ECOLE
POLYTECHNIQUE
DE BRUXELLES

ULB

UNIVERSITÉ LIBRE DE BRUXELLES

The Information Complexity of Quantum Non-locality

Mémoire présenté en vue de l'obtention du diplôme
d'Ingénieur Civil Physicien à finalité Physique Appliquée

Mathieu Cavenaile

Directeur
Jérémie Roland

Service
Center for Quantum Information and Computation

Année académique
2017 - 2018

Remerciements

JR, MB, famille, amis.

Résumé

La complexité d'information de la non-localité quantique.

Mémoire présenté en vue de l'obtention du diplôme d'Ingénieur Civil Physicien

Mathieu Cavenaile - Année académique 2017-2018

Ce travail traite des corrélations quantiques entre systèmes intriqués et des ressources nécessaires pour les reproduire au départ d'un modèle local.

Différentes notions sont introduites pour évaluer la complexité de ce problème: complexité de communication, complexité entropique, complexité amortie, et complexité d'information. Les outils théoriques nécessaires sont également présentés, ainsi que la dérivation de bornes inférieures et supérieurs sur ces grandeurs. On considère le cas à deux parties avec communication unidirectionnelle, principalement avec entrées et sorties binaires.

L'objectif de ce travail est de prouver l'optimalité de la borne sur la complexité d'information pour le problème à entrées binaires. Pour se faire, on dualise le problème d'optimisation convexe sous contrainte correspondant. La valeur de $1 - H_{[\mu]}$ démontrée par Jérémie Roland et Mario Szegedy s'avère être solution du problème dual, ce qui prouve son optimalité.

Contrairement à la complexité de communication, il s'avère que la complexité d'information est plus faible lorsqu'on autorise les deux joueurs à communiquer. Valeur

Mots-clés : intrication quantique, non-localité quantique, complexité de communication, complexité d'information, optimisation convexe.

Abstract

The Information Complexity of Quantum Non-locality

Master thesis submitted in order to be awarded of
the Master's Degree in Engineering Physics
Mathieu Cavenaile - Academic year 2017-2018

This work deals with the correlations between entangled systems and the classical communication necessary to reproduce them starting from a local model.

Various quantities are introduce to that end: communication complexity, entropic complexity, amortized complexity, and information complexity. The corresponding theoretical tools are also presented, as well as upper and lower bounds on these quantities and their derivation. The focus is mainly on the bipartite case with one-way communication, binary inputs and binary outputs.

The final goal is to prove the optimality of the current lower bound on the information complexity of the problem with binary inputs. This is done by dualizing the corresponding convex optimization problem. The current bound of $1 - H_{[\mu]}$ derived by Jérémie Roland and Mario Szegedy is solution of both the primal and dual problems and is therefore optimal.

Contrary to communication complexity, it is also shown that authorizing both parties to communicate lowers the information complexity.

Keywords : quantum entanglement, quantum non-locality, communication complexity, information complexity, convex optimization.

Contents

List of Figures

Introduction

The principle of locality, stating that any physical system is only affected by its direct surroundings, is a natural and intuitive notion. While instantaneous and distant actions like Coulomb's force or gravitation were initially part of physical theories, locality was restored with Einstein's Special Relativity in 1905 (ref). The velocity of light was established as an universal upper bound. Since any effect had to be carried by some kind of physical field, instantaneous actions were therefore forbidden between distant objects. However, the development of quantum mechanics and the study of correlations between entangled systems put locality at risk again.

A paradox was laid down in 1935 by Einstein, Podolski and Rosen: either quantum mechanics gives rise to non-locality, or its description of reality is incomplete (ref). They believed the later and hoped to use hidden variables to solve the problem. These variables would contain informations not yet predicted by physics at the time. They were supposed to complete quantum mechanics' description of reality and to get rid of both its apparent randomness and its non-locality.

John Bell showed in 1964 that a hidden variable model could not explain quantum correlations (ref). His theorem was later used to derive equations that hold true for the local hidden variables model, but can be violated in the quantum world (ref). His mathematical predictions were later confirmed experimentally (ref 5 to 9), endorsing non-locality as an actual, albeit counter-intuitive, property of quantum mechanics.

The local model was not abandoned. Its extensions granted a way to quantify non-locality by allowing additional resources like classical communication to come into play [10]. Shannon's theory of information is of great help in this context [11]. By Laudauer's principle [12], information is physical, meaning that it is also bounded by the velocity of light. Even though a local description of entangled systems is incomplete, non-locality can therefore not be used for faster-than-light communication [13].

Nevertheless, it was found in 1997 that non-locality and entanglement could help solve certain distributing problems. In the framework of communication complexity, two distant players try to compute a function $f(x, y)$, each of them only having access to one of the two inputs. Richard Cleve and Harry Buhrman found out that if the two players share entanglement, the amount of communication they have to exchange can be reduced [14].

In this context, various works evaluated communications costs and complexities of quantum entanglement in different settings [15] [16] [17]. The first objective of this thesis is to present the state of the art in this domain. The relevant concepts of quantum mechanics,

information theory, complexity theory will also be introduced. Jérémie Roland and Mario Szegedy demonstrated a lower bound of the information complexity of quantum entanglement in the simplest case: binary inputs and binary outputs [17]. We will in this work prove the optimality of this result by solving a convex optimization problem. This method could also be applied to more general versions of this problem. Finally it is also shown that contrary to communication complexity, information complexity is lowered when both players are allow to communicate.

Sections 1 through 4 present relevant notions of quantum mechanics, information theory, complexity theory and convex optimization. Section 5 lays down the basic principles of quantum non-locality. Section 6 presents previous theoretical results in this field and some of their mathematical derivations. The derivation of our result is done in section 7.

Chapter 1

Quantum mechanics

The basics of quantum mechanics are prerequisites for the understanding of this work. We recall a few useful notions to our problem here, most notably the formal definition of quantum entanglement.

1.1 States and bases

Definition 1.1 *The quantum state of a system is characterized by a state vector denoted by $|\cdot\rangle$ belonging to a complex vector space C_n of dimension n [18], where we used Dirac's ket notation.*

The column notation can also be used:

$$|v\rangle = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix} \tag{1.1}$$

The corresponding bra $\langle v|$ is the row vector:

$$\langle v| = \begin{pmatrix} z_1^* & z_2^* & ... & z_n^* \end{pmatrix} \tag{1.2}$$

Definition 1.2 *A spanning set of a vector space is a set of spanning vectors $|v_1\rangle, |v_2\rangle, ..., |v_n\rangle$ such that any vector $|v\rangle$ of that vector space can be written as a linear combination of the spanning vectors [19]:*

$$|v\rangle = \sum_{i=1}^{n} a_i |v_i\rangle \tag{1.3}$$

Definition 1.3 *If the spanning vectors are linearly independent, the spanning set is called a basis [19].*

A vector space generally has multiple bases.

Definition 1.4 *The density operator of a statistical ensemble of quantum states $|\psi_i\rangle$ is written [19]:*

$$\rho = \sum_{i=1}^{n} p_i \left|\psi_i\right\rangle \left\langle\psi_i\right| \tag{1.4}$$

where p_i is the probability for the system to be in the state $|\psi_i\rangle$.

The density operator ρ is used when the state of the system is imperfectly known. Its matrix representation is called the density matrix. The eigenvalues of the density matrix are the probabilities p_i to be in state $|\psi_i\rangle$.

Definition 1.5 *When the state ψ_i of a system is perfectly known, it is said to be in a pure state.*

The density operator of a pure state $|\psi\rangle$ is $\rho = |\psi\rangle \langle\psi|$. The corresponding eigenvalue of its density matrix is in that case 1, and all other eigenvalues are 0.

1.2 The qubit

Definition 1.6 *The smallest unit of quantum information is the quantum bit or qubit.*

The qubit is the quantum counterpart of the classical bit. It is the base unit of quantum information. A classical bit takes one of two distinct values, usually denoted by 0 and 1. Similarly, a qubit can exist in two distinct orthogonal vector states, conventionally written $|0\rangle$ and $|1\rangle$. Contrary to classical bits however, qubits can also exist in any complex superposition of those two states. The general description of a qubit can be written [19]:

$$|\psi\rangle = \alpha |0\rangle + \beta |1\rangle \tag{1.5}$$

Definition 1.7 *The two states $|0\rangle$ and $|1\rangle$ form a basis in a two-dimensional complex vector space called the computational basis.*

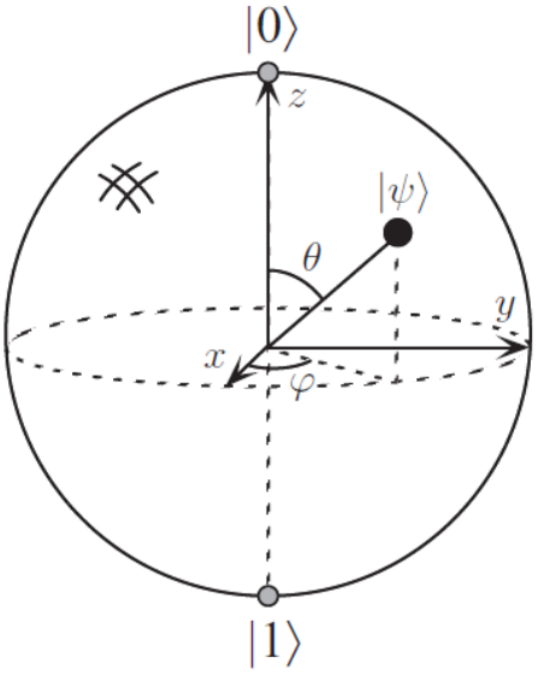
The normalization of the state imposes:

$$|\alpha|^2 + |\beta|^2 = 1 \tag{1.6}$$

with $\alpha, \beta \in \mathbb{C}$. Depending on the context and the application, other bases are also used. Physically speaking, a qubit can be represented by various systems. It can be for instance encoded in the polarization of a photon, with the horizontal or vertical polarizations $|\uparrow\rangle$ and $|\rightarrow\rangle$ representing the $|0\rangle$ and $|1\rangle$ states [20], or in the spin of a particle, with the up and down states $|\uparrow\rangle$ and $|\downarrow\rangle$ playing the same role [19]. Virtually any two-level system can serve as a qubit, as long as the quantum mechanical rules apply.

Theorem 1.1 *A measurement in the computational basis of the state of a qubit of the form* (1.5) *in the computational basis yields the $|0\rangle$ and $|1\rangle$ states with respective probability $|\alpha|^2$ and $|\beta|^2$ [19].*

Figure 1.1: Bloch sphere representation of the states of a qubit.



The Bloch sphere

Due to the normalization condition (1.6), equation (1.5) can be rewritten as follows [19]:

$$|\psi\rangle = e^{i\gamma}\left(\cos\frac{\theta}{2}|0\rangle + e^{i\phi}\frac{\theta}{2}|1\rangle\right) \tag{1.7}$$

with $\gamma, \theta, \phi \in \mathbb{R}$. The overall phase γ can be ignored since it does not have any observable effect, yielding:

$$|\psi\rangle = \cos\frac{\theta}{2}|0\rangle + e^{i\phi}\frac{\theta}{2}|1\rangle \tag{1.8}$$

The remaining vector state $|\psi\rangle$ can be interpreted as a unit vector in a three-dimensional space.

Definition 1.8 *Every possible state of a qubit can be seen as a point inside a unit sphere called the Bloch sphere. All the pure states correspond to points at the surface of the Bloch Sphere.*

The Bloch sphere is represented on fig.1.2.

Measurement

The principles of the measurement in quantum mechanics will not be recalled here. However, this section will presented the operators used to measure a qubit.

Even though a qubit can exist in a superposition of states, measuring it destroys that superposition. A measurement is done in a basis, and the two possible are the two corresponding basis vectors. Taking the example of a qubit in a state of the form (1.5), a

measurement in the computational basis will give either $|0\rangle$ or $|1\rangle$ with respective probability $|\alpha|^2$ and $|\beta|^2$.
As stated previously, a qubit can be represented by the spin of a particle. The spin angular momentum is actually nothing more than an application of the qubit formalism. The measurement of a qubit can therefore be treated using the Pauli matrices:

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{1.9}$$

$$\sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \tag{1.10}$$

$$\sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{1.11}$$

We take the north pole of the Bloch sphere as our $|0\rangle$ state, its south pole as our $|1\rangle$ state, and its vertical axis as our z-axis as done on fig. 1.2. In that case, the $|0\rangle$ and $|1\rangle$ states are the eigenstates of the operator S_z with matrix representation σ_z: measuring a qubit in the computational basis comes down to applying S_z to its state.

Any direction $\vec{u}$ on the Bloch sphere defines a basis in which the state of a qubit can be measured. The corresponding operator is $S_u = \vec{u} \cdot \vec{S}$ where the components of the vector operator $\vec{S}$ are the Pauly matrices σ_x, σ_y and σ_z.

Two qubits basis

The computational basis can be generalized for systems composed of more than one qubit.

Definition 1.9 *A product state of a multiple-qubit system is a tensor product of the states of each qubit.*

Each state of the basis is in that case a product state. Taking the example of a two-qubit system, we have the following four basis vectors [21]:

$$|00\rangle = |0\rangle \otimes |0\rangle \tag{1.12}$$

$$|01\rangle = |0\rangle \otimes |1\rangle \tag{1.13}$$

$$|00\rangle = |1\rangle \otimes |0\rangle \tag{1.14}$$

$$|11\rangle = |1\rangle \otimes |1\rangle \tag{1.15}$$

Again, any superposition of the four basis vectors is a valid state for the system. The overall state of the pair of qubits can be written [19]:

$$|\psi\rangle = \alpha\,|00\rangle + \beta\,|01\rangle + \zeta\,|10\rangle + \kappa\,|11\rangle\,. \tag{1.16}$$

Here the normalization of the state imposes $|\alpha|^2 + |\beta|^2 + |\zeta|^2 + |\kappa|^2 = 1$.

Definition 1.10 *Any pure state $|\psi\rangle$ of a bipartite system composed of two subsystems A and B can be written as a Schmidt decomposition [19]:*

$$|\psi\rangle = \sum_i \lambda_i\,|i_A\rangle\,|i_B\rangle\,, \tag{1.17}$$

where $\{|i_A\rangle\}$ and $\{|i_B\rangle\}$ are orthogonal bases for A and B and λ_i non-negative real numbers satisfying $\sum_i \lambda_i^2 = 1$.

Definition 1.11 *The number of values λ_i is the Schmidt number of the state $|\psi\rangle$.*

1.3 Entanglement

Definition 1.12 *The state of a composite system is separable when each subsystem can be described independently.*

Being product states, each state of the computational basis is separable. For instance, for a system in a $|01\rangle$ state, the first qubit is in the $|0\rangle$ state and the second in the $|1\rangle$ state. However, an independent description of each qubit is not always possible [19].

Definition 1.13 *If the states of subsystems cannot be defined independently, those subsystems are entangled.*

Definition 1.14 *Two systems are maximally entangled if evaluating the partial trace of their density matrix with respect to one of the subsystems yield a matrix proportional to the identity matrix [19].*

Entanglement is one of the most peculiar phenomena in quantum mechanics. A state $|\psi\rangle$ is entangled if its Schmidt number is greater than 1 [19] [21]. These entangled states exhibit properties at the foundation of many quantum information techniques, like quantum teleportation [22] and dense coding [23]. Non-local effects are also a manifestation of quantum entanglement.

The Bell basis

Definition 1.15 *The Bell basis is composed of the following four Bell states, written as a linear combination of the vectors of the computational basis [19]:*

$$\left|\phi^+\right\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \tag{1.18}$$

$$\left|\phi^-\right\rangle = \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle) \tag{1.19}$$

$$\left|\psi^+\right\rangle = \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle) \tag{1.20}$$

$$\left|\psi^-\right\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle) \tag{1.21}$$

$$\tag{1.22}$$

The Bell basis is the most natural basis to describe two-qubit entangled states. Each of the Bell states is maximally entangled.

