Universitat Autònoma
de Barcelona

BACHELOR'S THESIS

Degree in Physics

Quantum Change-Point Detection in Noisy Channels

Adrià Labay Mora
1420618

Supervisor
Ramón Muñoz Tàpia

Year
2018/19

To my colleagues,

with whom I've pleasurably suffered the last four years

Acknowledgement

I would like to give a special thank to my supervisor Ramon Muñoz, for sharing with me his enthusiasm on the topic and wasting time during lengthy conversations which helped me shed some light on the work.

Thanks to all my colleagues, to whom I dedicate this thesis, for making the last 4 years unforgettable but also for the mutual support given in the process of making our Bachelor Thesis, sharing multidisciplinary talks and methods that certainly made us push forward and learn from each other.

Also, I would like to say that all of it wouldn't have been possible if it wasn't for the open source journals and websites like arxiv.org that provide an extended database of open access scientific papers. But specially, to those altruistic scientist who publish their work for free making it available for a larger audience.

Finally, a special thank to my parents who, despite their ignorance on the topic, have always been there and supported me in some way.

A L M

Abstract

The identification of abrupt changes in the behaviour of a system is of crucial importance in areas such as medicine, climatology, biology... Here we will discuss the problem in the quantum world, we will consider a source generating a certain state which suffers an alteration and suddenly changes to another. The question is to find the position of the change based on measures carried out on the states with the greatest possible probability of success. Recently, the solution to the problem was found when the two states, before and after the change, were pure. In this thesis, we will study the case in which the states are mixed and therefore the uncertainty in the system increases. We will use both analytical and numerical techniques to find the optimal measure that maximises the probability of correctly identifying the point of change.

Resum

La identificació de canvis bruscs en el comportament d'un sistema es presenta de crucial importància en àmbits com la medicina, climatologia, biologia... Aquí discutirem el problema en el món quàntic, considerarem una font generadora d'un tipus d'estat que pateix un error i sobtadament canvia a un altre. La qüestió consisteix en trobar la posició del canvi a partir de mesures realitzades sobre els estats amb la màxima probabilitat d'encertar possible. Recentment, es va trobar la solució pel problema quan els dos estats, abans i després del canvi, eren purs. En aquesta tesi, estudiarem el cas en que els estats són barreja i per tant la incertesa en el sistema augmenta. Utilitzarem tant tècniques analítiques com numèriques per tal de trobar la mesura òptima que maximitza la probabilitat d'encertar el punt de canvi.

Preface

> In all cases, a quantum state is specifically and only a mathematical symbol for capturing a set of beliefs or gambling commitments
>
> — *Quantum Mechanics as Quantum Information*, Fuchs [2002]

Ever since the "discovery" of Quantum Mechanics (QM), the scientist community was shocked by the effects and strange phenomena emerging from this theory. Almost a century has passed and we still struggle to understand the surprising features of this theory, one of the most evolving of our present days, specially in the context of quantum information. This arises from the limits of classical computation that we face today. After all, information is encoded in a physical system, whatever it is and in the form we want, so the study of information and computation should be linked to the study of the underlying physical processes. This point of view is enriched in the well known statement "It from Bit" first pronounced by John Wheeler suggesting "the idea that every item of the physical world has at bottom — at a very deep bottom, in most instances — an immaterial source and explanation; that what we call reality arises in the last analysis from the posing of yes-no questions and the registering of equipment-evoked responses; in short, that all things physical are information-theoretic in origin and this is a participatory universe".

Today, we still don't have a unified view on what is a quantum state and what information does it encode. But this is a question that has been there since its origins, Einstein was one of the first who put into question the completeness of quantum mechanics [Einstein et al., 1935]. John Bell later proved that the hidden variable theory proposed by Einstein was not possible or it would otherwise violate local realism [Bell, 1964]. Nevertheless, even if we now take QM as a complete theory of reality, we haven't been able to discover what a quantum state is. We should content ourselves with its probabilistic nature.

Quantum Mechanics formalism is squeezed into 5 postulates (e.g. see Nielsen and Chuang [2000]) that cover the rules of a very big game. From these postulates, the scientist community has been able to move forward with the development of Quantum Field Theory with all its consequences and the creation of a new field in physics: Quantum Information. Many problems that are encountered in classical computation were brought to the quantum regime like cloning, teleportation, factoring, discrimination... and the main subject of this project, Quantum Change Point. "There is a feeling that the advent of quantum information theory heralds a new way of doing physics and supports the view that information should play a more central role in our world picture" says Fuchs [2002]. This is certainly a reality, for instance the European Union set a flagship in 2016 for the next 10 years with a founding of €1 billion[1] to investigate in the development of certain applications.

We seek to continue this trend and set the basics for future work in the topic, provided with the knowledge gained in the subjects of Quantum Physics and Quantum Information mainly, together with the experience build upon the other courses during the Bachelor in Physics.

[1] `https://qt.eu/about/`

Notation

I will try to use the common notation in quantum information theory and concretely the same as in Nielsen and Chuang [2000].

$\|\psi\rangle$	State vector or "ket" labelled by ψ
$\langle\psi\|$	Dual vector or "bra", the conjugate transpose of $\|\psi\rangle$
$\langle\psi\|\phi\rangle$	Inner product or "braket"
$\|\psi\rangle\langle\phi\|$	Outer product or "dyad"
$\mathcal{H}_n$	n-dimensional Hilbert space
ρ	Density matrix
$\otimes$	Tensor product or Kronecker product
$\oplus$	Orthogonal sum
$\rho^{\otimes n}$	Kronecker product of n times the state ρ
$A, B, \Pi, ...$	Matrix
$\mathbb{I}_n$	$n \times n$ identity matrix
$\mathbb{0}_n$	$n \times n$ zero matrix
$\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$	Vector of Pauli matrices
$A \geq 0$	A is a positive semi-definite matrix
$\|\bullet\|_1$	Trace-norm
$\|\bullet\|_2$	Euclidean norm
$\operatorname{tr} A$	Trace of A
$\hat{P}, \hat{\Pi}$	Projector, $\hat{P}^2 = \hat{P}$
$\mathscr{H}_n$	Space of $n \times n$ hermitian matrices
$\mathcal{E}(\rho)$	Trace preserving quantum operation on the state ρ
$\mathcal{M}$	Generalised measure
$p(j \mid k)$	Conditional probability of finding j given k
P_s / P_e	Success/Error probability
ξ_k	*A priori* probability of the k-th state
$\mathscr{B}(p)$	Bernoulli distribution with parameter $p \in [0, 1]$
$\mathscr{F}$	Space of feasible solutions of a SDP
$\{0,1\}^n$	Space of all possible binary numbers between 0 and $2^n - 1$
$i = (i_0 i_1 \dots i_{n-1})$	n-bit binary number.

Abbreviations

MED Minimum Error Discrimination.

POVM Positive Operator-Valued Measure.

QCP Quantum Change Point.

QM Quantum Mechanics.

QSD Quantum State Discrimination.

RW Random Walk.

SDP Semi-Definite Programming.

USD Unambiguous State Discrimination.

Contents

Abstract **v**

Preface **vii**

Notation **ix**

Abbreviations **xi**

1 Introduction **1**
1.1 Formalism 1

2 Quantum State Discrimination **4**
2.1 Quantum Hypothesis Testing 4
2.2 Optimality conditions 5
2.3 Ambiguous vs. unambiguous state discrimination 6
2.4 Two-state discrimination 7

3 Semi-Definite Programming **9**
3.1 SDP in state discrimination 10

4 Iterative method **12**

5 Quantum Change Point **13**
5.1 Pure vs. Pure 13
5.2 Pure vs. Mixed 14
5.2.1 Along the diameter 14
5.2.2 To any mixed state 15
5.3 Mixed vs. Mixed 19
5.3.1 Along the diameter 19
5.3.2 Equally mixed states 22

6 Conclusions **23**

References **25**

Appendix **27**

A Semi-Definite Programming **27**
A.1 Numerical solvers 28

B Simple Random Walk **29**

C Supplemental calculations for Section 5 **30**
C.1 Pure vs. Mixed 30
C.1.1 Along the diameter 30
C.1.2 To any mixed state 32
C.2 Mixed vs. Mixed 33
C.2.1 Along the diameter 33
C.2.2 Equally mixed states 36

1 Introduction

This thesis aims to study one of the fundamental topics in quantum information which is Quantum State Discrimination (QSD) in the particular case of the Quantum Change Point (QCP). A topic which is widely developed in the classical regime and it is of crucial importance in the detection of abrupt changes. It is used in a wide range of disciplines because of the intrinsic ability to the early warning of small deviations of a system with respect to a reference behaviour considered as normal [Basseville et al., 1993]. For instance, it can be used to make prediction of catastrophic natural phenomena or in the study of climate change [Reeves et al., 2007]. From a mathematical perspective, the change point problems tries to identify times when the probability distribution of a stochastic process or time series changes. The changes may be in the mean, standard deviation, dynamics... with one or multiple change points.

In the classical problem, we are given a series of data $\{x_k\}_{k=1}^n$ where x_k correspond to the result of an observation made at epoch k. The only knowledge on the system resides on the values $\{x_k\}_{k=1}^n$ which are imposed by the results of a measurement. The advantage of QM is that what is given are the states and there is the possibility of choosing the appropriate measure as to increase the probability of guessing right the change point. Yet, the disadvantage is that quantum mechanics doesn't allow to distinguish perfectly non-orthogonal states, it is our job to find how well we can do.

Recently, this problem has been brought to the quantum regime by Sentís, Bagan, Calsamiglia, Chiribella, and Munoz-Tapia [2016]. They studied the single abrupt equally-likely quantum change point problem where the state of a system drastically switches to another. The initial and the final states are known and assumed to be pure, no other information is given apart from the knowledge of the existence of a change point that can happen with equal probability in any location. Under these conditions, an analytical expression for the maximum probability of success is obtained that only depends on the overlap of the two states.

In this thesis, we will study a generalisation to the problem in which the states are no longer pure but mixed, that is, the initial and final states are a combination of pure states. As a consequence, the uncertainty increases, not only we ignore the position of the change point but also the initial and final states are ambiguous.

Section 2 provides the basic theory of states discrimination, Sections 3 and 4 explain the numerical methods that will be used in order to find the best measure and in Section 5 we present solutions for the QCP problem having definite initial and final states.

1.1 Formalism

In this section, I will present the basics of QM formalism that is needed for the project and will skip others aspects, although important are not relevant for the work.[2]

Together with each physical system $\mathcal{S}$ there is an associated d-dimensional complex space $\mathcal{H}_d$, the Hilbert space. The elements of this space constitute the possible states of such system and are represented by a normalised column vector or *ket* $|\psi\rangle$ labelled by the letter ψ, which may be a tuple encapsulating other properties like position, momentum, spin... All the information of that system is self-contained in the state $|\psi\rangle$ but not all the information is retrievable as we will see.

The smallest, non trivial, space with retrievable information is the two-dimensional Hilbert space $\mathcal{H}_2$. The space is spanned by two states labelled $\{|0\rangle, |1\rangle\}$ that form the *computational basis*. A state $|\psi\rangle \in \mathcal{H}_2$ is said to be a *qubit*, in complete analogy to the classical bit of information. A qubit can be written by a complex linear combination or *superposition* of them like

$$|\psi\rangle = \psi_0 |0\rangle + \psi_1 |1\rangle \qquad , \qquad |\psi_0|^2 + |\psi_1|^2 = 1 \tag{1.1}$$

where the components $\psi_j \in \mathbb{C}$ are determined through the inner product of $|\psi\rangle$ and $|j\rangle$. In Dirac notation, the inner product of a Hilbert space is represented by the product of a *bra* and a ket, being the bra the conjugate transpose of the ket, written as $\langle\psi| = |\psi\rangle^\dagger$. Then, the *braket* of two states is

$$\langle\phi|\psi\rangle = \sum_{j=0}^{1} \phi_j^* \psi_j = \langle\psi|\phi\rangle^* \tag{1.2}$$

[2]For a more detailed explanation, see the wonderful book of Nielsen and Chuang [2000] or the fantastic 20-paged summary of the basics of quantum information by two of its fathers, Bennett and Shor [1998].

The state $|\psi\rangle$, under a measurement that distinguishes the states $|0\rangle$ and $|1\rangle$, behaves like $|0\rangle$ with probability $p_0 = |\psi_0|^2$ and like $|1\rangle$ with probability $p_1 = |\psi_1|^2$. The normalisation follows from the conservation of $p_0 + p_1 = 1$.

The study of two of more systems $\mathcal{S}_0, \ldots, \mathcal{S}_{n-1}$ is made through the *tensor product* or *kronecker product* of the corresponding spaces $\mathcal{H}_0 \otimes \cdots \otimes \mathcal{H}_{n-1}$. When all the spaces are two-dimensional we will write $\mathcal{H}_2^{\otimes n}$, the computational basis in this big space of 2^n elements is spanned by the vectors $\{|j\rangle \mid j \in \{0,1\}^n\}$ where $\{0,1\}^n$ is the space of all the combinations of n zeros and ones. Sometimes, the state $|j\rangle$ will also be written as $|j_0 j_1 \ldots j_{n-1}\rangle = |j_0\rangle \otimes |j_1\rangle \otimes \cdots \otimes |j_{n-1}\rangle$ interchangeably, although the first notation is preferable for its simplicity. If at some point there is confusion on which space $|j\rangle$ belongs, a subscript on the ket will be written for clarification $|j\rangle_A$.

A measurement is made by an observable $\mathcal{A}$ that has an associated Hermitian matrix or *operator* $A \in \mathscr{H}_n$ which acts on the quantum states of some Hilbert space. In the computational basis, the operator reads $A = \sum_{j,k=1}^{n} a_{jk} |j\rangle\langle k|$ where $|j\rangle\langle k|$ is the outer product.

The hermiticity property of all observables allows a spectral decomposition as a sum $A = \sum_\lambda a_\lambda \hat{P}_\lambda$ being $\{a_\lambda\}$ the eigenvalues of A and $\hat{P}_\lambda$ the projector onto the eigenspace spanned by the eigenvectors corresponding to a_λ obeying the orthogonality and completeness relations

$$\hat{P}_\lambda \hat{P}_\mu = \hat{P}_\lambda \delta_{\lambda\mu} \tag{1.3a}$$

$$\sum_\lambda \hat{P}_\lambda = \mathbb{I}_d \tag{1.3b}$$

If A is a physical observable, then $\{a_\lambda\}$ are the physical values that we can observe after measuring a state $|\psi\rangle$. The probability that the result a_λ is obtained given that the state measured was $|\psi\rangle$ is

$$p(a_\lambda|\psi) = \langle\psi|\hat{P}_\lambda|\psi\rangle \tag{1.4}$$

and it holds that $\sum_\lambda p(a_\lambda|\psi) = 1$ on account of eq. (1.3b).

The measure is completely defined once the operators $\{\hat{P}_\lambda\}$ are given, then for each $\hat{P}_\lambda$ we associate the hypothesis that the value of the physical property $\mathcal{A}$ observed is a_λ. This measure is called projective or *von Neumann measure* because the elements are orthogonal projectors (eq. (1.3a)) [von Neumann, 1955]. The number of projectors is limited by the dimension of the space, otherwise the orthogonality condition wouldn't be satisfied. For this reason, we define a generalised measurement or Positive Operator-Valued Measure (POVM) as a set of positive operators $\{\Pi_j\}_{j=1}^n$ [Kraus, 1983], with n not necessarily equal to d, satisfying the completeness and positivity conditions

$$\sum_{j=1}^{n} \Pi_j = \mathbb{I}_d \tag{1.5a}$$

$$\Pi_j \geq 0 \qquad \forall j \tag{1.5b}$$

An observable of $\mathcal{H}_2$ is expressed, in the most general form, as a complex linear combination of the identity matrix $\mathbb{I}_2$ and the *Pauli matrices* $\{\sigma_i\}_{i=1}^3$[3] which span the space of 2×2 Hermitian matrices $\mathscr{H}_2$,

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad , \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad , \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{1.6}$$

The elements of the computational basis are by convention the eigenvectors of the third Pauli matrix such that $\sigma_3 |0\rangle = +|0\rangle$ and $\sigma_3 |1\rangle = -|1\rangle$.

Moreover, the 3 Pauli matrices are the generators of rotations in $SU(2)$, the symmetry group of qubits. A rotation of a qubit $|\psi\rangle$ along the direction $\boldsymbol{n}$ by an angle θ is performed by the unitary operator

$$U_{\theta,\boldsymbol{n}} = \exp\left(-i\frac{\theta}{2}\boldsymbol{n}\cdot\boldsymbol{\sigma}\right) \tag{1.7}$$

[3] Sometimes also expressed as $\{\sigma_x, \sigma_y, \sigma_z\}$.

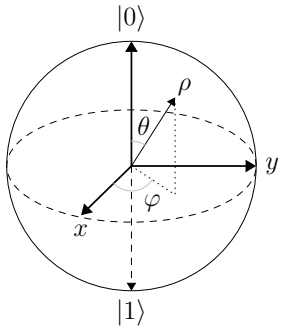


Figure 1: Three dimensional representation of a qubit ρ in the Bloch sphere.

The states we have been talking so far are called *pure states*, they represent situation of perfect knowledge on a system. However, there may be situations in that the state of the system is not well-defined and we have to make use of *mixed states* generated by an *ensemble of pure states* $\Xi = \{\xi_k, |\psi_k\rangle\}_{k=1}^n$ with $\sum_{k=1}^n \xi_k = 1$, meaning that the we can find the state $|\psi_k\rangle$ with probability ξ_k. A mixed states is represented by a positive-semidefinite hermitian *density matrix* ρ with trace equal to one,

$$\rho = \sum_{k=1}^{n} p_k \, |\psi_k\rangle\langle\psi_k| \tag{1.8}$$

The state ρ is called pure when there is only one state in the ensemble with probability 1, then $\rho = |\psi\rangle\langle\psi|$ and we return to the case above. Moreover, the ensemble of states Ξ might contain other mixed states ρ_k occurring with probability ξ_k. In any case, the probability of observing a_λ when measuring A given ρ in eq. (1.8) is

$$p(a_\lambda|\rho) = \mathrm{Tr}\Big(\hat{P}_\lambda \rho\Big) \tag{1.9}$$

which reduces to (1.4) when ρ is pure.

By construction, the density operator (1.8) is also hermitian and admits a spectral decomposition $\rho = \sum_{k=1}^n \xi_k \, |\xi_k\rangle\langle\xi_k|$ with positive eigenvalues ξ_k and corresponding eigenvector $|\xi_k\rangle$ satisfying $\sum_k \xi_k = 1$ and $\sum_k |\xi_k\rangle\langle\xi_k| = \mathbb{I}_d$. They constitute the "eigen-ensemble" $\{\xi_k, |\xi_k\rangle\}$ where the values ξ_k are interpreted as the probability that the system was in the state $|\xi_k\rangle$. An important property follows from this fact, given a density matrix ρ it is not possible to know from which ensemble Ξ it was constructed. Or equivalently, if two ensembles Ξ and Ξ' (with the same or different number of elements) lead to the same density matrix, the two systems $\mathcal{S}$ and $\mathcal{S}'$ are completely indistinguishable.

