A fundamental bound to quantum information processing

Marco Robbio
Supervisor: Davide Girolami

Contents

1 Quantum information basics **4**
- 1.1 Quantum mechanics postulates . . . 4
- 1.2 POVM measurements . . . 7
- 1.3 The density operator . . . 7
- 1.4 The Schmidt decomposition and purifications . . . 9
- 1.5 CPTP maps . . . 11

2 Quantum Distances **12**
- 2.1 Classical Distances . . . 12
  - 2.1.1 Trace distance . . . 12
  - 2.1.2 Fidelity . . . 12
- 2.2 Quantum Trace Distance . . . 13
- 2.3 Fidelity . . . 14
- 2.4 N-particles distance . . . 15
  - 2.4.1 Weighted distance . . . 16
  - 2.4.2 Bures weighted length . . . 17

3 Quantum Bound to state transformation **19**
- 3.1 Quantum process dynamic . . . 19
- 3.2 Discrete case . . . 20
- 3.3 Continuous case . . . 22

4 Quantum state preparation complexity **24**
- 4.1 Quantumness . . . 24
- 4.2 Optimal path . . . 25
  - 4.2.1 Qubit example . . . 27
  - 4.2.2 2-Qubits example . . . 27
- 4.3 Algorithmic complexity . . . 28
  - 4.3.1 GHZ preparation . . . 29

5 Appendix A **31**
- 5.1 Trace distance . . . 31
- 5.2 Fidelity . . . 31
- 5.3 Weighted distance . . . 33

Abstract

The main result of this thesis is a geometric bound which allows us to quantify the viability of quantum algorithms, in a simple yet effective way. On top of that, it can be used to understand if a path from an input state to a target state is optimal or not, allowing us to optimize the programming strategies used nowadays in quantum computing.

Quantum computers promise a huge improvement on our ability to solve hard and complex problems, but the technical challenges they present need new methodologies to deal with them. In this thesis we will focus our attention on different methodologies to bound the experimental costs (the energetic cost and the size) of quantum preparation algorithms. The approach we are going to focus on is a purely geometrical one, totally independent from the actual realization of the system since it is derived from the quantum mechanics postulates, thus, completely general.

The first part of our discussion will review the quantum mechanics formalism and geometric properties that can be defined on the state-space. In detail, the main topic of discussion is focused on the definition of distance (or metric) between states. The concept of metric defined on a space (the euclidean space or the Hilbert one) has as main purpose to help us understand how far the objects of the space are one from the other. Even if the mathematical world is full of different definitions of distances, each one has its well defined properties and usefulness. It turns out that the one defined in quantum mechanics struggles with N-particles systems, which are key ingredients when we deal with quantum computations for obvious reasons. To avoid such, not trivial, detail we employ a newly introduced metric called Weighted distance, which can be easily proven to be way more informative that its un-weighted counter part. The behaviour of the weighted distance is way more close to the concept of describing how far objects are the one from the other we stated before, making it a useful tool.

Once the weighted distance has been defined, we list its properties and discuss how to use it to bound the experimental cost of a transformation. When we talk about transformations, one must remember that they can always be divided into two components: a classical part and a quantum part. Since the classical part can be carried out in a well known way, we forget about it and suppose it costless, focusing on the quantum part alone. The first step is to define clearly the distance between the input and output state. It turns out that a good choice is the Bures length. From this result, we first discuss the discrete case in which we implement a generic unitary as a series of quantum gates (each involving a given number of qubits). It can be proven that the experimental cost in terms of quantum resources (taking into account energy cost, dimension of the gates and time needed for the transformation) is bounded from below by the, previously introduced, weighted Bures length. Once proved for discrete time transformations, we generalized the result to continuous time transformations obtaining a more general result. The bound we obtained is very powerful because: it is an exact bound; it works both for pure and mixed states; it has been obtained from the quantum mechanics postulates so it is independent on the actual realization of the system. For-

mally such bound can also be reinterpreted as: the ability to perform a transformation from a given input state to a different output state is never greater than the ability to distinguish them.

Last topic of our discussion is devoted to a closely related concept, the quantumness of a process. As said above, every process can be decoupled in a classical component and in a quantum component. The ability to quantify the energy cost of the latter can be used to optimize quantum algorithms, finding the optimal path from a given input state to a target state. The quantumness can be defined as the minimum energetic cost to perform a transformation. Since the quantumness it is obtained minimizing the quantum energetic cost, it can be find minimizing it on the set of all possible isospectral free states and all possible path from the input to the output. Even if this procedure seems very complex through geometrical properties we are able to bound the quantumness, as we did for the experimental cost. Once obtained such bound, we show a simple application of such result to find the optimal path. In detail, we focus on the best path from a state $(a|0\rangle + b|1\rangle)|0\rangle^{\otimes N}$ to an high entangled state restricting the dimension of the gates we use for the transformation.

1 Quantum information basics

> Quantum mechanics: Real Black Magic Calculus
>
> *Albert Einstein*

In this chapter we will focus on the basics of quantum mechanics and the useful tools for the later development of the subsequent topics.

1.1 Quantum mechanics postulates

Before the formal statements of the postulates, it is important to remember that quantum mechanics is a mathematical framework necessary to the development and the understanding of physical systems. The postulates below provide a connection between the laws which govern a system behavior and the math behind them[1].

Postulate 1: Any isolated physical system is associated to a complex vector space with an inner product known as state space. The system is completely described by its state vector, which is a unit vector in the system's state space.

Quantum mechanics does not state which is the state space or the space vector. Finding them usually is a very hard task. For the following discussion it is not really needed the actual computation of them, so we will focus on the most general possible framework and stick with our assumptions.
The simplest quantum system, and the most interesting one for quantum computation, is the qubit. A qubit state has two dimensional state space where for simplicity we identify $|0\rangle, |1\rangle$ as the element of the orthonormal basis. It follows that any qubit can be written as linear combination of the two:

$$|\Psi\rangle = a|0\rangle + b|1\rangle$$

Since $|\Psi\rangle$ must be a unit vector, we have that $|a|^2 + |b|^2 = 1$. The qubit, we just described, resembles a classical bit of information but there is a key difference between the two: a classical bit can only exists in one of the two possible states 0 or 1. On the other hand its quantum counter part can exist in a superposition of the two states. The states' superposition is one of the biggest differences between a classical environment and a quantum one.

Postulate 2: The evolution of a closed system is described by a unitary transformation. That is, if $|\Psi\rangle$ is the state of the system at time t_2 and $|\Psi'\rangle$ is the state of the system at time t_1, then

$$|\Psi\rangle = U|\Psi'\rangle \text{ with U only depending on } t_2 - t_1.$$

As before quantum mechanics does not state how U must be computed, the only important information is that U must be unitary. It is important for the later discussion to better explain a little detail about the second postulate. All closed system can be

thought as a system, on which we actually perform experiments and measurements, and the environment in which it is embedded. The postulate states clearly that a closed system will evolve through a unitary transformation U, but this does not imply that its components will do the same. So in a general scenario, we will have that the system under study will not evolve through a unitary transformation because it is coupled with the environment. The second postulates seems a bit difficult to apply when we consider that it is limited to closed systems which in practice can not be realized. Nevertheless in the quantum computation's environment, and not only, we can manage to approximate many physical systems to closed ones, which allows to simplify the description of the system under study.
The second postulate has also an alternative formulation which extend our knowledge on the dynamical evolution of our system.

Postulate 2(bis): The evolution of the state of a quantum system is described by the Schrodinger equation

$$i\hbar\frac{d}{dt}|\Psi\rangle = H|\Psi\rangle$$

In the previous equation we introduced $\hbar$, which is the Planck's constant (for our purposes will be set to 1, since it is a constant parameter which can be absorbed in H) and H which is the Hamiltonian of our system. H is an hermitian operator. The postulate does not differ from its previous form when we talk about closed systems since H can only be written for closed systems. Solving the equation is not simple, but once it is done we have a dynamical formulation for $|\Psi\rangle$.
To see the connection between the two statements we can simply verify that:

$$|\Psi(t_2)\rangle = e^{\frac{-iH(t_2-t_1)}{\hbar}}|\Psi(t_1)\rangle = U(t_2,t_1)|\Psi(t_1)\rangle$$

is a solution of the Schrodinger equation in the case H is time independent.
The matrix $U(t_2,t_1)$ is called time propagator and it satisfies the following equation

$$i\hbar\frac{d}{dt}U = H(t)U \quad \text{with } U(t_1,t_1) = I$$

where a general solution can be written as

$$U(t_2,t_1) = e^{-\frac{i}{\hbar}\int_{t_1}^{t_2} H(t')dt'}$$

The two versions of the second postulate coincide since it easy to prove that U is a unitary operator given that H is hermitian. Up to now we have stated the formal definition of our state vector and its evolution in the case the system does not interact with the outside world. We can create systems which are very similar to a close system but for a series of purposes it will be necessary to measure some of their properties at some point. In that case we can no longer treat our system as a close one. In this cases we need another postulate which clarifies the behavior of our system in response to a measure.

Postulate 3: quantum measurements are described by a collection $\{M_m\}$ of measurement operators. These are operator acting on the state space of the system

being measured. The index m refers to the measurement outcomes that may occur in the experiment. If the state of the quantum system is $|\Psi\rangle$ immediately before the measurement, then the probability that result m occurs is

$$p(m) = \langle\Psi|M_m^\dagger M_m|\Psi\rangle$$

and after the measurement the state of the system will be $|\Psi'\rangle$

$$|\Psi'\rangle = \frac{M_m|\Psi\rangle}{\sqrt{\langle\Psi|M_m^\dagger M_m|\Psi\rangle}}$$

The measurement operators satisfy completeness relation

$$\sum_m M_m^\dagger M_m = I$$

The completeness relation is just a consequence of the normalization of the probability, since

$$\sum_m M_m^\dagger M_m = I \iff 1 = \sum_m p(m) = \sum_m \langle\Psi|M_m^\dagger M_m|\Psi\rangle$$

This postulate simply states that measuring a property of our system has a direct effect on it. Let us make an example on a practical system like a qubit. As said before we can write a qubit as:

$$|\Psi\rangle = a|0\rangle + b|1\rangle$$

In the moment we perform a measure on the state of our system we modify it making it or $|0\rangle$ or $|1\rangle$. This behavior by itself it is peculiar of quantum mechanics. As said before, quantum superposition is a very fundamental concept in quantum computation, and it can be simply destroyed by performing a measurement on the system.
The last postulate deals with a very interesting topic. Up to now we have only considered one system at the time, but in general we have to study multiple system at the time. We need a mathematical tool to describe them effectively.

Postulate 4: The state space of a composite system is the tensor product of the state spaces of the components of such system.

