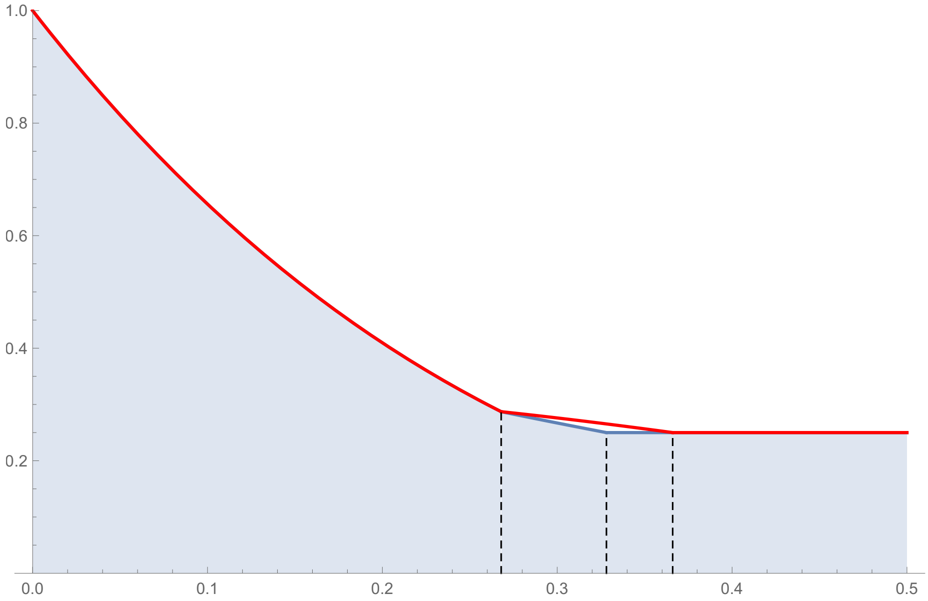
Local Simultaneous State Discrimination

Jaròn Has

June 24, 2022

Bachelor thesis Mathematics and Computer Science

Supervisors: dr. Maris Ozols, prof.dr. Christian Schaffner, dr. Mehrdad Tahmasbi



Informatics Institute
Korteweg-de Vries Institute for Mathematics
Faculty of Sciences
University of Amsterdam

Abstract

Local simultaneous state discrimination (LSSD) is a recently introduced problem. The classical form of LSSD is a non-local game played by non-communicating players against a referee. The referee generates one value for each of the players and one they keep for themselves. The players have to guess the referee's value and win if they all do so. In this thesis, we are interested in the winning probabilities, for certain LSSD settings, when we allow the players to share no-signalling resources versus classical resources.

We start by showing numerically that when there are three players and binary values, no-signalling resources can not provide any improvement over classical resources. We also take a look at a specific LSSD example defined by a binary symmetric channel, and find that when multiple simultaneous copies are played, no-signalling resources can improve on the optimal winning probability. Good classical strategies for this game can be defined by codes, and good no-signalling strategies by list-decoding schemes. We expand this example game to a class of games defined by an arbitrary channel and extend the idea of using codes and list decoding to define strategies for multiple simultaneous copies of these games. Finally, we give an expression for the limit of the exponent of the classical winning probabilities, and show that no-signalling strategies based on list-decoding schemes achieve the same limit.

Title: Local Simultaneous State Discrimination
Authors: Jaròn Has, jaron.has@student.uva.nl, 12851477
Supervisors: dr. Maris Ozols, prof.dr. Christian Schaffner, dr. Mehrdad Tahmasbi
Second grader: dr. Mikhail Isachenkov, dr. Florian Speelman
End date: June 24, 2022

Informatics Institute
University of Amsterdam
Science Park 904, 1098 XH Amsterdam
`http://www.ivi.uva.nl`

Korteweg-de Vries Institute for Mathematics
University of Amsterdam
Science Park 904, 1098 XH Amsterdam
`http://www.kdvi.uva.nl`

Contents

1. Introduction

Quantum mechanics is fundamental to our understanding of the universe and introduces new and very useful areas within mathematics and computer science, like quantum computing and quantum information theory. Research into quantum mechanics is important to expand our knowledge within these fields.

One topic that is often looked at in quantum research are non-local games. These games form small and isolated examples of how quantum resources (entanglement) can improve on classical resources (shared and local randomness). A non-local game is a game played by multiple non-communicating players against a referee. The referee asks each player a question in the form of a value. The players each give an answer to the referee, who then decides if the players win based on some pre-determined win conditions.

In researching non-local games we are mainly interested in optimal winning probabilities when we allow the players access to quantum resources. We are especially interested in whether quantum resources provide an improvement over classical resources. However, in this thesis we concern ourselves with no-signalling resources instead of quantum resources. No-signalling resources are stronger and easier to analyse than quantum resources, and therefore provide a good idea as to when quantum resources might outperform classical resources.

Local simultaneous state discrimination (LSSD) is a problem that was recently introduced by Majenz, Ozols, Schaffner and Tahmasbi [12]. In its simplest form, it is a non-local game played by two players, Alice and Bob, against a referee. The referee picks three values x, a and b according to some distribution P_{XAB} (which is known to Alice and Bob) and gives a to Alice and b to Bob, while keeping x for themselves. Alice and Bob try to guess the value x and win if they both succeed. As usual, Alice and Bob are not allowed to communicate, but they are allowed to share some resources.

In their paper, Majenz et al. give an explicit example of an LSSD game where the optimal winning probability using classical resources is strictly smaller than when using quantum resources, which in turn is smaller than the winning probability using no-signalling resources. They also showed that when x, a and b are all binary, the optimal winning probabilities for each of the types of shared resources are always the same.

In this thesis, we analyse some more LSSD settings, mainly focussing on the following questions: what are optimal classical and no-signalling strategies? And do no-signalling resources offer improvement in terms of winning probability versus classical resources? In Chapter 2 we discuss the necessary background information to understand this thesis. Next, in Chapter 3, we formally introduce the LSSD problem, including its quantum version. In Chapter 4, 5 and 6 we discuss our contributions to the LSSD problem, which are summarized below.

Our contributions In Chapter 4 we start by stating some results on optimal strategies. These results are extensions and generalizations of some results in the paper by Majenz et al. We make use of these results when we numerically show that in an LSSD game with three players and binary inputs, no-signalling resources cannot improve the winning probability of the players.

In chapter 5 we discuss an example of LSSD introduced by Majenz et al. In this example, the referee sends a bit x over a binary symmetric channel to Alice and Bob. We first show, in Theorem 5.1, that under certain conditions there is always an optimal classical strategy that is symmetric. We use this result to numerically find optimal classical strategies for two and three copies of the game. We also give optimal no-signalling strategies for two and three copies. Finally, we generalize the strategies to n simultaneous copies and argue how they can be defined by codes and list-decoding schemes.

In Chapter 6 we introduce channel games, which are an extension of the game in Chapter 5. We show that we can define classical strategies based on codes and no-signalling strategies based on list-decoding schemes. In Theorem 6.1 we provide an expression for the limit of the exponent of the classical winning probability, where we make use of strategies based on codes. In Theorem 6.4 we show that no-signalling strategies based on list-decoding schemes achieve the same limit as classical strategies. This last result makes use of Conjecture 6.5, which we leave unproven in this thesis. Finally, in Theorem 6.8 we show that the limit of the exponent of the no-signalling winning probability is equal to the classical limit. This result implies that no-signalling strategies based on list-decoding schemes are asymptotically optimal in the game defined by the binary symmetric channel.

2. Preliminaries

In this chapter, we will give some background information on certain mathematical topics necessary to fully understand this thesis. The background information on quantum is only necessary to understand the quantum version of LSSD, but is not used for any results. The concept of no-signalling is used throughout this whole thesis. Linear programming is used in Chapter 4 and Chapter 5, but only necessary to understand Chapter 4. Lastly, information theory is used in Chapter 5 and Chapter 6.

Notation For $n \in \mathbb{N}$, we denote the set $\{0, \dots, n-1\}$ by $[n]$ and the set of all permutations $\sigma\colon [n] \to [n]$ by S_n. By δ we denote the indicator function, which is 1 if its argument is true and 0 otherwise. In this thesis, by log we mean $\log_2$. We denote by $\oplus$ the bitwise XOR operator on bitstrings and finally, by 0^n and 1^n we denote the all-zero and all-one bitstrings of length n.

For a distribution P_{X} over $\mathscr{X}$, we denote by $P_{\mathsf{X}}^{\times n} = (P_{\mathsf{X}})^{\times n}$ the probability distribution defined by

$$P_{\mathsf{X}}^{\times n}(x^n) := \prod_{i=1}^{n} P_{\mathsf{X}}(x_i),$$

where x^n is an element of $\mathscr{X}^n$. We sometimes omit writing the subscript in P_{X}, when it is obvious over which set P is a distribution. Lastly, for $A \subset \mathscr{X}$, we denote by $P_{\mathsf{X}}(A)$ the probability of random variable X taking on a value in A:

$$P_{\mathsf{X}}(A) = \sum_{x \in A} P_{\mathsf{X}}(x).$$

2.1. Quantum

We mention only the necessary information. Interested readers are advised to read "Quantum Computing and Quantum Information" by Nielsen and Chuang [14], probably the best source on quantum information.

Quantum systems are described by a complex (finite-dimensional) Hilbert space, which is to say a complete complex inner product space. If we choose an orthonormal basis for a Hilbert space H of dimension d, we can identify it with the Hilbert space $\mathbb{C}^d$. Therefore, we will think of elements in H as vectors in $\mathbb{C}^d$.

We denote vectors in $\mathbb{C}^d$ by $|\psi\rangle$ and define $\langle\psi| = |\psi\rangle^\dagger$ (the complex conjugate transpose). With this notation, we can write the inner product and outer product of two vectors as $\langle\psi_1|\psi_2\rangle$ and $|\psi_1\rangle\langle\psi_2|$ respectively.

We denote the set of linear operators on a Hilbert space H as $L(H)$. Let $\{|e_i\rangle\}$ be an orthonormal basis for H. For $X \in L(H)$ the trace of X is given by

$$\mathrm{tr}(X) = \sum_i \langle e_i|X|e_i\rangle.$$

We say that $X \in L(H)$ is positive semi-definite if it is hermitian ($X = X^\dagger$) and has non-negative eigenvalues. The set of quantum states $D(H)$ is defined as the set of positive semi-definite matrices whose trace is equal to 1. We call a state ρ a pure state if we can write $\rho = |\psi\rangle\langle\psi|$ and a classical state if we can write

$$\rho = \sum_i P(i)|e_i\rangle\langle e_i|,$$

Where P is a probability distribution on $\{1, \dots, d\}$.

We define a measurement (POVM) M on a quantum system as a set of positive semi-definite matrices $\{M_1, \dots, M_n\}$ whose sum is the identity matrix: $\sum_{i=1}^n M_n = I$. When measuring a state ρ with measurement M, the probability of getting outcome $i \in \{1, \dots, n\}$ is given by $\mathrm{tr}(\rho M_i)$. We denote by $M(\mathbb{C}^d)$ the set of all measurements (the number of outcomes will always be clear).

Finally, we can combine quantum systems into a bigger quantum system using the tensor product. Let A and B be two matrices, the tensor product $A \otimes B$ of A and B is defined by

$$A \otimes B := \begin{pmatrix} A_{1,1}B & \cdots & A_{1,m}B \\ \vdots & \ddots & \vdots \\ A_{n,1}B & \cdots & A_{n,m}B \end{pmatrix}.$$

We call states in a combined system bipartite, tripartite etc. depending on how many systems are combined.

2.2. Non-locality and no-signalling

In the introduction it was mentioned that quantum and no-signalling resources can improve on classical resources in terms of winning probabilities in non-local games. It turns out that this effect occurs because quantum and no-signalling resources can achieve non-locality. In this section we will explain what this means and what no-signalling exactly is, taking inspiration from a paper by Barrett et al. [1, Section 2].

2.2.1. The local polytope

We can describe the actions of two parties in a non-local game by a conditional probability distribution $Q_{\mathsf{XY|AB}}$, called correlations or boxes, over $\mathscr{X} \times \mathscr{Y} \times \mathscr{A} \times \mathscr{B}$. Here, $\mathscr{A}$ and $\mathscr{B}$ are the input sets for Alice and Bob respectively and $\mathscr{X}$ and $\mathscr{Y}$ the output sets. In this case $Q(x, y|a, b)$ is the probability of Alice and Bob outputting x and y on inputs a and b.

We call the distribution $Q_{\mathsf{XY|AB}}$ local if we can write

$$Q_{\mathsf{XY|AB}} = \sum_{\lambda} P(\lambda) Q_{\mathsf{X|A}}(\lambda) Q_{\mathsf{Y|B}}(\lambda). \tag{2.1}$$

In other words, the distribution can be achieved by shared and local randomness. The shared randomness comes in the form of a variable λ with probability distribution $P(\lambda)$, and the local randomness comes in the form of two conditional probability distributions $Q_{\mathsf{X|A}}(\lambda)$ and $Q_{\mathsf{Y|B}}(\lambda)$, describing Alice's and Bob's actions depending on the shared randomness λ.

In fact, if we say that Alice and Bob are allowed to share classical resources, we mean that they are allowed to use shared and local randomness. Therefore, the set of all local correlations consists exactly of those correlations achievable by sharing classical resources. It is known that this set is a convex polytope (see Section 2.3) and it is often called the local polytope [1].

2.2.2. The no-signalling polytope

Using quantum resources it is possible to achieve correlations that cannot be written as in formula (2.1), we call these non-local correlations. In this thesis, however, we will not be considering quantum resources. Instead, we look at the strictly larger set of no-signalling correlations. This set contains all correlations achievable without communication between the parties, i.e., the correlations must satisfy the following no-signalling constraints:

$$\forall y, a, a', b: \sum_{x \in \mathscr{X}} Q_{\mathsf{XY|AB}}(x, y|a, b) = \sum_{x \in \mathscr{X}} Q_{\mathsf{XY|AB}}(x, y|a', b), \tag{2.2}$$

$$\forall x, a, b, b': \sum_{y \in \mathscr{Y}} Q_{\mathsf{XY|AB}}(x, y|a, b) = \sum_{y \in \mathscr{Y}} Q_{\mathsf{XY|AB}}(x, y|a, b'). \tag{2.3}$$

Since $\sum_x Q_{\mathsf{XY|AB}}(x, y|a, b) = Q_{\mathsf{Y|AB}}(y|a, b)$, these constraints basically mean that the output of Bob does not depend on the input of Alice and vice versa. The set of all no-signalling correlations also form a convex polytope. This last fact will become obvious in Section 2.3.

It is not hard to see that local correlations satisfy the no-signalling constraints:

$$\begin{aligned}
\sum_{x \in \mathscr{X}} Q_{\mathsf{XY|AB}}(x, y|a, b) &= \sum_{x \in \mathscr{X}, \lambda} P(\lambda) Q_{\mathsf{X|A}}(\lambda)(x|a) Q_{\mathsf{Y|B}}(\lambda)(y|b) \\
&= \sum_{\lambda} P(\lambda) Q_{\mathsf{Y|B}}(\lambda)(y|b) \sum_{x \in \mathscr{X}} Q_{\mathsf{X|A}}(\lambda)(x|a) \\
&= \sum_{\lambda} P(\lambda) Q_{\mathsf{Y|B}}(\lambda)(y|b).
\end{aligned}$$

The final expression does not depend on a, so (2.2) is satisfied. This observation means that the set of local correlations is a subset of the set of no-signalling correlations. In fact, apart from some trivial cases, the sets of correlations achievable by classical, quantum and no-signalling resources are strict subsets of each-other.

2.2.3. Multi-partite no-signalling correlations

Up until now, we have only looked at correlations between two parties. However, the concepts of locality and no-signalling can be extended to any finite number of parties. We show how to do this extension for no-signalling.

In the case of more than two parties, a correlation is no signalling if no subset of parties J can collectively signal to the rest of the parties I. So the output of the parties indexed by I cannot depend on the input to the parties indexed by J. This is worded formally in the next definition.

