Universidad de Concepción
Facultad de Ciencias Físicas y Matemáticas

Adaptive Standard Quantum Tomography in high dimensions

Tomografía Cuántica Estándar Adaptativa en altas dimensiones

Tesis para optar al grado de
Magíster en Ciencias con Mención en Física
por

Luciano Iván Pereira Valenzuela

Marzo, 2018

Director de Tesis: **Dr. Aldo Delgado Hidalgo**

Agradecimientos

En primer lugar, quiero agradecer a mi familia por todo el apoyo que me han brindado, en especial a mi hermano Matias, por ser el único en tratar de entender un poco las cosas que hago. También quiero agradecer a mis compañeros de carrera, a los exiliados del sexto piso, al grupo de investigación en informacion cuántica y a toda la gente de la facultad por su buena disposición y simpatía. Debo también agradecer a todos los profesores que me han enseñado durante la carrera, especialmente a mi tutor Aldo Delgado y a Sebastián Niklitschek, por su activa contribución en mi tesis. Quiero dar un particular agradecimiento a mi novia Nicole por estar conmigo todo este tiempo. Por último quiero agradecer a Conicyt y al Núcleo Milenio de Óptica Avanzada por el financiamiento.

Contents

List of Tables

List of Figures

Resumen

El desarrollo de métodos tomográficos precisos es de gran importancia para las tecnologías cuánticas, como lo son la información cuántica, la computación cuántica y la metrología cuántica. La tomografía cuántica ha mostrado ser especialmente difícil en el caso de sistemas de alta dimensión, donde el número de parámetros a determinar crece cuadráticamente con la dimensión. En esta tesis mostramos que una tomografía estándar adaptativa en dos pasos mejora la precisión de la tomografía de un qudit, reduciendo la infidelidad media desde $\mathcal{O}(1/\sqrt{N})$ a $\mathcal{O}(1/N)$ para toda clase de estado. Comparamos este método adaptativo de dos pasos con tomografía estándar y límites relevantes. En particular, mostramos mediante simulaciones numéricas que hasta la dimensión 16 este esquema tomográfico proporciona un mejor rendimiento que la tomografía estándar a igual recurso N para cualquier tipo de estado, lo que lleva a resultados cercanos al límite de Gill-Massar de la infidelidad para estados puros.

Abstract

The development of accurate tomographic methods is of great importance for quantum technologies such as quantum information, quantum computing and quantum metrology. Quantum tomography as shown to be especially difficult in the case of high-dimension systems where the number of parameters to be determined grows quadratically with the dimension. In this thesis, we show that a two-step adaptive standard tomography improves the accuracy of a single qudit tomography by reducing the mean infidelity from $\mathcal{O}(1/\sqrt{N})$ to $\mathcal{O}(1/N)$ for all classes of state. We compare this two-step adaptive method with standard tomography and relevant bounds. In particular, we show by numerical simulations that up to dimension 16 this tomographic scheme provides a better performance than standard tomography at equal resource N for any type of state, leading to results close to the Gill-Massar bound for the infidelity of pure states.

Chapter 1

Introduction

The study of quantum systems of high dimension is one of the great challenges of the quantum theory. These provide several advantages over two-dimensional quantum systems (qubits). For instance, it has been shown that with high dimensional quantum systems (qudits), quantum communication protocols can be implemented more efficiently, secure and resistant to noise [1–3]. Similarly, there are quantum computing algorithms which can be implemented more efficiently through qudits [4–6]. They also allow us to build Bell inequalities that are more resistant to noise and with larger violations of non-locality [7,8]. The reconstruction of unknown quantum states of high dimension becomes a problem of special interest for these applications [9].

The techniques that allow us to reconstruct an unknown state through measurements are known as quantum tomography. These techniques are fundamental in the growing field of quantum technologies. The concept of quantum tomography emerged for the reconstruction of functions of Wigner through quadratures [10, 11]. Nowadays, quantum tomography allows us to characterize photonic systems [12], trapped ions [13, 14], continuous variable states [15], cavity fields [16], atomic ensembles [17] and optical detectors [18, 19]. These systems are widely used in areas such as Quantum Computing [20, 21], Quantum Information [22], Quantum Metrology [23] and Quantum Simulation [24].

The common tomography schemes consist in to determine the d^2-1 independent real parameters which characterize a quantum state ρ by measuring a set of at least d^2-1 operators in an ensemble of many identically prepared copies of the unknown state. We denote the number of copies available as N. The state that minimizes the likelihood function of this data is chosen as the estimated state $\hat{\rho}$. Since the number of parameters to be determined grows quadratically with the dimension, the tomography of states becomes more complicated in high dimensions [14, 25].

Since we can only measure on finite ensembles, the estimated parameters by quantum tomography have uncertainty. This uncertainty, which also might depends on the experimental setup, is limited by a fundamental bound of quantum mechanics: the quantum Cramer-Rao bound [26, 27]. The lowest possible uncertainty is given by the set of measurements which achieve this bound. Finding these measurements in the multiparameter case is difficult and has only been found in some particular cases [27, 28]. Besides, there exist the Gill-Massar inequality [29], which impose another quantum limit for estimates got through individual (non-collective) measurements on each copy of the system. Generally, performing the quantum tomography of an unknown state on an arbitrary basis does not reach the optimum imposed by these limits, so it is not optimal accuracy tomography. The accuracy of the tomography is quantified through a notion of distance between the unknown state and its estimate [30], being the most used the infidelity, because this is inversely proportional to the resource N available for the tomography [31]. However, the infidelity presents a different behavior depending on the unknown state: scale as $\mathcal{O}(1/\sqrt{N})$ for states with small rank and as $\mathcal{O}(1/N)$ for full rank states [32, 33]. Therefore, the accuracy of the tomography also depends on the state to be reconstructed. Being able to improve the accuracy is vital for quantum tomography applications, especially for high-dimensional systems where the number of parameters to estimate grows quadratically with the dimension [14, 25]. It has been proposed to overcome this problem by means of adaptive quantum tomography [34]. This consists of performing a sequence of reconstructions, where we use the information

obtained in previous measurements to then perform a better tomography. In the case of quantum systems of dimension 2, it has been shown that a two-step adaptive tomography reduces infidelity between the unknown state and its estimate at order $\mathcal{O}(1/N)$ [32]: first perform a preliminary low-precision standard tomography; later adapt the measurement bases with the information of the preliminary estimate and perform a second standard tomography. Subsequently, the protocol was extended to reduce the weighted quadratic error for a family of metrics, in order to reach the Gill-Massar bound [35]. Other adaptive methods are the self-learning tomographies. In these, the post measurements are chosen as the positive operator value measurement (POVM) that minimizes a utility function [36–39].

In this thesis, we generalize this adaptive tomographic technique to the case of a single qudit and compare the precision it delivery us with respect to standard tomography , the Cramer-Rao bound and Gill-Massar bound. We show by simulations that until dimension 16 this tomography presents a better performance than the standard tomography at equal resource for any type of state, scaling as $\mathcal{O}(1/N)$, leading to results close to the Gill-Massar bound for the infidelity of pure states.

In chapter 2, we introduce the mathematical framework used in this thesis. In chapter 3, we introduce the foundation of quantum theory. In chapter 4, we briefly review some common tomography schemes. In chapter 5, we study the two-stage adaptive standard tomography in high dimensions and we perform computational simulations to contrast its mean infidelity with the standard tomography error, the Cramer-Rao bound and the Gill-Massar bound. In chapter 6, we present the conclusions of this work.

Chapter 2

Mathematical Framework

Quantum mechanics, like all physical theories, is built using mathematics as a foundation [30]. In this chapter, we review the mathematical framework that will allow us to formulate the quantum theory, and later, understand the main results of this thesis.

First, we review the basic concepts of *Hilbert space* and *linear operator*, emphasizing properties of finite dimensional spaces. These topics allow us to construct the notions of state, evolution and measurement of a quantum system. Then, we review *probability theory* and *statistics*. This is necessary because measurements in quantum mechanics are processes with random outcomes.

For vectors of a Hilbert space, which represents states quantum physical systems, we use the *Dirac notation* $|\psi\rangle$. For other kinds of vectors, like probabilities or coordinate vectors, we use *vector notation* $\boldsymbol{p}$, to distinguish them from quantum states.

2.1 Linear Algebra

In this section we define *Hilbert space*[1] and *linear operator*. We introduce properties like the *Gram-Schmidt procedure*, the *spectral decomposition theorem* and the *Fréchet derivative*. We finish reviewing *direct sum* and *tensor product* between Hilbert spaces.

2.1.1 Hilbert space

A set $\mathcal{H}$ of vectors is called *Hilbert space* if satisfies the following axioms:

1. The space $\mathcal{H}$ is a *linear vector space* on the body of the complex numbers $\mathbb{C}$. That is, there exists two operations called sum $(+)$ and scalar product $(\cdot)$

$$+ : \mathcal{H} \times \mathcal{H} \longrightarrow \mathcal{H}, \tag{2.1}$$

$$\cdot : \mathbb{C} \times \mathcal{H} \longrightarrow \mathcal{H}. \tag{2.2}$$

For $|a\rangle, |b\rangle, |c\rangle \in \mathcal{H}$ and $z, w \in \mathbb{C}$. These operations satisfy

$$\text{Commutativity of sum} \quad |a\rangle + |b\rangle = |b\rangle + |a\rangle . \tag{2.3}$$

$$\text{Associativity of sum} \quad (|a\rangle + |b\rangle) + |c\rangle = |a\rangle + (|b\rangle + |c\rangle). \tag{2.4}$$

$$\text{A zero vector exists} \quad |a\rangle + |0\rangle = |a\rangle . \tag{2.5}$$

$$\text{Each vector has an additive inverse} \quad |a\rangle + |-a\rangle = |0\rangle . \tag{2.6}$$

$$\text{Distributivity of vector sum} \quad z \cdot (|a\rangle + |b\rangle) = z \cdot |a\rangle + z \cdot |b\rangle . \tag{2.7}$$

$$\text{Distributivity of scalar sum} \quad (z + w) \cdot |a\rangle = z \cdot |a\rangle + w \cdot |a\rangle . \tag{2.8}$$

$$\text{Associativity of scalar product} \quad (zw) \cdot |a\rangle = z \cdot (w \cdot |a\rangle). \tag{2.9}$$

$$\text{An identity number exists} \quad 1 \cdot |a\rangle = |a\rangle . \tag{2.10}$$

For simplicity, we omit the dot of scalar product $z \cdot |a\rangle = z\,|a\rangle$ and write the zero vector as $|0\rangle = 0$.

[1] Named after David Hilbert, German mathematician (1862-1943).

2. An *inner scalar product* is defined in $\mathcal{H}$.

$$(\cdot,\cdot) : \mathcal{H} \times \mathcal{H} \longrightarrow \mathbb{C}. \tag{2.11}$$

We denote the scalar product between $|a\rangle$ and $|b\rangle$ by $(|a\rangle, |b\rangle)$. However, we usually employ the Dirac notation, that is, $\langle a|b\rangle$. The scalar product fulfills the following properties

$$\text{Linearity} \quad (|a\rangle, z\,|b\rangle + w\,|c\rangle) = z\,\langle a|b\rangle + w\,\langle a|c\rangle\,. \tag{2.12}$$

$$\text{Hermiticity} \quad \langle a|b\rangle = \langle b|a\rangle^*\,. \tag{2.13}$$

$$\text{Positivity} \quad \begin{cases} \langle a|a\rangle \geq 0, \quad \forall\,|a\rangle\,. \\ \langle a|a\rangle = 0 \Leftrightarrow |a\rangle = |0\rangle\,. \end{cases} \tag{2.14}$$

Using (2.12) and (2.13), we can show that

$$(z\,|a\rangle, |b\rangle) = (|b\rangle, z\,|a\rangle)^* = (z\,\langle b|a\rangle)^* = z^*\,\langle a|b\rangle\,. \tag{2.15}$$

The *norm* of the vector $|a\rangle$ is defined by

$$||\,|a\rangle\,|| = \sqrt{\langle a|a\rangle}, \tag{2.16}$$

which satisfies the *Schwartz's inequality*

$$|\,\langle a|b\rangle\,| \leq ||\,|a\rangle\,|| \quad ||\,|b\rangle\,||. \tag{2.17}$$

Two vectors $|a\rangle$ and $|b\rangle$ are said orthogonal if

$$\langle a|b\rangle = 0. \tag{2.18}$$

3. The space $\mathcal{H}$ is *complete*. That is, every *Cauchy sequence* $\{|a_n\rangle\}_{n\in\mathbb{N}}$ in $\mathcal{H}$ converges in $\mathcal{H}$.

A sequence $\{|a_n\rangle\}_{n\in\mathbb{N}}$ of $\mathcal{H}$ is convergent if

$$\forall\epsilon > 0, \quad \exists N > 0, \quad ||\,|a_n\rangle - |a\rangle\,|| < \epsilon, \quad \forall n > N, \tag{2.19}$$

where $|a\rangle = \lim_{n\to\infty} |a_n\rangle$ is called the limit of the sequence. A sequence $\{|a_n\rangle\}_{n\in\mathbb{N}}$ of $\mathcal{H}$ is called *Cauchy sequence* if

$$\forall \epsilon > 0, \quad \exists N > 0, \quad || \, |a_n\rangle - |a_m\rangle \, || < \epsilon, \quad \forall n, m > N. \tag{2.20}$$

Let us note that every convergent sequence in $\mathcal{H}$ is a Cauchy sequence but not every sequence of Cauchy is convergent.

Vectors of a Hilbert space can be represented using a *basis* $\{|e_n\rangle\}_{n\in\mathcal{N}}$, which is a linearly independent[2] subset of $\mathcal{H}$ that spans $\mathcal{H}$, that is

$$|a\rangle = \sum_{n\in\mathcal{N}} \alpha_n \, |e_n\rangle \;\; \forall \, |a\rangle \in \mathcal{H}. \tag{2.21}$$

The dimension $\dim(\mathcal{H})$ is the number of elements of a basis of $\mathcal{H}$. We say $\mathcal{H}$ is finite dimensional if $\dim(\mathcal{H}) < \infty$ and infinite dimensional if $\dim(\mathcal{H}) = \infty$. For Hilbert spaces with finite dimension, axiom 3 can be obtained from axioms 1 and 2. That is, all vector spaces with inner product of finite dimension are a Hilbert spaces. For simplicity, we call these spaces as finite Hilbert spaces.

A basis is *orthonormal* if

$$\langle e_n | e_m \rangle = \delta_{nm}, \tag{2.22}$$

where δ_{nm} is the *Kronecker delta* defined by

$$\delta_{nm} = \begin{cases} 1 & n = m. \\ 0 & n \neq m. \end{cases} \tag{2.23}$$

Let us consider a non-orthogonal basis $\{|w_n\rangle\}_{n=1,\ldots,d}$ of a d-dimensional Hilbert space. There is a useful method, called *Gram-Schmidt procedure*, that allows us to produce an orthonormal basis $\{|v_n\rangle\}_{n=1,\ldots,d}$ from $\{|w_n\rangle\}_{n=1,\ldots,d}$. These vectors are computed iteratively by

$$\begin{cases} |v_1\rangle = & \dfrac{|w_1\rangle}{|| \, |w_1\rangle \, ||} \\ |v_{k+1}\rangle = & \dfrac{|w_{k+1}\rangle - \sum_{i=1}^{k} \langle v_i | w_{k+1}\rangle \, |v_i\rangle}{|| \, |w_{k+1}\rangle - \sum_{i=1}^{k} \langle v_i | w_{k+1}\rangle \, |v_i\rangle \, ||}, \qquad k = 1, ..., d-1. \end{cases} \tag{2.24}$$

It is not difficult to verify that $\{|v_n\rangle\}_{n=1,\ldots,d}$ is an orthonormal basis by induction,

[2] A set of vectors is linearly independt if one of the vectors can not be written as a linear combination of the others.

- For $n = 1$,

$$\langle v_1|v_2\rangle = \frac{\langle v_1|w_2\rangle - \langle v_1|w_2\rangle \langle v_1|v_1\rangle}{|| \, |w_2\rangle - \langle v_1|w_1\rangle \, |v_1\rangle \, ||} = 0. \tag{2.25}$$

- Let us suppose that the vectors $\{|v_n\rangle\}_{n=1,\dots,d}$ are orthogonal until $n = m$. Testing for $n = m + 1$,

$$\langle v_m|v_{m+1}\rangle = \frac{\langle v_m|w_{m+1}\rangle - \sum_{i=1}^{m} \langle v_i|w_{m+1}\rangle \langle v_m|v_i\rangle}{|| \, |w_{m+1}\rangle - \sum_{i=1}^{m} \langle v_i|w_{m+1}\rangle \, |v_i\rangle \, ||} \tag{2.26}$$

$$= \frac{\langle v_m|w_{m+1}\rangle - \sum_{i=1}^{m} \langle v_i|w_{m+1}\rangle \, \delta_{mi}}{|| \, |w_{m+1}\rangle - \sum_{i=1}^{m} \langle v_i|w_{m+1}\rangle \, |v_i\rangle \, ||} \tag{2.27}$$

$$= \frac{\langle v_m|w_{m+1}\rangle - \langle v_m|w_{m+1}\rangle}{|| \, |w_{m+1}\rangle - \sum_{i=1}^{m} \langle v_i|w_{m+1}\rangle \, |v_i\rangle \, ||} \tag{2.28}$$

$$= 0. \tag{2.29}$$

Thus, any finite Hilbert space has an orthonormal basis. Therefore, the equation (2.21) becomes,

$$|a\rangle = \sum_{n=1}^{d} \alpha_n \, |v_n\rangle \, , \tag{2.30}$$

where $\alpha_n = \langle v_n|a\rangle$. Besides, it can be written in matrix representation, that is,

$$|a\rangle = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_d \end{pmatrix} = \begin{pmatrix} \langle v_1|a\rangle \\ \langle v_2|a\rangle \\ \vdots \\ \langle v_d|a\rangle \end{pmatrix} \in \mathbb{C}^d. \tag{2.31}$$

Moreover, the inner product on a finite Hilbert space in matrix representation is

$$\langle a|b\rangle = \left(\sum_{n=1}^{d} \alpha_n \, |v_n\rangle \, , \sum_{m=1}^{d} \beta_m \, |v_m\rangle \right) \tag{2.32}$$

$$= \sum_{n=1}^{d} \sum_{m=1}^{d} \alpha_n^* \beta_m \, \langle v_n|v_m\rangle \tag{2.33}$$

$$= \sum_{n=1}^{d} \sum_{m=1}^{d} \alpha_n^* \beta_m \delta_{nm} \tag{2.34}$$

$$= \sum_{n=1}^{d} \alpha_n^* \beta_n \tag{2.35}$$

$$= \begin{pmatrix} \alpha_1^* & \cdots & \alpha_d^* \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_d \end{pmatrix}. \tag{2.36}$$

Then, on finite Hilbert spaces, we can identity $\langle a|$ with the transposed conjugate of $|a\rangle$,

$$\langle a| = \begin{pmatrix} \alpha_1^* & \cdots & \alpha_d^* \end{pmatrix}. \tag{2.37}$$

The set of all vectors $\langle a|$ is called the *dual space* $\mathcal{H}^*$.

