New bounds for information complexity and quantum query complexity via convex optimization tools

Thèse présentée en vue de l'obtention du grade de

Docteur en Sciences de l'Ingénieur et Technologie

Auteur:
Mathieu Brandeho

Superviseur:
Jérémie Roland

Université libre de Bruxelles
École polytechnique de Bruxelles
Centre for Quantum Information and Communication

Année académique 2017-18

Mathieu Brandeho
Université libre de Bruxelles
École polytechnique de Bruxelles
50 av. F.D. Roosevelt - CP165/59
1050 Bruxelles
Belgique
Email: mbrandeh@ulb.ac.be

Jury: **Nicolas J. Cerf**, Président du jury, ULB

Stacey Jeffery, Centrum Wiskunde & Informatica, Netherlands

Sophie Laplante, Université Paris Diderot Paris 7, France

Serge Massar, ULB

Stephano Pironio, ULB

Jérémie Roland, Promoteur, ULB

Acknowledgments

Je souhaite remercier ma famille. Mon frère, ma sœur et ma mère, bien qu'absent puisqu'ils ont décidé de migrer vers le soleil, étaient présent moralement.

Bien sûr je remercie à Jérémie, mon superviseur, sans qui cette thèse n'aurait pu être possible. Pour sa présence, pour sa patience et sa clairvoyance. Pour ses années où il m'a financé, aider, encouragé, poussé et motivé.

Merci à Mathieu de m'avoir partagé ses passions pour l'archéologie, la Belgique, le taekwondo, l'ornithologie, le rock, le whisky, ses voyages, etc. Il ne fallait pas te donner autant de mal pour moi.

À toi Anaëlle dont les critiques blessantes et humiliantes durant toutes ces années ont fini d'achever mon petit cœur sensible.

Levon pour m'avoir laissé gagner au Badminton, et Pauline pour les repas gratuits. Zacharie qui est une source inépuisable de nourriture. Atul, un collègue, mais également un ami qui a donné de sa personne pour la correction de cette thèse. Shantanav et Leonardo à qui je dois deux ou trois bières, mais je ne sais plus trop combien, ni à qui. Zoé avec qui les pauses étaient toujours agréables, et parfois trop longues. Daria pour les bons moments passés ensemble à flâner et papillonner dans Bruxelles.

Je remercie le service du QuIC de m'avoir accueilli pendant ces cinq années et plus. Entre autres, je remercie Nicolas, Ognyan, Evgueni, Raúl, Christos, Joachim et Luc qui ont rendu ce séjour agréable.

Je n'oublie pas Pascale et les longues discussions à la pause café, dans le but précis de l'amadouer pour obtenir mes remboursements. Tandis qu'elle me répondait dans un dialecte rudimentaire.

Thank you to all the jury members for reading and reviewing this thesis.

Abstract

This thesis brings together three works on information complexity and quantum query complexity. These different complexities have in common mathematical tools to study them, i.e. optimization problems.

The first two works concern quantum query complexity, generalizing the important following result: In the article [LMR+11] the authors manage to characterize quantum query complexity, using the adversary method, a semi-definite program introduced by A. Ambainis in [Amb00]. However, this characterization is restricted to discrete time models with a bounded error. Thus in the first work, we generalize their results to continuous time models. While the second work is an approach, not completed, to characterize the quantum query complexity for the exact case and for an unbounded error.

In the **first work**, to characterize quantum query complexity for discrete-time models, we adapt the demonstration of the discrete-time model, by constructing a universal adiabatic quantum query algorithm. The principle of this algorithm is based on the adiabatic theorem [BF28], and on an optimal solution of the dual of the adversary method. Note that the analysis of the running time of our adiabatic algorithm is based on a proof that does not require a gap in the spectrum of the Hamiltonian.

In the **second work**, we want to characterize quantum query complexity for an unbounded error and exact case. To this end, we start from the adversary method and improve it with a Lagrangian mechanics approach, in which we build a Lagrangian indicating the number of queries necessary to move in the phase space. Thus we can define the “query action”. As this Lagrangian is expressed as a semi-definite program, its classical analysis via Euler-Lagrange equation requires the envelope theorem, a result from mathematical economics.

The **last work** concerns information complexity (and communication complexity by extension) to simulate non-local correlations. More precisely, the amount of information (according to Shannon) that two parts must share to obtain these correlations. For this purpose, we define a new complexity, called zero information complexity IC_0, via the zero communication model. This new complexity can be expressed as an optimization problem which makes it interesting. For CHSH correlations, we solved this optimization problem for the one-way scenario where we retrieve a known result. In the two-way case, we find a numerical bound and solve a relaxed form of IC_0 that is a new result.

Titre

Nouvelles bornes pour la complexité d'information et la complexité en requête quantique grâce aux outils d'optimisation convexe.

Résumé

Cette thèse rassemble trois travaux sur la complexité d'information et sur la complexité en requête quantique. Ces domaines d'études ont pour points communs les outils mathématiques pour étudier ces complexités, c'est-à-dire les problèmes d'optimisation.

Les deux premiers travaux concernent le domaine de la complexité en requête quantique, en généralisant l'important résultat suivant: dans l'article [LMR+11], leurs auteurs parviennent à caractériser la complexité en requête quantique, à l'aide de la méthode par adversaire, un programme semi-définie positif introduit par A. Ambainis dans [Amb00]. Cependant, cette caractérisation est restreinte aux modèles à temps discret, avec une erreur bornée. Ainsi, le premier travail consiste à généraliser leur résultat aux modèles à temps continu, tandis que le second travail est une démarche, non aboutie, pour caractériser la complexité en requête quantique dans le cas exact et pour erreur non bornée.

Dans ce **premier travail**, pour caractériser la complexité en requête quantique aux modèles à temps discret, nous adaptons la démonstration des modèles à temps discret, en construisant un algorithme en requête adiabatique universel. Le principe de cet algorithme repose sur le théorème adiabatique [BF28], ainsi qu'une solution optimale du dual de la méthode par adversaire. À noter que l'analyse du temps d'exécution de notre algorithme adiabatique est basée sur preuve qui ne nécessite pas d'écart dans le spectre de l'Hamiltonien.

Dans le **second travail**, on souhaite caractériser la complexité en requête quantique pour une erreur non bornée ou nulle. Pour cela on reprend et améliore la méthode par adversaire, avec une approche de la mécanique lagrangienne, dans laquelle on construit un Lagrangien indiquant le nombre de requêtes nécessaires pour se déplacer dans l'espace des phases, ainsi on peut définir l'"action en requête". Or ce lagrangien s'exprime sous la forme d'un programme semi-defini, son étude classique via les équations d'Euler-Lagrange nécessite l'utilisation du théorème de l'enveloppe, un puissant outils d'économathématiques.

Le **dernier travail**, plus éloigné, concerne la complexité en information (et par extension la complexité en communication) pour simuler des corrélations non-locales. Ou plus précisement la quantitié d'information (selon Shannon) que doive s'échanger deux parties pour obtenir ses corrélations. Dans ce but, nous définissons une nouvelle complexité, denommée la zero information complexity IC_0, via le modèle sans communication. Cette complexité a l'avantage de s'exprimer sous la forme d'une optimization convexe. Pour les corrélations CHSH, on résout le problème d'optimisation pour le cas à une seule direction où nous retrouvons un résultat connu. Pour le scénario à deux directions, on met numériquement en évidence la validité de cette borne, et on résout une forme relaxée de IC_0 qui est un nouveau résultat.

List of publications

- *A Universal Adiabatic Quantum Query Algorithm* Mathieu Brandeho and Jérémie Roland, 10th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2015), pages 163-179, doi:10.4230/LIPIcs.TQC.2015.163, arXiv:1409.3558.

Contents

Chapter 1

Introduction

Computational complexity is an interesting subfield of computer science. Since a Turing machine can simulate another one in polynomial time, polynomial classes such as **P** seem to belong to Platonic idealism: independent of time, of space, of technology. Then quantum mechanics decided to be the troublemaker. In other words, what happens if the universe and its physical laws, seen as a computer, are more powerful than a Turing machine? Many hypothesis have been proposed including the following,

> Quantum Church–Turing thesis [KLMK07]: “A quantum Turing machine can efficiently simulate any realistic model of computation.”

This thesis does not provide an answer to this statement, nevertheless, this hypothesis resumes that possibilities of physical world and quantum computation are still unknown. Topics of this work on **quantum query complexity** and **information complexity** both of which are subfields at the intersection of quantum mechanics and computational complexity. The objective of the former can be summarized as how quantum mechanics can improve algorithms. The latter addresses how information complexity helps to better understand Bell’s inequalities, and hence quantum mechanics.

A. Quantum query complexity

Query complexity is a natural lower bound of time complexity where instead of counting all actions done by an algorithm, we count a specific action called query. The query is the action of obtaining data from the input, typically one bit. Obviously, query complexity is limited by the length of the input but is often tight for functions that can be easily evaluated once their input is known.

Quantum query complexity first appears in the literature in the context of quantum algorithms like Shor’s algorithm [Sho97] and Grover’s algorithm [Gro96].To better understand quantum query complexity $Q(f)$ of a function f, several lower bound methods have been developed such as the adversary method originally introduced by A. Ambainis [Amb00]. This method is based on a semi-definite program where we optimize an objective value with a matrix M assigning weights to pairs of inputs. Later Høyer et al. showed [HLS07] that using negative weights also provides a lower bound, which is stronger for some functions. A series of works [Rei09, Rei11, RS12] then led to the breakthrough result that this *generalized* adversary method, which we will simply call adversary method, actually characterizes the quantum query complexity of any function f with Boolean output and binary input alphabet. This is shown by constructing a tight algorithm based on the dual of the semi-definite program corresponding to the adversary method. Finally, Lee et al. [LMR+11] have generalized this result to the quantum query complexity of state conversion where instead of computing a function $f(x)$, we convert a quantum state $|\rho_x\rangle$ into another quantum state $|\sigma_x\rangle$ for each input x.

All these results have been obtained in a discrete-time model where each query corresponds to applying a unitary operator $\mathcal{O}_x$, denoting the oracle operator. In this model, an algorithm is the concatenation of input-independent unitary operators $U_1, U_2, \ldots, U_T$, interleaved with the oracle operator $\mathcal{O}_x$. Another natural model is the continuous-time (or Hamiltonian-based) model where the algorithm evolves under Schrödinger's equation and hence an algorithm is completely described by a Hamiltonian. The oracle's action is represented by an *oracle* Hamiltonian $H_{\mathcal{Q}}(x)$ and input-independent evolution is represented by a *driver* Hamiltonian $H_D(t)$, their sum $H_x(t)$ characterizes a quantum query algorithm in the continuous-time model.

$$H_x(t) = H_D(t) + H_{\mathcal{Q}}(x). \tag{1.1}$$

These two models are related as the oracle operator $\mathcal{O}_x$ can be simulated by the oracle Hamiltonian H_x with a finite amount of time. This fact implies that the continuous-time model is at least as powerful as the discrete-time model. In the other direction, Cleve et al. [CGM09] have shown that the discrete-time model can simulate the continuous-time model up to at most a sub-logarithmic factor with a bounded error, which implies that continuous and discrete-time models are equivalent up to a sub-logarithmic factor. Later, Lee et al. [LMR+11] improved this result to a full equivalence of both models. They show that the fractional query model, an intermediate model defined in [CGM09], is equivalent to the continuous-time model. Note that this intermediate model is also lower bounded by the adversary method. Hence all these models are characterized by this same method.

A universal adiabatic quantum query algorithm

Even though these results imply that the continuous-time quantum query complexity is characterized by the adversary method, they do not provide an explicit continuous quantum query algorithm. The one obtained from the discrete-time algorithm by replacing each unitary oracle operator by the application of the Hamiltonian oracle for a constant amount of time is an exception. The evolution of this algorithm, however, involves many discontinuities this is not satisfying from the point of view of physics where Hamiltonians are smooth.

Thereby, the **first work** (Chapter 7) provides a direct proof that adversary method characterizes quantum query complexity but with several additional motivations. First, our algorithm is simple and easy to understand. Second, continuous-time models are more suitable for analysis. We give a new continuous-time quantum query algorithm for any state conversion problem based on an adiabatic theorem [FGGS00] where the Hamiltonian varies slowly. The soundness of the adiabatic evolution used in our algorithm relies on a lemma from Avron and Elgart [AE99a], which does not require the usual gap condition but only a weaker spectral condition (originally introduced to study atoms in quantized radiation field). To the best of our knowledge, it is the first time that such an adiabatic theorem without a gap condition is used in the context of quantum computation. We also provide an original proof of the adversary method for continuous-time models based on Ehrenfest's theorem. From this we conclude that the quantum query complexity of any state conversion problem is characterized by the adversary bound.

A new lower bound: the adversary action

The quantum query complexity has been characterized for a bounded error only, excluding the exact case and the unbounded error case. For example, we know that the adversary method is not tight for **OR** function in the exact case. In order to fill these gapes, the successful adversary method has been generalized to the multiplicative adversary method by R. Špalek [Špa08].

In the **second work** (Chapter 8) we introduce another generalization of the multiplicative adversary method called the adversary action. This method is based on [BSS03], where the authors constructed a semi-definite program that checked if a quantum state evolution is feasible with only one query. From this semi-definite program we construct the query Lagrangian that gives the infinitesimal number of queries for a quantum state (position) to evolve in a specific direction (momentum). Hence, integrating the query Lagrangian over a path in the "phase space" gives the amount of queries needed to follow this path with a quantum query algorithm. We obtain the adversary action by minimizing over all paths.

The adversary method is a semi-definite problem that defines a norm on a "position space". The norm of the difference between the initial and the final states defines a distance. And this distance lower bounds the query cost to travel from the initial state to the final state. The query Lagrangian, also a semi-definite problem, also defines a norm but extended to the phase space. Adding the "momentum" allows to obtain a more precise method where we are not only interested by *where* we want to go but also *how*.

Afterwards we give several properties of the query Lagrangian and prove explicitly that the action adversary subsumes both the adversary method and the multiplicative version. Finally, we apply the Euler-Lagrange equation to the action adversary to obtain the equation for the query Lagrangian. As our Lagrangian is a semi-definite program, we use the Envelope theorem [Mir71], an important result of mathematical economics, to derive our semi-definite program. More precisely, we use a recent version of this theorem [MS02] where the feasible choice is arbitrary. Thus we add the action adversary method to the tools for studying quantum query complexity.

B. Information complexity

The communication model introduced by A. Yao in [Yao79] is a model where two parties communicate to collaboratively perform a common task. From this model, the communication complexity is defined as the necessary communication that must be exchanged between the two parties, called Alice and Bob, to perform this task. In the one-way scenario where only Alice communicates to Bob, information theory [Sha48] allows one to characterize the communication complexity. Indeed, according to the *Slepian-Wolf coding theorem* [SW73] and the result of Braverman and Rao [BR11] the communication complexity is related to the compression of Alice's message M conditionally to Bob's input Y,

$$H(M|Y) = H(M,Y) - H(Y).$$

Nonetheless, applying information theory tools to characterize the two-way scenario is more complicated, since the compression must be interactive. This is the starting point of the information complexity model. In this model introduced in [CWY01, BYJKS04, BBCR10], two information complexities are defined depending on Alice's input X, Bob's input Y and the communication M. The internal information complexity is the information that Alice and Bob learn from each other, i.e.

$$I(M;X|Y) + I(M;Y|X).$$

The external information complexity is the information that a third party learns about Alice and Bob's inputs,

$$I(M;XY).$$

These two information complexities naturally define a lower bound on the communication complexity and important compression results have been found. M. Braverman showed how to compress a protocol with i internal information cost to a protocol with $2^{O(i)}$ communication cost [Bra12]. Barak *et al.* showed how to compress a protocol with i internal information cost and c communication cost to a protocol with $\tilde{O}(\sqrt{ic})$ communication cost, and $\tilde{O}(i)$ for a product input distribution [BBCR10].

Lower bound for simulating correlations

The communication model can be generalized to study correlations where at the end of the communication Alice and Bob generate respectively an output a and b such that we obtain a conditional probability distribution

$$\mathrm{p}(a,b|x,y),$$

according to their input x and y. Thus, the information complexity model allows us to study the information cost that Alice and Bob must share to generate specific correlations. This analysis is strongly related to Bell's inequality [Bel64], where two quantum states are separated then measured with outcomes a and b. In this experiment, if the observed correlations have non null information complexity then the locality assumption is violated.

In the **last work** (Chapter 9) we analyze the external information complexity applied to CHSH correlations with a new lower bound method, the zero information complexity denoted IC_0.

This method comes from the zero communication model where Alice and Bob cannot communicate. Instead they are independently allowed to abort the protocol to artificially raise the success probability. The probability that Alice and Bob don't abort is called the efficiency, and is related to the communication cost [Mas01, BHMR03, LLR12]. Similarly, in the zero communication model we define IC_0 and prove that it is a natural lower bound of the external information complexity. As the title suggests, IC_0 is an optimization program. We also derive two other lower bound methods: $\overline{\mathrm{IC}}_0$, a relaxed form of IC_0, and $\mathrm{IC}_0^{\rightarrow}$, a special case of IC_0 where only Alice can abort. These methods have a natural order,

$$\overline{\mathrm{IC}}_0 \leq \mathrm{IC}_0 \leq \mathrm{IC}_0^{\rightarrow}.$$

Finally, we apply these three new methods to CHSH correlations. For $\mathrm{IC}_0^{\rightarrow}$ we retrieve a known result [RS09]. For the relaxed form $\overline{\mathrm{IC}}_0$ we find a new lower bound. At last for IC_0 its analytic solution is left open and a numerical analysis is provided.

C. Structure

This thesis is divided in two parts. It starts with a brief review of general mathematical tools of interest for quantum computation, followed by the three aforementioned works on information complexity and quantum query complexity.

In Chapter 2 we introduce the basic mathematics used in this thesis, mainly to agree on notation and concepts. Chapter 3 is devoted to quantum mechanics and quantum computation, in particular to adiabatic algorithms. Chapter 4 is dedicated to quantum query complexity while Chapter 5 covers communication complexity, both with lower bound methods associated to those complexities. In Chapter 6, we present convex optimization as well as important results in this field, such as Slater's theorem, the KKT conditions and the Envelope theorem.

In Chapter 7 we introduce our universal quantum query algorithm, implying the characterization of quantum query complexity for bounded error in the continuous-time model. Chapter 8 presents our adversary action supplemented by its properties and relations with other lower bound methods. Finally in Chapter 9 we derive the zero information complexity and apply this new method to CHSH correlations where we obtain new results.

Part I

Mathematical tools for quantum computation

Chapter 2

Mathematical background

This chapter introduces basic but still important tools in this thesis such as linear algebra, basic topology, analysis, convexity and set theory. Although rigorous, this chapter is not a course but an aid to read this thesis.

Note that we only work with real or complex finite dimensional vector spaces in this thesis, thus we don't introduce specific properties of non-finite dimensional vector space. Some symbols, sub-scripts and upper-scripts are sometimes omitted when no confusion is possible.

2.1 Linear algebra

$\mathbb{N}$ is the set of natural numbers. $\mathbb{K}$ describe a field. $\mathbb{K}$ is either the set of real numbers $\mathbb{R}$, or the set of complex numbers $\mathbb{C}$. $\mathbb{K}^n$ is the direct product of n fields $\mathbb{K}$. $\mathcal{M}_{m,n}(\mathbb{K})$ is the set of m-by-n matrices with entries in $\mathbb{K}$. $\mathcal{M}_n(\mathbb{K})$ is the set of n-squared matrices with entries in $\mathbb{K}$. The **transpose** of an m-by-n matrix A is an n-by-m matrix $A^\top$ defined by $A^\top[x,y] = A[y,x]$. The **conjugate transpose** of an m-by-n matrix A with complex entries is an n-by-m matrix A^* defined by $A^*[x,y] = \overline{A[y,x]}$. A real matrix A is **symmetric** if $A^\top = A$. A complex matrix A is **Hermitian** if $A^* = A$. The **trace** of an n-by-n square matrix A is the sum of all diagonal terms, $\mathrm{tr}A = \sum_{i=1}^n A_{ii}$.

2.1.1 Hilbert space

A $\mathbb{K}$**-vector space** is a triplet $(\mathcal{V}, +, \times)$ where $\mathcal{V}$ is a set, $\mathbb{K}$ is a field, the vector addition $+ : \mathcal{V} \times \mathcal{V} \to \mathcal{V}$ and $\times_\mathcal{V}$ the scalar product $\times : \mathbb{K} \times \mathcal{V} \to \mathcal{V}$, that satisfies eight axioms. For all $u, v, w \in \mathcal{V}$ and $a, b \in \mathbb{K}$,

- **(Associativity of +)** $u + (v + w) = (u + v) + w$,
- **(Commutativity of +)** $u + v = v + u$,
- **(Identity of +)** there exists an element $0 \in \mathcal{V}$, such as $v + 0 = 0 + v = v$,
- **(Inverse of +)** for all $v \in \mathcal{V}$, there exists an element $-v$ such as $v + (-v) = 0$,
- **(Identity of ×)** there exists an element $1 \in \mathbb{K}$, such as for all $v \in V,\ 1.v = v$,
- **(Compatibility)** $a \times (bv) = (ab) \times v$,
- **(Distributivity of +)** $a \times (u + v) = a \times u + b \times v$,

- **(Distributivity of** $\times$**)** $(a+b)\times v = a\times v + b\times v$.

An **inner product space** is a $\mathbb{K}$-vector space with an additional structure called an inner product $\langle\,\cdot\,,\cdot\rangle : \mathcal{V}\times\mathcal{V}\to\mathbb{C}$ that satisfies four axioms. For all $u, v, w \in \mathcal{V}$ and $a\in\mathbb{K}$,

- **(Conjugate)** $\langle u, v\rangle = \overline{\langle v, u\rangle}$,
- **(Linearity)** $\langle w, u + a.v\rangle = \langle w, u\rangle + a.\langle w, v\rangle$,
- **(Positive-definiteness)** $\langle u, u\rangle$ is real positive,
- **(Zero vector)** $\langle u, u\rangle = 0 \quad\Leftrightarrow\quad u = 0$.

$\mathbb{R}^n$, $\mathbb{C}^n$, $\mathcal{M}_n(\mathbb{R})$ and $\mathcal{M}_n(\mathbb{C})$ are inner product spaces with the, respectively, inner product

- $\langle u, v\rangle_{\mathbb{R}^n} = \sum_{i=1}^n u_i v_i$,
- $\langle u, v\rangle_{\mathbb{C}^n} = \sum_{i=1}^n \overline{u_i} v_i$,
- $\langle U, V\rangle_{\mathcal{M}_n(\mathbb{R})} = \mathrm{tr}(U^\top V)$,
- $\langle U, V\rangle_{\mathcal{M}_n(\mathbb{C})} = \mathrm{tr}(U^* V)$.

The subscript of an inner product is omitted when there is no confusion.
A **Hilbert space** is a complete inner product space. (See Section 2.2 for the definition of 'complete'.)

2.1.2 Linear operator

Let $\mathcal{V}$ and $\mathcal{W}$ be two $\mathbb{K}$-vector spaces. A **linear map** $f : \mathcal{V}\to\mathcal{W}$ is a function such that for all $u, v\in\mathcal{V}$, and $a\in\mathbb{K}$ we have

$$f(u + a.v) = f(u) + a.f(v).$$

In the special case $\mathcal{V} = \mathcal{W}$, f is called a **linear operator**. Moreover for $\mathcal{V} = \mathbb{K}^n$, a linear operator can be represented by a matrix M in $\mathcal{M}_n(\mathbb{K})$ where $v\mapsto Mv$. An **affine map** ϕ is a generalization represented by a matrix M and a vector w, $\phi(v) : v\mapsto Mv + w$.
Let A be a linear operator on inner product space $\mathcal{V}$. A is **self-adjoint** if for all $u, v\in\mathcal{V}$

$$\langle u, Av\rangle = \langle Au, v\rangle.$$

For $\mathcal{V} = \mathbb{R}^n$, A is self-adjoint if and only if $A^\top = A$. For $\mathcal{V} = \mathbb{C}^n$, A is self-adjoint if and only if $A^* = A$. Let v be a vector in $\mathbb{K}^n$. Its **co-vector** v^* is the linear map $\langle v, \cdot\rangle$. Let $\mathcal{S}^n(\mathbb{K})$ be the **set of self-adjoint matrices** in $\mathcal{M}_n(\mathbb{K})$. A self-adjoint matrix $S\in\mathcal{S}^n(\mathbb{K})$ is **positive semi-definite** on $\mathcal{M}_n(\mathbb{K})$, if for all vectors on $v\in\mathbb{K}^n$

$$\langle vv^*, S\rangle_{\mathcal{M}_n(\mathbb{K})} \geq 0.$$

We define $S\in\mathcal{S}_+^n(\mathbb{K})$ to be the set of all positive semi-definite matrices in $\mathcal{M}_n(\mathbb{K})$.
Let $\mathcal{H}$ be a Hilbert space a **unitary operator** U is an operator that preserves the inner product for all u, $v\in\mathcal{H}$

$$\langle Uu, Uv\rangle_{\mathcal{H}} = \langle u, v\rangle_{\mathcal{H}}.$$

2.1.3 Gram matrix

The **Gram matrix** G of a set of the vectors $(v_i)_{i\in I}$ in an inner product space, is a self-adjoint $|I|$-square matrix defined by, $G[i,j] = \langle v_i, v_j \rangle$, also noted

$$G = \mathrm{Gram}(v_i : i \in I).$$

Indeed a matrix $G \in \mathcal{M}_n(\mathbb{K})$ is a Gram matrix if, there exists a set of the vectors in $\mathbb{K}^n$ where G is their Gram matrix. Moreover Gram matrices have several properties.

Fact 2.1.1. Let $M \in \mathcal{M}_n(\mathbb{C})$. The following properties are equivalent:

(a) M is positive semi-definite,

(b) M is a Gram matrix.

Let $(v_i)_i$ and $(w_i)_i$ be two sets of the vectors. These sets are **unitary equivalent** if there exists a unitary operator U such that $w_i = Uv_i$. Note that from the property of a unitary operator, two unitary equivalent sets of the vectors have the same Gram matrix. The following fact ensures that the reciprocal is true.

Fact 2.1.2. Let $(v_i)_i$ and $(w_i)_i$ be two sets of the vectors in $\mathbb{C}^n$ with the same Gram matrix. Therefore there exists a unitary operator U such that $w_i = Uv_i$ for all i.

2.1.4 Norm, distance and fidelity

A **norm** on a $\mathbb{K}$-vector space $\mathcal{V}$ is a positive function $\|\cdot\| : \mathcal{V} \to \mathbb{R}_+$ that satisfies three axioms. For all $u, v \in \mathcal{V}$ and $a \in \mathbb{K}$,

- **(Sub-additivity)** $\|u + v\| \leq \|u\| + \|v\|$,
- **(Absolutely homogeneous)** $\|a.u\| = |a|.\|u\|$,
- **(Zero vector)** if $\|u\| = 0$, then $u = 0$.

A vector space $\mathbb{K}^n$ with an inner product has a natural norm, $\|\cdot\|_{\mathbb{K}^n} = \sqrt{\langle\,\cdot\,,\cdot\,\rangle_{\mathbb{K}^n}}$.
A matrix space $\mathcal{M}_n(\mathbb{K})$ has several norms. We consider two important norms. Let A and B be n-by-n matrices

- Operator norm: $\|A\| = \max_{v\in\mathbb{K}^n} \frac{\|Av\|_{\mathbb{K}^n}}{\|v\|_{\mathbb{K}^n}} = \max_{u,v\in\mathbb{K}^n} \frac{\langle u,Av \rangle}{\|u\|_{\mathbb{K}^n}.\|v\|_{\mathbb{K}^n}}$,
- Trace norm: $\|A\|_{\mathrm{tr}} = \max_{B\in\mathcal{M}_n(\mathbb{K})} \frac{\langle A,B\rangle}{\|B\|}$.

There is a relation between these norms and the inner product for a matrix space $\mathcal{M}_n(\mathbb{K})$.

Lemma 2.1.3. *Let A and B be n-by-n matrices. We have* $\langle A, B\rangle \leq \|A\|_{\mathrm{tr}} \cdot \|B\|$.

A vector with a norm equal to one is a **unit vector**. A Gram matrix obtained from a set of unit vectors $(u_i)_i$ is a **unitary Gram matrix**, in particular all its diagonal entries are equal to one. A positive semi-definite matrix with a trace norm equal to one is called a **density matrix** in quantum mechanics.

A **distance** on a $\mathbb{K}$-vector space $\mathcal{V}$ is a positive function $d(\cdot,\cdot) : \mathcal{V} \times \mathcal{V} \to \mathbb{R}_+$ that satisfies three axioms. For all $u, v, w \in \mathcal{V}$,

- **(Identity)** $d(u,v) = 0 \Leftrightarrow u = v$,
- **(Symmetry)** $d(u,v) = d(v,u)$,
- **(Triangle inequality)** $d(u,v) + d(v,w) \leq d(u,w)$.

From the Trace norm we can derive a distance between two positive semi-definite matrices ρ and σ, called the **total distance** and defined as

$$\mathcal{D}(\rho,\sigma) = \|\rho - \sigma\|_{\mathrm{tr}}.$$

The next quantity, **fidelity**, is not a distance but can define a measure between two positive semi-definite matrices ρ and σ. This measure is symmetric and defined by

$$\mathcal{F}(\rho,\sigma) = \mathrm{tr}\sqrt{\sqrt{\rho}\,\sigma\sqrt{\rho}}.$$

The following theorem highlights the close relation between the distance and the fidelity.

Theorem 2.1.4. *[FvdG99] For any density matrices ρ, σ, we have*

$$1 - \mathcal{D}(\rho,\sigma) \leq \mathcal{F}(\rho,\sigma) \leq \sqrt{1 - \mathcal{D}^2(\rho,\sigma)}.$$

The **Hadamard product** of two n-by-m matrices is defined by the entry-wise product of these matrices, $(A \circ B)[i,j] = A[i,j] \cdot B[i,j]$. The Hadamard product has a natural property with an inner product of a matrix space.

Lemma 2.1.5. *Let A, B and C be n-by-n matrices. We have*

- $\langle A \circ C, B \rangle_{\mathcal{M}_n(\mathbb{R})} = \langle A, B \circ C^\top \rangle_{\mathcal{M}_n(\mathbb{R})}$,
- $\langle A \circ C, B \rangle_{\mathcal{M}_n(\mathbb{C})} = \langle A, B \circ C^* \rangle_{\mathcal{M}_n(\mathbb{C})}$.

Claim 2.1.6. Let A and B be two positive semi-definite matrices. $A \circ B$ is positive semi-definite.

The **Hadamard product fidelity** is introduced in [LR12] to characterize the output condition of quantum query problems. As the usual fidelity, the Hadamard product fidelity compares two semi-definite positive matrices ρ and σ and is defined as

$$\mathcal{F}_H(\rho,\sigma) = \min_{u:\|u\|=1} \mathcal{F}(\rho \circ uu^*, \sigma \circ uu^*). \tag{2.1}$$

We similarly define the **Hadamard product distance** of two semi-definite positive matrices ρ and σ as

$$\mathcal{D}_H(\rho,\sigma) = \max_{u:\|u\|=1} \mathcal{D}(\rho \circ uu^*, \sigma \circ uu^*). \tag{2.2}$$

Like the distance and the fidelity, the Hadamard product fidelity and the Hadamard product distance are closely related.

Corollary 2.1.7. *For any positive semi-definite matrices ρ and σ, we have*

$$1 - \mathcal{D}_H(\rho,\sigma) \leq \mathcal{F}_H(\rho,\sigma) \leq \sqrt{1 - \mathcal{D}_H^2(\rho,\sigma)}.$$

Let $A : V \to W$ be a linear operator with V and W two normed vector spaces. The **graph norm** $\|\cdot\|_A$ of $v \in V$ is defined as $\|v\|_A = \|v\| + \|Av\|$.

Definition 2.1.8 (γ_2 norm)**.** Let S be a finite set and A be a $|S|$-square matrix. The γ_2 norm of A is defined as

$$\begin{aligned}\gamma_2(A) &= \min_{\substack{m\in\mathbb{N}\\ \boldsymbol{u}_x,\boldsymbol{v}_y\in\mathbb{C}^m}} \left\{ \max_{x\in S} \max\left\{\|\boldsymbol{u}_x\|^2, \|\boldsymbol{v}_x\|^2\right\} \middle| \forall\, x,y\in S,\ A_{x,y}=\langle u_x, v_y\rangle \right\},\\ &= \max_{\substack{\boldsymbol{u}\colon \|\boldsymbol{u}\|=1\\ \boldsymbol{v}\colon \|\boldsymbol{v}\|=1}} \|A\circ \boldsymbol{u}\boldsymbol{v}^*\|_{\mathrm{tr}}.\end{aligned}$$

Fact 2.1.9. [LR12] Let A and B be two n-square matrices and their Hadamard product $A\circ B$. We have $\|A\circ B\| \le \gamma_2(A).\|B\|$.

2.2 Topology

A **topological space** is an ordered couple (X,τ) where X is a set and τ a family of subsets of X satisfying the following properties:

(a) the empty set $\emptyset$ and X are in τ,

(b) for every family of subsets $O_i\tau$, their union $\bigcup_i O_i \in \tau$,

(c) for every finite family of subsets $(O_i)_i$, their intersection $\bigcap_i O_i \in \tau$.

Every subset in τ is called an **open set**. $A^{\complement}$ the **complement** of a set A in a topological space X is the set of all point in X but not in A. A set is **closed**, if its complement is open.

Let V be a subset of X and x be in V. V is a **neighborhood of** x if there exists an open set O such that $x\in O\subset V$. Let $\mathcal{V}$ be a normed vector space and r be a real number. We define the set $B_r(x)$ as

$$B_r(x) = \left\{y\in\mathcal{V} : \|y-x\|\le r\right\}.$$

The set $B_r(x)$ is a **ball** with center the point x. It is a natural neighborhood of x.
The **closure** of a set A is defined as

$$\textbf{cl}\ A = \bigcap_{\substack{\text{F closed}\\ A\subset F}} F.$$

It is the smallest closed set containing A.
The **interior** of a set A is defined as

$$\textbf{int}\ A = \bigcup_{\substack{\text{O open}\\ O\subset A}} O.$$

It is the biggest open set inside A.
The **boundary** of a set A is defined as **bd** A = **cl** $A\setminus$ **int** A or **bd** $A =$ **cl** $A\cap$ **cl** $(A^{\complement})$.

From the point of view of sequences, we can have a better understanding of open/closed sets. A **sequence** (x_n) is an ordered collection of points in a set. A point x of a topological space (X,τ) is a **limit of a sequence** (x_n), if for every neighborhood V of x, there exists N such that for all $n>N$, $x_n\in V$. A sequence (x_n) is **convergent** if it has a limit x. A convergent sequence (x_n) with a limit x is denoted by $x_n\to x$. A topological space (X,τ) is **complete** if

all convergent sequences $(x_n \in X)$ have their limit in X.
Let O be an open set. For each point $x \in O$ there exists V a neighborhood of x such that $x \in V \subset O$. In other words for all $x \in O$ there exists a non trivial sequence $x_n \to x$ with $x_n \in O \setminus \{x\}$. Let F be a closed set. If a sequence (x_n) inside F has a limit x then $x \in F$.

Since in this thesis we are restricted to real and complex finite-dimensional spaces, we don't define compacity directly but we will use the following Theorem.

Theorem 2.2.1 (Heine–Borel theorem)**.** *[Sun15] Let $(\mathbb{R}^n, \tau)$ be a topological space and $S \subset \mathbb{R}^n$, therefore the two following properties are equivalent:*

(a) S is closed and bounded,

(b) S is compact.

If (x_n) is a sequence in a **compact set** then we can extract a sub-sequence with a limit.

2.3 Analysis

Let $f : X \to Y$ be a function. Its **domain** X is **dom** f and its **range** Y is **range** f. A **multi-valued function** or **correspondence** C between two sets X and Y is defined by the functions $C : X \to \mathcal{P}(Y)$, where $\mathcal{P}(Y)$ is the power set of Y. A **selection** of a correspondence C is a function $s : X \to Y$ such that for all $x \in X$, $s(x) \in C(x)$. A good example of correspondence is the inverse image of non-injective function.

Let $f : X \to Y$ be a function and, X and Y be two topological spaces. f is **continuous** at $x \in X$ if for any sequence (x_n) that converges to x, we have the sequence $f(x_n)$ converges to $f(x)$. In the case where X and Y are finite real-vector spaces, another definition of continuity is

$$\forall \varepsilon > 0,\ \exists \delta > 0,\ \text{such that } \forall y \in \mathbb{R}^n, \qquad \|x - y\|_{\mathbb{R}^n} < \delta \quad \Rightarrow \quad \|f(x) - f(y)\|_{\mathbb{R}^m} < \varepsilon.$$

A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is **differentiable** at x, if there exists a linear map $L : \mathbb{R}^n \to \mathbb{R}^m$ such that,

$$\forall \varepsilon > 0,\ \exists \delta > 0,\ \text{such that } \forall y \in \mathbb{R}^n, \quad \|x - y\|_{\mathbb{R}^n} < \delta \quad \Rightarrow \quad \|f(x) - f(y) - L(x - y)\|_{\mathbb{R}^m} < \varepsilon \|x - y\|_{\mathbb{R}^n},$$

and the linear map is defined as $D_x f$, the derivative of f at x. If f is a differentiable linear map, we define its **gradient** as $\boldsymbol{\nabla} f = Df^*$.

A function is **continuous/differentiable**, if it is continuous/differentiable at every point in its domain.

A family of functions $\{f_p(x)\}_{p \in \mathbb{N}}$ is **equi-differentiable** at x, if each function f_p is differentiable at x, and

$$\begin{aligned} &\forall \varepsilon > 0,\ \exists \delta > 0,\ \text{such that } \forall p \in \mathbb{N}, \forall y \in \mathbb{R}^n, \\ &\qquad\qquad \|x - y\|_{\mathbb{R}^n} < \delta \quad \Rightarrow \quad \|f_p(x) - f_p(y) - L_p(x - y)\|_{\mathbb{R}^m} < \varepsilon \|x - y\|_{\mathbb{R}^n}. \end{aligned}$$

In other words, all functions f_p converge uniformly, independently of $p \in \mathbb{N}$.

Let I be an interval in $\mathbb{R}$. A function $f : I \to \mathbb{R}$ is *absolutely continuous* on I, if for all $\varepsilon > 0$, there exists $\delta > 0$, any sub-interval $[a, b] \subset I$ which satisfies,

$$|a - b| < \delta \quad \Rightarrow \quad |f(a) - f(b)| < \varepsilon.$$

For a real-valued function on a compact interval $[a, b]$, the following properties are equivalent:

A. f is absolutely continuous,

B. there exists a Lebesgue integrable function g on $[a, b]$, such that

$$f(x) = f(a) + \int_a^x ds\, g(s), \qquad \forall x \in [a, b],$$

C. f has a derivative $D_x f$ almost everywhere and this derivative is Lebesgue integrable, i.e

$$f(x) = f(a) + \int_a^x ds\, D_x\, f(s), \qquad \forall x \in [a, b].$$

Let $f : X \times Y \to Z$ be a function. A **saddle-point** is a couple $(x_0, y_0) \in X \times Y$ such that

$$\sup_{y \in Y} f(x_0, y) \leq f(x_0, y_0) \leq \inf_{x \in X} f(x, y_0).$$

It is easy to check that the set of saddle-points is a product set. Note that this definition of a saddle-point is weaker than some definitions in the literature, where f is locally convex on x and locally concave on y at its saddle-point.

2.4 Probability theory

Let Ω be a set called the universe, and $\mathcal{A}$ be a subset of the power set of Ω, denoted $\mathcal{P}(A)$. The pair $(\Omega, \mathcal{A})$ is a **σ-algebra**, if the following conditions are satisfied:

- $\mathcal{A}$ is not empty,
- $\mathcal{A}$ is closed under complementation,
- $\mathcal{A}$ is closed under countable unions.

For a σ-algebra $(\Omega, \mathcal{A})$, a **probability distribution** is a map $p : \mathcal{A} \to [0, 1]$ such that:

- $p(\Omega) = 1$,
- $p(\cup_i U_i) = \sum_i p(U_i)$, where $(U_i)_i$ are disjoint sets.

The triplet $(\Omega, \mathcal{A}, p)$ defines a **probability space**.

Let $(\Omega_1, \mathcal{A}_1)$ and $(\Omega_2, \mathcal{A}_2)$ be two σ-algebras. A **measurable function** $f : (\Omega_1, \mathcal{A}_1) \to (\Omega_2, \mathcal{A}_2)$ satisfies

$$\forall E \in \mathcal{A}_2, \qquad f^{-1}(E) \subset \mathcal{A}_1.$$

Let $(\Omega_1, \mathcal{A}_1, p)$ be a probability space and $(\Omega_2, \mathcal{A}_2)$ a σ-algebra. A **random variable** X is a measurable function, $X : (\Omega_1, \mathcal{A}_1) \to (\Omega_2, \mathcal{A}_2)$. Hence X generates a probability distribution p_X on $(\Omega_2, \mathcal{A}_2)$ such that

$$\forall E \in \mathcal{A}_2, \qquad p_X(E) = p\big(X^{-1}(E)\big).$$

Let $(X_i)_{i\in\{1\dots n\}}$ be random variables defines on a probability space $(\Omega_i, \mathcal{A}_i, p_i)$ with probability distribution $(p_{X_i})_{i\in\{1\dots n\}}$ and n at least two. A **joint probability distribution** of $(X_i)_{i\in\{1\dots n\}}$ is a probability distribution $p_{X_1\dots X_n}$ such that,

$$\forall i \in \{1\dots n\} \text{ and } E \in \mathcal{A}_i \qquad p_{X_i}(E) = p_{X_1\dots X_n}(\Omega_1 \dots \Omega_{i+1}, E, \Omega_{i+1} \dots \Omega_n).$$

The probability distribution p_{X_i} is called the **marginal distribution** of $p_{X_1\dots X_n}$.
Let X and Y be two random variables defined with probability distributions p_X and p_Y. Let p_{XY} be a joint probability distribution of X, Y. The **conditional probability distribution** $p_{X|Y}$ is defined as

$$\forall (E,F), \qquad p_{X|Y}(X = E|Y = F) = \frac{p_{XY}(X = E, Y = F)}{p_Y(Y = F)}.$$

It is the probability to observe the event E knowing F is observed.
The **product distribution** $p_{X\times Y}$ of random variables X and Y is defined as

$$\forall (E,F), \qquad p_{X\times Y}(X = E, Y = F) = p_X(X = E).p_Y(Y = F).$$

A joint probability distribution p_{XY} is independent if it can be written as the product distribution of its marginal distributions p_X and p_Y.

Let S be a finite set. We define $\mathbb{P}(S)$ to be the set of all probability distributions on $(S, \mathcal{P}(S))$, and $\mathbb{B}(S)$ to be the set of all real functions on S. The **expectation** of a function $f \in \mathbb{B}(S)$ under the distribution $p \in \mathbb{P}(S)$ is

$$\mathbb{E}_p f = \langle f \rangle_p = \langle p, f \rangle = \sum_{s\in S} p(s)f(s).$$

A property P on a σ-algebra $(\Omega, \mathcal{A})$ with a measure μ is satisfied **almost everywhere** if,

$$\mu\big(\{\omega \in \Omega : \omega \text{ does not satisfy } P\}\big) = 0.$$

Let p, q be two probability distributions on X, we define the **total variation** as

$$|p - q|_{TV} = \sup_{S\subset X} \sum_{x\in S} \big|p(x) - q(x)\big|.$$

2.5 Information theory

Information theory was introduced by C. Shannon in [Sha48]. This theory provides two important theorems, the first for encoding a noiseless source and the second to encode a noisy channel. To obtain this result, C. Shannon introduced important mathematical tools such as the entropy and mutual information.

In this section, S is a finite set, X, Y, Z are random variables with respective probability distributions p_X, p_Y, p_Z in $\mathbb{P}(S)$, and their joint probability distributions p_{XYZ} in $\mathbb{P}(S^3)$. We define p_U to be the uniform probability distribution as $p_U(s) = 1/|S|$, for all $s \in S$. The function log is the binary logarithm ($\log 2 = 1$), and we use the convention $0 \log 0 = 0$.

The **entropy** $H(X)$ of a random variable X is a defined by

$$H(X) = -\sum_{s \in S} p_X(s) \log p_X(s).$$

Two properties of entropy are of interest: positivity $H(X) \geq 0$ and sub-additivity $H(X, Y) \leq H(X) + H(Y)$. The entropy achieves its maximum value $\log |S|$, for the uniform probability distribution p_U. The **conditional entropy** $H(X|Y)$ of random variables X, Y is a defined by

$$H(X|Y) = -\sum_{s_X \in S} \sum_{s_Y \in S} p_{XY}(s_X, s_Y) \log \frac{p_{XY}(s_X, s_Y)}{p_Y(s_Y)},$$

or more simply $H(X|Y) = H(X, Y) - H(Y)$, with $H(X, Y)$ defined to be the entropy of the joint probability distribution p_{XY}. Conditional entropy also satisfies similar properties, positivity $H(X|Y) \geq 0$ and strong sub-additivity $H(X|Y, Z) \leq H(X|Y)$. Conditional entropy achieves its maximum value $H(X)$, if X and Y are independent. The **mutual information** $I(X : Y)$ of random variables X, Y is defined by

$$I(X : Y) = \sum_{s_X \in S} \sum_{s_Y \in S} p_{XY}(s_X, s_Y) \log \frac{p_{XY}(s_X, s_Y)}{p_X(s_X) p_Y(s_Y)}.$$

or more simply as

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

The mutual information is symmetric, positive $I(X : Y) \geq 0$, and bounded by $H(X)$ and $H(Y)$. The mutual information of two random variables are null, if and only if there are independent.

The **Kullback-Leibler divergence** of X and Y is described as

$$D_{KL}(X||Y) = \sum_{s \in S} p_X(s) \log \frac{p_X(s)}{p_Y(s)}.$$

The Kullback-Leibler divergence has the important property of being positive, i.e. $D_{KL}(X||Y) \geq 0$. Where the equality is achieved if and only if $p_X = p_Y$. This is a direct consequence of the concavity of log. In particular, we can rewrite the entropy, the conditional entropy, and the mutual information as

(a) $H(X) = \log |S| - D_{KL}(X||U)$,

(b) $H(X|Y) = D_{KL}(X, Y||Y)$,

(c) $I(X : Y) = D_{KL}(XY||X \times Y)$.

Although D_{KL} is not symmetric, we can still construct a distance on $\mathbb{P}(S)$, with $2d(X, Y) = D_{KL}(X||Y) + D_{KL}(Y||X)$.

2.6 Convexity, cones and order

Let $\mathcal{V}$ be a vector space. An **affine combination** of the vectors v_i is a linear combination where all coefficients θ_i are reals and their sum is equal to one.

$$\theta_1 v_1 + \theta_2 v_2 + \ldots + \theta_n v_n, \qquad \text{with } \sum_{i=1}^{n} \theta_i = 1.$$

A **convex combination** $\mathbb{E}_p\, v$ of the vectors v_i is a linear combination where all real coefficients p_i are positive and their sum is equal to one.

$$\mathbb{E}_p\, v \ = \ \sum_{i=1}^{n} p_i v_i, \qquad \text{with } p_i \geq 0 \quad \text{and} \quad \sum_{i=1}^{n} p_i = 1$$

For example, all convex combinations of two points is the line between these points.
A **convex set** $\mathcal{C}$ is a stable set under all convex combinations of the vectors in it, such as the n-dimensional sphere. Note that convexity is a property stable for: intersection, scaling, element-wise sum, direct sum and direct product.
Let S be a subset of V. The **affine hull** of S is the set of all affine combinations of the vectors in S

$$\textbf{aff } S = \left\{ \sum_i \theta_i v_i : v_i \in S, \text{ and } \sum_i \theta_i = 1 \right\}.$$

The **conv hull** of S is the set of all convex combinations of the vectors in S

$$\textbf{conv } S = \left\{ \sum_i p_i v_i : v_i \in S, \ \sum_i \theta_i = 1, \text{ and } \forall i, \ p_i \geq 0 \right\}.$$

The interior of a set depends on its topological space. For example the interior of a disk is nonempty in $\mathcal{R}^2$ and empty in $\mathcal{R}^3$. Hence we define their interior relative to their affine hull. The **relative interior** of a set S is defined by

$$\textbf{relint } S = \{x \in S : \exists r > 0, B_r(x) \cap \textbf{aff } S \subset S\}.$$

2.6.1 Preorder, infimum and supremum

A **preorder** $\preceq$ on a set $\mathcal{S}$ is a relation with these properties:

- *(transitive)* if $x \preceq y$ and $y \preceq z$, then $x \preceq z$,
- *(reflexive)* for all $x \in \mathcal{S}$, $x \preceq x$,
- *(antisymmetric)* if $x \preceq y$ and $y \preceq x$, then $x = y$.

An order is *total* if for all $x, y \in \mathcal{S}$, $x \preceq y$ or $x \succeq y$. Let $(P, \leq)$ be a preorder and A be a subset of P. A **minimal element** of A is $m \in A$ such that for all $a \in A$, $m \leq a$. A **maximal element** of A is $m \in A$ such that for all $a \in A$, $a \leq m$. The **minimum element** of A exists if there is a unique minimal element. The **maximum element** of A exists if there is a unique maximal element. A **lower bound** of A is an $x \in P$ such that for all $a \in A$, $x \leq a$. An **upper bound** of A is an $x \in P$ such that for all $a \in A$, $x \geq a$. The **infimum** of A is the maximum of all lower bounds of A, if any exists. The **supremum** of A is the minimum of all upper bounds of A, if any exists. A subset is bounded if there is a lower bound and an upper bound.

2.6.2 Cone and generalized inequality

A **generalized inequality** on a vector space $\mathcal{V}$ is a preorder with the following additional properties:

- *(preserved under addition)* if $x_1 \preceq y_1$ and $x_2 \preceq y_2$, then $x_1 + x_2 \preceq y_1 + y_2$,
- *(preserved under nonnegative scaling)* if $x \preceq y$ and $\lambda \geq 0$, then $\lambda x \preceq \lambda y$,
- *(preserved under limits)* let (x_i) and (y_i) be sequences that converge to x and y, if $x_i \preceq y_i$ for $i \in \mathbb{N}$, then $x \preceq y$.

A **cone** is a set in a real vector space $\mathcal{V}$ stable under nonnegative scaling. A **proper cone** K is a cone with additional properties:

- K is closed and its interior is nonempty,
- K is convex,
- K is pointed, i.e. if $x \in K$ and $-x \in K$ then $x = 0$.

A proper cone K can define a generalized inequality:

$$y - x \in K \quad \Leftrightarrow \quad x \preceq_K y$$

Here is some examples of proper cones and its respectively inequality. The proper cone $\mathbb{R}^+$ of all positive reals defines the usual order $\leq$ on $\mathbb{R}$. The positive orthant $\mathbb{R}_n^+$ defines the product order $\leq_{\mathbb{R}_n^+}$ on $\mathbb{R}_n$. The set of positive semi-definite matrices $\mathcal{S}_+^n$ is a proper cone and defines the Loewner order $\leq_{\mathcal{S}_+^n}$ on $\mathcal{M}_n(\mathbb{K})$.
The subscript of $\leq_K$ can be omitted when the inner product space is well defined or unimportant.

2.6.3 Separating and supporting hyperplane theorems

Let V be a finite $\mathbb{R}$-vector space. An **affine hyperplan** h is an affine subspace of V described by a co-vector y and a real r such that

$$h = \{x \in V : \langle y, x \rangle = r\}.$$

A hyperplane h is a **supporting hyperplane** of a set S if, S is completely contained in one of the two closed half-spaces delimited by h and the intersection of S and h is nonempty.

Theorem 2.6.1 (Hyperplane separation theorem)**.** *[BV10] Let A and B be convex sets in V such that $A \cap B = \emptyset$. Then there exists a nonzero vector $c \in V$ and a real λ, such that $\langle c, x \rangle \geq \lambda$ if $x \in A$, and $\langle c, x \rangle \leq \lambda$ if $x \in B$.*

This theorem implies for two separated convex sets, there exists a hyperplane separating these sets. A corollary of this theorem is the Supporting hyperplane theorem which states that for every point x_0 in the boundary of a convex set A, there exists a supporting hyperplane for A in the point x_0.

Theorem 2.6.2 (Supporting hyperplane theorem)**.** *[BV10] Let A be a convex set in V and x_0 a point in* **bd** *A. Then there exists a nonzero vector $c \in V$ and a real λ, such that $\langle c, x_0 \rangle = \lambda$ and if $x \in A$ then $\langle c, x \rangle \geq \lambda$.*

This latter last theorem will be used in Chapter 6 to prove Slater's Theorem.

2.6.4 Convex and concave function

A **convex/concave function** is a real-valued function respecting the following Jensen's inequality

$$\textit{(convex)} \qquad f\big(\mathbb{E}_p v_i\big) \leq \mathbb{E}_p f(v_i) \tag{2.3}$$

$$\textit{(concave)} \qquad f\big(\mathbb{E}_p v_i\big) \geq \mathbb{E}_p f(v_i) \tag{2.4}$$

Note that f is concave *if and only if* $-f$ is convex. Convex functions and convex sets are deeply related. From a convex function we can construct a family of convex sets, $\mathcal{C}_\lambda = \{v \in \mathcal{V} : f(v) \leq \lambda\}$. On the other hand an indicator function of a convex set is a concave function.

Property 2.6.3 (First-order conditions)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function with* ***dom*** *f convex. Then f is convex if and only if for all $x, y \in$* **dom** *f*

$$f(y) \geq f(x) + Df_x(x) \cdot (y - x).$$

The above Property can be easily derived by looking at the epigraph of f and using the Supporting hyperplane Theorem 2.6.2.

Property 2.6.4 (Second-order conditions)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a double differentiable function with* **dom** *f open. Then f is convex if and only if* **dom** *f is convex and for all $x \in$* **dom** *f*

$$\boldsymbol{H}(x) \geq_{\mathcal{S}_+^n} 0,$$

where $\boldsymbol{H}(x)$ is the Hessian matrix of f, i.e $\boldsymbol{H}(x)[i,j] = \frac{\partial^2 f}{\partial x_i \partial x_j}(x)$.

Chapter 3

Quantum mechanics and quantum computation

This chapter is a brief introduction to quantum mechanics and quantum computing. As stated in Chapter 2 we work exclusively with finite dimensional Hilbert spaces and avoid dealing with infinite dimension particularities.