Definition 2.1. [2, Definition 11] An m-partite correlation $Q_{\mathsf{X}_1\cdots\mathsf{X}_m|\mathsf{A}_1\cdots\mathsf{A}_m}$ on $\mathscr{X}_1\times\cdots\times\mathscr{X}_m\times\mathscr{A}_1\times\cdots\times\mathscr{A}_m$ is called no-signalling if for any index set $I\subset\{1,\dots,m\}$ and its complement $J=\{1,\dots,m\}\setminus I$ it holds that

$$\sum_{x_J\in\mathscr{X}_J} Q(x_I,x_J\big|a_I,a_J)=\sum_{x_J\in\mathscr{X}_J} Q(x_I,x_J\big|a_I,a'_J) \tag{2.4}$$

for all $x_I\in\mathscr{X}_I, a_I\in\mathscr{A}_I$ and $a_J,a'_J\in\mathscr{A}_J$.

The next lemma states that we can loosen the constraints a little and still be left with an equivalent definition of no-signalling. Specifically, it states that it is enough to require that any single party can not signal to the rest.

Lemma 2.2. *Suppose Q is a m-partite correlation satisfying (2.4) for all index sets I such that their complements J have cardinality* 1 *and for all $x_I\in\mathscr{X}_I, a_I\in\mathscr{A}_I$ and $a_J,a'_J\in\mathscr{A}_J$. Then Q is a no-signalling correlation.*

Proof. We prove this by induction on the cardinality of the complement J of an index set I. If $|J|=1$, condition (2.4) holds by assumption. Now suppose $|J|=n$, and let $x_I\in\mathscr{X}_I, a_I\in\mathscr{A}_I$ and $a_J,a'_J\in\mathscr{A}_J$. Take $j\in J$ and let $J'=J\setminus\{j\}$. We now find

$$\begin{aligned}\sum_{x_J\in\mathscr{X}_J} Q(x_I,x_J\big|a_I,a_J) &= \sum_{x_{J'}\in\mathscr{X}_{J'}}\sum_{x_j\in\mathscr{X}_j} Q(x_I,x_{J'},x_j\big|a_I,a_{J'},a_j)\\ &\overset{(i)}{=} \sum_{x_{J'}\in\mathscr{X}_{J'}}\sum_{x_j\in\mathscr{X}_j} Q(x_I,x_{J'},x_j\big|a_I,a_{J'},a'_j)\\ &\overset{(ii)}{=} \sum_{x_{J'}\in\mathscr{X}_{J'}}\sum_{x_j\in\mathscr{X}_j} Q(x_I,x_{J'},x_j\big|a_I,a'_{J'},a'_j)\\ &= \sum_{x_J\in\mathscr{X}_J} Q(x_I,x_J\big|a_I,a'_J),\end{aligned}$$

where (i) follows by assumption on Q and (ii) by induction (we are free to exchange the sums). □

2.3. Linear programming

Linear programming is a technique in which we optimize a linear function over a domain that is a convex polytope. A polytope is a generalization of a polygon to any number of dimensions. There are two ways of describing a convex polytope: by giving its extreme points (and rays), called the vertex representation (V-representation), or by linear constraints, called the half-space representation (H-representation).

The H-representation of a convex polytope is a collection of (closed) half-spaces, such that their intersection is the convex polytope. A half-space can be described by a linear inequality

$$a_1 x_1 + \cdots + a_n x_n \leq c. \tag{2.5}$$

Using this description, the convex polytope can be represented as a system of linear inequalities, which can be written as a matrix inequality

$$Ax \leq d.$$

Here, A is the matrix containing all factors a_i and d the vector containing all constants c, for all inequalities as in (2.5) representing the polytope. Sometimes we include linear equalities as well, since they could also be described by two opposite inequalities.

We have already seen an example of a convex polytope described by linear (in)equalities: the no-signalling polytope. The variables, in this case, are each of the probabilities $Q(x, y|a, b)$, and must satisfy the no signalling constraints in (2.2) and (2.3) (or, more generally, (2.4)). The variables must also satisfy $Q(x, y|a, b) \geq 0$ for all x, y, a, b and finally,

$$\forall a, b: \sum_{x,y} Q(x, y|a, b) = 1.$$

Given a V-representation, the corresponding convex polytope is the convex hull of the extreme points. The convex hull of a set of points is the smallest convex set that contains all the points, or simply the set of all convex combinations of the points (i.e., all weighted averages). This representation is especially interesting, since a linear function always has a global maximum in (at least) one of the extreme points of a convex polytope. We make use of this fact in Section 4.2.

2.4. Information theory

The definitions in this section are largely based on the book by Csiszár and Körner [4].

Information theory is the study of communicating and storing information. The simplest setting in information theory is one with a single sender and a single receiver. The goal of the sender and the receiver, which could be separated by space or time, is to communicate information from the sender to the receiver in such a way that the receiver can be sure that the message they received was the message that was sent, up to some probability of error.

The communication between the sender and the receiver happens over a channel. A channel is a probabilistic function between two sets $\mathscr{X}$ and $\mathscr{A}$, represented by a conditional probability distribution $P_{\mathsf{A|X}}$. Here $P_{\mathsf{A|X}}(a|x)$ is the probability of the channel outputting a on input x. We call a channel memoryless if consecutive uses of the channel do not change its probability distribution. We can model n consecutive uses of a memoryless channel $P_{\mathsf{A|X}}$ by the channel $P_{\mathsf{A|X}}^{\times n}$.

Communication over a memoryless channel happens as follows: the sender has a message set M of possible messages (sometimes we denote the message set by $[M]$ instead, where M is the number of messages); they pick one message $m \in M$ to send; they encode m as an element x^n of $\mathscr{X}^n$, the codeword, using a function $\mathrm{Enc}\colon M \to \mathscr{X}^n$; next, they transmit each of the symbols x_i of this codeword to the receiver by consecutive uses of the channel; the receiver receives an a^n in $\mathscr{A}^n$ and decodes it to a message m', using a function $\mathrm{Dec}\colon \mathscr{A}^n \to M$. The communication was successful if $m = m'$. The encoding and decoding function together form a code (also called an n-block code, or error-correcting code).

The average success probability of a code is given by

$$\omega := \frac{1}{|M|} \sum_{m \in M} P_{\mathsf{A|X}}^{\times n}(\mathrm{Dec}^{-1}(m)| \,\mathrm{Enc}(m));$$

the minimum winning probability of a code is given by

$$\alpha := \min_{m \in M} P_{\mathsf{A|X}}^{\times n}(\mathrm{Dec}^{-1}(m)| \,\mathrm{Enc}(m))$$

and the rate of a code is $R = \frac{1}{n} \log |M|$.

Definition 2.3. We call a code $(\mathrm{Enc}, \mathrm{Dec})$ for a channel $P_{\mathsf{A|X}}$ an $(n, 2^{nR}, \alpha)$ code if

$$\begin{aligned} &\mathrm{Enc}\colon [2^{nR}] \to \mathscr{X}^n;\\ &\mathrm{Dec}\colon \mathscr{A}^n \to [2^{nR}];\\ &P_{\mathsf{A|X}}^{\times n}(\mathrm{Dec}^{-1}(m)| \,\mathrm{Enc}(m)) \geq \alpha. \end{aligned}$$

(Note that $Dec^{-1}(m)$ might be a set.)

There is another form of decoding, called list-decoding (described in the paper by Merhav [13]), that will prove to be useful in this thesis. In list decoding, the decoder outputs a set (list) of messages of size L, instead of a single message. The decoding is successful if the list contains the message that was sent. We denote the list outputted by the decoder on input a^n as C_{a^n}. The average success probability of a list-decoding scheme is given by

$$\omega := \frac{1}{|M|} \sum_{\substack{m \in M \\ a^n \in \mathscr{A}^n \colon C_{a^n} \ni m}} P_{\mathsf{A|X}}^{\times n}(a^n | \,\mathrm{Enc}(m))$$

and the minimum winning probability by

$$\alpha := \min_{m \in M} \sum_{a^n \in \mathscr{A}^n \colon C_{a^n} \ni m} P_{\mathsf{A|X}}^{\times n}(a^n | \,\mathrm{Enc}(m)).$$

Definition 2.4. We call a list code a $(n, 2^{nR}, L, \alpha)$ code if Dec maps elements a^n of $\mathscr{A}^n$ to subsets C_{a^n} of $[2^{nR}]$ of size L and

$$\begin{aligned} &\mathrm{Enc}\colon [2^{nR}] \to \mathscr{X}^n; \\ &\sum_{a^n \in \mathscr{A}^n \colon C_{a^n} \ni m} P^{\times n}_{\mathsf{A|X}}(a^n | \mathrm{Enc}(m)) \geq \alpha. \end{aligned}$$

Next, we take a look at an example of a channel, the binary symmetric channel, and some codes for this channel.

2.4.1. Binary symmetric channel

The binary symmetric channel (BSC) is probably the most well known example of a channel. It is a channel from $\{0,1\}$ and $\{0,1\}$, which takes an input bit x, flips it with probability $\alpha \in [0, 1/2]$ and outputs the result. A schematic is shown in Figure 2.1.

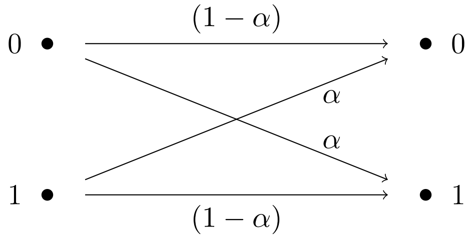


Figure 2.1.: Schematic of the binary symmetric channel.

The simplest code for this channel, called the repetition code, encodes two messages 0 and 1 to 0^n and 1^n respectively. Decoding works by taking the output a from the BSC and outputting the bit that occurs the most in a. This scheme is well-defined for odd n and has average success probability $\sum_{i=0}^{(n-1)/2} \binom{n}{i} \alpha^i (1-\alpha)^{n-i}$.

Hamming code Another code for the BSC, perhaps the most famous one, is the (7,4)-Hamming code, introduced by Richard Hamming [9]. This code encodes bitstrings $d_1 d_2 d_3 d_4$ of length 4 as bitstrings of length 7 by appending three parity bits: $d_1 d_2 d_3 d_4 p_1 p_2 p_3$. These bits represent the parity (XOR) of three of the original 4 bits (See Figure 2.2[1]).

Decoding works by checking if the parity bits are still correct (still equal to the parity of the corresponding 3 bits). If this is the case, we just remove the last three bits of the received bitstring. Now suppose an error occurred in exactly one bit.

- If the error occurred in d_4, all the parity bits are incorrect.
- If the error occurred in d_1, d_2 or d_3, two of the parity bits are incorrect (p_1 and p_2 for d_1, p_1 and p_3 for d_2 and p_2 and p_3 for d_3).

[1] Image by Cburnett on Wikipedia: `https://nl.wikipedia.org/wiki/Hamming-code`

- If the error occurred in one of the parity bits, only that parity bit will be incorrect.

Using the above, we can perfectly deduce in which bit the error occurred and correct it accordingly. If more than one error occurs, this method never decodes correctly.

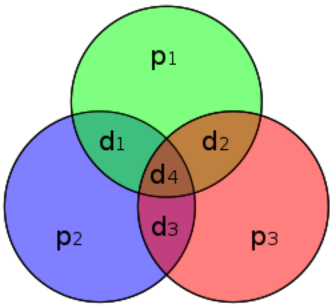


Figure 2.2.: The Hamming code visualized: The bitstring $d_1 d_2 d_3 d_4$ is encoded by appending the parity bits p_1, p_2 and p_3, where each parity bit represents the parity of the three bits inside their circle. A single error in one of the seven bits can be perfectly detected by checking which parity bits are incorrect.

Since the Hamming code corrects exactly 0 or 1 error, we can write the average success probability of this code as

$$(1-\alpha)^7 + 7\alpha(1-\alpha)^6.$$

Nearest neighbour decoding Nearest neighbour decoding is based on the concept of Hamming distance. The hamming distance $d(x, y)$ between two bitstrings x and y (of the same length) is the number of positions in which they differ:

$$d(x, y) = |\{i \mid x_i \neq y_i\}|.$$

In nearest neighbour decoding, we decode a bitstring a to the message whose codeword x has the smallest Hamming distance to a. Since a bit sent through a BSC is more likely to stay the same than to flip, nearest neighbour decoding optimizes the average success probability of a code. Now the question becomes: To which codewords do we send our messages such that the average success probability is maximized. Intuition suggests that we want to pick the codewords as far apart from each other, in terms of Hamming distance, as possible.

It is easy to see that if d is the minimal distance between any two codewords, nearest neighbour decoding can correct up to $\lfloor (d-1)/2 \rfloor$ bits. We call a code using nearest neighbour decoding an (n, k, d) code, if the encoding function maps messages from $\{0, 1\}^k$ to codewords in $\{0, 1\}^n$, such that the minimal distance between any two codewords is at least d. We can lower bound the average success probability of an (n, k, d) code by

$$\sum_{i=0}^{\lfloor (d-1)/2 \rfloor} \binom{n}{i} \alpha^i (1-\alpha)^{n-i}.$$

List decoding for the BSC To create a list decoding scheme for the binary symmetric channel, we can expand the idea of nearest neighbour decoding: the decoder outputs the L messages whose codewords are closest to the received bitstring. Note that when the message set is $\{0,1\}^n$, this corresponds to outputting the Hamming ball, with at most L elements, around the received bitstring. If there exists a radius d, such that the size of the corresponding Hamming ball is exactly L, this decoding scheme has average success probability

$$\sum_{i=0}^{d} \binom{n}{i} \alpha^i (1-\alpha)^{n-i}.$$

2.4.2. Entropy

In this section we quickly go over all definitions regarding entropy. However, since entropy only plays a small part in this thesis, we keep it quite concise. For a more in-depth explanation of entropy, one could look at "Elements of Information Theory" by Cover and Thomas [3, Chapter 2].

Let P be a probability distribution over $\mathscr{X}$ and let X be a random variable distributed according to P. We define the entropy $H(X) = H(P)$ of X as

$$H(X) := -\sum_{x\in\mathscr{X}} P(x)\log(P(x))$$

(wherever $P(x) = 0$, we say $P(x)\log(P(x)) = 0$). Note that $H(P) \in [0, \log|\mathscr{X}|]$. The entropy of a distribution can be seen as a measure of uncertainty, or of its information contents: the higher the entropy, the lower its information contents.

Now let X and Y be two random variables with joint probability distribution P_{XY}. We define the joint entropy as $H(X,Y) = H(P_{\mathsf{XY}})$ and the conditional entropy $H(X|Y)$ as

$$H(X|Y) := \sum_{y} P_{\mathsf{Y}}(y) H(X|Y=y).$$

It is not hard to show that $H(X|Y) = H(X,Y) - H(Y)$.

For two random variables X, Y, we define the mutual information $I(X;Y)$ as

$$I(X;Y) := H(X) + H(Y) - H(X,Y).$$

Next, for two probability distributions P and Q over $\mathscr{X}$ we define the relative entropy $D(P\|Q)$ as

$$D(P\|Q) := \sum_{x\in\mathscr{X}} P(x)\log\left(\frac{P(x)}{Q(x)}\right).$$

Finally, let $P^1_{\mathsf{X|Y}}$ and $P^2_{\mathsf{X|Y}}$ be two conditional distributions and let Q_{Y} be a distribution. We define the conditional relative entropy $D(P^1_{\mathsf{X|Y}}\|P^2_{\mathsf{X|Y}} \mid Q_Y)$ as

$$D(P^1_{\mathsf{X|Y}}\|P^2_{\mathsf{X|Y}} \mid Q_Y) := \sum_{y\in Y} Q_{\mathsf{Y}}(y) D(P^1_{\mathsf{X|Y=y}}\|P^2_{\mathsf{X|Y=y}}).$$

3. LSSD

In this chapter we deal with the definition of LSSD and the different forms of strategies (classical and no-signalling). We only discuss the case of two players, Alice and Bob, but all definitions can easily be generalized to any number of players.

The inputs to Alice and Bob can either be classical or quantum. In the first case, we denote by $\mathscr{X}, \mathscr{A}$ and $\mathscr{B}$ the sets of possible inputs for the referee, Alice and Bob respectively. When the inputs are quantum, we need to consider quantum systems. We denote by $\mathcal{X} = \mathbb{C}^{\mathscr{X}}, \mathcal{A} = \mathbb{C}^{\mathscr{A}}$ and $\mathcal{B} = \mathbb{C}^{\mathscr{B}}$ the quantum systems underlying the registers X, A and B, belonging to the referee, Alice and Bob.