In two dimensions, a mixed states can be expressed as linear combination of the identity matrix $\mathbb{I}_2$ and the Pauli matrices as

$$\rho = \frac{\mathbb{I}_2 + \boldsymbol{r} \cdot \boldsymbol{\sigma}}{2} \tag{1.10}$$

for some real coefficients $\boldsymbol{r} = (r_1, r_2, r_3)$. The eigenvalues of (1.10) are $(1 \pm \|\boldsymbol{r}\|_2)/2$, but because ρ is positive by definition we must have $\|\boldsymbol{r}\|_2 \leq 1$, with equality in the case of pure states. The value $r = \|\boldsymbol{r}\|_2$ is known as the purity of the state and represents the radial distance from the origin. The product $\boldsymbol{r} \cdot \boldsymbol{\sigma}$ symbolises the pseudo-scalar product between a vector and a vector of matrices: $\boldsymbol{r} \cdot \boldsymbol{\sigma} = \sum_{i=1}^3 r_i \sigma_i$.

From (1.10), it is clear the correspondence between a qubit and a point in a 3-dimensional sphere. The state vector $\boldsymbol{r}$ can always be expressed as $\boldsymbol{r} = r(\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta)$ with $r \in [0, 1]$ the modulus, $\theta \in [0, \pi)$ the angle with respect the z-axis and $\varphi \in [0, 2\pi)$ the angle with respect to the x-axis. The sphere where qubits are represented is called the *Bloch sphere*, see fig. 1, the peculiarity is that the angle between orthogonal states is π in contrast to the usual $\pi/2$ of the usual Euclidean space[4].

[4]The reason for this is a factor of two between the symmetry groups $SU(2)$ (where qubits live) and $SO(3)$ (the usual rotation group), in other words, to return to the same state one must make a rotation of 4π degrees.

2 Quantum State Discrimination

Fundamental properties of quantum mechanics make it impossible to perfectly distinguish non-orthogonal quantum states. Note that, if state discrimination was perfect, it would imply that quantum cloning could be done perfectly or that quantum entanglement would lead to instantaneous communication [Gisin, 1998]. For instance, the BB84 quantum key distribution scheme [Bennett and Brassard, 2014] is based in sending photons polarised in two non-orthogonal basis, usually the Z and X basis[5]. Because of the non-orthogonality and the impossibility to distinguish perfectly quantum states, an eavesdropper that intercepts the message without any knowledge on the basis the photons were encoded on, won't be able to read the full sentence. Indeed, if she receives a photon polarised in the X direction but she measures, unconsciously, in the Z direction there will be a 50% chance of mistake. Ultimately, one is forced to make a guess and it is the necessity of this guess that makes quantum mechanics intrinsically indeterministic.

Another example is quantum cloning, if that could be done perfectly then we would be able to generate n copies of two non-orthogonal quantum states $|\psi\rangle$ and $|\phi\rangle$. Since they are not orthogonal, we can't perfectly discriminate a single pair of them. However, if n copies are considered, then the overlap of the composite system goes as $|\langle\psi|\phi\rangle|^n$ which tends to 0 as the number of copies increases. Therefore, because a general quantum state cannot be cloned, state discrimination cannot be done perfectly.

The question now is, how can we *best* discriminate different quantum states? We can't certainly predict the result of a measurement, however, the foundations of quantum mechanics gives us with accuracy the probabilities of those outcomes. These follow some classical probability distribution and with the help of classical information theory we could find ways to distinguish them. The idea is to vary over the measurements that we make on a system to find the one that makes the classical distinguishability the best it can be [Fuchs, 1996].

First of all, it is not possible to go to the Hilbert space, put a ruler between quantum states and decide from this whether they are the same or not, just because a posterior measurement might change its nature. In any case, we can define a pseudo-distance between two general states ρ and ρ' as

$$D(\rho, \rho') = \|\rho - \rho'\|_1 \tag{2.1}$$

which is the so called *trace distance* [Nielsen and Chuang, 2000], denoting by $\|A\|_1$ the *trace norm* (or norm one)

$$\|A\|_1 = \operatorname{tr}\sqrt{AA^\dagger} = \sum_\lambda |a_\lambda| \tag{2.2}$$

where $\{a_\lambda\}$ are the eigenvalues of A.

Two quantum states are said to be close to each other if the trace distance is near zero. If the states are qubits, with state vector $\boldsymbol{r}$ and $\boldsymbol{r}'$ respectively, the expression (2.1) reduces to

$$D(\rho_{\boldsymbol{r}}, \rho_{\boldsymbol{r}'}) = \frac{\|\boldsymbol{r} - \boldsymbol{r}'\|_2}{2} \tag{2.3}$$

where $\|\boldsymbol{a}\|_2 = \sum_k |a_k|^2$ is the usual vector norm (or norm two). Notice that this pseudo-distance is half the ordinary distance between two points inside a sphere.

2.1 Quantum Hypothesis Testing

Consider that Alice prepares one state of some ensemble $\Xi = \{\xi_k, \rho_k\}_{k=1}^n$, all living in a d-dimensional Hilbert space $\mathcal{H}_d$. The probability that Alice chooses ρ_k is ξ_k, with $\sum_k \xi_k = 1$. After that, this state is send to Bob who is asked to distinguish it among the states inside the set Ξ. In Bob's hands, the system is described by the mixed state

$$\rho = \sum_{k=1}^n \xi_k \rho_k \tag{2.4}$$

[5]Defined as the eigenvalues of the Pauli matrices σ_z and σ_x respectively.

Bob can perform any measurement on the state, the most general form of such measurement is a POVM measure $\mathcal{M} = \{\Pi_j\}_{j=1}^m$ satisfying eqs. (1.5a) and (1.5b). Note that, m is not in general equal to n (the number of states), but it can be greater or smaller. This number is related to the number of hypothesis that can be made. For instance, if $m > n$ then we can assign to a combination of multiple outcomes of our measurement the same $\rho_k \in \Xi$. On the other hand, if $m < n$ then there will be some states for which we will have to make a guess, unless we know that they occur with 0 probability.

If the Π_j are orthogonal projectors ($\Pi_i\Pi_j = \delta_{ij}\Pi_i$), then $\mathcal{M}$ is a von Neumann measure and $m \leq n$, but they do not have to be. As an example, the operators $\Pi_j = \mathbb{I}/d$ associated to the no-measurement strategy are clearly not projectors.

Actually, it is found by Davies [1978] that the number of POVM elements m needed to distinguish n pure states is bounded between $n \leq m \leq n^2$ for linearly independent states. The number of POVM operators can be any inside this range but the process becomes an arduous task if the optimisation needs to be made also on the number of hypothesis. For simplicity, in our problem, we will fix $m = n$, i.e. the number of hypothesis is the same as the number of states, where j is the proposition that the state was ρ_j. We can do this because, even if m was greater than the number of states, we could group the operators from our hypothesis to build only n operators verifying eqs. (1.5a) and (1.5b).

On account of eq. (1.9), the probability of outcome j (Π_j) conditional that the given state was ρ_k is

$$p(j \mid k) = p(\mathcal{M} = \Pi_j \mid \Xi = \rho_k) = \operatorname{tr}(\Pi_j \rho_k) \tag{2.5}$$

Therefore, the state k will be successfully identified whenever the hypothesis Π_k is selected which happens with probability $p(k \mid k)$. Putting all together, it follows that the probability of correctly guessing the state is

$$P_s = \sum_{k=1}^{n} \xi_k \operatorname{tr}(\Pi_k \rho_k) \tag{2.6}$$

and because the states Ξ are not mutually orthogonal, there will be non-zero probability of failure (measure of an incorrect state): $0 \leq P_s \leq 1$. The expression for the error probability is just $P_e = 1 - P_s$.

2.2 Optimality conditions

In general, a measure $\mathcal{M}$ will give us some success probability (2.6) which will be suboptimal. We seek to find the POVM that maximises the success probability. It has been found by Holevo [1973] that the optimal operators must satisfy the conditions

$$\Pi_j(\xi_j\rho_j - \xi_k\rho_k)\Pi_k = 0 \qquad \forall j,k = 1,\dots,m \tag{2.7}$$

$$\Gamma - \xi_k\rho_k \geq 0 \qquad \forall k = 1,\dots,N \tag{2.8}$$

with the definition of the so called Lagrange operator

$$\Gamma = \sum_{k=1}^{n} \xi_k \Pi_k \rho_k \tag{2.9}$$

which places the role of a Lagrange multiplier taking account of the constraint (1.5a). It can be shown from the first condition that the Lagrange operator is hermitian. Take the sum over j and k in eq. (2.7), because $\Pi_j = \Pi_j^\dagger$ and $\rho_k = \rho_k^\dagger$, we are left with $\Gamma^\dagger - \Gamma = 0$ proving the hermiticity of the Lagrange operator. Indeed, eqs. (2.7) and (2.8) are not independent but the first can be derived from the second.

In fact, the first condition (2.7) can also be written, by summing over j, in terms of the Lagrange operator Γ as

$$(\Gamma - \xi_k\rho_k)\Pi_k = 0 \qquad \forall k = 1,\dots,m \tag{2.10}$$

which gives us a way to determine the operators Π_k once Γ is known. Indeed, both Π_k and $\Gamma - \xi_k\rho_k$ are positive operators, and thus eq. (2.10) can hold only if they are orthogonal, that is Π_k lays entirely within the kernel of $\Gamma - \xi_k\rho_k$ [Weir et al., 2017].

Equation (2.8) gives a necessary and sufficient condition for an optimal measurement, while eq. (2.7) gives only a necessary condition. In our posterior work, we will seek to find such measurement whose Lagrange operator Γ (2.9) is hermitian and for all the initial states in the ensemble Ξ we have that the second Holevo condition is verified. As we will see in Section 3, the Holevo condition eq. (2.8) is the same as the optimality conditions imposed by Semi-Definite Programming.

2.3 Ambiguous vs. unambiguous state discrimination

In the problem of quantum state discrimination, there are two major techniques: Minimum Error Discrimination (MED) and Unambiguous State Discrimination (USD). The former approach, also named ambiguous state discrimination, consist on minimising the probability of guessing a wrong result P_e, which can sometimes be achieved by not making any measurement at all and randomly guessing the result. In contrast, the latter has no error, if hypothesis Π_j is obtained we are 100% sure of that the state was ρ_j, yet we allow the possibility of an inconclusive result by introducing an extra operator $\Pi_?$. The two tasks are equally valid, the use of one or another only depends on the requirements of the problem. For example, in situations where we can't be wrong we should use USD instead of MED. In fact, there is a correspondence between both as it is possible to take a MED to a USD [Bagan et al., 2012].

In this thesis, the method used will be MED for a simple reason. Unambiguous state discrimination forces the operators to satify $\operatorname{tr}(\Pi_j \rho_k) = 0$ if $j \neq k$, this is not possible in general since both Π_j and ρ_k are positive operators. Only when the states $\{\rho_k\}_{k=1}^n$ have disjoint kernels, $\ker(\rho_k) \cap \ker(\rho_l) = \emptyset \ \forall k \neq l$, USD would be possible [Raynal, 2006; Rudolph et al., 2003] which is not the case in the QCP problem. Therefore, in what follows, we will be working in the context of MED which now proceed to explain in more detail.

In Section 2.1, the form of the success probability was deduced. It follows that the probability of error is $P_e = 1 - P_s$, so finding the minimum P_e is the same as maximising the success probability as a function of the measure,

$$P_s = \max_{\mathcal{M}} \sum_{k=1}^{n} \xi_k \operatorname{tr}(\Pi_k \rho_k) \tag{2.11}$$

under the conditions eqs. (1.5a) and (1.5b). Putting the sum inside the trace, we identify the operator to be maximised as Γ and from eq. (2.8) we can rewrite the problem as that of finding

$$P_s = \min_{\Gamma} \operatorname{tr} \Gamma \tag{2.12}$$

subject to the constraints $\Gamma - \xi_k \rho_k \geq 0$.

The meaning of this is that, for an arbitrary set of positive observables $\{\Pi_j\}$ that add up to the identity, we can construct the corresponding Lagrange operator. However, only the one which is optimal according to the relation eq. (2.8) will give the maximum probability. Even though two measures $\mathcal{M}$ and $\mathcal{M}'$, with Γ and Γ' respectively, are found to be optimal, the success probability will still be the same [Helstrom, 1969].

Equation (2.11) may look like a *tour de force* to the reader, having to maximise over all the possible measures. It happens that this is as complicated as it seems, very few analytical solutions are found while most of the results in quantum discrimination problems are found using numerical methods which will be described in section 3. The analytical solutions are only well known for the case of discrimination between two states or for geometrically uniform states. For example, the case of two states was first found by Helstrom who provided an exact value for the success probability [Helstrom, 1969] which we will reproduce in the following section. Then, Bae and Kwek [2015]; Barnett [2001] showed a minimum-error discrimination strategy between multiply symmetric states with a deeper study of the so called three mirror-symmetric states [Andersson et al., 2002; Chou, 2004; Ha and Kwon, 2013]. For a general number of states, there are unambiguous strategies found by Chefles and Barnett [1998] when the states are linearly independent and for minimum error discrimination, it is found that the discrimination between n qubit states can be divided into patches of only 4 qubits with a known optimal solution [Weir et al., 2017]. Also, Deconinck and Terhal [2010] provide a geometrical representation of the optimal measure in the Bloch sphere.

2.4 Two-state discrimination

It is instructive to work out the solution to the simplest problem in QSD following the process explained previously. We will evaluate the maximum success probability for the case of two general states $\Xi = \{(\xi_1, \rho_1), (\xi_2, \rho_2)\}$ and then give some simplified versions for when the states are pure, qubits[6]... The measure is made up of only two positive operators $\{\Pi_1, \Pi_2\}$ that satisfy $\Pi_1 + \Pi_2 = \mathbb{I}$. The maximum guess probability is given by eq. (2.12) where the Lagrange operator is

$$\Gamma = \xi_1 \Pi_1 \rho_1 + \xi_2 \Pi_2 \rho_2$$

but using the completeness relation, the dependence in one of the operators can be removed. Write

$$\begin{cases} \Gamma_+ &= \xi_2 \rho_2 + \Pi_1 X \\ \Gamma_- &= \xi_1 \rho_1 - \Pi_2 X \end{cases}$$

defining

$$X = \xi_1 \rho_1 - \xi_2 \rho_2 \tag{2.13}$$

Although the process can be done with Γ_+ or Γ_-, it is convenient to symmetrise those expressions and write the Lagrange operator for the problem as

$$\Gamma = \frac{1}{2}(\Gamma_+ + \Gamma_-) = \frac{1}{2}(\rho + \Lambda X) \tag{2.14}$$

where $\rho = \xi_1 \rho_1 + \xi_2 \rho_2$ and $\Lambda = \Pi_1 - \Pi_2$. The original POVM operators are related to Λ by $\Pi_1 = (\mathbb{I} + \Pi)/2$ and $\Pi_2 = (\mathbb{I} - \Pi)/2$; while $\Pi_1, \Pi_2 \geq 0$ the condition over Λ is that $-\mathbb{I} \leq \Lambda \leq \mathbb{I}$.

Putting all together, the success probability becomes

$$P_s = \max_{\Pi} \operatorname{tr} \Gamma = \frac{1}{2}\left(1 + \max_{\Lambda} \operatorname{tr} \Lambda X\right) \tag{2.15}$$

From the definition of X, because ρ_1 and ρ_2 are positive, its eigenvalues can be divided into positive and negative parts. Denoting by X_+ (X_-) the subspace spanned by the eigenspace of positive (negative) eigenvalues and λ_+ (λ_-, in absolute value) its sum, by the spectral theorem X reads $X = \lambda_+ X_+ - \lambda_- X_-$ where $X_+, X_- \geq 0$. Thus, the optimal measurement Λ is the one that projects the positive subspace to itself and flips the sign of the negative part, i.e. $\Lambda = X_+ - X_-$. Finally, the success probability is [Bae and Kwek, 2015]

$$P_s = \frac{1}{2}(1 + \lambda_+ + \lambda_-) = \frac{1}{2} + \frac{1}{2}\|X\|_1 \tag{2.16}$$

and the POVM consists on

$$\mathcal{M} = \{\Pi_1 = X_+,\ \Pi_2 = X_-\} \tag{2.17}$$

where we have used that $X_+ + X_- = \mathbb{I}$. Equation (2.16) is known as the Helstrom bound and establishes the best success probability to discriminate two mixed states [Helstrom, 1969] which depends only on the trace distance between the two.

It is easily checked that this measure is indeed optimal by constructing the Lagrange operator from eq. (2.14) using the measure found in eq. (2.17), it follows that

$$\Gamma = \frac{1}{2}[\rho + (X_+ - X_-)X] = \frac{1}{2}\rho + \frac{1}{2}(\lambda_+ X_+ + \lambda_- X_-) \tag{2.18}$$

Then, for the two states in Ξ the Holevo condition reads

$$\Gamma - \xi_1 \rho_1 = \frac{1}{2}(-\xi_1 \rho_1 + \xi_2 \rho_2) + \frac{1}{2}\Lambda X = -\frac{1}{2}X + \frac{1}{2}\Lambda X = \lambda_- X_- \geq 0$$
$$\Gamma - \xi_2 \rho_2 = \frac{1}{2}(\xi_1 \rho_1 - \xi_2 \rho_2) + \frac{1}{2}\Lambda X = \frac{1}{2}X + \frac{1}{2}\Lambda X = \lambda_+ X_+ \geq 0$$

Since the Holevo conditions are satisfied, we can be sure that the measure (2.17) is optimal.

[6]The following is not restricted to two dimensional spaces but is general to any two level system in a $\mathcal{H}_d$.

We should be careful with the previous result (2.17) since there may be cases where all eigenvalues are positive or negative if the *a priori* probabilities are different. Then, one of the eigenspaces will be the full space, in fact, it will correspond to the hypothesis of the state with maximum probability. The result is telling us not to waste any effort at all in measuring because we have enough information beforehand to achieve the maximum success probability by just guessing the state with maximum probability.

Of course, eq. (2.16) is much simplified when specific cases are considered. For example, if the states have *a priori* equal probabilities $\xi_1 = \xi_2 = 1/2$, the success probability is

$$P_s = \frac{1}{2} + \frac{1}{4}\|\rho_1 - \rho_2\|_1 \tag{2.19}$$

For qubits with state vector $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ respectively

$$P_s = \frac{1}{2} + \frac{1}{4}|p_1 - p_2 + \|\xi_1\boldsymbol{r}_1 - \xi_2\boldsymbol{r}_2\|_2| + \frac{1}{4}|p_1 - p_2 - \|\xi_1\boldsymbol{r}_1 - \xi_2\boldsymbol{r}_2\|_2| \tag{2.20}$$

which, for the case of equal *a priori* probabilities, reduces to

$$P_s = \frac{1}{2} + \frac{1}{2}\|\boldsymbol{r}_1 - \boldsymbol{r}_2\|_2 \tag{2.21}$$

Finally, if ρ_1 and ρ_2 are pure states $|\psi_1\rangle$ and $|\psi_2\rangle$ then [Barnett and Croke, 2009]

$$P_s = \frac{1}{2} + \frac{1}{2}\sqrt{1 - 4\xi_1\xi_2|\langle\psi_1|\psi_2\rangle|^2} \tag{2.22}$$

All of the above expressions contain a constant term, which doesn't depend at all of the states, and another that depends on the difference between them. Thus, whenever they are the same, i.e. X is the 0 matrix, the probability of success will be just $1/2$ which is to just pick one of the two possible hypothesis at random. In any other situation, the probability will increase, up to its maximum value.