2.1.2 Linear Operators

A *linear operator* A is a map from $\mathcal{H}$ to $\mathcal{H}'$, which satisfies linearity

$$A(z\,|a\rangle + w\,|b\rangle) = zA\,|a\rangle + wA\,|b\rangle\,. \tag{2.38}$$

We denote by $\mathcal{L}(\mathcal{H})$ the space of all linear operators from $\mathcal{H}$ to itself. A linear operator $A \in \mathcal{L}(\mathcal{H})$ is *bounded* if

$$||A\,|f\rangle\,|| \leq C||\,|f\rangle\,||, \qquad \forall\,|f\rangle \in \mathcal{H}, \qquad C \in \mathbb{R}. \tag{2.39}$$

Because in general the product between operators is not commutative, we define the *commutator* of A and B by

$$[A, B] = AB - BA. \tag{2.40}$$

If A commutes with B, we have $[A, B] = \Theta$, where Θ is the null matrix.

We can represent an operator using the *outer product*. The outer product between $|a\rangle$ and $|b\rangle$ is an operator $|a\rangle\langle b|$, whose action is defined by

$$\Big(\,|a\rangle\langle b|\,\Big)\,|c\rangle = \langle b|c\rangle\,|a\rangle\,. \tag{2.41}$$

Let us consider an orthonormal basis $\{|i\rangle\}_{i=1,\ldots,d}$ of a d-dimensional Hilbert space $\mathcal{H}$, then

$$\left(\sum_{i=1}^{d} |i\rangle\langle i|\right)|v\rangle = \sum_{i=1}^{d} \langle i|v\rangle\,|i\rangle = |v\rangle \tag{2.42}$$

Since the last equation is true for all $|v\rangle \in \mathcal{H}$, we obtain that

$$\sum_{i=1}^{d} |i\rangle\langle i| = \mathbb{I}_{\mathcal{H}}, \tag{2.43}$$

with $\mathbb{I}_{\mathcal{H}}$ the identity operator on $\mathcal{H}$. This is called the *completeness relation*. Thus, we can represent an operator $A \in \mathcal{L}(\mathcal{H})$ on a finite Hilbert space $\mathcal{H}$ as

$$\begin{aligned} A =& \mathbb{I}_{\mathcal{H}} A \mathbb{I}_{\mathcal{H}} && (2.44)\\ =& \left(\sum_{i=1}^{d} |i\rangle\langle i|\right) A \left(\sum_{j=1}^{d} |j\rangle\langle j|\right) && (2.45)\\ =& \sum_{i,j=1}^{d} \langle i|A|j\rangle \, |i\rangle\langle j| && (2.46)\\ =& \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1d} \\ A_{21} & A_{22} & \cdots & A_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ A_{d1} & A_{d2} & \cdots & A_{dd} \end{pmatrix} \in \mathcal{M}_{d\times d}(\mathbb{C}). && (2.47) \end{aligned}$$

where $A_{ij} = \langle i|A|j\rangle$ and $\mathcal{M}_{d\times d}(\mathbb{C})$ is the space of matrices of $d \times d$ over the complex numbers.

An *eigenvector* of an operator A is a non-zero vector $|a\rangle$ such that it is proportional to $A\,|a\rangle$,

$$A\,|a\rangle = a\,|a\rangle\,. \tag{2.48}$$

The proportionality constant $a \in \mathbb{C}$ is known as the eigenvalue of A corresponding to $|a\rangle$. They can be obtained by solving the *characteristic equation*,

$$\det(A - a\mathbb{I}) = 0, \tag{2.49}$$

where det is the *determinant* function for matrices. Two operators that commute (2.40) have the same set of eigenvectors but different sets of eigenvalues.

An important function of the matrix is the *trace*. The trace of A is defined be the sum of its diagonal elements,

$$\mathrm{Tr}(A) = \sum_{i=1}^{d} \langle i|A|i\rangle = \sum_{i=1}^{d} A_{ii}. \tag{2.50}$$

This function is characterized by being cyclic $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$ and linear $\mathrm{Tr}(\alpha A + \beta B) = \alpha\,\mathrm{Tr}(A) + \beta\,\mathrm{Tr}(B)$. Besides, this is equal to the sum of the eigenvalues of the matrix $\mathrm{Tr}(A) = \sum_{i=1}^{d} a_i$.

The *adjoint operator* $A^\dagger \in \mathcal{L}(\mathcal{H})$ of a operator $A \in \mathcal{L}(\mathcal{H})$ satisfies the following equation

$$(A^\dagger |a\rangle, |b\rangle) = (|a\rangle, A|b\rangle). \tag{2.51}$$

On finite Hilbert spaces, the adjoint operator $A^\dagger$ is the transposed conjugate of A,

$$A^\dagger = \begin{pmatrix} A_{11}^* & A_{21}^* & \cdots & A_{d1}^* \\ A_{12}^* & A_{22}^* & \cdots & A_{d2}^* \\ \vdots & \vdots & \ddots & \vdots \\ A_{1d}^* & A_{2d}^* & \cdots & A_{dd}^* \end{pmatrix}. \tag{2.52}$$

In particular, an operator $A \in \mathcal{B}(\mathcal{H})$ is called *hermitian* if

$$A = A^\dagger. \tag{2.53}$$

An eigenvalue a with eigenvector $|a\rangle$ of an hermitian operator A is real,

$$a = \langle a|A|a\rangle = \langle a|A^\dagger|a\rangle^* = \langle a|A|a\rangle^* \in \mathbb{R}. \tag{2.54}$$

An important class of hermitian operators are the *positive semidefinite operators*. An operator A is called positive semidefinite if

$$\langle v|A|v\rangle \geq 0, \qquad \forall\, |v\rangle \neq 0. \tag{2.55}$$

Then, these operators have non-negative eigenvalues. We denote by $\mathcal{P}(\mathcal{H})$ the set of all positive operators that maps $\mathcal{H}$ on itself.

A operator $U \in \mathcal{L}(\mathcal{H})$ is called *unitary* if it is invertible and $UU^\dagger = U^\dagger U = \mathbb{I}$. These matrices form the *special unitary group* $SU(d)$. Let $|u\rangle$ be an eigenvector of U with eigenvalue u, then

$$\langle u|U^\dagger U|u\rangle = \langle u|\mathbb{I}|u\rangle \Longrightarrow |u| = 1. \tag{2.56}$$

The trace of A is invariant under unitary transformations $\mathrm{Tr}\big(U^\dagger A U\big) = \mathrm{Tr}(A)$.

An operator $A \in \mathcal{L}(\mathcal{H})$ is called normal if $AA^\dagger = A^\dagger A$. These operators satisfy the following useful theorem,

Theorem 1. *(Spectral decomposition[3]) Any normal operator A on a d-dimensional Hilbert space $\mathcal{H}$ is diagonal with respect to the eigenvector basis of A. That is, A can be written as*

$$A = \sum_{i=1}^{d} a_i \left|a_i\right\rangle\!\left\langle a_i\right|, \tag{2.57}$$

where $\{|a_i\rangle\}_{i=1,\ldots,d}$ are eigenvectors of A with eigenvalues $\{a_i\}_{i=1,\ldots,d}$.

The equation (2.57) can be rewritten using *projective operators*. An operator P is projective if it satisfies

$$P^2 = P. \tag{2.58}$$

The operator $P_i = |a_i\rangle\langle a_i|$ is the projector onto the subspace with eigenvalue a_i. Then,

$$A = \sum_{i=1}^{d} a_i P_i. \tag{2.59}$$

Clearly, hermitian operators and unitary operators are normal. Therefore, these operators have a spectral decomposition. Furthermore, these also play an important role in the formulation of Quantum Mechanics.

The function f of a normal operator A is defined by the following operator

$$f(A) = \sum_{i=1}^{d} f(a_i) \left|a_i\right\rangle\!\left\langle a_i\right|, \tag{2.60}$$

that is, a function of operators is defined by the values of the function on the set of eigenvalues of the operators. One of the most commonly used functions is the exponential e^A. An unitary operator can be written as $U = e^{-\mathrm{i}H}$, where H is a hermitian operator. The exponential of an operator fulfill the Baker-Campbell-Hausdorff formula,

$$e^A B e^{-A} = B + [A, B] + \frac{1}{2!}[A, [A, B]] + \frac{1}{3!}[A, [A, [A, B]]] + \cdots. \tag{2.61}$$

[3] See proof in [30].

As a last observation, let us note that the set of linear operators $\mathcal{L}(\mathcal{H})$, endowed with the *Hilbert-Schmidt product*,

$$\langle A, B\rangle = \mathrm{Tr}\left(A^\dagger B\right), \tag{2.62}$$

as an inner product, is a Hilbert space. This product induces a norm on operators,

$$||A|| = \sqrt{\langle A, A\rangle} = \sqrt{\mathrm{Tr}(A^\dagger A)}. \tag{2.63}$$

2.1.3 Fréchet Derivative

In general, we can define notions of derivation on Hilbert spaces, like in $\mathcal{H} = \mathbb{C}^d$. Therefore, we can derivate functions of operators. Let us suppose a finite Hilbert space $\mathcal{H}$. A continuous function f from $\mathcal{U} \subseteq \mathcal{L}(\mathcal{H})$ to $\mathcal{L}(\mathcal{H})$ is said to be *differentiable* at a point $U \in \mathcal{U}$ if there exists $T \in \mathcal{L}(\mathcal{H})$ such that

$$\lim_{V\to\Theta} \frac{||f(U+V) - f(U) - TV||}{||V||} = 0. \tag{2.64}$$

The operator $T = Df(U)$ above is called the *Fréchet derivative*[4] of f at U [40]. Then, if f is differentiable at U, for all $V \in \mathcal{L}(\mathcal{H})$,

$$Df(U)(V) = \lim_{t_0\to 0} \frac{f(U+t_0V) - f(U)}{t_0V} \tag{2.65}$$

$$= \lim_{t_0\to 0} \frac{f\big(U+(t+t_0)V\big) - f(U+tV)}{t_0V}\bigg|_{t=0} \tag{2.66}$$

$$= \frac{\mathrm{d}}{\mathrm{d}t} f(U+tV)\bigg|_{t=0}. \tag{2.67}$$

This is called the *directional derivative* of f at U in the direction V. From (2.67), we can see that the Fréchet derivative inherits the usual rules of derivation:

$$\text{Linearity} \quad D(f+g)(U) = Df(U) + Dg(U). \tag{2.68}$$

$$\text{Chain rule} \quad D(g(f))(U) = Dg(f(U)) \cdot Df(U). \tag{2.69}$$

$$\text{Product rule} \quad D[f(U)\cdot g(U)](V) = Df(U)(V)\cdot g(U) + f(U)\cdot Dg(U)(V). \tag{2.70}$$

[4] Named after Maurice René Fréchet, French mathematician (1878-1973).

Derivatives of higher order can be defined by repeating derivatives. Thus, the pth derivative of f is calculated as

$$D^p f(U)(V_1, ..., V_p) = \frac{\partial^p}{\partial t_1 \cdots \partial t_p} f(U + t_1 V_1 + \cdots + t_p V_p)\Big|_{t_1=...=t_p=0}. \tag{2.71}$$

Using the Fréchet derivative, we can show the following useful theorem:

Theorem 2. *(Taylor's Theorem) Let f be a $(p+1)$ times differentiable function from $\mathcal{L}(H)$ to itself. Then, for all $X \in \mathcal{L}(H)$ and for small $H \in \mathcal{L}(H)$,*

$$\left|\left| f(X+H) - f(X) - \sum_{m=1}^{p} \frac{1}{m!} D^m f(X)([H]^m) \right|\right| = \mathcal{O}\Big(||H||^{p+1}\Big). \tag{2.72}$$

where $[H]^m$ is the m-tuple $(H, H, ..., H)$.

2.1.4 Direct Sum and Tensor Product

We can combine two Hilbert spaces to form larger spaces taking their *direct sum* or their *tensor product*. Let us consider a n-dimensional Hilbert space $\mathcal{H}$ and a m-dimensional Hilbert space $\mathcal{H}'$, with $\{|v_i\rangle\}_{i=1,...,n}$ and $\{|w_i\rangle\}_{i=1,...,m}$ their orthonormal bases, respectively.

First, the sum direct space $\mathcal{H} \oplus \mathcal{H}'$ is a $n+m$ dimensional Hilbert space. These spaces must only intersect at the zero vector, $\mathcal{H} \cap \mathcal{H}' = \{0\}$. An orthonormal basis of $\mathcal{H} \oplus \mathcal{H}'$ is $\{|v_i\rangle, |w_j\rangle\}$, with $i = 1, ..., n$ and $j = 1, ..., m$. Vectors of $\mathcal{H} \oplus \mathcal{H}'$ have the matrix form

$$|v\rangle \oplus |w\rangle = \begin{pmatrix} |v\rangle \\ |w\rangle \end{pmatrix}, \tag{2.73}$$

On the other hand, the tensor product space $\mathcal{H} \otimes \mathcal{H}'$ is a nm dimensional Hilbert space. If $|v\rangle \in \mathcal{H}$ and $|w\rangle \in \mathcal{H}'$, the tensor product vector $|v\rangle \otimes |w\rangle$ is on $\mathcal{H} \otimes \mathcal{H}'$. These vectors fulfill bilinearity

$$|v\rangle \otimes (z_1 |w_1\rangle + z_2 |w_2\rangle) = z_1 |v\rangle \otimes |w_1\rangle + z_2 |v\rangle \otimes |w_2\rangle, \tag{2.74}$$

$$(z_1 |v_1\rangle + z_2 |v_2\rangle) \otimes |w\rangle = z_1 |v_1\rangle \otimes |w\rangle + z_2 |v_2\rangle \otimes |w\rangle. \tag{2.75}$$

An orthonormal basis of $\mathcal{H} \otimes \mathcal{H}'$ is $\{|v_i\rangle \otimes |w_j\rangle\}$, with $i = 1, ..., n$ and $j = 1, ..., m$, and its vectors are

$$|a\rangle_{\mathcal{H}\otimes\mathcal{H}'} = \sum_{i=1}^{n}\sum_{j=1}^{m} \alpha_{ij} |v_i\rangle_{\mathcal{H}} \otimes |w_j\rangle_{\mathcal{H}'} . \tag{2.76}$$

The tensor product can be computed using the *Kronecker product*,

$$|a\rangle \otimes |b\rangle = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_d \end{pmatrix} \otimes |b\rangle = \begin{pmatrix} \alpha_1 |b\rangle \\ \alpha_2 |b\rangle \\ \vdots \\ \alpha_d |b\rangle \end{pmatrix}. \tag{2.77}$$

These operations are also valid for linear operators, because these are also Hilbert spaces. For finite dimensional operators,

$$A \oplus B = \begin{pmatrix} A & \Theta_{n,m} \\ \Theta_{m,n} & B \end{pmatrix}, \tag{2.78}$$

$$A \otimes B = \begin{pmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & \ddots & \vdots \\ A_{n1} & \cdots & A_{nn} \end{pmatrix} \otimes B = \begin{pmatrix} A_{11}B & \cdots & A_{1n}B \\ \vdots & \ddots & \vdots \\ A_{n1}B & \cdots & A_{nn}B \end{pmatrix}, \tag{2.79}$$

where $\Theta_{n,m}$ is the null matrix of $n \times m$.

2.2 Probability and Statistics

In this section, we introduce basic concepts of probability theory and classical statistics, such as Fisher Information. These are necessary due to the statistical nature of Quantum Mechanics.

2.2.1 Sample Space, Events and Probabilities

A random experiment is characterized by the impossibility of predicting with certainty the result of it. The set of all possible outcomes ω of a random experiment is called the *sample space* Ω. If this set is countable, we call it discrete, otherwise we call it continuous.

An event E is a subset of Ω. The set $\{E_i\}_{i=1,\ldots,n}$ of events is called mutually exclusive if $E_i \cap E_j = \emptyset$ for all i and j, with $i \neq j$. That is, if an outcome of event E_i occurs, this event occurs and none of the other events can happen.

The *probability* p is a measure which acts on a sample space Ω and assigns a number between 0 and 1 to each event. Probability must satisfy the following conditions

1. $p(E) \geq 0$ for all events E.
2. If $E = \Omega$, then $p(E) = 1$.
3. If the events $\{E_i\}_{i=1,\ldots,n}$ are mutually exclusive,

$$p\left(\cup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} p(E_i). \tag{2.80}$$

In many experiments, the probability of an event E_1 depends on another event E_2 having already occurred. This probability is called *conditional probability* and it is defined by

$$p(E_1|E_2) = \frac{P(E_1, E_2)}{p(E_2)}. \tag{2.81}$$

where $p(E_2) \neq 0$.

Theorem 3. *(Bayes' Theorem) The conditional probability $p(E_1|E_2)$ satisfies*

$$p(E_1|E_2) = \frac{p(E_2|E_1)p(E_1)}{p(E_2)}. \tag{2.82}$$

where $p(E_2) \neq 0$.