3.1. Definitions

In the most general setting, an LSSD game played by two players is defined by a cqq state ρ_{XAB}, which means that register X is classical, while registers A and B can be quantum. Such a state is of the form

$$\rho_{\mathsf{XAB}} = \sum_{x \in \mathscr{X}} P_{\mathsf{X}}(x) |x\rangle\langle x|_{\mathsf{X}} \otimes \rho^x_{\mathsf{AB}},$$

where P_{X} is a probability distribution over $\mathscr{X}$ and ρ^x_{AB} are bipartite quantum states. The referee gives register A to Alice and B to Bob and keeps register X for themselves. Alice and Bob know the state ρ_{XAB} and will try to guess the value x based on their received states. We denote their guesses by x_A and x_B. Alice and Bob are allowed to share some resources, but are not allowed to communicate with each other. Finally, they win the game if both guesses are correct: $x_A = x_B = x$.

In most of this thesis we are going to be looking at the case where ρ_{XAB} is completely classical. Meaning that there exist orthonormal bases $\{|a\rangle \mid a \in \mathscr{A}\}$ and $\{|b\rangle \mid b \in \mathscr{B}\}$ for $\mathcal{A}$ and $\mathcal{B}$ respectively, that are independent of $x \in \mathscr{X}$, and probability distributions P^x_{AB} over $\mathscr{A} \times \mathscr{B}$ such that

$$\rho^x_{\mathsf{AB}} = \sum_{\substack{a \in \mathscr{A} \\ b \in \mathscr{B}}} P^x_{\mathsf{AB}}(a,b) |a\rangle\langle a|_{\mathsf{A}} \otimes |b\rangle\langle b|_{\mathsf{B}}.$$

In this case, it is useful to reword the problem. Instead of the game being described by a cqq state, we can now describe it by a probability distribution P_{XAB} on $\mathscr{X} \times \mathscr{A} \times \mathscr{B}$. The referee picks elements $x \in \mathscr{X}, a \in \mathscr{A}$ and $b \in \mathscr{B}$ according to this distribution and gives a and b to Alice and Bob respectively. Alice and Bob know the distribution P_{XAB} and both try to guess the value x. Again, they can share some resources, but are not allowed to communicate, and they win if they both guess correctly. A schematic of LSSD is shown in Figure 3.1.

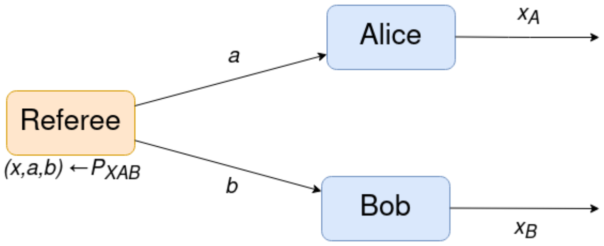


Figure 3.1.: A schematic of the LSSD game. On inputs a and b Alice and Bob make guesses x_A and x_B respectively, and win if $x = x_A = x_B$.

3.2. Strategies

In this section we will describe the different types of strategies based on the different possible shared resources: classical and no-signalling.

3.2.1. Classical strategies

Using classical resources, the players are allowed to share some randomness. However, sharing randomness does not help in increasing the winning probability. This can easily be seen by realizing that after a random value is generated, what is left is a strategy that does not depend on that randomness any more. So instead of using the randomness, the players can just use the strategy that achieves the highest winning probability. In the following, we assume that the players do not use shared randomness.

In the quantum case of the LSSD game (meaning that the game is described by a quantum state ρ_{XAB}), a strategy is completely defined two measurements $M = \{M_x \mid x \in \mathscr{X}\}$ and $N = \{N_x \mid x \in \mathscr{X}\}$. Alice and Bob perform measurements M and N respectively on their subsystem, which gives them their guess. Given the measurements M and N their winning probability is given by

$$\sum_{x\in\mathscr{X}} P_{\mathsf{X}}(x)\,\mathrm{tr}[\rho^x_{\mathsf{AB}}(M_x \otimes N_x)]$$

and the optimal winning probability is denoted by

$$\omega_{\mathrm{c}}(\mathsf{X}|\mathsf{A};\mathsf{B})_\rho := \sup_{\substack{M\in M(\mathcal{A})\\ N\in M(\mathcal{B})}} \sum_{x\in\mathscr{X}} P_{\mathsf{X}}(x)\,\mathrm{tr}[\rho^x_{\mathsf{AB}}(M_x \otimes N_x)].$$

In case ρ_{XAB} is purely classical and described by a probability distribution P_{XAB}, the strategy of Alice and Bob is given by two conditional probability distributions $Q_{\mathsf{X}_A|\mathsf{A}}$ and $Q_{\mathsf{X}_B|\mathsf{B}}$ describing their local behaviour. The winning probability is then given by

$$\sum_{\substack{x\in\mathscr{X}\\ a\in\mathscr{A},b\in\mathscr{B}}} P_{\mathsf{XAB}}(x,a,b)Q_{\mathsf{X}_A|\mathsf{A}}(x|a)Q_{\mathsf{X}_B|\mathsf{B}}(x|b).$$

The optimal winning probability can now be obtained by maximizing over all conditional probabilities. However, we can restrict this optimization to maximizing over all deterministic strategies, i.e., strategies that can be described by two functions $f\colon \mathscr{A} \to \mathscr{X}$ and $g\colon \mathscr{B} \to \mathscr{X}$. Similarly to shared randomness, Alice and Bob can condition any local randomness on the realization that maximizes their probability of winning. Now, the optimal winning probability is given by

$$\omega_{\mathrm{c}}(\mathsf{X}|\mathsf{A};\mathsf{B})_P = \max_{f,g} \sum_{\substack{x\in\mathscr{X}\\ a\in\mathscr{A},b\in\mathscr{B}}} P_{\mathsf{XAB}}(x,a,b)\delta[f(a)=g(b)=x].$$

3.2.2. No-signalling strategies

For no-signalling strategies we only look at the classical version of LSSD, i.e. games defined by a probability distribution P_{XAB}.

Using no-signalling resources, a strategy is given by a conditional probability distribution $Q_{\mathsf{X}_A\mathsf{X}_B|\mathsf{AB}}$ on $\mathscr{X}\times\mathscr{X}\times\mathscr{A}\times\mathscr{B}$ satisfying the no-signalling constraints. The winning probability of such a strategy is given by:

$$\sum_{\substack{x\in\mathscr{X}\\ a\in\mathscr{A},b\in\mathscr{B}}} P_{\mathsf{XAB}}(x,a,b)Q_{\mathsf{X}_A\mathsf{X}_B|\mathsf{AB}}(x,x|a,b).$$

The optimal winning probability can be found by taking the supremum over all possible no-signalling strategies:

$$\omega_{\mathrm{ns}}(\mathsf{X}|\mathsf{A};\mathsf{B})_P := \sup_{Q_{\mathsf{X}_A\mathsf{X}_B|\mathsf{AB}}} \sum_{\substack{x\in\mathscr{X}\\ a\in\mathscr{A},b\in\mathscr{B}}} P_{\mathsf{XAB}}(x,a,b)Q_{\mathsf{X}_A\mathsf{X}_B|\mathsf{AB}}(x,x|a,b)$$

Important to note here is that the winning probability given a no-signalling strategy is a linear function in the values $Q_{\mathsf{X}_A\mathsf{X}_B|\mathsf{AB}}(x_A,x_B|a,b)$. This observation, together with the fact that the set of no-signalling correlations form a convex polytope, means that we can use linear programming to find the optimal no-signalling winning probability of an LSSD game. It also means that there is always an optimal strategy at one of the extreme points of the no-signalling polytope.

This last fact is what Majenz et al. used to prove that there exists no probability distribution P_{XAB} with binary x, a and b, such that the corresponding LSSD game can be won with higher probability using no-signalling strategies [12, Proposition 3.3]. They showed that none of the no-signalling correlations at the extreme points of the no-signalling polytope could ever perform better than the simple classical strategy of outputting the most likely value for x. We do something similar in the next chapter. However, it turns out that this argument is not enough in the tripartite case, so we have to turn to numerical analysis to finish the argument.

4. Three-party binary LSSD

In this chapter, we will show (partially numerically) that there exist no probability distribution P_{XABC}, where x, a, b and c are all binary, such that the corresponding LSSD game can be won with higher probability using no-signalling strategies than with classical strategies. We will do this by showing that none of the no-signalling correlations at the extreme points of the no-signalling polytope can ever perform better than classical strategies.

In the next section we discuss some results on optimal classical and no-signalling strategies. These results allow us to discard some no-signalling strategies of which we know that they cannot perform better than classical strategies. For the strategies that are left, we turn to linear programming to numerically show that they also cannot perform better than classical.

4.1. Some results on optimal strategies

This first lemma is an extension of the classical part of Lemma 3.2 in the paper by Majenz et al. [12]. It gives a list of all deterministic strategies (or more accurately: winning probability thereof) we need to consider in finding the optimal classical winning probability. The proof of this lemma relies on the relatively simple observation that the players should have equal output sets (sets consisting of all things they could possibly output according to their strategy).

Lemma 4.1. *Let P_{XABC} be a probability distribution over $\mathscr{X} \times \mathscr{A} \times \mathscr{B} \times \mathscr{C}$ with $\mathscr{A} = \mathscr{B} = \mathscr{C} = \{0,1\}$ and $\mathscr{X} = [d]$, $d \geq 2$. The classical winning probability for P_{XABC} is given by*

$$\omega_c(\mathsf{X}|\mathsf{A};\mathsf{B};\mathsf{C})_P = \max_{\substack{s,t \\ s \neq t}} \max \left\{ \begin{matrix} P_{\mathsf{X}}(s), \\ P_{\mathsf{XABC}}(s,0,0,0) + P_{\mathsf{XABC}}(t,1,1,1), \\ P_{\mathsf{XABC}}(s,1,0,0) + P_{\mathsf{XABC}}(t,0,1,1), \\ P_{\mathsf{XABC}}(s,0,1,0) + P_{\mathsf{XABC}}(t,1,0,1), \\ P_{\mathsf{XABC}}(s,0,0,1) + P_{\mathsf{XABC}}(t,1,1,0) \end{matrix} \right\} \tag{4.1}$$

Proof. First, remember that we only have to consider deterministic strategies (see Subsection 3.2.1). Any deterministic strategy can be represented by three functions $f, g, h \colon \{0,1\} \to \mathscr{X}$. Given such a strategy, the probability of winning is given by

$$\sum_{x,a,b,c} P_{\mathsf{XABC}}(x,a,b,c)\delta[f(a) = g(b) = h(c) = x] = \sum_{a,b,c} P_{\mathsf{XABC}}(f(a),a,b,c)\delta[f(a) = g(b) = h(c)]. \tag{4.2}$$

Notice that there is always an optimal strategy such that $\{f(0), f(1)\} = \{g(0), g(1)\} = \{h(0), h(1)\}$. Suppose, for example, that for some a^*, we have that $f(a^*) \notin \{g(0), g(1)\}$. It follows that $\delta[f(a^*) = g(b) = h(c)] = 0$ for all b, c. Changing Alice's output on input a^*, such that $f(a^*) \in \{g(0), g(1)\}$, causes $\delta[f(a^*) = g(b) = h(c)]$ to possibly be equal to 1 for some b, c. This introduces non-negative terms in the sum of (4.2), while not losing any others, thereby increasing the winning probability.

There are 5 possible ways in which we have $\{f(0), f(1)\} = \{g(0), g(1)\} = \{h(0), h(1)\}$. The first is that all players ignore their input and always output some fixed s. In this case, the probability of winning is given by

$$\sum_{a,b,c} P_{\mathsf{XABC}}(s, a, b, c) = P_X(s).$$

This gives the first term in formula (4.1) The other 4 possibilities are when they all take their input into account:

- $f(0) = g(0) = h(0)$ and $f(1) = g(1) = h(1)$ or,
- $f(1) = g(0) = h(0)$ and $f(0) = g(1) = h(1)$ or,
- $f(0) = g(1) = h(0)$ and $f(1) = g(0) = h(1)$ or,
- $f(0) = g(0) = h(1)$ and $f(1) = g(1) = h(0)$.

defining $f(0) =: s$ and $f(1) =: t$, the winning probability in each of these cases is equal to a term in formula (4.1). □

Whereas the previous lemma reduced the number of interesting deterministic strategies, the next lemma and its corollary will do so for no-signalling strategies.

Lemma 4.2. *Let P be a probability distribution over $\mathscr{X} \times \mathscr{A}_1 \times \cdots \times \mathscr{A}_m$ with $|\mathscr{X}| = d$ and $d \geq 2$. Let Q be a no-signalling strategy for which*

$$Q(x, \ldots, x \big| a_1, \ldots, a_m) \leq \frac{1}{d}$$

holds for all $x \in \mathscr{X}$ and $a_1 \in \mathscr{A}_1, \ldots, a_m \in \mathscr{A}_m$. Then its winning probability in the LSSD game defined by P is at most the best classical winning probability:

$$\sum_{\substack{x \in \mathscr{X} \\ a_1 \in \mathscr{A}_1, \ldots, a_m \in \mathscr{A}_m}} P(x, a_1, \ldots, a_m) Q(x, \ldots, x \big| a_1, \ldots, a_m) \leq \omega_c(\mathsf{X} \big| \mathsf{A}_1; \ldots; \mathsf{A}_m)_P$$

Proof. The proof relies on the simple fact that the m players can always use deterministic strategies to win with at least probability $1/d$ by ignoring their inputs and guessing the value of x to be the one most likely in P. The probability that the referee picks a certain value x is given by $P(x) = \sum_{a \in \mathscr{A}_1 \times \cdots \times \mathscr{A}_m} P(x, a)$ and since $\sum_x P(x) = 1$, there exists an $x^* \in \mathscr{X}$ such that $P(x^*) \geq 1/d$. We conclude that $\omega_c(\mathsf{X} \big| \mathsf{A}_1; \ldots; \mathsf{A}_m)_P \geq 1/d$.

We use the previous argument to finish the proof:

$$\begin{aligned}&\sum_{\substack{x\in\mathscr{X}\\ a_1\in\mathscr{A}_1,\dots,a_m\in\mathscr{A}_m}} P(x,a_1,\dots,a_m)Q(x,\dots,x\big|a_1,\dots,a_m)\\ \leq&\frac{1}{d}\sum_{\substack{x\in\mathscr{X}\\ a_1\in\mathscr{A}_1,\dots,a_m\in\mathscr{A}_m}} P(x,a_1,\dots,a_m)=\frac{1}{d}\leq\omega_{\mathrm{c}}(\mathsf{X}\big|\mathsf{A}_1;\dots;\mathsf{A}_m)_P.\end{aligned}$$

□

Corollary 4.3. *Consider an LSSD problem with m players defined by a distribution P for which $\omega_{\mathrm{c}}(\mathsf{X}|\mathsf{A}_1;\dots;\mathsf{A}_m)_P < \omega_{\mathrm{ns}}(\mathsf{X}|\mathsf{A}_1;\dots;\mathsf{A}_m)_P$. There is an optimal no-signalling strategy Q at one of the vertices of the no-signalling polytope, such that there exist $x\in\mathscr{X}$, with $|\mathscr{X}|=d$, and $a_1\in\mathscr{A}_1,\dots,a_m\in\mathscr{A}_m$ for which $Q(x,\dots,x|a_1,\dots,a_m)>1/d$.*

Proof. Since the set of all no-signalling strategies is a convex polytope, and the winning probability of a no-signalling strategy is a linear function, we know that the optimal winning probability is achieved by a strategy Q at one of the vertices of the polytope (see Section 2.3). We also know that there exist $x\in\mathscr{X}$ and $a_1\in\mathscr{A}_1,\dots,a_m\in\mathscr{A}_m$ such that $Q(x,\dots,x|a_1,\dots,a_m)>1/d$, because otherwise this strategy would not achieve winning probability higher than $\omega_{\mathrm{c}}(\mathsf{X}|\mathsf{A}_1;\dots;\mathsf{A}_m)_P$ by Lemma 4.2. □

In the case of two players, we would now be done in showing that there is no binary LSSD game with a gap between no-signalling and classical winning probabilities, since all no-signalling correlations at the extreme points of the no-signalling polytope satisfy the conditions of Lemma 4.1 [1, Theorem 1]. We will see in the next section that for three players, this is not the case. However, Corollary 4.3 is still very useful as it eliminates many of the no-signalling strategies.