3.1 Quantum mechanics

A **quantum system** is defined by a Hilbert space $\mathcal{H}$. A **quantum state** is described by a unit vector $\boldsymbol{v}$ denoted $|v\rangle$ and called "ket". A co-vector $u^* : \mathcal{H} \to \mathbb{C}$ is denoted $\langle u\,|$ and called "bra". The evaluation of $\boldsymbol{u}^*$ on $\boldsymbol{v}$ is called "bra-ket" with the following notation

$$\langle u \mid v\rangle = \boldsymbol{u}^*(\boldsymbol{v}) = \big\langle \boldsymbol{u}, \boldsymbol{v}\big\rangle.$$

A physical quantity $\mathcal{M}$ is represented by a Hermitian linear operator M on $\mathcal{H}$, such an operator is called an **observable**. The measure of $\mathcal{M}$ can only give an eigenvalue of M. For a quantum state $|v\rangle$, an average value $\langle M\rangle$ of an observable M is given by

$$\langle M\rangle = \langle v\,|\,M\,|v\rangle\,.$$

Since M is a Hermitian operator acting on a finite Hilbert space, we can write it under the form

$$M = \sum_m m\boldsymbol{P}_m, \tag{3.1}$$

where m is an eigenvalue of M, and $\boldsymbol{P}_m$ is the projection on the subspace $\hat{V}_m$ spanned by eigenvectors with the eigenvalue m. Consequently, $\sum_m \boldsymbol{P}_m = \mathcal{I}d$ and $\mathcal{H} = \oplus_m \hat{V}_m$.

The probability to observe the outcome m on the quantum state $|v\rangle$ is described by the distribution probability $\boldsymbol{p}$ such that

$$\forall m, \quad p(m) = \langle v\,|\,\boldsymbol{P}_m\,|v\rangle\,. \tag{3.2}$$

A quantum state $|v(t)\rangle$ evolves over time with two different mechanisms. The first mechanism is the Schrödinger's equation

$$\frac{d}{dt}\,|v(t)\rangle = -iH(t)\,|v(t)\rangle\,, \tag{3.3}$$

where $H(t)$ is an observable called "Hamiltonian", and $\hbar$ the reduced Planck constant is fixed to one. From the above equation of motion, the evolution is described by a unitary operator $U(t',t)$ called **unitary evolution** such that $|v(t')\rangle = U(t',t)\,|v(t)\rangle$ for all times $t \leq t'$ and kets $|v\rangle$.
The second mechanism appears during a measurement. If we measure $\mathcal{M}$ on the quantum state $|v\rangle$ and obtain the outcome m, then $|v\rangle$ is projected to the quantum state

$$|v_m\rangle = \frac{1}{\sqrt{p(m)}} \boldsymbol{P}_m \,|v\rangle\,.$$

where $\boldsymbol{P}_m$ and $p(m)$ are defined in Equations (3.1) and (3.2).

A distribution $(q_k)_k$ of quantum states $|v_k\rangle$ cannot be described by a *ket*. Therefore we introduce a matrix to represent a distribution of kets, defined as

$$\rho = \sum_k q_k \,|v_k\rangle\!\langle v_k|\,.$$

The definition of the matrix above naturally implies following properties,

$$\rho \geq 0 \qquad \text{and} \qquad \mathrm{tr}\rho = 1. \tag{3.4}$$

Since every matrix that satisfies these properties has a positive spectrum normed to one, their spectrum can be interpreted as a distribution of kets. Therefore a **density matrix** is a matrix that satisfies both conditions (3.4).
The Schrödinger's equation can be rewritten as,

$$\frac{d}{dt}\rho(t) = -i[H(t), \rho(t)],$$

where $[\cdot,\cdot]$ is the commutator defined as $[A,B] = AB - BA$.
Similarly for an observable M as defined in Equation (3.1), we measure m with the probability,

$$p(m) = \mathrm{tr}(\boldsymbol{P}_m \rho),$$

and the density matrix ρ is projected to,

$$\rho_m = \frac{\boldsymbol{P}_m \rho \boldsymbol{P}_m}{p(m)}.$$

The distance between two quantum states $|u\rangle$ and $|v\rangle$ is defined from the norm on $\mathcal{H}$,

$$d(|u\rangle\,, |v\rangle) = \big\|\, |u\rangle - |v\rangle \,\big\|.$$

We say that $|u\rangle$ and $|v\rangle$ are **$\boldsymbol{\varepsilon}$-distant** if their distance is less than ε.

3.1.1 Adiabatic quantum theorem

The unitary evolution $U(t',t)$ can be easily derived by integrating Schrödinger's equation (3.3) when the Hamiltonian H is independent of time t.

$$U(t',t) = e^{-i(t'-t)H}. \tag{3.5}$$

Unfortunately for a time-dependent Hamiltonian the integration could become quite complicated. The quantum adiabatic theorem is a method to approximate this integration when the

Hamiltionian varies slowly in times.

The quantum adiabatic theorem, introduced by M. Born and V. Fock [BF28] and inspired from quasi-static process in thermodynamics, approximates the unitary evolution $U(t',t)$ when the Hamiltonian is continuous in time and varies slowly. In a few words, the intuition is

A quantum system with a time-dependent Hamiltonian remains in its instantaneous eigenvector if, the Hamiltonian variation is slowly enough and there is a large gap between the corresponding eigenvalue and the rest of the spectrum of the Hamiltonian.

Let clarify the main idea. Assume a time-dependent Hamiltonian $H(t)$ can be written as Equation (3.1) under its spectral form

$$\forall t, \qquad H(t) = \sum_n E_n(t) P_n(t),$$

with eigenvalues $E_n(t)$ and projections $P_n(t)$ both continuous in time, such that $H(t)P_n(t) = E_n(t)P_n(t)$ for each n. If functions $E_n(t)$ stay distant over the time (They never cross each other.) then a quantum state in the subspace $\hat{V}_n(t)$ remains predominantly in the subspace $\hat{V}_n(t+t')$ after a time t'. ($\hat{V}_n(t)$ is the range of $P_n(t)$.) The minimal distance over the time between functions $E_n(t)$ is called the gap and denoted g.

In order to formally describe adiabatic quantum computation, we define the definition of an adiabatic process.

Definition 3.1.1. An **adiabatic process** on the Hilbert space $\mathcal{H}$ is defined by a triplet $\{H(s), P(s), \tau\}$ with $s \in [0,1]$ where

(a) $H(s)$ is a double-differentiable map from $[0,1]$ to the space of bounded linear Hermitian operators on $\mathcal{H}$ equipped with the graph norm $\|\cdot\|_{H(0)}$,

(b) $P(s)$ is a rank-one projection onto an eigenvector of $H(s)$ where its corresponding eigenvalue $\lambda(s)$ is continuous in s,

(c) $\tau \in \mathbb{R}^+$ is the running time of the process.

The relation between the real time $t \in [0,\tau]$ and the time $s \in [0,1]$ used in the above Definition is defined by

$$t = s\tau.$$

For an adiabatic process $\{H(s), P(s), \tau\}$ we define $U_A(s)$ to be the **idealized evolution**, the unitary operator that maps the projection $P(0)$ onto the projection $P(s)$ for all s such that,

$$U_A(s)P(0)U_A^*(s) = P(s).$$

In the other hand we call the **physical evolution** the unitary operator $U_\tau(s)$ derived from the Schrödinger's equation

$$i\frac{d}{ds}U_\tau(s) = \tau H(s) U_\tau(s). \qquad (3.6)$$

Note that analytical conditions given in the above Definition 3.1.1 ensures existence and uniqueness of $U_\tau(s)$ as defined in Equation (3.6) with the initial condition $U_\tau(0) = \mathcal{I}d$. [RS75].

The quantum adiabatic theorem can now be summarized by the following statement

$$\lim_{\tau\to\infty} U_\tau(s)P(0) = U_A(s)P(0). \tag{3.7}$$

Since the adiabatic process gets slower with a large τ, $U_\tau(s)P(0)$ converges to $U_A(s)P(0)$ when τ converges to infinity. Nonetheless we need to analyze the error along the adiabatic process by looking at the norm of the difference between $U_\tau(s)P(0)$ and $U_A(s)P(0)$.

Definition 3.1.2. The **error** $\varepsilon_{AP}(s)$ of an adiabatic process $\{H(s), P(s), \tau\}$ is defined as

$$\varepsilon_{AP}(s) = \Big\| \big[U_\tau(s) - U_A(s)\big]P(0)\Big\|, \qquad \text{with} \quad \varepsilon_{AP} = \varepsilon_{AP}(1).$$

This definition implies that, every quantum state in the range of $P(0)$ will be ε_{AP}-distant from the range of $P(1)$ after the adiabatic process.

The main question is: How slow should be the adiabatic process to ensure an adiabatic error ε_{AP}?

A criterion often used is the **folk adiabatic condition**. It requires that

$$\tau \gg \int_0^1 \frac{\|\frac{d}{ds}H_\tau(s)\|}{g(s)^2} ds, \tag{3.8}$$

where the gap $g(s)$ represents the minimal distance between the eigenvalue $\lambda(s)$ and the rest of spectrum of $H(s)$. Unfortunately the folk adiabatic condition is only a criterion and it cannot be used to rigorously bound the adiabatic error ε_{AP}. Rigorous conditions have been found only recently [JRS07].

We now use the Newton's notation: $\dot{A}(s) = \frac{d}{ds}A(s)$ and $\ddot{A}(s) = \frac{d^2}{ds^2}A(s)$.

Theorem 3.1.3. *[JRS07]*
Let $\{H(s), P(s), \tau\}$ be an adiabatic process, $g = \min_{s\in[0,1]} g(s)$ be the minimum gap, $\dot{H}(s)$ and $\ddot{H}(s)$ bounded, and $\varepsilon > 0$.

$$\textit{If} \quad \tau \geq \frac{1}{\varepsilon}\left[\frac{\|\dot{H}(0)\| + \|\dot{H}(1)\|}{g^2} + \max_{s\in[0,1]}\left\{\frac{\|\ddot{H}(s)\|^2}{g^2} + 7\frac{\|\dot{H}(s)\|^2}{g^3}\right\}\right], \quad \textit{then} \quad \varepsilon_{AP} \leq \varepsilon.$$

Although the existence of a gap is required in the folk adiabatic condition (3.8) and Theorem 3.1.3, the following Lemma from J. Avron and A. Elgart [AE99a] shows that a gap is not always a necessary condition.

Lemma 3.1.4. *[AE99a]*
Let $\{H(s), P(s), \tau\}$ be an adiabatic process, $\varepsilon > 0$, $X(s)$ be a bounded operator satisfying the commutator equation

$$\dot{P}(s)P(s) = [H(s), X(s)], \tag{3.9}$$

and $\dot{X}(s)$ bounded.

$$\textit{If} \quad \tau \geq \frac{1}{\varepsilon}\Big[\|X(0)\| + \|X(1)\| + \max_{s\in[0,1]} \|\dot{X}(s)P(s)\|\Big], \quad \textit{then} \quad \varepsilon_{AP} \leq \varepsilon.$$

This Lemma is a special case of the statement proved in [AE99a] adapted to the case of continuous-time quantum computation. For completeness we provide a proof of Lemma 3.1.4 in Appendix A.

3.2 Quantum computation

The idea of quantum computation emerged at the end of 20th century when R. Feynman remarked that simulating a quantum system is hard, and could become quite easier if the computer would have quantum properties [Fey82]. For a more exhaustive introduction to this subject and to go further, I recommend the well-known "Quantum Computation and Quantum Information" from I. Chuang and M. Nielsen [NC11]

Without being comprehensive, in a quantum computer a bit 0 or 1 is replaced by a quantum system $\mathbb{C}^2$ called qubit with two levels $|0\rangle$ and $|1\rangle$. The description of a quantum algorithm is dependent of its computational model. Two important and polynomially equivalent models are the discrete-time model and the continuous-time model.

In the discrete-time model a quantum algorithm is a circuit constituted from quantum gates (unitary operators acting on one or two qubits at most). In the continuous-time model the quantum state evolves under the Schrödinger's equation with a Hamiltonian formed by the addition of locals Hamiltonians (Hamiltonian acting on one or two qubits at most).

In 2000, E. Farhi et al. [FGGS00] introduced the **adiabatic quantum computation**, a special case of the continuous-time model based on the quantum adiabatic theorem. In their article they solve instances of the satisfiability problem by constructing a final Hamiltonian H_f depending on satisfying assignment, such that a quantum state $|\psi_f\rangle$ encoding a solution of the problem has the lowest energy. Hence the quantum adiabatic theorem would allow that the linear interpolation $H(s)$

$$H(s) = (1-s)H_{in} + sH_f,$$

evolves the quantum state $|\psi_{in}\rangle$ with the lowest energy of H_{in} to the final state $|\psi_f\rangle$. H_{in} is a Hamiltonian with a ground state easily to construct by convenience. Of course, in this scheme the correctness of adiabatic algorithms relies on the existence of a spectral gap. It was later proved that the adiabatic model is polynomially equivalent to discrete-time model in term of time complexity.[AvDK+07].

In [FGGS00] Farhi and Gutmann also give the first example of adiabatic algorithm for unstructured search, a continuous-time analogue of Grover's algorithm based on a simple linear interpolation of two Hamiltonians (Later van Dam et al. [vDMV02], as well as Roland and Cerf [RC02], independently proposed an adiabatic version of this algorithm based on a slowly varying Hamiltonian). Algorithms were also developed in the continuous-time model for various problems such as spatial search [CG04a, CG04b, FGT14], oracle identification [Moc07], or element distinctness [Chi09]. In a seminal paper, Farhi et al. [FGG08] proposed a quantum algorithm for the NAND-tree based on scattering a wave incoming on the tree using a time-independent Hamiltonian. It is precisely this algorithm that, through successive extensions, led to the tight algorithm based on the adversary method for any function in [Rei11], but most of these extensions were using the discrete-time model.

Chapter 4

Query complexity

In this chapter and thereafter Σ is a finite set describing an alphabet, ϵ the blank character, and $\hat{\Sigma} = \Sigma \cup \{\epsilon\}$ the extended alphabet. $\mathcal{X} \subset \Sigma^n$ is a subset of n-length strings with $n \in \mathbb{N}$, and $N = |\mathcal{X}|$ is its size. Finally a string x is an element of $\mathcal{X}$.

$\mathcal{A}$ describes an algorithm and with some language abuse, also as a simple input/output map: a function for a classical algorithm and a unitary operator followed by a measurement for a quantum algorithm.

4.0.1 Query complexity

A query algorithm $\mathcal{A}(x)$ is a special algorithm where the input $x \in \mathcal{X}$ is unknown at the beginning of the computation, and can only be learned through a specific action called "query". This model of algorithm is studied to answer questions as: "Do I only need full or partial information on the input?", "Shall we distinguish all possible inputs?" or "Which queries are necessary?" This model of algorithm is used to lower bound the complexity of a function, since a query is a specific action.

A **query algorithm** evaluates a function $f : \mathcal{X} \to \mathcal{Y}$ with specific restrictions:

- the input $x \in \mathcal{X}$ is unknown at the beginning of the computation,
- each character of a string x can be known only through a function $\mathcal{O}_x$ called *oracle* defined by

$$\mathcal{O}_x : (k, b) \mapsto (k, b \oplus x_k), \tag{4.1}$$

where $k \in \{1 \ldots n\}$, $b \in \Sigma$, and $\oplus$ the addition modulo $|\Sigma|$.

A query is done by calling an oracle, thus we define the **query cost** $C(\mathcal{A}, x)$ to be the number of queries used by an algorithm $\mathcal{A}$ on input x. The query complexity of an algorithm $C(\mathcal{A})$ is the query cost of $\mathcal{A}$ on its worst input. The **query complexity** of a function is the minimum query cost over all query algorithms that evaluate f,

$$C(f) = \min_{\mathcal{A}:\mathcal{A}(\cdot)=f(\cdot)} \max_{x\in\mathcal{X}} C(\mathcal{A}, x).$$

Obviously these definitions imply that $C(f)$ is upper bounded by n, the length of inputs. For example for $\mathcal{X} = \Sigma^n$ the exact evaluation of the identity function needs n queries.

Remark. In classical computation, a query algorithm can be represented as a decision tree [BdW02]. Indeed, the algorithm can be represented as a tree where the top is the beginning of the algorithm, each vertex represents the choice of the query, and roots are the output. Clearly the depth of this tree corresponds to the query cost of the algorithm. Thus, the query complexity of a function f is the minimum depth of all decision trees computing the function f.

4.1 Quantum query complexity

A natural generalization of a query algorithm to the quantum world is to start with an arbitrary quantum state like $|0\rangle$ and to evolve it toward a state $|\sigma_x\rangle$ depending on x, where a "good" choice of measurement gives the desired output $f(x)$. Of course, evolution uses unitary transformations independent of x, and oracle $\mathcal{O}_x$ dependent of x.

We can generalize a little more our idea of a quantum query algorithm. Instead of evaluating a function we could generate a quantum state $|\sigma_x\rangle$. Furthermore, we could convert an input quantum state $|\rho_x\rangle$ to a final quantum state $|\sigma_x\rangle$. From this point of view, a quantum query algorithm can be seen as quantum state converter which convert the quantum state $|\rho_x\rangle$ to $|\sigma_x\rangle$, for each x. Hence evaluating a function f is the particular case where we start with an initial quantum state independent of x, and we end with identifiable orthonormal quantum states.

Remark. For any unitary transformation U, if a quantum query algorithm generates every state $|\sigma_x\rangle$, then it is simple to create a quantum query algorithm to generate states $U|\sigma_x\rangle$ without additive cost. Therefore we can represent a family $\{|\sigma_x\rangle\}_x$ by σ: the unitary Gram matrix of the set of unit vectors $|\sigma_x\rangle$.

$$\sigma = \mathrm{Gram}(|\sigma_x\rangle : x \in \mathcal{X}).$$

For families of quantum states $\{|\rho_x\rangle\}_x$ and $\{|\sigma_x\rangle\}_x$, we define respectively their unitary Gram matrix to be ρ and σ. We define $(\rho \to \sigma)$ to be a **state conversion problem**, the problem to convert for each $x \in \mathcal{X}$, the quantum state $|\rho_x\rangle$ to $|\sigma_x\rangle$.

A **quantum query algorithm** is a unitary operator $\mathcal{A}$ acting on a Hilbert space $\mathcal{H}_\mathcal{Q} \otimes \mathcal{H}_\mathcal{W}$, where $\mathcal{H}_\mathcal{Q}$ is the query register and $\mathcal{H}_\mathcal{W}$ is the workspace register. For each x, the quantum state $|\rho_x\rangle$ can be decomposed into

$$|\rho_x\rangle = |0\rangle_\mathcal{Q} \otimes |\rho_{x;0}\rangle_\mathcal{W} + \sum_{k\in\{1\ldots n\}} |k\rangle_\mathcal{Q} \otimes |\rho_{x;k}\rangle_\mathcal{W}\,, \tag{4.2}$$

where $\{|k\rangle\}_{k\in\{0\ldots n\}}$ is the canonical basis of $\mathcal{H}_\mathcal{Q}$, and $|\rho_{x;k}\rangle_\mathcal{W}$ is a non-normalized unit vector in $\mathcal{H}_\mathcal{W}$, constructed from the projection of $|\rho_x\rangle$ on $|k\rangle$. $|0\rangle_\mathcal{Q}$ is a special vector of $\mathcal{H}_\mathcal{Q}$ that remains unchanged after the oracle's action. Note that a vector $|k\rangle$ represents a query, so vectors $|\rho_{x;k}\rangle$ indicate for each x which fraction of query "k"we obtain after an oracle call.

Regarding the oracle it can be implemented in two different ways: either as a unitary operator $\mathcal{O}_x$ in the discrete-time model, or a Hamiltonian H_x in the continuous-time model. Also for each model there are several possible representations of an oracle, according to maps (4.1). But as long as two different representations can implement each other with a constant number of queries, their query complexity differ only by a constant factor.

4.1.1 Discrete-time model

In the discrete-time model, a **quantum query algorithm** $\mathcal{A}(x)$ is a sequence of input-independent unitary operators U_t interleaved with oracle calls $\mathcal{O}_x$. An oracle $\mathcal{O}_x$ is represented by a unitary operator and acts as the classical oracle (4.1), i.e. every letter of x can be queried with one and only one query. Hereafter we provide two possible representations for $\mathcal{O}_x$, the **phase-oracle** $\mathcal{O}_x^{\mathrm{ph}}$ and the **register-oracle** $\mathcal{O}_x^{\mathrm{reg}}$.

The register-oracle $\mathcal{O}_x^{\mathrm{reg}}$ acts on $\mathcal{H}_\mathcal{Q} \otimes \mathcal{H}_\mathcal{V} = \mathbb{C}^{|\Sigma|}$ where $\mathcal{H}_\mathcal{V}$ is a well-defined Hilbert subspace of $\mathcal{H}_\mathcal{W}$,

$$\mathcal{O}_x^{\mathrm{reg}} : \begin{cases} |k\rangle_\mathcal{Q} |\epsilon\rangle_\mathcal{V} & \mapsto |k\rangle_\mathcal{Q} |x_k\rangle_\mathcal{V} \quad \forall k \in \{1 \dots n\}, \\ |k\rangle_\mathcal{Q} |x_k\rangle_\mathcal{V} & \mapsto |k\rangle_\mathcal{Q} |\epsilon\rangle_\mathcal{V} \quad \forall k \in \{1 \dots n\}, \\ |k\rangle_\mathcal{Q} |l\rangle_\mathcal{V} & \mapsto |k\rangle_\mathcal{Q} |l\rangle_\mathcal{V} \quad \forall k \in \{1 \dots n\}, \forall l \in \Sigma \setminus \{x_k\}, \\ |0\rangle_\mathcal{Q} |l\rangle_\mathcal{V} & \mapsto |0\rangle_\mathcal{Q} |l\rangle_\mathcal{V} \quad \forall l \in \hat{\Sigma}. \end{cases}$$

Hence, the oracle $\mathcal{O}_x^{\mathrm{reg}}$ is an involution, and acts as identity everywhere else. In particular, the projection $|\rho_{x;0}\rangle$ is invariable under all $(\mathcal{O}_x^{\mathrm{reg}})_x$.

The phase-oracle $\mathcal{O}_x^{\mathrm{ph}}$, only defined for a binary alphabet ($|\Sigma| = 2$), is represented as a phase-operator,

$$\mathcal{O}_x^{\mathrm{ph}} : |k\rangle_\mathcal{Q} \mapsto (-1)^{x_k} |k\rangle_\mathcal{Q} ,$$

with the convention $x_0 = 0$. Note that for the register-oracle representation, $\mathcal{H}_\mathcal{V} = \mathbb{C} = \mathrm{span}\{|\epsilon\rangle\}$.

To characterize the difference between two oracles $\mathcal{O}_x$ and $\mathcal{O}_y$, we define $(\Delta_k)_k$ to be the set of matrices for each canonical vector $|k\rangle$ of $\mathcal{H}_\mathcal{Q}$. For each $k \in \{0 \dots n\}$, we define

$$\forall x, y \in \mathcal{X}, \qquad \Delta_k[x, y] = \langle k, \epsilon \mid \mathcal{O}_x^* \mathcal{O}_y | k, \epsilon\rangle .$$

These matrices are dependent of the representation of the oracle, for example

$$\Delta_k^{\mathrm{reg}}[x, y] = \delta[x_k, y_k] \qquad \text{and} \qquad \Delta_k^{\mathrm{ph}}[x, y] = (-1)^{y_k - x_k}.$$

The **discrete-time quantum query complexity** $Q_0^{\mathrm{dt}}(\rho \to \sigma)$ is the minimum number of queries over all algorithms converting exactly ρ to σ. The choice of the representation of the oracle affects the quantum query complexity $Q_0^{\mathrm{dt}}(\rho \to \sigma)$ but only by a constant factor at most.

4.1.2 Continuous-time model

In the continuous-time model the evolution is described by Schrödinger's equation, so unlike the previous model, we don't work directly with unitary operators but with Hamiltonians. Since a discrete-time algorithm is built with unitary operators independent of x and time-invariant oracle, then any Hamiltonian $H_x(t)$ of a continuous-time algorithm is the sum two parts: an unrestricted driver Hamiltonian $H_\mathcal{D}(t)$ independent of x and a time-invariant oracle Hamiltonian $H_\mathcal{Q}(x)$.

A **continuous-time quantum query algorithm** is described by a running time T, and a Hamiltonian $H_x(t)$ acting on the Hilbert space $\mathcal{H}_\mathcal{Q} \otimes \mathcal{H}_\mathcal{W}$, such that $H_x(t)$ has the following form

$$H_x(t) = H_{\mathcal{D}}(t) + \alpha(t) H_{\mathcal{Q}}(x), \qquad \text{with } \|H_{\mathcal{Q}}(x)\| \leq 1 \quad \forall x,$$
$$\text{and } \alpha : [0, T] \to [0, 1].$$

Note that the function α and the norm of the oracle Hamiltonian $H_{\mathcal{Q}}$ are both upper bounded by unity, otherwise we can obtain an arbitrary speed-up.

The representation of oracle Hamiltonian $H_{\mathcal{Q}}$ can be as general as in [Bel15], but in this thesis we restrict ourselves to the following standard representation of the Hamiltonian oracle acting on $\mathcal{H}_{\mathcal{Q}} \otimes \mathcal{H}_{\mathcal{V}}$ as

$$H_{\mathcal{Q}}(x) = \sum_{\substack{k \in \{1 \dots n\} \\ \tau = \pm 1}} \tau |k, \tau\rangle\langle k, \tau|_{\mathcal{Q}} \otimes h(x_k)_{\mathcal{V}}. \tag{4.3}$$

Where an $\mathcal{H}_{\mathcal{V}}$ is a well-defined Hilbert subspace of $\mathcal{H}_{\mathcal{W}}$, $h(l)$ is a Hamiltonian dependent of $l \in \Sigma$, and $\mathcal{H}_{\mathcal{Q}}$ has been extended to a $(2n+1)$-dimensional subspace with the canonical basis $\{|0\rangle, |k, \pm\rangle\}_{k \in \{1 \dots n\}}$. Since $H_{\mathcal{Q}}$ is bounded by unity, therefore $h(l)$ is also bounded by unity for all $l \in \Sigma$. Note that we add the factor τ to allow to rapidly un-compute a query. Evidently Hamiltonians $h(l)$'s must be chosen according to

$$\mathcal{O}_x = e^{-i\pi H_{\mathcal{Q}}(x)}, \quad \forall x \in \mathcal{X},$$

where $\mathcal{O}_x$ must be a representation of the oracle. We give two representations for the oracle Hamiltonian $H_{\mathcal{Q}}(x)$.

- The **register-Hamiltonian** $H_{\mathcal{Q}}^{\text{reg}}(x)$ acts on $\mathcal{H}_{\mathcal{Q}} \otimes \mathcal{H}_{\mathcal{V}}$ where an $\mathcal{H}_{\mathcal{V}} = \mathbb{C}^{\hat{\Sigma}}$ is a well-defined Hilbert subspace of $\mathcal{H}_{\mathcal{W}}$, and

$$\forall l \in \Sigma, \quad h(l) = |l^-\rangle\langle l^-|_{\mathcal{V}}, \text{ where } |l^-\rangle = \frac{1}{\sqrt{2}}(|\epsilon\rangle - |l\rangle). \tag{4.4}$$

- The **phase-Hamiltonian** $H_{\mathcal{Q}}^{\text{ph}}(x)$, only defined for a binary alphabet ($|\Sigma| = 2$), with $\mathcal{H}_{\mathcal{V}} = \mathbb{C}$ and represented as,

$$\forall l \in \Sigma, \quad h(l) = l. \tag{4.5}$$

As the infinitesimal difference between two oracles is defined by the difference between corresponding oracle Hamiltonians, we define $(\hat{\Delta}_k^\tau)_{k\tau}$ to be the set of matrices characterizing this difference. Each entry of the Hermitian matrix $\hat{\Delta}_k^\tau$ characterizes the difference between two different oracle Hamiltonians on each canonical vector $|k, \pm\rangle$ of $\mathcal{H}_{\mathcal{Q}}$. For each $k \in \{0 \dots n\}$ and $\tau \in \{+1, -1\}$,

$$\forall x, y \in \mathcal{X}, \qquad \hat{\Delta}_k^\tau[x, y] = i \langle k, \tau, \epsilon \mid H_{\mathcal{Q}}(y) - H_{\mathcal{Q}}(x) | k, \tau, \epsilon\rangle.$$

These matrices are dependent of the representation of the oracle, for example

$$\hat{\Delta}_k^{\tau,\text{reg}}[x, y] = 0 \qquad \text{and} \qquad \hat{\Delta}_k^{\tau,\text{ph}}[x, y] = i\tau(y_k - x_k).$$

Note that for the register-Hamiltonian $\hat{\Delta}_k^{\tau,\text{reg}}$ is null because the first order of the difference is null. To avoid a superfluous notation we remove the superscript such that $\hat{\Delta}_k^\tau = \hat{\Delta}_k^{\tau,\text{ph}}$.

In the Hilbert space $\mathcal{H}_{\mathcal{Q}} \otimes \mathcal{H}_{\mathcal{W}}$ every quantum state $|\rho_x\rangle$ can be decomposed into the form

$$|\rho_x\rangle = |0\rangle_{\mathcal{Q}} \otimes |\rho_{x;0}\rangle_{\mathcal{W}} + \sum_{\substack{k\in\{1\dots n\}\\ \tau=\pm1}} |k,\tau\rangle_{\mathcal{Q}} \otimes |\rho_{x;k\tau}\rangle_{\mathcal{W}}\,,$$

where $|\rho_{x;k\tau}\rangle$ is the projection of $|\rho_x\rangle$ on $|k,\tau\rangle$, therefore non necessarily normalized. In order to simplify the form of $|\rho_x\rangle$, we define the set $\hat{n} = \{0\} \cup \big(\{+1,-1\} \times \{1\dots n\}\big)$, and we have

$$|\rho_x\rangle = \sum_{\hat{k}\in\hat{n}} \left|\hat{k}\right\rangle_{\mathcal{Q}} \otimes \left|\rho_{x;\hat{k}}\right\rangle_{\mathcal{W}}. \tag{4.6}$$

The **query cost** $q(\mathcal{A})$ of a continuous-time quantum query algorithm $\mathcal{A}$ can be derived directly from the function $\alpha(t)$ and the running time T,

$$q(\mathcal{A}) = \frac{1}{\pi}\int_0^T dt\,\alpha(t). \tag{4.7}$$

The **continuous-time quantum query complexity** $Q_0^{\text{ct}}(\rho \to \sigma)$ is the minimum query cost over all algorithms converting exactly $|\rho_x\rangle$ to $|\sigma_x\rangle$, for all $x \in \mathcal{X}$.

4.1.3 Output conditions

For scenarios where we accept errors we must distinguish two cases : **coherent** and **non-coherent** quantum state conversion. Concretely, a computation will typically use some extra workspace and may therefore generate a state $|\sigma_x, J_x\rangle$, where $|J_x\rangle$ is the final state of the workspace. This might not be desirable if the state generation is used as a subroutine in a larger quantum algorithm, where we would like to use interferences between the states $|\sigma_x\rangle$ for different x. In that case, we would like to be able to reset the state $|J_x\rangle$ to a default state, so that it does not affect interferences. We therefore define the following output conditions (both for the discrete- and continuous-time models)

Definition 4.1.1 (Output condition)**.** A quantum query algorithm acting as unitary $\mathcal{A}(x)$ for input x converts ρ to σ with error at most ε if

- (coherent case) $\forall x \in \mathcal{X}$, $\text{Re}(\langle\sigma_x, 0 \mid \mathcal{A}(x)|\rho_x, 0\rangle) \geq \sqrt{1-\varepsilon}$,
- (non-coherent case) $\forall x \in \mathcal{X}$, $\exists\,|J_x\rangle$, $\text{Re}(\langle\sigma_x, J_x \mid \mathcal{A}(x)|\rho_x, 0\rangle) \geq \sqrt{1-\varepsilon}$.

Note that a sufficient condition for $\text{Re}(\langle\phi \mid \psi\rangle) \geq \sqrt{1-\varepsilon}$ is that these states are $\sqrt{\varepsilon}$-distant. Moreover, the output condition for the coherent case has been shown [LR12] to be equivalent to $\mathcal{F}_H(\sigma,\sigma') \geq \sqrt{1-\epsilon}$ where σ' is the Gram matrix of the output states $|\sigma'_x\rangle = \mathcal{A}(x)\,|\rho_x, 0\rangle$, and $\mathcal{F}_H$ the Hadamard product fidelity define in 2.1. Similarly, in the non-coherent case the output conditions can be rewritten as $\mathcal{F}_H(\sigma \circ J, \sigma') \geq \sqrt{1-\epsilon}$, where J is the Gram matrix of any set of unit vectors $|J_x\rangle$. This implies that bounded-error and zero-error quantum query complexities are related as follows.

Lemma 4.1.2 ([LR12])**.** *For any N-by-N Gram matrices ρ and σ we have*

$$Q_\varepsilon^{\bullet}(\rho,\sigma) = \min_{\sigma'}\left\{Q_0^{\bullet}(\rho,\sigma') : \mathcal{F}_H(\sigma,\sigma') \geq \sqrt{1-\epsilon}\right\}, \tag{4.8}$$

$$Q_\varepsilon^{\text{nc},\bullet}(\rho,\sigma) = \min_{\sigma'}\left\{Q_0^{\bullet}(\rho,\sigma') : \mathcal{F}_H(\sigma\circ J,\sigma') \geq \sqrt{1-\epsilon}, J \circ \mathcal{I}d = \mathcal{I}d\right\}, \tag{4.9}$$

where the superscript nc *denotes the non-coherent query complexity (otherwise we consider the coherent case by default), and the superscript $\bullet$ is either* dt *or* ct.

Computing a function f is equivalent to generating the Gram matrix $F[x,y]=\delta[f(x),f(y)]$ from the all-one Gram matrix $\mathbb{J}[x,y]=1$. In that case, it is not necessary to generate the state coherently, but for functions we can convert a non-coherent algorithm into a coherent algorithm, thereby we can consider the coherent case without loss of generality.

Lemma 4.1.3 ([LR12])**.** *For any function f and associated Gram matrix $F[x,y]=\delta[f(x),f(y)]$, we have $Q^{\bullet}_{\varepsilon}(f)=Q^{\mathrm{nc},\bullet}_{\varepsilon}(\mathbb{J},F)$ and*

$$Q^{\mathrm{nc},\bullet}_{\varepsilon}(\mathbb{J},F)\leq Q^{\bullet}_{\varepsilon}(\mathbb{J},F)\leq 2Q^{\mathrm{nc},\bullet}_{1-\sqrt{1-\varepsilon}}(\mathbb{J},F).$$

4.2 Gram matrix representation

In Section 4.1, we remarked that input/output quantum states can be represented by their unitary Gram matrix ρ and σ without loss of generality.

$$\rho=\mathrm{Gram}\big(\,|\rho_x\rangle : x\in\mathcal{X}\big),$$
$$\sigma=\mathrm{Gram}\big(\,|\sigma_x\rangle : x\in\mathcal{X}\big).$$

Indeed, this representation is even more convenient to study the evolution of a quantum state during the processing of a quantum query algorithm. More precisely, if we apply a unitary transformation U independent of the input on a set of quantum states $(\rho_x)_x$, their Gram matrix ρ does not change,

$$\forall x,y\in\mathcal{X}\qquad \big\langle U\rho_x,U\rho_y\big\rangle=\big\langle\rho_x,\rho_y\big\rangle=\rho[x,y].$$

In contrast the discrete-time model an oracle call $\mathcal{O}_x$, as dependent of x, does change ρ,

$$\forall x,y\in\mathcal{X}\qquad \big\langle\mathcal{O}_x\rho_x,\mathcal{O}_y\rho_y\big\rangle=\big\langle\rho_x,\mathcal{O}_x^*\mathcal{O}_y\rho_y\big\rangle.$$

Consequently in the discrete-time model where a quantum query algorithm is a sequence of unitary operator and oracle call, the **Gram matrix representation** allows to simplify analyses of the quantum query complexity, since we only consider oracle calls.

Remark. In the continuous-time model we show in Chapter 8 that the same behavior appears, i.e. the action of the driver Hamiltonian is canceled in the Gram matrix representation.

As remarked above, in Formula (4.2) vectors $|\rho_{x;k}\rangle$ indicate which "fraction" queries will be implemented after an oracle call. In other words, these vectors decide of the action of the oracle on ρ. These vectors are not fixed since we can modify them with a unitary operator independent of x, but they are still dependent of k. So we define for each k the following Gram matrix,

$$\forall x,y\in\mathcal{X},\qquad \rho_k=\mathrm{Gram}\big(\,|\rho_{x;k}\rangle : x\in\mathcal{X}\big),\tag{4.10}$$

and from Formula (4.2), they naturally satisfy

$$\rho=\sum_{k\in\{0\ldots n\}}\rho_k.\tag{4.11}$$

Some questions arise; Does every set of Gram matrices $(\sigma_k)_k$ satisfying Condition (4.11) represents an action of the oracle on ρ? Does there exist a unitary operator independent of x to change the action of the oracle from $(\rho_k)_k$ to $(\sigma_k)_k$?

The first answer is affirmative, from the definition of a Gram matrix in Section 2.1.3. The second answer is also affirmative, but it requires Fact 2.1.2 where we replace v_i with $\sum_{k\in K}|k\rangle\otimes|\rho_{x;k}\rangle$, and w_i with $\sum_{k\in K}|k\rangle\otimes|\sigma_{x;k}\rangle$.

Corollary 4.2.1. *Let $K, \mathcal{X}$ be finite sets, $\mathbb{C}^{|K|} \otimes \mathcal{H}$ be a finite Hilbert space, and $|k\rangle_{k\in K}$ be a basis of $\mathbb{C}^{|K|}$. For $(\rho_{x;k})_{x;k}$ and $(\sigma_{x;k})_{x;k}$ two families of the vectors in $\mathcal{H}$ indexed by $\mathcal{X} \times K$. If $\sum_{k\in K} Gram(\rho_{x;k} : x \in \mathcal{X}) = \sum_{k\in K} Gram(\sigma_{x;k} : x \in \mathcal{X})$, then there exists a unitary operator U such that, $U\big(\sum_{k\in K} |k\rangle \otimes \rho_{x;k}\big) = \sum_{k\in K} |k\rangle \otimes \sigma_{x;k}$, for all $x \in \mathcal{X}$.*

A main interest of the Gram matrix representation is that any discrete-time quantum query algorithm can be interpreted as a discrete path in the space of unitary Gram matrices:

$$G_N = \{\gamma \in \mathcal{M}_N(\mathbb{C}) : \gamma \circ \mathcal{I}d = \mathcal{I}d, \text{ and } \gamma \geq 0\}.$$

Hence, instead of looking at a quantum query algorithm $\mathcal{A}$ that converts ρ in σ, we can consider a discrete path $\rho(t) : \{0, \frac{1}{T} \dots \frac{T-1}{T}, 1\} \to G_N$, such that $\rho(0) = \rho$ and $\rho(T) = \sigma$. In the same way, a continuous quantum query algorithm that converts ρ in σ can be interpreted as a differentiable[1] path $\rho(t) : [0, T] \to G_N$, such that $\rho(0) = \rho$ and $\rho(T) = \sigma$. Depending on the time model, we either use a discrete path $\rho(t)$ with a finite domain $\{0, \frac{1}{T} \dots \frac{T-1}{T}, 1\}$, to describe a discrete-time algorithm with a running time T, or a differentiable path $\rho(t)$ to describe a continuous-time algorithm. Let's denote the set of all possible discrete paths,

$$\Gamma_{\text{dt}}[\rho \to \sigma] = \bigcup_{T\in\mathbb{N}} \left\{\gamma(t) \in F\big(\{0, \frac{1}{T} \dots \frac{T-1}{T}, 1\}, G_N\big) \;:\; \gamma(0) = \rho \text{ and } \gamma(1) = \sigma\right\},$$

and the set of all possible differentiable paths,

$$\Gamma_{\text{ct}}[\rho \to \sigma] = \bigcup_{T\in\mathbb{R}_+} \big\{\gamma(t) \in \mathcal{C}^1\big([0, T], G_N\big) \;:\; \gamma(0) = \rho \text{ and } \gamma(1) = \sigma\big\}.$$

If a query algorithm can be represented by a path, unfortunately the reciprocal is not true. A possible path in $\Gamma_\bullet[\rho \to \sigma]$ is not necessary "feasible", i.e. a path cannot ensure the existence of a quantum query algorithm which generates this path.

For the purpose of distinguishing feasible paths from unfeasible paths, we use the work of Barnum, Saks and Szegedy in [BSS03]. In this article, for a binary alphabet and the phase oracle representation, they construct a semi-definite program (cf. Chapter 6) that accepts a path, if and only if some conditions are satisfied. To obtain this semi-definite program, they prove that a unitary Gram matrix ρ can evolve into another ρ^+ with only one query, if and only if ρ and ρ^+ satisfy precise conditions.

Proposition 4.2.2. *[BSS03] Let ρ and ρ^+ be unitary Gram matrices. We can transform ρ to ρ^+ with one query of $\mathcal{O}_x^{\text{ph}}$, if and only if there exists a set of positive semi-definite matrices $(\rho_k)_{k\in\{0\dots n\}}$ such that*

$$\rho = \sum_{k\in\{0\dots n\}} \rho_k \qquad \textit{and} \qquad \rho^+ = \sum_{k\in\{1\dots n\}} \rho_k \circ \Delta_k^{\text{ph}}, \tag{4.12}$$

where for all $k \in \{1 \dots n\}$, $\Delta_k^{\text{ph}} = (-1)^{x_k - y_k}$.

In the above Proposition, ρ_k refers to the Gram matrix of the vectors $(|\rho_{x;k}\rangle)_x$ as defined in Formula (4.10). Hence, every query step described in a discrete path can be checked one by one using Proposition 4.2.2. If a step is feasible, then a solution of the semi-definite program gives a

[1] As the path is generated by integration of Schrödinger's equation, the path is differentiable according to the Fundamental theorem of calculus.

direction $(\rho_k)_k$ to implement this step, otherwise we reject the discrete path. Note that we know this step is implementable from Corollary 4.2.1.

Consequently, in the discrete model we can consider discrete paths satisfying Conditions 4.12 instead of a sequence of unitary operators and oracles $\mathcal{O}_x$. In Chapter 8, we show an equivalent Proposition (4.2.2) in the continuous-time model using the phase-Hamiltonian as representation.

Finally, the Gram matrix representation has also drawbacks since we don't work directly with an algorithm, so it may be difficult to reconstruct a quantum query algorithm from a path.

4.3 Lower bound methods

To the purpose of analyzing the quantum query complexity, several methods have been developed to lower bound the quantum query complexity of a problem; function, state generation, state conversion. In this section we introduce three main methods:

- polynomial method,
- adversary method,
- multiplicative adversary method.

4.3.1 Polynomial method

The polynomial method has been introduced in [BBC+01] for discrete-time model with a binary alphabet to evaluate a boolean function $f : \mathcal{X} \to \{0,1\}$. This method comes from the simple idea that the action of an oracle $\mathcal{O}_x$ can be described as a polynomial of degree one, with $\mathcal{O}_x$ acting as the map (4.1).

$$\mathcal{O}_x \sum_{\substack{k\in\{1\ldots n\}\\ b\in\{0,1\}}} \rho_{k,b}\,|k,b\rangle = \sum_{\substack{k\in\{1\ldots n\}\\ b\in\{0,1\}}} \left[\rho_{k,b}(1-x_k) + \rho_{k,b\oplus 1}x_k\right]|k,b\rangle\,.$$

On another side, U a unitary operator independent of x only mixes vectors $\rho_{k,b}$ without increasing polynomial degrees, since U is linear. Consequently, every quantum query algorithm using T queries, outputs a quantum state $|\sigma_x\rangle$ with the following form

$$|\sigma_x\rangle = \sum_{\substack{k\in\{1\ldots n\}\\ b\in\{0,1\}}} P_{k,b}[x_1\ldots x_n]\,|k,b\rangle\,,$$

where $P_{k,b}[x_1\ldots x_n]$ is a multi-linear polynomials with degree at most T. Let C be a strict subset of $\{1\ldots n\}\times\{0,1\}$, therefore the probability to observe $(k,b)\in C$ is

$$p[x_1\ldots x_n] = \sum_{(k,b)\in C} \left|P_{k,b}[x_1\ldots x_n]\right|^2,$$

a multi-linear polynomial with degree at most $2T$.

This result is quite powerful since, for each quantum query algorithm using T queries and computing exactly f, there exists a multi-linear polynomial p with degree at most $2T$ such that, $p = f$.

Definition 4.3.1. Let $f(x)$ be a boolean function. We denote $\deg(f)$ the minimum degree over all polynomials $p(x)$, such as $p(x) = f(x)$ for all x.
We define $\widetilde{\deg}(f)$ to be the minimum degree over all polynomials $p(x)$, such as $|p(x) - f(x)| < 1/3$ for all x.

Theorem 4.3.2. *[BBC+01] Let f be a boolean function, therefore*

$$\frac{1}{2}\deg(f) \leq Q_0^{dt}(f)$$
$$\frac{1}{2}\widetilde{\deg}(f) \leq Q_{1/3}^{dt}(f).$$

The proof is straightforward using the principle of contradiction. If the inequality is violated, then there exists a quantum query algorithm using T queries and computing exactly f and no multi-linear polynomial p with degree at most $2T$ such that, $p = f$.

4.3.2 Adversary method

The adversary method, denoted Adv^+, was originally introduced by A. Ambainis in [Amb00]. Later P. Høyer, T. Lee, and R. Špalek in [HLS07] improve the adversary method by adding negative weights, now called the 'general adversary method' and denoted $\mathrm{Adv}^{\pm}$. Finally in the article [LMR+11], the adversary method $\mathrm{Adv}^{\pm}$ has been adapted to state conversion problem by constructing another method, denoted Adv and called the 'query distance'. $\mathrm{Adv}^{\pm}$ and Adv are distinct, but restricted to function evaluation problems they are equivalent by a factor at most 2. A lot of adversary methods appear in the literature, but R. Špalek and M. Szegedy have proved that they are all equivalent [ŠS05].

This method can be explained in two steps. First, for every quantum query algorithm, we consider its path $\rho(t)$ in the Gram space $\Gamma[\rho \to \sigma]$, and we choose a unit vector v, therefore $\rho(t) \circ vv^*$ can be interpreted as a density operator. Second, we choose an observable M such that the change of its average value

$$\langle M \rangle_t = \langle M, \rho(t) \circ vv^* \rangle,$$

after an oracle call is bounded. The change of average value $\langle M \rangle_t$ is bounded differently depending on the model,

$$\text{(discrete model)} \qquad \left| \langle M \rangle_{t+1} - \langle M \rangle_t \right| \leq 1, \tag{4.13}$$

$$\text{(continuous model)} \qquad \left| \frac{d\langle M \rangle_t}{dt} \right| \leq 1. \tag{4.14}$$

Hence for every observable M that satisfies conditions above, we obtain by integrating the bound

$$|\langle M, (\sigma - \rho) \circ vv^* \rangle| = |\, \langle M \rangle_T - \langle M \rangle_0 \,| = \left| \sum_{t=0}^{T-1} \langle M \rangle_{t+1} - \langle M \rangle_t \right| \leq T,$$
$$|\langle M, (\sigma - \rho) \circ vv^* \rangle| = |\, \langle M \rangle_T - \langle M \rangle_0 \,| = \left| \int_0^T dt \frac{d\langle M \rangle_t}{dt} \right| \leq T.$$

This bound holds for all v and M satisfying Condition (4.13) or (4.14), therefore we can maximize over v and M.

This is the general idea, but we still need to reformulate Conditions (4.13) and (4.14) independently of the path $\rho(t)$. More precisely, we can show that these Conditions (4.13) and (4.14) are respected, if M satisfies particular conditions with Δ_k or $\hat{\Delta}_k$.

Here we derive the adversary method for a binary alphabet in the discrete-time model using the phase-oracle. From Proposition 4.2.2 and with matrices $\Delta_k^{\mathrm{ph}} = (-1)^{x_k - y_k}$, we show the following Lemma.

Lemma 4.3.3. *Let M be an observable, v be a unitary vector and $\rho(t)$ be a discrete feasible path. If*

$$M - \mathcal{I}d \leq M \circ \Delta_k^{\mathrm{ph}} \leq M + \mathcal{I}d \qquad \textit{for all } k \in \{1 \dots n\}, \tag{4.15}$$

then,

$$\left| \langle M \rangle_{t+1} - \langle M \rangle_t \right| \leq 1.$$

Proof. As $\rho(t)$ is feasible, from Proposition 4.2.2 we know that for all $t \in \{0 \dots T-1\}$ there exists $(\rho_k(t))_k$ such that

$$\rho(t) = \sum_{k \in \{0 \dots n\}} \rho_k(t) \qquad \text{and} \qquad \rho(t+1) = \sum_{k \in \{1 \dots n\}} \rho_k(t) \circ \Delta_k^{\mathrm{ph}}.$$

Therefore for all $t \in \{0 \dots t-1\}$,

$$\begin{aligned} \langle M \rangle_{t+1} &= \left\langle M, \rho(t+1) \circ vv^* \right\rangle, \\ &= \left\langle M, \sum_{k \in \{1 \dots n\}} \rho_k(t) \circ \Delta_k^{\mathrm{ph}} \circ vv^* \right\rangle, \\ &= \sum_{k \in \{1 \dots n\}} \left\langle M \circ \Delta_k^{\mathrm{ph}}, \rho_k(t) \circ vv^* \right\rangle, \\ &\leq \sum_{k \in \{1 \dots n\}} \left\langle M + \mathcal{I}d, \rho_k(t) \circ vv^* \right\rangle, \\ &\leq \left\langle M, \rho(t) \circ vv^* \right\rangle + \left\langle \mathcal{I}d, \rho(t) \circ vv^* \right\rangle, \\ &\leq \langle M \rangle_t + 1, \end{aligned}$$

the inequality comes from $\rho_k(t) \circ vv^* \geq 0$ and $M \circ \Delta_k^{\mathrm{ph}} \leq M + \mathcal{I}d$. Moreover, as $\rho_k(t) \circ vv^*$ is a density matrix, its trace is equal to one. The other inequality can be proved similarly. □

Since Condition (4.13) is implied by Conditions (4.15) in Lemma 4.3.3, we define the adversary method with and without error ε.

Definition 4.3.4 (Adversary method for discrete time)**.** [LR12]

$$\begin{aligned} \mathrm{Adv}_0^\star(\rho \to \sigma) = \sup_{\substack{M \\ v:\|v\|=1}} &\left\langle M \circ vv^*, \sigma - \rho \right\rangle, \\ \text{subject to} \quad &\forall k \in \{1 \dots n\}, \qquad M - \mathcal{I}d \leq M \circ \Delta_k^\star \leq M + \mathcal{I}d. \end{aligned}$$

$$\mathrm{Adv}_\varepsilon^\star(\rho \to \sigma) = \inf_{\sigma' : \mathcal{F}_H(\sigma, \sigma') \geq \sqrt{1-\epsilon}} \mathrm{Adv}^\star(\rho \to \sigma'),$$

where the superscript $\star$ is either ph or reg.

Remark. In the above definition the absolute value has been removed since Conditions (4.15) are symmetric under sign change. We have provided two definitions: $\mathrm{Adv}^{\mathrm{ph}}$ and $\mathrm{Adv}^{\mathrm{reg}}$ depending of the representation of the oracle. $\mathrm{Adv}^{\mathrm{reg}}$ is the adversary method introduced in [LMR⁺11], a proof that $\mathrm{Adv}^{\mathrm{reg}}$ lower bounds Q_0^{dt} can be found in their article. Finally, Condition (4.15) in Definition 4.3.4 can be re-written for each k's as,

$$M - \mathcal{I}d \leq M \circ \Delta_k^\star \leq M + \mathcal{I}d \qquad \Longleftrightarrow \qquad \left\| M \circ (\Delta_k^\star - \mathbb{J}) \right\| \leq 1. \tag{4.16}$$

Theorem 4.3.5. *[LMR⁺11] Let ρ and σ be two unitary Gram matrices, and $\varepsilon \in]0,1]$. We have*

$$\mathrm{Adv}_0^\star(\rho \to \sigma) \leq Q_0^{\mathrm{dt}}(\rho \to \sigma),$$
$$\mathrm{Adv}_\varepsilon^\star(\rho \to \sigma) \leq Q_\varepsilon^{\mathrm{dt}}(\rho \to \sigma).$$

Surprisingly, $\mathrm{Adv}^{\mathrm{reg}}$ is also a lower bound method for Q_0^{ct}.

Theorem 4.3.6. *[YM11]*
Let ρ and σ be two unitary Gram matrices, and $\varepsilon \in]0,1]$. We have

$$\frac{1}{2}\mathrm{Adv}_0^{\mathrm{reg}}(\rho \to \sigma) \leq Q_0^{ct}(\rho \to \sigma),$$
$$\frac{1}{2}\mathrm{Adv}_\varepsilon^{\mathrm{reg}}(\rho \to \sigma) \leq Q_\varepsilon^{ct}(\rho \to \sigma).$$

In Chapter 7, we provide an original proof independent of the choice $h(l)$'s in Formula (4.3).

An important result for this subsection. In [LMR⁺11], authors show that $\mathrm{Adv}_\varepsilon^{\mathrm{reg}}$ characterizes the bounded-error quantum query complexity $Q_\varepsilon^{dt}(\rho \to \sigma)$.

Theorem 4.3.7. *[LMR⁺11] Let ρ and σ be two unitary Gram matrices, and $\varepsilon > 0$. We have*

$$Q_\varepsilon^{dt}(\rho \to \sigma) = \Theta\left(\frac{\mathrm{Adv}_\varepsilon^{\mathrm{reg}}(\rho \to \sigma)}{\varepsilon}\right).$$

4.3.3 Multiplicative adversary method

The multiplicative adversary method $\mathrm{Madv}^{\mathrm{reg}}$ introduced by R. Špalek in [Špa08], subsumes both polynomial method and adversary methods [MR13]. Although this method is less convenient to use, it is quite powerful and allowed to prove a strong direct product theorem for quantum query complexity [LR12].

As for the adversary method, we define three versions: $\mathrm{Madv}^{\mathrm{ph}}$, $\mathrm{Madv}^{\mathrm{reg}}$ and $\mathrm{Madv}^{\mathrm{ct}}$. Here we only give a proof that $\mathrm{Madv}^{\mathrm{ph}}$ is a lower bound of Q^{dt}, with a binary alphabet and the phase-oracle representation. A demonstration for $\mathrm{Madv}^{\mathrm{reg}}$ can be found in the original article [Špa08]. Finally, as the multiplicative adversary method $\mathrm{Madv}^{\mathrm{ct}}$ for a continuous-time model is new, we only gives its definition for the moment, we will provide in Chapter 8 a demonstration that this new method is a lower bound of Q^{ct}.