Proof: From (2.81),

$$p(E_1|E_2) = \frac{P(E_1, E_2)}{p(E_2)}, \qquad p(E_2|E_1) = \frac{P(E_2, E_1)}{p(E_1)}. \tag{2.83}$$

Because $p(E_1, E_2) = p(E_2, E_1)$, we get

$$p(E_1|E_2)p(E_2) = p(E_2|E_1)p(E_1). \tag{2.84}$$

Finally, if $p(E_2) \neq 0$,

$$p(E_1|E_2) = \frac{p(E_2|E_1)p(E_1)}{p(E_2)}. \tag{2.85}$$

2.2.2 Random Variables

A random variable X is a function that, to each result $\omega \in \Omega$ assigns it a real number. When we perform an experiment on a discrete random variable X, it may take the value x with probability

$$p(X = x) = p(\{\omega : X(\omega) = x\}). \tag{2.86}$$

For simplicity, we define $p(x) = p(X = x)$. On the other hand, when the random variable X is continuous, it may take a value between x_1 and x_2 with probability

$$p(x_1 < X \leq x_2) = p(\{\omega : x_1 < X(\omega) \leq x_2\}). \tag{2.87}$$

A better form to express this probability is using the *probability density* f

$$p(x_1 < X \leq x_2) = \int_{x_1}^{x_2} f(x)dx. \tag{2.88}$$

This function must satisfy

$$f(x) \geq 0, \quad \forall x \in \mathbb{R}, \tag{2.89}$$

$$\int_{\mathbb{R}} f(x)dx = 1. \tag{2.90}$$

We can use the *Dirac Delta* to write the probability density of a discrete random variable,

$$f(x) = \sum_{x_0 \in \mathcal{I}} p(x_0)\delta(x - x_0). \tag{2.91}$$

where $\mathcal{I}$ is the set of all possible values for X. We say that two random variables X and Y are independent if

$$p(x, y) = p(x)p(y), \qquad \text{if } X \text{ is discrete.} \tag{2.92}$$

$$f(x, y) = f_X(x)f_Y(y), \qquad \text{if } X \text{ is continuous.} \tag{2.93}$$

The *expected value* of a random variable X is the weighted average of its values, using the probabilities as weights. The expected value of a random variable X is

defined by

$$\mathbb{E}(X) = \begin{cases} \sum_{x \in \mathcal{I}} xp(x), & \text{if } X \text{ is discrete.} \\ \int_{\mathbb{R}} xf(x)dx, & \text{if } X \text{ is continuous.} \end{cases} \tag{2.94}$$

Then, the expected value of the product between two independent random variables X and Y is

$$\mathbb{E}(XY) = \int_{\mathbb{R}} \int_{\mathbb{R}} xyf(x,y)dxdy \tag{2.95}$$

$$= \int_{\mathbb{R}} \int_{\mathbb{R}} xyf_X(x)f_Y(y)dxdy \tag{2.96}$$

$$= \int_{\mathbb{R}} xf_X(x)dx \int_{\mathbb{R}} yf_Y(y)dy \tag{2.97}$$

$$= \mathbb{E}(X)\mathbb{E}(Y). \tag{2.98}$$

The expected value of a function $g(X)$ is

$$\mathbb{E}(g(X)) = \begin{cases} \sum_{x \in \mathcal{I}} g(x)p(x), & \text{if } X \text{ is discrete.} \\ \int_{\mathbb{R}} g(x)f(x)dx, & \text{if } X \text{ is continuous.} \end{cases} \tag{2.99}$$

Using (2.81), we can define the *conditional probability* of discrete random variables and the *conditional density probability* of continuous random variables,

$$p(x|y) = \frac{p(x,y)}{p(y)}, \qquad p(y) \neq 0. \tag{2.100}$$

$$f_X(x|y) = \frac{f(x,y)}{f_Y(y)}, \qquad f_Y(y) \neq 0. \tag{2.101}$$

Then, the *conditional expected value* of X given Y is defined as

$$\mathbb{E}(X|Y) = \begin{cases} \sum_{x \in \mathcal{I}} xp(x|y), & \text{if } X \text{ is discrete.} \\ \int_{\mathbb{R}} xf_X(x|y)dx, & \text{if } X \text{ is continuous.} \end{cases} \tag{2.102}$$

In general, we can work with a set of n random variables

$$\boldsymbol{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \in \mathbb{R}^n, \tag{2.103}$$

and its corresponding expected value

$$\mathbb{E}(\boldsymbol{X}) = \begin{pmatrix} \mathbb{E}(X_1) \\ \vdots \\ \mathbb{E}(X_n) \end{pmatrix}. \tag{2.104}$$

Let us note that the expected value between sets of random variables $\boldsymbol{X}$ and $\boldsymbol{Y}$ is an inner product. Then, this satisfies the Cauchy-Schwarz inequality,

$$|\mathbb{E}(\boldsymbol{XY})|^2 \leq \mathbb{E}(\boldsymbol{X}^2)\mathbb{E}(\boldsymbol{Y}^2). \tag{2.105}$$

The *variance* of a random variable X is the expected value of the squared deviation of a random variable X from its expected value $\mathbb{E}(X)$, that is,

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}(X))^2]. \tag{2.106}$$

It can be rewritten as

$$\begin{aligned} \mathrm{Var}(X) =& \mathbb{E}[(X - \mathbb{E}(X))^2] && (2.107)\\ =& \mathbb{E}[X^2 - 2X\mathbb{E}(X) + \mathbb{E}(X)^2] && (2.108)\\ =& \mathbb{E}(X^2) - 2(X)\mathbb{E}(X) + \mathbb{E}(X)^2 && (2.109)\\ =& \mathbb{E}(X^2) - \mathbb{E}(X)^2. && (2.110) \end{aligned}$$

The square root of the variance is called the *standard deviation* $\sigma(X) = \sqrt{\mathrm{Var}(X)}$. This represents the dispersion of data around the expected value. The *confidence interval* is defined by

$$|X - \mathbb{E}(X)| \leq \epsilon\sigma(X) \tag{2.111}$$

with $\epsilon > 1$. The probability that X assumes a value in this interval is

$$p\Big(|X - \mathbb{E}(X)| \leq \epsilon\sigma(X)\Big) \geq 1 - \frac{1}{\epsilon^2}. \tag{2.112}$$

For instance, the random variable has a probability $p = 1 - 1/9 = 8/9$ of assuming a value in the interval $|X - \mathbb{E}(X)| \leq 3\sigma(X)$. This leads that the error in a random experiment is on the magnitude order of the standard deviation,

$$x - \mathbb{E}(X) \sim \mathcal{O}\big(\sigma(X)\big). \tag{2.113}$$

where we have supposed that the random variable X assumes the value x.

The *covariance* between two random variables X and Y is defined by

$$\mathrm{Cov}(X,Y) = \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))]. \tag{2.114}$$

It is a measurement of the correlation among the random variables X and Y. In particular $\mathrm{Cov}(X,X) = \mathrm{Var}(X)$. Analogous to the variance,

$$\begin{aligned}
\mathrm{Cov}(X,Y) =& \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))] && (2.115)\\
=& \mathbb{E}[XY - \mathbb{E}(X)Y - X\mathbb{E}(Y) + \mathbb{E}(X)\mathbb{E}(Y)] && (2.116)\\
=& \mathbb{E}(XY) - 2\mathbb{E}(X)\mathbb{E}(Y) + \mathbb{E}(X)\mathbb{E}(Y) && (2.117)\\
=& \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y). && (2.118)
\end{aligned}$$

The variance of a linear combination of two random variables X and Y

$$\begin{aligned}
\mathrm{Var}(aX + bY) =& \mathbb{E}([aX + bY]^2) - \mathbb{E}(aX + bY)^2 && (2.119)\\
=& a^2\mathbb{E}(X^2) + 2ab\mathbb{E}(XY) + b^2\mathbb{E}(Y^2) - a^2\mathbb{E}(X)^2 \\
& - 2ab\mathbb{E}(X)\mathbb{E}(Y) - b^2\mathbb{E}(Y)^2 && (2.120)\\
=& a^2[\mathbb{E}(X^2) - \mathbb{E}(X)^2] + b^2[\mathbb{E}(Y^2) - \mathbb{E}(Y)^2] \\
& + 2ab[\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)] && (2.121)\\
=& a^2\mathrm{Var}X + b^2\mathrm{Var}(Y) + 2ab\mathrm{Cov}(X,Y). && (2.122)
\end{aligned}$$

If we are working with a set of random variables $\boldsymbol{X}$, we have the *covariance matrix* $C(\boldsymbol{X})$, with elements

$$C_{ij}(\boldsymbol{X}) = \mathrm{Cov}(X_i, X_j). \tag{2.123}$$

Let us suppose a vectorial function $\boldsymbol{f}(\boldsymbol{X}) = A\boldsymbol{X}$, where $A \in \mathbb{C}^{n\times n}$. The elements of this function are

$$f_i(\boldsymbol{X}) = \sum_{l=1}^{n} A_{il}X_l. \tag{2.124}$$

Its covariance matrix is

$$C_{ij}(\boldsymbol{f}) = \mathrm{Cov}[f_i(X), f_j(X)] \tag{2.125}$$

$$=\mathbb{E}[f_i(X)f_j(X)] - \mathbb{E}[f_i(X)]\mathbb{E}[f_j(X)] \tag{2.126}$$

$$=\mathbb{E}\left[\sum_{l=1}^{n} A_{il}X_l \sum_{k=1}^{n} A_{jk}X_k\right] - \mathbb{E}\left[\sum_{l=1}^{n} A_{il}X_l\right]\mathbb{E}\left[\sum_{k=1}^{n} A_{ik}X_k\right] \tag{2.127}$$

$$=\sum_{l=1}^{n}\sum_{k=1}^{n} A_{il}A_{jk}\left[\mathbb{E}(X_lX_k) - \mathbb{E}(X_l)\mathbb{E}(X_k)\right] \tag{2.128}$$

$$=\sum_{l=1}^{n}\sum_{k=1}^{n} A_{il}A_{jk}\mathrm{Cov}(X_l, X_k) \tag{2.129}$$

$$=\sum_{l=1}^{n}\sum_{k=1}^{n} A_{il}\mathrm{Cov}(X_l, X_k)A_{jk} \tag{2.130}$$

$$=(ACA^{\mathrm{T}})_{ij}. \tag{2.131}$$

For a non-linear function, we can linearize employing the Taylor series up to the first order.

2.2.3 Fisher Information

Let us consider a random experiment where we want to estimate a set of n parameters $\boldsymbol{s} = \begin{pmatrix} s_1 & \cdots & s_n \end{pmatrix}^{\mathrm{T}}$. Let us consider that the probability distributions of this experiment is $\{p(x|\boldsymbol{s})\}_{x\in\mathcal{I}}$, with $\mathcal{I}$ the set of all outcomes x. From the definition of probability,

$$\sum_{x\in\mathcal{I}} p(x|\boldsymbol{s}) = 1 \qquad \forall \boldsymbol{s}. \tag{2.132}$$

Deriving the last equation

$$\frac{\partial}{\partial s_i}\left[\sum_{x\in\mathcal{I}} p(x|\boldsymbol{s})\right] = \sum_{x\in\mathcal{I}} p(x|\boldsymbol{s})\frac{\partial}{\partial s_i}\ln p(x|\boldsymbol{s}) = \mathbb{E}\left(\frac{\partial}{\partial s_i}\ln p(x|\boldsymbol{s})\right) = 0. \tag{2.133}$$

From the experiment we obtain estimations $\hat{\boldsymbol{s}}(x) = \begin{pmatrix} \hat{s}_1(x) & \cdots & \hat{s}_n(x) \end{pmatrix}^{\mathrm{T}}$, which are functions of the experimental outcomes. Besides,

$$\frac{\partial \hat{s}_j(x)}{\partial s_i} = 0. \tag{2.134}$$

Using the last equation we can define the following matrix

$$G_{ij} = \frac{\partial}{\partial s_i}\mathbb{E}\big(\hat{s}_j(x)\big) \tag{2.135}$$

$$= \frac{\partial}{\partial s_i} \left[\sum_{x \in \mathcal{I}} p(x|\boldsymbol{s}) \hat{s}_j(x) \right] \tag{2.136}$$

$$= \mathbb{E} \left(\frac{\partial}{\partial s_i} \ln p(x|\boldsymbol{s}) \hat{s}_j(x) \right), \tag{2.137}$$

Subtracting the equation (2.133),

$$G_{ij} = \mathbb{E} \left(\frac{\partial}{\partial s_i} \ln p(x|\boldsymbol{s}) \hat{s}_j(x) \right) - \mathbb{E} \left(\frac{\partial}{\partial s_i} \ln p(x|\boldsymbol{s}) \right) s_j \tag{2.138}$$

$$= \mathbb{E} \left(\frac{\partial}{\partial s_i} \ln p(x|\boldsymbol{s}) \big[\hat{s}_j(x) - s_j \big] \right). \tag{2.139}$$

For any vectors $\boldsymbol{u}$ and $\boldsymbol{w}$ in $\mathbb{R}^n$, we get

$$(\boldsymbol{u}^{\mathrm{T}} G \boldsymbol{w})^2 = \left(\sum_{i,j=1}^{n} u_i G_{ij} w_j \right)^2 \tag{2.140}$$

$$= \left[\sum_{i,j=1}^{n} u_i \mathbb{E} \left(\frac{\partial}{\partial s_i} \ln p(x|\boldsymbol{s}) [\hat{s}_j(x) - s_j] \right) w_j \right]^2 \tag{2.141}$$

$$= \left[\mathbb{E} \left(\sum_{i,j=1}^{n} u_i \frac{\partial}{\partial s_i} \ln p(x|\boldsymbol{s}) [\hat{s}_j(x) - s_j] w_j \right) \right]^2. \tag{2.142}$$

Using the Schwartz's inequality (2.105), we obtain

$$(\boldsymbol{u}^{\mathrm{T}} G \boldsymbol{w})^2 \leq \mathbb{E} \left(\left[\sum_{i=1}^{n} u_i \frac{\partial}{\partial s_i} \ln p(x|\boldsymbol{s}) \right]^2 \right) \mathbb{E} \left(\left[\sum_{j=1}^{n} [\hat{s}_j(x) - s_j] w_j \right]^2 \right) \tag{2.143}$$

$$= \left(\boldsymbol{u}^{\mathrm{T}} I \boldsymbol{u} \right) \left(\boldsymbol{w}^{\mathrm{T}} C \boldsymbol{w} \right), \tag{2.144}$$

where C is the covariance matrix and I is the Fisher information matrix[5] defined by

$$I_{ij} = \mathbb{E} \left(\frac{\partial}{\partial s_i} \ln p(x|\boldsymbol{s}) \frac{\partial}{\partial s_j} \ln p(x|\boldsymbol{s}) \right). \tag{2.145}$$

Therefore, for all $\boldsymbol{u}$ and $\boldsymbol{w}$, we have

$$\boldsymbol{w}^{\mathrm{T}} C \boldsymbol{w} \geq \frac{\boldsymbol{u}^{\mathrm{T}} G \boldsymbol{w} \boldsymbol{w}^{\mathrm{T}} G \boldsymbol{u}}{\boldsymbol{u}^{\mathrm{T}} I \boldsymbol{u}} \tag{2.146}$$

The right-hand side of this inequality is maximized when $\boldsymbol{u} = I^{-1} G \boldsymbol{w}$,

$$\boldsymbol{w}^{\mathrm{T}} C \boldsymbol{w} \geq \frac{\boldsymbol{u}^{\mathrm{T}} G \boldsymbol{w} \boldsymbol{w}^{\mathrm{T}} G I^{-1} G \boldsymbol{w}}{\boldsymbol{u}^{\mathrm{T}} I I^{-1} G \boldsymbol{w}} \tag{2.147}$$

[5]Named after Sir Ronald Fisher, British statistician and geneticist (1890-1962).

$$\boldsymbol{w}^{\mathrm{T}} C \boldsymbol{w} \geq \frac{\boldsymbol{u}^{\mathrm{T}} G \boldsymbol{w} \boldsymbol{w}^{\mathrm{T}} G I^{-1} G \boldsymbol{w}}{\boldsymbol{u}^{\mathrm{T}} G \boldsymbol{w}} \tag{2.148}$$

$$\boldsymbol{w}^{\mathrm{T}} C \boldsymbol{w} \geq \boldsymbol{w}^{\mathrm{T}} G I^{-1} G \boldsymbol{w}. \tag{2.149}$$

Thus, we obtain the following inequality

$$C \geq G I^{-1} G, \tag{2.150}$$

which is called the *Cramer-Rao inequality* [41, 42]. This inequality means that the matrix $C - GI^{-1}G$ is positive semidefinite. It expresses a lower bound for the covariance. A particular case of this inequality occurs when the estimator $\hat{\boldsymbol{s}}$ is unbiased,

$$\mathbb{E}[\hat{\boldsymbol{s}}(x)] = \boldsymbol{s}. \tag{2.151}$$

Here the matrix G is the identity matrix because

$$G_{ij} = \frac{\partial}{\partial s_i} \mathbb{E}[\hat{s}_j(x)] = \frac{\partial s_j}{\partial s_i} = \delta_{ij}. \tag{2.152}$$

The Cramer-Rao inequality for unbiased estimators is then

$$C \geq I^{-1}. \tag{2.153}$$

This establishes that the covariance matrix is bounded by below by the Fisher information matrix. An unbiased estimation method is optimal if it reaches the Cramer-Rao bound (2.153). It should be noted that there are not always unbiased techniques that achieve this lowe bound.

From (2.153), we can also find a bound for the *weighted mean square error* $\mathrm{Tr}(WC)$, where W is a positive matrix. For example, if $W = \mathbb{I}$, this is the mean square error,

$$\mathrm{Tr}(C) = \sum_{i=1}^{n} \mathbb{E}\big([\hat{s}_i(x) - s_i]^2\big). \tag{2.154}$$

The Cramer-Rao bound for the weighted mean square error is

$$\mathrm{Tr}(WC) \geq \mathrm{Tr}\big(WI^{-1}\big). \tag{2.155}$$

Chapter 3

Quantum Mechanics

Quantum Mechanics is a theory which allows us to describe the behavior of microscopic physical systems, such as atoms and photons. The quantum theory has deep differences with classical theories. The main one is the measurement process. Classical physics implicitly assumes the properties of systems to be unchanged by measurement processes. Thereby, different properties can be measured without any interference among them. This is not the case in quantum physics, where the measurements have random outcomes and they disturb the state of the system. Thereby, the measurement of a property typically precludes the knowledge about other properties. Quantum Mechanics has allowed the development of many technological applications, such as Quantum Information, Quantum Computation and Quantum Cryptography.

In this chapter, we briefly enunciate the *postulates of Quantum Mechanics*. Then, we present notions of distance between quantum states. We finish studying the quantum theory of estimation.

3.1 Postulates of Quantum Mechanics

In this section, we review shortly all the postulates of quantum mechanics. There are four basic postulates which tell us how to describe all we need to know about a quantum system [30, 43].