4.2. No gap

In this section, our goal is to show that $\omega_{\mathrm{ns}}(\mathsf{X}|\mathsf{A};\mathsf{B};\mathsf{C})_P=\omega_{\mathrm{c}}(\mathsf{X}|\mathsf{A};\mathsf{B};\mathsf{C})_P$ for all probability distributions P_{XABC} over binary numbers (see the code for this thesis [10]). This is obviously equivalent to showing that

$$\sup_P \omega_{\mathrm{ns}}(\mathsf{X}|\mathsf{A};\mathsf{B};\mathsf{C})_P-\omega_{\mathrm{c}}(\mathsf{X}|\mathsf{A};\mathsf{B};\mathsf{C})_P=0.$$

Now we have turned the problem into an optimization problem. It is, however, not possible to solve this problem using a single linear program, since the target function is not linear: the target function is the maximum of the difference between two sets. Luckily, using Corollary 4.3 and some smart tricks, we can solve this problem using multiple linear programs.

First of all, it should be noted that the set of all probability distributions P_{XABC} form a convex polytope in $\mathbb{R}^n$. The polytope is defined by the following linear constraints:

$$\forall x,a,b,c\;\; P_{\mathsf{XABC}}(x,a,b,c)\geq 0,$$

and

$$\sum_{x,a,b,c} P_{\mathsf{XABC}}(x,a,b,c) = 1.$$

Apart from the variables that describe a probability distribution, we also add two variables c_{d} and c_{ns} to the linear program, which represent $\omega_{\mathrm{c}}(\mathsf{X}|\mathsf{A};\mathsf{B};\mathsf{C})_P$ and $\omega_{\mathrm{ns}}(\mathsf{X}|\mathsf{A};\mathsf{B};\mathsf{C})_P$ respectively. These two variables should satisfy the following constraints:

$$c_{\mathrm{d}} \geq \sum_{x,a,b,c} P_{\mathsf{XABC}}(x,a,b,c) Q_{\mathrm{d}}(x,x,x|a,b,c)$$

for all deterministic strategies Q_{d} and

$$c_{\mathrm{ns}} \geq \sum_{x,a,b,c} P_{\mathsf{XABC}}(x,a,b,c) Q_{\mathrm{ns}}(x,x,x|a,b,c) \tag{4.3}$$

for all no-signalling strategies Q_{ns} at the vertices of the no-signalling polytope.

Now, the problem is to maximize $c_{\mathrm{ns}} - c_{\mathrm{d}}$, which is a linear function in two variables, so we can use a linear program. However, since we have not put an upper bound on c_{ns}, this problem is obviously unbounded. We can work around this issue by setting one of the constraints of (4.3) to an equality constraint. Solving the linear program with one of these constraints set to an equality constraint gives us the maximum gap under the assumption that the corresponding no-signalling strategy is the best strategy. By considering all no-signalling strategies in this way we can find the maximum gap between classical and no-signalling winning probabilities.

All that is left is to find the no-signalling strategies at the extreme points of the no-signalling polytope. We can find them using a python package called *cddlib*, which is based on a C package under the same name [7]. Similar to linear programs, we can define some linear constraints and then this package provides all the exact vertices of the corresponding polytope. In this case we need constraints on some strategy Q to be a conditional probability distribution on $\mathscr{X}^3 \times \mathscr{A} \times \mathscr{B} \times \mathscr{C}$ and constraints such that Q is no-signalling (where we can use Lemma 2.2 to omit redundant constraints). We find that this no-signalling polytope has 53856 extreme points, which is in line with the findings of the paper by Pironio et al. [15, Section 2.2].

Since the number of no-signalling strategies is quite large, we would like to reduce this number to reduce the number of linear programs we need to solve. Using Corollary 4.3, we can greatly reduce the number of relevant no-signalling strategies, since we know that there are always optimal strategies of a specific form. Corollary 4.3 reduces the number of relevant no-signalling strategies from 53856 to 174. We can also use Lemma 4.1 to reduce the number of relevant deterministic strategies from $2^6 = 64$ to 10.

Finally we have everything we need to find the maximum gap between classical and no-signalling probabilities. Solving the linear programs gives us numerical evidence that the maximum gap is 0, meaning that there is no binary LSSD game for three players such that no-signalling resources can improve the winning probability.

One could possibly prove this last statement by analysing each of the 174 remaining no-signalling strategies and arguing that they can never achieve higher winning probabilities than classical strategies. One argument could be to give an explicit classical strategy that performs better than the no-signalling strategy. However, such a classical strategy could depend on the specific probability distribution defining the LSSD game. For the purposes of this thesis, we are satisfied with the numerical argument.

5. The BSC game

In this chapter we will take a closer look at the BSC game, a specific example of a classical LSSD game described in the paper by Majenz et al. [12, Example 1]. Here we give a different but equivalent description of the game. The referee starts by generating a random bit x (0 or 1, both with probability $1/2$). The referee sends this bit to Alice and Bob over two identical and independent binary symmetric channels, so both with the same error probability α. Alice and Bob's inputs are the outputs of their binary symmetric channel.

The optimal winning probability for this game is given by: [12]

$$\omega_{\mathrm{c}}(\mathsf{X}|\mathsf{A};\mathsf{B}) = \omega_{\mathrm{ns}}(\mathsf{X}|\mathsf{A};\mathsf{B}) = \begin{cases} (1-\alpha)^2 & 0 \leq \alpha \leq 1 - \frac{1}{\sqrt{2}} \\ \frac{1}{2} & 1 - \frac{1}{\sqrt{2}} \leq \alpha \leq \frac{1}{2} \end{cases}$$

The winning probability in the first segment $\alpha \in [0, 1 - 1/\sqrt{2}]$ is achieved by Alice and Bob guessing x to be the same as their input. This strategy is not surprising, since when α is small, both their bits are likely to not have been flipped. In fact, if Alice were to be playing this game alone, this strategy would be optimal for all $\alpha \leq 1/2$. However, Alice and Bob should not just strive to be correct individually, but also simultaneously. That is the reason why, for $\alpha \in [1 - 1/\sqrt{2}, 1/2]$, the optimal strategy is to both output some fixed bit, regardless of the input.

Even more interesting things start happening when we play multiple simultaneous copies of this game, where the players need to win all of them. Playing n simultaneous copies can be thought of as the referee uniformly generating a bitstring x of length n and sending it to Alice and Bob by n consecutive uses of their channels. Remember from Section 2.4 that multiple consecutive uses of a channel $P_{\mathsf{A}|\mathsf{X}}$ can be modelled by a new channel $P_{\mathsf{A}|\mathsf{X}}^{\times n}$.

Majenz et al. have shown that when playing two of these games simultaneously, Alice and Bob can attain a better winning probability than playing two games consecutively. In the next sections we analyse the optimal classical and no-signalling winning probabilities for two and three simultaneous copies. We then discuss good classical and no-signalling strategies for n copies. To do these analyses, we first mention a useful result on optimal classical strategies. By a symmetric strategy we mean that Alice and Bob follow the same local strategy.

Theorem 5.1. *Let P_{XAB} be a distribution over $\mathscr{X} \times \mathscr{A} \times \mathscr{B}$, with $\mathscr{A} = \mathscr{B}$, satisfying the following:*

(i) The marginal distribution P_{X} over $\mathscr{X}$ is uniform.

(ii) $P_{\mathsf{AB|X}} = P_{\mathsf{A|X}} P_{\mathsf{B|X}}$.

(iii) $P_{\mathsf{A|X}} = P_{\mathsf{B|X}}$.

Then there is a symmetric deterministic strategy which is optimal for the classical LSSD game defined by P_{XAB}.

Proof. Let two functions $f\colon \mathscr{A} \to \mathscr{X}$ and $g\colon \mathscr{B} \to \mathscr{X}$ define a deterministic strategy. We prove that either Alice and Bob both performing f or both performing g can only increase the winning probability. Note that Alice and Bob can perform the same strategy, since $\mathscr{A} = \mathscr{B}$.

The winning probability of the strategy defined by f and g is given by

$$\begin{aligned}
&\sum_{\substack{x \in \mathscr{X} \\ a \in \mathscr{A}, b \in \mathscr{B}}} P_{\mathsf{XAB}}(x, a, b) \delta[f(a) = g(b) = x] \\
&\overset{(i)}{=} \frac{1}{|\mathscr{X}|} \sum_{\substack{x \in \mathscr{X} \\ a \in \mathscr{A}, b \in \mathscr{B}}} P_{\mathsf{AB|X}}(a, b|x) \delta[f(a) = g(b) = x] \\
&\overset{(ii)}{=} \frac{1}{|\mathscr{X}|} \sum_{x \in \mathscr{X}} \left(\sum_{a \in \mathscr{A}} P_{\mathsf{A|X}}(a|x) \delta[f(a) = x] \right) \left(\sum_{b \in \mathscr{B}} P_{\mathsf{B|X}}(b|x) \delta[g(b) = x] \right) \\
&\overset{(iii)}{=} \frac{1}{|\mathscr{X}|} \sum_{x \in \mathscr{X}} P_{\mathsf{A|X}}(f^{-1}(x)|x) P_{\mathsf{A|X}}(g^{-1}(x)|x).
\end{aligned}$$

($f^{-1}(x)$ and $g^{-1}(x)$ might be sets.)

Now write $q_f(x) := P_{\mathsf{A|X}}(f^{-1}(x)|x)$ and $q_g(x) := P_{\mathsf{A|X}}(g^{-1}(x)|x)$. Notice that q_f and q_g are vectors indexed by $x \in \mathscr{X}$, so we can write the winning probability as an inner product of these vectors:

$$\frac{1}{|\mathscr{X}|} \langle q_f, q_g \rangle . \tag{5.1}$$

The Cauchy-Schwarz inequality tells us that

$$| \langle q_f, q_g \rangle |^2 \leq \langle q_f, q_f \rangle \langle q_g, q_g \rangle ,$$

so we cannot have $\langle q_f, q_g \rangle > \langle q_f, q_f \rangle$ and $\langle q_f, q_g \rangle > \langle q_g, q_g \rangle$. Therefore, we can conclude that Alice and Bob either both performing f or both performing g does not decrease the winning probability given in formula (5.1). Now suppose we picked f and g to form an optimal strategy, then by the previous statement, we immediately find a symmetric deterministic strategy that is also optimal. □

5.1. Two copies

In this section we compare the winning probability for no-signalling strategies for two copies of the game to the deterministic winning probability. We calculated the winning probabilities numerically and the result is shown in Figure 5.1. Most notable is that both the graph with the winning probabilities for deterministic strategies and for no-signalling strategies consist of three segments (instead of two) and in the middle segment for no-signalling, there is a gap between classical and no-signalling winning probabilities.

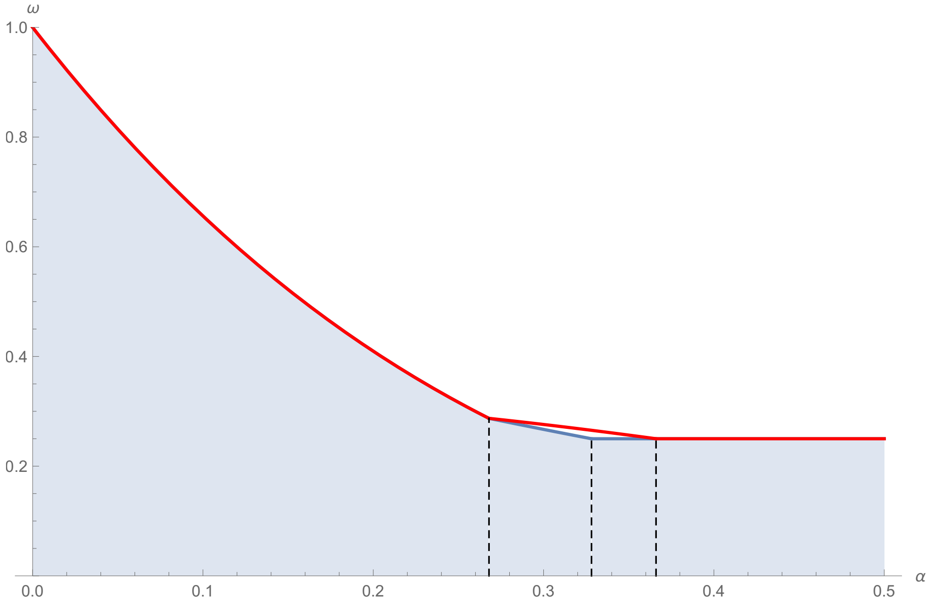


Figure 5.1.: This graph shows the optimal classical (blue) and no-signalling (red) winning probabilities for two copies of the BSC game. The results were found by brute-forcing over all symmetric strategies for the classical winning probability and solving a linear program for the no-signalling winning probability (see code [10]).

The first segment of both graphs completely overlaps. In this segment the best strategy for Alice and Bob is to output their input. Again, this strategy is not so surprising, since with low noise, one expects their input to be the initially chosen bitstring x. The winning probability in this segment is exactly the probability that no bits were flipped for either party, which happens with probability $(1-\alpha)^4$.

For large α (close to 1/2), both the classical and no-signalling winning probabilities are equal to 1/4. Just like when playing one copy, the strategy that achieves this winning probability is to output some fixed bitstring regardless of the input.

The middle segments of both graphs are the most interesting. We start with the classical case. In this segment it turns out that the best strategy is to output 00 if the input contains a 0 and 11 otherwise. The winning probability of this strategy is given by [12, Example 1]

$$\frac{1}{4}(1-\alpha^2)^2 + \frac{1}{4}(1-\alpha)^4.$$

Notice that for small α, the input of the players gives a lot of information on x, so the players take their input into account. As α grows larger, the input bitstrings contain less information on x, so it becomes more important for the players to output the same thing, while taking their input into account less. Following this reasoning, it becomes more intuitive as to why this strategy is optimal in segment 2.

In the second segment of the no-signalling graph, the best strategy is given by

$$Q_2(x, y|a, b) = \begin{cases} \frac{1}{3} & \text{when } (x = y \text{ or } x \oplus b = 11 = y \oplus a) \text{ and } (x \oplus a \neq 11 \neq y \oplus b) \\ 0 & \text{otherwise.} \end{cases} \tag{5.2}$$

This strategy has winning probability $(1-\alpha^2)^2/3$ (see Section 5.3.2, where we discuss this strategy more).

5.2. Three copies

Now let us consider three simultaneous copies of the BSC game. Expanding the problem to three copies creates problems in finding the best deterministic strategy. Even considering just symmetric strategies, there are $8^8 = 2^{24}$ possibilities. This enormous number makes it very slow to find an optimal deterministic strategy for any given α, let alone a large subset of possible α's. Therefore, we first consider no-signalling strategies. These can be efficiently found using a linear program. We will then use the results to hand-pick values for α for which we will find the best deterministic strategies.

Again, we can find the winning probabilities numerically. The results of this search are shown in figure 5.2. This time there are 4 distinct segments.

In the first segment the best strategy is still to output one's input and the best strategy in the last segment is to output a fixed bitstring. These strategies have winning probability $(1-\alpha)^6$ and $1/8$ respectively.

In the second segment we found two no-signalling strategies that achieved the same winning probability. We were unable to find an expression for these strategies, but in Section 5.3.2 we discuss a no-signalling strategy that also achieves the same winning probability. However, there is also a deterministic strategy that achieves this winning probability. In this deterministic strategy both players output the all-one string if they receive an input with more zeros than ones, and they output the all-zero string otherwise. The winning probability in this segment is

$$\frac{1}{4}\left((1-\alpha)^3 + 3\alpha(1-\alpha)^2\right)^2.$$

(See Section 5.3.1.)

Notice that the deterministic strategy just discussed is very similar to the one in segment two of the two-copy case. Only, in that case we had an even number of bits, so when there were an equal amount of ones as zeros, we picked the output to be 00.