Let us make an example to better understand the hidden meaning of the postulate. Suppose we have two systems described by the Hilbert spaces

$$H_1 : \{|\Psi_1\rangle, ..., |\Psi_N\rangle\} \text{ and } H_2 : \{|\Phi_1\rangle, ..., |\Phi_M\rangle\}$$

then the Hilbert space that describes the joint system is written as

$$H = \left\{ |\Psi_i\rangle \otimes |\Phi_j\rangle : \ i \in \{1, ...N, \}, j \in \{1, ..., M\} \right\}$$

It follows that a generic pure state that describe the joint system can always be written as:

$$|\Psi\rangle = \sum_{i=1}^{N}\sum_{j=1}^{M} c_{ij}|\Psi_i\rangle|\Phi_j\rangle \ \text{ such that } \sum_{i=1}^{N}\sum_{j=1}^{M} |c_{ij}|^2 = 1$$

1.2 POVM measurements

Postulate 3 gives a rule about measurements' statistics and the state of the system after we observe it. However, for some specific cases information about the post-measurement state of the system is very interesting. We are going to discuss now a mathematical tool called POVM (povitive operator valued measure) formalism, which is a very elegant way to approach such a problem.
Suppose a measurement described by M_m is performed upon a quantum system in the state $|\Psi\rangle$. From postulate 3 we know that the outcome m of a measurement has a probability $p(m) = \langle\Psi|M_m^\dagger M_m|\Psi\rangle$. Let us define now $E_m = M_m^\dagger M_m$ then we have

$$p(m) = \langle\Psi|E_m|\Psi\rangle \quad \text{and} \quad \sum_m E_m = I$$

E_m are positive operators and by definition are sufficient to determine the probabilities of different measurement outcomes. The operators E_m are known as the POVM elements associated to the measurement and the set $\{E_m\}$ is known as POVM[1].
As an example, let us consider the projective measurement operators P_m, where we have that

$$P_m P_{m'} = \delta_{m,m'} P_m \quad \text{and} \quad \sum_m P_m = I$$

In this peculiar case, all the POVM elements are the same measurement operator themselves, since $E_m = P_m^\dagger P_m = P_m$.
It is possible to show that there is a set $\{M_m\}$ of measurement operators defining a measurement described by the POVM $\{E_m\}$. Define $M_m = \sqrt{E_m}$ we have that

$$\sum_m M_m^\dagger M_m = \sum_m E_m = I$$

and therefore the set M_m describes a measurement with POVM $\{E_m\}$. For this reason it is convenient to define a POVM to be any set of operators $\{E_m\}$ such that:

- each operator E_m is positive
- the set $\{E_m\}$ satisfies the completeness relation

POVM are really powerful tool when we are interested in distinguish different states only from measurement statistics, that is why we are going to make use of them in the following discussion.

1.3 The density operator

Up to now we have discussed all the postulates of quantum mechanics focusing only on the state vector formalism. There is an equivalent version which is based on the so called density matrix operator. It is important to state that both descriptions are equivalent, but the density matrix operator provides a more general mathematical tool to deal with than the other.
Suppose a quantum system is in one state $|\Psi_i\rangle$ with $i \in \{1, ..., n\}$ with respective

probabilities p_i. The set $\{p_i, |\Psi_i\rangle\}$ is an ensemble of pure states. The density operator is defined as :

$$\rho = \sum_i p_i |\Psi_i\rangle\langle\Psi_i|$$

The postulates of quantum mechanics can be reformulated in terms of the density operator matrix.
If the time evolution of a system is described by a unitary transformation U such that: $|\Psi_i\rangle \implies U|\Psi_i\rangle$ then we have that

$$\rho = \sum_i p_i |\Psi_i\rangle\langle\Psi_i| \implies U\rho U^\dagger = \sum_i p_i U|\Psi_i\rangle\langle\Psi_i|U^\dagger$$

Measurements are also easily described in the density matrix formalism. Suppose we perform a measurement described by a measurement operators M_m. If the initial state was $|\Psi_i\rangle$, then the probability of getting m as a result is:

$$p(m|i) = \langle\Psi_i|M_m^\dagger M_m|\Psi_i\rangle = Tr(M_m^\dagger M_m|\Psi_i\rangle\langle\Psi_i|)$$

By simply applying the laws of probability we have that :

$$p(m) = \sum_i p(m|i) = \sum_i Tr(M_m^\dagger M_m|\Psi_i\rangle\langle\Psi_i|) = Tr(M_m^\dagger M_m \rho)$$

After the measurement the state of the system will simply be

$$|\Psi_i^m\rangle = \frac{M_m|\Psi_i\rangle}{\sqrt{\langle\Psi_i|M_m^\dagger M_m|\Psi_i\rangle}}$$

So in terms of our density operators we will have:

$$\rho_m = \sum_i p(m|i)|\Psi_i^m\rangle\langle\Psi_i^m| = \sum_i p_i \frac{M_m|\Psi_i\rangle\langle\Psi_i|M_m^\dagger}{\langle\Psi_i|M_m^\dagger M_m|\Psi_i\rangle} = \frac{M_m\rho M_m^\dagger}{Tr(M_m^\dagger M_m\rho)}$$

So the first three postulates can be rewritten also in this formalism, and also the fourth one, but before doing it, we need to clarify a bit the jargon.

A quantum system whose state $|\Psi\rangle$ is known exactly is said to be a **pure state**. In this case $\rho = |\Psi\rangle\langle\Psi|$. If this is not the case, we call it a **mixed state**. Suppose we now prepare a system in the state ρ_i with probability p_i. The density matrix will have the form $\sum_i p_i\rho_i$ and we say that ρ is a mixture of the states ρ_i with probabilities p_i.
The density matrix ρ satisfies the following properties:

- $Tr(\rho) = 1$;
- ρ is Hermitian;
- ρ is positive semi-definite.

$$|\Psi\rangle = \begin{pmatrix} c_1 \\ c_2 \\ \dots \\ c_n \end{pmatrix} \iff \rho = |\Psi\rangle\langle\Psi| = \begin{pmatrix} |c_1|^2 & c_2^* c_1 & \dots & c_n^* c_1 \\ c_1^* c_2 & |c_2|^2 & \dots & c_n^* c_2 \\ \dots & \dots & \dots & \dots \\ c_1^* c_n & c_2^* c_n & \dots & |c_n|^2 \end{pmatrix}$$

Figure 1: It is important to remember that the two approaches describe both fully the system and there is no additional information hidden in the matrix formalism. As shown above, once we choose a basis for the vector space and find the coefficients $c_\alpha \in \mathbb{C}$, such that they satisfy the normalization condition, we can derive the density matrix and viceversa.

There is also another property which is interesting to recall : for pure state $Tr(\rho^2) = 1$ while for mixed state $Tr(\rho^2) \leq 1$ (it is saturated when a pure state is written a trivial mixed state).

Last concept we will introduce about density matrix is the **reduced density matrix**. This tool is fundamental to the study of composite systems.
Suppose we have two physical systems A and B, whose composite state is described by ρ_{AB}. Then the reduced density matrix for system A is defined as:

$$\rho_A = Tr_B(\rho_{AB})$$

where we introduced $Tr_B(\cdot)$ which is the partial trace over state B and is defined as:

$$Tr_B(|a_1\rangle\langle a_2| \otimes |b_1\rangle\langle b_2|) = |a_1\rangle\langle a_2| \cdot Tr(|b_1\rangle\langle b_2|)$$

with $|a_1\rangle, |a_2\rangle$ state vector of A and $|b_1\rangle, |b_2\rangle$ state vector of B.
As defined, it is not obvious that the reduced matrix ρ_A is a good description for A, but it can be proven that it provides the correct measurement statistics for measurements made on system A[1].

1.4 The Schmidt decomposition and purifications

Theorem: Suppose $|\Psi\rangle$ is a pure state of a composite system AB. Then there exist orthonormal states $|i_A\rangle$ for system A and orthonormal states $|i_B\rangle$ for system B s.t. $|\Psi\rangle = \sum_i \lambda_i |i_A\rangle|i_B\rangle$ with $\sum_i \lambda_i^2 = 1$ where the λ_i are called Schmidt coefficients

The above theorem is called Schmidt's decomposition theorem and by itself provides a wide range of powerful result. For example let $|\Psi\rangle$ be state vector of the composite system AB. Then we can rewrite the two reduced density matrices as follows:

$$\rho_A = \sum_i \lambda_i |i_A\rangle\langle i_A| \quad \text{and} \quad \rho_B = \sum_i \lambda_i |i_B\rangle\langle i_B|$$

So we have that the eigenvalue of A and B are identical, and as we know, eigenvalues determine many property of a physical system. This implies that if a composite system

is described by a pure state then both A and B must share some properties.
The two bases $|i_A\rangle, |i_B\rangle$ are called **Schmidt bases** for A and B respectively and the number of non zero λ_i is called **Schmidt number** for $|\Psi\rangle$. The Schmidt's number is a very important property of a composite system which allows us to quantify the entanglement between the two[1].
Another vital concept related to quantum computation are **purifications**. Suppose we introduce a state ρ_A of a quantum system A. It is possible to introduce another system, we denote it by R, and define a pure state $|AR\rangle$ for the joint system AR such that $\rho_A = Tr_R(|AR\rangle\langle AR|)$.
The above procedure is pure math, which allows us to associate pure states to mixed states. It is known as purification. We will call the system R the reference system without any physical meaning. It can be proven that purification can be done for any state thanks to Schmidt's theorem. Suppose that A can be written as $\rho_A = \sum_i p_i |i_A\rangle\langle i_A|$. To purify it we introduce a system R and define the pure state of the combined system

$$|AR\rangle = \sum_i \sqrt{p_i}|i_A\rangle|i_R\rangle$$

We can now evaluate the reduced density matrix for system A corresponding to the state $|AR\rangle$:

$$Tr_R(|AR\rangle\langle AR|) = \sum_{i,j} \sqrt{p_i p_j}|i_A\rangle\langle j_A| \cdot Tr(|i_R\rangle\langle j_R|) = \sum_{i,j} \sqrt{p_i p_j}|i_A\rangle\langle j_A|\delta_{i,j} = \rho_A$$

So $|AR\rangle$ is a purification of ρ_A.
It must be noticed the closed relation between the two mathematical tools we introduced. Purifying a mixed state of a system A consists of defining a pure state whose Schmidt's basis for system A is just the basis in which the mixed state is diagonal, with the Schmidt's coefficients being the square root of the eigenvalues of the density operator being purified.
It is interesting to notice a little detail about purification and mixed state. When we gave the definition of mixed state it seemed obvious to think that to describe the system A, we had an ensemble of states $\{|\Psi_i\rangle\}$, in which every states had its probability to be generated $\{p_i\}$. This allowed us to write $\rho = \sum_i p_i |\Psi_i\rangle\langle\Psi_i|$. The problem can also be seen from a different prospective using the Schmidt's decomposition. Let us consider a purification of A, such that:

$$|AB\rangle = \sum_i \sqrt{p_i}|\Psi_i\rangle|\Phi_i\rangle$$

Performing a measurement on system B will generate the same ensemble we started from in our discussion, even if a purification is purely a mathematical tool. Or in other words, taking the partial trace over the system B will generate the same ensemble:

$$Tr_B\Big(|AB\rangle\langle AB|\Big) = \sum_{i,j} \sqrt{p_i p_j}|\Psi_i\rangle\langle\Psi_j| Tr_B\Big(|\Phi_i\rangle\langle\Phi_j|\Big) = \sum_i p_i |\Psi_i\rangle\langle\Psi_i| = \rho$$

This allows us to have a way clearer interpretation of what a purification really stand for and how to think about it.

1.5 CPTP maps

When we defined the density matrix operator and its properties, we stated that is semi-definite positive and its trace is unitary. So it is obvious that any quantum transformation should preserve these properties. Such transformations are called positive and trace preserving or PTP maps. But it is not true that all the PTP maps are valid quantum operations. This can be easily proven with an example.
Let us consider two subsystem such that the joint system $|\Omega\rangle = \frac{1}{\sqrt{d}} \sum_i |i\rangle|i\rangle$ (with d a normalization factor) and let us apply the so called partial transpose $I \otimes T$ (we apply the identity operator on the first subsystem and transpose the second).

$$(I \otimes T)|\Omega\rangle\langle\Omega| = \frac{1}{d} \sum_{i,j} |i\rangle\langle j| \otimes |j\rangle\langle i|$$

The resulting matrix is not positive semi-definite, which implies that it is not a density matrix. We need to impose a further constrain on the set of possible operations. We need to enforce not only that the operation is positive by itself but also that the result is a positive matrix. This set of operation is called **CPTP maps (completely positive trace preserving maps)** and by definition all the quantum operations belong to this class of mapping.

2 Quantum Distances

> Quantum theory was split up into dialects. Different people describe the same experiences in remarkably different languages. This is confusing even to physicists.
>
> *David Finkelstein*

In this section we are going to introduce some really useful tools needed for the rest of the discussion. Many of them require a formal mathematical demonstration which will be given in the Appendix A. Distance is a numerical or occasionally qualitative measurement of how far apart objects or points are. In math this is an abstract concept which bases its existence on formal definition which makes it not unique. Once we state the formal definition of distance we are going to use, it will be clear that each definition has its clear meaning and well defined property which can (as we will show later) go beyond the geometry definition we give.