3.1.1 State Space

The first postulate teaches us how to mathematically represent quantum systems.

Postulate 1. *Every physical system is associated with a complex vector space with inner product $\mathcal{H}$ (called a Hilbert space). The state vector of the system is a ray in said Hilbert space. This vector completely describes the system.*

A ray is a one-dimensional subspace of $\mathcal{H}$. Then, the states $|\psi\rangle$ can be identified with an equivalence class[1] represented by a unitary vector $|\langle\psi|\psi\rangle| = 1$. Also, the state $|\psi\rangle$ is equal to $e^{i\varphi}|\psi\rangle$ up the *global phase* $e^{i\varphi}$. We call a 2-dimensional quantum system a *qubit (quantum bit)*[2],

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle \tag{3.1}$$

where $\{|0\rangle, |1\rangle\}$ is the basis of the space and $||\,|\psi\rangle\,|| = |\alpha|^2 + |\beta|^2 = 1$. In general, a d dimensional state is called a *qudit (quantum dit)*.

An alternative formulation is possible using a tool known as the *density operator* or *density matrix*. The language of the density matrices is convenient to describe systems that are not known in their entirety. This is a system that can't be described only with a state $|\psi\rangle$, instead we need a set of n states $|\psi_i\rangle$ with their respective probabilities p_i, called the *ensemble of pure states* $\{p_i, |\psi_i\rangle\}_{i=1,\ldots,n}$. The density operator of the system is defined by the equation

$$\rho = \sum_{i=1}^{n} p_i |\psi_i\rangle\langle\psi_i| . \tag{3.2}$$

[1] An equivalence class is a set which satisfies an equivalence relation. A relation is of equivalence if it fulfills: reflexivity, symmetry and transitivity.

[2] Inspired by the bit notion of classical computing.

Note that the expansion (3.2) does not necessarily coincide with the spectral decomposition of ρ and n can be different than $\dim(\mathcal{H})$. The class of operators that are density operators are characterized by the following useful theorem:

Theorem 4. *An operator ρ is a density operator associated to some ensemble $\{p_i, |\psi_i\rangle\}_{i=1,\dots,n}$ if and only if it satisfies the conditions:*

1. *The trace of ρ is equal to 1. That is, $\mathrm{Tr}(\rho) = 1$,*

2. *$\rho \in \mathcal{P}$. That is, ρ is positive semidefinite operator.*

Proof: Suppose $\rho = \sum_{i=1}^{n} p_i \, |\psi_i\rangle\langle\psi_i|$ is a density operator. Then

$$\mathrm{Tr}(\rho) = \sum_{i=1}^{n} p_i \, \mathrm{Tr}(|\psi_i\rangle\langle\psi_i|) = \sum_{i=1}^{n} p_i = 1, \tag{3.3}$$

so the trace condition 1 is satisfied. Suppose $|\varphi\rangle$ is an arbitrary vector in the state space. Then

$$\langle\varphi|\rho|\varphi\rangle = \sum_{i=1}^{n} p_i \, \langle\varphi|\psi_i\rangle \, \langle\psi_i|\varphi\rangle = \sum_{i=1}^{n} p_i |\, \langle\varphi|\psi_i\rangle \,|^2 \geq 0, \tag{3.4}$$

so the condition 2 is satisfied.

Density matrices are classified into two types: if there exists a vector $|\psi\rangle$ such that $\rho = |\psi\rangle\langle\psi|$, then we say that the state is pure; otherwise we say that the state is mixed. A state ρ can be determined as pure or mixed by evaluating the *purity* $\mathrm{Tr}(\rho^2)$. Let us expand ρ in its spectral descomposition,

$$\rho = \sum_{i=1}^{n} \lambda_i \, |\lambda_i\rangle\langle\lambda_i| \, . \tag{3.5}$$

Thus, calculating the purity,

$$\rho^2 = \sum_{i=1}^{n} \lambda_i^2 \, |\lambda_i\rangle\langle\lambda_i| \, , \qquad \mathrm{Tr}\big(\rho^2\big) = \sum_{i=1}^{n} \lambda_i^2. \tag{3.6}$$

Because $p_i^2 \leq p_i$, equality will be achieved if and only if one of the λ_i equals 1 and the rest are 0. So a state ρ is pure if and only if $\mathrm{Tr}(\rho^2) = 1$.

There are many representations of a density matrix. For example, a qubit can be expressed by

$$\rho = \frac{1}{2}\left(\mathbb{I} + x\sigma_x + y\sigma_y + z\sigma_z\right) = \frac{1}{2}\left(\mathbb{I} + \boldsymbol{r}^{\mathrm{T}}\boldsymbol{\sigma}\right), \tag{3.7}$$

where $\boldsymbol{r} = \begin{pmatrix} x & y & z \end{pmatrix}^{\mathrm{T}}$ is a 3-dimensional real vector and $\boldsymbol{\sigma} = \begin{pmatrix} \sigma_x & \sigma_y & \sigma_z \end{pmatrix}^{\mathrm{T}}$ are the *Pauli Matrices*[3],

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma_y = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}, \qquad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{3.8}$$

If we calculate the eigenvalues of ρ we can known the domain of $\boldsymbol{r}$ such that ρ is positive semidefinite.

$$\det[\rho - \lambda\mathbb{I}] = 0 \tag{3.9}$$

$$\det\left[\frac{1}{2}\begin{pmatrix} 1+z-2\lambda & x-\mathrm{i}y \\ x+\mathrm{i}y & 1-z-2\lambda \end{pmatrix}\right] = 0 \tag{3.10}$$

$$\left(\frac{1}{2}\right)^2\left[(1+z-2\lambda)(1-z-2\lambda) - (x+\mathrm{i}y)(x+\mathrm{i}y)\right] = 0 \tag{3.11}$$

$$(1-2\lambda)^2 - z^2 - x^2 - y^2 = 0 \tag{3.12}$$

$$(1-2\lambda)^2 - |\boldsymbol{r}|^2 = 0. \tag{3.13}$$

Using that $|\boldsymbol{r}| = \sqrt{x^2+y^2+z^2}$, we have $\lambda_\pm = \frac{1}{2}(1 \pm |\boldsymbol{r}|)$. Now, because $\lambda_\pm \geq 0$, we have $|\boldsymbol{r}| \leq 1$. That is a 3-dimensional unitary sphere, where each point represents a different density matrix. This sphere is called the Bloch Sphere, and gives us a graphical representation of the qubits. From the purity we can identify two cases depending on the value of $|\boldsymbol{r}|$,

$$\mathrm{Tr}\left(\rho^2\right) = \lambda_+^2 + \lambda_-^2 = \frac{1}{2}\left(1 + |\boldsymbol{r}|^2\right). \tag{3.14}$$

Then, pure states are on the surface the sphere $|\boldsymbol{r}| = 1$ and mixed states are inside the sphere $|\boldsymbol{r}| < 1$.

[3]Named after Wolfang Pauli, Austrian physicist (1900-1958).

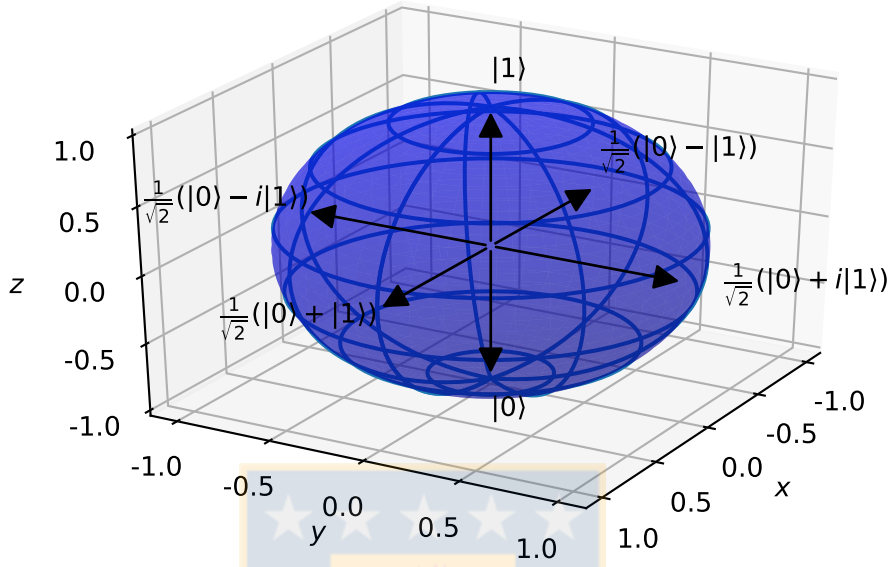

Figure 3.1: Graph of Bloch Sphere. Source: Made by the author

3.1.2 Evolution

The second postulate describes how the state changes with time.

Postulate 2. *The evolution of a closed quantum system is described by a unitary transformation. For an initial time t_0, at a any later time t the relationship between the initial and final states is given by:*

$$|\psi(t)\rangle = U(t, t_0)\,|\psi(t_0)\rangle\,. \tag{3.15}$$

A *closed quantum system* is a system that does not interact with any other system. We can calculate the evolution of a density operator,

$$\rho(t) = \sum_{i=1}^{n} p_i\,|\psi_i(t)\rangle\langle\psi_i(t)| \tag{3.16}$$

$$= \sum_{i=1}^{n} p_i U(t, t_0)\,|\psi_i(t_0)\rangle\langle\psi_i(t_0)|\,U(t, t_0)^\dagger \tag{3.17}$$

$$= U(t, t_0)\left(\sum_{i=1}^{n} p_i\,|\psi_i(t_0)\rangle\langle\psi_i(t_0)|\right)U(t, t_0)^\dagger \tag{3.18}$$

$$= U(t, t_0)\rho(t_0)U(t, t_0)^\dagger. \tag{3.19}$$

Furthermore, the general evolution of a density matrix is a completely positive map that preserves the trace, which can be represented by a set of $n \leq d^2$ operators $\{K_i\}_{i=1,\ldots,n}$ which satisfies $\sum_{i=1}^{n} K_i^\dagger K_i = \mathbb{I}$, called *Kraus Operators* [30],

$$\rho(t) = \sum_{i=1}^{n} K_i \rho(t_0) K_i^\dagger. \tag{3.20}$$

3.1.3 Measurements

The third postulate describes how to obtain information from a quantum system. The processes that allow us this are called *measurements*. This postulate is the main difference between classical physics and quantum physics. Classical physics implicitly assumes the properties of systems to be unchanged by measurement processes. Otherwise, in Quantum Physics the measurements have random outcomes and they modify the state of the system.

Postulate 3. *Quantum measurements are described by a set* $\{M_m\}_{m=1,\ldots,n}$ *called measurement operators. The index* m *refers to the measurement's outcomes that may occur in the experiment. The measurement operators satisfy the completeness relation*

$$\sum_{m=1}^{n} M_m^\dagger M_m = \mathbb{I}. \tag{3.21}$$

If the state of the quantum system is $|\psi\rangle$ *immediately before the measurement, then the probability that result* m *occurs is*

$$p(m) = \langle\psi|M_m^\dagger M_m|\psi\rangle, \tag{3.22}$$

and the state of the system immediately after the measurement is

$$|\psi^m\rangle = \frac{M_m|\psi\rangle}{\sqrt{\langle\psi|M_m^\dagger M_m|\psi\rangle}}. \tag{3.23}$$

For a density matrix and a measurement $\{M_m\}_{m=1,\ldots,n}$ on the ensemble $\{p_i, |\psi_i\rangle\}_{i=1,\ldots,n'}$ we have that the probability of obtaining outcome m is

$$p(m) = \sum_{i=1}^{n'} p(m|i)p_i \tag{3.24}$$

$$= \sum_{i=1}^{n'} \langle\psi_i| M_m^\dagger M_m |\psi_i\rangle p_i \tag{3.25}$$

$$= \sum_{i=1}^{n'} \mathrm{Tr}\big(M_m^\dagger M_m |\psi_i\rangle\langle\psi_i|\big)p_i \tag{3.26}$$

$$= \mathrm{Tr}\left(M_m^\dagger M_m \sum_{i=1}^{n'} p_i |\psi_i\rangle\langle\psi_i|\right) \tag{3.27}$$

$$= \mathrm{Tr}\big(M_m^\dagger M_m \rho\big), \tag{3.28}$$

where $p(m|i)$ is the conditional probability of m if the initial state was $|\psi_i\rangle$. The states of the ensemble after the measurement are

$$|\psi_i^m\rangle = \frac{M_m |\psi_i\rangle}{\sqrt{\langle\psi_i| M_m^\dagger M_m |\psi_i\rangle}}. \tag{3.29}$$

with respective probabilities $p(i|m)$. Then, the density matrix after the measurement is

$$\rho^m = \sum_{i=1}^{n'} p(i|m) |\psi_i^m\rangle\langle\psi_i^m| \tag{3.30}$$

$$= \sum_{i=1}^{n'} p(i|m) \left(\frac{M_m |\psi_i\rangle}{\sqrt{\langle\psi_i| M_m^\dagger M_m |\psi_i\rangle}}\right) \left(\frac{\langle\psi_i| M_m^\dagger}{\sqrt{\langle\psi_i| M_m^\dagger M_m |\psi_i\rangle}}\right) \tag{3.31}$$

$$= \sum_{i=1}^{n'} \frac{p(m|i)p_i}{p(m)} \frac{M_m |\psi_i\rangle\langle\psi_i| M_m^\dagger}{\langle\psi_i| M_m^\dagger M_m |\psi_i\rangle} \tag{3.32}$$

$$= \sum_{i=1}^{n'} \frac{\langle\psi_i| M_m^\dagger M_m |\psi_i\rangle p_i}{\mathrm{Tr}\Big(M_m^\dagger M_m \rho\Big)} \frac{M_m |\psi_i\rangle\langle\psi_i| M_m^\dagger}{\langle\psi_i| M_m^\dagger M_m |\psi_i\rangle} \tag{3.33}$$

$$= \frac{M_m \sum_{i=1}^{n'} p_i |\psi_i\rangle\langle\psi_i| M_m^\dagger}{\mathrm{Tr}\Big(M_m^\dagger M_m \rho\Big)} \tag{3.34}$$

$$= \frac{M_m \rho M_m^\dagger}{\mathrm{Tr}\Big(M_m^\dagger M_m \rho\Big)}. \tag{3.35}$$

where we used Bayes' theorem (3).

The simplest type of measurement is the *projective measurement*. A set of d projective measurements is associated to a hermitian operator called an *observable*. An observable O has a spectral decomposition

$$O = \sum_{m=1}^{d} \lambda_m \left|\phi_m\right\rangle\!\left\langle\phi_m\right|, \tag{3.36}$$

where λ_m and $|\phi_m\rangle$ are its respective eigenvalues and eigenvectors. Then, because the projectors $\pi_m = |\phi_m\rangle\langle\phi_m|$ are hermitian and idempotent, the probability to obtain outcome λ_m (or m for short) upon measuring the state $|\psi\rangle$ is, according to (3.28),

$$p(m) = \mathrm{Tr}\left(\pi_m^\dagger \pi_m \rho\right) = \mathrm{Tr}(\pi_m \rho) = \left\langle\phi_m\right|\rho\left|\phi_m\right\rangle, \tag{3.37}$$

and the state after the measurement is, according to Eq. (3.35), given by

$$\rho^m = \frac{\pi_m \rho \pi_m^\dagger}{\mathrm{Tr}\left(\pi_m^\dagger \pi_m \rho\right)} = \frac{\left\langle\phi_m\right|\rho\left|\phi_m\right\rangle \left|\phi_m\right\rangle\!\left\langle\phi_m\right|}{\left\langle\phi_m\right|\rho\left|\phi_m\right\rangle} = \left|\phi_m\right\rangle\!\left\langle\phi_m\right|. \tag{3.38}$$

The expected value of O in the state ρ is

$$\mathbb{E}(O|\rho) = \sum_{m=1}^{d} m p(m) = \sum_{m=1}^{d} m \,\mathrm{Tr}(\pi_m \rho) = \mathrm{Tr}\left[\left(\sum_{m=1}^{d} m \pi_m\right)\rho\right] = \mathrm{Tr}(O\rho). \tag{3.39}$$

The expected values of two observables O_1 and O_2 can be obtained simultaneously if they share their base of eigenstates. This occur if the two observables commute $[O_1, O_2] = \Theta$. Thereby, the expected values of two observables that do not commute $[O_1, O_2] \neq \Theta$ can not be obtained simultaneously.

The most general measurement on a density matrix is the *Positive Operators-Values Measure* or *POVM*. The POVMs are used when the post-measurement state is unimportant and the probabilities are the main interest. A POVM is a set $\{E_m\}_{m=1,\ldots,n}$ such that the operators E_m are all positive and satisfy the relation $\sum_{m=1}^{n} E_m = \mathbb{I}$. We can consider $M_m = \sqrt{E_m}$ such that $E_m = M_m^\dagger M_m$. Then we see that the POVM satisfies the equation (3.21) $\sum_{m=1}^{n} M_m^\dagger M_m = \sum_{m=1}^{n} E_m = \mathbb{I}$, and therefore the set $\{M_m\}_{m=1,\ldots,n}$ describes a measurement with POVM $\{E_m\}_{m=1,\ldots,n}$. The probability of outcome m is given by

$$p(m) = \mathrm{Tr}(E_m \rho). \tag{3.40}$$

3.1.4 Composite Systems

The last postulate allows to describe quantum systems that are formed by two or more subsystems.

Postulate 4. *The state space of a composite physical system is the tensor product of the state spaces of the component physical systems. If the subsystems are numbered from* 1 *to* n*, the state space is*

$$\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2 \otimes ... \otimes \mathcal{H}_n. \tag{3.41}$$

In particular, if the subsystems are prepared in the states $\{|\psi_i\rangle\}_{i=1,...,n}$, the state of the system is

$$|\Psi\rangle = |\psi_1\rangle \otimes |\psi_2\rangle \otimes ... \otimes |\psi_n\rangle \,. \tag{3.42}$$

It is important to remark that not all states in the composite space are of this form. If a state can be written in the form (3.42) we call it *separable*. Otherwise, we say it is *entangled*.