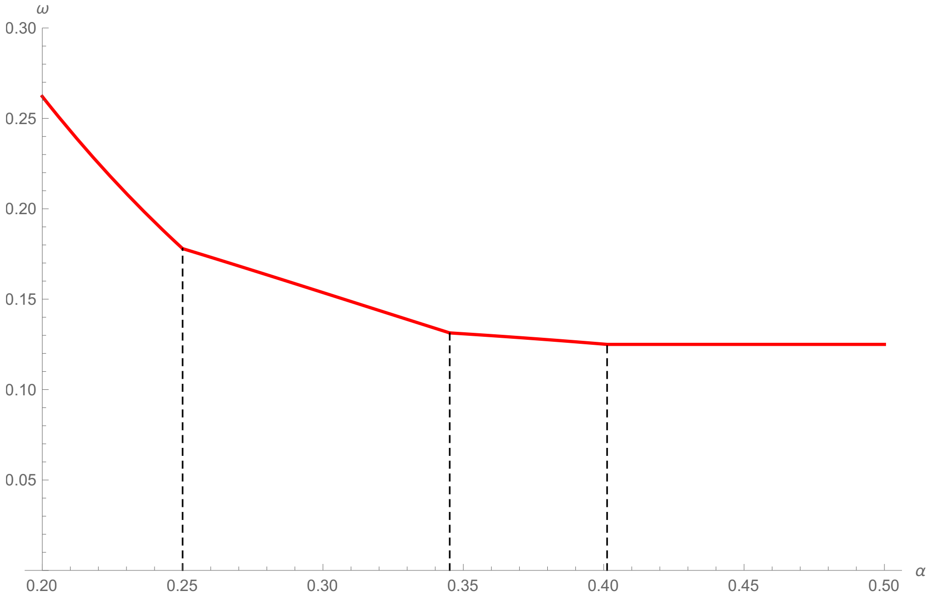


Figure 5.2.: The winning probabilities of no-signalling strategies in three simultaneous copies of the BSC game as a function of the error probability $\alpha \in [0.2, 0.5]$ (see code [10]).

In segment three we found a no-signalling strategy very similar to the one we found for two copies of the game:

$$Q_3(x, y|a, b) = \begin{cases} \frac{1}{7} & \text{when } (x = y \text{ or } x \oplus b = 111 = y \oplus a) \text{ and } (x \oplus a \neq 111 \neq y \oplus b) \\ 0 & \text{otherwise.} \end{cases} \tag{5.3}$$

This strategy achieves winning probability $(1 - \alpha^3)^2/7$. In this segment, we were unable to find a deterministic strategy that achieves the same winning probability. So, again, there is likely a gap between the classical and no-signalling winning probabilities.

5.3. n Copies

In this section, we will look to find classes of good strategies, both classical and no-signalling, for n simultaneous copies of the BSC game.

5.3.1. Classical strategies

We have already seen some similarities in classical strategies between one, two and three copies of the game. For small α, the best strategy is always to output the input (identity strategy). For α close to $1/2$ the best strategy is to output some fixed bitstring regardless of the input (constant strategy). The winning probabilities of these strategies for n copies are $(1-\alpha)^{2n}$ and 2^{-n}. For two and three copies, we also found similar strategies "in between" the identity and constant strategies. These strategies can also be extended to n copies: outputting 0^n if the input contains at least as many zeros as ones and outputting 1^n otherwise (majority strategy). For odd n, the winning probability of the majority strategy is given by

$$\frac{1}{2^{n-1}}\left(\sum_{i=0}^{(n-1)/2}\binom{n}{i}\alpha^i(1-\alpha)^{n-i}\right)^2 . \tag{5.4}$$

Notice how the majority strategy is very similar to the repetition code for the BSC described in Section 2.4.1. In fact, the majority strategy is exactly the same as first decoding and then encoding the result, or in other words, decoding directly to the codeword. Also, the term in between brackets in (5.4) is exactly the average success probability of this code.

In the next example we explore the idea of using error-correcting codes to define classical strategies some more, by considering the hamming code for 7 simultaneous copies of the BSC game.

Example 5.2. Consider the following strategy for 7 copies of the BSC game based on the Hamming code: both players perform the correction part of the Hamming code on their input and output the result (this is the same as decoding and then encoding again). It is obvious that the players win if and only if the initial bitstring x is in the range of the encode function and the decoding of both players was successful. This observation results in the following winning probability:

$$\frac{2^4}{2^7}\left((1-\alpha)^7+7\alpha(1-\alpha)^6\right)^2 .$$

It turns out that this Hamming code strategy is strictly better for a large range of α than the identity, constant and majority strategy for 7 copies of the game. This confirms the idea that error correcting codes define good classical strategies. In general, an (n,k,d) code for the BSC defines a classical strategy that achieves winning probability of at least

$$\frac{2^k}{2^n}\left(\sum_{i=0}^{\lfloor (d-1)/2\rfloor}\binom{n}{i}\alpha^i(1-\alpha)^{n-i}\right)^2 .$$

5.3.2. No-signalling strategies

For two and three copies of the BSC game, we found the optimal no-signalling strategies Q_2 and Q_3 (described in (5.2) and (5.3)). We can extend these no-signalling strategies to n copies as follows:

$$Q(x,y|a,b)=\begin{cases}\frac{1}{2^n-1} & \text{when } (x=y \text{ or } x\oplus b=1^n=y\oplus a) \text{ and } (x\oplus a\neq 1^n\neq y\oplus b)\\ 0 & \text{otherwise.}\end{cases} \tag{5.5}$$

There is, however, a more intuitive way to describe this no-signalling correlation. Alice and Bob both have a set of possible outputs, which consists of every bitstring apart from the one opposite of their input. We then create pairs of elements of their output sets (each pair consists of an element of Alice's output set and an element of Bob's output sets), such that each element occurs in exactly one pair and every element that occurs in both sets is paired up with itself. We then uniformly pick one of the pairs to be Alice's and Bob's guesses. An example of this process is shown in Figure 5.3.

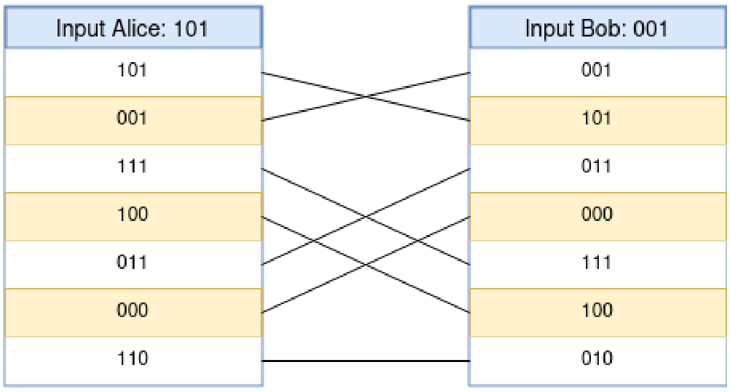


Figure 5.3.: An example of a pairing of elements between the output sets of Alice and Bob, for three simultaneous copies. Each line represents a pair, and at the end of the process on pair is chosen uniformly.

This formulation makes it obvious that we can define a more general class of no-signalling strategies: instead of the output sets consisting of everything apart from the opposite of the input, we can let the output sets consist of all bitstrings within Hamming distance d from the input. We can then pair up the elements from the output sets and say that each of those pairs is output with equal probability. Again, if an element occurs in both lists, we pair it with itself. This description defines a no-signalling strategy, since Alice and Bob always output each of the elements of their output sets with the same probability, regardless of the input of the other. We denote a no-signalling strategy for n copies of the BSC game defined by Hamming distance d by Q_n^d. Note that for $d \in \{1, \ldots, n-2\}$ the strategy Q_n^d is not unique, but they all achieve the same winning probability.

Let us find the winning probability of a strategy Q_n^d. Suppose that x is the bitstring generated by the referee. The only way the players could output the combination (x, x) is if both $d(x, a) \leq d$ and $d(x, b) \leq d$, in which case it is outputted with probability $\left(\sum_{i=0}^{d} \binom{n}{d}\right)^{-1}$, since the sum is the size of their output sets. The probability that a lies within distance d from x is $\sum_{i=0}^{d} \binom{n}{i} \alpha^i (1-\alpha)^{n-i}$. We conclude that the winning probability of Q_n^d is given by

$$\frac{1}{\sum_{i=0}^{d} \binom{n}{i}} \left(\sum_{i=0}^{d} \binom{n}{i} \alpha^i (1-\alpha)^{n-i} \right)^2 .$$

It turns out that all the optimal winning probabilities for one, two and three simultaneous copies of the BSC game can be achieved by a strategy of the form Q_n^d. If we pick $d = 0$ we get exactly the identity strategy. If we pick $d = n$, we get the average of all possible constant strategies (and by linearity, this achieves the same winning probability as a constant strategy). If we pick $d = n-1$, we get exactly the strategy defined in (5.5). This strategy achieves winning probability

$$\frac{1}{2^n - 1}\left(\sum_{i=0}^{n}\binom{n}{i}\alpha^i(1-\alpha)^{n-i} - \alpha^n\right)^2 = \frac{1}{2^n-1}\left(1-\alpha^n\right)^2 .$$

We are left with segment two for three copies. The strategy Q_3^1 achieves winning probability

$$\frac{1}{4}\left((1-\alpha)^3 + 3\alpha(1-\alpha)^2\right)^2 .$$

This probability is exactly the same winning probability as the majority strategy, which we found to be optimal in this segment. We conclude that all optimal winning probabilities for one, tow and three copies of the game can be achieved by a strategy of the form Q_n^d

Notice that this class of no-signalling strategies is very similar to list decoding: on their input, both players create a list of bitstrings that were most likely sent. The eventual output is just one element of this list.

In the next chapter, we consider LSSD games defined by a general channel $P_{\mathsf{A}|\mathsf{X}}$.

6. Channel games

In the previous chapter, we have seen how the BSC defines an LSSD game. In this chapter, we will see that any channel defines an LSSD game. For n simultaneous copies of these games, we will discuss classical strategies based on error-correcting codes and no-signalling strategies based on list-decoding schemes. We will also take a look at what happens when the number of simultaneous copies approaches infinity. For this last part it is important to note that for any non-local game, with optimal no-signalling winning probability smaller than 1, the winning probability when playing n simultaneous copies goes to 0 exponentially [2, Theorem 16]. This is why we will be considering the limit of the exponent of the winning probability.

Let $\mathscr{X}$ and $\mathscr{A}$ be finite sets and let $P_{\mathsf{A|X}}$ be a channel from $\mathscr{X}$ to $\mathscr{A}$. The channel game defined by this channel is given by the probability distribution

$$P_{\mathsf{XAB}} = P_{\mathsf{X}} P_{\mathsf{A|X}} P_{\mathsf{B|X}}$$

with P_{X} the uniform distribution over $\mathscr{X}$, $\mathscr{A} = \mathscr{B}$, and $P_{\mathsf{B|X}} = P_{\mathsf{A|X}}$. Playing n simultaneous copies of this channel game is the same as playing the channel game defined by the channel $P_{\mathsf{A|X}}^{\times n}$, which can be thought of as the referee generating a string $x^n \in \mathscr{X}^n$ and sending it to Alice and Bob by n uses of their channels. Note that channel games satisfy the conditions of Theorem 5.1, which means that we only need to consider symmetric deterministic strategies.

Throughout this chapter, we assume that all encoding functions are injective.

6.1. Classical strategies

Let us now expand the idea of using codes to define classical strategies for multiple simultaneous copies, just like we did for the BSC game, to this more general channel game. Let $(\mathrm{Enc}, \mathrm{Dec})$ be an (n, M, α) code for the $P_{\mathsf{A|X}}$ channel, which means that for each $m \in [M]$ we have

$$\sum_{\substack{a^n \in \mathscr{A}^n: \\ \mathrm{Dec}(a^n)=m}} P_{\mathsf{A|X}}^{\times n}(a^n | \, \mathrm{Enc}(m)) = P_{\mathsf{A|X}}^{\times n}(\mathrm{Dec}^{-1}(m) | \, \mathrm{Enc}(m)) \geq \alpha \tag{6.1}$$

We define the strategy f, for n copies of the channel game, used by both players as $f := \mathrm{Enc} \circ \mathrm{Dec}$. This strategy can be interpreted as the players decoding directly to the codeword of a message instead of to the message itself.

Using (6.1) and some analysis we did in Theorem 5.1, we can find a lower bound on the winning probability of the strategy given by f:

$$\frac{1}{|\mathscr{X}|^n} \sum_{x^n \in \mathscr{X}^n} P^{\times n}_{\mathsf{A}|\mathsf{X}}(f^{-1}(x^n)|x^n)^2 = \frac{1}{|\mathscr{X}|^n} \sum_{x^n \in \mathscr{X}^n} \left(\sum_{a^n \in \mathscr{A}^n : f(a^n) = x^n} P^{\times n}_{\mathsf{A}|\mathsf{X}}(a^n|x^n) \right)^2 \tag{6.2}$$

$$\overset{(6.1)}{\geq} \frac{1}{|\mathscr{X}|^n} \sum_{x^n \in Im(\text{Enc})} \alpha^2 = \frac{M}{|\mathscr{X}|^n} \alpha^2. \tag{6.3}$$

Notice that there is a trade-off between the success probability and the number of messages. We simultaneously want the success probability and the number of messages to be large. However, Increasing one necessarily means decreasing the other.

The lower bound given in (6.3) is useful for the proof of the result in the next section.

6.2. Limit behaviour of classical strategies

In this section we give an explicit expression for the limit of the exponent of the optimal classical winning probabilities for n simultaneous copies of a channel game.

Theorem 6.1. *Let $P_{\mathsf{A}|\mathsf{X}}$ be a channel and let $P^{\times n}_{\mathsf{XAB}}$ be the probability distribution defining the channel game corresponding to the channel $P^{\times n}_{\mathsf{A}|\mathsf{X}}$. We have*

$$\lim_{n\to\infty} \frac{\log(\omega_{\mathrm{c}}(\mathsf{X}^n|\mathsf{A}^n;\mathsf{B}^n)_{P^{\times n}})}{n} = \max_{Q_{\mathsf{XA}}} I(X;A)_Q - 2D(Q_{\mathsf{A}|\mathsf{X}} \| P_{\mathsf{A}|\mathsf{X}} \mid Q_{\mathsf{X}}) - \log(|\mathscr{X}|)$$

To prove Theorem 6.1, we need two lemmas. The first lemma will help us prove achievability of the limit. We leave its proof to Appendix A. The second lemma is a lemma from Dueck and Körner [5, Lemma 5], and will help us prove that the limit is optimal.

Lemma 6.2. *Let $P_{\mathsf{A}|\mathsf{X}}$ be a channel, Q_{XA} a probability distribution over $\mathscr{X} \times \mathscr{A}$ and $\delta > 0$. For large enough n, there exists an*

$$\left(n, 2^{n(I(X;A)_Q - \delta)}, \frac{2^{-nD(Q_{\mathsf{A}|\mathsf{X}} \| P_{\mathsf{A}|\mathsf{X}} | Q_{\mathsf{X}})}}{\text{poly}(n)} \right)$$

code for the channel $P_{\mathsf{A}|\mathsf{X}}$.

Lemma 6.3. *For any $(n, 2^{nR}, 2^{-n\zeta})$ code for $P_{\mathsf{A}|\mathsf{X}}$, we have*

$$\zeta \geq \min_{Q_{\mathsf{XA}}} D(Q_{\mathsf{A}|\mathsf{X}} \| P_{\mathsf{A}|\mathsf{X}} \mid Q_{\mathsf{X}}) + \max\{R - I(X;A)_Q, 0\} + o(1).$$

Proof of Theorem 6.1. We first prove achievability of the limit using Lemma 6.2. After this part, we prove optimality using Lemma 6.3.