One of these four states is generally used in a theoretical Bell experiment, usually the singlet state $|\psi^-\rangle$ because of its isotropy: measuring the spin along any axis will yield the same result [19]. The demonstration of this property is done in appendix A.1.

Chapter 2

Information theory

Information theory was originally developed by Claude Shannon in "A Mathematical Theory of Communication", published in 1948 [11]. While information is a mathematical concept, it has to be contained in a physical system and therefore obeys the laws of Physics [12]. Specifically, due to Einstein's Special Relativity, information cannot be exchanged faster than light [13] [20].

Another corollary is the necessity to generalize information theory to include the effects of quantum concepts like entanglement, superposition of states and the quantum no-cloning theorem [19]. Quantum informations theory also gives rise to peculiar properties and schemes like superdense coding [23] and quantum teleportation [22].

The present chapter introduces some notions of information theory relevant to the problem at hand. While we consider an experiment involving quantum mechanics and entangled qubits in particular, we will only be interested in classical information and classical entropy. The main definitions presented here are extracted from ref. [24].

2.1 Shannon's entropy

Shannon's entropy function is the core concept of information theory. The entropy of a random variable can be interpreted as the uncertainty on that variable, or the amount of information gained on average by one observation of that variable. It is also the minimum number of bits needed to represent that random variable, i.e. how to most efficiently code and potentially compress information [25].

Definition 2.1 *The entropy of a discrete random variable X with alphabet $\mathcal{X}$ and probability mass function $p(x) = Pr[X = x]$ for $x \in \mathcal{X}$ is the expected value of the random variable $\dfrac{1}{\log p(X)}$ [24]:*

$$H(X) = -\mathbb{E}\{\log(X)\} = -\sum_{x\in\mathcal{X}} p(x)\log p(x) \tag{2.1}$$

$H(X) = 0$ when X is deterministic. The entropy is conventionally expressed in bits. The logarithm is to the base 2 in that case. Other units exist, for instance the nats (when

using a log to the base e) or a hartley (when using a log to the base 10) [24]. In this work, all logarithms are to the base 2.

The entropy of a random variable X with two outcomes with respective probability p and $(1-p)$ is given by the function:

$$H(X) = -p \log p - (1-p) \log(1-p) \tag{2.2}$$

This function will be denoted by $H_{[p]}$.

Definition 2.2 *The differential entropy of a continuous random variable X with probability density function $\rho(x)$ is defined as:*

$$h(X) = -\int_{S_x} \rho(x) \log \rho(x) dx, \tag{2.3}$$

where S is the support set of $\rho(x)$, i.e. the domain such that $\rho(x) \neq 0 \ \forall x \in S_x$.

Differential entropy is the counterpart of entropy when dealing with continuous random variables. The distinction between them will not explicitly be made in this work. Both will be denoted by $H(X)$ and the nature of X in each case will determine the one that is being evaluated.

2.2 Further definitions

The following section presents quantities related to the entropy used throughout this thesis.

Definition 2.3 *The joint entropy $H(X,Y)$ of two random variables X and Y is given by:*

$$H(X,Y) = -\mathbb{E}\{\log p(X,Y)\} = -\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x,y) \log p(x,y), \tag{2.4}$$

where $p(x,y)$ is the joint probability density of X and Y.

Definition 2.4 *The conditional entropy is written:*

$$H(X|Y) = -\mathbb{E}\{\log p(X|Y)\} = -\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x,y) \log p(x|y), \tag{2.5}$$

where $p(x|y)$ is the probability density of X conditioned on Y.

$H(X|Y) = 0$ when X deterministically depends on Y, i.e. when knowing the value of Y implies knowing the value of X. We note the use the conditional probability in the logarithm but the joint probability $p(x,y)$ in front of it. The conditional entropy can also be written:

$$\sum_{x \in \mathcal{X}} p(x) H(Y|X=x) \tag{2.6}$$

Definition 2.5 *The mutual information (or mutual entropy) of two random variables X and Y is written:*

$$I(X;Y) = \mathbb{E}\{\log \frac{p(X,Y)}{p(X)p(Y)}\} = \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log\frac{p(x,y)}{p(x)p(y)} \tag{2.7}$$

Definition 2.6 *The Kullback-Leibler divergence (or relative entropy) between two probability mass functions $p(x)$ and $q(x)$ is defined as follows:*

$$D(p(x)||q(x)) = \sum_{x\in\mathcal{X}} p(x)\log\frac{p(x)}{q(x)} \tag{2.8}$$

This function is not symmetric, i.e. $D(p(x)||q(x)) \neq D(q(x)||p(x))$. Additionally, $D(p(x)||q(x)) \geq 0$ and $D(p(x)||q(x)) = 0$ only if $p(x) = q(x)$.

The entropy of a random variable can be expressed in term of its Kullback-Leibler divergence with a uniform distribution. A uniform distribution has a probability mass function $u(x)$ such that:

$$u(x) = \frac{1}{|\mathcal{X}|}\ \forall x \in \mathcal{X}, \tag{2.9}$$

where $|\mathcal{X}|$ is number of elements in $\mathcal{X}$.

This yields:

$$\begin{aligned} H(X) &= -\sum_{x\in\mathcal{X}} p(x)\log p(x) && (2.10)\\ &= -\sum_{x\in\mathcal{X}} p(x)\log p(x) - \sum_{x\in\mathcal{X}} p(x)\log u(x) + \sum_{x\in\mathcal{X}} p(x)\log u(x) && (2.11)\\ &= -\sum_{x\in\mathcal{X}} p(x)\log u(x) - D(p(x)||u(x)) && (2.12)\\ &= \log|\mathcal{X}| - D(p(x)||u(x)). && (2.13) \end{aligned}$$

This form and the properties of $D(p(x)||u(x))$ confirm that the entropy is maximized by a uniform distribution.

The mutual information can also be written in term of a Kullback-Leibler divergence:

$$I(X;Y) = D(p(x,y)||p(x)p(y)) \tag{2.14}$$

Definition 2.7 *Considering three random variables X, Y, Z, the conditional mutual information is written [24]:*

$$\begin{aligned} I(X;Y|Z) &= \mathbb{E}\{\log\frac{p(X,Y|Z)}{p(X|Z)p(Y|Z)}\} && (2.15)\\ &= H(X|Z) - H(X|Y,Z) && (2.16) \end{aligned}$$

These various quantities exist for continuous random variables as well. Similarly to the entropy, sums become integrals, probability mass functions become probability density, and alphabets become support sets. Their expressions are written as follows:

$$H(X,Y) = \int_{S_x}\int_{S_y} \rho(x,y)\log\rho(x,y) \tag{2.17}$$

$$H(X|Y) = \int_{S_x}\int_{S_y} \rho(x,y)\log\rho(x|y)dxdy \tag{2.18}$$

$$D(\rho(x)||\xi(x)) = \int_{S_x} \rho(x)\log\frac{\rho(x)}{\xi(x)}dx \tag{2.19}$$

$$I(X;Y) = \int_{S_x}\int_{S_y} p(x,y)\log\frac{\rho(x,y)}{\rho(x)\rho(y)}dxdy \tag{2.20}$$

$$= D(\rho(x,y)||\rho(x)\rho(y)) \tag{2.21}$$

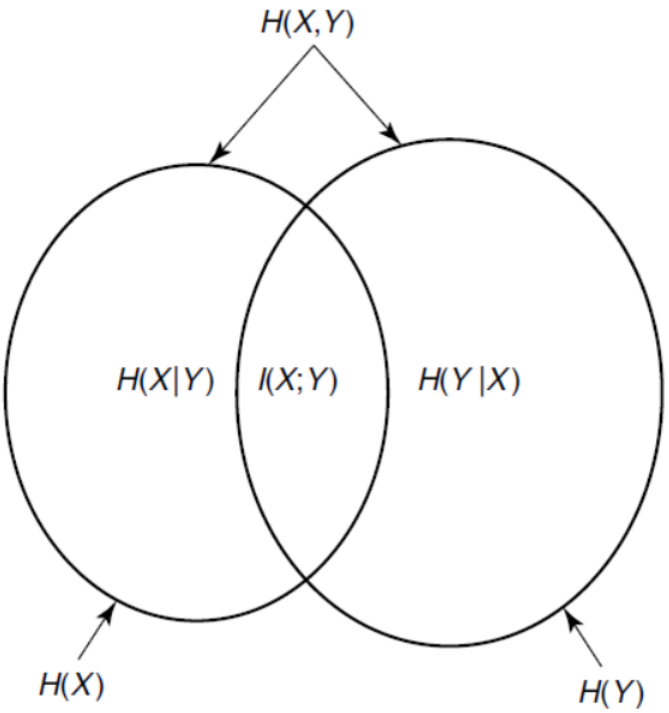
The relations and theorems derived in this chapter are valid for both discrete and differential entropies. Again, the distinction between them will not be made explicitly.
Joint, conditional and mutual entropies obey relations such as:

$$H(X,Y) = H(X) + H(Y|X) = H(Y) + H(X|Y) \tag{2.22}$$

$$I(X;Y) = H(X) - H(X|Y) \tag{2.23}$$

These relations can be summarized by a Venn diagram, as represented on fig.2.2.

Figure 2.1: Representation of the relations between entropies.



Theorem 2.1 (Chain rule) *The entropy of n random variables X_1, X_2, ..., X_n follows a chain rule [24]:*

$$H(X_1, X_2, ..., X_n) = \sum_{i=1}^{n} H(X_i|X_{i-1}, ..., X_1) \tag{2.24}$$

Chain rules for conditional entropy and mutual information can be written similarly.

Definition 2.8 *Three random variables X, Y, Z form a Markov chain $X \to Y \to Z$ (in that order) if the conditional distribution of Z depends only on Y and is conditionally independent of X. In that case we have for their respective probability mass functions:*

$$p(x, y, z) = p(x)p(y|x)p(z|y) \tag{2.25}$$

In particular, if Z is a function of Y, $X \to Y \to Z$ form a Markov chain.

Theorem 2.2 (Data processing inequality) *The mutual informations of random variables forming a Markov chain $X \to Y \to Z$ obeys the data processing inequality [24]:*

$$I(X; Y) \geq I(X; Z) \tag{2.26}$$

2.3 Shannon's theorems

Various theorems establish relations between random variables, their entropy, their optimal coding, and the capacity of a channel to transmit them reliably. Most of them were demonstrated by Claude Shannon in his *"Mathematical Theory of Information"*, and some of them are used in this work.

Theorem 2.3 (Noiseless coding theorem) *The optimal coding of a random variable X has an expected length in bits $L(X)$ bounded by its entropy:*

$$H(X) \leq L(X) \leq H(X) + 1. \tag{2.27}$$

$L(X)$ is the cost in bits of sending a sample distributed according to X. A useful corollary is the limit:

$$\lim_{n\to\infty} \frac{L(X^n)}{n} = H(X), \tag{2.28}$$

where $L(X^n)$ is the cost of sending n independent samples distributed according to X.

Aside from the entropy, a sizeable part of Shannon's work is dedicated to the transmission of data via channels.

Definition 2.9 *A discrete channel is a system consisting in discrete input and output alphabets $\mathcal{X}$ and $\mathcal{Y}$ and a probability transition matrix $\{p(y|x)\}$.*

Definition 2.10 *A channel is memoryless if the output at a given time only depends on the input at that time, and not on previous inputs and outputs.*

All channels will be considered memoryless in what follows. All the behaviour of a channelis dictated by the conditional probabilities $\{p(y|x)\}$.

Definition 2.11 *The channel capacity marks the maximal achievable rate of information transmission through a channel.*

Theorem 2.4 (Channel coding theorem) *Given inputs and outputs extracted from random variables X and Y, the capacity of a channel is given by:*

$$C = \max_{p(x)} I(X;Y) \tag{2.29}$$

This theorem gives the minimum cost of transmitting information through a communication channel. The converse statement, the reverse Shannon theorem, was proven with quantum applications in mind [27]. We will however use a corollary:

Theorem 2.5 (Reverse Shannon Theorem) *Considering a large number of repetitions, the use of shared randomness and only one-way communication, a channel $X \to Y$ can be simulated with $I(X;Y)$ bits on average per repetition:*

$$\lim_{n\to\infty} \frac{C_n^{\to}}{n} \leq I(X;Y), \tag{2.30}$$

where $^{\to}$ denotes one-way communication.

Chapter 3

Communication complexity

The notion of communication complexity was first introduced by Andrew Yao in 1979 [28]. His bipartite model will be defined here, following the presentation from ref. [29].

3.1 Bipartite model

We consider two players, Alice and Bob. They wish to evaluate a function $f(x, y)$ for inputs $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, with Alice only knowing x and Bob only knowing y. A more general problem is to produce outputs a and b according to a probability distribution $\mathbf{p}(a, b|x, y)$, that we will write $\mathbf{p}$ for short in the general case.
They therefore have to exchange information following a given protocol $\mathcal{P}$ determined beforehand. Since we are only interested in these exchanges, Alice and Bob computational power is supposed to be infinite. They can also share randomness in the form of a potentially infinite set Λ. A random string $\lambda \in \Lambda$ with distribution $q(\lambda)$ is used during each run to determine the outcome of every random variable.

Definition 3.1 *The transcript of the exchanged messages, noted $M_{\mathcal{P}}(x, y, \lambda)$, is the content of all the messages exchanged between Alice and Bob during the execution of the protocol $\mathcal{P}$.*

Definition 3.2 *The length of the transcript $|M_{\mathcal{P}}(x, y, \lambda)|$ is the number of bits exchanged between Alice and Bob during the execution of the protocol $\mathcal{P}$.*

For a given protocol, the transcript is only a function of the inputs and the shared randomness. Each protocol runs in rounds, during which either Alice is sending a message to Bob, Bob is sending a message to Alice, or they both produce their outputs.

Definition 3.3 *The communication cost of $\mathcal{P}$ on input (x, y) is the number of bits exchanged by Alice and Bob to produce a and b according to $\mathbf{p}$ given the input (x, y).*

Definition 3.4 *The maximal cost of $\mathcal{P}$ over all possible inputs is called the worst-case communication cost C_w of $\mathcal{P}$:*

$$C_w(\mathcal{P}) = \max_{(x,y)} |M_{\mathcal{P}}(x, y, \lambda)|. \tag{3.1}$$

Definition 3.5 *The minimum worst-case cost over all protocols that computes $\mathbf{p}$ is the worst-case communication complexity C_w of $\mathbf{p}$:*

$$C_w(\mathbf{p}) = \min_{\mathcal{P}} C_w(\mathcal{P}) \tag{3.2}$$

Complexity is used to describe how hard it is to solve a problem, without real interest in the actual solution. Other types of cost and complexities can be defined in a similar fashion.