3 Semi-Definite Programming

Semi-Definite Programming (SDP) is a mathematical technique used to solve linear programming problems, that is the maximisation or minimisation of a linear convex function $f(x)$ with linear constraints over the set of all $x \geq 0$ [Wolkowicz et al., 2012]. It can be used to solve a problem analytically but it is mostly used in numerical analysis as solutions can be found efficiently. The utility of SDP is that a lot of problems in quantum information theory (and many other) can be cast into the form of a standard SDP which is then solved to the consumer's pleasure.

Let's start with some definitions.

Definition 1. *A set $\mathscr{S}$ is convex if for $X, Y \in \mathscr{S}$ then $\lambda X + (1-\lambda)Y \in \mathscr{S}$ with $\lambda \in [0,1]$.*

Similarly, a function $f(X)$ is said to be convex if $f(\lambda X + (1-\lambda)Y) = \lambda f(X) + (1-\lambda)f(Y)$ for all $X, Y \in \mathscr{S}$.

Definition 2. *A map Φ is hermiticity preserving if $\Phi(X) \in \mathscr{H}(\mathcal{Y})$ for all $X \in \mathscr{H}(\mathcal{X})$.*

As said, a linear programming problem consist on the maximisation of a convex linear function $f(x)$ called the *goal function*. This, in matrix form, can always be written as the inner product of two matrices: $f(X) = \operatorname{tr} AX$; where $X \geq 0$ and $A \in \mathscr{H}(\mathcal{X})$.

Definition 3. *A semi-definite program is a triplet (Φ, A, B), where*

1. *$\Phi[\bullet] : \mathcal{X} \mapsto \mathcal{Y}$ is a hermiticity preserving convex linear map.*
2. *$A \in \mathscr{H}(\mathcal{X})$ and $B \in \mathscr{H}(\mathcal{Y})$ are hermitian operators.*

for some complex Euclidean spaces $\mathcal{X} \subset \mathbb{C}^{m\times m}$ and $\mathcal{Y} \subset \mathbb{C}^{n\times n}$.

The standard or *primal* problem associated with the triplet (Φ, A, B) reads

$$\begin{aligned} \max_X \ & \operatorname{tr} AX \\ & \Phi[X] = B \\ & X \geq 0 \end{aligned} \tag{3.1}$$

Our goal is to find the X that maximises the goal function among the X that satisfy the constraints. Thus, we define the set of feasible solutions of the primal problem as

$$\mathscr{F}_P(\mathcal{X}) = \{X \in \mathcal{X} \mid X \geq 0 \text{ and } \Phi[X] = B\} \tag{3.2}$$

Note that this is a convex set, any feasible solution may be obtained as a convex combination of two other. The optimal solution is the $X \in \mathscr{F}_P(\mathcal{X})$ for which the objective function achieves a maximum. The optimal primal value of the SDP is defined as

$$\alpha = \sup_{X \in \mathscr{F}_P(\mathcal{X})} \operatorname{tr} AX \tag{3.3}$$

It may happen that there is no supremum when $\mathscr{F}_P(\mathcal{X}) = \emptyset$, in such a case, $\alpha = -\infty$.

Together with the primal problem there exist a *dual* problem which reads

$$\begin{aligned} \min_Y \ & \operatorname{tr} YB \\ & \Phi^*[Y] - A \geq 0 \\ & Y \in \mathscr{H}(\mathcal{Y}) \end{aligned} \tag{3.4}$$

where $\Phi^*[\bullet]$ is the dual map defined as

$$\operatorname{tr}(\Phi[X]Y) = \operatorname{tr}(X\Phi^*[Y]) \tag{3.5}$$

In practice, this relation implies that if $\Phi[\bullet] : \mathbb{C}^{m\times m} \to \mathbb{C}^{n\times n}$ then $\Phi^*[\bullet] : \mathbb{C}^{n\times n} \to \mathbb{C}^{m\times m}$.

We can also construct the convex set of feasible dual solutions as

$$\mathscr{F}_D(\mathcal{Y}) = \{Y \in \mathcal{Y} \mid Y = Y^\dagger \text{ and } \Phi^*[Y] \geq A\} \tag{3.6}$$

and the optimal dual value as

$$\beta = \inf_{Y \in \mathscr{F}_D(\mathcal{Y})} \operatorname{tr} YB \tag{3.7}$$

In the case there is no feasible solution, $\mathscr{F}_D(\mathcal{Y})$ is empty, the optimal value is defined to be $\beta = \infty$.

The dual problem provides us with a way to check if a solution to the primal is optimal.

Theorem 1 (Slater)**.** *For every semi-definite program (Φ, A, B) it is true that:*

1. *If $\mathscr{F}_P \neq \emptyset$ and there exist an hermitian operator Y for which $\Phi^*[Y] \geq A$, then $\alpha = \beta$ and there exist a primal feasible solution $X \in \mathscr{F}_P$ for which $\operatorname{tr} AX = \alpha$.*
2. *If $\mathscr{F}_D \neq \emptyset$ and there exist an positive semi-definite operator X for which $\Phi[X] = B$, then $\alpha = \beta$ and there exist a dual feasible solution $Y \in \mathscr{F}_D$ for which $\operatorname{tr} YB = \beta$.*

The proof of this theorem is not difficult but it is needed to introduce the concept of hyperplanes and closed sets which are out of the scope of this thesis. For a complete proof see the original paper by Slater [2014] or, for a simplified and user friendly version, the lecture notes by Watrous [2011].

The important corollary of this theorem is that we have a way of checking if $X \in \mathscr{F}_P$ is indeed the operator that maximises the objective function, just by checking that its dual Y satisfies $\Phi^*[Y] \geq A$.

3.1 SDP in state discrimination

Now that we know the basics of SDP and its formulation, it is time to start applying this knowledge to the problem of state discrimination. In Section 2, we discussed the form of a state discrimination problem which involved the maximisation of the success probability eq. (2.6) for all valid measures $\mathcal{M}$ satisfying eqs. (1.5a) and (1.5b). Without effort, we can write the SDP primal problem

$$\begin{aligned} \max_{\{\Pi_j\}} \ & \sum_{k=1}^{n} \xi_k \operatorname{tr}(\Pi_k \rho_k) \\ & \sum_{j=1}^{n} \Pi_j = \mathbb{I} \\ & \Pi_j \geq 0 \end{aligned} \tag{3.8}$$

However, this does not have exactly the standard form as presented in eq. (3.1) but it is easily recovered. Define the following quantities

$$A \equiv \bigoplus_{k=1}^{n} \xi_k \rho_k \tag{3.9}$$

$$X \equiv \bigoplus_{j=1}^{n} \Pi_j \tag{3.10}$$

$$B \equiv \mathbb{I}_{2^n} \tag{3.11}$$

and the map function

$$\Phi[X] = \Phi\left[\oplus_{j=1}^{n} \Pi_j\right] = \sum_{j=1}^{n} \Pi_j \tag{3.12}$$

Then, the triplet (A, B, Φ), defined as in eqs. (3.9), (3.11) and (3.12) respectively, form the general SDP problem in quantum state discrimination. Of course, in concrete examples, we will see that it can sometimes be reduced to smaller matrices to reduce computational time.

The dual of this problem is a bit tricky. We first note that the new variable Y must live in a space $\mathbb{C}^{2^n \times 2^n}$ while $X \in \mathbb{C}^{n2^n \times n2^n}$. Having this set, we see that from a $2^n \times 2^n$ matrix we have to write a $n2^n \times n2^n$ matrix which satisfies eq. (3.5). The simplest possibility is the dual map

$$\Phi^*[Y] = \bigoplus_{j=1}^{m} Y = \widetilde{Y} \tag{3.13}$$

Then, the dual problem reads

$$\begin{aligned} &\min_{Y} \ \operatorname{tr} \Gamma \\ &\quad \widetilde{Y} - A \geq 0 \\ &\quad Y = Y^\dagger \end{aligned} \tag{3.14}$$

It is of great importance to notice that eq. (3.14) is equivalent to the Holevo condition (2.8) for an optimal measurement in the state discrimination problem. From the block form of the matrices $\widetilde{Y}$ and A, $\widetilde{Y} - A \geq 0$ is the same as comparing the block matrices one by one, i.e. $Y - \xi_k \rho_k \geq 0 \ \forall k = 1, \ldots, n$. In fact, the space of all feasible dual solutions $\mathscr{F}_D$ is nothing else than the space of all Lagrange operators for a given measure, then the optimal Y is just the Lagrange operator Γ generated by the optimal measures. Indeed, whenever a solution is obtained using SDP, we are certain that this is optimal as it automatically feasible.

To summarise, we have two ways to calculate the optimal measure. Either by solving the primal problem which gives the optimal measures in block form or via the dual problem which returns the optimal Lagrange operator from which we can recover the observables Π_j using eq. (2.10). However, by virtue of Theorem 1, the two solutions are linked and provides a way to check for the optimality. Indeed, if the operators $\{\Pi_j\}$ have been found using the primal problem, by constructing the associated Lagrange operator from eq. (2.9) and checking that $\widetilde{\Gamma} - A \geq 0$ implies that $\{\Pi_j\}$ is optimal and the success probability is $P_s = \alpha = \beta$. The inverse process can also be done theoretically, although it might be difficult in some cases to recover the observables Π_j from Γ but, if this is possible, then checking that $\sum_{j=1}^{n} \Pi_j = \mathbb{I}$ proves the optimality of the measure and that $P_s = \alpha = \beta$.

4 Iterative method

SDP provides a robust and efficient way for solving many problems in a wider range of disciplines. In particular, this method is guaranteed to converge to the real solution. However, in the QCP discrimination problem, the number of variables increases as $\mathcal{O}(n2^n)$ for the primal and $\mathcal{O}(2^n)$ for the dual producing a numerical overhead that makes it impossible to solve this kind of problems for more than a few qubits. For instance, for 4 qubits it requires to solve around 600 variables. One could use the symmetry of ρ_k to reduce the number of variables, but in any case the computational cost would still be prohibitive. Instead, we can rely on an efficient iterative method. For low n, we will still use SDP to benchmark the iterative algorithm, which in contrast to the SDP, it is not rigorously guaranteed to yield the exact solution.

The method was proposed by Ježek et al. [2002] and consists in iterating the matrix Γ (see eq. (2.9)) from an initial choice of the operators Π_j. The advantage over SDP is that it doesn't have to solve for unknown variables but only involves elemental operations between matrices which are efficiently implemented in all programming languages. Also, at each step, the POVM conditions eqs. (1.5a) and (1.5b) are automatically satisfied.

Consider the first guess for the measure $\mathcal{M}^{(0)} = \{\Pi_j^{(0)}\}_{j=1}^n$, the easiest choice is the no-measurement with $\Pi_j = \mathbb{I}/n$. With them, the first correction to the Lagrange operators is evaluated using the relation

$$\Gamma^{(1)} = \left[\sum_{j=0}^{n} \xi_j^2 \rho_j \Pi_j^{(0)} \rho_j\right]^{1/2} \tag{4.1}$$

and with it, the first correction to the operators

$$\Pi_j^{(1)} = \xi_j^2 [\Gamma^{(1)}]^{-1} \rho_j \Pi_j^{(0)} \rho_j [\Gamma^{(1)}]^{-1} \tag{4.2}$$

with $\{\Pi_j^{(1)}\}$ still satisfying the completeness relation. This can be seen by taking the sum over j on the previous equation and substituting by the first correction to the Lagrange operator $\Gamma^{(1)}$.

In general, the k-th approximation of the Lagrange operator and the measure is given by

$$\Gamma^{(k+1)} = \left[\sum_{j=0}^{n} \xi_j^2 \rho_j \Pi_j^{(k)} \rho_j\right]^{1/2} \tag{4.3a}$$

$$\Pi_j^{(k+1)} = \xi_j^2 [\Gamma^{(k+1)}]^{-1} \rho_j \Pi_j^{(k)} \rho_j [\Gamma^{(k+1)}]^{-1} \tag{4.3b}$$

After some iterations (n), the solutions should tend to a stationary point (within some interval $\pm\epsilon$, $\epsilon > 0$ small) for which the last values of $\Gamma^{(n+1)}$ and $\{\Pi_j^{(n+1)}\}$ will correspond to the extreme. Nevertheless, convergence is not proven for this method, therefore we should check for the optimality condition (eq. (2.8)) after a stationary solution has been found in order to find out whether the extreme corresponds to the maximum. If no stationary point is reached or the final values are not optimal we should try again with a different initial guess for the measure $\mathcal{M}^{(0)}$.

The program will be executed until an accuracy of $\epsilon \approx 10^{-6}$ is reached in the probability. The operators, however, may have a worst accuracy than ϵ, this will be seen while checking for the optimality of the resulting measure, up to which level eq. (2.8) is satisfied.

To check the optimality, consider that the program outputs the set of operators $\{\widetilde{\Pi}_j\}$, the numerical test will consist on computing the minimum eigenvalue $\lambda_k = \min_\lambda \Gamma - \xi_k \rho_k$ for each $\rho_k \in \Xi$. Only when $\lambda_k(+\epsilon) \geq 0 \ \forall k$ the solution will satisfy (2.8) and it will be taken as optimal. There is no need to check for Γ being hermitian as it is by construction.

5 Quantum Change Point

The case in study is the single equiprobable QCP in a sequence of n non-pure quantum states. The initial state is characterised by the density matrix ρ which changes abruptly to $\widetilde{\rho}$ at some point in the n possible positions.

The states are sent by Alice one at each time and Bob may choose a measurement to perform on the states, either local (online) or global (offline). In a local strategy, Bob measures the qubits one at each time applying a POVM which may depend on the outcome of the previous measurements. Whereas, in a global strategy, the states are stored in a quantum memory and a measurement is performed on the composite system of n qubits. If the states are pure, it is found that a global measurement outperforms any online measurement [Sentís et al., 2016]. Yet, it is still not clear whether this is true for mixed states. In any case, an online measurement can be of high utility and easier to implement in reality as it can give you the maximum likelihood position on the go and there is no need to store the states. We will for now leave this discussion apart, until the expression for the POVM elements are obtained, since the methods are the same for both.

Bob's knowledge resides uniquely on the states ρ and $\widetilde{\rho}$ and the guarantee that there is a change point. Effectively, he sees an ensemble of states $\Xi = \{\rho_k\}_{k=0}^{n-1}$ each happening with the same probability $\xi_k = 1/n$ for $k = 0, \ldots, n-1$, , where n counts the number of qubits send by Alice. The composite state of a system with a change point in k is expressed as

$$\rho_k = \rho^{\otimes k} \otimes \widetilde{\rho}^{\otimes(n-k)} \quad \in \quad (\mathcal{H}_2)^{\otimes n} \tag{5.1}$$

The proposition of a change point excludes the state $\rho_n = |0\rangle\langle 0|^{\otimes n}$ from the ensemble, which tells us that the change hasn't occurred yet. However, all the states $\rho_0, \ldots, \rho_{n-1}$ have $\widetilde{\rho}$ as the last qubit in the sequence, providing us with no extra information. Consequently, we can safely remove it to reduce the length of the sequence to $n-1$ qubits and work with an ensemble $\Xi = \{\rho_k\}_{k=0}^{n}$ of $n+1$ states by dropping the proposition of a change point. In the succeeding, we will work with n qubits and $n+1$ states ρ_k $k = 0, \ldots, n$ that can happen with *a priori* probability $\xi_k = 1/(n+1)$.

Bob is allowed to perform any measure on the whole system in order to determine k^* (the change point position) with highest probability of success or, in other words, Bob has to distinguish ρ_{k^*} from the others in Ξ with maximum success probability. We know from Section 2 that the success probability, in the SDP formalism of Section 3.1, is

$$\begin{aligned} P_s = \max_{\{\Pi_j\}} \; & \frac{1}{n+1} \sum_{k=0}^{n} \mathrm{tr}(\Pi_k \rho_k) \\ & \sum_{j=0}^{n} \Pi_j = \mathbb{I} \\ & \Pi_j \geq 0 \end{aligned} \tag{5.2}$$

In the following sections we will discuss the solution of some SDP problems for specific ρ and $\widetilde{\rho}$ analytically, if possible, but mainly we will perform a numerical analysis. Although the goal was to find a general method to distinguish the change for two general qubits, this was found to be beyond the scope of this thesis. However, we have found interesting results considering the case of a pure state going to different mixed states inside the Bloch sphere and between two equally mixed states.

5.1 Pure vs. Pure

The already solved problem considers the change from a pure state to another pure state

$$|0\rangle \longrightarrow \quad |\phi\rangle = \cos\frac{\theta}{2}|0\rangle + \sin\frac{\theta}{2}|1\rangle \tag{5.3}$$

for any $\theta \in [0, 2\pi)$.

The states ρ_k are projectors onto the corresponding state $\psi_k = |0\rangle^{\otimes k} \otimes |\phi\rangle^{\otimes n-k}$. Since the states $\{|\psi_k\rangle\}_{k=0}^{n}$ are linearly independent (except for $\theta = 0$, both states are the same), the optimal

measurement is a projective measurement $\mathcal{M} = \{|\varphi_n\rangle\langle\varphi_n|\}_{j=0}^{n}$ [Belavkin, 1975]. It is proven by Sentís et al. [2016] that the success probability, in the asymptotic limit of large n, is

$$P_s \approx \frac{4(1-c^2)}{\pi^2} K^2(c^2) \tag{5.4}$$

where $c = |\langle 0|\phi\rangle|$ is the overlap between the initial and final state and $K(x)$ the complete elliptic function of the first kind. This limit is achieved by a non-local measurement on the whole set of particles.

In the following work, we will lay aside this result since the states considered won't be pure. However, it will be useful as way to check ours in the limit of the mixed state going to pure, in the Bloch sphere this is just taking $\|\boldsymbol{r}\|_2 \to 1$.

5.2 Pure vs. Mixed

This section studies the change from a pure state to a mixed state. The initial pure state can be considered to be $\rho = |0\rangle\langle 0|$ without loss of generality as any other pure state is related with this one by a unitary transformation which will not change the success probability, but only the POVM elements.