Let Q_{XA} be a distribution over $\mathscr{X} \times \mathscr{A}$ and $\delta > 0$. By Lemma 6.2, for large enough n, there exists an $\left(n, 2^{n(I(X;A)_Q - \delta)}, \frac{2^{-nD(Q_{\mathsf{A|X}} \| P_{\mathsf{A|X}} | Q_{\mathsf{X}})}}{\mathrm{poly}(n)}\right)$ code for $P_{\mathsf{A|X}}$. Let $f = \mathrm{Enc} \circ \mathrm{Dec}$ be the strategy defined by this code. The winning probability of this strategy is at most the optimal classical winning probability, so by using (6.3) we find

$$\omega_{\mathrm{c}}(\mathsf{X}^n|\mathsf{A}^n;\mathsf{B}^n)_{P^{\times n}} \geq \frac{2^{n(I(X;A)_Q - \delta - D(Q_{\mathsf{A|X}} \| P_{\mathsf{A|X}} | Q_{\mathsf{X}}))}}{|\mathscr{X}|^n \, \mathrm{poly}(n)}$$

and therefore

$$\frac{\log(\omega_{\mathrm{c}}(\mathsf{X}^n|\mathsf{A}^n;\mathsf{B}^n)_{P^{\times n}})}{n} \geq I(X;A)_Q - \delta - 2D(Q_{\mathsf{A|X}} \| P_{\mathsf{A|X}} \mid Q_{\mathsf{X}}) - \log(|\mathscr{X}|) - \frac{\log(\mathrm{poly}(n))}{n}. \tag{6.4}$$

Since (6.4) holds for any Q_{XA} and $\delta > 0$, and $\lim_{n\to\infty} \frac{\log(\mathrm{poly(n)})}{n} = 0$ (irrespective of the polynomial), we conclude

$$\lim_{n\to\infty} \frac{\log(\omega_{\mathrm{c}}(\mathsf{X}^n|\mathsf{A}^n;\mathsf{B}^n)_{P^{\times n}})}{n} \geq \max_{Q_{\mathsf{XA}}} I(X;A)_Q - 2D(Q_{\mathsf{A|X}} \| P_{\mathsf{A|X}} \mid Q_{\mathsf{X}}) - \log(|\mathscr{X}|)$$

Now we prove the inverse inequality. By Theorem 5.1, We can assume that Alice and Bob use the same strategy $f \colon \mathscr{A}^n \to \mathscr{X}^n$. For $x^n \in \mathscr{X}^n$, we define

$$q(x^n) := P_{\mathsf{A|X}}^{\times n}(f^{-1}(x^n)|x^n).$$

($f^{-1}(x^n)$ might be a set.) We can write the winning probability of the strategy defined by f by (see the proof of Theorem 5.1)

$$\frac{1}{|\mathscr{X}|^n} \sum_{x^n \in \mathscr{X}^n} q(x^n)^2. \tag{6.5}$$

Let $\delta > 0$, for each $i \geq 0$, we define

$$\mathcal{R}_i := \{x^n \in \mathscr{X}^n \mid 2^{-n\delta(i+1)} \leq q(x^n) < 2^{-n\delta i}\}.$$

We define a code $(\mathrm{Enc}_i, \mathrm{Dec}_i)$ by $\mathrm{Enc}_i \colon \mathcal{R}_i \to \mathscr{X}^n, x \mapsto x$ (so the messages are the elements of $\mathcal{R}_i$) and

$$\mathrm{Dec}_i(a^n) = \begin{cases} f(a^n) & \text{if } f(a^n) \in \mathcal{R}_i \\ \hat{x}^n & \text{otherwise,} \end{cases}$$

for some $\hat{x}^n \in \mathcal{R}_i$. For $x^n \in \mathcal{R}_i$, we have $P_{\mathsf{A|X}}^{\times n}(\mathrm{Dec}_i^{-1}(x^n)|\,\mathrm{Enc}_i(x^n)) \geq q(x^n) \geq 2^{-n\delta(i+1)}$, so this code is an $(n, |\mathcal{R}_i|, 2^{-n\delta(i+1)})$ code. Now, according to Lemma 6.3, we have

$$\delta(i+1) \geq \min_{Q_{\mathsf{XA}}} D(Q_{\mathsf{A|X}} \| P_{\mathsf{A|X}} \mid Q_{\mathsf{X}}) + \max\left\{\frac{\log|\mathcal{R}_i|}{n} - I(X;A)_Q, 0\right\} + o(1).$$

Suppose that $q(x^n) > 0$, then there exists an $a^n \in \mathscr{A}^n$ such that $q(x^n) \geq P_{\mathsf{A|X}}^{\times n}(a^n|x^n)$, which means that there exist $x \in \mathscr{X}$ and $a \in \mathscr{A}$ such that $q(x^n) \geq (P_{\mathsf{A|X}}(a|x))^n$. We conclude that $q(x^n) \geq 2^{-n\mu}$ where $\mu = \max_{x,a:P_{\mathsf{A|X}}(a|x)>0} -\log(P_{\mathsf{A|X}}(a|x))$.

From the above, it follows that if $i \geq t := \lfloor \frac{\mu}{\delta} \rfloor$, then $\mathcal{R}_i$ is empty. Now we find

$$\begin{aligned}
\sum_{x^n \in \mathscr{X}^n} q(x^n)^2 &= \sum_{i=0}^{t} \sum_{x^n \in \mathcal{R}_i} q(x^n)^2 \\
&\leq \sum_{i=0}^{t} |\mathcal{R}_i| 2^{-2n\delta i} \\
&\leq \sum_{i=0}^{t} 2^{n\left(\frac{\log|\mathcal{R}_i|}{n} - 2\min_{Q_{\mathsf{XA}}}\left(D(Q_{\mathsf{A|X}} \| P_{\mathsf{A|X}} | Q_{\mathsf{X}}) + \max\left\{\frac{\log|\mathcal{R}_i|}{n} - I(X;A)_Q, 0\right\}\right) + o(1) + \delta\right)} \\
&= \sum_{i=0}^{t} 2^{n\left(\frac{\log|\mathcal{R}_i|}{n} + \max_{Q_{\mathsf{XA}}}\left(2\min\left\{I(X;A)_Q - \frac{\log|\mathcal{R}_i|}{n}, 0\right\} - 2D(Q_{\mathsf{A|X}} \| P_{\mathsf{A|X}} | Q_{\mathsf{X}})\right) + o(1) + \delta\right)} \\
&\leq \sum_{i=0}^{t} 2^{n\left(\frac{\log|\mathcal{R}_i|}{n} + \max_{Q_{\mathsf{XA}}}\left(\min\left\{I(X;A)_Q - \frac{\log|\mathcal{R}_i|}{n}, 0\right\} - 2D(Q_{\mathsf{A|X}} \| P_{\mathsf{A|X}} | Q_{\mathsf{X}})\right) + o(1) + \delta\right)} \\
&\leq \sum_{i=0}^{t} 2^{n\left(\max_{Q_{\mathsf{XA}}}\left(I(X;A)_Q - 2D(Q_{\mathsf{A|X}} \| P_{\mathsf{A|X}} | Q_{\mathsf{X}})\right) + o(1) + \delta\right)} \\
&= (t+1) 2^{n\left(\max_{Q_{\mathsf{XA}}}\left(I(X;A)_Q - 2D(Q_{\mathsf{A|X}} \| P_{\mathsf{A|X}} | Q_{\mathsf{X}})\right) + o(1) + \delta\right)} \\
&\leq \left(\frac{\mu}{\delta} + 1\right) 2^{n\left(\max_{Q_{\mathsf{XA}}}\left(I(X;A)_Q - 2D(Q_{\mathsf{A|X}} \| P_{\mathsf{A|X}} | Q_{\mathsf{X}})\right) + o(1) + \delta\right)}.
\end{aligned}$$

Since the previous holds for any strategy f and using (6.5), we find, with $\delta = 1/n$,

$$\omega_{\mathrm{c}}(\mathsf{X}^n | \mathsf{A}^n; \mathsf{B}^n)_{P^{\times n}} \leq \frac{1}{|\mathscr{X}|^n} (n\mu + 1)\, 2^{n\left(\max_{Q_{\mathsf{XA}}}\left(I(X;A)_Q - 2D(Q_{\mathsf{A|X}} \| P_{\mathsf{A|X}} | Q_{\mathsf{X}})\right) + o(1) + 1/n\right)}.$$

Now we have

$$\frac{\log(\omega_{\mathrm{c}}(\mathsf{X}^n | \mathsf{A}^n; \mathsf{B}^n)_{P^{\times n}})}{n} \leq \max_{Q_{\mathsf{XA}}} \left(I(X;A)_Q - 2D(Q_{\mathsf{A|X}} \| P_{\mathsf{A|X}} \mid Q_{\mathsf{X}})\right) - \log|\mathscr{X}| + o(1) + \frac{\log(n\mu + 1) + 1}{n}.$$

Obviously, $\frac{\log(n\mu+1)+1}{n} + o(1)$ goes to 0 as n goes to infinity, so we find

$$\lim_{n \to \infty} \frac{\log(\omega_{\mathrm{c}}(\mathsf{X}^n | \mathsf{A}^n; \mathsf{B}^n)_{P^{\times n}})}{n} \leq \max_{Q_{\mathsf{XA}}} \left(I(X;A)_Q - 2D(Q_{\mathsf{A|X}} \| P_{\mathsf{A|X}} \mid Q_{\mathsf{X}})\right) - \log|\mathscr{X}|,$$

which concludes our proof. □

Next, we take a look at no-signalling strategies.

6.3. No-signalling strategies

Let us define no-signalling strategies in terms of list-decoding schemes. Let (Enc, Dec) be an (n, M, L, α) code, where Dec maps elements $a^n \in \mathscr{A}^n$ to subsets $C_{a^n} \subset [M]$ of size L.

On inputs a^n and b^n, we generate two output sets $C'_{a^n} := \mathrm{Enc}(C_{a^n})$ and $C'_{b^n} := \mathrm{Enc}(C_{b^n})$ for Alice and Bob respectively. Next, we create a pairing of elements in the sets C'_{a^n} and C'_{b^n}, where elements that are in both sets are always paired up. This process results in a set $\{(x_A, x_B) \mid x_A \in C'_{a^n}, x_b \in C'_{b^n}\}$, where each element of C'_{a^n} and C'_{b^n} occurs in exactly one pair. Finally, we choose the output of Alice and Bob according to a uniform distribution over the set of pairs. Since each player outputs a uniform element from their output set, regardless of the input to the other player, this strategy is obviously no-signalling. See Section 5.3.2 and specifically Figure 5.3, for an example of such a strategy.

Using this strategy, the players win if $x^n \in C'_{a^n} \cap C'_{b^n}$ and the pair (x^n, x^n) is chosen (with x^n, a^n and b^n the values in the LSSD game). It follows that the winning probability of a no-signalling strategy defined by a list-decoding scheme is given by

$$\frac{1}{L|\mathscr{X}|^n} \sum_{\substack{x^n, a^n, b^n: \\ x^n \in C'_{a^n} \cap C'_{b^n}}} P^{\times n}_{\mathsf{A|X}}(a^n|x^n) P^{\times n}_{\mathsf{A|X}}(b^n|x^n) \tag{6.6}$$

$$= \frac{1}{L|\mathscr{X}|^n} \sum_{x^n} \left(\sum_{a^n : C'_{a^n} \ni x^n} P^{\times n}_{\mathsf{A|X}}(a^n|x^n) \right) \left(\sum_{b^n : C'_{b^n} \ni x^n} P^{\times n}_{\mathsf{A|X}}(b^n|x^n) \right) \tag{6.7}$$

$$= \frac{1}{L|\mathscr{X}|^n} \sum_{x^n} \left(\sum_{a^n : C'_{a^n} \ni x^n} P^{\times n}_{\mathsf{A|X}}(a^n|x^n) \right)^2 . \tag{6.8}$$

Note that when $L = 1$, the list-decoding scheme is just a regular error-correcting code, so the corresponding no-signalling strategy is the same as the classical strategy of the error correcting code.

This time, there is a trade-off between three things: the success probability of the code, the size of the lists and the number of messages. We want the list size to be small, to increase the probability of picking the right pair, but a smaller list size means a smaller success probability. Similarly, we want the number of messages to be large, but a larger number of messages means a smaller success probability.

6.4. Limit behaviour of no-signalling strategies based on list-decoding schemes

In this section we prove that no-signalling strategies based on list-decoding schemes do not achieve a better winning probability exponent when $n \to \infty$ than classical strategies. To this extent, let $\omega'_{\mathrm{ns}}(\mathsf{X}^n|\mathsf{A}^n;\mathsf{B}^n)_{P^{\times n}}$ be the optimal winning probability in n copies of a channel game using only no-signalling strategies based on list-decoding schemes.

Theorem 6.4. *Let $P_{\mathsf{A|X}}$ be a channel and let $P^{\times n}_{\mathsf{XAB}}$ be the probability distribution defining the channel game corresponding to the channel $P^{\times n}_{\mathsf{A|X}}$. We have*

$$\lim_{n\to\infty} \frac{\log(\omega'_{\mathrm{ns}}(\mathsf{X}^n|\mathsf{A}^n;\mathsf{B}^n)_{P^{\times n}})}{n} = \max_{Q_{\mathsf{XA}}} I(X;A)_Q - 2D(Q_{\mathsf{A|X}} \| P_{\mathsf{A|X}} \mid Q_{\mathsf{X}}) - \log(|\mathscr{X}|)$$

Again, we need a result to prove that this limit is the highest attainable. We conjecture that we can expand Lemma 5 from Dueck and Körner [5] (Lemma 6.3 in this thesis). Although we do not have a formal proof, we do have a strong suspicion that Conjecture 6.5 is true.

Conjecture 6.5. *For any* $(n, 2^{nR}, 2^{nR_L}, 2^{-n\zeta})$ *code for* $P_{\mathsf{A}|\mathsf{X}}$*, we have*

$$\zeta \geq \min_{Q_{\mathsf{XA}}} D(Q_{\mathsf{A}|\mathsf{X}} \| P_{\mathsf{A}|\mathsf{X}} \mid Q_{\mathsf{X}}) + \max\{R - R_L - I(X;A)_Q, 0\} + o(1).$$

Proof of Theorem 6.4. Note first that, since classical strategies based on error-correcting codes are a subset of the no-signalling strategies based on list-decoding schemes, we have

$$\begin{aligned}\lim_{n\to\infty} \frac{\log(\omega'_{\mathrm{ns}}(\mathsf{X}^n|\mathsf{A}^n;\mathsf{B}^n)_{P^{\times n}})}{n} &\geq \lim_{n\to\infty} \frac{\log(\omega_{\mathrm{c}}(\mathsf{X}^n|\mathsf{A}^n;\mathsf{B}^n)_{P^{\times n}})}{n} \\ &= \max_{Q_{\mathsf{XA}}} I(X;A)_Q - 2D(Q_{\mathsf{A}|\mathsf{X}} \| P_{\mathsf{A}|\mathsf{X}} \mid Q_{\mathsf{X}}) - \log(|\mathscr{X}|)\end{aligned}$$

For the inverse inequality, we follow the optimality part of the proof of Theorem 6.1 closely. Suppose Alice and Bob use a strategy defined by a list-decoding scheme (Enc, Dec), where Enc is injective and Dec maps a^n to the set C_{a^n} of size L. We define

$$q(x^n) := \sum_{a^n : C'_{a^n} \ni x^n} P^{\times n}_{\mathsf{A}|\mathsf{X}}(a^n|x^n).$$

(Remember that $C'_{a^n} = \mathrm{Enc}(C_{a^n})$.) By (6.8), we can write the winning probability of the strategy as

$$\frac{1}{L|\mathscr{X}|^n} \sum_{x^n} q(x^n)^2.$$

Let $\delta > 0$, for each $i \geq 0$, we define

$$\mathcal{R}_i := \{x^n \in \mathscr{X}^n \mid 2^{-n\delta(i+1)} \leq q(x^n) < 2^{-n\delta i}\}.$$

We define a list-decoding scheme $(\mathrm{Enc}_i, \mathrm{Dec}_i)$ as follows: $\mathrm{Enc}_i \colon \mathcal{R}_i \to \mathscr{X}^n$ is the identity function and

$$\mathrm{Dec}_i(a^n) = C'_{a^n} \cap \mathcal{R}_i.$$

Note that intersecting C'_{a^n} with $\mathcal{R}_i$ only decreases the size of the list, making the code weaker. This observation means that we will still be able to use Conjecture 6.5 for a list decoding with list size L. For each $x^n \in \mathcal{R}_i$, we have

$$\sum_{a^n \colon \mathrm{Dec}_i(a^n) \ni x^n} P^{\times n}_{\mathsf{A}|\mathsf{X}}(a^n|x^n) \geq q(x^n) \geq 2^{-n\delta(i+1)},$$

so $(\mathrm{Enc}_i, \mathrm{Dec}_i)$ defines a $(n, |\mathcal{R}_i|, L, 2^{-\delta(i+1)})$ code. By Conjecture 6.5, we have

$$\delta(i+1) \geq \min_{Q_{\mathsf{XA}}} D(Q_{\mathsf{A}|\mathsf{X}} \| P_{\mathsf{A}|\mathsf{X}} \mid Q_{\mathsf{X}}) + \max\left\{\frac{\log|\mathcal{R}_i|}{n} - \frac{\log(L)}{n} - I(X;A)_Q, 0\right\} + o(1).$$

Just like in the proof of Theorem 6.1, we find that if $q(x^n) > 0$, then $q(x^n) \geq 2^{-n\mu}$, with $\mu := \max_{x,a:P(a|x)>0} - \log(P(a|x))$ and if $i \geq t := \lfloor \frac{\mu}{\delta} \rfloor$, then $\mathcal{R}_i$ is empty. Now, we find

$$\begin{aligned}
\frac{1}{L} \sum_{x^n \in \mathscr{X}^n} q(x^n)^2 &= \sum_{i=0}^{t} \sum_{x^n \in \mathcal{R}_i} \frac{1}{L} q(x^n)^2 \\
&\leq \sum_{i=0}^{t} \frac{|\mathcal{R}_i|}{L} 2^{-2n\delta i} \\
&\leq \sum_{i=0}^{t} 2^{n\left(\frac{\log|\mathcal{R}_i|}{n} - \frac{\log(L)}{n} - 2\min_{Q_{\mathsf{XA}}}\left(D(Q_{\mathsf{A|X}} \| P_{\mathsf{A|X}} | Q_{\mathsf{X}}) + \max\left\{\frac{\log|\mathcal{R}_i|}{n} - \frac{\log(L)}{n} - I(X;A)_Q, 0\right\}\right) + o(1) + \delta\right)} \\
&\leq \sum_{i=0}^{t} 2^{n\left(\max_{Q_{\mathsf{XA}}}\left(I(X;A)_Q - 2D(Q_{\mathsf{A|X}} \| P_{\mathsf{A|X}} | Q_{\mathsf{X}})\right) + o(1) + \delta\right)}.
\end{aligned}$$

From this point, the proof is exactly the same as in Theorem 6.1. □

6.5. Exponent for the BSC game

We return to the game defined by a binary symmetric channel. We first show that in the BSC game, the expression in Theorem 6.4 also holds if we allow any kind of no-signalling strategies. This result means that, asymptotically, no-signalling strategies do not perform better than classical strategies. We then calculate what the value of the expression is. Throughout this section, $P_{\mathsf{A|X}}$ is a BSC.