2.1 Classical Distances

Before introducing the correct way to measure distances in a quantum environment, it is useful to recall some classical methods.

2.1.1 Trace distance

Given p and q two probability distributions defined on the same subset X, then the trace distance $D(p,q)$ is defined as:

$$D(p,q) = \frac{1}{2}\sum_{x\in X}|p_x - q_x|$$

This quantity is also known as L_1 Kolmogorov distance and it is a proper metric, so it satisfies the usual properties (it is symmetric in its arguments, it is zero if and only if p=q and it satisfies the triangular inequality).
The trace distance has a clearly defined physical interpretation: it can be proven that

$$D(p,q) = \frac{1}{2}\max_S |p(S) - q(S)| = \frac{1}{2}\max_S \left|\sum_{x\in S} p_x - \sum_{x\in S} q_x\right|$$

where S is a subset of X. So the trace distance can be interpreted as the maximum distance between the probability that the event S occurs if it is drawn from the probability distribution p and the probability that the same event occurs if drawn from the probability distribution q[1].

2.1.2 Fidelity

Given p and q two probability distributions defined on the same subset X, then the fidelity $D(p,q)$ is defined as:

$$F(p,q) = \sum_{x\in X}\sqrt{p_x q_x}$$

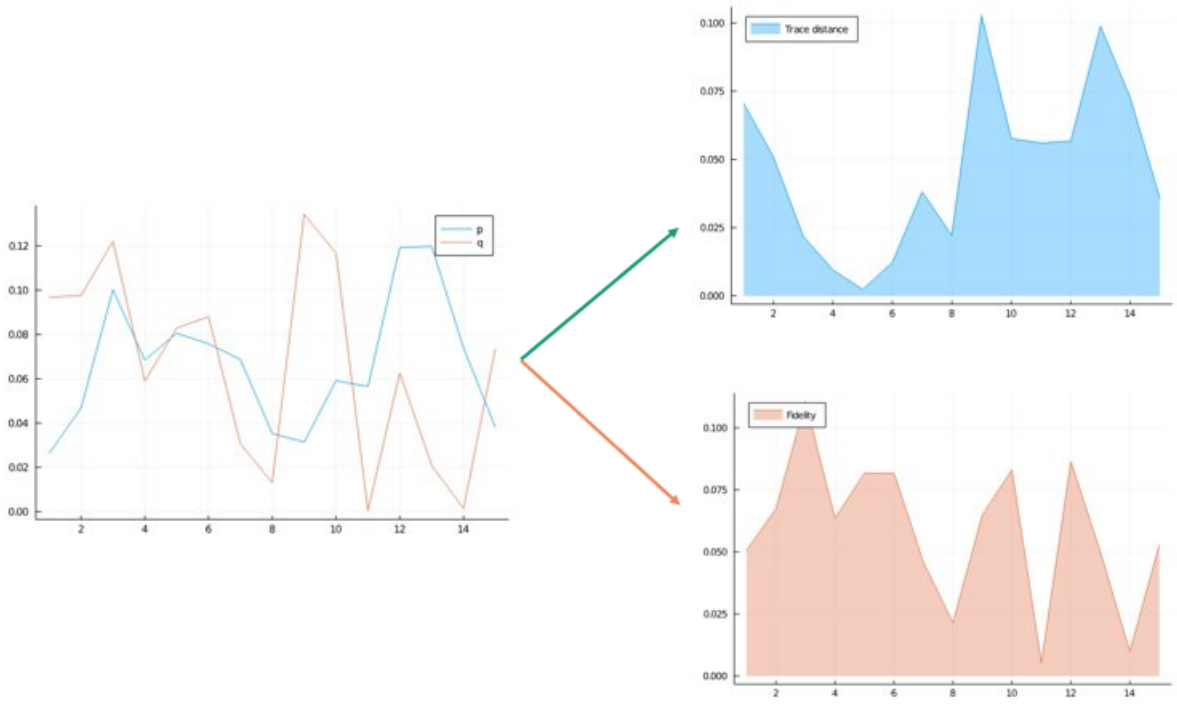

Figure 2: Given two probability distributions p and q (randomly generated in the image above), we can clearly see that the fidelity and the trace distance behave in different way. The choice of one with respect to the other is connected with the properties they have, as we will see in the next chapter.

This quantity is not a metric although later we will define a metric closely related to the fidelity.

2.2 Quantum Trace Distance

The quantum trace distance is a generalization of its classical counterpart and shares with it some nice properties. Before introducing it, we define $|A| = \sqrt{A^\dagger A}$ and let ρ and σ be two density matrices. Then the quantum trace distance between ρ and σ is defined as[1]

$$D(\rho, \sigma) = \frac{1}{2} Tr|\rho - \sigma|$$

A peculiar case is the one in which ρ and σ commute:

$$\rho = \sum_x p_x |x\rangle\langle x| \qquad \sigma = \sum_x q_x |x\rangle\langle x|$$

So we have that the quantum distance coincide with its classical counter part :

$$D(\rho, \sigma) = \frac{1}{2} Tr \left| \sum_x (p_x - q_x)|x\rangle\langle x| \right| = \frac{1}{2} \sum_x \left| p_x - q_x \right| = D(p, q)$$

Even if it is an abstract concept in the simplest case of a quantum bit, we can use the Bloch sphere representation to grasp a deeper understanding of the properties of D.

Given ρ_1 and ρ_2 two density matrices and $\vec{\sigma}$ the usual Pauli matrices vector, we can write

$$\rho_1 = \frac{I + \vec{r_1} \cdot \vec{\sigma}}{2} \qquad \rho_2 = \frac{I + \vec{r_2} \cdot \vec{\sigma}}{2} \implies$$

$$D(\rho_1, \rho_2) = \frac{1}{2} Tr \left| \rho_1 - \rho_2 \right| = \frac{1}{4} Tr \left| (\vec{r_1} - \vec{r_2}) \cdot \vec{\sigma} \right| = \frac{|\vec{r_1} - \vec{r_2}|}{2}$$

where we use the fact that $\vec{\sigma}$ has two eigenvalues ± 1.
The result states that the trace distance between two single qubits is half the Euclidean distance on the Bloch sphere[1].
Once shown the intuitive geometric representation of the trace distance we can deal with its properties:

- $D(\rho, \sigma) = D(U\rho U^\dagger, U\sigma U^\dagger)$
  The trace distance is invariant under unitary transformation

- Let ϵ be a trace preserving quantum operation, then

$$D(\epsilon(\rho), \epsilon(\sigma)) \leq D(\rho, \sigma)$$

- Strong convexity of the trace distance

$$D\left(\sum_i p_i \rho_i, \sum_i q_i \sigma_i \right) \leq D(p, q) + \sum_i p_i D(\rho_i, \sigma_i)$$

- Given P a positive operator

$$D(\rho, \sigma) = \max_P Tr(P(\rho - \sigma))$$

  This is a generalization of the classical case. We can look at the result in another light recalling that POVMs are positive operators so we have that: The trace distance is equal to the difference in probability that measuring the POVMs element P may occur depending on whether the state is ρ or σ and maximized over all P[1].

2.3 Fidelity

The fidelity of the state ρ and the state σ is defined as[1]

$$F(\rho, \sigma) = Tr\left(\sqrt{\rho^{\frac{1}{2}} \sigma \rho^{\frac{1}{2}}} \right)$$

There are two interesting cases

- $\rho = \sum_i r_i |i\rangle\langle i|$, $\sigma = \sum_i s_i |i\rangle\langle i|$ (ρ and σ commute) $\implies F(\rho, \sigma) = \sum_i \sqrt{r_i s_i} = F(r, s)$

- Given $|\Psi\rangle$ a pure state and an arbitrary density matrix ρ, we have

$$F(|\Psi\rangle, \rho) = Tr \sqrt{\langle\Psi|\rho|\Psi\rangle\langle\Psi|\Psi\rangle} = Tr \sqrt{\langle\Psi|\rho|\Psi\rangle}$$

The fidelity as its classical counter part satisfies some useful properties:

- $F(U\rho U^\dagger, U\sigma U^\dagger) = F(\rho, \sigma)$
  The fidelity is invariant under unitary transformation of its arguments
- Ulhmann's Theorem: Suppose ρ and σ are states of a quantum system Q. Introduce a quantum system R which is a copy of Q . Then

  $$F(\rho, \sigma) = \max_{|\Psi\rangle, |\Phi\rangle} |\langle\Psi|\Phi\rangle|$$

  where $|\Psi\rangle$ and $|\Phi\rangle$ are the purifications of ρ and σ from R into Q.
- Strong concavity

  $$F\left(\sum_i p_i\rho_i, \sum_i q_i\sigma_i\right) \geq \sum_i \sqrt{p_i q_i} F(\rho_i, \sigma_i)$$

- Suppose ϵ is a trace preserving quantum operation. Let ρ and σ be density matrices. Then

  $$F(\epsilon(\rho), \epsilon(\sigma)) \geq F(\rho, \sigma)$$

  This is also called the monotonicity of the fidelity.

As already told the fidelity is not a metric but we can turn into one using a geometric arguments. From Ulhmann's theorem we derive an interesting information: the fidelity is the maximum inner product between purifications of those states. This suggests that we can define an angle between ρ and σ such as

$$B(\rho, \sigma) = arcos(F(\rho, \sigma))$$

B satisfies symmetry of the inputs, non negativity and $B(\rho, \sigma) = 0 \iff \rho = \sigma$. The last property needed to define a metric is the triangular inequality, which can be proven using Ulhmann's Theorem. This is also called **Bures length (or Bures angle)** and we will see later an improved version of it. It is important, since it is often used in literature, that the Bures length is different from the Bures distance. Given two density matrices ρ and σ, the Bures distance is defined as:

$$D(\rho, \sigma) = 2\left(1 - \sqrt{F(\rho, \sigma)}\right)$$

It can be proven that is a proper metric (as the Bures length).

2.4 N-particles distance

Up to now we have only dealt with the formal definition of distances in a quantum environment. It is important to state that we never stated any property of our system, so the definitions are totally general. It is necessary to introduce a little example to better understand why this distances (even if formally correct) are not fully informative. Let us consider for example the case of three different states:

$$|\Psi_1\rangle = a|0\rangle^{\otimes N} + b|1\rangle^{\otimes N} \quad |\Psi_2\rangle = |0\rangle^{\otimes N} \quad |\Psi_3\rangle = |1\rangle|0\rangle^{\otimes N-1}$$

From the practical point of view it seems obvious that $|\Psi_2\rangle$ and $|\Psi_3\rangle$ are similar states, but our definition of distances will return the maximum possible distance between them, since their overlap is zero (independently on the system size N). The theory we described up to now claims that $|\Psi_1\rangle$ is closer to $|\Psi_2\rangle$ and $|\Psi_3\rangle$ than they are one to the other. So we need to introduce a new mathematical tool to distinguish N-particles quantum systems.