This method is based on the same idea that adversary methods, we choose a unitary vector v and an observable M, but we chose different restrictions for $\langle M \rangle_t$,

$$|\langle M \rangle_{t+1}| \leq c \langle M \rangle_t\,, \tag{4.17}$$

where $c > 1$ is a real, and $M \geq 0$ such as $\langle M \rangle_t \geq 0$ for all t. From this relation we derive the following lower bound,

$$\frac{1}{\ln c} \big[\ln \langle M \rangle_T - \ln \langle M \rangle_0 \big] \leq T.$$

As in the previous Subsection, we must express Condition 4.17 independently of $\rho(t)$. From Proposition 4.2.2 for a binary alphabet and using a phase-oracle representation, we can show that

Lemma 4.3.8. *Let $M \geq 0$ be an observable, $c > 1$ be a real, v be a unitary vector and $\rho(t)$ be a discrete feasible path. If*

$$c^{-1} M \leq M \circ \Delta_k^{\mathrm{ph}} \leq cM \qquad \textit{for all } k \in \{1 \dots n\}, \tag{4.18}$$

then,

$$\langle M \rangle_{t+1} \leq c \, \langle M \rangle_t \, .$$

Proof. From Proposition 4.2.2, we know that for all $t \in \{0 \dots T-1\}$ there exists $(\rho_k(t))_k$ such that

$$\rho(t) = \sum_{k \in \{0 \dots n\}} \rho_k(t) \qquad \text{and} \qquad \rho(t+1) = \sum_{k \in \{1 \dots n\}} \rho_k(t) \circ \Delta_k^{\mathrm{ph}}.$$

Then for all $t \in \{0 \dots t-1\}$,

$$\begin{aligned}
\langle M \rangle_{t+1} &= \big\langle M, \rho(t+1) \circ vv^* \big\rangle, \\
&= \Big\langle M, \sum_{k \in \{1 \dots n\}} \rho_k(t) \circ \Delta_k^{\mathrm{ph}} \circ vv^* \Big\rangle, \\
&= \sum_{k \in \{1 \dots n\}} \big\langle M \circ \Delta_k^{\mathrm{ph}}, \rho_k(t) \circ vv^* \big\rangle, \\
&\leq \sum_{k \in \{1 \dots n\}} \big\langle cM, \rho_k(t) \circ vv^* \big\rangle, \\
&\leq c \big\langle M, \rho(t) \circ vv^* \big\rangle, \\
&\leq c \, \langle M \rangle_t \, ,
\end{aligned}$$

the inequality comes from $\rho_k(t) \circ vv^* \geq 0$ and $M \circ \Delta_k^{\mathrm{ph}} \leq cM$. The other inequality can be proved similarly. □

Since Condition 4.17 is satisfied by Condition 4.18 in Lemma 4.3.8, we define the multiplicative adversary method with and without error ε.

Definition 4.3.9 (Multiplicative adversary method for discrete time)**.** [LR12]

$$\begin{aligned}
\mathrm{Madv}_0^\star(\rho \to \sigma) = &\sup_{c>1} \frac{1}{\ln c} \sup_{\substack{M \geq 0 \\ v : \|v\|=1}} \big[\ln \big\langle M \circ vv^*, \sigma \big\rangle - \ln \big\langle M \circ vv^*, \rho \big\rangle \big], \\
&\text{subject to} \quad \forall\, k \in \{1 \dots n\}, \qquad c^{-1} M \leq M \circ \Delta_k^\star \leq cM.
\end{aligned}$$

$$\mathrm{Madv}_\varepsilon^\star(\rho \to \sigma) = \inf_{\sigma' : \mathcal{F}_H(\sigma, \sigma') \geq \sqrt{1-\epsilon}} \mathrm{Madv}_0^\star(\rho \to \sigma'),$$

where the superscript $\star$ is either ph or reg.

So from Proposition 4.2.2 and Lemma 4.3.8 we conclude that,

Theorem 4.3.10. *[LR12]*
Let ρ and σ be two unitary Gram matrices, and $\varepsilon \in]0,1]$. We have

$$\begin{aligned}\mathrm{Madv}_0^\star(\rho \to \sigma) &\leq Q_0^{\mathrm{dt}}(\rho \to \sigma),\\ \mathrm{Madv}_\varepsilon^\star(\rho \to \sigma) &\leq Q_\varepsilon^{\mathrm{dt}}(\rho \to \sigma).\end{aligned}$$

Remark. It is easy to see that Madv beats Adv, from Conditions (4.13) and (4.17). By choosing $c = 1 + \delta$ with δ closed to zero, such as $T\delta$ stays small,

$$\begin{aligned}\frac{\langle M \rangle_{t+1}}{\langle M \rangle_t} &\leq c = 1 + \delta,\\ \frac{\langle M \rangle_T}{\langle M \rangle_0} &\leq c^T \simeq 1 + T\delta,\\ \ln \langle M \rangle_T - \ln \langle M \rangle_0 &\lesssim T\delta,\\ \frac{1}{\ln c}\big[\ln \langle M \rangle_T - \ln \langle M \rangle_0\big] &\lesssim T,\end{aligned}$$

with $\ln c \simeq \delta$. As we can see, for c closed to one the multiplicative method acts the Adversary method, where each multiplication by c can be approximated by adding $\ln c$.

Chapter 5

Communication complexity

The communication model has been introduced in '79 by A. Yao [Yao79]. It describes a system that computes a task distributed over several parties of the system, where each part owns resources to complete this task. In the mathematics formalism, a task over n parties is represented as a function $f : X_1 \times \ldots \times X_n \to Z$ where Z and all X_i's are finite sets. The model is simple, at the beginning each part or player has an element $x_i \in X_i$, then they start to communicate and share information to know $f(x_1, \ldots, x_n)$. The number of communications needed to compute a well-defined task is the resource we wish to estimate in this model.

In the thesis, we only work with the two party model $f : X \times Y \to Z$, called Alice and Bob by convention. The communication between Alice and Bob is done through binary messages. To go further, the book "Communication complexity" from Kushilevitz and Nisan [KN97] is a good reference.

5.0.1 Communication model

In the communication model, the two players computing a function $f(x, y)$ do not speak arbitrarily, but one bit after one bit, and moreover they follow a protocol.
A **deterministic protocol** $\mathcal{P}$ with domain $X \times Y$ and range Z is a binary decision tree, where each internal node v is attached to Alice with function $a_v : X \to \{0, 1\}$ or to Bob with function $b_v : Y \to \{0, 1\}$. These functions decide respectively which bit Alice or Bob sends to each other. Finally, each leaf of $\mathcal{P}$ is labeled by an element in Z.

The **communication cost** $CC(\mathcal{P})$ of a deterministic protocol $\mathcal{P}$ is equal to its depth denoted $|\mathcal{P}|$. A deterministic protocol $\mathcal{P}_f$ implements the function f, if $\mathcal{P}_f(x, y) = f(x, y)$ for all x, y. The **communication complexity** $CC(f)$ of a function f is the minimal depth over all deterministic protocols that implements f

$$CC(f) = \min_{\mathcal{P}_f} CC(\mathcal{P}_f).$$

Let μ be a probability distribution over $X \times Y$, called **input distribution**, we define $\mathcal{P}^{\varepsilon}_{\mu,f}$ to be a deterministic protocol that computes correctly f on a fraction of $X \times Y$ with a measure at least $1-\varepsilon$, as

$$\begin{aligned}\Pr[\mathcal{P}^{\varepsilon}_{\mu,f}(x,y)=f(x,y)] &= \sum_{x,y}\mu(x,y)\delta[\mathcal{P}^{\varepsilon}_{\mu,f}(x,y),f(x,y)],\\ \Pr[\mathcal{P}^{\varepsilon}_{\mu,f}(x,y)=f(x,y)] &\geq 1-\varepsilon.\end{aligned}$$

The **distributional complexity** $CC^{\mu}_{\varepsilon}(f)$ of a function f is defined by

$$CC^{\mu}_{\varepsilon}(f)=\min_{\mathcal{P}^{\varepsilon}_{\mu,f}} CC(\mathcal{P}^{\varepsilon}_{\mu,f}).$$

A **private coin randomized protocol** $\mathcal{P}^{\varepsilon,priv}$ with ε-error is a binary decision tree similar to a deterministic protocol, except that its domain is $X\times Y\times R_A\times R_B$ where R_A and R_B are random variables, and different functions $a_v : X\times R_A\to\{0,1\}$ and $b_v : Y\times R_B\to\{0,1\}$. We observe that a private coin randomized protocol acts exactly as a deterministic protocol $\mathcal{P}^{\varepsilon,priv}(\,\cdot\,,\,\cdot\,,r_A,r_B)$, once random variables R_A and R_B has been fixed. A private coin randomized protocol $\mathcal{P}^{\varepsilon,priv}_f$ implements the function f, if for all $(x,y)\in X\times Y$

$$\begin{aligned}\Pr[\mathcal{P}^{\varepsilon,priv}_f(x,y)=f(x,y)] &= \sum_{r_A,r_B} p_{R_A}(r_A)p_{R_B}(r_B)\delta[\mathcal{P}^{\varepsilon,priv}_f(x,y,r_A,r_B),f(x,y)],\\ \Pr[\mathcal{P}^{\varepsilon,priv}_f(x,y)=f(x,y)] &\geq 1-\varepsilon.\end{aligned}$$

Similarly, we define $CC(\mathcal{P}^{\varepsilon,priv})$ to be the communication cost of a private coin randomized protocol as the worst communication cost of $\mathcal{P}^{\varepsilon,priv}(\,\cdot\,,\,\cdot\,,r_A,r_B)$ over all r_A,r_B, likewise we define

$$R^{priv}_{\varepsilon}(f)=\min_{\mathcal{P}^{\varepsilon,priv}_f}\max_{r_A,r_B} CC\big(\mathcal{P}^{\varepsilon,priv}_f(\,\cdot\,,\,\cdot\,,r_A,r_B)\big).$$

A **public coin randomized protocol** $\mathcal{P}^{\varepsilon}$ with ε-error is a binary decision tree similar to a deterministic protocol, except that its domain is $X\times Y\times\Omega$ where Ω is random variable, different functions $a_v : X\times\Omega\to\{0,1\}$ and $b_v : Y\times\Omega\to\{0,1\}$. A public coin randomized protocol is like the private version, but Alice and Bob share a unique random variable Ω. A public coin randomized protocol $\mathcal{P}^{\varepsilon}_f$ implements the function f, if for all $(x,y)\in X\times Y$

$$\Pr[\mathcal{P}^{\varepsilon}_f(x,y)=f(x,y)] = \sum_{\omega} p_{\Omega}(\omega)\delta[\mathcal{P}^{\varepsilon}_f(x,y,\omega),f(x,y)], \tag{5.1}$$

$$\Pr[\mathcal{P}^{\varepsilon}_f(x,y)=f(x,y)] \geq 1-\varepsilon. \tag{5.2}$$

The communication cost $CC(\mathcal{P}^{\varepsilon})$ of a public coin randomized protocol is the worst communication cost of $\mathcal{P}^{\varepsilon}(\,\cdot\,,\,\cdot\,,\omega)$ over all ω's

$$R_{\varepsilon}(f)=\min_{\mathcal{P}^{\varepsilon}_f}\max_{\omega} CC\big(\mathcal{P}^{\varepsilon}_f(\,\cdot\,,\,\cdot\,,\omega)\big).$$

From these definitions, we can directly establish relations between these complexity. Let $f : X\times Y\to Z$ be a function,

$$CC^{\mu}_{\varepsilon}(f)\leq R_{\varepsilon}(f)\leq R^{priv}_{\varepsilon}(f)\leq CC(f). \tag{5.3}$$

Theorem 5.0.1. *(Yao's principle)[Yao77]*
Let $f : X\times Y\to Z$ be a function and μ be an input distribution over $X\times Y$. We have

$$R_{\varepsilon}(f)=\max_{\mu} CC^{\mu}_{\varepsilon}(f).$$

Yao's principle allows us to lower bound $R_\varepsilon(f)$ by lower bounding $CC^\mu_\varepsilon(f)$, then maximizing over all input distributions μ's. We do not provide a proof, just an overview.

The proof is based on von-Neumann's minimax theorem. A randomized protocol can be interpreted as a two party zero-sum game, where the first player $\mathcal{A}$ choses an input (x, y) and the second player $\mathcal{B}$ choses a deterministic protocol $\mathcal{P}_{f,\omega}$, where $\mathcal{A}$ wins when $\mathcal{P}_{f,\omega}(x, y) = f(x, y)$ and loses otherwise. Thus a strategy for $\mathcal{A}$ is an input distribution μ over $X \times Y$, and a strategy for $\mathcal{B}$ is a probability distribution p_Ω over deterministic protocols.

5.1 Lower bound methods

In this section, we show that a deterministic protocol $\mathcal{P}_f$ naturally induces a partition of monochromatic rectangles on **dom** f.

Partition representation

A subset $R \subset X \times Y$ is a **rectangle**, if there exists $X' \subset X$ and $Y' \subset Y$ such that $R = X' \times Y'$. For the same rectangle R, we denote $R[1] = X'$ and $R[2] = Y'$. Moreover we introduce the notation: $x \in_1 R$ if $x \in R[1]$, and $y \in_2 R$ if $y \in R[2]$. For a function f, a subset $S \subset \mathbf{dom} f$ is f**-monochromatic**, if f is constant on S. If R is f-monochromatic, the notation $f(R)$ indicates the unique value of f on R.

Lemma 5.1.1. *[Yao79]*
Let $f : X \times Y \to Z$ be a function. A deterministic protocol $\mathcal{P}_f$ induces a partition of f-monochromatic rectangles on $X \times Y$.

To prove Lemma 5.1.1 we use the fact that the intersection of two rectangles is still a rectangle. Indeed for two rectangles R_1 and R_2 such that $R_1 = X_1 \times Y_1$ and $R_2 = X_2 \times Y_2$, we have $R_1 \cap R_2 = (X_1 \cap X_2) \times (Y_1 \cap Y_2)$.

Proof. Let v be a node in the binary decision tree $\mathcal{P}_f$. We define R_v to be the set of inputs that reach this node. For any deterministic protocol is deterministic, every input $(x, y) \in X \times Y$ reaches a unique leaf l, then $(R_l)_{l \in L}$ is a partition of $X \times Y$, where L is the set of leaves. Moreover, from the definition of $\mathcal{P}_f$ every R_l is f-monochromatic.
We must still show that every R_l is a rectangle, to prove it we use a recursion over the binary tree from the root down to leaves.

- let t be the root of the decision tree, then $R_t = X \times Y$,
- let v be an internal node attached to a function a_v with the assumption that R_v is rectangle, and v_0, v_1 be child-nodes of v, therefore

  $$R_{v_0} = R_v \cap \Big(\big\{x : a_v(x) = 0\big\} \times Y \Big),$$
  $$R_{v_1} = R_v \cap \Big(\big\{x : a_v(x) = 1\big\} \times Y \Big),$$

  as the rectangle property is preserved under intersection, then R_{v_0} and R_{v_1} are rectangles,
- if a internal node v is attached to a function b_v, we can use the same method that for a_v.

□

This simple Lemma allows to derive a lower bound on the communication complexity. Since a deterministic protocol induces a partition of $X \times Y$ with m rectangles, then this deterministic protocol must use at least $\lceil \log m \rceil$ bits of communication to discriminate all rectangles. Let $C^D(f)$ be the smallest f-monochromatic partition on $X \times Y$, called the **partition number**. We have

$$\log C^D(f) \leq CC(f). \tag{5.4}$$

Lemma 5.1.1 is not reciprocal. A f-monochromatic partition on $X \times Y$ does not imply the existence of a deterministic protocol $\mathcal{P}_f$ that induces this partition. Hence, the lower bound (5.4) is not tight [GPW15].

5.1.1 Discrepancy

The **discrepancy** is a lower bound method based on Lemma (5.1.1). Let $S(f)$ be the set of all f-monochromatic rectangles on $\mathbf{dom} f$, we define

$$\mathbf{disc}(f) = \frac{1}{|\mathbf{dom}| f} \max_{R \in S(f)} |R|,$$

the size of the largest f-monochromatic rectangle divided by $\mathbf{dom} f$. Thus, every partition of f-monochromatic rectangles needs at least $\lceil \mathbf{disc}(f) \rceil$ rectangles, and using the same argument as for (5.4), we obtain

$$-\log \mathbf{disc}(f) \leq CC(f). \tag{5.5}$$

This method can be adapted for $R_\varepsilon(f)$ using the Yao's principle 5.0.1. Without giving details, we define

$$\mathbf{disc}_\mu(f) = \max_{\substack{z \in Z \\ R \in S(f)}} \Big| \sum_{(x,y) \in R} \mu(x,y).(-1)^{\delta[z,f(x,y)]} \Big|.$$

Where we choose a rectangle with the best trade-off between its size and its f-monochromaticity. Then we obtain

$$\log \frac{(1-2\varepsilon)}{\mathbf{disc}_\mu(f)} \leq CC_\varepsilon^\mu(f).$$

5.1.2 Partition bound

The partition bound $\mathbf{prt}_\varepsilon(f)$ is lower bound method for $R_\varepsilon(f)$, introduced by R. Jain and H. Klauck in [JK10]. This method is like the lower bound (5.4) but adapted for $R_\varepsilon(f)$, and before going further we should find an equivalent Lemma 5.1.1.

Lemma 5.1.2. *Let $f : X \times Y \to Z$ be a function, and S_R be the set of all rectangle subsets of $X \times Y$. A public coin randomized protocol $\mathcal{P}_f^\varepsilon$ induces a weight $\xi_{R,z}$ on each couple $(R,z) \in S_R \times Z$ satisfying:*

(a) $\forall R \in S_R, \forall z \in Z, \qquad \xi_{R,z} \geq 0,$

(b) $\forall (x,y) \in \boldsymbol{dom}\ f, \qquad \sum_{R \in S_R : (x,y) \in R} \xi_{R,f(x,y)} \geq 1 - \varepsilon,$

(c) $\forall (x,y) \in X \times Y, \qquad \sum_{z \in Z} \sum_{R \in S_R : (x,y) \in R} \xi_{R,z} = 1.$

Proof. A public coin randomized protocol $\mathcal{P}_f^\varepsilon$ can be seen as a distribution of deterministic protocols $\mathcal{P}(\,\cdot\,,\,\cdot\,,\omega)$ that we simplify by $\mathcal{P}_\omega$. $p_\Omega(\omega)$ is a distribution relative to deterministic protocols $\mathcal{P}_\omega$. Hence, $\mathcal{P}_f^\varepsilon$ we can interpret as a distribution $p_\Omega(\omega)$ of partitions $P(\omega)$ of rectangles. For each rectangle $R \in P(\omega)$, we associated an output z such that $z = \mathcal{P}_\omega(R)$. Then we use the notation $(R,z) \in \hat{P}(\omega)$, where $\hat{P}(\omega)$ is a subset of $P(\omega \times Z)$ and implies that $R \in P(\omega)$ and $\mathcal{P}_\omega(R)$ outputs z.

As a protocol $\mathcal{P}_f^\varepsilon$ must satisfy Condition (5.1), we reformulate this condition under the following form

$$\forall (x,y) \in \mathbf{dom}\ f, \qquad \sum_{\substack{\omega\in\Omega \\ (R,z)\in\hat{P}(\omega):(x,y)\in R}} p_\Omega(\omega).\delta[z, f(x,y)] \geq 1-\varepsilon. \tag{5.6}$$

To clarify more precisely Equation 5.6 above, we eliminate the variable ω by defining $\hat{P} = \cup_\omega \hat{P}(\omega)$ and the weight $\lambda_{R,z}$ as,

$$\lambda_{R,z} = \sum_{\omega\in\Omega:(R,z)\in\hat{P}(\omega)} p_\Omega(\omega). \tag{5.7}$$

The weight $\lambda_{R,z}$ represents the probability of choosing a deterministic protocol $\mathcal{P}_\omega$ where the rectangle R is in the partition $P(\omega)$ and output z on R. Using this notation we can reformulate again Condition (5.6) as

$$\forall (x,y) \in \mathbf{dom}\ f, \qquad \sum_{\left(R,f(x,y)\right)\in\hat{P}:(x,y)\in R} \lambda_{R,f(x,y)} \geq 1-\varepsilon. \tag{5.8}$$

Also as a public coin randomized protocol always gives an output, we can show that the following equality is respected,

$$\forall (x,y), \qquad \sum_{(R,z)\in\hat{P}:(x,y)\in R} \lambda_{R,z} = \sum_{\omega\in\Omega} p_\Omega(\omega) \sum_{(R,z)\in\hat{P}(\omega):(x,y)\in R} 1 = 1. \tag{5.9}$$

Finally, as $\lambda_{R,z} \geq 0$ by definition, the choice $\xi_{R,z} = \lambda_{R,z}$ if $(R,z) \in \hat{P}$ and $\xi_{R,z} = 0$ otherwise, satisfies all required conditions. □

Now, we define the partition bound $\mathbf{prt}_\varepsilon(f)$.

Definition 5.1.3. (Partition bound)[JK10]
The partition bound of a function f with error ε, denoted $\mathbf{prt}_\varepsilon(f)$, is defined by the linear program

$$\mathbf{prt}_\varepsilon(f) = \min_{\xi_{R,z}\geq 0} \sum_{R\in S_R}\sum_{z\in Z} \xi_{R,z} \quad \text{s.t.} \quad \forall (x,y)\in \mathbf{dom}\ f, \quad \sum_{R:(x,y)\in R} \xi_{R,f(x,y)} \geq 1-\varepsilon,$$
$$\forall (x,y) \in X\times Y, \quad \sum_{z\in Z}\sum_{R:(x,y)\in R} \xi_{R,z} = 1.$$

Theorem 5.1.4. *[JK10]*
Let $f : X \times Y \to Z$ be a function. We have

$$\log \boldsymbol{prt}_\varepsilon(f) \leq R_\varepsilon(f).$$

Proof. Let $\mathcal{P}_f^\varepsilon$ be a public coin randomized protocol, therefore weights $\lambda_{R,z}$ in the proof of Lemma 5.1.2 is a feasible solution of the linear program $\mathbf{prt}_\varepsilon(f)$, according to Equations (5.8) and (5.9). Moreover, the sum of all $\lambda_{R,z}$ is equal to the average size of the partition $\hat{P}(\omega)$,

$$\sum_{(R,z)\in\hat{P}} \lambda_{R,z} = \sum_{\omega\in\Omega} p_\Omega(\omega) \sum_{(R,z)\in\hat{P}(\omega)} 1 = \sum_{\omega\in\Omega} p_\Omega(\omega)|\hat{P}(\omega)| = \left\langle |\hat{P}(\omega)| \right\rangle_{p_\Omega}.$$

Since the solution $\lambda_{R,z}$ is not necessary optimal,

$$\mathbf{prt}_\varepsilon(f) \leq \langle |P(\omega)| \rangle_{p_\Omega} \leq \max_\omega |\hat{P}(\omega)|.$$

Finally we conclude,

$$\log \mathbf{prt}_\varepsilon(f) \leq \max_\omega CC(\mathcal{P}_\omega) \leq CC(\mathcal{P}_f^\varepsilon).$$

For the first inequality we use the argument that, at least $\lceil \log |\hat{P}(\omega)| \rceil$ bits of communication is needed to discriminate all rectangles in $\hat{P}(\omega)$. The second inequality comes from the definition of $CC(\mathcal{P}_f^\varepsilon)$. □

The partition bound $\mathbf{prt}_\varepsilon(f)$ can be also be adapted to $CC_\varepsilon^\mu(f)$,

Definition 5.1.5. (Partition bound with distribution)[KLL+12]
The partition bound with distribution of f with error ε and input distribution μ, denoted $\mathbf{prt}_\varepsilon^\mu(f)$, is defined by the linear program

$$\mathbf{prt}_\varepsilon^\mu(f) \;=\; \min_{\xi_{R,z}\geq 0} \sum_{R\in S_R} \sum_{z\in Z} \xi_{R,z} \qquad \text{subject to,}$$

- $$\sum_{(x,y)\in \mathbf{dom}\ f} \mu(x,y) \sum_{R:(x,y)\in R} \xi_{R,f(x,y)} + \sum_{(x,y)\notin \mathbf{dom}\ f} \mu(x,y) \sum_{\substack{z\in Z \\ R:(x,y)\in R}} \xi_{R,z)} \geq 1-\varepsilon,$$
- $$\forall (x,y) \in X\times Y, \qquad \sum_{z\in Z} \sum_{R:(x,y)\in R} \xi_{R,z} = 1.$$

Although we do not provide a proof, we can observe that the error condition on each input (x,y) has relaxed to an average error condition relative to input distribution μ, and for each input outside of **dom** f Alice and Bob automatically succeed.

5.1.3 Information complexity

The communication exchanged between Alice and Bob can be represented by a **transcript** m: the concatenation of bits sent. For a randomized protocol $\mathcal{P}^\varepsilon$, this transcript is determined by the input (x,y) and, public and private coins Ω, then we can define a function $g : X\times Y\times\Omega \to M$, where M is the set of all possible transcripts. Moreover, when $X\times Y$ are random variables with the joint probability distribution μ, the function $g(x,y,\omega)$ is measurable function that induces a probability distribution π over M, such that

$$\forall m, \qquad \pi(m) = \sum_{\omega\in\Omega} p_\Omega(\omega).\mu\Big(\{(x,y)\in X\times Y : \mathcal{P}_\varepsilon(x,y,\omega) \text{ sends } m\}\Big).$$

Hence M is random variable with the distribution π. Hence, we can extend this probability distribution to $\Omega\times M\times X\times Y$, such that for all ω, m, x, y we have

$$\pi(x,y,\omega,m) = \mu(x,y).p_\Omega(\omega).\delta[\mathcal{P}_\varepsilon(x,y,\omega) \text{ sends } m]. \tag{5.10}$$

The definition of π can be generalized naturally to different protocols, as $\mathcal{P}^{\varepsilon}_{\mu}$ and $\mathcal{P}^{\varepsilon,priv}$.

In Information theory, the Shannon information of a message is always shorter than its size. Hence we can lower bound the communication complexity by a new resource, called **information complexity**. We define the **internal information cost** by,

$$\mathrm{IC}^{\mu}_{int}(\pi) = I(Y : M|X, \Omega) + I(X : M|Y, \Omega),$$

where the first term is the information than Alice learns from Bob's input, once the protocol over and the transcript known, and *vice-versa* for the second term.
Similarly, we define the information that a third party observing Alice and Bob learned of (x, y), once the transcript known.

$$\mathrm{IC}^{\mu}_{ext}(\pi) = I(X, Y : M, \Omega).$$

This is the **external information cost**.
The following Proposition shows that external information cost and internal information cost have a natural order.

Proposition 5.1.6. *[BR11]*
Let μ be an input distribution and π be a distribution induced by a protocol $\mathcal{P}^{\varepsilon}_{\mu}$. We have

$$IC^{\mu}_{int}(\pi) \le IC^{\mu}_{ext}(\pi) \le CC(\mathcal{P}^{\varepsilon}_{\mu}),$$

where $IC^{\mu}_{int}(\pi)$ and $IC^{\mu}_{ext}(\pi)$ are equal, if μ is a product distribution over $X \times Y$.

For a function f and a distribution μ over $X \times Y$, we define

$$\mathrm{IC}^{\mu,\varepsilon}_{\bullet}(f) = \min_{\pi:\, \exists \mathcal{P}^{\varepsilon}_{f} \text{ which induces } \pi} \mathrm{IC}^{\mu}_{\bullet}(\pi), \tag{5.11}$$

where the superscript $\bullet$ means either *int* or *ext*. Finally by maximization over all input distributions μ's, we define

$$\mathrm{IC}^{\varepsilon}_{\bullet}(f) = \max_{\mu \text{ a distribution over } X\times Y} \mathrm{IC}^{\mu,\varepsilon}_{\bullet}(f). \tag{5.12}$$

From the definition of communication cost $CC(\mathcal{P}^{\varepsilon}_{\mu,f})$, Yao's principle 5.0.1 and Proposition 5.1.6, we obtain

$$\begin{aligned} \mathrm{IC}^{\mu,\varepsilon}_{int}(f) \le \mathrm{IC}^{\mu,\varepsilon}_{ext}(f) \le CC^{\mu}_{\varepsilon}(f), \\ \mathrm{IC}^{\varepsilon}_{int}(f) \le \mathrm{IC}^{\varepsilon}_{ext}(f) \le R_{\varepsilon}(f). \end{aligned}$$

Information complexity and partition bound are the two main lower bound methods for communication complexity, since they both subsume all norm based methods, such as the discrepancy method [JK10]. This result has been proved by using a new communication model, called the zero-communication model.

5.2 Zero-communication model

In the **zero-communication model** Alice and Bob want to compute a function f without communication, but with shared randomness and aborting allowed. They both receive respectively input x and y, then they respectively output a value a and b, or they can decide to abort by sending $\perp$. Alice and Bob succeed to compute a function $f(x,y)$, if $f(x,y)=a=b$.

For this model, we define $\mathcal{P}^{\perp}$ to be a protocol with zero-communication and shared randomness where Alice an Bob can abort. A protocol $\mathcal{P}_f^{\perp}$ outputs $f(x,y)$ for all (x,y) when Alice and Bob do not abort. The **efficiency** of a protocol $\mathcal{P}^{\perp}$, denoted $\text{eff}(\mathcal{P}^{\perp})$, is the minimum probability that this protocol does not abort over all inputs (x,y). Of course, a good protocol for this model has a large efficiency. Hence for a function f, we define $\text{eff}(f)$ to be the minimum efficiency of a zero-communication protocol with shared randomness, as

$$\text{eff}(f) = \min_{\mathcal{P}_f^{\perp}} \text{eff}(\mathcal{P}^{\perp}).$$

The advantage of the efficiency is that it can be expressed as a linear program. We define $Z^{\perp} = Z \cup \{\perp\}$ to be the extension of the output set, and $\mathcal{F}_A = \{f : A \to Z^{\perp}\}$ the set of all functions that characterizes Alice's deterministic action. ($\mathcal{F}_B$ similarly for Bob.)

Definition 5.2.1. (Efficiency)
The efficiency of a function f denoted $\mathbf{eff}(f)$ is defined by the linear program

$$\text{eff}(f) = \max_{\substack{\eta\geq 0\\ p_{f_A f_B}\geq 0}} \eta \quad \text{s.t.} \quad \forall (x,y)\in \mathbf{dom}\, f, \quad \sum_{\substack{f_A\in\mathcal{F}_A:\, f_A(x)=f(x,y)\\ f_B\in\mathcal{F}_B:\, f_B(y)=f(x,y)}} p_{f_A f_B} = \eta, \tag{5.13}$$

$$\forall (x,y)\in X\times Y, \quad \sum_{\substack{f_A\in\mathcal{F}_A:\, f_A(x)\neq\perp\\ f_B\in\mathcal{F}_B:\, f_B(y)\neq\perp}} p_{f_A f_B} = \eta, \tag{5.14}$$

$$\sum_{\substack{f_A\in\mathcal{F}_A:\\ f_B\in\mathcal{F}_B}} p_{f_A f_B} = 1. \tag{5.15}$$

In a similar way, the maximum efficiency of a zero-communication with private randomness can be defined replacing the joint distribution $p(f_A f_B)$ by a product distribution $p_A(f_A).p_B(f_B)$.

We have introduced this model because the efficiency $\text{eff}(f)$ is related to the partition bound $\mathbf{prt}^0(f)$ with error null, and is a natural lower bound for communication complexity $CC(f)$ [LLR12].

Theorem 5.2.2. *[LLR12]*
Let $f : X\times Y \to Z$ be a function. We have

$$\frac{\boldsymbol{prt}_0(f)}{|Z|} \leq \boldsymbol{eff}(f)^{-1} \leq \boldsymbol{prt}_0(f) \leq 2^{CC(f)}.$$

Proof. We prove inequalities one by one.

(A) $\frac{\mathbf{prt}_0(f)}{|Z|} \leq \text{eff}(f)^{-1}$.
From an optimal solution of $\text{eff}(f)$ we construct a feasible solution of $\text{prt}_0(f)$ with an optimal value less than $|Z|\text{eff}(f)^{-1}$.
Let η and $p_{f_A f_B}$ be an optimal solution of $\text{eff}(f)$ satisfying Conditions (5.13), (5.14) and (5.15). Functions f_A and f_B provide a unique partition $P(f_A, f_B)$ of $X\times Y$, with at most $(|Z|+1)^2$ rectangles. Among this partition Alice and Bob agree on the same value $z\in Z$ without aborting

on at most $|Z|$ rectangles. We define for all $R \in S_R$ and $z \in Z$,

$$\xi_{R,z} = \frac{1}{\eta} \sum_{\substack{f_A, f_B \colon R \in P(f_A, f_B) \\ f_A(R[1]) = z \\ f_B(R[2]) = z}} p_{f_A f_B}.$$

Where for each f_A and f_B, if $f_A(R[1])$ and $f_B(R[2])$ both output z, and $R \in P(f_A, f_B)$ then we add the weight $p_{f_A f_B}$ to $\xi_{R,z}$. The solution $\xi_{R,z}$ satisfies both conditions of $\mathbf{prt}^0(f)$ in Definition (5.1.3),

- $\forall (x, y) \in \mathbf{dom}\ f$,

$$\begin{aligned}
\sum_{R:(x,y)\in R} \xi_{R,f(x,y)} &= \frac{1}{\eta} \sum_{R:(x,y)\in R} \sum_{\substack{f_A, f_B \colon R \in P(f_A, f_B) \\ f_A(R[1]) = f(x,y) \\ f_B(R[2]) = f(x,y)}} p_{f_A f_B}, \\
&= \frac{1}{\eta} \sum_{\substack{f_A, f_B \\ f_A(R[1]) = f(x,y) \\ f_B(R[2]) = f(x,y)}} \sum_{R \in P(f_A, f_B) \colon (x,y) \in R} p_{f_A f_B}, \\
&= \frac{1}{\eta} \sum_{\substack{f_A, f_B \\ f_A(x) = f(x,y) \\ f_B(y) = f(x,y)}} p_{f_A f_B}, \\
&= 1,
\end{aligned}$$

- $\forall (x, y) \in X \times Y$,

$$\begin{aligned}
\sum_{z \in Z} \sum_{R:(x,y)\in R} \xi_{R,z} &= \frac{1}{\eta} \sum_{z \in Z} \sum_{R:(x,y)\in R} \sum_{\substack{f_A, f_B \colon R \in P(f_A, f_B) \\ f_A(R[1]) = z \\ f_B(R[2]) = z}} p_{f_A f_B}, \\
&= \frac{1}{\eta} \sum_{z \in Z} \sum_{\substack{f_A, f_B \\ f_A(x) = z \\ f_B(y) = z}} p_{f_A f_B}, \\
&= \frac{1}{\eta} \sum_{\substack{f_A, f_B \\ f_A(x) \neq \perp \\ f_B(y) \neq \perp}} p_{f_A f_B}, \\
&= 1.
\end{aligned}$$

Since $\xi_{R,z}$ is a feasible solution of $\mathbf{prt}_0(f)$, we have

$$\mathbf{prt}_0(f) \leq \sum_{z \in Z} \sum_{R \in S_R} \xi_{R,z} \leq \frac{|Z|}{\eta} \sum_{f_A, f_B} p_{f_A f_B} = \frac{|Z|}{\eta} = |Z| \mathrm{eff}(f)^{-1}.$$

Where each weight $p_{f_A f_B}$ has been added at most Z times.

(B) $\mathrm{eff}(f)^{-1} \leq \mathbf{prt}_0(f)$.
From an optimal solution of $\mathbf{prt}_0(f)$ we construct a feasible solution of $\mathrm{eff}(f)$ with an optimal value less than $\mathbf{prt}_0(f)$.
Let $\xi_{R,z}$ be an optimal solution of $\mathbf{prt}_0(f)$ satisfying both conditions in Definition (5.1.3). We

define functions $f_A^{R,z}$ and $f_B^{R,z}$ for each rectangle R and output z,

$$f_A^{R,z}(x) = \begin{cases} z & \text{if } x \in_1 R, \\ \perp & \text{otherwise,} \end{cases} \qquad \text{and} \qquad f_B^{R,z}(y) = \begin{cases} z & \text{if } y \in_2 R, \\ \perp & \text{otherwise.} \end{cases}$$

We define η and $p_{f_A f_B}$ by,

$$\frac{1}{\eta} = \sum_{z \in Z} \sum_{R \in S_R} \xi_{R,z} \qquad \text{and} \qquad p_{f_A f_B} = \begin{cases} \eta.\xi_{R,z} & \text{if } f_A = f_A^{R,z} \text{ and } f_B = f_B^{R,z}, \\ 0 & \text{otherwise.} \end{cases}$$

We show that η and $p_{f_A f_B}$ satisfy Conditions (5.13), (5.14) and (5.15),

- $\forall (x,y) \in \mathbf{dom}\ f$, $$\sum_{\substack{f_A \colon f_A(x) = f(x,y) \\ f_B \colon f_B(y) = f(x,y)}} p_{f_A f_B} = \eta \sum_{R:(x,y) \in R} \xi_{R,f(x,y)} = \eta,$$
- $\forall (x,y) \in X \times Y$, $$\sum_{\substack{f_A \colon f_A(x) \neq \perp \\ f_B \colon f_B(y) \neq \perp}} p_{f_A f_B} = \eta \sum_{z \in Z} \sum_{R:(x,y) \in R} \xi_{R,z} = \eta,$$
- $$\sum_{f_A, f_B} p_{f_A f_B} = \eta \sum_{z \in Z} \sum_{R \in R} \xi_{R,z} = 1.$$

From the definition of η we conclude that, $\frac{1}{\eta} \leq \mathbf{prt}_0(f)$.

(C) $\mathbf{prt}_0(f) \leq 2^{CC(f)}$.

The proof directly comes from Theorem 5.1.4 and inequalities (5.3).

□

5.3 More lower bound methods

Definition 5.3.1. (Relaxed partition bound with distribution)[KLL+12]
The relaxed partition bound with distribution of f with error ε and input distribution μ, denoted $\overline{\mathbf{prt}}_\varepsilon^\mu(f)$, is defined by the linear program

$$\overline{\mathbf{prt}}_\varepsilon^\mu(f) = \min_{\substack{\eta \geq 0 \\ p_{R,z} \geq 0}} \frac{1}{\eta} \qquad \text{subject to,}$$

- $$\sum_{(x,y) \in \mathbf{dom}\ f} \mu(x,y) \sum_{R:(x,y) \in R} p_{R,f(x,y)} + \sum_{(x,y) \notin \mathbf{dom}\ f} \mu(x,y) \sum_{\substack{z \in Z \\ R:(x,y) \in R}} p_{R,z)} \geq \eta(1-\varepsilon),$$
- $\forall (x,y) \in X \times Y$, $$\sum_{z \in Z} \sum_{R:(x,y) \in R} p_{R,z} \leq \eta,$$
- $$\sum_{z \in Z} \sum_{R} p_{R,z} = 1.$$

Fact 5.3.2. Let μ be an input distribution, ε be an error and f be a function. We have

$$\overline{\mathbf{prt}}_\varepsilon^\mu(f) \leq \mathbf{prt}_\varepsilon^\mu(f)$$

.

Proof. Let $\xi_{R,z}$ be a feasible solution of $\mathbf{prt}^{\mu}_{\varepsilon}(f)$. Then the choice

$$\bullet\ \frac{1}{\eta} = \sum_{R,z} \xi_{R,z}, \tag{5.16}$$

$$\bullet\ p_{R,z} = \eta\xi_{R,z}, \qquad \forall R, z, \tag{5.17}$$

is a feasible solution of $\overline{\mathbf{prt}}^{\mu}_{\varepsilon}(f)$. □

Indeed, we can derive the relaxed partition bound with distribution from the partition bound with both substitution (5.16), (5.17), and by "relaxing" the equality condition to an inequality condition. Note that $p_{R,z}$ can be interpreted as a probability distribution over $S_R \times Z$.

Definition 5.3.3. (Relaxed partition bound)[KLL+12]
The relaxed partition bound of f with error ε, denoted $\overline{\mathbf{prt}}_{\varepsilon}(f)$, is defined by the linear program

$$\begin{aligned}
\overline{\mathbf{prt}}_{\varepsilon}(f) \;=\; \min_{\substack{\eta \geq 0 \\ p_{R,z} \geq 0}} \frac{1}{\eta} \quad \text{s.t.} \qquad & \forall (x,y) \in \mathbf{dom}\ f, \quad \sum_{R:(x,y)\in R} p_{R,f(x,y)} \geq \eta(1-\varepsilon), \\
& \forall (x,y) \in X \times Y, \quad \sum_{z\in Z} \sum_{R:(x,y)\in R} p_{R,z} \leq \eta, \\
& \sum_{z\in Z} \sum_{R} p_{R,z} = 1.
\end{aligned}$$

The relaxed partition bound with error $\mathbf{prt}_{\varepsilon}(f)$ is directly related to the relaxed partition bound with distribution $\overline{\mathbf{prt}}^{\mu}_{\varepsilon}(f)$.

Fact 5.3.4. [KLL+12]

$$\overline{\mathbf{prt}}_{\varepsilon}(f) = \max_{\mu \text{ a distribution over } X \times Y} \overline{\mathbf{prt}}^{\mu}_{\varepsilon}(f).$$

The relaxed partition bound with error, as indicated by its name, is weaker than the partition bound with error.

Fact 5.3.5. [KLL+12] Let f be a function and ε an error. We have

$$\overline{\mathbf{prt}}_{\varepsilon}(f) \leq \mathbf{prt}_{\varepsilon}(f),$$

with equality when ε is null.

Finally, we show an important theorem that links the relaxed partition bound with information complexity.

Theorem 5.3.6. *[KLL+12]*
Let $f : X \times Y \to Z$ be a function, ε and δ be two errors, and μ an input distribution. Then there exists a positive constant C such that

$$IC^{\mu,\varepsilon}_{int}(f) \geq \frac{\delta^2}{C} \cdot \Big(\log \overline{\boldsymbol{prt}}^{\mu}_{\varepsilon+3\delta}(f) - \log |Z| \Big) - \delta. \tag{5.18}$$

Hence, the above Theorem implies the relaxed partition bound $\overline{\mathbf{prt}}(f)$ is subsumed by the information complexity $\mathrm{IC}_{int}(f)$. Moreover, from this Theorem the relaxed partition can be used to lower bound the information complexity. However, in Inequality (5.18) if negative terms $\log |Z|$ or δ are too large and $\overline{\mathbf{prt}}(f)$ too small, then the lower bound from Theorem 5.3.6 will not be relevant.

5.4 Simulation model

A way to generalize previous models introduced in this chapter, is to replace the function $f(x, y)$ by a conditional distribution $p(a, b|x, y)$, such that for an external observer, Alice and Bob receive separately x and y, then output respectively a and b according to $p(a, b|x, y)$. In this model Alice and Bob share randomness and communication is allowed.

We define $\mathbb{P}$ to be the set of all conditional distributions $\big(p(a, b|x, y)\big)_{x,y,a,b}$ with $(x, y, a, b) \in X \times Y \times A \times B$, where $p(\,\cdot\,, \cdot\,|x, y)$ is a probability distribution over $A \times B$ for each input (x, y). We simply denoted by $\mathbf{p}$ an element of $\mathbb{P}$.

The special case of computing a function f is represented by,

$$\forall x, y, a, b, \qquad p_f(a, b|x, y) = \begin{cases} 1 & \text{if } f(x, y) = a = b, \\ 0 & \textit{otherwise.} \end{cases}$$

A **deterministic distribution p** in $\mathbb{P}$ is determined by two functions $f_A : X \to A$ and $f_B : Y \to B$ such that

$$\forall x, y, a, b, \qquad p(a, b|x, y) = \delta[a = f_A(x)].\delta[b = f_B(y)].$$

We denote $\mathcal{L}_{det}$ the set of all deterministic distributions. A **private randomness distribution p** in $\mathbb{P}$ is described by random variables R_A and R_B, and sets of functions $\{f_A : X \times R_A \to A\}$, $\{f_B : Y \times R_B \to B\}$ such that for all x, y, a, b,

$$p(a, b|x, y) = \sum_{r_A \in R_A} p_{R_A}(r_A)\delta[a = f_A(x, r_A)] \quad \cdot \quad \sum_{r_B \in R_B} p_{R_B}(r_B)\delta[b = f_B(x, r_B)]. \tag{5.19}$$

We define $\mathcal{L}_{priv}$ to be the set of all private randomness distributions. A **local distribution p** in $\mathbb{P}$ is determined by random variable Ω, and two functions $f_A : X \times \Omega \to A$, $f_B : Y \times \Omega \to B$ such that

$$\text{for all } x, y, a, b, \qquad p(a, b|x, y) = \sum_{\omega \in \Omega} p_\Omega(\omega)\delta[a = f_A(x, \omega)].\delta[b = f_B(y, \omega)].$$

We denote $\mathcal{L}$ the set of all local distributions. Note that $\mathbb{P}$ is the convex hull of all deterministic distributions. So we have $\mathcal{L}_{det} \subset \mathcal{L}_{priv} \subset \mathcal{L} \subset \mathbb{P}$. We define $\|\cdot\|_{TV}$ to be a distance between two conditional distributions as,

$$\big\|\mathbf{p} - \mathbf{q}\big\|_{TV} = \max_{x,y} \big|p(a, b|x, y) - q(a, b|x, y)\big|_{TV},$$

where $|\cdot|_{TV}$ is the total variance.

A protocol $\mathcal{P}_{\mathbf{p}}$ simulates $\mathbf{p}$, if for each input (x, y) Alice and Bob output (a, b) with probability $p(a, b|x, y)$. So the communication cost for $\mathbf{p}$ is defined by

$$R_0(\mathbf{p}) = \min_{\mathcal{P}_{\mathbf{p}}} \quad CC(\mathcal{P}_{\mathbf{p}}),$$

$$R_\varepsilon(\mathbf{p}) = \min_{q \in \mathbb{P} : \|\mathbf{p}, \mathbf{q}\|_{TV} \leq \varepsilon} R_0(\mathbf{q}).$$

For a input distribution μ, the internal/external information cost for $\mathbf{p}$ is defined by

$$\mathrm{IC}^{\mu}_{\bullet}(\mathbf{p}) = \min_{\pi :\, \exists \mathcal{P}_{\mathbf{p}} \text{ which induces } \pi} \mathrm{IC}^{\mu}_{\bullet}(\pi),$$

$$\mathrm{IC}_{\bullet}(\mathbf{p}) = \max_{\mu \text{ a distribution over } X \times Y} \mathrm{IC}^{\mu}_{\bullet}(\mathbf{p}),$$

Zero communication case

The zero communication model can be also extended to the simulation model. In this model Alice and Bob cannot communicate but they can abort ($\perp$). As sets A and B have been extended to $A \cup \{\perp\}$ and $B \cup \{\perp\}$, respectively, we define equivalently sets $\mathbb{P}^{\perp}$, $\mathcal{L}^{\perp}_{det}$ and $\mathcal{L}^{\perp}$. Obviously their relations are preserved,

$$\mathcal{L}^{\perp}_{det} \subset \mathcal{L}^{\perp}_{priv} \subset \mathcal{L}^{\perp} \subset \mathbb{P}^{\perp}.$$

The efficiency defined in Definition 5.2.1 can be adapted to the simulation model.

Definition 5.4.1. [LLR12]

$$\text{eff}(\mathbf{p}) = \max_{\substack{\eta \geq 0 \\ q_l \geq 0}} \eta \qquad \text{subject to,}$$

- $\displaystyle\sum_{l \in \mathcal{L}_{det}} q_l \cdot l(a,b|x,y) = \eta \,.\, p(a,b|x,y), \qquad \forall x,y,a,b \in X \times Y \times A \times B,$
- $\displaystyle\sum_{l \in \mathcal{L}_{det}} q_l = 1.$

5.5 One-way model

Finally, we end this chapter by introducing the one-way model. This model is a special case where only one player communicates (We choose Alice by convention.) More precisely, a **one-way protocol** $\mathcal{P}^{\rightarrow}$ is a binary decision tree where each decision at a vertex v is made by a function a_v. As the definition of communication cost CC can be applied to one-way protocols without need to generalize, we automatically obtain definition for $CC^{\rightarrow}(f)$, $CC^{\mu,\rightarrow}_{\varepsilon}(f)$, $R^{priv,\rightarrow}_{\varepsilon}(f)$, $R^{\rightarrow}_{\varepsilon}(f)$, $R^{\rightarrow}_{\varepsilon}(\mathbf{p})$, $\text{eff}^{\rightarrow}(f)$ and $\text{eff}^{\rightarrow}(\mathbf{p})$. Except for information cost $\text{IC}^{\mu}_{\bullet}(\pi)$, therefore we similarly define

$$\text{IC}^{\mu,\rightarrow}(\pi) = I(X : M, \Omega),$$

where μ is an input distribution on $X \times Y$, and π the distribution over the transcript set induced by the marginal distribution μ_X.

Chapter 6

Convex optimization

In this chapter, we do a brief introduction to the theory of convex optimization. We define optimization problems and give some important tools of this field like: the dualization method, Lagrangian, and Slater's condition. These tools will be used in Chapter 9. We also show important results as the *Karush-Kuhn-Tucker* conditions and the Envelope theorem. These results will mainly be used in Chapter 8.

Although orignal, this chapter has been written with the help of the book [BV10] of S. Boyd and L. Vandenberghe. If you wish to learn about optimization theory, I strongly recommend this book.

6.1 Optimization problems

Intuitively, an *optimization problem* is a problem where we are not only looking for a solution, but for the best solution. A point x of $\mathbb{R}^n$ is a solution or a feasible point, if it satisfies some well-defined conditions. The set of all solutions $\mathcal{C}$, called *feasible set*, is a subset of $\mathbb{R}^n$. To compare each solution we use a function mapping $\mathcal{C}$ on a total ordered set, mostly the real line $\mathbb{R}$, this function is called the *objective function* f.
An optimization problem can be represented under its minimization or maximization form:

$$\inf_{x\in\mathcal{C}} f(x) \qquad \text{or} \qquad \sup_{x\in\mathcal{C}} -f(x).$$

In this thesis, we choose to represent optimization problems under its minimization form.

For the moment the set of feasible points $\mathcal{C}$ is arbitrary. A set $\mathcal{C}$ too difficult to identify could make the problem harder, for example the set $\mathbb{R}\setminus\mathbb{Q}$. Therefore, we restrict to a set $\mathcal{C}$ described by a finite number of inequality and equality constraints.
Inequality constraints are represented by inequality functions $g_i : \mathbb{R}^n \to \mathbb{R}$ for $i \in \{1,\dots,p\}$. Equality constraints are represented by equality functions $h_j : \mathbb{R}^n \to \mathbb{R}$ for $j \in \{1,\dots,q\}$. We summarize respectively these constraints by the inequality vector $\mathbf{g}(x)$ and the equality vector $\mathbf{h}(x)$.
The domain $\mathcal{D}$ of an optimization problem is defined by domains of the objective function and

all constraints functions,

$$\mathcal{D} = \operatorname{dom} f \bigcap_{i=1}^{p} \operatorname{dom} g_i \bigcap_{j=1}^{q} \operatorname{dom} h_j \,.$$

As equality and inequality functions have a definition, we define the feasible set to be,

$$\mathcal{C} = \left\{ x \in \mathcal{D} \;\middle|\; \mathbf{g}(x) \leq 0, \text{ and } \mathbf{h}(x) = 0 \right\}.$$

Finally, we rewrite an optimization problem under its most known form

$$\begin{aligned} \inf_{x \in \mathcal{D}} f(x) \qquad \text{subject to} \qquad & \forall i \in \{1, \ldots, p\}, \quad g_i(x) \leq 0, \\ & \forall j \in \{1, \ldots, q\}, \quad h_j(x) = 0. \end{aligned} \tag{6.1}$$

If there is no solution that satisfied all constraints ($\mathcal{C}$ is empty), then the optimization problem is called unfeasible. If there exists a sequence $x_n \in \mathcal{C}$ such that $f(x_n) \to -\infty$, then the optimization problem is called unbounded. Otherwise the optimization problem has a well-defined value p^*, called *optimal value*, and defined as,

$$p^\star = \inf \left\{ f(x) : x \in \mathcal{C} \right\}.$$

By convention, $p^\star$ takes the value $+\infty$ if the optimization problem is unfeasible, and the value $-\infty$ if it is unbounded.
An *optimal point* $x^\star$ is a feasible point with $f(x^\star) = p^\star$. The set of all optimal points is defined as,

$$X^\star = \left\{ x \in \mathcal{C} : p^\star = f(x) \right\}.$$

Remark. If $\mathcal{C}$ is bounded, closed and non-empty then $X^\star$ is non-empty.

6.2 Category of optimization problems

Optimization problems can be sorted in several categories. Some category have useful properties, to solve them. Here we introduce several categories of optimization problems from the general form (6.1) described at the beginning of this chapter,

$$\begin{aligned} \inf_{x \in \mathcal{D}} f(x) \qquad \text{subject to} \qquad & \forall i \in \{1, \ldots, p\}, \quad g_i(x) \leq 0, \\ & \forall j \in \{1, \ldots, q\}, \quad h_j(x) = 0. \end{aligned} \tag{6.2}$$

6.2.0.1 Linear program

A *linear program* is an optimization problem where all functions f, g_i and h_j are affine.
Linear programs are well-known problems. There exists several algorithms to solve linear programs. The most famous is the *Simplex algorithm* [GA11] by G. Dantzig that is notably efficient in practice.

6.2.0.2 Quadratic program

A *quadratic program* is an optimization problem where the objective function f and all constraint functions g_i, h_j are quadratic forms.
For a positive quadratic form the *ellipsoid method* solves the problem efficiently [Kha80], otherwise it can be NP-hard.

6.2.0.3 Convex optimization problem

A *convex optimization problem* is an optimization problem where functions f and g_i are convex, and functions h_j are affine.
Further in this chapter, we will show that these problems have interesting properties. For example, every locally optimal point is a global optimal point and the set $X^\star$ is convex. Moreover, the strong duality is easy to prove with the Slater's conditions.

6.2.0.4 Generalized optimization problem

The objective function f and equality functions h_j are linear forms like for a linear program, except that inequality constraints are generalized. Let K be a proper cone inside $\mathbb{R}^n$. We can replace the order $\leq$ on $\mathbb{R}$ by a generalized order $\leq_K$, and inequality functions are now defined as $g_i : \mathbb{R}^n \to V$.
Note that all proofs in this chapter stay valid for these generalized optimization problems.