6.5.1. Optimality of the exponent

Throughout this subsection, we denote by x, y, a, b elements of $\{0,1\}^n$. The next two lemmas tell us that there is always an optimal strategy of a certain form. This fact will help us prove Theorem 6.8.

Lemma 6.6. *For n simultaneous copies of the BSC game, defined by $P_{\mathsf{XAB}}^{\times n}$, there is always an optimal no-signalling strategy Q such that*

$$\forall s \in \{0,1\}^n : \quad Q(x \oplus s, y \oplus s | a \oplus s, b \oplus s) = Q(x, y | a, b). \tag{6.9}$$

Proof. Let Q be an optimal strategy and $s \in \{0,1\}^n$. Consider the strategy Q_s defined by $Q_s(x,y|a,b) = Q(x \oplus s, y \oplus s | a \oplus s, b \oplus s)$. The strategy Q_s has the same winning probability as Q, since $P_{\mathsf{XAB}}^{\times n}(x,a,b) = P_{\mathsf{XAB}}^{\times n}(x \oplus s, a \oplus s, b \oplus s)$ (the probability $P_{\mathsf{XAB}}^{\times n}(x,a,b)$ only depends on Hamming distances $d(x,a)$ and $d(x,b)$). We define

$$\hat{Q} := \frac{1}{2^n} \sum_{s \in \{0,1\}^n} Q_s.$$

The strategy $\hat{Q}$ satisfies (6.9):

$$\begin{aligned}\hat{Q}(x \oplus t, y \oplus t | a \oplus t, b \oplus t) &= \frac{1}{2^n} \sum_{s \in \{0,1\}^n} Q_s(x \oplus t, y \oplus t | a \oplus t, b \oplus t) \\ &= \frac{1}{2^n} \sum_{s \in \{0,1\}^n} Q(x \oplus t \oplus s, y \oplus t \oplus s | a \oplus t \oplus s, b \oplus t \oplus s) \\ &= \frac{1}{2^n} \sum_{r \in \{0,1\}^n} Q(x \oplus r, y \oplus r | a \oplus r, b \oplus r) \\ &= \frac{1}{2^n} \sum_{r \in \{0,1\}^n} Q_r(x, y | a, b) \\ &= \hat{Q}(x, y | a, b).\end{aligned}$$

Finally, by linearity of the winning probability, $\hat{Q}$ also achieves the same winning probability as Q, which means that it is optimal. □

For the next lemma we introduce some notation: for a permutation $\sigma \in S_n$ and a bitstring $x \in \{0,1\}^n$ we denote by $\sigma(x) \in \{0,1\}^n$ the bitstring obtained from x by permuting its bits according to σ. We omit the proof of the next lemma, as it is completely analogous to the proof of the previous lemma.

Lemma 6.7. *For n simultaneous copies of the BSC game, defined by $P^{\times n}_{\mathsf{XAB}}$, there is always an optimal no-signalling strategy Q such that*

$$\forall \sigma \in S_n : \quad Q(\sigma(x), \sigma(y) | \sigma(a), \sigma(b)) = Q(x, y | a, b). \tag{6.10}$$

Now suppose that Q is an optimal strategy satisfying (6.10) and let $\hat{Q}$ be the constructed from Q, according to the proof of Lemma 6.6, satisfying (6.9). We show that $\hat{Q}$ still satisfies (6.10). Let $\sigma \in S_n$, then

$$\begin{aligned}\hat{Q}(\sigma(x), \sigma(y) | \sigma(a), \sigma(b)) &= \frac{1}{2^n} \sum_{s \in \{0,1\}^n} Q(\sigma(x) \oplus s, \sigma(y) \oplus s | \sigma(a) \oplus s, \sigma(b) \oplus s) \\ &= \frac{1}{2^n} \sum_{s \in \{0,1\}^n} Q(\sigma(x) \oplus \sigma(s), \sigma(y) \oplus \sigma(s) | \sigma(a) \oplus \sigma(s), \sigma(b) \oplus \sigma(s)) \\ &= \frac{1}{2^n} \sum_{s \in \{0,1\}^n} Q(\sigma(x \oplus s), \sigma(y \oplus s) | \sigma(a \oplus s), \sigma(b \oplus s)) \\ &= \frac{1}{2^n} \sum_{s \in \{0,1\}^n} Q(x \oplus s, y \oplus s | a \oplus s, b \oplus s) \\ &= \hat{Q}(x, y | a, b).\end{aligned}$$

We conclude that there is always an optimal strategy satisfying both (6.9) and (6.10). We use this to prove the following theorem.

Theorem 6.8. *Let $P_{\mathsf{XAB}}^{\times n}$ define n copies of the BSC game. Then*

$$\lim_{n\to\infty} \frac{\log(\omega_{\mathrm{ns}}(\mathsf{X}^n|\mathsf{A}^n;\mathsf{B}^n)_{P^{\times n}})}{n} = \max_{Q_{\mathsf{XA}}} I(X;A)_Q - 2D(Q_{\mathsf{A|X}}\|P_{\mathsf{A|X}} \mid Q_{\mathsf{X}}) - \log(|\mathscr{X}|)$$

Proof. Let Q be an optimal strategy satisfying (6.9) and (6.10). From these two properties, it follows that the marginal distributions $Q(x|a)$ and $Q(y|b)$ only depend on the hamming distances $d(x,a)$ and $d(y,b)$ respectively (by picking the right $\sigma \in S_n$ and $s \in \{0,1\}^n$, one can show $Q(x|a) = Q(0^n|0^{n-d(x,a)}1^{d(x,a)})$). We can write the winning probability of Q as follows:

$$\sum_{x,a,b} P_{\mathsf{XAB}}^{\times n}(x,a,b)Q(x,x|a,b) = \sum_{i,j=0}^{n} \sum_{\substack{x,a,b:\\ d(x,a)=i\\ d(x,b)=j}} P_{\mathsf{XAB}}^{\times n}(x,a,b)Q(x,x|a,b). \tag{6.11}$$

Since the winning probability of Q is equal to $\omega_{\mathrm{ns}}(\mathsf{X}^n|\mathsf{A}^n;\mathsf{B}^n)_{P^{\times n}}$ and there are $(n+1)^2$ terms in the first sum of the RHS of (6.11), we know that there exist i,j such that

$$\sum_{\substack{x,a,b\\ d(x,a)=i\\ d(x,b)=j}} P_{\mathsf{XAB}}^{\times n}(x,a,b)Q(x,x|a,b) \geq \frac{\omega_{\mathrm{ns}}(\mathsf{X}^n|\mathsf{A}^n;\mathsf{B}^n)_{P^{\times n}}}{(n+1)^2}. \tag{6.12}$$

Now consider the following strategy $\tilde{Q}$:

- on input (a,b) Alice and Bob generate (x,y) according to Q;
- Alice checks if $d(x,a) = i$ and if not, uniformly generates a new output $\tilde{x}$ such that $d(\tilde{x},a) = i$;
- Bob checks if $d(y,a) = j$ and if not, uniformly generates a new output $\tilde{y}$ such that $d(\tilde{y},a) = j$;

This strategy is no-signalling and has winning probability of at least $\frac{\omega_{\mathrm{ns}}(\mathsf{X}^n|\mathsf{A}^n;\mathsf{B}^n)_{P^{\times n}}}{(n+1)^2}$, by (6.12). We also have that $\tilde{Q}(x|a)$ is uniform over $C_a := \{x \mid d(x,a) = i\}$, since $Q(x|a)$ only depends on $d(x,a)$. Similarly, $\tilde{Q}(y,b)$ is uniform over $D_b := \{y \mid d(y,b) = j\}$.

Defining $L_A = |C_a|$ and $L_B = |D_b|$ (these sizes are independent of a and b), we find

$$\begin{aligned}
\frac{\omega_{\mathrm{ns}}(\mathsf{X}^n|\mathsf{A}^n;\mathsf{B}^n)_{P^{\times n}}}{(n+1)^2} &\leq \sum_{x,a,b} P_{\mathsf{XAB}}^{\times n}(x,a,b)\tilde{Q}(x,x|a,b)\\
&\leq \sum_{x,a,b} P_{\mathsf{XAB}}^{\times n}(x,a,b)\min\{\tilde{Q}(x|a),\tilde{Q}(x|b)\}\\
&\leq \frac{1}{\max\{L_A,L_B\}} \sum_{x,a,b} P_{\mathsf{XAB}}^{\times n}(x,a,b)\delta(x\in C_a)\delta(x\in D_b)\\
&= \frac{1}{\max\{L_A,L_B\}|\mathscr{X}|^n} \sum_{x} \left(\sum_{a:\, C_a\ni x} P_{\mathsf{A|X}}^{\times n}(a|x)\right)\left(\sum_{b:\, D_b\ni x} P_{\mathsf{A|X}}^{\times n}(b|x)\right).
\end{aligned}$$

Now define

$$q_A(x) := \sum_{a\colon C_a \ni x} P^{\times n}_{\mathsf{A|X}}(a|x);$$
$$q_B(x) := \sum_{b\colon D_b \ni x} P^{\times n}_{\mathsf{A|X}}(b|x).$$

Using a similar argument as in the proof of Theorem 5.1, we can choose that $q_A = q_B$, which also means we can choose $C_a = D_a$ for all a. This results in an upper bound on the winning probability That is similar to the winning probability of strategies based on list decoding schemes (see (6.8)). In other words, this upper bound is achievable by a no-signalling strategy based on a list-decoding scheme. From this observation it follows that

$$\frac{\omega_{\mathrm{ns}}(\mathsf{X}^n|\mathsf{A}^n;\mathsf{B}^n)_{P^{\times n}}}{(n+1)^2} \leq \omega'_{\mathrm{ns}}(\mathsf{X}^n|\mathsf{A}^n;\mathsf{B}^n)_{P^{\times n}}. \tag{6.13}$$

Combining (6.13) with Theorem 6.4 and using that $\lim_{n\to\infty} \frac{\log((n+1)^2)}{n} = 0$, we find

$$\begin{aligned}\lim_{n\to\infty} \frac{\log(\omega_{\mathrm{ns}}(\mathsf{X}^n|\mathsf{A}^n;\mathsf{B}^n)_{P^{\times n}})}{n} &\leq \lim_{n\to\infty} \frac{\log(\omega'_{\mathrm{ns}}(\mathsf{X}^n|\mathsf{A}^n;\mathsf{B}^n)_{P^{\times n}})}{n} \\ &= \max_{Q_{\mathsf{XA}}} I(X;A)_Q - 2D(Q_{\mathsf{A|X}} \| P_{\mathsf{A|X}} \mid Q_{\mathsf{X}}) - \log(|\mathscr{X}|).\end{aligned}$$

The opposite inequality holds by Theorem 6.1, which concludes the proof. □

6.5.2. Calculating the exponent

We calculate, for the binary symmetric channel, the value of the limit of the exponent in theorems 6.1 and 6.8: $\max_{Q_{\mathsf{XA}}} I(X;A)_Q - 2D(Q_{\mathsf{A|X}} \| P_{\mathsf{A|X}} \mid Q_{\mathsf{X}}) - \log(|\mathscr{X}|)$. To this extent, let Q_{XA} be a distribution over $\{0,1\} \times \{0,1\}$. Now let us calculate the exponent one step at a time. First of all, we have

$$I(X;A)_Q = H(X)_Q + H(A)_Q - H(X,A)_Q.$$

We have

$$H(X)_Q = -\sum_{x=0}^{1} Q_{\mathsf{X}}(x)\log(Q_{\mathsf{X}}(x)) = -\sum_{x=0}^{1}\left(\sum_{a=0}^{1} Q_{\mathsf{XA}}(x,a)\right)\log\left(\sum_{a=0}^{1} Q_{\mathsf{XA}}(x,a)\right).$$

We can find $H(A)_Q$ in a similar way. We also have

$$H(X,A)_Q = -\sum_{x,a=0}^{1} Q_{\mathsf{XA}}(x,a)\log(Q_{\mathsf{XA}}(x,a)).$$

Now let us find the value of $D(Q_{\mathsf{A|X}} \| P_{\mathsf{A|X}} \mid Q_{\mathsf{X}})$:

$$\begin{aligned}D(Q_{\mathsf{A|X}} \| P_{\mathsf{A|X}} \mid Q_{\mathsf{X}}) &= \sum_{x=0}^{1} Q_{\mathsf{X}}(x) D(Q_{\mathsf{A|X}=x} \| P_{\mathsf{A|X}=x}) \\ &= \sum_{x=0}^{1}\left(\sum_{a=0}^{1} Q_{\mathsf{XA}}(x,a)\right)\left(\sum_{a=0}^{1} Q_{\mathsf{A|X}}(a|x)\log\left(\frac{Q_{\mathsf{A|X}}(a|x)}{P_{\mathsf{A|X}}(a|x)}\right)\right)\end{aligned}$$

Using numerical analysis we found that the maximum $\max_{Q_{\mathsf{XA}}} I(X;A)_Q - 2D(Q_{\mathsf{A|X}} \| P_{\mathsf{A|X}} \mid Q_{\mathsf{X}}) - \log(|\mathscr{X}|)$ is always achieved by a distribution Q_{XA} for which $Q_{\mathsf{XA}}(0,0) = Q_{\mathsf{XA}}(1,1) =: c$ and $Q_{\mathsf{XA}}(0,1) = Q_{\mathsf{XA}}(1,0) =: d$. Using this property, we have

$$H(X)_Q = H(A)_Q = -2(c+d)\log(c+d)$$

and

$$H(X,Y)_Q = -2c\log(c) - 2d\log(d).$$

We also find

$$D(Q_{\mathsf{A|X}} \| P_{\mathsf{X|A}} \mid Q_{\mathsf{X}}) = 2\left(c\log\left(\frac{c}{(c+d)(1-\alpha)}\right) + d\log\left(\frac{d}{(c+d)\alpha}\right)\right)$$

Combining the expressions above, we find the value $I(X;A)_Q - 2D(Q_{\mathsf{A|X}} \| P_{\mathsf{A|X}} \mid Q_{\mathsf{X}}) - \log(|\mathscr{X}|)$. Note that for Q_{XA} to be a distribution, we need $d = \frac{1}{2} - c$. This observation means that we only need to maximize with respect to the variable c (we see α as a constant). We can do this by calculating the derivative, setting it to 0 and solving for c. Using a computer algebra system, we find

$$\max_{Q_{\mathsf{XA}}} I(X;A)_Q - 2D(Q_{\mathsf{A|X}} \| P_{\mathsf{A|X}} \mid Q_{\mathsf{X}}) - \log(|\mathscr{X}|) = \log(1 - 2(1-\alpha)\alpha). \tag{6.14}$$

In Figure 6.1 we plotted this expression together with exponent of the optimal winning probability achieved by the strategies Q_n^d for some n (see Section 5.3.2). We can clearly see how this exponent approaches the limit calculated in (6.14).

