Quantum detection in time

Change Point and Sequential Analysis

by
Esteban Martínez Vargas

This page is intentionally left blank

Universitat Autònoma de Barcelona

Quantum detection in time

Change Point and Sequential Analysis

by
Esteban Martínez Vargas

under the supervision of

Prof. Ramón Muñoz Tapia
and
Dr. Gael Sentís Herrera

A thesis submitted in partial fulfillment for the degree of
Doctor of Philosophy in Physics

Física Teòrica: Informació i Fenòmens Quàntics
Departament de Física, Facultat de Ciències

Bellaterra, 11 of February 2022

Resum

Aquesta tesi tracta sobre detectar propietats de conjunts d'estats quàntics ordenats en el temps. No obstant, no tracta sobre la mesura del temps, sinó sobre la detecció de un senyal fet d'estats quàntics. S'aborden aquests problemes estudiant com extreure informació d'un senyal que pressuposem que té ordre, anomenem temps a aquest ordre. Sent més precisos, aquesta tesi tracta sobre dos temes principals en estadística quàntica: el punt de canvi i l'anàlisi seqüencial. El problema de punt de canvi tracta sobre un senyal que canvia en un cert temps que és desconegut a l'observador. La tasca llavors és detectar un canvi abrupte de la manera més precisa i ràpida possible. Aquest és un camp destudi en estadística clàssica el que només recientement ha estat analitzat en un escenari quàntic. Aquí estenem l'estudi d'aquest camp en sistemes quàntics. Primer analitzem la versió exacta del problema, que correspon a discriminació no ambigua però restringint-nos a estratègies de mesura de partícules una a una en oposició a estratègies quàntiques més generals. Trobem que per a un rang de valors del solapament entre l'estat inicial i el mutat, l'estratègia en línia té la mateixa probabilitat èxit que lestratègia global. Després estudiem protocols nous que són útils per a aquest problema general, fan ús explícit de l'ordre dels estats en el temps. El protocol que proposem permet interpolar entre els protocols unambigu i d'error mínim.

La segona part d'aquesta tesi tracta sobre contrast d'hipòtesis. Estudiem protocols seqüencials de contrast d'hipòtesis per a sistemes quàntics. Això representa un angle nou al problema de contrast d'hipòtesis quàntic ja que fixem en el nostre enfoc les taxes d'error que volem implementar i prenem el nombre de còpies com a variable aleatòria, cosa que contrasta amb l'esquema usual on es té un nombre fix de còpies i es minimitza l'error. Estudiem cotes

inferiors en el rendiment que es pot assolir quan es vol distingir dos estats quàntics i qualsevol estratègia quàntica de mesurament permesa. El rendiment en aquest cas està donat pel mínim nombre promig de còpies necessàries per obtenir una decisió per una hipòtesi amb els límits d'error que es demanen. Ens restringim al cas més senzill de dues hipòtesis quàntiques i estats barreja de dimensió finita. També estudiem el cas de dos estats purs i obtenim resultats pel nombre mitjà òptim de còpies necessàries. Els nostres resultats suggereixen de forma natural l'estudi de protocols de discriminació amb estats purs en línia més enllà del cas binari. Estudiem el problema de discriminació unambigua per a tres estats simètrics que és un case molt natural i simple però on les estratègies en línia, en general, tenen un rendiment diferent que el global quan més duna còpia és disponible. No obstant determinarem els casos pel que el rendiment és el mateix.

Acabem aquesta tesi a les conclusions amb alguns pensaments sobre els tòpics presentats i en general sobre el camp d'informació quàntica i la relació amb els fonaments de la física quàntica. Incloem també noves línies de recerca prometedores que deriven dels resultats d'aquesta tesi.

Resumen

Esta tesis trata sobre detectar propiedades de conjuntos de estados cuánticos ordenados en el tiempo. Sin embargo, no trata sobre medir el tiempo, sino sobre la detección de una señal hecha de estados cuánticos. Se abordan estos problemas estudiando como extraer información de una señal que presuponemos que tiene orden, llamamos tiempo a dicho orden. Siendo más precisos, esta tesis trata sobre dos temas principales en estadística cuántica: punto de cambio y análisis secuencial. El problema de punto de cambio trata sobre una señal que cambia en cierto tiempo que es desconocido al observador. La tarea entonces es detectar un cambio abrupto de la manera más certera y rápida posible. Este es un campo de estudio en estadística clásica y recientemente ha sido primeramente analizado en el caso cuántico. Aquí extendemos el estudio de este campo en sistemas cuánticos. Primero analizamos la versión exacta del problema, que corresponde a discriminación no ambigua pero restringiéndonos a estrategias de medida de partículas una a una en oposición a estrategias cuánticas más generales, encontramos que para un rango de valores del traslape entre el estado inicial y el mutado, la estrategia en línea tiene la misma probabilidad de éxito que la estrategia global. Después estudiamos protocolos novedosos que son útiles para este problema general, hacen uso explícito del orden de los estados en el tiempo. El protocolo que proponemos permite interpolar entre los protocolos unambiguo y de error mínimo.

La segunda parte de esta tesis trata sobre contraste de hipótesis. Estudiamos protocolos secuenciales de contraste de hipótesis para sistemas cuánticos. Ésto representa un nuevo ángulo al problema de contraste de hipótesis cuántico ya que en nuestro enfoque fijamos las tasas de error que queremos implementar y tomamos el número de copias como una variable aleatoria, lo que contrasta con

el esquema usual donde se tiene un número fijo de copias y se minimiza el error. Estudiamos cotas inferiores en el rendimiento que puede alcanzarse cuando se quiere distinguir dos estados cuánticos y cualquier estrategia cuántica de medición es permitida. El rendimiento en este caso está dado por el mínimo número promedio de copias que necesarias para obtener una decisión por una hipótesis con los límites de error que se piden. Nos restringimos al caso más sencillo de dos hipótesis cuánticas y estados mezcla de dimensión finita. También estudiamos el caso de dos estados puros y obtenemos resultados para el número promedio óptimo de copias necesarias. Nuestros resultados sugieren de forma natural el estudio de protocolos en línea de discriminación con estados puros más allá del caso binatio. Estudiamos el problema de discriminación unambigua para tres estados simétricos que es un case muy natural y simple pero donde las estrategias en línea, en general, tienen un rendimiento distinto que el global cuando más de una copia es disponible. Sin embargo, determinamos los casos en los que el rendimiento es el mismo.

Terminamos esta tesis en las conclusiones con algunos pensamientos sobre los tópicos presentados y en general sobre el campo de información cuántica y sobre su relación con los fundamentos de la física cuántica. Incluímos también nuevas líneas de investigación prometedoras que se derivan de los resultados de esta tesis.

Abstract

This thesis is about detecting properties of sets of quantum states ordered in time. However, it is not about measuring time but about the detection of a signal made of quantum states. We address these issues studying how to extract information from a signal that we presupose that has order; we call time such order. Being more specific, this thesis deals with two major themes in quantum statistics: change point and sequential analysis. Change point analyzes with a signal that changes abruptly at a certain time which is unknown to the observer. The task then is to detect the abrupt change as accurate and fast as possible. This is a field of study in classical statistics and was first analyzed in the quantum setting recently. Here we extend the study of this field for quantum systems. We address firstly the exact version of the problem, which corresponds to unambiguous discrimination but restrict to measuring copies as they are available as opposed to more general quantum strategies. We find that for a given range of the values of the overlap between the initial and the mutated state, the strategy that we study can reach the performance of the global one. We then study novel protocols that are useful in this problem in the more general setting, they make explicit use of the ordering of the states in time. The protocol that we propose interpolates between unambiguous and minimum error discrimination.

The second part of the thesis turns to hypothesis testing. We study sequential hypothesis testing protocols for quantum systems. This represents a novel approach to quantum hypothesis testing because in our approach we fix the error rates that we want to reach and take the number of copies as a random variable, which contrasts with the usual scheme of having a fixed number of copies and minimize the error rates. We study lower bounds on

the performance that can be achieved when trying to distinguish two quantum states and any quantum measurement strategy is allowed in the limit of small errors. The performance in this case is given by the minimum average number of samples that one needs to achieve a decision for one hypothesis with the error thresholds that are asked for. We restrict to the case of two quantum hypotheses represented by finite dimensional mixed states. We also study the case of two pure states and obtain exact results for the optimal average of samples needed. Our results naturally suggested the study of unambiguous online protocols of discrimination with pure states beyond the binary case. We studied the problem of unambiguous discrimination of three symmetric states, which is a very natural simple case of three hypotheses but where online strategies have a lower performance than global ones when more than one copy is available. Nevertheless, we determined the cases where the performance is the same for both cases.

We finish this thesis in the conclusions with thoughts about the topics presented and in general about the field of quantum information and its relation with quantum foundations. We also include promising new lines of research derived from the results of this thesis.

This page is intentionally left blank

Declaration

I declare that the thesis has been composed by myself and that the work has not been submitted for any other degree or professional qualification. I confirm that the work submitted is my own, except where work which has formed part of jointly-authored publications has been included. My contribution and those of the other authors to this work have been explicitly indicated below. I confirm that appropriate credit has been given within this thesis where reference has been made to the work of others.

List of Publications

- G. Sentís, E. Martínez-Vargas and Ramon Muñoz-Tapia, ‘Online strategies for exactly identifying a quantum change point’. Phys. Rev. A 98, 052305 (2018)
- E. Martínez-Vargas and Ramon Muñoz-Tapia, ‘Certified answers for ordered quantum discrimination problems’. Phys. Rev. A 100, 042331 (2019)
- E. Martínez-Vargas C. Hirche, G.Sentís, M. Skotiniotis, M. Carrizo, Ramon Muñoz-Tapia and J. Calsamiglia, ‘Quantum Sequential Hypothesis Testing’. Phys. Rev. Lett. 126, 180502 (2021)
- G. Sentís, E. Martínez-Vargas and Ramón Muñoz-Tapia, ‘Online identification of symmetric pure states’. ArXiv:2107.02127

This page is intentionally left blank

Acknowledgements

The PhD is an adventure. My case has not been an exception in this sense. It has been an experience that has changed me profoundly and has opened me horizons that I didn't know before. In the middle, I encountered hardship as anything in life but in the end I look back at the journey and I find it quite an experience.

I am profoundly thankful to Ramón Muñoz Tapia for accepting me as a PhD student and becoming my advisor. He also introduced me to a lot of concepts that I didn't know. I am grateful for his patience with me, his enthusiasm, and of course, his scientific rigour. Ramón introduced me to a former student of his, Gael Sentís, with whom I am very grateful as well. Gael returned to GIQ and afterwards became my other advisor. I have worked closely with both of them and I have learnt a lot about physics and life from them. They have supported me and nourished my way of thinking deeply.

I also want to thank John Calsamiglia Costa and Michalis Skotiniotis for their scientific guidance and their teachings on a lot of topics within physics and not strictly physics. Also, I thank their friendship and support.

I want to thank Anna Sanpera and Andreas Winter for their mentorship and the vivid "sobremesas". They represented a big support for me emotionally and taught me a thing or two about a wide array of interesting topics. I also want to thank Emili Bagan, that although we did not interact much, he always had a positive attitude.

I also want to thank the friendship of several people on the group: Matteo, Abel, Christoph, Zahra, Farzin, María, Carlo, Andreu and Jessica. Special thanks to Philipp and Matías with whom I could share a some more beers.

I want to thank Carlos Pineda and Pablo Barberis whom I worked telemat-

ically in a publication (not included in this thesis) and supported me during the bachelor and master thesis.

I want to thank the Mexican Consejo Nacional de Ciencia y Tecnología (CONACyT) for giving me a scholarship to be able to do my PhD in Universitat Autónoma de Barcelona.

I want to thank to my family, which has always been a fundamental part in my life. They are always present in my heart.

The pandemic forced almost all people to connect via internet. I could reconnect with friends from Mexico. Special thanks to my friends Ernesto and Raúl. I reconnected with the members of the baptist church Jerusalén from Mexico city, who accompained me and have been an important part of my life.

This thesis could not have been made without Grecia Guijarro, who accompained me and shared with me most of the experience of the PhD. I thank her affection and her presence in my life.

This page is intentionally left blank

Contents

“*Conté una extraña experiencia que había tenido junto al Sena, no lejos del Quai St-Michel. Había entrado en una librería que, ya desde sus dos escaparates simétricos, exhibía su esquizofrenia. De un lado, obras sobre computadoras y sobre el futuro de la electrónica, del otro, sólo ciencias ocultas. Y lo mismo en el interior: Apple y Cábala.*

-Increíble-dijo Belbo.

-Obvio- dijo Diotallevi-. O al menos, tú eres el último que debería asombrarse Jacopo. El mundo de las máquinas trata de encontrar el secreto de la creación: letras y números.”

El péndulo de Foucault, Umberto Eco

1

Introduction

1.1 Science and technology

Science and technology have an intertwined history. Science produces knowledge that sometimes is transformed into technological advances. Nevertheless, science needs technology to produce knowledge [AAB+19, Pot21]. We are in a very peculiar moment in this regard in the realm of quantum science as it has mixed a lot with technology [ABB+18]. This relatively new paradigm brings new problems to be thought about. In particular, it is not clear a priori what is exactly the path that this mixture with technology is dictating to science.

There is a current trend to call quantum technologies as part of a "second quantum revolution" [Jae18, Ací16]. This term, from my point of view, comes from two revolutions in human history: the industrial revolution and the discovery of quantum theory, which changed a lot of paradigms in physics. This statement of course, deserves a much larger study but I am just pointing it out. Now, the first one is a revolution for industry and means of production and the second one was a breakthrough in science (could be called the first quantum revolution). The second quantum revolution, being born amidst the development of quantum technologies is more of an industry revolution than a scientific one. This is a strange situation because the existence of this revolution is a claim made mainly by scientists [Uni16], not people from industry (the people in industry however, have become interested in these technologies but *a posteriori*). It is therefore important for science to ask: What is the *scientific* part of the second quantum revolution? I do not think that this question has been seriously addressed. Neither do I claim to have the answer but at least in this introduction I convey my personal view in this matter. I think that the sensation of paradigm shift that brings to mind a revolution is because there is a change in our epistemological approach to nature. To support this claim, I will contextualize quantum technologies as a result of developments in statistics, computation and science.

This serves as a motivation for the thesis. We will focus on the study of the optimal detection of sets of quantum states which are ordered in time. This might sound as an academic question at first glance but it has practical and theoretical implications. For instance, there is a paradigmatic experiment that has this setting: the Young double slit experiment with quantum particles. In that experiment, electrons are shot, one at a time through a double slit and then collide on a screen. This experiment shows a lot of mysterious properties from quantum theory. Feynman even said that the double slit experiment contains "the only mystery" from quantum theory [FLS11]. This sentence suggests that many of the puzzling phenomena from quantum theory, which constitute foundational issues can be revealed in this setting.

1.2 Epistemology and determinism

One of the ideals of enlightenment was to put a system known as modern science as the central epistemological tool to know facts of the world[1]. The mathematician Pierre-Simon Laplace was certainly a debtor of enlightenment's tradition. He made a name of his own in pre-Napoleonic France as a professor in the École Militaire. He made important contributions to astronomy and he published his treatises in Celestial Mechanics [dL99]. One of the triumphs of classical mechanics is that it allows the complete description of a set of particles given initial conditions. It is curious, however, that in order to do so he invented a creature that would remain in the collective imagination to this day. Laplace called into being a "supreme intelligence" that could know everything in the universe at any time [dL14]:

[1] This section is based on Jimena Canales' book "Bedeviled: A Shadow History of Demons in Science" 2020 Princeton University Press.

> An intellect which at any given moment knew all of the forces that animate nature and the mutual positions of the beings that compose it, if this intellect were vast enough to submit the data to analysis, could condense into a single formula the movement of the greatest bodies of the universe and that of the lightest atom; for such an intellect nothing could be uncertain and the future just like the past would be present before its eyes.

Afterwards, this creature received the denomination of a "demon" [Kov15]. This supernatural being was the saint patron of determinism [Day67]. It yielded the idea that every phenomenon was in principle predictable *ad infinitum*. This means, all the information for all future events is already out there somewhere, as the universe is a giant mechanical engine. Before Laplace's demon appearance, the notion of determinism existed in classical mechanics. The prediction of events from initial conditions was in principle unbounded. This intelligence extrapolated the logical consequences of determinism to every possible aspect of science and the universe.

The existence of such intelligence poses a problem for free will and ethics. If everything can be calculated from a past state, there is no place for choosing, as it could have been predetermined by a previous step. Not even crimes would be outside the grasp of the Laplacian demon.

The spirit of determinism crossed the sea into United Kingdom, however, where it originated with Isaac Newton [JS98]. Charles Babbage and Ada Lovelace had in mind an endeavour that had similar objectives than that of Laplace's demon. They wanted to construct a machine that could make calculations mechanically. If it is possible to calculate in principle, then it is a matter of arduous work possibly made by a machine. Although not very rewarding, the work by Babbage and Lovelace became the predecessor of computers.

By the same time Darwin proposed a theory of living beings that related species between them with some dynamics: species came about other species. Although there is randomness involved through mutations, the overall theory regards life in a deterministic way [dar59] as evolution explained a lot of the mysteries in the variety of the fossil records. In a diametrically opposite way from Babbage, who regarded a machine as a being that calculated, Darwin saw the living beings as mechanisms. Evolution can be viewed as a mechanical view of nature. This proved to be a powerful idea that clarified much of biology. However, obviously, the parallelism is not perfect here. Living beings depart much from what then was regarded as a machine. Nevertheless, the universe as a mechanism was a powerful picture that inspired thinking in new directions.

The Laplacian mechanical universe included human affairs, of course. In a groundbreaking book titled *Philosophical Essay on Probabilities* from 1814 [dL14] Laplace introduced statistics into politics and culture. This meant to study human behavior in the aggregate, which inaugurated the quantitative study of society. The tools of statistics could be used to understand phenomena that were complex to grasp by conventional wisdom. Not only the sciences gave man a sense of empowerment over nature but also over itself in society.

The ideal that some being could grasp the events of the universe was so inspiring that approximations to it came into existence with Babbage's calculating engine and by applying the tools of statistics to society. It also inspired somehow evolution theory. Would determinism and therefore, secularism have had the same impact without the allegory of a being? Perhaps it is the supernatural what excited the imagination.

1.3 Statistics and Computers

A new manifestation of Laplace's demon is thriving today. Certainly, the ideal that society's phenomena can be known empirically is a common place. The complex societies of the beginning of the XXI century produce large quantities of data as they keep record of many facts of the world. Also, thanks to a widespread use of the internet, much of these data has a very high degree of availability and is easily extracted from the users. This is the ideal place for a Laplacian demon.

The actual existence of such a being was never a central point for Laplace to make a point and incite the imagination. An ideal is shown and approximations to that ideal can prove some use. An *intelligent being* is not crucial for such a task, non-intelligent machines capable of some finite computational power can exist and have been constructed.

Newton's equations of motion require a precise input to give a precise output. However, precise knowledge of the whole universe is a very big luxury that in practice is not available, not even for a specific given system! Therefore, statistical analysis of human societies was addressed from the standpoint that the perfect knowledge about them was not available. Not because these aspects did not exist in reality, but because a system so complex and diverse as the world has large quantities of knowledge and we are confronted with the reality of limited resources to transform it into information.

This partial knowledge produced an imperfect image of the world that is *good enough* for practical applications. We can define an ideal Laplacian demon as an agent that knows everything about a system at a given time. With this it is not extravagant to think in an approximate Laplacian demon, which knows part of the information contained in a system and has limited computing power.

The ideal of deterministic prediction subjected to limitations persisted in the form of statistics. The XX century produced a large body of research in statistics and its tools. Also, aided by computer machines new methods were available. There was a lot of research devoted to artificial intelligence and statistics of neural networks.

The form of the Laplacian ideal of knowledge nowadays is given by transforming large amounts of data into concise knowledge using computers and statistics. Being more concise, Machine Learning is the designation that several statistical techniques have received [The15, Bar12]. It is a paradigm that works under the presupposition that models of systems can be learned in a controlled loop. There is a learning agent, a computer program, that can be modified after processing data such that in the end it contains a model that describes well enough the observed data. The applications of this techniques has in fact resulted in a very profitable industry of knowledge [Eco17].

A crucial point that I want to stress here is that statistics has been useful historically as an epistemological tool. Many phenomena that are too complex to assess directly can be reduced to few statistical parameters. Not only that, it can help to efficiently extract information from given data. For example, data that is ordered in time can contain a lot of information that can be extracted as e.g. the change of the variance of a distribution through time.

Two statistical fields will be the focus of this thesis: Change Point detection and Sequential Analysis [TNB14]. These are statistical methods developed around the half of the XX century. These methods analyze data that is ordered in time.

Time is a fundamental concept that permeates human understanding [KGW98]. We could dive deep into the philosophical considerations that this entails but in this thesis we will stand somewhat distant. For us, time will be a parameter that *orders* a set of discrete data. Our data can be given at once in an ordered list or can be given to us in separated moments. We can have the whole history of the data upon our eyes or we can see it as it evolves. For example, in Figure (1.1) we have an example of a data series in time. A priori, we can ask different questions about given data. We could address a question of hypothesis testing, that is, of guessing if the data we are given comes from a source with given properties or another possible one. For example, is climate change from human

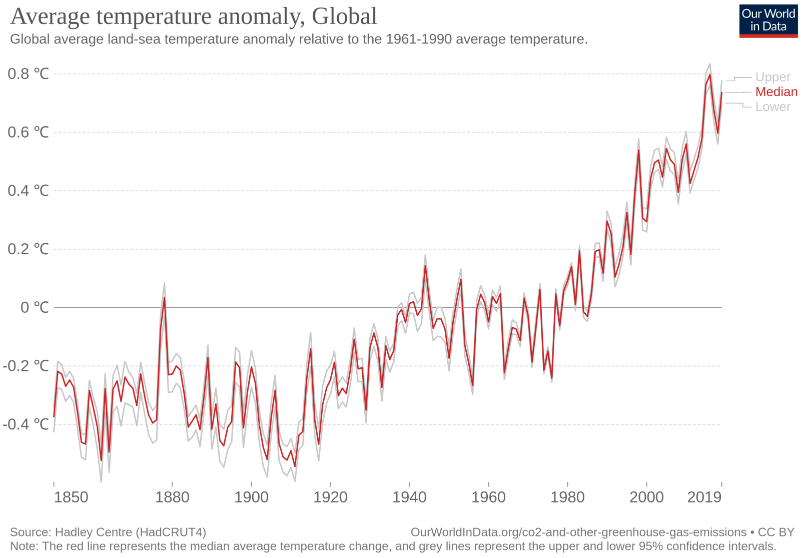


Figure 1.1: Average global temperature anomaly in time.

origin or is it natural?

Perhaps the experience of *change* is as ubiquitous as the experience of time itself. How could one experience the passage of time if it was not because something has changed? Changes can involve large periods of time like those of geology where we see a displacement in the tectonic plates, or fast like an economic crash. What is important is that both require the knowledge of averages in position and configuration. The idea of change point detection is to precisely locate one statistic change in a signal. The basic scheme of detection is that of a base signal that is “stable” and then changes into another stable state. Our task is to detect where this change took place with the highest accuracy, given a series of observations about a system (our signal).

Returning to Figure (1.1), observe that the graph in itself doesn’t tell us much, we have to interpret the data. We can use the baggage from statistics and apply the change point analysis to infer a change in the probability distribution. This means that the source of the random data must have changed somehow. Here we have all the data available as a whole because it was taken through some time and stored.

We can take this series as a whole and infer from it. We can think, however, in a process that is also ordered in time but such that the data becomes available as the phenomenon evolves. This means we are given the data on the fly. If the data is accumulated and processed in batches there can be a difference in the cost function one normally tries to optimize. With quantum resources available there can be an advantage in making a global measurement on the whole quantum data that constitutes a time series. We call the scenario of having all the data at once and allowing global measurements an *offline* scenario and the scenario of being given samples one by one we call it *online*.

The second part of this thesis is concerned with the problem of hypothesis testing. There is a substantial modification of the first part problem as here, the source machine does *not* change, we are given several samples of the same state, we try to figure out which state we are being given. Sequential Analysis is a method of statistical inference that takes place with several samples that are processed as they are collected. Several points should be remarked. First of all, we have an agent trying to make a decision based on observations. This agent has to decide between two (or more) hypotheses based on its observations within a reasonable error. The sequential method introduced by Abraham Wald [Wal73] has the advantage that uses less number of observations on average compared to other methods. It also has the practical advantage of being online, taking a decision right after each observation, in other words, it is an on-the-fly method.

1.4 Quantum science and technology

A major leap in science was taken at the beginning of the XX century: the discovery of quantum mechanics and the development of quantum theory. As far as we know, quantum theory is the most precise description of the world. Quantum mechanics poses a fundamental limit on what can be known by the ideal Laplacian demon. Heisenberg's uncertainty principle impedes the perfect knowledge of the position x and momentum p of a particle:

$$\Delta x \Delta p \geq \frac{\hbar}{2}, \tag{1.1}$$

where Δ denotes uncertainty.

If we use statistical knowledge then one might assume that a source of "noise" like quantum uncertainty (Eq. 1.1) would only make things worse. In a strange turn of events, statistics will be benefited by the same theory that prevents perfect knowledge. In some cases, as in metrology one can get better precision with the same number of resources if they are quantum particles [GLM11, MVPLBB17]: a *quantum* Laplacian demon would surpass its classical counterpart.

Quantum statistics [WM09] is the established theory that arises when quantum systems are used to store and retrieve information. It was born amidst the development of quantum technologies [ABB^{+}18].

Much of the efforts surrounding Quantum Theory have been devoted to the description of the matter itself, its constituents. This is what we know as particle physics and high energy physics. This research yielded several important technological developments. Quantum information turns away from this paradigm and formulates fundamental questions from the viewpoint of practical tasks to fulfill.

One practical task can be developing highly precise detection devices. A large research field involves estimating parameters with quantum systems. Also, very precise clocks can be achieved by measuring the oscillations of an atom, therefore giving rise to atomic clocks. Using generalized quantum measurements one can achieve an overhead in precision with respect to what conventional, classical approaches do.

GPS is a very common technology nowadays that is used in boats and planes to navigate [ftSP18]. However it depends on satellites orbiting Earth and having a good signal with them. Another type of navigation was developed in the XX century, which needs to keep precise track of the changes in momentum of the ship. With an accurate map and clock we can integrate the trajectory of the ship in time. This is called *inertial* navigation and it has the advantage that can operate independently of external signals [Coo11]. This is very convenient for example in submarines where stealth is a very important factor. There are projects to fabricate a quantum inertial navigation system [Lon21] which benefits from the precision that quantum systems can provide. Quantum technology becomes very real in this scenario since a good function of the accelerometer and clock is a matter of life and death for a submarine crew.

Observe that to navigate we would need the information about the inertia of the submarine which is a set of ordered data in time. One would also have precise time detection associated with this. Therefore any information that we can extract from this time series would be beneficial for us. The methods analyzed in this thesis treat precisely with this general scenario: transforming data into knowledge with precise mathematical interpretation in such a way that the error is minimized. Also, it would be beneficial to make the inference of the position *in real time*, that is, as data become available. One of the challenges for using inertial navigation is that it has drift errors which makes it less precise [Coo11]. A gain in precision from quantum measurements for example would be very significant.

The search for precision confronts us with the question of navigating a ship using quantum apparatuses. This might seem an innocent question but it has foundational echoes. This is because quantum theory corresponds to our foundational theoretical understanding of the world. A “quantum submarine” is system whose only means of direction come from quantum detection. Therefore, one may ask how to navigate using “quantum senses”. This is by no means innocent, it goes deep into epistemology, which has been a subject of philosophy since always. The search for precision leaves us with an epistemological question: what can be known when one has at hand *only* quantum systems? This is a guiding question for developing quantum detection technologies. For me the availability of new tools for knowing the world represent an epistemological shift, which is the main topic that I refer to in the first paragraph.

1.5 This thesis

Under this perspective I study here two main problems that deal with the task of getting information using quantum systems. We study in this thesis the quantum version of the Change Point problem and Sequential Analysis. Both problems are related with statistical analysis of time series, however, the questions they address differ from each other.

1.5.1 Change Point

The first section of the thesis deals with change point detection. Basically, the problem is to detect an abrupt change from a set of observations in time in the sampling distribution. The problem can be regarded as a machine that produces samples with a given distribution and then it produces samples of another distribution. In the *classical case* it is very close to the sequential setting, this means, when one has access to samples one at a time. The problem to address classically is that we can have false detection events that can be also called false alarms. Imagine a large sample and we want to detect the change point as fast as possible. If we have an extremely reactive detector in order to detect changes quickly then there would be false alarms frequently. On the other hand, a detector that is too cautious will delay between the time of occurrence of a real change and its detection. So the problem becomes one of finding a detection scheme that minimizes the average delay to detection subject to a constraint on the tolerable frequency of false alarms.

In this thesis we focus on the quantum case, where there can be an advantage if we allow large batches of samples states from a machine and perform a collective measurement. The idea is then to let the machine produce states in a batch and collect it and make a global measurement. The machine prepares states of one type and then it changes and produces states of another type. We are interested in detecting when this change happens. This problem was introduced in [SBC$^+$16] and [SCMnT17]. In this thesis we further addressed this questions in two publications:

1. Optimal online identification of a quantum change point.

The first publication dealing with Change Point (not presented in this thesis) deals with the problem in its simplest form [SBC$^+$16] with no constraints and the other demands exact identification [SCMnT17] of the change point. The latter one is an extremely constrained scheme as we ask for answers with no error. In the first work presented here we study an identification protocol that allows no error and uses strictly *local* measurements. It turns out that, surprisingly one can achieve the same probability of error that the one achieved in the global case, at least in a given regime of the parameters. In fact, our results go further than just local measurements, the global optimal can be achieved with an *online* protocol that is the one we consider here.

2. Certified answers for ordered quantum systems.

In this work we introduce a new kind of discrimination protocol that interpolates between the minimum error and the unambiguous protocols using semidefinite programming (SDP) [BV04, Eld03]. If we define the unambiguous protocol as one that allows no error, our interpolating protocol allows a given number of errors around the true value, but no more. One could say that it gives a *certified answer* when the change point is measured. This means, it gives an answer with a certificate that no more than certain errors are made. Unambiguous discrimination would be a perfect certificate: no errors allowed. Minimum error would be the worst certificate: any error is allowed. This protocol is only possible because the set of states we are dealing with is ordered. We map the complicated SDP of the protocol into a simplified version that reveals in an easier manner valuable information. We obtain an analytical lower bound on the average probability of error. As we mentioned above the order of the samples is important, however, this ordering does not necessarily have to be related to time ordering. We illustrate this fact by addressing another problem different from change point: the anomaly detection [SHCM18] which is also ordered but the ordering is more related to the spatial distribution of the quantum states. In both cases we have a problem with states that are in a graph with a linear open topology (a chain).

1.5.2 Sequential Analysis

The second part of the thesis deals with sequential analysis, an area of research in statistics that was born with the works of Abraham Wald from the half last century [WW48]. We address this problem for the first time its quantum version. We address the simplest problem: try to identify one of two hypotheses from the statistics of the observed samples and stop as soon as given error criteria are met. The approach taken here is different than that of what is found in the quantum statistics literature. The number of samples is considered a random variable, the aim is to minimize the average of this random variable. Basically one defines a statistical test named Sequential Probability Ratio Test (SPRT) and studies its statistical properties. The SPRT precisely minimizes the average number of samples needed to certify an hypothesis within some given error thresholds, it is therefore the optimal sequential test [WW48]. If one compares the SPRT with other tests, for example we get a significantly better performance with respect the Neyman-Pearson test [Wal73]. We derive general lower bounds that can be achieved using quantum protocols.

At the end of the first publication of this section [MVHS$^+$21], a question relating unambiguous discrimination and online measurements was raised. It turns out that the unambiguous protocol achieves the same average number of samples with online measurements than with a global protocol in the zero-error scenario. This fact instigates us to think about the study of unambiguous discrimination of general sources of quantum states. For two hypotheses it is equivalent to use online and global protocols [CY01], however for three hypotheses the question becomes highly nontrivial. We study this case using our knowledge of Gram matrices and semidefinite programming related to previous works. This part contains two publications:

3. Quantum Sequential Hypothesis testing

Here we study the quantum version of sequential analysis by Wald. The idea here is to change the merit function to minimum number of samples needed (as stated above). First we address the SPRT using qubits with a fixed POVM. We observe that using sequential analysis one can get better performance with respect to the average number of samples, when comparing with the optimal deterministic strategy given by the quantum Chernoff bound [ACMnT+07]. Then we address a very general problem of hypothesis testing where we allow a feedback on the information we get from measurements. This also implies that any kind of quantum measurement strategy is allowed with this knowledge: weak measurements, collective, etc. We do not solve this problem exactly as it seems extremely challenging, however, we are able to give an ultimate lower bound on the average number of samples needed in the asymptotic regime of small errors. Finally we observe that in the case of pure states we can solve the problem exactly because the optimal global protocol performance coincides with the unambiguous local one. This motivates the following publication of this thesis.

4. Online unambiguous identification of three symmetric hypotheses

If N samples of a state are given, for two hypotheses, the global unambiguous discrimination protocol yields the same probability of error than a local protocol [CY01]. Is this the case for more hypotheses? The answer is negative, however, there are interesting cases where the global unambiguous is equal to the local protocol. In general, this is a nontrivial problem for more than two hypotheses. We address a very simple discrimination problem with three hypotheses, where the states form a symmetric ensemble. We parametrize this set of states via the Gram matrix. We observe the regions where the probability of error for online protocols give the same probability of error as the global protocol. Our approach is that we do online measurements, a special class of local operations. We also consider sets of states that are not linearly dependent that become linearly independent when multiple samples are available.

2

Preliminaries

The purpose of this chapter is to present some fundamental concepts that will be used throughout the rest of the thesis. Also I will give my personal view of the motivation of Quantum Information theory.

2.1 The Lobster & The Quantum

In Yorgos Lanthimos' film "The Lobster" people are literally expelled from their community if they cannot be in a romantic relationship [Int15]. If someone has the misfortune of ending a relationship that person can go to a hotel where people in the same disgraceful bachelor state is enforced to find a partner.In case that people do not find a couple within 45 days, they are turned into an animal of their preference forever. The protagonist of the movie mentions that if such would be his case, he would like to be turned into a lobster. In other aspects, the context of the movie is very similar to a contemporary country.

There are dissidents of this ruling system: they live isolated from the community in the forest. They are idealists, they are strictly bachelor. Also, as they reject companionship they have the mandate to do most things alone.The main community hunts them and turns them into animals, one of the activities of the people in the hotel is to hunt these dissidents.

Can we learn a deeper lesson from this nonsense? Perhaps about freedom. A free society will always contain dissidents of the ideal of a happy person. There must be a place for “non-happy” ways of living. Whatever the intention of the movie’s plot really is, one thing is clear, it creates a sensation of strangeness and otherness with respect to the universe of the spectators of a contemporary country. It contains sufficient basic elements to seem like a familiar setting, however, essential elements are different, like the freedom of not having a romantic partner. This seemingly small difference in the grand context of affairs produces a very unsettling and just plain odd reality.

A similar sensation of oddness arises when trying to understand Quantum Theory deeply. Quantum Theory poses a challenge to our knowledge of reality. Reasonably or not, the fact that we take Quantum Theory as the fundamental reality of our world asks for deep understanding. Quantum Theory presents us a reality that follows specific rules. These rules are not so extravagant at first glance, however they produce a weird world.

To assess this weirdness we can follow its rules in situations that are familiar and learn something from its consequences. There is a big area of research nowadays in Quantum Information Theory and technologies that follows these inquiries. My point here is that another main *scientific* objective for developing quantum technologies is to immerse oneself in the world that quantum theory creates. Not only see the movie but kind of experience it.

In what follows, I will introduce the basic constituents from Quantum Theory and some elements of Quantum Information theory. My exposition will be closer to that of another classic book of Nielsen and Chuang [NC11]. I want to notice the change of exposition strategy, the same concepts are translated into succinct postulates. As if it is easier to see it as a “thinking system” rather than just a necessary description of phenomena.

2.2 The postulates of quantum mechanics

2.2.1 Postulate 1: Quantum state

The basic assumption in Quantum Mechanics is that the state of every physical system in the universe is described by a wave function [Gri17]. For a given physical system there exists states which are represented by vectors $|\varphi\rangle$ in a Hilbert space $H(\mathbb{C})$. This Hilbert space can have arbitrary dimension but in this thesis we will restrict to Hilbert spaces of finite dimension.

Note that two states in a Hilbert space can be put in linear combination, and the result, as its definition demands, lies in the Hilbert space itself. This seemingly simple fact gives rise to the phenomenon of quantum coherence, which arises from the fact that wave functions which describe physical systems can be superimposed. The usual classical theory supposes that states are one or the other, but not both at the same time. The allegory by Schrödinger of a cat that is in a superimposed state of being alive and being dead is one of the first expositions of this phenomenon.

There is a very interesting discussion about the ontology of the quantum state. What *is* a quantum state? Information about an unknown substrate? Or the "thing-in-itself" [PBR12]? The most general quantum states that one can consider are not the vectors $|\psi\rangle$ but statistical mixtures. When there is a classical uncertainty, described by a probability distribution of being given one of several pure states. Therefore we will consider not only vectors $|\varphi\rangle$ but aggregates that convey an explicit ignorance. If we are given the state $|\varphi_0\rangle$ with probability η_0 and the state $|\varphi_1\rangle$ with probability $\eta_1 = 1 - \eta_0$ then the state

$$\rho = \eta_0 |\varphi_0\rangle\langle\varphi_0| + \eta_1 |\varphi_1\rangle\langle\varphi_1| \,. \tag{2.1}$$

is the one that describes our knowledge.

In general, states will be represented by linear operators in a (finite-dimensional) Hilbert space that have the properties of being positive semidefinite, $\rho \geq 0$, and normalized, $\operatorname{tr}\rho = 1$. We will call states of rank 1 *pure* and any higher rank will be called a *mixed* state.

2.2.2 Qubits

The warhorse of quantum computation and information is one of the simplest quantum states: a two level system usually called qubit. We are considering then a two-dimensional Hilbert space which has a basis of orthogonal states which we will denote $\{|0\rangle, |1\rangle\}$ and any state in this Hilbert space can be written as $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ with α and β complex numbers that fulfill $|\alpha|^2 + |\beta|^2 = 1$. This constrains the numbers α and β and we can write any qubit in terms of two angles $\{\theta, \phi\}$. The parametrization therefore can be chosen such that

$$|\psi\rangle = \cos\frac{\theta}{2}|0\rangle + e^{i\phi}\sin\frac{\theta}{2}|1\rangle \,, \tag{2.2}$$

with $\theta \in [0, \pi)$ and $\phi \in [0, 2\pi)$. These pair of angles parametrize a sphere of unit radius with $\vec{n} = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$. In general, we can have

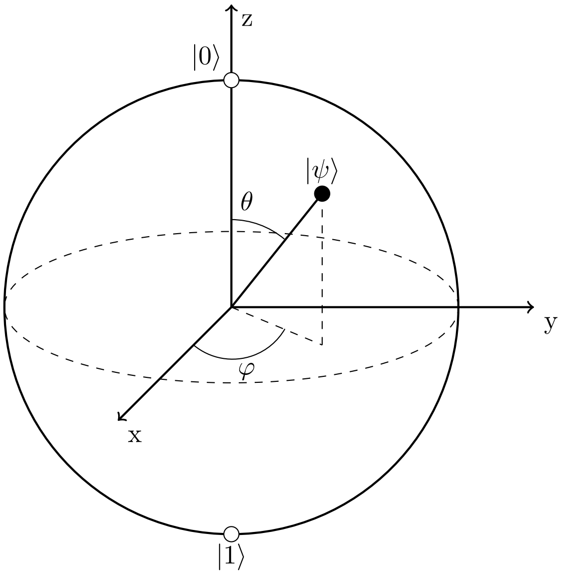


Figure 2.1: Bloch sphere representation of qubits.

mixtures of qubits and any density matrix ρ can be written in a compact form

$$\rho = \frac{1}{2}(\mathbb{1} + r\mathbf{v} \cdot \sigma), \tag{2.3}$$

with $\vec{n}$ as defined above, $r \in [0, 1]$ is called the purity and $\sigma = (\sigma_{\mathbf{x}}, \sigma_{\mathbf{y}}, \sigma_{\mathbf{z}})$ is a vector of the traceless Pauli matrices. Observe that this forms a solid sphere with points determined by the vector $\vec{n}$ and the length of the vector by r. See Figure (2.1) for the representation of pure states in the so-called Bloch sphere.

2.2.3 Postulate 2: Unitary evolution

Time is inscribed into quantum theory as a real continuous parameter. The states that we talked about evolve through time given some problem-specific Hamiltonian H. Basically, this yields the Schrödinger equation:

$$i\hbar \frac{\partial \Psi}{\partial t} = \hat{H}\Psi. \tag{2.4}$$

This is equivalent to saying that the states evolve with unitaries:

$$|\Psi(t)\rangle = e^{-\frac{it}{\hbar}\hat{H}}|\Psi_0\rangle = U(t)|\Psi_0\rangle. \tag{2.5}$$

2.2.4 Postulate 3: Quantum measurements

Perhaps this is the most problematic postulate [WM09]. When a measurement is performed in a system, the state of this system changes, it "collapses" into another state. This is what is known as the Copenhague interpretation. There is a large literature written with respect to this point. We will not enter in this discussion. We will take this interpretation with its vagueness.

We need to define some operators in order to be able to talk about measurements in quantum mechanics. For $m = 1, 2, \ldots, n$ let $\hat{M}_m$ be a set of operators that act on the Hilbert space $H(\mathbb{C})$ of dimension d where we defined our states $|\Psi\rangle$. The set of operators cannot be arbitrary, it has to fulfill a completeness property:

$$\sum_m \hat{M}_m^\dagger \hat{M}_m = \mathbb{I}. \tag{2.6}$$

A system described by a state $|\Psi\rangle$ is described (after obtaining a measurement measurement outcome m) by

$$|\Psi'\rangle = \frac{\hat{M}_m|\Psi\rangle}{\sqrt{\langle\Psi|\hat{M}_m^\dagger \hat{M}_m|\Psi\rangle}}. \tag{2.7}$$

Equation (2.6) implies that

$$\sum_m \langle\Psi|\hat{M}_m^\dagger \hat{M}_m|\Psi\rangle = 1. \tag{2.8}$$

With $\langle\Psi|\hat{M}_m^\dagger \hat{M}_m|\Psi\rangle \geq 0$ [NC11] therefore we can assume that we have a probability density function for the subscript m,

$$p(m) = \langle\Psi|\hat{M}_m^\dagger \hat{M}_m|\Psi\rangle. \tag{2.9}$$

In general we will designate $\hat{M}_m^\dagger \hat{M}_m = \hat{E}_m$ and ask the completeness property of the set (2.6). Also, we will ask the set of operators to be positive semidefinite $\hat{E}_m \geq 0$.

2.2.5 Postulate 4: Aggregates or tensor products

We observe several independent systems in our universe. Quantum Mechanics can consider the situation when there is more than one system at a given time. However, a priori it is not evident how to do this. If we have n states of different systems $\{|\Psi_1\rangle, \dots, |\Psi_n\rangle\}$ the overall quantum state of the system aggregate is given by [Gri17] its *tensor product*:

$$|\Psi_{tot}\rangle = |\Psi_1\rangle \otimes \dots \otimes |\Psi_n\rangle. \tag{2.10}$$

Usually the symbol $\otimes$ is omitted and we take the convention that the juxtaposition of several quantum states denotes its tensor product.

2.2.6 Quantum entanglement

Let the state of two quantum particles be

$$|\varphi\rangle = |\psi_1\rangle|\psi_2\rangle. \tag{2.11}$$

One might ask what is the most general state that describes both particles at the same time. In general, Quantum Theory admits states that *cannot* be written as a decomposition into tensor products of independent states, as in Equation (2.11). This kind of states will be called entangled states.

In general, for mixed states, as considered above, we have that if a state can be written as

$$\rho = \sum_{ij} \sigma_A^i \sigma_B^j \tag{2.12}$$

then we say that ρ is separable. If ρ is *not* separable then we say that it is entangled. Basically, entanglement is a phenomenon that arises in Hilbert spaces of composite systems.