5.2.1 Along the diameter

In the first case, Alice prepares initially the state $|0\rangle$ which is affected by an uncorrelated or white noise which depolarises the state. In the Bloch sphere, the noisy state would correspond to a point in the diameter joining $|0\rangle$ and $|1\rangle$. From eq. (1.10), this change is represented as

$$\rho = |0\rangle\langle 0| \quad \longrightarrow \quad \widetilde{\rho} = \frac{\mathbb{I}_2 + r\sigma_z}{2} \tag{5.5}$$

for some value of the parameter $r \in [-1, 1]$. When $r = -1$, the state is $|1\rangle$ which is orthogonal to $|0\rangle$ and therefore we expect to distinguish the change point without ambiguity. On the other hand, if $r = 1$, all the states ρ_k are the same, leaving us with no choice but to randomly guess the position. The reader may question whether in this situation the discrimination task does even make sense, the answer follows from the fact that Bob's inability to distinguish the states doesn't mean that the change point hasn't occurred at all. The state $|0\rangle$ is just some label used to identify some complicated state of the system which is then changed to another state still seen as $|0\rangle$, since you don't have the means to notice the difference[7]. Therefore, the discrimination problem still makes sense although the probability of correctly finding the change point cannot be bigger than the initial *a priori* probabilities $1/(n+1)$.

Without needing to turn to the mathematics, we can intuitively make a guess on the success probability by performing local measurements on the system. Because both states ρ and $\widetilde{\rho}$ are diagonal, they can be understood as classical probability distributions, specifically, they resemble that of a Bernoulli distribution $\mathscr{B}(p) = \{p, 1-p\}$ where p is the probability of obtaining 0 and $1-p$ that of getting 1. Think of ρ as a coin with distribution $\mathscr{B}(1)$, one that always tosses heads and $\widetilde{\rho}$ as a coin with distribution $\mathscr{B}((1+r)/2)$. When will you guess correctly that the change was in k? Obviously, whenever in the k-th toss a 1 is obtained which happens with probability $(1-r)/2$. Only if the state ρ_n is given then you will success without error which happens $1/(n+1)$ times. Thus, for $k = 0, \ldots, n-1$ the success probability is $p(1|k)p(k) = (1-r)/[2(n+1)]$ and $p(n) = 1/(n+1)$ for the extra state ρ_n, as a result the change is identified with probability

$$P_s = \frac{1}{n+1}\left[1 + n\frac{1-r}{2}\right] \tag{5.6}$$

[7] Think of a factory producing perfectly spherical red and green marbles which we label as $|\bullet\rangle$ and $|\bullet\rangle$ respectively. Alice gives to Bob a marble for him to classify and separate in two boxes depending on the colour. Classically, Bob can measure the marbles in different ways but the most effective is by looking at them and identifying its colour. However, if the lights go off, he can no longer use the vision and the two marbles look the same for him, $|\bullet\rangle$ and $|\bullet\rangle$, because there is no other measure (label) he can use to distinguish. Bob is then forced to just guess each time the colour with a 50% change of error.

This function has the behaviour we expected to find: when $r = -1$, it is equal to unity and when $r = 1$ gives the uniform probability $1/(n+1)$. In the limit $n \gg 1$, the success probability tends to the probability of finding a 1 in the k-th position, $(1-r)/2$. Furthermore, in the limit of large n, the probability of correctly identifying the change is finite, with value $(1-r)/2$, which is remarkable since a priori one would have expected that the probability of correctly guessing the change point vanishes for large n.

What we have done in this quick *gedankenexperiment* is to perform the best local measurement on the states which is given by Helstrom (see eq. (2.17)) that tells you to do a projective measurement $\{\Pi_0, \Pi_1\}$ into the states $|0\rangle$ and $|1\rangle$. This result is completely natural as the state $\widetilde{\rho}$ has a non-zero probability of projecting into the state $|1\rangle$ (unless $r = 1$), so whenever the hypothesis Π_1 is selected, we can be 100% that the state was $\widetilde{\rho}$.

Indeed, the systematical analysis together with the numerical results confirm our guess for the probability (5.6) (see Appendix C.1.1) with the expression for the observables

$$\begin{cases} \Pi_j = |0\rangle\langle 0|^{\otimes j} \otimes |1\rangle\langle 1| \otimes \mathbb{I}_2^{\otimes(n-1-j)} & \forall j = 0, \dots, n-1 \\ \Pi_n = |0\rangle\langle 0|^{\otimes n} \end{cases} \tag{5.7}$$

which constitute the optimal measure $\mathcal{M} = \{\Pi_k\}_{k=0}^n$ with $\Pi_j \Pi_k = \delta_{jk} \Pi_j$. Each Π_j is build to measure a one in the position $j+1$, the first position where the state is $\widetilde{\rho}$, the position of the change point, and leaves the rest untouched. It is quite remarkable that the measures do not depend on the value of r but are valid for all the possible values of the noise, no matter how strong it is or how distinguishable are the two states ρ and $\widetilde{\rho}$.

5.2.2 To any mixed state

The first generalisation of the previous results is to consider a QCP between a pure state and any other mixed state, this is

$$\rho = |0\rangle\langle 0| \quad \longrightarrow \quad \widetilde{\rho} = \frac{\mathbb{I}_2 + \boldsymbol{r} \cdot \boldsymbol{\sigma}}{2} \tag{5.8}$$

for some $\boldsymbol{r} = (r_1, r_2, r_3)$ such that $\|\boldsymbol{r}\|_2 \leq 1$. However, because of rotational symmetry, we can consider without lack of generality that $r_2 = 0$, since all the points lying on the plane $z = r_3$ have the same success probability as they are related by a global unitary transformation.

The state vector of a qubit lying on the xz plane can be written as $\boldsymbol{r} = r(\sin\theta, 0, \cos\theta)$ with $r \in [0, 1]$ and $\theta \in [0, 2\pi)$. Then, the structure of the states is

$$\rho_k = \frac{1}{2^{n-k}} |0\rangle\langle 0|^{\otimes k} \otimes \begin{pmatrix} 1 + r\cos\theta & r\sin\theta \\ r\sin\theta & 1 - r\cos\theta \end{pmatrix}^{\otimes(n-k)} \tag{5.9}$$

The matrices are no longer diagonal, this is the first encounter with a genuine quantum mechanical problem where we start having correlations between states.

Due to the difficulty to find the exact expression for the measure, we will make a guess based on the one previously found for the case of two states lying on the diameter. Taking as reference eq. (5.7), we propose the measure

$$\begin{cases} \Pi_j = \mu\, |0\rangle\langle 0|^{\otimes j} \otimes \Lambda \otimes \mathbb{I}_2^{\otimes(n-j-1)} & \forall j = 0, \dots, n-1 \\ \Pi_n = \mathbb{I}_{2^n} - \sum_{j=0}^{n-1} \Pi_j \end{cases} \tag{5.10}$$

where Λ is a positive operator. Recall that a POVM is a set of positive operators that add up to the identity. The last operator Π_n has to complete the measure, the crucial point here is to check that the construction yields a positive operator. We have done so by introducing a positive parameter μ which will be determined by imposing $\Pi_n \geq 0$ once the operators are found. With this choice, we are guaranteed to have at least a feasible measure. To determine the operator Λ, substitute into the expression for the success probability (2.6). The traces in there give

$$\operatorname{tr}(\Pi_k \rho_k) = \mu \operatorname{tr}(|0\rangle\langle 0|)^k \operatorname{tr}(\Lambda\widetilde{\rho}) \operatorname{tr}(\widetilde{\rho})^{(n-k-1)} = \mu \operatorname{tr}(\Lambda\widetilde{\rho})$$

where we have used that $\mathrm{tr}(\rho) = 1\ \forall\rho$ in the last equality. This result holds for all $k = 0, \ldots, n-1$, for $k = n$ we have

$$\mathrm{tr}(\Pi_n \rho_n) = \mathrm{tr}\left[\left(\mathbb{I}_{2^n} - \mu \sum_{j=0}^{n-1} \Pi_j\right) \rho_n\right] = \mathrm{tr}(\rho_n) - \mu \sum_{j=0}^{n-1} \mathrm{tr}(\Pi_j \rho_n) = 1 - n\mu\,\mathrm{tr}(\Lambda\rho)$$

Putting the previous two results together, setting $\xi_k = 1/(n+1)$, we obtain

$$P_s = \max_{\Lambda,\mu}\ \frac{1}{n+1}\left(1 + n\mu\,\mathrm{tr}[\Lambda(\widetilde{\rho} - \rho)]\right) \tag{5.11}$$

subject to the conditions $\Lambda \geq 0$ and $\mu \geq 0$. This problem has already been solved this in section 2, the solution is given by Λ being the projector onto the positive subspace of $X = \widetilde{\rho} - \rho$. This imposes $0 \leq \mu \leq 1$, otherwise the probability could be greater than one when the states are orthogonal, so we take $\mu = 1$ as it is the maximum value it can take. Consequently, the probability of successfully identifying the change is

$$P_s = \frac{1}{n+1}\left(1 + n\lambda_+\right) = \frac{1}{n+1}\left(1 + \frac{n}{2}\|\widetilde{\rho} - \rho\|_1\right) \tag{5.12}$$

being λ_+ the positive eigenvalue of X and Λ the projector onto the positive subspace X_+. The last equality holds since $\widetilde{\rho} - \rho$ is trace-less and thus the eigenvalues must be the same with opposite sign. The reader may argue why we take only the positive subspace while in eq. (2.15), for the same maximisation function, the operator is chosen to be $X_+ - X_-$. The reason is simple, in eq. (2.15) Λ is restricted to be $-\mathbb{I}_2 \leq \Lambda \leq \mathbb{I}_2$ while here $0 \leq \Lambda \leq \mathbb{I}_2$, therefore it can only be X_+ or $-X_-$, but because we want to detect $\widetilde{\rho}$ we must choose $\Lambda = X_+$.

If the states ρ and $\widetilde{\rho}$ have state vector $\boldsymbol{r}$ and $\widetilde{\boldsymbol{r}}$, then the eigenvalues of $\widetilde{\rho} - \rho$ are $\pm\|\widetilde{\boldsymbol{r}} - \boldsymbol{r}\|_2/2$ and thus

$$P_s = \frac{1}{n+1}\left(1 + \frac{n}{2}\|\widetilde{\boldsymbol{r}} - \boldsymbol{r}\|_2\right) \tag{5.13}$$

Considering $\boldsymbol{r} = (0, 0, 1)$ and $\widetilde{\boldsymbol{r}} = r(\sin\theta, 0, \cos\theta)$ one obtains the function

$$P_s = \frac{1}{n+1}\left(1 + \frac{n}{2}\sqrt{1 + r^2 - 2r\cos\theta}\right) \tag{5.14}$$

which reduces to (5.6) in the specific cases $\theta = 0, \pi$.

This strategy takes into account that the initial state is pure so projecting onto itself is the best we can do to distinguish it from $\widetilde{\rho}$, for the rest we consider that the best way to detect the change is to measure in the POVM $\{\Lambda, \mathbb{I}_2 - \Lambda\}$ that best discriminates the states $\widetilde{\rho}$ and ρ, that is the Helstrom measure (2.17). We can in turn give an analytical expression for Λ, given that it is a projector it can be written in the form $|\lambda_+\rangle\langle\lambda_+|$ with $|\lambda_+\rangle = \cos(\varphi/2)\,|0\rangle + \sin(\varphi/2)\,|1\rangle$ a qubit in the xz plane. The angle φ can be calculated in terms of r and θ from the eigenvector corresponding to the positive eigenvalue of $\widetilde{\rho} - \rho$ which gives

$$\sin\frac{\varphi}{2} = \frac{r\sin\theta}{\sqrt{r^2\sin^2\theta + \left(1 - r\cos\theta - \sqrt{1 + r^2 - 2r\cos\theta}\right)^2}} \tag{5.15}$$

The characteristic behaviour of the measurement angle is plotted in fig. 2, where we see that it ranges from $\pi/2$ to π. For $r = 0$, the function has the constant angle $\varphi = \pi$ as expected which tells us to measure in the orthogonal direction to $|0\rangle$. For $0 < r < 1$, the angle φ first decreases as $\widetilde{\rho}$ separates from $|0\rangle$ until a minimum is reached at $\theta_{min} = \arccos r$, after this point the measurement direction returns smoothly to the orthogonal position. The minimum tells us that we won't obtain further information by making the projector and the state parallel, but by making it orthogonal with respect to the initial state. Finally, for $r = 1$, the function diverges at $\theta = 0$ because any angle φ will give us the same success probability and grows linearly with θ up to π for the other values.

We will compare the bound (5.14) with the numerical values obtained using the iterative method. The convergence of the solution is exponentially fast, a precision of $\epsilon = 10^{-6}$ is achieved after seven iterations. SDP will be used only for three qubits as a way to validate and compare the results

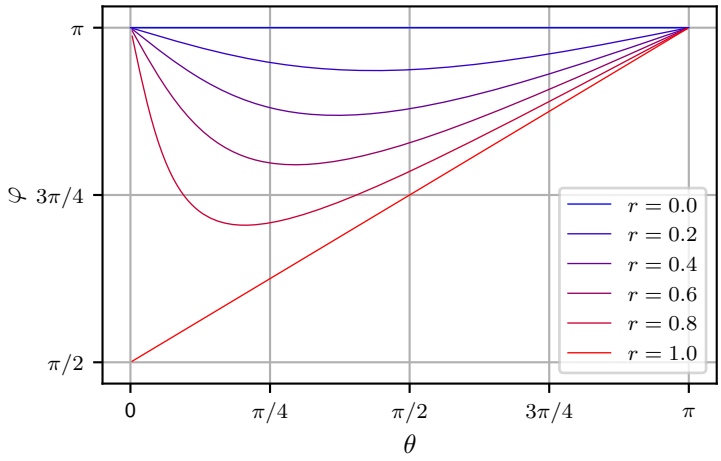


Figure 2: Evolution of the measurement direction as function of the angle θ of the second qubit for various purities r.

obtained via the latter method. Here, in fig. 3, we present the case of $n = 9$ qubits which is the largest sequence that we could solve numerically. The results are plotted as function of the angle θ and the purity r (more plots can be seen in Appendix C.1 for other n). We restrict to the range $\theta \in [0, \pi)$ as the probability (5.14) is symmetrical with respect to the z axis.

On the one hand, the dependence on the angle is somehow the expected, as $r \to 0$ the function flattens up to the limit $P_s = 0.5499998 \pm 37 \cdot 10^{-7}$ in complete agreement with eq. (5.6) and approach the limiting function (5.4) as $r \to 1$. The surprising feature is the change in the concavity of the probability at $\theta_{inf} = \arccos r$, the same at which the measurement angle is minimum, going from positive to negative concavity, instead of being linear or monotone as in fig. 7, this is due to the contributions coming from the correlations. That is, when, $\theta \to 0$ the two states ρ and $\widetilde{\rho}$ become almost parallel but with different amplitudes which promote an extra factor of distinguishability that increases P_s. On the other hand, when $\widetilde{\rho}$ becomes antiparallel with ρ (always in the sense of the Bloch sphere), the states should be maximally distinguishable but, due to the impurity of the second state, P_s reduces.

On the other hand, the dependence on r is also characteristic, for $\theta \leq \pi/2$ the function presents a minimum at $r_{min} = \cos\theta$. As a result, for a given angle, there exist an amplitude which is less distinguishable than any other, specifically, less distinguishable than the completely mixed state $\widetilde{\rho} = \mathbb{I}_2/2$. After θ goes over $\pi/2$, the function grows monotonously as expected.

Equation (5.14) is an exact result which follows from the specific form of the measure (5.10) considered but it can be seen that it gives a good bound to the success probability. We can say that it fits with great accuracy the behaviour under a change of θ but it fails to adjust for some fixed θ, like is seen in the right hand side plots of figs. 3, 10a and 10b with a difference of at most a 10% with respect to the numerical results. This small error is expected since the measure (5.10) is not optimal, despite being feasible.

The optimality of the measure can be checked by constructing the corresponding Lagrange operator

$$\Gamma = \frac{1}{n+1}\left[\mathbb{I}_{2^n} + \sum_{j=0}^{n-1} \Pi_j(\rho_j - \rho_n)\right] \tag{5.16}$$

which is not guaranteed to be hermitian, neither positive, because of the second term. To see this, substitute with the expression for the measure (5.10) and the states (5.9),

$$\sum_{j=0}^{n-1} |0\rangle\langle 0|^{\otimes j} \left[(\Lambda\widetilde{\rho}) \otimes \widetilde{\rho}^{\otimes(n-j-1)} - (\Lambda\,|0\rangle\langle 0|) \otimes |0\rangle\langle 0|^{\otimes(n-j-1)}\right]$$

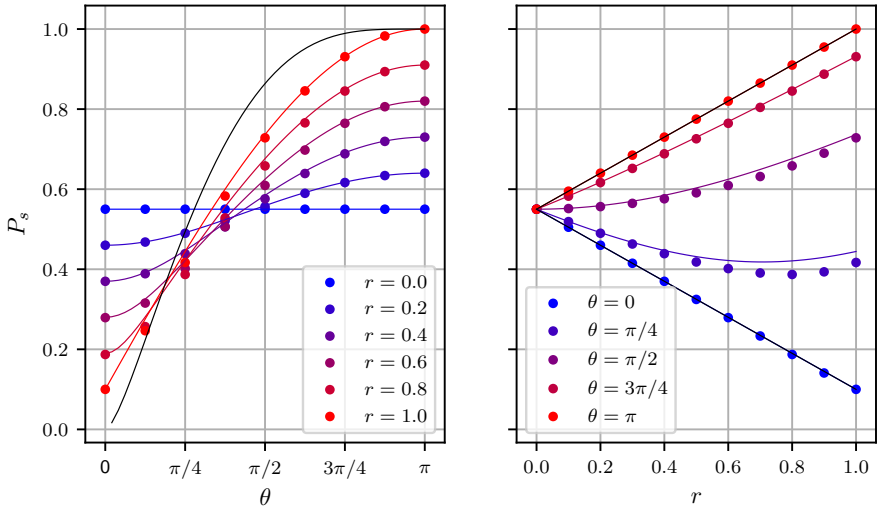


Figure 3: Success probability for the change pure to mixed with $n = 9$. The leftward plot shows the dependence on θ for various purities r (dots), the asymptotic solution for n large when $r = 1$. The rightward plot shows the dependence on r for various representative θ, also the analytical solutions found previously when $\theta = 0, \pi$. The solid colour lines gives the Helstrom bound.

Taking the hermitian conjugate of the previous leads to

$$\sum_{j=0}^{n-1} |0\rangle\langle 0|^{\otimes j} \left[(\widetilde{\rho}\Lambda) \otimes \widetilde{\rho}^{\otimes(n-j-1)} - (|0\rangle\langle 0|\,\Lambda) \otimes |0\rangle\langle 0|^{\otimes(n-j-1)} \right]$$

So only when $[\Lambda, \widetilde{\rho}] = [\Lambda, |0\rangle\langle 0|] = 0$ we will have $\Gamma = \Gamma^\dagger$ which is not the case in general since the three matrices Λ, $\widetilde{\rho}$ and $|0\rangle\langle 0|$ must commute and therefore share a common basis where the 3 are diagonal. This will happen when $\widetilde{\rho}$ lays over the z axis, the three will be diagonal and the measure (5.10) will be optimal, whereas any other state does not satisfy the requirement to be optimal. However, because of the simplicity and accuracy of the one proposed, it should not be rejected at all.