2.4.1 Weighted distance

In this section, it will be developed a theory which is fully general and does not depend on the choice of the actual measure d_{cl} we are going to use in the following sections.
Let ρ_N and σ_N be arbitrary density matrices of a N-particles quantum systems and suppose we can perform all the possible POVMs on the system such that

$$M = \left\{ M_i \geq 0 : \sum_i M_i = I_N \right\}$$

and let $\tilde{M} = \{\tilde{M}_i\}$ the set of most informative measures. Then we define

$$d(\rho_N, \sigma_N) = \max_M \sum_i d_{cl}\Big(Tr[M_i \rho_N], Tr[M_i \sigma_N] \Big) = \sum_i d_{cl}\Big(Tr[\tilde{M}_i \rho_N], Tr[\tilde{M}_i \sigma_N] \Big)$$

with d_{cl} a classical statistical distance.
Given our definition of weighted distance, it is to prove that it inherits all the properties of the unweighted one:

- Non-negativity: $d(\rho_N, \sigma_N) \geq 0$
- Faithfulness: $\rho_N = \sigma_N \iff d(\rho_N, \sigma_N) = 0$
- Contractivity: $d(\rho_N, \sigma_N) \geq d(\Lambda(\rho_N), \Lambda(\sigma_N))$ with Λ a trace preserving quantum operation
- Triangular inequality: $d(\rho_N, \sigma_N) \leq d(\rho_N, \tau_N) + d(\tau_N, \sigma_N)$

We ask furthermore that our distance is normalized

$$Tr(\rho_N \sigma_N) = 0 \iff d(\rho_N, \sigma_N) = M_d$$

Which means that it exists a value M_d such that our distance will always be lower than it, so it has a maximum.
The contractivity under trace preserving operations implies that $d(\rho_N, \sigma_N) \geq d(\rho_k, \sigma_k)$ in which ρ_k and σ_k are the states of $k \leq N$ particles. So in other words, since the distance is not increasing under partial trace, the ability to extract information from a quantum system depends on the size of the measurement setup[2].
Let us now consider a general scenario, in which we define a set of cooperating observers that want to discriminate between ρ_N and σ_N. Each of them performs the optimal

measurements $\tilde{M}^{k_\alpha}$ to discriminate ρ_{k_α} and σ_{k_α} of subsystems composed by $k_\alpha \leq N$ particles, then computing $d(\rho_{k_\alpha}, \sigma_{k_\alpha})$. The setup defines a partition P_{k_α} such that

$$P_{k_\alpha} := \left\{ \tilde{M}^{k_\alpha} : \sum_\alpha k_\alpha = N \right\}$$

Recall that the measurements defined on different subsystems are independent one from the other and compatible:

$$[\tilde{M}^{k_{\alpha_i}}, \tilde{M}^{k_{\alpha_j}}] = 0 \quad \text{for} \quad \tilde{M}^{k_{\alpha_j}}, \tilde{M}^{k_{\alpha_i}} \in P_{k_\alpha}$$

Now it seems obvious to write a new distance as $\max_{P_{k_\alpha}} \sum_\alpha d(\rho_{k_\alpha}, \sigma_{k_\alpha})$. This choice is indeed not correct because it does not take into account that the measures are taken on subsets of particles of different dimensions[2]. So the idea is to add a weight to consider this last detail. We define the weighted quantum distance between the two states ρ_N and σ_N as

$$D_d(\rho_N, \sigma_N) := \max_{P_{k_\alpha}} \sum_\alpha \frac{d(\rho_{k_\alpha}, \sigma_{k_\alpha})}{k_\alpha}$$

This new distance inherits the properties of the d_{cl} we used to define it and, on top of that, it can be bounded both from above and below

$$\frac{d(\rho_N, \sigma_N)}{N} \leq D_d(\rho_N, \sigma_N) \leq N d(\rho_N, \sigma_N) \leq N M_d$$

It is interesting to notice that the importance of the largest measurement setup does not increase under trivial extensions of the system. To be clearer let us make an example: let us consider an N particles system and two possible states $|0\rangle^{\otimes N}$ and $|x_1, ..., x_N\rangle$ and let us add to both of them a register of dimension Q-particles such that we end up with $|0\rangle^{\otimes N+Q}$ and $|x_1, ..., x_N\rangle|0\rangle^{\otimes Q}$. We have that

$$(N+Q) d(\rho_{N+Q}, \sigma_{N+Q}) \geq N d(\rho_N, \sigma_N)$$

$$D_d(\rho_{N+Q}, \sigma_{N+Q}) = D_d(\rho_N, \sigma_N)$$

So the new distance is not affected by the addition of a new register since the measurement setup is such that an N-particles detection is still maximally informative[2].

2.4.2 Bures weighted length

After introducing the weighted distance, we can make use of the Bures length to exploit some useful result. We define

$$D_B(\rho_N, \sigma_N) := \max_{P_{k_\alpha}} \sum_\alpha \frac{1}{k_\alpha} B(\rho_{k_\alpha}, \sigma_{k_\alpha})$$

In the following table, it is shown in a few relevant examples the difference between the Bures length and its weighted counterpart. In general the full knowledge of the quantum state under study is required for exact calculation of both of them, but statistical methods for extimating the standard distance from incomplete data can be used to overcome this problem.

ρ_N	σ_N	$B(\rho_N, \sigma_N)$	$D_B(\rho_N, \sigma_N)$
$\|0\rangle^{\otimes N}$	$\|1\rangle^{\otimes k}\|0\rangle^{\otimes N-k}$	$\frac{\pi}{2}$	$k \cdot \frac{\pi}{2}$
$\|0\rangle^{\otimes N}$	$ghz_k\|0\rangle^{\otimes N-k}$	$cos^{-1}(\|a\|)$	$k \cdot cos^{-1}(\|a\|)$
$\|0\rangle^{\otimes N}$	$ghz_l^{\otimes k}\|0\rangle^{\otimes N-kl}$	$cos^{-1}(\|a\|^k)$	$k \cdot l \cdot cos^{-1}(\|a\|)$
$\|0\rangle\langle 0\|^{\otimes N}$	$class_k \otimes \|0\rangle\langle^{\otimes N-k}$	$cos^{-1}(\|a\|)$	$k \cdot cos^{-1}(\|a\|)$
$\|0\rangle\langle 0\|^{\otimes N}$	$class_l^{\otimes k} \otimes \|0\rangle\langle^{\otimes N-kl}$	$cos^{-1}(\|a\|^k)$	$k \cdot l \cdot cos^{-1}(\|a\|)$
$\|0\rangle^{\otimes N}$	$\|dicke_{Nk}\rangle$	$\frac{\pi}{2}$	$N \cdot cos^{-1}(1 - \frac{k}{N})$
$\|0\rangle\langle 0\|^{\otimes N}$	$I_k/2^k \otimes \|0\rangle\langle 0\|^{\otimes N-k}$	$cos^{-1}(\frac{1}{\sqrt{2^k}})$	$k \cdot cos^{-1}(\frac{1}{\sqrt{2}})$
$\|ghz_N\rangle\langle ghz_N\|$	$I_N/2^N with \|a\|, \|b\| \neq \frac{1}{\sqrt{2}}$	$cos^{-1}(\frac{\|a\|+\|b\|}{\sqrt{2^N}})$	$N \cdot cos^{-1}(\frac{\|a\|+\|b\|}{\sqrt{2}})$
$class_N$	$I_N/2^N with \|a\|, \|b\| \neq \frac{1}{\sqrt{2}}$	$cos^{-1}(\frac{\|a\|+\|b\|}{\sqrt{2^N}})$	$N \cdot cos^{-1}(\frac{\|a\|+\|b\|}{\sqrt{2}})$
$\|ghz_N\rangle\langle ghz_N\|$	$I_N/2^N with Neven \|a\|, \|b\| = \frac{1}{\sqrt{2}}$	$cos^{-1}(\frac{1}{\sqrt{2^{N-1}}})$	$\frac{N\pi}{16}$
$class_N$	$I_N/2^N with Neven \|a\|, \|b\| = \frac{1}{\sqrt{2}}$	$cos^{-1}(\frac{1}{\sqrt{2^{N-1}}})$	$\frac{N\pi}{16}$
$class_N$	$\|ghz_N\rangle\langle ghz_N\|$	$cos^{-1}(\sqrt{a^4 + b^4})$	$\frac{cos^{-1}(\sqrt{a^4+b^4})}{N}$

Where we adopted the following : $|ghz_k\rangle = (a|0\rangle^{\otimes k} + b|1\rangle^{\otimes k})$, $class_k = (|a|^2|0\rangle\langle 0|^{\otimes k} + |b|^2|1\rangle\langle 1|^{\otimes k})$, $|dicke_{N,k}\rangle = \frac{1}{\sqrt{\binom{N}{k}}} \sum_i P_i|0\rangle^{\otimes N-k}|1\rangle^{\otimes k}$[2].

As we can see from the table above, the weighted distance carries an amount of information which is greater than its counterpart. In the next section we will also prove that it can be used to bound the experimental cost of a given (quantum) transformation.

3 Quantum Bound to state transformation

> The art of doing mathematics consists in finding that special case which contains all the germs of generality.
>
> *David Hilbert*

As shown in the previous chapter, the weighted Bures length has a clear metrological interpretation, which allows us to better discriminate different states of N-particles systems. In this section we will also shown that the weighted Bures length provides a bound to the cost of generating different configurations of our system. In particular turning an initial state $|0\rangle^{\otimes N}$ in a very different state is a basic requirement for a quantum algorithm. Providing a lower bound to the experimental cost can show the physical limits of quantum programming.

3.1 Quantum process dynamic

A quantum dynamics from an N-qubit input state ρ_N to a final state σ_N is a path in the space of N-particles density matrices. We can rewrite $\rho_{N,t}$ using its spectral decomposition :

$$\rho_{N,t} = \sum_{r=1}^{2^N} \lambda_r(t)|r(t)\rangle\langle r(t)| \quad \forall t \in [0,T]$$

with $\rho_{N,0} = \rho_N$, $\rho_{N,T} = \sigma_N$ and T the time required to the transformation .
We can define the rate of change as the time derivative of our density matrix $\rho_{N,t}$. This is a reminder of classical processes in which the rate of change is associated to the first derivative of the position, or in other words to the velocity. Let us also define the distance measure between two quantum states ρ_N and σ_N as:

$$\min_{\dot{\rho}_{N,t}} \int_0^T \|\dot{\rho}_{N,t}\| \; dt \quad \text{for a given norm}$$

In particular the input/output Bures length can be written as follows:

$$B(\rho_N, \sigma_N) = \min_{\dot{\rho}_{N,t}} \int_0^T \|\dot{\rho}_{N,t}\|_F \; dt$$

where we introduced the Fischer norm as follows:

$$\|\dot{\rho}_{N,t}\|_F^2 := \sum_r \frac{\dot{\lambda}_r^2(t)}{4\lambda_r(t)} + \sum_{r<s} \frac{|\langle r(t)|\dot{\rho}_{N,t}|s(t)\rangle|^2}{\lambda_r(t) + \lambda_s(t)}$$

Where the former term is a classical contribution, while the latter is a purely quantum term[2].
The classical term takes into account the eigenvalue change which can be carried out

by a classical algorithm, which is not interesting for our purpose. So from now on we will assume that the classical computations are free in terms of resources.
The quantum term is, on the other hand, responsible for the eigenbasis transformation. The transformation from $\rho_N \implies \tau_N \implies \sigma_N$, in which

$$\tau_N = \sum_r \lambda_r(T)|r(0)\rangle\langle(0)|$$

can be implemented via two steps :

- eigenvalues change: which can be assumed free from the point of view of the cost;
- eigenbasis change.