Let us consider the two qubit space, which is a 4-dimensional Hilbert space. A separable basis of this space is

$$|00\rangle = |0\rangle \otimes |0\rangle \quad |01\rangle = |0\rangle \otimes |1\rangle \quad |10\rangle = |1\rangle \otimes |0\rangle \quad |11\rangle = |1\rangle \otimes |1\rangle \,. \tag{3.43}$$

On the other hand, an entangled basis is the *Bell Basis*,

$$\left|\psi^{\pm}\right\rangle = \frac{1}{\sqrt{2}} \left(|01\rangle \pm |10\rangle\right) \qquad \left|\phi^{\pm}\right\rangle = \frac{1}{\sqrt{2}} \left(|00\rangle \pm |11\rangle\right). \tag{3.44}$$

The entangled states are used in Quantum Teleportation [44], Quantum Cryptography [45] and fundamental tests of Quantum Mechanics [46].

To define entanglement of density matrices is a bit more complicated. Suppose we have two systems $\mathcal{H}_1$ and $\mathcal{H}_2$, whose state is $\rho^{(12)}$. We say that it is separable if it is a convex combination[4] of products of density matrices,

$$\rho^{(12)} = \sum_{i=1}^{n} p_i \rho_i^{(1)} \otimes \rho_i^{(2)}. \tag{3.45}$$

[4] A convex combination of vectors $\{\boldsymbol{x}_i\}$ is $\sum_i \alpha_i \boldsymbol{x}_i$, where $\alpha_i \geq 0$ and $\sum_i \alpha_i = 1$.

Otherwise, the state is entangled. The entanglement measures of density matrices are current research topics [47, 48].

At last, we define the *reduced density matrix* as the state that allows us to describe a subsystem of a larger system. We eliminate the subspaces in which we are not interested tracing over them. Therefore, the reduced density matrix of $\mathcal{H}_1$ for $\rho^{(12)} \in \mathcal{H}_1 \otimes \mathcal{H}_2$ is

$$\rho^{(1)} = \mathrm{Tr}_2\left(\rho^{(12)}\right), \tag{3.46}$$

where $\mathrm{Tr}_2(\rho)$ is the *partial trace* of ρ over $\mathcal{H}_2$,

$$\mathrm{Tr}_2(\rho^{(12)}) = \mathrm{Tr}_2\left(\sum_{i,j=1}^{d} p_{ij}\rho_i^{(1)} \otimes \rho_j^{(2)}\right) \tag{3.47}$$

$$= \sum_{i,j=1}^{d} p_{ij}\rho_i^{(1)}\,\mathrm{Tr}\left(\rho_j^{(2)}\right) \tag{3.48}$$

$$= \sum_{i,j=1}^{d} p_{ij}\rho_i^{(1)}. \tag{3.49}$$

3.2 Distance between Quantum States

In many protocols in quantum information it is important to quantify how close two quantum states are. Some examples of this are state discrimination, state tomography and processes tomography.

A distance measure is a function from density operators to the real numbers. It must satisfy the metric axioms

$$\text{Positivity} \quad \begin{cases} 0 \leq d(\rho,\sigma), & \forall \rho,\sigma. \\ d(\rho,\sigma) = 0 \iff \rho = \sigma. \end{cases} \tag{3.50}$$

$$\text{Symmetry} \quad d(\rho,\sigma) = d(\sigma,\rho). \tag{3.51}$$

$$\text{Triangle inequality} \quad d(\rho,\sigma) \leq d(\rho,\tau) + d(\tau,\sigma). \tag{3.52}$$

We discuss two distance measures, the *trace distance* and the *fidelity*, which are especially useful in quantum information.

3.2.1 Trace Distance

The most natural distance measure is the *trace distance*. The trace distance between ρ and σ is defined by

$$D(\rho,\sigma)=\frac{1}{2}\operatorname{Tr}|\rho-\sigma|, \tag{3.53}$$

with $|A|=\sqrt{A^{\dagger}A}$. It consists of the sum of the absolute values of the eigenvalues of $\rho-\sigma$. Let A_+ and A_- two positive operators such as $A=A_+-A_-$. We call them positive and negative parts of A, respectively. Therefore $|A|=A_++A_-$. Because $\operatorname{Tr}(\rho-\sigma)=0$, we obtain $\operatorname{Tr}(\rho-\sigma)_+=\operatorname{Tr}(\rho-\sigma)_-$. Therefore

$$D(\rho,\sigma)=\frac{1}{2}\operatorname{Tr}\Big[(\rho-\sigma)_++(\rho-\sigma)_-\Big]=\operatorname{Tr}\Big[(\rho-\sigma)_+\Big]. \tag{3.54}$$

Clearly, the trace distance satisfies positivity and symmetry. Moreover, it fulfills the triangle inequality. Let P be a projective operator to $(\rho-\sigma)_+$,

$$\begin{aligned}
D(\rho,\sigma)=&\operatorname{Tr}[P(\rho-\sigma)] && (3.55)\\
=&\operatorname{Tr}[P(\rho-\tau+\tau-\sigma)] && (3.56)\\
=&\operatorname{Tr}[P(\rho-\tau)]+\operatorname{Tr}[P(\tau-\sigma)] && (3.57)\\
\leq&\operatorname{Tr}[(\rho-\tau)_+]+\operatorname{Tr}[(\tau-\sigma)_-] && (3.58)\\
=&D(\rho,\tau)+D(\tau,\sigma). && (3.59)
\end{aligned}$$

Thereby, the trace distance is a metric. This is important in state discrimination [49], because the probability of distinguishing two states with equal a priori probability is

$$P_d=\frac{1}{2}\Big[1+D(\rho,\sigma)\Big]. \tag{3.60}$$

3.2.2 Quantum Fidelity

The second distance between quantum states is the *fidelity*. This is defined as

$$F(\rho,\sigma)=\operatorname{Tr}\left(\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}}\right)^2. \tag{3.61}$$

If $\sigma = |\psi\rangle\langle\psi|$ is pure, it reduces to the transition probability from ρ to $|\psi\rangle$,

$$F(\rho, \sigma) = \langle\psi|\rho|\psi\rangle\,. \tag{3.62}$$

Inspired by this, we define the *infidelity* as

$$\mathrm{Inf}(\rho, \sigma) = 1 - F(\rho, \sigma). \tag{3.63}$$

The fidelity (or infidelity) is not actually a metric on the space of density matrices. However, we can construct a metric from it using the following theorem,

Theorem 5. *(Uhlmann's) Let us suppose ρ and σ are states of a quantum system $\mathcal{H}_1$. Introduce a second quantum system $\mathcal{H}_2$, equal to $\mathcal{H}_1$. Then*

$$F(\rho, \sigma) = \max_{|\psi\rangle, |\varphi\rangle \in \mathcal{H}_1 \otimes \mathcal{H}_2} |\langle\psi|\phi\rangle|^2, \tag{3.64}$$

where the maximization is over all purifications $|\psi\rangle$ of ρ and $|\varphi\rangle$ of σ, that is

$$\mathrm{Tr}_2\left(|\psi\rangle\langle\psi|\right) = \rho, \qquad \mathrm{Tr}_2\left(|\varphi\rangle\langle\varphi|\right) = \sigma. \tag{3.65}$$

The states $|\psi\rangle$ and $|\varphi\rangle$ belong to $\mathbb{C}^d$, so we can calculate their distance using the distance of $\mathbb{C}^d$. Since quantum states are invariant up to global phases, we must include a minimization on all possible global phases,

$$\begin{aligned} d_{\mathbb{C}^d} &= \min_{|\epsilon|=1} || \, |\psi\rangle - \epsilon \, |\varphi\rangle \, || && (3.66)\\ &= \min_{|\epsilon|=1} \sqrt{\Big(\langle\psi| - \epsilon^* \langle\varphi| \Big) \Big(|\psi\rangle - \epsilon \, |\varphi\rangle \Big)} && (3.67)\\ &= \min_{|\epsilon|=1} \sqrt{2 - 2\,\mathrm{Re}[\epsilon \, \langle\psi|\varphi\rangle]} && (3.68)\\ &= \sqrt{2 - 2|\,\langle\psi|\varphi\rangle\,|}. && (3.69) \end{aligned}$$

This is called the *Fubini-Study distance* between pure states. Then, using (3.64) and the last equation, we can define the *Bures distance*,

$$B(\rho, \sigma) = \sqrt{2 - 2\sqrt{F(\rho, \sigma)}}. \tag{3.70}$$

It inherits the metric properties of $d_{\mathbb{C}^d}$, so it is a metric too.

For a 2-dimensional system, the fidelity can be calculated as,

$$F(\rho,\sigma) = \mathrm{Tr}(\rho\sigma) + \sqrt{1-\mathrm{Tr}(\rho^2)}\sqrt{1-\mathrm{Tr}(\sigma^2)}. \tag{3.71}$$

Let us suppose ρ and σ are two infinitesimally close states, that is

$$F(\rho,\sigma) = 1 - \mathrm{d}F + \mathcal{O}(\mathrm{d}F^2). \tag{3.72}$$

We have that the infidelity and the Bures distance coincide up to first order,

$$\begin{aligned}
\mathrm{Inf}(\rho,\sigma) =& 1 - F(\rho,\sigma) && (3.73)\\
=& 1 - 1 + \mathrm{d}F + \mathcal{O}(\mathrm{d}F^2) && (3.74)\\
=& \mathrm{d}F + \mathcal{O}(\mathrm{d}F^2), && (3.75)
\end{aligned}$$

$$\begin{aligned}
B(\rho,\sigma)^2 =& 2\Big(1 - \sqrt{F(\rho,\sigma)}\Big) && (3.76)\\
=& 2\Big(1 - 1 + \frac{\mathrm{d}F}{2} + \mathcal{O}(\mathrm{d}F^2)\Big) && (3.77)\\
=& \mathrm{d}F + \mathcal{O}(\mathrm{d}F^2). && (3.78)
\end{aligned}$$

For this reason, we will use infidelity and Bures distance interchangeably. Besides, fidelity and trace distance are related through the following inequality [30],

$$1 - \sqrt{F(\rho,\sigma)} \leq D(\rho,\sigma) \leq \sqrt{1 - F(\rho,\sigma)}. \tag{3.79}$$

Therefore, infidelity and trace distance are equivalent for infinitesimally close states,

$$\begin{aligned}
\frac{1}{2}\mathrm{Inf}(\rho,\sigma) \leq& D(\rho,\sigma) \leq \sqrt{\mathrm{Inf}(\rho,\sigma)}. && (3.80)
\end{aligned}$$

3.3 Quantum Fisher Information

In this section we generalize the Fisher information matrix and the Cramer Rao bound to the quantum case.

Let us consider a d-dimensional quantum system in a state $\rho(\boldsymbol{s})$ characterized by the parameters $\boldsymbol{s} = \begin{pmatrix} s_1 & \cdots & s_n \end{pmatrix}^{\mathrm{T}}$, which we want to obtain. This is known as the

quantum estimation problem. Given a measurement set $\{E_i\}_{i=1,\dots,m}$, the probability of obtaining outcome i is

$$p(i|\boldsymbol{s}) = \mathrm{Tr}\left(E_i\rho(\boldsymbol{s})\right). \tag{3.81}$$

The classical Fisher information (2.145) is

$$I_{ij} = \sum_{k=1}^{m} p(k|\boldsymbol{s})\left(\frac{\partial}{\partial s_i}\ln p(k|\boldsymbol{s})\frac{\partial}{\partial s_j}\ln p(k|\boldsymbol{s})\right) \tag{3.82}$$

$$= \sum_{k=1}^{m}\frac{1}{p(k|\boldsymbol{s})}\frac{\partial p(k|\boldsymbol{s})}{\partial s_i}\frac{\partial p(k|\boldsymbol{s})}{\partial s_j} \tag{3.83}$$

$$= \sum_{k=1}^{m}\frac{1}{p(k|\boldsymbol{s})}\mathrm{Tr}\left(E_k\frac{\partial\rho(\boldsymbol{s})}{\partial s_i}\right)\mathrm{Tr}\left(E_k\frac{\partial\rho(\boldsymbol{s})}{\partial s_j}\right). \tag{3.84}$$

Let $\boldsymbol{w}\in\mathbb{R}^n$,

$$\boldsymbol{w}^{\mathrm{T}}I\boldsymbol{w} = \sum_{i,j=1}^{n} w_i I_{ij} w_j \tag{3.85}$$

$$= \sum_{k=1}^{m}\frac{1}{p(k|\boldsymbol{s})}\mathrm{Tr}\left(E_k\sum_{i=1}^{n}w_i\frac{\partial\rho(\boldsymbol{s})}{\partial s_i}\right)\mathrm{Tr}\left(E_k\sum_{j=1}^{n}\frac{\partial\rho(\boldsymbol{s})}{\partial s_j}w_j\right) \tag{3.86}$$

$$= \sum_{k=1}^{m}\frac{1}{p(k|\boldsymbol{s})}\mathrm{Tr}\left(E_k\sum_{i=1}^{n}w_i\frac{\partial\rho(\boldsymbol{s})}{\partial s_i}\right)^2. \tag{3.87}$$

We define operators $L_i(\boldsymbol{s})$ such that they satisfy the equation

$$\frac{\partial\rho(\boldsymbol{s})}{\partial s_i} = \frac{1}{2}\Big(\rho(\boldsymbol{s})L_i(\boldsymbol{s}) + L_i(\boldsymbol{s})\rho(\boldsymbol{s})\Big). \tag{3.88}$$

These operators are called symmetric logarithmic derivative of $\rho(\boldsymbol{s})$ with respect to s_i. Then

$$\sum_{i=1}^{n} w_i\frac{\partial\rho(\boldsymbol{s})}{\partial s_i} = \frac{1}{2}\sum_{i=1}^{n} w_i\Big(\rho(\boldsymbol{s})L_i(\boldsymbol{s}) + L_i(\boldsymbol{s})\rho(\boldsymbol{s})\Big) \tag{3.89}$$

$$= \frac{1}{2}\left(\rho(\boldsymbol{s})\sum_{i=1}^{n}w_iL_i(\boldsymbol{s}) + \sum_{i=1}^{n}w_iL_i(\boldsymbol{s})\rho(\boldsymbol{s})\right). \tag{3.90}$$

$$\tag{3.91}$$

Writing $L = \sum_{i=1}^{n} w_i L_i$ and replacing in (3.87),

$$\boldsymbol{w}^{\mathrm{T}} I \boldsymbol{w} = \frac{1}{2} \sum_{k=1}^{m} \frac{1}{p(k|\boldsymbol{s})} \operatorname{Tr}\Big(E_k \rho(\boldsymbol{s}) L(\boldsymbol{s}) + E_k L(\boldsymbol{s}) \rho(\boldsymbol{s})\Big)^2 \tag{3.92}$$

$$= \sum_{k=1}^{m} \frac{1}{p(k|\boldsymbol{s})} \operatorname{Re}\Big[\operatorname{Tr}\Big(E_k L(\boldsymbol{s}) \rho(\boldsymbol{s})\Big)\Big]^2 \tag{3.93}$$

$$\leq \sum_{k=1}^{m} \frac{1}{p(k|\boldsymbol{s})} \Big|\operatorname{Tr}\Big(E_k L(\boldsymbol{s}) \rho(\boldsymbol{s})\Big)\Big|^2 \tag{3.94}$$

$$= \sum_{k=1}^{m} \frac{1}{p(k|\boldsymbol{s})} \left|\operatorname{Tr}\left(\Big[\sqrt{E_k}\sqrt{\rho(\boldsymbol{s})}\Big]^{\dagger}\Big[\sqrt{E_k} L(\boldsymbol{s}) \sqrt{\rho(\boldsymbol{s})}\Big]\right)\right|^2 \tag{3.95}$$

$$= \sum_{k=1}^{m} \frac{1}{p(k|\boldsymbol{s})} \left|\left\langle \sqrt{E_k}\sqrt{\rho(\boldsymbol{s})}, \sqrt{E_k} L(\boldsymbol{s}) \sqrt{\rho(\boldsymbol{s})}\right\rangle\right|^2, \tag{3.96}$$

where $\operatorname{Re}(z)$ is the real part of z and $\langle \cdot, \cdot \rangle$ is the Hilbert-Schmidt product. Using Schwartz's inequality (2.17),

$$\boldsymbol{w}^{\mathrm{T}} I \boldsymbol{w} \leq \sum_{k=1}^{m} \frac{1}{p(k|\boldsymbol{s})} \Big|\Big|\sqrt{E_k}\sqrt{\rho(\boldsymbol{s})}\Big|\Big|^2 \quad \Big|\Big|\sqrt{E_k} L(\boldsymbol{s}) \sqrt{\rho(\boldsymbol{s})}\Big|\Big|^2 \tag{3.97}$$

$$= \sum_{k=1}^{m} \frac{1}{p(k|\boldsymbol{s})} \operatorname{Tr}\Big(E_k \rho(\boldsymbol{s})\Big) \operatorname{Tr}\Big(E_k L(\boldsymbol{s}) \rho(\boldsymbol{s}) L(\boldsymbol{s})\Big) \tag{3.98}$$

$$= \sum_{k=1}^{m} \operatorname{Tr}\Big(E_k L(\boldsymbol{s}) \rho(\boldsymbol{s}) L(\boldsymbol{s})\Big) \tag{3.99}$$

$$= \operatorname{Tr}\Big(\rho(\boldsymbol{s}) L(\boldsymbol{s})^2\Big) \tag{3.100}$$

$$= \operatorname{Tr}\left(\rho(\boldsymbol{s}) \sum_{i=1}^{n} w_i L_i(\boldsymbol{s}) \sum_{j=1}^{n} w_j L_j(\boldsymbol{s})\right) \tag{3.101}$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} w_i \frac{1}{2} \operatorname{Tr}\left(\rho(\boldsymbol{s})\Big[L_i(\boldsymbol{s}) L_j(\boldsymbol{s}) + L_j(\boldsymbol{s}) L_i(\boldsymbol{s})\Big]\right) w_j \tag{3.102}$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} w_i \big[J(\boldsymbol{s})\big]_{ij} w_j \tag{3.103}$$

$$= \boldsymbol{w}^{\mathrm{T}} J(\boldsymbol{s}) \boldsymbol{w}. \tag{3.104}$$

Thus, we obtain an upper bound of I,

$$I \leq J, \tag{3.105}$$

where J is the *Quantum Fisher Information matrix*,

$$J_{ij}(\boldsymbol{s}) = \frac{1}{2}\,\mathrm{Tr}\left(\rho(\boldsymbol{s})\Big[L_i(\boldsymbol{s})L_j(\boldsymbol{s}) + L_j(\boldsymbol{s})L_i(\boldsymbol{s})\Big]\right). \tag{3.106}$$

This matrix establishes the *Quantum Cramer-Rao Bound* for the covariance matrix [26, 27],

$$C \geq I^{-1} \geq J^{-1}. \tag{3.107}$$

Let us note that the Quantum Fisher Information matrix does not depend explicitly on the set of measurement operators $\{E_i\}_{i=1,\ldots,m}$ but on the state $\rho(\boldsymbol{s})$ and on the parameters to be estimated. Besides, this inequality is only valid for an unbiased estimate.