To finish this discussion, we can compare this expression in the limit $n \gg 1$ and $r = 1$ with that obtained by Sentís et al. [2016]. In this regime, the probability takes the simple form $P_s \approx |\sin(\theta/2)|$. Plotting both functions together (see fig. 4) we see that only when $\theta < \pi/2$ the functions differs in a significant amount. Thus, we may conclude that eq. (5.14) is a good approximation to the success probability of correctly identifying a change point when the initial state is pure and the final mixed.

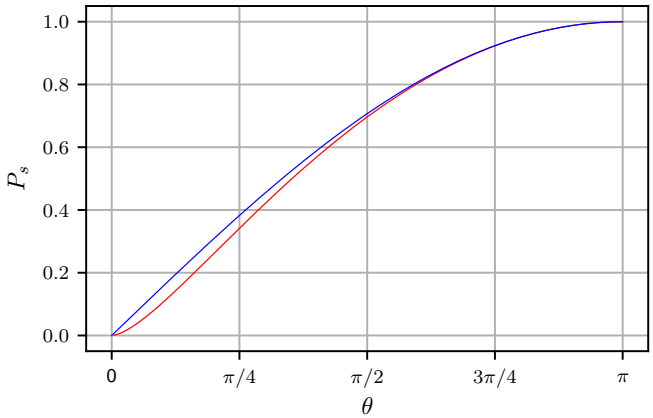


Figure 4: Comparison of the QCP discrimination between two pure states in the asymptotic limit of large n. In red, the solution (5.4) and in blue, the solution (5.14).

5.3 Mixed vs. Mixed

In this part, we will take the first step towards the generalisation of the QCP detection between two mixed states by considering two cases where the symmetry of the system simplifies the resolution of the problem.

We do not seek to find the expression for the measure but to comment the behaviour of the probability function and compare it with the Helstrom measure.

5.3.1 Along the diameter

Consider that Alice is sending a mixed state, both before and after the change, polarised in the z-direction, i.e.

$$\rho = \frac{\mathbb{I}_2 + r_i\sigma_z}{2} \quad \longrightarrow \quad \widetilde{\rho} = \frac{\mathbb{I}_2 + r_f\sigma_z}{2} \tag{5.17}$$

with $r_i, r_f \in [-1, 1]$.

This is a quite realistic case as it is very difficult to engineer a machine that generates perfectly pure states, usually it will be created with a small noise represented by the parameters r_i and r_f in eq. (5.17). A first consequence of this initial noise is that there exist no state $\widetilde{\rho}$ orthogonal to ρ if $-1 < r_i < 1$, therefore we know before hand that the measure (5.7) would not be optimal here.

Taking as reference the decomposition of $\widetilde{\rho}^{\otimes(n-k)}$ done in eq. (C.1), the expression for the states ρ_k gives

$$\rho_k = \sum_{x=0}^{2^n-1} \left(\frac{1+r_i}{2}\right)^{k-S_0^k(x)} \left(\frac{1-r_i}{2}\right)^{S_0^k(x)} \left(\frac{1+r_f}{2}\right)^{n-k-S_k^n(x)} \left(\frac{1-r_f}{2}\right)^{S_k^n(x)} |x\rangle\langle x| \tag{5.18}$$

where the function $S_a^b(x)$ with $0 \leq a \leq b \leq n$ sums all the ones in the binary number $x = (x_0 x_1 \dots x_{n-1})$ from x_a to x_{b-1}, if no indices are specified the function is understood to mean the sum over the whole range from 0 to $n-1$. The coefficient that goes together with each state x is understood as the conditional probability $p(x \mid k)$ that given the change position k the outcome x is obtained.

On account of what was done to solve analytically the pure to mixed case, we again consider the POVM operators $\{\Pi_j\}$ fully diagonal, nothing else can be set to 0 in advance. The product $\Pi_k\rho_k$ is then trivial, denoting by π_x^k the x diagonal element of Π_k, the expression for the success probability is

$$P_s = \frac{1}{n+1} \max_{\{\Pi_j\}} \sum_{k=0}^{n} \sum_{x=0}^{2^n-1} \pi_x^k p(x \mid k) \tag{5.19a}$$

$$\text{subject to} \quad \sum_{k=0}^{n} \pi_x^k = 1 \qquad \forall x \in \{0,1\}^n \tag{5.19b}$$

The measure is constructed as follows: the function in eq. (5.19a) will be maximum when each of the 2^n components of the sum are the greatest they can be, thus, given an x, the only non-vanishing component $\pi_x^{k^*}$ corresponds to

$$k^* = \arg\max_k \, p(x \mid k) \tag{5.20}$$

Then, by the completeness relation (5.19b) we conclude $\pi_x^{k^*} = 1$. Clearly, a measure constructed in this way automatically satisfies the Holevo conditions of optimality as the minimum eigenvalue of Γ is at least equal than the maximum of ρ_k.

Keeping eq. (5.20) in mind, we can rewrite the success probability (5.19a) as

$$P_s = \frac{1}{n+1} \sum_{x=0}^{2^n-1} \max_k p(x \mid k) \tag{5.21}$$

Let us pause for a moment and meditate about the result found in eq. (5.21). The sum over all the states ρ_k has been replaced by the sum over all of its possible eigenstates $\{|x\rangle\}_{x=0}^{2^n-1}$ and we are saying that the contribution to the total success probability is given by the k that maximises the

conditional probability $p(x|k)$. The states $|x\rangle$ represent the possible outcomes of a local measurement performed by Bob in the Z basis and the number $p(x|k)$ tells us the state that is more likely to have a change in k. The maximum over all k for a fixed x is the one that fits best with the characteristics of the problem.

These outcomes of a quantum measurements follow classical probability distributions, can we find a classical analogue to the problem? Of course, the answer is yes, just read the previous paragraph again replacing the word state by path, our problem becomes that of a Random Walk (RW). The states $|x\rangle$ are just paths on a one dimensional euclidean space, "0" meaning "move one unit upwards" and 1 meaning "move one unit downwards". The probability of moving up/down (measuring a 0 or 1) are those of the states in eq. (5.17), before and after the change point, which constitute two Bernoulli distributions $\mathscr{B}(p)$ and $\mathscr{B}(q)$ respectively. The change point that best divides a path x would be given by the maximum distance from the initial point, i.e. the step k that maximises $p(x|k)$. This last statement was proven by Lorden et al. [1971] and it sets the basics for the CUSUM algorithm in the classical theory of change point detection [Basseville et al., 1993]. The theorem states that, for a series of random variables $x_0, x_1, \ldots, x_n$ distributed according to some probability distributions $p_{\theta_0}(x)$ and $p_{\theta_1}(x)$, before and after the change respectively, the best guess for the detection of the change is the one that maximises the likelihood function. In our case, the likelihood function can be interpreted as the distance from the origin in the random walk.

Symmetric solution The general problem doesn't have an analytical solution but some work can be done when one considers a mirror change, in which the probability of 0 and 1 are exchanged,

$$\rho = \frac{\mathbb{I}_2 + r\sigma_z}{2} \quad \longrightarrow \quad \widetilde{\rho} = \frac{\mathbb{I}_2 - r\sigma_z}{2} \tag{5.22}$$

Then, the expression for the states simplifies to

$$\rho_k = \sum_{x=0}^{2^n-1} \left(\frac{1+r}{2}\right)^{k-S_0^k(x)+S_k^n(x)} \left(\frac{1-r}{2}\right)^{n-k+S_0^k(x)-S_k^n(x)} |x\rangle\langle x| \tag{5.23}$$

and using that $S_0^k(x) + S_k^n(x) = S(x)$ we can factor out the terms independent of k so that condition (5.20) takes the form

$$k^* = \arg\max_k \left(\frac{1-r}{1+r}\right)^{\nu_k(x)} \tag{5.24}$$

with the definition $\nu_k(x) \equiv 2S_0^k(x) - k$. The maximum will be achieved at the same time that $\nu_k(x)$ presents a maximum, in other words, the distance from the origin is larger than at any other step. This can be calculated in the asymptotic limit of large n as then all the positions are approximately equally likely to have a change point and thus eq. (5.21) reduces to

$$P_s \approx \sum_{x=0}^{2^n-1} p(x \mid k) \tag{5.25}$$

which is just the probability that given the hypothesis of a change point in k, there is a maximum in x.

The sum in eq. (5.25) represents an arduous task of counting all the path that have a maximum in k, in Appendix C we provide some insight of how would it be done. For now, remember we are in the limit of large n and we can take profit of it. Each term is equivalent to moving our reference point to the maximum and calculating the probability that a RW to the left never returns to the origin and a RW to the right never crosses it. The multiplication of the two expressions gives the total probability to have a maximum. Considering $r \geq 0$, we end up with

$$P_s \approx \frac{2r^2}{1+r} \tag{5.26}$$

The previous function has the desired properties: it vanishes at $r = 0$ as the two states are the same and is maximum at $r = 1$ when they are orthogonal. If we wanted to consider $r < 0$, we

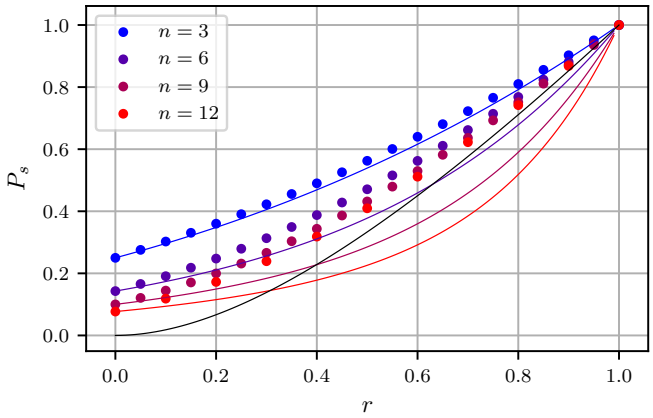


Figure 5: Solutions for a change point with symmetrical states in the z direction, numerically (dots) for $n = 3, 6, 9, 12$ using SDP and analytically (solid black line) in the asymptotic limit of large n. The solid colour lines correspond to the Helstrom bound.

should reconsider the problem, searching for a minimum instead of a maximum, but because it is completely symmetrical, this corresponds to replacing r with $-r$ in eq. (5.26).

The numerical analysis can be performed in the same way as in Section 5.2.1, using the corresponding states in eq. (5.23). The results are plotted in fig. 5 for multiple number of qubits with the limiting function (5.26). Unfortunately, the computational limits do not allow us to achieve this limit but we can observe that as n increases, the values approach the asymptotic limit.

The Helstrom bound is obtained from the measure (5.7), calling $\Lambda_0 = |0\rangle\langle 0|$ and $\Lambda_1 = |1\rangle\langle 1|$ and substituting into the expression for P_s we encounter the following two terms

$$\operatorname{tr}(\Pi_k \rho_k) = \operatorname{tr}(\Lambda_0 \rho)^k \operatorname{tr}(\pi_1 \widetilde{\rho}) \operatorname{tr}(\widetilde{\rho})^{(n-k-1)} = \operatorname{tr}(\Lambda_0 \rho)^k \operatorname{tr}(\Lambda_1 \widetilde{\rho}) \qquad k = 0, \ldots, n-1$$
$$\operatorname{tr}(\Pi_n \rho_n) = \operatorname{tr}(\rho_n) - \sum_{j=0}^{n-1} \operatorname{tr}(\Pi_j \rho_n) = 1 - \sum_{j=0}^{n-1} \operatorname{tr}(\Lambda_0 \rho)^k \operatorname{tr}(\Lambda_1 \rho)$$

Joining both result and inserting them in eq. (2.6), with $\xi_k = 1/(n+1)$, gives

$$P_s = \max_{\{\Lambda_0, \Lambda_1\}} \frac{1}{n+1} \left(1 + \operatorname{tr}[\Lambda_1(\widetilde{\rho} - \rho)] \sum_{k=0}^{n-1} \operatorname{tr}(\Lambda_0 \rho)^k \right)$$

The sum is a geometric series with ratio $\operatorname{tr}(\Lambda_0 \rho)$ which adds to

$$P_s = \max_{\{\Lambda_0, \Lambda_1\}} \frac{1}{n+1} \left(1 + \operatorname{tr}[\Lambda_1(\widetilde{\rho} - \rho)] \frac{1 - \operatorname{tr}(\Lambda_0 \rho)^n}{1 - \operatorname{tr}(\Lambda_0 \rho)} \right) \tag{5.27}$$

This is the general expression for the Helstrom bound where $\{\Lambda_0, \Lambda_1 = \mathbb{I}_2 - \Lambda_0\}$ is the POVM that best discriminates the states ρ and $\widetilde{\rho}$. In our concrete example, substituting for Λ_0, Λ_1, ρ and $\widetilde{\rho}$ gives the expression

$$P_s = \frac{1}{n+1} \left[1 + \frac{2r}{1-r} \left(1 - \left(\frac{1+r}{2} \right)^n \right) \right] \tag{5.28}$$

Once plotted (see fig. 5), this is found to be suboptimal as compared with the values found numerically. In fact, in the limit $n \gg 1$ the success probability tends to 0 (except for $r = 1$ which still we can discriminate without error) instead of leading to (5.26).

An improvement to this method would be to measure locally the states in the Helstrom POVM $\{\Lambda_0, \Lambda_1\}$ and then to apply the technique of the random walk to the measure outcomes. Then, by Lorden et al. [1971], the best guess to the change point position won't be the first position at which the hypothesis Λ_1 is obtained but that corresponding to the maximum of the RW.

5.3.2 Equally mixed states

To imagine this case, consider that Alice sends first k states of the first kind $|0\rangle$ but there is an error in the generating machine which rotates this state to $|\phi\rangle$ in eq. (5.3). For now, the two states are pure and the change point problem is already solved. Yet, we have to remember that the real world is noisy and mixes the states during the communication process in the same amount. Therefore, Bob no longer receives $|0\rangle$ or $|\phi\rangle$ but an effective ρ and $\widetilde{\rho}$, where

$$\rho = \frac{\mathbb{I}_2 + r\sigma_z}{2} \qquad \longrightarrow \qquad \widetilde{\rho} = U_\theta \rho U_\theta^\dagger = \frac{\mathbb{I}_2 + \boldsymbol{r}\cdot\boldsymbol{\sigma}}{2} \tag{5.29}$$

with $\boldsymbol{r} = r(\sin\theta, 0, \cos\theta)$[8] with $\theta \in [0, 2\pi)$. This two states are said to have the same purity, i.e. they lay over the same surface in the Bloch sphere with radius r, and consequently are related by a unitary transformations.

The values for the success probability have been found using the iterative algorithm and they are compared with the von Neumann measure

$$\begin{cases} \Pi_j = \mu\Lambda_0^{\otimes j} \otimes \Lambda_1 \otimes \mathbb{I}_2^{\otimes(n-j-1)} & \forall j = 0, \dots, n-1 \\ \Pi_n = \mathbb{I}_{2^n} - \sum_{j=0}^{n-1} \Pi_j \end{cases} \tag{5.30}$$

where $\{\Lambda_0, \Lambda_1 = \mathbb{I}_2 - \Lambda_0\}$ is the POVM that best distinguishes the states ρ and $\widetilde{\rho}$ locally given by the Helstrom measure (2.17) and μ a positive parameter to ensure $\Pi_n \geq 0$.

The results obtained by the iterative method and using the measure (5.30) are plotted in fig. 6, showing the dependence on the angle and the purity separately. The typical behaviour is obtained, the probability is minimum when the two states are the same, with a value that goes as $1/(n+1)$, and maximum when the states are pure and completely orthogonal. It is possible to see that the values tends to the limit function (5.4) for $r = 1$ and to (5.26) for $\theta = \pi$. The bound given by the projective measure, with $\mu = 1$ as shown by numerical methods, results in a poor approximation to the real value. This was expected from the previous section, as it should agree with eq. (5.25) when the states are over the same diameter. Furthermore, the bound doesn't even follow the same behaviour as it can be seen from a change of concavity around $\pi/2$ and $r > 0.4$ that doesn't show up in the tendency followed by the numerical values. We should conclude that the measure (5.30) is not optimal.

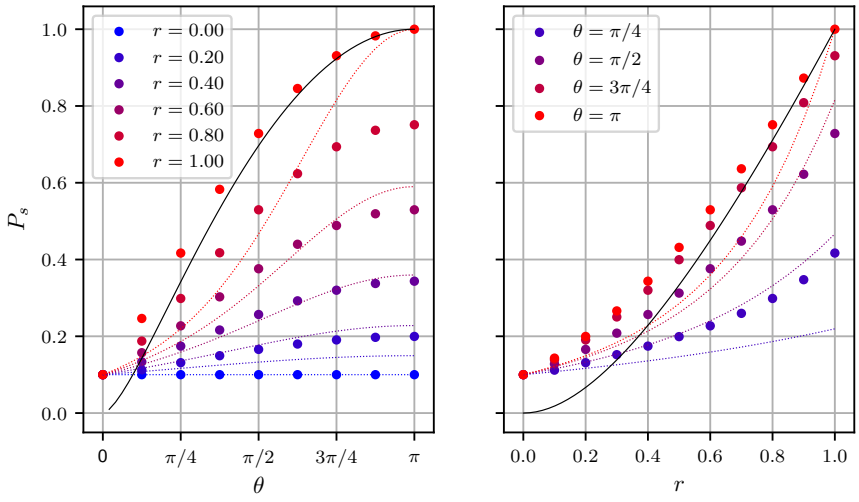


Figure 6: Success probability for the change mixed to mixed with same purity and $n = 9$. The leftward plot shows the dependence on θ for various purities r (dots) and the asymptotic solution for n large when $r = 1$. The rightward plot shows the dependence on r for various representative θ, also the analytical solutions found previously when $\theta = \pi$. The dashed colour lines gives the Helstrom bound.

[8] Again, the y component is set to 0 as all the states in a plane perpendicular to the z direction have the same success probability.

6 Conclusions

We have learnt about one of the most intrinsic problems of quantum mechanics which is Quantum State Discrimination, providing a consistent technique build upon a measurement strategy to best discriminate any set of states. Reviewed the analytical conditions to be met by a discrimination strategy to be optimal which showed to be the same as in the SDP formulation. This numerical technique allows to find exact solutions to the discrimination problem. However, the complexity of the problem in the presence of noise grows exponentially with the number of qubits, in contrast to the case of pure states which grows linearly. Therefore, we considered an iterative method which showed to be efficient and used SDP to benchmark their performance when possible. This last method, despite not being exact, converged very rapidly to the solution.

The study of the QCP started by considering the change from a pure to a mixed state. In this regime, an analytical expression for the probability and the measures where found when both states are aligned over the same diameter of the Bloch sphere. The best guess to be made was the best classical guess since both states diagonalise in the same basis and thus the quantum states could be interpreted as classical probability distributions. The numerical values obtained adjusted perfectly with the theoretical result.