6.2.0.5 Semi-definite program

A *semi-definite program* is an example of generalized optimization problem where the usual order on $\mathbb{R}$ is replaced by the Loewner order $\leq_{\mathcal{S}_+^n}$.

$$\inf_{x\in\mathcal{M}_n(\mathbb{K})} f(x) \qquad \text{subject to} \qquad \begin{aligned} &\forall i, \quad g_i(x) \leq_{\mathcal{S}_+} 0,\\ &\forall j, \quad h_j(x) = 0.\end{aligned}$$

In other words, a semi-definite program is a linear program with the Loewner generalized order. In quantum computation complexity, we often study problems under this form. Especially in this thesis where lower bound methods, as the Adversary method, are semi-definite programs.

6.3 Lagrangian and duality

From an optimization problem written under Form (6.1) we can construct its *Lagrangian*

$$\begin{aligned} L(x,\boldsymbol{\lambda},\boldsymbol{\mu}) &= f(x) + \big\langle \boldsymbol{\lambda}, \mathbf{g}(x)\big\rangle + \big\langle \boldsymbol{\mu}, \mathbf{h}(x)\big\rangle,\\ &= f(x) + \sum_{i=1}^{p} \lambda_i g_i(x) + \sum_{j=1}^{p} \mu_j h_j(x), \end{aligned} \tag{6.3}$$

where vectors $\boldsymbol{\lambda} \in \mathbb{R}^p$ and $\boldsymbol{\mu} \in \mathbb{R}^q$ are named *dual vectors* or *dual variables*, and a couple $(\boldsymbol{\lambda},\boldsymbol{\mu})$ is called a *dual point*. Real numbers λ_i and μ_j are called *Lagrange multipliers*. Note that $L(x,\boldsymbol{\lambda},\boldsymbol{\mu})$ is an affine function in $(\boldsymbol{\lambda},\boldsymbol{\mu})$ for $x \in \mathcal{D}$.

The following property helps to understand the utility of the Lagrangian.

Property 6.3.1.

$$x \in \mathcal{C} \quad \Leftrightarrow \quad \forall\ \boldsymbol{\lambda} \in \mathbb{R}_+^p,\ \mu \in \mathbb{R}^q, \quad \sum_i \lambda_i g_i(x) + \sum_j \mu_j h_j(x) \leq 0. \tag{6.4}$$

Proof. If $x \in \mathcal{C}$ then $\mathbf{g}(x) \leq 0$ and $\mathbf{h}(x) = 0$. Since $\boldsymbol{\lambda} \geq 0$ and $\boldsymbol{\mu}$ then $\big\langle \boldsymbol{\lambda}, \mathbf{g}(x)\big\rangle \leq 0$ and $\big\langle \boldsymbol{\mu}, \mathbf{h}(x)\big\rangle = 0$. In the opposite direction, if $x \notin \mathcal{C}$ then at least one of (in)equalities constraints are violated, such that there exists i_0 or j_0 where $g_{i_0}(x) > 0$ or $h_{j_0}(x) \neq 0$. Then it suffices to take $\lambda_i = \delta_{i,i_0}$ or $\mu_j = h_j(x)\delta_{j,j_0}$. □

A direct result of the above Property is a lower bound for the objective function.

$$\forall x \in \mathcal{C}, \boldsymbol{\lambda} \in \mathbb{R}_+^p, \boldsymbol{\mu} \in \mathbb{R}^q, \qquad L(x, \boldsymbol{\lambda}, \boldsymbol{\mu}) \leq f(x), \tag{6.5}$$

This inequality is tight by choosing the dual point $(0, 0)$.

Property 6.3.1 shows how Lagrange multipliers λ_i and μ_j can be seen as "penalties", when x diverges from $\mathcal{C}$. Indeed for $x \notin \mathcal{C}$, we can choose $(\boldsymbol{\lambda}, \boldsymbol{\mu})$ such that $L(x, \boldsymbol{\lambda}, \boldsymbol{\mu})$ becomes as large as we wish. Therefore, by maximizing over all "penalties"$(\boldsymbol{\lambda}, \boldsymbol{\mu})$, any point x outside $\mathcal{C}$ become unfeasible:

$$\sup_{\substack{\boldsymbol{\lambda} \geq 0 \\ \boldsymbol{\mu}}} L(x, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \begin{cases} f(x) & \text{if} \quad x \in \mathcal{C}, \\ +\infty & \text{otherwise.} \end{cases}$$

From this simple observation, the form (6.1) can be written under two news forms. A min-max form,

$$\inf_{x \in \mathcal{D}} \sup_{\substack{\boldsymbol{\lambda} \geq 0 \\ \boldsymbol{\mu}}} L(x, \boldsymbol{\lambda}, \boldsymbol{\mu}) \quad = \quad \inf_{x \in \mathcal{D}} f(x) \quad \text{s.t.} \quad x \in \mathcal{C}. \tag{6.6}$$

And a minimization form where $x^\star$ is an optimal point,

$$\sup_{\substack{\boldsymbol{\lambda} \geq 0 \\ \boldsymbol{\mu}}} L(x^\star, \boldsymbol{\lambda}, \boldsymbol{\mu}) \;=\; f(x^\star) \;=\; p^\star. \tag{6.7}$$

6.4 Lagrange dual function

A Lagrangian is strongly related to its optimization problem, by definition. From the Lagrangian $L(x, \boldsymbol{\lambda}, \boldsymbol{\mu})$ we introduce the *Lagrange dual function*, a function related to an optimization problem, but independent of x.

Definition 6.4.1 (Lagrange dual function)**.**

$$d(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \inf_{x \in \mathcal{D}} L(x, \boldsymbol{\lambda}, \boldsymbol{\mu}). \tag{6.8}$$

The Lagrange dual function lower bounds the optimal value $p^\star$, when $\boldsymbol{\lambda} \geq 0$.

Property 6.4.2. *For all* $\boldsymbol{\lambda} \in \mathbb{R}_+^p$ *and* $\boldsymbol{\mu} \in \mathbb{R}^q$, $\qquad d(\boldsymbol{\lambda}, \boldsymbol{\mu}) \leq p^\star$.

Proof. The proof is quick.

$$d(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \inf_{x \in \mathcal{D}} L(x, \boldsymbol{\lambda}, \boldsymbol{\mu}) \leq \inf_{x \in \mathcal{C}} L(x, \boldsymbol{\lambda}, \boldsymbol{\mu}) \leq \inf_{x \in \mathcal{C}} f(x) = p^\star.$$

The first inequality holds because $\mathcal{C} \subseteq \mathcal{D}$. The second inequality is implied by the result (6.5). □

The above Property 6.4.2 highlights that $d(\boldsymbol{\lambda}, \boldsymbol{\mu})$ defines a lower bound for $p^\star$ when $\boldsymbol{\lambda} \geq 0$. Finally, by maximizing over Lagrange multipliers we construct another optimization problem, called the *Lagrange dual problem.*

Definition 6.4.3 (Lagrange dual problem)**.**

$$\sup_{\boldsymbol{\lambda}, \boldsymbol{\mu}} \quad d(\boldsymbol{\lambda}, \boldsymbol{\mu}) \qquad \text{subject to} \qquad \boldsymbol{\lambda} \geq 0. \tag{6.9}$$

Remark. In this thesis, this maximization form is referred as the *dual form* or the *dual problem*. And the initial optimization problem is now called, in opposition, the *primal form* or the *primal problem*. In other literature, we can find the opposite convention: the primal form refers to the maximization form and vice-versa. The choice is not important but we have to make one.

Similarly to the primal form, we define the *dual optimal value* $d^\star$ as,

$$d^\star = \sup \left\{ d(\boldsymbol{\lambda}, \boldsymbol{\mu}) : \boldsymbol{\lambda} \in \mathbb{R}^p_+,\ \boldsymbol{\mu} \in \mathbb{R}^q \right\}.$$

We define $(\boldsymbol{\lambda}^\star, \boldsymbol{\mu}^\star)$ to be a dual point that maximizes the dual form, such that $d(\boldsymbol{\lambda}^\star, \boldsymbol{\mu}^\star) = d^\star$, also called *optimum Lagrange multipliers* or *dual optimal point* $(\boldsymbol{\lambda}^\star, \boldsymbol{\mu}^\star)$.

An important consequence of Property 6.4.2 is, $d^\star \leq p^\star$. This property is called *weak duality*.

Definition 6.4.4 (Weak duality)**.**

$$d^\star \leq p^\star. \tag{6.10}$$

The weak duality implies a direct order relation between the primal and dual problem:

(a) If the primal problem is unbounded ($p^\star < -\infty$), then the dual problem is unfeasible.

(b) If the dual problem is unbounded ($d^\star > +\infty$), then the primal problem is unfeasible.

Another way to understand the weak duality is to express an optimization under its min-max form (6.6), and its Lagrange dual function under its max-min form, composed from (6.8) and (6.9).

$$\sup_{\substack{\boldsymbol{\lambda} \geq 0 \\ \boldsymbol{\mu}}} \inf_{x \in \mathcal{D}} L(x, \boldsymbol{\lambda}, \boldsymbol{\mu}) \quad \leq \quad \inf_{x \in \mathcal{D}} \sup_{\substack{\boldsymbol{\lambda} \geq 0 \\ \boldsymbol{\mu}}} L(x, \boldsymbol{\lambda}, \boldsymbol{\mu}). \tag{6.11}$$

The above inequality is the trivial part of the Minimax theorem [Von28]. The equality holds when the Lagrangian L has a saddle-point.

6.5 Strong duality

From the weak duality the Lagrange dual function is a lower bound of its optimization problem. But if the weak duality is tight, these two optimization problem are equal.

Definition 6.5.1 (Strong duality)**.** An optimization problem satisfies the strong duality if

$$d^\star = p^\star. \tag{6.12}$$

If an optimization problem satisfies the strong duality then:

(a) The primal problem is unbounded, if and only if the dual problem is unfeasible.

(b) The dual problem is unbounded, if and only if the primal problem is unfeasible.

Assume there exists a primal optimal point $x^\star$ and a dual optimal point $(\boldsymbol{\lambda}^\star, \boldsymbol{\mu}^\star)$. From Definition 6.4.3 and Equation (6.5), we have

$$d(\boldsymbol{\lambda}^\star, \boldsymbol{\mu}^\star) = \inf_{x \in \mathcal{D}} L(x, \boldsymbol{\lambda}^\star, \boldsymbol{\mu}^\star) \leq L(x^\star, \boldsymbol{\lambda}^\star, \boldsymbol{\mu}^\star) \leq f(x^\star), \tag{6.13}$$

where the first inequality comes from the choice $x = x^\star$. Therefore, the strong duality can also be represented with this equation,

$$f(x^\star) = d(\boldsymbol{\lambda}^\star, \boldsymbol{\mu}^\star).$$

If the strong duality holds then the inequality (6.13) becomes an equality. Moreover, if a dual optimal point $(\boldsymbol{\lambda}^\star, \boldsymbol{\mu}^\star)$ exists, then the primal problem can be expressed by a minimization of $L(\,\cdot\,, \boldsymbol{\lambda}^\star, \boldsymbol{\mu}^\star)$ over $\mathcal{D}$, instead of minimizing the objective function f over $\mathcal{C}$.

If the strong duality holds and both primal and dual optimal points exists, then $(x^\star, \boldsymbol{\lambda}^\star, \boldsymbol{\mu}^\star)$ is a saddle-point of the Lagrangian $L(x, \boldsymbol{\lambda}, \boldsymbol{\mu})$.

Property 6.5.2. *Let $x^\star$ be an optimal primal point and $(\boldsymbol{\lambda}^\star, \boldsymbol{\mu}^\star)$ be an optimal dual point. Assume that the strong duality is valid. Then for all $x \in \mathcal{D}$, $\boldsymbol{\lambda} \in \mathbb{R}^p_+$, $\boldsymbol{\mu} \in \mathbb{R}^q$,*

$$L(x^\star, \boldsymbol{\lambda}, \boldsymbol{\mu}) \leq L(x^\star, \boldsymbol{\lambda}^\star, \boldsymbol{\mu}^\star) \leq L(x, \boldsymbol{\lambda}^\star, \boldsymbol{\mu}^\star). \tag{6.14}$$

Moreover $L(x^\star, \boldsymbol{\lambda}^\star, \boldsymbol{\mu}^\star)$ is equal to the optimal value $p^\star$ and $d^\star$.

Proof. The first inequality comes from Equation (6.7). The second inequality comes from Inequality (6.13). Finally, we know that $x^\star$ minimizes $L(x, \boldsymbol{\lambda}^\star, \boldsymbol{\mu}^\star)$ over $\mathcal{D}$ when the strong duality holds. □

The above Property implies important properties on optimal primal point $x^\star$ and optimal dual point $(\boldsymbol{\lambda}^\star, \boldsymbol{\mu}^\star)$.

Corollary 6.5.3 (Complementary slackness)**.** *Let $x^\star$ be an optimal primal point and $(\boldsymbol{\lambda}^\star, \boldsymbol{\mu}^\star)$ be an optimal dual point. Assume that the strong duality is valid. Then*

$$\forall\ i \in \{1 \dots p\}, \qquad \lambda_i^\star = 0 \qquad \textit{or} \qquad g_i(x^\star) = 0.$$

Proof. From Property 6.5.2 we know,

$$p^\star - L(x^\star, \boldsymbol{\lambda}^\star, \boldsymbol{\mu}^\star) = 0, \qquad \text{such that,} \qquad \big\langle \mathbf{g}(x^\star), \boldsymbol{\lambda}^\star \big\rangle + \big\langle \mathbf{h}(x^\star), \boldsymbol{\mu}^\star \big\rangle = 0.$$

Since $x^\star$ is a feasible point, $\mathbf{h}(x^\star)$ is null and $\mathbf{g}(x^\star)$ is non positive. As $\boldsymbol{\lambda}^\star \geq 0$, each term $g_i(x^\star)\lambda_i$ is non positive. However, their sum is null then each product term in the sum $\big\langle \boldsymbol{\lambda}^\star, \mathbf{g}(x^\star) \big\rangle$ is null. □

6.6 KKT conditions

Hereafter, we assume that functions f, g_i and h_j are differentiable with open domains.

In the previous Section we have proved that the strong duality implies that a optimization problem can be written under a minimization form over $\mathcal{D}$ for objective function the Lagrangian $L(x, \boldsymbol{\lambda}^\star, \boldsymbol{\mu}^\star)$, where $(\boldsymbol{\lambda}^\star, \boldsymbol{\mu}^\star)$ are dual optimal point. Therefore, a point x is optimal only if the derivative of L at x is null.

$$Df_x(x^\star) + \big\langle \boldsymbol{\lambda}^\star, D_x g(x^\star) \big\rangle + \big\langle \boldsymbol{\mu}^\star, D_x h(x^\star) \big\rangle = 0. \tag{6.15}$$

The above Equation is a necessary condition for a point x to be optimal, but not sufficient. Adding several conditions such as Conditions in Corollary 6.5.3, we obtain the *Karush-Kuhn-Tucker* conditions.

Definition 6.6.1 (KKT conditions)**.**

$$g_i(x^\star) \leq 0 \quad \forall\, i = \{1,\dots,p\} \tag{6.16}$$

$$h_j(x^\star) = 0 \quad \forall\, j = \{1,\dots,q\} \tag{6.17}$$

$$\lambda_i^\star \geq 0 \quad \forall\, i = \{1,\dots,p\} \tag{6.18}$$

$$\lambda_i^\star g_i(x^\star) = 0 \quad \forall\, i = \{1,\dots,p\} \tag{6.19}$$

$$Df(x^\star) + \big\langle \boldsymbol{\lambda}^\star, Dg(x^\star)\big\rangle + \big\langle \boldsymbol{\mu}^\star, Dh(x^\star)\big\rangle = 0 \tag{6.20}$$

If an optimization problem satisfies the strong duality and has a saddle-point $(x^\star, \boldsymbol{\lambda}^\star, \boldsymbol{\mu}^\star)$, then this optimization problem satisfies the KKT conditions. However, the KKT conditions are not sufficient. Indeed, let $x_0 \in \mathcal{C}$ be a local minimum of the objective function f, then the triplet $(x_0, 0)$ satisfies the KKT conditions.

In the next Section we restrict to particular optimization problems where the KKT conditions are necessary and sufficient; convex optimization problems.

6.7 Convex optimization problems

A raw and exhaustive method to solve an optimization problem is to:

(1) find all local optima, solution of the equation, $Df(x) = 0$,

(2) eliminate solutions that are not local minimal or feasible,

(3) take the global minimum among all remaining solutions.

From this method, we understand that the presence of several local optima complicates greatly an optimization problem. Considering an convex optimization problem eludes this problem.

Definition 6.7.1 (Convex optimization problem)**.** A *convex optimization problem* is a particular optimization problem, where:

- the objective function f is convex,
- for all $i = \{1,\dots,p\}$, inequality functions g_i are convex,
- for all $j = \{1,\dots,q\}$, equality functions g_i are affine.

From this definition several properties arise.

Property 6.7.2. *The domain $\mathcal{D}$ and the feasible set $\mathcal{C}$ are convex.*

Proof. Since all functions f, g and h are convex, their domain too, then $\mathcal{D}$.
For $\mathcal{C}$, let $x_0, x_1 \in \mathcal{C}$, $\lambda \in [0,1]$, and $x_\lambda = (1-\lambda)x_0 + \lambda x_1$. Using convexity of $\mathbf{g}(x)$, $\mathbf{g}(x_\lambda) \leq (1-\lambda)\mathbf{g}(x_0) + \lambda\mathbf{g}(x_1) \leq 0$. And $\mathbf{h}(x_\lambda) = (1-\lambda)\mathbf{h}(x_0) + \lambda\mathbf{h}(x_1) = 0$. □

Property 6.7.3. *For all $\boldsymbol{\lambda} \geq 0$ and $\boldsymbol{\mu}$, the Lagrangian L is convex.*

Proof. As λ_i is positive, then $\lambda_i g_i$ stays a convex function. For all $\mu_j \in \mathbb{R}$, the function $\mu_j h_j$ stays affine f. We conclude knowing that a sum of convex functions is a convex function. □

We define precisely a locally optimal point of an optimization problem.

Definition 6.7.4. Let x_0 be a point in $\mathcal{C}$ and R a positive real. A point x_0 is a *locally optimal point* of an optimization problem (6.2), if there exists R where x_0 is an optimal point of the optimization problem,

$$\inf_{x\in\mathcal{D}} f(x) \qquad \text{subject to} \qquad \begin{aligned} &\forall i \in \{1,\dots,p\}, \quad g_i(x) \le 0, \\ &\forall j \in \{1,\dots,q\}, \quad h_j(x) = 0, \\ &\|x - x_0\| \le R. \end{aligned} \tag{6.21}$$

Property 6.7.5. *Any locally optimal point of a convex optimal problem is a globally optimal point.*

Proof. We prove by contradiction.
Let x_0 and x_1 be two locally optimal points with $f(x_0) > f(x_1)$. Since x_0 is a locally optimal point, then there exists a neighborhood $R > 0$ where for all $y \in \mathcal{B}(x_0, R)$, $f(y) \ge f(x_0)$. From Property 6.7.2, the convex combination $x_\lambda = (1-\lambda)x_0 + \lambda x_1$ is also in $\mathcal{C}$, for all $\lambda \in [0,1]$. We choose $\bar{\lambda}$ such that $x_{\bar{\lambda}} \in \mathcal{B}(x_0, R)$, then

$$f(x_{\bar{\lambda}}) \le (1-\bar{\lambda})f(x_0) + \bar{\lambda}f(x_1) < f(x_0).$$

Since $x_{\bar{\lambda}}$ is in neighborhood of x_0, contradiction. □

This is an important property of convex optimization problems. Therefore, if the objective functions f is differentiable with an open domain, every solution of $Df(x) = 0$ is an optimal point.

Another important property of convex optimization problems, is that the KKT conditions 6.6.1 are now sufficient when the strong duality holds.

Property 6.7.6. *Let be a convex optimization problem where the objective function and constraint functions are differentiable with open domains, and the strong duality holds. Then every triplet $(x, \boldsymbol{\lambda}, \boldsymbol{\mu})$ that satisfies the KKT conditions is an optimal primal-dual point.*

Proof. Let $(\bar{x}, \bar{\boldsymbol{\lambda}}, \bar{\boldsymbol{\mu}})$ be a solution of the KKT conditions. From Conditions (6.16), (6.17), (6.18) in Definition 6.6.1, $\bar{x}$ and $(\bar{\boldsymbol{\lambda}}, \bar{\boldsymbol{\mu}})$ are respectively feasible primal and dual points. Since the strong duality holds, we have

$$d(\bar{\boldsymbol{\lambda}}, \bar{\boldsymbol{\mu}}) \le d^\star = p^\star \le f(\bar{x}).$$

As $\bar{\boldsymbol{\lambda}} \ge 0$, the Lagrangian $L(x, \bar{\boldsymbol{\lambda}}, \bar{\boldsymbol{\mu}})$ is convex in $x \in \mathcal{D}$, and Condition (6.20) implies than $\bar{x}$ minimizes the Lagrangian on $\mathcal{D}$. Then,

$$\begin{aligned} d(\bar{\boldsymbol{\lambda}}, \bar{\boldsymbol{\mu}}) &= \inf_{x\in\mathcal{D}} L(x, \bar{\boldsymbol{\lambda}}, \bar{\boldsymbol{\mu}}), \\ &= L(\bar{x}, \bar{\boldsymbol{\lambda}}, \bar{\boldsymbol{\mu}}), \\ &= f(\bar{x}) + \langle \bar{\boldsymbol{\lambda}}, \mathbf{g}(\bar{x}) \rangle + \langle \bar{\boldsymbol{\mu}}, \mathbf{h}(\bar{x}) \rangle, \\ &= f(\bar{x}), \end{aligned}$$

where the two last terms are null from Conditions (6.17) and (6.19). □

Another important result for convex optimization problems is the Slater's condition. This condition allows to easily check the strong duality of a convex optimization problem.

Definition 6.7.7 (Slater's condition)**.** Let $\mathcal{D}$ be the domain of an optimization problem. Slater's condition is satisfied, if there exists $x \in \textbf{relint}\,\mathcal{D}$, a *strictly feasible point*, such that $\mathbf{g}(x) < 0$ and $\mathbf{h}(x) = 0$.

Theorem 6.7.8 (Slater's theorem)**.** *[Sla14] For a convex optimization problem, if Slater's condition is satisfied then the strong duality too.*

The proof is in Appendix B.

6.8 Envelope theorem

An optimization problem gives the "best" value of an objective function evaluated over feasible point. But what happens if the objective function or the feasible set dependent on a parameter t? How will $p^\star$ evolve on s? continuously? deferentially? And if the late question is true, can we find an expression of its derivative? An example that appears on mathematical economics, is when the market fixes prices that restrained feasible actions, but prices can later change. The Envelope theorem answers these questions.

The Enveloppe theorem has been introduced by R.B. Mirrlees [Mir71], but nowadays there exists several Envelope theorems in mathematical literature. In this Section the Envelope theorem presented is a recent result of P. Milgrom and I. Segal [MS02]. A main difference from the original theorem is that there is only one condition on domain sets, relieving the first theorem. The other difference is: We don't study the variation of the minimization of an objective function, $p^\star$, but the variation of the min-max of its Lagrangian, which is equivalent if the strong duality is valid.

Let $\mathcal{A}$ and $\mathcal{B}$ be two non-empty sets. We define $F : \mathcal{A} \times \mathcal{B} \times [0,1] \to \mathbb{R}$ to be a function, such that for almost all $t \in [0,1]$, $F(a,b,t)$ has a saddle-point $(a^\star, b^\star)$ in $\mathcal{A} \times \mathcal{B}$. In other words, for almost all $t \in [0,1]$,

$$\forall\ a \in \mathcal{A}, b \in \mathcal{B}, \qquad F(a^\star, b, t) \le F(a^\star, b^\star, t) \le F(a, b^\star, t).$$

From previous Sections, $\mathcal{A}$ and $\mathcal{B}$ is interpreted, respectively, as $\mathcal{D}$ and $\mathbb{R}^p_+ \times \mathbb{R}^q$. The function F is interpreted as the Lagrangian of an optimization problem where the strong duality holds, such that

$$p^\star(t) = \inf_{a \in \mathcal{A}} \sup_{b \in \mathcal{B}} F(a,b,t) = \sup_{b \in \mathcal{B}} \inf_{a \in \mathcal{A}} F(a,b,t).$$

Note that for each $t \in [0,1]$, the set of all saddle-points is the product set $\mathcal{A}^\star(t) \times \mathcal{B}^\star(t)$ defined as

$$\mathcal{A}^\star(t) = \big\{ a \in \mathcal{A} \,\big|\, \sup_{b \in \mathcal{B}} F(a,b,t) = p^\star(t) \big\},$$
$$\mathcal{B}^\star(t) = \big\{ b \in \mathcal{B} \,\big|\, \inf_{a \in \mathcal{A}} F(a,b,t) = p^\star(t) \big\}.$$

Theorem 6.8.1 (Envelope theorem)**.** *[MS02]*
Let $\mathcal{A}$ and $\mathcal{B}$ be two non-empty sets and $F : \mathcal{A} \times \mathcal{B} \times [0,1] \to \mathbb{R}$ be a function. Assume that:

(1) *for almost all $t \in [0,1]$, $\mathcal{A}^\star(t) \times \mathcal{B}^\star(t)$ is non-empty,*

(2) *for all $a \in \mathcal{A}$, $b \in \mathcal{B}$, $F(a,b,t)$ is absolutely continuous in t,*

(3) *for all $(a,b) \in \mathcal{A} \times \mathcal{B}$, and almost all $t \in [0,1]$, there exists an integrable function $c : [0,1] \to \mathbb{R}$ that bounds $|D_t\, F(a,b,t)| \le c(t)$.*

Then $p^\star(t)$ is absolutely continuous.
In addition assume that:

(4) *$\mathcal{A}$ and $\mathcal{B}$ are topological spaces satisfying the second axiom of countability*[1]*,*

(5) *$D_t\, F(a, b, t)$ is continuous in each of $a \in \mathcal{A}$ and $b \in \mathcal{B}$,*

(6) *the family $\big(F(a, b, t)\big)_{(a,b)\in\mathcal{A}\times\mathcal{B}}$ is equi-differentiable in t.*

Then for any selection $\big(a^\star(t), b^\star(t)\big) \in \mathcal{A}^\star(t) \times \mathcal{B}^\star(t)$,

$$p^\star(t) = p^\star(0) + \int_0^t ds\, D_s F\big(a^\star(s), b^\star(s), s\big).$$

The proof is in the appendix C.

[1] A topological space satisfies the second axiom of countability if it has a countable base.

Part II

Contributions

Chapter 7

A universal adiabatic quantum query algorithm

The adversary method $\mathrm{Adv}^{\mathrm{reg}}$ defined in Chapter 4 (Definition 4.3.4) is a method for lower bounding the quantum query complexity $Q_\varepsilon^{\mathrm{dt}}$. In their article [LMR+11], T. Lee et al. have constructed a quantum query algorithm with a query cost linear in $\mathrm{Adv}^{\mathrm{reg}}$ and a bounded error. This implies that $\mathrm{Adv}^{\mathrm{reg}}$ characterizes the bounded-error quantum query complexity $Q_\varepsilon^{\mathrm{dt}}$ (Theorem 4.3.7).

Theorem. *Let ρ and σ be two unitary Gram matrices, and $\varepsilon > 0$. Then*

$$Q_\varepsilon^{dt}(\rho \to \sigma) = \Theta\left(\frac{\mathrm{Adv}_\varepsilon^{\mathrm{reg}}(\rho \to \sigma)}{\varepsilon}\right).$$

As $\mathrm{Adv}^{\mathrm{reg}}$ also lower bounds the quantum query complexity $Q_\varepsilon^{\mathrm{ct}}$ [YM11], a corollary of this Theorem is that $\mathrm{Adv}^{\mathrm{reg}}$ also characterizes the bounded-error quantum query complexity $Q_\varepsilon^{\mathrm{ct}}$ for the continuous time model. Since an algorithm in the discrete time model can easily be converted to the continuous time model by replacing each unitary operator by a Hamiltonian. But this conversion not very satisfying from the point of view of physics, where a reasonable Hamiltonian is smooth. Hence, a first motivation is to construct a continuous query algorithm more adapted to the continuous time model. Moreover, we directly prove the characterization of $Q_\varepsilon^{\mathrm{ct}}$.

The algorithm constructed in [LMR+11] is well-defined and simple, since it is constructed from the "phase-detection" procedure. However, its evolution is difficult to describe. Another motivation is with a clearer evolution to better understand the original algorithm, as the error grows.

First, we provide an original proof that $\mathrm{Adv}^{\mathrm{reg}}$ lower bounds Q_0^{ct} based on the described method as in Subsection 4.3.2; i.e. we introduce an observable M, a unit vector $\boldsymbol{v}$, and analyze the average value $\langle M \rangle_t$ over time. To do this last step, we use the well known Ehrenfest's theorem [Ehr]

$$\frac{d\langle M\rangle_t}{dt} = -i\left\langle [M, H(t)]\right\rangle_t + \left\langle \frac{\partial M}{\partial t}\right\rangle_t.$$

Secondly, we introduce a universal adiabatic quantum query algorithm, denoted by **Adia-Convert**, based on the Adiabatic theorem [BF28] of M. Born and V. Fock. The Hamiltonian of this algorithm is constructed with optimal vectors of the dual form of $\mathrm{Adv}^{\mathrm{reg}}$. We bound the adiabatic error with Lemma 3.1.3 showing that a query cost linear in $\mathrm{Adv}^{\mathrm{reg}}$ is sufficient to obtain a bounded error. We recall the definition of $\mathrm{Adv}^{\pm}$ and give its dual from.

$$\mathrm{Adv}^{\mathrm{reg}}(\rho \to \sigma) = \sup_{\substack{M \\ v:\|v\|=1}} \quad \langle M \circ vv^*, \sigma - \rho \rangle, \tag{7.1}$$

$$\text{subject to} \quad \forall\, k \in \{1 \ldots n\}, \qquad M - \mathcal{I}d \leq M \circ \Delta_k^{\mathrm{reg}} \leq M + \mathcal{I}d. \tag{7.2}$$

$$\mathrm{Adv}^{\mathrm{reg}}(\rho \to \sigma) = \inf_{\substack{m \in \mathbb{N} \\ \boldsymbol{u}_{x,k}, \boldsymbol{v}_{y,k} \in \mathbb{C}^m}} \quad \max \Big\{ \max_{x \in \mathcal{X}} \sum_k \|\boldsymbol{u}_{x,k}\|^2, \max_{y \in \mathcal{X}} \sum_k \|\boldsymbol{v}_{y,k}\|^2 \Big\}, \tag{7.3}$$

$$\text{subject to} \quad \forall x, y \in \mathcal{X}, \qquad (\sigma - \rho)[x, y] = \sum_k \langle \boldsymbol{u}_{x,k}, \boldsymbol{v}_{y,k} \rangle \,.\Delta_k^{\mathrm{reg}}[x, y]. \tag{7.4}$$

Combining these two results leads to the following theorem.

Theorem 7.0.1. *[LMR+11, YM11] Let ρ and σ be two unitary Gram matrices, and $\varepsilon > 0$. Then*

$$Q_\varepsilon^{\mathrm{ct}}(\rho \to \sigma) = \Theta\bigg(\frac{\mathrm{Adv}_\varepsilon^{\mathrm{reg}}(\rho \to \sigma)}{\varepsilon} \bigg).$$

7.1 Adversary lower bound in the continuous-time model

In this section we give a direct proof that the adversary method $\mathrm{Adv}^{\mathrm{reg}}$ is a lower bound for Q_0^{ct}, the zero-error quantum query complexity in the continuous-time model.

Theorem 7.1.1. *[YM11] Let ρ and σ be two unitary Gram matrices, and $\varepsilon \in]0, 1]$. Then,*

$$Q_0^{\mathrm{ct}}(\rho \to \sigma) \geq \frac{1}{2} \mathrm{Adv}_0^{\mathrm{reg}}(\rho \to \sigma),$$
$$Q_\varepsilon^{\mathrm{ct}}(\rho \to \sigma) \geq \frac{1}{2} \mathrm{Adv}_\varepsilon^{\mathrm{reg}}(\rho \to \sigma).$$

Proof. Let $|\rho_x(t)\rangle$ be the state of the algorithm on input x at time $t \in [0, T]$, and $\rho(t)$ be the unitary Gram matrix of those states. Let M be an N-by-N Hermitian matrix and $\boldsymbol{v}$ be a N-dimensional unit vector. We consider the following superposition of states:

$$|\hat{\rho}_t\rangle = \sum_x v_x \, |x\rangle_{\mathcal{I}} \, |\rho_x(t)\rangle_{\mathcal{A}} \qquad \text{with} \qquad \mathrm{tr}_{\mathcal{A}} \, |\hat{\rho}_t\rangle\langle\hat{\rho}_t\,| = \rho(t) \circ \boldsymbol{v}\boldsymbol{v}^*,$$

where $\mathcal{H}_{\mathcal{A}}$ is the actual register of the algorithm, while $\mathcal{H}_{\mathcal{I}}$ is a (virtual) input register that has been introduced for the sake of analysis.

Since each state $|\rho_x(t)\rangle$ evolves under the Hamiltonian $H_x(t)$ as defined by Equation (4.3), the state $|\hat{\rho}_t\rangle$ evolves under the following global Hamiltonian

$$H(t) = \sum_x |x\rangle\langle x| \otimes H_x(t). \tag{7.5}$$

Similar to Subsection 4.3.2, we consider an observable M, a unit vector $\boldsymbol{v}$ and the average value $\langle M\rangle_t$, defined as

$$\begin{aligned}\langle M\rangle_t &= \langle M, \rho(t)\circ \boldsymbol{v}\boldsymbol{v}^*\rangle\,,\\ &= \mathrm{tr}_{\mathcal{I}}\left[M(\rho(t)\circ \boldsymbol{v}\boldsymbol{v}^*)\right],\\ &= \langle \hat{\rho}_t \,|\, M\otimes \mathcal{I}d_{\mathcal{A}} \,|\hat{\rho}_t\rangle\,,\\ &\equiv \langle M\otimes \mathcal{I}d_{\mathcal{A}}\rangle_{|\hat{\rho}(t)\rangle}\,.\end{aligned}$$

From Ehrenfest's theorem [Ehr], this average value evolves as

$$\frac{d\langle M\rangle_t}{dt} = -i\,\langle [M\otimes \mathcal{I}d_{\mathcal{A}}, H(t)]\rangle_{|\hat{\rho}(t)\rangle} + \left\langle \frac{\partial M\otimes \mathcal{I}d_{\mathcal{A}}}{\partial t}\right\rangle_{|\hat{\rho}(t)\rangle}\,,$$

where the second term is zero since $M\otimes \mathcal{I}d_{\mathcal{A}}$ is time-independent. Therefore, we have

$$\begin{aligned}\frac{d\langle M\rangle_t}{dt} &= -i\,\langle [M\otimes \mathcal{I}d_{\mathcal{A}}, H(t)]\rangle_{|\hat{\rho}(t)\rangle}\,,\\ &= -i\sum_{x,y} v_x v_y^* M[y,x]\,\langle \rho_y(t)\,|\, H_x(t)-H_y(t)\,|\rho_x(t)\rangle\,,\\ &= -i\alpha(t)\sum_{x,y} v_x v_y^* M[y,x]\sum_{k:x_k\neq y_k}\langle \rho_y(t)\,|\,|k\rangle\langle k|\otimes[h(x_k)-h(y_k)]\,|\rho_x(t)\rangle\,,\\ &= -i\alpha(t)\sum_k\sum_{x,y}(1-\delta[x_k,y_k])v_x v_y^* M[y,x]\hat{\rho}_k(t)[x,y],\\ &= i\alpha(t)\sum_k\left\langle M\circ(\Delta_k^{\mathrm{ph}}-\mathbb{J}), \hat{\rho}_k(t)\circ \boldsymbol{v}\boldsymbol{v}^*\right\rangle,\end{aligned}$$

where we define for each k, the matrix

$$\hat{\rho}_k(t)[x,y] = \langle \rho_y(t)\,|\;|k\rangle\langle k|\otimes[h(x_k)-h(y_k)]|\rho_x(t)\rangle\,.$$

The matrices $\hat{\rho}_k$ are different from matrices ρ_k introduced in Subsection 4.3.2 In particular they are not necessarily positive semi-definite, and hence we cannot use the Conditions in Definition 4.3.4 to complete the proof. Instead, we use the γ_2 norm and its properties.
Knowing that $|\alpha(t)|\leq 1$, we bound the variation of the average value by

$$\begin{aligned}\left|\frac{d\langle M\rangle_t}{dt}\right| &\leq \left|\sum_k\left\langle M\circ(\Delta_k^{\mathrm{ph}}-\mathbb{J}), \hat{\rho}_k(t)\circ \boldsymbol{v}\boldsymbol{v}^*\right\rangle\right|\\ &\leq \sum_k \|M\circ(\Delta_k^{\mathrm{ph}}-\mathbb{J})\|.\|\hat{\rho}_k(t)\circ \boldsymbol{v}\boldsymbol{v}^*\|_{\mathrm{tr}},\\ &\leq \sum_k \|M\circ(\Delta_k^{\mathrm{ph}}-\mathbb{J})\|.\gamma_2\big(\hat{\rho}_k(t)\big),\\ &\leq \max_k \|M\circ(\Delta_k^{\mathrm{ph}}-\mathbb{J})\|\cdot\Big[\sum_k\gamma_2\big(\hat{\rho}_k(t)\big)\Big],\end{aligned}$$

where we use Lemma 2.1.3 to deduce the second equality, and Fact 2.1.9 for the third equality.

Now, we show that $\sum_k \gamma_2\big(\hat{\rho}_k(t)\big)\leq 2$. First, as $\big(|k\rangle\langle k|\big)_k$ is a set of orthogonal projectors defined from the orthogonal basis $(|k\rangle)_k$, we have $\sum_k\gamma_2\big(\hat{\rho}_k(t)\big) = \gamma_2\big(\sum_k\hat{\rho}_k(t)\big)$.

Using the minimization form in Definition 2.1.8, we show that there exists $\{\boldsymbol{u}_x, \boldsymbol{w}_x\}_{x\in\mathcal{X}}$ such that $\sum_k \hat{\rho}_k(t)[x,y] = \langle u_x, w_y\rangle$ and $\max_x \big\{ \max\{\|\boldsymbol{w}_x\|^2, \|\boldsymbol{u}_x\|^2\}\big\} \leq 2$.

$$\begin{aligned} \boldsymbol{u}_x &= -H_{\mathcal{Q}}(x)\left|\rho_x(t)\right\rangle\left|0\right\rangle + \left|\rho_x(t)\right\rangle\left|1\right\rangle, \\ \boldsymbol{w}_x &= \left|\rho_x(t)\right\rangle\left|0\right\rangle + H_{\mathcal{Q}}(x)\left|\rho_x(t)\right\rangle\left|1\right\rangle. \end{aligned}$$

Then, we have $\langle \boldsymbol{u}_x, \boldsymbol{w}_y\rangle = \sum_k \rho_k(t)[x,y]$, and the upper bound on the norms of these vectors follows from conditions $\|h(l)\| \leq 1$ for all $l \in \Sigma$, which imply $\|H_{\mathcal{Q}}(x)\| \leq 1$ for all x. Since $\sum_k \gamma_2\big(\rho_k(t)\big) \leq 2$, the last bound is reduced to

$$\left|\frac{d\left\langle M\right\rangle_t}{dt}\right| \leq 2\max_k \big\|M \circ (\Delta_k^{\mathrm{ph}} - \mathbb{J})\big\|.$$

Moreover, for a zero-error algorithm we also have

$$\begin{aligned} \big|\left\langle M \circ (\sigma - \rho), \boldsymbol{v}\boldsymbol{v}^*\right\rangle\big| &= \big|\left\langle M\right\rangle_T - \left\langle M\right\rangle_0\big|, \\ &= \left|\int_0^T d\left\langle M\right\rangle_t\right|, \\ &\leq T \sup_{t\in[0,T]}\left|\frac{d\left\langle M\right\rangle_t}{dt}\right|, \\ &\leq 2T\max_k \big\|M \circ (\Delta_k^{\mathrm{ph}} - \mathbb{J})\big\|. \end{aligned}$$

As remarked, Equivalence (4.16) implies that conditions of the maximization form of $\mathrm{Adv}^{\mathrm{reg}}$ (Definition 4.3.4) are equivalent to

$$\big\|M \circ (\Delta_k^{\mathrm{ph}} - \mathbb{J})\big\| \leq 1.$$

Finally, by maximization over observables M and unit vectors $\boldsymbol{v}$, we obtain the lower bound,

$$T \geq \frac{1}{2}\mathrm{Adv}^{\mathrm{reg}}(\rho \to \sigma).$$

□

7.2 Adiabatic quantum query algorithm

In this section, we build an adiabatic quantum query algorithm, denoted by **AdiaConvert**$(\rho, \sigma, \varepsilon)$, for solving the quantum state conversion problem $(\rho \to \sigma)$, with an error ε and a running time,

$$T = O\Big(\frac{\mathrm{Adv}^{\mathrm{reg}}(\rho \to \sigma)}{\varepsilon}\Big).$$

Together with Theorem 4.3.6, this results implies that the adversary method $\mathrm{Adv}_\varepsilon^{\mathrm{reg}}$ characterizes the quantum query complexity in the continuous-time model for a bounded error (Theorem 7.0.1).

Description of AdiaConvert The algorithm acts on a Hilbert space,

$$\mathcal{H} = \mathcal{H}_{\mathcal{O}} \oplus \mathcal{H}_{\mathcal{Q}} \otimes \mathcal{H}_{\mathcal{V}} \otimes \mathcal{H}_{\mathcal{W}},$$

where $\mathcal{H}_{\mathcal{O}}$ is the input/output register, $\mathcal{H}_{\mathcal{Q}}$ the query register, $\mathcal{H}_{\mathcal{W}}$ the workspace register and $\mathcal{H}_{\mathcal{V}}$ receives the oracle's answer. Without loss of generality, we can make the initial and target states orthogonal by adding an ancilla qubit: state $|0\rangle$ for $|\rho_x\rangle$ and $|1\rangle$ for $|\sigma_x\rangle$. We define a smooth path from $|\rho_x\rangle\,|0\rangle$ to $|\sigma_x\rangle\,|1\rangle$:

$$\left|\phi_x^+(s)\right\rangle_{\mathcal{O}} = \quad \cos\theta(s)\,|0,\rho_x\rangle_{\mathcal{O}} + \sin\theta(s)\,|1,\sigma_x\rangle_{\mathcal{O}}\,,$$

with $\theta(s) = \frac{\pi}{2}s$ and $s \in [0,1]$. Moreover we define an orthogonal vector to this path,

$$\left|\phi_x^-(s)\right\rangle_{\mathcal{O}} = -\sin\theta(s)\,|0,\rho_x\rangle_{\mathcal{O}} + \cos\theta(s)\,|1,\sigma_x\rangle_{\mathcal{O}}\,,$$

with the relation $|\phi_x^-(s)\rangle = \frac{2}{\pi}\partial_s\,|\phi_x^+(s)\rangle$.

From the dual form of $\mathrm{Adv}^{\mathrm{reg}}$ (7.3), let $\big(\,|u_{x,k}\rangle\,, |v_{x,k}\rangle\,\big)_{(x,k)}$ be an optimal solution of the dual of $\mathrm{Adv}^{\mathrm{reg}}(\rho \to \sigma)$, and $\mathrm{Adv}^{\mathrm{reg}}$ its optimal value, . We use those states to define the following non-normalized vectors:

$$\begin{aligned}
\left|\Psi_x^+(s,\varepsilon)\right\rangle &= \left|\phi_x^+(s)\right\rangle_{\mathcal{O}} + \frac{\varepsilon}{\sqrt{\mathrm{Adv}^{\mathrm{reg}}}}\sum_k |k\rangle_{\mathcal{Q}}\left|x_k^+\right\rangle_{\mathcal{V}}|u_{x,k}\rangle_{\mathcal{W}}\,,\\
\left|\Psi_x^-(s,\varepsilon)\right\rangle &= \left|\phi_x^-(s)\right\rangle_{\mathcal{O}} + \xi(s)\frac{\sqrt{\mathrm{Adv}^{\mathrm{reg}}}}{\varepsilon}\sum_k |k\rangle_{\mathcal{Q}}\left|x_k^-\right\rangle_{\mathcal{V}}|v_{x,k}\rangle_{\mathcal{W}}\,,
\end{aligned}$$

where $\left|x_k^\pm\right\rangle$ is defined as (4.4), and $\xi(s) = 2\cos\theta(s)\sin\theta(s)$. Note that we have $\left\langle x_k^- \mid y_k^+\right\rangle = \frac{1}{2}\big(1-\delta[x_k,y_k]\big)$. Also we define $|\psi_x^\pm(s,\varepsilon)\rangle$ to be the normalized version of $|\Psi_x^\pm(s)\rangle$.

The Hamiltonian of the algorithm is described by its driver Hamiltonian and oracle Hamiltonian. The driver Hamiltonian is the projection $\Lambda(s,\varepsilon)$ on the vector subspace $V(s,\varepsilon)$ defined as

$$V(s,\varepsilon) = \mathrm{span}\Big\{\,\left|\Psi_x^-(s,\varepsilon)\right\rangle\,:\,x \in \mathcal{X}\Big\}.$$

The oracle Hamiltonian is defined by

$$\Pi_x = \sum_{k\in\{1\ldots n\}} |k\rangle\langle k|_{\mathcal{Q}} \otimes |x_k^-\rangle\langle x_k^-|_{\mathcal{V}} \otimes \mathcal{I}d_{\mathcal{W}},$$

where we note that the condition $\|\Pi_x\| \le 1$ is respected.

```
AdiaConvert($\rho, \sigma, \varepsilon$)

1 Prepare the state $|0, \rho_x\rangle$.

2 If $\mathrm{Adv}^{\mathrm{reg}} < \varepsilon/2$, do nothing.

3 Otherwise apply the Hamiltonian $H_x(s,\varepsilon) = \Lambda(s,\varepsilon) - \Pi_x$,
     where $s = t/T$ and $T = 15\frac{\mathrm{Adv}^{\mathrm{reg}}}{\varepsilon}$, from $t = 0$ to $t = T$.
```

The action of the algorithm is simple. First, if $\mathrm{Adv}^{\mathrm{reg}} < \varepsilon/2$, then we claim that ρ and σ are close enough and satisfy the coherent output condition given in Definition 4.1.1.

Proposition 7.2.1. *Let ρ and σ be two unitary Gram matrices. Then,*

$$\mathcal{D}_H(\rho,\sigma) \le \mathrm{Adv}^{\mathrm{reg}}(\rho \to \sigma).$$

Proof. Since the trace distance may be rewritten as

$$\mathcal{D}(\rho', \sigma') = \max_{M : \|M\| \leq 1} \frac{1}{2} \langle M, (\rho' - \sigma') \rangle ,$$

we can reformulate the Hadamard product distance (2.2) as

$$\mathcal{D}_H(\rho, \sigma) = \max_{\substack{M : \|M\| \leq 1/2 \\ \boldsymbol{u} : \|\boldsymbol{u}\| = 1}} \langle M, (\rho - \sigma) \circ \boldsymbol{u}\boldsymbol{u}^* \rangle = \max_{M : \|M\| \leq 1/2} \|M \circ (\rho - \sigma)\|.$$

We observe that this form is similar to $\mathrm{Adv}^{\mathrm{reg}}$ in Definition 4.3.4, except for the constraints on M. We conclude the proof by showing that conditions on M are weaker for $\mathrm{Adv}^{\mathrm{reg}}$, i.e. if $\|M\| \leq 1/2$, then $\|M \circ (\mathbb{J} - \Delta_k^{\mathrm{reg}})\| \leq 1$ for all $k \in \{1 \ldots n\}$.

For each $k \in \{1 \ldots n\}$, we have

$$\|M \circ (\mathbb{J} - \Delta_k^{\mathrm{reg}})\| \leq \|M\| + \|M \circ \Delta_k^{\mathrm{reg}}\| \leq \Big(1 + \gamma_2(\Delta_k^{\mathrm{reg}})\Big) \|M\|,$$

where the inequalities follows from the triangle inequality and Fact 2.1.9, respectively. We finally upper bound $\gamma_2(\Delta_k^{\mathrm{reg}})$ using the minimization form in Definition 2.1.8 with an appropriate choice. For each k, we choose $\boldsymbol{u}_x = \boldsymbol{v}_x = e_{x_k}$, with $(e_i)_i$ a canonical basis. We thus have, $\langle \boldsymbol{u}_x, \boldsymbol{v}_y \rangle = \Delta_k^{\mathrm{reg}}[x, y] = \delta[x_k, y_k]$, which entails $\gamma_2(\Delta_k) \leq 1$. □

Using Proposition 7.2.1 and first inequality in Corollary 2.1.7, we conclude that $\mathrm{Adv}^{\mathrm{reg}} < \varepsilon/2$ implies that $\mathcal{F}_H(\rho, \sigma) > 1 - \varepsilon/2 > \sqrt{1 - \varepsilon}$.

If we reach step **3**, in order to convert the initial state $|0, \rho_x\rangle$ into a state close enough to the target state $|1, \sigma_x\rangle$, we consider the state $|\psi_x^+(s, \varepsilon)\rangle$ which is ε-distant to the state $|\phi_x^+(s)\rangle$ interpolating between the initial and target state. We use the adiabatic process $\{H_x(s, \varepsilon), P_x(s, \varepsilon), T\}$ with failure ε, where $P_x(s, \varepsilon)$ is the rank-one orthogonal projection on the state $|\psi_x^+(s, \varepsilon)\rangle$. The correctness of the adiabatic evolution is based on Lemma 3.1.4, where the solution of Equation (3.9) follows from **Item 5** in the next Proposition 7.2.2. Then the final state is 3ε-distant from the target state, since the algorithm incurs error ε at the initial state, during the adiabatic process, and at the target state. This implies that we solve the quantum state generation problem with error at most $9\varepsilon^2$, and in turn that

$$Q_{9\varepsilon^2}^{\mathrm{ct}}(\rho \to \sigma) \leq 15 \frac{\mathrm{Adv}^{\mathrm{reg}}(\rho \to \sigma)}{\varepsilon^2}.$$

The proof of Theorem 7.0.1 is the consequence of the existence of the optimal quantum query algorithm, i.e. **AdiaConvert**. As the number of queries involved is given by the time scale T, the demonstration relies on the derivation of an adiabatic bound linear in $\mathrm{Adv}^{\mathrm{reg}}$.

In order to prove Theorem 7.0.1, we first derive several useful properties of the algorithm **AdiaConvert**.

Proposition 7.2.2. *For all $s \in [0, 1]$, $\varepsilon > 0$ and for all $x \in \mathcal{X}$. We have*

1) $N_x(\varepsilon) \overset{\mathrm{def}}{=} \| \, |\Psi_x^+(s, \varepsilon)\rangle \, \| \leq 1 + \varepsilon^2/2,$

2) $|\phi_x^+(s)\rangle$ *and* $|\psi_x^+(s, \varepsilon)\rangle$ *are ε-distant,*

3) $\Lambda(s, \varepsilon) \, |\psi_x^+(s, \varepsilon)\rangle = 0,$

4) $|\psi_x^+(s,\varepsilon)\rangle$ *is an eigenvector of* $H_x(s,\varepsilon)$ *with eigenvalue* $\lambda_x(s,\varepsilon) = 0$,

5) $\langle\psi_x^+(s,\varepsilon)|\left(\partial_s|\psi_x^+(s,\varepsilon)\rangle\right) = 0$,

6) $\partial_s|\Psi_x^+(s,\varepsilon)\rangle = \frac{\pi}{2}H_x(s,\varepsilon)|\Psi_x^-(s,\varepsilon)\rangle$,

7) $\| |\Psi_x^-(s,\varepsilon)\rangle \|^2 \leq 1 + \left(\mathrm{Adv}^\star/\varepsilon\right)^2$.

Remark. **Item 5** is the key property that prevents the instantaneous state $|\psi_x^+(s,\varepsilon)\rangle$ from leaking to degenerate subspaces with the same eigenvalue.

Proof. **1)** By Definition 7.3, we have $\sum_i \| |u_{x,i}\rangle \|^2 \leq \mathrm{Adv}^{\mathrm{reg}}$, then,

$$N_x^2(\varepsilon) = \left\| |\Psi_x^+(s,\varepsilon)\rangle \right\|^2 = 1 + \frac{\varepsilon^2}{\mathrm{Adv}^{\mathrm{reg}}}\sum_i \left\| |u_{x,i}\rangle \right\|^2 \leq 1 + \varepsilon^2.$$

Then **Item 1** follows from the inequality $\sqrt{1+\delta} \leq 1 + \delta/2$, for $\delta \in [0,1]$.

2) The scalar product of these vectors gives

$$\left\langle \psi_x^+(s,\varepsilon) \,\middle|\, \phi_x^+(s) \right\rangle = \frac{1}{N_x(\varepsilon)}\left\langle \Psi_x^+(s,\varepsilon) \,\middle|\, \phi_x^+(s) \right\rangle = \frac{1}{N_x(\varepsilon)} \geq 1 - \varepsilon^2/2.$$

Since this scalar product is real, we have

$$\left\| |\phi_x^+(s)\rangle - |\psi_x^+(s,\varepsilon)\rangle \right\|^2 = 2 - 2\left\langle \psi_x^+(s,\varepsilon) \,\middle|\, \phi_x^+(s) \right\rangle \leq \varepsilon^2.$$

3) As $\Lambda(s,\varepsilon)$ is the projection on subspace $V(s,\varepsilon) = \mathrm{span}\{|\Psi_x^-(s,\varepsilon)\rangle : x \in \mathcal{X}\}$. Then, it suffices to show that for all $x,y \in \mathcal{X}$, $\left\langle \Psi_x^+(s,\varepsilon) \,\middle|\, \Psi_y^-(s,\varepsilon) \right\rangle = 0$. By definition of $|\Psi_x^+(s,\varepsilon)\rangle$ and $|\Psi_x^-(s,\varepsilon)\rangle$, we have

$$\left\langle \Psi_x^+(s,\varepsilon) \,\middle|\, \Psi_y^-(s,\varepsilon) \right\rangle = -\cos\theta(s)\sin\theta(s)\left[\rho[x,y] - \sigma[x,y] - \sum_{k:x_k\neq y_k}\langle u_{x,k} \mid v_{y,k}\rangle\right].$$

The right hand side is then zero due to properties of $\left\{ |u_{x,k}\rangle, |v_{x,k}\rangle \right\}_{(x,k)}$ in Definition 7.4.