3.2 Complexities

Average complexity

Definition 3.6 *The average cost $C^D(\mathcal{P})$ is the expected number of bits exchanged between Alice and Bob given an input distribution D on $\mathcal{X} \times \mathcal{Y}$. The expectation value is taken over the shared randomness and the inputs (x, y):*

$$C^D(\mathcal{P}) = \sum_{\lambda \in \Lambda} q(\lambda) \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} D(x, y) |M_{\mathcal{P}}(x, y, \lambda)|. \tag{3.3}$$

Definition 3.7 *The minimum average cost over all protocols for a given distribution is written:*

$$C^D(\mathbf{p}) = \min_{\mathcal{P}} C^D(\mathcal{P}). \tag{3.4}$$

Definition 3.8 *The average communication complexity is the maximum over all distributions of $C^D_($**p**$)$:*

$$C(\mathbf{p}) = \max_{D} C^D(\mathbf{p}). \tag{3.5}$$

Entropic complexity

Definition 3.9 *The entropic cost is the conditional entropy of the transcript given the shared randomness:*

$$C^D_H(\mathcal{P}) = H(M|\Lambda). \tag{3.6}$$

Definition 3.10 *The entropic complexity is defined as the maximum on all distributions of the minimal entropic cost over the protocols:*

$$C^D_H(\mathbf{p}) = \max_{D} \min_{\mathcal{P}} C^D_H(\mathcal{P}). \tag{3.7}$$

Amortized complexity The notion of asymptotic or amortized complexity appears when running a high number n (taking the limit to infinity) of instances of the problem in parallel. In that case we use a distribution of the inputs and a n-fold parametrization of **p** written:

$$D^{\otimes n}(\vec{x}, \vec{y}) = \prod_{i=1}^{n} D(x_i, y_i) \tag{3.8}$$

$$\mathbf{p}^{\otimes n}(\mathbf{a}, \mathbf{b}|\vec{x}, \vec{y}) = \prod_{i=1}^{n} \mathbf{p}(a_i, b_i|x_i, y_i) \tag{3.9}$$

The communication cost of producing the distribution **p** n times is written $C^{D^{\otimes n}}(\mathbf{p}^{\otimes n})$.

Definition 3.11 *The distributional amortized communication complexity is the quantity:*

$$C_{\infty}^{D}(\mathbf{p}) = \lim_{n\to\infty} \frac{C^{D^{\otimes n}}(\mathbf{p}^{\otimes n})}{n} \tag{3.10}$$

Definition 3.12 *The maximum of the distributional amortized communication complexity over all possible input distributions is the amortized communication complexity:*

$$C_{\infty}(\mathbf{p}) = \max_{D} C_{\infty}^{D}(\mathbf{p}) \tag{3.11}$$

Information complexity The information complexity is the quantity we wish to evaluate in this work. Similarly to the entropy giving a lower bound on the communication needed to transmit a random variable, the information complexity gives a lower bound on the communication complexity of a distribution [30]. One has to distinguish internal and external information costs.

Definition 3.13 *The internal information cost of a protocol P is the minimum amount of new information that Alice and Bob learn about each others input during a run of $\mathcal{P}$ [31]:*

$$C_{I}^{int,D}(\mathcal{P}) = I(M;X|Y) + I(M;Y|X) \tag{3.12}$$

Definition 3.14 *The internal information complexity is the maximum on all distributions of the minimum internal information cost on all protocols:*

$$C_{I}^{int}(\mathbf{p}) = \max_{D} \min_{\mathcal{P}} C_{I}^{int,D}(\mathcal{P}) \tag{3.13}$$

Definition 3.15 *The external information cost is the information about the output revealed to a outside observer by the message:*

$$C_{I}^{ext,D}(\mathcal{P}) = I(M,\Lambda;X,Y) \tag{3.14}$$

with M being the concatenation of M_0 and λ.

Definition 3.16 *The external information complexity is defined accordingly:*

$$C_{I}^{ext}(\mathbf{p}) = \max_{D} \min_{\mathcal{P}} C_{I}^{ext,D}(\mathcal{P}) \tag{3.15}$$

It was shown by Braverman in ref.[31] that internal information complexity equals communication complexity, we will therefore only write $C_I = C_I^{ext}$ and $C_\infty = C_I^{ext}$.

Distributions

Definition 3.17 *A product input distribution D on $\mathcal{X} \times \mathcal{Y}$ is a distribution that can be written: where D_A (respectively D_B) is the marginal distribution of x on $\mathcal{X}$ (respectively y on $\mathcal{Y}$).*

All costs have to be maximized over all input distributions to evaluate the corresponding complexity. We therefore have to identify which distribution is the most detrimental to the problem at hand.

Theorem 3.1 *For a given distribution* $\mathbf{p}(a,b|x,y)$ *and considering only one-way communication, two input distributions* D, D_0 *with the same marginal distributions on* x *give rise to the same communication complexity.*

This theorem allows us to only worry about the marginal distributions D_A and D_B, as long as the communication is one-way. This theorem applies to worst-case, average, entropic, amortized and information complexities.

3.3 Theorems and relations

Theorem 3.2 *(Direct sum theorem) For a product distribution* D*:*

$$C_I^{D^{\otimes n}}(\mathbf{p}^{\otimes n}) = nC_I^D(\mathbf{p}) \tag{3.16}$$

This result is due to Chakrabarti, Shi, Wirth, and Yao [32]. A corollary presented in ref. [17] is that for one-way communication:

$$C_I^{\rightarrow,D^{\otimes n}}(\mathbf{p}^{\otimes n}) = nC_I^{\rightarrow,D}(\mathbf{p}) \tag{3.17}$$

This relation is useful due to the fact that the best known protocols for our problem are one-way [17].

Theorem 3.3 *The communication complexities satisfy the following relations:*

$$C_\infty(\mathbf{p}) \leq C_I(\mathbf{p}) \leq C_H(\mathbf{p}) \leq C(\mathbf{p}) \leq C_w(\mathbf{p}) \tag{3.18}$$

Proofs: The proofs follow the approach presented in [17]:

- $C(\mathbf{p}) \leq C_w(\mathbf{p})$ is trivial given the definitions of $C(\mathbf{p})$ and $C_w(\mathbf{p})$;

- $C_H(\mathbf{p}) \leq C(\mathbf{p})$: by Shannon's noiseless coding theorem 2.4, for a protocol $\mathcal{P}$, $H(M|\Lambda) \leq |M_\mathcal{P}(x,y,\lambda)|$ and therefore $C_H(\mathbf{p}) \leq C(\mathbf{p})$;

- $C_I(\mathbf{p}) \leq C_H(\mathbf{p})$: we suppose that with shared randomness λ an optimized protocol $\mathcal{P}$ has an entropic cost C_H^D. This means that a protocol $\mathcal{P}'$ starting with Alice sending λ to Bob, and then proceeding as $\mathcal{P}$ would reproduce $\mathbf{p}$. Its transcript would be composed of λ and $M_\mathcal{P}(x,y,\lambda)$, and its information cost $C_I^D(\mathcal{P}') = I(\lambda, M; X, Y)$. Combining the relations (2.6) and the facts that:

  - the shared randomness is independent of the inputs:

$$I(X,Y;\Lambda) = 0 \Leftrightarrow H(X,Y) = H(X,Y|\Lambda) \tag{3.19}$$

  - the message is fully determined by the inputs and the shared randomness:

$$H(M|\Lambda,X,Y) = 0 \Leftrightarrow H(M,\Lambda,X,Y) = H(\Lambda,X,Y), \tag{3.20}$$

we have:

$$\begin{aligned} I(\Lambda, M; X, Y) &= H(X,Y) - H(X,Y|\Lambda, M) && (3.21)\\ &= H(X,Y,\Lambda) - H(\Lambda) - H(X,Y|\Lambda, M) && (3.22)\\ &= H(M|\Lambda) + H(\Lambda, M, X, Y) - H(\Lambda, M) - H(X,Y|\Lambda, M) && (3.23)\\ &= H(M|\Lambda), && (3.24) \end{aligned}$$

yielding:

$$C_I^D(\mathcal{P}') = C_H^D(\mathcal{P}). \tag{3.25}$$

Since $\mathcal{P}$ is defined as optimal but $\mathcal{P}'$ might not be, we can conclude that:

$$C_I(\mathbf{p}) \leq C_H(\mathbf{p}). \tag{3.26}$$

- $C_\infty(\mathbf{p}) \leq C_I(\mathbf{p})$: again suppose we achieve $\mathbf{p}$ with a protocol $\mathcal{P}$ with transcript of messages $M_\mathcal{P}$ that does not use shared randomness. If the messages are sent by Alice, $X \to M$ defines a channel and we have:

  $$C_I(\mathcal{P}) = I(M; X). \tag{3.27}$$

  These specifications allow us to apply the reverse Shannon theorem 2.5. We therefore know that a large number n of repetitions of $\mathbf{p}$ can be simulated using a protocol $\mathcal{P}_n$ with average cost such that:

  $$\lim_{n\to\infty} \frac{C(\mathcal{P}_n)}{n} = I(M; X), \tag{3.28}$$

  meaning that the amortized complexity is at most equal to the information cost:

  $$C_\infty(\mathbf{p}) \leq C_I(\mathbf{p}). \tag{3.29}$$

Theorem 3.4 *For a product input distribution D, we have:*

$$C_\infty^D(\mathbf{p}) = C_I^D(\mathbf{p}). \tag{3.30}$$

In other words, internal and external information complexities are equal for a product input distribution.

Proof: : From (3.1) and (3.3):

$$\begin{aligned} C_I^{D^{\otimes n}}(\mathbf{p}^{\otimes n}) &= n C_I^D(\mathbf{p}) && (3.31)\\ C_I^{D^{\otimes n}}(\mathbf{p}^{\otimes n}) &\leq C^{D^{\otimes n}}(\mathbf{p}^{\otimes n}) && (3.32)\\ C_I^D(\mathbf{p}) &\leq C_I^{D^{\otimes n}}(\mathbf{p}^{\otimes n})/n && (3.33) \end{aligned}$$

Taking the limit for $n \to \infty$, we get:

$$C_I^D(\mathbf{p}) \leq C_\infty^D(\mathbf{p}). \tag{3.34}$$

Since $C_\infty(\mathbf{p}) \leq C_I(\mathbf{p})$, we have $C_\infty(\mathbf{p}) = C_I(\mathbf{p})$ for a product distribution.

Thanks to these relations, we have access to various types of tools to prove upper and lower bounds on complexities.

Chapter 4

Convexity and optimization

We present here relevant notions of convex optimization

4.1 Convexity and Jensen's inequality

Definition 4.1 *A convex function f(x) on an interval I is such that:*

$$f(\alpha x + (1-\alpha)y) \geq \alpha f(x) + (1-\alpha)f(y) \quad \forall \alpha \in [0;1] \ \& \ \forall x, y \in I \tag{4.1}$$

Definition 4.2 *A function f is concave on I if -f is convex on I.*

Convex and concave functions have useful features regarding optimization. They most notably have a single extremum (minimum for convex functions and maximum for concave functions). The convexity of a function is straightforward to test: as long as the second derivative with respect to x exists and is positive on I, $f(x)$ is convex on I. Conversely, $f(x)$ is concave on I if its second derivative exists and is negative on I. The second derivative of linear functions vanishes, therefore they are both convex and concave, but neither strictly convex nor strictly concave.

Convexity is relevant to the present work since many entropies are either strictly convex or strictly concave.

Theorem 4.1 *The Kullblack-Leibler divergence $D(p(x)||q(x))$ is a convex function of a pair $p(x)$, $q(x)$.*

Proof: We start from the log sum inequality:

$$\sum_{i=1}^{n} a_i \log \frac{a_i}{b_i} \geq \left(\sum_{i=1}^{n} a_i \right) \log \frac{\sum_{i=1}^{n} a_i}{b_i}, \tag{4.2}$$

holding for non-negative $a_1, ..., a_n$ and $b_1, ..., b_n$. Considering probability mass functions $p_1(x)$, $p_2(x)$, $q_1(x)$ and $q_2(x)$ and $\alpha \in [0;1]$, the log sum inequality yields:

$$\alpha(p_1(x) + (1-\alpha)p_2(x)) \log \frac{p_1(x) + (1-\alpha)p_2(x)}{q_1(x) + (1-\alpha)q_2(x)} \tag{4.3}$$

$$\leq \alpha p_1(x) \log \frac{\alpha p_1(x)}{\alpha q_1(x)} + (1-\alpha)p_2(x) \log \frac{(1-\alpha)p_2(x)}{(1-\alpha)q_2(x)} \tag{4.4}$$

We find that:

$$D(\alpha p_1 + (1-\alpha)p_2||\alpha q_1 + (1-\alpha)q_2) \leq \alpha D(p_1||q_1) + (1-\alpha)D(p_2||q_2) \tag{4.5}$$

by summing over all $x \in \mathcal{X}$.

Theorem 4.2 *Shannon's entropy $H(X)$ is a concave function of x.*

Proof: The convexity of $D(p(x)||u(x))$ combined with the previously derived expression of $H(X)$:

$$H(X) = \log|\mathcal{X}| - D(p(x)||u(x)) \tag{4.6}$$

shows that Shannon's entropy is concave.

A consequence of the concavity of Shannon's entropy is Jensen's inequality.

Theorem 4.3 (Jensen's inequality) *The following inequality holds for a random variable X and a convex function f:*

$$\mathbb{E}\{f(X)\} \geq f\left(\mathbb{E}\{X\}\right) \tag{4.7}$$

This can be proven recursively. The full proof is derived in ref.[CoverThomas].

Theorem 4.4 *The mutual information $I(X;Y)$ is a concave function of $p(x)$ for fixed $p(y|x)$ and a convex function of $p(x|y)$ for fixed $p(x)$.*

Proof: By definition we have:

$$p(y) = \sum_{x\in\mathcal{X}} p(x)p(y|x), \tag{4.8}$$

meaning that $p(y)$ is a linear function of $p(x)$ for fixed $p(x|y)$. $H(Y)$ is a concave function of $p(y)$ and therefore a concave function of $p(x)$ in that case. $H(X|Y)$ being a linear function of $p(x)$, $I(X;Y) = H(Y) - H(Y|X)$ is a concave function of $p(x)$ for fixed $p(x|y)$.
The mutual information can also be written as a divergence $D(p(x)p(x|y)||p(x)p(y))$, meaning that for fixed $p(x)$, $I(X;Y)$ is a convex function of $p(x|y)$.

4.2 Convex optimization

We consider the general form of an optimization problem, written as follows:

$$\begin{aligned} \inf_{\vec{x}} \quad & f(\vec{x}) && \text{(4.9a)} \\ \text{subject to} \quad & g(\vec{x}) \leq 0 && \text{(4.9b)} \\ & h(\vec{x}) = 0 && \text{(4.9c)} \end{aligned}$$

We call $f(\vec{x})$ the objective function and (4.9b) and (4.9c) the constraints. Problem (4.9) is a convex optimization problem if the objective and the inequality constraints are convex functions and the equality constraints are affine functions.

Note that when the inf (respectively sup) of an expression is attainable, we can just write min (resp. max).

Definition 4.3 *The solution p^* to problem* (4.9) *is the minimal value of $f(\vec{x})$ with $\vec{x}$ respecting the constraints* (4.9b) *and* (4.9c)*:*

$$p^* := \inf_{\vec{x}} f(\vec{x}) \tag{4.10}$$

$$:= f(\vec{x}^*) \tag{4.11}$$

x^* is the optimizer of problem (4.9)*.*

Definition 4.4 *The Lagrangian of problem* (4.9) *is the quantity:*

$$\mathcal{L}(\vec{x}, \vec{\mu}, \vec{\nu}) := f(\vec{x}) + \sum_i \mu_i g(x_i) + \sum_j \nu_j h(x_j), \tag{4.12}$$

where $\vec{\mu}$ and $\vec{\nu}$ are vectors of Lagrange multipliers and the sums are taken over the size of $\vec{\mu}$ and $\vec{\nu}$ respectively. We suppose $\mu_i \geq 0 \quad \forall i$. The Lagrangian is build by inserting the constraints into the objective function. The general idea is that this quantity cannot be optimized unless these constraints are satisfied. Indeed we can make $\mathcal{L}(\vec{x}, \vec{\mu}, \vec{\nu})$ as big or as small as we want if the second and third terms of the expression do not vanish.