The previous example led to the generalisation of a pure state going to another mixed state. First, we provided a bound on the probability by building a projective measure with the best local discrimination protocol given by Helstrom. This showed to be a good approximation, despite not being optimal, when comparing with the numerical results obtained by the iterative algorithm with a precision up to the 6-th floating point position. Moreover, taking the second state as pure and for a large number of qubits, the bound proposed provided a good approximation to the asymptotic solution already found in literature for the QCP with two pure states.

The next logical step was to remove the purity of the first state and consider the change point problem among two mixed steps. At first, by taking both of them over the same diameter. As before, this problem could be understood using classical means with the aid of the theory of random walks. This was found to have no known analytical solution in the general case. However, for a symmetric change, we could find an analytical expression in the asymptotic regime. Under this specific situation, the numerical results showed to tend to the asymptotic solution while the Helstrom measurement gave a lower bound for the probability. Lastly, for completeness, we generalised the previous to post change states outside the diameter, comparing the numerical results with the best projective measurement. The latter showed clearly to be suboptimal.

To conclude, we can state that a projective measurement provides a good approximation to the probability of success for QCP problems when the initial state is pure but poorly fits the numerical results when both are mixed states, before and after the change.

The future work to be done is somehow clear, find the solution for the general problem between two states. Nevertheless, a more feasible goal would be to provide the optimal measure to the pure to mixed case and next to perform further simulations in order to find a good expression for the mixed to mixed case.

References

E. Andersson, S. M. Barnett, C. R. Gilson, and K. Hunter. Minimum-error discrimination between three mirror-symmetric states. *Phys. Rev. A*, 65:052308, Apr 2002. doi: 10.1103/PhysRevA.65.052308.

J. Bae and L.-C. Kwek. Quantum state discrimination and its applications. *Journal of Physics A: Mathematical and Theoretical*, 48(8):083001, 2015.

E. Bagan, R. Muñoz-Tapia, G. Olivares-Rentería, and J. Bergou. Optimal discrimination of quantum states with a fixed rate of inconclusive outcomes. *Physical Review A*, 86(4):040303, 2012.

S. M. Barnett. Minimum-error discrimination between multiply symmetric states. *Phys. Rev. A*, 64:030303, Aug 2001. doi: 10.1103/PhysRevA.64.030303.

S. M. Barnett and S. Croke. Quantum state discrimination. *Adv. Opt. Photon.*, 1(2):238–278, Apr 2009. doi: 10.1364/AOP.1.000238.

M. Basseville, I. V. Nikiforov, et al. *Detection of abrupt changes: theory and application*, volume 104. Prentice Hall Englewood Cliffs, 1993.

V. P. Belavkin. Optimal multiple quantum statistical hypothesis testing. *Stochastics: An International Journal of Probability and Stochastic Processes*, 1(1-4):315–345, 1975.

J. S. Bell. On the einstein podolsky rosen paradox. *Physics Physique Fizika*, 1(3):195, 1964.

C. H. Bennett and G. Brassard. Quantum cryptography: Public key distribution and coin tossing. *Theoretical Computer Science*, 560:7–11, 2014.

C. H. Bennett and P. W. Shor. Quantum information theory. *IEEE transactions on information theory*, 44(6):2724–2742, 1998.

A. Chefles and S. M. Barnett. Optimum unambiguous discrimination between linearly independent symmetric states. *Physics Letters A*, 250(4):223 – 229, 1998. ISSN 0375-9601. doi: https://doi.org/10.1016/S0375-9601(98)00827-5.

C.-L. Chou. Minimum-error discrimination among mirror-symmetric mixed quantum states. *Phys. Rev. A*, 70:062316, Dec 2004. doi: 10.1103/PhysRevA.70.062316.

E. Davies. Information and quantum measurement. *IEEE Transactions on Information Theory*, 24 (5):596–599, 1978.

M. E. Deconinck and B. M. Terhal. Qubit state discrimination. *Physical Review A*, 81(6):062304, 2010.

A. Einstein, B. Podolsky, and N. Rosen. Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.*, 47:777–780, 5 1935. doi: 10.1103/PhysRev.47.777.

W. Feller. *An introduction to probability theory and its applications*. Number v. 1 in Wiley series in probability and mathematical statistics. Probability and mathematical statistics. Wiley, 1968. ISBN 9780471257080.

C. A. Fuchs. *Distinguishability and accessible information in quantum theory*. PhD thesis, University of New Mexico, 1996.

C. A. Fuchs. Quantum mechanics as quantum information (and only a little more). *arXiv preprint quant-ph/0205039*, 2002.

N. Gisin. Quantum cloning without signaling. *Physics Letters A*, 242(1):1 – 3, 1998. ISSN 0375-9601. doi: https://doi.org/10.1016/S0375-9601(98)00170-4.

G. Grimmett and D. Stirzaker. *Probability and Random Processes*. Probability and Random Processes. OUP Oxford, 2001. ISBN 9780198572220.

D. Ha and Y. Kwon. Complete analysis for three-qubit mixed-state discrimination. *Physical Review A*, 87(6):062302, 2013.

C. W. Helstrom. Quantum detection and estimation theory. *Journal of Statistical Physics*, 1(2): 231–252, 1969.

A. Holevo. Statistical decision theory for quantum systems. *Journal of Multivariate Analysis*, 3(4): 337 – 394, 1973. ISSN 0047-259X. doi: https://doi.org/10.1016/0047-259X(73)90028-6.

M. Ježek, J. Řeháček, and J. Fiurášek. Finding optimal strategies for minimum-error quantum-state discrimination. *Physical Review A*, 65(6):060301, 2002.

K. Kraus. *States, effects and operations: fundamental notions of quantum theory*. Springer, 1983.

J. Löfberg. Yalmip: A toolbox for modeling and optimization in matlab. In *Proceedings of the CACSD Conference*, volume 3. Taipei, Taiwan, 2004. URL `https://github.com/yalmip/YALMIP`.

G. Lorden et al. Procedures for reacting to a change in distribution. *The Annals of Mathematical Statistics*, 42(6):1897–1908, 1971.

M. Nielsen and I. Chuang. *Quantum Computation and Quantum Information*. Cambridge Series on Information and the Natural Sciences. Cambridge University Press, 2000. ISBN 9780521635035.

P. Raynal. Unambiguous state discrimination of two density matrices in quantum information theory. *arXiv preprint quant-ph/0611133*, 2006.

J. Reeves, J. Chen, X. L. Wang, R. Lund, and Q. Q. Lu. A review and comparison of changepoint detection techniques for climate data. *Journal of Applied Meteorology and Climatology*, 46(6): 900–915, 2007. doi: 10.1175/JAM2493.1. URL `https://doi.org/10.1175/JAM2493.1`.

T. Rudolph, R. W. Spekkens, and P. S. Turner. Unambiguous discrimination of mixed states. *Physical Review A*, 68(1):010301, 2003.

G. Sentís, E. Bagan, J. Calsamiglia, G. Chiribella, and R. Munoz-Tapia. Quantum change point. *Physical review letters*, 117(15):150502, 2016.

M. Slater. Lagrange multipliers revisited. In *Traces and Emergence of Nonlinear Programming*, pages 293–306. Springer, 2014.

J. F. Sturm. Using sedumi 1.02, a matlab toolbox for optimization over symmetric cones. *Optimization methods and software*, 11(1-4):625–653, 1999. URL `https://github.com/sqlp/sedumi`.

K.-C. Toh, M. J. Todd, and R. H. Tütüncü. Sdpt3—a matlab software package for semidefinite programming, version 1.3. *Optimization methods and software*, 11(1-4):545–581, 1999. URL `https://github.com/sqlp/sdpt3`.

L. Vandenberghe and S. Boyd. Semidefinite programming. *SIAM review*, 38(1):49–95, 1996.

J. von Neumann. *Mathematical Foundations of Quantum Mechanics*. Princeton University Press, 1955. ISBN 9780691028934.

J. Watrous. Semidefinite programming in quantum information, 2011. URL `https://cs.uwaterloo.ca/~watrous/CS867.Winter2017/`.

G. Weir, S. M. Barnett, and S. Croke. Optimal discrimination of single-qubit mixed states. *Phys. Rev. A*, 96:022312, Aug 2017. doi: 10.1103/PhysRevA.96.022312.

H. Wolkowicz, R. Saigal, and L. Vandenberghe. *Handbook of Semidefinite Programming: Theory, Algorithms, and Applications*. International Series in Operations Research & Management Science. Springer US, 2012. ISBN 9781461543817.

A Semi-Definite Programming

The form of the primal problem given in the main text can be generalised to

$$\begin{aligned} \max_X \ & \operatorname{tr} AX \\ & \Phi[X] \leq B \\ & X \geq 0 \end{aligned} \tag{A.1}$$

It differs in the constraint where the equality has been relaxed to a inequality. Consequently, the space of all feasible solutions increases and its optimal value might change.

However, we can rephrase the previous problem with inequality to a problem with equality by the introduction of a *slack variable*. Thus, the inequality $\Phi[X] \leq B$ is the same as $\Phi[X] + Z = B$ for some $Z \geq 0$ such that $Z \in \mathcal{Y}$. Then, the whole problem can be reformulated with the introduction of Z, define

$$\widetilde{A} = A \oplus \mathbb{0}$$
$$\widetilde{X} = X \oplus Z$$

and the map

$$\widetilde{\Phi}[\widetilde{X}] = \Phi[X] + Z$$

Using these definitions, the triplet $(\widetilde{A}, B, \widetilde{\Phi})$ constitute the equivalent SDP problem to (A.1) with an equality constraint.

The set of feasible solutions may easily be checked to be convex. Take $X, Y \in \mathscr{F}_P(\mathcal{X})$ then $Z = \lambda X + (1-\lambda)Y$ satisfies

- $\Phi[Z] = \Phi[\lambda X + (1-\lambda)Y] = \lambda\Phi[X] + (1-\lambda)\Phi[Y] = \lambda B + (1-\lambda)B = B$.
- For a general $|u\rangle$, $\langle u|Z|u\rangle = \lambda \langle u|X|u\rangle + (1-\lambda)\langle u|Y|u\rangle \geq 0$ because, by hypothesis, X, Y are both positive and $\lambda \in [0,1]$. □

Therefore, Z is also inside $\mathscr{F}_P(\mathcal{X})$ and we conclude that any feasible solution can be constructed by a convex combination of two others.

A geometric interpretation of the problem follows from the linear programming problem. The feasible region consist of the boundary curve that encloses the region where $X \geq 0$, i.e. those positive matrix which satisfy $\Phi[X] = B$. Roughly speaking, the semi-definite program tries to move towards A while staying in the feasible region [Vandenberghe and Boyd, 1996]. For this reason, it is important that the feasible region is a convex set, so we can move from one solution to another without problems. Therefore, provided that the problem is feasible, i.e. $\mathscr{F}_P(\mathcal{X}) \neq \emptyset$, there is always an optimal solution on the boundary X_{opt}.

The relationship between the primal and dual problem is obtained by considering the Lagrangian

$$\mathcal{L} = \operatorname{tr} AX + \operatorname{tr}[Y(B - \Phi[X])] + \operatorname{tr} ZX \tag{A.2}$$

where Y, Z act as Lagrange multipliers taking into account the constraints in eq. (3.1), we also require $Z \geq 0$ and $Y = Y^\dagger$. Then, using eq. (3.5), the Lagrangian can be rewritten as

$$\mathcal{L} = \operatorname{tr} AX + \operatorname{tr} YB - \operatorname{tr}(\Phi^*[Y]X) + \operatorname{tr} ZX = \operatorname{tr} YB + \operatorname{tr}[(A + Z - \Phi^*[Y])X] \tag{A.3}$$

The minimum is then found by extremizing the function with respect to X,

$$\frac{\partial \mathcal{L}}{\partial A} = A + Z - \Phi^*[Y] = 0 \Longrightarrow \Phi^*[Y] = A + Z \geq A \tag{A.4}$$

which is the condition set in eq. (3.4) for the dual problem. □

The dual problem provides us with a way to check if a solution to the primal is optimal. First, let's formulate a theorem stating a relation between α and β [Watrous, 2011].

Theorem 2 (Weak duality)**.** *For every semi-definite program* (Φ, A, B) *it holds that* $\alpha \leq \beta$*.*

Proof. The proof is trivial in the case $\mathscr{F}_P = \emptyset$ or $\mathscr{F}_D = \emptyset$ because either $\alpha = -\infty$ and $\beta = \infty$ which automatically hold. For non empty sets, if $X \in \mathscr{F}_P$ and $Y \in \mathscr{F}_D$ it holds that

$$\operatorname{tr} AX \leq \operatorname{tr} \Phi^*[Y]X = \operatorname{tr} Y\Phi[X] = \operatorname{tr} YB$$

Taking the supremum over all $X \in \mathscr{F}_P$ and the infimum over all $Y \in \mathscr{F}_D$ we obtain $\alpha \leq \beta$. □

As a consequence, because $\alpha \geq \operatorname{tr} AX$ and $\beta \leq \operatorname{tr} YB$, it also holds that

$$\operatorname{tr} AX \leq \alpha \leq \beta \leq \operatorname{tr} YB \tag{A.5}$$

for all $X \in \mathscr{F}_P$ and $Y \in \mathscr{F}_D$. Indeed, if one finds a feasible solution for the primal problem X and another for the dual Y for which $\operatorname{tr} AX = \operatorname{tr} YB$ then $\alpha = \beta$ and therefore, X and Y must be optimal solutions. This is called *strong duality* but it may not hold for all SDP problems. However, for physical problems it is usually possible to find such X and Y so that $\alpha = \beta$ [Watrous, 2011].

A.1 Numerical solvers

In order to solve SDP problems numerically we will use the MATLAB solver YALMIP designed to solve this type of problems among many others [Löfberg, 2004]. Numerical optimisation of SDP problems comes with a weighty package of methods from which YALMIP chooses the suitable one depending on the form of the conditions. Also, it includes an option to use other solvers apart from the default YALMIP solver, which we will use as a test to check if a given solution is indeed a solution or may be a consequence of numerical or algorithmic errors. For it, we will use the SDPT3 solver [Toh et al., 1999] and the SeDuMi solver [Sturm, 1999].[9]

Even if we program only the primal problem, each solver will compute both primal and dual individually and return a solution if any of the two conditions of Theorem 1 are satisfied. Even though the program considers them as optimal solutions, we will check numerical reliability with the Lagrange operator constructed from the output of the program. Consider that the program outputs the set of operators $\{\widetilde{\Pi}_j\}$, we will do this in two steps:

1. Check that the output is feasible and indeed satisfy the conditions of the SDP problem. This is done through the method check() implemented by YALMIP which evaluates the smallest eigenvalue of the conditions. For instance, in the SDP problem eq. (3.14) it will compute the minimum eigenvalue of $\lambda_1 = \min_\lambda \widetilde{Y} - A$ and the maximum $\lambda_2^{\max}$ and minimum $\lambda_2^{\min}$ eigenvalue of $Y - Y^\dagger$. The solution $\{\widetilde{\Pi}_j\}$ is considered feasible if $\lambda_1 > 0$ and $\lambda_2^{\max} = \lambda_2^{\min} = 0$.

   Of course, numerical methods are never exact and there may be approximation errors. Consequently, we will add a small tolerance ϵ when we check this conditions. Thus, a solution will be considered approximately feasible if $\lambda_1 + \epsilon > 0$ and $\lambda_2^{\max}, \lambda_2^{\min} \in (-\epsilon, \epsilon)$.

   If this is not the case, $\{\widetilde{\Pi}_j\} \notin \mathscr{F}_P$ and we may investigate why.

2. The previous check is successful and we conclude that $\{\widetilde{\Pi}_j\} \in \mathscr{F}$. We must now find out whether it is optimal or not by using the optimality condition eq. (2.8). This will be done in a similar way as before, computing the minimum eigenvalue $\lambda_k = \min_\lambda \Gamma - \xi_k\rho_k$ for each $\rho_k \in \Xi$. Only when $\lambda_k(+\epsilon) \geq 0\ \forall k$ the solution will be taken as optimal.

[9] The three libraries used are open source and the url to the GitHub repositories may be found in the references.

B Simple Random Walk

A Random Walk is a random process by which objects moves according to some probability distribution $\mathscr{P}$. In the present work, we are interested in the simplest of all, a one dimensional RW with only two possible movements which we identify by up and down. A step is characterised by a random variable x that can take the value $+1$ with probability p and -1 with probability $1-p$. The random variables x_i are said to follow a Bernoulli distribution $\mathscr{B}(p) = \{p, 1-p\}$ where $0 \leq p \leq 1$. A path $\mathbb{X}_n$ of length n is just a sequence of steps $\{x_1, \ldots, x_n\}$. We define the accumulated sum as

$$S_k(\mathbb{X}_n) = a + \sum_{l=1}^{k} x_l \tag{B.1}$$

where a is our starting point. In simpler words, S_n gives us the position of the walker from a at the n-th step. The motion of the particle following $\mathbb{X}_n$ is recorded as the list of points $\{(k, S_k)\}_{k=0}^{n}$ which constitute a path on a plane.