4) From **Item 3** we already know that $\Lambda(s,\varepsilon)|\psi_x^+(s,\varepsilon)\rangle = 0$. Then,

$$\Pi_y\left|\psi_x^+(s,\varepsilon)\right\rangle \propto \sum_k \left(1 - \delta[x_k,y_k]\right)\left|k, x_k^+, u_{x,k}\right\rangle,$$

which is null for $x = y$.

5) The property follows from

$$\partial_s\left|\psi_x^+(s,\varepsilon)\right\rangle = \frac{1}{N_x(\varepsilon)}\partial_s\left|\Psi_x^+(s,\varepsilon)\right\rangle = \frac{\pi}{2N_x(\varepsilon)}\left|\phi_x^-(s)\right\rangle$$

and the fact that,

$$\left\langle \psi_x^+(s,\varepsilon) \,\middle|\, \phi_x^-(s) \right\rangle \propto \left\langle \phi_x^+(s) \,\middle|\, \phi_x^-(s) \right\rangle = 0.$$

6)

$$\begin{aligned}\partial_s \left|\Psi_x^+(s,\varepsilon)\right\rangle &= \frac{\pi}{2}\left|\phi_x^-(s)\right\rangle, \\ &= \frac{\pi}{2}\Big(\mathcal{I}d - \Pi_x\Big)\left|\Psi_x^-(s,\varepsilon)\right\rangle, \\ &= \frac{\pi}{2}\Big[\Big(\Lambda(s,\varepsilon)-\Pi_x\Big)+\Big(\mathcal{I}d-\Lambda(s,\varepsilon)\Big)\Big]\left|\Psi_x^-(s,\varepsilon)\right\rangle, \\ &= \frac{\pi}{2}H_x(s,\varepsilon)\left|\Psi_x^-(s,\varepsilon)\right\rangle.\end{aligned}$$

In the second line, Π_x acts as the identity on $\left|k, x_k^-\right\rangle$. In the third line, the second term is zero by definition of $\Lambda(s,\varepsilon)$.

7) Similarly to the proof of **Item 1** all vectors $|v_{x,k}\rangle$ have their norm bounded by $\mathrm{Adv}^{\mathrm{reg}}$, then we obtain

$$\Big\| \left|\Psi_x^-(s,\varepsilon)\right\rangle \Big\|^2 = 1 + \xi^2(s)\frac{\mathrm{Adv}^\star}{\varepsilon^2}\sum_i \Big\| |v_{x,i}\rangle \Big\|^2 \leq 1 + \left(\frac{\mathrm{Adv}^\star}{\varepsilon}\right)^2,$$

since $\xi(s) = 2\sin\big(\theta(s)\big)\cos\big(\theta(s)\big) = \sin\big(2\theta(s)\big)$.

□

Proof of Theorem 7.0.1.

We denote $\mathrm{Adv}^{\mathrm{reg}} = \mathrm{Adv}^{\mathrm{reg}}(\rho \to \sigma)$. We show that **AdiaConvert** solves the quantum state conversion in time $T = 15\frac{\mathrm{Adv}^\star}{\varepsilon^2}$ with error at most $9\varepsilon^2$. Let us first consider the case where $\mathrm{Adv}^{\mathrm{reg}} < \varepsilon/2$. Then, Proposition 7.2.1 implies $\mathcal{D}_H(\rho,\sigma) < \varepsilon/2$, and Corollary 2.1.7 concludes that $\mathcal{F}_H(\rho,\sigma) > 1-\varepsilon/2 > \sqrt{1-\varepsilon}$, so that the coherent output condition is already satisfied by the unitary Gram matrix ρ.

We now assume that $\mathrm{Adv}^\star \geq \varepsilon/2$. Before we go any further, we must justify that the triplet $\{H_x(s,\varepsilon),\ P_x(s,\varepsilon), T\}$ is an adiabatic process as defined in Definition 3.1.1. First by definition, the state $|\psi_x^\pm(s,\varepsilon)\rangle$ is smooth on s. It follows that $H_x(s,\varepsilon)$ and $P_x(s,\varepsilon)$ are also smooth on s. Moreover, by **Item 4** of Proposition 7.2.2, $|\psi_x^+(s,\varepsilon)\rangle$ is an eigenstate of $H_x(s,\varepsilon)$ with a constant eigenvalue $\lambda_x(s,\varepsilon) = 0$.

In order to bound the error of the adiabatic process ε_{AP} with Lemma 3.1.4, we define an operator $X_x(s,\varepsilon)$ to be a solution of Equation (3.9), with $X_x(s,\varepsilon)$ and $\dot{X}_x(s,\varepsilon)P_x(s,\varepsilon)$ both bounded.

$$\forall x \in \mathcal{X}, \quad X_x(s,\varepsilon) = \frac{\pi}{2N_x(\varepsilon)}\left|\Psi_x^-(s,\varepsilon)\middle\rangle\!\middle\langle\psi_x^+(s,\varepsilon)\right|.$$

Items 4 and **6** of Proposition 7.2.2 imply that,

$$[H_x(s,\varepsilon), X_x(s,\varepsilon)] = H_x(s,\varepsilon)X_x(s,\varepsilon) = \dot{P}_x(s,\varepsilon)P_x(s,\varepsilon).$$

To obtain ε_{AP} we derive a bound for $X_x(s,\varepsilon)$ and $\dot{X}_x(s,\varepsilon)P_x(s,\varepsilon)$.

- First, we have

$$\|X_x(s,\varepsilon)\|^2 = \left[\frac{\pi}{2N_x(\varepsilon)}\right]^2 \Big\| \left|\Psi_x^-(s,\varepsilon)\right\rangle \Big\|^2.$$

From **Item 7** of Proposition 7.2.2 and the fact that $\mathrm{Adv}^\star \geq \varepsilon/2$, we obtain

$$\big\| \left|\Psi_x^-(s,\varepsilon)\right\rangle \big\|^2 \leq 1 + \left(\frac{\mathrm{Adv}^\star}{\varepsilon}\right)^2 \leq 5\left(\frac{\mathrm{Adv}^\star}{\varepsilon}\right)^2,$$

knowing that $N_x(\varepsilon) \geq 1$ we obtain the bound : $\|X_x(s,\varepsilon)\| \leq \frac{\pi\sqrt{5}}{2}\frac{\mathrm{Adv}^\star}{\varepsilon}$.

- Secondly, to bound $\|\dot{X}_x(s,\varepsilon)P_x(s,\varepsilon)\|$ we derive $X_x(s,\varepsilon)$

$$\dot{X}_x(s,\varepsilon) = \frac{\pi}{2N_x(\varepsilon)}\partial_s\big(\left|\Psi_x^-(s,\varepsilon)\right\rangle\big)\left\langle\psi_x^+(s,\varepsilon)\right| + \frac{\pi^2}{4N_x(\varepsilon)}\left|\Psi_x^-(s,\varepsilon)\right\rangle\!\left\langle\phi_x^-(s)\right|.$$

After adding $P_x(s,\varepsilon)$ on the right side, the second term disappears following **Item 5** of Proposition 7.2.2, and we have

$$\begin{aligned}
\|\dot{X}_x(s,\varepsilon)P_x(s,\varepsilon)\|^2 &= \left[\frac{\pi}{2N_x(\varepsilon)}\right]^2 \Big\|\partial_s\left|\Psi_x^-(s,\varepsilon)\right\rangle\Big\|^2 \\
&\leq \left[\frac{\pi}{2}\right]^2\Big(\frac{\pi^2}{4} + \pi^2\cos^2(\pi s)\frac{\mathrm{Adv}^\star}{\varepsilon^2}\sum_k \|\left|v_{x,k}\right\rangle\|^2\Big) \\
&\leq \left[\frac{\pi}{2}\right]^2\pi^2\Big(\frac{1}{4} + \frac{\mathrm{Adv}^{\star 2}}{\varepsilon^2}\Big) \\
&\leq \left[\frac{\pi}{2}\right]^2 2\pi^2\frac{\mathrm{Adv}^{\star 2}}{\varepsilon^2}.
\end{aligned}$$

Thereby we have all the required conditions to use Lemma 3.1.4 for the adiabatic process $\{H_x(s,\varepsilon), P_x(s,\varepsilon), T\}$, which ensures that $\varepsilon_{AP} \leq \varepsilon$, if

$$T \geq \frac{15\mathrm{Adv}^\star}{\varepsilon^2} \geq \frac{1}{\varepsilon}\Big[\frac{\mathrm{Adv}^\star}{\varepsilon}\Big(\pi\sqrt{5} + \frac{\pi^2}{\sqrt{2}}\Big)\Big].$$

Let $|\tilde{\sigma}_x\rangle$ be the output state. Since the initial state $|0,\rho_x\rangle$ and the target state $|1,\sigma_x\rangle$ are ε-distant from $|\psi_x^+(0,\varepsilon)\rangle$ and $|\psi_x^+(1,\varepsilon)\rangle$ (**Item 2** of Proposition 7.2.2) and the adiabatic process introduces an additional error of $\varepsilon_{AB} \leq \varepsilon$, the output state $|\tilde{\sigma}_x\rangle$ and the target state $|1,\sigma_x\rangle$ are 3ε-distant, which implies that $\mathrm{Re}(\langle\tilde{\sigma}_x \mid 1,\sigma_x\rangle) \geq \sqrt{1-9\varepsilon^2}$. Therefore, we obtain

$$Q^{\mathrm{ct}}_{9\varepsilon^2}(\rho,\sigma) \leq 15\frac{\mathrm{Adv}^\star}{\varepsilon^2},$$

which implies the theorem by setting $\varepsilon' = 9\varepsilon^2$. □

Chapter 8

A new lower bound for Q^{ct}

Throughout this Chapter we use the notation introduced in Chapter 4, in particular the Gram matrix representation from Section 4.2. We only work in the continuous-time model with a binary alphabet[1]*, together with the phase-Hamiltonian oracle representation with $\hat{\Delta}_k^{\tau}$. We also switch frequently between notation, $\hat{k}$ and (k, τ).*

The adversary method introduced in Section 4.3.2 characterizes the bounded-error quantum query complexity Q_ε for discrete et continuous time models (Theorems 4.3.7 and 7.0.1), but it left open its characterization for the unbounded-error and the zero-error cases. The characterization of Q_0 would be useful, since Lemma 4.1.2 characterizes Q_ε for any ε in term of Q_0 (including unbounded error) using the Hadamard product distance (2.2). This characterization will be useful in domains such as cryptography where one might want to prove hardness results even for very low success probabilities.
Also, the characterization of Q_ε for the unbounded-error case could allow to demonstrate composition theorems as strong direct product theorem. For example, a function composed of k other functions with non-zero error may have its error depending on k.

A good candidate to characterize the quantum query complexity is the multiplicative adversary method (Definition 4.3.9), since this method subsumes the polynomial method (Subsection 4.3.1) and the adversary method $\mathrm{Adv}^{\mathrm{reg}}$ [MR13]. Moreover, the multiplicative adversary method inherently satisfies a strong direct product theorem [LR12].

The main idea is to define a lower bound method powerful enough to characterize Q_ε. The method is called **Adversary action**, Sadv, and it is a generalization of the adversary method. However, this method is too complicated to be used in practice. Then, we hope to simplify Sadv to $\mathrm{Madv}^{\mathrm{ct}}$ without loss of generality.

First, we introduce an adapted version of Proposition 4.2.2 to the continuous time model, that allows to check if a differentiable path $\gamma \in \Gamma_N$ is feasible or not. From this new Proposition 8.1.1, we adapt adversary methods $\mathrm{Adv}^{\mathrm{dt}}$ and $\mathrm{Madv}^{\mathrm{dt}}$ to the continuous time model by construction new adversary methods $\mathrm{Adv}^{\mathrm{ct}}$ and $\mathrm{Madv}^{\mathrm{ct}}$, introduced earlier in Definitions 8.2.1 and 8.2.4. Then we prove that these new methods lower bound the quantum query complexity Q_0^{ct}. We also prove that $\mathrm{Madv}^{\mathrm{ct}}$ subsumes $\mathrm{Adv}^{\mathrm{ct}}$.

[1] This choice is not a restriction. It only make the analysis simple.

After, we construct our new adversary method Sadv, the adversary action. This method relies on the query Lagrangian $\mathcal{L}$ based on a semi-definite program that derive from the adapted version of Proposition 4.2.2. We prove several properties of $\mathcal{L}$ such as strong duality and optimal points existence. As the adversary action is constructed as a physical action, we use the Euler-Lagrange equation to derive conditions on a locally optimal path. Finally, we show that Sadv naturally subsumes $\mathrm{Adv}^{\mathrm{ct}}$ and $\mathrm{Madv}^{\mathrm{ct}}$.

We also construct a new adversary method Sadv called the adversary action. This method relies on the query Lagrangian $\mathcal{L}$ based on Proposition 8.1.1. Thereafter, we use these properties, we establish several conditions that a locally optimal solution of Sadv must satisfy. Finally, we show that Sadv naturally subsumes $\mathrm{Adv}^{\mathrm{ct}}$ and $\mathrm{Madv}^{\mathrm{ct}}$.

8.1 Feasible differentiable paths

We recall several notations introduced in Section 4.1.2.A quantum query algorithm $\mathcal{A}$ is described by,

$$H_x(t) = H_{\mathcal{D}}(t) + \alpha(t) H_{\mathcal{Q}}(x), \qquad \text{with } \|H_{\mathcal{Q}}(x)\| \leq 1 \forall x, \\ \text{and } \alpha : [0,T] \to [0,1].$$

and a time-dependent quantum state $|\gamma_x(t)\rangle$ is defined as

$$|\gamma_x(t)\rangle = \sum_{(k,\tau)\in\hat{n}} |k\rangle_{\mathcal{Q}} \otimes |\gamma_{x;k\tau}(t)\rangle_{\mathcal{W}}.$$

In Section 4.2, we have introduced the Gram matrix representation and the set of possible differentiable paths as

$$\Gamma_{\mathrm{ct}}[\rho \to \sigma] = \bigcup_{T\in\mathbb{R}_+} \Big\{ \gamma(t) \in \mathcal{C}^1\big([0,T], G_N\big) \,:\, \gamma(0) = \rho \text{ and } \gamma(1) = \sigma \Big\}.$$

In order to simplify the notation, we discard the running time T with the substitution $\hat{H}_x(s) = T \cdot H_x(T \cdot s)$. We also define $\beta(s)$ to be equal to $\alpha(s)/\pi$ such that a query can be implemented with a unit of time. Hence the Hamiltonian of the algorithm becomes

$$\hat{H}_x(s) = T \cdot H_{\mathcal{D}}(s \cdot T) + \pi \beta(s) H_{\mathcal{Q}}(x), \qquad \text{with } \|H_{\mathcal{Q}}(x)\| \leq 1 \forall x, \\ \text{and } \beta(s) = T \cdot \alpha(s \cdot T)/\pi.$$

Hence, the query cost of an algorithm $\mathcal{A}$ defined by Equation 4.7 becomes

$$q(\mathcal{A}) = \int_0^1 ds\, \beta(s). \tag{8.1}$$

From the definition of $\beta(s)$, the running time is now defined by

$$T = \frac{1}{\pi} \sup_{s\in[0,1]} |\beta(s)|.$$

And the definition of Γ_{ct} is simplified by

$$\Gamma_{\mathrm{ct}}[\rho \to \sigma] = \Big\{ \gamma(t) \in \mathcal{C}^1\big([0,1], G_N\big) \,:\, \gamma(0) = \rho \text{ and } \gamma(1) = \sigma \Big\}.$$

Proposition 8.1.1. *Let ρ be unitary Gram matrix, δ be a Hermitian matrix and β_0 a positive real. Then, we can transform infinitesimally ρ to $\rho+\delta$ with an infinitesimal query of the Hamiltonian $\beta_0 H_Q^{\mathrm{ph}}$, if and only if there exists a set of positive semi-definite matrices $(\gamma_{\hat{k}})_{\hat{k}\in\hat{n}}$, such that*

$$\rho = \sum_{\hat{k}\in\hat{n}} \gamma_{\hat{k}} \qquad \text{and} \qquad \delta = \beta_0 \sum_{\substack{k\in\{0\cdots n\}\\ \tau\in\{+1,-1\}}} \gamma_k^\tau \circ \hat{\Delta}_k^\tau, \tag{8.2}$$

where for all $(k,\tau)\in\hat{n}$, $\hat{\Delta}_k^\tau = i\tau(y_k - x_k)$ and $\gamma_{\hat{k}} = \gamma_k^\tau$ for $\hat{k} = \{k,\tau\}$.

Proof. The path $\rho(s)$ is generated by a quantum query algorithm defined by $H_x(s)$ with the oracle Hamiltonian $H_Q^{\mathrm{ph}}(x)$ as defined in 4.5, and $s\in[0,1]$. We must prove that the derivative of $\rho(s)$ evolving under $H_x(s)$ is equal to $d\rho$.

$$\begin{aligned}
\frac{d}{ds}\rho[x,y](s) &= \frac{d}{ds}\Big(\langle \rho_x(s) \mid \rho_y(s)\rangle \Big),\\
&= i\,\langle \rho_x(s) \,|\, H_x(s) - H_y(s)\, | \rho_y(s)\rangle\,,\\
&= i\beta(s)\,\langle \rho_x(s) \,|\, H_Q^{\mathrm{ph}}(x) - H_Q^{\mathrm{ph}}(y)\, | \rho_y(s)\rangle\,,\\
&= i\beta(s) \sum_{\substack{k\in\{0\ldots n\}\\ \tau=\pm}} \tau\,\langle k,\tau,\rho_{x;k\tau}(s) \,|\, h(x_k) - h(y_k)\, | k,\tau,\rho_{y;k\tau}(s)\rangle\,,\\
&= i\beta(s) \sum_{\substack{k\in\{0\ldots n\}\\ \tau=\pm}} \tau\,\big[x_k - y_k\big]\,\langle \rho_{x;k\tau}(s) \mid \rho_{y;k\tau}(s)\rangle\,,\\
&= \beta(s) \sum_{\substack{k\in\{0\ldots n\}\\ \tau=\pm}} \hat{\Delta}_k^\tau[x,y].\rho_{k\tau}[x,y](s),
\end{aligned}$$

where $|\rho_{y;k\tau}(s)\rangle$ is defined by (4.6), and $\rho_{k\tau} = \mathrm{Gram}\big(\,|\rho_{x;k\tau}\rangle : x\in\mathcal{X}\big)$. In other words,

$$\frac{d}{ds}\rho(s) = \beta(s)\sum_{k\in\hat{n}} \rho_k(s)\circ\hat{\Delta}_k. \tag{8.3}$$

The other direction comes from Corollary 4.2.1. □

Hence, a possible differentiable path γ is feasible, if for each $s\in[0,1]$ there exists $(\gamma_{\hat{k}})$ and $\beta_0(s)$ that satisfy conditions in Proposition 8.1.1. Moreover, the query cost $q(\gamma)$ and the running time $T(\gamma)$ of this path is defined as,

$$\begin{aligned}
q(\gamma) &= \int_0^1 ds\,\beta_0(s),\\
T(\gamma) &= \sup_{s\in[0,1]} \beta_0(s).
\end{aligned}$$

Remark. Is every feasible path $\gamma\in\Gamma_{\mathrm{ct}}[\rho\to\sigma]$ is associated to a quantum query algorithm? This is an open question. However, from a feasible path we can construct a family of discrete-time quantum query algorithm $(\mathcal{A}_n)_n$ with increasing precision.
For a nonnegative integer n, we can convert a differentiable path γ to a discrete path where the interval $[0,1]$ has been splitting into n equal subintervals $([\frac{m}{n},\frac{m+1}{n}])_i$. Then for all $m\in$

$\{0\dots n{-}1\}$, we can convert $\gamma(\frac{m}{n})$ to $\gamma(\frac{m+1}{n})$, with error, by implementing the oracle Hamiltonian $\hat{H}_x(\frac{m}{n})$ with a fraction $\beta_0(\frac{m}{n})/n$. For all $x \in \mathcal{X}$ and $m \in \{0\dots n-1\}$, we obtain

$$\left|\gamma_x^+(\frac{m}{n})\right\rangle = e^{-\frac{i}{n}\beta_0(\frac{m}{n})H_Q(x)}\left|\gamma_x(\frac{m}{n})\right\rangle.$$

Then, between each oracle call we use the driver Hamiltonian to transport $\left|\gamma_x^+(\frac{m}{n})\right\rangle$ as close as possible to $\left|\gamma_x(\frac{m+1}{n})\right\rangle$, for all $x \in \mathcal{X}$. Hence, although a feasible path γ may not be converted to a continuous-time quantum query algorithm, we can construct a sequence of query algorithms $(\mathcal{A}_n)_n$ with error going to zero and query cost going to $q(\gamma)$ when n goes to infinity.

8.2 Adversary methods

In this Section, we define new Adversary methods $\mathrm{Adv}^{\mathrm{ct}}$ and $\mathrm{Madv}^{\mathrm{ct}}$, adapted version of $\mathrm{Adv}^{\mathrm{dt}}$ and $\mathrm{Madv}^{\mathrm{dt}}$ for Q^{dt}. We prove that these new methods lower bound Q^{dt} using Proposition 8.1.1. Finally, we prove that $\mathrm{Adv}^{\mathrm{dt}}$ and $\mathrm{Madv}^{\mathrm{dt}}$ respectively subsume $\mathrm{Adv}^{\mathrm{ct}}$ and $\mathrm{Madv}^{\mathrm{ct}}$.

8.2.1 Adversary method $\mathrm{Adv}^{\mathrm{ct}}$

The following definition of $\mathrm{Adv}^{\mathrm{ct}}$ is constructed with the same reasoning that $\mathrm{Adv}^{\mathrm{dt}}$, except that we use the phase-Hamiltonian with $\hat{\Delta}_k^\tau[x,y] = i\tau\pi(x_k - y_k)$.

Definition 8.2.1 (Adversary method for continuous time)**.**

$$\begin{aligned}\mathrm{Adv}_0^{\mathrm{ct}}(\rho\to\sigma) = \sup_{\substack{M\\ v:\|v\|=1}} \quad & \big\langle M\circ vv^*, \sigma-\rho\big\rangle,\\ \text{subject to} \quad & \forall\, k\in\{0\dots n\},\ \tau\in\{+1,-1\} \qquad -\mathcal{I}d \le M\circ\hat{\Delta}_k^\tau \le \mathcal{I}d.\end{aligned}$$

Note that the above definition of $\mathrm{Adv}_0^{\mathrm{ct}}$ is similar to $\mathrm{Adv}_0^{\mathrm{dt}}$, only the conditions change.

Theorem 8.2.2. *Let ρ and σ be two unitary Gram matrices. Then,*

$$\mathrm{Adv}_0^{ct}(\rho\to\sigma) \le Q_0^{ct}(\rho\to\sigma).$$

The proof is a consequence of the following Lemma.

Lemma 8.2.3. *Let M be an observable, v be a unitary vector and $\rho(t)$ be a feasible differentiable path. We define $\langle M\rangle_t = \langle M, \rho(t)\circ vv*\rangle$. If*

$$-\mathcal{I}d \le M\circ\hat{\Delta}_k^\tau \le +\mathcal{I}d, \qquad \textit{for all } k\in\{0\dots n\}, \tau\in\{+1,-1\}, \tag{8.4}$$

then,

$$\left|\frac{d\langle M\rangle_t}{dt}\right| \le 1.$$

Proof. As $\rho(t)$ is feasible, from Proposition 8.1.1 we know that for all $s\in[0,1]$ there exists $\big(\gamma_{\hat{k}}(s)\big)_{\hat{k}}$ and $\beta_0(s)$, such that

$$\rho(s) = \sum_{\hat{k}\in\hat{n}} \gamma_{\hat{k}}(s) \qquad \text{and} \qquad \frac{d\rho}{ds}(s) = \beta_0(s)\sum_{\substack{k\in\{0\cdots n\}\\ \tau\in\{+1,-1\}}} \gamma_k^\tau(s)\circ\hat{\Delta}_k^\tau. \tag{8.5}$$

Therefore for all $t \in [0, T]$,

$$\begin{aligned}
\frac{d\langle M\rangle_t}{dt} &= \frac{1}{T}\frac{d\langle M\rangle_s}{ds},\\
&= \frac{1}{T}\big\langle M, \frac{d\rho}{ds}(s) \circ vv^*\big\rangle,\\
&= \big\langle M, \beta_0(s) \sum_{\substack{k\in\{0\cdots n\}\\ \tau\in\{+1,-1\}}} \gamma_k^\tau(s) \circ \hat{\Delta}_k^\tau \circ vv^*\big\rangle,\\
&= \frac{\beta_0(s)}{T} \sum_{\substack{k\in\{0\cdots n\}\\ \tau\in\{+1,-1\}}} \big\langle M \circ \hat{\Delta}_k^\tau, \gamma_k^\tau(s) \circ vv^*\big\rangle,\\
&\le \frac{\beta_0(s)}{T} \sum_{\substack{k\in\{0\cdots n\}\\ \tau\in\{+1,-1\}}} \big\langle \mathcal{I}d, \gamma_k^\tau(s) \circ vv^*\big\rangle,\\
&\le \frac{\beta_0(s)}{T}\big\langle \mathcal{I}d, \rho(s) \circ vv^*\big\rangle,\\
&\le \frac{\beta_0(s)}{T},\\
&\le 1,
\end{aligned}$$

the first inequality uses the fact that $\gamma_k^\tau(s) \circ vv^* \ge 0$ and $M \circ \hat{\Delta}_k^\tau \le +\mathcal{I}d$. The last inequality comes from $T = \sup_s |\beta_0(s)|$.
Using the other inequality $M \circ \hat{\Delta}_k^\tau \ge -\mathcal{I}d$, we can prove similarly that

$$\frac{d\langle M\rangle_t}{dt} \ge -1.$$

□

8.2.2 Multiplicative adversary method $Madv^{\mathrm{ct}}$

The following definition of $\mathrm{Madv}^{\mathrm{ct}}$ is constructed with the same reasoning that $\mathrm{Madv}^{\mathrm{dt}}$, except that we use the phase-Hamiltonian with $\hat{\Delta}_k^\tau[x,y] = i\tau\pi(x_k - y_k)$.

Definition 8.2.4 (Multiplicative adversary method for continuous time)**.**

$$\mathrm{Madv}_0^{\mathrm{ct}}(\rho \to \sigma) = \sup_{b>0} \frac{1}{b} \sup_{\substack{M\ge 0\\ v:\|v\|=1}} \big[\ln\big\langle M \circ vv^*, \sigma\big\rangle - \ln\big\langle M \circ vv^*, \rho\big\rangle\big],$$

$$\text{subject to} \quad \forall\, k \in \{1\dots n\}, \forall \tau \in \{+,-\}, \qquad -bM \le M \circ \hat{\Delta}_k^\tau \le bM.$$

Note that the above definition of $\mathrm{Madv}_0^{\mathrm{ct}}$ is similar to $\mathrm{Madv}_0^{\mathrm{dt}}$, only the conditions change.

Theorem 8.2.5. *Let ρ and σ be two unitary Gram matrices. Then*

$$\mathrm{Madv}_0^{\mathrm{ct}}(\rho \to \sigma) \le Q_0^{\mathrm{ct}}(\rho \to \sigma).$$

The proof is a consequence of the following Lemma.

Lemma 8.2.6. *Let $M \geq 0$ be an observable, b be a strictly positive real, v be a unitary vector and $\rho(t)$ be a feasible differentiable path. We define $\langle M \rangle_t = \langle M, \rho(t) \circ vv* \rangle$. If*

$$-bM \leq M \circ \hat{\Delta}_k^\tau \leq bM, \qquad \textit{for all } k \in \{0 \dots n\}, \tau \in \{+1, -1\}, \tag{8.6}$$

then,

$$\left| \frac{d \langle M \rangle_t}{dt} \right| \leq b \langle M \rangle_t$$

Proof. As $\rho(t)$ is feasible, from Proposition 8.1.1 we know that for all $s \in [0,1]$ there exists $\left(\gamma_{\hat{k}}(s)\right)_{\hat{k}}$ and $\beta_0(s)$, such that

$$\rho(s) = \sum_{\hat{k} \in \hat{n}} \gamma_{\hat{k}}(s) \qquad \text{and} \qquad \frac{d\rho}{ds}(s) = \beta_0(s) \sum_{\substack{k \in \{0 \cdots n\} \\ \tau \in \{+1,-1\}}} \gamma_k^\tau(s) \circ \hat{\Delta}_k^\tau. \tag{8.7}$$

Therefore all $t \in [0, T]$,

$$\begin{aligned}
\frac{d \langle M \rangle_t}{dt} &= \frac{1}{T} \frac{d \langle M \rangle_s}{ds}, \\
&= \frac{1}{T} \big\langle M, \rho(s) \circ vv^* \big\rangle, \\
&= \frac{1}{T} \big\langle M, \beta_0(s) \sum_{\substack{k \in \{0 \cdots n\} \\ \tau \in \{+1,-1\}}} \gamma_k^\tau(s) \circ \hat{\Delta}_k^\tau \circ vv^* \big\rangle, \\
&= \frac{\beta_0(s)}{T} \sum_{\substack{k \in \{0 \cdots n\} \\ \tau \in \{+1,-1\}}} \big\langle M \circ \hat{\Delta}_k^\tau, \gamma_k^\tau(s) \circ vv^* \big\rangle, \\
&\leq \frac{\beta_0(s)}{T} \sum_{\substack{k \in \{0 \cdots n\} \\ \tau \in \{+1,-1\}}} \big\langle bM, \gamma_k^\tau(s) \circ vv^* \big\rangle, \\
&\leq b \frac{\beta_0(s)}{T} \big\langle M, \rho(s) \circ vv^* \big\rangle, \\
&\leq b \frac{\beta_0(s)}{T} \langle M \rangle_s, \\
&\leq b \langle M \rangle_t,
\end{aligned}$$

the first inequality uses the fact that $\gamma_k^\tau(s) \circ vv^* \geq 0$ and $M \circ \hat{\Delta}_k^\tau \leq bM$. The last inequality comes from $T = \sup_s |\beta_0(s)|$.
Using the other inequality $M \circ \hat{\Delta}_k^\tau \geq -bM$, we can prove similarly that

$$\frac{d \langle M \rangle_t}{dt} \geq -b \langle M \rangle_t.$$

□

8.2.3 Relation between adversary methods

A quantum algorithm in continuous-time model can simulate a quantum algorithm in discrete-time model, but the opposite is not straightforward. Thus, this is not surprising to have a relation between adversary methods for discrete-time and continuous-time. The following theorem gives two relations for Adv and Madv. We recall that, for all $x, y \in \mathcal{X}$

$$\Delta_k^{\mathrm{ph}}[x,y] = (-1)^{y_k - x_k}, \quad \text{and} \quad \hat{\Delta}_k^{\tau}[x,y] = i\tau(y_k - x_k).$$

Theorem 8.2.7. *Let ρ and σ be two unitary Gram matrices, and an error ε. Then*

$$\begin{aligned} \frac{1}{\pi}\mathrm{Adv}_{\varepsilon}^{\mathrm{ct}}(\rho \to \sigma) &\leq \mathrm{Adv}_{\varepsilon}^{\mathrm{ph}}(\rho \to \sigma), \\ \frac{1}{\pi}\mathrm{Madv}_{\varepsilon}^{\mathrm{ct}}(\rho \to \sigma) &\leq \mathrm{Madv}_{\varepsilon}^{\mathrm{ph}}(\rho \to \sigma). \end{aligned}$$

The demonstration of Theorem 8.2.7 is simple, we just show that the conditions of $\mathrm{Adv}^{\mathrm{ct}}$ and $\mathrm{Madv}^{\mathrm{ct}}$ are stronger than $\mathrm{Adv}^{\mathrm{ph}}$ and $\mathrm{Madv}^{\mathrm{ph}}$, respectively. We prove this assertion with the following lemma.

Lemma 8.2.8. *For any Hermitian matrix M, $b \in \mathbb{R}_+$, $k \in \{0 \dots n\}$, and $\tau \in \{+1, -1\}$,*

(a) $-\mathcal{I}d \leq M \circ \hat{\Delta}_k^{\tau} \leq \mathcal{I}d \quad \Rightarrow \quad M - \pi\mathcal{I}d \leq M \circ \Delta_k^{\mathrm{ph}} \leq M + \pi\mathcal{I}d,$

(b) $-b\,M \leq M \circ \hat{\Delta}_k^{\tau} \leq b\,M \quad \Rightarrow \quad e^{-b\pi}\,M \leq M \circ \Delta_k^{\mathrm{ph}} \leq e^{b\pi}\,M$

Proof. As inequalities are invariant under the sign of τ,

$$\begin{aligned} -\mathcal{I}d \leq M \circ \hat{\Delta}_k^{+1} \leq \mathcal{I}d \quad &\Longleftrightarrow \quad -\mathcal{I}d \leq M \circ \hat{\Delta}_k^{-1} \leq \mathcal{I}d, \\ -b\,M \leq M \circ \hat{\Delta}_k^{+1} \leq b\,M \quad &\Longleftrightarrow \quad -b\,M \leq M \circ \hat{\Delta}_k^{-1} \leq b\,M, \end{aligned}$$

then we only prove the lemma for $\tau = +1$.
Let define for $s \in [0,1]$,

$$\Delta_k(s) = e^{s\pi\hat{\Delta}_k^{+1}},$$

So that,

$$\begin{aligned} \Delta_k(0) &= \mathbb{J}, \\ \Delta_k(1) &= \Delta_k^{\mathrm{ph}}, \\ \partial_s\Delta_k(s) &= \pi\hat{\Delta}_k^{+1} \circ \Delta_k(s). \end{aligned}$$

Note that $\Delta_k(s)$ is the unitary Gram matrix: $\mathrm{Gram}(e^{s\pi x_k} : x \in \mathcal{X})$, then $\Delta_k(s) \geq 0$.
Let ρ be a density matrix, we define $\langle M \rangle_s = \langle \rho, M \circ \Delta_k(s) \rangle$.

(a) $M \circ \hat{\Delta}_k^{+1} \leq \mathcal{I}d \quad \Rightarrow \quad M \circ \Delta_k^{\mathrm{ph}} \leq M + \pi\mathcal{I}d$

$$\begin{aligned} \partial_s \langle M \rangle_s &= \pi \left\langle \rho, M \circ \hat{\Delta}_k^{+1} \circ \Delta_k(s) \right\rangle, \\ &= \pi \left\langle \rho \circ \Delta_k(s), M \circ \hat{\Delta}_k^{+1} \right\rangle, \\ &\leq \pi \left\langle \rho \circ \Delta_k(s), \mathcal{I}d \right\rangle, \\ &\leq \pi \left\langle \rho, \mathcal{I}d \right\rangle, \end{aligned}$$

where $\rho \circ \Delta_k(s) \geq 0$ from Claim 2.1.6. By integration this inequality over $[0,1]$ we have,

$$\begin{aligned} \langle M \rangle_1 - \langle M \rangle_0 &\leq \pi \langle \rho, \mathcal{I}d \rangle, \\ \langle \rho, M \circ (\Delta_k - \mathbb{J}) \rangle &\leq \pi \langle \rho, \mathcal{I}d \rangle, \\ \langle \rho, M \circ \Delta_k \rangle &\leq \pi \langle \rho, M + \mathcal{I}d \rangle. \end{aligned}$$

Since this inequality holds for all density matrices ρ, we have

$$M \circ \Delta_k \leq M + \pi \mathcal{I}d.$$

The other inequality is proved similarly.

(b) $M \circ \hat{\Delta}_k^{+1} \leq b\,M \quad \Rightarrow \quad M \circ \Delta_k^{\text{ph}} \leq e^{b\pi}\,M$

$$\begin{aligned}\partial_s \langle M \rangle_s &= \pi \left\langle \rho, M \circ \hat{\Delta}_k^{+1} \circ \Delta_k(s) \right\rangle \\ &= \pi \left\langle \rho \circ \Delta_k(s), M \circ \hat{\Delta}_k^{+1} \right\rangle \\ &\leq \pi \left\langle \rho \circ \Delta_k(s), b\,M \right\rangle \\ &\leq \pi b \left\langle M \right\rangle_s .\end{aligned}$$

Dividing by $\langle M \rangle_s$, then integrating over $[0, 1]$, we have

$$\begin{aligned}\ln \langle M \rangle_1 - \ln \langle M \rangle_0 &\leq \pi b, \\ \langle \rho, M \circ \Delta_k \rangle &\leq e^{b\pi} \langle \rho, M \circ \mathbb{J} \rangle, \\ \langle \rho, M \circ \Delta_k \rangle &\leq e^{b\pi} \left\langle \rho, e^{b\pi} M \right\rangle .\end{aligned}$$

Since, this inequality holds for all density matrices ρ, we have

$$M \circ \Delta_k \leq e^{b\pi} M.$$

The other inequality is proved similarly. □

8.3 Adversary action Sadv

In this Section we define Sadv to be the adversary action, a lower bound method for Q_0^{ct}. This method is based on a semi-definite program $\mathcal{L}$ called the query Lagrangian. Once Sadv defined, we show that Sadv lower bounds Q_0^{ct}. We then prove several useful properties of $\mathcal{L}$ as strong duality. Afterwards we show several conditions that an optimal locally path γ of Sadv must satisfy. Finally, we prove that Sadv subsumes both $\text{Madv}_0^{\text{ct}}$ and Adv_0^{ct}.

8.3.1 Definition of Sadv

In this Subsection we describe a new lower bound for the quantum query complexity of state conversion in continuous time $Q_0^{ct}(\rho \to \sigma)$: the adversary action Sadv. This new method relies on Proposition 8.1.1. In this proposition we have described a semi-definite program that checks the feasibility of the evolution of a differentiable path $\gamma \in \Gamma_{\text{ct}}$ for a position $s \in [0, 1]$. From this semi-definite program, we define another semi-definite program $\mathcal{L}$ that still checks feasibility of γ on $s \in [0, 1]$, but also outputs the minimum β. We recall that the parameter β in Proposition 8.1.1 represents the infinitesimal number of queries.

Hence, for $\gamma \in G_N$ and $\eta \in T_\gamma G_N$, the tangent space of G_N, $\mathcal{L}$ outputs the infinitesimal minimal number of queries to go in direction η from the position γ. $\mathcal{L}$ has the same conditions that Proposition 8.1.1 and we minimize over β.

Definition 8.3.1. The **query Lagrangian** $\mathcal{L}$ is a semi-definite program with two parameters $\gamma \in S_+^N$ and $\eta \in S^N$, defined as

$$\mathcal{L}(\gamma,\eta) = \inf_{\substack{q\in\mathbb{R}\\ \gamma_k^\pm \in S_+^N\\ \gamma_0\in S_+^N}} q \quad \text{s.t.} \qquad q\gamma = \gamma_0 + \sum_{k\in\{1\ldots n\}} \gamma_k^+ + \gamma_k^-, \tag{8.8}$$

$$\eta = \sum_{k\in\{1\ldots n\}} \left(\gamma_k^+ - \gamma_k^-\right) \circ \hat{\Delta}_k. \tag{8.9}$$

We have changed some notation: β becomes q, $\hat{\Delta}_k^{+1}$ becomes $\hat{\Delta}_k$. We use the fact that $\hat{\Delta}_k^{+1} = -\hat{\Delta}_k^{-1}$ and $\hat{\Delta}_0 = 0$. We have also change the position of q in Equality constraints without loss of generality.

Since $\mathcal{L}$ defines a scalar field on $G_N \times T_\gamma G_N$, we define the integration of a differentiable path γ from ρ to σ as

$$\mathcal{L}[\gamma] = \int_\rho^\sigma ds\, \mathcal{L}\big(\gamma(s), \dot{\gamma}(s)\big),$$

where we use the Newton's notation, $\dot{\gamma}(s) = \frac{d\gamma}{ds}(s)$. From Formula 8.1, $\mathcal{L}[\gamma]$ gives the number of queries needed to follows the differentiable path γ. Note that $\mathcal{L}[\gamma]$ is a functional defined on Γ_{ct}.

Hence, we define $\mathrm{Sadv}(\rho \to \gamma_1)$ by minimizing over all differentiable paths in $\Gamma_{\mathrm{ct}}[\rho \to \sigma]$.

Definition 8.3.2. Let ρ and σ be two unitary Gram matrices. The **adversary action** Sadv is defined as

$$\mathrm{Sadv}(\rho \to \sigma) = \inf_{\gamma\in\Gamma_{\mathrm{ct}}[\rho\to\sigma]} \int_\gamma ds\, \mathcal{L}\big(\gamma(s), \dot{\gamma}(s)\big).$$

We introduce two notations to simplify it. For $\gamma \in \mathcal{C}^1\big([0,1], G_N\big)$

$$\mathcal{L}_\gamma(s) = \mathcal{L}\big(\gamma(s), \dot{\gamma}(s)\big),$$
$$\mathrm{Sadv}[\gamma] = \int_\gamma ds\, \mathcal{L}_\gamma(s).$$

Now that the adversary action Sadv has been defined, we prove that Sadv is a lower bound method for $Q_0^{ct}(\rho \to \sigma)$. To simplify the proof we use the following lemma.

Lemma 8.3.3. *Let $\mathcal{A}$ be a continuous-time quantum query algorithm with the following Hamiltonian,*

$$\hat{H}_x(s) = H_{\mathcal{D}}(s) + \pi\beta(s)H_{\mathcal{Q}}(x) \qquad \text{where} \qquad s\in[0,1] \text{ and } \|H_{\mathcal{Q}}\| \le 1.$$

If $\mathcal{A}$ converts ρ to σ through the differentiable path $\gamma(s)$, then for all $s \in [0,1]$, we have

$$\mathcal{L}_\gamma(s) \le \beta(s),$$
$$\mathrm{Sadv}[\gamma] \le q(\gamma).$$

Since Sadv is defined as a minimization over all γ in $\Gamma_{\mathrm{ct}}[\rho \to \sigma]$, we can conclude.

Theorem 8.3.4. *For any $N \in \mathbb{N}$, and $\rho, \sigma \in G_N$,*

$$\mathrm{Sadv}(\rho \to \sigma) \le Q_0^{ct}(\rho \to \sigma).$$

Proof of Lemma 8.3.3. Formula 8.3 implies that the algorithm $\mathcal{A}$ provide a feasible solution of $\mathcal{L}_\gamma(s)$ for all $s \in [0,1]$ with the objective value $\beta(s)$, then $\mathcal{L}_\gamma(s) \leq \beta(s)$. By integrating over γ we obtain $\text{Sadv}[\gamma] \leq q(\gamma)$.
Corollary 8.3.4 is obtained by minimizing over $\Gamma_{\text{ct}}[\rho \to \sigma]$. □

***Remark.* 1.** The query Lagrangian $\mathcal{L}(\gamma,\eta)$ can be interpreted as a Lagrangian in the phase space $G_N \times TG_N$, where TG_N is the tangent bundle of G_N and "query" has replaced "action". $\text{Sadv}[\gamma]$ is the necessary number of queries to follow the path γ and by minimizing over all paths in $\Gamma[\rho \to \sigma]$, we obtain $\text{Sadv}(\rho \to \sigma)$.

***Remark.* 2.** From Remark at the end of Section 8.1. We can conjecture that Sadv is tight for Q_0^{ct}.

8.3.2 Query Lagrangian and its properties

We begin this Subsection by providing a more practical form for the query Lagrangian $\mathcal{L}$. Afterwards we also dualize it to obtain the Lagrange dual problem $\mathcal{L}_d$. We denote $\mathcal{D}$ the domain of $\mathcal{L}$ and Λ the domain of $\mathcal{L}_d(\gamma,\eta)$.

Thereafter, to prove important properties of $\mathcal{L}$ we show that, if $\mathcal{L}$ is feasible then there exists an optimal solution, likewise for $\mathcal{L}_d$. From this proposition, we can prove these important properties: $\mathcal{L}$ is a norm and satisfies the strong duality if $\mathcal{L}$ is feasible. In order to prove this proposition we demonstrate that for each (γ,η) where $\mathcal{L}(\gamma,\eta)$ is feasible then we can restrict $\mathcal{D}$, the feasible set of $\mathcal{L}$, to a compact set $\hat{\mathcal{D}}_{\gamma,\eta}$ without loss of generality, likewise for $\mathcal{L}_d$.

We refine Definition 8.3.1 of $\mathcal{L}(\gamma,\eta)$. In Equality constraint (8.8) as γ, γ_0 and $\left(\gamma_k^+,\gamma_k^-\right)_k$ are semi-definite positive matrices then q is necessarily non-negative. Afterwards we remove γ_0 by replacing Equality constraint (8.8) by an inequality constraint.

Definition. (Primal form of $\mathcal{L}$)
The query Lagrangian $\mathcal{L}$ is a semi-definite program with two parameters $\gamma \in S_+^N$ and $\eta \in S^N$, defined as

$$\mathcal{L}(\gamma,\eta) = \inf_{\substack{q\in\mathbb{R}_+\\ \gamma_k^\pm \in S_+^N}} q \quad \text{s.t.} \qquad q\gamma \geq \sum_{k\in\{1\ldots n\}} \gamma_k^+ + \gamma_k^-, \tag{8.10}$$

$$\eta = \sum_{k\in\{1\ldots n\}} \left(\gamma_k^+ - \gamma_k^-\right) \circ \hat{\Delta}_k. \tag{8.11}$$

Claim 8.3.5. (Dual form of $\mathcal{L}$)
The dual form of $\mathcal{L}$, denoted $\mathcal{L}_d$, is a semi-definite program with two parameters $\gamma \in S_+^N$ and $\eta \in S^N$, defined as

$$\mathcal{L}_d(\gamma,\eta) = \sup_{\substack{\mathcal{U}\in\mathcal{S}^N\\ \mathcal{V}\in\mathcal{S}_+^N}} \langle \mathcal{U},\eta \rangle \quad \text{s.t.} \qquad \forall k \in \{1\ldots n\}, \quad -\mathcal{V} \leq \mathcal{U} \circ \hat{\Delta}_k \leq \mathcal{V}, \tag{8.12}$$

$$\langle \mathcal{V},\gamma \rangle \leq 1. \tag{8.13}$$

Proof. We construct $L(\gamma,\eta,q,\gamma_k^\pm,\mathcal{V},\mathcal{U})$ the Lagrange dual function of $\mathcal{L}(\gamma,\eta)$ from Equation

(6.3) we have,

$$\begin{aligned}
L(\gamma,\eta,q,\gamma_k^{\pm},\mathcal{V},\mathcal{U}) &= q + \langle \mathcal{V}, \sum_{k\in\{1\ldots n\}} (\gamma_k^+ + \gamma_k^-) - q\gamma\rangle + \langle \mathcal{U}, \eta - \sum_{k\in\{1\ldots n\}} (\gamma_k^+ - \gamma_k^-)\circ \hat{\Delta}_k\rangle,\\
&= \langle \mathcal{U},\eta\rangle + q(1-\langle \mathcal{V},\gamma\rangle) + \sum_{k\in\{1\ldots n\}} \langle \gamma_k^+, \mathcal{V} - \mathcal{U}\circ\hat{\Delta}_k\rangle\\
&\qquad + \sum_{k\in\{1\ldots n\}} \langle \gamma_k^-, \mathcal{V} + \mathcal{U}\circ\hat{\Delta}_k\rangle.
\end{aligned}$$

From Definition 6.4.1, the Lagrange dual function is

$$\begin{aligned}
d(\gamma,\eta,\mathcal{V},\mathcal{U}) &= \inf_{\substack{q\in\mathbb{R}_+\\ \gamma_k^{\pm}\in S_+^N}} L(\gamma,\eta,q,\gamma_k^{\pm},\mathcal{V},\mathcal{U}),\\
&= \begin{cases} \langle \mathcal{U},\eta\rangle & \text{if } \langle \mathcal{V},\gamma\rangle \le 1 \text{ and } \forall k\in\{1\ldots n\},\ -\mathcal{V}\le \mathcal{U}\circ\hat{\Delta}_k \le \mathcal{V},\\ -\infty & \text{otherwise.}\end{cases}
\end{aligned}$$

Finally the Lagrange dual problem is derived from Definition 6.4.3,

$$\mathcal{L}_d(\gamma,\eta) = \sup_{\substack{\mathcal{V}\in S_+^N\\ \mathcal{U}\in S^N}} d(\gamma,\eta,\mathcal{V},\mathcal{U}).$$

□

By a slight abuse of notation, $\mathcal{L}(\gamma,\eta)$ describes a semi-definite program and its optimal value $q^\star(\gamma,\eta)$. Similarly for $\mathcal{L}_d(\gamma,\eta)$ and its dual optimal value $d^\star(\gamma,\eta)$.

Remark.
- From the dual form $\mathcal{L}_d$, we can observe that the optimal value $q^\star$ is necessarily positive. For each feasible solution $(\mathcal{U},\mathcal{V})$ of $\mathcal{L}_d(\gamma,\eta)$ then $(-\mathcal{U},\mathcal{V})$ is also feasible. This fact implies that $\mathcal{L}_d(\gamma,\eta)\ge 0$, as well for $\mathcal{L}(\gamma,\eta)$ by weak duality.
- γ_0 can be interpreted as a query that does not provide any data.
- From (8.10), if $q^\star$ is the optimal value of the primal form $\mathcal{L}$, then there exists a feasible solution for each value $q' > q$.
- From the dual form $\mathcal{L}_d$, since $\mathcal{V}\ge 0$ and $\gamma\ge 0$ the second constraint (8.13) is lower bounded by 0.

In the next Proposition 8.3.6 we show that if $\mathcal{L}(\gamma,\eta)$ and $\mathcal{L}_d(\gamma,\eta)$ are feasible, then their respective domains $\mathcal{D}$ and Λ can be restricted to compact domains $\hat{\mathcal{D}}_{\gamma,\eta}$ and $\hat{\Lambda}_{\gamma,\eta}$ without loss of generality.

Proposition 8.3.6. *Let $\gamma\in S_+^N$ and $\eta\in S^N$.*
- *If $\mathcal{L}(\gamma,\eta)$ is feasible, then there exists a compact set $\hat{\mathcal{D}}_{\gamma,\eta}\subset\mathcal{D}$ such that*

$$\mathcal{L}(\gamma,\eta) = \min_{(q,\gamma_k^{\pm})\in\hat{\mathcal{D}}_{\gamma,\eta}} q \quad s.t. \quad \begin{aligned} q\gamma &\ge \sum_{k\in\{1\ldots n\}} \gamma_k^+ + \gamma_k^-,\\ \eta &= \sum_{k\in\{1\ldots n\}} (\gamma_k^+ - \gamma_k^-)\circ\hat{\Delta}_k.\end{aligned}$$

- *If $\mathcal{L}_d(\gamma,\eta)$ is feasible, then there exists a compact set $\hat{\Lambda}_{\gamma,\eta} \subset \Lambda$ such that*

$$\mathcal{L}_d(\gamma,\eta) = \max_{(\mathcal{V},\mathcal{U})\in\hat{\Lambda}_{\gamma,\eta}} \langle \mathcal{U},\eta\rangle \quad s.t. \qquad \forall k \in \{1\dots n\}, \quad -\mathcal{V} \leq \mathcal{U}\circ\hat{\Delta}_k \leq \mathcal{V}, \qquad \langle \mathcal{V},\gamma\rangle \leq 1.$$

Corollary 8.3.7. *Let $\gamma \in S_+^N$ and $\eta \in S^N$. $\mathcal{L}(\gamma,\eta)$ is feasible if and only if $X^\star(\gamma,\eta)$ is non-empty. $\mathcal{L}_d(\gamma,\eta)$ is feasible if and only if $Y^\star(\gamma,\eta)$ is non-empty.*

The proof of Proposition 8.3.6 uses the following lemma. From Inequality constraint (8.11), we observe that matrices $(\hat{\Delta}_k)_k$ induces a restriction on η. Indeed if $(\hat{\Delta}_k)_k$ are nulls then $\mathcal{L}(\,\cdot\,,\eta)$ is unfeasible for $\eta \neq 0$. The following lemma generalizes this idea.

Lemma 8.3.8. *Let $\gamma \in S_+^N$ and $\eta \in S^N$. If supp$(\eta) \nsubseteq \Delta$ then $\mathcal{L}(\gamma,\eta)$ is unfeasible, and $\mathcal{L}_d(\gamma,\eta)$ is unbounded, where $\Delta = \cup_{k\in\{1\dots n\}}$ supp$(\hat{\Delta}_k)$.*

Proof. Proof by contradiction.
Let (x,y) be in *supp*$(\eta)\setminus\Delta$, λ be a real number and $(e_x)_{x\in\mathcal{X}}$ the standard basis.
Since η and $(\hat{\Delta}_k)_k$ are in S^n, then (y,x) is also in *supp*$(\eta)\setminus\Delta$. Therefore

$$\mathcal{V} = 0 \qquad \text{and} \qquad \mathcal{U} = \lambda\big(\eta[x,y]e_x^* e_y + \bar{\eta}[x,y]e_y^* e_x\big),$$

is a feasible solution of $\mathcal{L}_d(\gamma,\eta)$ since $\mathcal{U}\circ\hat{\Delta}_k$ is a null matrix for all $k \in \{1\dots n\}$. Moreover its value is

$$\langle \mathcal{U},\eta\rangle = 2\lambda|\eta[x,y]|^2.$$

Since $\eta[x,y]$ is not null, then $\mathcal{L}_d(\gamma,\eta)$ is unbounded. $\mathcal{L}(\gamma,\eta)$ is unfeasible by weak duality. □

Proof of Proposition 8.3.6. Let be $\gamma \in S_+^N$ and $\eta \in S^N$.