With this concept we are entering the weird world of Quantum Theory. The states that are entangled have effects that puzzle the classical mind. This is because the states that are entangled present correlations that are stronger than those from classical distributions. To see the correlations one has to think in terms of a task. It is simpler to see the case of Bell inequalities for qubits [NC11]. For example, suppose that Alice and Bob have each two classical random variables: Q, R and S, T respectively as in Figure (2.2), each of which can yield the results ±1. The following quantity can be defined $QS + RS + RT - QT = (Q + R)S + (R - Q)T$. It follows that $(Q + R)S = 0$ or $(R - Q)T = 0$, in either case $QS + RS + RT - QT = \pm 2$. Taking the expectation value we can get the Bell inequality,

$$\mathbb{E}(QS) + \mathbb{E}(RS) + \mathbb{E}(RT) - \mathbb{E}(QT) \leq 2 \tag{2.13}$$

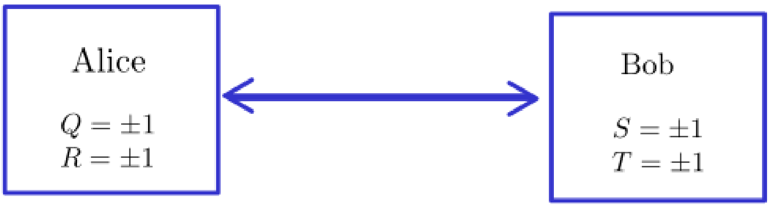


Figure 2.2: Schematic for Bell inequalities.

where $\mathbb{E}$ denotes expectation value. If the systems that Alice and Bob have is quantum mechanical things can be different. For example, suppose each one has a particle such that they share the entangled state

$$|\psi\rangle = \frac{|01\rangle - |10\rangle}{\sqrt{2}}. \tag{2.14}$$

If they perform measurements of the following observables in terms of the Pauli matrices Z, X:

$$Q = Z_1 \tag{2.15}$$
$$R = X_1 \tag{2.16}$$
$$S = \frac{-Z_2 - X_2}{\sqrt{2}} \tag{2.17}$$
$$T = \frac{Z_2 - X_2}{\sqrt{2}}. \tag{2.18}$$

We thus have

$$\langle QS\rangle + \langle RS\rangle + \langle RT\rangle - \langle QT\rangle = 2\sqrt{2}. \tag{2.19}$$

Here we have a clear numerical difference between the classical case and the quantum one, because $2 < 2\sqrt{2}$. This is an example of what we mean with stronger correlations.

Observe that in general we can see this inequality as follows: optimize the expectation of the quantity $QS + RS + RT - QT$ over input states. Quantum mechanics allows for states that give a larger expectation value of this quantity. Often problems in quantum information consist in cost functions that depend on quantum objects (states and measurements) that have to be optimized. Quantum mechanics offers the optimization over entangled states and measurements.

From Postulate 1, observe that mixed states are positive semidefinite operators. We observe also from Postulate 3 that measurements in general are described by positive semidefinite operators. As seen in Equation (2.12) there exists the notion of entanglement for these kind of operators. It is natural then to ask for the structure of the measurement operators. It turns out that this is a very rich question with deep implications. The measurements that can be done with entangled operators yield different kind of information than those with separable ones. In fact, usually when we optimize over measurements a cost function, the optimal measurement turns out to be an entangled one. It is possible, however, that in some cases, a separable one can yield the same results [ABB+05].

We can make a classification of the possible measurements that can be implemented in a given system. Suppose we have an $N-$partite quantum system, with an underlying Hilbert space given by $H = H^{A_1} \otimes H^{A_2} \otimes \ldots \otimes H^{A_N}$, which means, we have the parties $A_1, A_2, \ldots, A_N$ each of which have a part of the system.

First, the most general measurement that can be implemented in H could be an entangled operator, i.e., an operator that does not have a decomposition as in (2.12). We will call them entangled operators.

Then we can have operators that can have a decomposition of the form

$$E_m = E^{A_1} \otimes E^{A_2} \otimes \ldots \otimes E^{A_N}. \tag{2.20}$$

with a suboperator on each Hilbert subspace. We will call this operator separable.

Now, we can consider more restrictive measurements, where there is only local operations and classical communication allowed between the parties $A_1, A_2, \ldots, A_N$. This means that each party applies a complete measurement: $\sum_i E_i^{A_i} = \mathbb{1}$, and that the outcomes of the separable measurements are shared between the parties and this information is used to make measurements over copies of states. We will call this kind of measurements LOCC (local operations and classical communication) [CLM+14].

One could consider a special case of LOCC measurements where only a series of measurements is allowed, which means, the system A_1 is measured, then the result is informed to the party that holds the system A_2 to modify its measurement apparatus and so on. This kind of measurements will be called one-way LOCC. If, furthermore we ask that the measurement each party applies is optimal at each step, we call these measurements *online*.

Online measurements are different from LOCC when we consider the optimization on cost functions. LOCC would allow for communication between parties $A_1, A_2, \ldots, A_N$ and making measurements between. In contrast, online measurements have to be optimal at each step. LOCC allows for suboptimality in the measurements until the last measurement is made.

3

Quantum state discrimination

In this chapter I include the basic mathematical tools and results that support the relevance and validity of the articles presented in this thesis. Here we will present a very fundamental task that will prove to be extremely versatile: quantum state discrimination. This is an essential part of many tasks of quantum information theory.

We have two agents: Alice and Bob. Alice has promised to give Bob a quantum system prepared in a quantum state out of two possible ones. In classical physics it is always possible *in principle* to distinguish two states 0 and 1. However, in quantum mechanics we can have two pure states which are nonorthogonal and thus, not perfectly distinguishable.

Now, notice that Alice has promised one of two states, this means that Bob will have in mind two possible states but knows that one state is the real one. Bob prepares a measurement for figuring out which of these states have actually been given. Notice here that the problem is defined by the set of hypothesis states.

3.1 Classical hypothesis testing

Our task stems from the study of hypothesis testing. Once we define the classical theory of hypothesis testing, the quantum part will appear naturally.

Suppose Alice gives Bob an apple. It would be very easy for Bob to tell if the apple is red or yellow. Bob just looks at it (counting Fuji apples as red). There are however relevant questions about complex systems where asking a simple question can be more complicated. Take a district in Barcelona for example, is it right wing or left wing? This is a relevant question for democracy to work and it is non-trivial to answer in many cases (e.g. how to consider the position of the people who is not "political"?).

To address this kind of non-trivial questions statistics gives us tools to extract knowledge from the data that is available. A particular one is hypothesis testing. The basic objective is to identify if a system follows one of several hypothesis from a given set. In terms of probability, we frame it as the question that a system follows one of several possible probability distributions.

We analyze the case of two hypotheses. The process is self-explanatory: to answer the question of which of the hypotheses we believe as true we test observed data. We can make a wrong guess of course, this is what makes this problem interesting in the first place! We can therefore see the inference process as a decision problem. An observer samples once from either distribution $p_0(x)$ or $p_1(x)$. The outcome of the sampling will reveal something about the identity of the distribution from which it was drawn.

For two hypotheses, let us call them H_0 and H_1 there are only two types of errors, that we assess hypothesis H_0 as true when is false which is called type-I error or that we wrongly assess hypothesis H_1 as true when is false which we will call type-II error. We will associate the probabilities $p(0|H_1)$ and $p(1|H_0)$ respectively to these errors. However, it is also useful to think of an average error. If prior information is available, let's say, that we know that H_0 is true with probability $0 \leq \eta_0 \leq 1$ and H_1 with probability η_1 then we can consider the average Bayesian probability of error

$$P_e = \eta_0 p(1|H_0) + \eta_1 p(0|H_1). \tag{3.1}$$

We still need a procedure to carry out a suitable guess from the possible hypothesis. A natural way to establish a guess is to choose the most probable hypothesis through a Bayesian updating process [Jay03]. Given an outcome x, we update our estimate about an hypothesis as

$$p(0|x) = \frac{\eta_0 p_0(x)}{p(x)} = \frac{\eta_0 p_0(x)}{\eta_0 p_0(x) + \eta_1 p_1(x)}. \tag{3.2}$$

Where $p(x)$ is the probability to have the outcome x. The Bayes estimate is given by comparing the quantities $\eta_0 p_0(x)$ and $\eta_0 p_1(x)$. We can define the Bayes' decision function as follows,

$$\delta_B(x) = \begin{cases} 0, & \text{if } \eta_0 p_0(x) > \eta_1 p_1(x) \\ 1, & \text{if } \eta_0 p_0(x) < \eta_1 p_1(x) \\ \text{anything} & \text{if } \eta_0 p_0(x) = \eta_1 p_1(x). \end{cases} \tag{3.3}$$

One can prove that this decision method is optimal in terms of the error probability in the sense that any other decision function would yield a larger error probability i.e. $P_e(\delta) > P_e(\delta_B)$ for $\delta \neq \delta_B$ [Fuc95]. We denote the probability of error with respect to Bayes' decision method as P_e.

Observe that in case that both hypotheses yield the same value then the only possible guess is random. This rule tells us how to calculate the error in the estimation as it is the minimum of the conditional probabilities $\min\{p(0|x), p(1|x)\}$ as our guess corresponds to the hypothesis with the highest probability. Observe that we have two hypotheses but the number of outcomes can be arbitrary. Let us consider a $l \in \mathbb{N}$ outcome experiment, therefore the probability of error is given by

$$P_e = \sum_{x=1}^{l} p(x) \min\{p(0|x), p(1|x)\} \tag{3.4}$$

$$= \sum_{x=1}^{l} \min\{\eta_0 p_0(x), \eta_1 p_1(x)\}. \tag{3.5}$$

Using the Bayes rule, we can update the priors with the information that we obtain from samples. Therefore, the inference process can be iterated n times. The estimation would obviously depend on the number of iterations. However, notice that it always depends on the number of samples one considers. It would be useful to relate probability distributions through a geometric distance between them. Such a well-defined distance would *not* depend on the number of measurements one takes, i.e. it would only depend on the distributions that we consider.

The dependence on the number of measurements will go away when we consider the number of samples going to infinity. In this regime (large sample size) it is intuitive that the probability of error decays exponentially and it can be shown that the optimal exponent is given by the Chernoff bound [Che52].

3.1.1 Chernoff bound

The most natural scenario for testing is a symmetric one. Here we consider the Chernoff case as a contraposition of the asymmetric discrimination one. We repeat the Bayesian inference process N times and observe the probabilities we obtain. Given the outcomes

$$x^{(N)} = (x_1, x_2, \dots, x_N), \tag{3.6}$$

which consists on a vector of N outcomes. We therefore have two possible distributions assuming one hypothesis or the other:

$$p_0(x^{(N)}) = p_0(x_1)p_0(x_2)\dots p_0(x_N) \tag{3.7}$$
$$p_1(x^{(N)}) = p_1(x_1)p_1(x_2)\dots p_1(x_N). \tag{3.8}$$

Using the inequality [Che52, Fuc95]

$$\min\{a, b\} \leq a^s b^{1-s}, \quad s \in [0,1], \tag{3.9}$$

in Equation (3.5) we obtain

$$\begin{aligned}
P_e &= \sum_{i=1}^{l} \min\{\eta_0 p_0(x^{(N)}), \eta_1 p_1(x^{(N)})\} && (3.10)\\
&\leq \eta_0^s \eta_1^{1-s} \sum_{i=1}^{l} \left(\prod_{k=1}^{N} p_0(x_k)^s p_1(x_k)^{1-s} \right) && (3.11)\\
&= \eta_0^s \eta_1^{1-s} \prod_{k=1}^{N} \left(\sum_{i=1}^{l} p_0(x_k)^s p_1(x_k)^{1-s} \right) && (3.12)\\
&= \eta_0^s \eta_1^{1-s} \left(\sum_{i=1}^{l} p_0(x_k)^s p_1(x_k)^{1-s} \right)^N && (3.13)\\
&\leq \min_{s\in[0,1]} \eta_0^s \eta_1^{1-s} \left(\sum_{i=1}^{l} p_0(x_k)^s p_1(x_k)^{1-s} \right)^N . && (3.14)
\end{aligned}$$

Therefore $P_e(N)$ is upper bounded by the so-called Chernoff bound. We can obtain a quantity that is independent of N that yields a notion of distance between probability distributions. The probability of error when $N \to \infty$ is given by

$$P_e(N \to \infty) \sim e^{-NC(p_0,p_1)}, \tag{3.15}$$

where we define the Chernoff distance as

$$C(p_0, p_1) \equiv -\log \min_{s\in[0,1]} \sum_{i=1}^{l} p_0(i)^s p_1(i)^{1-s}. \tag{3.16}$$

3.1.2 Kullback-Liebler divergence

In contraposition to the Chernoff case, we could consider an asymmetric distinction of hypotheses: we will choose H_0 as true unless evidence convince us that H_1 is true. The Kullback-Liebler divergence (or relative entropy) $D(p_0||p_1)$ is a measure of the inefficiency of assuming that the distribution is p_1 when the true distribution is p_0 [CT12]. This quantity compares also two probability distributions, however in an asymmetric way. Therefore, it does not define a distance in general. The relative entropy between two probability distributions p_0 and p_1 is defined

$$D(p_0||p_1) = \sum_x p_0(x) \log \frac{p_0(x)}{p_1(x)}, \tag{3.17}$$

The Kullback-Liebler becomes important in hypothesis testing because of the following theorem, called Stein's Lemma [CT12]

Theorem 1. *Let $X_1, X_2, \ldots, X_n$ be i.i.d.$\sim Q$. Consider the hypothesis test between two alternatives, $Q = p_0$ and $Q = p_1$, where $D(p_0||p_1) < \infty$. Let $A_n \subseteq \mathcal{X}^n$ be an acceptance region for hypothesis H_1. Let the probabilities of error be*

$$\alpha_n = p_0^n(A_n) \tag{3.18}$$

$$\beta_n = p_1^n(A_n), \tag{3.19}$$

and for $0 < \epsilon < 1/2$, define

$$\beta_n^\epsilon = \min_{A_n\subseteq\mathcal{X}^n,\quad \alpha_n<\epsilon} \beta_n. \tag{3.20}$$

Then,

$$\lim_{n\to\infty} \frac{1}{n} \log \beta_n^\epsilon = -D(p_0||p_1). \tag{3.21}$$

3.2 Minimum error quantum discrimination

As mentioned earlier in this chapter, Alice gives Bob a quantum system in a state and Bob has to figure out which one out of the two promised states Alice actually gave him.

First we shall examine the natural case of two hypotheses: ρ and σ. These states can be pure or mixed. An important figure of merit to consider is the average error an estimation process makes.

We will see the average error (3.1) when considering quantum systems and measurements. First of all, we can write in general a pair of qubits in terms of an orthogonal basis $\{|0\rangle, |1\rangle\}$ as

$$\begin{aligned}|\psi_0\rangle &= \cos\theta|0\rangle + \sin\theta|1\rangle,\\ |\psi_1\rangle &= \cos\theta|0\rangle - \sin\theta|1\rangle.\end{aligned} \tag{3.22}$$

Recall that the possible states are given by $|\psi_0\rangle$ and $|\psi_1\rangle$ and a POVM is given by a complete set of positive operators. We will optimize for 2-outcome POVMs, $E_0 + E_1 = \mathbb{1}$ therefore we get the equation

$$P_e = \eta_0\langle\psi_0|E_1|\psi_0\rangle + \eta_1\langle\psi_1|E_0|\psi_1\rangle, \tag{3.23}$$

$$= \eta_0 - \mathrm{tr}\left[(\eta_0\,|\psi_0\rangle\langle\psi_0| - \eta_1\,|\psi_1\rangle\langle\psi_1|)E_0\right]. \tag{3.24}$$

Now, we know that *any* operator $0 \leq E_0 \leq \mathbb{1}$ will fulfill Equation (3.24). However, knowing that we have a specific function we can ask for which operator E_0 is the average probability of error P_e minimized, which is the optimization that we mentioned above. One can show that the minimum is reached when E_0 is a projector onto the subspace spanned by the eigenvectors corresponding to the positive eigenvalues of the operator $\eta_0\,|\psi_0\rangle\langle\psi_0| - \eta_1\,|\psi_1\rangle\langle\psi_1|$ usually called the Helstrom operator.

As we have two pure states then this defines a two-dimensional space. Without loss of generality we consider $\eta_0 \geq \eta_1$, then one can calculate the eigenvalues of this operator:

$$\lambda_\pm = \frac{1}{2}\left(\eta_0 - \eta_1 \pm \sqrt{1 - 4\eta_0\eta_1\cos^2 2\theta}\right). \tag{3.25}$$

then, from Equation (3.24) the minimum probability of error is given by

$$P_e = \frac{1}{2}\left(1 - \sqrt{1 - 4\eta_0\eta_1|\langle\psi_0|\psi_1\rangle|^2}\right). \tag{3.26}$$

Equation (3.26) is usually called the Helstrom bound [Hel76] and the Helstrom measurement given in terms of the Helstrom operator, which achieves this bound.

The minimum error measurement will sometimes indicate a state incorrectly. We will see in the next section that protocols that make no errors are also possible, albeit not always.

Square Root Measurement

The problem of quantum state discrimination can be generalized to the case when we have more than 2 hypotheses. The general problem can be addressed numerically but analytically in general it becomes highly complicated. In fact, no general method has been found to solve minimum error state discrimination analytically [Bae13]. Even the simple case of three states if no symmetry is known among given states, no analytic solution is known. One measurement that results very useful in this general setting is the Square Root Measurement (SRM). It is a specific measurement over an arbitrary finite set of states but its versatility earned the name of pretty good measurement [HW94]. Suppose we are given n states ρ_i with respective probabilities η_i. We define

$$\bar{\rho} = \sum_{i=1}^{n} \eta_i \rho_i. \tag{3.27}$$

Then the SRM can be defined as a POVM with elements

$$E_i = \eta_i \bar{\rho}^{-1/2} \rho_i \bar{\rho}^{-1/2}. \tag{3.28}$$

It is clear that the operators E_i are positive semidefinite and form a complete set

$$\sum_{i=1}^{n} \eta_i \bar{\rho}^{-1/2} \rho_i \bar{\rho}^{-1/2} = \bar{\rho}^{-1/2} \bar{\rho} \bar{\rho}^{-1/2} = \mathbb{1}. \tag{3.29}$$

It is known that there are cases where this measurement is optimal [DPP15]. Also, it might be suboptimal but optimal in an asymptotic limit [HJS+96, SBC+16].

The cases where the SRM is optimal are very symmetric ones, we will see the condition for optimality in a following section (3.3.1).

3.3 Unambiguous discrimination

Minimum error discrimination deals with the most unrestricted problem of quantum discrimination in the sense that the only thing that it is asked is for the probability of error to be minimal. However, we can ask for more restricted problems that naturally arise. A very useful scheme is that of unambiguous discrimination [BFF12].

For pure states we can obtain a measurement scheme that does *not* yield erroneous answers, this means, it yields a conclusive answer if and only if we have a linearly independent set [Che98]. This would suggest that we have surpassed the minimum error scheme. However this scheme will succeed perfectly with certain probability but can sometimes fail altogether and yield no conclusive answer.

Let us see the simplest case of two hypotheses. For the case of two qubits, given the states (3.22) we have the operators

$$\begin{aligned} E_0 &= a_0(\sin\theta\,|0\rangle + \cos\theta\,|1\rangle)(\sin\theta\,\langle 0| + \cos\theta\,\langle 1|), \\ E_1 &= a_1(\sin\theta\,|0\rangle - \cos\theta\,|1\rangle)(\sin\theta\,\langle 0| - \cos\theta\,\langle 1|). \end{aligned} \tag{3.30}$$

We choose these operators so that $\langle\psi_0|\,E_1\,|\psi_0\rangle = \langle\psi_1|\,E_0\,|\psi_1\rangle = 0$ and with $0 \leq a_0, a_1 \leq 1$. Therefore, if we get the outcome E_0 we know that we had the state $|\psi_0\rangle$ for sure and the same goes for the 1 outcome. The operators in Equation (3.30) do not sum up to $\mathbb{1}$ so there is a positive semidefinite operator that we need to add to fulfill the completeness requirement (2.6). We have therefore an inconclusive outcome operator,

$$E_? = \mathbb{1} - E_0 - E_1. \tag{3.31}$$

Knowing this we can ask for the protocol that yields an inconclusive answer with the smallest probability.

Normally the probability of success which is the probability that our measurement scheme yields a correct answer will normally will be *lower* than that of minimum error. The reason of this is that unambiguous is a very restricted case of discrimination. It asks a lot from the measurement, that it yields no error.

The probability to get the inconclusive outcome is given by

$$P_? = \eta_0\langle\psi_0|E_?|\psi_0\rangle + \eta_1\langle\psi_1|E_?|\psi_1\rangle = 1 - \sin^2 2\theta(\eta_0 a_0 + \eta_1 a_1). \tag{3.32}$$

The optimal operator $E_?$ is found when $P_?$ is minimized. The parameter θ is fixed, as are p_0 and p_1. What is left to optimize are the parameters a_0 and

a_1 under the constraint that they are positive and that the operator $Pi_? \geq 0$. For equal a priori probabilities $p_0 = p_1 = \frac{1}{2}$, the minimum probability of unambiguous is given by $P_? = \cos 2\theta = |\langle\psi_0|\psi_1\rangle|$ with the parameters and operator,

$$a_0 = a_1 = \frac{1}{2\cos^2\theta}, \tag{3.33}$$

$$E_? = (1 - \tan^2\theta)\,|0\rangle\langle 0|\,. \tag{3.34}$$

Now, unambiguous discrimination is also possible for sets of mixed states. However, the condition is stronger, it needs the support of the states to be non-identical. However, this is a hard condition to have experimentally [ZFY06].

3.3.1 Gram Matrix

The Gram Matrix is a mathematical tool that is very useful for decoding problems of discrimination of pure states.

Although the concept of Gram matrix is very general as it is defined for any vector space with inner product (normally denoted $\langle\cdot,\cdot\rangle$) we will focus on finite dimensional Hilbert spaces H. Suppose we have a set of states $\{|\psi_1\rangle, \ldots, |\psi_n\rangle\}$ in a Hilbert space of finite dimension H. We define a Gram matrix for this set of vectors component-wise as [HJ12]

$$G_{ij} = \langle\psi_i|\psi_j\rangle, \tag{3.35}$$

which is an $n \times n$ matrix with complex entries. There are interesting properties of this matrix as it condenses properties of the whole set of states. To illustrate some of these properties we have the following theorem,

Theorem 2. *Let $\{|\psi_1\rangle, \ldots, |\psi_n\rangle\}$ be states in a Hilbert space H and let $G = [\langle\psi_j|\psi_i\rangle]_{i,j=1}^m$ then,*

- *G is positive semidefinite if and only if the states $\{|\psi_1\rangle, \ldots, |\psi_n\rangle\}$ are linearly independent.*
- $\operatorname{rank} G = \dim \operatorname{span}\{|\psi_1\rangle, \ldots, |\psi_n\rangle\}$.

The proof can be found in [HJ12].

Also, observe that if we have positive semidefinite matrix $A \in M_m$ then we could see the square root matrix $A^{1/2}$ as a square matrix of columns $A^{1/2} = [|\psi_1, \ldots, \psi_n\rangle]$ and therefore $(A^{1/2})^\dagger A^{1/2} = A$ and $[A]_{ij} = \langle\psi_i|\psi_j\rangle$. Therefore, every positive semidefinite matrix is a Gram matrix.

The reason we introduce the Gram matrix here is that it is a useful tool not only to check linear independence of a set but also to study multi-hypotheses discrimination problems. It contains all the information necessary for solving a discrimination problem with pure states. As it only depends on the overlaps $G_{ij} = \langle\psi_i|\psi_j\rangle$ it is therefore independent of the bases of the states $|\psi_i\rangle$.

The square root measurement is generally considered to be a suboptimal measurement that "scales well" with a growing number of hypothesis n [HJS+96]. However, there are cases when this measurement is optimal. They have to be very symmetric as we will see.

We have the following theorem by Dalla Pozza et. al for linearly independent sets of states [DPP15].

Theorem 3. *Given a Gram matrix G and its square root $G^{1/2}$ that are block diagonal, which means $G = G_1 \oplus \ldots \oplus G_n$ then the square root measurement is optimal if and only if the square root $G_i^{1/2}$ of each block has equal diagonal entries.*

Basically the SRM needs a very uniform set of states. A known example for which the SRM is the optimal measurement is one that has *geometrical uniform symmetry* [DPP15]. Actually one can see that the more symmetric the better from the following theorem that is related to the Change Point Problem [SBC+16],

Theorem 4. *Let $\{|\psi_i\rangle\}_{k=1}^n$ be a linearly independent set of states with the Gram matrix $G_{ij} = \langle\psi_i|\psi_j\rangle$. The maximum probability of correctly identifying a state drawn uniformly at random from the set $\{|\psi_i\rangle\}_{k=1}^n$ satisfies the bounds*

$$P_{max} \geq \left(\frac{\operatorname{tr}\sqrt{G}}{n}\right)^2 \tag{3.36}$$

and

$$P_{max} \leq \left(\frac{\operatorname{tr}\sqrt{G}}{n}\right)^2 + \sqrt{\lambda_{max}} \parallel \mathbf{q} - \mathbf{u} \parallel_\mathbf{1}, \tag{3.37}$$

where λ_{max} is the maximum eigenvalue of G, $\mathbf{q} = \{\mathbf{q_k}\}$ is defined as $q_k := (\sqrt{G})_{kk}/\operatorname{tr}[\sqrt{G}]$ and $\mathbf{u} = \{\mathbf{u_k}\}$ is the uniform distribution, and $\parallel \cdot \parallel$ denotes the trace norm.

The idea behind this theorem is that using a decomposition of the Gram matrix one can observe the dependence of the probability of success on its square root matrix. However, in general it also depends on a unitary transformation that, when the SRM is optimal, it turns out to be the identity.

What this last theorem tells us is that the closer to a uniform distribution the diagonal of $G^{1/2}$ gets, the closer it is to the optimal value of P_{max}.

Although the Gram matrix is useful to study some cases of multi-hypotheses pure state discrimination, except for very special and symmetric cases there is no way to address this problem [Bae13]. However, we can investigate this problem using a very powerful mathematical concept: Semidefinite programming. In the next section we review this useful technique.

3.4 Semidefinite programming

Semidefinite programming is a very powerful tool to address optimization problems in quantum information theory. This technique has advantages for understanding optimization problems and also it is very useful for numerical calculations.

A semidefinite program (SDP) is a convex optimization problem. It is closely related to linear programming (LP) [BV04]. LP is an optimization problem over a polytope that is given by the possible solutions of a linear system of inequalities. Linear programs use vectors as the variable to optimize. Semidefinite programs, generalize this kind of problems to positive semidefinite matrices. There are linear constraints for SDPs as for the LPs; the difference with SDPs is that the constraint to positive semidefinite operators is not linear in the components of the matrix variables. Nevertheless, there exist algorithms to solve SDPs in polynomial time [GB14].

We define SDPs following Watrous [Wat18],

Definition 1. *A semidefinite program is a triple* (Φ, A, B) *where*

- $\Phi \in T(X, Y)$ *is a Hermiticity-preserving map.*
- $A \in \mathrm{Herm}\,(X)$ *and* $B \in \mathrm{Herm}\,(Y)$ *are Hermitian operators.*

for some choice of complex Euclidean spaces X *and* Y*. The final part associated with the triple* (Φ, A, B) *is two optimization problems called* primal *and* dual*, as follows:*

$$\begin{aligned} \text{maximize} \quad & \langle A, X\rangle \\ \text{subject to} \quad & \Phi(X) = B, \\ & X \in \mathrm{Pos}(X). \end{aligned} \tag{3.38}$$

$$\begin{aligned} \text{minimize} \quad & \langle B, Y\rangle \\ \text{subject to} \quad & \Phi^*(Y) \geq A, \\ & Y \in \mathrm{Herm}(Y). \end{aligned} \tag{3.39}$$

There are several remarks of the previous definition: the $\langle \cdot, \cdot \rangle$ denote an inner product for operators, $\mathrm{Pos}(X)$ denotes the set of positive semidefinite operators and $\Phi^*(Y)$ is the adjoint map of $\Phi(X)$, which is the unique map that fulfills [Wat18]

$$\langle \Phi^*(Y), X\rangle = \langle Y, \Phi(X)\rangle. \tag{3.40}$$

Eq. (3.38) is normally called the *primal* problem and Eq. (3.39) is called the *dual* problem. Now, the interesting part of the semidefinite programs defined above is their interrelation. Finding the optimal values might be a challenge but it will be useful to study the solutions of both programs.

Any operator that fulfills the constraints of Eq. (3.38), which means the set A defined as

$$\mathcal{A} = \{X \in \mathrm{Pos}(X) : \Phi(X) = B\}, \tag{3.41}$$

is said to be *primal feasible*. In the same vein operators in the set

$$\mathcal{B} = \{Y \in \mathrm{Herm}(Y) : \Phi^*(Y) \geq A\}, \tag{3.42}$$

is said to be *dual feasible*. Analogously, we can define the functions $X \rightarrow \langle A, X\rangle$ and $Y \rightarrow \langle B, Y\rangle$ as the primal and dual objective functions. The primal optimum and dual optimum are the values that satisfy,

$$\alpha = \sup_{X\in\mathcal{A}} \langle A, X\rangle \tag{3.43}$$

$$\beta = \inf_{Y\in\mathcal{B}} \langle B, Y\rangle. \tag{3.44}$$

The values α and β may be infinite or finite.

Both the primal and dual programs solve, under certain circumstances, *the same* problem. Now, the having two versions of the same problem is helpful because we have two ways to observe the same problem. There are two types of correspondence (usually also called duality) between the SDPs, one *weak* and one *strong*, which correspond to the following proposition and theorem [Wat18]:

Proposition 1. *For every semidefinite program (Φ, A, B) it holds that $\alpha \leq \beta$.*

This implies that every dual-feasible operator $Y \in B$ yields an upper bound of $\langle B, Y\rangle$ on the optimal value α. More generally,

$$\langle A, X\rangle \leq \alpha \leq \beta \leq \langle B, Y\rangle, \tag{3.45}$$

for every $X \in A$ and $Y \in B$. The condition $\alpha = \beta$ is known as *strong duality*. Strong duality does not hold for every semidefinite program but it does for most of them [Wat18].

Semidefinite programming is a very useful tool to study quantum state discrimination [EMV03, Eld03]. We can translate the problem into a SDP easily. We observe that the trace can be regarded as an inner product, therefore, for minimum error, we can write the programs

$$\begin{aligned} \text{maximize} \quad & \operatorname{tr}[\rho E] \\ \text{subject to} \quad & \textstyle\sum_i \hat{E}_i = \mathbb{1}, \\ & \hat{E}_i \geq 0 \;\; \forall\, i. \end{aligned} \tag{3.46}$$

$$\begin{aligned} \text{minimize} \quad & \operatorname{tr}[Y] \\ \text{subject to} \quad & Y \geq \rho_i \;\; \forall\, i, \\ & Y \geq 0. \end{aligned} \tag{3.47}$$

where we encode the hypotheses states into a block matrix as

$$\rho = \begin{pmatrix} \rho_1 & 0 & \dots & 0 \\ 0 & \rho_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \rho_N \end{pmatrix}. \tag{3.48}$$

Analogously the POVM elements are arranged in a block matrix as well:

$$E = \begin{pmatrix} \hat{E}_1 & 0 & \dots & 0 \\ 0 & \hat{E}_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \hat{E}_N \end{pmatrix}. \tag{3.49}$$

Notice that the dual program yields the Holevo conditions of optimality in the constraint $Y - \rho_i \geq 0 \;\forall\;\; i$ [Hol73]. We can write a correspondent program for unambiguous discrimination, in that case the POVM operators are essentially fixed. The only thing that remains to be optimized is a scalar factor as we will see next. It is nice to write it for pure states only as we can make use of the Gram matrix and SDP [SCMnT17].

Suppose we have N linearly independent pure states $\{|\psi_k\rangle\}_{k=1}^N$. If we have an orthonormal basis $\{|i\rangle\}$ of dimension N we can define the operator

$$R = \sum_k |\psi_k\rangle\langle k| \,. \tag{3.50}$$

We observe that therefore,

$$G = \sum_{i,j=1}^N \langle\psi_i|\psi_j\rangle \, |i\rangle\langle j| = R^\dagger R. \tag{3.51}$$

The inverse of R exists because the set is linearly independent and is given by

$$R^{-1} = \sum_{k=1}^N \left|k\rangle\langle\tilde{\Phi}_k\right| , \tag{3.52}$$

where $\langle k|R^{-1}R|l\rangle = \langle\tilde{\Phi}_k|\psi_l\rangle = \delta_{kl}$ and $|\tilde{\Phi}_k\rangle$ is not normalized in general. From the above considerations we observe that the POVM satisfying the required characteristics for unambiguous discrimination is one given by

$$E_k = \gamma_k \left|\tilde{\Phi}_k\rangle\langle\tilde{\Phi}_k\right| . \tag{3.53}$$

We can call the $0 \leq \gamma_k \leq 1$ efficiencies [SCMnT17]. Observe that if the a priori probability distribution of the hypotheses is given by $p(i)$ then the probability of success is given by $P_s = \sum_{k=1}^N p(k)\gamma_k$. If we take the diagonal matrix given by $\Gamma_D = \text{diag}\{\gamma_1, \gamma_2, \ldots, \gamma_N\}$, we can transform the condition of having the ambiguous result positive

$$E_? = \mathbb{1} - \sum_{k=1}^N \gamma_k \left|\tilde{\Phi}_k\rangle\langle\tilde{\Phi}_k\right| , \tag{3.54}$$

by multiplying by $R^\dagger$ from the left and R from right into a very simple equation. We can therefore obtain the following (primal) SDP,

$$\begin{aligned} \text{maximize}_{\,\Gamma} \quad & \text{tr}\left[\Gamma\eta\right] \\ \text{subject to} \quad & G - \Gamma_D \geq 0, \\ & \Gamma \geq 0. \end{aligned} \tag{3.55}$$

The dual program is thus,

$$\begin{aligned} \text{minimize}_Z \quad & \operatorname{tr}[GZ] \\ \text{subject to} \quad & Z \geq \eta, \\ & Z \geq 0, \end{aligned} \tag{3.56}$$

where $\eta = \operatorname{diag}\{p(1), p(2), \ldots, p(N)\}$, $p(i)$ is a probability distribution and G is the Gram matrix formed by the hypotheses.

We have treated the problem in general for a large (finite) set of states. However, we can still have interesting consequences in the two-state case. We can have many copies of the states and the states can be mixed. In this situation we need a characterization which requires sophisticated (however standard) tools that we present in the next section.

3.5 Many qudits

There are useful representations of the state $\rho^{\otimes N}$. There is a particular decomposition that comes from using representation theory. It was introduced in the context of purification and estimation of qubits [CEM99, BBG$^+$06]. It is also used in learning protocols such as [SCMTB12]. For pure qubits one has the Clebsch-Gordan decomposition in $SU(2)$ which gives us a basis for sum of angular momentum. More generally, we have density matrices, which is the case we review here. Basically, what we do next is investigate the Hilbert space of the aggregate of copies of qubits. We observe that when we consider N copies the collective information is distributed in a particular way given by the decomposition at hand. Then, we can access the information by implementing general quantum measurements.

In particular, using representation theory one can write the product of $\frac{1}{2}$ spaces in terms of invariant subspaces

$$\left(\frac{1}{2}\right)^{\otimes N} = \bigoplus_{j,\alpha} \mathbf{j}^{(\alpha)}, \tag{3.57}$$

where $j = 0(1/2), \ldots, J = N/2$ for even (odd) N, and α labels the different equivalent irreducible representations $\mathbf{j}$. The density operator $\rho^{\otimes N}$ written in the invariant subspaces can be expressed in the block-diagonal form

$$\rho^{\otimes N} = \bigoplus_{j,\alpha} \rho_j^{(\alpha)}, \tag{3.58}$$

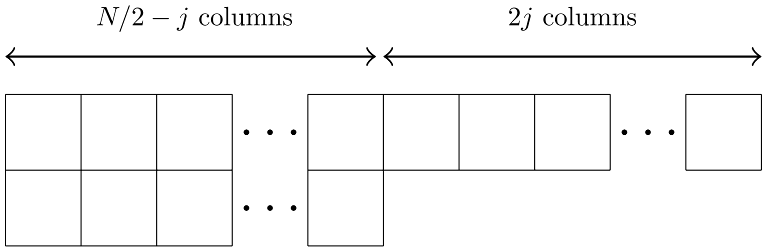


Figure 3.1: A Young diagram with N boxes.

where $\rho_j^{(\alpha)}$ represents the block associated with the subspace $\mathbf{j}^{(\alpha)}$.

The explicit form of the blocks can be easily obtained by analysing the Young diagrams that can be constructed with N boxes, one for each qubit as, for example, the one shown in Figure (3.1). There will be as many different $\mathbf{j}$ as Young diagrams.

Specifically, what this means is that we can write the tensor product of n qubits as

$$\rho^{\otimes n} = \sum_j p_j^n \rho_j \otimes \frac{\mathbb{I}_j}{\nu_j^n}. \tag{3.59}$$

What happens is that the subspaces repeat each other certain number of times. We will call multiplicity the number of times a subspace repeats itself. We write $j = 0(1/2), \ldots, n/2$ if n is even (odd), $\mathbb{I}_j$ is the identity in the multiplicity space $\mathbb{C}^{\nu_j^n}$. The multiplicity ν_j^n is given by

$$\nu_j^n = \binom{n}{n/2 - j} \frac{2j+1}{n/2 + j + 1}. \tag{3.60}$$

The multiplicity can be calculated from the Young tableaux of a given number of qubits [SN17].

The normalized state ρ_j which is in a subspace $S_j = \operatorname{span}\{|j,m\rangle\}$ of dimension $2j + 1 = d_{2j}$ is

$$\rho_j = U_s \left(\sum_{m=-j}^{j} a_m^j [j,m] \right) U_s^\dagger, \tag{3.61}$$

with

$$a^j_m = \frac{1}{c_j}\left(\frac{1-r}{2}\right)^{j-m}\left(\frac{1+r}{2}\right)^{j+m} \tag{3.62}$$

$$c_j = \frac{1}{r}\left\{\left(\frac{1+r}{2}\right)^{2j+1} - \left(\frac{1-r}{2}\right)^{2j+1}\right\}, \tag{3.63}$$

with this $\sum_{m=-j}^{j} a^j_m = 1$ and we use the notation $[j,m] \equiv |j,m\rangle\langle j,m|$. We have that U_s is given by a rotation in SU(2) which are given by the Wigner matrices. If we measure on the various subspaces S_j we will have the state ρ_j as posterior state with probability

$$p^n_j = \nu^n_j c_j \left(\frac{1-r^2}{4}\right)^{n/2-j}. \tag{3.64}$$

3.5.1 Asymptotic quantum relative entropy

A particular application of the structure of an aggregate of states will be to see that the quantum relative entropy can be reached asymptotically in the number of copies. This becomes useful in asymmetric quantum discrimination.

Umegaki defined a relative information quantity between two states ρ and σ as

$$D(\rho||\sigma) = \operatorname{tr}\left[\rho(\ln\rho - \ln\sigma)\right]. \tag{3.65}$$

This concept was introduced in 1962 [Ume62]. It is an analogue with the Kullback-Leibler relative entropy between probability distributions,

$$D(p||q) = \sum_x p(x)\log\frac{p(x)}{q(x)}. \tag{3.66}$$

This also has been named divergence as it is a kind of distance between probability distributions. However, it is not a distance because it is not symmetric as in general $D(p||q) \neq D(q||p)$. Given a POVM $\{E_x\}$ and the states ρ and σ one would have two probability distributions and therefore we define a Kullback-Liebler relative entropy for quantum systems dependent of a POVM as

$$D^{E^n}(\rho||\sigma) = \sum_x \operatorname{tr}\left[\rho E_x\right]\ln\left(\frac{\operatorname{tr}\left[\rho E_x\right]}{\operatorname{tr}\left[\sigma E_x\right]}\right). \tag{3.67}$$

The quantum relative entropy defined in (3.65) is an upper bound to any Kullback-Liebler divergence obtained this way [CT12].

Nevertheless, the quantum relative entropy can be attained asymptotically specifically, it can be shown as a theorem [Hay01]:

Theorem 5. *Let k be the dimension of H and let σ be a state on H. Then there exists a POVM M^n on the tensored space $H^{\otimes n}$ which satisfies*

$$D(\rho||\sigma) - \frac{(k-1)\log(n+1)}{n} \leq \frac{1}{n} D^{M^n}(\rho^{\otimes n}||\sigma^{\otimes n}) \leq D(\rho||\sigma) \ \ \forall \ \rho. \tag{3.68}$$

A measurement that achieves this behavior comes from knowing the decomposition of the tensor product $\rho^{\otimes n}$ into irreps. In the qubit case it means to measure the whole spin of the aggregate and then the projection into the z-axis [Hay01].

Therefore, the quantum relative entropy is defined asymptotically, as when $n \to \infty$. This means, that it is a quantity that tells us the rate of the error in the limit of many copies. For the Sequential Analysis article [MVHS+21] we are also in this asymptotic limit and the quantum relative entropy naturally appears as the rate of the average number of copies needed when the error rates are asymptotically small.

“In economics, things take longer than you think, and then they happen faster than you thought they could.”

— Rüdiger Dornbusch

4

Quantum detection in time I

Here we introduce the problem of Change Point detection from a quantum mechanics perspective. This problem originates classically only considering the statistics of a source; we will only see some aspects of the classical version of this problem. We will center the exposition of the Change Point problem to the quantum case as it is simple enough. Basically, the whole problem stands in a phenomenological point of view, which means, that we are given a black box source that we want to characterize from its outputs and nothing more. However, minimal information is given about the source: we know that it produces a base state $|0\rangle$ and, suddenly, produces another state $|\phi\rangle$ which has nonzero overlap with the original state, i.e. $\langle 0|\phi\rangle = c \neq 0$.

What we have then is a machine that *changes* the state it produces. The problem then is, having N states produced by this kind of machine, figure out when does the change happen. Our task for detection is to devise the best possible measurement apparatus to figure out the position of the change in a sample of N particles in the given states.

The problem can be cast as an N-hypotheses discrimination problem when we write each hypothesis global state of N particles as

$$|\psi_k\rangle = |0\rangle^{\otimes k-1} |\phi\rangle^{\otimes N-k+1}, \tag{4.1}$$

which would mean that the change happens at position k. It is very convenient

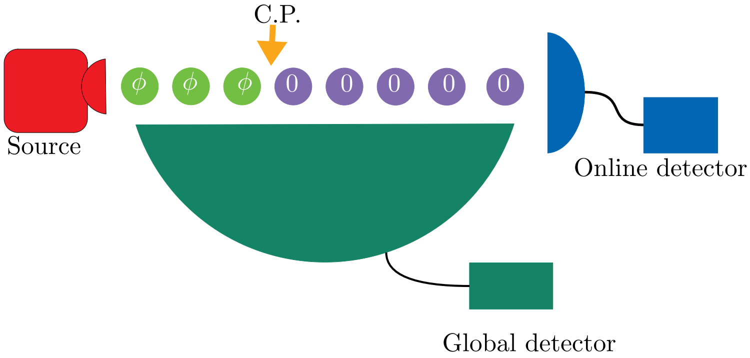


Figure 4.1: Schematics of the Change Point problem.

to write this problem in terms of its Gram matrix, as defined in section (3.3.1). Observe that

$$\langle \psi_l | \psi_k \rangle = c^{|k-l|}. \tag{4.2}$$

Therefore, this Gram matrix has a very specific form.

$$G = \begin{pmatrix} 1 & c & c^2 & \dots & c^{N-1} \\ c & 1 & c & \dots & c^{N-2} \\ c^2 & c & 1 & \dots & c^{N-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c^{N-1} & c^{N-2} & c^{N-3} & \dots & 1 \end{pmatrix}$$

In Figure (4.1) we have a depiction of two cases of the Change Point problem. The first one is with an online detector and the most general case that considers the sample of N outcomes as a whole. Now, what is the optimal measurement that we are capable of getting to discriminate the hypotheses? For a given sample size N we can observe that the square root measurement is a pretty good one and tends to the optimal when $N \to \infty$. Using theorem (4) from section (3.3.1) we can get the lower bound

$$P_{max} \leq (\text{tr}\,\sqrt{G}/n)^2 + 4\left(\frac{1+c}{1-c}\right)^{3/2} \frac{1}{n^{1-\epsilon}}, \tag{4.3}$$

where $\epsilon > 0$ is an arbitrary constant and $\sqrt{G}$ denotes the square root matrix of G [SBC+16]. This is the protocol without any constraint, which corresponds to doing a Minimum Error protocol. The case when the protocol is unambiguous has also been studied [SCMnT17]. In the unambiguous case, the problem is highly constrained as the POVM elements are fixed with the exemption of an overall numeric factor which can be optimized.