Denote by $N_k(a,b)$ the number of paths from a to b in k steps, the total number of steps is exactly the number of up and down steps, u and d respectively, and from (B.1) it can be seen that $b - a = u - d$. Thus, the number of up and down steps is

$$u = \frac{1}{2}(k + b - a) \tag{B.2a}$$

$$d = \frac{1}{2}(k - b + a) \tag{B.2b}$$

Then, $N_k(a,b)$ is the total number of permutations of exactly u up steps and d down steps [Feller, 1968]

$$N_k(a,b) = \binom{k}{\frac{1}{2}(k+b-a)} = \binom{k}{\frac{1}{2}(k-b+a)} \tag{B.3}$$

so the probability of ending at $S_k = b$ if the walker follows the distribution $\mathscr{B}(p)$ is

$$P(S_k = b) = N_k(a,b) p^{(k+b-a)/2} (1-p)^{(k-b+a)/2} \tag{B.4}$$

The binomial coefficient should be understood to be 0 unless $(n+b-a)/2$ is an integer, for which the binomial has the expression

$$\binom{k}{j} = \frac{k!}{j!(n-k)!} = \binom{k}{k-j} \tag{B.5}$$

In the process of going to b, the walker may or may not have gone through the origin. Let $N_k^0(a,b)$ be the number of paths within $N_k(a,b)$ that visit the origin, i.e. $S_j = 0$ for some $0 < j \leq k$. The reflection or mirror theorem tells us that

Theorem 3 (Mirroring)**.** *If $a, b > 0$ then $N_k^0(a,b) = N_k(-a,b)$.*

By virtue of the reflection theorem we can calculate many results of random walks. For instance, consider the so called *Ballot problem*, how many paths are there from $(0,0)$ to (k,b) $(b > 0)$ such that the walker never returns to the origin? Under this conditions, the first step is already fixed to go to $(1,1)$, then we must calculate the number of paths of $k-1$ steps from $(1,1)$ to (k,b) and subtract those which touch the origin [Grimmett and Stirzaker, 2001]. Using Theorem 3, we end up with

$$N_k^{>0}(0,b) = N_{k-1}(1,b) - N_{k-1}^0(1,b) = N_{n-1}(1,b) - N_{n-1}(-1,b) = \frac{b}{n} N_n(0,b) \tag{B.6}$$

The probability for this to happens is

$$P(S_j > 0 \ \forall j > 0, S_k = b) = \frac{b}{n} \ P(S_k = b) \tag{B.7}$$

C Supplemental calculations for Section 5

C.1 Pure vs. Mixed

C.1.1 Along the diameter

First of all, let's write the form of the density matrices ρ_k for this problem. From eq. (5.1), their expression in the computational basis $\{|j\rangle\}_{j=0}^{2^n-1}$ is

$$\begin{aligned}\rho_k &= |0\rangle\langle 0|^{\otimes k} \otimes \left[\sum_{j=0}^{1} \frac{1+(-1)^j r}{2} |j\rangle\langle j|\right]^{\otimes(n-k)} \\ &= \sum_{j_0,j_2,\ldots,j_{n-k-1}=0}^{1} \left(\frac{1+r}{2}\right)^{n-k-\sum_{l=0}^{n-k-1} j_l} \left(\frac{1-r}{2}\right)^{\sum_{l=0}^{n-k-1} j_l} \left|0\overset{k}{\cdots}0 j_0 \ldots j_{n-k-1}\right\rangle\!\left\langle 0\overset{k}{\cdots}0 j_0 \ldots j_{n-k-1}\right| \\ &= \sum_{j=0}^{2^{n-k}-1} \left(\frac{1+r}{2}\right)^{n-k-S(j)} \left(\frac{1-r}{2}\right)^{S(j)} |j\rangle\langle j| \end{aligned} \tag{C.1}$$

where $S(j) = \sum_{l=0}^{n-1} j_l = \sum_{l=0}^{n-k-1} j_l$ counts the number of ones in the n-bit binary number $j = (j_0 \ldots j_{n-k-1} 0\overset{k}{\ldots}0)$. In other words, $S(j)$ counts the number of states $|1\rangle$ in each state $|j\rangle = |j_0\rangle \cdots |j_{n-k-1}\rangle\, |0\rangle^{\otimes k}$. It is important to see that ρ_k has a block form with only the first 2^{n-k} diagonal terms different from zero. This fact allows a straightforward calculation of the operators in the measure as seen in Appendix C.1.1. Their exact form is

From the expression of the states in eq. (C.1), the form of the measures can be simplified by seeing that the only contribution to the success probability comes from the trace of the product $\Pi_k \rho_k$. Writing Π_k in the computational basis in the most general form, we have

$$\mathrm{tr}[\Pi_k \rho_k] = \mathrm{tr}\left[\sum_{i,j} \pi_{ij}^k \rho_{jj}^k |i\rangle\langle j|\right] = \sum_{j=0}^{2^k-1} \pi_{jj}^k \rho_{jj}^k \tag{C.2}$$

So the success probability only depends on the diagonal elements of the operators Π_k, consequently we can make all the other components identically 0 since they do not have any effect on the final result and express them as

$$\Pi_k = \sum_{j=0}^{2^{n-k}-1} \pi_j^k |j\rangle\langle j| \qquad k = 1, \ldots, n \tag{C.3}$$

with $\pi_j^k \geq 0\ \forall j, k$ as restricted by the problem.

We also have that the observables Π_j must satisfy the completeness relation (1.5a). The only operator which is not partially filled with zeros is Π_0, the next one Π_1 has only 2^{n-1} non-vanishing variables on the first half of the diagonal, Π_2 has 2^{n-2} and so on. In general, the Π_j operator has 2^{n-j} non-vanishing elements starting from the top of the diagonal. Thus, the completeness relation induces 2^n equations

$$\sum_{j=0}^{n-k} \pi_{i_k}^j = 1 \qquad i_k = 2^{k-1}+1, \ldots, 2^k \quad \forall k = 1, \ldots, n \tag{C.4}$$

When $k = n$, only Π_0 takes part in the sum so $[\pi_0]_{ii} = 1$ for $i = 2^{n-1}+1, \ldots, 2^n$. The next one, $k = n-1$, has two terms in the sum corresponding to the operators Π_0 and Π_1: $\pi_i^0 + \pi_i^1 = 1\ i = 2^{n-1}+1, \ldots, 2^n$. The exact value for the two variables can be evaluated by maximising the success function over them, indeed, their contribution to the probability is proportional to $f_i(x_0, x_1) = \rho_i^0 \pi_i^0 + \rho_i^1 \pi_i^1$. This problem is a simple SDP that can be analytically solved and gives that the maximum, under the constraints, is attained when $\pi_1 = 1$ and $\pi_0 = 0$ since $\rho_i^1 \geq \rho_i^0\ \forall i$ and $\forall t$. This process can be done for all k but, at the end of the day, one obtains that the operators are separable in the subspaces $(\mathcal{H}_2)^{\otimes n}$ and have the form as shown in eq. (5.7).

Yet, in order to be sure that this is indeed optimal, we need to check the Holevo condition (2.8). The Lagrange operator associated to the measure (5.7) reads

$$\Gamma = \frac{1}{n+1}\,|0\rangle\langle 0| + \frac{1}{n+1}\sum_{j=1}^{2^n-1}\left(\frac{1+r}{2}\right)^{\lfloor \log_2 j\rfloor+1-S(j)}\left(\frac{1-r}{2}\right)^{S(j)}|j\rangle\langle j| \tag{C.5}$$

where $\lfloor x\rfloor$ rounds x to the greatest integer less than or equal to x.

The operator in eq. (C.5) is hermitian and positive as required for a valid measure. Using the expression for ρ_k in eq. (C.1), we have

$$\begin{aligned}(n+1)\Gamma - \rho_k = &\left[1-\left(\frac{1+r}{2}\right)^k\right]|0\rangle\langle 0| \\ &+\sum_{j=1}^{2^{n-k}-1}\left[\left(\frac{1+r}{2}\right)^{\lfloor \log_2 j\rfloor+1}-\left(\frac{1+r}{2}\right)^{n-k}\right]\left(\frac{1-r}{1+r}\right)^{S(j)}|j\rangle\langle j| \\ &+\sum_{j=2^{n-k}}^{2^n-1}\left(\frac{1+r}{2}\right)^{\lfloor \log_2 j\rfloor-1-S(j)}\left(\frac{1-r}{2}\right)^{S(j)}|j\rangle\langle j|\end{aligned}$$

and since $0 \le (1-r)/2 \le 1$ and $0 \le (1+r)/2 \le 1$, we automatically see that the elements in the first and third line are positive. The elements in the second line will be non-negative if $n-k \ge \lfloor \log_2 j\rfloor + 1$, because the logarithm is a monotonous continuous function, the maximum value it can take in the sum is $\lfloor \log_2(2^{n-k}-1)\rfloor = n-k-1$ and the condition is satisfied identically. Thus, the Holevo condition holds, concluding that the POVM formed by the operators in eq. (5.7) is optimal and eq. (5.6) gives the maximum success probability.

There are only two special cases at the poles of the Bloch sphere, where $\widetilde{\rho}$ is a pure state. At this points, the state after the change is either indistinguishable ($r=1$) or orthogonal ($r=-1$) and so, perfectly distinguishable. The measure in both cases is degenerate, there is another POVM, apart from the one found in eq. (5.7), that achieve the same success probability and it is optimal. In the first case, $r=1$, the measure is just given by $\Pi_n = \mathbb{I}$ with the others set to $\Pi_k = \mathbb{0}_{2^n}$ for $k = 0,\dots,n-1$. Clearly, the SDP is telling us not to waste any effort in identifying the other states as, effectively, there is only one possible outcome. In the second case, $r=-1$, the states are orthogonal $\rho_j\rho_k = \rho_k\delta_{jk}$ and the measure is given by the projection onto each of them

$$\begin{cases}\Pi_j = |0\rangle\langle 0|^{\otimes j}\otimes|1\rangle\langle 1|^{\otimes(n-j)} & \forall j = 0,\dots,n-1\\ \Pi_n = \mathbb{I}_{2^n} - \sum_{j=0}^{n-1}\Pi_j\end{cases} \tag{C.6}$$

The operators $\{\Pi_j\}_{j=0}^n$ can also be written, as well as the respective state ρ_j, as $\Pi_j = \left|2^j-1\right\rangle\!\left\langle 2^j-1\right|$ in the computational basis $\{|j\rangle\}_{j=0}^{2^n-1}$. These constitute a POVM with Lagrange operator $\Gamma = \sum_k \xi_k\rho_k$ that trivially satisfies the Holevo condition.

The results of the SDP are presented in fig. 7. The numerical values fit perfectly over the analytical line for all n. The primal and dual solutions are the same as expected from Theorem 1 for an optimal solution. In addition, the increase in the number of qubits makes the slope tend to the limiting value of $-1/2$, which is never reached.

Numerical solution Before computing the solution of the problem, we can further simplify its declaration to reduce the number of operations. As we saw, only the diagonal element of Π_j and ρ_k take part in the computation, which are all positive and non-negative. By identifying $\mathrm{diag}(M)$ as the column vector containing the diagonal elements of matrix M, we define the following objects

$$X = \Big(\mathrm{diag}(\Pi_0)\quad \mathrm{diag}(\Pi_1)\quad \cdots\quad \mathrm{diag}(\Pi_n)\Big) \tag{C.7}$$

$$A = \Big(\mathrm{diag}(\rho_0)\quad \mathrm{diag}(\rho_1)\quad \cdots\quad \mathrm{diag}(\rho_n)\Big)^t \tag{C.8}$$

$$B = \mathrm{diag}(\mathbb{I}_{2^n}) = \mathbf{1}_{2^n} \tag{C.9}$$

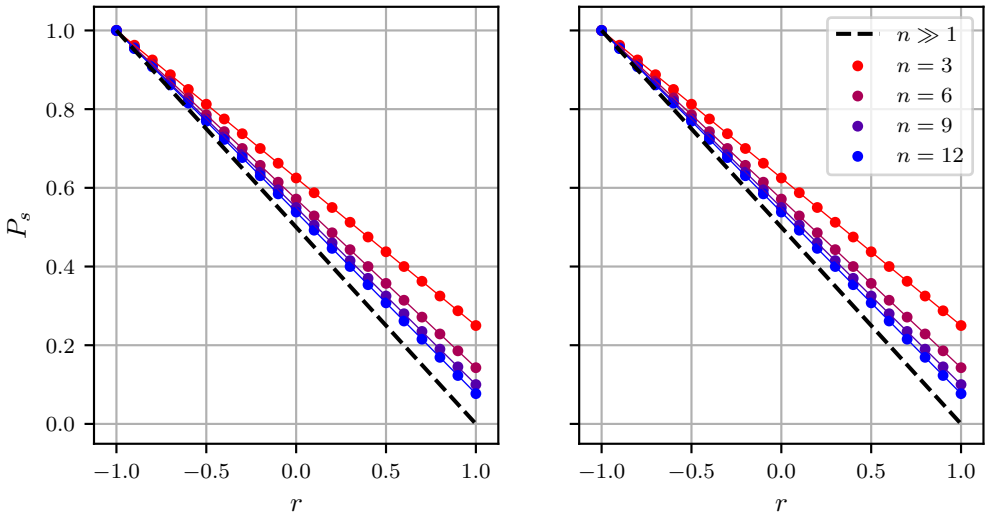


Figure 7: Numerical solution to the QCP problem between a pure and mixed state along the diameter.

where $\mathbf{1}_d$ is a vector full of ones of dimension d. The primal map associated to the problem can be expressed as

$$\Phi[X] = X\mathbf{1}_{n+1} = \sum_{k=0}^{n} X_k \tag{C.10}$$

which sums, for each row, all the columns.

Then, the triplet (A, B, Φ) as defined in eqs. (C.8) to (C.10) composes the primal and dual SDP problem

$$\begin{aligned} &\underline{\text{Primal}} && &&\underline{\text{Dual}} \\ &\max_X \ \frac{1}{(n+1)}\operatorname{tr}(XA) && &&\min_{\boldsymbol{y}} \ \|\boldsymbol{y}\|_1 \\ &\quad X\mathbf{1}_{n+1} = \mathbf{1}_{2^n} && &&\quad \boldsymbol{y} \geq \frac{1}{n+1}\operatorname{diag}(\rho_k) \quad \forall k \\ &\quad X_{jk} \geq 0 \quad \forall j,k \end{aligned} \tag{C.11}$$

In the definition of the dual, $\boldsymbol{y}$ contains the diagonal elements of the Lagrange operator (2.9). Also, there is no need to impose the positivity condition $\boldsymbol{y} \geq 0$ as it is implicit in the previous one, just because the diagonal elements of ρ_k are positive. The objective function, in complete analogy with eq. (3.4), should be $\boldsymbol{y} \cdot \mathbf{1}_{2^n}$ but we have simplified this by noting that this product sums all the elements of $\boldsymbol{y}$ and since the last condition in eq. (C.11) imposes that all the components of the vector must be positive: $\boldsymbol{y} \cdot \mathbf{1}_{2^n} = \sum_i y_i = \sum_i |y_i| = \|\boldsymbol{y}\|_1$.

C.1.2 To any mixed state

The inflexion point as a function of θ is evaluated by equating the second derivative of P_s with respect to θ to 0,

$$\begin{aligned} \frac{\partial P_s}{\partial \theta} &= \frac{nr\sin\theta}{2(n+1)\sqrt{1+r^2-2r\cos\theta}} \\ \frac{\partial^2 P_s}{\partial \theta^2} &= \frac{nr\cos(\theta)}{2(n+1)\sqrt{1+r^2-2r\cos\theta}} - \frac{nr^2\sin^2(\theta)}{2(n+1)\left(1+r^2-2r\cos\theta\right)^{3/2}} \end{aligned}$$

Thus, the solutions to $\partial^2 P_s/\partial\theta^2 = 0$ are

$$\theta = \pm\arccos r \qquad \& \qquad \theta = \pm\arccos\frac{1}{r}$$

The second is not valid because arccos is not defined for $1/r \in [1, \infty)$, therefore the only possibilities are $\theta_{inf} = \pm\arccos r$. The choice of the sign will depend on the concavity of P_s before and after the

change, indeed we know that it must go from positive to negative, which shows that the correct sign is the positive. This makes sense from the same structure of the graphs, for $r \to 0 \Rightarrow \theta_{inf} \to \pi/2$ and for $r \to 1 \Rightarrow \theta_{inf} = 0$, i.e. the concavity of P_s doesn't change in this range.

The minimum on r can be evaluated in a similar way, the first derivative of eq. (5.14) gives

$$\frac{\partial P_s}{\partial r} = \frac{n(2r - 2\cos\theta)}{4(n+1)\sqrt{1 + r^2 - 2r\cos\theta}}$$

which vanishes at $r_{min} = \cos\theta$. In fact, both solutions are the same, the minimum in the probability appears when r and θ satisfy the relation $r = \cos\theta$.

The limit $n \gg 1$ is simply the term multiplying n in eq. (5.12),

$$P_s \approx \frac{\|\widetilde{\boldsymbol{r}} - \boldsymbol{r}\|_2}{2} \tag{C.12}$$

Figure 10 as well as fig. 3 show the numerical solutions together with this approximation. As said in the main text, they where obtained by imposing a precision of 10^{-6} in the probability. After doing the experiment, this was seen to happen before 10 iterations, see fig. 8, the error between the first and second iterations is large as the initial guess is far from being optimal but just with the second iteration the relative error is not larger than 0.1% for $n = 3$. Shortly, after the 6-th iteration, the error becomes smaller than 10^{-6} as required and the convergence of the method is proved. Furthermore, the convergence is seen to be exponentially fast. A linear regression with the data in fig. 8 show that the error e decays with the iterations i as

$$e \approx 10^{-\left(\frac{i}{a} + b\right)} \tag{C.13}$$

with $a = 2.15 \pm 0.23$ and $b = 2.68 \pm 0.17$. This tells us that the iterative method achieves a new digit of precision every two iterations. Therefore, to achieve a numerical accuracy up to the 6-th decimal position, the number of iterations should be at least seven. However, we will leave a standard of 10 iterations in order to reduce as maximum as possible this error, more iterations would not make sense as we will start having errors due to the floating point precision of the computer used.

A final test was made by running the SDP program for $n = 3$. Comparing those values (see fig. 9) with the ones in fig. 10a showed a discrepancy of at most a 0.02%, validating the results obtained through the iterative method.

C.2 Mixed vs. Mixed

C.2.1 Along the diameter

The expression for the states is a little bit tricky to find. First, taking as reference eq. (C.1), we can write the expression for the two separate parts of each $\rho_k = \rho^{\otimes k} \otimes \widetilde{\rho}^{\otimes(n-k)}$,

$$\rho^{\otimes k} = \sum_{i=0}^{2^k-1} \left(\frac{1+t_i}{2}\right)^{k-S(i)} \left(\frac{1-t_i}{2}\right)^{S(i)} |i\rangle\langle i|$$

$$\widetilde{\rho}^{\otimes(n-k)} = \sum_{j=0}^{2^{n-k}-1} \left(\frac{1+t_f}{2}\right)^{n-k-S(j)} \left(\frac{1-t_f}{2}\right)^{S(j)} |j\rangle\langle j|$$

where $S(i)$ is a before the number of ones in i. Then, the tensor product of both parts gives

$$\begin{aligned}\rho_k &= \sum_{i=0}^{2^k-1} \sum_{j=0}^{2^{n-k}-1} \left(\frac{1+t_i}{2}\right)^{k-S(i)} \left(\frac{1-t_i}{2}\right)^{S(i)} \left(\frac{1+t_f}{2}\right)^{n-k-S(j)} \left(\frac{1-t_f}{2}\right)^{S(j)} \left|i2^{n-k}+j\right\rangle\!\left\langle i2^{n-k}+j\right| \\ &= \sum_{x=0}^{2^n-1} \left(\frac{1+t_i}{2}\right)^{k-S_{n-k}^n(x)} \left(\frac{1-t_i}{2}\right)^{S_{n-k}^n(x)} \left(\frac{1+t_f}{2}\right)^{n-k-S_0^{n-k}(x)} \left(\frac{1-t_f}{2}\right)^{S_0^{n-k}(x)} |x\rangle\langle x|\end{aligned}$$

Defining $S_a^b(x) \equiv \sum_{l=a}^{b-1} x_l$ as the partial sum of the binary number x. The previous expression is similar to the one shown in the main text but not exactly, to go from this one to the other we need

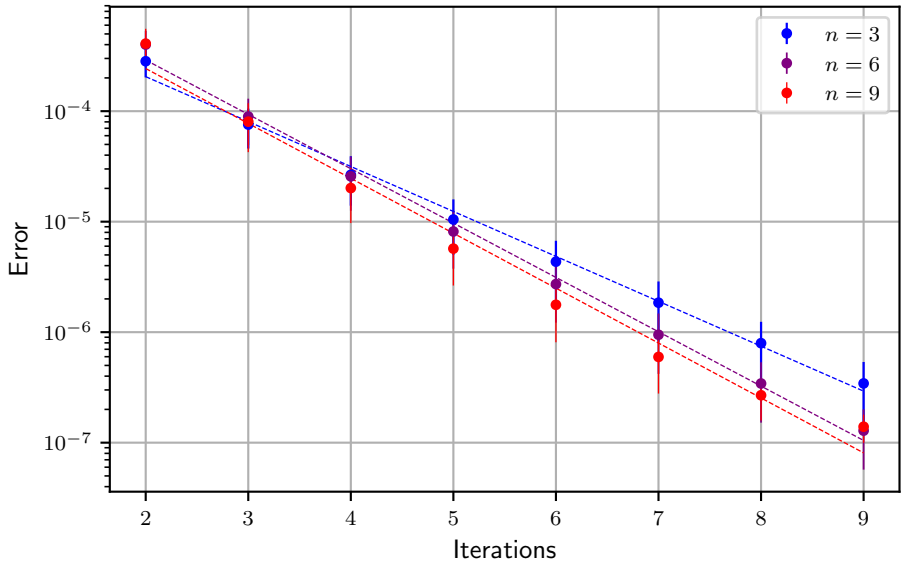


Figure 8: Convergence of the iterative algorithm for the pure to mixed case within 10 iterations (a logarithmic scale is used for the y axis). The dashed lines represent the linear fit of the corresponding values.