The second step can be implemented through a unitary path[2] $\tau_{N,t}$ such as :

$$\tau_{N,0} = \tau_N \text{ and } \tau_{N,T} = \sigma_N$$

For a unitary process the first step is redundant, so $\rho_N = \tau_N$. Then the quantum cost for a transformation $\rho_N \implies \sigma_N$ (even if not unitary) is defined as:

$$B^q(\rho_N, \sigma_N) := \min_{\text{unitary path } :\tau_{N,t}} \int_0^T \|\dot{\tau}_{N,t}\|_F dt$$

3.2 Discrete case

Let us now suppose to carry out the eigenbasis transformation through a sequence of unitary quantum gates:

$$U = \prod_l U_l = \prod_l e^{-iH_l T_l}$$

with T_l the running time of each gate.
The spectral decomposition of each time independent Hamiltonian is:

$$H_l = \sum_{x_l=1}^{2^{x_l}} h_{x_l}|h_{x_l}\rangle\langle h_{x_l}| \text{ with } h_{x_m} \geq h_{x_l} \text{ if } m > l$$

Note that each Hamiltonian H_l acts on $k_l \leq N$ particles. Let's call τ^l_{N,t_l} the intermediate state at time $t_l \in [0, T_l]$ while implementing U_l with:

$$\tau^l_{N,0} = \tau^l_N \text{ and } \tau^1_{N,0} = \tau_N$$

Since a time independent Hamiltonian gives rise to a constant speed process we have that

$$B^q(\rho_N, \sigma_N) \leq \int_0^{\sum_l T_l} \|\dot{\tau}_{N,t}\|_F dt = \sum_l \int_0^{T_l} \|\dot{\tau}_{N,t_l}\|_F dt = \sum_l \|\dot{\tau}_N\|_F T_l$$

The inequality can be saturated when σ_N (and τ_N) is a pure state[2]. The squared speed of the process lower bounds the variance of the generating Hamiltonian, which is also constant in time:

$$V_{\tau^l_N}(H_l) := Tr\left\{H_l^2 \tau_N\right\} - Tr^2\left\{H_l \tau_N\right\} \geq \|\dot{\tau}_N\|_F^2 \ \forall l$$

Let us introduce $E_l := \frac{1}{2}\left(h_{x_l=2^{k_l}} - h_{x_l=1}\right)$ and we can quantify the experimental cost of the process which transform ρ_N in σ_N in terms of physical resources:

$$R_{U_l} := k_l E_l T_l \implies R_U = \sum_l R_{U_l} \text{ with } k_l \text{ the size of the } U_l's \text{ gate}$$

It must be noticed that :

$$E_l^2 \geq V_{\rho_l}(H_l) \quad \forall l$$

It is important to be able to factorize the gate size, since a single qubit Hamiltonian is easier to implement in some given T_l than a k >1-particle interaction, even if the eigenvalue gap E_l is the same. Now we can retrieve a powerful result using some properties of the weighted Bures Length:

$$R_{U_l} \geq k_l B^q(\tau_N^l, \tau_N^{l+1}) \geq D_B(\tau_N^l, \tau_N^{l+1}) \ \forall l$$

So we have that:

- if the process is unitary: $R_U \geq D_B(\rho_N, \sigma_N)$;
- for a generic process: $R_U \geq D_B(\tau_N, \sigma_N)$.

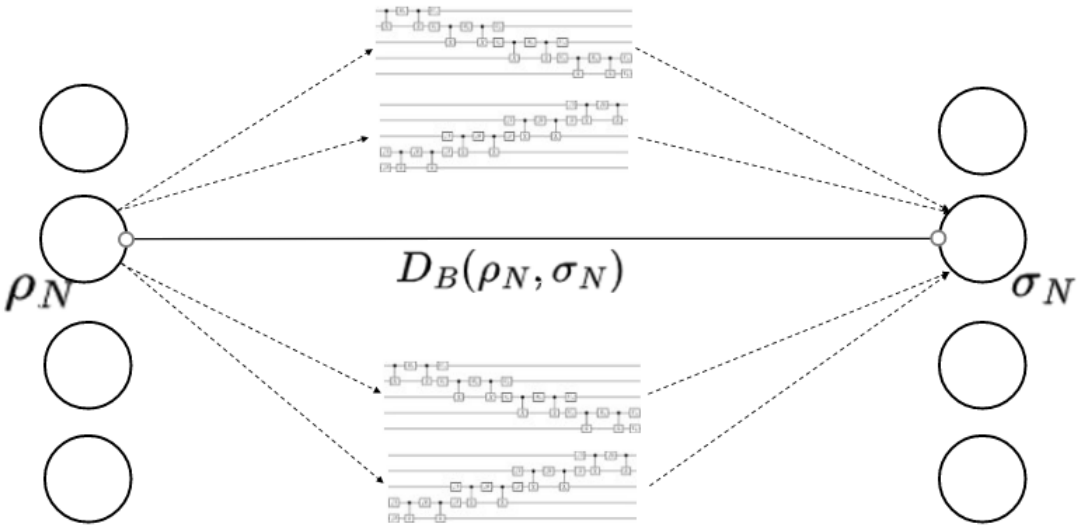

Figure 3: We proved that the weighted Bures Length $D_B(\rho_N, \sigma_N)$ is a lower bound to quantum transformation experimental cost. The bound is also valid for nonunitary quantum processes.

The bounds above allow us to relate the cost to implement a transformation to the distance between the initial and the final state. Note that for unitary processes $[\rho_N, \sigma_N] = 0 \iff D_B(\rho_N, \sigma_N) = 0$, i.e. the cost is zero if it exists a classical transformation from the initial state to the final one (coherent with our assumption of zero cost for classical transformations). Note that R_{U_l} is maxed out for the most sensible states to unitary perturbations, i.e. if they are coherent superpositions. The above result has many implications:

- it provides a lower limit to the difficulty of running quantum computations in terms of an exact analytical bound, rather than an order of magnitude;
- it is not limited to pure states: it also works in the case of mixed states and non unitary processes;
- the right hand-side of the inequality, D_B, is not only a mathematical tool but it has a clear physical meaning;
- the inequality is independent of the way the actual physical system is realized, thus completely general.

Being able to know at priori the cost of a quantum algorithm, without even looking to its physical realization, will make us able to analyze even better the possibility of quantum computing. On top of that, the result is not only related on quantum computing. This result has been obtained from quantum mechanics knowledge, thus totally general.

3.3 Continuous case

We can now extend the previous result to continuous processes. To do so we first state the theorem:
Theorem: Let f be a Lebesgue integrable function defined over the interval $[a, b]$. Then it exists a set $t_0, t_1, ..., t_{n+1}$ such that

$$a = t_0 < ... < t_{n+1} = b \text{ such that } \delta = \max_i(t_{i+1} - t_i) \text{ and } \epsilon > 0$$

Then we have that:

$$\lim_{\delta \to 0} \left[\int_0^T f(t)dt - \sum_{i=0}^{n} f(t_i)(t_{i+1} - t_i) \right] \leq \epsilon$$

Proof : Suppose f a Lebesgue integrable function defined over $[a, b]$ and let us call $R_t[f, t_0, ..., t_{n+1}]$ its Riemann sum such that

$$a = t_0 < ... < t_{n+1} = b \text{ with } \delta = \max_i(t_{i+1} - t_i)$$

$$R_t[f, t_0, ..., t_{n+1}] = \sum_{i=0}^{n} f(t_i)(t_{i+1} - t_i)$$

R_t is not guaranteed to be a good approximation of $\int_a^b f(t)dt$ in the limit $\delta \to 0$, since the Riemann integrability is not required. Let us introduce the set $s_0, ..., s_{n+1}$ such that

$$s_j = t_j + t \text{ if } t \in [0, b - a]$$

where we suppose $s_j + t \neq b$ and $s_j + t - (b - a)$ if $s_j + t \geq b$. Let us substitute f with

$$f(s) = f(t - (b - a)) \ : \ b \leq t \leq 2b - a$$

So we can introduce a new Riemann series

$$R_{t'}[f, s_0, ..., s_{n+1}] = \sum_{i=0}^{n} f(s_i)(s_{i+1} - s_i)$$

Then it is trivial to verify that

$$\lim_{\delta\to 0}\int_0^{b-a}\big|R_t - R_{t'}\big|dt = 0$$

Let $\epsilon \geq 0$ and $f_\epsilon \in C^0[a, 2b-a]$ such that

$$\int_0^{2b-a}\Big|f(t) - f_\epsilon(t)\Big| \leq \epsilon$$

Then

$$\int_0^{b-a}\left|R_{t'} - \int_a^b f(s)ds\right|dt = \int_0^{b-a}\left|R_{t'} - \int_a^b f(s+t)ds\right|dt =$$

$$\leq \sum_{j=1}^{n-1}\int_{s_j}^{s_{j+1}} ds\int_0^{b-a}\left|f(s_j+t)-f(s+t)\right|dt \leq \sum_{j=1}^{n-1}\int_{s_j}^{s_{j+1}} ds\int_0^{b-a}\left|f_\epsilon(s_j+t)-f_\epsilon(s+t)\right|dt+2\epsilon(b-a)$$

It implies that:

$$\int_0^{b-a}\left|R_{t'} - \int_a^b f(s)ds\right|dt \leq 2(b-a)\epsilon \text{ for } \delta\to 0$$

Since ϵ can be taken arbitrarily near to 0, then $R_{t'}$ is a good replacement of R_t. This implies that must exist a choice of the set $s_0, ..., s_{n+1}$ such that the theorem is satisfied[3]. (Q.E.D.)

The theorem simply states that under some minor regularity conditions, every continuous process can be approximated to a discrete one. This result is particularly useful since we can rewrite :

$$\left|\int_0^T k(t)E(t)dt - \sum_{i=0}^{n} k(t_i)E(t_i)(t_{i+1}-t_i)\right| = O(\epsilon)$$

As stated before, it is true that $R_{U_l} \geq D_B(\tau_N^l, \tau_N^{l+1})$, then it follows that

$$\sum_{i=0}^{n} k(t_i)E(t_i)(t_{i+1}-t_i) \geq \sum_{i=0}^{n} D_B(\tau_N^i, \tau_N^{i+1}) \geq D_B(\tau_N, \sigma_N)$$

where the last inequality is just a consequence of the triangular inequality. Then

$$\int_0^T k(t)E(t)dt \geq D_B(\tau_N, \sigma_N) + O(\epsilon)$$

We extended the bound to time continuous processes. The bound we derived allows us to compute, independently from the physical realization and the way the process is carried out (time dependent or not), the minimal/optimal cost of a process.
So in conclusion we proved that the weighted Bures length not only is a better way to distinguish between quantum physical systems, but also bounds the size of preparation algorithm. This leads to a nice consequence: the ability to prepare a given transformation to an initial state ρ to a final one σ will never be better than the ability to distinguish them.

4 Quantum state preparation complexity

> Geometry is almost the only subject as to which we find truths wherein all men agree; and one cause of this is, that geometers alone regard the true laws of demonstration.
>
> *Blaise Pascal*

In this section we will focus on the difficulty of driving a quantum state into a given target state. In particular we will focus our attention on the problem of finding the optimal path which leads to the minimum experimental cost through a geometric method. This allow us to find the most efficient path independently from the actual realization of the system, thus in a completely general way. As said before, the ability of predict the cost of a transformation can be used to measure the performance of a preparation algorithm, thus its efficiency.