In this thesis, it will be very important to study the semidefinite program associated with the unambiguous discrimination problem. From the study of this program we can get more information from the output of a Change Point machine. The SDP for unambiguous discrimination was introduced in section (3.4).

4.1 Online strategies for exactly identifying a quantum change point

After N outputs of the machine, the result of unambiguous detection of the Change Point using the SDP (3.55) gives an answer with global measurements which means that one needs to consider in principle the whole sample of N outputs as in Eq. (4.1). Notice that Figure (4.1) suggests that the change point could also be detected in an online way. In the first publication we asked how well optimal online strategies compare to the optimal global one. The POVM elements $E_i \in H^{\otimes N}$ of the optimal measurement can be entangled operators. It is in fact nontrivial to differentiate when an entangled operator has an advantage over a protocol that has access only to LOCC [CLM+14].

The global unambiguous protocol is an N-hypotheses discrimination problem. However, the LOCC measurement consists on a local two-outcome POVM. In either case, the POVM operators are fixed and we have to optimize the scalar weights that we assign to them. The LOCC protocol is an optimal two outcome unambiguous discrimination one at each step, what is not trivial is how to modify optimally the priors and the weights of the POVM elements after each step. We find the optimal priors after each measurement. Notice that it is by no means obvious a priori that the online two-outcome protocol should coincide to the global N-outcome one.

However, the result here is that, in fact, the unambiguous version of the change point problem is one of the particular instances where there is *no gain* with an entangled POVM over an online one for an overlap of the states in the interval $[0, 1/2]$. Outside this range the online protocol does very well but not optimally. This means, the optimal protocol is also implementable in an online fashion in a wide range of overlaps which is very convenient in the practical sense. Notice here that a protocol being online is a stronger requirement than being a local one as it requires optimality at each step.

Outside the discussed interval, i.e. $c > 1/2$, the problem is analyzed for a fixed local strategy, which means, the weights do not change after each measurement. It is found that this strategy is optimal asymptotically in the number of copies.

4.2 Certified answers for ordered quantum discrimination problems

The other question that we address here concerns the implicit order of the set of states of the change point problem. It turns out that one can use this knowledge to make an analysis of the quality of the answers that a measurement protocol can yield.

Here we return to a global setup like that of Eq. (3.55). What happens is that the unambiguous protocol has also a spatial coordinate for a special kind of problems. There is a one-to-one relation between the states and the possible position of the change point. If we take the set of N states of an output of a machine ordered in a line then there is a concrete sense of the position of the first altered state $|\phi\rangle$.

The unambiguous protocol asks for a POVM that makes no mistake in the position of the change point. We can ask for a less restrictive protocol by modifying the error function. We make explicit use of the spatial relationship of the hypotheses states. The modified error will be a cost function that for the change point at hypothesis k it assigns zero cost to erroneously guessing that the change point happened at positions $k+1$ or $k-1$. All other erroneous hypotheses are not allowed (or have an infinite cost). The protocol thus changes and one can obtain a higher probability of success in this scheme.

After going away from the known unambiguous protocol by one position, we can then ask for cost functions that allow errors of length 2 and more. We define the SDP for these cases until we allow all possible errors, which coincides with the minimum error problem. These SDPs yield a “certified answer” in the sense that they allow a constrained amount of errors for the change point. These kind of constrictions are illustrated in Figure (4.2).

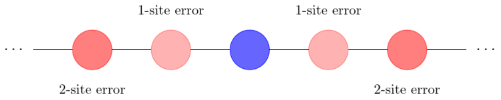


Figure 4.2: Types of errors considered by our schemes.

The protocol that we introduce is more related to the error-margin protocols [HHH08, SBCMnT13], however, here we only deal with position instead of limits on the admissible probability of making an error. The protocols that we introduce generalize the SDP in Equation (3.55).

Having the problem in an SDP form is useful not only for numerical calculations but also because we can get theoretical insight for an analytical solution. In our case, the insight we get is that if one knows the solution to the minimum error protocol (where all errors are allowed) one can get a nontrivial lower bound on the general solution for the SDP. The trick is to give a more restrictive auxiliar SDP. This auxiliar SDP is not optimized, but a feasible solution is given, therefore we get a lower bound on the original problem. This is a very general method that applies to any ordered problem.

An important aspect of the change point problem is that it is an ordered one. Therefore we could consider problems that are ordered in another sense. Here, we also take the example of the Anomaly Detection problem, which considers an array of states that does not necessarily imply time ordering. The Anomaly Detection problem is like the change point problem just that when the machine produces the state ϕ, the next state it produces is 0 again as all the following ones. We also obtain an analytic lower bound to the probability of success in this problem.

Online strategies for exactly identifying a quantum change point

Gael Sentís[1],* Esteban Martínez-Vargas[2],† and Ramon Muñoz-Tapia[2]‡
[1] *Naturwissenschaftlich-Technische Fakultät, Universität Siegen, 57068 Siegen, Germany*
[2] *Física Teòrica: Informació i Fenòmens Quàntics, Departament de Física,*
Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona), Spain

We consider online detection strategies for identifying a change point in a stream of quantum particles allegedly prepared in identical states. We show that the identification of the change point can be done without error via sequential local measurements while attaining the optimal performance bound set by quantum mechanics. In this way, we establish the task of exactly identifying a quantum change point as an instance where local protocols are as powerful as global ones. The optimal online detection strategy requires only one bit of memory between subsequent measurements, and it is amenable to experimental realization with current technology.

I. INTRODUCTION

The ability to process streaming data on-the-fly and promptly detect changes in trends has become a most desirable feature of modern data-analysis algorithms. Change point detection is a vast branch of statistical analysis [1, 2] devoted to techniques for uncovering abrupt changes in the underlying probability distribution that generates a stream of stochastic data. Applications are far-reaching, including quality control [3], medical diagnosis [4], and robotics [5]. Generically, there are two distinct approaches for detecting change points: *offline* strategies that require availability of a complete time series of data, and *online* strategies that are able to process data sequentially. Naturally, having access to the full data history of a given stochastic process typically results in higher change-point identification rates. On the other hand, online strategies enable real-time decision making, are more versatile, and require less memory. These are most relevant in machine learning, for devising online algorithms with effective mechanisms to address learning in the context of non-stationary distributions, a problem known as concept drift [6].

The first extension of the change point identification problem, in its simplest formulation, into a quantum setup was recently introduced in Refs. [7, 8]. The problem can be stated as follows. A source assumed to prepare a sequence of quantum particles in identical states suffers a sudden alteration at some unspecified point, after which the particles are prepared in a mutated state. Given a sequence of particles, one aims at detecting when the mutation took place. In the most fundamental setting, that we also consider here, the initial and final states are assumed to be pure and known and, for a given sequence of length n, all potential positions of the change point in the sequence are expected to happen with equal probability. In Ref. [7], the minimum probability of erroneously identifying a quantum change point and a strategy that achieves it were obtained. This optimal strategy consists in a quantum measurement acting coherently on the given sequence of n particles. It was also shown that a fairly general class of online strategies, based on sequential adaptive measurements on each individual quantum particle, underperform the optimal protocol, and strong numerical evidence that this is the case for all online strategies was provided. The experimental implementation of adaptive online strategies for change point detection has been very recently demonstrated [9]. In contrast, Ref. [8] addressed the quantum change point problem from a different approach: when no identification errors are allowed. The identification protocol then has two possible outcomes, either a correct answer or an inconclusive one [10], and optimality means achieving a minimal rate of inconclusive outcomes. This scenario covers situations where, after the identification of a change point, a response action shall be taken only in conditions of absolute certainty. The optimal procedure and its associated optimal success probability were derived analytically for any length n and arbitrary states [8]. Again, this optimal protocol would in principle require a coherent quantum measurement over the full sequence of particles, and hence also quantum memories to store them, which may render the protocol impractical in some scenarios. In this paper, we look into online strategies for exactly identifying a change point in streaming quantum data. Some simple online protocols were already considered in Ref. [8] and shown to significantly underperform the optimal global protocol. Here, we deepen the analysis and address more general online strategies by allowing classical communication between local measurements. Contrary to our initial conjecture [8], we find the striking result that there is an online strategy that does achieve optimal performance up to a critical value of the overlap between the reference and mutated state. We also obtain that only one bit of memory is required at each measurement step to achieve optimality: it is enough to know whether the previous result was inconclusive. Our results hence imply that the exact optimal identification of a quantum change point is a readily implementable task with current technology, thus prone to integration within diverse quantum information processing protocols.

* gael.sentis@uni-siegen.de
† esteban.martinez@uab.cat
‡ ramon.munoz@uab.cat

We begin by setting the notation and briefly reviewing the results for the optimal (global) strategy in Section II. Then, we turn to online strategies in Section III and present our core results. We show that these provide optimal performance in a given range of the overlap parameter. Beyond this range, we show that the best online strategy is, albeit suboptimal, very close to optimality. We finish in Section IV with a short discussion.

II. OPTIMAL GLOBAL STRATEGY

Let us denote by $|0\rangle$ the default state, $|\phi\rangle$ the mutated state, and $c=\langle 0|\phi\rangle$ their overlap. Without loss of generality, we take c real and non-negative. Given a sequence of n particles, the change point identification corresponds to identifying a state within the set of equally likely source states $\{|\Psi_k\rangle\}_{k=1}^n$, where

$$|\Psi_k\rangle = |\underbrace{0\ldots 0}_{k-1}\underbrace{\phi\ldots\phi}_{n-k+1}\rangle \qquad (1)$$

is associated with the change point occurring at position k. A strategy that unambiguously identifies the correct source state is characterized by a positive operator valued measure (POVM) with $n+1$ elements $\{E_l\geq 0\}_{l=0}^n$. The outcomes $l=1,\ldots,n$ detect without error each possible source state, i.e., the corresponding POVM elements fulfill $\langle\Psi_k|\,E_l\,|\Psi_k\rangle=0$ for $k\neq l$, and the remaining element $E_0=\mathbb{1}-\sum_{k=1}^n E_k\geq 0$ corresponds to the inconclusive outcome. Since the source states (1) are linearly independent, it is possible to find a set of orthogonal states $\{|\tilde\Phi_k\rangle\}_{k=1}^n$ such that $\langle\tilde\Phi_l|\Psi_k\rangle=\delta_{kl}$ (the tilde indicates that these states are not normalized in general). These states can be compactly written as $|\tilde\Phi_k\rangle=\Omega^{-1}|\Psi_k\rangle$, with $\Omega=\sum_k|\Psi_k\rangle\!\langle\Psi_k|$, where the inverse Ω^{-1} has to be understood in the pseudoinverse sense [12] if necessary. Then, the POVM elements of the unambiguous measurement simply read $E_l=\gamma_l|\tilde\Phi_l\rangle\!\langle\tilde\Phi_l|$, $l=1,\ldots,n$, where the parameters $0\leq\gamma_l\leq 1$ are the conditional success probabilities of identifying each source state. We will refer to γ_l as *efficiencies* [13]. The success probability of identifying a change point without error is given by $P_s=\frac{1}{n}\sum_{k=1}^n\gamma_k$, and the efficiencies γ_k, the only free parameters left to be optimized, are constrained by the condition $E_0\geq 0$.

The optimal efficiencies, up to a certain critical value c^* of the overlap, are [8]

$$\gamma_n(k)=\sum_{j=1}^n(-c)^{|k-j|}\,,\quad k=1,\ldots,n\,, \qquad (2)$$

where we have explicitly included the dependence on the number n of particles and written γ_k as $\gamma_n(k)$. The corresponding optimal success probability reads

$$P_s=\frac{1}{n}\sum_{k=1}^n\gamma_n(k)=\frac{1-c}{1+c}+\frac{1}{n}\frac{2c\,[1-(-c)^n]}{(1+c)^2}\,. \qquad (3)$$

This expression is valid in the range $0\leq c\leq c^*$, where $c^*\approx(\sqrt{5}-1)/2$ is determined by the equation $\gamma_n(2)=0$. In the rest of the range, $c^*\leq c\leq 1$, the optimal efficiencies and success probability read [8]

$$\gamma'_n(k)=\gamma_n(k)-\gamma_n(2)\frac{(-c)^{|k-2|}+(-c)^{|n-k-1|}}{1+(-c)^{n-3}} \qquad (4)$$

and

$$P'_s=\frac{1}{n}\sum_{k=1}^n\gamma'_n(k)=P_s-\frac{2}{n}\frac{\gamma_n^2(2)}{1+(-c)^{n-3}}\,, \qquad (5)$$

respectively.

III. ONLINE STRATEGIES

The optimal solution, comprised by Eqs. (3) and (5), in principle requires a global measurement on the whole set of n particles that may be infeasible to implement in practice. It is therefore of interest to elucidate whether the task can be achieved with online strategies that act locally on each particle, possibly assisted by classical communication between measurements, and how does their performance compare to the optimal one. Such strategies are far easier to implement in practice, and, additionally, would allow for the detection of a change point in a stream of quantum particles as soon as it occurs. We will show that, quite extraordinarily, there is a simple online protocol that performs optimally for $0\leq c\leq 1/2$ and needs to store only the outcome of the last measurement at each step. In this overlap range, this result holds true for sequences of arbitrary length n. For $c>1/2$ the best online protocol does not attain the optimal success probability, although it is remarkably close.

A change point at position k can be exactly identified by a local protocol only if there are two successive unambiguous detections: $|0\rangle$ at position $k-1$, followed by $|\phi\rangle$ at position k. For the end-point case $k=1$ one only requires the detection of state $|\phi\rangle$ at the first position, while for the last change point position, $k=n$, detecting $|0\rangle$ at position $n-1$ suffices since it is assumed that a change point has always occurred and, hence, the state of the last particle is necessarily $|\phi\rangle$ [?]. To lighten the presentation, from now on we will simply write 'detection' or 'detect' for 'unambiguous detection' or 'unambiguously detect'.

Let $\mathcal{M}_n$ be a local measurement strategy for strings of n particles, where each local measurement has three possible outcomes: 0, ϕ, and I, which correspond to detecting $|0\rangle$, $|\phi\rangle$, and an inconclusive result, respectively. Let Θ_j be a particular set of outcomes of the first j measurements. Then, the sequence of outcomes $(\Theta_{k-2},0_{k-1},\phi_k)$ leads to the detection of a change point at position k. The probability of a successful detection of the change point given the source state $|\Psi_k\rangle$ and a measurement strategy

$\mathcal{M}_n$, that we name local efficiency for position k, reads

$$D_n(k) := \sum_{\Theta_{k-2}} \Pr[(\Theta_{k-2}, 0_{k-1}, \phi_k)|\Psi_k, \mathcal{M}_n] \,, \qquad (6)$$

and the average success probability is given by

$$P_s^{\mathrm{L}} = \frac{1}{n}\sum_{k=1}^{n} D_n(k) \,. \qquad (7)$$

We characterize next the local measurements that comprise a strategy $\mathcal{M}_n$. An optimal measurement that unambiguously discriminates between two states $|0\rangle$ and $|\phi\rangle$ that are assumed to occur with prior probabilities η_0 and η_ϕ, respectively, succeeds with conditional probability $1-c\sqrt{\eta_\phi/\eta_0}$ if the state was $|0\rangle$ and $1-c\sqrt{\eta_0/\eta_\phi}$ if it was $|\phi\rangle$ [8]. Therefore each local measurement is determined by a strength parameter $x := \sqrt{\eta_\phi/\eta_0}$ that specifies its bias towards detecting $|0\rangle$ or $|\phi\rangle$. In terms of x and c, the local conditional probabilities Pr(outcome|state) read $\Pr(0|0) = 1 - cx$, $\Pr(I|0) = cx$, $\Pr(\phi|\phi) = 1 - c/x$, and $\Pr(I|\phi) = c/x$. Obviously, $\Pr(0|\phi) = \Pr(\phi|0) = 0$. The positivity of these probabilities bounds the strength parameter to the interval $c \le x \le 1/c$. The extreme value $x = c$ ($x = 1/c$) corresponds to an effective two-outcome measurement that either detects $|0\rangle$ ($|\phi\rangle$) or yields an inconclusive answer, and any other intermediate value of x represents a three-outcome measurement. An optimal local measurement strategy $\mathcal{M}_n$ is a sequence of $n-1$ unambiguous measurements that maximizes Eq. (7).

We address the problem of finding the optimal $\mathcal{M}_n$ by considering general adaptive strategies that take into account the information learned in previous measurements. We introduce this feature by letting the strength of the measurement over particle j generically depend on all past outcomes $\mathbf{r}_{j-1} = \{r_1, \ldots, r_{j-1}\}$, that is, $x(j; \mathbf{r}_{j-1})$. Note that $\mathbf{r}_{j-1}$ cannot contain any outcome ϕ, as the procedure stops after obtaining the first ϕ. Thus, $\mathbf{r}_{j-1}$ is a binary string of 0's and I's. This is the most general one-way local-operations-and-classical-communication (LOCC) protocol that one can devise [16]. Optimizing LOCC protocols is in general unfeasible, since the number of parameters grows exponentially with n. However, for the problem at hand, this number is effectively reduced to $n-1$ and thus the optimization can be tackled efficiently. This exponential reduction is a direct consequence of the logic behind unambiguous measurements: after obtaining an outcome 0 at position j, one knows for a fact that all particles of the string up to the jth position were in the state $|0\rangle$, therefore any information that previous outcomes may provide is superseded. Further, if the outcome of the measurement over the jth particle is I, the following optimal measurement strength is fixed to detect only the state $|0\rangle$, since the sequence of outcomes $I\phi$ irremediably implies the failure of the protocol. These observations are condensed in the equations $x(j; r_{j-1} = 0) =: x(j)$, $x(j; r_{j-1} = I) = c$, hence the free parameters of a general adaptive strategy $\mathcal{M}_n$ is just the set of strengths $\{x(j)\}_{j=1}^{n-1}$ of measurements that are preceded by an outcome 0.

To gain intuition on the general solution, we first show the explicit construction of the optimal strategy for $n = 4$. The conditional detection probabilities are

$$D_4(1) = 1 - \frac{c}{x(1)} \,, \qquad (8)$$

$$D_4(2) = [1 - c\,x(1)] \left[1 - \frac{c}{x(2)}\right] \,, \qquad (9)$$

$$D_4(3) = [1 - c\,x(1)]\,[1 - c\,x(2)] \left[1 - \frac{c}{x(3)}\right] + c\,x(1)(1-c^2) \left[1 - \frac{c}{x(3)}\right] \,, \qquad (10)$$

$$\begin{aligned} D_4(4) = {} & [1 - c\,x(1)]\,[1 - c\,x(2)]\,[1 - c\,x(3)] \\ & + c\,x(1)(1-c^2)\,[1 - c\,x(3)] \\ & + [1 - c\,x(1)]\,c\,x(2)(1-c^2) + c\,x(1)c^2(1-c^2) \,. \end{aligned} \qquad (11)$$

Each summand in $D_4(k)$ corresponds to the probability of a string of outcomes leading to detection of the change point at position k. For instance, $D_4(4)$ comprises the strings 000, I00, 0I0, and II0. The maximization of Eq. (7) leads to the optimal strengths

$$x(1) = \frac{1}{1-c+c^2},\; x(2) = \frac{1}{1-c},\; x(3) = 1 \,. \qquad (12)$$

The first key observation is that the optimal local efficiencies match the optimal global efficiencies for each change point, and, therefore, $P_s^{\mathrm{L}} = P_s$. Indeed, inserting Eq. (12) into Eqs. (8) to (11), one obtains $D_4(k) = \gamma_4(k)$ for $k = 1, \ldots, 4$, where $\gamma_4(k)$ is given in Eq. (2). The second key observation is that this solution only holds for $c \le 1/2$, since outside this range $x(2) > 1/c$ and hence it does not yield a valid measurement. Further, the optimal value $x(3) = 1$ can be easily understood: conditioned to having obtained $r_2 = 0$, the probability of the third particle being in the state $|0\rangle$ or $|\phi\rangle$ is $1/2$, hence the optimal choice is a symmetric measurement. We will see that these features remain valid in the general case.

A. Optimal online protocol

Let us now present the solution for the optimal strength parameters and detection probabilities for arbitrary n. It is convenient to write the explicit dependence on the total number of particles, i.e., $x(j)$ as $x_n(j)$. As discussed before, obtaining an outcome 0 at position $j-1$ discards all hypotheses with a change point at position $k \le j-1$, effectively resetting the problem to one with a change point in a string of $n-j+1$ particles. Hence, we have that $x_n(j) = x_{n-j+1}(1)$ holds for optimal strengths. Now, we follow the intuition from the $n = 4$ problem that, in case there is no performance gap between the optimal global and local strategies, the global and local efficiencies should match one by one. This leads us to the equation $D_m(1) = 1 - c/x_m(1) = \gamma_m(1)$ [cf. Eq. (8)]. Using the explicit value of $\gamma_{n-j+1}(1)$ from Eq. (2), we obtain

$$x_n(j) = x_{n-j+1}(1) = \frac{1+c}{1-(-c)^{n-j}}\,, \quad j = 1, \ldots, n-1\,. \tag{13}$$

Note that this formula reduces to Eq. (12) for $n = 4$, and that it is a solution for the set of equations $\{D_n(k) = \gamma_n(k) : k = 1, \ldots, n-1\}$. In Appendix A we provide a proof by induction of Eq. 13. We also note that $x_n(n-2) = 1/(1-c)$ is the first strength to saturate at the extreme value $1/c$ with increasing c, hence this general solution is still only valid up to $c = 1/2$. In summary, for overlaps $0 \leq c \leq 1/2$, the optimal online strategy consists in a sequence of unambiguous measurements of strengths $x_n(j)$ given by Eq. (13) if the outcome of the measurement on the previous particle is 0, and fixed strengths c if the previous outcome is I. This online protocol attains the performance of the optimal (global) strategy, given by Eq. (3).

B. Beyond $c > 1/2$

We now analyze the optimal local strategy for $c > 1/2$. It is clear that local strategies cannot reach optimal performance in this range of overlaps, as this would require the expressions of the strengths (13) remain valid beyond their upper limit $1/c$. The optimization of local protocols is much more constrained than that of a global strategy and, hence, a smaller feasibility region is to be expected. As c increases, there is a progressive saturation of the strengths, starting with $x_n(n-2)$. The exact saturation point for each strength can be computed from the techniques shown in Appendix B, as well as the point $c_S \approx 0.69$ where all strengths but the last one [always fixed to be $x_n(n-1) = 1$] are saturated at $x_n(k) = 1/c$. Beyond c_S, the optimal online strategy is a sequence of two-outcome unambiguous measurements, aiming at the detection of 0 (ϕ) if the previous outcome was I (0).

The exact expressions for the optimal local protocol in the intermediate region $1/2 < c < c_S$ are rather impractical. Instead, we provide a simpler local protocol that we prove to be optimal for large n. By doing so, we also discover that online protocols can still attain optimal global performance beyond $c = 1/2$ and up to $c = c^*$, precisely the value that divides the two regimes in the global approach. We consider the simple local strategy with constant strengths $x_n(k) = x$ after a 0 outcome and, of course, a strength c after an I outcome. In Appendix C we show that the success probability of such strategy for large n reads

$$P_s \simeq \frac{1-c^2}{1+cx-c^2}\left(1-\frac{c}{x}\right)\,, \tag{14}$$

which is maximal for $x = 1 + c$. Note that we could have anticipated this result, as it corresponds to the approximation of Eq. (13) for large n. The maximal success probability for local strategies with constant strengths then reads

$$P_s^{\mathrm{FL}} \simeq \frac{1-c}{1+c}\,, \tag{15}$$

which coincides with the optimal asymptotic value in Eqs. (3) and (5) (the superscript FL stands for fixed local). The choice $x = 1 + c$ yields a valid a measurement up to $c^* \approx 0.61$, a value that is determined by the saturation condition $1 + c = 1/c$.

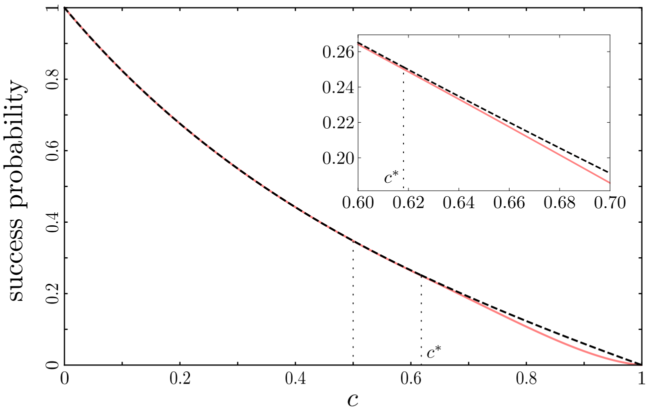


FIG. 1. Probability of exact identification of a change point as a function of $c = |\langle 0|\phi\rangle|$ for a string of $n = 31$ particles. The black dashed line is the optimal success probability when global strategies are considered, given by the piecewise function composed by Eqs. (3) and (5). The pink solid line corresponds to the success probability of different online detection strategies, depending on the value of c. In the interval $0 \leq c \leq 1/2$, the online strategy characterized by the strengths in Eq. (13) matches exactly the optimal performance. Beyond $c = 1/2$ the FL strategy is optimal only asymptotically and up to $c = c^*$, although the difference for $n = 31$ of around $\sim 0.1\%$ is hardly appreciated. For $c > c^*$ the success probability of the SL strategy starts to deviate from optimality and shows a finite gap even in the asymptotic limit. The inset plot highlights this regime transition of online strategies around c^*.

For $c > c^*$, the constant strengths saturate to $x = 1/c$ and Eq. (14) reads

$$P_s^{\mathrm{SL}} \simeq \frac{(1-c^2)^2}{2-c^2}\,, \tag{16}$$

where SL stands for saturated local. The success probability deviates from the optimal value given in Eq. (15), but the difference is no larger than 2.2% in the worst case [**?**]. The closeness of online protocols to optimal performance is patent in Fig. 1, where we represent the average success probability of the best online strategy in each overlap regime together with the optimal one for a string of length $n = 31$.

IV. CONCLUSIONS

Let us conclude by reviewing our results and its implications. In this work, we have derived the optimal local protocol that unambiguously detects a quantum change point. We have shown that it attains exactly the performance of an optimal global protocol for arbitrary string lengths and for values of the overlap c below $1/2$. Our results provide not only one of the few non-trivial examples of state identification tasks where the optimal protocol can be found, but also the first instance with arbitrarily many hypotheses where one-way LOCC measurements match optimal performance (see Refs. [14, 15] and [17, 18] for such instances for binary and ternary discrimination, respectively). Besides this remarkable feature, the LOCC protocol has several attractive aspects: (i) it is an online protocol, i.e., in case the change point is detected, it is as soon as it appears; (ii) no quantum memories are required; (iii) the necessary measurements are all local and, hence, easy to implement experimentally; and (iv) the memory required for the adaptive selection of subsequent measurements amounts to just one bit (encoding whether the previous outcome was conclusive or not), which may benefit the stability and robustness of an experimental setup.

We have also analyzed how above $c = 1/2$ the problem becomes too constrained for any online strategy to attain global optimality for strings of arbitrary finite length, although the performance gap is very small. Despite this, for large n, we have shown that optimal global performance can still be attained for overlaps up to $c^* \approx 0.61$ by an online fixed-strength strategy. Beyond c^*, the best local protocol essentially consists in a sequence of two-outcome measurements that detect just one of the local states: either $|0\rangle$ or $|\phi\rangle$. We have shown that such protocol deviates from the optimal performance only by 2.2% in the worst case.

Finally, it is worth mentioning that our results are amenable to experimental realization with current technology, as the experimental implementation of the necessary unambiguous measurements has already been demonstrated in optical platforms [19–22].

ACKNOWLEDGMENTS

This research was supported by the Spanish MINECO through contract FIS2016-80681-P, the ERC Consolidator Grant 683107/TempoQ and the DFG. EMV acknowledges financial support from CONACYT.

[1] B.E. Brodsky and B.S. Darkhovsky, *Non-Parametric Statistical Diagnosis* (Springer-Science+Business Media, B.V, Dordrecht, 2000).

[2] M. Baseville and I.V. Nikiforov, *Detection of Abrupt Changes: Theory and Application*, Prentice Hall Information and System Science Series, (Prentice Hall, New Jersey, 1993).

[3] T.L. Lai, Journal of the Royal Statistical Society. Series B (Methodological), **57**, 613 (1995).

[4] D. Rosenfield, E. Zhou, F.H. Wilhelm, A. Conrad, W.T. Roth,, and A.E. Meuret, Biol Psychol. 84, 112 (2010).

[5] A. Ranganathan, *Pliss: Detecting and labeling places using online change-point detection*, Robotics: Science and Systems VI (MIT Press, Cambridge, 2011).

[6] G.I. Webb, R. Hyde, H. Cao, H.L. Nguyen, and F. Petitjean, Data Min. Knowl. Discov. 30, 964 (2016).

[7] G. Sentís, E. Bagan, J. Calsamiglia, G. Chiribella, and R. Munoz-Tapia, Phys. Rev. Lett. **117**, 150502 (2016).

[8] G. Sentís, J. Calsamiglia and R. Munoz-Tapia, Phys. Rev. Lett. **119**, 140506 (2017).

[9] Shang Yu *et al.*, arXiv:1801.07508.

[10] I. D. Ivanovic, Phys. Lett. A **123**, 257 (1987); D. Dieks, Phys. Lett. A **126**, 303 (1988); A. Peres, Phys. Lett. A **128**, 19 (1988); A. Chefles, Phys. Lett. A **239**, 339 (1998).

[11] S. Boyd and L. Vandenberghe, *Convex Optimization* (Cambridge University Press, Cambridge, 2004).

[12] R.A. Horn and C.R. Johnson, *Matrix Analysis* 2nd. Ed. (Cambridge University Press, Cambridge, 2013).

[13] L.M. Duan and G.C. Guo, Phys. Rev. Lett. **80**, 4999 (1998).

[14] A. Acin, E. Bagan, M. Baig, L. Masanes, and R. Munoz-Tapia, Phys. Rev. A **71**, 032338 (2005).

[15] Z. Ji, H. Cao, and M. Ying, Phys. Rev. A **71**, 032323 (2005).

[16] E. Bagan, M. Baig, and R. Munoz-Tapia, Phys. Rev. Lett. **89**, 277904 (2002).

[17] A. Chefles, Phys. Rev. A **64**, 062305 (2001).

[18] K. Nakahira, K. Kato, and T.S. Usuda, arXiv:1806.08784 [quant-ph] (2018).

[19] M. Mohseni, A.M. Steinberg, and J. A. Bergou, Phys. Rev. Lett. **93**, 200403 (2004).

[20] F.E. Becerra, J. Fan, and A. Migdall, Nat. Commun. **4**, 2028 (2013)

[21] Y.Y. Zhao, N.K. Yu, P. Kurzynski, G.Y. Xiang, C.F. Li, and G.C. Guo, Phys. Rev. A **91**, 042101 (2015).

[22] Z. Bian, J. Li, H. Qin, X. Zhan, R. Zhang, B.C. Sanders, and P. Xue, Phys. Rev. Lett. **114**, 203602 (2015).

Appendix A: Proof of Eq. (13)

In this section we provide a proof by induction of the optimal form of the strengths $x_n(k)$, given by Eq. (13). This optimal form is

$$x_n(k) = \frac{1+c}{1-(-c)^{n-k}} \,. \tag{A1}$$

Let us first establish some additional notation. Recalling Eq. (6), given a local strategy $\mathcal{M}_n$, the efficiency of detection of the change point at position k reads

$$\begin{aligned} D_n(k) &= \sum_{\Theta_{k-2}} \Pr[(\Theta_{k-2}, 0_{k-1}, \phi_k)|\Psi_k, \mathcal{M}_n] \\ &=: \mathbf{P}_n(\Sigma_{k-2}, 0_{k-1}, \phi_k)\,, \end{aligned} \tag{A2}$$

where the sum runs over the 2^{k-2} sets that only contain outcomes 0 and I, and Σ_j denotes all such sets of j outcomes. The argument of $\mathbf{P}_n$ is always to be understood as an ordered, consecutive sequence of outcomes, so we will omit the position subscripts when no confusion arises. As argued in the main text, the optimal local protocol can be obtained by equating each local efficiency to the corresponding global one, i.e., $D_n(k) = \gamma_n(k)$, and solving the resulting system of equations. Recall that the optimal global detection efficiencies read

$$\gamma_n(k) = \sum_{j=1}^{n} (-c)^{|k-j|} = \frac{1 - c - (-c)^k - (-c)^{n-k+1}}{1+c} \tag{A3}$$

for $k = 1, \ldots, n$. We first observe that, for $k < n$, these equations read

$$\mathbf{P}_n(\Sigma_{k-2}, 0, \phi) = \mathbf{P}_n(\Sigma_{k-2}, 0)\left[1 - \frac{c}{x_n(k)}\right] = \gamma_n(k)\,. \tag{A4}$$

We also have that

$$\begin{aligned} \mathbf{P}_n(\Sigma_{k-1}, 0) = \mathbf{P}_n(\Sigma_{k-2}, 0)&[1 - c\,x_n(k)] \\ &+ \mathbf{P}_n(\Sigma_{k-2}, I)(1 - c^2)\,, \end{aligned} \tag{A5}$$

and recall that $\mathbf{P}_n(\Sigma_{k-2}, I) = 1 - \mathbf{P}_n(\Sigma_{k-2}, 0)$. Then, using Eqs. (A4) and (A5) we get the relation

$$x_n(k+1) = c\left[1 - \frac{\gamma_n(k+1)}{(1-c^2) - c\,\gamma_n(k)\,x_n(k)}\right]^{-1}. \tag{A6}$$

The first strength is immediate to derive from the equation $D_n(1) = 1 - c/x_n(1) = \gamma_n(1)$:

$$x_n(1) = \frac{c}{1 - \gamma_n(1)} = \frac{1+c}{1 - (-c)^{n-1}}\,, \tag{A7}$$

where we have used Eq. (A3) for $k = 1$. Using Eqs. (A6) and (A3), we arrive by induction at the formula for the optimal strengths Eq. (A1).

The attentive reader should have noticed that the system of equations $D_n(n) = \gamma_n(n)$ for $k = 1, \ldots, n$ is overconstrained: there are n equations but $n-1$ unknowns $x_n(k)$. The first $n-1$ equations determine univocally all the unknowns, and the last equation, $D_n(n) = \gamma_n(n)$, should be automatically satisfied. This could seem at first sight a rather non-trivial requirement as $D_n(n)$ contains 2^{n-2} summands (such is the size of the set Σ_{n-2}), but the proof is quite straightforward. We recall that $\gamma_n(n-1) = D_n(n-1) = \mathbf{P}_n(\Sigma_{n-3}, 0_{n-2}, \phi_{n-1}) = \mathbf{P}_n(\Sigma_{n-3}, 0_{n-2}, 0_{n-1}) = (1-c)\mathbf{P}_n(\Sigma_{n-3}, 0_{n-2})$, because the last strength takes the symmetric value $x_n(n-1) = 1$. Then,

$$\begin{aligned} D_n(n) &= \mathbf{P}_n(\Sigma_{n-3}, 0, 0) + \mathbf{P}_n(\Sigma_{n-3}, I, 0) \\ &= (1-c)\mathbf{P}_n(\Sigma_{n-3}, 0) + (1-c^2)\mathbf{P}_n(\Sigma_{n-3}, I) \\ &= (1-c)\mathbf{P}_n(\Sigma_{n-3}, 0) + (1-c^2)[1 - \mathbf{P}_n(\Sigma_{n-3}, 0)] \\ &= (1-c^2) - (1-c)\mathbf{P}_n(\Sigma_{n-3}, 0) \\ &= (1-c^2) - \gamma_n(n-1) = \gamma_n(n), \end{aligned} \tag{A8}$$

where the last equality can be easily checked from Eq. (A3).

Appendix B: Construction of optimal local strategies

Here we show a general method to construct an optimal set of strengths for any given n. This method is particularly useful in the range of overlaps $1/2 < c \leq c_S \approx 0.69$, where a mixture of saturated and unsaturated strengths coexist. Given an arbitrary local strategy determined by the set of strengths $\{x_n(k)\}_{k=1}^{n-1}$, we write the maximization conditions $\partial P_s^{\mathrm{L}}/\partial x_n(k) = 0$, where $P_s^{\mathrm{L}} = (1/n)\sum_{k=1}^{n} D_n(k)$ and $D_n(k)$ is given by Eq. (A2). Starting from the last strength, we note that all the terms of P_s^{L} that depend on $x_n(n-1)$ can be written as

$$\begin{aligned} &\mathbf{P}_n(\Sigma_{n-3}, 0_{n-2})\,[\Pr(\phi_{n-1}|0_{n-2}) + \Pr(0_{n-1}|0_{n-2})] \\ &= \mathbf{P}_n(\Sigma_{n-3}, 0_{n-2})\left[1 - \frac{c}{x_n(n-1)} + 1 - c x_n(n-1)\right], \end{aligned} \tag{B1}$$

where the last factor takes the same form for any value of n. The two probabilities inside the brackets in the first line of Eq. (B1) are, respectively, the probabilities of obtaining outcomes ϕ and 0 at position $n-1$ conditioned on an outcome 0 at position $n-2$. Note that both events successfully identify change points at positions $n-1$ and n, respectively. The maximization of Eq. (B1) yields $x_n(n-1) = 1$. This value intuitively makes sense, since measuring the state of the particle at position $n-1$ means that only two equally likely possible change points remain, either at position $n-1$ or at position n. In such binary identification case, it is clear that a balanced measurement is optimal.

Next we write all the terms of P_s^{L} that depend on $x_n(n-2)$, and substitute the value $x_n(n-1) = 1$. We obtain

$$\begin{aligned} \mathbf{P}_n(\Sigma_{n-4}, 0_{n-3})\Big\{1 - \frac{c}{x_n(n-2)} + 2(1-c)[1 - c x_n(n-2)] \\ + (1-c^2) c x_n(n-2)\Big\}. \end{aligned} \tag{B2}$$

Again, the term in brackets is the same for any n, and it determines the optimal value $x_n(n-2) = 1/(1-c)$.

One can proceed recursively, and realize that the optimal value of $x_n(n-j)$ satisfies $x_n(n-j) = x_{n+1}(n+1-j)$. This observation provides an alternative proof that $x_n(k) = x_{n-1}(k-1)$ is not only a feature of the optimal unsaturated strengths [cf. Eq. (A1)], but also holds for optimal strengths in general, even when the extremal conditions $\partial P_s^{\rm L}/\partial x_n(k) = 0$ are not satisfied (which happens when the feasibility constraint $x_n(k) \leq 1/c$ is hit). In this situation, one substitutes the strengths for its extremal value and carries on with the maximization of the next strength.

For the subsequent strengths it is convenient to use the notation $\Sigma_{k_1}^{k_2}$ to denote all the possible strings of outcomes between particle k_2 and k_1. For $x_n(n-3)$, we have $\Sigma_{n-1}^{n-3} = \{II0, I00, 000, I0\phi, 00\phi\}$, and we have to consider the conditional probability $\Pr(\Sigma_{n-1}^{n-3}|0_{n-4})$ with $x_n(n-2) = 1/c$ and $x_n(n-1) = 1$. Solving $\partial \Pr(\Sigma_{n-1}^{n-3}|0_{n-4})/\partial x_n(n-3) = 0$, we obtain

$$x_n(n-3) = \frac{1}{\sqrt{c(2-c)(1-c^2)}} \,. \tag{B3}$$

The saturation condition $x_n(n-3) = 1/c$ has the solution $c =: c_S \approx 0.69$. Checking for several values of n, one sees that $x_n(n-3)$ is always the last strength that reaches the saturation point. This is in accordance with the intuition that the smallest unsaturated strength [of the form in Eq. (A1)] should be the last one to reach $1/c$. Then, c_S corresponds to the total saturation point, defined as the point beyond which all strengths are saturated [naturally with the exception $x_n(n-1) = 1$].

For the following strengths $x_n(k)$ for $k = n-4, n-5, \ldots$, one proceeds by recursively maximizing $\Pr(\Sigma_{n-1}^{n-k}|0_{n-k-1})$, taking into account if the saturation condition is hit for any of the strengths. Notice that only one variable, $x_n(k)$, is maximized at each step, because the strengths $x_n(k+1), x_n(k+2), \ldots, x_n(n-1)$ have been already fixed at their optimized value obtained in previous optimization steps.

Appendix C: The fixed local strategy

Here we include the calculation of the success probability for the fixed local (FL) strategy, and explicitly show that, in its range of validity, it is asymptotically optimal. Let us first rename $\mathbf{P}_n(\Sigma_{k-1}, I) =: G(k)$, and note that $\mathbf{P}_n(\Sigma_{k-1}, 0) = 1 - G(k)$. Then, for a generic local strategy with fixed strengths $x_n(k) =: x$, we can write

$$\begin{aligned} G(k+1) =& \mathbf{P}_n(\Sigma_k, I) = \mathbf{P}_n(\Sigma_{k-1}, I, I) + \mathbf{P}_n(\Sigma_{k-1}, 0, I) \\ =& c^2 \mathbf{P}_n(\Sigma_{k-1}, I) + c\,x\,\mathbf{P}_n(\Sigma_{k-1}, 0) \\ =& c^2 G(k) + c\,x\,[1 - G(k)] \\ =& c\,x - (c\,x - c^2) G(k) \,, \end{aligned} \tag{C1}$$

where we have used that $\mathbf{P}_n(I, I) = c^2 \mathbf{P}_n(I)$ (recall that, after an outcome I, we always apply an extreme two-outcome unambiguous measurement completely biased towards detection of the local state $|0\rangle$), and $\mathbf{P}_n(0, I) = c\,x\,\mathbf{P}_n(0)$. The recursion relation (C1) can be readily solved using the initial condition $G(1) = c\,x$ to give

$$G(k) = c\,x \frac{1 - (c^2 - c\,x)^k}{1 + c\,x - c^2} \,. \tag{C2}$$

We can relate the local detection efficiencies $D_n(k)$ to the function $G(k)$ by looking at Eq. (A2). We obtain

$$\begin{aligned} D_n(k) =& G(k-2)(1-c^2)\left(1 - \frac{c}{x}\right) \\ &+ [1 - G(k-2)](1 - c\,x)\left(1 - \frac{c}{x}\right) \,, \end{aligned} \tag{C3}$$

which is valid for $k = 1, \ldots, n-2$. Note that, while the original definition $G(k) = \mathbf{P}_n(\Sigma_{k-1}, I)$ does not hold physical meaning in the cases $G(-1)$ and $G(0)$, using the functional expression (C2) in Eq. (C3) we recover the correct local efficiencies for the first and second positions, namely $D_n(1) = 1 - c/x$ and $D_n(2) = (1 - c\,x)(1 - c/x)$. The last two local efficiencies have a slightly different expression and cannot be recovered from Eq. (C3). This is so because the last strength is always fixed to $x_{n-1} = 1$ conditioned to having obtained $r_{n-2} = 0$ as a previous outcome, as argued in the main text. In addition, recall that the nth particle is in the state $|\phi\rangle$ by definition and hence there is no need to measure it. Taking this into account, the success probability for the FL strategy can be written as

$$\begin{aligned} P_s^{\rm FL} =& \frac{1}{n} \Bigg\{ \sum_{k=1}^{n-2} G(k-2)(1-c^2)\left(1 - \frac{c}{x}\right) \\ &+ [1 - G(k-2)]\,(1 - cx)\left(1 - \frac{c}{x}\right) \\ &+ \left\{G(n-3)(1-c^2) + [1 - G(n-3)](1 - c\,x)\right\}(1-c) \\ &+ G(n-2)(1-c^2) + [1 - G(n-2)](1-c) \Bigg\} \\ =& \frac{1}{n} \Bigg\{ \sum_{k=1}^{n-2} G(k-2)(1-c^2)\left(1 - \frac{c}{x}\right) \\ &+ [1 - G(k-2)]\,(1 - cx)\left(1 - \frac{c}{x}\right) \\ &+ (1-c)\,[2 - (1-c)G(n-2)] \Bigg\} \,. \end{aligned} \tag{C4}$$

For large n, the leading order of the success probability is

$$\begin{aligned} P_s^{\rm FL} \simeq& \frac{cx}{1 + cx - c^2}(1-c^2)\left(1 - \frac{c}{x}\right) \\ &+ \frac{1-c^2}{1 + cx - c^2}(1 - cx)\left(1 - \frac{c}{x}\right) \\ =& \frac{1}{1 + c\,x - c^2} - \frac{c}{x} \,, \end{aligned} \tag{C5}$$

which just amounts to neglect the exponential terms in Eq. (C2) and the slightly different last term in Eq. (C4).