We also note:

$$k(\vec{\mu}, \vec{\nu}) := \inf_{\vec{x}} \mathcal{L}(\vec{x}, \vec{\mu}, \vec{\nu}) \tag{4.13}$$

Lemma 4.1 *For $\mu_i \geq 0$ we have:*

$$k(\vec{\mu}, \vec{\nu}) \leq p^* \tag{4.14}$$

Indeed for all $\vec{x}$ respecting the constraints and $\mu_i \geq 0$:

$$\mu_i g(x_i) \leq 0; \tag{4.15}$$

$$\nu_j h(x_j) = 0; \tag{4.16}$$

$$k(\vec{\mu}, \vec{\nu}) \leq f(x). \tag{4.17}$$

In particular for $\vec{x} = \vec{x}^*$, $k(\vec{\mu}, \vec{\nu}) \leq p^*$. We can therefore write p^* as:

$$p^* = \inf_{\vec{x}} \sup_{\substack{\vec{\nu}, \vec{\mu}, \\ \mu_i \geq 0}} \mathcal{L}(\vec{x}, \vec{\mu}, \vec{\nu}) \tag{4.18}$$

Let us now consider

$$d^* := \sup_{\substack{\vec{\nu}, \vec{\mu}, \\ \mu_i \geq 0}} k(\vec{\mu}, \vec{\nu}) = \sup_{\substack{\vec{\nu}, \vec{\mu}, \\ \mu_i \geq 0}} \inf_{\vec{x}} \mathcal{L}(\vec{x}, \vec{\mu}, \vec{\nu}) \tag{4.19}$$

A corollary of lemme 4.1 is that $d^* \leq p^*$:

$$\sup_{\substack{\vec{\nu}, \vec{\mu}, \\ \mu_i \geq 0}} \inf_{\vec{x}} \mathcal{L}(\vec{x}, \vec{\mu}, \vec{\nu}) \leq \inf_{\vec{x}} \sup_{\substack{\vec{\nu}, \vec{\mu}, \\ \mu_i \geq 0}} \mathcal{L}(\vec{x}, \vec{\mu}, \vec{\nu}) \tag{4.20}$$

It is generally possible to extract constraints on $\vec{\mu}$ and $\vec{\nu}$ to rewrite d^* as the solution of another optimization problem.

$$\sup_{\substack{\vec{\mu}, \vec{\nu} \\ \mu_i \geq 0}} \quad \tilde{f}(\vec{\mu}, \vec{\nu}) \tag{4.21a}$$

$$\text{subject to} \quad \tilde{g}(\vec{\mu}) \leq 0 \tag{4.21b}$$

$$\tilde{h}(\vec{\mu}) = 0 \tag{4.21c}$$

$$\tilde{g'}(\vec{\nu}) \leq 0 \tag{4.21d}$$

$$\tilde{h'}(\vec{\nu}) = 0 \tag{4.21e}$$

Definition 4.5 *Problem* (4.21) *is the dual problem of problem* (4.9)*. Conversely,* (4.9) *is called the primal problem of* (4.21)*.*

The bound (4.20) is referred to as weak duality and holds for all types of optimization problems. It is saturated under various sets of conditions. We will use the following in this work.

Lemma 4.2 (Slater's condition) *Strong duality between problems* (4.9) *and* (4.21) *holds if there exist* $\vec{x}$ *such that* $g(\vec{x}) < 0$ *and* $h(\vec{x}) = 0$*.*

When strong duality holds, the solutions of the primal and dual problems are the same: $d^* = p^*$. The dual problem is therefore a valuable tool if a candidate solution for the primal is known: if the candidate is also solution to the dual problem, it is optimal.

The last result that we will present in this chapter is presented in ref[opti]. It gives the solution and optimizer of optimization problem involving a Kullback-Leibler divergence.

Lemma 4.3 *The unique optimizer of a problem of the form:*

$$\min_{\rho} D(\rho||\xi) - < \rho, c >, \tag{4.22}$$

with:

- ρ *and* ξ *probability densities on the space* $\mathbb{K}$*, i.e.* $\int_{\mathbb{K}} \rho(dx) = \int_{\mathbb{K}} \xi(dx) = 1$*;*
- $D(\rho||\xi)$ *the Kullback-Leibler divergence between* ρ *and* ξ*;*
- $< \rho, c >= \int_{\mathbb{K}} c(x)\rho(dx)$ *a bilinear form,*

is given by:

$$\rho^* = \frac{2^{c(x)}\xi(dx)}{\int_{\mathbb{K}} 2^{c(x)}\xi(dx)}. \tag{4.23}$$

The optimal value of the program is:

$$-\log \int_{\mathbb{K}} 2^{c}(x)\xi(dx). \tag{4.24}$$

Additionally, ref[opti] also shows that for this type of problems, Slater's condition is satisfied and we have strong duality between the primal and dual problems.

Chapter 5

Quantum non-locality

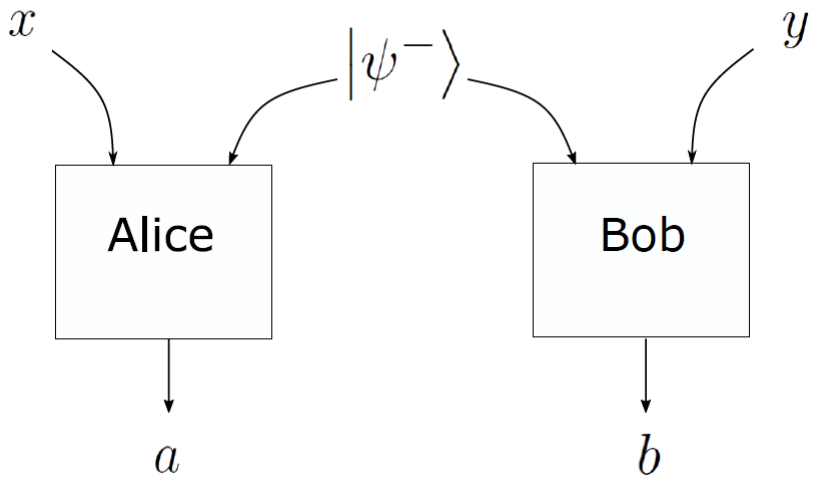
John Bell showed in 1964 the implications of the EPR paradox and the mathematical proof that quantum mechanics exhibits non-local behaviours [3]. This chapter presents the mathematical description of correlations between entangled systems and shows the incompatibility between quantum mechanics and local models.

5.1 Bell experiment

The following model describes the Bell experiment at the center of the EPR paradox [2].

We consider two systems which may have interacted in the past. They are then spatially separated to prevent any unwanted exchange between them. Two players, Alice and Bob, each perform a local measurement on one of the systems. These measurements are parametrized by settings, or inputs, denoted by x and y for Alice and Bob respectively. They yield results, or outputs, a and b. Fig.5.1 presents this model, considering a system composed of entangled qubits in the $|\psi^-\rangle$ state.

Figure 5.1: Mathematical model of a Bell experiment.



Note that these are conventional labels. A measurement can for instance be the measure of the polarization of a photon. The input would then define the basis in which the measurement is performed, and the result would be the output.

Definition 5.1 *The results of a Bell experiment are governed by the probability to obtain the outputs conditioned on the inputs, namely* $\mathbf{p}(a, b|x, y)$*, referred to as correlations.*

Since the two observed systems have interacted in the past, the outputs on each system are generally not independent, i.e. $\mathbf{p}(a, b|x, y) \neq \mathbf{p}(a|x)\mathbf{p}(b|y)$, even in a local theory [33].

Multiple variations of the Bell experiment exist. We will consider two in this work.

CHSH experiment The simplest case of Bell experiment is called the CHSH experiment, from Clauser, Horne, Shimony, and Holt [4]. The experiment has four possible inputs $(x, y) \in \{0, 1\}^2$ and four possible outputs $(a, b) \in \{-1, 1\}^2$. Alice and Bob share a pair of entangled qubits in the singlet state $|\psi^-\rangle$. The values $x = 0$ and $x = 1$ (respectively $y = 0$ and $y = 1$) define two orthogonal directions on the Bloch sphere $\vec{x}_0$ and $\vec{x}_1$ (resp. $\vec{y}_0$ and $\vec{y}_1$). Alice and Bob each perform a measurement of the state of their qubit along the direction specified by their input. The results of the measurement are the states $|0\rangle$, $|1\rangle$ and correspond to the outputs -1 or 1.

EPR-Bohm experiment This more general version of the Bell experiment was introduced by David Bohm in 1957 [34]. Alice and Bob share the same state $|\psi^-\rangle$ the inputs are this time unit vectors $\vec{x}$ and $\vec{y}$ on the surface of the d-dimensional sphere $\mathbb{S}_{d-1}$. A CHSH experiment is just a particular case of an EPR-Bohm experiment.

5.2 Correlations

Quantum mechanics allow us to evaluate the correlations between a and b in an EPR-Bohm experiment. The singlet state being isotropic, we do not have to worry about the overall directions of the measurements, only the angle between them. In the trivial case where we measure the qubits in the same basis, $x = y$ and $\vec{x}_i = \vec{y}_i \ \forall i \in \{0, 1\}$. Given the specification of the singlet state (1.21), the results of the measurement on Alice's and Bob's sides are perfectly anti-correlated and we randomly get either $|01\rangle$ or $|10\rangle$ with probability $1/2$:

$$p(a = 1, b = 1|x, y) = \quad p(a = -1, b = -1|x, y) \quad = 0 \tag{5.1}$$

$$p(a = -1, b = 1|x, y) = \quad p(a = 1, b = -1|x, y) \quad = \frac{1}{2} \tag{5.2}$$

When considering an angle θ between x and y, the overall probability of a and b being equal is $(1 - \cos\theta)/2$ and we get:

$$p(a = 1, b = 1|x, y) = \quad p(a = -1, b = -1|x, y) \quad = \frac{1 - \cos\theta}{4} \tag{5.3}$$

$$p(a = -1, b = 1|x, y) = \quad p(a = 1, b = -1|x, y) \quad = \frac{1 + \cos\theta}{4} \tag{5.4}$$

These results can be rewritten in a more compact forms.

Theorem 5.1 *For an EPR-Bohm experiment, the correlations assume the form:*

$$\mathbf{p}_{dim=d}(a, b|\vec{x}, \vec{y}) = \frac{1 - ab\ \vec{x} \cdot \vec{y}}{4} \tag{5.5}$$

For a CHSH experiment, an angle of $\pi/4$ is generally chosen between x_0 and y_1, giving the following expression:

Theorem 5.2 *In the case of a CHSH experiment, the quantum correlations take the convenient form:*

$$\mathbf{p}_\mu(a,b|x,y) = \frac{1+\mu(-1)^{xy}}{4}, \tag{5.6}$$

with $\mu = \sqrt{2}/2$, thanks to the specifications of the inputs and outputs in 4.1.

We have used the parameter $\mu \in [0;1]$ as in ref.[17]. Classical correlations take the same form as (5.6), but with $\mu \in [0;1/2]$. This expression is again a particular case of (5.5).

Alternatively, the correlations can also be written in terms of expectation values. As specified by quantum mechanics, the results of measurements on Alice's or Bob's qubit appear random. To respect causality, Alice's expectation value is also independent of Bob's input. We can therefore write $\mathbf{p}(a|x,y) = \mathbf{p}(a|x), \mathbf{p}(b|x,y) = \mathbf{p}(b|y)$ and:

$$\langle a_{x,y}\rangle = \langle a_x\rangle = \sum_a a\ p(a|x) \tag{5.7}$$

$$\langle b_{x,y}\rangle = \langle b_y\rangle = \sum_a b\ p(b|x) \tag{5.8}$$

Theorem 5.3 *The quantum expectation value of the product ab gives:*

$$\langle a_x b_y\rangle = \sum_{a,b} ab\ p(a,b|x,y) = -\vec{x}\cdot\vec{y}, \tag{5.9}$$

meaning that the choice of inputs has an influence on the overall expected results of the experiment. However, these results appear random to the two players and they have to compare them (i.e., communicate or be reunited) to observe the correlations. Causality is therefore respected. This is the essence of quantum non-locality.

5.3 Local model

The local model was originally introduced as a classical way to solve the EPR paradox [2].

In this version of the Bell experiment, instead of an entangled pair of qubits Alice and Bob share a random variable λ from a possibly infinite set Λ. As in chapter 3, λ is a random string of bits. It is supposed to represent any randomness affecting the final result of the measurements. Thus we can replace Alice's and Bob's outputs by deterministic functions of their inputs and the shared randomness:

$$a = A(\lambda, x) \tag{5.10}$$

$$b = B(\lambda, y) \tag{5.11}$$

To maintain the locality of the model, Alice's output cannot depend on Bob's input, and vice-versa. In Bell's model, the inputs are determined after the last interaction between the two systems: the shared randomness λ therefore does not depend on the inputs.

Every potential dependencies between a and b being included in λ, we can assume that for a given λ, the correlations admit a decomposition of the form:

$$p(a,b|x,y) = p(a|\lambda,x)p(b|\lambda,y) \tag{5.12}$$

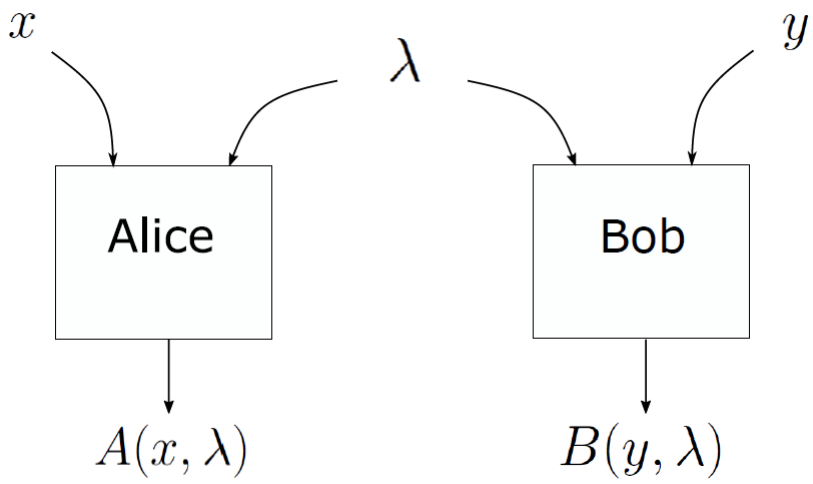
Considering potentially different λ from run to run, the overall correlations can be written:

$$p(a,b|x,y) = \int_\Lambda p(a|\lambda,x)p(b|\lambda,y)d\lambda \tag{5.13}$$

Definition 5.2 *Models respecting the aforementioned specifications are called Local Hidden Variable(LHV) models.*

The choice of the type of inputs x and y and functions A and B allows a simulation of a CHSH experiment or an EPR-Bohm experiment. The LHV model is presented on fig.5.3.

Figure 5.2: Local Hidden Variable model.



5.4 Bell's theorem and Bell inequalities

What John Bell showed in 1965 is that locality is incompatible with the predictions of quantum mechanics [3]. Conversely, it also means than quantum correlations do not admit a decomposition of the form (5.13).

Theorem 5.4 *(Bell's Theorem) A LHV model is not able to reproduce correlations of the form* (5.5).

The proof of Bell's theorem in the CHSH case is presented in appendix A.1. It is based on the evaluation the quantity:

$$S = \langle a_0 b_0 \rangle + \langle a_0 b_1 \rangle + \langle a_1 b_0 \rangle - \langle a_1 b_1 \rangle \tag{5.14}$$

It can be shown that S is bounded by 2 in LHV model, whereas there exist quantum settings such that $S = 2\sqrt{2}$. Conveniently, this introduces the concept of Bell inequalities.

Definition 5.3 *A Bell inequality is an linear expression of the form:*

$$\sum_{x,y,a,b} B_{xyab}\ p(a,b|x,y) \leq B_0 \tag{5.15}$$

that is always satisfied for a LHV model but can be violated in the quantum case.

Theorem 5.5 *(CHSH inequality) In a CHSH experiment, the inequality:*

$$\sum_{x,y,a,b} ab(-1)^{xy} p(a,b|x,y) \leq 2 \tag{5.16}$$

always holds for the local correlations but not for quantum correlations.