4.1 Quantumness

Let us introduce a useful concept that will help us in finding the optimal path. As already said before, the energy cost for a quantum process (from an initial state ρ to a final state σ) can be written as

$$E^{\gamma_t} = \int_0^T ||\dot{\gamma_t}||^2 dt : \left\{ \gamma_t(0) = \rho, \gamma_t(T) = \sigma \right\} = E^{\gamma_t}_{classical}(\rho, \sigma) + E^{\gamma_t}_{quantum}(\rho, \sigma)$$

where the classical contribution takes into account the eigenvalues evolution and the quantum one the eigenvectors change. The definition is valid for any norm induced by a Riemannian Fisher metric, but we will focus on the Fisher norm introduced before. As already stated, the classical contribution can be neglected. So we introduce the quantity $Q_\rho(\sigma)$ called the **quantumness of a computation from ρ to σ** as

$$Q_\rho(\sigma) := E_q^{\overline{\gamma}_t}(\overline{\rho}, \sigma)$$

$$E_q^{\overline{\gamma}_t}(\overline{\rho}, \sigma) := \min_{\gamma_t, \rho} E_q^{\gamma_t}(\rho, \sigma), \text{ such that } \gamma_t : \rho \to \sigma, \ \rho \in \tilde{M}_{i_R}$$

The quantumness measures the minimum quantum energy cost to drive a system from an initial state into a given target state in a given time T. The definition of the energy cost is, as it was for the rate of change of the process, a reminiscence of the classic definition of the work–energy principle (where our first derivative acts as a velocity of the process, so its modulo square acts as an energy). If we define $\Gamma(\gamma_t) : \Gamma(\rho) \to \Gamma(\sigma)$ the dynamics of a state under a CPTP map at any time, then the quantumness is contractive, i.e. $Q_\rho(\sigma) \geq Q_{\Gamma(\rho)}(\Gamma(\sigma))$[4]. As said in the previous section, the quantum part of the process can be carried out through unitary gates, so we have that

$$E^{\gamma_t^u}(\rho^u, \sigma) = E_q^{\gamma_t^u}(\rho^u, \sigma)$$

Given a target state σ of a d-dimensional system with spectrum multiplicity m_i, there are $\frac{d!}{\prod_i m_i!}$ isospectral free states which can transform freely one into the other via permutation of the eigenvalues. So the minimum cost of the quantum part is obtained by minimizing through the states isospectral to the target

$$Q^u_\rho(\sigma) := E_q^{\overline{\gamma}^u_t}(\overline{\rho^u}, \sigma)$$

$$E_q^{\overline{\gamma}^u_t}(\overline{\rho}^u, \sigma) = \min_{\gamma^u_t, \rho^u_p} E_q^{\gamma^u_t}(\rho^u_p, \sigma), \text{ such that } \gamma^u_t : \rho^u_p \to \sigma$$

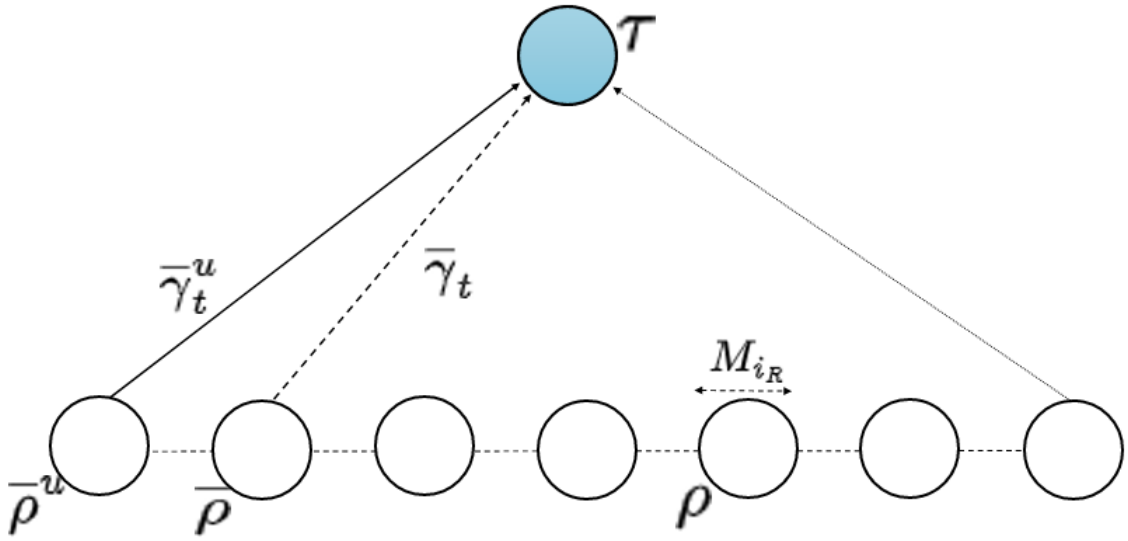

Figure 4: The optimal path $\overline{\gamma}_t$ to drive a system from $\rho \in M_{i_R}$ into τ is the one that minimize the quantum cost $Q_\rho(\tau) = E_q^{\overline{\gamma}_t}(\overline{\rho} \to \tau)$ over all the possible free states, since classical computation costless. The unitary path $\overline{\gamma}^u_t$ generates the energy minimizing path from a free state $\overline{\rho}^u$ isospectral to the target.

By construction we have $Q^u_\rho(\sigma) \geq Q_\rho(\sigma)$. This bound also satisfies contractivity ($Q^u_\rho(\sigma) \geq Q^u_{\Gamma(\rho)}(\Gamma(\sigma))$), faithfulness ($Q_\rho(\sigma) = 0 \iff \sigma \in M_{i_R}$) and invariance ($Q^u_\rho(\sigma) = Q^u_{\tilde{\rho}}(\sigma), \forall \tilde{\rho} \in M_{i_R}$). Note that the two-step, classical-quantum split is optimal by construction. A classical map is, by definition, a transformation in M_{i_R}. Hence, the path corresponding to an arbitrary sequence of multiple classical and quantum steps returns to M_{i_R} multiple times, requiring more energy.
It is interesting to notice that for time independent Hamiltonian

$$T \cdot V_\rho(H) \geq E^{\gamma^u_t}(\rho, \sigma) \text{ with } V_\rho(H) := Tr\Big\{H^2\rho\Big\} - \Big(Tr\Big\{H\rho\Big\}\Big)^2$$

4.2 Optimal path

Our goal is now to compute the best path $\overline{\gamma}^u_t$ to reach the target from an isospectral free state. Let us recall from chapter 2 the definition of Bures length

$$D_B(\rho, \sigma) = cos^{-1}\Big(Tr\Big|\sqrt{\rho}\sqrt{\sigma}\Big|\Big)$$

The map between two states which minimizes the energy $E^{\gamma_t}(\rho,\sigma)$ is the length minimizer at constant speed. The energy minimizing path from a pure isospectral state $\rho^u = |\Psi_{\rho^u}\rangle\langle\Psi_{\rho^u}|$ to a target pure state $\tau = |\Psi_\tau\rangle\langle\Psi_\tau|$ is the length minimizing unitary. In this specific case the Bures length reduces to the Fubini-Study Distance $D_{FS}(\Psi_{\rho^u},\Psi_\sigma) := cos^{-1}|\langle\Psi_{\rho^u}|\Psi_\sigma\rangle|$. The closest free pure state to the target is, of course, the one with maximal overlap. This allows us to write the energy minimizing constant speed path:

$$\overline{\gamma}_t^u = |\Psi_{\overline{\gamma}_t^u}\rangle\langle\Psi_{\overline{\gamma}_t^u}| \quad \text{with} \quad |\Psi_{\overline{\gamma}_t^u}\rangle = \left(cos\theta - \frac{sin\theta}{tan(d)}\right)|\Psi_{\overline{\rho}^u}\rangle + \left(\frac{sin\theta}{sin(d)}\right)|\Psi_\sigma\rangle$$

where $d = D_{FS}(\Psi_{\overline{\rho}^u},\Psi_\sigma)$ and $\theta = d\frac{t}{T}$[5, 6, 7].
It can be proven that the energy minimizing path is unique up to affine transformations $t \to t' = at + b$ with $a, b \in R$. The problem is way more challenging for a target mixed state, but the previous result for pure states yields a lower bound to the quantumness $Q_\rho^u(\sigma)$ for arbitrary target states. This is a consequence of the fact that the distance between two mixed states is the minimum distance between their purifications. The closest isospectral free state to the target is then the one whose purification $|\Psi_{\overline{\rho}^u}^{puri}\rangle$ is closer to the target purification $|\Psi_\sigma^{puri}\rangle$. We can write them in a compact form

$$|\Psi_{\overline{\rho}^u}^{puri}\rangle = \sum_i \sqrt{\overline{\rho}^u}|i_R\rangle \otimes |i_R\rangle$$

$$|\Psi_\sigma^{puri}\rangle = \sum_i \frac{1}{\sqrt{\overline{\rho}^u}}\sqrt{\sqrt{\overline{\rho}^u}\sigma\sqrt{\overline{\rho}^u}}|i_R\rangle \otimes |i_R\rangle$$

The length/energy minimizing (generally not unitary) path between two mixed states is obtained by partial trace along the shortest (unitary) path between the closest purifications. So we have that $Q_\rho^u(\sigma) \geq Q_{|\Psi_{\overline{\rho}^u}^{puri}\rangle}(|\Psi_\sigma^{puri}\rangle)$, which is saturated when the target is a pure state. The above bound satisfies by construction faithfulness, invariance and monotonicity.[4]

We can make use of the fact that the variance of the generating Hamiltonian acts as a upper bound to the energy cost. So let us consider a process carried out by a series of quantum gates described by an Hamiltonian H_l such that each of the acts on k_l qubits for an interval of time t_l. Then we have that

$$\sum_l E_l^2 t_l \geq \sum_l V(H_l)t_l \geq \sum_l t_l||\dot{\tau}_l||^2 = E_{quantum}^{\gamma_t}(\rho,\sigma)$$

Where E_l is the halved seminorm associated to H_l. This result can be easily recast in the continuous form remembering the theorem used in chapter 3. So we end up with

$$\min_{\gamma_t}\left\{\int_0^T E^2(t)dt : \gamma_t = \rho \to \sigma\right\} \geq Q_\rho^u(\sigma) \geq Q_{|\Psi_{\overline{\rho}^u}^{puri}\rangle}(|\Psi_\sigma^{puri}\rangle)$$

This gives also an upper bound to the quantumness of a process.

4.2.1 Qubit example

Let us consider a single example in which we drive a simple qubit from an input state $\rho = \frac{1}{2}\big(I + q_0\sigma_z\big)$ given $q_0 \neq 0$, to a target state $\tau = \frac{1}{2}\big(I + \vec{r}\cdot\vec{\sigma}\big)$. The isospectral free state to the target share the property $|\vec{r}| = q_0$. Which leads to

$$Q_{|\Psi^{puri}_{\rho^u}\rangle}(|\Psi^{puri}_{\tau}\rangle) = \frac{1}{T}\left\{cos^{-1}\left[\frac{\sqrt{f_+}+\sqrt{f_-}}{2}\right]\right\}^2$$

with $f_\pm = 1 + q_0 r_z \pm (q_0 + r_z)$.

The process is classical for $r_x = r_y = 0$, while it reaches the maximum energy to be implemented ($\frac{\pi^2}{16T} \approx \frac{0.6168}{T}$) when $r_x^2 + r_y^2 = 1$. The same method applies for system with d>2, as their states admit a representation in Bloch form as $\frac{1}{d}\left[I + \vec{r}\cdot\vec{\Sigma}\right]$, with $\vec{\Sigma}$ a vector of $d \otimes d$ traceless matrices.[4]

4.2.2 2-Qubits example

Let us consider now a system made of two qubits which must be driven from an initial state ρ to a final state τ. To keep it general we introduce the vector $\vec{\sigma} = (\sigma_0, \sigma_1, \sigma_2, \sigma_3) = (I_{2\text{x}2}, \sigma_x, \sigma_y, \sigma_z)$, whose element form a basis for the space of 2x2 complex matrices. So we can write the density matrix of ρ and τ as :

$$\rho = \frac{1}{4}\sum_{i,j=0}^{3} R^{\rho}_{i,j}\sigma_i \otimes \sigma_j \ , \ \tau = \frac{1}{4}\sum_{i,j=0}^{3} R^{\tau}_{i,j}\sigma_i \otimes \sigma_j$$