Note that this asymptotic limit of the success probability would remain invariant if we would choose the same fixed strength $x_{n-1} = x$ for the last measurement too, as opposed to the slightly better choice of a symmetric strength $x_{n-1} = 1$. Eq. (C5) can be easily maximized to obtain

$$x_{max} = 1 + c \quad \rightarrow \quad P_s^{\rm FL} \simeq \frac{1-c}{1+c} \,. \tag{C6}$$

Hence, as anticipated, we obtain that a protocol that measures a particle with a local measurement of fixed strength $x = 1 + c$ if the previous outcome was 0, and $x = c$ if the previous outcome was inconclusive, is asymptotically optimal up to a threshold value of the overlap $c^* \approx 0.61$. This threshold is the solution of the boundary constraint on the fixed strength x, i.e., $1 + c = 1/c$.

Appendix D: Success probability of the Saturated Local strategy

The saturated local (SL) strategy is defined by the fixed strengths $x_n(k) = 1/c$, at the boundary of their physicality interval. The corresponding local efficiencies, $D_n(k)$, up to $k = n - 2$ read

$$D_n(1) = (1 - c^2), \quad D_n(2) = 0\,, \tag{D1}$$

and

$$D_n(k) = (1 - c^2)^2 F(k-2), \quad k = 3, \ldots, n-2, \tag{D2}$$

where the function $F(k)$ can be directly read off of Eq. (C2) particularizing for $x = 1/c$. The exact expression for the success probability and the leading order term in the asymptotic regime of large n are derived likewise from Eqs. (C4) and (C5). The latter reads

$$P_s^{\rm SL} \simeq \frac{(1-c^2)^2}{2-c^2}. \tag{D3}$$

Notice that this value is smaller than the leading term $(1 - c)/(1 + c)$ of the optimal success probability [cf. Eqs. (3), (5), and Eq. (C6)]. The difference is however very small, with a maximal value of 0.022 at $c \approx 0.89$. The asymptotic success probability for the SL strategy, Eq. (D3), equals the optimal value precisely at $c^* = (\sqrt{5} - 1)/2$, below which the FL strategy is asymptotically optimal, as discussed in the main text.

Certified answers for ordered quantum discrimination problems

Esteban Martínez Vargas* and Ramon Muñoz-Tapia†
Física Teòrica: Informació i Fenòmens Quàntics, Departament de Física,
Universitat Autònoma de Barcelona, 08193 Bellatera (Barcelona) Spain
(Dated: August 14, 2019)

We investigate the quantum state discrimination task for sets of linear independent pure states with an intrinsic ordering. This structured discrimination problems allow for a novel scheme that provides a certified level of error, that is, answers that never deviate from the true value more than a specified distance and hence a control of the desired quality of the results. We obtain an efficient semidefinite program and also find a general lower bound valid for any error distance that only requires the knowledge of optimal minimum error scheme. We apply our results to the quantum change point and quantum state anomaly detection cases.

I. INTRODUCTION

State discrimination plays a fundamental role in quantum information sciences as it determines the capacity of quantum systems to carry information. The task consists in identifying in which of some known set of states a system was prepared by some source. If the possible states are mutually orthogonal this task can be done perfectly. However, if the states are not mutually orthogonal the problem is very nontrivial and it requires optimization with respect to some reasonable criteria.

The most studied discrimination schemes are minimum error (ME) and unambiguous discrimination (UD). In ME after a measurement is performed on the system the experimenter must give an answer about its state. Naturally, some of the answers will be erroneous, and the optimal ME strategy is the one that yields the minimum probability of committing an error [1]. In contrast, in UD, no errors are allowed, i.e, the answers of the experimenter must be absolutely certain. This can only be achieved at the expense of permitting inconclusive measurement outcomes. The optimal strategy is the one that minimizes the probability of inconclusive answers. It is known that UD is only possible for sets of linearly independent states [2]. For mixed states UD is also possible as long as they do not have identical supports [3].

Some extensions of these fundamental schemes have also been considered. Discrimination with maximum confidence [4] can be applied to states that are not necessarily independent and can be regarded as a generalized UD strategy. Strategies that interpolate between ME and UD have also been studied [5]. In those a given maximum value for the error probability (or equivalently a maximum value for inconclusive probability) is enforced. Varying this value yields a continuous set of strategies between UD (or maximum confidence) and ME.

Despite being such a fundamental task, analytical solutions for optimal discrimination schemes in the multi-hypothesis case remains a challenge (see [6] for recent developments). Essentially only the two state [1] and symmetric states cases [7–9] have been solved (see [10–12] for reviews on state discrimination).

In this work we consider a novel multi-hypothesis scheme for sources that prepare states with intrinsic structure. In particular, we consider linear independent states that can be represented as a linear chain (see Fig 1) of n local states. This type of sources includes the interesting cases of change point [13–15] and state anomaly detection [16] problems. In these structured sources the hypotheses are labelled by some position in the chain, Hence the errors have a natural distance, i.e., we can have have a one-site error, two-site error, etc-., if the outcome of the protocol is an answer that is at distance of one, two, etc., units from the site labelling the true hypothesis. This scheme is interesting not only from the theoretical point of view, but also for practical purposes. In many circumstances not any error can be tolerated, however small deviations from the true hypothesis may have only a limited impact on our decisions. So, it may prove useful to find optimal schemes under the constraint that no outcome can differ from the true hypothesis more than a given threshold distance Δ. Doing so, we have certified answers that will not spoil decisions that we may take upon the outcome of the protocol. We therefore call this scheme certified answer discrimination (CAD). Also if we relax the UA condition and allow some errors, the success probability of guessing the correct hypothesis can increase substantially as we will show. For $\Delta = 0$ we recover the UD scheme while for $\Delta = n - 1$ we get the ME scheme, thus CAD also provides an interpolation between UD and ME. The interpolating scheme discussed in [5] also yield a significant increase in the success probability, but, contrasting the CAD scheme, it may give erroneous answers that are very far from the true value. As it will become clear, CAD is a more natural scheme, closer to the notion of Hamming distances between states (i.e, the sum of positional mismatches [17]).

In this paper we give a convenient and efficient semidefinite program (SDP) [18–20] formulation of CAD schemes for linearly independent states. The SDP also enables us to find an analytical lower bound for the probability of success for any allowed error distance Δ. Interestingly, this lower bound only requires to calculate the ME suc-

*Electronic address: Esteban.Martinez@uab.cat
†Electronic address: Ramon.Munoz@uab.cat

cess probability, It provides an approximation on how much the success probability is reduced as we increase the requirements on the quality of the answers of the discrimination protocol.

The paper is organized as follows. In the next Section we present the CAD scheme and its SDP formulation. In Section III we obtain a lower bound for the success probability for any value of Δ. In Section IV we apply our results to the paradigmatic case of the change point and also discuss the state anomaly detection problem. Section IV contains the conclusions of our findings. We also include an appendix with some technical details.

II. CERTIFIED ANSWER DISCRIMINATION Δ-SCHEMES

Consider a quantum state multi-hypothesis discrimination problem where the source quantum states have an intrinsic ordering such as a one dimensional chain as depicted in Fig. 1. In this case it is possible to define a natural distance between the states.

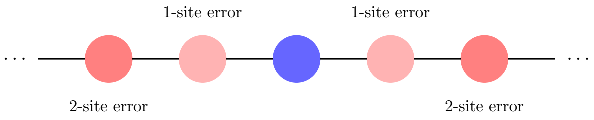


Figure 1: Structure of the source states. The position of the blue dot labels the state.

If we are given a state $|\Psi_k\rangle$, where k is the position that labels the state, our aim is to find a measurement, generally given by a Positive Operator Value (POVM), that returns the value k with the highest probability. The POVM has to satisfy the additional constraints that no errors beyond some distance Δ can be committed. For states given by a chain of n states, the POVM contains $n+1$ elements $\{E_k \geq 0\}_{k=0}^n$, where $E_0 = \mathbb{1} - \sum_{k=1}^n E_k$ is the element corresponding to an inconclusive answer. As in UD this element has to be introduced in order to satisfy the constraints. Naturally as Δ increases, i.e, more and more type of errors are allowed, we have $\langle\Psi_k| E_0 |\Psi_k\rangle \to 0$ for $k = 1, 2, \dots, n$.

The optimization problem can be written as the following SDP:

$$\begin{aligned} \underset{E}{\text{maximize}} \quad & \frac{1}{n}\sum_{i=1}^{n} \langle\Psi_i| E_i |\Psi_i\rangle \\ \text{subject to} \quad & \langle\Psi_j| E_i |\Psi_j\rangle = 0 \quad \forall |i-j| > \Delta \\ & \sum_{i=1}^{n} E_i \leq \mathbb{1} \\ & E_i \geq 0 \; \forall i, \end{aligned} \tag{1}$$

where for simplicity we assume that the prior probability is the same for all source states. We will also assume that the source states are linear independent, as naturally happens in the examples considered here (see section IV). Observe that each value $\Delta = 0, 1, 2, \dots, n-1$ defines a discrimination scheme that we will call a Δ-scheme. Note also that E_0 is a slack variable that it is taken into account by the inequality $\sum_{i=1}^n E_i \leq \mathbb{1}$ in the POVM condition.

For a given value of Δ we have a probability of success P_s^Δ, a probability of error P_e^Δ and a probability of inconclusive outcome P_I^Δ, and they satisfy the unitarity condition $P_s^\Delta + P_e^\Delta + P_I^\Delta = 1$. The value $\Delta = 0$ corresponds to the unambiguous case for which the error probability vanishes, $P_e^{\Delta=0} = 0$, and the outcome can either perfectly identify the state or be inconclusive, but not erroneous. For $\Delta = n-1$ the are no constraints on the errors and we recover the minimum error scheme, i.e the inconclusive probability vanishes, $P_I^{\Delta=n-1} = 0$. As we will see later, the minimum error limit can be effectively achieved for much smaller values of Δ.

If the source states are linearly independent, we can transform the SDP (1) into an equivalent and more useful program. From the n linearly independent estates $\{|\Psi_i\rangle\}_{i=1}^n$ we construct the R matrix,

$$R = \sum_{i=1}^{n} |\Psi_i\rangle\langle i|, \tag{2}$$

where $|i\rangle$ is any orthonormal basis (note that linear independence implies that R is invertible) and consider the new operators $F_r^\Delta = R^\dagger E_r^\Delta R$. Observe that the diagonal elements of F_r^Δ are the expectation values $\langle\Psi_i| E_r^\Delta | |\Psi_i\rangle = [F_r^\Delta]_{i,i}$. Thus, the first constraint in Eq. (1) translates into the condition that all diagonal elements $[F_r^\Delta]_{i,i}$ vanish except those with $|i-r| \leq \Delta$. Note also that $E_r^\Delta \geq 0 \to F_r^\Delta \geq 0$ [21]. The off-diagonal terms $[F_r^\Delta]_{i,j}$ are then also constrained by positivity, and hence we have $[F_r^\Delta]_{i,j} = 0$ for $|r-i| > \Delta$ and $|r-j| > \Delta$. The structure of the matrix F_r^Δ is illustrated in Fig. (2).

The second constraint in Eq. (1) can be recast as

$$G - \sum_{r=1}^{n} F_r^\Delta \geq 0, \tag{3}$$

by applying the matrix $R^\dagger$ on the left and the matrix R on the right. Here $G = R^\dagger R$ is the Gram matrix [22] whose elements are

$$G_{i,j} = \langle\Psi_i|\Psi_j\rangle. \tag{4}$$

Thus the SDP (1) is transformed onto

$$\begin{aligned} \underset{Z}{\text{maximize}} \quad & \frac{1}{n}\text{Tr}[ZA] \\ \text{subject to} \quad & \Phi_\Delta[Z] \leq G \\ & Z \geq 0. \end{aligned} \tag{5}$$

The matrix variable Z has a block diagonal structure containing the non-vanishing elements of F_r^Δ. In Fig. 3

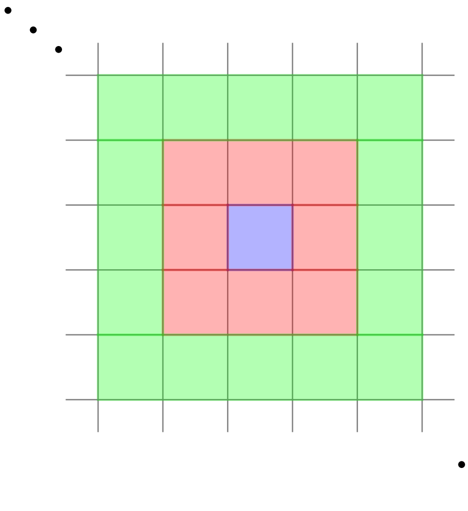


Figure 2: Example of the structure of $F_r^\Delta = R^\dagger E_r^\Delta R$. The matrices F_r^Δ have dimensions $n \times n$ and their non-vanishing elements are depicted as colored boxes. For $\Delta = 0$ (no errors), the central blue box is the only non-vanishing element. For $\Delta = 1$ (one error) the non-vanishing elements are contained in the red 3×3 block, and in the 5×5 green block for $\Delta = 2$, etc. The remaining entries of F_r^Δ are all zero.

we explicitly depict it for $\Delta = 1$. The elements highlighted are the ones that appear in the objective function $\mathrm{Tr}[ZA]$. The constant matrix A depends on the number n of hypothesis and maximum distance Δ of the allowed errors (we do not add these labels to avoid cluttering too much the notation). Matrix A "selects" the elements of the matrix variable Z that have to be optimized, i.e., the central elements of the Z blocks. For $\Delta = 1$ one has $A = \mathrm{diag}\{1, 0, 0, 1, 0, 0, 1, \ldots, 1\}$ and Z and A are $(3n-2) \times (3n-2)$ matrices. The generalization for any Δ is straightforward. Note that the appearance of the Gram matrix G in the second constraint of (5) showcases that all the discrimination properties of sets of linearly independent states are encapsulated in the Gram matrix.

The linear map Φ_Δ that incorporates the constraints (3) can be regarded as the action of two linear maps: $\Phi_\Delta = \Phi^2 \circ \Phi^1_\Delta$. The first map, Φ^1_Δ, embeds each block into a $n \times n$ sub-matrix and pads the remaining elements with zeros. The embedding is such that the k'th sparse sub-matrix has the central (highlighted) element in the kth position of the diagonal, as can be seen in Fig 4. With all the sub-matrices we have an $n^2 \times n^2$ block diagonal matrix. The second map, Φ^2, adds the sub-matrices to get a final $n \times n$ matrix, also illustrated in Fig. 4. Notice that this map is independent of Δ.

We note that the variable Z from SDP (5) has dimensions $[n(2\Delta+1) - \Delta(\Delta+3)] \times [n(2\Delta+1) - \Delta(\Delta+3)]$ which is significantly lower than $n^2 \times n^2$ of the original SDP (1). The size of the variables is similar only for $\Delta \to n$. However, as we will see in the quantum change point, the ME limit can be effectively reached for small values of Δ, and then the number of variables remains low for all meaningful values of Δ.

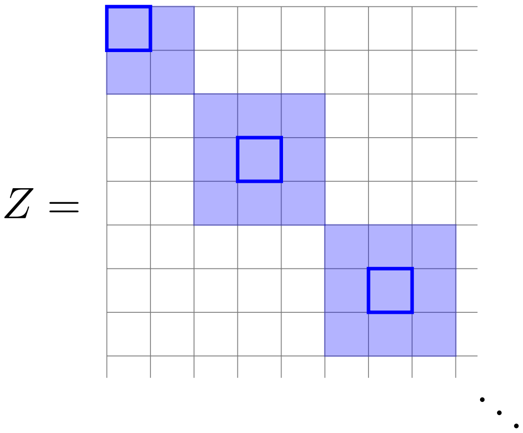


Figure 3: Structure of the matrix variable Z for $\Delta = 1$. The blue boxes correspond to the free matrix elements and the blank ones are fixed to be zero. The highlighted boxes are the elements that appear in the objective function $(1/n)\,\mathrm{Tr}[ZA]$ of Eq. (5).

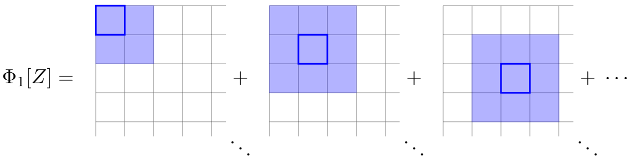


Figure 4: The correspondent map takes the non-zero parts of the variable Z and accommodates it in $n \times n$ matrices with zeros in the remaining places. Observe that matrix sum is defined only for matrices of the same dimensions.

There is no general mathematical method for solving a given SDP analytically, only problems with high degree of symmetry are known to be solvable. In some cases the primal or the dual version of the SDP can suggest an ansatz that may provide the solution (see [14] for a nice example). Therefore, any understanding of the form of the solutions of SDPs at hand is of interest. The transformation of the SDP made above proves to be beneficial not only for the numerical advantage but also to obtain insight into how the probability of success behaves in the intermediate regime between unambiguous and minimum error schemes. In particular, it enables us that find useful analytical lower bounds of the probability of success for any Δ that we discuss in the next section.

III. A LOWER BOUND FOR P_s^Δ

The main idea is to obtain a feasible solution of the SDP (5). Any ansatz matrix $\tilde{Z}$ that satisfies the constraints of an SDP is by construction a lower bound to the optimal solution. The method depends heavily on

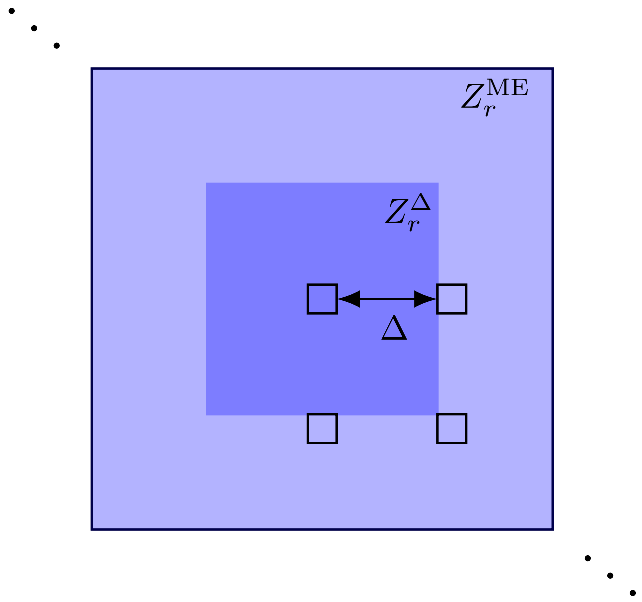


Figure 5: We depict a specific block r of $Z^{\mathrm{ME}}\ - Z^{\Delta}$. The light blue block corresponds to Z_r^{ME} and the darker one to Z_r^{Δ}. The small black boxes show the elements of the minor of interest to obtain the bound (9).

having previously solved the ME scheme, i.e. we have at our disposal the success probability P_s^{ME}, and the corresponding Z^{ME} . Fortunately, in many cases the minimum error scheme can be computed or well approximated with a square root measurement [23, 24].

As discussed in previous section the mapping $\Phi_\Delta[Z]$ in the SDP (5) can be understood as two step mapping that first transforms the variable Z into a $n^2 \times n^2$ variable that has zeros in appropriate places and a second step that sums all the individual blocks into a $n \times n$ matrix. If we only apply the first map, we get the following SDP:

$$\begin{aligned} &\underset{Z}{\text{maximize}} && \frac{1}{n}\,\mathrm{Tr}[ZA] \\ &\text{subject to} && \Phi^1_\Delta[Z] \leq Z^{\mathrm{ME}} \\ &&& Z \geq 0. \end{aligned} \tag{6}$$

Observe that any variable Z that satisfies $\Phi^1_\Delta[Z] \leq Z^{\mathrm{ME}}$ also satisfies $\Phi_\Delta[Z] \leq G$ (just apply the map Φ^2, on both sides of the first inequality). Hence, any feasible solution of the SDP (6) is in the feasible set of the SDP (5), but not vice versa, and it provides a lower bound for the probability of success.

For simplicity, let us call $\Phi^1_\Delta[Z] = Z^{\Delta}$ and $[Z_r^{\Delta}]_{i,j}$ the i,j element of its r-th sub-matrix Z_r^{Δ}. The positivity condition $Z^{\mathrm{ME}}\ - Z^{\Delta} \geq 0$ in Eq. (6) implies that any principal minor of $Z^{\mathrm{ME}}\ - Z^{\Delta}$ has to be positive [22]. To get a bound in terms of the known Z^{ME} , the elements of the principal minor have to be outside the central blocks of Z^{Δ}, as depicted in Fig. 5. The choice of this minor is such that it contains only one non-vanishing diagonal element of Z^{Δ} and three remaining elements are at a Δ distance and hence take the (known) Z^{ME} values. We take the minimum distance Δ as larger distances will give less stringent bounds.

The positivity condition then gives

$$\begin{aligned} \left([Z^{\mathrm{ME}}{}_{i}]_{i,i} - [Z_i^{\Delta}]_{i,i}\right) [Z^{\mathrm{ME}}{}_{i}]_{i+\Delta+1,i+\Delta+1} \geq \\ \left|[Z^{\mathrm{ME}}{}_{i}]_{i,i+\Delta+1}\right|^2 . \end{aligned} \tag{7}$$

Using the fact that the arithmetic mean is bigger than the geometric mean we finally have that

$$\begin{aligned} \left([Z^{\mathrm{ME}}{}_{i}]_{i,i} - [Z_i^{\Delta}]_{i,i}\right) + [Z^{\mathrm{ME}}{}_{i}]_{i+\Delta+1,i+\Delta+1} \\ \geq 2\left|[Z^{\mathrm{ME}}{}_{i}]_{i,i+\Delta+1}\right| . \end{aligned} \tag{8}$$

As we will be dealing with problems having some symmetry it is convenient to choose this lower minor for the first $\lceil n/2 \rceil$ and the corresponding upper minor for the rest of blocks. For these upper minors we get the same inequality (8) with the change $\Delta \to -\Delta$.

In order to calculate the bound of the success probability only the diagonal elements of $\tilde{Z}$ have to be specified. The best choice is to take them to saturate the inequalities (8), i.e.,

$$[\tilde{Z}_i]_{i,i} = \begin{cases} [Z^{\mathrm{ME}}{}_{i}]_{i,i} - H_i(\Delta) & \text{for } 1 \leq i \leq \lceil n/2 \rceil \\ [Z^{\mathrm{ME}}{}_{i}]_{i,i} - H_i(-\Delta) & \text{for } i > \lceil n/2 \rceil \end{cases} , \tag{9}$$

where

$$H_i(\Delta) = 2\left|[Z^{\mathrm{ME}}{}_{i}]_{i,i+\Delta+1}\right| - [Z^{\mathrm{ME}}{}_{i}]_{i+\Delta+1,i+\Delta+1} . \tag{10}$$

Adding all the terms in Eq. (9), the lower bound $\tilde{P}_s$ for the success probability reads

$$\begin{aligned} P_s^{\Delta} \geq & \tilde{P}_s = P_s^{\mathrm{ME}} \\ & - \frac{1}{n}\left[\sum_{i=1}^{\lceil n/2 \rceil} H_i(\Delta) + \sum_{i=\lceil n/2 \rceil + 1}^{n} H_i(-\Delta)\right] , \end{aligned} \tag{11}$$

which depends only on Z^{ME} . The bound (11) has two parts, the first is just the success probability of the minimum error case (i.e., the unrestricted case), while the second takes into account how much this value is diminished by the additional constraints imposed by the value Δ. The main virtue of this bound is that given the solution for the minimum error case it provides an expression on how much this probability is lessened by increasing the quality of the answers, i.e., by reducing the maximum allowed distance of the answers to the true state.

IV. APPLICATIONS

In this section we apply our findings to two paradigmatic multi-hypothesis cases. We first discuss the Quantum Change Point (QCP) problem [13–15] and then

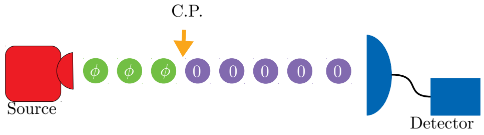


Figure 6: A machine produces a signal state and suddenly it produces another signal. Our task is to determine by measurements the exact moment when this change happens.

briefly discuss the Quantum State Anomaly Detection (QSAD) problem [16].

The QCP problem is depicted in figure (6). A source prepares systems in a default state $|0\rangle$ for some time and suddenly it changes and prepares systems in a mutated state $|\phi\rangle$. Both states are assumed to be known and the change is also assumed to occur at any time with the same probability. The total number of systems is n. The goal is to identify the position of the mutation with the highest probability. This is a multi-hypothesis case for which the optimal ME and UD probabilities of success are known [13, 14].

The global states can be written as

$$|\Psi_k\rangle = |0\rangle^{\otimes k-1}|\phi\rangle^{\otimes n-k+1}. \tag{12}$$

The Gram matrix has elements $G_{i,j} = \langle\Psi_i|\Psi_j\rangle = c^{|i-j|}$, where $c = \langle 0|\phi\rangle$ and w.l.o.g. can be taken to be in the interval $0 \leq c \leq 1$. Note that for $c \neq 0, 1$ the off diagonal elements of the G decay exponentially as they depart from the diagonal , which shows that in the QCP the Hamming distance between states is directly related to the overlap between states.

The CAD scheme is particularly pertinent for this problem. It is reasonable to assume that here some deviations of the output guess from the true change point can be tolerated, but not too many in order to avoid jeopardizing the validity of the identification task. In Fig. 7 we show the success probability as a function of Δ as given by the SDP (5) for $c = 0.6$ and $n = 25$. We note a remarkable increase in the success probability by just allowing one error deviation of the guess. The value of P_s jumps almost a factor of two, from 0.27 for $\Delta = 0$, to 0.50 for $\Delta = 1$. Also the inconclusive probability drops from 0.73 to 0.4, while only 10% of the answers will be erroneous (and just by one position). If these are counted as satisfactory answers, the total success probability goes up to 60%. We have checked that these values of the probabilities essentially remain constant for any $n > 25$. We also observe that the probability of success stabilizes to the ME value for $\Delta \gtrsim 8$ (again this threshold value remains the same for larger values of n). This just shows that the ME protocol effectively does not yield answers that are at distance greater than eight space units from the true state, as can explicitly be seen in Fig. 8.

We next calculate the bound (11). As discussed in previous section, the bound requires to have the solution

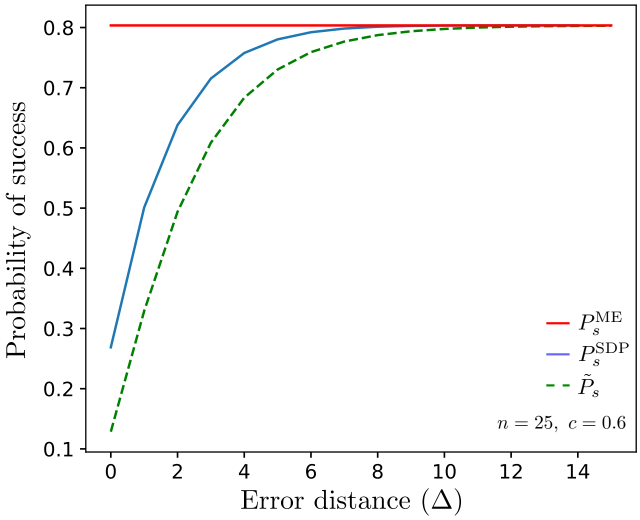


Figure 7: Probability of success versus the allowed error distance Δ for QCP with $n = 25$ and $c = 0.6$. The blue solid curve are the exact numerical SDP values (5). The dotted green curve is the analytical lower bound (21). We also show as a reference the red straight line with the value of the minimum error scheme.

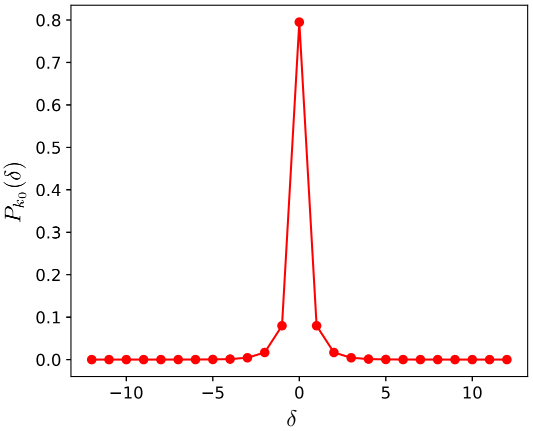


Figure 8: Outcome probability profile of the minimum error scheme of the QCP. The parameter $\delta = \hat{k} - k_0$ is the distance of the output guess $\hat{k}$ with respect to the position k_0 of the true change point. Here $n = 25$, $c = 0.6$ and we take the change point to occur at the central position $k_0 = 13$.

Z^{ME} , but, as shown in [13], for the QCP can be very well approximated by he square root measurement, i.e. by a projective POVM $\{E_k = |m_k\rangle\langle m_k|\}_{k=1}^n$, with $S\,|m_k\rangle = |\Psi_k\rangle$ and $S = \sqrt{G} = \sum_k \sqrt{\lambda_k}\,|v_k\rangle\langle v_k|$, where λ_k and $|v_k\rangle$ are the eigenvalues and eigenvectors of G, respectively.

The matrix Z^{ME} in terms of the square root S simply reads

$$Z^{\mathrm{ME}} = \bigoplus_{k=1}^{n} |s_k\rangle\langle s_k| \quad \text{with} \quad \langle m_l|s_k\rangle = S_{l,k}, \tag{13}$$

i.e. $|s_k\rangle$ are the column vectors of S.

The crucial point to obtain a useful bound is to prove that the elements of S away from the diagonal decay exponentially. From the supplemental material of [13] we have

$$S_{k,l} \approx \frac{\sqrt{1-c^2}}{\pi}\int_0^{\pi} d\theta \frac{(\sin k\theta - c\sin(k-1)\theta)(\sin l\theta - c\sin(l-1)\theta)}{(1-2c\cos\theta + c^2)^{3/2}}. \tag{14}$$

After some straightforward algebra Eq. (14) reads

$$\begin{aligned} S_{k,l} \approx \frac{\sqrt{1-c^2}}{4\pi}\int_{-\pi}^{\pi} d\theta \left[\frac{\cos(k-l)\theta - \cos(k+l)\theta}{(1-2c\cos\theta+c^2)^{1/2}}\right. \\ \left. + \chi(k+l,c)\right], \end{aligned} \tag{15}$$

where $\chi(k+l,c)$ contains terms that oscillate rapidly and will be considered later (observe that the second term also oscillates more rapidly than the first). We also note that the explicit terms terms shown in Eq. (15) correspond to the Fourier series of of the function

$$\mu(\theta,c) = \frac{1}{(1-2c\cos\theta+c^2)^{1/2}}, \tag{16}$$

so we consider

$$\widehat{\mu}(k,c) = \int_{-\pi}^{\pi} \mu(\theta,c)e^{ik\theta}d\theta \tag{17}$$

for $k \in \mathbb{N}$. We prove in Appendix A that $\widehat{\mu}(k,c)$ exhibits an exponential decay in k given by

$$|\widehat{\mu}(r,c)| \leq M_0(c)e^{k\log(c)}. \tag{18}$$

where $M_0 = \int_{-\pi}^{\pi}\mu(\theta,c)d\theta$. The other terms included in $\chi(k+l,c)$ of Eq. (15))are proportional to $\mu^3(r,c)$ and can be tackled in a similar fashion. Including the term proportional to $\cos(k+l)\theta$ and the terms coming from $\chi(k+l,c)$ we get

$$S_{k.l} \leq \frac{\sqrt{1-c^2}}{4\pi}\left(M_0 e^{|k-l|\log(c)} + \sum_{i=-1}^{2} M_i e^{(k+l+i)\log(c)}\right). \tag{19}$$

We can now calculate Z^{ME} inserting (19) into Eq. (13). We further just take into consideration the (first) dominant term to obtain

$$\begin{aligned} \left|[Z^{\mathrm{ME}}\ {}_i]_{i,i\pm\Delta+1}\right| &\leq c\, e^{\Delta\log c}\left|[Z^{\mathrm{ME}}\ {}_i]_{i,i}\right| \\ \left|[Z^{\mathrm{ME}}\ {}_i]_{i\pm\Delta+1,i\pm\Delta+1}\right| &\leq c^2\, e^{2\Delta\log c}\left|[Z^{\mathrm{ME}}\ {}_i]_{i,i}\right|. \end{aligned} \tag{20}$$

Finally from equation (11) we get

$$\tilde{P}_s \geq (1 - 2ce^{\Delta\log c} + c^2 e^{2\Delta\log c})P_s^{\mathrm{ME}}, \tag{21}$$

which shows that the success probability approaches at least exponentially P_s^{ME} for sufficiently large Δ. Note also that in the limit $c \to 0$ we recover the obvious result that $P^{\mathrm{ME}} = P^{\mathrm{UA}}$. We show the bound (21) along with the exact numerical results in figure (7). We observe that indeed the bound approaches the minimum error value for large Δ.

To end this section we study the Quantum State Anomaly Detection (QSAD) problem [16, 25] , which will provide some further insight of the features of the our certified answers protocol. QSAD can be regarded as a simplified case of the QCP. The source is assumed to prepare systems in a given default sate $|0\rangle$, however one (and just one) of the local systems was prepared in a different anomalous state $|\phi\rangle$. As in the QCP we assume both states to be known and equal probability for the position of the anomalous state. The task consists in identifying the position of the faulty state with the highest probability when a string of n systems has been prepared. Also here we may consider a protocol that yields guesses not deviating more than Δ units from the true position of the anomaly.

The set of hypothesis is is given by

$$|\Psi_k\rangle = |0\rangle^{\otimes k-1}|\phi\rangle|0\rangle^{\otimes n-k}. \tag{22}$$

and again we define $c = \langle\phi|0\rangle$ that w.l.o.g. can be taken to be in the interval $0 \leq c \leq 1$. Notice that we have a very simple Gram matrix in this case

$$G_{i,j} = \langle\psi_i|\psi_j\rangle = (1-c^2)\delta_{ij} + c^2. \tag{23}$$

This Gram matrix is circulant [26] , and hence the square root measurement is optimal [13, 25]. It is straightforward to find $S = \sqrt{G}$:

$$S_{i,j} = (a-b)\delta_{ij} + b \tag{24}$$

where

$$\begin{aligned} a =& \frac{\sqrt{1+(n-1)c^2} + (n-1)\sqrt{1-c^2}}{n} \\ b =& \frac{\sqrt{1+(n-1)c^2} - \sqrt{1-c^2}}{n} \end{aligned} \tag{25}$$

Note that the success probability for the minimum error scheme is simply [25]

$$P_s^{\mathrm{ME}} = \frac{1}{n}\sum_{i=1}^{n} S_{i,i}^2 = a^2. \tag{26}$$

The fact that all source states have the same overlap, or equivalently have equal Hamming distance, makes the distance to the true anomaly a less natural parameter in this case and we have different behaviors for $\Delta < \lfloor n/2 \rfloor$ and $\Delta \geq \lfloor n/2 \rfloor$. It is easy to convince oneself that the symmetry of the problem implies that the condition $\langle \psi_i | E_j | \psi_i \rangle = 0$ for $|i-j| \geq \Delta$ for any $\Delta < \lfloor n/2 \rfloor$ is in fact equivalent to impose $\Delta = 0$. Whence for $0 \leq \Delta < \lfloor n/2 \rfloor$ we have a constant probability of success, as can be seen in Fig. 9, and the protocol is equivalent to unambiguous discrimination. It is interesting to calculate the bound (11) in this regime. We have

$$[Z^{\mathrm{ME}}\ _i]_{ii} = a^2, \quad [Z^{\mathrm{ME}}\ _i]_{i,i\pm\Delta+1} = ab$$
$$[Z^{\mathrm{ME}}\ _i]_{i\pm\Delta+1,i\pm\Delta+1} = b^2. \tag{27}$$

From equation (11) we get

$$\tilde{P}_s = (a-b)^2 = 1 - c^2. \tag{28}$$

This value is exactly the unambiguous success probability. Notice that for $\Delta = 0$, the matrix A in Eq. (5) is $\mathbb{1}_n$ and that by symmetry $Z = z\mathbb{1}_n$, with z a real parameter. Then the SDP reads

$$\begin{aligned} &\text{maximize} && z \\ &\text{subject to} && z\mathbb{1}_n \leq G \\ & && z \geq 0, \end{aligned} \tag{29}$$

which is the SDP for the minimum eigenvalue of G. From (23) it is direct to obtain $z = 1 - c^2$, as expected.

For $\Delta \geq \lfloor n/2 \rfloor$ we can start having some errors, and the success probability starts to increase from UA to ME as seen in Fig. 9. We also see that the lower bound (11) in this regime departs from the P_s^{UA} value. Now at least one block of Z^{ME} can be completely covered by Z^{Δ} which allows for larger contributions to the bound. So $[\tilde{Z}_j]_{ii}$ has some elements constrained to be $(a-b)^2$ and as Δ increases new ones equal to the larger value a^2. Defining $d := \Delta - \lfloor n/2 \rfloor$ and recalling that $P_s^{\mathrm{ME}} = a^2$ and $P_s^{\mathrm{UA}} = (a-b)^2$, we obtain from Eq. (11)

$$\tilde{P}_s = \begin{cases} \frac{n-(2d+1)}{n} P_s^{\mathrm{UA}} + \frac{2d+1}{n} P_s^{\mathrm{ME}} & \text{for } n \text{ odd} \\ \frac{n-2d}{n} P_s^{\mathrm{UA}} + \frac{2d}{n} P_s^{\mathrm{ME}} & \text{for } n \text{ even,} \end{cases} \tag{30}$$

which exhibits a nice linear behavior interpolating between UA and ME.

V. CONCLUSIONS

We have introduced a novel scheme of quantum discrimination for ordered hypothesis of linearly independent states that gives certified answers that do not depart from the true hypothesis more than a given distance Δ. Our scheme may be of practical importance in cases where small deviations from the true hypothesis can be tolerated without compromising the effectiveness of the discrimination task. The scheme allows to tune at will the quality versus the quantity of the answers.

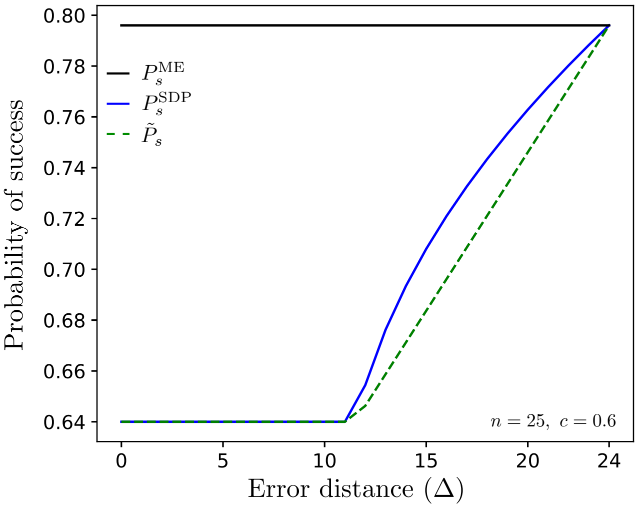


Figure 9: Probability of success against the error distance Δ for QSAD with $n = 25$ and $c = 0.6$. The blue solid line shows the numerical results from SDP (5) and the dashed green line the lower bound Eqs. (28) and (30). We also show as a reference a black straight line with the value of the minimum error scheme.

We have shown that all the discrimination properties of a given set of hypotheses are contained in the Gram matrix of the set. We have obtained a compact SDP for the optimal solution that can be solved very efficiently. We have also obtained a lower bound of the success probability for any value of the deviation that only requires the knowledge of the minimum error solution. The bound gives an analytical expression of how much the minimum error success probability is reduced as the maximum distance error Δ is decreased.

We have applied our findings to the quantum change point problem and the quantum state anomaly detection. For the former, we have shown that allowing a small departure from the true change point increases quite dramatically the success probability. We have computed the lower bound and shown that the increase of the success of probability is exponential in the allowed distance of the errors. For the QSAD we see that up to $n/2$ the protocol is equivalent to unambiguous discrimination. The lower bound for $\Delta \geq n/2$ gives a linear interpolation between UA and ME error protocols.

Our scheme is versatile enough to address other interesting situations. For instance, one might consider non-symmetric errors, i.e the tolerated distance of forward and backward errors may be different. Also one can consider incompatibilities, i.e, given some hypothesis the protocol is required to avoid some specific answers. One important extension of our protocol would be to consider sets of linearly dependent and noisy states. The main difficulty here is how to extend the Gram matrix formalism in these settings. We are currently exploring these scenarios.

Acknowledgments

We thank useful discussions with Gael Sentís. We acknowledge the financial support of the Spanish MINECO, ref. FIS2016-80681-P (AEI/FEDER, UE), and Generalitat de Catalunya CIRIT, ref. 2017-SGR-1127. EMV thanks financial support from CONACYT.

Appendix A

In this Appendix we prove that the Fourier coefficients $\widehat{\mu}(k,c)$ of Eq. (17) decay exponentially with k as

$$|\widehat{\mu}(k,c)| \leq M(c)e^{k\log(c)}, \tag{A1}$$

where

$$M(c) = \int_{-\pi}^{\pi} \mu(\theta,c)d\theta. \tag{A2}$$

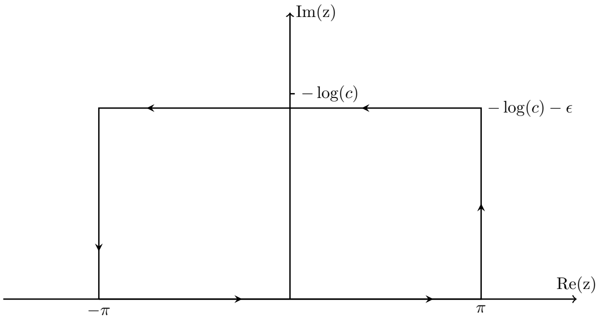


Figure 10: Path of the contour integral of $\mu(z,c)$ in the complex plane.