The CHSH inequality is the simplest Bell inequality. Explicitly written, the upper-bounded quantity in (5.16) is nothing else than the S defined in (4.18). The upper bound is $2\sqrt{2}$ in the quantum case [35].

The CHSH inequality is actually a consequence of the fact that a LHV model exhibits correlations of the form:

$$\mathbf{p}_\mu(a,b|x,y) = \frac{1-\mu(-1)^{xy}}{4}, \tag{5.17}$$

this time with $0 \leq \mu \leq 1/2$ [17]. A parameter can also be introduced in the $\mathbf{p}_{dim=d}$ correlations (4.9):

$$\mathbf{p}_{dim=d,\nu}(a,b|\vec{x},\vec{y}) = \frac{1-\nu ab\ \vec{x}\cdot\vec{y}}{4} \tag{5.18}$$

Finding the maximum value of ν such that $\mathbf{p}_{dim=d,\nu}$ can be simulated by a LHV model is still an open problem [36].

5.5 Polytopes

A geometric approach can be used to represent the sets of local and quantum correlations. A bipartite Bell experiment with i settings and j outcomes will be described by a finite set P of i^2j^2 conditional probabilities $\{p(a,b|x,y)\}$ [33]. This set can be seen as a point P in a $\mathbb{R}^{i^2j^2}$ space [37].

For instance, taking the simplest case where $x,y \in \{0,1\}$ and $a,b \in \{-1,1\}$, the correlations are composed of the following 16 probabilities:

$$\begin{array}{cccc} p(1,1|0,0) & p(1,-1|0,0) & p(-1,1|0,0) & p(-1,-1|0,0) \\ p(1,1|0,1) & p(1,-1|0,1) & p(-1,1|0,1) & p(-1,-1|0,1) \\ p(1,1|1,0) & p(1,-1|1,0) & p(-1,1|1,0) & p(-1,-1|1,0) \\ p(1,1|1,1) & p(1,-1|1,1) & p(-1,1|1,1) & p(-1,-1|1,1) \end{array} \tag{5.19}$$

These probabilities are constrained by:

- positivity :

$$p(a,b|x,y) \geq 0\ \forall x,y,a,b\ ; \tag{5.20}$$

- normalization :

$$\sum_{a,b} p(a,b|x,y) = 1. \tag{5.21}$$

Due to the normalization constraint, P lies in a space of dimension $i^2j^2 - 1$.

As stated previously, quantum non-locality cannot be used for instantaneous transfer of classical information. Mathematically, it means that Alice's marginal probabilities cannot depend on Bob's measurement settings y, and vice versa. These constraints can be formally written [35]:

$$\sum_b p(a,b|x,y) = \sum_b p(a,b|x,y') \ \forall a,x,y,y' \tag{5.22}$$

$$\sum_a p(a,b|x,y) = \sum_a p(a,b|x',y) \ \forall a,x,x',y \tag{5.23}$$

Definition 5.4 *The set of all correlations p satisfying this condition is called the no-signaling set $\mathcal{NL}$.*

It is shown in ref. [38] that $\mathcal{NL}$ is an affine subspace of $\mathbb{R}^{i^2j^2}$ of dimension $2(j-1)i + (j-1)^2i^2$

Definition 5.5 *A compact and convex set with a finite number of extreme points (vertices) and delimited by a finite set of hyperplanes (facets) is called a polytope.*

Definition 5.6 *The set of $P \in \mathbb{R}^{i^2j^2}$. admitting a local decomposition like in (4.17) is a polytope [33]. This local polytope is denoted by $\mathcal{L}$.*

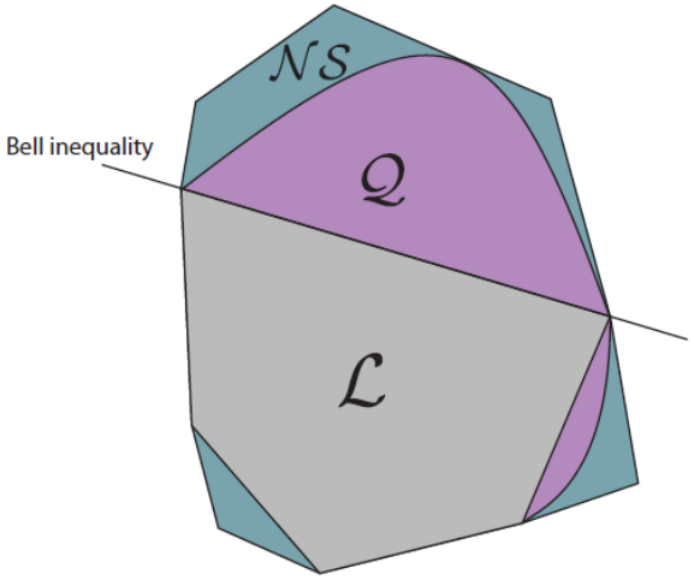
The set of local correlations is a strictly smaller subset of the no-signaling set: $\mathcal{L} \subset \mathcal{NL}$.

The facets of the local polytope are hyperplanes characterized by inequalities. The facets determined by the positivity constraints (5.20) are also facets of the $\mathcal{NL}$ set [33]. If the correlations admit a decomposition of the form (5.13), all other facet inequalities are satisfied. If not, at least one of these inequalities is violated. The inequalities not derived from the positivity constraints are therefore Bell inequalities by definition [37].

Definition 5.7 *The set of quantum correlations is denoted by $\mathcal{Q}$.*

Quantum correlations do not admit a decomposition of the form (4.17). The set $\mathcal{Q}$ is a subset of $\mathcal{NL}$. It is not a polytope in most cases: it does not have a finite number of vertices and cannot be described by a finite number of linear inequalities. $\mathcal{Q}$ still respects the positivity constraints, therefore its boundary contains flat regions [33]. These flat boundary regions are shared with the local set $\mathcal{L}$. All extremal points of $\mathcal{L}$ are extremal points of $\mathcal{Q}$. Similarly to Bell inequalities, there exists a maximum value achievable by quantum behaviour, called Tsirelson's bound. This value is $\sqrt{2}/2$ in the CHSH case [35] [33]. Fig.5.5 shows a representation of the local $\mathcal{L}$, quantum $\mathcal{Q}$ and no-signaling $\mathcal{NL}$ sets.

Figure 5.3: Geometric representation of the $\mathcal{L}$, $\mathcal{Q}$ and $\mathcal{NL}$ sets.



Chapter 6

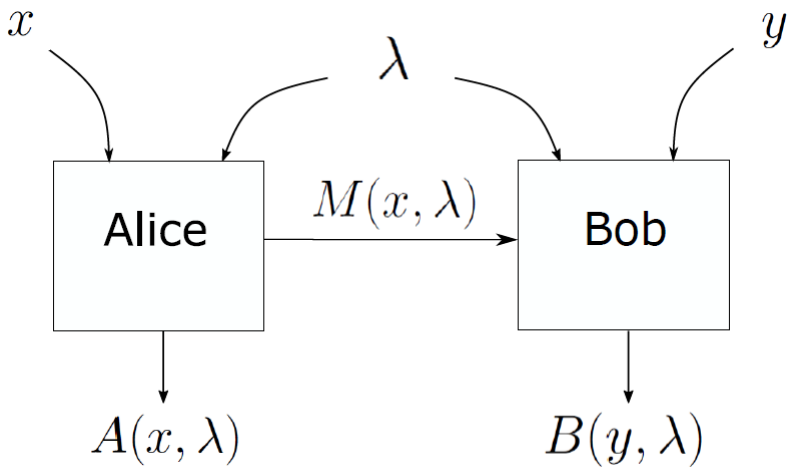
State of the art

6.1 The problem

Although we have seen that a LHV model is not able to reproduce the quantum correlations by itself, the distribution can be simulated with the help of additional resources. The amount of resources necessary to simulate the quantum correlations gives us a way to quantify non-locality. Various resources can be used [10], like a non-local box [39], or post-selection [10] (i.e. authorizing the protocol to abort in some cases).

The resource we are interested in here is classical communication. Similarly to the model presented in chapter 3, we can define protocols, or strategies, to that end. We will mostly consider one-way protocols with communication exchanges from Alice to Bob for simplification. Note that the most efficient known protocols are actually one-way [17]. Fig.6.1 shows a model considering a one-way protocol $\mathcal{P}$ with transcript $M_{\mathcal{P}}(\lambda, x)$.

Figure 6.1: LHV model with one-way communication.



We already know that reproducing $\mathbf{p}_{\mu=1/2}$ with a LHV does not need any exchange of communication. What we are looking for is upper and lower bounds for the complexities of $\mathbf{p}_{\mu=\sqrt{2}/2}$, $\mathbf{p}_{dim=3}$ or more generally $\mathbf{p}_{dim=d}$. As presented in section 3.2, we also have to identify the input distribution maximizing the communication costs. It can be shown that for $\mathbf{p}_{\sqrt{2}/2}$ a uniform distribution is the most detrimental. For $\mathbf{p}_{dim\geq 3}$, the costs are maximised by distributions with uniform marginals [40]. Thanks to theorem (3.1), we do not have to precisely determine which distribution is the worst.

As any possible randomness in the protocol can be included in the random string λ, we can limit ourselves to deterministic protocols without loss of generality. Any protocol $\mathcal{P}$ can be written as a probabilistic combination of deterministic protocols $\mathcal{D}^\lambda$ with distribution $g(\lambda)$ [41] :

$$\mathcal{P} = \sum_\lambda g(\lambda)\mathcal{D}^\lambda \tag{6.1}$$

These protocols can be regrouped in subsets indexed by λ^i, in which every protocol require the same amount of communication c_i[41]:

$$C(\mathcal{D}^{\lambda_i}) = c_i \ \forall \lambda_i \tag{6.2}$$

The local protocols therefore belong to the subset $\mathcal{D}_0$ for which $c_0 = 0$.

A Bell inequality can be seen as a vector $\mathbf{b}$ associating to any probability distribution $\mathbf{p}$ number:

$$B(\mathbf{p}) = \mathbf{b} \cdot \mathbf{p} \tag{6.3}$$

We can write B_0 the maximum value taken by $B(\mathbf{p})$ for protocols $\mathcal{P} \in \mathcal{D}_0$. Considering the decomposition in deterministic protocols, Bell inequalities are violated if $\mathbf{p}$ requires the use of a protocol belonging to a set $\mathcal{D}_i, c_i > 0$.

Previous results

Results bounding the communication complexities of $\mathbf{p}_{\mu=\sqrt{2}/2}$, $\mathbf{p}_{dim=3}$ and $\mathbf{p}_{dim=d}$ are presented here.

Three-dimensional inputs The simulation of $\mathbf{p}_{dim=3}$ as been researched in details:

- Toner and Bacon developed a protocol solving $\mathbf{p}_{dim=3}$ with 1 bit of communication on average [40]. At most 1 bit needs to be exchanged, giving $C_w(\mathbf{p}_{dim=3}) \leq 1$ bit;
- from a claim by Elitzur [42], it was shown by Pironio in ref. [43] that $C(\mathbf{p}_{dim=3}) \geq 1$, establishing together with Toner and Bacon's protocol that $C(\mathbf{p}_{dim=3}) = 1$ bit;
- Roland and Szegedy proved that $C_H(\mathbf{p}_{dim=3}) \leq \sqrt{2} - 1 \approx 0.41$ bit and that $C^{\rightarrow}_{\infty}(\mathbf{p}_{dim=3}) \geq 1 - H_{[\sqrt{2}/2]} \approx 0.13$ bit;
- using a protocol by Degorre, Laplante and Roland [10], it is also shown in ref. [17] that $C_\infty(\mathbf{p}_{dim=3}) \leq 1 - 1/(2\ln 2) \approx 0.28$ bit.

These various bounds actually apply even for $\mathbf{p}_{\mu=\sqrt{2}/2}$. It is also of note that the bound on $C^{\rightarrow}_{\infty}$ was derived with a CHSH experiment in mind. This means that it is not necessarily optimized for $\mathbf{p}_{dim=d}$. The technique that we will use to prove the optimality of this bound for $\mathbf{p}_{\mu=\sqrt{2}/2}$ could in fact be useful to derive one on $\mathbf{p}_{dim=3}$.

Arbitrary number of dimensions Considering an EPR-Bohm experiment with d-dimensional vectors as inputs, the following bounds where derived:

- Regev and Toner provided a protocol reproducing $\mathbf{p}_d$ with at most 2 bits and 1.82 bits on average, yielding $C_w(\mathbf{p}_d) \leq 2$ and $C(\mathbf{p}_d) \leq 1.82$;

- again in ref[17] it was shown that $C_H(\mathbf{p}_d) \leq \sqrt{2}-1$ and $C_\infty(\mathbf{p}_d) \leq 1/2 \ln 2(\ln \pi - \gamma)$, where γ is the Euler-Mascheroni constant.

The proofs for some of these bounds are presented in the following subsections. The emphasis is put on $\mathbf{p}_{\mu=\sqrt{2}/2}$, since it is the problem that we are working on.

6.2 Worst-case and average complexity

Upper bound

The following protocol is due to Toner and Bacon [40] and simulates $\mathbf{p}_{\sqrt{2}/2}$ using 1 bit of communication.

In this protocol, Alice and Bob share two independent unit vectors $\vec{\lambda_1}$ and $\vec{\lambda_2}$ distributed uniformly over the unit sphere $\mathbb{S}_2$. They follow the subsequent steps:

- Alice outputs $a = \sigma(\vec{x} \cdot \vec{\lambda_1})$, where $\sigma(x)$ is the sign function:

$$\sigma(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ -1 & \text{if } x < 0; \end{cases} \tag{6.4}$$

- Alice sends $c = \sigma(\vec{x} \cdot \vec{\lambda_1})\sigma(\vec{x} \cdot \vec{\lambda_2})$ to Bob;

- Bob outputs $b = \sigma\left[(\vec{y} \cdot (\vec{\lambda_1} + c\vec{\lambda_2})\right]$.

Fig.6.2 shows a representation of Alice's output depending on $\vec{\lambda_1}$ and the position of her input on the Bloch sphere. Alice outputs -1 if $\vec{x}$ is in the darkened area and 1 if not. Fig.6.3 presents similar constructions for the sent bit c as well as Bob's output, functions of $\vec{\lambda_1}$ and $\vec{\lambda_2}$, and $\vec{\lambda_1}$, $\vec{\lambda_2}$ and c respectively.

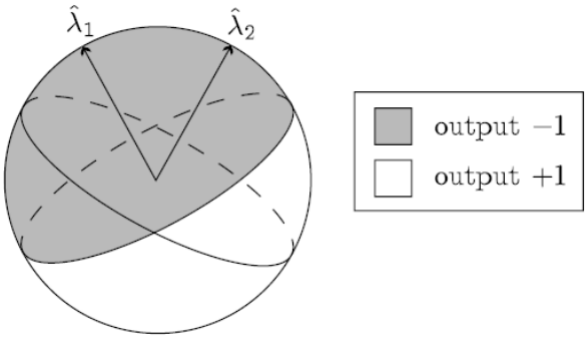


Figure 6.2: Alice's output for a given $\vec{\lambda_1}$.

As expected, $\langle a \rangle = \langle b \rangle = 0$. We also have $\langle ab \rangle = -\vec{x} \cdot \vec{y}$. The details of the calculations are presented in appendix A.2.