From Eq. (3.107) we can see that the reciprocal of the Quantum Fisher Information matrix provides a smaller lower bound for the Covariance matrix than the reciprocal of the Classical Fisher Information matrix. Thereby, the laws of Quantum Mechanics allow for an improvement in the accuracy of estimations. In this regard, the use of quantum systems for estimation purposes might surpass the accuracy achieved by solely classical means. This important result led to the new field of Quantum Metrology [23], which aims at improving the accuracy in the measurements by exploiting quantum systems.

To estimate the set $\boldsymbol{s}$ of parameters we have at our disposal N quantum systems equally prepared in the state $\rho(\boldsymbol{s})$, that is,

$$\rho^{\otimes N}(\boldsymbol{s}) = \rho(\boldsymbol{s}) \otimes \rho(\boldsymbol{s}) \otimes \cdots \otimes \rho(\boldsymbol{s}). \tag{3.108}$$

This sample of size N is our primary resource. The symmetric logarithmic derivatives $L_i^{(N)}(\boldsymbol{s})$ of $\rho^{\otimes N}(\boldsymbol{s})$ are defined by

$$\frac{\partial \rho^{\otimes N}(\boldsymbol{s})}{\partial s_i} = \frac{1}{2}\Big(\rho(\boldsymbol{s})L_i^{(N)}(\boldsymbol{s}) + L_i^{(N)}(\boldsymbol{s})\rho(\boldsymbol{s})\Big). \tag{3.109}$$

Using the chain rule,

$$\frac{\partial \rho}{\partial s_i} \otimes \rho \otimes \cdots \otimes \rho + \cdots + \rho \otimes \cdots \otimes \frac{\partial \rho}{\partial s_i} \otimes \rho = \frac{1}{2}\Big(\rho L_i^{(N)} + L_i^{(N)}\rho\Big). \tag{3.110}$$

This corresponds to the equation of L_i per subsystem,

$$L_i^{(N)} = L_i \otimes \mathbb{I} \otimes \cdots \otimes \mathbb{I} + \cdots + \mathbb{I} \otimes \cdots \otimes L_i \otimes \mathbb{I}. \tag{3.111}$$

Then, the quantum Fisher information matrix becomes

$$\begin{aligned} J_{ij}^{(N)} =& \frac{1}{2} \operatorname{Tr}\left(\rho^{\otimes N}\left[L_i^{(N)} L_j^{(N)} + L_j^{(N)} L_i^{(N)}\right]\right) && (3.112)\\ =& \frac{N}{2} \operatorname{Tr}\left(\rho\left[L_i L_j + L_j L_i\right]\right) && (3.113)\\ =& N J_{ij}. && (3.114) \end{aligned}$$

From this ensemble of N copies we can obtain information about the set $\boldsymbol{s}$ of parameters by performing two different types of processes:

- Measuring each copy separately, or
- Measuring globally the collective system formed by all copies.

Let $C^{(N)}(\boldsymbol{s})$ the covariance matrix and $I^{(N)}(\boldsymbol{s})$ the classical Fisher information matrix, for an estimate with sample size N. The Quantum Cramer-Rao Bound becomes

$$C^{(N)} \geq \left[I^{(N)}\right]^{-1} \geq \frac{1}{N} J^{-1}, \tag{3.115}$$

where J corresponds to the quantum Fisher information of a single copy.

Here arises the question whether the Quantum Cramer Rao bound can be saturated or not. One would like to know what quantum strategy allows us to attain the Quantum Cramer-Rao bound. Such a strategy represents the optimal estimation technique providing the highest possible accuracy allowed by the laws of Quantum Mechanics. It has been shown that in order to approach the Cramer-Rao bound, collective measurements must be used [28, 50]. Nevertheless, these measures are difficult to implement experimentally. This leads to another important constraint [29].

Theorem 6. *(Gill-Massar) When we measure a state ρ with separable measurements, the Fisher information matrices I and J must satisfy the Gill-Massar inequality*

$$\operatorname{Tr}\left(I^{(N)} J^{-1}\right) \leq N(d-1). \tag{3.116}$$

This theorem implies that it is generally impossible to construct a measurement that is optimal for all parameters by separable measurements only. But, we can try to find the best possible measurement by the following optimization problem

$$\min_{\boldsymbol{s}} \operatorname{Tr}\left(WC^{(N)}(\boldsymbol{s})\right), \\ \operatorname{Tr}\left(I^{(N)}(\boldsymbol{s})J(\boldsymbol{s})^{-1}\right) \leq N(d-1). \tag{3.117}$$

where W is a positive defined matrix. The function $\operatorname{Tr}\left(WC^{(N)}\right)$ is called weight mean square error. The solution to this problem is [51]

$$\operatorname{Tr}\left(WC^{(N)}\right) = \frac{1}{N(d-1)}\left(\operatorname{Tr}\left[\sqrt{\sqrt{W}J^{-1}\sqrt{W}}\right]\right)^2, \tag{3.118}$$

$$I^{(N)} = N(d-1)\sqrt{J}\frac{\sqrt{J^{-1/2}WJ^{-1/2}}}{\operatorname{Tr}\left(\sqrt{J^{-1/2}WJ^{-1/2}}\right)}\sqrt{J}. \tag{3.119}$$

It reaches the classical Cramer-Rao bound $C = I^{-1}$ and the Gill-Massar Inequality $\operatorname{Tr}\left(I^{(N)}J^{-1}\right) = N(d-1)$, but generally not the quantum Cramer-Rao bound. Thereby, there could be a difference between the optimal separate measurements and the quantum Cramer-Rao bound.

Chapter 4

Quantum State Tomography

In the previous section. we saw that a quantum system is described by a state. Many quantum systems in Nature are in states that are unknown to us, so it becomes necessary to formulate processes to estimate these states. Due to the probabilistic character of the outcomes of a quantum measurement, the experimental determination of all properties that specify a quantum state becomes a difficult enterprise. In order to achieve this goal several schemes have been proposed, the so called *quantum tomographic methods* [25].

The tomographic methods are based on measuring a set of at least $d^2 - 1$ operators in an ensemble of many copies of the unknown state, to then determine the $d^2 - 1$ independent real parameters which characterize ρ by inverting a linear system. Finally, we perform a post-processing by maximum likelihood to get a positive estimation. In this chapter, we briefly enunciate and review some techniques for quantum state tomography.

4.1 Preliminaries Aspects

In the first place, it is important to emphasize some preliminaries aspects about the resources of the Quantum Tomography:

- Since a density matrix has $d^2 - 1$ independent coefficients, at least we can perform an experiment with $d^2 - 1$ outcomes. Due to the number of parameters to be estimated grows quadratically with the dimension, the quantum tomographic problem inevitably becomes more complicated in high dimensions. This is called the *curse of dimensionality* [14, 25].

- The reconstruction of an unknown state be achieved through a single POVM or through the concatenation of several projective measurements [9, 52, 53].

- Since the outcomes of quantum measurements are random, we must repeat the measurements for estimating the parameters of the state. This means that the estimated density matrix will always have a statistical error, which causes problems on its positivity [27, 54].

- Since the initial state cannot be cloned [55] and it changes after the measurement, we must have access to more than one state preparation of the initial state to be able to estimate it [30].

Thereby, the ensemble of copies available to perform the tomography must be distributed in each of the different set of measurements. As example, let us suppose we have access to N state preparations to perform its tomography by m sets of measurements. The amount of sample per set of measurement is

$$N_m = \frac{N}{m}. \tag{4.1}$$

4.2 Linear Inversion Tomography

Let us consider a d-dimensional Hilbert space. Any density matrix ρ of a qudit can be uniquely represented by d^2 independent coefficients, so we need a set of d^2

linearly independent hermitian matrices to expand the density matrix, that is

$$\rho = \sum_{i=0}^{d^2-1} s_i \Lambda_i \tag{4.2}$$

where $\{s_i\}_{i=0,\ldots,d^2-1}$ are real parameters. These parameters are such that the matrix ρ is positive. Note that because $\mathrm{Tr}(\rho) = 1$, actually only $d^2 - 1$ measurements will be necessary to reconstruct the state. In general, we can perform m measurements $\{M_i\}_{i=1,\ldots,m}$ (these can be projectors, observables or POVM elements) on the density matrix ρ, obtaining results with probabilities $\{p_i\}_{i=1,\ldots,m}$,

$$p_j = \mathrm{Tr}(M_j \rho) = \mathrm{Tr}\left(M_j \sum_{i=0}^{d^2-1} s_i \Lambda_i \right) = \sum_{i=0}^{d^2-1} \mathrm{Tr}(M_j \Lambda_i) s_i. \tag{4.3}$$

Defining $\boldsymbol{p} = \begin{pmatrix} p_1 & \cdots & p_m \end{pmatrix}^{\mathrm{T}}$, $\boldsymbol{s} = \begin{pmatrix} s_1 & \cdots & s_{d^2} \end{pmatrix}^{\mathrm{T}}$ and $B_{ij} = \mathrm{Tr}(M_i \Lambda_j)$, the previous equation becomes a linear system

$$\boldsymbol{p} = B\boldsymbol{s}. \tag{4.4}$$

If $m = d^2 - 1$ we get the solution by inverting B, that is,

$$\boldsymbol{s} = B^{-1}\boldsymbol{p}. \tag{4.5}$$

If $m \geq d^2 - 1$ the equation system is overcomplete. In this case, we search for the least square solution,

$$\boldsymbol{s} = \arg\min_{\boldsymbol{s'}} ||B\boldsymbol{s'} - \boldsymbol{p}||^2. \tag{4.6}$$

This minimization can be easily solved

$$\frac{\partial}{\partial s_i} ||B\boldsymbol{s} - \boldsymbol{p}||^2 = \frac{\partial}{\partial s_i} \sum_{j=1}^{m} \left[\sum_{k=1}^{d^2} B_{jk} s_k - p_j \right]^2 = 2 \sum_{j=1}^{m} \left[\sum_{k=1}^{d^2} B_{jk} s_k - p_j \right] B_{ji}. \tag{4.7}$$

$$\sum_{j=1}^{m} B_{ji} \sum_{k=1}^{d^2} B_{jk} s_k = \sum_{j=1}^{m} B_{ji} p_j \tag{4.8}$$

$$B^{\mathrm{T}} B\boldsymbol{s} = B^{\mathrm{T}} \boldsymbol{p} \tag{4.9}$$

Finally, the least square solution is

$$\boldsymbol{s} = \left(B^{\mathrm{T}}B\right)^{-1}B^{\mathrm{T}}\boldsymbol{p}, \tag{4.10}$$

where $B^{+} = \left(B^{\mathrm{T}}B\right)^{-1}B^{\mathrm{T}}$ is called the *pseudoinverse* of B. Then, by estimating $\boldsymbol{p}$ experimentally, it is possible to reconstruct the density matrix ρ. In the next sections we present some common tomography schemes that use the linear inversion.

4.2.1 Standard Qudit Tomography

Standard Tomography of a qudit consists in using *generalized Gell-Mann*[1] *matrices* [9, 56] to reconstruct the state (4.2). This are given by

$$\sigma_{ij}^{x} = |i\rangle\langle j| + |j\rangle\langle i|\,, \tag{4.11}$$

$$\sigma_{ij}^{y} = -\,\mathrm{i}\Big(\,|i\rangle\langle j| - |j\rangle\langle i|\,\Big), \tag{4.12}$$

$$\sigma_{k}^{z} = \sqrt{\frac{2}{k(k+1)}}\left(\sum_{j=1}^{k}|j\rangle\langle j| - k\,|k+1\rangle\langle k+1|\right), \tag{4.13}$$

with $1 \leq i < j \leq d$ | and $1 \leq k \leq d-1$. In total there are $d^2 - 1$ matrices. All of them are hermitian and traceless. We can label these matrices employing a single letter by

$$\sigma_{(j-1)^2+2(i-1)} = \sigma_{ij}^{x}, \tag{4.14}$$

$$\sigma_{(j-1)^2+2i-1} = \sigma_{ij}^{y}, \tag{4.15}$$

$$\sigma_{j^2-1} = \sigma_{j-1}^{z}. \tag{4.16}$$

These matrices satisfy the following products (A),

$$\mathrm{Tr}(\sigma_i\sigma_j) = 2\delta_{ij}. \tag{4.17}$$

The identity operator together with the matrices $\{\sigma_i\}_{i=1,\ldots,d^2-1}$ form a basis to d-dimensional Hilbert space. Then, we can write any density matrix as

$$\rho = s_0\mathbb{I} + \sum_{i=1}^{d^2-1} s_i\sigma_i. \tag{4.18}$$

[1] Named after Murray Gell-Mann, American physicist (1929-).

From the normalization,

$$\mathrm{Tr}(\rho) = s_0\,\mathrm{Tr}(\mathbb{I}) + \sum_{i=1}^{d^2-1} s_i\,\mathrm{Tr}(\sigma_i) = s_0 d. \tag{4.19}$$

Then $s_0 = 1/d$. Performing d^2-1 measurements on the observables $\{\sigma_i\}_{i=1,\ldots,d^2-1}$ we can determinate the parameters $\{s_i\}_{i=1,\ldots,d^2-1}$,

$$\mathbb{E}(\sigma_i) = \mathrm{Tr}(\sigma_i\rho) = \frac{1}{d}\,\mathrm{Tr}\left(\sigma_i + \sum_{j=1}^{d^2-1} s_j\sigma_i\sigma_j\right) = \sum_{j=1}^{d^2-1} s_j\,\mathrm{Tr}(\sigma_i\sigma_j). \tag{4.20}$$

Experimentally there are $d(d^2-1)$ different outcomes of the eigenvectors of these matrices. The matrices (4.11) and (4.12) have non-null eigenvalues $\{1,-1\}$ with eigenvectors

$$\sigma_{ij}^x \Longrightarrow \begin{cases} +1, & \left|+_{ij}^x\right\rangle = \dfrac{1}{\sqrt{2}}\Big(\,|i\rangle + |j\rangle\,\Big), \\ -1, & \left|-_{ij}^x\right\rangle = \dfrac{1}{\sqrt{2}}\Big(\,|i\rangle - |j\rangle\,\Big). \end{cases} \tag{4.21}$$

$$\sigma_{ij}^y \Longrightarrow \begin{cases} +1, & \left|+_{ij}^y\right\rangle = \dfrac{1}{\sqrt{2}}\Big(\,|i\rangle + \mathrm{i}\,|j\rangle\,\Big), \\ -1, & \left|-_{ij}^y\right\rangle = \dfrac{1}{\sqrt{2}}\Big(\,|i\rangle - \mathrm{i}\,|j\rangle\,\Big). \end{cases} \tag{4.22}$$

respectively. The matrices (4.13) are diagonal and its eigenvectors are the canonical basis $\{|k\rangle\}_{k=1,\ldots,d}$ with eigenvalues

$$\sigma_k^z \Longrightarrow \begin{cases} \sqrt{\dfrac{2}{k(k+1)}}, & \text{for } |1\rangle, \ldots, |k\rangle\,. \\ -k\sqrt{\dfrac{2}{k(k+1)}}, & \text{for } |k+1\rangle\,. \\ 0, & \text{for } |k+2\rangle, \ldots, |d\rangle\,. \end{cases} \tag{4.23}$$

The parameters can be obtained inverting $B_{ij} = \mathrm{Tr}(\sigma_i\sigma_j)$ (4.5). From (4.17), we have $B_{ij} = 2\delta_{ij}$ and $B_{ij}^{-1} = (1/2)\delta_{ij}$. Then,

$$s_i = \sum_{j=1}^{d^2-1} B_{ij}^{-1} p_j = \frac{1}{2}\mathbb{E}(\sigma_i). \tag{4.24}$$

Thus, substituting in (4.18) and defining $S_i = \mathbb{E}(\sigma_i)$, we get

$$\rho = \frac{1}{d}\mathbb{I} + \frac{1}{2}\sum_{i=1}^{d^2-1} S_i\sigma_i \tag{4.25}$$

$$= \frac{1}{d}\mathbb{I} + \frac{1}{2}\sum_{1\leq i<j\leq d} S_{ij}^x\sigma_{ij}^x + \frac{1}{2}\sum_{1\leq i<j\leq d} S_{ij}^y\sigma_{ij}^y + \frac{1}{2}\sum_{k=1}^{d-1} S_k^z\sigma_k^z. \tag{4.26}$$

Therefore, we can do the tomography of an unknown state by obtaining the coefficients $\{S_i\}_{i=1,\ldots,d^2-1}$ by d^2-1 independent experiments.