Proof. We first extend the function $\mu(\theta,c)$ to the complex plane as

$$\mu(z,c) = \frac{1}{\sqrt{(c-e^{iz})(c-e^{-iz})}}. \tag{A3}$$

If we take the principal branch of the logarithm as a domain of $z \to \sqrt{z}$ the function $\mu(z,c)$, is analytic in $\mathbb{C}/\{-i\log(c), i\log(c)\}$ because it is the composition of several analytic functions. It is a known fact that the Fourier coefficients of analytic functions decay exponentially [27]. We next compute

$$\widehat{\mu}(k,c) = \int_{-\pi}^{\pi} \mu(\theta,c)e^{ik\theta}d\theta, \tag{A4}$$

for $k \in \mathbb{N}$. Notice that due to the symmetry $\mu(\theta,c) = \mu(-\theta,c)$, only the cosine term of $e^{ik\theta}$ survives. We consider the contour integral in the complex plane shown in Fig. 10. We will call γ^c the part of the contour that does not lie in the real line. By analyticity of $\mu(z,c)$ in this region we have that,

$$\int_{-\pi}^{\pi} \mu(\theta,c)e^{ik\theta}d\theta + \int_{\gamma^c} \mu(z,c)e^{ikz}dz = 0. \tag{A5}$$

Notice that $\mu(\theta,c) \geq 0\ \forall\ \theta \in [-\pi,\pi]$ and $0 \leq c \leq 1$, i.e, $\mu(\theta,c) = |\mu(\theta,c)|$. We also see that the contributions of the right and left vertical sections of the path cancel out. Thus, we have

$$\int_{-\pi}^{\pi} |\mu(\theta,c)|d\theta = \int_{-\pi}^{\pi} |\mu(x+i(-\log(c)-\epsilon),c)|dx, \tag{A6}$$

and from Eq. (A5) we get

$$\begin{aligned}\left|\int_{-\pi}^{\pi} \mu(\theta,c)e^{ik\theta}d\theta\right| &= \left|\int_{\gamma^c} \mu(z,c)e^{ikz}dz\right| \\ &\leq \int_{\gamma^c} \left|\mu(z,c)e^{ikz}\right| dz \\ &= \int_{\gamma^c} |\mu(z,c)|\, e^{-ky}dz \\ &= e^{k(\log(c)+\epsilon)} \\ &\times \int_{-\pi}^{\pi} |\mu(x+i(-\log(c)-\epsilon),c)|dx \\ &= M(\epsilon,c)e^{k(\log(c)+\epsilon)},\end{aligned}$$

where in going from the second to the third r.h.s expression we use the fact that the the right and left arms contributions of the contour γ^c cancel out. Note that the constant $M(\epsilon,c)$ does not depend on k. Taking the limit $\epsilon \to 0$ and recalling Eq. (A6), we get

$$|\widehat{\mu}(k,c)| \leq e^{k\log(c)} \int_{-\pi}^{\pi} \mu(\theta,c)d\theta, \tag{A7}$$

i.e., $M(c) = \int_{-\pi}^{\pi} \mu(\theta,c)d\theta$. □

We can calculate in a completely analogous fashion the Fourier coefficients for other powers of $\mu(\theta,c)$. For instance, the function $\chi(k+l,c)$ in Eq. (15) includes terms proportional to $\mu^3(\theta,c)$ and these will also decay exponentially.

All the elements of $S = \sqrt{G}$ of the QCP can thus be expressed as

$$\begin{aligned}S_{k,l} \leq \frac{\sqrt{1-c^2}}{4\pi} &\Big(M_0 e^{|k-l|\log(c)} \\ &+ \sum_{i=-1}^{2} M_i e^{(k+l+i)\log(c)}\Big),\end{aligned} \tag{A8}$$

where M_i are constants that only depend on c and not on k or l.

[1] C. W. Helstrom, *Quantum detection and estimation theory* (Academic press, 1976).

[2] A. Chefles, Unambiguous discrimination between linearly independent quantum states, Physics Letters A **239**, 339 (1998).

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

[4] S. Croke, E. Andersson, S. M. Barnett, C. R. Gilson, and J. Jeffers, Maximum confidence quantum measurements, Phys. Rev. Lett. **96**, 070401 (2006).

[5] E. Bagan, R. Muñoz-Tapia, G. A. Olivares-Rentería, and J. A. Bergou, Optimal discrimination of quantum states with a fixed rate of inconclusive outcomes, Phys. Rev. A **86**, 040303 (2012).

[6] T. Singal, E. Kim, and S. Ghosh, Structure of minimum error discrimination for linearly independent states, Phys. Rev. A **99**, 052334 (2019).

[7] M. Ban, K. Kurokawa, R. Momose, and O. Hirota, Optimum measurements for discrimination among symmetric quantum states and parameter estimation, International Journal of Theoretical Physics **36**, 1269 (1997).

[8] S. M. Barnett, Minimum-error discrimination between multiply symmetric states, Phys. Rev. A **64**, 030303 (2001).

[9] H. Krovi, S. Guha, Z. Dutton, and M. P. da Silva, Optimal measurements for symmetric quantum states with applications to optical communication, Phys. Rev. A **92**, 062333 (2015).

[10] S. M. Barnett and S. Croke, Quantum state discrimination, Adv. Opt. Photon. **1**, 238 (2009).

[11] A. Chefles, Quantum state discrimination, Contemporary Physics **41**, 401 (2000), https://doi.org/10.1080/00107510010002599 .

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

[13] G. Sentís, E. Bagan, J. Calsamiglia, G. Chiribella, and R. Muñoz-Tapia, Quantum change point, Phys. Rev. Lett. **117**, 150502 (2016).

[14] G. Sentís, J. Calsamiglia, and R. Muñoz-Tapia, Exact identification of a quantum change point, Phys. Rev. Lett. **119**, 140506 (2017).

[15] G. Sentís, E. Martínez-Vargas, and R. Muñoz-Tapia, Online strategies for exactly identifying a quantum change point, Phys. Rev. A **98**, 052305 (2018).

[16] M. Skotiniotis, R. Hotz, J. Calsamiglia, and R. Muñoz-Tapia, Identification of malfunctioning quantum devices, arXiv:1808.02729 (2018).

[17] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition*, 10th ed. (Cambridge University Press, New York, NY, USA, 2011).

[18] L. Vandenberghe and S. Boyd, Semidefinite programming, SIAM Review **38**, 49 (1996).

[19] Y. C. Eldar, A semidefinite programming approach to optimal unambiguous discrimination of quantum states, IEEE Transactions on Information Theory **49**, 446 (2003).

[20] J. Watrous, *The Theory of Quantum Information* (Cambridge University Press, 2018).

[21] R. Bhatia, *Matrix Analysis*, Graduate Texts in Mathematics (Springer New York, 1996).

[22] R. A. Horn and C. R. Johnson, *Matrix Analysis*, 2nd ed. (Cambridge University Press, New York, NY, USA, 2012).

[23] P. Hausladen and W. K. Wootters, A 'pretty good' measurement for distinguishing quantum states, Journal of Modern Optics **41**, 2385 (1994).

[24] P. Hausladen, R. Jozsa, B. Schumacher, M. Westmoreland, and W. K. Wootters, Classical information capacity of a quantum channel, Phys. Rev. A **54**, 1869 (1996).

[25] N. Dalla Pozza and G. Pierobon, Optimality of square-root measurements in quantum state discrimination, Phys. Rev. A **91**, 042334 (2015).

[26] R. M. Gray, Toeplitz and circulant matrices: A review, Foundations and Trends in Communications and Information Theory **2**, 155 (2006).

[27] Y. Katznelson, *An Introduction To Harmonic Analysis*, Cambridge Mathematical Library (Cambridge University Press, 2004).

5

Quantum detection in time II

5.1 Quantum Sequential Hypothesis Testing

We stay still in the scheme where we have a set of states ordered in time but we want to address a different problem: hypothesis testing of quantum states. We take a different approach than the usual[Hay01] as we are now incursing into the area of Sequential Analysis, a wide topic of research in statistics that stems from the classic works of Abraham Wald [Wal73, WW48].

Related to our work is that of Slussarenko et. al. [SWL+17]. In that publication the authors also take a fixed error and minimize the average number of samples needed, however, we consider a more general case as they restrict the type of error they are allowing.

The methodology in sequential analysis is somewhat distinct from what one finds in literature known as hypothesis testing. Normally the number of samples of random variables of an unknown distribution is a fixed number N and we try to minimize the possible error in our guess. The sequential scheme was inspired on a test by Neyman and Pearson [NPP33] that compares two possible hypotheses, however, it presents advantages over this one.

Sequential Analysis proposes a testing procedure called Sequential Probability Ratio Test (SPRT) that can be made as samples are available "on the fly".

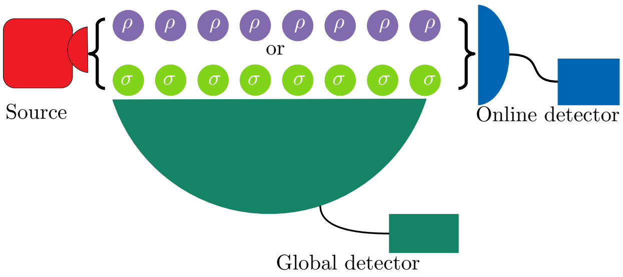


Figure 5.1: We consider an i.i.d. source in this problem. Here the source produces all the time one of two possible states.

It is notable that the methodology changes, because in the SPRT the number of samples is a random variable. What stays fixed is the maximum error one asks for. In hypothesis testing there are two types of error, as we mention in Section (3.1). We can fix both types of error in Sequential analysis and then ask for the number of copies it takes on average to yield an answer with the error requirements. The SPRT is based on a cumulative sum of a quantity that depends on the outcomes of a sequence of measurements. It is an instance of a martingale [MU05].

If we use a fixed POVM to measure each copy of a quantum state produced by an i.i.d. source then we will get i.i.d. samples from the same classical probability distribution and therefore this setting is within the classical SPRT. Suppose that our hypotheses H_0 and H_1 correspond to having one of two possible states ρ and σ respectively and we define a POVM $\{E_x\}$. Then, following the postulate 3 we have two probability distributions

$$p(x) = \mathrm{tr}\,[E_x\rho] \qquad\qquad q(x) = \mathrm{tr}\,[E_x\sigma]. \tag{5.1}$$

Wald theory tells us that the average number of samples needed to finish the process of testing in the limit of small errors is given by

$$\langle N\rangle_0 \sim -\frac{\log\epsilon_0}{D(p||q)} \qquad\qquad \langle N\rangle_1 \sim -\frac{\log\epsilon_1}{D(q||p)}, \tag{5.2}$$

where ϵ_0 and ϵ_1 are the Type-I error and Type-II errors which are fixed. Defining a probability distribution of n outcomes $\mathbf{x}_n = \{x_1, x_2, \ldots, x_n\}$ as

$$P(\mathbf{x}_n) = \prod_{k=1}^{n} p(x_k) \qquad\qquad Q(\mathbf{x}_n) = \prod_{k=1}^{n} q(x_k), \tag{5.3}$$

a log likelihood ratio is calculated after each online measurement as

$$Z_n = \log \frac{Q(x_n)}{P(x_n)} = \sum_{k=1}^{n} z_k. \tag{5.4}$$

In the first part, we analyzed the case where ρ and σ are qubits and a two outcome POVM that is parameterised with an angle θ, but it is the same for all samples. We show an advantage in terms of a reduction of the number of resources required with respect to the optimal deterministic test, which shows that the advantages of the classical SPRT translate well into the quantum setting.

However, quantum mechanics allows for more complicated strategies that involve changing the measurement apparatus with feedback from the outcome. These set of strategies can become very complex: weak measurements, collective, etc. This moves us to ask if there is an improvement over classical sequential methods. We would like to be able also to quantify this advantage. Also, given that quantum theory generalizes what can be done with classical resources, we would like to know what is the maximum advantage (maximum saving of copies) possible with quantum resources. The complete problem seems very complicated and we do not address it directly. Nevertheless, we found a lower bound for the average number of copies needed to do hypothesis testing with given error bounds in the asymptotic limit of small errors. This bound is a constraint on the minimum number of copies needed. We call it "ultimate" because any other protocol with a lower average number of samples would not fulfill the error constraints that the problem asks. Recently it has been proved that that in general this bounds are attainable [LTT]. We also give upper bounds for the average number of copies giving a specific strategy. We also give an upper bound for a worst-case scenario and give regions of saturability for qubits.

This is the first extension of sequential analysis in the quantum case, in the paradigmatic and simplest case of binary quantum hypothesis testing. This is a promising start of a whole research program to apply quantum sequential methodologies to all sorts of quantum statistical inference tasks like parameter estimation, channel discrimination etc. It promises to exploit the advantage of the sequential method that a decision can be made dynamically as copies are available.

The case for pure states specially reveals the novelty of our approach as it presents interesting unexpected behavior. First of all, we find the optimal global strategy because it coincides with the online one: doing unambiguous discrimination at each step. This is very useful because we are able to obtain closed expressions. A remarkable result is that we can ask for zero error protocols and find that there is a finite number of copies needed; in contraposition with the mixed case where the number of copies needed grows indefinitely as the error goes to zero.

5.2 Online identification of symmetric pure states

In the sequential analysis scheme we noticed that for pure states the behaviour is very peculiar if compared with the general mixed state case. Maintaining ourselves in the case of the i.i.d. machine, i.e. Fig. (5.1) we asked the question of when unambiguous discrimination is equivalent to the case one uses online or global measurements with many instances of the machine, which mean many copies of one state of the two hypotheses. We return to the usual scheme of a fixed number of copies given and minimizing the average error.

We first study the discrimination of two hypotheses in the unambiguous and minimum error schemes. In the literature it was observed that for several copies these schemes are equivalent in terms of the probability of success. However, for more hypotheses this was largely unexplored. We addressed the question of unambiguous discrimination of 3 symmetric pure states as it presents sufficient complications. Having in mind the equivalence between online strategies and global ones for multiple copies in the two-hypotheses case, it is only natural to ask if this behavior still persists with 3 hypotheses. We know in general that discrimination of multiple hypotheses presents difference between local and global measurements even if the states are separable [BDF+99].

Even in the symmetric case that we consider we find that doing a parametrization of the overlaps the cases where online strategies equal the global ones are very particular. We study the multi-copy problem and are able to include cases that are not available with only one copy because the states are not linearly independent. We also observe a case when multihypothesis (more than three) discrimination is equal with online and global protocols.

Quantum Sequential Hypothesis Testing

Esteban Martínez Vargas,[1,*] Christoph Hirche,[2,†] Gael Sentís,[1,‡] Michalis Skotiniotis,[1,§] Marta Carrizo,[1] Ramon Muñoz-Tapia,[1,¶] and John Calsamiglia[1,**]

[1] *Física Teòrica: Informació i Fenòmens Quàntics, Departament de Física, Universitat Autònoma de Barcelona, 08193 Bellatera (Barcelona) Spain*

[2] *QMATH, Department of Mathematical Sciences, University of Copenhagen, Universitetsparken 5, 2100 Copenhagen, Denmark*

(Dated: May 10, 2021)

We introduce sequential analysis in quantum information processing, by focusing on the fundamental task of quantum hypothesis testing. In particular our goal is to discriminate between two arbitrary quantum states with a prescribed error threshold, ϵ, when copies of the states can be required on demand. We obtain ultimate lower bounds on the average number of copies needed to accomplish the task. We give a block-sampling strategy that allows to achieve the lower bound for some classes of states. The bound is optimal in both the symmetric as well as the asymmetric setting in the sense that it requires the least mean number of copies out of all other procedures, including the ones that fix the number of copies ahead of time. For qubit states we derive explicit expressions for the minimum average number of copies and show that a sequential strategy based on fixed local measurements outperforms the best collective measurement on a predetermined number of copies. Whereas for general states the number of copies increases as $\log 1/\epsilon$, for pure states sequential strategies require a finite average number of samples even in the case of perfect discrimination, i.e., $\epsilon = 0$.

Introduction. Statistical inference permeates almost every human endeavor, from science and engineering all the way through to economics, finance, and medicine. The perennial dictum in such inference tasks has been to optimize performance—often quantified by suitable cost functions—given a fixed number, N, of relevant resources [1, 2]. This approach often entails the practical drawback that all N resources need to be batch-processed before a good inference can be made. Fixing the number of resources ahead of time does not reflect the situation that one encounters in many real-life applications that might require an online, early, inference–such as change-point detection [3–6], or where additional data may be obtained on demand if the required performance thresholds are not met.

Sequential analysis [7] is a statistical inference framework designed to address these shortcomings. Resources are processed on-the-fly, and with each new measured unit a decision to stop the experiment is made depending on whether prescribed tolerable error rates (or other cost functions) are met; the processing is continued otherwise. Since the decision to stop is solely based on previous measurement outcomes, the size N of the experiment is not predetermined but is, instead, a random variable. A sequential protocol is deemed optimal if it requires the least *average number of resources* among all statistical tests that guarantee the same performance thresholds. For many classical statistical inference tasks it is known that sequential methods can attain the required thresholds with substantially lower average number of samples than any statistical test based on a predetermined number of samples [7]. The ensuing savings in resources, and the ability to take actions in real-time, have found applications in a wide range of fields [3, 8]. Extending sequential analysis to the quantum setting is of fundamental interest, and with near-term quantum technologies on the verge of impacting the global market, the versatility and resource efficiency that sequential protocols provide for quantum information processing is highly desirable.

In this paper we consider the discrimination of two arbitrary finite dimensional quantum states [9], ρ (corresponding to the null hypothesis H_0) and σ (corresponding to the alternative hypothesis H_1), in a setting where a large number of copies can be used in order to meet a desired error threshold ϵ. A first step in this direction was taken in Ref. [10], which considers the particular case where ρ and σ are pure states and restricts the analysis to specific local measurement strategies. Here, we address the problem in full generality, including arbitrary states, weak and collective measurements. For collective strategies involving a large *fixed* number of copies the relation between this number and the error ϵ is $N \sim -\frac{1}{\xi}\log\epsilon$ [11], where the rate ξ depends on the pair of hypotheses and on the precise setting as we explain shortly. We show that one can significantly reduce the expected number of copies, $\langle N\rangle$, by considering sequential strategies where copies are provided on demand. We give the ultimate lower-bounds as a single-letter expression of the form,

$$\langle N\rangle_0 \geq -\frac{\log\epsilon}{D(\rho\|\sigma)} + O(1)\ ,\langle N\rangle_1 \geq -\frac{\log\epsilon}{D(\sigma\|\rho)} + O(1) \quad (1)$$

for $\epsilon \ll 1$, where $\langle N\rangle_\nu$ is the mean number of copies given the true hypothesis is $\nu \in \{0,1\}$ and $D(\rho\|\sigma) = \mathrm{tr}\rho(\log\rho - \log\sigma)$ is the quantum relative entropy. In addition, we provide upper bounds which, for the worst case $N_{\mathrm{wc}} = \max\{\langle N\rangle_0, \langle N\rangle_1\}$, are achievable for some families of states.

Specifically, we consider quantum hypothesis testing in a scenario where one can guarantee that for *each* realization of the test the conditional probability of correctly identifying each of the hypotheses is above a given threshold. This scenario, first introduced in [10], can be considered genuinely sequential since such *strong error* conditions cannot be generally met in a deterministic setting. The proof method can be easily extended to the more common asymptotic *symmetric* and *asymmetric* scenarios involving the usual *type I* (or *false positive*) and *type II* (or *false negative*) errors. We give the optimal scaling of the mean number of copies when the thresholds for either one or both types of errors are asymptotically small.

Before proceeding, let us briefly review these fundamental hypothesis testing scenarios, which come about from the relative importance one places between type I error—the error of guessing the state to be σ when the true state is ρ whose probability we denote by $\alpha = P(\hat{H}_1|\rho)$—and type II error—the error of guessing ρ when the state is σ whose probability is $\beta = P(\hat{H}_0|\sigma)$. Often, the two types of errors are put on equal footing (*symmetric* scenario) and one seeks to minimise the mean probability of error $\bar{\epsilon} = \eta_0\alpha + \eta_1\beta$ with $\eta_0, \eta_1 = 1 - \eta_0$ the prior probabilities for each hypothesis. The mean error decays exponentially with the number of copies with an optimal rate given by the Chernoff distance [12, 13], $\xi_{\rm Ch} = -\inf_{0\leq s\leq 1}\log\operatorname{tr}(\rho^{1-s}\sigma^s)$.

Yet, there are asymmetric instances, e.g., in medical trials, where the effect of approving an ineffective treatment (type-II) is far worse than discarding a potentially good one (type-I). In such cases it is imperative to minimise the type II error whilst maintaining a finite probability of successfully identifying the null hypothesis, i.e., $p(\hat{H}_0|\rho) = 1 - \alpha \geq p_s > 0$. The corresponding optimal error rate for quantum hypotheses is given by quantum Stein's lemma [14, 15], $\beta \sim \mathrm{e}^{-\xi_S N}$ where $\xi_S = D(\rho\|\sigma)$ is the quantum relative entropy. If, on the other hand, we require the type I error to decay exponentially, i.e., $\alpha \leq \mathrm{e}^{-rN}$ for some rate r, then the optimal rate is given by the quantum Hoeffding bound [16, 17]. These optimal error rates for strategies with fixed number of copies have found applications in quantum Shannon theory [18], quantum illumination [19], and provide operational meaning to abstract information measures [20–22].

What the above results also show is that for N fixed there is a trade-off between the probabilities of committing either error. The advantage of sequential analysis is that it provides strategies capable of minimising the average number of copies when *both* errors are bounded, and yields higher asymptotic rates in each of the settings described above.

Fixed local measurements. We begin by considering the case when each quantum system is measured with the same measurement apparatus E, giving rise to identically distributed samples of a classical probability distribution. This strategy has the advantage of being easily implementable, and that it lets us introduce the classical sequential analysis framework. Specifically, the optimal classical sequential test, for both the strong error as well as the symmetric and asymmetric setting, is known to be the *Sequential Probability Ratio Test* (SPRT)[23] which we now review.

After n measurements have been performed, we have a string of outcomes $\mathbf{x}_n = \{x_1, x_2, \ldots, x_n\}$, where each element has been sampled effectively from a probability distribution determined by the POVM $E = \{E_x\}$ and the true state of the system, i.e., either $p(x) := \operatorname{tr}(E_x\rho)$, or $q(x) := \operatorname{tr}(E_x\sigma)$. For given error thresholds ϵ_0, ϵ_1, the strong condition demands that for each conclusive sequence the conditional probabilities obey either

$$P(\rho|\mathbf{x}_n) = \frac{\eta_0 p(\mathbf{x}_n)}{\eta_0 p(\mathbf{x}_n) + \eta_1 q(\mathbf{x}_n)} \geq 1 - \epsilon_0, \text{ or} \tag{2}$$

$$P(\sigma|\mathbf{x}_n) = \frac{\eta_1 q(\mathbf{x}_n)}{\eta_0 p(\mathbf{x}_n) + \eta_1 q(\mathbf{x}_n)} \geq 1 - \epsilon_1 \tag{3}$$

where $p(\mathbf{x}_n) = \prod_{k=1}^n p(x_k)$ since the copies are identical and independent (the same holds for q). If neither condition is met, a new copy needs to be requested and we continue measuring. That is, starting at $n = 1$ at every step n we check whether

1. $P(\rho|\mathbf{x}_n) \geq 1 - \epsilon_0$, then STOP and accept H_0, with guaranteed probability of success $s_0 = 1 - \epsilon_0$.

2. $P(\sigma|\mathbf{x}_n) \geq 1 - \epsilon_1$, then STOP and accept H_1, with guaranteed probability of success $s_1 = 1 - \epsilon_1$.

3. If neither 1 nor 2 hold, continue sampling.

Using (2) and (3), the condition to continue sampling can be written in terms of a single sample statistic, the log-likelihood ratio

$$Z_n = \log\frac{q(\mathbf{x}_n)}{p(\mathbf{x}_n)} = \sum_{k=1}^n z_k \text{ with } z_k = \log\frac{q(x_k)}{p(x_k)} \tag{4}$$

as $b := \log B \leq Z_n \leq \log A =: a$, where $A = \frac{\eta_0}{\eta_1}\frac{1-\epsilon_1}{\epsilon_1}$, $B = \frac{\eta_0}{\eta_1}\frac{\epsilon_0}{1-\epsilon_0}$.

It is convenient to interpret Z_n as a random walk (see Fig. 1) that at every instance performs a step of length z_k with probability $p(x_k)$, if H_0 holds, or with probability $q(x_k)$, if H_1 holds. Under H_1 the mean position of the walker at step n is given by $\langle Z_n\rangle_1 = \sum_{k=1}^n\langle z_k\rangle_1 = n\langle z\rangle_1 = nD(q\|p) > 0$ where $D(q\|p) = \sum_x q(x)\log\frac{q(x)}{p(x)}$ is the relative entropy; while for H_0, $\langle Z_n\rangle_0 = -nD(p\|q) < 0$. That is, under H_1 the walker has a drift towards the positive axis, while under H_0 it drifts towards the negative axis. We define as the stopping time N the first instance in which the walker steps out of the region (a, b), i.e., $N := \inf\{n : Z_n \notin (b, a)\}$, and note that it is a stochastic variable that *only* depends on the current as well as the past measurement

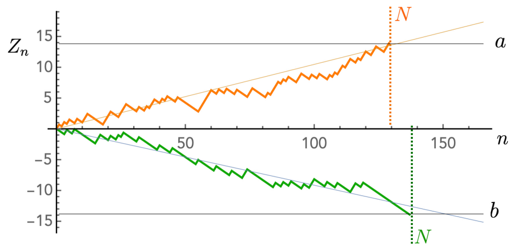


Figure 1. Random walk describing the likelihood function Z_n under H_0 (green) and H_1 (orange). When the value of Z_n crosses $b \sim \log \epsilon_0$ ($a \sim \log \frac{1}{\epsilon_1}$) we decide in favour of H_0 (H_1). N indicates the corresponding stopping time.

record. The stochastic variable $Z := Z_N$ is the position of the walker at N. The mean value of this position can be related to the mean number of steps by Wald's identity [24],

$$\langle Z\rangle_1 = \langle \sum_{k=1}^{N} z_k\rangle_1 = \langle z\rangle_1 \langle N\rangle_1 = D(q\|p)\langle N\rangle_1 \qquad (5)$$

under hypothesis H_1, and likewise $\langle Z\rangle_0 = -D(p\|q)\langle N\rangle_0$. In order to estimate $\langle N\rangle_i$ from (5) we need to provide a good estimate for $\langle Z\rangle_i$. For this purpose let us first define $\mathcal{X}_1$ as the set of strings $\mathbf{x}$ such that $b < Z_j < a$ for all $j < n$ *and* $Z_n \geq a$, and $\mathcal{X}_0$ as the set of strings $\mathbf{x}$ such that $b < Z_j < a$ for all $j < n$ *and* $Z_n \leq b$. Then, the following relations hold:

$$\begin{aligned} \alpha =& P_0(Z \geq a) = \sum_{\mathbf{x}\in\mathcal{X}_1} p(\mathbf{x}) \leq \sum_{\mathbf{x}\in\mathcal{X}_1} \frac{q(\mathbf{x})}{A} = \frac{1-\beta}{A} \\ \beta =& P_1(Z \leq b) = \sum_{\mathbf{x}\in\mathcal{X}_0} q(\mathbf{x}) \leq \sum_{\mathbf{x}\in\mathcal{X}_0} p(\mathbf{x}) B = (1-\alpha)B \end{aligned} \qquad (6)$$

where in the first (second) inequality we used that $\frac{q(\mathbf{x})}{p(\mathbf{x})} \geq A$ for strings in $\mathcal{X}_1$ ($\frac{q(\mathbf{x})}{p(\mathbf{x})} \leq B$ for strings in $\mathcal{X}_2$), and in the last equality we have used that $\lim_{n\to\infty} P(Z_n \in (b,a)) = 0$ [23], i.e. the walker eventually stops. The above equations are an instance of so-called Wald's likelihood ratio identity [25]. We note that the above inequalities can be taken to be approximate equalities if we assume that the process ends close to the prescribed boundary, i.e. there is no *overshooting*. In particular, this will be valid in our asymptotically small error settings where the boundaries are far relative to the (finite) step size z_k. This allows us to establish a one-to-one correspondence between the thresholds A, B and the type I & II errors: $\alpha \approx \frac{1-B}{A-B} = \frac{\epsilon_1(\eta_1-\epsilon_0)}{(1-\epsilon_0-\epsilon_1)\eta_0}$ and $\beta \approx \frac{B(A-1)}{A-B} = \frac{\epsilon_0(\eta_0-\epsilon_1)}{(1-\epsilon_0-\epsilon_1)\eta_1}$. [33, 34] Neglecting the overshooting also allows us to consider Z as a stochastic variable that takes two values $Z \in \{a, b\}$. Under hypothesis H_0, a occurs with probability $P_0(Z = a) = \alpha$ and b with $P_0(Z = b) = 1 - \alpha$; while under hypothesis H_1, a and b occur with probabilities $P_1(Z = a) = 1 - \beta$ and $P_1(Z = b) = \beta$. So,

$$\langle Z\rangle_0 = a\alpha + b(1-\alpha) \text{ and } \quad \langle Z\rangle_1 = a(1-\beta) + b\beta. \quad (7)$$

Making use of (5) one can now write a closed expression for $\langle N\rangle_0$ and $\langle N\rangle_1$ in terms of ϵ_1, ϵ_0 and the priors. A remarkable property of the SPRT with error probabilities α and β is that it minimizes *both* $\langle N\rangle_0$ and $\langle N\rangle_1$ among all tests (sequential or otherwise) with bounded type I and type II errors. This optimality result due to Wald and Wolfowitz [23] allows us to extend the above results to the asymmetric scenario. For the symmetric scenario, the SPRT has also been shown [26] to be optimal among all tests respecting a bounded mean error $\bar{\epsilon}' \leq \bar{\epsilon}$. In the asymptotic limit of small error bounds, $\epsilon_0, \epsilon_1 \ll 1$, the threshold values are $a \sim -\log \epsilon_1$ and $b \sim \log \epsilon_0$, which correspond to $\alpha \sim \frac{\eta_1}{\eta_0}\epsilon_1$ and $\beta \sim \frac{\eta_0}{\eta_1}\epsilon_0$, yielding

$$\langle N\rangle_0 \sim -\frac{\log \epsilon_0}{D(p\|q)} \text{ and } \langle N\rangle_1 \sim -\frac{\log \epsilon_1}{D(q\|p)}. \qquad (8)$$

The same expressions hold at leading order in the asymmetric scenario when the type I & II errors are vanishingly small, replacing $\log \epsilon_1$ and $\log \epsilon_0$ by $\log \alpha$ and $\log \beta$ respectively—and in the symmetric scenario replacing both quantities by $\log \bar{\epsilon}$. If one of the error thresholds, say α, is kept finite while the second is made vanishingly small $\beta \ll 1$, $\langle N\rangle_1$ remains finite, while the other conditional mean scales as $\langle N\rangle_0 \sim -\frac{(1-\alpha)\log\beta}{D(p\|q)}$.

In the supplemental material [27] we apply these results to the discrimination of qubit states using projective measurements and give closed expressions for the optimal Bayesian mean number of copies $\langle N\rangle := \eta_0\langle N\rangle_0 + \eta_1\langle N\rangle_1$. Figure 2 shows that in the symmetric setting these restricted sequential strategies already require on average between 25-50% less resources than the best deterministic strategy that uses a fixed number of copies $N_{\text{Ch}} \sim -\log \bar{\epsilon}/\xi_{\text{Ch}}$ [12, 21], and requires non-trivial collective measurements [35].

Ultimate quantum limit. Quantum mechanics allows for much more sophisticated strategies. For a start, performing a non-projective generalized measurement already gives important advantages (see below). One can also adapt the measurements depending on the previous measurement outcomes and, importantly, measurements may be weak so that each new measurement acts on a fresh copy but also on the preceding, already measured, copies. Without loss of generality we can assume that at every step k we perform a measurement with three outcomes $x_k \in \{0, 1, 2\}$: the first two must fulfill conditions (2) and (3) and trigger the corresponding guess (H_0 or H_1 respectively), while the third outcome signals to continue measuring having an additional fresh copy available. The measurement at step k is characterized by a quantum instrument $\mathcal{M}^k = \{\mathcal{M}_0^k, \mathcal{M}_1^k, \mathcal{M}_2^k\}$, and the sequential strategy is given by a sequence of instruments $\mathcal{M} = \{\mathcal{M}^k\}_{k=1}^{\infty}$. With this, given hypothesis $\nu = \{0, 1\}$,

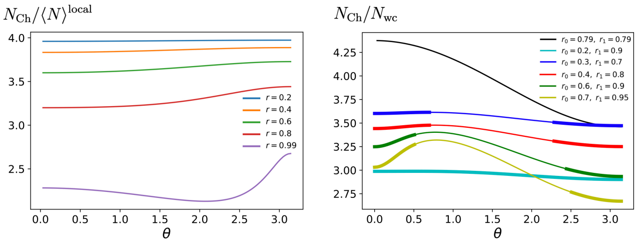


Figure 2. Left: ratio between the number of copies required by the best deterministic strategy, $N_{\rm Ch}$, and the Bayesian mean number of copies for a sequential strategy based on fixed local unbiased measurements, $\langle N\rangle_{\rm local}$, for pairs of states of purity r and relative angle θ [27]. Right: ratio between $N_{\rm Ch}$ and the worst-case $N_{\rm wc}$, for pairs of states of different purities r_0, r_1. The thin lines use the expression (12) for $N_{\rm wc}$, whereas the thick lines represent the cases for which this ultimate limit of $N_{\rm wc}$ is attained by a block-sampling strategy.

the probability of getting outcome x_k at step k can be written as

$$P_\nu(x_k) := \mathrm{tr}[\mathcal{M}^k_{x_k} \circ \mathcal{M}^{k-1}_2 \ldots \circ \mathcal{M}^1_2(\rho_\nu^{\otimes k})] = \mathrm{tr}(E^k_{x_k}\rho_\nu^{\otimes k}) \tag{9}$$

where we have used that in order to arrive to step k a "continue" outcome must be triggered in all previous steps, and in the last equality we have defined the effective POVM $E^k = \{E^k_i\}_{i=0}^2$. Making use of the indicator function $\mathbb{1}_{k\leq N}$, the mean number of steps under hypothesis ν can be computed as

$$\langle N\rangle_\nu = \langle\sum_{k=1}^{N} 1\rangle_\nu = \langle\sum_{k=1}^{\infty} \mathbb{1}_{k\leq N}\rangle_\nu = \langle\sum_{n=0}^{\infty} \mathbb{1}_{n<N}\rangle_\nu = \sum_{n=0}^{\infty} T^n_\nu \tag{10}$$

where $T^n_\nu = P_\nu(n < N)$ is the probability that the sequence does not stop at step n, which from (9) is given by $T^n_\nu = P_\nu(x_n = 2)$. Optimizing $\langle N\rangle_\nu$ over all quantum sequential strategies $\mathcal{M}$ is daunting, as all terms T^n_ν are strongly interrelated through the intricate structure of E^n. However, a lower bound to each T^n_ν can be found by relaxing such structure and only imposing minimal requirements on the effective POVM; namely the error bounds (6), positivity and completeness:

$$\min_{E^n} \mathrm{tr}(E^n_2\rho_\nu^{\otimes n}) \text{ s.t. } E^n_i \geq 0,\ \sum_{i=0}^{2} E^n_i = \mathbb{1}, \text{ and} \tag{11}$$
$$\mathrm{tr}[E^n_1\left(\sigma^{\otimes n} - A\rho^{\otimes n}\right)] \geq 0,\ \mathrm{tr}[E^n_0\left(\rho^{\otimes n} - B^{-1}\sigma^{\otimes n}\right)] \geq 0.$$

This semi-definite program, which can be considered a two-sided version of the quantum Neyman-Pearson test [20], is an interesting open problem in its own right. Our focus, however, is the asymptotic regime of small error bounds. In these asymptotic scenarios we are able to show, exploiting some recent strong converse results in hypothesis testing [31, 36], that for all $n < n^* = -\frac{\log\epsilon_0(1-A^{-1})}{D(\rho\|\sigma)}$, $T^n_0 \geq 1 - O(\epsilon_0^\kappa)$ for some $\kappa \in (0,1)$ [27], which leads to the desired bound:

$$\langle N\rangle_0 \geq \sum_{n=0}^{\lfloor n^*\rfloor} T^n_0 \geq -\frac{\log\epsilon_0(1-A^{-1})}{D(\rho\|\sigma)} + O(1)\,. \tag{12}$$

An analogous bound holds for $\langle N\rangle_1$. The bounds for asymmetric (symmetric) scenarios (see [27]) take the same form, replacing $\log\epsilon_0$ by $\log\beta$ ($\log\bar{\epsilon}$) and A^{-1} by α ($\bar{\epsilon}$). In the asymmetric scenario where ϵ_1 or α is kept finite, it also holds that $\langle N\rangle_1 = O(1)$ and $\langle N\rangle_0$ is given by the appropriate version of (12).

Attainability and upper bounds. Consider a sequential strategy that involves a fixed, collective measurement $K = \{K_i\}$, acting on consecutive blocks of ℓ copies, yielding two possible distributions p^ℓ_K, q^ℓ_K. Using the classical SPRT we get that

$$\langle N\rangle_0 = \ell\inf_K\langle M\rangle_0 \sim \ell\inf_K\frac{-\log\epsilon_0}{D(p^\ell_K\|q^\ell_K)} \sim \frac{-\log\epsilon_0}{D(\rho\|\sigma)} \tag{13}$$

where M is the number of blocks used at the stopping time. In the last relation of (13) we have used the fact that we are in the asymptotic setting where $\epsilon_0 \ll 1$ and therefore we can take arbitrarily long block lengths $\ell \gg 1$. We also exploit the following property of the measured relative entropy [14, 37]: $\sup_K D(q^\ell_K\|p^\ell_K) \sim \ell D(\sigma\|\rho)$.

Notice, however, that for arbitrary states ρ and σ block sampling can attain either $\langle N\rangle_0$ or $\langle N\rangle_1$, but it is unknown whether one can attain in general both bounds simultaneously, i.e., whether a measurement achieving the supremum of $\lim_{\ell\to\infty}\frac{1}{\ell}D(q^\ell_K\|p^\ell_K)$ can also attain the supremum of $\lim_{\ell\to\infty}\frac{1}{\ell}D(p^\ell_K\|q^\ell_K)$. For instance, if we wish to optimize the Bayesian mean number of copies $\langle N\rangle$, we can use block sampling to attain

$$\langle N\rangle_{\rm block} \sim \lim_{\ell\to\infty}\inf_K\left(\frac{-\ell\eta_0\log\epsilon_0}{D(p^\ell_K\|q^\ell_K)} - \frac{\ell\eta_1\log\epsilon_1}{D(q^\ell_K\|p^\ell_K)}\right). \tag{14}$$

However, this strategy might be sub-optimal and hence it only provides an upper bound to the optimal Bayesian mean $\langle N\rangle \leq \langle N\rangle_{\rm block}$. This notwithstanding, there are at least two cases when this upper bound coincides with the lower bound provided by (12): when ρ and σ commute, and when the two states do not have common support. If, say, $\mathrm{supp}(\sigma)\cap\ker(\rho) \neq 0$, one can use block-sampling to attain (12) for $\langle N\rangle_0$ and always detect ρ with a finite number of copies—note that since $D(\sigma\|\rho) = \infty$, the lower bound $\langle N\rangle_1 = O(1)$ is also attained.

We can also give achievable lower bounds for a worst-case type figure of merit $N_{\rm wc} := \max\{\langle N\rangle_0, \langle N\rangle_1\}$. If, say, $\langle N\rangle_0 > \langle N\rangle_1$, then in [27] we give some instances of qubit pairs where a specific block-sampling strategy [37] saturates (12) for $\langle N\rangle_0$, while at the same time $\lim_{\ell\to\infty}\frac{1}{\ell}D(q^\ell_K\|p^\ell_K) \geq D(\rho\|\sigma)$, and hence (12) provides the ultimate attainable limit for $N_{\rm wc}$. In Fig. 2 we compare $N_{\rm wc}$ with $N_{\rm Ch}$ for several pairs of states, highlighting the achievable cases, and show a consistent advantage of sequential protocols over deterministic ones [38].

Finally, we note that, in an asymmetric scenario where $\langle N\rangle_1$ is finite and the value of $\langle N\rangle_0$ achieves the lower bound (12), sequential protocols provide a strict advantage over Stein's limit for deterministic protocols by a factor $(1-\alpha)$.

The curious case of pure states. If the two states are pure, the behavior of $\langle N\rangle_\nu$ changes drastically: it is possible to reach a decision with guaranteed zero error using a finite average number of copies. To see this, consider again Eq. (18). Under a zero-error condition, the minimal (unrestricted) T_ν^n is achieved by a global *unambiguous* three-outcome POVM [39–41] on n copies, which identifies the true state with zero error when the first or the second outcome occurs —at the expense of having a third, inconclusive outcome. For a single-copy POVM over pure states, the probabilities c_ν of the inconclusive outcome under H_ν are subject to the tradeoff relation $c_0 c_1 \geq \mathrm{tr}\rho\sigma$ [6], where equality can always be attained by suitable POVM that maximizes the probability of a successful identification. Likewise, for a global measurement on n copies we have $T_0^n T_1^n \geq (\mathrm{tr}\rho\sigma)^n$. Now, it is evident that a sequence of n locally optimal unambiguous POVMs applied on every copy, for which $T_\nu^n = c_\nu^n$, also fulfills the global optimality condition. Hence, we have

$$\langle N\rangle_\nu \geq \sum_{n=0}^{\infty} T_\nu^n = \sum_{n=0}^{\infty} c_\nu^n = \frac{1}{1-c_\nu} =: \langle N\rangle_\nu^{\mathrm{local}} . \tag{15}$$

This shows that, for pure states, it suffices to perform local unambiguous measurements to attain the optimal (finite) average number of copies with zero error under hypothesis H_ν. Note that because of the tradeoff $c_0 c_1 \geq \mathrm{tr}\rho\sigma$ one cannot attain the minimal values of $\langle N\rangle_0^{\mathrm{local}}$ and $\langle N\rangle_1^{\mathrm{local}}$ for general states ρ, σ, simultaneously. For instance, one can reach the minimal value $c_0 = \mathrm{tr}(\rho\sigma)$ for one hypothesis, but then having a maximal value $c_1 = 1$ for the second; or choose the optimal symmetric setting, $c_0 = c_1 = \sqrt{\mathrm{tr}\rho\sigma}$, that achieves the minimum value of both the worst-case N_{wc} and the Bayesian mean $\langle N\rangle$ with equal priors (see [27]). This is in stark contrast with the behavior found in [10], where all strategies considered were based on two-outcome projective measurements, for which the average number of copies scaled as $\langle N\rangle \propto -\log\epsilon$.

Acknowledgments. We acknowledge financial support from the Spanish Agencia Estatal de Investigación, project PID2019-107609GB-I00, from Secretaria d'Universitats i Recerca del Departament d'Empresa i Coneixement de la Generalitat de Catalunya, co-funded by the European Union Regional Development Fund within the ERDF Operational Program of Catalunya (project QuantumCat, ref. 001-P-001644), and Generalitat de Catalunya CIRIT 2017-SGR-1127. CH acknowledge financial support from the European Research Council (ERC Grant Agreement No. 81876), VILLUM FONDEN via the QMATH Centre of Excellence (Grant No.10059) and the QuantERA ERA-NET Cofund in Quantum Technologies implemented within the European Union Horizon 2020 Programme (QuantAlgo project) via Innovation Fund Denmark, JC acknowledges from the QuantERA grant C'MON-QSENS!, via Spanish MICINN PCI2019-111869-2, EMV thanks financial support from CONACYT. JC also acknowledges support from ICREA Academia award.

* Esteban.Martinez@uab.cat
† christoph.hirche@gmail.com
‡ Gael.Sentis@uab.cat
§ michail.skoteiniotis@uab.cat
¶ Ramon.Munoz@uab.cat
** John.Calsamiglia@uab.cat

[1] E. L. Lehmann and G. Casella, *Theory of point estimation*, 2nd ed. (Springer Science & Business Media, 2006).

[2] C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic Press, 1976).

[3] A. G. Tartakovsky, I. V. Nikiforov, and M. Basseville, *Sequential Analysis: hypothesis testing and changepoint detection*, edited by F. Bunea, V. Isham, N. Keiding, T. Louis, R. L. Smith, and H. Tong (CRC Press, Taylor and Francis, 2014).

[4] G. Sentís, E. Bagan, J. Calsamiglia, G. Chiribella, and R. Muñoz Tapia, Phys. Rev. Lett. **117**, 150502 (2016).

[5] G. Sentís, J. Calsamiglia, and R. Muñoz Tapia, Phys. Rev. Lett. **119**, 140506 (2017).

[6] G. Sentís, E. Martínez-Vargas, and R. Muñoz Tapia, Phys. Rev. A **98**, 052305 (2018).