(Primal form) If $\mathcal{L}(\gamma,\eta)$ is feasible, then there exists a sequence of feasible points $(q_r,\gamma_{k,r}^{\pm})_{r\in\mathbb{N}}$ satisfying constraints of $\mathcal{L}(\gamma,\eta)$. As $\mathcal{L}(\gamma,\eta)$ is finite, q_r converge to q_0 with $q_0 = \mathcal{L}(\gamma,\eta)$.
From Inequality constraint (8.10), each $\gamma_{k,r}^{\pm} \geq 0$ is upper bounded by $q_r\gamma \geq \gamma_{k,r}^{\pm}$. Therefore, for each k there exists a sub-sequential of $(\gamma_{k,r}^{\pm})_r$ with the limit in the compact set

$$\big\{M \in S^N : 0 \leq M \leq q_0\gamma\big\}.$$

(Dual form) If $\mathcal{L}_d(\gamma,\eta)$ is feasible, then there exists a sequence of feasible dual points $(\mathcal{V}_r,\mathcal{U}_r)_{r\in\mathbb{N}}$ satisfying constraints (8.12), (8.13) of $\mathcal{L}_d(\gamma,\eta)$. **Assume** that $\gamma \geq 0$ is full rank, then the second inequality constraint $\langle \mathcal{V}_r,\gamma\rangle \leq 1$ implies that $(\mathcal{V}_r)_r$ is in the compact set

$$\big\{M \in S^N : 0 \leq M \leq \frac{1}{N}\gamma^{-1}\big\}.$$

From first inequality constraints and the fact that $\mathcal{V}_r$ is upper bounded, we have for each r and $k \in \{1\dots n\}$

$$-\frac{1}{N}\gamma^{-1} \leq \mathcal{U}_r\circ\hat{\Delta}_k \leq \frac{1}{N}\gamma^{-1}.$$

While $\mathcal{U}_r \circ \hat{\Delta}_k$ is bounded for all $k \in \{1 \dots n\}$, this is not sufficient to imply that $\mathcal{U}_r$ is bounded because it may exist free variables eliminated by $(\hat{\Delta}_k)_k$. Instead, we can nullify free variables of $\mathcal{U}_r$ using Lemma 8.3.8. Let define $1_\Delta \in \mathcal{M}_N(\{0,1\})$

$$1_\Delta[x,y] = \begin{cases} 1 & \text{if } \exists k \in \{1 \dots n\},\ \hat{\Delta}_k[x,y] \neq 0, \\ 0 & \text{otherwise.} \end{cases}$$

From $(\mathcal{U}_r)_r$ we construct the sequence $(\mathcal{U}_r \circ 1_\Delta)_r$ bounded for all r. The new sequence satisfies first constraint (8.12) since $\hat{\Delta}_k = 1_\Delta \circ \hat{\Delta}_k$ for all $k \in \{1 \dots n\}$ and with the objective value $\lim_r \langle \mathcal{U}_r \circ 1_\Delta, \eta \rangle$. Using the contrapositive of Lemma 8.3.8, as $\mathcal{L}_d(\gamma,\eta)$ is finite then $supp(\eta) \subseteq \Delta$, in other words $\eta \circ 1_\Delta = \eta$.
If $\gamma \geq 0$ is not full rank, then there exists free variables in $\mathcal{V}_r$ non restricted by $\langle \mathcal{V}_r, \gamma \rangle \leq 1$, but those free variables don't appear in $\mathcal{U}_r$ otherwise $\mathcal{L}_d$ would be unbounded, then we can set these free-variables to zero. □

From Corollary 8.3.7 we can prove that $\mathcal{L}$ satisfies strong duality if $\mathcal{L}$ is feasible. We also show that L, the Lagrange dual function introduced in the proof of Claim 8.3.5, has a saddle-point if $\mathcal{L}$ feasible.

Proposition 8.3.9. *For all $\gamma \in S_+^N$ and $\eta \in S^N$. If $\mathcal{L}(\gamma,\eta)$ is feasible. then $\mathcal{L}(\gamma,\eta)$ satisfies the strong duality. Moreover, there exists an optimal point $(q^\star, \gamma_k^{\pm,\star})$ and a dual optimal point $(\mathcal{U}^\star, \mathcal{V}^\star)$ such that L, the Lagrangian of $\mathcal{L}(\gamma,\eta)$, has a saddle-point*

$$\forall q, \gamma_k^\pm, \mathcal{V}, \mathcal{U}, \qquad L(\gamma,\eta,q^\star,\gamma_k^{\pm,\star},\mathcal{V},\mathcal{U}) \leq L(\gamma,\eta,q^\star,\gamma_k^{\pm,\star},\mathcal{V}^\star,\mathcal{U}^\star) \leq L(\gamma,\eta,q,\gamma_k^\pm,\mathcal{V}^\star,\mathcal{U}^\star).$$

Proof. **(Strong duality)** From Corollary 8.3.7 if $\mathcal{L}(\gamma,\eta)$ is feasible there exists an optimal solution $\left(q^\star, \gamma_k^{\pm,\star}\right)$. Therefore, for $s > q$, the feasible solution $\left(s, \gamma_k^{\pm,\star}\right)$ is a strictly feasible solution, hence the Slater's condition is satisfied.

(Optimal point) Existence of a dual optimal point $(\mathcal{U}^\star, \mathcal{V}^\star)$ comes from Corollary 8.3.7.

(Saddle point) Directly from Property 6.5.2.

□

We conclude that $\mathcal{L}(\gamma,\eta)$ satisfies all conditions to be a norm in η for all γ semi-definite positive.

Proposition 8.3.10. *For all $\gamma \in S_+^N$ and $\eta \in S^N$, we have*

*(a) **(positive)** $\mathcal{L}(\gamma,\eta) \geq 0$,*

*(b) **(absolutely homogeneous)** for all $\lambda \in \mathbb{R}$, $\mathcal{L}(\gamma,\lambda\eta) = |\lambda|.\mathcal{L}(\gamma,\eta)$,*

*(c) **(triangle inequality)** $\mathcal{L}(\gamma,\eta_1+\eta_2) \leq \mathcal{L}(\gamma,\eta_1) + \mathcal{L}(\gamma,\eta_2)$,*

*(d) **(zero matrix)** $\mathcal{L}(\gamma,\eta) = 0 \quad \Leftrightarrow \quad \eta = 0$.*

Corollary 8.3.11. *For all $\gamma \in S_+^N$, $\mathcal{L}(\,\gamma\,,\cdot)$ is a norm in S^N.*

Proof of Proposition 8.3.10. (a) The point $(0,0)$ is always feasible for the dual form and its value is 0. By weak duality $\mathcal{L}$ is always positive.

(b) Let λ be a strictly positive real and $\left(q, \gamma_k^{\pm}\right)$ be an optimal point of $\mathcal{L}(\gamma, \lambda\eta)$. Then $\left(\lambda q, \lambda^{-1}\gamma_k^{\pm}\right)$ is a feasible point of $\mathcal{L}(\gamma, \eta)$. Moreover, by reciprocity this is also an optimal point.
Let $\left(q, \gamma_k^{\pm}\right)$ be an optimal point of $\mathcal{L}(\gamma, \eta)$, then $\left(q, \gamma_k^{\mp}\right)$ is a feasible point of $L(\gamma, -\eta)$. Moreover, by reciprocity this is also an optimal point.
For $\mathcal{L}(\gamma, 0)$, the optimal value is reached by the optimal point $(0, 0)$.

(c) Let $\left(q_1, \gamma_{k,1}^{\pm}\right)$ and $\left(q_2, \gamma_{k,2}^{\pm}\right)$ respective optimal points of $\mathcal{L}(\gamma, \eta_1)$ and $\mathcal{L}(\gamma, \eta_2)$. Then $\left(q_1 + q_2, \gamma_{k,1}^{\pm} + \gamma_{k,2}^{\pm}\right)$ is a feasible point of $\mathcal{L}(\gamma, \eta_1 + \eta_2)$.

(d) If $\mathcal{L}(\gamma, \eta) = 0$, the first constraint of the primal form implies that all $\gamma_k^{\pm}$ are null matrices, hence $\eta = 0$. Reciprocally, if $\eta = 0$, the optimal value is 0 reached by $(0, 0)$.

□

8.3.3 Necessary conditions on Sadv

Let ρ and σ be in G_N. We define $\gamma^\star$ to be a double-differentiable path in $\Gamma_{\mathrm{ct}}[\rho \to \sigma]$ that is a local minimal of $\mathrm{Sadv}(\rho \to \sigma)$ For all $s \in [0,1]$, we associate to the path $\gamma^\star$ an optimal point selection of $\mathcal{L}_{\gamma^\star}(s)$ defined as,

$$\left(q^\star(s), \gamma_k^{\pm,\star}(s), \mathcal{U}^\star(s), \mathcal{V}^\star(s)\right).$$

In this subsection, we derive from the Euler-Lagrange theorem the necessary conditions that must satisfy a local minimal $\gamma^\star$ of $\mathrm{Sadv}(\rho \to \sigma)$. More precisely, these conditions are on the optimal point selection $\left(q^\star(s), \gamma_k^{\pm,\star}(s), \mathcal{U}^\star(s), \mathcal{V}^\star(s)\right)$ that describes $\mathcal{L}_{\gamma^\star}(s)$. Note that $(\mathcal{U}^\star(s), \mathcal{V}^\star(s))$ describes $\mathcal{L}_{\gamma^\star}(s)$ almost everywhere since, if $\mathrm{Sadv}(\rho \to \sigma)$ is finite then $\mathcal{L}_{\gamma^\star}(s)$ is integrable on $[0,1]$. Hence, Proposition 8.3.9 implies that the strong duality holds almost everywhere.

At first, we use the KKT conditions 6.6.1, Complementary slackness from Corollary 6.5.3 and Strong duality 6.7.8. Latter we use the Euler-Lagrange equation D.0.2 on $\mathrm{Sadv}(\rho \to \sigma)$ to obtain another condition on $(\mathcal{U}^\star(s), \mathcal{V}^\star(s))$. Since we must derive $\mathcal{L}_{\gamma^\star}$ we apply the Envelope theorem 6.8.1 to prove the existence of derivatives of $\mathcal{L}_{\gamma^\star}$ almost everywhere on $[0,1]$.

From Proposition 6.7.6, we know that the KKT conditions hold for almost all $s \in [0,1]$ since the objective function and inequality constraints of $\mathcal{L}_{\gamma^\star}(s)$ are differentiable, and the duality holds almost everywhere, likewise the Complementary slackness Corollary 6.5.3. Moreover, the KKT conditions are sufficient since $\mathcal{L}$ is a semi-definite program, thus a convex optimization problem.

Fact 8.3.12. Let $\left(q^\star(s), \gamma_k^{\pm,\star}(s), \mathcal{U}^\star(s), \mathcal{V}^\star(s)\right)$ be an optimal point selection of $\mathcal{L}_{\gamma^\star}$. Therefore for almost all $s \in [0,1]$

- $$\dot{\gamma}^\star(s) = \sum_{k\in\{1\dots n\}} \Big(\gamma_k^{+,\star}(s) - \gamma_k^{-,\star}(s)\Big) \circ \hat{\Delta}_k \tag{8.14}$$
- $$q^\star(s)\gamma^\star(s) \geq \sum_{k\in\{1\dots n\}} \Big(\gamma_k^{+,\star}(s) + \gamma_k^{-,\star}(s)\Big) \tag{8.15}$$
- $$\langle \mathcal{V}^\star(s), \gamma^\star(s)\rangle = 1 \tag{8.16}$$
- $$\mathcal{V}^\star(s) \mp \mathcal{U}^\star(s) \circ \hat{\Delta}_k \geq 0 \qquad \forall k \in \{1\dots n\} \tag{8.17}$$
- $$\Big\langle \gamma_k^{\pm,\star}(s), \mathcal{V}^\star(s) \mp \mathcal{U}^\star(s) \circ \hat{\Delta}_k \Big\rangle = 0 \qquad \forall k \in \{1\dots n\} \tag{8.18}$$
- $$\left\langle \mathcal{V}^\star(s), q^\star(s)\gamma^\star(s) - \sum_{k\in\{1\dots n\}} \Big(\gamma_k^{+,\star}(s) + \gamma_k^{-,\star}(s)\Big) \right\rangle = 0 \tag{8.19}$$
- $$\langle \mathcal{U}^\star(s), \dot{\gamma}^\star(s)\rangle = q^\star(s) \tag{8.20}$$

Proof. (8.14) and (8.15) are conditions of the primal form $\mathcal{L}$. (8.16) and (8.17) are conditions of the primal form $\mathcal{L}_d$ where (8.16) is now tight from the KKT conditions. (8.18) and (8.19) are from the Complementary slackness. (8.20) is from the strong duality. □

The following Envelope theorem proves that $\mathcal{L}_{\gamma^\star}$ is differentiable almost everywhere on $[0,1]$, in other words $\mathcal{L}_{\gamma^\star}$ is absolutely continuous. Then, we apply the Euler-Lagrange equation.

Theorem 8.3.13. *Let ρ and σ be in G_N, $\gamma^\star$ be a double differentiable path in $\Gamma[\rho \to \sigma]$ that is a locally optimal path of* $\mathrm{Sadv}(\rho \to \sigma)$ *with $\mathcal{L}[\gamma^\star]$ finite, and $\mathcal{U}^\star(s)$ and $\mathcal{V}^\star(s)$ be a dual optimal point selection of $\mathcal{L}_{\gamma^\star}(s)$ for almost all $s \in [0,1]$. Then the function $\mathcal{L}_{\gamma^\star}(s)$ is absolutely continuous on $[0,1]$, and for almost all $s \in [0,1]$, we have*

$$\mathcal{L}_{\gamma^\star}(1) = \mathcal{L}_{\gamma^\star}(0) + \int_0^1 ds \Big[- q^\star(s) \big\langle \mathcal{V}^\star(s), \frac{d\gamma^\star}{ds}(s) \big\rangle + \big\langle \mathcal{U}(s)^\star, \frac{d^2\gamma^\star}{ds^2}(s) \big\rangle \Big],$$

where $q^\star(s) = \mathcal{L}_{\gamma^\star}(s) = \langle \mathcal{U}^\star(s), \dot{\gamma}^\star(s) \rangle$.

Proof. Since $\mathcal{L}[\gamma^\star]$ is finite then $\mathcal{L}_{\gamma^\star}(s)$ is finite for almost all s. From Proposition 8.3.9, we have

$$\mathcal{L}_{\gamma^\star}(s) = \min_{(q,\gamma_k^\pm) \in \hat{\mathcal{D}}_{\gamma^\star(s),\dot{\gamma}^\star(s)}} \max_{(\mathcal{V},\mathcal{U}) \in \hat{\Lambda}_{\gamma^\star(s),\dot{\gamma}^\star(s)}} L(\gamma^\star(s), \dot{\gamma}^\star(s), q, \gamma_k^\pm, \mathcal{V}, \mathcal{U}).$$

Where the Lagrangian is defined as,

$$\begin{aligned} L(\gamma^\star(s), \dot{\gamma}^\star(s), q, \gamma_k^\pm, \mathcal{V}, \mathcal{U}) = {} & \big\langle \mathcal{U}, \dot{\gamma}^\star(s) \big\rangle + q\big(1 - \big\langle \mathcal{V}, \gamma^\star(s) \big\rangle\big) \\ & + \sum_{k \in \{1 \dots n\}} \big\langle \gamma_k^+, \mathcal{V} - \mathcal{U} \circ \hat{\Delta}_k \big\rangle \\ & + \sum_{k \in \{1 \dots n\}} \big\langle \gamma_k^-, \mathcal{V} + \mathcal{U} \circ \hat{\Delta}_k \big\rangle. \end{aligned}$$

First, we prove that $\mathcal{L}_{\gamma^\star}(s)$ is absolutely continuous on $[0,1]$ using the Enveloppe theorem 6.8.1. To make this we show that the Lagrangian satisfies all six conditions **(1)-(6)** from the Envelope theorem.

(1) Since $\mathcal{L}_{\gamma^\star}(s)$ is finite for almost all s, Proposition 8.3.9 implies that the existence of optimal primal/dual solution for almost all s .

(2) The Langragian is is absolutely continuous in s for all $\mathcal{U}$, $\mathcal{V}$, q and $\gamma_k^\pm$.

(5) The derivative of the Lagrange has the following form

$$D_s\, L(\gamma^\star(s), \dot{\gamma}^\star(s), q, \gamma_k^\pm, \mathcal{V}, \mathcal{U}) = \big\langle \mathcal{U}, \frac{d^2\gamma^\star}{ds^2}(s) \big\rangle - q \big\langle \mathcal{V}, \frac{d\gamma^\star}{ds}(s) \big\rangle.$$

As $\gamma^\star$ is double differentiable then L is differentiable in $s \in [0,1]$ for all $\mathcal{U}$, $\mathcal{V}$, q and $\gamma_k^\pm$. Hence, $D_s\, L$ is continuous in $s \in [0,1]$, and then absolutely continuous.

(3) From the point **(5)**, we observe that $|D_s\, L|$ increased absolutely linearly in $\mathcal{U}$, $\mathcal{V}$ or q. Since Proposition 8.3.6 shows that $\mathcal{D}$ and Λ can restricted to compact sets $\hat{\mathcal{D}}$ and $\hat{\Lambda}$ without loss of generality, we define

$$\hat{\mathcal{D}} = \bigcup_{s \in [0.1]} \hat{\mathcal{D}}_{\gamma^\star(s),\dot{\gamma}^\star(s)} \qquad \text{and} \qquad \hat{\Lambda} = \bigcup_{s \in [0.1]} \hat{\Lambda}_{\gamma^\star(s),\dot{\gamma}^\star(s)},$$

where $\hat{\mathcal{D}}_{\gamma,\eta}$ and $\hat{\Lambda}_{\gamma,\eta}$ are described in the proof of Proposition 8.3.6. Hence, $|D_s\, L|$ can be bounded for almost all s.

(6) From point **(2)**, the Lagrangian L is differentiable in $s \in [0,1]$ for all $\mathcal{U}$, $\mathcal{V}$, q and $\gamma_k^\pm$ in $\mathcal{D} \times \Lambda$. if we restrict the domain of the Lagrangian to $\hat{\mathcal{D}}_{\gamma,\eta} \times \hat{\Lambda}_{\gamma,\eta}$, then its family is equi-differentiable.

(4) $\mathcal{D} \times \Lambda$ satisfies the axiom of countability, since they are included in $\mathbb{R}^n$.

Now we conclude by derivation the Lagrangian in γ and $\frac{d\gamma}{ds}$.

$$\frac{dL(s)}{ds} = \big\langle \frac{\partial L(s)}{\partial \gamma^\star(s)}, \frac{d\gamma^\star(s)}{ds} \big\rangle + \big\langle \frac{\partial L(s)}{\partial \dot{\gamma}^\star(s)}, \frac{d^2\gamma^\star(s)}{ds^2} \big\rangle,$$

where $L(s)$ abbreviates $L(\gamma^\star(s), \dot{\gamma}^\star(s), q, \gamma_k^\pm, \mathcal{V}, \mathcal{U})$.

$$\frac{\partial L(s)}{\partial \gamma^\star(s)} = \big(\mathcal{U}^\star(s)\big)^*$$
$$\frac{\partial L(s)}{\partial \dot{\gamma}^\star(s)} = -q^\star(s)\big(\mathcal{V}^\star(s)\big)^*.$$

□

Now, we can use the Euler-Lagrange equation on the Lagrangian, since the Enveloppe theorem ensures that the Lagrangian is differentiable almost everywhere when Sadv is finite.

Theorem 8.3.14. *Let ρ and σ be in G_N, $\gamma^\star$ be a double differentiable path in $\Gamma[\rho \to \sigma]$ that is a locally optimal path of* $\mathrm{Sadv}(\rho \to \sigma)$ *with $\mathcal{L}[\gamma^\star]$ finite, and $\mathcal{U}^\star(s)$ and $\mathcal{V}^\star(s)$ be a dual optimal point selection of $\mathcal{L}_{\gamma^\star}(s)$ for almost all $s \in [0,1]$. Then, the following equation is satisfied for almost all $s \in [0,1]$*

$$\mathcal{V}^\star(s) = -\frac{\frac{d}{ds}\mathcal{U}^\star(s)}{\big\langle \mathcal{U}^\star(s), \frac{d\gamma^\star}{ds}(s) \big\rangle}.$$

Proof. This is just a simple application of the Euler-Lagrange equation D.0.2,

$$\frac{\partial L}{\partial \gamma^\star} - \frac{d}{ds}\frac{\partial L}{\partial \dot{\gamma}^\star} = 0,$$

with the fact that $q^\star(s) = \langle \mathcal{U}^\star(s), \dot{\gamma}^\star(s) \rangle$ for almost all s. □

8.3.4 Relation with $\mathrm{Adv}^{\mathrm{ct}}$ and $\mathrm{Madv}^{\mathrm{ct}}$

To complete this Chapter, we show that the adversary action Sadv upper bounds $\mathrm{Adv}^{\mathrm{ct}}$ and $\mathrm{Madv}^{\mathrm{ct}}$.

Theorem 8.3.15. *Let ρ and σ be two unitary Gram matrices. Therefore*

$$\mathrm{Adv}_0^{\mathrm{ct}}(\rho \to \sigma) \leq \mathrm{Sadv}(\rho \to \sigma),$$
$$\mathrm{Madv}_0^{\mathrm{ct}}(\rho \to \sigma) \leq \mathrm{Sadv}(\rho \to \sigma).$$

To make the proof more convenient, we include several reminders.

Definition. (Adversary action)

$$\mathrm{Sadv}(\rho \to \sigma) = \inf_{\gamma \in \Gamma_{\mathrm{ct}}[\rho \to \sigma]} \int_\gamma ds\, \mathcal{L}\big(\gamma(s), \frac{d}{ds}\gamma(s)\big),$$

where the query Lagrangian $\mathcal{L}_d$ is defined as,

$$\mathcal{L}_d(\gamma,\eta) \;=\; \sup_{\substack{\mathcal{U}\in\mathcal{S}^N\\ \mathcal{V}\in\mathcal{S}^N_+}} \langle \mathcal{U},\eta\rangle \quad \text{subject to} \quad \forall k\in\{1\dots n\},\quad -\mathcal{V}\le \mathcal{U}\circ\hat{\Delta}_k\le\mathcal{V}, \qquad \langle\mathcal{V},\gamma\rangle\le 1.$$

Definition (Adversary method for continuous time)**.**

$$\mathrm{Adv}_0^{\mathrm{ct}}(\rho\to\sigma) = \sup_{\substack{M\\ v:\|v\|=1}} \quad \langle M\circ vv^*,\sigma-\rho\rangle\,,$$
$$\text{subject to}\quad \forall\, k\in\{0\dots n\},\ \tau=\pm \qquad -\mathcal{I}d\le M\circ\hat{\Delta}_k^{\tau}\le\mathcal{I}d.$$

Proof for $\mathrm{Adv}_0^{\mathrm{ct}}\le\mathrm{Sadv}$. Let γ be a path in $\Gamma_{\mathrm{ct}}[\rho\to\sigma]$, M a Hermitian matrix and v be a unit vector. For each $s\in[0,1]$, we choose

$$\mathcal{U}(s)=M\circ vv^*,\qquad\text{and}\qquad \mathcal{V}(s)=\mathcal{I}d\circ vv^*,$$

as feasible solution of $\mathcal{L}\big(\gamma(s),\frac{d}{ds}\gamma(s)\big)$, since $\langle\gamma(s),\mathcal{V}(s)\rangle\le 1$. Hence,

$$\begin{aligned}
\inf_{\gamma\in\Gamma_{\mathrm{ct}}[\rho\to\sigma]}\int_\gamma ds\,\mathcal{L}\big(\gamma(s),\frac{d}{ds}\gamma(s)\big) &= \inf_{\gamma\in\Gamma_{\mathrm{ct}}[\rho\to\sigma]}\int_\gamma ds\,\sup_{\mathcal{U}\in\mathcal{S}^N}\left\langle\mathcal{U},\frac{d}{ds}\gamma(s)\right\rangle,\\
&\ge \inf_{\gamma\in\Gamma_{\mathrm{ct}}[\rho\to\sigma]}\sup_{\mathcal{U}\in\mathcal{S}^N}\int_\gamma ds\left\langle\mathcal{U},\frac{d}{ds}\gamma(s)\right\rangle,\\
&\ge \inf_{\gamma\in\Gamma_{\mathrm{ct}}[\rho\to\sigma]}\sup_{\mathcal{U}\in\mathcal{S}^N}\langle\mathcal{U},\sigma-\rho\rangle\,,\\
&\ge \sup_{\mathcal{U}\in\mathcal{S}^N}\langle\mathcal{U},\sigma-\rho\rangle\,,\\
&\ge \sup_{\substack{M\\ v:\|v\|=1}}\ \langle M\circ vv^*,\sigma-\rho\rangle\,,
\end{aligned}$$

where the minimization over Γ_{ct} disappears after the integration. Note that conditions are weaker since

$$-\mathcal{I}d\le M\circ\hat{\Delta}_k^{+1}\le\mathcal{I}d,\qquad\Longrightarrow\qquad -\mathcal{I}d\circ vv^*\le M\circ\hat{\Delta}_k\circ vv^*\le\mathcal{I}d\circ vv^*,$$

where we use Claim 2.1.6 and the fact that $vv^*\ge 0$. □

Definition (Multiplicative adversary method for continuous time)**.**

$$\mathrm{Madv}_0^{\mathrm{ct}}(\rho\to\sigma) = \sup_{b>0}\frac{1}{b}\sup_{\substack{M\ge 0\\ v:\|v\|=1}}\quad\big[\ln\big\langle M\circ vv^*,\sigma\big\rangle-\ln\big\langle M\circ vv^*,\rho\big\rangle\big],$$
$$\text{subject to}\quad\forall\, k\in\{1\dots n\},\forall\tau\in\{+,-\},\qquad -bM\le M\circ\hat{\Delta}_k^{\tau}\le bM.$$

Proof for $\mathrm{Madv}_0^{\mathrm{ct}}\le\mathrm{Sadv}$. Let γ be a path in $\Gamma_{\mathrm{ct}}[\rho\to\sigma]$, M a Hermitian matrix, v be a unit vector and b be a strictly positive real. For each $s\in[0,1]$, we chose

$$\mathcal{U}(s)=\frac{M\circ vv^*}{b\,\langle M\circ vv^*,\gamma(s)\rangle},\qquad\text{and}\qquad \mathcal{V}(s)=\frac{M\circ vv^*}{\langle M\circ vv^*,\gamma(s)\rangle},$$

as feasible solution of $\mathcal{L}\big(\gamma(s), \frac{d}{ds}\gamma(s)\big)$, since $\langle \gamma(s), \mathcal{V}(s)\rangle = 1$. Hence,

$$\begin{aligned}
\inf_{\gamma\in\Gamma_{\mathrm{ct}}[\rho\to\sigma]} \int_\gamma ds\, \mathcal{L}\big(\gamma(s), \frac{d}{ds}\gamma(s)\big) &= \inf_{\gamma\in\Gamma_{\mathrm{ct}}[\rho\to\sigma]} \int_\gamma ds \sup_{\mathcal{U}\in\mathcal{S}^N} \left\langle \mathcal{U}, \frac{d}{ds}\gamma(s)\right\rangle, \\
&\geq \inf_{\gamma\in\Gamma_{\mathrm{ct}}[\rho\to\sigma]} \sup_{\mathcal{U}\in\mathcal{S}^N} \int_\gamma ds \left\langle \mathcal{U}, \frac{d}{ds}\gamma(s)\right\rangle, \\
&\geq \inf_{\gamma\in\Gamma_{\mathrm{ct}}[\rho\to\sigma]} \sup_{b>0} \frac{1}{b} \sup_{\substack{M\geq 0\\ v:\|v\|=1}} \int_\gamma ds \frac{\langle M\circ vv^*, \frac{d}{ds}\gamma(s)\rangle}{\langle M\circ vv^*, \gamma(s)\rangle}, \\
&\geq \inf_{\gamma\in\Gamma_{\mathrm{ct}}[\rho\to\sigma]} \sup_{b>0} \frac{1}{b} \sup_{\substack{M\geq 0\\ v:\|v\|=1}} \big[\ln\langle M\circ vv^*, \gamma(1)\rangle - \ln\langle M\circ vv^*, \gamma(0)\rangle\big], \\
&\geq \sup_{b>0} \frac{1}{b} \sup_{\substack{M\geq 0\\ v:\|v\|=1}} \big[\ln\langle M\circ vv^*, \gamma(1)\rangle - \ln\langle M\circ vv^*, \gamma(0)\rangle\big],
\end{aligned}$$

where the minimization over Γ_{ct} disappears after the integration. Note that conditions are weaker since

$$-\mathcal{I}d \leq M\circ\hat{\Delta}_k^{+1} \leq \mathcal{I}d, \qquad \Longrightarrow \qquad -\mathcal{I}d\circ vv^* \leq M\circ\hat{\Delta}_k\circ vv^* \leq \mathcal{I}d\circ vv^*,$$

where we use Claim 2.1.6 and the fact that $vv^* \geq 0$. □

Chapter 9

Lower bound for Information complexity

Throughout this chapter, we use the notation introduced in Chapters 3 and 5. We add the superscript $\perp$ for a set S, such that $S^{\perp} = S \cup \{\perp\}$, similarly we add the subscript π to entropy H_{π} and information I_{π}, to specify which probability distribution π is used for the entropy or the information.

The communication complexity of a function f could be a complicated quantity to determine. Several lower bound methods have been constructed such as the discrepancy, the partition bound, the efficiency and the external information complexity presented in Section 5.1. The last one is particularly interesting, since it lower bounds the communication complexity, but it can also be interpreted as the amount of information that Alice and Bob reveal about their inputs via the transcript used in the protocol. This quantity is also interesting in scenarios where multiple copies of the problem are solved in parallel since [BR11] have shown that amortized communication complexity is equal to internal information complexity

The external information could also be quite complicated to evaluate, since we must minimize over all distribution π from Formula (5.11). Hence, a first motivation was to provide a simple method to characterize IC_{ext}. In the first part of this chapter, we introduce a new method IC_0, called Zero information complexity, which lower bounds IC_{ext}, and by extension the communication complexity. IC_0 is an optimization problem expressed a minimal form, therefore any feasible point of this optimization problem provides a lower bound.

The new method IC_0 can be applied to the simulation model. In the second part of this chapter, we provide an application of IC_0 to CHSH correlations $\mathbf{p}_{\chi}$, for several reasons.

Definition 9.0.1. (CHSH correlations)
For $\chi \in [0,1]$, $a, b \in \{-1,+1\}$ and $x, y \in \{0,1\}$. We define

$$\forall a, b, x, y, \qquad p_{\chi}(a, b|x, y) = \frac{1 + \chi ab(-1)^{xy}}{4}.$$

For $\chi = 0$, $\mathbf{p}_{\chi}$ is uniform, and for $\chi = 1$, Alice and Bob evaluate the AND function,

$$x \wedge y = a \oplus b.$$

Therefore, χ represents the noise. For $\chi \leq 0.5$, $\mathbf{p}_\chi$ can be simulated using shared randomness but no communication. For χ less than $\sqrt{2}/2$, $\mathbf{p}_\chi$ can also be simulated without communication with the additional help of entanglement. Indeed, $\mathbf{p}_\chi$ for $\chi = \sqrt{2}/2$ corresponds to the quantum correlations obtained in a Bell experiment testing the violation of the CHSH inequality which results from performing projective measurements on a Bell pair. Otherwise, Alice and Bob need to communicate to simulate $\mathbf{p}_\chi$.

They are several motivations in the choice of the CHSH correlations. A first motivation is to provide a better grasp on quantum non locality. Indeed, the violation of a Bell's inequality is just the value of an affine function whereas the communication and information complexities of $\mathbf{p}_\chi$ have a clear operational interpretation: this is the amount of communication that Alice and Bob need to use to simulate quantum non locality using a classical model (in a one-shot or amortized scenario, respectively). As such the information complexity can be interpreted as the amount of information shared between two entangled quantum systems. Another motivation is that the simulation $\mathbf{p}_\chi$ is equivalent to evaluating the AND function on x and y, which can be used as a primitive to compute any other function. A lower bound for the information complexity of this primitive might therefore be used as a starting point to prove lower bounds for other functions.

First, we define the zero information cost IC_0. The efficiency eff is defined from the communication cost CC in the zero communication model, and similarly IC_0 is defined from the external information cost in the same model. Next, we dualize IC_0 to obtain an optimization form under a more useful minimization form. In third Section we simplify IC_0 for the special case where the input distribution is a product distribution. In fourth Section we define $\mathrm{IC}_0^{\rightarrow}$, a particular case of IC_0, in the one-way scenario where Bob cannot abort. In fifth Section we define $\overline{\mathrm{IC}}_0$, a relaxed form of IC_0, where we remove an equality constraint of the optimization program IC_0.

Finally, in last Section we apply our new methods $\mathrm{IC}_0^{\rightarrow}(p_\chi)$ and $\overline{\mathrm{IC}}_0(p_\chi)$ on CHSH correlations. For IC_0, we only provide a numerical analysis.

9.1 Zero information complexity IC_0

In Section 5.4, we have introduced the simulation model, as well as zero communication protocols $\mathcal{P}^\perp$ with private and public coins, for simulating a conditional distribution $\mathbf{p} \in \mathbb{P}$. A zero communication protocol $\mathcal{P}^\perp_{\mathbf{p}}$ that simulates $\mathbf{p}$, satisfies for all $a \in A, b \in B, x \in X, y \in Y$,

$$\sum_{l^\perp_\omega \in \mathcal{L}^\perp_{priv}} p_\Omega(\omega) l^\perp_\omega(a,b|x,y) = \eta p(a,b|x,y), \tag{9.1}$$

where η is the efficiency of the protocol, denoted $\mathrm{eff}(\mathbf{p})$, and p_Ω, a distribution over $\mathcal{L}^\perp_{priv}$ that represents the strategy of Alice and Bob.

Also, for an input distribution μ over $X \times Y$, and public coin Ω, we can induce π_0 the probability distribution over $\Omega \times X \times Y \times A^\perp \times B^\perp$, such that

$$\forall x, y, a, b, \qquad \pi_0(\omega, a, b, x, y) = p_\Omega(\omega).l^\perp_\omega(a,b|x,y).\mu(x,y), \tag{9.2}$$

Therefore from this distribution π_0, we define the **zero information cost** IC_0^μ by

$$\mathrm{IC}_0^\mu(\pi_0) = I_{\pi_0}(X, Y : \Omega | A \neq \perp, B \neq \perp).$$

The definition of the zero information complexity follows naturally.

$$\mathrm{IC}_0^\mu(\mathbf{p}) = \min_{\pi:\,\exists\mathcal{P}_{\mathbf{p}} \text{ which induces } \pi} \mathrm{IC}_0^\mu(\pi),$$
$$\mathrm{IC}_0(\mathbf{p}) = \max_{\mu \text{ a distribution over } X\times Y} \mathrm{IC}_0^\mu(\mathbf{p}).$$

As observed in (9.1), a zero communication protocol $\mathcal{P}_{\mathbf{p}}^\perp$ can be completely characterized by the distribution $p_\Omega(\omega)$, now denoted $\pi_0(\omega)$, and its efficiency η. Using this representation, we can provide a better definition of IC_0 depending of the distribution π_0 and the efficiency η.

Definition 9.1.1. (Zero information complexity)
Let μ be an input distribution, The **zero information complexity** IC_0^μ is an convex optimization program, defined as

$$\mathrm{IC}_0^\mu(\mathbf{p}) = \inf_{\substack{\eta\geq 0\\ \pi_0(\omega)\geq 0}} \quad I_{\pi_0}(X,Y:\Omega|A\neq\perp,B\neq\perp) \qquad \text{subject to,}$$

- $$\sum_{l_\omega\in\mathcal{L}_{priv}^\perp} \pi_0(\omega) l_\omega^\perp(a,b|x,y) = \eta p(a,b|x,y), \qquad \forall a\in A,\, \forall b\in B,\, \forall(x,y)\in\mathrm{supp}(\mu).$$

We introduce the notation η_{xy}^ω, the efficiency of the private randomness distribution l_ω on the input (x,y), such that

$$\eta_{xy}^\omega = \sum_{\substack{a\in A^\perp:\,a\neq\perp\\ b\in B^\perp:\,b\neq\perp}} l_\omega(a,b|x,y), \tag{9.3}$$

as well as

$$\eta^\omega = \left\langle \eta_{xy}^\omega \right\rangle_\mu = \sum_{x,y} \mu(x,y).\eta_{xy}^\omega. \tag{9.4}$$

From Equation (9.1), the average value $\left\langle \eta_{xy}^\omega \right\rangle_{\pi_0(\omega)}$ is equal to η,

$$\left\langle \eta_{xy}^\omega \right\rangle_{\pi_0(\omega)} = \sum_\omega \pi_0(\omega)\eta_{xy}^\omega = \sum_{\substack{a\in A^\perp:\,a\neq\perp\\ b\in B^\perp:\,b\neq\perp}} \sum_\omega \pi_0(\omega) l_\omega(a,b|x,y) = \sum_{\substack{a\in A\\ b\in B}} \eta p(a,b|x,y) = \eta.$$

As indicated in Equation (5.19), a private randomness distribution is written as a product two private randomness distributions, denoted l_ω^A and l_ω^B, so we define η_x^ω and η_y^ω as

$$\eta_{xy}^\omega = \sum_{\substack{a\in A^\perp:\,a\neq\perp\\ b\in B^\perp:\,b\neq\perp}} l_\omega(a,b|x,y) = \sum_{a\in A^\perp:\,a\neq\perp} l_\omega^A(a|x)\;.\sum_{b\in B^\perp:\,b\neq\perp} l_\omega^B(b|y) = \eta_x^\omega.\eta_y^\omega.$$

In the following Proposition, we prove that the zero information complexity $\mathrm{IC}_0^\mu(\mathbf{p})$ is a lower bound of the external information complexity $\mathrm{IC}_{ext}^{\mu,0}(\mathbf{p})$ with error null, therefore by extension of $\mathrm{CC}(\mathbf{p})$.

Proposition 9.1.2. *Let $\boldsymbol{p}$ be a conditional distribution and μ be an input distribution.*
If there exists a deterministic protocol $\mathcal{P}_{\boldsymbol{p}}$ with a communication cost $\mathbf{c}$, then there exists a zero communication protocol $\mathcal{P}_{\boldsymbol{p}}^\perp$ with efficiency 2^{-c}.
Moreover, if there exists a deterministic protocol $\mathcal{P}_{\boldsymbol{p}}$ with an information cost $IC\,_{\mathrm{ext}}^{\mu}(\pi)$, then there exists a zero communication protocol $\mathcal{P}_{\boldsymbol{p}}^\perp$ with a zero information cost $IC\,_0^{\mu}(\pi_0)$.

Corollary 9.1.3. *Let $\boldsymbol{p}$ be a conditional distribution. Then*

$$-\log \mathrm{eff}(\boldsymbol{p}) \leq CC(\boldsymbol{p}) \qquad \textit{and} \qquad IC_0(\boldsymbol{p}) \leq IC^0_{ext}(\boldsymbol{p}).$$

Proof. From Lemma 5.1.1, every deterministic protocol $\mathcal{P}_{\mathbf{p}}$ induces a partition P, such that $|P| \leq 2^c$. We enlarge this partition to a partition P^+, such that $|P^+| = 2^c$, by separating one f-monochromatic rectangle into two f-monochromatic rectangles, as many times as necessary. From this partition we can create a zero communication protocol $\mathcal{P}^{\perp}_{\mathbf{p}}$.

1. With the public coin, Alice and Bob choose uniformly a rectangle R in P^+,

2. If $x \in_1 R$, Alice outputs $f(R)$, otherwise Alice outputs $\perp$,

3. If $y \in_2 R$, Bob outputs $f(R)$, otherwise Bob outputs $\perp$.

In this zero communication protocol, Alice and Bob succeed if they choose the good rectangle. As they choose uniformly, the efficiency is $\eta = 2^{-c}$.
For an input distribution μ, the deterministic protocol $\mathcal{P}_{\mathbf{p}}$ induces

$$\pi(R, x, y) = \mu(x, y).\delta[(x, y) \in R], \qquad \forall (x, y) \in X \times Y, \forall R \in P^+.$$

π is the distribution defined in (5.10), except that a transcript m is characterized by a rectangle R, and Ω absent since the protocol is deterministic. We also have π_0 induces by $\mathcal{P}^{\perp}_{\mathbf{p}}$, such that

$$\pi_0(R, a, b, x, y) = \pi_0(R).\mu(x, y).l^{\perp}_R(a, b|x, y), \qquad \forall (x, y) \in X \times Y, \forall R \in P^+, \forall l^{\perp}_R \in \mathcal{L}^{\perp}_{priv},$$

where R replaces Ω, since we choose the rectangle randomly to the distribution Ω, $\pi_0(R) = \eta$, and

$$l^{\perp}_R(a, b|x, y) = l^{A,\perp}_R(a|x).l^{B,\perp}_R(b|y),$$

with,

$$l^{A,\perp}_R(a'|x) = \begin{cases} 1 & \text{if } x \in_1 R \text{ and } a' = a, \\ 1 & \text{if } x \notin_1 R \text{ and } a' = \perp, \\ 0 & \text{otherwise.} \end{cases} \quad \text{and,} \quad l^{B,\perp}_R(b'|x) = \begin{cases} 1 & \text{if } y \in_2 R \text{ and } b' = b, \\ 1 & \text{if } y \notin_2 R \text{ and } b' = \perp, \\ 0 & \text{otherwise.} \end{cases}$$

where R a is (a, b)-monochromatic rectangle. Moreover, from Equation (9.1), we have

$$\begin{aligned} \pi_0(a \neq \perp, b \neq \perp) &= \sum_{\substack{a \in A \\ b \in B}} \sum_{x,y} \sum_{R \in P^+} \pi_0(R, a, b, x, y), \\ &= \sum_{\substack{a \in A \\ b \in B}} \sum_{x,y} \sum_{R \in P^+} \pi_0(R).\mu(x, y).l^{\perp}_R(a, b|x, y), \\ &= \sum_{\substack{a \in A \\ b \in B}} \sum_{x,y} \mu(x, y).\eta p(a, b|x, y), \\ &= \eta. \end{aligned}$$

We obtain,

$$\pi_0(R, x, y|a \neq \perp, b \neq \perp) = \frac{\pi_0(R, a, b, x, y)}{\pi_0(a \neq \perp, b \neq \perp)}$$

Now, we have the expression of $\pi_0(R,x,y|a\neq\bot,b\neq\bot)$ and $\pi(R,x,y)$, we calculate

$$\begin{aligned}
IC^{\mu}_{\text{ext}}(\pi) &= I_\pi(R:X,Y),\\
&= H_\pi(R)-H_\pi(R|X,Y),\\
&= H_\pi(R).\\
IC^{\mu}_0(\hat{\pi}_0) &= I_{\pi_0}(R:X,Y|A\neq\bot,B\neq\bot),\\
&= H_{\pi_0}(R|A\neq\bot,B\neq\bot)-H_{\pi_0}(R|X,Y,A\neq\bot,B\neq\bot),\\
&= H_{\pi_0}(R|A\neq\bot,B\neq\bot),
\end{aligned}$$

where each second term of the second line is null, since R is determined when the input is known and Alice and Bob don't abort. We conclude that these quantities are equal, since

$$\begin{aligned}
\pi_0(R|a\neq\bot,b\neq\bot) &= \frac{1}{\eta}\pi_0(R,a\neq\bot,b\neq\bot),\\
&= \frac{1}{\eta}\sum_{\substack{a\in A\\ b\in B}}\sum_{x,y}\pi_0(R,a,b,x,y),\\
&= \frac{1}{\eta}\sum_{\substack{a\in A\\ b\in B}}\sum_{x,y}\pi_0(R).\mu(x,y).l^{\bot}_{R}(a,b|x,y),\\
&= \sum_{\substack{a\in A\\ b\in B}}\sum_{x,y}\mu(x,y).l^{\bot}_{R}(a,b|x,y),\\
&= \sum_{(x,y)\in R}\mu(x,y),\\
&= \pi(R).
\end{aligned}$$

□

9.2 Dualization of IC_0

This section is completely dedicated to dualize the minimization form of Definition 9.1.1, to obtain

Theorem 9.2.1. *(Zero information complexity)*

$$\mathrm{IC}^{\mu}_0(\mathbf{p}) = \sup_{B_{abxy}}\sum_{abxy}B_{abxy}\mu(x,y)\mathrm{p}(a,b|x,y) \qquad \textit{subject to,}$$

- $\forall l\in\mathcal{L}_{det}, \qquad \sum_{abxy}\nu(x,y)B_{abxy}l(a,b|x,y)\leq D(\nu||\mu),$
- $\forall x\in X,\forall y\in Y,\forall\eta_x,\eta_y\in[0,1], \quad \nu(x,y)=\frac{\eta_x\eta_y}{\sum_{x',y'}\eta'_x\eta'_y\mu(x',y')}\mu(x,y).$

Proof. Starting from Definition 9.1.1, we define $\lambda_\omega=\frac{\hat{\pi}_0}{\eta}$, the first constraint then becomes

$$\sum_{l_\omega\in\mathcal{L}^{\bot}_{priv}}\lambda_\omega l^{\bot}_{\omega}(a,b|x,y)=p(a,b|x,y).$$

From the second constraint, the joint distribution probability π_0 is defined as

$$\pi_0(\omega, a, b, x, y) = \hat{\pi}_0(\omega).l_\omega^\perp(a, b|x, y).\mu(x, y).$$

From this distribution we can obtain,

$$\begin{aligned}
\pi_0(a \neq \perp, b \neq \perp, x, y) &= \sum_\omega \sum_{\substack{a \in A \\ b \in B}} \pi_0(\omega, a, b, x, y) = \sum_\omega \hat{\pi}_0(\omega).\eta_{xy}^\omega.\mu(x, y) = \eta.\mu(x, y), \\
\pi_0(a \neq \perp, b \neq \perp) &= \sum_{x,y} \eta.\mu(x, y) = \eta, \\
\pi_0(x, y|a \neq \perp, b \neq \perp) &= \frac{\pi_0(a \neq \perp, b \neq \perp, x, y)}{\pi_0(a \neq \perp, b \neq \perp)} = \mu(x, y), \\
\pi_0(\omega, x, y|a \neq \perp, b \neq \perp) &= \frac{\pi_0(\omega, a \neq \perp, b \neq \perp, x, y)}{\pi_0(a \neq \perp, b \neq \perp)} = \frac{\hat{\pi}_0(\omega).\eta_{xy}^\omega.\mu(x, y)}{\eta} = \lambda_\omega.\eta_{xy}^\omega.\mu(x, y), \\
\pi_0(\omega|a \neq \perp, b \neq \perp) &= \sum_{x,y} \lambda_\omega.\eta_{xy}^\omega.\mu(x, y) = \lambda_\omega.\eta^\omega,
\end{aligned}$$

We use the last three equations to calculate $I_{\pi_0}(X, Y : \Omega|A \neq \perp, B \neq \perp)$,

$$\begin{aligned}
&I_{\pi_0}(X, Y : \Omega|A \neq \perp, B \neq \perp) \\
&\qquad = \sum_{x,y,\omega} \pi_0(\omega, x, y|a \neq \perp, b \neq \perp) \log \frac{\pi_0(\omega, x, y|a \neq \perp, b \neq \perp)}{\pi_0(\omega|a \neq \perp, b \neq \perp).\pi_0(x, y|a \neq \perp, b \neq \perp)}, \\
&\qquad = \sum_{x,y,\omega} \lambda_\omega.\eta_{xy}^\omega.\mu(x, y) \log \frac{\eta_{xy}^\omega}{\eta^\omega}.
\end{aligned}$$

Then, the initial Definition 9.1.1 of IC_0^μ becomes

$$\begin{aligned}
\mathrm{IC}_0^\mu(\mathbf{p}) = \inf_{\lambda_\omega \geq 0} \sum_{x,y,\omega} \lambda_\omega.\eta_{xy}^\omega.\mu(x, y) \log \frac{\eta_{xy}^\omega}{\eta^\omega} \qquad & \text{subject to,} \\
\bullet \sum_\omega \lambda_\omega l_\omega^\perp(a, b|x, y) = p(a, b|x, y), \qquad & \forall a \in A,\ \forall b \in B,\ \forall (x, y) \in \mathrm{supp}(\mu), \forall l_\omega^\perp \in \mathcal{L}_{priv}^\perp.
\end{aligned}$$

Note that the zero information method is now a linear program. Hence, let's check the strong duality with the Slater's condition (Definition 6.7.7) by looking for a strictly feasible. Since there is no inequality constraints, we must only find a strictly feasible $(\hat{\lambda}_\omega)_\omega$ such that $\hat{\lambda}_\omega < \lambda_\omega$ for all

ω where $\lambda_\omega \neq 0$. Which is trivial.

$$\begin{aligned}
\mathrm{IC}_0^\mu(\mathbf{p}) = \inf_{\lambda_\omega \geq 0} \sup_{B_{abxy}} & \sum_{x,y,\omega} \lambda_\omega.\eta^\omega_{xy}.\mu(x,y) \log \frac{\eta^\omega_{xy}}{\eta^\omega} \\
& + \sum_{a,b,x,y} \mu(x,y).B_{abxy}\Big[p(a,b|x,y) - \sum_\omega \lambda_\omega l^\perp_\omega(a,b|x,y)\Big], \\
\mathrm{IC}_0^\mu(\mathbf{p}) = \sup_{B_{abxy}} \inf_{\lambda_\omega \geq 0} & \sum_{a,b,x,y} \mu(x,y) B_{abxy} p(a,b|x,y) \\
& + \sum_{x,y,\omega} \lambda_\omega.\mu(x,y)\Big[\eta^\omega_{xy} \log \frac{\eta^\omega_{xy}}{\eta^\omega} - \sum_\omega B_{abxy}.l^\perp_\omega(a,b|x,y)\Big], \\
\mathrm{IC}_0^\mu(\mathbf{p}) = \sup_{B_{abxy}} & \sum_{a,b,x,y} \mu(x,y) B_{abxy} p(a,b|x,y) \qquad \text{subject to,}
\end{aligned}$$

- $\displaystyle\sum_{x,y} \mu(x,y).B_{abxy}.l^\perp_\omega(a,b|x,y) \leq \sum_{x,y} \mu(x,y).\eta^\omega_{xy} \log \frac{\eta^\omega_{xy}}{\eta^\omega}, \qquad \forall a \in A,\ \forall b \in B,\ \forall l^\perp_\omega \in \mathcal{L}^\perp_{priv}.$

To make equations clearer, we replace η^ω_{xy} by $\eta^l xy$ since there is an equivalence between the variable ω and a conditional distribution l .
Note that only the condition of the dual form of IC_0^μ must be satisfied for all private coin randomized protocols $l^\perp \in \mathcal{L}^\perp_{priv}$ without abort. Therefore, we can restrict each protocol $l^\perp \in \mathcal{L}^\perp_{priv}$ to $l^* \in \mathcal{L}_{priv}$ with,

$$l^*(a,b|x,y) = \begin{cases} \frac{1}{\eta^l_{xy}} l^\perp(a,b|x,y) & \text{if } \eta^l_{xy} \neq 0, \\ \frac{1}{|A||B|} & \text{otherwise.} \end{cases}$$

Then, the only condition of the dual form of IC_0^μ becomes

$$\begin{aligned}
\sum_{x,y} \mu(x,y).B_{abxy}.l^\perp(a,b|x,y) &\leq \sum_{x,y} \mu(x,y).\eta^l_{xy} \log \frac{\eta^l_{xy}}{\eta^l} & \forall l^\perp \in \mathcal{L}^\perp_{priv}, \\
\sum_{x,y} \mu(x,y).B_{abxy}.\eta^l_{xy}.l^*(a,b|x,y) &\leq \sum_{x,y} \mu(x,y).\eta^l_{xy} \log \frac{\eta^l_{xy}}{\eta^l} & \forall l^\perp \in \mathcal{L}^\perp_{priv}, \\
\sum_{x,y} \mu_l(x,y).B_{abxy}.l^*(a,b|x,y) &\leq \sum_{x,y} \mu_l(x,y). \log \frac{\mu_l(x,y)}{\mu(x,y)} & \forall l^\perp \in \mathcal{L}^\perp_{priv}, \\
\sum_{x,y} \mu_l(x,y).B_{abxy}.l^*(a,b|x,y) &\leq D(\mu_l||\mu) & \forall l^\perp \in \mathcal{L}^\perp_{priv},
\end{aligned}$$

where we define $\mu_l(x,y) = \mu(x,y)\frac{\eta^l_x \eta^l_y}{\eta^l}$. Note that from Equation (9.4), μ_l is still a probability distribution over $X \times Y$.
As every protocol with abort $l^\perp \in \mathcal{L}^\perp_{priv}$, can be constructed from a protocol $l \in \mathcal{L}_{priv}$ without abort and coefficients $(\eta^l_x)_x$ for Alice and $(\eta^l_y)_y$ for Bob. Then we decompose a protocol in $l \in \mathcal{L}^\perp_{priv}$ in a protocol $l \in \mathcal{L}_{priv}$ and coefficients η^l_x's and η^l_y's, both independent. Then

$$\begin{aligned}
\sum_{x,y} \nu(x,y).B_{abxy}.l(a,b|x,y) &\leq D(\nu||\mu) & \forall l \in \mathcal{L}_{priv}, \\
\frac{\eta_x \eta_y}{\sum_{x',y'} \eta_{x'} \eta_{y'} \mu(x',y')} \mu(x,y) &= \nu(x,y) & \forall x \in X, \forall y \in Y, \forall \eta_x, \eta_y \in [0,1].
\end{aligned}$$

Finally, as the left term of the constraint is linear in $l \in \mathcal{L}_{priv}$, a private randomness distribution, then this is sufficient to only consider deterministic distributions, since $\mathcal{L}_{det} \subset \mathcal{L}_{priv} \subset \textbf{conv } \mathcal{L}_{det}$. □

9.3 IC_0 with a product input distribution

Let μ be a input product distribution with $\mu(x,y) = \mu_x \mu_y$. Then

$$\begin{aligned}
\nu(x,y) &= \mu(x,y).\frac{\eta_{xy}}{\sum_{x',y'} \eta_{x'y'}\mu'_x\mu'_y},\\
&= \mu_x\mu_y\frac{\eta_x\eta_y}{\sum_{x',y'} \mu_{x'}\mu_{y'}\eta_{x'}\eta_{y'}},\\
&= \frac{\mu_x\eta_x}{\sum_{x'} \mu_{x'}\eta_{x'}}\frac{\mu_y\eta_y}{\sum_{y'} \mu_{y'}\eta_{y'}},\\
&= \nu_x\nu_y.
\end{aligned}$$

Hence, in the dual form of the zero information complexity 9.2.1, the last condition can be separated into two conditions.

Corollary 9.3.1. *(Zero information complexity for product input distribution)*

$$\mathrm{IC}_0^\mu(\boldsymbol{p}) = \sup_{B_{abxy}} \sum_{a,b,x,y} B_{abxy}\mu_x\mu_y \mathrm{p}(a,b|x,y) \qquad \textit{subject to,}$$

- $\sum_{a,b,x,y} \nu_x\nu_y B_{abxy} l(a,b|x,y) \leq D(\nu_x||\mu_x) + D(\nu_y||\mu_y), \qquad \forall l \in \mathcal{L}_{det},$
- $\nu_x = \frac{\mu_x\eta_x}{\sum_{x'} \mu_{x'}\eta_{x'}}, \qquad \forall \eta_x \in [0,1], \forall x \in X,$
- $\nu_y = \frac{\mu_x\eta_y}{\sum_{y'} \mu_{y'}\eta_{y'}}, \qquad \forall \eta_y \in [0,1], \forall y \in Y.$

We can observe in the above corollary that ν_x and ν_y are not restricted anymore. I.e. for all distributions $\mu_x \in \mathbb{P}\big(\mathrm{supp}(\mu_x)\big)$, there exits $(\eta_x)_x$ that satisfies the condition in the corollary. Same for μ_y. Hence, we can simplify again the zero information complexity.