4.2.2 Local Multiqudit Tomography

Let us consider a composite system of m qudits [9, 30, 56]. The Hilbert space $\mathcal{H}$ of this system has dimension $\dim(\mathcal{H}) = d^m$, because it is a tensor product space $\mathcal{H} = \mathcal{H}_1 \otimes ... \otimes \mathcal{H}_m$, where $\dim(\mathcal{H}_i) = d$ for $i = 1, ..., m$. The measurements that we choose to perform on this system are only local measurements $M = M_1 \otimes ... \otimes M_m$, because it is difficult to perform entangled measurements experimentally. Then, since some observables have entangled eigenvectors, we can not perform Standard Qudit Tomography to estimate the density matrix of this system. However, we can represent the density matrix in terms of the local observables $\{\sigma_0, \sigma_i\}_{i=1,\ldots,d^2-1}$, with $\sigma_0 = \mathbb{I}$,

$$\rho = \sum_{i_1=0}^{d^2-1} \cdots \sum_{i_m=0}^{d^2-1} s_{i_1\cdots i_m}\sigma_{i_1} \otimes \cdots \otimes \sigma_{i_m}. \tag{4.27}$$

Performing local measurements $\{\sigma_{j_1} \otimes \cdots \otimes \sigma_{j_m}\}_{j_1,\cdots,j_m=1,\ldots,d^2-1}$ on the system,

$$\mathbb{E}\big(\sigma_{j_1} \otimes \cdots \otimes \sigma_{j_m}\big) = \mathrm{Tr}\left(\rho\sigma_{j_1} \otimes \cdots \otimes \sigma_{j_m}\right) \tag{4.28}$$

$$= \sum_{i_1=0}^{d^2-1} \cdots \sum_{i_m=0}^{d^2-1} s_{i_1\cdots i_m}\,\mathrm{Tr}\left(\sigma_{i_1}\sigma_{j_1} \otimes \cdots \otimes \sigma_{i_m}\sigma_{j_m}\right). \tag{4.29}$$

Using $\mathrm{Tr}(\sigma_i\sigma_j) = c_i\delta_{ij}$, where $c_0 = d$ and $c_j = 2$, with $j = 1, 2, ..., d$,

$$\mathbb{E}\big(\sigma_{j_1} \otimes \cdots \otimes \sigma_{j_m}\big) = \sum_{i_1=1}^{d^2-1} \cdots \sum_{i_m=1}^{d^2-1} c_{i_1}\cdots c_{i_m}\delta_{i_1j_1}\cdots\delta_{i_mj_m}s_{i_1\cdots i_m} = c_{j_1}\cdots c_{j_m}s_{j_1\cdots j_m}. \tag{4.30}$$

It is a linear system (4.4), where its solution is

$$s_{j_1\cdots j_m} = \frac{1}{c_{j_1}\cdots c_{j_m}}\mathbb{E}\big(\sigma_{j_1}\otimes\cdots\otimes\sigma_{j_m}\big). \tag{4.31}$$

Because $\mathrm{Tr}(\rho)=1$ we have $S_{0_1\cdots 0_m}=1/d^m$. Thus, we can express a multiqubit state with local measurements as

$$\rho = \sum_{i_1=0}^{d^2-1}\cdots\sum_{i_m=0}^{d^2-1}\frac{1}{c_{i_1}\cdots c_{i_m}}S_{i_1\cdots i_m}\sigma_{i_1}\otimes\cdots\otimes\sigma_{i_m}. \tag{4.32}$$

where $S_{i_1\cdots i_m}=\mathbb{E}\big(\sigma_{j_1}\otimes\cdots\otimes\sigma_{j_m}\big)$. In particular for two qudits, the density matrix is

$$\rho = \frac{1}{d^2}\mathbb{I}\otimes\mathbb{I}+\frac{1}{2d}\sum_{i=1}^{d^2-1}\Big(S_{0i}\mathbb{I}\otimes\sigma_i+S_{i0}\sigma_i\otimes\mathbb{I}\Big)+\frac{1}{2^2}\sum_{i=1}^{d^2-1}\sum_{j=1}^{d^2-1}S_{ij}\sigma_i\otimes\sigma_j. \tag{4.33}$$

Therefore, measuring on the local observables $\{\sigma_{j_1}\otimes\cdots\otimes\sigma_{j_m}\}_{j_1,\cdots,j_m=1,\ldots,d^2-1}$ we can perform the quantum tomography of an unknown state locally.

4.2.3 Mutually Unbiased Bases Tomography

Let us consider a Hilbert space with dimension d and D orthonormal bases $\{|a\alpha\rangle\}$, where $a=1,\ldots,D$ labels each basis and $\alpha=1,\ldots,d$ labels each element of a given basis. We call those bases a set of *mutually unbiased bases (MUBs)* if and only if they satisfy the condition [53]

$$|\langle a\alpha|b\beta\rangle|^2=\frac{1}{d}(1-\delta_{ab})+\delta_{ab}\delta_{\alpha\beta}. \tag{4.34}$$

From this equation for different basis $a\neq b$, we obtain $|\langle a\alpha|b\beta\rangle|=1/d$. For the case where the dimension d is a integer power of a prime number, the system has $D=d+1$ MUBs [53,57,58]. Otherwise, it is not known how many MUBs the system allows.

We can use MUBs to perform the tomography of a state ρ on a prime power dimensional space. We denote the projector onto the state $|a\alpha\rangle$ as $\Pi_{a\alpha}=|a\alpha\rangle\langle a\alpha|$ and their respective probabilities $p_{a\alpha}=\mathrm{Tr}(\rho\Pi_{a\alpha})$. From orthogonality,

$$\mathrm{Tr}\left(\Pi_{a\alpha}\Pi_{b\beta}\right)=\frac{1}{d}(1-\delta_{ab})+\delta_{ab}\delta_{\alpha\beta}. \tag{4.35}$$

$$\sum_{a=1}^{d+1}\sum_{\alpha=1}^{d-1}\Pi_{a\alpha} = (d+1)\mathbb{I}. \tag{4.36}$$

Thus, the operators $E_{a\alpha} = (d+1)^{-1}\Pi_{a\alpha}$ form a POVM. Since the MUBs elements are linearly independent, the identity operator together with d^2-1 operators $\{\Pi_{a\alpha}\}$ with $a = 1, ..., d+1$ and $\alpha = 1, ..., d-1$ form a basis in the d-dimensional Hilbert space. Then, any density operator can be resolved through these bases,

$$\rho = c_{\mathbb{I}}\mathbb{I} + \sum_{a=1}^{d+1}\sum_{\alpha=1}^{d-1} c_{a\alpha}\Pi_{a\alpha}, \tag{4.37}$$

where $c_{\mathbb{I}}$ and $c_{a\alpha}$ are real parameters. Using the trace condition $\mathrm{Tr}(\rho) = 1$,

$$\mathrm{Tr}(\rho) = \mathrm{Tr}\left(c_{\mathbb{I}}\mathbb{I} + \sum_{a=1}^{d+1}\sum_{\alpha=1}^{d-1} c_{a\alpha}\Pi_{a\alpha}\right) = c_{\mathbb{I}}d + \sum_{a=1}^{d+1}\sum_{\alpha=1}^{d-1} c_{a\alpha}, \tag{4.38}$$

$$c_{\mathbb{I}} = \frac{1}{d}\left(1 - \sum_{a=1}^{d+1}\sum_{\alpha=1}^{d-1} c_{a\alpha}\right). \tag{4.39}$$

Thus, it is possible to write the density matrix as

$$\rho = \frac{1}{d}\mathbb{I} + \sum_{a=1}^{d+1}\sum_{\alpha=1}^{d-1} c_{a\alpha}\left(\Pi_{a\alpha} - \frac{1}{d}\mathbb{I}\right). \tag{4.40}$$

Performing measurements in the $\Pi_{a\alpha}$ operators,

$$p_{a\alpha} = \mathrm{Tr}(\Pi_{a\alpha}\rho) = \frac{1}{d} + \sum_{b=1}^{d+1}\sum_{\beta=1}^{d-1} c_{b\beta}\,\mathrm{Tr}\left[\Pi_{a\alpha}\left(\Pi_{b\beta} - \frac{1}{d}\mathbb{I}\right)\right]. \tag{4.41}$$

It is a linear system (4.4), then we can obtain the parameters $c_{b\beta}$ inverting

$$B_{a\alpha,b\beta} = \mathrm{Tr}\left[\Pi_{a\alpha}\left(\Pi_{b\beta} - \frac{1}{d}\mathbb{I}\right)\right] \tag{4.42}$$

$$= \mathrm{Tr}\left(\Pi_{a\alpha}\Pi_{b\beta}\right) - \frac{1}{d}\mathrm{Tr}(\Pi_{a\alpha}) \tag{4.43}$$

$$= \frac{1}{d}(1-\delta_{ab}) + \delta_{ab}\delta_{\alpha\beta} - \frac{1}{d} \tag{4.44}$$

$$= \delta_{ab}\left(\delta_{\alpha\beta} - \frac{1}{d}\right). \tag{4.45}$$

This is a block matrix with blocks $(B_{block})_{\alpha\beta} = (\delta_{\alpha\beta} - 1/d)$,

$$B = \begin{pmatrix} B_{block} & 0 & \dots & 0 \\ 0 & B_{block} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & B_{block} \end{pmatrix}_{d+1\times d+1}, \tag{4.46}$$

$$B_{block} = \begin{pmatrix} 1-\frac{1}{d} & -\frac{1}{d} & \dots & -\frac{1}{d} \\ -\frac{1}{d} & 1-\frac{1}{d} & \cdots & -\frac{1}{d} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{d} & -\frac{1}{d} & \cdots & 1-\frac{1}{d} \end{pmatrix}_{d-1\times d-1}. \tag{4.47}$$

It can be easily checked than its inverse matrix is $B^{-1}_{b\beta,a\alpha} = \delta_{ba}\left(\delta_{\beta\alpha} + 1\right)$. This is also a block matrix,

$$B^{-1} = \begin{pmatrix} B^{-1}_{block} & 0 & \dots & 0 \\ 0 & B^{-1}_{block} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & B^{-1}_{block} \end{pmatrix}_{d-1\times d-1}, \tag{4.48}$$

$$B^{-1}_{block} = \begin{pmatrix} 2 & 1 & \dots & 1 \\ 1 & 2 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 2 \end{pmatrix}_{d+1\times d+1}. \tag{4.49}$$

$$(BB^{-1})_{a\alpha,c\gamma} = \sum_{b=1}^{d+1}\sum_{\beta=1}^{d-1} B_{a\alpha,b\beta}B^{-1}_{b\beta,c\gamma} \tag{4.50}$$

$$= \sum_{b=1}^{d+1}\sum_{\beta=1}^{d-1} \delta_{ab}\left(\delta_{\alpha\beta} - \frac{1}{d}\right)\delta_{bc}\left(\delta_{\beta\gamma}+1\right) \tag{4.51}$$

$$= \sum_{\beta=1}^{d+1} \delta_{ab}\delta_{bc}\sum_{\beta=1}^{d-1}\left(\delta_{\alpha\beta} - \frac{1}{d}\right)\left(\delta_{\beta\gamma}+1\right) \tag{4.52}$$

$$= \delta_{ab}\left(\delta_{\alpha\gamma} - \frac{1}{d} + 1 - \frac{d-1}{d}\right) \tag{4.53}$$

$$= \delta_{ac}\delta_{\alpha\gamma} \tag{4.54}$$

$$= \mathbb{I}_{a\alpha,c\gamma}. \tag{4.55}$$

Then, the coefficients are

$$c_{b\beta} = \sum_{a=1}^{d+1}\sum_{\alpha=1}^{d-1} B^{-1}_{b\beta,a\alpha}\left(p_{a\alpha} - \frac{1}{d}\right) \tag{4.56}$$

$$= \sum_{a=1}^{d+1} \sum_{\alpha=1}^{d-1} \delta_{ba} \left(\delta_{\beta\alpha} + 1\right) \left(p_{a\alpha} - \frac{1}{d}\right) \tag{4.57}$$

$$= \sum_{a=1}^{d+1} \sum_{\alpha=1}^{d-1} \delta_{ba} \left(\delta_{\beta\alpha} p_{a\alpha} - \frac{1}{d}\delta_{\beta\alpha} + p_{a\alpha} - \frac{1}{d}\right) \tag{4.58}$$

$$= p_{b\beta} - \frac{1}{d} + \sum_{\alpha=1}^{d-1} p_{b\alpha} - \frac{d-1}{d} \tag{4.59}$$

$$= p_{b\beta} - p_{bd}. \tag{4.60}$$

Substituting in (4.40), we have

$$\rho = \frac{1}{d}\mathbb{I} + \sum_{a=1}^{d+1} \sum_{\alpha=1}^{d-1} (p_{a\alpha} - p_{ad}) \left(\Pi_{a\alpha} - \frac{1}{d}\mathbb{I}\right) \tag{4.61}$$

$$= \frac{1}{d(d+1)} \sum_{a=1}^{d+1} \sum_{\alpha=1}^{d} p_{a\alpha}\mathbb{I} + \sum_{a=1}^{d+1} \sum_{\alpha=1}^{d-1} (p_{a\alpha} - p_{ad}) \left(\Pi_{a\alpha} - \frac{1}{d}\mathbb{I}\right) \tag{4.62}$$

$$= \frac{1}{d(d+1)} \sum_{a=1}^{d+1} p_{ad}\mathbb{I} + \sum_{a=1}^{d} \sum_{\alpha=1}^{d-1} p_{a\alpha} \left(\Pi_{a\alpha} - \frac{1}{d}\mathbb{I} + \frac{1}{d(d+1)}\mathbb{I}\right) - \sum_{a=1}^{d+1} p_{ad} \sum_{\alpha=1}^{d-1} \left(\Pi_{a\alpha} - \frac{1}{d}\mathbb{I}\right) \tag{4.63}$$

$$= \sum_{a=1}^{d+1} \sum_{\alpha=1}^{d-1} p_{a\alpha} \left(\Pi_{a\alpha} - \frac{1}{d+1}\mathbb{I}\right) - \sum_{a=1}^{d+1} p_{ad} \left(\mathbb{I} - \Pi_{ad} - \frac{(d-1)}{d}\mathbb{I} - \frac{1}{d(d+1)}\mathbb{I}\right) \tag{4.64}$$

$$= \sum_{a=1}^{d+1} \sum_{\alpha=1}^{d-1} p_{a\alpha} \left(\Pi_{a\alpha} - \frac{1}{d+1}\mathbb{I}\right) - \sum_{a=1}^{d+1} p_{ad} \left(\frac{1}{d+1}\mathbb{I} - \Pi_{ad}\right) \tag{4.65}$$

$$= \sum_{a=1}^{d+1} \sum_{\alpha=1}^{d} p_{a\alpha} \left(\Pi_{a\alpha} - \frac{1}{d+1}\mathbb{I}\right). \tag{4.66}$$

Therefore, we can perform the tomography of an unknown state by measuring $D = d + 1$ independent basis on a prime power dimensional Hilbert space.

4.2.4 SIC-POVM Tomography

For a Hilbert space of dimension d, a *symmetrically informatically complete POVM* or *SIC-POVM* is a set of d^2 operators $E_m = \pi_m/d$, where $\{\pi_m\}_{m=1,\ldots,d^2}$

are projectors that satisfy

$$\mathrm{Tr}(\pi_i\pi_j) = \frac{1}{d+1}, \qquad \text{for all } i \neq j. \tag{4.67}$$

It is not known if SIC-POVMs exist in all dimensions, but it has been conjectured so [52, 59–61].

Let us consider a Hilbert space endowed with a *SIC-POVM*. We can use the *SIC-POVM* to reconstruct the density matrix ρ because it contains d^2 elements,

$$\rho = \sum_{i=1}^{d^2} c_i \pi_i, \tag{4.68}$$

where the $\{c_i\}_{i=1,\ldots,d^2}$ are real coefficients. From the trace condition, we have

$$\mathrm{Tr}(\rho) = \sum_{i=1}^{d^2} c_i \,\mathrm{Tr}(\pi_i) \tag{4.69}$$

$$1 = \sum_{i=1}^{d^2} c_i. \tag{4.70}$$

Now, if we measure the *SIC-POVM* $\{E_j\}$ on the state ρ, we obtain

$$p_j = \mathrm{Tr}(\rho E_j) \tag{4.71}$$

$$= \frac{1}{d}\sum_{i=1}^{d^2} c_i \,\mathrm{Tr}(\pi_i\pi_j) \tag{4.72}$$

$$= \frac{1}{d}\left[\sum_{\substack{i=1\\ i\neq j}}^{d^2} c_i \,\mathrm{Tr}(\pi_i\pi_j) + c_j \,\mathrm{Tr}\left(\pi_i^2\right)\right] \tag{4.73}$$

$$= \frac{1}{d}\left[\sum_{\substack{i=1\\ i\neq j}}^{d^2} \frac{c_i}{d+1} + c_j\right] \tag{4.74}$$

$$= \frac{1}{d}\left[\frac{1-c_j}{d+1} + c_j\right] \tag{4.75}$$

$$= \frac{1}{1+d}\left[\frac{1}{d} + c_j\right]. \tag{4.76}$$

This is a linear system (4.4) that can be easily solved isolating the coefficients $\{c_j\}_{j=1,\ldots,d^2}$,

$$c_j = (d+1)p_j - \frac{1}{d}. \tag{4.77}$$

Thus, the expression (4.67) becomes

$$\rho = \sum_{i=1}^{d^2} \left[(d+1)p_j - \frac{1}{d} \right] \pi_i, \tag{4.78}$$

and the unknown density matrix can be estimated with an experiment in the SIC-POVM $\{E_j\}_{j=1,\ldots,d^2}$.