[7] A. Wald, *Sequential Analysis*, Dover books on advanced mathematics (Dover Publications, 1973).

[8] T. Leung Lai, Statistica Sinica **11**, 303 (2001).

[9] J. Bae and L.-C. Kwek, Journal of Physics A: Mathematical and Theoretical **48**, 083001 (2015).

[10] S. Slussarenko, M. M. Weston, J.-G. Li, N. Campbell, H. M. Wiseman, and G. J. Pryde, Physical Review Letters **118**, 030502 (2017).

[11] By $f(\epsilon) \sim g(\epsilon)$ we mean asymptotic equivalence $\lim_{\epsilon\to 0} f(\epsilon)/g(\epsilon) = 1$.

[12] K. M. R. Audenaert, J. Calsamiglia, R. Muñoz-Tapia, E. Bagan, L. Masanes, A. Acín, and F. Verstraete, Physical Review Letters **98**, 160501 (2007).

[13] M. Nussbaum and A. Szkoła, Annals of Statistics **37**, 1040 (2009).

[14] F. Hiai and D. Petz, Communications in Mathematical Physics **143**, 99 (1991).

[15] T. Ogawa and H. Nagaoka, IEEE Transactions on Information Theory **46**, 2428 (2000).

[16] M. Hayashi, Physical Review A **76**, 062301 (2007).

[17] H. Nagaoka, arXiv:quant-ph/0611289 (2006).

[18] M. M. Wilde, *Quantum Information Theory*, 2nd ed. (Cambridge University Press, 2017).

[19] S. Lloyd, Science **321**, 1463 (2008).

[20] K. M. R. Audenaert, M. Nussbaum, A. Szkola, and F. Verstraete, Communications in Mathematical Physics **279**, 251 (2008).

[21] J. Calsamiglia, R. Muñoz-Tapia, L. Masanes, A. Acin, and E. Bagan, Physical Review A **77**, 032311 (2008).

[22] M. Berta, F. G. Brandao, and C. Hirche, arXiv:1709.07268 (2017).

[23] A. Wald and J. Wolfowitz, The Annals of Mathematical Statistics **19**, 326 (1948).

[24] M. Mitzenmacher and E. Upfal, *Probability and Computing: Randomized Algorithms and Probabilistic Analysis* (Cambridge University Press, 2005).

[25] T. L. Lai, Sequential Analysis **23**, 467 (2004).

[26] G. Simons, The Annals of Statistics **4**, 1240 (1976).

[27] See Supplemental Material [url] where we provide the proof of the lower bound for the mean number of copies under each hypothesis, we apply our general results to the case of qubit states, we compute the exact optimal mean number of copies (worst-case and Bayesian) required for the perfect discrimination of pure states, and which includes [28–34].

[28] M. Mosonyi and T. Ogawa, Communications in Mathematical Physics **334**, 1617 (2015).

[29] M. Müller-Lennert, F. Dupuis, O. Szehr, S. Fehr, and M. Tomamichel, Journal of Mathematical Physics **54**, 122203 (2013).

[30] M. M. Wilde, A. Winter, and D. Yang, Communications in Mathematical Physics **331**, 593 (2014).

[31] S. Beigi, N. Datta, and C. Rouzé, Communications in Mathematical Physics **376**, 753 (2020).

[32] A. Acín, E. Bagan, M. Baig, L. Masanes, and R. Muñoz-Tapia, Physical Review A **71**, 032338 (2005).

[33] B. Gendra, E. Ronco-Bonvehi, J. Calsamiglia, R. Muñoz-Tapia, and E. Bagan, New Journal of Physics **14**, 105015 (2012).

[34] K. M. Audenaert and J. Eisert, Journal of Mathematical Physics **46**, 102104 (2005) .

[35] J. Calsamiglia, J. I. de Vicente, R. Muñoz-Tapia, and E. Bagan, Physical Review Letters **105**, 080504 (2010).

[36] T. Cooney, M. Mosonyi, and M. M. Wilde, Communications in Mathematical Physics **344**, 797 (2016).

[37] M. Hayashi, Journal of Physics A: Mathematical and General **34**, 3413 (2001).

[38] Note that the comparison with $N_{\rm Ch}$ is unfavorable to sequential strategies. Substituting ϵ_0 and ϵ_1 by $\bar\epsilon$ in Eq. (12) implies that *each* type of error is independently constrained, whereas $N_{\rm Ch}$ refers to a deterministic (symmetric) protocol where the *mean* error is $\bar\epsilon$ and thus to a weaker version of the problem. In spite of this, the sequential scenario displays a significant advantage.

[39] I. Ivanovic, Physics Letters A **123**, 257 (1987).

[40] D. Dieks, Physics Letters A **126**, 303 (1988).

[41] A. Peres, Physics Letters A **128**, 19 (1988).

SUPPLEMENTAL MATERIAL

This supplemental material contains some technical details as well as some extensions for the interested reader. In Sec. we state and prove a theorem that provides the lower bounds for the average number of copies, Eq. (12) in the main text. In Sec. explicit expressions for the optimal sequential test and attainability regions of the worst case bound are provided for the qubit case. In Sec. we present a second theorem that provides a general lower bound for the deviation of the measured entropy from its maximum value for arbitrary finite dimensions. Finally, in Sec. we give the optimality proof for the zero-error protocol for pure states.

CONVERSE PROOF

Our aim here is to prove that if one of the two error bounds is vanishingly small, say $\epsilon_0 \ll 1$ under the strong error condition, or $\beta \ll 1$ in the asymmetric scenario (see main text for the definitions), then the mean number of copies under hypothesis H_0 is always lower-bounded by

$$\langle N\rangle_0 \geq \sum_{n=0}^{\lfloor n^*\rfloor} T_0^n \geq -\frac{\log \epsilon_0(1-1/A)}{D(\rho\|\sigma)} + O(1) \qquad \text{or} \tag{16}$$

$$\langle N\rangle_0 \geq \sum_{n=0}^{\lfloor n^*\rfloor} T_0^n \geq -\frac{\log \beta(1-\alpha)}{D(\rho\|\sigma)} + O(1)\,, \tag{17}$$

respectively. Analogous bounds also hold for $\langle N\rangle_1$ when $\epsilon \ll 1$ or $\alpha \ll 1$, replacing $\epsilon_0 \leftrightarrow \epsilon_1$, $A \leftrightarrow B$, $\alpha \leftrightarrow \beta$, and $\rho \leftrightarrow \sigma$. We will first provide the proof for the strong error condition and indicate how to adapt it to the other hypothesis testing scenarios considered here.

In the main text (MT) we have shown that under hypothesis H_0 the mean number of sampled copies is given by

$$\langle N\rangle_0 = \sum_{n=0}^{\infty} T_0^n \geq \sum_{n=0}^{n^*} T_0^n\,, \tag{18}$$

where the last inequality holds for all values of n^* since $T_0^n \geq 0$. The nth term in sum, $T_0^n = P_0(n < N)$, is the probability of getting a "continue" outcome at step n, corresponding to the POVM element E_2^n implicitly defined in (12) in MT. The continue probability at a particular step n obeys a lower bound given by the following semidefinite program (SDP)

$$T_0^n \geq \tilde{T}_0^n := \min_{E^n} \mathrm{tr}(E_2^n \rho_\nu^{\otimes n}) \text{ s.t.} \begin{cases} 0) \quad \{E_i^n \geq 0\}_{i=0}^2 \text{ and} \sum_{i=0}^2 E_i^n = \mathbb{1} \\ 1) \quad \mathrm{tr}[E_1^n \left(\sigma^{\otimes n} - A\rho^{\otimes n}\right)] \geq 0 \\ 2) \quad \mathrm{tr}[E_0^n \left(\rho^{\otimes n} - B^{-1}\sigma^{\otimes n}\right)] \geq 0 \end{cases} . \tag{19}$$

The conditions in 0) have to hold for any valid POVM, while the second and third conditions are an alternative way of writing the strong errors conditions, (1) and (2) in the MT, as SDP constrains. For small error bounds $\epsilon_0 \ll 1$ (i.e. $A = \frac{\eta_0}{\eta_1}\frac{1-\epsilon_1}{\epsilon_1} \gg 0$) the solution of the SDP program in (19) has a characteristic dependence on n as illustrated in Figure 3. When the number of sampled copies n is small it is not possible to meet the low error bound and the probability of getting a "continue" outcome is $\tilde{T}_0^n = 1$. This probability remains constant as n increases until it approaches the critical point $n^* \sim \log(1/\epsilon)/D(\rho\|\sigma)$, at which point it rapidly drops to zero. Note that this drop becomes more abrupt as ϵ_0 decreases. These observations suggest that $\sum_{n=0}^{n^*} \tilde{T}_0^n$ is a very tight lower bound to $\sum_{n=0}^{\infty} \tilde{T}_0^n$ (area under the curve in Figure 3) and is given to a very good approximation by $\sum_{n=0}^{n^*} \tilde{T}_0^n \approx n^*$.

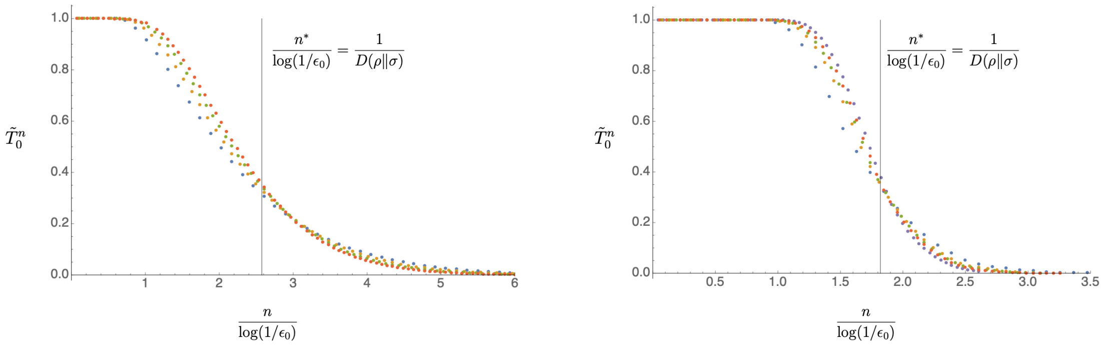


Figure 3. Lower bound $\tilde{T}_0^n$ on the continue probability as a function of the step number n for: (Left) two qubits of equal purity $r = 0.9$ and relative angle $\theta = \pi/4$, for error bounds $\epsilon_0 = \{10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}\}$ (from bottom to top, looking at the left side of the figure); (Right) two commuting qubits with $r = 0.5$ and $\theta = \pi$, for $\epsilon_0 = \{10^{-8}, 10^{-12}, 10^{-14}, 10^{-16}, 10^{-20}\}$ (from bottom to top, looking at the left side of the figure). The values on the left plot have been obtained by numerically solving the SDP program of (19), exploiting the block-diagonal structure of iid quantum states $\rho^{\otimes n}$ and $\sigma^{\otimes n}$ (see Section).

With this at hand we can now carry on with the formal presentation and proof of the lower bound.

Theorem 1. *Given two finite-dimensional states, ρ (H_0) and σ (H_1), occurring with prior probabilities η_0 and η_1 respectively, the most general quantum sequential strategy that satisfies the strong error conditions $P(H_0|x_N = 0) \geq 1 - \epsilon_0$ and $P(H_1|x_N = 1) \geq 1 - \epsilon_1$, where $x_N \in \{0, 1\}$ is the output of the measurement at the stopping time N, necessarily fulfills the following asymptotic lower bound for the mean number of sampled copies when $\epsilon_0 \ll 1$:*

$$\langle N \rangle_0 \geq -\frac{(1 - A^{-1}) \log \epsilon_0}{D(\rho\|\sigma)} + O(1) , \qquad \textit{where} \qquad A = \frac{\eta_0}{\eta_1}\frac{1 - \epsilon_1}{\epsilon_1}. \tag{20}$$

Similarly, the most general quantum sequential strategy that satisfies the (weak) error conditions $P(\hat{H}_1|\rho) \leq \alpha$ and $P(\hat{H}_0|\sigma) \leq \beta$, where $\hat{H}_0$ and $\hat{H}_1$ are the events of accepting hypothesis 0 and 1 respectively, necessarily fulfills the following asymptotic lower bound for the mean number of sampled copies when $\beta \ll 1$:

$$\langle N \rangle_0 \geq -\frac{(1 - \alpha) \log \beta}{D(\rho\|\sigma)} + O(1) . \tag{21}$$

Proof. We start by noting that the strong error conditions [see (19)] imply

$$\mathrm{tr}(E_1^n\rho^{\otimes n}) \leq \frac{1}{A}\mathrm{tr}(E_1^n\sigma^{\otimes n}) \leq \frac{1}{A} = \frac{\eta_1}{\eta_0}\frac{\epsilon_1}{1-\epsilon_1}\,, \tag{22}$$

$$\mathrm{tr}(E_0^n\sigma^{\otimes n}) \leq \frac{1}{B}\mathrm{tr}(E_0^n\rho^{\otimes n}) \leq \frac{1}{B} = \frac{\eta_0}{\eta_1}\frac{\epsilon_0}{1-\epsilon_0}\,. \tag{23}$$

Next we form a two-outcome POVM by binning two outcomes of the effective POVM at step n, as defined in (12) in MT, $F^n = \{F_0^n = E_0^n, F_1^n := E_1^n + E_2^n\}$. This measurement can be used to discriminate between $\rho^{\otimes n}$ and $\sigma^{\otimes n}$ and the associated type-I and type-II errors will be denoted by $\tilde{\alpha}_n = \mathrm{Tr}[F_1^n\rho^{\otimes n}]$ and $\tilde{\beta}_n = \mathrm{Tr}[F_0^n\sigma^{\otimes n}]$. From (23) and the above definitions it follows that $\tilde{\beta}_n \leq \frac{\eta_0}{\eta_1}\epsilon_0 + O(\epsilon_0^2)$. In addition, using (23) we find that the probability of continuing at step n when H_0 holds satisfies

$$T_0^n = \mathrm{tr}[E_2^n\rho^{\otimes n}] = \tilde{\alpha}_n - \mathrm{tr}[E_1^n\rho^{\otimes n}] \geq \tilde{\alpha}_n - \frac{1}{A}\,. \tag{24}$$

Now we use Lemma 1 stated below, which uses the recently developed methods for strong converse exponents [36] in order to establish a lower bound on $\tilde{\alpha}_n$ when the type-II error $\tilde{\beta}_n$ is bounded by ϵ and when n is below a critical threshold n^*. In particular, applying Lemma 1 to the test defined by F_n above, with type-I & II errors $\tilde{\alpha}_n$ and $\tilde{\beta}_n \leq \frac{\eta_0}{\eta_1}\epsilon_0 + O(\epsilon_0^2) =: \epsilon$, we have that (24) reads

$$T_0^n \geq \tilde{\alpha}_n - \frac{1}{A} \geq 1 - \epsilon^{\kappa_n(\rho,\sigma)} - \frac{1}{A} \quad \forall\; n < n^*\,, \tag{25}$$

where $\kappa_n(\rho,\sigma) > 0$. The proof for the strong error conditions ends by inserting this lower bound in (18).

For the weak form of error bounds one can follow the same steps as above by writing the type-I and type-II errors of the sequential strategy as $P(\hat{H}_1|\rho) = \sum_{k=1}^{\infty}\tilde{\alpha}_n$ and $P(\hat{H}_0|\sigma) = \sum_{k=1}^{\infty}\tilde{\beta}_n$. Since $\tilde{\beta}_n \geq 0$, the error bound $P(\hat{H}_0|\sigma) \leq \beta$ translates to $\tilde{\beta}_n \leq \beta$, and similarly $\tilde{\alpha}_n \leq \alpha$. The former is directly of the form required for Lemma 1, while the latter can be used instead of (22), i.e., $\mathrm{tr}(E_1^n\rho^{\otimes n}) = \alpha_n \leq \alpha$, hence A in (20) becomes α in (21).

□

Lemma 1. *Let ρ and σ be finite-dimensional density operators associated to hypotheses H_0 and H_1, respectively. For any quantum hypothesis testing strategy that uses n copies of the states and that respects the type-II error bound $\beta_n \leq \epsilon$, with $\epsilon \ll 1$, the type-I error will converge to one at least as*

$$\alpha_n \geq 1 - \epsilon^{\kappa_n(\rho,\sigma)} \quad \textit{for all} \quad n < n^* = \frac{-\log\epsilon}{D(\rho\|\sigma)}\,, \tag{26}$$

where $0 < \kappa_n(\rho,\sigma) < 1$ is given by

$$\kappa_n(\rho,\sigma) = \sup_{s>1}\frac{s-1}{s}\frac{\xi_n - \tilde{D}_s(\rho\|\sigma)}{\xi_n} = \frac{H(\xi_n)}{\xi_n}\,, \quad \text{with} \quad \xi_n = -\frac{\log\epsilon}{n} > -\frac{\log\epsilon}{n^*} = D(\rho\|\sigma)\,, \tag{27}$$

where $H(\xi_n)$ is the strong converse exponent [28] and where the sandwiched Renyi relative entropy [29, 30] is given by

$$\tilde{D}_s(\rho\|\sigma) = \frac{1}{s-1}\log\mathrm{tr}\Big(\sigma^{\frac{1-s}{2s}}\rho\sigma^{\frac{1-s}{2s}}\Big)^s\,, \tag{28}$$

taking $\tilde{D}_s(\rho\|\sigma) = \infty$ when $\mathrm{supp}\,\rho \not\subseteq \mathrm{supp}\,\sigma$.

Proof. The proof makes use of the following strong converse result by Mosonyi and Ogawa [28] that relates the type I and type II errors for an arbitary n by means of the sandwiched Renyi relative entropy:

$$\frac{1}{n}\log(1-\alpha_n) \leq \frac{s-1}{s}(\tilde{D}_s(\rho\|\sigma) + \frac{1}{n}\log\beta_n), \quad s > 1\,. \tag{29}$$

Note that in order to avoid confusion with the type-I error, here we use s instead of the traditional α used in the Renyi entropies. Among the number of properties that make the sandwiched Renyi relative entropy such a formidable quantity, here we will use two: i) it increases monotonically with s, and ii) $\lim_{s\to 1}\tilde{D}_s(\rho\|\sigma) = D(\rho\|\sigma)$.

Since $\beta_n \leq \epsilon$,

$$1 - \alpha_n \leq \mathrm{e}^{n \frac{s-1}{s}(\tilde{D}_s(\rho\|\sigma) + \frac{1}{n}\log\epsilon)} \,. \tag{30}$$

Observe that $\forall n < n^*$ we can define $\xi_n > D(\rho\|\sigma)$ such that

$$n = -\frac{\log\epsilon}{\xi_n} \,. \tag{31}$$

Using this parametrization of n in (30), we have

$$\alpha_n \geq 1 - \epsilon^{\frac{s-1}{s}\frac{\xi_n - \tilde{D}_s(\rho\|\sigma)}{\xi_n}} \,. \tag{32}$$

Hence, if we define the supremum of the exponent

$$\kappa_n(\rho,\sigma) := \sup_{s>1} \frac{s-1}{s}\frac{\xi_n - \tilde{D}_s(\rho\|\sigma)}{\xi_n} \,, \quad \text{with} \quad \xi_n = -\frac{\log\epsilon}{n} > -\frac{\log\epsilon}{n^*} = D(\rho\|\sigma) \,, \tag{33}$$

we arrive to the desired result

$$\alpha_n \geq 1 - \epsilon^{\kappa_n(\rho,\sigma)} \qquad \forall\, n < n^* \,. \tag{34}$$

Taking into account that $0 < \frac{s-1}{s} = 1 - \frac{1}{s} < 1$ for all $s > 1$, that $\xi_n > D(\rho\|\sigma)$ for $n < n^*$, and from conditions i) and ii) above that $\tilde{D}_s(\rho\|\sigma) > D(\rho\|\sigma)$, it follows that there will always be an s' realizing the supremum in (33) such that $\xi_n > \tilde{D}_{s'}(\rho\|\sigma)$, and therefore $0 < \kappa_n(\rho,\sigma) < 1$. □

An alternative way to arrive to the result in Lemma 1 is provided in Beigi *et al.* [31] where, using quantum reverse hypercontractivity, a second order strong converse result on hypothesis testing is derived.

We finally note that Lemma 1 assures that, below n^*, the continue probability is $T^n \sim 1$. On the other hand, from Stein's Lemma we know that, for fixed (large) n, the optimal type-II error rate is given by the relative entropy, i.e., $\beta_n \sim \mathrm{e}^{-nD(\rho\|\sigma)}$. This explains why one does not need to continue measuring after $n > n^* = -\log\epsilon/D(\rho\|\sigma)$, and $T^{n>n^*} \sim 0$ (see Fig. 3), and why we may expect the lower bound to be tight, in the sense that we are not dropping significant contributions by truncanting the sum in (18). Of course, this still does not imply the attainability of the lower bound, and even less the simultaneous attainability of the bound for $\langle N\rangle_0$ and the analogous bound for $\langle N\rangle_1$.

SEQUENTIAL HYPOTHESIS TESTING FOR QUBITS

In this section we study the discrimination of qubit states using sequential methodologies, deriving explicit formulae for the mean number of copies using different measurement strategies.

Optimal sequential test for fixed projective measurements

We will first study the optimal performance under the simplest type of measurement apparatus, i.e. a fixed Stern-Gerlach-type measurement. The main purpose of this section is to show that using sequential strategies a simple projective measurement can determine the correct hypothesis with guaranteed bounded error requiring an expected number of copies significantly lower than the most general collective measurement acting on a fixed number of copies. In addition, we provide closed expressions for the optimal asymptotic performance.

Without loss of generality we characterize the two hypotheses by

$$\begin{aligned} \rho &= r_0\,|\psi_0\rangle\langle\psi_0| + (1-r_0)\mathbb{1}/2 \\ \sigma &= r_1\,|\psi_1\rangle\langle\psi_1| + (1-r_1)\mathbb{1}/2\,, \end{aligned} \tag{35}$$

where $|\psi_i\rangle = \cos\frac{\theta}{4}\,|0\rangle + (-1)^i \sin\frac{\theta}{4}\,|1\rangle$, $0 \le \theta \le \pi$, $0 \le r_i \le 1$ and the (fixed) local measurement as $E_0 = |\phi\rangle\langle\phi|$ and $E_1 = \mathbb{1} - E_0$, with $|\phi\rangle = \cos\frac{\phi}{2}\,|0\rangle + \sin\frac{\phi}{2}\,|1\rangle$, and $0 \le \phi \le \pi$. With these parametrizations, the probabilities of obtaining outcome $i = 0, 1$ are $p_\phi(i) = P(i|H_0) = \frac{1}{2}[1 + (-1)^i\cos(\theta/2 - \phi)]$ and $q_\phi(i) = P(i|H_1) = \frac{1}{2}[1 + (-1)^i \cos(\theta/2 + \phi)]$, depending on which hypothesis is true. For simplicity we take equal priors $\eta_0 = \eta_1 = 1/2$ and study the Bayesian mean number of copies under the same strong error bounds $\epsilon_0 = \epsilon_1 = \epsilon \ll 1$. In the main text we show that the optimal test for a given choice of measurement angle is given by Wald's SPRT strategy, which according to (11) in MT leads to

$$\langle N \rangle = \eta_0 \langle N \rangle_0 + \eta_1 \langle N \rangle_1 \sim -\frac{1}{2} \log\epsilon \left(\frac{1}{D(p_\phi \| q_\phi)} + \frac{1}{D(q_\phi \| p_\phi)} \right). \tag{36}$$

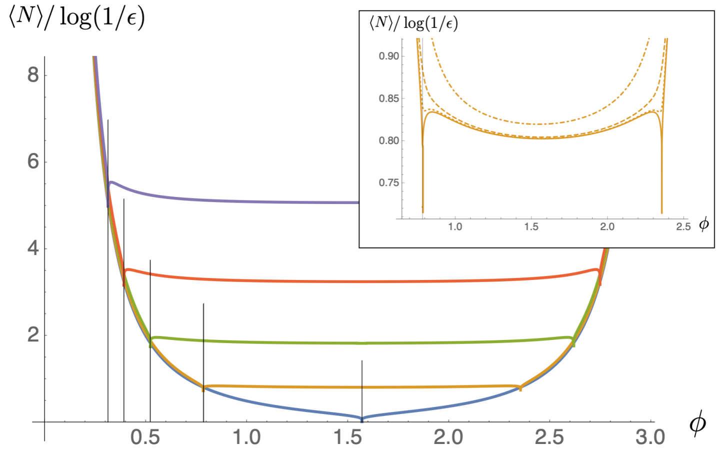


Figure 4. Bayesian mean number of copies $\langle N\rangle$ as a function of the measurement angle ϕ for different pairs of pure states: from bottom to top $\theta = \{\pi, \frac{\pi}{2}, \frac{\pi}{3}, \frac{\pi}{4}, \frac{\pi}{5}\}$. The vertical lines show the corresponding optimal measurement angles $\phi = \theta/2$. The inset shows in more detail the case with $\theta = \frac{\pi}{8}$, including the curves for noisy states with $1 - r = \{10^{-2}, 10^{-3}, 10^{-4}\}$ (dashed lines, from top to bottom).

In Figure 4 we show the Bayesian mean number of copies required to have a guaranteed, asymptotically small bounded error ϵ for all outcomes of the experiment. For pure states ($r = 1$), we observe that the optimal angle is a singular point located at $\phi = \theta/2$, that corresponds to the fully biased measurement for which outcome 1 can only occur under hypothesis $H_1 : p(1|H_0) = 0$ while $p(1|H_1) = \cos^2(\theta/2) > 0$. Hence, H_1 is detected with certainty after a small number of steps $\langle N\rangle_1 \sim \cos^{-2}(\theta/2)$ (independent of the error bound ϵ), and therefore the leading contribution to the expected number of copies when hypothesis H_0 is true is

$$\langle N\rangle_{\text{local}} \sim \frac{\log \epsilon}{2\log(\cos^2 \frac{\theta}{2})} \,. \tag{37}$$

Note that this is exactly half of the number of copies that the most general collective deterministic strategy would require to attain this error bound, since $\epsilon = \frac{1}{2}(1 - \sqrt{1-\cos^{2N}(\theta/2)}) \sim \cos^{2N}(\theta/2)$. This error bound can be attained with local adaptive measurements for finite N [32] and fixed local measurements for asymptotically large N. The result in (37) is in agreement with that derived in [10] for the fully biased strategy, which we have shown to be optimal in the limit of small error bounds (among fixed local measurement strategies). In Figure 4 we also note that a small change around the optimal value $\phi = \theta/2$ produces a very rapid increase of the effective number of copies while the local minimum at $\phi = \pi/2$, which corresponds to the fully unbiased measurement, is much more shallow and hence more robust to a possible measurement misalignment.

We now proceed to study what happens in the presence of noise, when both states are mixed, in particular when $r_0 = r_1 = r$. As shown in the inset of Figure 4, the presence of noise makes the two states more indistinguishable and a higher number of samples are required to meet the error bound. It is also apparent that in presence of noise the fully unbiased measurement, $\phi = \pi/2$, becomes optimal (except for extremely high values of the purity $1 - r \sim 10^{-5}$ for which fully biased performs slightly better). The unbiased measurement is straightforward to compute:

$$\langle N\rangle_{\text{local}} \sim \frac{\log \epsilon}{r \sin\theta \log\left(\frac{1-r\sin\frac{\theta}{2}}{1+r\sin\frac{\theta}{2}}\right)} \,. \tag{38}$$

We can again compare $\langle N\rangle$ reached by the local measurements (38)with the sample size, N, required by the optimal deterministic protocol using a predetermined number of copies to achieve the same error ϵ. When N is large, i.e. ϵ is small, this can be obtained from the asymptotic error exponent in the quantum Chernoff bound [12, 21]. We find

$$N_{\text{Ch}} \sim \frac{\log \epsilon}{\log\left(1 - (1-\sqrt{1-r^2})\sin^2\frac{\theta}{2}\right)} \,. \tag{39}$$

In Figure 2 in MT we compare Eqs. (38) and (39) and observe a reduction of the required number of copies of at least 50% on average if we employ the sequential test instead of the deterministic one. The reduction goes up to 75% if ρ and σ are very mixed.

For illustration purposes, in Figure 5 we show explicitly the results of several runs of a SPRT using unbiased local measurements. We observe how the mean trajectories that the cummulative log-likelihood ratio Z_n follows point upwards or downwards depending on the underlying hypothesis. In this simulation, the state ρ, corresponding to H_0, is identified quicker than σ (the decision boundary b is closer than a), despite being more mixed. A histogram of stopping times under each hypothesis shows us that the distributions of N are well-centered around their empirical mean, with right tails that are slightly longer; this is also apparent on the left figure from the cross-sections of the trajectories with the decision boundaries. Finally, we observe that the mean number of copies increases linearly with $\log\epsilon$ for $\epsilon \ll 1$, as predicted.

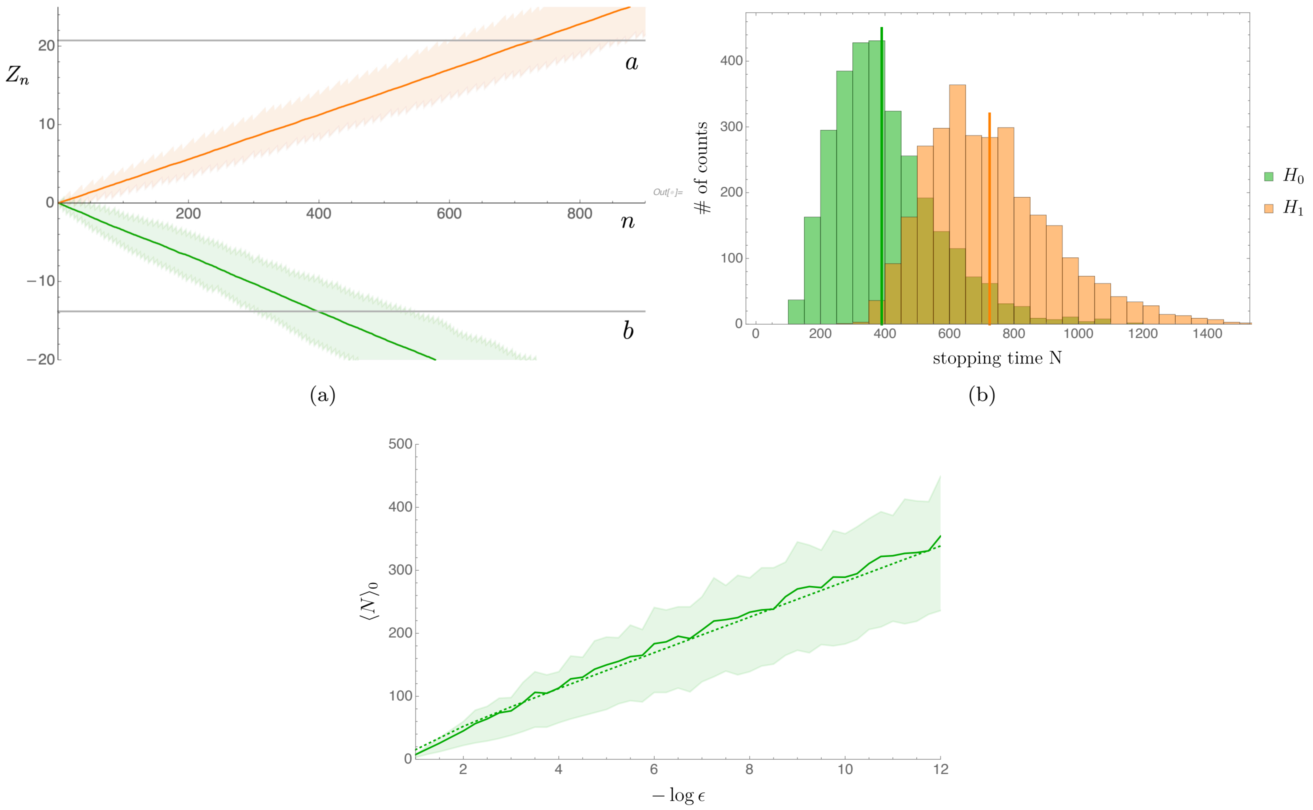
Block-sampling and irrep projection

Here we study the mean number of copies under both hypotheses using a block-sampling strategy where the same collective measurement is repeated on batches of ℓ copies. In particular we will consider a collective measurement for which Hayashi [37] showed that the (classical) relative entropy of the distributions that arise from it, attains the quantum relative entropy when the block length ℓ is large. Denoting by $M = \{M_k\}$ such collective POVM and by p^ℓ_M, q^ℓ_M the probability distributions of the outcomes, i.e., $\{p^\ell_M(k) = \mathrm{tr}(\rho^{\otimes\ell}M_k)\}$ and $\{q^\ell_M(k) = \mathrm{tr}(\sigma^{\otimes\ell}M_k)\}$, in Ref. [37] it is shown that

$$\frac{D\left(p^\ell_M \| q^\ell_M\right)}{\ell} \leq D(\rho\|\sigma) \leq \frac{D\left(p^\ell_M \| q^\ell_M\right)}{\ell} + (d-1)\frac{\log(\ell+1)}{\ell} \,, \tag{40}$$

where d is the dimension of the underlying Hilbert space, from where

$$\frac{D\left(p^\ell_M \| q^\ell_M\right)}{\ell} \to D(\rho\|\sigma) \quad \text{as } \ell \to \infty \,. \tag{41}$$



(a)

(b)

(c)

Figure 5. (a) Behaviour of the likelihood ratio Z_n using unbiased local measurements as a function of the step n, for 3000 realizations of the hypothesis test with $\theta = \frac{\pi}{10}$, $r_0 = 0.7$, $r_1 = 0.9$, $\epsilon_0 = 10^{-6}$, $\epsilon_1 = 10^{-9}$, under hypotheses H_0 (green) and H_1 (orange). The solid lines show the mean trajectory, and the shaded areas correspond to the 2nd and 3rd quartiles (25% above and below the mean). The decision boundaries a and b are also plotted. (b) Histogram of stopping times under each hypothesis. The vertical lines mark the mean of each distribution. (c) The mean number of copies $\langle N\rangle_0$ as a function of $\log\epsilon$, for the same state parameters as in (a) and equal error bounds $\epsilon_0 = \epsilon_1 = \epsilon$. The solid line represents the mean trajectory of 500 runs, the shaded area shows the 2nd and 3rd quartiles, and the dashed line is the analytic expression $\langle Z\rangle_0/D(p_\phi\|q_\phi)$, where recall that $\langle Z\rangle_0 \sim -\log\epsilon$ for $\epsilon \ll 1$.

Quite remarkably, the measurement M in Eq. (40) depends solely on state σ.

As explained in the MT such a strategy allows one to attain the lower bound for one of the hypotheses, say $\langle N\rangle_0 \sim -\log\epsilon_0(1 - A^{-1})/D(\rho\|\sigma)$ for the strong error bounds [cf. Eq. (15) in MT], or

$$\langle N\rangle_0 \sim \frac{\log\beta(1-\alpha)}{D(\rho\|\sigma)} \tag{42}$$

for the asymmetric setting.

In what follows we compute the (sub-optimal) performance of this very same measurement under the other hypothesis, i.e., $\langle N\rangle_1$.

For qubit systems, the POVM that achieves the quantum relative entropy [37] corresponds to the simultaneous measurement of the total angular momentum J^2 (eigenspaces labeled by j) and its component along the axis J_z (eigenspaces labeled by m), where $\hat{z}$ is picked to be the axis along which the state σ points, i.e., $\sigma = 1/2(\mathbb{1} + r_1\sigma_z)$. The quantum number j labels the $SU(2)$ irreducible representations (irreps), and since it is invariant under the action of any rigid rotation $U^{\otimes n}$ it will only provide information about the spectra of ρ or σ —which we denote $\lambda_i^\pm = \frac{1}{2}(1 \pm r_i)$ for $i = 0, 1$ respectively. The second measurement J_z is clearly not $SU(2)$ invariant and provides information about relative angle between both hypotheses, and additional information on their spectrum.

Due to the permutational invariance of the ℓ copies it is possible to write the states in a block-diagonal form in terms of the $\{j, m\}$ quantum numbers (see e.g. [33]):

$$\sigma^{\otimes \ell} = \sum_{j=j_{\min}}^{\ell/2} q_j \sigma_j \otimes \frac{\mathbb{1}_j}{\nu_j} \quad \text{with} \quad \sigma_j = \sum_{m=-j}^{j} q(m|j)\, |j,m\rangle\langle j,m|\,, \tag{43}$$

where $\mathbb{1}_j$ are projectors over the subspaces of dimension $\nu_j = \binom{\ell}{\ell/2-j}\frac{2j+1}{\ell/2+j+1}$ that host the irreps of the permutation group (i.e. multiplicity space of spin j), $j_{\min} = 0$ for ℓ even ($j_{\min} = 1/2$ for ℓ odd) and

$$q_j = \left(\frac{1-r_1^2}{4}\right)^{\ell/2} \nu_j Z_j\,, \tag{44}$$

$$q(m|j) = \frac{1}{Z_j} R_1^m \quad \text{with} \quad Z_j = \frac{R_1^j - R_1^{-j}}{R_1 - 1} \quad \text{and} \quad R_1 = \frac{1+r_1}{1-r_1} > 1 \tag{45}$$

are normalized probability distributions.

Under hypothesis H_0 the state has exactly the same structure except for a global rotation around the $\hat{z}$ axis by an angle θ,

$$\rho^{\otimes \ell} = \sum_{j=j_{\min}}^{\ell/2} p_j \rho_j \otimes \frac{\mathbb{1}_j}{\nu_j} \quad \text{with} \quad \rho_j = \sum_{m=-j}^{j} p(m|j)\; U_\theta\, |j,m\rangle\langle j,m|\, U_\theta^\dagger, \tag{46}$$

where p_j and $p(m|j)$ take the form of (44) and (45) replacing r_1 and R_1 by r_0 and R_0.

The outcomes of the J^2 and J_z measurements lead to probability distributions

$$p(j,m) = p_j\, \tilde{p}(m|j) \quad \text{with} \quad \tilde{p}(m|j) = \sum_{m'=-j}^{j} p(m'|j) |\langle j,m|\, U_\theta\, |j,m'\rangle|^2\,, \tag{47}$$

$$q(j,m) = q_j\, q(m|j)\,, \tag{48}$$

whose relative entropy can be written as

$$D(q^\ell \| p^\ell) = \sum_{j=j_{\min}}^{\ell/2} q_j \sum_{m=-j}^{j} q(m|j) \log \frac{q_j\, q(m|j)}{p(j,m)} \sim \log \frac{q_{j^*}\, q(j^*|j^*)}{p(j^*, j^*)} \tag{49}$$

where we have used the fact that for $\ell \gg 1$, q_j is strongly peaked at $j^* = r_1\ell/2$ and $q(m|j^*)$ decays exponentially, and hence it is peaked at $m = j^*$. In addition we note that

$$\begin{aligned} p(j, m=j) &= p_j \sum_{m'=-j}^{j} p(m'|j) |\langle j,j|\, U_\theta\, |j,m'\rangle|^2 \\ &= \langle j,j| \left(p_j \sum_{m'=-j}^{j} p(m'|j) U_\theta\, |j,m'\rangle\langle j,m'|\, U_\theta^\dagger \right) |j,j\rangle = \left(\frac{1-r_0^2}{4}\right)^{l/2-j} \nu_j\, \langle j,j|\, \rho^{\otimes 2j}\, |j,j\rangle \\ &= \left(\frac{1-r_0^2}{4}\right)^{l/2-j} \nu_j\, \langle \uparrow|\, \rho\, |\uparrow\rangle^{2j} = \left(\frac{1-r_0^2}{4}\right)^{l/2-j} \nu_j \left(\frac{1+r_0\cos\theta}{2}\right)^{2j}, \end{aligned} \tag{50}$$

where we used the general decomposition of (46), and $|\uparrow\rangle$ is short-hand notation for $|j = 1/2, m = 1/2\rangle$. Inserting this expression in (49) and using the definitions in (44) and (45) we finally arrive at

$$\begin{aligned} D_{M_\sigma}(\sigma\|\rho) := \lim_{\ell\to\infty} \frac{1}{\ell} D(q^\ell\|p^\ell) &= \frac{1}{2}(1-r_1) \log \frac{1-r_1^2}{1-r_0^2} + r_1 \log \frac{1+r_1}{1+r_0\cos\theta} \\ &= D(\lambda_1\|\lambda_0) + r_1 \log \frac{1+r_0}{1+r_0\cos\theta}\,, \end{aligned} \tag{51}$$

where the symbol M_σ recalls that we have chosen M as a measurement over the eigenbasis of σ, which maximizes the relative entropy $D(q^l\|p^l)$, and

$$D(\lambda_1\|\lambda_0) = \frac{1}{2}(1+r_1)\log\frac{1+r_1}{1+r_0} + \frac{1}{2}(1-r_1)\log\frac{1-r_1}{1-r_0} \tag{52}$$

is the relative entropy between the spectra of σ and ρ. Hence, the second term in (51) can be associated to the distinguishability caused by the different orientation (non-commutativity) of the states.

On the other hand, from (41) it follows

$$\begin{aligned} D_{M_\sigma}(\rho\|\sigma) &:= \lim_{\ell\to\infty}\frac{1}{\ell}D(p^\ell\|q^\ell) = D(\rho\|\sigma) \\ &= \frac{1}{2}(1+r_0)\log\frac{1+r_0}{1+r_1} + \frac{1}{2}(1-r_0)\log\frac{1-r_0}{1-r_1} + r_0\log\frac{1+r_0}{1-r_1}\sin^2\frac{\theta}{2} \\ &= D(\lambda_0\|\lambda_1) + r_0\log\frac{1+r_0}{1-r_1}\sin^2\frac{\theta}{2}\,. \end{aligned} \tag{53}$$

From the above results we conclude that applying the measurement that reaches the ultimate bound for one hypothesis

$$\langle N\rangle_0 = -\frac{\log\beta}{D(\rho\|\sigma)} \tag{54}$$

will result in a sub-optimal value

$$\langle N\rangle_1 \sim -\frac{\log\alpha}{D_{M_\sigma}(\sigma\|\rho)} \tag{55}$$

for the other hypothesis, with $D_{M_\sigma}(\sigma\|\rho)$ given in (51).

We observe that, as expected, when the states commute we can reach the ultimate bound for both $\langle N\rangle_0, \langle N\rangle_1$. We also note that, when ρ is pure, one can also preserve asymptotic optimality for both means, since when $\lambda_0^\pm \in \{1,0\}$, $D_{M_\sigma}(\sigma\|\rho)$ diverges and the leading contribution in $\langle N\rangle_1$ vanishes, while $\langle N_0\rangle$ reaches the optimal value. These results hold for the block-sampling strategy that uses blocks of large length $l \gg 1$, so one needs to find other ways to compute the finite $O(1)$ contribution to $\langle N\rangle_1$, as we shall show next.

We have already shown [see (18) in MT] that when *both* states are pure we can detect both hypotheses with a finite mean number of copies. If only one of the states is pure, say ρ, it is easy to notice that the J^2 measurement alone guarantees a constant value for $\langle N\rangle_1$: $\rho^{\otimes n}$ lies in the fully symmetric space (with $j = n/2$) and therefore any measurement outcome $j < n/2$ will unambiguously identify σ. The above block-sampling might have an important overhead when ℓ is large. A way of reducing this overhead can be devised by leveraging the fact that the measurement of J^2 on n copies commutes with the measurement of J^2 on $n+1$ for all $n \geq 1$: starting at $n = 2$, we measure J^2 sequentially on all available copies until we get an outcome $j < n/2$, at which moment we stop and accept H_1. Note that each step of this sequence uses the already measured copies, increasing the number of jointly-measured systems by one. Since the probability of not detecting σ at step n (continue measuring) is given by $T_1^n = P(j = n/2) = \left(\frac{1+r_1\cos\theta}{2}\right)^n =: q_+^n$, we can write

$$\langle N\rangle_1 = \sum_{n=0}^{\infty} T_1^n = \sum_{n=0}^{\infty} q_+^n = \frac{1}{1-q_+} = \frac{2}{1-r_1\cos\theta}\,. \tag{56}$$

Note, however, that measuring J^2 sequentially is on its own not enough to reach the optimal mean number of copies also under hypothesis H_0, $\langle N\rangle_0$. For this reason, after every batch of ℓ copies, $\ell \gg \langle N\rangle_1$ ($\ell = o(\log 1/\epsilon)$), we interrupt the sequence of J^2 measurements with a measurement of J_z on the last batch of ℓ copies (and then continue again with the J^2 sequential measurement). The measurement statistics obtained by this procedure mimicks the block-sampling method described above and hence we are guaranteed to converge to (54).