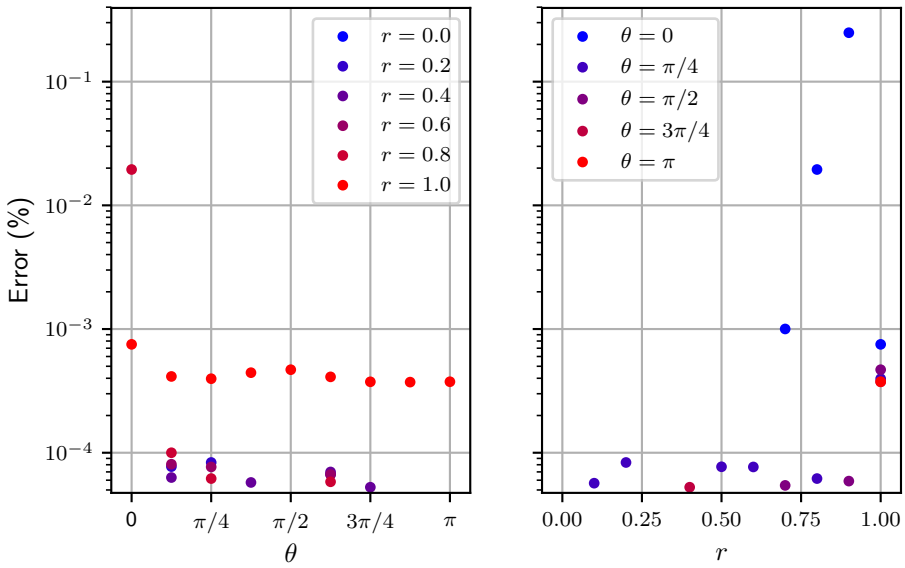
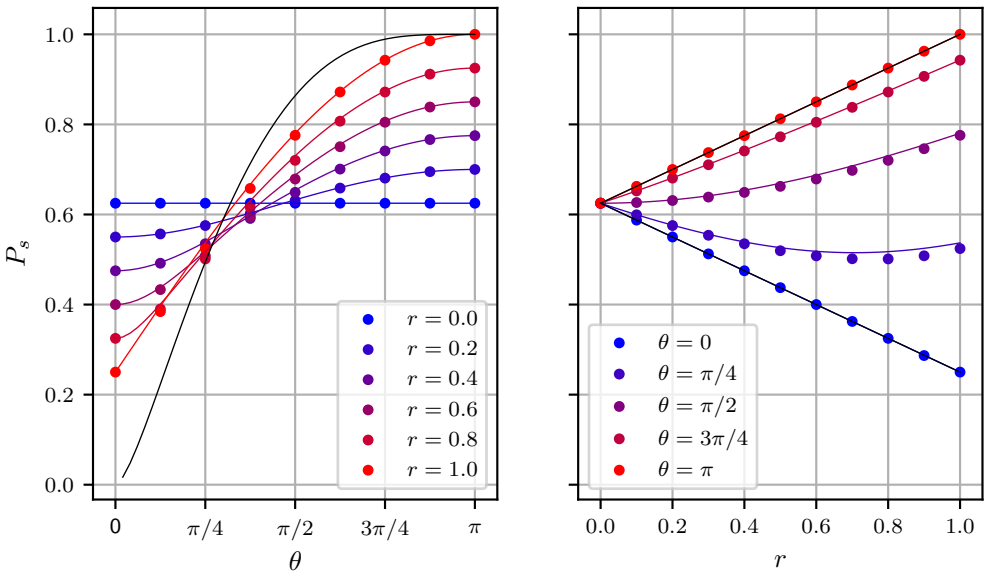
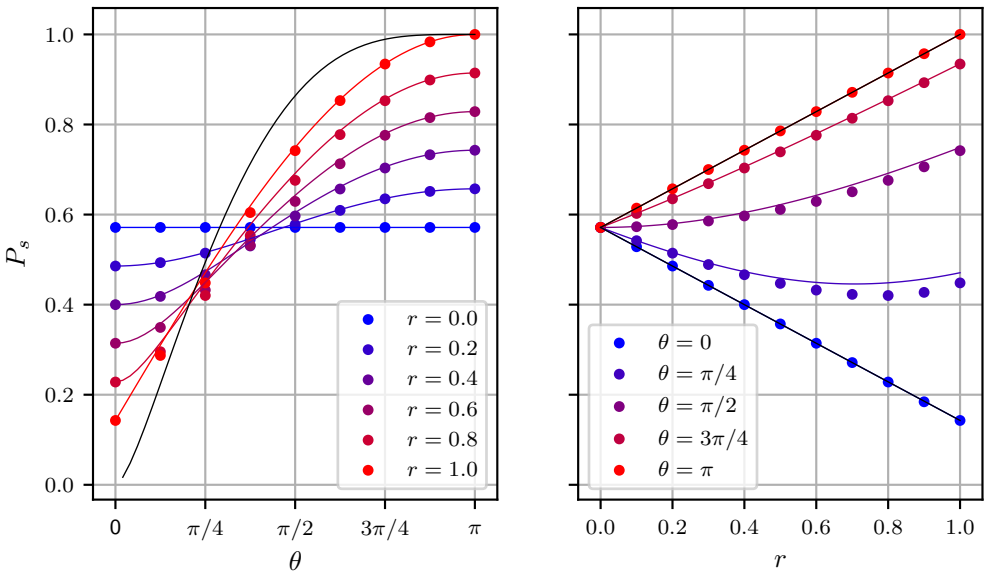


Figure 9: Relative error between SDP and the iterative method for $n = 3$ in the case of a pure going to any mixed state.



(a) $n = 3$



(b) $n = 6$

Figure 10: Numerical solution (dots) for various n in the pure to mixed case together with the Helstrom bound (solid lines).

to take into account that the sum is over all $x \in \{0,1\}^n$. Therefore, we can take the last k bits to the beginning of the sequence and define $x' = i + 2^k j$ in such a way that $S_0^{n-k}(x) = S_k^n(x')$ and $S_{n-k}^n(x) = S_0^k(x')$. After performing this change and relabelling x' as x, eq. (5.18) is obtained.

We seek to find the probability of having a maximum at some position of the random walk which we have seen that is equivalent to finding the probability of a RW not crossing and not returning to the origin. The exact form corresponds to the sum over all paths having a maximum in k, that is, for a fixed number of up steps there are $N_k^{>0}(0, u-d)$ paths and the probability for this is given by eq. (B.7). After the change has occurred, we only impose that the walker should not go further than this maximum, by the mirror theorem, the number of such paths is the same as the number of paths that cross the maximum starting at $b+1$ and ending at some point $c \leq b$ in $n-k-1$ steps. The sum is then taken for all the change point positions and all the possible maximums. Unfortunately, there is no analytical solution for this but we can work out in the limit of a large number of steps for a symmetric walker.

Symmetric Take the right hand part, the steps of the walker are distributed according to $\mathscr{B}(1-p) = \{1-p, p\}$. Note that the no crossing probability is 1 minus the probability of crossing. Denote by T_{jk} the time (number of steps) it takes to go from j to k. Then, $P[T_{01} < \infty]$ gives the probability of crossing in a finite time which is given by

$$P[T_{01} < \infty] = (1-p) + pP[T_{-11} < \infty] = (1-p) + pP[T_{01} < \infty]^2 \tag{C.14}$$

The first part takes into account the probability of going up (and crossing) in one step and the second considers the case that the walker goes down in one step times the probability of going from -1 to 1 in finite time. This last value is just the probability of going from $-1 \to 0$ and then from $0 \to 1$, i.e. two times that of going from $0 \to 1$.

The previous relation is just a quadratic equation whose solutions are

$$P[T_{01} < \infty] = \begin{cases} 1 & p \geq 1/2 \\ p/(1-p) & p < 1/2 \end{cases} \tag{C.15}$$

which is to say that a walker biased in going up will eventually pass through the origin, even in the case of a symmetric walker with $p = 1/2$. We have skipped a lot of formalities in this derivation, for a complete explanation I refer to Feller [1968, Chapter XIV, Section 2] in terms of *difference equations* or to Grimmett and Stirzaker [2001, Chapter 5, Section 3] using the generating function of the Bernoulli distribution.

Take now the left hand part, this walker follows the Bernoulli distribution $\mathscr{B}(p) = \{p, 1-p\}$ but is restricted to be always on the negative side without ever returning to the origin, otherwise it would mean that there is a maximum before this one. As before, the probability of not returning is 1 minus the probability of returning, which by eq. (C.15) is

$$P[T_{00} < \infty] = pP[T_{10} < \infty] + (1-p)P[T_{-10} < \infty] = p\frac{1-p}{p} + (1-p) = 2p = 1 - |2p-1| \tag{C.16}$$

where the last equality follows when considering the two cases $p \geq 1/2$ and $p < 1/2$.

Putting together the results in eqs. (C.15) and (C.16), considering $p > 1/2$, we end up with

$$P_s \approx (1 - P[T_{00} < \infty])\,(1 - P[T_{01} < \infty]) = \frac{(2p-1)^2}{p} \tag{C.17}$$

Finally, using $p = (1+r)/2$ we obtain eq. (5.26).

C.2.2 Equally mixed states

We do not provide the exact expression for the success probability for this case because it is not illustrative given the poor approximation that it gives to the numerical solution. In any case, it would be evaluated from eq. (5.27) with the corresponding expressions for ρ and $\widetilde{\rho}$.

The success probability for $n = 3$ and $n = 6$ can be seen in fig. 13. In the case $n = 3$, the values are compared to those obtained via SDP. Figure 11 shows the relative error between the two values which doesn't go over a 0.1%, again validating the results of the iterative algorithm.

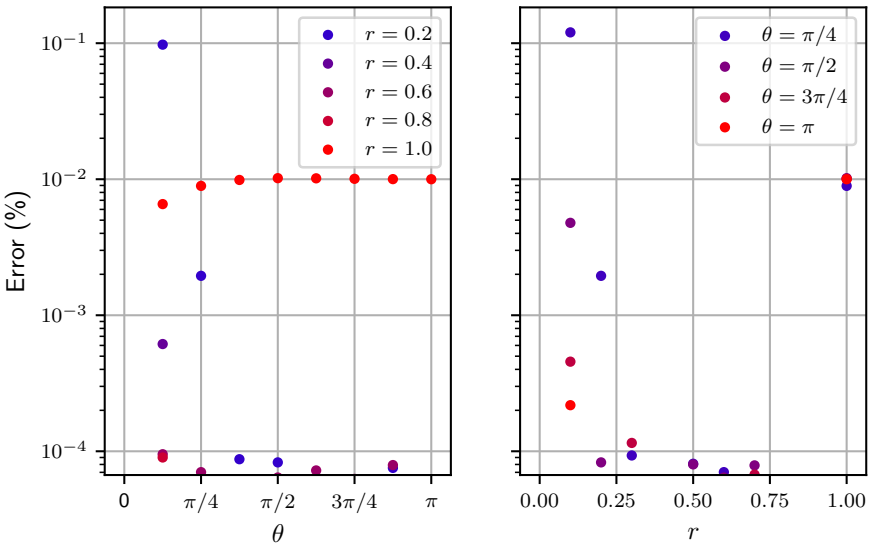


Figure 11: Relative error between SDP and the iterative method for $n = 3$ in the case of equally mixed states.

In a similar way as we did previously, we will compute the error between iterations. The data is shown in fig. 12 with the linear fit. The error decays exponentially fast with the iterations, a linear fit shows that the dependence on the number of iterations decays like in eq. (C.13) with $a = 4.069 \pm 0.020$ and $b = 0.798 \pm 0.096$. As compared to the pure to mixed, the decay rate is two times small, that is, we need to double the number of iterations to achieve the same precision as before.

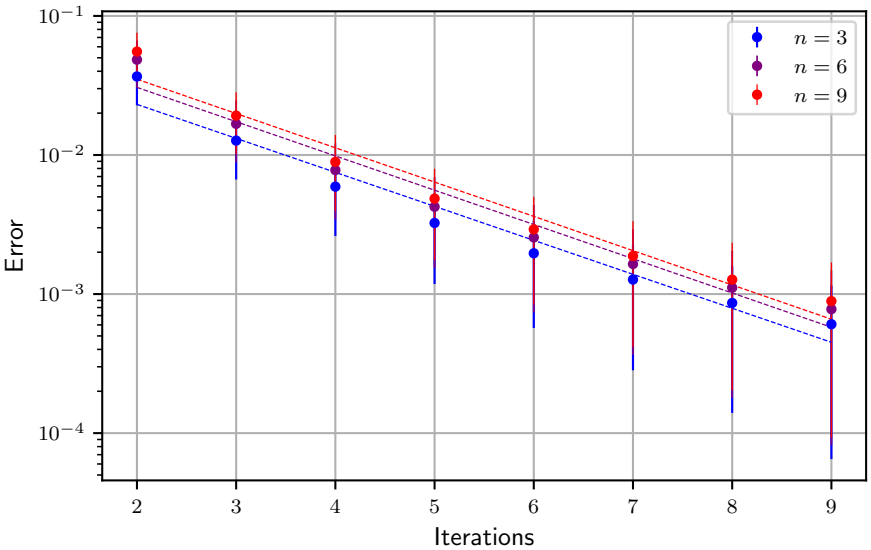
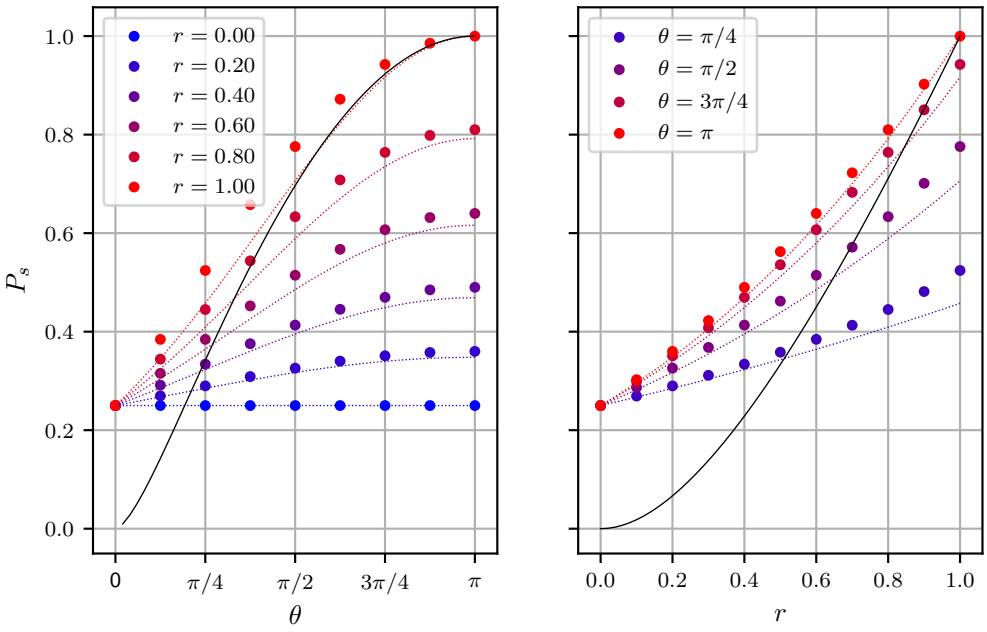
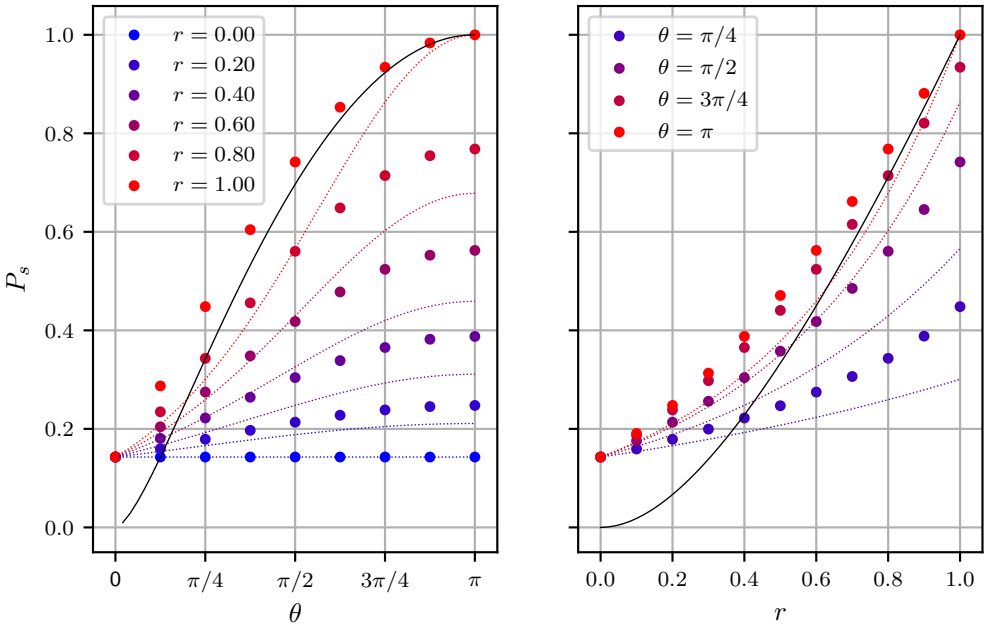


Figure 12: Convergence of the iterative algorithm for the mixed to mixed case within 10 iterations (a logarithmic scale is used for the y axis). The dashed lines represent the linear fit of the corresponding values.



(a) $n = 3$



(b) $n = 6$

Figure 13: Numerical solution (dots) for various n in the equally mixed case together with the approximate solutions (dashed lines).