Where $R^{\tau}_{i,j} = Tr\left[\tau\left(\sigma_i \otimes \sigma_j\right)\right]$[8]. Consequently we have that $R_{0,0} = 1$ and we can rewrite our state in a more meaningful way as

$$\tau = \frac{1}{4}\left(I_{4\text{x}4} + \sum_{i=1}^{3} x_i\sigma_i \otimes I_{2\text{x}2} + \sum_{i=1}^{3} y_i I_{2\text{x}2} \otimes \sigma_i + \sum_{i,j=1}^{3} t_{i,j}\sigma_i \otimes \sigma_j\right) =$$

$$= \frac{1}{4}\begin{pmatrix} 1+t_{33}+x_3+y_3 & t_{31}+y_1-i(t_{32}+y_2) & t_{13}+x_1-i(t_{23}+x_2) & t_{11}-t_{22}-i(t_{12}+t_{21}) \\ t_{31}+y_1+i(y_2+t_{32}) & 1-t_{33}+x_3-y_3 & t_{11}+t_{22}+i(t_{12}-t_{21}) & x_1-t_{13}-i(x_2-t_{23}) \\ t_{13}+x_1+i(t_{23}+x_2) & t_{11}+t_{22}+i(t_{21}-t_{12}) & 1+y_3-t_{33}-x_3 & y_1-t_{31}+i(t_{32}-y_2) \\ t_{11}-t_{22}+i(t_{12}+t_{21}) & x_1-t_{13}+i(x_2-t_{23}) & y_1-t_{31}-i(t_{32}-y_2) & 1+t_{33}-x_3-y_3 \end{pmatrix}$$

The vectors $\vec{x}$ and $\vec{y}$ are the Bloch vectors associated with each of the two qubits, while $t_{i,j}$ are the elements of the correlation matrix $\boldsymbol{T}$. To be coherent with the previous example we start from a state $\rho = \frac{1}{4}(I_{2x2} + q_1\sigma_z) \otimes (I_{2x2} + q_2\sigma_z)$ such as

$$\rho = \frac{1}{4}\begin{pmatrix} 1+q_1+q_2+q_1q_2 & 0 & 0 & 0 \\ 0 & 1+q_1-q_2-q_1q_2 & 0 & 0 \\ 0 & 0 & 1-q_1+q_2-q_1q_2 & 0 \\ 0 & 0 & 0 & 1-q_1-q_2+q_1q_2 \end{pmatrix}$$

So we can introduce a useful notation where

$$||\boldsymbol{R}||^2 = |\vec{x}|^2 + |\vec{y}|^2 + ||\boldsymbol{T}||_2^2$$

Asking that ρ and τ are isospectral reduces to $||\boldsymbol{R_\tau}|| = ||\boldsymbol{R_\rho}|| = \sqrt{q_1^2 + q_2^2 + q_1^2 q_2^2}$.
So as before we have that

$$Q^u_\rho(\sigma) \geq Q_{|\Psi^{puri}_\rho\rangle}(|\Psi^{puri}_\sigma\rangle) = \frac{1}{T}\left(D_B(|\Psi^{puri}_{\overline{\rho}}\rangle, |\Psi^{puri}_\tau\rangle)\right)^2$$

4.3 Algorithmic complexity

It is important for a quantum computer being able to easily transform $|0\rangle^{\otimes N}$ in a high entangled state or in a superposition of state, which are the key features which allows quantum computer to perform "better" than their classical counter part. The quantumness measure can be used to bound the size of preparation algorithms. Suppose the unitary map $\overline{\gamma}^u_t := \rho^u \to \sigma$ is synthesized by N commuting unitary operations (logic gates) such that

$$\overline{\gamma}^u_t = U_t \rho^u U_t^\dagger \text{ with } U_t = e^{iH_l t_l} : \left\{ \sum_l H_l = H, [H_l, H_k] = 0, \ \forall l, k \right\}$$

Consider the halved semi-norm of each Hamiltonian $|H_l| = \frac{h_{l,M} - h_{l,m}}{2}$, already defined before as the difference between the largest and the smallest eigenvalues. We can introduce now $\overline{|H|} = \frac{1}{N}\sum_l |H_l|$ as the average semi-norm. This allows us to obtain the following bound

$$V_\rho(H) \leq |H|^2 \leq \left(\sum_l |H_l|^2\right)$$

So using the previous bound between the speed of the process and the quantumness we end up with

$$Q^u_\rho(\sigma) \leq E^{\gamma^u_t}(\rho, \sigma) \leq T \cdot V_\rho(H) \leq T N^2 \overline{|H|}^2$$

The bound is saturated for superpositions of the largest and the smallest eigenstates of H, which are the most sensitive inputs to the map.
This result can be extended also to the case of time dependent Hamiltonians. If our Hamiltonian **H** is time dependent and commute, e.g. $[H_l(t), H_k(t)] = 0$ $(\forall l, k, t)$, then the path can be approximated by a discrete process such that

$$\gamma^u_t = U \rho^u U^\dagger + O(\epsilon) \ \text{ with } U = \prod_k e^{-i\left(\sum_l H_l(t_k)\right)(t_{k+1} - t_k)}$$

where $(t_{k+1} - t_k)$ is the interval of time in which our discrete time independent Hamiltonian $\sum_l H_{l,T_k}$ acts on the system. This is a mathematical consequence of the theorem used in chapter 3, since the quantity we want to take into consideration are integral whose argument is a function of γ^u_t. The correction of order ϵ is negligible since is taken as $\epsilon = \min_k(t_{k+1} - t_k)$, so we will always consider the limit $\epsilon \to 0^+$. So we have

$$E^{\gamma^u_t}(\rho, \sigma) = \sum_{i=1}^n ||\dot{\tau}(t_i)||^2 (t_{i+1} - t_i) = \sum_{i=1}^n E^{\gamma^u_t}(\tau(t_i), \tau(t_{i+1}))$$

Which leads to

$$E^{\gamma^u_t}(\rho, \sigma) \leq \sum_k V_{\tau(t_k)}(t_{k+1} - t_k) \leq \sum_k |H(t_k)|^2 (t_{k+1} - t_k)$$

The last sum can be seen as a Riemann sum which allows us to write

$$\sum_k |H(t_k)|^2(t_{k+1}-t_k) = \sum_k N^2\overline{|H(t_k)|}^2(t_{k+1}-t_k) = \int_0^T N^2\overline{|H(t)|}^2 dt + O(\epsilon)$$

So in the end we have that

$$Q^u_\rho(\sigma) \leq \int_0^T N^2\overline{|H(t)|}^2 dt + O(\epsilon)$$

So the same bound found for discrete cases can be extended to continuous ones, allowing as to bound from above, even more generally, the energetic cost of a computation.

4.3.1 GHZ preparation

Let us consider a common scenario in which we want to prepare an highly entangled symmetric state ($|\Psi_\rho\rangle = (a|0\rangle+b|1\rangle)|0\rangle^{\otimes N} \rightarrow |\Psi_\sigma\rangle = a|0\rangle^{\otimes N+1}+b|1\rangle^{\otimes N+1}$ such that $|a|^2+|b|^2=1$), via N-controlled gates between the first and the $(l+1)$-th qubit. Such final state $|\Psi_\sigma\rangle$ is also called GHZ state[9] (Greenberger–Horne–Zeilinger state, from the name of the three physicists who showed first its peculiar non classical behaviour) and it is used in several protocols in quantum communication and cryptography, for example, in secret sharing or in the quantum Byzantine agreement. We are going to prove that this is the optimal path using N gates.

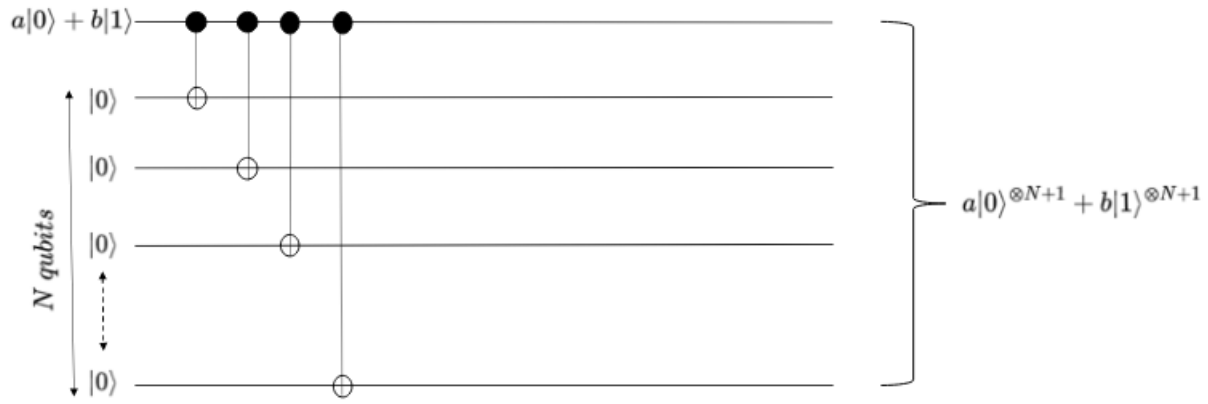

Figure 5: Quantum circuit representation of the process $|\Psi_\rho\rangle = (a|0\rangle + b|1\rangle)|0\rangle^{\otimes N} \rightarrow |\Psi_\sigma\rangle = a|0\rangle^{\otimes N+1} + b|1\rangle^{\otimes N+1}$ using a series of control not gates.

We can notice a little detail about the result we used in the previous discussion. For pure states the bound we express before reads as:

$$\int_0^T ||\dot{\tau}||^2 dt \int_0^T 1^2 dt \geq \left(\int_0^T (||\dot{\tau}|| \cdot 1) dt\right)^2$$

Which is a peculiar case of the Cauchy-Schwarz Inequality for L^2 function. The inequality states that for all elements u and v of an inner product space is true that:

$$|\langle u|v\rangle|^2 \leq |\langle u|u\rangle| \cdot |\langle v|v\rangle| = |u|^2 \cdot |v|^2$$

Theorem: given u, v then the Cauchy-Schwarz Inequality is saturated if u and v are proportional one to the other.
Proof: The case $u = 0$ (or $v = 0$) saturates the inequality at priori, so let us assume v and u both different from zero. Let us write the decomposition of u in a parallel component to v and an orthogonal to v (let us call it w):

$$u = \frac{\langle u|v\rangle}{|v|^2}v + w$$

Then we have that

$$|u|^2 = \left|\frac{\langle u|v\rangle}{|v|^2}v\right|^2 + |w|^2 \geq \left|\frac{\langle u|v\rangle}{|v|^2}\right|^2$$

If we multiply both side by $|v|^2$ we retrieve the inequality. The equal sign holds only in the case $w = 0$ or in other words u and v proportional the one to the other. (Q.E.D.)
So we can rewrite the bound above as:

$$\int_0^T 1^2 dt \int_0^T ||\dot{\tau}||^2 dt = \left(\int_0^T ||\dot{\tau}|| dt\right)^2 \text{ if } ||\dot{\tau}|| \propto 1$$

Or in other words $||\dot{\tau}||$ is constant in time (coherent with what we said in section 4.2). Let us start from the simplest case in which N=1, so we want to transform $(a|0\rangle+b|1\rangle)|0\rangle$ into $a|0\rangle^{\otimes 2} + b|1\rangle^{\otimes 2}$ in a time t_0 then we have that:

$$t_0 \int_0^{t_0} ||\dot{\tau}||^2 dt = t_0^2 ||\dot{\tau}||^2 = \left(cos^{-1}|a|^2\right)^2$$

Which implies $||\dot{\tau}|| = \frac{cos^{-1}|a|^2}{t_0}$. But if $||\dot{\tau}||$ is constant in time, it implies that must be carried out through a single gate. The only gate that satisfies the requirement is the CNOT gate.
Now it is easy to see that once we have the state $a|0\rangle^{\otimes 2} + b|1\rangle^{\otimes 2}$ the same approach can be derived to prove that the best path from $|\Psi_\rho\rangle = (a|0\rangle^{\otimes 2} + b|1\rangle^{\otimes 2})|0\rangle \rightarrow |\Psi_\sigma\rangle = a|0\rangle^{\otimes 3} + b|1\rangle^{\otimes 3}$ is a CNOT gate between the first and the third gate.
The interest now is focused on proving that any other path carried out with 2 gates will lead to an higher energetic cost. It is obvious that once we choose the intermediate state $|\Psi_\alpha\rangle$ we can find the optimal path from $|\Psi_\rho\rangle \rightarrow |\Psi_\alpha\rangle$ and $|\Psi_\alpha\rangle \rightarrow |\Psi_\sigma\rangle$. The sum of the two path will be optimal only if the two are by themselves optimal. If we write our intermediate state as:

$$|\Psi_\alpha\rangle = (\alpha_{00}|00\rangle + \alpha_{01}|01\rangle + \alpha_{10}|10\rangle + \alpha_{11}|11\rangle)|0\rangle$$

and remember the isospectral condition written before, the maximum overlap is found for $\alpha_{00} = a, \alpha_{11} = b, \alpha_{10} = \alpha_{01} = 0$, which correspond to the state

$$|\Psi_\alpha\rangle = (a|00\rangle + b|11\rangle)|0\rangle$$

so the qubit entangled state. Which means that the best path can be carried out through 2 CNOT gates. Then the generalization to N gates can be done in the same way. As we stated before it is important to focus our attention on the dimension of the gates we are using. Even if two gates might share the same energy cost a small one (in the number of involved qubits) is easier to implement than a k-dimensional one in some given time.