Alternatively, in order to attain (56) one can directly measure the system in the basis that diagonalizes $\rho = |\lambda_0^+\rangle\langle\lambda_0^+|$, so that an $|\lambda_0^-\rangle$ outcome unambiguously detects σ with probability $q_- = 1 - q_+$. It is immediate to check that the sequential application of this measurement also leads to (56). Again after having measured a sufficiently large number of copies $\ell = o(\log 1/\epsilon)$ one can adopt the block-sampling strategy in order to achieve the bound (54).

Ultimate limit for $N_{\rm wc}$. Attainability regions

In this section we study the achievability of the lower bound on the worst-case mean number of copies,

$$N_{\rm wc} := \max\{\langle N\rangle_0, \langle N\rangle_1\} \geq \max\left\{-\frac{\log\epsilon}{D(\rho\|\sigma)}, -\frac{\log\epsilon}{D(\sigma\|\rho)}\right\}, \tag{57}$$

where for simplicity we assume that both error bounds are equal, i.e., $\epsilon_0 = \epsilon_1 = \epsilon \ll 1$.

In the previous section we have shown that the block-sampling with a given POVM M_σ can reach the optimal value under H_0, but it does so at the expense of attaining a sub-optimal value under H_1. Making use of the results of (51) and (53) one can show that there are pairs of states $\{\rho, \sigma\} \in \mathcal{R}$ for which either

$$D(\rho\|\sigma) \leq D_{M_\sigma}(\sigma\|\rho) \leq D(\sigma\|\rho) \quad \text{or} \quad D(\sigma\|\rho) \leq D_{M_\rho}(\rho\|\sigma) \leq D(\rho\|\sigma)\,. \tag{58}$$

When this happens we can assert that the bound in (57) is attainable, since the worst-case value is attained, i.e.,

$$N_{\rm wc} \sim \max\left\{-\frac{\log\epsilon}{D(\rho\|\sigma)}, -\frac{\log\epsilon}{D(\sigma\|\rho)}\right\}, \tag{59}$$

Figure 6 shows some representative regions where (58) is fulfilled, and (59) holds. We observe that for small relative angles θ almost all states attain the ultimate bound, except for a region around the pairs of equal purity. It is easy to check that for states with $r_0 = r_1$, $D(\rho\|\sigma) = D(\sigma\|\rho)$, and therefore (58) cannot be satisfied, independently of the relative angle θ. When $\theta = \pi/2$, i.e., when the pair of states exhibits more non-classicality, only pairs comprised by a highly pure and a highly mixed state can attain the bound.

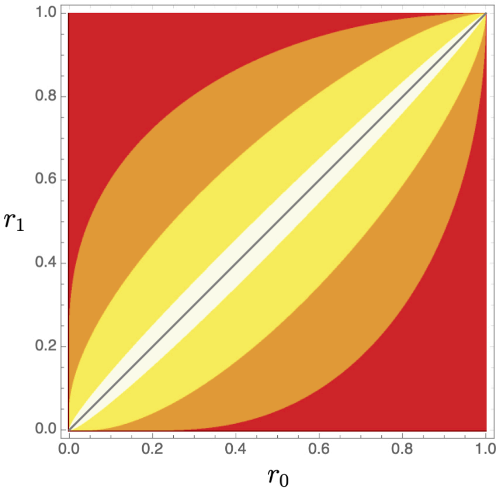


Figure 6. Regions $\mathcal{R}$ of purities (r_0, r_1) for different relative angles θ, where the ultimate bound for $N_{\rm wc}$ (59) is achievable. The red region corresponds to $\theta = \pi/2$, orange and red to $\theta = \pi/10$, yellow, orange and red to $\theta = \pi/100$, and all admissible values of (r_0, r_1) are achievable for commuting states ($\theta = 0$).

.

THE OVERHEAD FOR ARBITRARY DIMENSIONS

In this section we would like to explore how conditions (58) look when states ρ and σ have arbitrary dimension $d > 2$. In this case, exactly quantifying $D_{M_\rho}(\rho\|\sigma)$ [recall that we denote by M_ρ the block-sampling measurement on ℓ copies that attains $D(\sigma\|\rho)$ when $\ell \to \infty$] is more involved. Here instead we provide a general lower bound for the deviation of $D_{M_\rho}(\rho\|\sigma)$ from its maximum value $D(\rho\|\sigma)$. We follow closely Ref. [37].

First, consider the following operation on a state ρ for a given a projective measurement $E = \{E_j\}$ (i.e., $E_j^2 = E_j$ and $E_jE_k = \delta_{jk}E_j$),

$$\varepsilon_E(\rho) := \sum_j E_j\rho E_j\,. \tag{60}$$

When E commutes with states ρ and σ we have

$$\varepsilon_E(\rho) = \rho\,, \quad \varepsilon_E(\sigma) = \sigma\,. \tag{61}$$

Then, consider a projective measurement $F(\rho) = \{F_k\}$ that consists of rank-one projectors in the eigenbasis of ρ, i.e., a measurement of the spectrum of ρ. Note that we have $\varepsilon_F(\rho) = \rho$, but $\varepsilon_F(\sigma) \neq \sigma$ for a generic state σ that does not commute with ρ. Note also that E (which commutes with ρ and σ) is a coarse-grained measurement of $F(\rho)$; we can then say that $F(\rho)$ is stronger than E. In [37] this fact is denoted by $F(\rho) \geq E$. We then have the following lemma:

Lemma 2. *Let ρ and σ be states and let $F(\rho) \geq E$. The quantum relative entropy between ρ and σ can be expressed as*

$$D(\rho\|\sigma) = D(\varepsilon_F(\rho)\|\varepsilon_F(\sigma)) + \operatorname{Tr}\rho(\log\varepsilon_F(\sigma) - \log\sigma)\,. \tag{62}$$

Proof. Recalling that $\varepsilon_F(\rho) = \rho$, we have $\operatorname{Tr}\varepsilon_F(\rho)\log\varepsilon_F(\sigma) = \operatorname{Tr}\rho\log\varepsilon_F(\sigma)$, thus

$$D(\varepsilon_F(\rho)\|\varepsilon_F(\sigma)) - D(\rho\|\sigma) = \operatorname{Tr}\varepsilon_F(\rho)(\log\varepsilon_F(\rho) - \log\varepsilon_F(\sigma)) - \operatorname{Tr}\rho(\log\rho - \log\sigma) = \operatorname{Tr}\rho(\log\sigma - \log\varepsilon_F(\sigma))\,. \tag{63}$$

Hence, it follows that

$$D(\varepsilon_F(\rho)\|\varepsilon_F(\sigma)) = D(\rho\|\sigma) - \operatorname{Tr}\rho(\log\varepsilon_F(\sigma) - \log\sigma)\,. \tag{64}$$

□

We also need the following lemma:

Lemma 3. *For a given projective measurement E such that $E \leq F$, if E commutes with σ and ρ we have that*

$$\operatorname{Tr}\rho(\log\varepsilon_F(\sigma) - \log\sigma) \leq \sup_i\{\operatorname{Tr}\rho_i(\log\varepsilon_F(\sigma_i) - \log\sigma_i)\}\,, \tag{65}$$

where we define $\rho_i := \frac{1}{a_i}E_i\rho E_i$, $\sigma_i := \frac{1}{b_i}E_i\sigma E_i$, $a_i := \operatorname{Tr}E_i\rho E_i$, and $b_i := \operatorname{Tr}E_i\sigma E_i$.

Proof. Starting from the left side of inequality (65), the following steps hold:

$$\operatorname{Tr}\rho(\log\varepsilon_F(\sigma) - \log\sigma) = \operatorname{Tr}[\sum_i E_i\rho(\log\varepsilon_F(\sigma) - \log\sigma)] = \operatorname{Tr}[\sum_i E_i\rho E_i(E_i\log\varepsilon_F(\sigma)E_i - E_i\log\sigma E_i)] \tag{66}$$

$$= \operatorname{Tr}[\sum_i a_i\rho_i(\log\varepsilon_F(\sigma_i) - \log\sigma_i)] \leq \sup_i\{\operatorname{Tr}\rho_i(\log\varepsilon_F(\sigma_i) - \log\sigma_i)\} =: \omega(\sigma). \tag{67}$$

□

We are now ready to derive a bound on $D_{M_\rho}(\rho\|\sigma)$ with the following theorem:

Theorem 2. *Let us define the projective measurement $M_\rho = F(\rho^{\otimes\ell}) \times E^\ell$ acting on ℓ copies, where E^ℓ, applied first, is a measurement that projects onto the irreps of $SU(d)^{\otimes\ell}$. Then, $F(\rho^{\otimes\ell})$ is a spectral measurement of $\rho^{\otimes\ell}$, i.e., a projective measurement on the basis that diagonalizes ρ. For this measurment, we have*

$$D(\rho\|\sigma) - \frac{\omega(\sigma)}{\ell} \leq \frac{1}{\ell}D_{M_\rho}(\rho^{\otimes\ell}\|\sigma^{\otimes\ell}) \leq D(\rho\|\sigma)\,. \tag{68}$$

Proof. We simply use Lemma 2 and Lemma 3 with the change $\rho \to \rho^{\otimes\ell}$ and $\sigma \to \sigma^{\otimes\ell}$, and we recall the property of the quantum relative entropy $D(\rho^{\otimes\ell}\|\sigma^{\otimes\ell}) = \ell D(\rho\|\sigma)$. The result follows from applying Lemma 2 to all the terms in Lemma 3. □

A very generous bound can be obtained by dropping the negative term $\log\varepsilon_F(\sigma_i)$ in $\omega(\sigma)$ [34]:

$$\omega(\sigma) \leq \sup_i -\operatorname{Tr}[\rho_i(\log\sigma_i)] \leq \max_i -\log[\lambda_{\min}(\sigma_i)]\,. \tag{69}$$

ZERO-ERROR PROTOCOL FOR PURE STATES

As we have seen in the MT, as the error ϵ goes to 0, the average number of copies goes to infinity. However, for pure states $\{\rho = |\psi_0\rangle\langle\psi_0| \, , \sigma = |\psi_1\rangle\langle\psi_1|\}$ there are sequential strategies with local measurements that give a strictly zero error with a finite average number of samples. Here we detail the protocol already mentioned in the MT and prove its optimality for equal priors for the Bayesian mean and worst-case number of copies.

To this end, consider a sequence of fixed unambiguous measurements on each copy with inconclusive probabilities c_ν if the given state is $|\psi_\nu\rangle$, $\nu = 0, 1$. We notice that these probabilities satisfy the 'uncertainty' relation $c_0 c_1 \geq s^2$, where $s = |\langle\psi_0|\psi_1\rangle|$ [**?**]. The protocol stops only if one of the states is identified with no error. Hence, at each step n there are only two possibilities: continue, with conditional probability (after having arrived at step n) c_ν, or stop, with conditional probability $1 - c_\nu$. The probability of exactly stopping at step n is $P_\nu^n = c_\nu^{n-1}(1 - c_\nu)$. Then, the average number of copies required to get a zero-error outcome is

$$\langle N\rangle_\nu = \sum_{n=1}^{\infty} n c_\nu^{n-1}(1 - c_\nu) = \sum_{n=0}^{\infty} c_\nu^n = \frac{1}{1 - c_\nu} \, . \tag{70}$$

Notice that both means are finite if one performs an unambiguous measurement with $c_0 < 1$ and $c_1 < 1$, which is allowed by the relation $c_0 c_1 \geq s^2$.

In the case of equal priors, we now show that the symmetric choice $c_0 = c_1 = s$ gives the optimal Bayesian mean $\langle N\rangle = (\langle N\rangle_0 + \langle N\rangle_1)/2$. We observe that the inconclusive probability attained by the optimal global measurement on n copies of $|\psi_\nu\rangle$, T_ν^n, cannot be beaten by any local strategy, hence, using (70) we have

$$\langle N\rangle = \frac{1}{1 - s} = \sum_{n=0}^{\infty} s^n \geq \sum_{n=0}^{\infty} \frac{1}{2} \left(T_0^n + T_1^n\right) =: \langle N^*\rangle \, . \tag{71}$$

Analogously to the derivation of (15) in MT, the r.h.s. of (71) corresponds to a relaxation of the original problem in which we have independently optimized each term in the sum, considering the action of optimal n-copy unambiguous measurements for each n, hence $\langle N^*\rangle$ is a lower bound to the most general protocol. Since n-copy pure states are simply pure states of larger dimension, we have $|\langle\psi_0^{\otimes n}|\psi_1^{\otimes n}\rangle| = s^n$, and the equivalent relation $T_0^n T_1^n \geq s^{2n}$ holds for global strategies. Then, the symmetric choice $T_0^n = T_1^n = s^n$ minimizes each summand in (71), and we obtain $\langle N^*\rangle = \sum_{n=0}^{\infty} s^n = (1 - s)^{-1} = \langle N\rangle$.

Finally, we also note that the symmetric choice also optimizes the figure of merit given by the worst-case number of copies $N_{\mathrm{wc}} = \max\{\langle N\rangle_0, \langle N\rangle_1\}$. Because of the relation $T_0^n T_1^n \geq s^{2n}$, if $T_0^n > s^n$ then $T_1^n < s^n$ and $N_{\mathrm{wc}} = \langle N\rangle_0 > 1/(1 - s)$. The same argument applies if $T_1^n > s^n$, hence it follows that the optimal measurement has $T_0^n = T_1^n = s^n$ and $N_{\mathrm{wc}} = 1/(1 - s)$.

Online identification of symmetric pure states

Gael Sentís, Esteban Martínez-Vargas, and Ramon Muñoz-Tapia

Física Teòrica: Informació i Fenòmens Quàntics, Departament de Física, Universitat Autònoma de Barcelona, 08193 Bellatera (Barcelona) Spain

We consider online strategies for discriminating between symmetric pure states with zero error when n copies of the states are provided. Optimized online strategies involve local, possibly adaptive measurements on each copy and are optimal at each step, which makes them horizon independent, hence robust in front of particle losses or an abrupt termination of the discrimination process. We first review previous results on binary minimum and zero error discrimination with local measurements that achieve the maximum success probability set by optimizing over global measurements, highlighting their online features. We then extend these results to the case of zero error identification of three symmetric states with constant overlap. We provide optimal online schemes that attain global performance for any n if the state overlaps are positive, and for odd n if overlaps have a negative value. For arbitrary complex overlaps, we show compelling evidence that online schemes fail to reach optimal global performance. The online schemes that we describe only require to store the last outcome obtained in a classical memory, and adaptiveness of the measurements reduce to at most two changes, regardless of the value of n.

1 Introduction

The task of discriminating among non-orthogonal quantum states [1–4] underlies many prominent applications of quantum information sciences. A basic primitive in quantum communication [5, 6], it also has fundamental implications in quantum key distribution [7–10], in the design of quantum algorithms [11], and in foundations of quantum theory [12–15]. Due to the no-cloning theorem [16], it is not possible to perfectly and deterministically identify which is the state of a given quantum system out of a known finite set of possible ones, unless these are mutually orthogonal. If copies of identically prepared systems in the same unknown state are provided, we may extract more information and increase our chances of identifying it correctly. However, in order to take full advantage of these extra resources, one generally needs to apply a collective quantum measurement on all the provided systems, which requires performing entangling operations and keeping all systems to be measured in a coherent quantum memory. Such collective measurement, once optimized, is guaranteed to yield the best performance in the discrimination task allowed by quantum mechanics, but the necessary requirements to implement it are hardly met in practical situations.

More experimentally viable (albeit generally sub-optimal) schemes are those that only involve local measurements on each system, thus removing the need of quantum memories and quantum correlations in the measurement apparatus. Such schemes fall under the paradigm of *local operations and classical communication* (LOCC). The question of when can LOCC schemes attain optimal (global) performance in a state discrimination task has been considered in the literature under different angles [17–26].

The motivation behind this topic is not only practical, but also foundational: a performance gap between optimal local and global schemes in discriminating separable states is a signature of the phenomenon called "quantum nonlocality without entanglement" [27], which has implications in the capacities of quantum channels [28], in the ability to hide information to classical observers [29], and in distinguishing quantum theory from other generalized probabilistic theories [30].

In this paper, we take a step further from LOCC and consider *online* strategies for state discrimination, that is, feed-forward local measurement schemes that do not depend on knowing beforehand the number of copies of the states available (the horizon), and are optimal at each step of the process. In contrast to horizon-dependent LOCC, online schemes do not loose optimality if some of the systems are lost or if the procedure stops at an earlier time than planned, thus making them the most desirable schemes for robust realistic implementations. This sort of data processing can be regarded as a self-learning process [31], and it is the natural procedure in sequential analysis algorithms [32, 33].

When trying to discriminate between two states, it is known that online strategies attain optimal global performance, regardless of whether one considers minimum error discrimination [22] or unambiguous identification [19, 21], the two usual approaches to state discrimination. Discriminating more than two hypotheses is a much harder problem: optimal protocols are only known for certain special cases [4, 34–39], and results on local distinguishability are even more scarce [24, 25, 40–42]. Here we tackle the problem of unambiguous (zero error) identification of three symmetric pure quantum states with constant (but arbitrary) overlap c when n copies are provided, characterizing for which parameter ranges do online schemes attain global performance. We first rederive the case of binary discrimination, highlighting the online features of the optimal local protocols, and then we extend our formalism to three hypotheses. Specifically, we show that online strategies based on Bayesian updating are globally optimal for any n if $c \geq 0$, and for odd n if $c < 0$. Our analysis straightforwardly extends to the case of tensor products of n trines with constant but different overlaps. Importantly, the choice of each measurement in these strategies depends only on the last outcome obtained, thus greatly limiting the size of the classical memory required. For complex-valued overlaps, we provide strong evidence of a gap between online and global strategies.

The paper is organized as follows. In Sections 2 and 3 we review online binary minimum-error discrimination and unambiguous identification, respectively, and extend these results to non-identical copies of the states. This serves us to set notation and techniques that we use later. Section 4 contains our main results for three symmetric states, and we finish with some conclusions of our analysis.

2 Two-state minimum error discrimination

Here we briefly review binary discrimination for minimum error [22, 43] and its extension to the multi-hypothesis case.

Any two pure states can be written w.l.o.g. as

$$\left|\psi_{0/1}\right\rangle = \sqrt{\frac{1+c}{2}}\,|0\rangle \pm \sqrt{\frac{1-c}{2}}\,|1\rangle\,, \qquad (1)$$

where $|0\rangle$ and $|1\rangle$ is a basis of the space spanned by $\{|\psi_0\rangle\,, |\psi_1\rangle\}$ and $c = |\langle\psi_0|\psi_1\rangle|$. For later reference it is convenient to view this parametrization as $|\psi_0\rangle = \xi_0\,|0\rangle + \xi_1\,|1\rangle$, where $|0\rangle$ and $|1\rangle$ are the eigenstates of the unitary operation $U = |0\rangle\!\langle 0| + e^{\frac{2i\pi}{2}}\,|1\rangle\!\langle 1| = |0\rangle\!\langle 0| - |1\rangle\!\langle 1|$, $|\psi_1\rangle = U\,|\psi_0\rangle$, and $\xi_i = \sqrt{\lambda_i(G)/2}$, $i = 0, 1$, where $\lambda_i(G)$ are the eigenvalues of the Gram matrix whose elements are $g_{ij} = \langle\psi_i|\psi_j\rangle$. With this parametrization the operator $\Omega = \sum_k |\psi_k\rangle\!\langle\psi_k|$, which plays a key role in the extension of larger sets of symmetric states (Sec 4), is diagonal, i.e., $\Omega = 2\,\mathrm{diag}\{|\xi_0|^2, |\xi_1|^2\}$.

We assume that the two states can occur with arbitrary a priori probabilities η_0 and η_1, respectively. The aim is to minimize the average error probability $P_e = \eta_0 p(1|\psi_0) + \eta_1 p(0|\psi_1)$, or equivalently maximize the success probability $P_s = \eta_0 p(0|\psi_0) + \eta_1 p(1|\psi_1)$, where $p(r|\psi_i)$, $r = 0, 1$, is the probability of making the guess $|\psi_r\rangle$ when the state was $|\psi_i\rangle$. These conditional probabilities are determined by the measurement $\mathcal{M}$ performed on the system, which is described mathematically as a positive operator-valued measure (POVM). Here the POVM has only two elements $\mathcal{M} = \{E_0, E_1\}$, with $E_r \geq 0$ and $E_0 + E_1 = \mathbb{1}$. The Born rule dictates that $p(r|\psi_i) = \mathrm{tr}\,[E_r\,|\psi_i\rangle\!\langle\psi_i|]$. The optimal success probability has the well known expression [5]

$$P_s = \frac{1+\sqrt{1-4\eta_0\eta_1 c^2}}{2}. \tag{2}$$

It is also well known that this success probability is attained with a POVM with elements that are the projectors on the positive and negative spectrum of the operator $\Gamma = \eta_0\,|\psi_0\rangle\!\langle\psi_0| - \eta_1\,|\psi_1\rangle\!\langle\psi_1|$, the so-called Helstrom measurement [5].

The generalization to the multi-copy case is straightforward. The optimal value of the success probability $P_s(n) = \eta_0 p(0|\psi_0{}^{\otimes n}) + \eta_1 p(1|\psi_1{}^{\otimes n})$ is obtained by simply replacing $c \to c^n$ in Eq. (2), i.e.,

$$P_s^G(n) = \frac{1+\sqrt{1-4\eta_0\eta_1 c^{2n}}}{2}, \tag{3}$$

where the superscript G stands for global. The global measurement attaining this bound acts jointly on the n copies, hence a quantum memory to store the systems is required. Note also that it may involve entangling operations between the systems.

Let us now succinctly show that there exists a scheme where each system is measured locally and still achieves the optimal success probability given by Eq. (3). It consists of a sequence of Helstrom measurements on each system where prior probabilities are updated at each step k according to the Bayes rule

$$\begin{aligned}\eta_i^{(k)}(r_k) =: \eta_i^{(k)} &= p(\psi_i|r_k)\\ &= \frac{\eta_i^{(k-1)} p(r_k|\ \psi_i)}{\eta_0^{(k-1)} p(r_k|\psi_0) + \eta_1^{(k-1)} p(r_k|\psi_1)}.\end{aligned} \tag{4}$$

Here $r_k = 0, 1$ is the outcome value of the k'th measurement and we have streamlined the notation when no confusion arises.

The crucial property is that the Helstrom measurements yield the relation $\eta_0^{(k)}\eta_1^{(k)} = \eta_0^{(k-1)}\eta_1^{(k-1)}c^2$ for any value of the outcome r_k (see [22]), and thus $\eta_0^{(k)}\eta_1^{(k)} = \eta_0\eta_1 c^{2k}$ that, once inserted in Eq. (2) for $k = n-1$, precisely gives Eq. (3).

The Bayes rule (4) can be seen as a learning process that updates our belief on the occurrence of each state. Observe that the optimal value of the success probability is obtained at each step. This is an online procedure as the knowledge of the total number of systems that are available for measurement is not required, in contrast, e.g., to dynamic programming problems where the knowledge of the horizon is needed to carry out an optimization in reverse [44]. Furthermore, measurements in this local scheme only depend on the previous outcome (as opposed to the whole sequence of previous outcomes), thus the size of the required classical memory is minimal.

Interestingly, the same Bayesian updating local protocol turns out to be optimal in the non-*i.i.d.* case, i.e., for two arbitrary multipartite product states $|\Phi\rangle = |\phi_1\rangle \otimes |\phi_2\rangle \otimes \cdots \otimes |\phi_n\rangle$ and $|\Psi\rangle = |\psi_1\rangle \otimes |\psi_2\rangle \otimes \cdots \otimes |\psi_n\rangle$ with arbitrary priors η_Φ and η_Ψ, respectively [45]. The overlap in this case is $C = |\langle\Phi|\Psi\rangle| = c_1 c_2 \cdots c_n$, with $c_k = |\langle\phi_k|\psi_k\rangle|$. We proceed as in the *i.i.d.* case, that is, we perform a series of local Helstrom measurements with sequentially updated priors and get $\eta_\Phi^k \eta_\Psi^k = \eta_0\eta_1 c_1^2 c_2^2 \cdots c_k^2$. The success probability then reads

$$P_s = \frac{1+\sqrt{1-4\eta_\Phi\eta_\Psi c_1^2 c_2^2\cdots c_n^2}}{2}, \tag{5}$$

i.e., the optimal success probability, Eq. (2) with $c \to C$.

Going beyond the binary case is much more involved as there are no closed expressions for the success probability for arbitrary priors. Optimal solutions (single or multi-copy) are known only in very few cases, that essentially correspond to symmetric instances (see e.g. [4]). Notice that in any local protocol, even with symmetric sources, the updating rule will necessarily bias the priors and hence render the problem intractable analytically. One can nevertheless carry out a numerical study. It has been recently shown numerically that local measurements supplemented with the Bayesian updating rule do not yield the optimal global success probability in the minimum error approach already in the case of three symmetric states [46] (see also [47] for an analysis with symmetric coherent states). However, it remains an open question whether this feature also holds for zero-error protocols, which we discuss next.

3 Two-state zero-error identification

We now turn our attention to protocols that identify a state without errors at the expense of having inconclusive outcomes, a task also known as unambiguous discrimination [48, 49]. Here we show that in the binary case there is also a local online procedure that gives the maximum success probability provided by the most general global POVM acting on all systems.

The zero-error POVM in principle has three elements: F_0 and F_1, that unambiguously detect $|\psi_0\rangle$ and $|\psi_1\rangle$, respectively, and F_I, which we associate to an inconclusive outcome. In order to achieve optimality the success probability $P_s^u = \eta_0 P(0|\psi_0) + \eta_1 P(1|\psi_1) =: \eta_0 p_0 + \eta_1 p_1$ is maximized or, equivalently, the inconclusive probability $Q = \eta_0 P(I|\psi_0) + \eta_1 P(I|\psi_1) =: \eta_0 q_0 + \eta_1 q_1$ is minimized while keeping the condition that no errors are committed, i.e $P(1|\psi_0) = P(0|\psi_1) = 0$. Notice that necessarily $F_0 \propto \left|\psi_1^\perp\middle\rangle\!\middle\langle\psi_1^\perp\right|$ and $F_1 \propto \left|\psi_0^\perp\middle\rangle\!\middle\langle\psi_0^\perp\right|$, therefore the two proportionality constants are the only free parameters. It proves useful to cast the problem as a semidefinite program [50] and use the conditional success probabilities p_0, p_1 as the parameters to be optimized. The program reads [39, 51]

$$\begin{aligned} \max \quad & \eta_0 p_0 + \eta_1 p_1 \\ \text{s.t.} \quad & G - \Gamma \geq 0 \\ & \Gamma \geq 0\,, \end{aligned} \tag{6}$$

where recall that G is the Gram matrix whose elements are given by the overlaps $g_{ij} = \langle\psi_i|\psi_j\rangle$, and Γ is a diagonal matrix of the conditional success probabilities, $\Gamma = \operatorname{diag}\{p_0, p_1\}$. The first constraint stems from the POVM condition $\mathbb{1} - F_0 - F_1 = F_I \geq 0$. We note that this condition does not depend on the priors, only on G. This is a general feature that applies to any number of hypotheses. In the binary case it yields the interesting uncertainty relation

$$q_0 q_1 \geq c^2, \tag{7}$$

from which the solution of the SDP (6) follows directly:

$$q_0 = c\sqrt{\frac{\eta_1}{\eta_0}}, \qquad q_1 = c\sqrt{\frac{\eta_0}{\eta_1}} \tag{8}$$

if

$$c^2 \leq \frac{\eta_0}{\eta_1} \leq \frac{1}{c^2}, \tag{9}$$

and either $q_0 = 1$ and $q_1 = c^2$ if $\eta_0/\eta_1 \leq c^2$, or $q_1 = 1$ and $q_0 = c^2$ if $\eta_0/\eta_1 \geq 1/c^2$. In these extremal cases the priors are so biased that the optimal measurement discards detecting the state with the lowest prior and the POVM changes from having three to two elements. For instance, in the case $q_0 = 1$ we only have elements F_1 and F_I with $F_1 + F_I = \mathbb{1}$. The symmetric case $\eta_0 = \eta_1 = 1/2$ falls inside the range (9) for any value of the overlap and yields the well-known minimum inconclusive probability $Q = c$ (see, e.g., [3]).

The generalization to arbitrary n amounts to do the change $c \to c^n$ in Eq. (8). Note that this replacement also widens the range of validity of the three outcome POVM

$$c^{2n} \leq \frac{\eta_0}{\eta_1} \leq \frac{1}{c^{2n}}. \tag{10}$$

This fact plays an important role when discussing local protocols. The minimum average success probability finally reads (here and thereof we assume w.l.o.g. that $\eta_0 \leq \eta_1$)

$$Q(n) = \begin{cases} 2\sqrt{\eta_0\eta_1}c^n & \text{if} \quad \sqrt{\frac{\eta_0}{\eta_1}} \geq c^n \\ \eta_0 + \eta_1 c^{2n} & \text{if} \quad \sqrt{\frac{\eta_0}{\eta_1}} \leq c^n \end{cases}. \tag{11}$$

We next show that the optimal performance given by Eq. (11) can always be attained with local measurements. At first glance this result may seem a bit surprising because, for a given n and the same pair of priors, the global optimal POVM has three outcomes [i.e., Eq. (10) is satisfied], while a local one has only two [i.e., Eq. (9) is not fulfilled]. This mismatch could lead us to think that a local strategy could not attain global optimal performance. However, we note that upon obtaining an inconclusive outcome in a two element local POVM, the priors get updated in such a way that they become more equilibrated. In fact, there is a step where the updated priors become sufficiently balanced as to satisfy Eq. (9).

From there on local POVMs also have three outcomes.

The proof of the agreement between the local and global procedures for any n and any initial value of the priors goes as follows. We have to consider the three different ranges of values where the ratio of the priors may lie:

$$(i)\ \frac{\eta_0}{\eta_1} \le c^{2n},\ (ii)\ c^{2n} \le \frac{\eta_0}{\eta_1} \le c^2,\ (iii)\ c^2 \le \frac{\eta_0}{\eta_1} \le 1\,. \tag{12}$$

We start addressing range (iii) (note that the symmetric case of equal priors falls in this range). Here both conditions (9) and (10) are satisfied for any n, i.e., both global and local POVMs give a non-zero probability of detecting any of the states. The first local measurement is the optimal one yielding the inconclusive probabilities given by Eq. (8). After this measurement, if we have not been successful, it is straightforward to see that the priors are updated to $\eta_0^1 = \eta_1^1 = 1/2$. The next measurement is hence optimized for equal priors, which gives an inconclusive outcome with probability c for both sates. Upon failing we repeat the symmetric measurement in all subsequent copies. The overall inconclusive probability of this local strategy then reads

$$\begin{aligned} Q^L(n) &= \eta_0 \Pi_{k=1}^n q_0^k + \eta_1 \Pi_{k=1}^n q_1^k \\ &= \eta_0 c \sqrt{\frac{\eta_1}{\eta_0}} c^{n-1} + \eta_1 c \sqrt{\frac{\eta_0}{\eta_1}} c^{n-1} \\ &= 2\sqrt{\eta_0\eta_1} c^n, \end{aligned} \tag{13}$$

i.e., the optimal value in the first case of Eq. (11).

In the range (i) the priors are so biased that, even for a global measurement, it is not worth detecting the state $|\psi_0\rangle$. The local procedure consists of a series of measurements $\{F_1 = |\psi_0^\perp\rangle\langle\psi_0^\perp|, F_I = |\psi_0\rangle\langle\psi_0|\}$ that either detect unambiguously $|\psi_1\rangle$ or fail. In this case we have

$$Q^L(n) = \eta_0 \times (1)^n + \eta_1 (c^2)^n = \eta_0 + \eta_1 c^{2n}, \tag{14}$$

which coincides with the second line of Eq. (11). Note that, for large n, the region (i) is increasingly small. We would like to stress that, while all the measurements are identical, the updated priors are not. Each time one gets an inconclusive result the belief that the state is $|\psi_1\rangle$ diminishes and the belief in favor of $|\psi_0\rangle$ increases. This balances the priors, however not enough to be worth testing the state $|\psi_0\rangle$. Indeed, Bayesian updating gives that, for all $k \le n-1$,

$$\frac{\eta_0^{(k)}}{\eta_1^{(k)}} = \frac{1}{c^2} \frac{\eta_0^{(k-1)}}{\eta_1^{(k-1)}} \to \frac{\eta_0^{(k)}}{\eta_1^{(k)}} = \frac{1}{c^{2k}} \frac{\eta_0}{\eta_1} \le c^2, \tag{15}$$

since $\eta_0/\eta_1 \le c^{2n}$ in this range.

The most interesting range is (ii). While the global strategy uses a three outcome POVM, the local strategy starts with a fully biased two outcome measurement (because $\eta_0/\eta_1 \le c^2$). Upon obtaining an inconclusive outcome, the priors are updated according to Eq. (15) and get more balanced, i.e., our belief that the state is $|\psi_0\rangle$ increases. We keep doing the same measurement until a step k_0 that yields $\eta_0^{(k_0)}/\eta_1^{(k_0)} \ge c^2$. This step is guaranteed to be reached before n, i.e., $k_0 < n$. Simply observe that

$$\frac{\eta_0^{(k_0)}}{\eta_1^{(k_0)}} = \frac{1}{c^{2k_0}} \frac{\eta_0}{\eta_1} \ge c^2 \to \frac{\eta_0}{\eta_1} \ge c^{2(k_0+1)}, \tag{16}$$

which is always compatible with the initial condition of beginning in range (ii) for some $k_0 < n$ (the actual value of k_0 depends on the particular ratio η_0/η_1). Therefore, the protocol consists in performing a sequence of fixed two-outcome measurements until the k_0 step, when we do a three-outcome measurement for biased priors $\eta_0^{(k_0)}$ and $\eta_1^{(k_0)}$, and continue with a sequence of three-outcome measurements for balanced priors as in region (iii) (of course, for as long as we keep on failing). The probability for n failures is

$$\begin{aligned} Q^L(n) =& \eta_0 \left[(1)^{k_0} \times c \sqrt{\frac{\eta_1^{(k_0)}}{\eta_0^{(k_0)}}} \times c^{n-k_0-1} \right] \\ &+ \eta_1 \left[(c^2)^{k_0} \times c \sqrt{\frac{\eta_0^{(k_0)}}{\eta_1^{(k_0)}}} \times c^{n-k_0-1} \right], \end{aligned} \tag{17}$$

were we have explicitly displayed the terms corresponding to the three different stages of the procedure. Now, taking into account the expression of the updated priors ratio Eq. (15), we get

$$Q^L(n) = \eta_0 \sqrt{\frac{\eta_1}{\eta_0}} c^n + \eta_1 \sqrt{\frac{\eta_0}{\eta_1}} c^n = 2\sqrt{\eta_0\eta_1} c^n, \tag{18}$$

which again coincides with the global bound, Eq. (11).

We can summarize the procedure in all three regions by the position k_0 of the first three-outcome local measurement in the sequence. In region (iii), $k_0 = 0$ and we already start with a three outcome local measurement. In region (ii), $k_0 \leq n-1$, i.e., the accumulated balance of the priors given by the inconclusive outcomes induces to start a three-outcome measurement at some point before reaching n. Finally, in region (i), for very biased priors the number of copies is not enough to abandon the strategy that only detects one of the states.

As in the minimum error case, this local protocol works also in the non-$i.i.d.$ case of product states. One just needs to take into account that at each step k we have a different overlap c_k and also a different validity range Eq. (9). The minimum failure probability is simply Eq. (11) with c^n replaced by $C = c_1 c_2 \cdots c_n$.

It is worth emphasizing that the local procedure described yields the optimal success probability at each step, regardless of total number of systems that are finally available for measurement. Besides not requiring quantum memories, the local measurement at any given step depends only on the outcome of the previous measurement, hence the size of the classical memory required is minimal. Furthermore, the measurement setting at most changes two times.

4 Zero-error identification of symmetric multiple hypotheses

In this section we extend our results to multi-hypothesis scenarios. Rather surprisingly, the performance of online sequential strategies and their comparison with the global optimal values for zero-error identification have hardly been explored. Although even the simplest case of three symmetric states (TSS) is quite a big challenge, as discussed in [42], the constraints imposed by the zero-error requirement provide more chances to obtain analytical results. Here we will mainly focus our attention in the TSS case, and also address some straightforward generalizations.

The problem we address consists in doing a zero-error identification of a set of states that have equal prior probabilities $\eta_i = 1/3$, $i = 0,1,2$, and symmetric overlaps $\langle\psi_0|\psi_1\rangle = \langle\psi_1|\psi_2\rangle = \langle\psi_2|\psi_0\rangle = c$. We first analyze the case of positive values of c, and then we address the negative range. We finally consider the sequential performance for complex values of c.

The positive range, $0 \leq c \leq 1$, can actually be solved for any number r of hypotheses as we show below. Note that the anomaly identification problem [38] falls under this case. The Gram matrix, G, together with the priors encapsulate all the discrimination properties of an ensemble, and no explicit form of the states is even needed, although the very existence of a valid Gram, i.e., $G \geq 0$, imposes some restrictions on the states that can give rise to G. For instance, if $0 \leq c < 1$ the states are necessarily linearly independent (a requisite to have zero-error discrimination [48]) and therefore the dimension d of the Hilbert space of the states must be at least $d \geq r$. The Gram matrix of a set of three states with equal overlap c reads

$$G = \begin{pmatrix} 1 & c & c \\ c & 1 & c \\ c & c & 1 \end{pmatrix}. \tag{19}$$

In this symmetric setting the optimal conditional success probabilities must be identical, $p_i = p$, hence the SDP (6) reads

$$\begin{aligned} \max \quad & p \\ \text{s.t.} \quad & G - p\mathbb{1} \geq 0 \\ & p \geq 0\,. \end{aligned} \tag{20}$$

This optimization gives the minimum eigenvalue of G,

$$p = \lambda_{\min}(G) = 1 - c\,, \tag{21}$$

i.e., $q = c$. Note that this solution is the same for any number of symmetric hypotheses. Given n copies of the states, the minimum inconclusive probability for any set of symmetric states with constant positive overlaps is $Q = c^n$.

Next we would like to know if the global performance can also be reached with an online protocol. This way, no quantum memory would be required and the identification process can be completed at much earlier times without compromising the probability of success [52]. The online protocol consists simply in a local optimal unambiguous measurement at each step k. One stops as soon as a conclusive outcome is obtained. This protocol can be regarded as a Bayesian updating procedure: if the identification is successful, the priors become 1 for the identified state and zero for the rest of states. If one fails, the updated priors are again symmetric. The proof follows directly from the fact that the inconclusive probability at each step is c and n consecutive failures have probability c^n.

The particular form of the unambiguous POVM that we need depends on the specific source states at hand. We present the TSS case ($r = 3$) in detail, but the generalization to an arbitrary number of symmetric source states is straightforward. As already introduced in Section 2, the most convenient parametrization is to use the eigenbasis of the unitary $U = |0\rangle\langle 0| + e^{2i\pi/3}\,|1\rangle\langle 1| + e^{4i\pi/3}\,|2\rangle\langle 2|$, and write the states as $|\psi_0\rangle = \xi_0\,|0\rangle + \xi_1\,|1\rangle + \xi_2\,|2\rangle$, $|\psi_1\rangle = U\,|\psi_0\rangle$ and $|\psi_2\rangle = U^2\,|\psi_0\rangle$. Here the amplitudes ξ_i are related to the eigenvalues of G, λ_i, through

$$\xi_i = \sqrt{\frac{\lambda_i}{3}}, \quad i = 0, 1, 2, \tag{22}$$

which is the direct extension of Eq. (1). This parametrization can be regarded as the canonical form of symmetric states for any overlap c (real or complex), and generalizes trivially to any number of symmetric states. It is useful to note that the operator $\Omega = \sum_{k=0}^{2} |\psi_k\rangle\langle\psi_k|$ is diagonal in this basis:

$$\Omega = 3\begin{pmatrix} |\xi_0|^2 & 0 & 0 \\ 0 & |\xi_1|^2 & 0 \\ 0 & 0 & |\xi_2|^2 \end{pmatrix} \tag{23}$$

(this property holds true for any set of three symmetric states, normalized or not). The specific values of ξ_i are

$$\xi_0 = \sqrt{\frac{1+2c}{3}}\,, \quad \xi_1 = \xi_2 = \sqrt{\frac{1-c}{3}}\,. \tag{24}$$

The POVM has elements $F_i = p|\tilde{\phi}_i\rangle\langle\tilde{\phi}_i|$, $i = 0, 1, 2$, and $F_I = \mathbb{1} - \sum_{i=0}^{2} F_i$, where $p = 1 - c$, as given in Eq. (21). The unnormalized states $|\tilde{\phi}_i\rangle$ satisfy the unambiguous condition $\langle\tilde{\phi}_i|\psi_j\rangle = \delta_{ij}$ and are constructed from a state $|\tilde{\phi}_0\rangle$ as $|\tilde{\phi}_k\rangle = U^k|\tilde{\phi}_0\rangle$. With this parametrization the fiducial state simply reads

$$|\tilde{\phi}_0\rangle = \sum_{i=0}^{2} \sqrt{\frac{1}{3\lambda_i}}\,|i\rangle\ . \tag{25}$$

Let us next complete the analysis for negative values of the overlap. We note that $G \geq 0$ implies that $c \geq -1/2$. In the range $c \in [-1/2, 0]$, the minimum eigenvalue of Eq. (20) changes to $\lambda_{\min} = 1 - 2|c|$. For a given number of copies n the minimum eigenvalue alternates between $1 - 2|c|^n$ and $1 - |c|^n$ depending on whether n is odd or even, respectively. This means that the minimum inconclusive probability is

$$Q(n) = \begin{cases} 2|c|^n & \text{if } n \text{ is odd} \\ |c|^n & \text{if } n \text{ is even} \end{cases}. \tag{26}$$

Note that indeed $Q(n)$ is a decreasing function of n since $|c|^{2k} \geq 2|c|^{2k+1} \geq |c|^{2k+2}$ if $|c| \leq 1/2$.

A local protocol based on fixed unambiguous measurements gives a failure probability $Q^L = 2^n|c|^n$, which is away from the optimal value by an exponential factor. Given such a large gap, one expects that there exist better local protocols. The analysis of the extremal value $c = -1/2$ gives us the clues on how to proceed. For this value one has $\det G = 0$, i.e., the three states are linearly dependent. This means that zero-error identification is not possible [48] with only one copy. Of course, given $n > 1$ copies, the tensored states become linearly independent with a global Gram matrix $G > 0$. The global inconclusive probability is given by Eq. (26) with $c = -1/2$. It is remarkable that $Q(n)$ is the same for $2k$ and $2k + 1$ copies of the state, $Q(2k) = Q(2k + 1) = 2^{-2k}$, i.e., having an additional copy is of no use (a result already noticed in [49]).

Although with only one copy it is impossible to unambiguously identify the state, one can still gather useful information to be used in the following measurements. In particular, it is possible to perform a measurement that is able to exclude one of the states [53] with 100% probability. It is easy to see that a POVM with elements $E_k = \frac{2}{3}|\psi_k^\perp\rangle\langle\psi_k^\perp|$, $k = 0, 1, 2$, does the job, as indeed it constitutes a POVM: $\sum_{k=0}^{2} E_k = \mathbb{1}$. Then, from the second step onwards, one can proceed with two-state discrimination measurements as in Section 3 with equal priors. The failure probability then reads

$$Q^L(n) = \left(\frac{1}{2}\right)^{n-1}, \tag{27}$$

which coincides with the optimal value for odd n, Eq. (26). Hence, this protocol is optimal for any odd number of states. For even n it does not reach global optimality, but we conjecture that also in this case no local protocol can do better than this one.