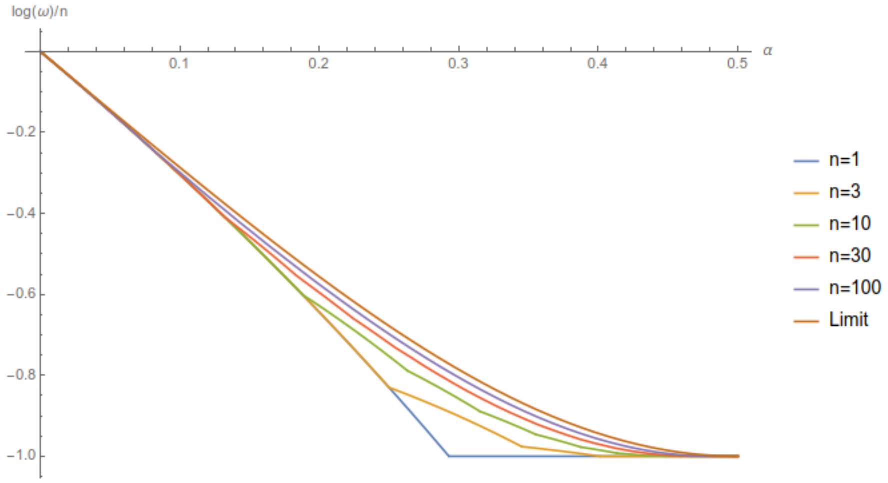


Figure 6.1.: This figure shows $\log(\omega)/n$ for different values of n and the limit of this expression, given by (6.14), against α. We calculated ω as the optimal winning probability achieved by the strategies Q_n^d (see Section 5.3.2).

7. Conclusion

In Chapter 4, we extended a result from Majenz et al., stating that there is no gap in winning probabilities using no-signalling resources versus classical resources when there are two players and all values are binary, to the case of three players. We achieved this result by numerical calculations, but could this also be done analytically? Another open problem is whether this result holds for any number of players. However, extending our numerical analysis to a larger number of players requires enumerating over all extrema of the corresponding no-signalling polytope. This polytope will quickly grow in number of vertices, making the analysis very slow.

In Chapter 5, we discussed an example of an LSSD game. We numerically found optimal classical and no-signalling strategies for two and three simultaneous copies of this game. We found that the strategies could be defined by (list-) decoding schemes and used that observation to define strategies for any number of simultaneous games. In Chapter 6 we showed that these strategies are asymptotically optimal (for no-signalling, this result depended on Conjecture 6.5). We also showed that classical and no-signalling strategies achieve the limit of the exponent of the success probability. Although the classical and no-signalling winning probabilities are the same when the number of simultaneous games approaches infinity, we found examples of a finite number of games where there is a gap between the optimal winning probabilities.

Again, we ask whether the numerically found optimal strategies could be proven to be optimal. We also suggest examining exactly when there is a gap between the no-signalling and classical winning probabilities in the BSC game. Wherever there is a gap, it is interesting to look for a quantum strategy that also performs better than classical.

In Chapter 6, we considered a class of LSSD games defined by any channel and considered playing multiple simultaneous copies. We extended the idea of using codes and list-decoding schemes to define classical and no-signalling strategies to this more general class of games. We gave an expression for the limit of the exponent of the classical winning probability and showed that no-signalling strategies based on list-decoding schemes do not achieve a better exponent. However, this last result depended on a conjecture, which we have not proven.

An obvious suggestion for future work is to prove Conjecture 6.5. Other open questions are: can we show, like for the BSC, that no-signalling strategies based on list-decoding schemes are asymptotically optimal? Are there more examples of channels for which there is a gap in winning probability between classical and no-signalling strategies in a finite number of simultaneous copies? Can the results be extended to classical-quantum channels, where Alice and Bob receive a quantum state? For this last question, we would need to extend the idea of no-signalling to the case where the inputs and outputs can be quantum states.

Chapter 6 also gives rise to a new area within information theory: simultaneous decoding. Within this setting, a sender tries to send a message to two receivers using identical channels and the communication is successful if both receivers decode correctly. We can allow the receivers to share some quantum or no-signalling resources and examine whether this leads to better coding schemes. There are similar settings that have already been researched. In one such setting, the messages sent to the receivers are not necessarily the same, or two different channels are used (like in the book by El Gamal [8, Part 2]). In another similar setting we allow the sender and the receiver to share some entanglement (like in the book by Holevo [11, Section 9]). There is even very recent research in a setting with two senders and one receiver that all share a no-signalling box (see the paper by Fawzi and Fermé [6]).

Ethical aspects We believe that none of the contributions made by this paper have any ethical aspects directly connected to them. However, there are some indirect risks connected to researching quantum mechanics. Especially when we advertise the research to the world, telling everyone how great quantum mechanics are. The capabilities of quantum mechanics and specifically things like quantum computing and non-locality could be blown out of proportions by people that are not well-read into the topics, or even by people with bad intentions. As an example, a future scammer might be able to convince others into buying a certain app or device (maybe a non-local box) that promises the ability to do things that it realistically cannot do. This might seem like a stretch, but in recent years we have seen how, for example, the blockchain technology caused a huge hype around cryptocurrency (and most recently NFT's). This hype has led to many people losing money.

We believe that, while research into quantum mechanics is important, we should treat the subject carefully. Mostly, we should always make sure to be honest about the limits of quantum mechanics, even if there are a lot of reasons to be enthusiastic about quantum in general.

Bibliography

[1] Jonathan Barrett, Noah Linden, Serge Massar, Stefano Pironio, Sandu Popescu, and David Roberts. "Nonlocal correlations as an information-theoretic resource". In: *Physical Review A* 71.2 (2005), p. 022101. DOI: 10.1103/physreva.71.022101.

[2] Harry Buhrman, Serge Fehr, and Christian Schaffner. "On the Parallel Repetition of Multi-Player Games: The No-Signaling Case". In: *9th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2014)*. Ed. by Steven T. Flammia and Aram W. Harrow. Vol. 27. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2014, pp. 24–35. DOI: 10.4230/LIPIcs.TQC.2014.24.

[3] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley, 2005. DOI: 10.1002/047174882x.

[4] Imre Csiszár and János Körner. *Information theory: coding theorems for discrete memoryless systems*. Cambridge University Press, 2011. DOI: 10.1017/cbo9780511921889.

[5] Gunter Dueck and János Körner. "Reliability function of a discrete memoryless channel at rates above capacity (Corresp.)" In: *IEEE Transactions on Information Theory* 25.1 (1979), pp. 82–85. DOI: 10.1109/tit.1979.1056003.

[6] Omar Fawzi and Paul Fermé. *Beating the Sum-Rate Capacity of the Binary Adder Channel with Non-Signaling Correlations*. 2022. DOI: 10.48550/ARXIV.2206.10968.

[7] Komei Fukuda. URL: https://people.inf.ethz.ch/fukudak/cdd_home/.

[8] Abbas El Gamal and Young-Han Kim. *Network Information Theory*. Cambridge University Press, 2011. DOI: 10.1017/cbo9781139030687.

[9] Richard W. Hamming. "Error Detecting and Error Correcting Codes". In: *Bell System Technical Journal* 29.2 (1950), pp. 147–160. DOI: 10.1002/j.1538-7305.1950.tb00463.x.

[10] Jaron Has. URL: https://github.com/Djerren/Jaron_Has_Bachelor_Project/.

[11] Alexander S. Holevo. *Quantum Systems, Channels, Information*. De Gruyter, 2019. DOI: 10.1515/9783110642490.

[12] Christian Majenz, Maris Ozols, Christian Schaffner, and Mehrdad Tahmasbi. *Local simultaneous state discrimination*. 2021. DOI: 10.48550/ARXIV.2111.01209.

[13] Neri Merhav. "List decoding - Random coding exponents and expurgated exponents". In: *2014 IEEE International Symposium on Information Theory*. Vol. 60. 11. IEEE, 2014, pp. 6749–6759. DOI: 10.1109/isit.2014.6875276.

[14] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2012. DOI: 10.1017/cbo9780511976667.

[15] Stefano Pironio, Jean-Daniel Bancal, and Valerio Scarani. “Extremal correlations of the tripartite no-signaling polytope”. In: *Journal of Physics A: Mathematical and Theoretical* 44.6 (2011), p. 065303. DOI: 10.1088/1751-8113/44/6/065303.

Popular summary

Consider the following scenario: two people, Alice and Bob, play a game against a referee. In this game, the referee generates three values, which we call x, a and b. Alice and Bob know how the referee generates these three values. The referee gives the value a to Alice and b to Bob, who then both have to guess the third value x without communicating with each other. Alice and Bob win the game if they both guess correctly. We call this game an LSSD game (see Figure 7.1).

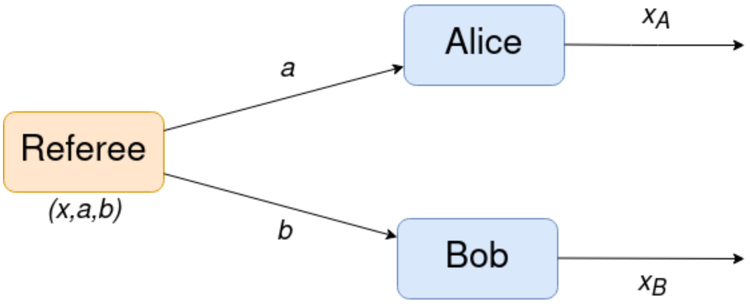


Figure 7.1.: Diagram of an LSSD game. Alice and Bob's guesses are named x_A and x_B. They win the game if they both guess correctly, so if $x_A = x_B = x$.

In this thesis, we are interested in what happens when we give the players access to a magic box (usually called a no-signalling box). The players both input their value into this box and they both get an output from this box. The magic box is allowed to do anything as long as it satisfies the following constraint: based on the output Alice gets from the box, she should not be able to obtain information on what Bob gave as input and vice versa. Now, the question is whether such a box could help the players win the game.

The main result of this thesis is centered around a specific example of an LSSD game. In this example, the referee generates a sequence of n bits (zeros and ones), where choosing a zero and choosing a one both happen with probability 1/2. The referee then generates a copy of this sequence by going over all elements of the sequence one by one and changing them from 0 to 1 or from 1 to 0 with a probability α (we assume that the probability α of such a flip is less than 1/2). The referee gives this copy to Alice, and generates another copy, which they give to Bob. Alice and Bob need to guess the initially generated sequence of bits.

Important to realize is that, in this game, Alice and Bob try to achieve two things: they want to individually be correct and they want to guess the same thing (if they guess something different, at least one of them will be incorrect). They can achieve the first goal by always guessing a sequence that is similar to the sequence they received, since, especially for small error probability α the sequence they receive will usually be similar to the original. The second goal can be achieved by limiting the number of sequences they might guess. This way, there is a higher probability they guess the same thing.

It turns out that good strategies (without a magic box) are for Alice and Bob to both perform the same "error-correcting code" on their sequences and output the result as their guess. An error correcting code is an algorithm that can detect and correct errors (flips) in sequences of bits. A code can only correct to a part of all possible sequences. The smaller this part is, the more likely Alice and Bob guess the same thing.

Strategies that use a magic box work a bit differently. The players input their sequences into the box. The box generates a list of all sequences that are different from Alice's sequence in at most d positions, and does the same for Bob's sequence. Next, The box pairs up items of these two lists, such that two items that are the same are always paired. Finally, the box picks one of the pairs at random and outputs it to Alice and Bob. Alice and Bob give the output from the box as their guess (see Figure 7.2).

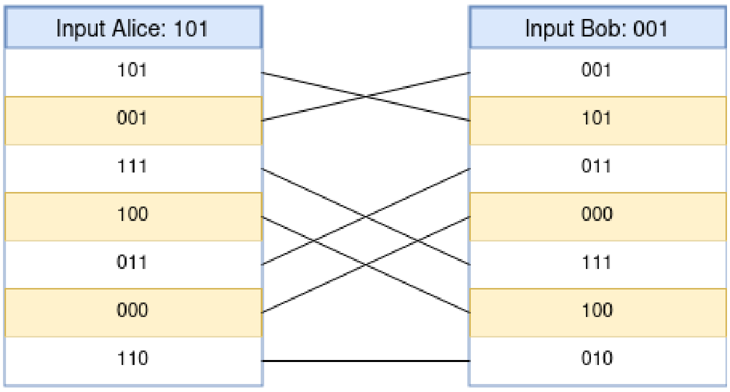


Figure 7.2.: An example of how a magic box strategy works when the length of the sequences is 3, $d = 2$ and the inputs of Alice and Bob are 101 and 001. The box generates two lists of sequences that are different in at most d positions. It then pairs up elements of the lists (by drawing lines), making sure to always pair elements that occur in both lists. One pair is chosen to be the output.

Probably the most interesting result in this thesis is that there are examples where strategies using a magic box increases the probability of winning compared to not using a magic box.

A. Proof of Lemma 6.2

The proof of Lemma 6.2 relies on concepts and theorems from the book by Csiszar [4]. We will not be discussing these concepts here.

Lemma 6.2. *Let $P_{\mathsf{A|X}}$ be a channel, Q_{XA} a probability distribution over $\mathscr{X} \times \mathscr{A}$ and $\delta > 0$. For large enough n, there exists an*

$$\left(n, 2^{n(I(X;A)_Q - \delta)}, \frac{2^{-nD(Q_{\mathsf{A|X}} \| P_{\mathsf{A|X}} | Q_{\mathsf{X}})}}{\mathrm{poly}(n)}\right)$$

code for the channel $P_{\mathsf{A|X}}$.

Proof. Let $R = I(X;A)_Q - \delta$. By the packing lemma (Lemma 10.1 in Csiszar's book), there exists a function $\mathrm{Enc}\colon [2^{nR}] \to \mathscr{X}^n$ such that

- $\mathrm{Enc}(m)$ is of type Q_X for all $m \in [2^{nR}]$;
- $|T_{Q_{A|X}}(\mathrm{Enc}(m)) \cap \bigcup_{m' \neq m} T_{Q_{A|X}}(\mathrm{Enc}(m'))| \leq |T_{Q_{A|X}}(\mathrm{Enc}(m))| 2^{-n\frac{\delta}{2}}$

(Note that the conditions of the packing lemma are satisfied, because $H(X)_Q \geq I(X;A)_Q$).

Now define $\mathrm{Dec}\colon \mathscr{A}^n \to [2^{nR}]$ by $\mathrm{Dec}(a^n) = m$ if m is the unique message such that $a^n \in T_{Q_{A|X}}(\mathrm{Enc}(m))$, otherwise we set $\mathrm{Dec}(a^n) = 0$. For all $m \in [2^{nR}]$, we have

$$\sum_{a^n\colon \mathrm{Dec}(a^n) = m} P_{\mathsf{A|X}}^{\times n}(a^n | \mathrm{Enc}(m)) = |\mathrm{Dec}^{-1}(m)| 2^{-n(D(Q_{\mathsf{A|X}} \| P_{\mathsf{A|X}} | Q_{\mathsf{X}}) + H(A|X)_Q)} \tag{A.1}$$

by Lemma 2.6 in Csiszar's book (Using that $\mathrm{Enc}(m)$ are all of type Q_{X}). By definition of the decoder, we also have

$$\begin{aligned} |\mathrm{Dec}^{-1}(m)| &\geq |T_{Q_{A|X}}(\mathrm{Enc}(m)) \setminus \bigcup_{m' \neq m} T_{Q_{A|X}}(\mathrm{Enc}(m'))| && \text{(A.2)} \\ &\geq |T_{Q_{A|X}}(\mathrm{Enc}(m))|(1 - 2^{-n\frac{\delta}{2}}) && \text{(A.3)} \\ &\geq (n+1)^{-|\mathscr{A}|}(1 - 2^{-n\frac{\delta}{2}}) 2^{nH(A|X)_Q} && \text{(A.4)} \\ &\geq \frac{2^{nH(A|X)_Q}}{\mathrm{poly}(n)}, && \text{(A.5)} \end{aligned}$$

where (A.3) follows from the second property of Enc and (A.4) follows from Lemma 2.3 of Csiszar's book. By combining (A.5) with (A.1) we conclude that $(\mathrm{Enc}, \mathrm{Dec})$ is a

$$\left(n, 2^{n(I(X;A)_Q - \delta)}, \frac{2^{-nD(Q_{\mathsf{A|X}} \| P_{\mathsf{A|X}} | Q_{\mathsf{X}})}}{\mathrm{poly}(n)}\right)$$

code. □