Theorem 9.3.2. *(Zero information complexity for product input distribution)*

$$\mathrm{IC}_0^\mu(\boldsymbol{p}) = \sup_{B_{abxy}, B_0} \sum_{a,b,x,y} B_{abxy}\mu_x\mu_y \mathrm{p}(a,b|x,y) - B_0 \qquad \textit{subject to,}$$

- $\inf_{\substack{\nu_x \in \mathbb{P}(X_\mu))\\ \nu_y \in \mathbb{P}(Y_\mu)}} \Big[D(\nu_x||\mu_x) + D(\nu_y||\mu_y) - \sum_{a,b,x,y} \nu_x\nu_y B_{abxy} l(a,b|x,y)\Big] \geq -B_0, \qquad \forall l \in \mathcal{L}_{det}$

where $X_\mu = \mathrm{supp}(\mu_x)$ *and* $Y_\mu = \mathrm{supp}(\mu_y)$.

Proof. We simply change the variable B_{abxy} to $B_{abxy} - B_0$. After we substitute η_x's to ν_x, which is possible only if μ_x is not null. Similarly for μ_y. □

Now, we introduce a lemma which we will be useful to simplify optimization program.

Lemma 9.3.3. *[SSL17]*
Let q be a probability distribution in $\mathbb{P}(S)$, f be a function in $\mathbb{B}(S)$ and $\langle p, f\rangle$ be the expectation of f under the distribution p. Then, the optimization problem,

$$\min_{p\in\mathbb{P}(S)} D_{KL}(p||q) - \langle p, f\rangle,$$

has for optimal value, $-\log\langle q, 2^f\rangle$, achieved by the unique optimal probability distribution

$$\hat{p}(s) = \frac{q(s)2^{f(s)}}{\langle q, 2^f\rangle}.$$

The next proof was not the first, but it is original.

Proof. This proof uses tools of convex optimization. Since $D_{KL}(p||q)$ is convex in p, the Slater's condition holds.

$$\begin{aligned}
T(q,f) &= \min_{\forall s,\, p(s)\geq 0} D_{KL}(p||q) - \langle p, f\rangle, \qquad \text{such that} \quad \sum_{s\in S} p(s) = 1,\\
T(q,f) &= \min_{\forall s,\, p(s)\geq 0} \max_{\lambda\in\mathbb{R}} \sum_s p(s)\log\frac{p(s)}{q(s)} - \sum_{s\in S} p(s)f(s) + \lambda\big(1 - \sum_{s\in S} p(s)\big),\\
T(q,f) &= \max_{\lambda\in\mathbb{R}} \left\{\lambda + \sum_{s\in S} \min_{p(s)\geq 0} \Big(p(s)\big[\log\frac{p(s)}{q(s)} - f(s) - \lambda\big]\Big)\right\},
\end{aligned}$$

as the term between parentheses is strictly convex on $p(s)$, we can derive optimal points $p^\star(s) = e^{-1}q(s)2^{f(s)+\lambda}$, with e the base of natural logarithm. Thus we obtain,

$$T(q,f) = \max_{\lambda\in\mathbb{R}} \left\{\lambda - \frac{2^\lambda}{e\ln 2}\sum_{s\in S} q(s)2^{f(s)}\right\}.$$

Since the function between accolades is concave on λ, we can derive the optimal point $\lambda^\star = \log\frac{e}{\langle q, 2^f\rangle}$. Inserting $\lambda^\star$ in $T(q,f)$ and $p^\star(s)$, we obtain

$$T(q,f) = -\log\langle q, 2^f\rangle \qquad \text{and} \qquad p^\star(s) = \frac{q(s)2^{f(s)}}{\langle q, 2^f\rangle}.$$

□

Moreover, the unique condition in Theorem 9.3.2 can be expressed under the minimization of ν_x or ν_y.

Lemma 9.3.4. *The optimal value of the minimization problem*

$$\begin{aligned}
&\inf_{\substack{\nu_x\geq 0\\ \nu_y\geq 0}} D(\nu_x||\mu_x) + D(\nu_y||\mu_y) - \sum_{a,b,x,y} \nu_x\nu_y B_{abxy} l(a,b|x,y)\\
&\qquad = \inf_{\nu_y\geq 0} D(\nu_y||\mu_y) - \log\left[\sum_x \mu_x 2^{\sum_{a,b,y}\nu_y B_{abxy} l(a,b|x,y)}\right],\\
&\qquad = \inf_{\nu_x\geq 0} D(\nu_x||\mu_x) - \log\left[\sum_y \mu_y 2^{\sum_{a,b,x}\nu_x B_{abxy} l(a,b|x,y)}\right],
\end{aligned}$$

where optimal points are respectively,

$$\nu_x^\star(\nu_y) = \frac{\mu_x 2^{\sum_{a,b,y} \nu_y B_{abxy} l(a,b|x,y)}}{\sum_x \mu_x 2^{\sum_{a,b,y} \nu_y B_{abxy} l(a,b|x,y)}},$$
$$\nu_y^\star(\nu_x) = \frac{\mu_y 2^{\sum_{a,b,x} \nu_x B_{abxy} l(a,b|x,y)}}{\sum_y \mu_y 2^{\sum_{a,b,x} \nu_x B_{abxy} l(a,b|x,y)}}.$$

Proof. The proof is a direct consequence of Lemma 9.3.3. □

9.4 Relaxed information cost $\overline{\mathrm{IC}}_0$

In this Section, we provide $\overline{\mathrm{IC}}_0^\mu$ a relaxed version of IC_0^μ, such that for all input distributions μ,

$$\overline{\mathrm{IC}}_0^\mu \leq \mathrm{IC}_0^\mu.$$

Theorem 9.4.1. *(Relaxed zero information complexity)*

$$\overline{\mathrm{IC}}_0^\mu(\boldsymbol{p}) = \sup_{\substack{B_{abxy}\\ B_0}} \sum_{a,b,x,y} B_{abxy}\mu(x,y)p(a,b|x,y) - B_0 \qquad \textit{subject to,}$$
$$\sum_{x,y} \mu(x,y) 2^{\sum_{a,b} B_{abxy} l(a,b|x,y)} \leq 2^{B_0}, \qquad \forall l \in \mathcal{L}_{det}.$$

This bound is derived from Claim 9.4.2, that proves an information cost equivalence between a zero communication protocol with efficiency η and a zero communication protocol with full efficiency ($\eta = 1$) with restrictions.

Claim 9.4.2. Let $\mathbf{p}$ be a conditional probability distribution and μ an input distribution. There exists a zero communication protocol for $\mathbf{p}$ with efficiency η and distribution π_0 with $\mathrm{IC}_0^\mu(\pi_0) = i$, if and only if, there exists a zero communication protocol for $\mathbf{p}$ with full efficiency ($\eta = 1$), a joint distribution $\pi(\omega, x, y)$ with $\mathrm{IC}^\mu(\pi) = i$, and functions $u : X \times \Omega \to \mathbb{R}_+$, $v : Y \times \Omega \to \mathbb{R}_+$ such that,

$$\begin{aligned} \forall a \in A,\, \forall b \in B,\, \forall (x,y) \in \mathrm{supp}(\mu), &\qquad \sum_{l_\omega \in \mathcal{L}_{priv}} \pi(\omega|x,y) l_\omega(a,b|x,y) = \mathrm{p}(a,b|x,y),\\ \forall \omega \in \Omega,\, \forall (x,y) \in \mathrm{supp}(\mu), &\qquad \pi(\omega|x,y) = u(x,\omega).v(y,\omega),\\ &\qquad \mathrm{IC}^\mu(\pi) = i. \end{aligned}$$

With this equivalence in term of information complexity, we rewrite Definition 9.1.1.

Corollary 9.4.3. *(Zero information complexity)*

$$\mathrm{IC}_0^\mu(\boldsymbol{p}) = \inf_{\substack{\pi(\omega,x,y)\geq 0\\ u(x,\omega)\geq 0\\ v(y,\omega)\geq 0}} I_\pi(X,Y:\Omega) \qquad \textit{subject to,}$$

- $\displaystyle\sum_{l_\omega \in \mathcal{L}_{priv}} \pi(\omega,x,y) l_\omega(a,b|x,y) = \mu(x,y)p(a,b|x,y),$ $\quad \forall \omega \in \Omega, \forall x \in X, \forall y \in Y,$ $\forall a \in A,\, \forall b \in B,$
- $\pi(\omega|x,y) = u(x,\omega).v(y,\omega),$ $\quad \forall \omega \in \Omega, \forall (x,y) \in \mathrm{supp}(\mu),$
- $\pi(\omega,x,y) = \pi(\omega|x,y).\mu(x,y),$ $\quad \forall \omega \in \Omega, \forall x \in X, \forall y \in Y.$

Proof of Claim 9.4.2. $\boxed{\Rightarrow}$
For a zero communication protocol with efficiency η and distribution π_0 with $\mathrm{IC}_0^{\mu}(\pi_0) = i$, such that

$$\sum_{l_\omega^\perp \in \mathcal{L}_{priv}^\perp} p_\Omega(\omega) l_\omega^\perp(a,b|x,y) = \eta p(a,b|x,y).$$

We define the joint distribution $\pi(\omega,x,y)$ as

$$\begin{aligned} \pi(\omega,x,y) &= \pi(\omega|x,y)\mu(x,y), \\ \pi(\omega|x,y) &= \pi_0(\omega|x,y,a \neq \perp, b \neq \perp). \end{aligned}$$

This definition directly implies $\mathrm{IC}_0(\pi_0)$ =IC(π). Moreover, we have

$$\begin{aligned} \pi(\omega|x,y) &= \pi_0(\omega|x,y,a \neq \perp, b \neq \perp), \\ &= \frac{1}{\eta} \sum_{\substack{a \in A \\ b \in B}} \pi_0(\omega,a,b|x,y), \\ &= \frac{1}{\eta} \sum_{\substack{a \in A \\ b \in B}} \pi_0(\omega) l_\omega^\perp(a,b|x,y), \\ &= \frac{1}{\eta} \pi_0(\omega) \eta_{xy}^\omega, \\ &= \frac{1}{\eta} \pi_0(\omega) \eta_x^\omega . \eta_y^\omega, \end{aligned}$$

then we can choose $u(x,\omega) = \eta_x^\omega \sqrt{\frac{\pi_0(\omega)}{\eta}}$ and $v(y,\omega) = \eta_y^\omega \sqrt{\frac{\pi_0(\omega)}{\eta}}$.
Furthermore, for each $l_\omega^\perp \in \mathcal{L}_{priv}^\perp$ we define its not aborting version $l_\omega \in \mathcal{L}_{priv}$, as

$$l_\omega(a|x) = \begin{cases} \frac{1}{\eta_x^\omega} l_\omega^{A,\perp}(a|x) & \text{if } \eta_x^\omega \neq 0, \\ \frac{1}{|A|} & \text{otherwise}, \end{cases} \quad \text{and} \quad l_\omega(b|y) = \begin{cases} \frac{1}{\eta_y^\omega} l_\omega^{B,\perp}(b|y) & \text{if } \eta_y^\omega \neq 0, \\ \frac{1}{|B|} & \text{otherwise}, \end{cases}$$

where $l_\omega^\perp = l_\omega^{A,\perp} . l_\omega^{B,\perp}$ and $l_\omega = l_\omega^A . l_\omega^B$. Finally, we obtain

$$\begin{aligned} \sum_{l_\omega \in \mathcal{L}_{priv}} \pi(\omega|x,y) l_\omega(a,b|x,y) &= \frac{1}{\eta} \sum_{l_\omega \in \mathcal{L}_{priv}} \pi_0(\omega) . \eta_x^\omega . \eta_y^\omega . l_\omega^A(a|x) . l_\omega^B(b|y), \\ &= \frac{1}{\eta} \sum_{l_\omega \in \mathcal{L}_{priv}^\perp} \pi_0(\omega) . l_\omega^{A,\perp}(a|x) . l_\omega^{B,\perp}(b|y), \\ &= \mathrm{p}(a,b|x,y). \end{aligned}$$

$\boxed{\Leftarrow}$
For a zero communication protocol with full efficiency, a joint distribution $\pi(\omega,x,y)$ with $\mathrm{IC}^\mu(\pi) = i$ and $l_\omega \in \mathcal{L}_{priv}$, and functions $u : X \times \Omega \to \mathbb{R}_+$, $v : Y \times \Omega \to \mathbb{R}_+$.
We set $\pi_0(\omega)$ to be uniform distribution over Ω, $\pi_0(\omega) = \frac{1}{|\Omega|}$, and private randomness protocol with abort $l_\omega^\perp \in \mathcal{L}_{priv}^\perp$ such that $l_\omega^\perp = l_\omega^{A,\perp} . l_\omega^{B,\perp}$,

$$l_\omega^{A,\perp}(a|x) = \begin{cases} \frac{u(x,\omega)}{U} l_\omega^A(a|x) & \text{if } a \neq \perp, \\ 1 - \frac{u(x,\omega)}{U} & \text{if } a = \perp, \end{cases} \quad \text{and} \quad l_\omega^{B,\perp}(b|y) = \begin{cases} \frac{v(y,\omega)}{V} l_\omega^B(b|y) & \text{if } b \neq \perp, \\ 1 - \frac{v(y,\omega)}{V} & \text{if } b = \perp, \end{cases}$$

where $U = \max_{(x,\omega)} u(x,\omega)$ and $V = \max_{(y,\omega)} v(y,\omega)$. Therefore we obtain,

$$\begin{aligned}
\sum_{l_\omega^\perp \in \mathcal{L}_{priv}^\perp} \pi_0(\omega) l_\omega^\perp(a,b|x,y) &= \frac{1}{|\Omega|} \sum_{l_\omega \in \mathcal{L}_{priv}} \frac{u(x,\omega)}{U} \frac{v(y,\omega)}{V} l_\omega^A(a|x).l_\omega^B(b|y), \\
&= \frac{1}{|\Omega| UV} \sum_{l_\omega \in \mathcal{L}_{priv}} \pi(\omega|x,y) l_\omega(a,b,x,y), \\
&= \frac{1}{|\Omega| UV} \mathrm{p}(a,b|x,y),
\end{aligned}$$

where the efficiency is $\eta = \frac{1}{|\Omega|UV}$. To conclude, we show that $\mathrm{IC}_0(\pi_0)$ $=\mathrm{IC}(\pi)$,

$$\begin{aligned}
\pi_0(\omega|x,y,a \neq \perp, b \neq \perp) &= \frac{1}{\eta} \sum_{\substack{a \in A \\ b \in B}} \pi_0(\omega,a,b|x,y) \\
&= \frac{1}{\eta} \sum_{\substack{a \in A \\ b \in B}} \pi_0(\omega) l_\omega^\perp(\omega,a,b|x,y), \\
&= \frac{1}{\eta|\Omega|} \sum_{\substack{a \in A \\ b \in B}} \frac{u(x,\omega)}{U} l_\omega^A(a|x). \frac{v(y,\omega)}{V} l_\omega^B(b|y), \\
&= \frac{1}{\eta|\Omega| UV} \sum_{\substack{a \in A \\ b \in B}} \pi(\omega|x,y) l_\omega(a,b|x,y), \\
&= \pi(\omega|x,y).
\end{aligned}$$

□

From this version of IC_0, we give a relaxed version $\overline{\mathrm{IC}}_0$ by removing the second constraint in 9.4.3, that is

$$\pi(\omega,x,y) = u(x,\omega)v(y,\omega).\mu(x,y) \qquad \forall x,y,\omega.$$

Definition 9.4.4. (Relaxed zero information complexity)

$$\overline{\mathrm{IC}}_0^\mu(\mathbf{p}) = \inf_{\pi(\omega,x,y)\geq 0} \quad I_\pi(X,Y:\Omega) \qquad\qquad \text{subject to,}$$

- $\displaystyle\sum_{l_\omega \in \mathcal{L}_{priv}} \pi(\omega,x,y) l_\omega(a,b|x,y) = \mu(x,y) p(a,b|x,y),$ $\quad \forall \omega \in \Omega, \forall x \in X, \forall y \in Y,$ $\forall a \in A,\ \forall b \in B,$
- $\pi(\omega,x,y) = \pi(\omega|x,y).\mu(x,y),$ $\quad \forall \omega \in \Omega, \forall (x,y) \in \mathrm{supp}(\mu).$

Obviously, we trivially have the relation $\overline{\mathrm{IC}}_0^\mu \leq \mathrm{IC}_0^\mu$ for all input distributions. Now, we reformulate $\overline{\mathrm{IC}}_0^\mu$ in a more practical form.

Proof of Theorem 9.4.1. First, we use the notation $\pi_\omega(\omega) = \pi(\omega)$ and $\pi_\omega^{xy} = \pi(\omega|x,y)$ to rewrite

$I_\pi(X,Y:\Omega)$ as,

$$\begin{aligned} I_\pi(X,Y:\Omega) &= \sum_{x,y,\omega} \pi(\omega,x,y)\log\frac{\pi(\omega,x,y)}{\pi(\omega)\mu(x,y)}, \\ &= \sum_{x,y,\omega} \mu(x,y)\pi(\omega|x,y)\log\frac{\pi(\omega|x,y)}{\pi(\omega)}, \\ &= \sum_{x,y} \mu(x,y) D(\pi^{xy}_\omega|\pi_\omega). \end{aligned}$$

Now, we make $\pi(\omega|x,y)$ and $\pi(\omega)$ independent by adding a constraint, then the linear program becomes,

$$\overline{\mathrm{IC}}^{\mu}_0(\mathbf{p}) = \inf_{\substack{\pi^{xy}_\omega(\omega)\geq 0\\ \pi_\omega(\omega)\geq 0}} \sum_{x,y} \mu(x,y) D(\pi^{xy}_\omega|\pi_\omega) \qquad \text{subject to,}$$

- $\displaystyle\sum_{l_\omega\in\mathcal{L}_{priv}} \mu(x,y)\pi^{xy}_\omega(\omega) l_\omega(a,b|x,y) = \mu(x,y)p(a,b|x,y),$ $\quad\forall\omega\in\Omega, \forall x\in X, \forall y\in Y, \forall a\in A, \forall b\in B,$
- $\displaystyle\pi_\omega(\omega) = \sum_{x,y}\pi^{xy}_\omega(\omega).\mu(x,y),$ $\quad\forall\omega\in\Omega, \forall x\in X, \forall y\in Y.$

Now, we dualize,

$$\begin{aligned} \overline{\mathrm{IC}}^{\mu}_0(\mathbf{p}) = &\inf_{\substack{\pi^{xy}_\omega(\omega)\geq 0\\ \pi_\omega(\omega)\geq 0}} \sup_{\substack{B_{abxy}\\ \lambda_\omega}} \sum_{x,y}\mu(x,y)D(\pi^{xy}_\omega|\pi_\omega) \\ &+ \sum_{a,b,x,y} B_{abxy}\mu(x,y)\left[\, p(a,b|x,y) - \sum_{l_\omega\in\mathcal{L}_{priv}} \pi^{xy}_\omega(\omega) l_\omega(a,b|x,y)\right] - \sum_\omega \lambda_\omega\left[\sum_{x,y}\pi^{xy}_\omega(\omega).\mu(x,y) - \pi_\omega(\omega)\right], \\ \overline{\mathrm{IC}}^{\mu}_0(\mathbf{p}) = &\inf_{\pi_\omega(\omega)\geq 0} \sup_{\substack{B_{abxy}\\ \lambda_\omega}} \inf_{\pi^{xy}_\omega(\omega)\geq 0} \sum_{x,y,\omega} \mu(x,y)\pi^{xy}_\omega(\omega)\left[\log\frac{\pi^{xy}_\omega}{\pi_\omega} - \sum_{a,b} B_{abxy} l_\omega(a,b|x,y) - \sum_\omega \lambda_\omega\right] \\ &+ \sum_{a,b,x,y} B_{abxy}\mu(x,y)\ p(a,b|x,y) + \sum_\omega \lambda_\omega \pi_\omega(\omega). \end{aligned}$$

The swap between inf and sup is allowed, because the Lagrangian is convex in B_{abxy}, λ_ω and π^{xy}_ω, then there is a strong duality from Slater Theorem 6.7.8. Using Lemma 9.3.3 on $\mu(x,y)\pi^{xy}_\omega$, we obtain

$$\begin{aligned} \overline{\mathrm{IC}}^{\mu}_0(\mathbf{p}) = &\inf_{\pi_\omega(\omega)\geq 0} \sup_{\substack{B_{abxy}\\ \lambda_\omega}} -\log\left[\sum_{\omega,x,y}\mu(x,y)\pi_\omega(\omega)2^{\lambda_\omega + \sum_{a,b} B_{abxy} l_\omega(a,b|x,y)}\right] \\ &+ \sum_{a,b,x,y} B_{abxy}\mu(x,y)\ p(a,b|x,y) + \sum_\omega \lambda_\omega\pi_\omega(\omega), \\ \overline{\mathrm{IC}}^{\mu}_0(\mathbf{p}) = &\sup_{B_{abxy}} \sum_{a,b,x,y} B_{abxy}\mu(x,y)\ p(a,b|x,y) \\ &+ \inf_{\pi_\omega(\omega)\geq 0}\sup_{\lambda_\omega}\left[\sum_\omega \lambda_\omega\pi_\omega(\omega)\ - \log\sum_{\omega,x,y}\mu(x,y)\pi_\omega(\omega)2^{\lambda_\omega + \sum_{a,b} B_{abxy} l_\omega(a,b|x,y)}\right], \end{aligned}$$

Here again, the swap between inf and sup is allowed, because the Lagrangian is concave in B_{abxy} and π_ω, then there is a strong duality from Slater Theorem 6.7.8. To maximization the last term over λ_ω, we introduce the notation

$$D(\omega) = \sum_{x,y} \mu(x,y) 2^{\sum_{a,b} B_{abxy} l_\omega(a,b|x,y)} \qquad \text{and} \qquad C(\boldsymbol{\lambda}) = \sum_{\omega} \pi_\omega(\omega) D(\omega) 2^{\lambda_\omega},$$

where $\boldsymbol{\lambda}$ is a vector over Ω. Thus, the maximization becomes,

$$\sup_{\lambda_\omega} \ f(\pi_\omega) = \sup_{\lambda_\omega} \ \Big[\langle \boldsymbol{\lambda}, \pi_\omega \rangle - \log C(\boldsymbol{\lambda}) \Big].$$

Note that the objective function $f(\pi_\omega)$ is strictly concave, since

$$\begin{aligned} \frac{df(\pi_\omega)}{d\lambda_\omega} &= \pi_\omega - \frac{\pi_\omega(\omega) D(\omega) 2^{\lambda_\omega}}{C(\boldsymbol{\lambda})}, \\ \frac{d^2 f(\pi_\omega)}{d\lambda_\omega^2} &= -\ln 2 \frac{\pi_\omega(\omega) D(\omega) 2^{\lambda_\omega}}{C(\boldsymbol{\lambda})^2} \big[C(\boldsymbol{\lambda}) - \pi_\omega(\omega) D(\omega) 2^{\lambda_\omega} \big], \end{aligned}$$

where all terms $\pi_\omega(\omega) D(\omega) 2^{\lambda_\omega}$ are positive and from the definition of $C(\boldsymbol{\lambda})$. As, $f(\pi_\omega)$ is strictly concave, its maximum value is,

$$\sup_{\lambda_\omega} \ f(\pi_\omega) = - \sum_{\omega} \pi_\omega(\omega) \log D(\omega),$$

with maximum attained for $\lambda^\star_\omega = \log \frac{C(\boldsymbol{\lambda})}{D(\omega)}$.
So, our optimization problem becomes

$$\begin{aligned} \overline{\mathrm{IC}}_0^\mu(\mathbf{p}) = &\sup_{B_{abxy}} \sum_{a,b,x,y} B_{abxy} \mu(x,y) \ p(a,b|x,y) \\ &+ \inf_{\pi_\omega(\omega) \geq 0} \Big[-\log \sum_{\omega,x,y} \mu(x,y) \pi_\omega(\omega) 2^{\sum_{a,b} B_{abxy} l_\omega(a,b|x,y)} \Big], \\ \overline{\mathrm{IC}}_0^\mu(\mathbf{p}) = &\sup_{\substack{B_{abxy} \\ B_0}} \sum_{a,b,x,y} B_{abxy} \mu(x,y) \ p(a,b|x,y) - B_0 \qquad \text{subject to,} \\ &\bullet \quad -B_0 \leq \inf_{\pi_\omega(\omega) \geq 0} \Big[-\log \sum_{\omega,x,y} \mu(x,y) \pi_\omega(\omega) 2^{\sum_{a,b} B_{abxy} l_\omega(a,b|x,y)} \Big], \\ \overline{\mathrm{IC}}_0^\mu(\mathbf{p}) = &\sup_{\substack{B_{abxy} \\ B_0}} \sum_{a,b,x,y} B_{abxy} \mu(x,y) \ p(a,b|x,y) - B_0 \qquad \text{subject to,} \\ &\bullet \quad \sup_{\pi_\omega(\omega) \geq 0} \Big[\sum_{\omega} \pi_\omega(\omega) \sum_{x,y} \mu(x,y) 2^{\sum_{a,b} B_{abxy} l_\omega(a,b|x,y)} \Big] \leq 2^{B_0}. \end{aligned}$$

As the restriction is linear in π_ω, the maximization over π_ω can be replaced, without loss of generality, by

$$\begin{aligned} \overline{\mathrm{IC}}_0^\mu(\mathbf{p}) = &\sup_{\substack{B_{abxy} \\ B_0}} \sum_{a,b,x,y} B_{abxy} \mu(x,y) \ p(a,b|x,y) - B_0 \qquad \text{subject to,} \\ &\bullet \quad \sum_{x,y} \mu(x,y) 2^{\sum_{a,b} B_{abxy} l(a,b|x,y)} \leq 2^{B_0} \qquad \text{for all } l \in \mathcal{L}_{priv}. \end{aligned}$$

Finally, as a private randomness protocol is a convex combination of deterministic protocols, and the exponential function is convex, then Jensen's inequality 2.3 implies that the maximum is attained for a deterministic protocol. □

9.5 $\mathrm{IC}_0^{\to}$ in the one-way model

In the one-way model, only one players (Alice) communicates and aborts, so we define a zero information complexity $\mathrm{IC}_0^{\to}$ only depending of variables X, A, Ω.

Definition 9.5.1. (One-way zero information complexity)

$$\mathrm{IC}_0^{\mu,\to}(\mathbf{p}) = \inf_{\substack{\eta\geq 0\\ \hat{\pi}_0(\omega)\geq 0}} \quad I_{\pi_0}(X:\Omega|A\neq\perp) \qquad \text{subject to,}$$

- $$\sum_{l_\omega\in\mathcal{L}_{priv}^{\perp,\to}} \hat{\pi}_0(\omega)l\omega(a,b|x,y) = \eta p(a,b|x,y), \qquad \forall a\in A^\perp,\ \forall b\in B^\perp,\ \forall (x,y)\in \mathrm{supp}(\mu),$$
- $$\pi_0(\omega,a,b,x,y) = \hat{\pi}_0(\omega).l_\omega^{\perp,\to}(a,b|x,y).\mu(x,y), \qquad \forall\omega, \forall a\in A^\perp, \forall b\in B^\perp, \forall x\in X, \forall y\in Y.$$

Since, only Alice aborts $\eta_x^l = \eta_{xy}^l$ and $\eta_y^l = 1$, for all $l\in\mathcal{L}_{priv}^{\perp,\to}$. We also define,

$$\hat{\eta}^l = \sum_{x\in\mathcal{X}} \mu(x).\eta_x^l \qquad \text{and} \qquad \hat{\mu}_l(x) = \mu(x)\frac{\eta_x^l}{\hat{\eta}^l}. \tag{9.5}$$

As in the previous Section, we dualize $\mathrm{IC}_0^{\mu,\to}$ to obtain a linear program.

Theorem 9.5.2. *(Zero information complexity)*

$$\mathrm{IC}_0^{\mu,\to}(\boldsymbol{p}) = \sup_{B_{abxy}} \sum_{a,b,x,y} B_{abxy}\mu(x,y)\mathrm{p}(a,b|x,y) \qquad \textit{subject to,}$$

- $$\sum_{a,b,x,y} \nu_x\mu(y|x)B_{abxy}l(a,b|x,y) \leq D(\nu_x||\mu), \qquad \forall l\in\mathcal{L}_{det}^{\to},$$
- $$\nu_x = \frac{\eta_x}{\sum_x' \eta_{x'}\mu(x')}\mu(x), \qquad \forall \eta_x\in[0,1], \forall x\in X.$$

Proof. This proof is a special case of the proof of Theorem 9.2.1.
Starting from Definition 9.1.1, we define $\lambda_\omega = \frac{\hat{\pi}_0}{\eta}$, then the first constraint becomes

$$\sum_{l_\omega\in\mathcal{L}_{priv}^{\perp,\to}} \lambda_\omega l_\omega(a,b|x,y) = p(a,b|x,y).$$

From the second constraint, the joint distribution probability π_0 is defined as

$$\pi_0(\omega,a,b,x,y) = \hat{\pi}_0(\omega).l_\omega^{\perp,\to}(a,b|x,y).\mu(x,y).$$

From this distribution we can obtain,

$$\begin{aligned}
\pi_0(a \neq \perp, x) &= \sum_{\omega}\sum_{b}\sum_{\substack{a\in A\\ b\in B^{\perp}}} \pi_0(\omega, a, b, x, y) = \sum_{\omega} \hat{\pi}_0(\omega).\eta_x^{\omega}.\mu(x) = \eta.\mu(x),\\
\pi_0(a \neq \perp) &= \sum_{x,y} \eta.\mu(x,y) = \eta,\\
\pi_0(x|a \neq \perp) &= \frac{\pi_0(a \neq \perp, x)}{\pi_0(a \neq \perp)} = \mu(x),\\
\pi_0(\omega, x|a \neq \perp) &= \frac{\pi_0(\omega, a \neq \perp, x)}{\pi_0(a \neq \perp,)} = \frac{\hat{\pi}_0(\omega).\eta_x^{\omega}.\mu(x)}{\eta} = \lambda_{\omega}.\eta_x^{\omega}.\mu(x),\\
\pi_0(\omega|a \neq \perp) &= \sum_{x} \lambda_{\omega}.\eta_x^{\omega}.\mu(x) = \lambda_{\omega}.\hat{\eta}^{\omega},
\end{aligned}$$

We use the last three equations to calculate $I_{\pi_0}(X, Y : \Omega|A \neq \perp)$,

$$\begin{aligned}
&I_{\pi_0}(X, Y : \Omega|A \neq \perp)\\
&\qquad = \sum_{x,\omega} \pi_0(\omega, x|a \neq \perp) \log \frac{\pi_0(\omega, x|a \neq \perp)}{\pi_0(\omega|a \neq \perp).\pi_0(x|a \neq \perp)},\\
&\qquad = \sum_{x,\omega} \lambda_{\omega}.\eta_x^{\omega}.\mu(x,y) \log \frac{\eta_x^{\omega}}{\hat{\eta}^{\omega}}.
\end{aligned}$$

Therefore the initial Definition 9.1.1 of IC_0^{μ} becomes,

$$\begin{aligned}
&\mathrm{IC}_0^{\mu,\rightarrow}(\mathbf{p}) = \inf_{\lambda_{\omega}\geq 0} \sum_{x,\omega} \lambda_{\omega}.\eta_x^{\omega}.\mu(x) \log \frac{\eta_x^{\omega}}{\hat{\eta}^{\omega}} \qquad \text{subject to,}\\
&\bullet \sum_{\omega} \lambda_{\omega} l_{\omega}^{\perp}(a, b|x, y) = p(a, b|x, y), \qquad \forall a \in A,\ \forall b \in B,\ \forall (x, y) \in \mathrm{supp}(\mu), \forall l_{\omega}^{\perp} \in \mathcal{L}_{priv}^{\perp,\rightarrow}.
\end{aligned}$$

Note that the one-way zero information method is now a linear program. Hence, let's check the strong duality with the Slater's condition (Definition 6.7.7) by looking for a strictly feasible. Since there is no inequality constraints, we must only find a strictly feasible $(\hat{\lambda}_{\omega})_{\omega}$ such that $\hat{\lambda}_{\omega} < \lambda_{\omega}$ for all ω where $\lambda_{\omega} \neq 0$. Which is trivial.

$$\begin{aligned}
\mathrm{IC}_0^{\mu,\rightarrow}(\mathbf{p}) = \inf_{\lambda_{\omega}\geq 0} \sup_{B_{abxy}} &\sum_{x,\omega} \lambda_{\omega}.\eta_x^{\omega}.\mu(x) \log \frac{\eta_x^{\omega}}{\hat{\eta}^{\omega}}\\
&+ \sum_{a,b,x,y} \mu(x,y).B_{abxy}\Big[p(a, b|x, y) - \sum_{\omega} \lambda_{\omega} l_{\omega}^{\perp}(a, b|x, y)\Big],\\
\mathrm{IC}_0^{\mu}(\mathbf{p}) = \sup_{B_{abxy}} \inf_{\lambda_{\omega}\geq 0} &\sum_{a,b,x,y} \mu(x,y) B_{abxy} p(a, b|x, y)\\
&+ \sum_{x,y,\omega} \lambda_{\omega}.\mu(x)\Big[\eta_x^{\omega} \log \frac{\eta_x^{\omega}}{\hat{\eta}^{\omega}} - \mu(y|x).\sum_{\omega} B_{abxy}.l_{\omega}^{\perp}(a, b|x, y)\Big],\\
\mathrm{IC}_0^{\mu}(\mathbf{p}) = \sup_{B_{abxy}} &\sum_{a,b,x,y} \mu(x,y) B_{abxy} p(a, b|x, y) \qquad \text{subject to,}
\end{aligned}$$

$$\bullet \sum_{x,y} \mu(x,y).B_{abxy}.l_{\omega}^{\perp}(a, b|x, y) \leq \sum_{x} \mu(x).\eta_x^{\omega} \log \frac{\eta_x^{\omega}}{\hat{\eta}^{\omega}}, \qquad \forall a \in A,\ \forall b \in B,\ \forall l_{\omega}^{\perp} \in \mathcal{L}_{priv}^{\perp,\rightarrow}.$$

Remark. After the dualization, the variable ω is not useful. Henceforth, we remove it and use the notation η^l_{xy} instead of η^ω_{xy}.

Note that only the condition of the dual form of IC_0^μ must be satisfied for all private coin randomized protocols $l^{\perp,\rightarrow} \in \mathcal{L}^{\perp,\rightarrow}_{priv}$ without abort. Therefore we can restrict each protocol $l^{\perp,\rightarrow} \in \mathcal{L}^{\perp,\rightarrow}_{priv}$ to $l^{*,\rightarrow} \in \mathcal{L}^{\rightarrow}_{priv}$ with,

$$l^{*,\rightarrow}(a,b|x,y) = \begin{cases} \frac{1}{\eta^l_x} l^{\perp,\rightarrow}(a,b|x,y) & \text{if } \eta^l_x \neq 0, \\ \frac{1}{|A||B|} & \text{otherwise.} \end{cases}$$

Then, the only condition of the dual form of IC_0^μ becomes

$$\sum_{x,y} \mu(x,y).B_{abxy}.l^{\perp,\rightarrow}(a,b|x,y) \leq \sum_{x,y} \mu(x).\eta^l_x \log \frac{\eta^l_x}{\hat{\eta}^l} \qquad \forall l^\perp \in \mathcal{L}^\perp_{priv},$$

$$\sum_{x,y} \mu(x,y).B_{abxy}.\eta^l_x.l^{*,\rightarrow}(a,b|x,y) \leq \sum_{x} \mu(x).\eta^l_x \log \frac{\eta^l_x}{\hat{\eta}^l} \qquad \forall l^\perp \in \mathcal{L}^\perp_{priv},$$

$$\sum_{x,y} \mu(y|x)\hat{\mu}_l(x).B_{abxy}.l^{*,\rightarrow}(a,b|x,y) \leq \sum_{x} \mu_l(x).\log \frac{\hat{\mu}_l(x)}{\mu(x)} \qquad \forall l^\perp \in \mathcal{L}^\perp_{priv},$$

$$\sum_{x,y} \mu(y|x)\hat{\mu}_l(x).B_{abxy}.l^{*,\rightarrow}(a,b|x,y) \leq D(\hat{\mu}_l||\mu_x) \qquad \forall l^\perp \in \mathcal{L}^\perp_{priv},$$

where we define $\hat{\mu}_l(x) = \mu(x)\frac{\eta^l_x}{\hat{\eta}^l}$. Note that from Equation (9.5), $\hat{\mu}_l$ is still a probability distribution over X. Moreover, as every protocol with abort $l^\perp \in \mathcal{L}^{\perp,\rightarrow}_{priv}$ can be construct from a protocol $l \in \mathcal{L}^{\rightarrow}_{priv}$ with abort $(\eta^l_x)_x$, then we can consider η^l_x's independent of $l \in \mathcal{L}^{\perp,\rightarrow}_{priv}$, such that,

$$\sum_{x,y} \nu(x)\mu(y|x).B_{abxy}.l(a,b|x,y) \leq D(\nu||\mu) \qquad \forall l \in \mathcal{L}^{\rightarrow}_{priv},$$

$$\frac{\eta_x}{\sum_{x'} \eta_{x'}\mu(x')}\mu(x) = \nu(x) \qquad \forall x \in X, \forall \eta_x \in [0,1].$$

Finally, as the left term of the constraint is linear in $l \in \mathcal{L}^{\rightarrow}_{priv}$, a private randomness distribution, then this is sufficient to consider deterministic distributions, since $\mathcal{L}^{\rightarrow}_{det} \subset \mathcal{L}^{\rightarrow}_{priv} \subset$ **conv** $\mathcal{L}^{\rightarrow}_{det}$. □

To anticipate the future application, by giving to $\mathrm{IC}_0^{\mu,\rightarrow}$ a more practical form,

Theorem 9.5.3. *(Zero information complexity for one-way model)*

$$\mathrm{IC}_0^{\mu,\rightarrow}(\boldsymbol{p}) = \sup_{B_{abxy},B_0} \sum_{a,b,x,y} B_{abxy}\mu(x,y)\mathrm{p}(a,b|x,y) - B_0 \qquad \textit{subject to,}$$

- $$\sum_x \mu(x)2^{\sum_{a,b,y}\mu(y|x)B_{abxy}l(a,b|x,y)} \leq 2^{B_0}, \qquad \forall l \in \mathcal{L}^{\rightarrow}_{det}.$$

Proof. Changing the variable B_{abxy} to $B_{abxy} - B_0$, and substituting η_x's to ν_x, which is possible only if μ_x is not null. We obtain,

$$\mathrm{IC}_0^{\mu,\rightarrow}(\mathbf{p}) = \sup_{B_{abxy},B_0} \sum_{a,b,x,y} B_{abxy}\mu(x,y)\mathrm{p}(a,b|x,y) - B_0 \qquad \text{subject to,}$$

- $$\inf_{\nu_x \in \mathbb{P}(X_\mu)} \Big[D(\nu_x||\mu_x) - \sum_{a,b,x,y} \nu_x\mu(y|x)B_{abxy}l(a,b|x,y)\Big] \geq -B_0, \qquad \forall l \in \mathcal{L}^{\rightarrow}_{det},$$

where $X_\mu = \mathrm{supp}(\mu_x)$. Using Lemma 9.3.3 on the condition, we obtain

$$\inf_{\nu_x \geq 0} \Big[D(\nu_x || \mu_x) - \sum_{a,b,x,y} \nu_x \mu(y|x) B_{abxy} l(a,b|x,y) \Big] = -\log \sum_x \mu(x) 2^{\sum_{a,b,y} \mu(y|x) B_{abxy} l(a,b|x,y)},$$

which concludes the proof. □

9.6 Application

In this section, we apply the three different versions of the zero information cost, defined previously, to CHSH correlations $\mathbf{p}_\chi$. Originally, CHSH is a game where Alice and Bob receive respectively an input x and y among $\{0,1\}$, then they output a and b among $\{+1,-1\}$ according to

$$(-1)^{xy} = a.b \qquad \text{for all } a,b,x,y.$$

CHSH correlations $\mathbf{p}_\chi$ is a generalization where we simulate the CHSH game for $\chi = 1$, and progressively adding noise when χ goes to zero. Indeed, these correlations can be simulated without communication for $\chi = 0.5$, and with one bit for $\chi = 1$. For $\chi - \frac{\sqrt{2}}{2}$, $\mathbf{p}_\chi$ corresponds to quantum correlations obtained from projective measurement on a Bell pair. We recall that for $\chi \in [0,1]$, $a,b \in \{-1,+1\}$ and $x,y \in \{0,1\}$,

$$p_\chi(a,b|x,y) = \frac{1 + \chi ab(-1)^{xy}}{4}.$$

To calculate IC_0, we don't maximize over all input distributions μ, instead we choose the uniform distribution π_u over $X \times Y$. Thus

$$\mathrm{IC}_0^{\pi_u} \leq \mathrm{IC}_0.$$

Likewise, for each application we use coefficients B_{abxy} with B a real, such that

$$B_{abxy} = B.ab.(-1)^{xy} \qquad \forall x,y \in \{0,1\}, \forall a,b \in \{+1,-1\}. \tag{9.6}$$

Therefore,

$$\begin{aligned} \sum_{\substack{a,b \in \{+1,-1\} \\ x,y \in \{0,1\}}} B_{abxy} \mu(x,y) p_\chi(a,b|x,y) &= \frac{1}{4} \sum_{\substack{a,b \in \{+1,-1\} \\ x,y \in \{0,1\}}} B.ab.(-1)^{xy} \frac{1 + \chi ab(-1)^{xy}}{4}, &(9.7) \\ &= \frac{1}{16} \sum_{\substack{a,b \in \{+1,-1\} \\ x,y \in \{0,1\}}} B\chi, &(9.8) \\ &= B\chi. &(9.9) \end{aligned}$$

Note that all deterministic strategies for Alice and Bob are characterized by functions

$$f_A : \{0,1\} \to \{+1,-1\}, \qquad \text{and} \qquad f_B : \{0,1\} \to \{+1,-1\},$$

then there exists only 16 different deterministic strategies for Alice and Bob. For a conditional distribution $l \in \mathcal{L}$, we define several average values, such that

$$\begin{aligned}
\langle A_x \rangle &= \sum_{a \in A} a\, l(a|x), \\
\langle B_y \rangle &= \sum_{b \in B} b\, l(b|y), \\
\langle A_x B_y \rangle &= \sum_{\substack{a \in A \\ b \in B}} ab\, l(a,b|x,y),
\end{aligned}$$

with $\langle A_x B_y \rangle = \langle A_x \rangle \langle B_y \rangle$, if $l \in \mathcal{L}_{priv}$. Moreover for $l \in \mathcal{L}_{det}$, $\langle A_x \rangle = f_A(x)$ and $\langle B_y \rangle = f_B(y)$.

9.6.1 $\mathrm{IC}_0^{\rightarrow}$ for CHSH correlations

In a previous article [RS09] Szegedy and Roland have shown that

$$\mathrm{IC}_0^{\rightarrow}(\mathbf{p}_\chi) \geq 1 - H(\chi), \qquad \forall \chi \in [0.5, 1].$$

In this Subsection we find this lower bound.

We recall the optimization form of $\mathrm{IC}_0^{\mu,\rightarrow}(\mathbf{p})$.

$$\mathrm{IC}_0^{\mu,\rightarrow}(\mathbf{p}) = \sup_{B_{abxy}, B_0} \sum_{a,b,x,y} B_{abxy}\mu(x,y)\mathrm{p}(a,b|x,y) - B_0 \qquad \text{subject to,}$$

- $$\sum_x \mu(x) 2^{\sum_{a,b,y} \mu(y|x) B_{abxy} l(a,b|x,y)} \leq 2^{B_0}, \qquad \forall l \in \mathcal{L}_{det}^{\rightarrow}.$$

Theorem 9.6.1. *[RS09] For* $\chi \in [0.5, 1]$,

$$\mathrm{IC}_0^{\rightarrow}(\mathbf{p}_\chi) \geq \mathrm{IC}_0^{\pi_u,\rightarrow}(\mathbf{p}_\chi) \geq 1 - H(\chi).$$

Proof. From Equation 9.9, we already know that the objective value is, $B\chi - B_0$. Let characterize the condition for all $l \in \mathcal{L}_{det}$.

$$\begin{aligned}
\sum_x \mu(x) 2^{\sum_{a,b,y} \mu(y|x) B_{abxy} l(a,b|x,y)} &= \frac{1}{2} \sum_x 2^{\frac{B}{2} \sum_{a,b,y} (-1)^{xy} ab.l(a,b|x,y)}, \\
&= \frac{1}{2} \sum_x 2^{\frac{B}{2} \sum_y (-1)^{xy} \langle A_x B_y \rangle}, \\
&= \frac{1}{2} 2^{\frac{B}{2} \langle A_0 \rangle \left(\langle B_0 \rangle + \langle B_1 \rangle \right)} + \frac{1}{2} 2^{\frac{B}{2} \langle A_1 \rangle \left(\langle B_0 \rangle - \langle B_1 \rangle \right)}.
\end{aligned}$$

As $\langle B_y \rangle \in \{+1, -1\}$, at least on term disappear, and by symmetry we only have two different conditions,

$$\frac{1 + 2^B}{2} \leq 2^{B_0} \qquad \text{or} \qquad \frac{1 + 2^{-B}}{2} \leq 2^{B_0}.$$

As B is real these conditions are equivalent, then we choose the first condition and we restraint B to $\mathbb{R}_+$. Thus, we obtain the following optimization program,

$$\mathrm{IC}_0^{\pi_u,\rightarrow}(\mathbf{p}_\chi) \geq \sup_{\substack{B \geq 0 \\ B_0}} B\chi - B_0 \qquad \text{subject to,} \qquad \frac{1 + 2^B}{2} \leq 2^{B_0},$$

$$\mathrm{IC}_0^{\pi_u,\rightarrow}(\mathbf{p}_\chi) \geq \sup_{B \geq 0} B\chi - \log \frac{1 + 2^B}{2},$$

since we want B_0 as small as possible.
We show the objective function, denoted $g(B)$, is strictly concave,

$$\frac{dg(B)}{dB} = \chi - \frac{1}{1+2^{-B}},$$
$$\frac{d^2g(B)}{dB^2} = -\ln 2 \frac{2^{-B}}{(1+2^{-B})^2},$$

where the maximum is,

$$B^\star = \log \frac{\chi}{1-\chi}.$$

Therefore, we conclude

$$\begin{aligned} \mathrm{IC}_0^{\pi_u,\rightarrow}(\mathbf{p}_\chi) &\geq 1 + \chi B^\star - \log\big(1+2^{B^\star}\big), \\ &= 1 + \chi \log \frac{\chi}{1-\chi} - \log\big(1 + \frac{\chi}{1-\chi}\big), \\ &= 1 + \chi \log \frac{\chi}{1-\chi} - \log\big(1 + \frac{\chi}{1-\chi}\big), \\ &= 1 - H(\chi). \end{aligned}$$

Finally, since there exists a protocol for $\mathbf{p}_\chi$ with information cost $1 - H(\chi)$ (Theorem 11 in [RS09]), then this bound is tight. □

9.6.2 $\overline{\mathrm{IC}}_0$ for CHSH correlations

We recall the optimization form of $\overline{\mathrm{IC}}_0^\mu(\mathbf{p})$.

$$\overline{\mathrm{IC}}_0^\mu(\mathbf{p}) = \sup_{\substack{B_{abxy}\\B_0}} \sum_{a,b,x,y} B_{abxy}\mu(x,y)p(a,b|x,y) - B_0 \qquad \text{subject to,}$$
$$\sum_{x,y} \mu(x,y) 2^{\sum_{a,b} B_{abxy} l(a,b|x,y)} \leq 2^{B_0}, \qquad \forall l \in \mathcal{L}_{det}.$$

Theorem 9.6.2. *For $\chi \in [0.5, 1]$. We have*

$$\overline{\mathrm{IC}}_0(\mathbf{p}_\chi) \geq \overline{\mathrm{IC}}_0^{\pi_u}(\mathbf{p}_\chi) = 1 + \frac{1+\chi}{2}\log\frac{1+\chi}{3} + \frac{1-\chi}{2}\log(1-\chi).$$

The above theorem provides a lower bound of approximately 0.046 bits of information complexity of $\mathbf{p}_{\sqrt{2}/2}$, which is a new result. To prove this theorem we use the following fact

Fact 9.6.3. For $a, b \in \{+1, -1\}$ fixed, $(-1)^{xy}ab = 1$ for exactly one or three couples of (x, y).

Proof. From Equation 9.9, we already know that the objective value is, $B\chi - B_0$. Let us characterize the condition for all $l \in \mathcal{L}_{det}$.

$$\begin{aligned} \sum_{x,y} \mu(x,y) 2^{\sum_{a,b} B_{abxy} l(a,b|x,y)} &= \frac{1}{4}\sum_{x,y} 2^{B(-1)^{xy}\sum_{a,b} ab.l(a,b|x,y)}, \\ &= \frac{1}{4}\sum_{x,y} 2^{B(-1)^{xy}\langle A_x B_y\rangle}. \end{aligned}$$

Using Fact 9.6.3, we only have two different conditions,

$$\frac{3.2^{B}+2^{-B}}{4} \leq 2^{B_0} \qquad \text{or} \qquad \frac{3.2^{-B}+2^{B}}{4} \leq 2^{B_0}.$$

Note that the two inequalities are symmetric under the sign change of B. So we can choose a inequality and the sign of B without loss of generality. Then we restrict B to be a nonnegative real, and the left inequality since it is the hardest for B nonnegative. Hence, we obtain the following optimization program,

$$\overline{\mathrm{IC}}_0^{\pi_u}(\mathbf{p}_\chi) = \sup_{\substack{B\geq 0\\ B_0}} \; B\chi - B_0 \qquad\qquad \text{subject to,} \qquad \frac{3.2^{B}+2^{-B}}{4} \leq 2^{B_0},$$

$$\overline{\mathrm{IC}}_0^{\pi_u}(\mathbf{p}_\chi) = \sup_{B\geq 0} \; B\chi - \log\frac{3.2^{B}+2^{-B}}{4},$$

since we want B_0 as small as possible.
We show that the objective function, denoted $g(B)$, is strictly concave,

$$\begin{aligned}\frac{dg(B)}{dB} &= \chi - \frac{3.2^{B}-2^{-B}}{3.2^{B}+2^{-B}},\\ \frac{d^2g(B)}{dB^2} &= -\frac{\ln 2}{(3.2^{B}+2^{-B})^2},\end{aligned}$$

where the maximum is,

$$B^\star = \frac{1}{2}\log\left(\frac{1}{3}\left[\frac{1+\chi}{1-\chi}\right]\right).$$

Finally, we conclude

$$\begin{aligned}\overline{\mathrm{IC}}_0^{\pi_u}(\mathbf{p}_\chi) &= 2+\chi B^\star - \log\left(3.2^{B^\star}-2^{-B^\star}\right),\\ &= 2+\frac{\chi}{2}\log\left(\frac{1}{3}\left[\frac{1+\chi}{1-\chi}\right]\right) - \log\left(3\sqrt{\frac{1}{3}\frac{1+\chi}{1-\chi}}+\sqrt{3\frac{1-\chi}{1+\chi}}\right),\\ &= 2-\frac{\chi+1}{2}\log 3+\frac{\chi}{2}\log\left[\frac{1+\chi}{1-\chi}\right]-\log\frac{2}{\sqrt{(1-\chi)(1+\chi)}},\\ &= 1-\frac{\chi+1}{2}\log 3+\frac{\chi}{2}\log\left[\frac{1+\chi}{1-\chi}\right]+\frac{1}{2}\log(1-\chi)+\frac{1}{2}\log(1+\chi),\\ &= 1-\frac{\chi+1}{2}\log 3+\frac{1+\chi}{2}\log(1+\chi)+\frac{1-\chi}{2}\log(1-\chi),\\ &= 1+\frac{1+\chi}{2}\log\frac{1+\chi}{3}+\frac{1-\chi}{2}\log(1-\chi).\end{aligned}$$

□

9.6.3 IC$_0$ for CHSH correlations

From Theorem 9.3.2 and Lemma 9.3.4, we know

$$\mathrm{IC}_0^{\mu}(\mathbf{p}) = \sup_{B_{abxy},B_0} \sum_{a,b,x,y} B_{abxy}\mu_x\mu_y \mathrm{p}(a,b|x,y) - B_0, \qquad\qquad \text{subject to,}$$

$$\inf_{\nu_x\geq 0} D(\nu_x||\mu_x) - \log\left[\sum_y \mu_y 2^{\sum_{a,b,x}\nu_x B_{abxy} l(a,b|x,y)}\right] \geq -B_0, \qquad\qquad \forall l\in\mathcal{L}_{det}.$$

Unfortunately, the derivation of $\mathrm{IC}_0^{\pi_u}(\mathbf{p}_\chi)$ involves a optimization problem that stays unsolved, therefore we have done a numerical analysis (Figure 9.6.3).

In Figure 9.6.3, we observe that the order between all three zero information methods is respected.

$$\overline{\mathrm{IC}}_0^{\pi_u}(\mathbf{p}_\chi) \leq \mathrm{IC}_0^{\pi_u}(\mathbf{p}_\chi) \leq \mathrm{IC}_0^{\rightarrow}(\mathbf{p}_\chi).$$

Since $\overline{\mathrm{IC}}_0$ is a relaxation of IC_0. And $\mathrm{IC}_0^{\rightarrow}$ is above IC_0, as the one-way model is a restriction of the communication model.

We observe that the curves for $\mathrm{IC}_0^{\rightarrow}(\mathbf{p}_\chi)$ and $\overline{\mathrm{IC}}_0^{\pi_u}(\mathbf{p}_\chi)$ converges for χ going to one. From this numerical evidence we conjecture that

$$\mathrm{IC}_0^{\pi_u}(\mathbf{p}_\chi) \geq 1 - H(\chi) - O\big(H(\chi)^2\big).$$

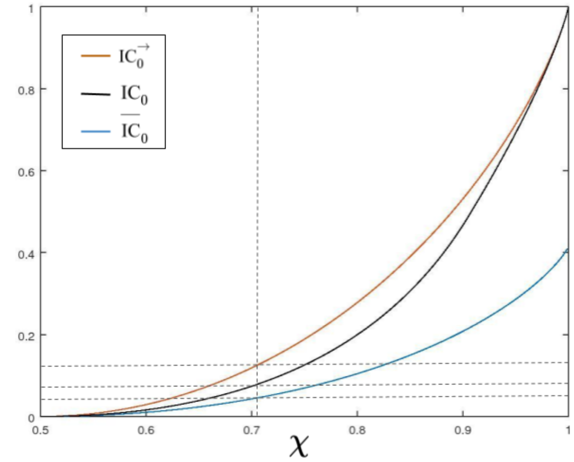

Figure 9.1: Lower bounds obtain from $\mathrm{IC}_0^{\rightarrow}(\mathbf{p}_\chi)$ (red curve), $\mathrm{IC}_0^{\pi_u}(\mathbf{p}_\chi)$ (black curve) and $\overline{\mathrm{IC}}_0^{\pi_u}(\mathbf{p}_\chi)$ (blue curve) for CHSH correlations $\mathbf{p}_\chi$ with B_{abxy} as defined in (9.6) and an uniform input distribution.