5 Appendix A

We are going to provide some of the useful proof of the properties discussed before.

5.1 Trace distance

- $D(\rho,\sigma) = D(U\rho U^\dagger, U\sigma U^\dagger)$
  The trace distance is invariant under unitary transformation
  Proof

$$D(U\rho U^\dagger, U\sigma U^\dagger) = \frac{1}{2}Tr\left|U\rho U^\dagger - U\sigma U^\dagger\right| = \frac{1}{2}Tr\left|U(\rho-\sigma)U^\dagger\right| = \frac{1}{2}Tr\left|\rho-\sigma\right| = D(\rho,\sigma)$$

- Let ϵ be a trace preserving quantum operation then

$$D(\epsilon(\rho),\epsilon(\sigma)) \leq D(\rho,\sigma)$$

  Proof
  Let us use the spectral decomposition $\rho - \sigma = Q - S$ with Q and S positive matrices with orthogonal support, and let P be a projector such that

$$D(\epsilon(\rho),\epsilon(\sigma)) = Tr\left[P\left(\epsilon(\rho) - \epsilon(\sigma)\right)\right]$$

  and let's observe that $Tr(Q) - Tr(S) = Tr(\rho) - Tr(\sigma) = 0$ so this implies $Tr(Q) = Tr(S)$ thus $Tr(\epsilon(Q)) = Tr(\epsilon(S))$

$$D(\rho,\sigma) = \frac{1}{2}Tr|\rho-\sigma| = \frac{1}{2}|Q-S| = \frac{1}{2}Tr(Q) + \frac{1}{2}Tr(S) =$$

$$= Tr(\epsilon(Q)) \geq Tr(P(\epsilon(Q))) \geq Tr(P(\epsilon(Q) - \epsilon(S))) = Tr(P(\epsilon(\rho) - \epsilon(\sigma))) = D(\epsilon(\rho),\epsilon(\sigma))$$

- Strong convexity of the trace distance

$$D\left(\sum_i p_i\rho_i, \sum_i q_i\sigma_i\right) \leq D(p,q) + \sum_i p_i D(\rho_i,\sigma_i)$$

  Proof

$$D\left(\sum_i p_i\rho_i, \sum_i q_i\sigma_i\right) = \sum_i p_i Tr(P\rho_i) - \sum_i q_i Tr(P\sigma_i) =$$

$$\sum_i p_i Tr(P(\rho_i - \sigma_i)) + (p_i - q_i)Tr(P(\sigma_i)) =\leq \sum_i p_i D(\rho_i,\sigma_i) + D(p,q)$$

5.2 Fidelity

- $F(U\rho U^\dagger, U\sigma U^\dagger) = F(\rho,\sigma)$
  The fidelity is invariant under unitary transformation of its arguments.
  Proof
  Remember that $\sqrt{UAU^\dagger} = U\sqrt{A}U^\dagger$ if U is a unitary matrix, so we have :

$$F(U\rho U^\dagger, U\sigma U^\dagger) = Tr\sqrt{\sqrt{U\rho U^\dagger}U\sigma U^\dagger\sqrt{U\rho U^\dagger}} = Tr\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}} = F(\rho,\sigma)$$

- Strong concavity property

$$F\left(\sum_i p_i\rho_i, \sum_i q_i\sigma_i\right) \geq \sum_i \sqrt{p_i q_i} F(\rho_i, \sigma_i)$$

Proof
Let $|\Psi_i\rangle$ and $|\Phi_i\rangle$ be a purification of ρ_i and σ_i such that $F(\rho_i, \sigma_i) = \langle\Psi_i|\Phi_i\rangle$. Introduce now a system which has an orthonormal basis state $|i\rangle$ corresponding to the index i of the probability distributions. Define

$$|\Psi\rangle = \sum_i \sqrt{p_i}|\Psi_i\rangle|i\rangle \ , \ |\Phi\rangle = \sum_i \sqrt{q_i}|\Phi_i\rangle|i\rangle$$

Note that $|\Psi\rangle$ and $|\Psi\rangle$ are purifications of $\sum_i p_i\rho_i$ and $\sum_i q_i\sigma_i$, respectively. So we have that

$$F\left(\sum_i p_i\rho_i, \sum_i q_i\sigma_i\right) \geq |\langle\Psi|\Phi\rangle| = \sum_i \sqrt{p_i q_i}|\langle\Psi_i|\Phi_i\rangle| = \sum_i \sqrt{p_i q_i} F(\rho_i, \sigma_i)$$

- Ulhmann's Theorem : Suppose ρ and σ are states of a quantum system Q. Introduce a quantum system R which is a copy of Q. Then

$$F(\rho, \sigma) = \max_{|\Psi\rangle, |\Phi\rangle} |\langle\Psi|\Phi\rangle|$$

where $|\Psi\rangle$ and $|\Phi\rangle$ are the purification of ρ and σ from R into Q.
Proof
Fix an orthonormal base $|i_R\rangle$ and $|i_Q\rangle$ in the systems R and Q. Because R and Q are of the same dimension the index i may be assumed to run over the same set of values . Define

$$|m\rangle = \sum_i |i_R\rangle|i_Q\rangle$$

Let $|\Psi\rangle$ be a purification of ρ , then the Schmidt's decomposition will be :

$$|\Psi\rangle = \left(U_R \otimes \sqrt{\rho} U_Q\right)|m\rangle$$

For some unitary operation U_R and U_Q on system R and system Q. Similarly , if $|\Phi\rangle$ is any purification of σ , then there exist unitary operators V_R and V_Q such that :

$$|\Phi\rangle\left(V_R \otimes \sqrt{\sigma} V_R\right)|m\rangle$$

So in the end we can write

$$\langle\Psi|\Phi\rangle = |\langle m|U_R^\dagger V_R \otimes U_Q^\dagger \sqrt{\rho\sigma} V_Q|m\rangle| = |Tr(V_R^\dagger U_R U_Q^\dagger \sqrt{\rho\sigma} V_Q)|$$

If we set $U = V_Q V_R^\dagger U_R U_Q^\dagger$ we obtain

$$|\langle\Psi|\Phi\rangle| = |Tr(\sqrt{\rho\sigma}U)| \leq Tr|\sqrt{\rho\sigma}| \leq Tr(\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}}) = F(\rho, \sigma)$$

- Suppose ϵ a trace preserving quantum operation. Let ρ and σ be density matrices then
$$F(\epsilon(\rho), \epsilon(\sigma)) \geq F(\rho, \sigma)$$
Proof
Let $|\Psi\rangle$ and $|\Phi\rangle$ be purifications of ρ and σ into a joint system RQ such that $F(\rho, \sigma) = |\langle\Psi|\Phi\rangle|$. Introduce a model environment E for the quantum operation ϵ which starts from $|0\rangle$ and interacts with Q via a unitary matrix U. $U|\Psi\rangle|0\rangle$ is a purification of $\epsilon(\rho)$. By Ulhmann's theorem
$$F(\epsilon(\rho), \epsilon(\sigma)) \geq |\langle\Psi|\langle 0|U^\dagger U|0\rangle|\Phi\rangle| = |\langle\Psi|\Phi\rangle| \geq F(\rho, \sigma)$$
So we have
$$F(\epsilon(\rho), \epsilon(\sigma)) \geq F(\rho, \sigma)$$

5.3 Weighted distance

- Non-negativity : $D_d(\rho_N, \sigma_N) \geq 0$
Proof
Since D_d is a weighted sum of the distance measures d which are all non negative, D_d has to be non negative.

- Faithfulness : $\rho_N = \sigma_N \iff D_d(\rho_N, \sigma_N) = 0$
Proof
If $\sigma_N = \rho_N$ then it must be true that $\sigma_{k_\alpha} = \rho_{k_\alpha}, \forall\ k_\alpha \implies D_d(\rho_{k_\alpha}, \sigma_{k_\alpha}) = 0, \forall k_\alpha \implies D_d(\rho_N, \sigma_N) = 0$. If on the other hand $D_d(\rho_N, \sigma_N) = 0 \implies d(\rho_{k_\alpha}, \sigma_{k_\alpha}) = 0, \forall k_\alpha$ because D_d is maximal between the possible weighted sums so $\rho_{k_\alpha} = \sigma_{k_\alpha}, \forall k_\alpha \implies \rho_N = \sigma_N$

- Contractivity : $D_d(\rho_N, \sigma_N) \geq D_d(\Lambda(\rho_N), \Lambda(\sigma_N))$ with Λ any CPTP maps on a single subsystem.
Proof
Each of the element of the sum satisfies the contractivity property since we chose d such that it has that property. So the overall sum inherit such property, since it is a weighted sum in which each weight $\frac{1}{k_\alpha}$ is positive and greater than zero.

- Triangular inequality : $D(\rho_N, \sigma_N) \leq D(\rho_N, \tau_N) + D(\tau_N, \sigma_N)$
Proof
Since d is a proper metric it satisfies the triangular inequality . D is a weighted sum over distances d, but the weight are all positive numbers which do not modify the relation
$$\frac{1}{k_\alpha} d(\rho_{k_\alpha}, \sigma_{k_\alpha}) \leq \frac{1}{k_\alpha}\Big(d(\rho_{k_\alpha}, \tau_{k_\alpha}) + d(\tau_{k_\alpha}, \sigma_{k_\alpha})\Big)$$
So in the end D satisfies that relation since all the elements of the sum satisfies such relation.

References

[1] M. A. Nielsen and I. L. Chuang. "Quantum Computation and Quantum Information". In: *Cambridge University Press, Cambridge, England* (2000).

[2] F.Anzà and D. Girolami. "Quantifying the Difference between Many-Body Quantum States". In: *PHYSICAL REVIEW LETTERS* 126 (2021). DOI: https://doi.org/10.1103/PhysRevLett.126.170502.

[3] Joseph L. Doob. "Stochastic processes". In: (1957).

[4] D. Girolami. "How Difficult is it to Prepare a Quantum State?" In: *PHYSICAL REVIEW LETTERS* 122 (2019). DOI: https://doi.org/10.1103/PhysRevLett.122.010505.

[5] H. N. Barnum. "PhD thesis". In: *University of New Mexico* (1998).

[6] Åsa Ericsson. "Geodesics and the best measurement for distinguishing quantum states". In: *Journal of Physics A: Mathematical and General* 38 (2005).

[7] A. Uhlmann. "The "transition probability" in the state space". In: *Rep.Math Phy.* 9 (1976). DOI: https://doi.org/10.1016/0034-4877(76)90060-4.

[8] G.Adesso and D. Girolami. "Quantum discord for general two–qubit states: Analytical progress". In: *PHYSICAL REVIEW LETTERS* 83 (2011). DOI: https://doi.org/10.48550/arXiv.1103.3189.

[9] Daniel M. Greenberger, Michael A. Horne, and Anton Zeilinger. "Going Beyond Bell's Theorem". In: *PHYSICS REVIEW LETTERS* (2007). DOI: https://doi.org/10.48550/arXiv.0712.0921.