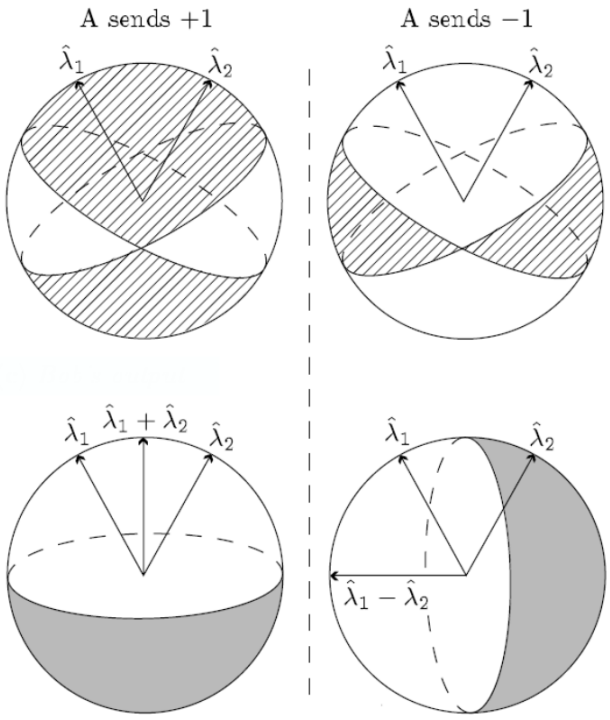


Figure 6.3: Sent bit c and Bob's output for given $\vec{\lambda_1}$ and $\vec{\lambda_2}$.

Lower bound

The following proof of the 1 bit lower bound of $C_w(\mathbf{p})$ was derived by Pironio in ref.

The idea is to write the correlations as a mixture of local and non-local components:

$$\mathbf{p}^{QM}(a,b|x,y) = q\mathbf{p}^{L}(a,b|x,y) + (1-q)\mathbf{p}^{NL}(a,b|x,y), \tag{6.5}$$

where $q \in [0;1]$. Elitzur, Popescu and all showed in ref. that the Bell inequality eqref always hold for such a mixture unless $q = 0$, i.e. unless the correlations are completely non-local. This implies that every protocol reproducing $\mathbf{p}$ needs at least 1 bit of communication.

6.3 Entropic complexity

Lower bound

The following technique was developed by Pironio to derive a lower bound on the average communication complexity. While this lower bound is now obsolete, the approach can still be used to derive a lower bound on the entropic complexity.

We recall that any protocol $\mathcal{P}$ can be written as a combination of deterministic protocols, with the shared randomness λ determining which one is used in a particular instance of $\mathcal{P}$. Reproducing a non-local probability distribution requires protocols with communication, i.e. protocols in $D_i, c_i > 0$. For each of these protocols, a given communication (resp.

entropic) cost produces a violation of a Bell inequality. The deterministic protocol with the highest violation per cost gives a lower bound on the communication (resp. entropic) complexity.

This idea was used in ref.[] to derive the following bound.

Theorem 6.1 *The following bound on entropic cost holds:*

$$C_H^D(\mathbf{p}) \geq \frac{B(\mathbf{p}) - B_0}{B(\mathbf{p}_{\lambda*}) - B_0} C_H^D(\mathcal{P}_{\lambda*}), \tag{6.6}$$

with:

- $\mathbf{p}$ *any probability distribution;*
- $\mathbf{p}_\lambda$ *probability distributions simulated by deterministic protocols* $\mathcal{P}_\lambda$ *and such that* $\mathbf{p} = \sum_\lambda p(\Lambda = \lambda)\mathbf{p}_\lambda$*;*
- $B(\mathbf{p}) = \mathbf{b} \cdot \mathbf{p} \leq B_0$ *a Bell inequality and* B_0 *the maximum value of* $B(\mathbf{p})$ *for local distributions;*
- $\mathbf{p}_{\lambda^*}$ *the probability density maximizing:*

$$\frac{\mathbf{b} \cdot \mathbf{p}_\lambda - B_0}{C_H(\mathcal{P}_\lambda)}. \tag{6.7}$$

Proof The full derivation of this expression is done in appendix A.3.

In a CHSH setting, $B(\mathbf{p})$ is the CHSH inequality (5.16) and we have $B_0 = 2$. The most detrimental distribution is the uniform distribution: $C_H(\mathbf{p}_\mu) = C_H^U(\mathbf{p}_\mu)$. The maximum violation per entropy is attained when Alice sends Bob her output, with $C_H^U(\mathcal{P}_{\lambda*}) = 1$ and $B(\mathbf{p}_{\lambda*}) = 4$.
In that case we finally obtain the bounds:

$$C_H(\mathbf{p}_\mu) \geq \frac{4\mu - 2}{4 - 2} = 2\mu - 1 \tag{6.8}$$

$$C_H(\mathbf{p}_{\sqrt{2}/2}) \geq \sqrt{2} - 1 \tag{6.9}$$

6.4 Information complexity

Upper bound

The protocol used in this proof was presented in ref. [10]. Following an idea by Feldmann ([45]), ref.[10] shows that if Alice and Bob are able to share a biased variable following the distribution:

$$\rho(\vec{\lambda}|\vec{x}) = \frac{|\vec{x} \cdot \vec{\lambda}|}{2\pi}, \tag{6.10}$$

then the correlations $\mathbf{p}_{dim=3}$ can be reproduced. Indeed if Alice and Bob set their outputs to:

$$A(x, \lambda) = -\sigma(\vec{x} \cdot \vec{\lambda}) \tag{6.11}$$

$$B(y, \lambda) = \sigma(\vec{x} \cdot \vec{\lambda}), \tag{6.12}$$

they get $\langle A\rangle = \langle B\rangle = 0$ as desired. Additionally, as expected:

$$\begin{aligned}\langle AB\rangle &= \int_{\mathbb{S}_2} \rho(\lambda|x)A(\vec{x},\vec{\lambda})B(\vec{y},\vec{\lambda})d\lambda &(6.13)\\ &= -\frac{1}{2\pi}\int_{\mathbb{S}_2} \vec{x}\cdot\vec{\lambda}\ \sigma(\vec{y}\cdot\vec{\lambda})d\lambda &(6.14)\\ &= -\vec{x}\cdot\vec{y} &(6.15)\end{aligned}$$

The protocol can be split in two steps:

- Alice used the shared randomness to sample a bias variable following the density (6.10);
- Alice shares the sampled variable with Bob using communication.

Thanks to the bound (3.18), we only have to compute the information cost of this protocol $I(X;\Lambda) = H(\Lambda) - H(\Lambda|X)$.

X being uniformly distributed on the sphere, we have:

$$\begin{aligned}\rho(\vec{x}) &= \frac{1}{4\pi} &(6.16)\\ \rho(\vec{\lambda}|\vec{x}) &= \frac{1}{4\pi}, &(6.17)\end{aligned}$$

yielding:

$$\begin{aligned}H(\Lambda) &= -\int_{\mathbb{S}_2} \rho(\lambda)\log\rho(\lambda)d\lambda &(6.18)\\ &= \log 4\pi. &(6.19)\end{aligned}$$

The conditional entropy is also independent of the value of $\vec{x}$, giving:

$$\begin{aligned}H(\Lambda|X) &= \int_{\mathbb{S}_2} H(\Lambda|X=\vec{x})d\lambda &(6.20)\\ &= H(\Lambda|X=\vec{x_0}). &(6.21)\end{aligned}$$

Taking x_0 along the z-axis, we evaluate the conditional entropy:

$$\begin{aligned}H(\Lambda|X) &= -\frac{1}{2\pi}\int_0^{2\pi} d\phi \int_0^{\pi} d\theta \sin\theta|\cos\theta| \log\frac{|\cos\theta|}{2\pi} &(6.22)\\ &= -2\int_0^1 du\ u\log\frac{u}{2\pi} = \log 2\pi - 2\int_0^1 du\ u\log u &(6.23)\\ &= \log 2\pi + \frac{1}{2\ln 2} &(6.24)\end{aligned}$$

We finally have:

$$I(\Lambda;X) = \log 4\pi - \log 2\pi - \frac{1}{2\ln 2} = 1 - \frac{1}{2\ln 2}, \quad (6.25)$$

and therefore:

$$C_I(\mathbf{p}_{dim=3}) \leq 1 - \frac{1}{2\ln 2} \approx 0.28 \text{ bit}. \quad (6.26)$$

Lower bound

The following derivation of a lower bound on the information complexity of $\mathbf{p}_{\mu=\sqrt{2}/2}$. It was presented by Roland and Szegedy in ref.[17]. It is based on the fact that Bob can evaluate his output as a function his input and Alice's message.

We denote by b Bob's output when $y = 0$ and b' his output when $y = 1$. Since $a \in \{-1, 1\}$, we have:

$$(ab)(ab') = bb' \tag{6.27}$$

The expression of $\mathbf{p}_\mu$ (5.6) yields:

$$p_\mu(ab = 1|x, y = 0) = \frac{1+\mu}{2} \tag{6.28}$$

$$p_\mu(ab' = (-1)^x|x, y = 1) = \frac{1+\mu}{2}, \tag{6.29}$$

meaning that:

$$p(bb' = (-1)^x) \geq \mu. \tag{6.30}$$

We note E the random variable with realization $e = bb'$. Since $x \in \{0, 1\}$ we have:

$$e = (-1)^x \tag{6.31}$$

$$= 1 - 2x \tag{6.32}$$

$$\Leftrightarrow x = \frac{1-e}{2} \tag{6.33}$$

for $\mu \in \left[\frac{1}{2}; 1\right]$. From (6.30) we get:

$$p(x|e) \geq \mu \tag{6.34}$$

and:

$$I(X; E) = H(X) - H(X|E) \tag{6.35}$$

$$\geq 1 - H_{[\mu]} \tag{6.36}$$

Since E is a function of the message $m \in M$, $X \to M \to E$ form a Markov chain and the data processing inequality gives:

$$I(X; M) \geq I(X; E) \geq 1 - H_{[\mu]}, \tag{6.37}$$

therefore proving the bound:

$$C_I(\mathbf{p}_\mu) \geq 1 - H_{[\mu]} \approx 0.13 \text{ bit}. \tag{6.38}$$

This bound is actually the optimal value, as we will show in the next chapter:

$$C_I(\mathbf{p}_\mu) = 1 - H_{[\mu]}. \tag{6.39}$$

For quantum correlations, $\mu = \sqrt{2}/2$ and:

$$C_I(\mathbf{p}_{\mu=\sqrt{2}/2}) = 1 - H_{[\sqrt{2}/2]} \approx 0.13 \text{ bit}. \tag{6.40}$$

Chapter 7

CHSH optimization problem

We will prove in this section the optimality of bound (6.39). Considering binary inputs and binary outputs, we are trying to solve the following optimization problem:

$$\min_{\substack{p(\lambda)\\p(x,y,\lambda)}} \quad I(\Lambda; X, Y) \tag{7.1a}$$

$$\text{subject to} \quad p(\lambda) - \sum_{x,y} p(x,y,\lambda) = 0 \quad \forall\lambda \tag{7.1b}$$

$$p(a,b,x,y) - \sum_{\lambda} p(x,y,\lambda) l_\lambda(a,b|x,y) = 0, \quad \forall a,b \in \{-1,1\} \text{ and } \forall x,y \in \{0,1\} \tag{7.1c}$$

where we note $l_\lambda(a,b|x,y)$ the deterministic local distributions determining the outputs as a function of the inputs. Constraint (7.1b) comes from a general property of probability distributions. Constraint (7.1c) translates the fact that the distribution of local functions governed by $p(x,y,\lambda)$ has to reproduce the joint distribution $p(a,b,x,y) = p_\mu(a,b|x,y)p(x,y)$. Note that the normalization constraint of $p(x,y,\lambda)$ is not explicitly written. In other words, we optimize $I(\Lambda; X, Y)$ on $p(x,y,\lambda)$ belonging to the probability density space.

We are going to dualize this problem and show that the bound (6.39) is also solution of the dual.

One-way problem

If we suppose only one-way communication, with Alice sending messages to Bob, Λ no longer depends on Y. We can write $p(x,y,\lambda) = p(x,\lambda)p(y)$ and our problem becomes:

$$\min_{\substack{p(\lambda)\\p(x,\lambda)}} \quad I(\Lambda; X) \tag{7.2a}$$

$$\text{subject to} \quad p(\lambda) - \sum_{x} p(x,\lambda) = 0 \quad \forall\lambda \tag{7.2b}$$

$$p(a,b,x,y) - \sum_{\lambda} p(x,\lambda)p(y) l_\lambda(a,b|x,y) = 0 \quad \forall a,b \in \{-1,1\} \text{ and } \forall x,y \in \{0,1\}, \tag{7.2c}$$

The two constraints respectively hold for all possible λ and all possible inputs and outputs. The Lagrangian of problem (7.2) is therefore written:

$$I(\Lambda;X)+\sum_{\lambda}\alpha_{\lambda}\left(p(\lambda)-\sum_{x}p(x,\lambda)\right)+\sum_{abxy}\beta_{abxy}\left(p(a,b,x,y)-\sum_{\lambda}p(x,\lambda)p(y)l_{\lambda}(a,b|x,y)\right). \tag{7.3}$$

We are trying to maximize this quantity over the Lagrange multipliers α and β and minimize it over the probability distributions $p(x,\lambda)$ and $p(\lambda)$.

Minimization over $p(x,\lambda)$ Lemma 4.3 gives the following result:

$$\min_{p(x,\lambda)} I(\Lambda;X)-\sum_{\lambda}\alpha_{\lambda}\sum_{x}p(x,\lambda)-\sum_{abxy}\beta_{abxy}\sum_{\lambda}p(x,\lambda)p(y)l_{\lambda}(a,b|x,y)=-\log\sum_{\lambda}p(\lambda)D(\lambda)2^{\alpha_{\lambda}}, \tag{7.4}$$

where we wrote:

$$D(\lambda):=\sum_{x}p(x)2^{\sum_{aby}\beta_{abxy}p(y)l_{\lambda}(a,b|x,y)} \tag{7.5}$$

Since strong duality holds, we are also free to exchange max and min as we wish. We are therefore left with the following expression:

$$\min_{p(\lambda)}\max_{\alpha,\beta}\sum_{abxy}\beta_{abxy}p(a,b,x,y)+\sum_{\lambda}\alpha_{\lambda}p(\lambda)-\log\sum_{\lambda}p(\lambda)D(\lambda)2^{\alpha_{\lambda}} \tag{7.6}$$

Note that because of the sum over all λ, the last term does not depend on λ. Hence we can write:

$$C:=\sum_{\lambda}p(\lambda)D(\lambda)2^{\alpha_{\lambda}}. \tag{7.7}$$

Maximization over α

Claim 1 *The function $f(\alpha)$ defined by:*

$$f(\alpha):=\sum_{\lambda}\alpha_{\lambda}p(\lambda)-\log C \tag{7.8}$$

has a single maximum given by:

$$f(\alpha^{*}):=\max_{\alpha}f(\alpha)=-\sum_{\lambda}p(\lambda)\log D(\lambda) \tag{7.9}$$

and attained for:

$$\alpha_{\lambda}^{*}=\log C-\log D(\lambda) \tag{7.10}$$

Proof: Each α_{λ} is independent of the others, we can therefore derive $f(\alpha)$ with respect to each of them:

$$\frac{\partial}{\partial\alpha_{\lambda}}f(\alpha) = p(\lambda)-\frac{p(\lambda)D(\lambda)2^{\alpha_{\lambda}}}{C} \tag{7.11}$$

$$\frac{\partial^{2}}{\partial\alpha_{\lambda}^{2}}f(\alpha) = -\frac{\ln 2(p(\lambda)D(\lambda)2^{\alpha_{\lambda}})(C-p(\lambda)D(\lambda)2^{\alpha_{\lambda}})}{C^{2}}. \tag{7.12}$$

Since $C \geq p(\lambda)D(\lambda)2^{\alpha_\lambda}$, the second derivative of $f(\alpha)$ is negative. $f(\alpha)$ is therefore strictly concave and has a single maximum attained for the α^*_λ such that:

$$\frac{\partial}{\partial \alpha_\lambda} f(\alpha)\Big|_{\alpha^*_\lambda} = 0 \quad \forall \lambda \tag{7.13}$$

Hence:

$$\begin{aligned} \frac{\partial}{\partial \alpha_\lambda} f(\alpha)\Big|_{\alpha^*_\lambda} &= 0 && (7.14) \\ \Leftrightarrow \qquad \alpha^*_\lambda &= \log C - \log D(\lambda) && (7.15) \end{aligned}$$

and:

$$\begin{aligned} f(\alpha^*) &= \sum_\lambda p(\lambda)(\log C - \log D(\lambda)) - \log C && (7.16) \\ &= -\sum_\lambda p(\lambda) \log D(\lambda) && (7.17) \end{aligned}$$

Problem (7.1) becomes:

$$\min_{p(\lambda)} \max_\beta \sum_{abxy} \beta_{abxy} p(a,b,x,y) - \sum_\lambda p(\lambda) \log D(\lambda) \tag{7.18}$$

Minimization over $p(\lambda)$ Notice that expression (7.18) can be written:

$$\max_\beta \sum_{abxy} \beta_{abxy} p(a,b,x,y) - \max_{p(\lambda)} \sum_\lambda p(\lambda) \log D(\lambda) \tag{7.19}$$

The second term is the expected value of $\log D(\lambda)$ over all the λ. It is therefore maximized by a deterministic distribution such that:

$$p(\lambda) = \begin{cases} 1 \text{ if } D(\lambda) \geq D(\lambda') \ \forall \lambda' \\ 0 \text{ else.} \end{cases} \tag{7.20}$$

Our problem therefore comes down to the expression:

$$\max_\beta \sum_{abxy} \beta_{abxy} p(a,b,x,y) - \log D(\lambda^*), \tag{7.21}$$

with the constraint:

$$\log D(\lambda) \leq \log D(\lambda^*) \quad \forall \lambda. \tag{7.22}$$

CHSH bound

The first term in (7.21) has the form a Bell inequality. We know that it is maximized by coefficients written:

$$\beta_{abxy} = \beta(-1)^{xy} ab \tag{7.23}$$

The multiplication factor β was omitted in previous derivations. Since our problem is non-linear, it is of importance here.