Conclusion

This thesis mainly focuses on studying information complexity and quantum query complexity via convex optimization tools. Indeed, optimization problems naturally arise while studying these complexities, and tools such as the Envelope theorem 6.8.1 or the recent Lemma 9.3.3 allow to progress in these subfields of computational complexity.

Firstly in Chapter 7.2, the main result is our universal adiabatic quantum query algorithm. Although Theorem [LMR+11] was already proved, this algorithm provides a direct proof, as well as a simple description as an adiabatic process. Also, we have provided an original proof that $\mathrm{Adv}_0^{\mathrm{reg}}$ is a lower bound of Q_0^{ct} (Theorem 7.1.1).

To go further, some functions only have a discrete quantum algorithm, our universal quantum query algorithm allows to construct continuous quantum an algorithm for these functions.

In Chapter 8, we have enlarged our understanding of how a set of quantum states evolves while querying an oracle. From this knowledge we have constructed a new norm, the query Lagrangian $\mathcal{L}_d(\gamma, \dot{\gamma})$, a semi-definite program that defines the infinitesimal number of query for an infinitesimal motion from γ to $\gamma + \delta\dot{\gamma}$ for δ infinitesimal. Thus, we define our new method, the adversary action Sadv, a minimization program over all possible paths. Afterwards, from KKT conditions 6.6.1, Euler-Lagrange equation D.0.2 and Envelope theorem 6.8.1, we derive necessary conditions for optimal points $\mathcal{U}$ and $\mathcal{V}$ of Sadv. Also, we have shown that this new method subsumes both the adversary method and the multiplicative method.

Thus, we have provided a refined method to better characterize the quantum query complexity in the exact case or in the unbounded error case.

In the last Chapter 9, we have two important results. Firstly, we have constructed a new method IC_0 to lower bound the external information complexity. Secondly, we have successfully applied these methods to CHSH correlations. For $\mathrm{IC}_0^{\rightarrow}$, we retrieve the known result of [RS09], the relaxed form $\overline{\mathrm{IC}}_0$ provides a new lower bound of 0.046 bits for the quantum correlations p_χ with $\chi = \sqrt{2}/2$ in the two-way case. For IC_0, we provide numerical evidence that this new method is a good lower bound, while its analytic solution is left open.

To go further, we have good hope to solve the conjecture on $\mathrm{IC}_0^{\pi_u}$ for CHSH correlations. Finally, as IC_0 is a new method we can apply it to other correlations such as those appearing in the EPR-Bohm experiment[1].

[1] Correlations obtained from all projective measurements on a Bell pair.

Part III

Appendices

Appendix A

Adiabatic theorem without a gap condition

In this Appendix we give an adapted version of the proof of Lemma A.0.1 in [AE99a]. We derive an upper bound on the error ε_{AP} caused by the adiabatic process without a gap condition. We use the same notations as in Subsection 3.1.1 in Chapter 3.

Lemma A.0.1. *[AE99a]*
Let $\{H(s), \boldsymbol{P}(s), \tau\}$ be an adiabatic process, $\varepsilon > 0$, and $X(s)$ be an operator satisfying the commutator equation

$$\dot{P}(s)P(s) = [H(s), X(s)], \tag{A.1}$$

and both $X(s)$ and $\dot{X}(s)$ bounded.

$$\textit{If} \quad \tau \geq \frac{1}{\varepsilon}\Big[\|X(0)\| + \|X(1)\| + \max_{s\in[0,1]} \|\dot{X}(s)P(s)\|\Big], \quad \textit{then} \quad \varepsilon_{AP} \leq \varepsilon.$$

Proof of Lemma A.0.1 In order to bound the quantity ε_{AP}, we would like to describe an idealized adiabatic evolution $U_A(s)$ that transports the projector $P(0)$ to $P(s)$, such that $U_A(s)P(0) = P(s)U_A(s)$. To achieve this, we use a technique given by [Kat50] (later improved in [AE99b]), and define $H_A(s)$ as the *adiabatic Hamiltonian*

$$H_A(s) = \lambda(s)\mathcal{I}d + \frac{i}{\tau}[\dot{P}(s), P(s)], \tag{A.2}$$

where $[\,\cdot\,,\,\cdot\,]$ is the commutator. We define $U_A(s)$ to be the solution of the Schrödinger's equation for this Hamiltonian, that is,

$$i\partial_s U_A(s) = \tau H_A(s)U_A(s), \tag{A.3}$$

with the initial condition $U_A(0) = \mathcal{I}d$. The existence and uniqueness of $U_A(s)$ follows from the analytical properties in Definition 3.1.1. Moreover we show that $U_A(s)$ has the desired property.

Lemma A.0.2. *[Kat50] (Intertwining property)*

$$U_A(s)P(0) = P(s)U_A(s). \tag{A.4}$$

The proof of this property uses the following Fact.

Fact A.0.3. For any orthogonal projector P we have $P = P^2$, so that $\dot{P} = \dot{P}P + P\dot{P}$ and $P\dot{P}P = 0$.

Proof of Lemma A.0.2. Since $U_A(s)$ is the solution of the differential equation $i\partial_s Y(s) = \tau H_A(s)Y(s)$ with $Y(0) = \mathcal{I}d$, then every other solution of this equation has the form $Y(s) = U_A(s)Y(0)$. All we need to do is prove that $P(s)U_A(s)$ is also a solution. Indeed, this implies that $P(s)U_A(s) = U_A(s)Y(0)$, and by setting $s = 0$ we obtain $P(0) = Y(0)$. Using Fact A.0.3, we have

$$\begin{aligned}
i\partial_s\big(P(s)U_A(s)\big) &= i\dot{P}(s)U_A(s) + P(s)\tau H_A(s)U_A(s) \\
&= i\dot{P}(s)U_A(s) + \tau\lambda(s)P(s)U_A(s) + iP(s)[\dot{P}(s), P(s)]U_A(s) \\
&= \tau\lambda(s)P(s)U_A(s) + i\big(\dot{P}(s) - P(s)\dot{P}(s)\big)U_A(s) \\
&= \tau\lambda(s)P(s)U_A(s) + i\dot{P}(s)P(s)U_A(s) \\
&= \big(\tau\lambda(s)\mathcal{I}d + i[\dot{P}(s), P(s)]\big)P(s)U_A(s) \\
&= \tau H_A(s)P(s)U_A(s)
\end{aligned}$$

In the third and fifth lines we use $P\dot{P}P = 0$. In the fourth line we use $\dot{P} - P\dot{P} = \dot{P}P$. □

In order to prove Lemma A.0.1, we need two more claims.

Note that $\varepsilon_{AP}(s)$ can be rewritten as $\|\big(\Omega(s) - \mathcal{I}d\big)P(0)\|$, where $\Omega(s) = U_\tau^*(s)U_A(s)$.

Claim A.0.4. $\dot{\Omega}(s)P(0) = U_\tau^*(s)\dot{P}(s)U_A(s)P(0)$

Proof. Using (3.6) and (A.2), we obtain

$$\dot{\Omega}(s) = U_\tau^*(s)\Big[i\tau\big(H(s) - \lambda(s)\mathcal{I}d\big) + [\dot{P}(s), P(s)]\Big]U_A(s).$$

The claim follows from the intertwining property (Lemma A.0.2), Fact A.0.3 and $H(s)P(s) = \lambda(s)P(s)$. □

Claim A.0.5. Let $\Phi(s) = e^{-i\tau\lambda(s)}\mathcal{I}d$ and $V_A(s) = \Phi^*(s)U_A(s)$. Then $V_A(s)$ satisfies the intertwining property (A.4), that is, $V_A(s)P(0) = P(s)V_A(s)$, as well as the Schrödinger's equation $\dot{V}_A(s) = [\dot{P}(s), P(s)]V_A(s)$.

Proof. The fact that $V_A(s)$ satisfies the intertwining property is immediate since $U_A(s)$ satisfies this property and $\Phi(s)$, being proportional to the identity, commutes with any operator. The fact that it satisfies the Schrödinger's equation follows from the fact that $\Phi(s)$ satisfies $i\dot{\Phi}(s) = \tau\lambda(s)\Phi(s)$, $U_A(s)$ satisfies $i\dot{U}_A(s) = \tau H_A(s)U_A(s)$, and both terms of $H_A(s) = \lambda(s)\mathcal{I}d + \frac{i}{\tau}[\dot{P}(s), P(s)]$ commute. □

Let $X(s)$ an operator solution of $\dot{P}(s)P(s) = [H(s), X(s)]$, then

$$\begin{aligned}
\big(\Omega(s) - \mathcal{I}d\big)P(0) &= \int_0^s \dot{\Omega}(s')ds'P(0) \\
&= \int_0^s U_\tau^*(s')\dot{P}(s')U_A(s')ds'P(0) \\
&= \int_0^s U_\tau^*(s')\Phi(s')\dot{P}(s')V_A(s')ds'P(0) \\
&= \int_0^s U_\tau^*(s')\Phi(s')[H(s'), X(s')]V_A(s')ds'P(0) \\
&= \int_0^s U_\tau^*(s')\Phi(s')\big(H(s') - \lambda(s')\big)X(s')V_A(s')ds'P(0) \\
&= \frac{1}{i\tau}\int_0^s \partial_{s'}[U_\tau^*(s')\Phi(s')]X(s')V_A(s')ds'P(0) \\
&= \frac{1}{i\tau}\Big[U_\tau^*(s')\Phi(s')X(s')V_A(s')\Big]_0^s P(0) - \frac{1}{i\tau}\int_0^s U_\tau^*(s')\Phi(s')\partial_{s'}[X(s')V_A(s')]ds'P(0) \\
&= \frac{1}{i\tau}\Big[U_\tau^*(s')X(s')U_A(s')\Big]_0^s P(0) - \frac{1}{i\tau}\int_0^s U_\tau^*(s')\big(\dot{X}(s') + X(s')\dot{P}(s')\big)U_A(s')ds'P(0)
\end{aligned}$$

We explain line by line:

(1 → 2) We use Claim A.0.4.

(2 → 3) We rearrange the expression using $U_A(s) = \Phi(s)V_A(s)$ and the fact that $\Phi(s)$ commutes with any operator.

(3 → 4) We use the intertwining property for $V_A(s)$ (Claim A.0.5) and Equation (A.1).

(6 → 7) We integrate by parts.

The third term in the last line is null, because $X(s) = X(s)P(s)$ and the intertwining property (Lemma A.0.2) yields the expression $P\dot{P}P$, which is zero by Fact A.0.3. Using the triangle inequality, the fact that a norm is preserved by unitary operations and can only decrease under projections, we finally have

$$\begin{aligned}
\varepsilon_{AP}(s) &= \|\big(\Omega(s) - \mathcal{I}d\big)P(0)\| \\
&\leq \frac{1}{\tau}\Big[\|X(0)\| + \|X(s)\| + s \sup_{s'\in[0,s]} \|\dot{X}(s')P(s')\|\Big]
\end{aligned}$$

This conclude the proof. ■

Appendix B

Slater's theorem

Theorem B.0.1. *[Sla14] For a convex optimization problem, if Slater's condition holds, then strong duality holds.*

Let's summarize the demonstration.

In the first part [1], we construct a map of the domain $\mathcal{D}$ to a set $\mathcal{A}$ in the vector space $\mathbb{R}\times\mathbb{R}^p\times\mathbb{R}^q$. In this new representation, a Lagrangian can be written as an inner product, and the Lagrange dual function $d(\lambda,\mu)$ represents a hyperplane supporting the set $\mathcal{A}$. Especially, we show that $\mathcal{A}$ is convex, and an optimal point is mapped in $\mathbf{bd}\,\mathcal{A}$.
In the second part [2], we use the *supporting hyperplane theorem* to show than there exists a hyperplane supporting an optimal point in $\mathbf{bd}\,\mathcal{A}$, and we assume we can interpret this hyperplane as a Lagrangian.
Finally in [3] we use the Slater's condition to get rid-of the previous assumption.

Before to start we make several assumptions.
We **can assume** that $p^\star$ is finite. Indeed the convex optimization problem is not unfeasible since Slater's condition is satisfied, there exists x_S a strictly feasible point. Moreover if the problem is unbounded, then the weak duality directly implies the strong duality.
As equality functions h_j are affine, we define $H\in\mathcal{M}_{q,n}(\mathbb{R})$ and $k\in\mathbb{R}^q$ such that $\boldsymbol{h}(x)=Hx-k$. We **assume** without loss of generality that $\mathbf{rank}\,H=q$, and $\mathbf{int}\,\mathcal{D}$ is non-empty (by choosing $\mathbb{R}^n=\mathbf{aff}\,\mathcal{D}$).

[1] Let's define $\mathcal{G}\subset\mathbb{R}\times\mathbb{R}^p\times\mathbb{R}^q$,

$$\mathcal{G}=\big\{\big(f(x),g_1(x),\dots,g_p(x),h_1(x),\dots,h_q(x)\big)\big|x\in\mathcal{D}\big\},$$

then from $\mathcal{G}$ we define $\mathcal{A}$,

$$\mathcal{A}=\mathcal{G}+\big(\mathbb{R}_+\times\mathbb{R}_+^p\times\{0\}^q\big),$$

where the addition is entry-wise by entry-wise. The relation between $\mathcal{D}$ and $\mathcal{A}$ is straightforward. For each $(t,\boldsymbol{u},\boldsymbol{v})\in\mathcal{A}$, there exists $x\in\mathcal{D}$, such that $f(x)\le t$, $\mathbf{g}(x)\le\boldsymbol{u}$ and $\mathbf{h}(x)=\boldsymbol{v}$.
In the other direction, for each $x\in\mathcal{D}$, every point $(t,\boldsymbol{u},\boldsymbol{v})$ satisfying $f(x)\le t$, $\mathbf{g}(x)\le\boldsymbol{u}$ and $\boldsymbol{h}(x)=\boldsymbol{v}$, is in $\mathcal{A}$.

In the vector space $\mathbb{R}\times\mathbb{R}^p\times\mathbb{R}^q$, the Lagrangian can be represented as an inner product between a vector $(1, \boldsymbol{\lambda}, \boldsymbol{\mu})$ and a vector in $\mathcal{G}$,

$$\begin{aligned} L(x, \boldsymbol{\lambda}, \boldsymbol{\mu}) &= f(x) + \langle \boldsymbol{\lambda}, \mathbf{g}(x) \rangle + \langle \boldsymbol{\mu}, \mathbf{h}(x) \rangle, \\ &= \Big\langle \big(1, \boldsymbol{\lambda}, \boldsymbol{\mu}\big), \big(f(x), \mathbf{g}(x), \mathbf{h}(x)\big) \Big\rangle. \end{aligned}$$

Likewise the Lagrange dual function in Definition (6.4.1), the minimization over $\mathcal{D}$ can be replaced by the minimization over $\mathcal{G}$, and even extended over $\mathcal{A}$ if $\lambda \geq 0$.

$$\begin{aligned} d(\boldsymbol{\lambda}, \boldsymbol{\mu}) &= \inf_{x \in \mathcal{D}} L(x, \boldsymbol{\lambda}, \boldsymbol{\mu}), \\ &= \inf_{x \in \mathcal{D}} \Big\langle \big(1, \boldsymbol{\lambda}, \boldsymbol{\mu}\big), \big(f(x), \mathbf{g}(x), \mathbf{h}(x)\big) \Big\rangle, \\ &= \inf_{(t,u,v) \in \mathcal{G}} \Big\langle \big(1, \boldsymbol{\lambda}, \boldsymbol{\mu}\big), \big(t, \boldsymbol{u}, \boldsymbol{v}\big) \Big\rangle, \\ &= \inf_{(t,u,v) \in \mathcal{A}} \Big\langle \big(1, \boldsymbol{\lambda}, \boldsymbol{\mu}\big), \big(t, \boldsymbol{u}, \boldsymbol{v}\big) \Big\rangle. \end{aligned}$$

The last equality comes from the definition of $\mathcal{A}$, since for all $(t, \boldsymbol{u}, \boldsymbol{v}) \in \mathcal{A}$ with $\lambda \geq 0$, there exists $x \in \mathcal{D}$, $L(x, \boldsymbol{\lambda}, \boldsymbol{\mu}) \leq \langle (1, \boldsymbol{\lambda}, \boldsymbol{\mu}), (t, \boldsymbol{u}, \boldsymbol{v}) \rangle$.
Together with Property 6.4.2, this implies,

$$\text{if}\, \boldsymbol{\lambda} \geq 0\,, \qquad \text{then} \quad p^\star \geq \inf_{(t,\boldsymbol{u},\boldsymbol{v}) \in \mathcal{A}} \Big\langle \big(1, \boldsymbol{\lambda}, \boldsymbol{\mu}\big), \big(t, \boldsymbol{u}, \boldsymbol{v}\big) \Big\rangle.$$

The proof of the convexity of $\mathcal{A}$ is straightforward. Let $(t_0, \boldsymbol{u_0}, \boldsymbol{v_0})$ and $(t_1, \boldsymbol{u_1}, \boldsymbol{v_1}) \in \mathcal{A}$, then there exists respectively x_0 and x_1 in $\mathcal{D}$ with the following properties: $f(x_0) \leq t_0$, $\mathbf{g}(x_0) \leq \boldsymbol{u_0}$, $\boldsymbol{h}(x_0) = \boldsymbol{v_0}$, and $f(x_1) \leq t_1$, $\mathbf{g}(x_1) \leq \boldsymbol{u_1}$, $\boldsymbol{h}(x_1) = \boldsymbol{v_1}$, Let $\alpha \in [0, 1]$, then each α-convex combination of x_0 and x_1 implies that the α-convex combination of $(t_0, \boldsymbol{u_0}, \boldsymbol{v_0})$ and $(t_1, \boldsymbol{u_1}, \boldsymbol{v_1})$ is in $\mathcal{A}$.

Now we show that $(p^\star, 0, 0)$ is in $\mathbf{bd}\,\mathcal{A}$. There exists a sequence x_n of feasible points such that $f(x_n) \to p^\star$, then there exists the sequence $(f(x_n), 0, 0) \in \mathcal{A}$. Then $(p^\star, 0, 0)$ is in $\mathbf{cl}\,\mathcal{A}$. So those $s < p^\star$, $(s, 0, 0) \notin \mathcal{A}$ otherwise there exists a feasible point x, with $f(x) < p^\star$. Therefore $(p^\star, 0, 0)$ is also in $\mathbf{cl}\,\mathcal{A}^C$.

[2] Since $\mathcal{A}$ is convex and $(p^\star, 0, 0) \in \mathbf{cl}\,\mathcal{A}$, the *supporting hyperplane theorem* 2.6.1 implies the existence of a supporting hyperplane $(\hat{\nu}, \hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\mu}}) \neq 0$, such that

$$\text{for all}\, (t, \boldsymbol{u}, \boldsymbol{v}) \in \mathcal{A}, \qquad \Big\langle \big(\hat{\nu}, \hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\mu}}\big), \big(t, \boldsymbol{u}, \boldsymbol{v}\big) \Big\rangle \geq \Big\langle \big(\hat{\nu}, \hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\mu}}\big), \big(p^\star, 0, 0\big) \Big\rangle, \tag{B.1}$$

$$\hat{\nu} t + \langle \hat{\boldsymbol{\lambda}}, \boldsymbol{u} \rangle + \langle \hat{\boldsymbol{\mu}}, \boldsymbol{v} \rangle \geq \hat{\nu} p^\star. \tag{B.2}$$

From the definition of $\mathcal{A}$, t and $\boldsymbol{u}$ can be as large as possible, then this inequality implies that $\hat{\nu} \geq 0$ and $\hat{\boldsymbol{\lambda}} \geq 0$, otherwise the lower bound can be violated.

Assume that $\hat{\nu} > 0$, and let's minimize the above lower bound over $\mathcal{A}$, we obtain

$$\inf_{(t,\boldsymbol{u},\boldsymbol{v}) \in \mathcal{A}} t + \langle \hat{\boldsymbol{\lambda}} / \hat{\nu}, u \rangle + \langle \hat{\boldsymbol{\mu}} / \hat{\nu}, v \rangle \geq p^\star. \tag{B.3}$$

As $\boldsymbol{\lambda} \geq 0$, Equation (B) implies $d(\hat{\boldsymbol{\lambda}}/\hat{\nu}, \hat{\boldsymbol{\mu}}/\hat{\nu}) = p^{\star}$; the strong duality.

[3] Assume that $\hat{\nu} = 0$, then the inequality (6.19) becomes

$$\text{for all}\, x \in \mathcal{D}, \qquad \left\langle \hat{\boldsymbol{\lambda}}, \boldsymbol{g}(x) \right\rangle + \left\langle \boldsymbol{\mu}, Hx - k \right\rangle \geq 0.$$

Since Slater's condition implies the existence of a strictly feasible point x_S, with $\boldsymbol{g}(x_S) < 0$ and $Hx_S - k = 0$. Therefore $\hat{\boldsymbol{\lambda}} = 0$, otherwise the above inequality is violated by x_S.

Afterwards the inequality becomes,

$$\text{for all}\, x \in \mathcal{D}, \qquad \left\langle \hat{\boldsymbol{\mu}}, Hx - k \right\rangle \geq 0.$$

From the definition of a strictly feasible point $Hx_S - k$ is null. Since $x_S \in \mathbf{int}\, \mathcal{D}$ there exists a point x_V in a neighborhood of x_S, such that $\left\langle \hat{\boldsymbol{\mu}}, Hx_V - k \right\rangle < 0$, unless $\hat{\boldsymbol{\mu}}^T H = 0$. However at the beginning we have assumed $\mathbf{rank}\, H = q$. Contradiction.

Appendix C

Envelope theorem

Let $\mathcal{A}$ and $\mathcal{B}$ be two non-empty sets. We define $F : \mathcal{A} \times \mathcal{B} \times [0,1] \to \mathbb{R}$ to be a function, such that for almost all $t \in [0,1]$, $F(a,b,t)$ has a saddle-point $(a^\star, b^\star)$ in $\mathcal{A} \times \mathcal{B}$. In other words, for almost all $t \in [0,1]$,

$$\forall\ a \in \mathcal{A}, b \in \mathcal{B}, \qquad F(a^\star, b, t) \leq F(a^\star, b^\star, t) \leq F(a, b^\star, t).$$

The function F is interpreted as the Lagrangian of an optimization problem where the strong duality holds, such that

$$p^\star(t) = \inf_{a \in \mathcal{A}} \sup_{b \in \mathcal{B}} F(a,b,t) = \sup_{b \in \mathcal{B}} \inf_{a \in \mathcal{A}} F(a,b,t).$$

Note that for each $t \in [0,1]$, the set of all saddle-points is the product set $\mathcal{A}^\star(t) \times \mathcal{B}^\star(t)$ defined as

$$\mathcal{A}^\star(t) = \big\{ a \in \mathcal{A} \,\big|\, \sup_{b \in \mathcal{B}} F(a,b,t) = p^\star(t) \big\},$$
$$\mathcal{B}^\star(t) = \big\{ b \in \mathcal{B} \,\big|\, \inf_{a \in \mathcal{A}} F(a,b,t) = p^\star(t) \big\}.$$

Theorem C.0.1 (Envelope theorem)**.** *[MS02] Assume that:*

(1) *for almost all $t \in [0,1]$, $\mathcal{A}^\star(t) \times \mathcal{B}^\star(t)$ is non-empty,*

(2) *for all $a \in \mathcal{A}$, $b \in \mathcal{B}$, $F(a,b,t)$ is absolutely continuous in t,*

(3) *there exists $c : [0,1] \to \mathbb{R}$, an integrable function that bounds $|D_t\, F(a,b,t)| \leq c(t)$, for all $(a,b) \in \mathcal{A} \times \mathcal{B}$, and almost all $t \in [0,1]$.*

Then $p^\star(t)$ is absolutely continuous. Assume, in addition, that:

(4) *$\mathcal{A}$ and $\mathcal{B}$ are topological spaces satisfying the second axiom of countability,*

(5) *$D_t\, F(a,b,t)$ is continuous in each of $a \in \mathcal{A}$ and $b \in \mathcal{B}$,*

(6) *the family $\{F(a,b,t)\}_{(a,b)}$ is equi-differentiable in t, for all $(a,b) \in \mathcal{A} \times \mathcal{B}$.*

Then for any selection $\big(a^\star(t), b^\star(t)\big) \in \mathcal{A}^\star(t) \times \mathcal{B}^\star(t)$,

$$p^\star(t) = p^\star(0) + \int_0^t ds\, D_t F\big(a^\star(s), b^\star(s), s\big).$$

First we demonstrate that $p^\star(t)$ is absolutely continuous.
Let $t_0, t_1 \in [0,1]$ with $t_0 < t_1$, if $F(\cdot,\cdot,t_0)$ and $F(\cdot,\cdot,t_1)$ have respectively saddle points (a_0,b_0) and (a_1,b_1) then

$$F(a_1,b_0,t_1) - F(a_1,b_0,t_0) \leq p^\star(t_1) - p^\star(t_0) \leq F(a_0,b_1,t_1) - F(a_0,b_1,t_0).$$

Consequently for almost all $t_0 < t_1$,

$$|p^\star(t_1) - p^\star(t_0)| \leq \sup_{\substack{a\in\mathcal{A}\\ b\in\mathcal{B}}} \big|F(a,b,t_1) - F(a,b,t_0)\big|.$$

From Assumptions **(2)** and **(3)**, we have

$$\begin{aligned}\sup_{\substack{a\in\mathcal{A}\\ b\in\mathcal{B}}} \big|F(a,b,t_1) - F(a,b,t_0)\big| &= \sup_{\substack{a\in\mathcal{A}\\ b\in\mathcal{B}}} \left|\int_{t_0}^{t_1} ds\, D_t\, F\big(a,t,s\big)\right|, \\ &\leq \sup_{\substack{a\in\mathcal{A}\\ b\in\mathcal{B}}} \int_{t_0}^{t_1} ds\, \big|D_t\, F\big(a,t,s\big)\big|, \\ &\leq \int_{t_0}^{t_1} ds\, c(s), \\ &\leq \sup_{I:|I|=t_1-t_0} \int_I ds\, c(s).\end{aligned} \tag{C.1}$$

Since $c(t)$ is an integrable function the last term is bounded independently of any subinterval of size $t_1 - t_0$, implying that $p^\star(t)$ is absolutely continuous.

Now to prove the **second** part, if $p^\star(t)$ is differentiable at t_0, we must show that $D_t\, p^\star(t_0) = D_t\, F\big(a_0,b_0,t_0\big)$, with $(a_0,b_0) \in \mathcal{A}^\star(t_0) \times \mathcal{B}^\star(t_0)$.
Consider a saddle-point selection $(a_t^\star, b_t^\star)$, and its graph $G = \{(t, a_t^\star, b_t^\star) : t \in [0,1]\} \subset [0,1] \times \mathcal{A} \times \mathcal{B}$. Since the product topological space $[0,1] \times \mathcal{A} \times \mathcal{B}$ satisfies the second axiom of countability, by Assumption **(4)**, then the set of isolated points of G is at most countable. Hence the set of S of points $t \in [0,1]$, such that $(t, a_t^\star, b_t^\star)$ is not isolated in G, has full measure on $[0,1]$.
For any point $t_0 \in S$, $(a_{t_0}^\star, b_{t_0}^\star) = (a_0,b_0)$ is not isolated. Then there exists a sequence $\{(t_k,a_k,b_k)\}_{k=1}^\infty \subset G$, such that $(t_k,a_k,b_k) \to (t_0,a_0,b_0)$, as $k \to \infty$ and $t_k \neq t_0$. Moreover we can choose the sequence, such that $t_0 < t_k$ for all k.
By the definition of a saddle-point, we have

$$\frac{F(a_k,b_0,t_k) - F(a_k,b_0,t_0)}{t_k - t_0} \leq \frac{p^\star(t_k) - p^\star(t_0)}{t_k - t_0} \leq \frac{F(a_0,b_k,t_k) - F(a_0,b_k,t_0)}{t_k - t_0}.$$

From Assumption **(6)** the family $\{F(a,b,t)\}_{(a,b)}$ is equi-differentiable in t for all $(a,b) \in \mathcal{A} \times \mathcal{B}$, then we have

$$D_t\, F(a_k,b_0,t_0) + \frac{o(t_k - t_0)}{t_k - t_0} \leq \frac{p^\star(t_k) - p^\star(t_k)}{t_k - t_0} \leq D_t\, F(a_0,b_k,t_0) + \frac{o(t_k - t_0)}{t_k - t_0}.$$

As k goes to ∞, by the continuity of $D_t\, F(a,b,t)$ for each of $a \in \mathcal{A}$ and $b \in \mathcal{B}$, both bounds converge to $D_t\, F(a_0,b_0,t_0)$. Hence we have $D_t\, p^\star(t_0) = D_t\, F\big(a_0,b_0,t_0\big)$. Since this result holds for each $t_0 \in S$ with full measure in $[0,1]$, this concludes the proof.

Appendix D

Euler-Lagrange equation

In this Appendix, we introduce the Euler-Lagrange equation used in the field of Calculus of variations field. For a functional S, this equation allows to derive necessary conditions that a local optimal function $\boldsymbol{x}$ satisfied. In the proof of the Euler-Lagrange equation, we use the Fundamental lemma of calculus of variations. Here we introduce a vectorial version of this lemma, more suitable to our needs.

Lemma D.0.1 (Fundamental lemma of calculus of variations (vectorial version))**.** *Let U be an open of $\mathbb{R}^N$, E be a pre-hilbertian space and E' its topological dual space. If a locally continuous function $f : U \to E$ satisfies the equality,*

$$\int_U \langle f, h \rangle = 0, \qquad \textit{for all compactly supported smooth functions } \; h : U \to E,$$

then f is identically null almost everywhere.

Proof. Proof by contradiction.
Let $x_0 \in U$ such that $f(x_0) \neq 0$ then there exists $y \in E'$ with $\langle f(x_0), y \rangle > 0$. As f is locally continuous there exists $R, \delta > 0$ such that

$$\langle f(x), y \rangle \geq \delta \qquad \text{for } x \in B_R(x_0).$$

We choose the compactly supported smooth function $h(x) = \rho(\|x - x_0\|)y$ where

$$\rho(r) = \left\{ \begin{array}{cl} e^{-\frac{1}{R^2 - r^2}} & \text{if } r < R, \\ 0 & \text{if } r \geq R. \end{array} \right.$$

Therefore,

$$\int_U \langle f, h \rangle = \int_U dx \, \langle f(x), h(x) \rangle \geq \delta \int_{B_R(x_0)} dx \rho(\|x - x_0\|) > 0.$$

□

Let $[a, b]$ be a real interval, E be a pre-hilbertian space and U_1, U_2 be two open sets of E. We define L a continuously differentiable function called Lagrangian as

$$L : \quad [a, b] \times U_1 \times U_2 \to \mathbb{R} : \quad (t, \boldsymbol{x}, \boldsymbol{u}) \mapsto L(t, \boldsymbol{x}, \boldsymbol{u}),$$

where $\boldsymbol{x}$ is a differentiable function $\boldsymbol{x} : [a, b] \to U_1$ such that $\dot{\boldsymbol{x}} \in U_2$ for all $t \in [a, b]$.

Theorem D.0.2 (Euler-Lagrange equation)**.** *Let S be a functional defined for all functions $\boldsymbol{x}$ such that,*

$$S[\boldsymbol{x}] = \int_a^b dt\, L\big(\boldsymbol{x}(t), \dot{\boldsymbol{x}}(t), t\big).$$

Then the Euler-Lagrange equation given by,

$$\frac{\partial L}{\partial \boldsymbol{x}} - \frac{d}{dt}\frac{\partial L}{\partial \dot{\boldsymbol{x}}} = 0,$$

is a necessary condition for L to be a local optimal of S.

Proof. Let $\boldsymbol{x}$ be a local optimal of S. We define $\boldsymbol{x}_\varepsilon(t) = \boldsymbol{x}(t) + \varepsilon \boldsymbol{h}(t)$ where $\boldsymbol{h} : [a, b] \to U_1$ with $\boldsymbol{h}(a) = \boldsymbol{h}(b) = 0$. As $\boldsymbol{x}$ is a local optimal we have

$$\begin{aligned}
0 &= \frac{dS[\boldsymbol{x}]}{d\varepsilon}, \\
&= \int_a^b dt \frac{dL}{d\varepsilon}\big(\boldsymbol{x}(t), \dot{\boldsymbol{x}}(t), t\big), \\
&= \int_a^b dt \bigg[\boldsymbol{h}(t)\frac{\partial L}{\partial \boldsymbol{x}}\big(\boldsymbol{x}(t), \dot{\boldsymbol{x}}(t), t\big) + \dot{\boldsymbol{h}}(t)\frac{\partial L}{\partial \dot{\boldsymbol{x}}}\big(\boldsymbol{x}(t), \dot{\boldsymbol{x}}(t), t\big)\bigg], \\
&= \boldsymbol{h}(b)\frac{\partial L}{\partial \dot{\boldsymbol{x}}}\bigg|_{t=b} - \boldsymbol{h}(a)\frac{\partial L}{\partial \dot{\boldsymbol{x}}}\bigg|_{t=a} + \int_a^b dt\, \boldsymbol{h}\bigg(\frac{\partial L}{\partial \boldsymbol{x}} - \frac{d}{dt}\frac{\partial L}{\partial \dot{\boldsymbol{x}}}\bigg), \\
&= \int_a^b dt\, \boldsymbol{h}\bigg(\frac{\partial L}{\partial \boldsymbol{x}} - \frac{d}{dt}\frac{\partial L}{\partial \dot{\boldsymbol{x}}}\bigg).
\end{aligned}$$

We have integrated by parts and used the fact that $\boldsymbol{h}(a) = \boldsymbol{h}(b) = 0$. We conclude using the *Fundamental lemma of calculus of variations.* □

Bibliography

[AE99a] J.E. Avron and A. Elgart. Adiabatic Theorem without a Gap Condition. *Commun. Math. Phys.*, 203(2):445–463, 1999. arXiv:9805022, doi:10.1007/s002200050620. [1, 3.1.1, 3.1.4, 3.1.1, A, A.0.1]

[AE99b] J.E. Avron and A. Elgart. Adiabatic theorem without a gap condition. *Commun. Math. Phys.*, 1999. arXiv:9805022v4, doi:10.1007/s002200050620. [A]

[Amb00] A. Ambainis. Quantum lower bounds by quantum arguments. *Proc. thirty-second Annu. ACM*, 64(4):1–14, 2000. arXiv:0002066v1, doi:10.1006/jcss.2002.1826. [(document), 1, 4.3.2]

[AvDK$^+$07] D. Aharonov, W. van Dam, J. Kempe, Z. Landau, S. Lloyd, and O. Regev. Adiabatic Quantum Computation is Equivalent to Standard Quantum Computation. *SIAM J. Comput.*, 37(1):166–194, 2007. doi:10.1137/S0097539705447323. [3.2]

[BBC$^+$01] R. Beals, H. Buhrman, R. Cleve, M. Mosca, and R. de Wolf. Quantum lower bounds by polynomials. *J. ACM*, 48(4):778–797, 2001. arXiv:9802049, doi:10.1145/502090.502097. [4.3.1, 4.3.2]

[BBCR10] B. Barak, M. Braverman, X. Chen, and A. Rao. How to compress interactive communication. *Proc. 42nd ACM Symp. Theory Comput. - STOC '10*, page 67, 2010. doi:10.1145/1806689.1806701. [1]

[BdW02] H. Buhrman and R. de Wolf. Complexity measures and decision tree complexity: A survey. *Theor. Comput. Sci.*, 288(1):21–43, 2002. doi:10.1016/S0304-3975(01)00144-X. [4.0.1]

[Bel64] J.S. Bell. On the Einstein-Podolsky-Rosen paradox. *Physics (College. Park. Md).*, 1:195–200, 1964. [1]

[Bel15] A. Belovs. Variations on Quantum Adversary. *Unpublished*, page 33, 2015. arXiv:1504.06943. [4.1.2]

[BF28] M. Born and V. Fock. Beweis des Adiabatensatzes. *Zeitschrift für Phys.*, 51(3-4):165–180, 1928. doi:10.1007/BF01343193. [(document), 3.1.1, 7]

[BHMR03] H. Buhrman, P. Høyer, S. Massar, and H. Röhrig. Combinatorics and Quantum Nonlocality. *Phys. Rev. Lett.*, 91(4):047903, 2003. doi:10.1103/PhysRevLett.91.047903. [1]

[BR11] M. Braverman and A Rao. Information equals amortized communication. *2011 IEEE 52nd Annu. Symp.*, 2011. arXiv:1106.3595v1, doi:10.1109/TIT.2014.2347282. [1, 5.1.6, 9]

[BR14] M. Brandeho and J. Roland. A universal adiabatic quantum query algorithm. 2014. arXiv:1409.3558, doi:10.4230/LIPIcs.TQC.2015.163. []

[Bra12] M. Braverman. Interactive information complexity. *Proc. forty-fourth Annu. ACM Symp.*, 123(123), 2012. doi:10.1137/130938517. [1]

[BSS03] H. Barnum, M. Saks, and M. Szegedy. Quantum query complexity and semi-definite programming. *18th IEEE Annu. Conf. Comput. Complexity, 2003. Proceedings.*, pages 179–193, 2003. doi:10.1109/CCC.2003.1214419. [1, 4.2, 4.2.2]

[BV10] S. Boyd and L. Vandenberghe. *Convex Optimization*, volume 25. 2010. arXiv:1111.6189v1, doi:10.1080/10556781003625177. [2.6.1, 2.6.2, 6]

[BYJKS04] Z. Bar-Yossef, T.S. Jayram, R. Kumar, and D. Sivakumar. An information statistics approach to data stream and communication complexity. *J. Comput. Syst. Sci.*, 68(4):702–732, 2004. doi:10.1016/j.jcss.2003.11.006. [1]

[CG04a] A. Childs and J. Goldstone. Spatial search by quantum walk. *Phys. Rev. A*, 70(2):022314, 2004. arXiv:0306054v2, doi:10.1103/PhysRevA.70.022314. [3.2]

[CG04b] A.M. Childs and J. Goldstone. Spatial search and the Dirac equation. *Phys. Rev. A*, 70(4):042312, 2004. arXiv:0405120, doi:10.1103/PhysRevA.70.042312. [3.2]

[CGM09] R. Cleve, D. Gottesman, and M. Mosca. Efficient discrete-time simulations of continuous-time quantum query algorithms. *Proceedings of the forty-first annual ACM symposium on Theory of computing*, pages 1–12, 2009. arXiv:0811.4428v1, doi:10.1145/1536414.1536471. [1]

[Chi09] A.M. Childs. On the Relationship Between Continuous- and Discrete-Time Quantum Walk. *Commun. Math. Phys.*, 294(2):581–603, 2009. arXiv:0810.0312, doi:10.1007/s00220-009-0930-1. [3.2]

[CWY01] A. Chakrabarti, A. Wirth, and A. Yao. Informational complexity and the direct sum problem for simultaneous message complexity. In *Proc. 2001 IEEE Int. Conf. Clust. Comput.*, pages 270–278. IEEE Comput. Soc, 2001. doi:10.1109/SFCS.2001.959901. [1]

[Ehr] P. Ehrenfest. Bemerkung über die angenäherte gültigkeit der klassischen mechanik innerhalb der quantenmechanik. *Zeitschrift fur Phys.*, pages 455–457. doi:10.1007/BF01329203. [7, 7.1]

[Fey82] R.P. Feynman. Simulating physics with computers. *Int. J. Theor. Phys.*, 21(6-7):467–488, 1982. doi:10.1007/BF02650179. [3.2]

[FGG08] E. Farhi, J. Goldstone, and S. Gutmann. A Quantum Algorithm for the hamiltonian NAND Tree. *Theory Comput.*, 4:169–190, 2008. doi:10.4086/toc.2008.v004a008. [3.2]

[FGGS00] E. Farhi, J. Goldstone, S. Gutmann, and M. Sipser. Quantum Computation by Adiabatic Evolution. *Unpublished*, 2000. arXiv:0001106. [1, 3.2]

[FGT14] I. Foulger, S. Gnutzmann, and G. Tanner. Quantum Search on Graphene Lattices. *Phys. Rev. Lett.*, 112(7):070504, 2014. arXiv:1312.3852, doi:10.1103/PhysRevLett.112.070504. [3.2]

[FvdG99] C.A. Fuchs and J. van de Graaf. Cryptographic distinguishability measures for quantum-mechanical states. *Inf. Theory, IEEE*, (1):12–33, 1999. arXiv:9712042v2, doi:10.1109/18.761271. [2.1.4]

[GA11] S.I. Gass and A.A. Assad. In *Profiles in Operations Research*, pages 217–240. Springer, Boston, MA, 2011. doi:10.1007/978-1-4419-6281-2_13. [6.2.0.1]

[GPW15] Mika Goos, Toniann Pitassi, and Thomas Watson. Deterministic communication vs. partition number. In *Proceedings of the 2015 IEEE 56th Annual Symposium on Foundations of Computer Science (FOCS)*, FOCS '15, pages 1077–1088. IEEE Computer Society, 2015. doi:10.1109/FOCS.2015.70. [5.1]

[Gro96] L.K. Grover. A fast quantum mechanical algorithm for database search. In *Proc. twenty-eighth Annu. ACM Symp. Theory Comput. - STOC '96*, pages 212–219, New York, New York, USA, 1996. ACM Press. doi:10.1145/237814.237866. [1]

[HLS07] P. Hoyer, T. Lee, and R. Spalek. Negative weights make adversaries stronger. *Proc. thirty-ninth Annu. ACM*, pages 1–29, 2007. arXiv:0611054v2, doi:10.1145/1250790.1250867. [1, 4.3.2]

[JK10] R. Jain and H. Klauck. The partition bound for classical communication complexity and query complexity. *CCC '10 Proceedings of the 2010 IEEE 25th Annual Conference on Computational Complexity*, pages 1–28, 2010. arXiv:0910.4266v2, doi:10.1109/CCC.2010.31. [5.1.2, 5.1.3, 5.1.4, 5.1.3]

[JRS07] S. Jansen, M.-B. Ruskai, and R. Seiler. Bounds for the adiabatic approximation with applications to quantum computation. *J. Math. Phys.*, 48(10):102111, 2007. doi:10.1063/1.2798382. [3.1.1, 3.1.3]

[Kat50] T. Kato. On the Adiabatic Theorem of Quantum Mechanics. *J. Phys. Soc. Japan*, 5(6):435–439, 1950. doi:10.1143/JPSJ.5.435. [A, A.0.2]

[Kha80] L.G. Khachiyan. Polynomial algorithms in linear programming. *USSR Computational Mathematics and Mathematical Physics*, 20(1):53–72, 1980. doi:10.1016/0041-5553(80)90061-0. [6.2.0.2]

[KLL$^+$12] I. Kerenidis, S. Laplante, V. Lerays, J. Roland, and D. Xiao. Lower Bounds on Information Complexity via Zero-Communication Protocols and Applications. In *2012 IEEE 53rd Annu. Symp. Found. Comput. Sci.*, pages 500–509. IEEE, 2012. arXiv:1204.1505v2, doi:10.1109/FOCS.2012.68. [5.1.5, 5.3.1, 5.3.3, 5.3.4, 5.3.5, 5.3.6]

[KLMK07] P. Kaye, R. Laflamme, M. Mosca, and P.R. Kaye. *An Introduction to Quantum Computing*. 2007. doi:10.1007/978-94-007-0080-2_10. [1]

[KN97] E. Kushilevitz and N. Nisan. *Communication Complexity*. Cambridge University Press, New York, NY, USA, 1997. [5]

[LLR12] S. Laplante, V. Lerays, and J. Roland. Classical and quantum partition bound and detector inefficiency. *Autom. Lang. Program.*, pages 617–628, 2012. arXiv:1203.4155v1, doi:doi.org/10.1007/978-3-642-31594-7_52. [1, 5.2, 5.2.2, 5.4.1]

[LMR$^+$11] T. Lee, R. Mittal, B.W. Reichardt, R. Špalek, and M. Szegedy. Quantum query complexity of state conversion. In *Proc. - Annu. IEEE Symp. Found. Comput. Sci. FOCS*, pages 344–353, 2011. arXiv:1011.3020, doi:10.1109/FOCS.2011.75. [(document), 1, 1, 4.3.2, 4.3.2, 4.3.5, 4.3.2, 4.3.7, 7, 7.0.1, 9.6.3]

[LR12] T. Lee and J. Roland. A Strong Direct Product Theorem for Quantum Query Complexity. In *2012 IEEE 27th Conf. Comput. Complex.*, pages 236–246. IEEE, 2012. arXiv:1104.4468, doi:10.1109/CCC.2012.17. [2.1.4, 2.1.9, 4.1.3, 4.1.2, 4.1.3, 4.3.4, 4.3.3, 4.3.9, 4.3.10, 8]

[Mas01] S. Massar. Non locality, closing the detection loophole and communication complexity. *Phys. Rev. A*, 65(3):032121, 2001. arXiv:0109008, doi:10.1103/PhysRevA.65.032121. [1]

[Mir71] J.A. Mirrlees. An Exploration in the Theory of Optimum Income Taxation. *Rev. Econ. Stud.*, 38:175–208, 1971. doi:10.2307/2296779. [1, 6.8]

[Moc07] C. Mochon. Hamiltonian oracles. *Phys. Rev. A*, 75(4):042313, 2007. arXiv:0602032, doi:10.1103/PhysRevA.75.042313. [3.2]

[MR13] L. Magnin and J. Roland. Explicit relation between all lower bound techniques for quantum query complexity. *Int. J. Quantum Inf.*, 2013. arXiv:1209.2713v2, doi:10.1142/S0219749913500597. [4.3.3, 8]

[MS02] P. Milgrom and I. Segal. Envelope Theorems for Arbitrary Choice Sets. *Econometrica*, 70(2):583–601, 2002. doi:10.1111/1468-0262.00296. [1, 6.8, 6.8.1, C.0.1]

[NC11] M.A. Nielsen and I.L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition.* Cambridge University Press, New York, NY, USA, 10th edition, 2011. [3.2]

[RC02] J. Roland and N.J. Cerf. Quantum search by local adiabatic evolution. *Phys. Rev. A*, 2002. arXiv:0107015v1, doi:10.1103/PhysRevA.65.042308. [3.2]

[Rei09] B.W. Reichardt. Span Programs and Quantum Query Complexity: The General Adversary Bound Is Nearly Tight for Every Boolean Function. In *Found. Comput. Sci. 2009. FOCS'*, pages 544–551. IEEE, 2009. arXiv:0904.2759, doi:10.1109/FOCS.2009.55. [1]

[Rei11] B.W. Reichardt. Reflections for quantum query algorithms. In *Proceedings of the Twenty-second Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '11, pages 560–569. Society for Industrial and Applied Mathematics, 2011. [1, 3.2]

[RS75] M. Reed and B. Simon. *Methods of modern mathematical physics. 2. Fourier analysis, self-adjointness.* 1975. doi:B978-0-12-585001-8.X5001-6. [3.1.1]

[RS09] J. Roland and M. Szegedy. Amortized Communication Complexity of Distributions. In *Autom. Lang. Program.*, pages 738–749. 2009. doi:10.1007/978-3-642-02927-1_61. [1, 9.6.1, 9.6.1, 9.6.3]

[RS12] B.W. Reichardt and R. Spalek. Span-program-based quantum algorithm for evaluating formulas. *Theory Comput.*, 8(13):291–319, 2012. arXiv:0710.2630, doi:10.4086/toc.2012.v008a013. [1]

[Sha48] C.E. Shannon. A Mathematical Theory of Communication. *Bell Syst. Tech. J.*, 27:623–656, 1948. [1, 2.5]

[Sho97] P.W. Shor. Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer. *SIAM J. Comput.*, 26(5):1484–1509, 1997. arXiv:9508027v2, doi:10.1137/S0097539795293172. [1]

[Sla14] M. Slater. Lagrange multipliers revisited. In *Traces Emerg. Nonlinear Program.*, pages 293–306. 2014. doi:10.1007/978-3-0348-0439-4_14. [6.7.8, B.0.1]

[Špa08] R. Špalek. The Multiplicative Quantum Adversary. In *Proc. 23rd Annual IEEE Conference Computational Complexity*, pages 237–248. IEEE Computer Society, 2008. doi:10.1109/CCC.2008.9. [1, 4.3.3]

[ŠS05] R. Špalek and M. Szegedy. All quantum adversary methods are equivalent. *Autom. Lang. Program.*, 2:18, 2005. arXiv:0409116, doi:10.1007/11523468_105. [4.3.2]

[SSL17] T. Sutter, P.M. Sutter, D. Esfahani, and J. Lygeros. Generalized maximum entropy estimation. pages 1–16, 2017. arXiv:1708.07311, doi:TheStataJournal. [9.3.3]

[Sun15] R. Sundström. A Pedagogical History of Compactness. *Am. Math. Mon.*, 122(7):619, 2015. arXiv:1006.4131, doi:10.4169/amer.math.monthly.122.7.619. [2.2.1]

[SW73] D. Slepian and J. Wolf. Noiseless coding of correlated information sources. *IEEE Trans. Inf. Theory*, 19(4):471–480, 1973. doi:10.1109/TIT.1973.1055037. [1]

[vDMV02] W. van Dam, M. Mosca, and U. Vazirani. How powerful is adiabatic quantum computation? In *Proc. 2001 IEEE Int. Conf. Clust. Comput.*, pages 279–287. IEEE Comput. Soc, 2002. arXiv:0206003, doi:10.1109/SFCS.2001.959902. [3.2]

[Von28] J. Von Neumann. Zur Theorie der Gesellschaftsspiele. *Math. Ann.*, 100(1):295–320, 1928. doi:10.1007/BF01448847. [6.4]

[Yao77] A. Yao. Probabilistic computations: Toward a unified measure of complexity. In *18th Annu. Symp. Found. Comput. Sci. (sfcs 1977)*, pages 222–227. IEEE, 1977. doi:10.1109/SFCS.1977.24. [5.0.1]

[Yao79] A. Yao. Some complexity questions related to distributive computing. *Annu. ACM Symp. Theory Comput.*, page 209, 1979. doi:10.1145/800135.804414. [1, 5, 5.1.1]

[YM11] D. Yonge-Mallo. Adversary lower bounds in the Hamiltonian oracle model. *Unpublished*, pages 1–6, 2011. arXiv:1108.2479v1. [4.3.6, 7, 7.0.1, 7.1.1]

Notation

Important sets

$\mathbb{R}$	Real number.
$\mathbb{R}^n$	Real vector of length n.
$\mathbb{R}_+$	Positive real number.
$\mathbb{R}^n_+$	Positive real vector of lenght n.
$\mathbb{C}$	Complex number.
$\mathbb{C}^n$	Complex vector of length n.
$\mathbb{K}$	$\mathbb{R}$ or $\mathbb{C}$.
$\mathcal{M}_{n,m}(\mathbb{R})$	n-by-m matrices with entries in $\mathbb{K}$.
$\mathcal{M}_n(\mathbb{R})$	n-by-n matrices with entries in $\mathbb{K}$.
S^n	Hermitian n-by-n matrices.
S^n_+	Positive semi-definite n-by-n matrices.
$F(A,B)$	Function space.
$C^0(A,B)$	Continuous function space.
$C^k(A,B)$	Continuously k-differentiable function space.
$\emptyset$	The empty set.

Set theory

$\lvert A\rvert$	Cardinal of set A.
$A^{\complement}$	The complement of set$A$
$A \cap B$	Intersection of sets A, B.
$A \cup B$	Union of sets A, B.
$\mathcal{P}(A)$	Power set of set A.
$A \setminus B$	The relative complement of B in A.

Probability

$\langle X\rangle_p$	Average value of random vector X.
$\mathcal{D}(\cdot,\cdot)$	The total distance.
$\mathcal{F}(\cdot,\cdot)$	The fidelity.
$\lvert\cdot\rvert_{TV}$	The total variance.
$\mathbb{P}(S)$	Set of all probability distributions on S.
$\mathbb{B}(S)$	Set of all real functions on S.

Linear algebra

$\mathcal{I}d$	Identity matrixreal line.
$\mathbb{J}$	Matrix with all entries equal to one.
$(e_i)_i$	Canonic basis vectors .
$\delta[a,b]$	Kronecker delta.
X^T	Transpose of matrix X.
X^*	Conjugate transpose of matrix X.
$\mathrm{tr}X$	Trace of matrix X.
$\mathbf{rank}\ X$	Rank of matrix X.
$\Vert \cdot \Vert$	A norm.
$\Vert \cdot \Vert_{\mathrm{tr}}$	The trace norm.
$d(\cdot,\cdot)$	A distance.
$X \circ Y$	The Hadamard product.
$V \oplus W$	The direct sum between vector spaces V and W.
$V \otimes W$	The tensor product between vector spaces V and W.

Topology

$B_r(x)$	Closed ball of radius r and center x.
$\mathbf{cl}\ A$	Closure of set A.
$\mathbf{int}\ A$	Interior of set A.
$\mathbf{relint}\ A$	Relative interior of set A.
$\mathbf{bd}\ A$	Boundary of set A.
$\mathbf{aff}\ A$	Affine hull of set A.
$\mathbf{conv}\ A$	Convex hull of set A.

Analysis

$f : A \to B$	A function with domain include in A and range include in B.
$f : A \to \mathcal{P}(B)$	A multi-valued function with domain $\subset A$ and range $\subset B$.
$\mathbf{dom}\ f$	Domain of function f.
$\mathbf{rg}\ f$	Range of function f.
Df	Derivative of function f.
$\boldsymbol{\nabla} f$	Gradient of function f.

Convex and order

$x \leq y$	Inequality between reals x and y.
$\boldsymbol{x} \leq \boldsymbol{y}$	Composen-twise inequality between vectors $\boldsymbol{x}$ and $\boldsymbol{y}$.
$X \leq Y$	Loewner order between matrices X and Y.
$x \leq_K y$	Generalized inequality between x and y relative to proper cone K.