We can now tackle the whole negative range $-1/2 < c < 0$ with local protocols. The idea is to combine unambiguous identification with the state-excluding measurement that has been the key idea to solve the extremal point $c = -1/2$. The unambiguous POVM elements are $F_k = (1-2|c|)|\tilde{\phi}_k\rangle\langle\tilde{\phi}_k|$, $k = 0,1,2$, where $|\tilde{\phi}_k\rangle$ are given in Eq. (25) and above, and $1-2|c|$ is the minimum eigenvalue of G in this range of c. The crucial observation is that it is possible to construct three additional operators E_l that exclude one of the states and satisfy $E := \sum_{l=0}^{2} E_l = \mathbb{1} - \sum_{k=0}^{2} F_k =: \mathbb{1} - F$. Thus, with the first measurement, either a state is identified with certainty (operators F_i) or a state is excluded also with certainty (operators E_l). In other words, either we stop or we continue with a two-state unambiguous measurement (with equal priors after their update). Using Eq. (25) and Eq. (23) with the ordering $\lambda_0 = \lambda_1 = 1 + |c|$ and $\lambda_2 = 1 - 2|c|$, we have

$$\mathbb{1} - F = \frac{3|c|}{1+|c|}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{28}$$

The operators

$$E_k = \frac{3|c|}{1+|c|}|\tilde{\varphi}_k\rangle\langle\tilde{\varphi}_k|, \tag{29}$$

where $|\tilde{\varphi}_k\rangle = U^k|\tilde{\varphi}_0\rangle$ and $|\tilde{\varphi}_0\rangle = |0\rangle - |1\rangle$, satisfy the desired conditions

$$\begin{aligned} &\langle\psi_k| E_k |\psi_k\rangle = 0, \quad k = 0,1,2, \\ &E + F = \mathbb{1}. \end{aligned} \tag{30}$$

With this measurement, the success probability of unambiguously detecting the state is $1-2|c|$, and hence the probability of excluding one state is $2|c|$. The following measurements are binary symmetric which give an optimal inconclusive probability $|c|$, Eq. (8) with $\eta_0/\eta_1 = 1$. Therefore, after n measurements the overall inconclusive probability reads

$$Q^L(n) = 2|c|^n, \tag{31}$$

which again coincides with the optimal value Eq.(26) for n odd. This result also proves that, for negative values of c, this protocol is the optimal one among all local procedures when the number of states measured is odd. For even numbers of states, although we do no have a rigorous proof, there are strong evidences that this is also the case. A measurement can provide three types of information: (i) exclude two states (unambiguous identification), (ii) exclude one of the states (exclusion) or (iii) update our belief over the different states (learning). Naturally (i) is the most valuable information. In a convex combination of POVM elements that achieve (ii) and (iii), note that the overall failure probability with two copies decreases if one puts the maximum weight in the elements leading to (ii). The POVM $\{F_{0,1,2}, E_{0,1,2}\}$ maximizes the contribution to the success probabilities of (i) and (ii) by construction, hence it is presumably the optimal local measurement for any n.

Finally, for complex overlaps $c = se^{i\theta}$, the eigenvalues of the Gram matrix read

$$\lambda_k = 1 + 2s\cos\left[\theta + \frac{2k\pi}{3}\right], \quad k = 0,1,2. \tag{32}$$

The minimum eigenvalue is λ_1 for $0 \le \theta \le 2\pi/3$, λ_0 for $2\pi/3 \le \theta \le 4\pi/3$, and λ_2 for $4\pi/3 \le \theta \le 2\pi$. The positivity of the Gram matrix imposes some restrictions on the phase θ for $s > 1/2$. The region allowed by the physical restriction $G \ge 0$ is the triangle depicted in Fig. 1. Note that, by symmetry, values of the overlap c differing in a phase of $2\pi/3$ are equivalent. In particular this holds true for the three lines with $\theta = 0, 2\pi/3, 4\pi/3$ and the dashed lines with $\theta = \pi/3, \pi, 5\pi/3$. That is, for values of c lying in the "Mercedes-Benz" lines of Fig. 1 a protocol of repeated unambiguous local measurements provide the same success probability as gathering all the copies and performing an optimal global measurement, for any n. For values in the dashed lines of Fig. 1 this is only true for odd n.

For complex values of the overlap and for $n = 2$ copies, it is possible to find a region with a

"Mitsubishi-logo" shape where a sequence of two local measurements yields the same success probability as the global measurement [42]. However, the strategy proposed in [42] is not online, since it requires knowing the horizon. Indeed, it sacrifices optimality in the first step (by not putting the maximum possible weight on the POVM elements $F_{0,1,2}$) in order to match global performance at the second step. We have carried out numerical checks by optimizing over online strategies with local POVMs of the form $\{F_{0,1,2}, E_{0,1,2}, \mathbb{1} - F - E\}$. Our results indicate that there is no online protocol yielding the global optimal success probability outside the dark blue and magenta lines of Fig. 1.

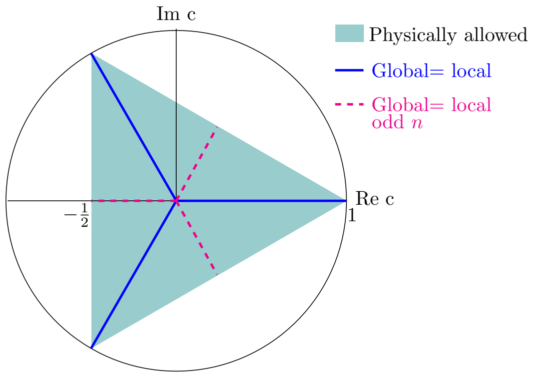


Figure 1: Complex plane of the overlap values. Horizontal and vertical axis correspond to real and imaginary parts, respectively. The shaded triangular region is the physically allowed range. The Mercedes-Benz lines of length one (solid blue) are the values for which there is an online protocol that matches the optimal performance of global schemes. The rotated lines of length 1/2 (dashed magenta) are the values for which optimality is also attained for odd numbers of copies.

Our results naturally extend to the case of product states that are not necessarily identical, but where each local state comes from a different symmetric trine $\{|\psi_0^{(k)}\rangle, |\psi_1^{(k)}\rangle, |\psi_2^{(k)}\rangle\}$ with overlap c_k, $k = 1, \ldots, n$. This case corresponds to a non-*i.i.d.* source that produces three possible global hypotheses of the form $|\psi_i^{(1)}\rangle|\psi_i^{(2)}\rangle \cdots |\psi_i^{(n)}\rangle$, $i = 0, 1, 2$. For instance, as in the case of identical copies, our online scheme yields the optimal global success probability if $c_k \geq 0, \quad \forall k$. Also, if the local trines have positive and negative values of the overlap, the online scheme matches optimal performance if $\Pi_{k=1}^{n} c_k < 0$. Notice that in this case there must be a first trine with negative overlap, say at step k. Recall that the local measurement for this trine either identifies the state with probability $1 - 2|c_k|$ or excludes one of the possibilities with probability $2|c_k|$ and thereafter one has a symmetric binary problem. Thus, the total inconclusive probability reads $Q = c_1 c_2 \cdots 2|c_k||c_{k+1}| \cdots |c_n|$ which coincides with the global optimum $2|c_1 c_2 \cdots c_n|$ since $c_i > 0$ for $i < k$.

5 Conclusions

The tasks of binary pure state identification for minimum and zero error can be carried out in an online fashion with optimal performance. The scheme has no horizon, i.e., the information about the number of states available does not affect the measurement scheme. Optimality is attained at each step regardless of whether systems are lost or one has to stop at an earlier time than planned.

Extending the analysis beyond the binary case is a much more challenging task. Already the minimum extension of three symmetric states is a highly non-trivial case. For minimum error the direct application of local measurements with Bayesian updating for two copies of the states does not give the optimal global performance [46, 47]. As far as we are aware, there is no proof that this is the case for more general one-way local protocols.

The zero-error identification task, still being quite involved, offers more possibilities to be tackled as most of the structure of the POVM is already fixed by the zero-error constraints. We have formulated the problem as a semidefinite program that greatly simplifies the optimization task and also provides a very useful tool, not only for numerical calculations but, as we exploit here, also for obtaining analytical results. It also opens the path for addressing more complex instances as, e.g., non-symmetric overlaps or different priors. We have given a canonical way of writing r symmetric states in terms of the eigenvalues of their Gram matrix. For $r = 3$, we have obtained the optimal online protocol for arbitrary positive values of the overlap and any n, and for negative values for odd n. We have proven that these protocols attain the optimal global performance. These results directly extend to symmetric complex values of the overlap with phases $2\pi/3$. Our findings for positive overlaps also hold for any number of hypotheses r. Unlike [42], we are not restricted to sources of linearly independent states. We are able to find, e.g., online optimal protocols for trines of symmetric qubits.

For arbitrary complex values of the overlap, our results also suggest that there is no online protocol achieving the same performance as global protocols outside the three symmetric lines of Fig. 1. The existence of this gap could be exploited in several ways. For instance, one could consider an extension of the B92 protocol [7] with trine states to produce keys of trits. If Alice were to use multiple copies of a trine state for which such gap exists with the objective of increasing the key rate, Bob would take advantage by measuring collectively, while Eve would be forced to measure in an online manner (thus suboptimally) to keep the rate of communication. Another direct application of our results is probabilistic cloning of states from a finite set [54] in the asymptotic limit of producing many clones: if the set is a trine where there is no gap between online and global strategies, the task could be optimally performed by an online measure and prepare strategy, thus saving resoures with respect to measuring several copies collectively.

Acknowledgments

We acknowledge discussions with E. Bagan and K. Nakahira. We acknowledge financial support from the Spanish Agencia Estatal de Investigación, project PID2019-107609GB-I00, from Secretaria d'Universitats i Recerca del Departament d'Empresa i Coneixement de la Generalitat de Catalunya, co-funded by the European Union Regional Development Fund within the ERDF Operational Program of Catalunya (project QuantumCat, ref. 001-P-001644), and Generalitat de Catalunya CIRIT 2017-SGR-1127. EMV thanks financial support from CONACYT.

References

[1] A. Chefles, *Quantum state discrimination*, Contemporary Physics **41**, 401 (2000), DOI: 10.1080/00107510010002599.

[2] S. M. Barnett and S. Croke, *Quantum state discrimination*, Advances in Optics and Photonics **1**, 238 (2009), DOI: 10.1364/AOP.1.000238.

[3] J. A. Bergou, *Discrimination of quantum states*, Journal of Modern Optics **57**, 160 (2010), DOI: 10.1080/09500340903477756.

[4] J. Bae and L.-C. Kwek, *Quantum state discrimination and its applications*, Journal of Physics A: Mathematical and Theoretical **48**, 083001 (2015), DOI: 10.1088/1751-8113/48/8/083001.

[5] C. W. Helstrom, *Quantum detection and estimation theory* (Academic press) (1976).

[6] N. Gisin and R. Thew, *Quantum communication*, Nature Photonics **1**, 165 (2007), DOI: 10.1038/nphoton.2007.22.

[7] C. H. Bennett, *Quantum cryptography using any two nonorthogonal states*, Physical Review Letters **68**, 3121 (1992), DOI: 10.1103/PhysRevLett.68.3121.

[8] N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden, *Quantum cryptography*, Reviews of Modern Physics **74**, 145 (2002), DOI: 10.1103/RevModPhys.74.145.

[9] A. Acín, J. Bae, E. Bagan, M. Baig, L. Masanes, and R. Muñoz-Tapia, *Secrecy properties of quantum channels*, Physical Review A **73**, 012327 (2006), DOI: 10.1103/PhysRevA.73.012327.

[10] R. Renner, *Security of quantum key distribution*, International Journal of Quantum Information **6**, 1 (2008), DOI: 10.1142/S0219749908003256.

[11] D. Bacon, A. Childs, and W. van Dam, *From optimal measurement to efficient quantum algorithms for the hidden subgroup problem over semidirect product groups*, in *46th Annual IEEE Symposium on Foundations of Computer Science (FOCS'05)*, vol. 2005, 469–478 (IEEE) (2005), ISBN 0-7695-2468-0, DOI: 10.1109/SFCS.2005.38.

[12] J. Bae, W.-Y. Hwang, and Y.-D. Han, *No-Signaling Principle Can Determine Optimal Quantum State Discrimination*, Physical Review Letters **107**, 170403 (2011), DOI: 10.1103/PhysRevLett.107.170403.

[13] R. Takagi and B. Regula, *General Resource Theories in Quantum Mechanics and Beyond: Operational Characterization via Discrimination Tasks*, Physical Review X **9**, 031053 (2019), DOI: 10.1103/PhysRevX.9.031053.

[14] M. Oszmaniec and T. Biswas, *Operational relevance of resource theories of quantum measurements*, Quantum **3**, 133 (2019), DOI: 10.22331/q-2019-04-26-133.

[15] R. Uola, T. Kraft, J. Shang, X.-D. Yu, and O. Gühne, *Quantifying Quantum Resources with Conic Programming*, Physical Review Letters **122**, 130404 (2019), DOI: 10.1103/PhysRevLett.122.130404.

[16] W. K. Wootters and W. H. Zurek, *A single quantum cannot be cloned*, Nature **299**, 802 (1982), DOI: 10.1038/299802a0.

[17] J. Walgate, A. J. Short, L. Hardy, and V. Vedral, *Local Distinguishability of Multipartite Orthogonal Quantum States*, Physical Review Letters **85**, 4972 (2000), DOI: 10.1103/PhysRevLett.85.4972.

[18] S. Virmani, M. Sacchi, M. Plenio, and D. Markham, *Optimal local discrimination of two multipartite pure states*, Physics Letters A **288**, 62 (2001), DOI: 10.1016/S0375-9601(01)00484-4.

[19] Y.-X. Chen and D. Yang, *Optimal conclusive discrimination of two nonorthogonal pure product multipartite states through local operations*, Physical Review A **64**, 064303 (2001), DOI: 10.1103/PhysRevA.64.064303.

[20] Y.-X. Chen and D. Yang, *Optimally conclusive discrimination of nonorthogonal entangled states by local operations and classical communications*, Physical Review A **65**, 022320 (2002), DOI: 10.1103/PhysRevA.65.022320.

[21] Z. Ji, H. Cao, and M. Ying, *Optimal conclusive discrimination of two states can be achieved locally*, Physical Review A **71**, 032323 (2005), DOI: 10.1103/PhysRevA.71.032323.

[22] A. Acín, E. Bagan, M. Baig, L. Masanes, and R. Muñoz-Tapia, *Multiple-copy two-state discrimination with individual measurements*, Physical Review A **71**, 032338 (2005), DOI: 10.1103/PhysRevA.71.032338.

[23] S. Croke, S. M. Barnett, and G. Weir, *Optimal sequential measurements for bipartite state discrimination*, Physical Review A **95**, 052308 (2017), DOI: 10.1103/PhysRevA.95.052308.

[24] A. Peres and W. K. Wootters, *Optimal detection of quantum information*, Physical Review Letters **66**, 1119 (1991), DOI: 10.1103/PhysRevLett.66.1119.

[25] E. Chitambar and M.-H. Hsieh, *Revisiting the optimal detection of quantum information*, Physical Review A **88**, 020302 (2013), DOI: 10.1103/PhysRevA.88.020302.

[26] H.-C. Cheng, A. Winter, and N. Yu, *Discrimination of quantum states under locality constraints in the many-copy setting*, in *2021 IEEE International Symposium on Information Theory (ISIT)*, 1188–1193 (IEEE) (2021), DOI: 10.1109/ISIT45174.2021.9518100.

[27] C. H. Bennett, D. P. DiVincenzo, C. A. Fuchs, T. Mor, E. Rains, P. W. Shor, J. A. Smolin, and W. K. Wootters, *Quantum nonlocality without entanglement*, Physical Review A **59**, 1070 (1999), DOI: 10.1103/PhysRevA.59.1070.

[28] C. A. Fuchs, *Just two nonorthogonal quantum states*, in P. Kumar, G. M. D'Ariano, and O. Hirota (eds.), *Quantum Communication, Computing, and Measurement 2*, 11–16 (Springer) (2002), DOI: 10.1007/0-306-47097-7˙2.

[29] T. Eggeling and R. F. Werner, *Hiding classical data in multipartite quantum states*, Physical Review Letters **89**, 097905 (2002), DOI: 10.1103/PhysRevLett.89.097905.

[30] S. S. Bhattacharya, S. Saha, T. Guha, and M. Banik, *Nonlocality without entanglement: Quantum theory and beyond*, Physical Review Research **2**, 012068 (2020), DOI: 10.1103/PhysRevResearch.2.012068.

[31] D. G. Fischer, S. H. Kienle, and M. Freyberger, *Quantum-state estimation by self-learning measurements*, Physical Review A **61**, 032306 (2000), DOI: 10.1103/PhysRevA.61.032306.

[32] A. Wald, *Sequential Analysis*, Dover books on advanced mathematics (Dover Publications) (1973), ISBN 9780486615790.

[33] E. Martínez Vargas, C. Hirche, G. Sentís, M. Skotiniotis, M. Carrizo, R. Muñoz-Tapia, and J. Calsamiglia, *Quantum Sequential Hypothesis Testing*, Physical Review Letters

126, 180502 (2021), DOI: 10.1103/PhysRevLett.126.180502.

[34] S. M. Barnett, *Minimum-error discrimination between multiply symmetric states*, Physical Review A **64**, 030303 (2001), DOI: 10.1103/PhysRevA.64.030303.

[35] J. A. Bergou, U. Futschik, and E. Feldman, *Optimal unambiguous discrimination of pure quantum states*, Physical Review Letters **108**, 250502 (2012), DOI: 10.1103/PhysRevLett.108.250502.

[36] N. Dalla Pozza and G. Pierobon, *Optimality of square-root measurements in quantum state discrimination*, Physical Review A **91**, 042334 (2015), DOI: 10.1103/PhysRevA.91.042334.

[37] H. Krovi, S. Guha, Z. Dutton, and M. P. da Silva, *Optimal measurements for symmetric quantum states with applications to optical communication*, Physical Review A **92**, 062333 (2015), DOI: 10.1103/PhysRevA.92.062333.

[38] M. Skotiniotis, R. Hotz, J. Calsamiglia, and R. Muñoz Tapia, *Identification of malfunctioning quantum devices*, arXiv:1808.02729 (2018).

[39] G. Sentís, J. Calsamiglia, and R. Muñoz Tapia, *Exact identification of a quantum change point*, Physical Review Letters **119**, 140506 (2017), DOI: 10.1103/PhysRevLett.119.140506.

[40] E. Chitambar, R. Duan, and M.-H. Hsieh, *When Do Local Operations and Classical Communication Suffice for Two-Qubit State Discrimination?*, IEEE Transactions on Information Theory **60**, 1549 (2014), DOI: 10.1109/TIT.2013.2295356.

[41] G. Sentís, E. Martínez-Vargas, and R. Muñoz Tapia, *Online strategies for exactly identifying a quantum change point*, Physical Review A **98**, 052305 (2018), DOI: 10.1103/PhysRevA.98.052305.

[42] K. Nakahira, K. Kato, and T. S. Usuda, *Local unambiguous discrimination of symmetric ternary states*, Physical Review A **99**, 022316 (2019), DOI: 10.1103/PhysRevA.99.022316.

[43] D. Brody and B. Meister, *Minimum decision cost for quantum ensembles*, Physical Review Letters **76**, 1 (1996), DOI: 10.1103/PhysRevLett.76.1.

[44] G. L. Nemhauser, *Introduction to dynamic programming* (John Wyley and Sons, New York) (1966).

[45] S. Brandsen, M. Lian, K. D. Stubbs, N. Rengaswamy, and H. D. Pfister, *Adaptive procedures for discriminating between arbitrary tensor-product quantum states*, in *2020 IEEE International Symposium on Information Theory (ISIT)*, 1933–1938 (IEEE) (2020), DOI: 10.1109/ISIT44484.2020.9174234.

[46] S. Brandsen, K. D. Stubbs, and H. D. Pfister, *Reinforcement learning with neural networks for quantum multiple hypothesis testing*, in *2020 IEEE International Symposium on Information Theory (ISIT)*, 1897–1902 (IEEE) (2020), DOI: 10.1109/ISIT44484.2020.9174150.

[47] K. Nakahira, K. Kato, and T. S. Usuda, *Optimal discrimination of optical coherent states cannot always be realized by interfering with coherent light, photon counting, and feedback*, Physical Review A **97**, 022320 (2018), DOI: 10.1103/PhysRevA.97.022320.

[48] A. Chefles, *Unambiguous discrimination between linearly independent quantum states*, Physics Letters A **239**, 339 (1998), DOI: 10.1016/S0375-9601(98)00064-4.

[49] A. Chefles, *Unambiguous discrimination between linearly dependent states with multiple copies*, Physical Review A **64**, 062305 (2001), DOI: 10.1103/PhysRevA.64.062305.

[50] Y. C. Eldar, *A semidefinite programming approach to optimal unambiguous discrimination of quantum states*, IEEE Transactions on Information Theory **49**, 446 (2003), DOI: 10.1109/TIT.2002.807291.

[51] E. Martínez Vargas and R. Muñoz Tapia, *Certified answers for ordered quantum discrimination problems*, Physical Review A **100**, 042331 (2019), DOI: 10.1103/PhysRevA.100.042331.

[52] E. Martínez Vargas, C. Hirche, G. Sentís, M. Skotiniotis, M. Carrizo, R. Muñoz Tapia, and J. Calsamiglia, *Quantum sequential hypothesis testing*, Physical Review Letters **126**, 180502 (2021), DOI: 10.1103/PhysRevLett.126.180502.

[53] C. M. Caves, C. A. Fuchs, and R. Schack, *Conditions for compatibility of quantum-state assignments*, Physical Review A

66, 062111 (2002), DOI: 10.1103/PhysRevA.66.062111.

[54] L.-M. Duan and G.-C. Guo, *Probabilistic cloning and identification of linearly independent quantum states*, Physical Review Letters **80**, 4999 (1998), DOI: 10.1103/PhysRevLett.80.4999.

“The limits of my language mean the limits of my world.”

— Tractatus logico-philosophicus, Ludwig Wittgenstein

6

Conclusions

This thesis is divided in two parts as the topics itself ask. This notwithstanding, there is a clear unity in the topics treated here: the interrelationship between states ordered in time and their optimal detection.

First we consider a machine that produces a change point. We addressed two problems that extend previous work on the exact identification of the change point problem. Having a full solution of the global problem is very useful as it is easier to address variations. Firstly we ask if the global solution can be attained with more a practical scheme: online. Surprisingly, the answer turns out to be yes, in a given region and very close in another one. This is remarkable because it is not the case for the minimum error scheme. Depending on the function to optimize, one can get different answers on the resources needed. Also, this suggests addressing the question of when there is a gain with measurements that have entangled operators in them in sources that are not necessarily i.i.d.

The second problem is one that remains in the global case of unambiguous discrimination for the change point problem but realises that a less constrained variation of this scheme is possible. These schemes are possible because the set of copies given to us is ordered in time. Quantum discrimination depends on the fact that the possible states given by Alice (to Bob) are known beforehand. In this case, not only are the states known, also the knowledge of their ordering is available. This allows for a different protocol that takes this into account. We introduce a certified answer protocol that interpolates between unambiguous and minimum error for this kind of ordered cases. We are also able to give an analytic lower bound for the certified error scheme that depend on the solution of the well known minimum error problem. Our scheme can be generalized to graphs of states as we will see in the outlook. The scheme and techniques introduced in this publication have a lot of versatility and could be applied to other interesting problems.

For the second part we change to an iid source. This part of the thesis addresses a very widespread technique in statistics: Sequential Analysis. Remarkably, in the quantum case, the kind of perspective for sequential analysis (minimizing the average number of copies) had been addressed very little before. We investigated this problem for special cases and then, for two hypotheses addressed the problem in a very general manner (for mixed states of finite dimension). We were able to give an ultimate lower bound on the average required samples to fulfill error constraints. We also obtain upper bounds using a specific strategy. We study a worst case scenario and the attainability of an upper bound with qubits. Recently it has been shown that our ultimate bound can be saturated [LTT]. The special case of pure states is also addressed. This is by no means a trivial case, it shows very nontrivial behavior. For instance, we observe that the optimal scheme here is to perform online unambiguous discrimination. Remarkably, there a finite number of states is needed to make a decision with zero error. Our results hint towards a special relationship between the online case and unambiguous discrimination for pure states. For mixed states there is a wide range of problems that can be addressed with the problem treated here.

The study of the pure case in the previous problem (and the first one of this thesis) suggested the study of unambiguous discrimination in an online or sequential fashion. We noticed that the problem of online unambiguous discrimination had already been addressed for two hypothesis but very few results were known for more hypotheses. We study the online case of three hypotheses for multiple copies. We characterize the regions where online unambiguous is equal to the global for a symmetric case of three hypotheses. Our results suggests developing a general theory for having criteria of gain with entangled operators over LOCC ones. We observe a general example for multihypotheses: the anomally detection matrix. This suggest studying what makes this example special in terms of no gain with entangled operators. Also, our results can be applied to study cryptographic protocols based on unambiguous discrimination of three hypotheses.

6.1 Outlook

Here I present some ideas that stem from the work done in this thesis. This is unfinished work as they still have to be developed in the near future.

6.1.1 Graph order

An interesting perspective from our SDP approach in the certified answers is that it can be generalized for any configuration of graphs of states. That is, one can regard the linear chain of states as a type of graph. In general, the relationship of the states can be encoded into a Gram matrix. A certified error discrimination scheme can be studied in this special case. This would generalize the notion of distance of the errors. This would be useful to distinguish more complicated structures of states.

6.1.2 Markov source machines

This project is related with the change point problem. Normally, the sources one considers theoretically are i.i.d., which means that they produce a the same state once and again. The change point problem takes a non-i.i.d. source because it changes at some point. Are there more general sources? The answer to this question turns out to be positive, we can have a "Markov source", one that produces states according to the state of a Markov chain. Basically, as in Figure (6.1).

One might follow the usual analysis of Markov chains and try to apply it. For example, a two state Markov chain as in Figure (6.1) would be described by a stationary state given by

$$\nu = \begin{pmatrix} \frac{\delta}{\delta+\epsilon} \\ \frac{\epsilon}{\delta+\epsilon} \end{pmatrix}$$

which is a probability distribution. This means that the machine produces with probability $\delta/(\delta+\epsilon)$ the state $|0\rangle$ and with probability $\epsilon/(\delta+\epsilon)$ the state $|\phi\rangle$. This might suggest that the machine produces the mixed state

$$\rho_{\epsilon,\delta} = \frac{\delta}{\delta+\epsilon} |0\rangle\langle 0| + \frac{\epsilon}{\delta+\epsilon} |\phi\rangle\langle\phi| \, . \tag{6.1}$$

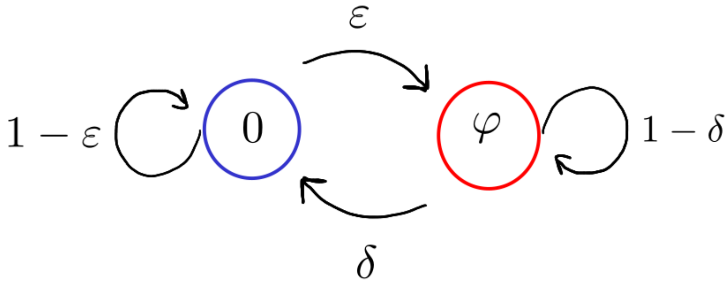


Figure 6.1: The natural generalization to the Change Point problem yields the Stable Source problem here presented as a Hidden Markov Model. We recover the Change Point problem when $\delta = 0$.

However, observe that the machine is not an i.i.d. machine, therefore it does not produce after N iterations the state $\rho_{\epsilon,\delta}^{\otimes N}$. It produces always a state with certain probability and another one with certain probability but following a Markov chain. We can write a density matrix for the output of this Markov source but we don't have the terminology to reduce it as in the i.i.d. case. Specifically it takes 3 symbols to denote that we have an i.i.d. machine producing an output of N states: $1 : \rho, 2 : \otimes$ and $3 : N$. This notwithstanding, the idea of the Markov source machine is very compact itself, it is a simple diagram shown in Figure (6.1).

6.1.3 Optimal chunks for sequential analysis

As we saw, taking samples sequentially can be beneficial for hypothesis testing with very low errors. However, the usual scheme of hypothesis testing with quantum states presupposes that one has N of them given at some point. Considering the benefits of the sequential method then one might wonder if it would be beneficial always.

Given N copies of a state one might ask if it would be more beneficial to make a global measurement than to *order* the set of N states into equal chunks of l states. One interesting thing is that sequential analysis deals with the overall estimation procedure. For a given error ordering the states can imply less number of copies needed. There may be the case that just one chunk of N copies is needed, but that would be for a specific error. In general, one would expect that the larger the chunks the better to distinguish them. However, if the chunks are too big then there might not be enough to stop as the average number of copies needed to stop would be larger than what

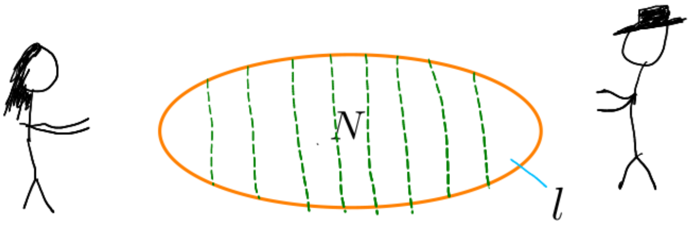


Figure 6.2: Alice gives Bob N states and Bob cuts this set into chunks of size l.

is available. Therefore, the probability of stopping at step r grows with l because we can use quantum collective measurements, reaches a maximum, and decreases with l because the constraint of the total number of available copies. The existence of a maximum l, the size of the chunks points out the benefit of ordering the states. More or less like in Figure (6.2), where Alice sends Bob N copies of a state and then Bob splits the N chunk into smaller chunks of size l and then processes them.

The question asked here is for the optimal size of chunks that would make an arbitrary set of N states into an ordered set of $R = N/l$ states.

6.2 A critique on language for the foundations

We have seen several examples of applications from our results. It has been fruitful to bring the terminology of the change point and sequential analysis to quantum information science. My personal appreciation is that the problems studied here are not only practical but also have a relation to the foundations of quantum mechanics. This might seem not so clear at first but has make me rethink the concept of foundations of quantum mechanics in certain way.

Quantum information theory shows us new approaches to nature. Nevertheless, the strangeness of quantum theory remains. It seems like a void in the understanding of reality. Can we overcome this strangeness within quantum theory? or an extension of it is necessary? More importantly, are the questions that we have asked the ones that will reveal us something new? There is no way to know. Nevertheless, there are things we know and that we can name precisely.

The search to fill the void in our understanding of quantum physics is normally labeled "foundations" [Nor17]. They constitute a set of specific topics that range from nonlocality and entanglement, interpretations of quantum mechanics, quantum effects (quantum Zeno effect, quantum erasure, etc) among others. It is my impression that these topics don't have a clear unified structure. The concept "foundations of physics" presupposes that there are specific topics that are fundamental and others that are accessory.

However, we don't know what are the questions that will reveal us something new about nature. Labeling a set of problems foundational is in my view a metaphysical, or religious way of looking at physics problems. I think there is no such thing as the foundations of physics, specifically, there is not an area that is more foundational than another. Quantum Mechanics itself was born as a way to describe some systems, it was only afterwards that it was taken as a fundamental way to describe nature. I see the concept of "foundations of physics" as a practical concept to designate a specific set of problems that have historically revealed important aspects of nature. For me there are only physics problems. That one problem is foundational is something that is very context-dependent, in this case, the context is the capability of understanding of the physicists. Perhaps quantum information is revealing us fundational insights for our understanding.

Bibliography

[AAB+19] Frank Arute, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph C. Bardin, et al. Quantum supremacy using a programmable superconducting processor. *Nature*, 574(7779):505–510, Oct 2019.

[ABB+05] Antonio Acín, Emili Bagan, Maria Baig, Lluis Masanes, and Ramon Muñoz-Tapia. Multiple-copy two-state discrimination with individual measurements. *Physical Review A*, 71(3):032338, 2005.

[ABB+18] Antonio Acín, Immanuel Bloch, Harry Buhrman, Tommaso Calarco, Christopher Eichler, Jens Eisert, Daniel Esteve, Nicolas Gisin, Steffen J Glaser, Fedor Jelezko, Stefan Kuhr, Maciej Lewenstein, Max F Riedel, Piet O Schmidt, Rob Thew, Andreas Wallraff, Ian Walmsley, and Frank K Wilhelm. The quantum technologies roadmap: a european community view. *New Journal of Physics*, 20:080201, 8 2018.

[Ací16] Antonio Acín. TEDxBarcelona: La segunda revolución cuántica. https://www.youtube.com/watch?v=9kHAKwcRhtY\&t=279s, 2016.

[ACMnT+07] K. M. R. Audenaert, J. Calsamiglia, R. Muñoz Tapia, E. Bagan, Ll. Masanes, A. Acin, and F. Verstraete. Discriminating states: The quantum chernoff bound. *Phys. Rev. Lett.*, 98:160501, Apr 2007.

[Bae13] Joonwoo Bae. Structure of minimum-error quantum state discrimination. *New Journal of Physics*, 15(7):073037, jul 2013.

[Bar12] D. Barber. *Bayesian Reasoning and Machine Learning*. Cambridge University Press, 2012.

[BBG$^+$06] E. Bagan, M. A. Ballester, R. D. Gill, A. Monras, and R. Muñoz Tapia. Optimal full estimation of qubit mixed states. *Phys. Rev. A*, 73:032301, Mar 2006.

[BDF$^+$99] Charles H. Bennett, David P. DiVincenzo, Christopher A. Fuchs, Tal Mor, Eric Rains, Peter W. Shor, John A. Smolin, and William K. Wootters. Quantum nonlocality without entanglement. *Phys. Rev. A*, 59:1070–1091, Feb 1999.

[BFF12] János A. Bergou, Ulrike Futschik, and Edgar Feldman. Optimal unambiguous discrimination of pure quantum states. *Physical Review Letters*, 108:250502, Jun 2012.

[BV04] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, New York, NY, USA, 2004.

[CEM99] J. I. Cirac, A. K. Ekert, and C. Macchiavello. Optimal purification of single qubits. *Phys. Rev. Lett.*, 82:4344–4347, May 1999.

[Che52] Herman Chernoff. A Measure of Asymptotic Efficiency for Tests of a Hypothesis Based on the sum of Observations. *The Annals of Mathematical Statistics*, 23(4):493 – 507, 1952.

[Che98] Anthony Chefles. Unambiguous discrimination between linearly independent quantum states. *Physics Letters A*, 239(6):339–347, 1998.

[CLM$^+$14] Eric Chitambar, Debbie Leung, Laura Mančinska, Maris Ozols, and Andreas Winter. Everything you always wanted to know about locc (but were afraid to ask). *Communications in Mathematical Physics*, 328(1):303–326, May 2014.

[Coo11] G. Cook. *Mobile Robots: Navigation, Control and Remote Sensing*. Wiley, 2011.

[CT12] T.M. Cover and J.A. Thomas. *Elements of Information Theory*. Wiley, 2012.

[CY01] Yi-Xin Chen and Dong Yang. Optimal conclusive discrimination of two nonorthogonal pure product multipartite states through local operations. *Physical Review A*, 64:064303, Nov 2001.

[dar59] *On the Origin of Species by Means of Natural Selection, Or, The Preservation of Favoured Races in the Struggle for Life*. The World's Classics. J. Murray, 1859.

[Day67] J. P. Day. The philosophy of science: A systematic account. by peter caws. (london: Van nostrand. 1965. pp. 354. price 52s. 6d.). *Philosophy*, 42(160):181–183, 1967.

[dL99] Pierre-Simon Marquis de Laplace. *Celestial Mechanics Vol I*. 1799.

[dL14] Pierre-Simon Marquis de Laplace. *Essai philosophique sur les probabilités*. 1814.

[DPP15] Nicola Dalla Pozza and Gianfranco Pierobon. Optimality of square-root measurements in quantum state discrimination. *Physical Review A*, 91, 4 2015.

[Eco17] The Economist. The world's most valuable resource is no longer oil, but data. https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data, 2017.

[Eld03] Y. C. Eldar. A semidefinite programming approach to optimal unambiguous discrimination of quantum states. *IEEE Transactions on Information Theory*, 49(2):446–456, Feb 2003.

[EMV03] Y.C. Eldar, A. Megretski, and G.C. Verghese. Designing optimal quantum detectors via semidefinite programming. *IEEE Transactions on Information Theory*, 49(4):1007–1012, 2003.

[FLS11] R.P. Feynman, R.B. Leighton, and M. Sands. *The Feynman Lectures on Physics, Vol. III: The New Millennium Edition: Quantum Mechanics*. The Feynman Lectures on Physics. Basic Books, 2011.

[ftSP18] European Union Agency for the Space Programme. Galileo reference centre now officially open. https://www.euspa.europa.eu/newsroom/news/galileo-reference-centre-now-officially-open, 2018.

[Fuc95] Christopher A. Fuchs. *Distinguishability and Accessible Information in Quantum Theory*. University of New Mexico, PhD thesis, 1995.

[GB14] Michael Grant and Stephen Boyd. CVX: Matlab software for disciplined convex programming, version 2.1. http://cvxr.com/cvx, March 2014.

[GLM11] Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone. Advances in quantum metrology. *Nature Photonics*, 5:222–229, 4 2011.

[Gri17] D.J. Griffiths. *Introduction to Quantum Mechanics*. Cambridge University Press, 2017.

[Hay01] Masahito Hayashi. Asymptotics of quantum relative entropy from a representation theoretical viewpoint. *Journal of Physics A: Mathematical and General*, 34(16):3413–3419, apr 2001.

[Hel76] C. W. Helstrom. *Quantum Detection and Estimation Theory*. Academic Press, 1976.

[HHH08] A. Hayashi, T. Hashimoto, and M. Horibe. State discrimination with error margin and its locality. *Phys. Rev. A*, 78:012333, Jul 2008.

[HJ12] Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, New York, NY, USA, 2nd edition, 2012.

[HJS$^+$96] Paul Hausladen, Richard Jozsa, Benjamin Schumacher, Michael Westmoreland, and William K. Wootters. Classical information capacity of a quantum channel. *Physical Review A*, 54:1869–1876, Sep 1996.

[Hol73] A.S Holevo. Statistical decision theory for quantum systems. *Journal of Multivariate Analysis*, 3(4):337–394, 1973.

[HW94] Paul Hausladen and William K. Wootters. A ‘pretty good’ measurement for distinguishing quantum states. *Journal of Modern Optics*, 41(12):2385–2390, 1994.

[Int15] Internet Movie Data Base. The Lobster. `https://www.imdb.com/title/tt3464902/`, 2015.

[Jae18] Lars Jaeger. *The Second Quantum Revolution From Entanglement to Quantum Computing and Other Super-Technologies.* Springer, 2018.

[Jay03] E. T. Jaynes. *Probability Theory: The Logic of Science.* Cambridge University Press, 2003.

[JS98] J.V. José and E.J. Saletan. *Classical Dynamics: A Contemporary Approach.* Cambridge University Press, 1998.

[KGW98] I. Kant, P. Guyer, and A.W. Wood. *Critique of Pure Reason.* Cambridge Edition of the Works. Cambridge University Press, 1998.

[Kov15] Boris Kovznjak. Who let the demon out? Laplace and Boscovich on determinism. *Studies in History and Philosophy of Science Part A*, 51:42–52, 2015.

[Lon21] Imperial College London. Developing a quantum inertial navigation system. `https://www.quantumcity.org.uk/project/developing-quantum-inertial-navigation-system,https://www.imperial.ac.uk/centre-for-cold-matter/research/quantum-navigation/`, 2021.

[LTT] Yonglong Li, Vincent Y. F. Tan, and Marco Tomamichel. Optimal adaptive strategies for sequential quantum hypothesis testing. *arXiv:2104.14706.*

[MU05] Michael Mitzenmacher and Eli Upfal. *Probability and Computing: Randomized Algorithms and Probabilistic Analysis.* Cambridge University Press, 2005.

[MVHS$^+$21] Esteban Martínez Vargas, Christoph Hirche, Gael Sentís, Michalis Skotiniotis, Marta Carrizo, Ramon Muñoz Tapia, and John Calsamiglia. Quantum sequential hypothesis testing. *Phys. Rev. Lett.*, 126:180502, May 2021.

[MVPLBB17] Esteban Martínez-Vargas, Carlos Pineda, François Leyvraz, and Pablo Barberis-Blostein. Quantum estimation of unknown parameters. *Phys. Rev. A*, 95:012136, Jan 2017.

[NC11] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, New York, NY, USA, 10th edition, 2011.

[Nor17] Travis Norsen. *Foundations of Quantum Mechanics*. Undergraduate Lecture Notes in Physics. Springer Nature, 2017.

[NPP33] Jerzy Neyman, Egon Sharpe Pearson, and Karl Pearson. Ix. on the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 231(694-706):289–337, 1933.

[PBR12] Matthew F. Pusey, Jonathan Barrett, and Terry Rudolph. On the reality of the quantum state. *Nature Physics*, 8(6):475–478, Jun 2012.

[Pot21] Sean Potter. Nasa's webb telescope launches to see first galaxies, distant worlds. https://www.nasa.gov/press-release/nasas-webb-telescope-launches-to-see-first-galaxies-distant-worlds, 2021.

[SBC$^+$16] Gael Sentís, Emilio Bagan, John Calsamiglia, Giulio Chiribella, and Ramon Muñoz Tapia. Quantum change point. *Physical Review Letters*, 117:150502, Oct 2016.

[SBCMnT13] G. Sentís, E. Bagan, J. Calsamiglia, and R. Muñoz Tapia. Programmable discrimination with an error margin. *Physical Review A*, 88:052304, Nov 2013.

[SCMnT17] Gael Sentís, John Calsamiglia, and Ramon Muñoz Tapia. Exact identification of a quantum change point. *Physical Review Letters*, 119:140506, Oct 2017.

[SCMTB12] G. Sentís, J. Calsamiglia, R. Muñoz-Tapia, and E. Bagan. Quantum learning without quantum memory. *Scientific Reports*, 2(1):708, Oct 2012.

[SHCM18] Michalis Skotiniotis, Ronja Hotz, John Calsamiglia, and Ramon Muñoz-Tapia. Identification of malfunctioning quantum devices. *arXiv:1808.02729*, Aug 2018.

[SN17] J.J. Sakurai and J. Napolitano. *Modern Quantum Mechanics*. Cambridge University Press, 2017.

[SWL$^+$17] Sergei Slussarenko, Morgan M. Weston, Jun-Gang Li, Nicholas Campbell, Howard M. Wiseman, and Geoff J. Pryde. Quantum state discrimination using the minimum average number of copies. *Physical Review Letters*, 118:030502, 2017.

[The15] S. Theodoridis. *Machine Learning: A Bayesian and Optimization Perspective*. .NET Developers Series. Elsevier Science, 2015.

[TNB14] A. G. Tartakovsky, I. V. Nikiforov, and M. Basseville. *Sequential Analysis: hypothesis testing and changepoint detection*. CRC Press, Taylor and Francis, 2014.

[Ume62] Hisaharu Umegaki. Conditional expectation in an operator algebra. IV. Entropy and information. *Kodai Mathematical Seminar Reports*, 14(2):59 – 85, 1962.

[Uni16] European Union. Quantum manifesto. https://qt.eu/app/uploads/2018/04/93056_Quantum-Manifesto_WEB.pdf, 2016.

[Wal73] A. Wald. *Sequential Analysis*. Dover books on advanced mathematics. Dover Publications, 1973.

[Wat18] John Watrous. *The Theory of Quantum Information*. Cambridge University Press, 2018.

[WM09] Howard M. Wiseman and Gerard J. Milburn. *Quantum Measurement and Control*. Cambridge University Press, 2009.

[WW48] A. Wald and J. Wolfowitz. Optimum character of the sequential probability ratio test. *The Annals of Mathematical Statistics*, 19(3):326–339, 1948.

[ZFY06] Chi Zhang, Yuan Feng, and Mingsheng Ying. Unambiguous discrimination of mixed quantum states. *Physics Letters A*, 353(4):300–306, 2006.