Evaluation of $D(\lambda^*)$

Claim 2 *The maximum value of $D(\lambda)$ is given by:*

$$D(\lambda^*) = \frac{1}{2}(2^\beta + 1) \tag{7.24}$$

Proof: Injecting β_{abxy} of $D(\lambda)$ gives:

$$D(\lambda^*) = \sum_x p(x) 2^{\beta \sum_y p(y)(-1)^{xy} \sum_{ab} ab \; l_{\lambda^*}(a,b|x,y)} \tag{7.25}$$

The $l_{\lambda^*}(a,b|x,y)$ are deterministic distributions governing the outputs. They represent all sixteen possible strategies Alice and Bob can use. They are regrouped in the following table:

λ	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8	λ_9	λ_{10}	λ_{11}	λ_{12}	λ_{13}	λ_{14}	λ_{15}	λ_{16}
$A(0)$	1	−1	1	−1	1	−1	1	−1	1	−1	1	−1	1	−1	1	−1
$A(1)$	1	1	−1	−1	1	1	−1	−1	1	1	−1	−1	1	1	−1	−1
$B(0)$	1	1	1	1	−1	−1	−1	−1	1	1	1	1	−1	−1	−1	−1
$B(1)$	1	1	1	1	1	1	1	1	−1	−1	−1	−1	−1	−1	−1	−1

We can therefore write the sum over a and b as a function of the inputs:

$$\sum_{ab} ab \; l_{\lambda^*}(a,b|x,y) = A(x,\lambda^*)B(y,\lambda^*) = A^*(x)B^*(y), \tag{7.26}$$

Note that the deterministic distributions are local, meaning that A (respectively B) only depends on x (respectively y) and λ. This gives:

$$D(\lambda^*) = \sum_x p(x) 2^{\beta \sum_y p(y)(-1)^{xy} A^*(x)B^*(y)}. \tag{7.27}$$

We are still looking for A^* and B^* maximizing $D(\lambda)$. In the exponent at most three of the four $(-1)^{xy}A^*(x)B^*(y)$ terms are positive. We also know that the distribution of inputs that maximizes the information cost is uniform. Hence:

$$\begin{aligned} D(\lambda^*) &= \frac{1}{2}(2^{(A^*(0)B^*(0)+A^*(0)B^*(1))\beta/2} + 2^{(A^*(1)B^*(0)-A^*(1)B^*(1))\beta/2}) && (7.28)\\ &= \frac{1}{2}(2^\beta + 1) && (7.29) \end{aligned}$$

Maximization over β We recall the expression of the CHSH correlations:

$$\mathbf{p}_\mu(a,b|x,y) = \frac{1 + \mu ab(-1)^{xy}}{4}, \tag{7.30}$$

with $\mu = \sqrt{2}/2$ in the quantum case and $\mu = 1/2$ in the classical case. Injecting the expression of β_{abxy} in the first term of expression (7.21) yields:

$$\begin{aligned} \sum_{abxy} \beta_{abxy} p(a,b,x,y) &= \sum_{abxy} \beta_{abxy} p_\mu(a,b|x,y) p(x,y) && (7.31)\\ &= \beta \sum_{xy} p(x,y)(-1)^{xy} \sum_{ab} ab \; p(a,b|x,y) && (7.32)\\ &= \beta\mu. && (7.33) \end{aligned}$$

Claim 3 *The function $g(\beta)$ defined by:*

$$g(\beta) := \beta\mu - \log \frac{1}{2}(2^\beta + 1) \tag{7.34}$$

has a single maximum given by:

$$g(\beta^*) := \max_\beta g(\beta) = \mu \log \frac{\mu}{1-\mu} - \log \frac{1}{2}\left(\frac{\mu}{1-\mu} + 1\right) \tag{7.35}$$

and attained for:

$$\beta^* = \log \frac{\mu}{1-\mu} \tag{7.36}$$

Proof: We optimize $g(\beta)$ by differentiating it with respect to β:

$$\begin{aligned} g(\beta) &:= \beta\mu - \log \frac{1}{2}(2^\beta + 1) && (7.37)\\ \frac{\partial}{\partial\beta} g(\beta) &= \mu - \frac{2^\beta}{2^\beta + 1} && (7.38)\\ \frac{\partial^2}{\partial\beta^2} g(\beta) &= -\frac{\ln 2\left((2^\beta+1)2^\beta + 2^{2\beta}\right)}{(2^\beta+1)^2}. && (7.39) \end{aligned}$$

Once again $g(\beta)$ is strictly concave and has a single maximum:

$$\begin{aligned} \frac{\partial}{\partial\beta} g(\beta)\Big|_{\beta^*} &= 0 && (7.40)\\ \Leftrightarrow\ \beta^* &= \log \frac{\mu}{1-\mu}, && (7.41) \end{aligned}$$

yielding the final optimal value of our program:

$$\mu \log \frac{\mu}{1-\mu} - \log \frac{1}{2}\left(\frac{\mu}{1-\mu} + 1\right) \tag{7.42}$$

We first note that for $\mu = 1/2$ we get an optimized value of 0 as expected: it is in that case possible to reproduce the correlation with local model without communication and the information complexity is 0.

Claim 4 *The information complexity $C_I^{\rightarrow}(\mathbf{p}_\mu)$ is given by:*

$$C_I(\mathbf{p}_\mu) = 1 - H_{[\mu]} \tag{7.43}$$

where $H_{[\mu]}$ is the binary entropy:

$$H_{[\mu]} = -\mu \log \mu - (1-\mu)\log(1-\mu) \tag{7.44}$$

Proof: A closer look at our expression gives:

$$\begin{aligned} & \mu \log \frac{\mu}{1-\mu} - \log \frac{1}{2}\left(\frac{\mu}{1-\mu} + 1\right) && (7.45)\\ =\ & 1 + \mu \log \mu + (1-\mu)\log(1-\mu) && (7.46)\\ =\ & 1 - H_{[\mu]} && (7.47) \end{aligned}$$

This means that the expression found in 6.4 is solution of both problem (7.2) and its dual. Since we have strong duality between these problems, this expression is optimal.

Two-way problem

We now come back to problem (7.1):

$$\min_{\substack{p(\lambda)\\p(x,y,\lambda)}} \quad I(\Lambda; X, Y) \tag{7.48a}$$

$$\text{subject to} \quad p(\lambda) - \sum_{x,y} p(x,y,\lambda) = 0 \tag{7.48b}$$

$$p(a,b,x,y) - \sum_{\lambda} p(x,y,\lambda) l_\lambda(a,b|x,y) = 0, \tag{7.48c}$$

We will again derive the optimal value of this program, this time without any assumption, i.e. Λ can depend on both X and Y. The Lagrangian is this time written:

$$I(\Lambda; X, Y) + \sum_{\lambda} \alpha_\lambda \left(p(\lambda) - \sum_{xy} p(x,y,\lambda) \right) + \sum_{abxy} \beta_{abxy} \left(p(a,b,x,y) - \sum_{\lambda} p(x,y,\lambda) l_\lambda(a,b|x,y) \right). \tag{7.49}$$

Once again we want to maximize this quantity over the Lagrange multipliers α and β and minimize it over the probability distributions $p(x,\lambda)$ and $p(\lambda)$.

Claim 5 *The optimal value of expression* (7.49) *is given by:*

$$\max_{\beta} \sum_{a,xy} \beta_{abxy} p(a,b,x,y) - \log D'(\lambda^*), \tag{7.50}$$

with the constraint:

$$\log D'(\lambda) \leq \log D'(\lambda^*) \quad \forall \lambda. \tag{7.51}$$

Proof: Using lemma 4.3, this problem is reduced to:

$$\min_{p(\lambda)} \max_{\alpha,\beta} \sum_{abxy} \beta_{abxy} p(a,b,x,y) + \sum_{\lambda} \alpha_\lambda p(\lambda) - \log \sum_{\lambda} p(\lambda) D'(\lambda) 2^{\alpha_\lambda}, \tag{7.52}$$

this time with:

$$D'(\lambda) := \sum_{xy} p(x,y) 2^{\sum_{a,b} \beta_{abxy} l_\lambda(a,b|x,y)} \tag{7.53}$$

The maximization over α is handled exactly as before, yielding once again:

$$\max_{\beta} \sum_{abxy} \beta_{abxy} p(a,b,x,y) - \max_{p(\lambda)} \sum_{\lambda} p(\lambda) \log D'(\lambda) \tag{7.54}$$

This expression is minimized by the same $p(\lambda)$ and we have:

$$\max_{\beta} \sum_{abxy} \beta_{abxy} p(a,b,x,y) - \log D'(\lambda^*), \tag{7.55}$$

with $D'(\lambda)$ such that:

$$\log D'(\lambda) \leq \log D'(\lambda^*) \quad \forall \lambda. \tag{7.56}$$

Two-way bound

Claim 6 *The information complexity $C_I(\mathbf{p}_\mu)$ is given by:*

$$C_I(\mathbf{p}_\mu) = 2 + \frac{(\mu+1)}{2}\log\chi_\mu - \log(3\chi_\mu + 1), \tag{7.57}$$

with:

$$\chi_\mu := \frac{1-\mu^2}{3(1-\mu)^2} \tag{7.58}$$

Proof: Injecting the CHSH coefficients β_{abxy} in the first term of (7.55) gives $\beta\mu$ as before. In the second term, D' becomes:

$$D'(\lambda^*) = \sum_{xy} p(x,y) 2^{\beta(-1)^{xy}\sum_{ab} ab\ l^*_\lambda(a,b|x,y)} \tag{7.59}$$

Replacing the l_{λ^*} gives:

$$D'(\lambda^*) = \sum_{xy} p(x,y) 2^{\beta(-1)^{xy}A^*(x)B^*(y)}. \tag{7.60}$$

We know by experience that at most three of the four terms have a positive exponent. Considering uniformly distributed input, we get:

$$-\log D'(\lambda^*) = -\log\left[\frac{3}{4}(2^\beta) + \frac{1}{4}(2^{-\beta})\right] \tag{7.61}$$

We now optimize our expression by differentiating with respect to β:

$$\begin{aligned}
g'(\beta) &:= \beta\mu - \log\left[\frac{3}{4}(2^\beta) + \frac{1}{4}(2^{-\beta})\right] && (7.62)\\
\frac{\partial}{\partial\beta}\, g'(\beta) &= \mu - \frac{3(2^\beta) - 2^{-\beta}}{3(2^\beta) + 2^{-\beta}} && (7.63)\\
\frac{\partial^2}{\partial\beta^2}\, g'(\beta) &= -\frac{\ln 2\left[(3(2^\beta)+2^{-\beta})^2 - (3(2^\beta)-2^{-\beta})^2\right]}{3(2^\beta)+2^{-\beta}}. && (7.64)
\end{aligned}$$

$g'(\beta)$ is strictly concave and has a single maximum:

$$\begin{aligned}
\frac{\partial}{\partial\beta}\, g'(\beta)\Big|_{\beta^*} &= 0 && (7.65)\\
\Leftrightarrow \qquad \beta^* &= \frac{1}{2}\log\frac{\mu+1}{3(1-\mu)} && (7.66)\\
&= \frac{1}{2}\log\frac{1-\mu^2}{3(1-\mu)^2} && (7.67)\\
&= \frac{1}{2}\log\chi_\mu, && (7.68)
\end{aligned}$$

with:

$$\chi_\mu := \frac{1-\mu^2}{3(1-\mu)^2} \tag{7.69}$$

This gives the final optimal value of our program:

$$\frac{\mu}{2}\log\chi_\mu - \log\left[\frac{3}{4}\chi_\mu^{1/2} + \frac{1}{4}\frac{1}{\chi_\mu^{1/2}}\right] \tag{7.70}$$

$$= 2 + \frac{\mu}{2}\log\chi_\mu - \log\left[3\chi_\mu^{1/2} + \frac{1}{\chi_\mu^{1/2}}\right] \tag{7.71}$$

$$= 2 + \mu\log\chi_\mu^{1/2} - \log\left[\frac{3\chi_\mu + 1}{\chi_\mu^{1/2}}\right] \tag{7.72}$$

$$= 2 + \frac{(\mu+1)}{2}\log\chi_\mu - \log(3\chi_\mu + 1) \tag{7.73}$$

As expected, the information cost vanishes for $\mu = 1/2$. Indeed we have:

$$\chi_{1/2} = \frac{3/4}{3(1/2)^2} = 1, \tag{7.74}$$

hence:

$$2 + \frac{(\mu+1)}{2}\log\chi_\mu - \log(3\chi_\mu + 1) \tag{7.75}$$

$$= 2 - \log 4 \tag{7.76}$$

$$= 0 \tag{7.77}$$

We can now compute the information complexity of $\mathbf{p}_{\sqrt{2}/2}$ with two-communication. First we have:

$$\chi_{\sqrt{2}/2} = \frac{1/2}{3(1-\sqrt{2}/2)^2} = \frac{1}{6(3/2-\sqrt{2})} = \frac{3/2-\sqrt{2}}{6} \tag{7.78}$$

$$= \frac{3-\sqrt{2}}{12}. \tag{7.79}$$

Injecting into our expression gives:

$$2 + \frac{\mu+1}{2}\log\chi_\mu - \log(3\chi_\mu + 1) \tag{7.80}$$

$$= 2 + \frac{2+\sqrt{2}}{4}\left[\log\left(1-\frac{\sqrt{2}}{3}\right) - 2\right] - \log\left(7-\sqrt{2}\right) + 2 \tag{7.81}$$

$$= 3 - \frac{\sqrt{2}}{2} + \frac{2+\sqrt{2}}{4}\log\left(1-\frac{\sqrt{2}}{3}\right) - \log\left(7-\sqrt{2}\right) \tag{7.82}$$

Chapter 8

Multidimensional problem

We will in this chapter study the EPR-Bohm problem, in particular for $\mathbf{p}_{dim=3}$. The goal is to pave the way for a potential generalization of our previous results. We will only consider one-way communication, with Alice sending messages to Bob.

The corresponding optimization problem is written as follows:

$$\min_{\substack{\rho(\lambda),\rho(\lambda,\vec{x}) \\ A,B}} \quad I(\Lambda; X) \tag{8.1a}$$

$$\text{subject to} \int_\Lambda \rho(\lambda,\vec{x})\rho(\vec{y})A(\vec{x},\lambda)B(\vec{y},\lambda)d\lambda \qquad = -\vec{x}\cdot\vec{y} \quad \forall\vec{x},\vec{y} \tag{8.1b}$$

with:

- $\vec{x},\vec{y} \in \mathbb{S}_2$ Alice's and Bob's inputs;
- $\lambda \in \Lambda$ an output strategy;
- $\rho(\lambda,\vec{x})$ the joint probability density function of λ and $\vec{x}$;
- $A(\vec{x},\lambda)$ and $B(\vec{y},\lambda)$ the functions specifying respectively Alice's and Bob's output.

Difficulties arise from the fact that the strategies λ are this time much more difficult to define.

8.1 General case

Similarly to the proof in subsection 6.4, Bob can evaluate his output for any potential message sent by Alice to determine the strategy. With unit vectors as inputs, this evaluation gives him a Bloch sphere split in a number of regions, and depending on which region his input lies in, his output will either be 1 or -1.

Claim 7 *The result of Bob's evaluation of his output for all possible λ is a sphere with central symmetry.*

Proof: Given the form of the probability distribution (5.5), a and b are anti-correlated if $x = y$ and correlated when $x = -y$, meaning that:

$$A(y,\lambda) = B(-y,\lambda) \tag{8.2}$$
$$A(y,\lambda) = -B(y,\lambda), \tag{8.3}$$

and therefore:

$$B(y,\lambda) = -B(-y,\lambda). \tag{8.4}$$

This result holds for every possible output functions $A(\lambda,\vec{x})$ and $B(\lambda,\vec{y})$. As is however, it seems insufficient to derive a bound on information complexity given the variety of possible ways to define Λ.

8.2 Hemisphere hypothesis

In the simplest case, the sphere will be divided in two hemispheres, each corresponding to one of Bob's outputs. The content of Alice's message in that situation comes down to a direction specifying the separation plane between the two regions, and a sign for Bob to know which hemisphere corresponds to which output, i.e. a unit vector on $\mathbb{S}_2$ that we will write $\vec{\lambda}$.

It is under this assumption that bound (6.26) was derived:

$$C_I^{\rightarrow}(\mathbf{p}_{dim=3}) \leq 1 - \frac{1}{2\ln 2} \tag{8.5}$$

This result was formulated in term of the conditional probability density $\rho(\vec{\lambda}|\vec{x})$ and the density $\rho(\vec{x})$ was supposed to be constant. The optimizers were:

$$\rho(\vec{\lambda}|\vec{x}) = \frac{|\vec{x}\cdot\vec{\lambda}|}{4\pi} \tag{8.6}$$

$$A(\vec{x},\vec{\lambda}) = -\sigma(\vec{x}\cdot\vec{\lambda}) \tag{8.7}$$

$$B(\vec{y},\vec{\lambda}) = \sigma(\vec{y}\cdot\vec{\lambda}), \tag{8.8}$$

with $\sigma(x)$ the sign function. Using the same argument as in 8.1, if Bob's output is specified by $\sigma(\vec{y}\cdot\vec{\lambda})$, we necessarily have $A = -\sigma(\vec{x}\cdot\vec{\lambda})$, justifying the choice of the functions A and B (8.7) and (8.8).

Proving that the bound (6.26) is optimal can therefore by divided in two steps:

- showing that due to the symmetries of the problem, any division of the Bloch sphere by Bob can be brought back to the simplest case of two hemispheres;
- showing that under this assumption, the minimum mutual information $I(\Lambda;X)$ is obtained with the conditional density (8.6) and the output specifications (8.7) and (8.8).

8.3 Multidimensional optimization problem

Under the hemisphere hypothesis, the multidimensional optimization problem can be written as follows:

$$\min_{\substack{\rho(\vec{\lambda}),\rho(\vec{\lambda},\vec{x}) \\ A,B}} \quad I(\Lambda;X) \tag{8.9a}$$

$$\text{subject to} \int_{\mathbb{S}_2} \rho(\vec{\lambda},\vec{x})\rho(\vec{y})\sigma(\vec{x}\cdot\vec{\lambda})\sigma(\vec{y}\cdot\vec{\lambda})d\vec{\lambda} \qquad = -\vec{x}\cdot\vec{y} \quad \forall\vec{x},\vec{y} \tag{8.9b}$$

As we did in the CHSH case, we can used lemma 4.3 to derive a general expression of the dual problem.

Conclusion

Appendix A

Proofs

A.1 Isotropy of the singlet state

The following proof is derived from ref[].

Let us consider a pair of qubits in the singlet state $|\psi^-\rangle$. We measure the spin of both qubits along an arbitrary direction $\vec{u}$. The eigenstates of the observable $\vec{u}\cdot\vec{S}$ are denoted $|a\rangle$ and $|b\rangle$.

The states $|0\rangle$ and $|1\rangle$ can be written as a superposition of $|a\rangle$ and $|b\rangle$:

$$|0\rangle = \alpha\,|a\rangle + \beta\,|b\rangle \tag{A.1}$$
$$|1\rangle = \xi\,|a\rangle + \kappa\,|b\rangle\,, \tag{A.2}$$

with the normalization imposing $|\alpha|^2 + |\beta|^2 = |\xi|^2 + |\kappa|^2 = 1$.

The singlet state can be rewritten:

$$\frac{|01\rangle - |10\rangle}{\sqrt{2}} = (\alpha\kappa - \beta\xi)\frac{|ab\rangle - |ba\rangle}{\sqrt{2}}. \tag{A.3}$$

Thanks to the normalization conditions, the term $(\alpha\kappa - \beta\xi)$ can be interpreted as the determinant of a unitary matrix and therefore an overall phase factor. As states in section ref chap 1, this factor has no observable consequence and can be ignored. Hence the choice of direction $\vec{u}$ has no observable effect and $|\psi^-\rangle$ is isotropic.

A.2 CHSH inequality

This proof of the CHSH inequality follows the approach presented in [] and [].

Local predictions The local expectations are defined as follows:

$$\langle a_x\rangle_\lambda = \sum_a a\; p(a|x,\lambda) \tag{A.4}$$
$$\langle b_y\rangle_\lambda = \sum_b b\; p(b|y,\lambda), \tag{A.5}$$

where $x, y \in \{0, 1\}$ and $a, b \in \{-1, 1\}$. The expectation value of the product $\langle a_x b_y \langle$ can be written as an average on the shared randomness of the local expectations:

$$\begin{aligned} \langle a_x b_y \rangle &= \sum_{a,b} ab\ p(a, b|x, y) && \text{(A.6)} \\ &= \int_\Lambda q(\lambda) \langle a_x \rangle_\lambda \langle b_y \rangle_\lambda d\lambda, && \text{(A.7)} \end{aligned}$$

with $q(\lambda)$ the probability density function of Λ.

Since $\langle a_x \rangle_\lambda \in [-1; 1]\ \forall x$, we have:

$$\begin{aligned} |S_\lambda| &= |\langle a_0 \rangle_\lambda \langle b_0 \rangle_\lambda + \langle a_1 \rangle_\lambda \langle b_0 \rangle_\lambda + \langle a_0 \rangle_\lambda \langle b_1 \rangle_\lambda - \langle a_1 \rangle_\lambda \langle b_1 \rangle_\lambda| && \text{(A.8)} \\ &\leq |\langle b_1 \rangle_\lambda - \langle b_0 \rangle_\lambda| + |\langle b_0 \rangle_\lambda \langle b_1 \rangle_\lambda| && \text{(A.9)} \\ &\leq 2|\langle b_0 \rangle_\lambda|, && \text{(A.10)} \end{aligned}$$

yielding in the local case:

$$|S| = |\int_\Lambda S_\lambda d\lambda| \leq 2. \tag{A.11}$$

Quantum predictions The measurement settings $x, y \in \{0, 1\}$ now each denote orthogonal vectors $\vec{x}_0$, $\vec{x}_1$, $\vec{y}_0$ and $\vec{y}_1$ and are associated with measurements $\vec{x}_i \cdot \vec{\sigma}$ and $\vec{y}_j \cdot \vec{sigma}$, where $\vec{\sigma}$ is the Pauly vector.

By taking a $\frac{\pi}{4}$ between $\vec{x}_1$ and $\vec{y}_0$, we get:

$$\begin{aligned} \langle a_0 b_0 \rangle = \langle a_1 b_0 \rangle &= \langle a_0 b_1 \rangle = \frac{\sqrt{2}}{2} && \text{(A.12)} \\ \langle a_1 b_1 \rangle &= -\frac{\sqrt{2}}{2}. && \text{(A.13)} \end{aligned}$$

We therefore have:

$$|S| = |\langle a_0 \rangle \langle b_0 \rangle + \langle a_1 \rangle \langle b_0 \rangle + \langle a_0 \rangle \langle b_1 \rangle - \langle a_1 \rangle \langle b_1 \rangle|, \tag{A.14}$$

and the Bell inequality (5.16) does not hold.

A.3 Lower bound on entropic complexity: detailed proof

This approach was originally used by Stefano Pironio in ref[]. This proof is extracted from ref[]. We consider the Bell inequality $B(\mathbf{p}) \leq B_0$ valid for local protocols and a probability distributions $\mathbf{p}$ with decomposition $\mathbf{p} = \sum_\lambda p(\lambda)\mathbf{p}_\lambda$, where $\mathbf{p}_\lambda$ are distributions simulated by deterministic protocols $\mathcal{P}_\lambda$.

We have:

$$\mathbf{b} \cdot \mathbf{p} - B_0 = \sum_\lambda (\mathbf{b} \cdot \mathbf{p}_\lambda - B_0) \tag{A.15}$$

The distribution $\mathbf{p}_\lambda$ maximizing $\dfrac{\mathbf{b}\cdot\mathbf{p}_\lambda - B_0}{C_H(\mathcal{P}_\lambda)}$ can be isolated:

$$\mathbf{b}\cdot\mathbf{p} = (\mathbf{b}\cdot\mathbf{p}_{\lambda^*} - B_0)p(\lambda^*) + \sum_{\lambda\neq\lambda^*}(\mathbf{b}\cdot\mathbf{p}_\lambda - B_0)p(\lambda), \tag{A.16}$$

giving:

$$p_{\lambda^*} = \frac{\mathbf{b}\cdot\mathbf{p} - B_0}{\mathbf{b}\cdot\mathbf{p}_{\lambda^*} - B_0} + \sum_{\lambda\neq\lambda^*}\frac{\mathbf{b}\cdot\mathbf{p} - B_0}{\mathbf{b}\cdot\mathbf{p}_{\lambda^*} - B_0}p(\lambda). \tag{A.17}$$

The entropic complexity can be written:

$$C_H(\mathbf{p}) = \sum_\lambda p(\lambda)C_H(\mathcal{P}_\lambda). \tag{A.18}$$

Again isolating the terms related to $\mathbf{p}_{\lambda^*}$ we get:

$$C_H(\mathbf{p}) = \left[\frac{\mathbf{b}\cdot\mathbf{p} - B_0}{\mathbf{b}\cdot\mathbf{p}_{\lambda^*} - B_0} + \sum_{\lambda\neq\lambda^*}\frac{\mathbf{b}\cdot\mathbf{p} - B_0}{\mathbf{b}\cdot\mathbf{p}_{\lambda^*} - B_0}p(\lambda)\right]C_H(\mathcal{P}_{\lambda^*}) + \sum_{\lambda\neq\lambda^*}p(\lambda)C_H(\mathcal{P}_\lambda) \tag{A.19}$$

$$= \frac{\mathbf{b}\cdot\mathbf{p} - B_0}{\mathbf{b}\cdot\mathbf{p}_{\lambda^*} - B_0}C_H(\mathcal{P}_{\lambda^*}) + \sum_{\lambda\neq\lambda^*}p(\lambda)\left[C_H(\mathcal{P}_\lambda) - \frac{\mathbf{b}\cdot\mathbf{p}_\lambda - B_0}{\mathbf{b}\cdot\mathbf{p}_{\lambda^*} - B_0}C_H(\mathcal{P}_{\lambda^*})\right] \tag{A.20}$$

Since:

$$C_H(\mathcal{P}_\lambda) \geq \frac{\mathbf{b}\cdot\mathbf{p}_\lambda - B_0}{\mathbf{b}\cdot\mathbf{p}_{\lambda^*} - B_0}C_H(\mathcal{P}_{\lambda^*}) \tag{A.21}$$

we have:

$$C_H(\mathbf{p}) \geq \frac{\mathbf{b}\cdot\mathbf{p} - B_0}{\mathbf{b}\cdot\mathbf{p}_{\lambda^*} - B_0}C_H(\mathcal{P}_{\lambda^*}). \tag{A.22}$$

Appendix B

Multidimensional case

$$\min_{\substack{p(\vec{\lambda}) \\ p(\vec{x},\vec{y},\vec{\lambda})}} \quad I(\Lambda;X) \tag{B.1a}$$

$$\text{subject to} \quad < \vec{x},\vec{y} > + \int_{\mathbb{S}_2} \frac{\rho(\vec{x},\vec{\lambda})}{\rho(\vec{x})} \sigma(< \vec{x},\vec{\lambda} >)\sigma(< \vec{y},\vec{\lambda} >) d\vec{\lambda} = \; 0 \quad \forall \vec{x},\vec{y} \tag{B.1b}$$

$$\tag{B.1c}$$

Lagrangian:

$$\mathcal{L}(\vec{x},\vec{\lambda},\alpha) \;=\; D(\rho(\vec{x},\vec{\lambda})||\rho(\vec{x})\rho(\vec{\lambda})) - \int_{\mathbb{S}_2} \alpha \left[< \vec{x},\vec{y} > + \int_{\mathbb{S}_2} \frac{\rho(\vec{x},\vec{\lambda})}{\rho(\vec{x})} \sigma(< \vec{x},\vec{\lambda} >)\sigma(< \vec{y},\vec{\lambda} >) d\vec{\lambda} \right] d\vec{y} \tag{B.2}$$

$$= \; D(\rho(\vec{x},\vec{\lambda})||\rho(\vec{x})\rho(\vec{\lambda})) - \iint_{\mathbb{S}_2^2} \frac{\rho(\vec{x},\vec{\lambda})}{\rho(\vec{x})} \sigma(< \vec{x},\vec{\lambda} >)\sigma(\vec{y},\vec{\lambda} >) d\vec{\lambda} d\vec{y} - \int_{\mathbb{S}_2} \alpha < \vec{x},\vec{y} > d\vec{y} \tag{B.3}$$

By lemma 4.3:

$$\max_{\alpha} \min_{\substack{\rho(\vec{x},\vec{\lambda}) \\ \rho(\vec{\lambda})}} \mathcal{L}(\vec{x},\vec{\lambda},\alpha) \tag{B.4}$$

$$= \max_{\alpha} \min_{\rho(\vec{x})} - \textstyle\int_{\mathbb{S}_2} \alpha < \vec{x},\vec{y} > d\vec{y} - \log \rho(\vec{x})\rho(\lambda) 2^{\int_{\mathbb{S}_2} \frac{\alpha}{\rho(\vec{x})} \sigma(<\vec{x},\vec{\lambda}>)\sigma(<\vec{y},\vec{\lambda}>) d\vec{y}} d\vec{x} d\vec{\lambda} \tag{B.5}$$