4.3 Maximum Likelihood Estimation

The reconstruction of ρ via linear inversion has a problem; although it preserves hermiticity and the trace of the states, these might not be positive semidefinite. This is because we can only obtain estimates of the probabilities experimentally. Let us consider an experiment where we have N identically prepared copies of a state ρ and we perform measurements $\{E_i\}_{i=1,\ldots,m}$. In the experiment we estimate the probabilities $p_i = \mathrm{Tr}(\rho E_i)$ using the numbers of detections $\hat{n}_i$ of E_i as

$$\hat{p}_i = \frac{\hat{n}_i}{N}. \tag{4.79}$$

Since experimental measurements are random variables, they necessarily fluctuate for every finite sample size N [54]. This experiment is analog to the roll of one d sided die. Then, the number of detection $n_i = Np_i$ has multinomial noise,

$$\mathrm{Var}(\hat{n}_i) = Np_i(1-p_i). \tag{4.80}$$

Because of the above, when we employ linear tomography the estimated density matrix $\hat{\rho}$ has error Δ. This error affects the positivity of the density matrix,

$$\langle\varphi|\hat{\rho}|\varphi\rangle = \langle\varphi|(\rho+\Delta)|\varphi\rangle = \langle\varphi|\rho|\varphi\rangle + \langle\varphi|\Delta|\varphi\rangle\,. \tag{4.81}$$

Thereby, when $\langle\varphi|\rho|\varphi\rangle$ is smaller than $-\langle\varphi|\Delta|\varphi\rangle$, the mean value might be negative $\langle\varphi|\hat{\rho}|\varphi\rangle < 0$. Thus, the solution $\hat{\rho}$ might not be a positive semidefinite operator. To solve this problem we use the well-known concept of *maximum likelihood estimation (MLE)* [56, 62, 63]. If we have observed a set of detections $\hat{\boldsymbol{n}} = \begin{pmatrix} \hat{n}_1 & \cdots & \hat{n}_m \end{pmatrix}^{\mathrm{T}}$ from m measurements $\{E_i\}_{i=1,\ldots,m}$, the likelihood function is

$$\mathcal{L}(\rho|\hat{\boldsymbol{n}}) = \prod_{i=1}^{m} \mathrm{Tr}(\rho E_i)^{\hat{n}_i} \tag{4.82}$$

For computational simplicity, we work with the log-likelihood function,

$$\mathcal{F}(\rho|\hat{\boldsymbol{n}}) = \log\Big(\mathcal{L}(\rho|\hat{\boldsymbol{n}})\Big) = \sum_i \hat{n}_i \log\Big(\operatorname{Tr}(\rho E_i)\Big). \tag{4.83}$$

It can be approximated by a Gaussian distribution [63] for large sample size,

$$\mathcal{F}(\rho|\hat{\boldsymbol{n}}) = -\frac{1}{2}\sum_{i=1}^{m} \frac{[n_i(\rho) - \hat{n}_i]^2}{N_i p_i(1-p_i)} = -\frac{1}{2}\sum_{i=1}^{m} \frac{N_i\,[\operatorname{Tr}(E_i\rho) - \hat{p}_i]^2}{\operatorname{Tr}(E_i\rho)[1-\operatorname{Tr}(E_i\rho)]}, \tag{4.84}$$

where N_i is the sample on which E_i was measured. Then, the estimated density matrix is

$$\hat{\rho}(\hat{\boldsymbol{n}}) = \arg\max_{\rho\in\mathcal{P}} \mathcal{F}(\rho|\hat{\boldsymbol{n}}). \tag{4.85}$$

This optimization is usually done by a gradient descent algorithm by parameterizing ρ by *Cholesky Decomposition*

$$\rho(\boldsymbol{t}) = \frac{T^\dagger(\boldsymbol{t})T(\boldsymbol{t})}{\operatorname{Tr}\left(T^\dagger(\boldsymbol{t})T(\boldsymbol{t})\right)}, \tag{4.86}$$

with $\boldsymbol{t} = \begin{pmatrix} t_1 & \cdots & t_{d^2} \end{pmatrix}^{\mathrm{T}}$ a vector of reals parameters and $T(\boldsymbol{t})$ a lower triangular matrix,

$$T(\boldsymbol{t}) = \begin{pmatrix} t_1 & 0 & \cdots & 0 \\ t_{d+1} + \mathrm{i}t_{d+2} & t_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ t_{d^2-1} + \mathrm{i}t_{d^2} & t_{d^2-3} + \mathrm{i}t_{d^2-2} & \cdots & t_d \end{pmatrix}. \tag{4.87}$$

As our initial guess, we consider our tomographic density matrix by linear tomography,

$$\rho(\boldsymbol{t}_{guess}) = \frac{|\hat{\rho}_{lin}|}{\operatorname{Tr}|\hat{\rho}_{lin}|} = \frac{\sqrt{\hat{\rho}_{lin}^\dagger\hat{\rho}_{lin}}}{\operatorname{Tr}\left(\sqrt{\hat{\rho}_{lin}^\dagger\hat{\rho}_{lin}}\right)}. \tag{4.88}$$

Thereby, the tomography by linear inversion together with post-processing by MLE provide a complete solution for the problem of Quantum Tomography.

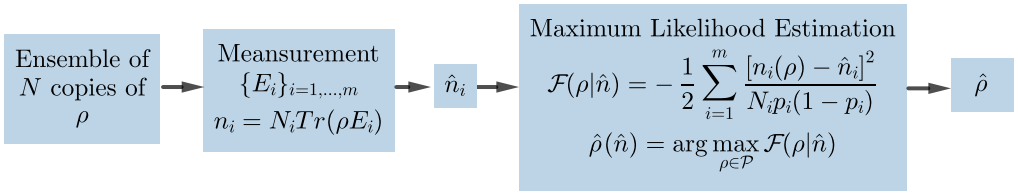

Figure 4.1: Diagram of the quantum state tomography. Source: Made by the author.

The disadvantage of this method is that it quickly becomes intractable in high dimensions. For example, a post-processing time of two weeks has been reported for the reconstruction of a state in dimension $d = 2^8$ [14, 64].

4.4 Research in Quantum Tomography

Quantum tomography is a current topic of research. Due to the high number of experimental results of the standard tomography $d(d^2-1)$, initially the research focused on the formulation of tomographic methods with the least possible number of measurement outcomes [37, 52, 53, 65]. Examples of this are MUBs tomography with $d(d+1)$ outcomes and tomography by SIC-POVM with d^2 outcomes. Nowadays, it is also investigated tomographic methods that increase the accuracy of the estimate for a given ensemble size [32, 35, 66] and tomographic methods with a priori information [66–68]. Besides, experimental realizations of all these quantum tomography protocols are investigated [14, 69–72]. In particular, methods that involve POVMs and collective measurements are especially difficult to implement. From the computational point of view, it seeks to formulate new post-processing methods and faster algorithms to reduce computing times [63, 64, 73].

Chapter 5

Adaptive Standard Quantum Tomography

The quantum tomographic protocols described in the previous chapter are based on a fixed set of measurements. These methods have different accuracy depending on the state. This unwanted feature can be overcome by performing a sequence of tomographic reconstructions, where we use the information about the state obtained in previous measurements to then perform a better tomography in the next round. This is called *adaptive quantum tomography* [34,36–39]. The simplest protocol consists in performing a two-stage standard tomography [32,33,35]. In the first stage, a low accuracy estimate is obtained. The basis of eigenvectors of this estimate is then used to adapt the measurement base of the Pauli matrices for a second, higher accuracy tomography. Other more elaborated methods are the self-learning tomographic methods [36–39], which optimizes a utility function between each step. These methods are characterized by a high computational expense.

In this chapter, briefly review two-stage adaptive quantum tomography for a single qubit. Afterward, we generalize the two-stage adaptive standard tomography to the case of a single qudit, that is, higher dimensional quantum systems. Finally, we study the accuracy provided by adaptive quantum tomography in higher dimensions and compare it with the relevant bounds.

5.1 Motivation: Qubit Adaptive Tomography

In this section we describe the adaptive tomography of a qubit in the state ρ [32,33]. The standard tomography of a qubit consists in measuring the observables $\{\sigma_x, \sigma_y, \sigma_z\}$, since we have that

$$\rho = \frac{1}{2}\Big(\mathbb{I} + S_x\sigma_x + S_y\sigma_y + S_z\sigma_z\Big). \tag{5.1}$$

Each observable has two outcomes $\{-,+\}$. Let $\hat{p}_i^j$ the estimate of the probability of outcome $j = \pm$ when we measure observable σ_i with $i = x, y, z$. The experimentally obtained probabilities have, in the case of a perfect detection procedure, multinomial noise (4.79,4.80)

$$\hat{p}_i^j = \frac{n_i^j}{N_i} = \frac{3n_i^j}{N}, \qquad \mathrm{Cov}\Big(\hat{p}_i^j, \hat{p}_i^{j'}\Big) = \frac{1}{N_i}p_i^j(\delta_{j,j'} - p_i^{j'}) = \frac{3}{N}p_i^j(\delta_{j,j'} - p_i^{j'}), \tag{5.2}$$

where N is the total number of detections and N_j is the fraction of N employed in the measurement of the observable σ_j. The estimate $\hat{S}_i$ of $S_i = \mathbb{E}(\sigma_i) = Tr(\rho\sigma_i)$ is given by

$$\hat{S}_i = \hat{p}_i^+ - \hat{p}_i^-. \tag{5.3}$$

Then, the variance of $\hat{S}_i$ is (2.122) given by

$$\begin{aligned}
\mathrm{Var}\Big(\hat{S}_i\Big) =& \mathrm{Var}\Big(\hat{p}_i^+\Big) - \mathrm{Var}\Big(\hat{p}_i^-\Big) + 2\mathrm{Cov}\Big(\hat{p}_i^+, \hat{p}_i^-\Big) && (5.4)\\
=& \frac{3}{N}\Big[p_i^+(1-p_i^+) + p_i^-(1-p_i^-) + 2p_i^+p_i^-\Big] && (5.5)\\
=& \frac{3}{N}\Big[1 - p_+^2 - p_-^2 + 2p_+p_-\Big] && (5.6)\\
=& \frac{3}{N}\Big[1 - (p_i^+ - p_i^-)^2\Big] && (5.7)\\
=& \frac{3}{N}\Big[1 - S_i^2\Big]. && (5.8)
\end{aligned}$$

Clearly, the variance depends on the state ρ.

On the other hand, the fidelity between states of a single qubit can be calculated as (3.71)

$$F(\rho,\sigma) = \mathrm{Tr}(\rho\sigma) + \sqrt{1 - \mathrm{Tr}(\rho^2)}\sqrt{1 - \mathrm{Tr}(\sigma^2)}. \tag{5.9}$$

Let us suppose that $\sigma = \rho + \Delta$ is an estimate of ρ with error Δ, which is hermitian and traceless.

$$\Delta = \sigma - \rho = \frac{1}{2}\sum_{i=1}^{3}(\hat{S}_i - S_i)\sigma_i. \tag{5.10}$$

In general, the error matrix scale as the uncertainty (2.113),

$$||\Delta|| = \sqrt{\mathrm{Tr}(\Delta^2)} \tag{5.11}$$

$$= \mathrm{Tr}\left(\left[\frac{1}{2}\sum_{i=1}^{3}(\hat{S}_i - S_i)\sigma_i\right]\left[\frac{1}{2}\sum_{j=1}^{3}(\hat{S}_j - S_j)\sigma_j\right]\right)^{1/2} \tag{5.12}$$

$$= \left[\frac{1}{4}\sum_{i,j=1}^{3}(\hat{S}_i - S_i)(\hat{S}_j - S_j)\,\mathrm{Tr}\,(\sigma_i\sigma_j)\right]^{1/2} \tag{5.13}$$

$$= \left[\frac{1}{2}\sum_{i=1}^{3}(\hat{S}_i - S_j)^2\right]^{1/2} \tag{5.14}$$

$$\sim \mathcal{O}\Big(1/\sqrt{N}\Big). \tag{5.15}$$

Then

$$F(\rho, \rho+\Delta) = \mathrm{Tr}(\rho[\rho+\Delta]) + \sqrt{1-\mathrm{Tr}(\rho^2)}\sqrt{1-\mathrm{Tr}([\rho+\Delta]^2)} \tag{5.16}$$

$$= \mathrm{Tr}\big(\rho^2\big) + \mathrm{Tr}(\rho\Delta) + \sqrt{1-\mathrm{Tr}(\rho^2)}\sqrt{1-\mathrm{Tr}(\rho^2+\rho\Delta+\Delta\rho+\Delta^2)}. \tag{5.17}$$

If ρ has a high purity, that is, $1-\mathrm{Tr}(\rho^2) \approx 0$, the the fidelity approximately becomes

$$F(\rho, \rho+\Delta) = 1 + \mathrm{Tr}(\rho\Delta) + \mathcal{O}\Big(||\Delta||^2\Big). \tag{5.18}$$

Otherwise,

$$F(\rho, \rho+\Delta) = \mathrm{Tr}\big(\rho^2\big) + \mathrm{Tr}(\rho\Delta) + \Big(1-\mathrm{Tr}\big(\rho^2\big)\Big)\sqrt{1-\frac{\mathrm{Tr}(2\rho\Delta+\Delta^2)}{1-\mathrm{Tr}(\rho^2)}} \tag{5.19}$$

$$= \mathrm{Tr}\big(\rho^2\big) + \mathrm{Tr}(\rho\Delta) + \Big(1-\mathrm{Tr}\big(\rho^2\big)\Big)\left(1-\frac{\mathrm{Tr}(2\rho\Delta+\Delta^2)}{2[1-\mathrm{Tr}(\rho^2)]}\right) + \mathcal{O}\Big(||\Delta||^2\Big) \tag{5.20}$$

$$= \mathrm{Tr}\big(\rho^2\big) + \mathrm{Tr}(\rho\Delta) + 1 - \mathrm{Tr}\big(\rho^2\big) - \mathrm{Tr}(\rho\Delta) - \frac{\mathrm{Tr}(\Delta^2)}{2} + \mathcal{O}\Big(||\Delta||^2\Big) \tag{5.21}$$

$$=1-\frac{\mathrm{Tr}(\Delta^2)}{2}+\mathcal{O}\Big(||\Delta||^2\Big). \tag{5.22}$$

Therefore, the fidelity for high purity states scale as $||\Delta|| \sim \mathcal{O}\Big(1/\sqrt{N}\Big)$ and for mixed states scale as $||\Delta||^2 \sim \mathcal{O}(1/N)$. This occurs because the fidelity is sensitive to the zero-eigenvalue. To minimize the infidelity, we must accurately estimate this eigenvalue,

$$\lambda_{\pm}=\frac{1}{2}\Big(1\pm\sqrt{S_x^2+S_y^2+S_z^2}\Big). \tag{5.23}$$

On the basis of eigenvectors of ρ the eigenvalues only depend on $\Big|S_z^{(diag)}\Big|$,

$$\lambda_{\pm}=\frac{1}{2}\big(1\pm\big|S_z^{(diag)}\big|\big). \tag{5.24}$$

If the state has hight purity $\Big|S_z^{(diag)}\Big| \approx 1$, the uncertainty of $S_z^{(diag)}$ is

$$\mathrm{Var}\Big(\hat{S}_z^{(diag)}\Big) \approx \frac{3}{N}\Big[1-\big|S_z^{(diag)}\big|^2\Big] \approx 0. \tag{5.25}$$

Thereby, the uncertainty of $S_z^{(diag)}$ and the eigenvalues vanish in this basis. That is, the infidelity is minimized measuring the observables on the basis of eigenvectors of ρ.

This tomography is not feasible since it requires knowing the base of eigenstates of an unknown state. However, we can perform standard tomography on a sample of size $N_0 < N$ to generate a first estimate $\hat{\rho}_0$ of ρ. Thereafter, we perform a second standard tomography on the remaining sample of size $N - N_0$, writing the observables σ_i in the basis of eigenvector of $\hat{\rho}_0$. This new tomographic scheme provides an infidelity that scale as $||\Delta||^2 \sim \mathcal{O}(1/N)$ for all states, at the expense of increasing the total number of measurement outcomes. The choice of the fraction of the sample N_0 is not trivial. However, it had been suggested that $N_0 = N^{2/3}$ would be sufficient [28], but numerical simulations have shown that the choice $N_0 = N/2$ provides the best results [32, 33].

This adaptive quantum tomographic scheme can be easily generalized to minimize the weighted mean square error $\mathrm{Tr}(WC)$ of the parameters $\{S_x, S_y, S_z\}$ [35]. This is equivalent to the optimization problem with J the Fisher information matrix of

(5.1). The solution to his problem is (3.118), which attain the upper bound of Gill-Massar inequality. Supposing W diagonal and rotating the observables to the basis of eigenvectors of ρ,

$$W = \begin{pmatrix} w_x & 0 & 0 \\ 0 & w_y & 0 \\ 0 & 0 & w_z \end{pmatrix}, \qquad J^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1-S^2 \end{pmatrix}. \tag{5.26}$$

The classical Fisher information matrix becomes

$$I = \frac{N}{\sqrt{w_x}+\sqrt{w_y}+\sqrt{w_z(1-S^2)}} \begin{pmatrix} \sqrt{w_x} & 0 & 0 \\ 0 & \sqrt{w_y} & 0 \\ 0 & 0 & \sqrt{\frac{w_z}{1-S^2}} \end{pmatrix}, \tag{5.27}$$

and

$$IJ^{-1} = \frac{N}{\sqrt{w_x}+\sqrt{w_y}+\sqrt{w_z(1-S^2)}} \begin{pmatrix} \sqrt{w_x} & 0 & 0 \\ 0 & \sqrt{w_y} & 0 \\ 0 & 0 & \sqrt{w_z(1-S^2)} \end{pmatrix}. \tag{5.28}$$

Clearly, the maximum of the Gill-Massar inequality is reached $\mathrm{Tr}(IJ^{-1}) = N$ (3.116). The optimal measurement protocol has weighted mean square error

$$\mathrm{Tr}(WC) = \frac{1}{N}\left[\sqrt{w_x}+\sqrt{w_y}+\sqrt{w_z(1-S^2)}\right]^2, \tag{5.29}$$

and it is attained measuring the observables with probabilities

$$p_x = \frac{\sqrt{w_x}}{\sqrt{w_x}+\sqrt{w_y}+\sqrt{w_z(1-S^2)}}, \tag{5.30}$$

$$p_y = \frac{\sqrt{w_y}}{\sqrt{w_x}+\sqrt{w_y}+\sqrt{w_z(1-S^2)}}, \tag{5.31}$$

$$p_z = \frac{\sqrt{w_z(1-S^2)}}{\sqrt{w_x}+\sqrt{w_y}+\sqrt{w_z(1-S^2)}}. \tag{5.32}$$

Therefore, the second tomography must be performed with different samples per observable,

$$N_i = Np_i, \qquad i = x, y, z. \tag{5.33}$$

In particular, in the case $W = \mathbb{I}$ we have that

$$\mathrm{Tr}(C) = \frac{1}{N}\left(2+\sqrt{1-S^2}\right)^2, \tag{5.34}$$

$$p_x = p_y = \frac{1}{2+\sqrt{1-S^2}}, \qquad p_z = \frac{\sqrt{1-S^2}}{2+\sqrt{1-S^2}}. \tag{5.35}$$

If $W = J$, we obtain the adaptive quantum tomography scheme for optimizing the infidelity,

$$\mathrm{Tr}(JC) = \frac{9}{N}, \tag{5.36}$$

and thus

$$p_x = p_y = p_z = \frac{1}{3}. \tag{5.37}$$

Figures (5.1) and (5.3) compare the mean infidelity over 10^3 randomly chosen qubits with respect to the resource N (E). The mean infidelities had been obtained by standard tomography, diagonal tomography, adaptive tomography with $N_0 = N/2$ and $N_0 = N^{2/3}$, and weighted adaptive tomography $W = \mathbb{I}$ with $N_0 = N/2$. The dot-dashed lines are the best fit $\mathrm{Inf}(N, \alpha, \beta) = \beta/N^{\alpha}$ by least squares. Tables (5.1) and (5.3) show the fit coefficients.

From the simulations, it can be seen that with the same resource, adaptive tomography and weighted adaptive tomography delivers better infidelities than standard tomography. Moreover, adaptive tomography and weighted adaptive tomography scale better with the sample size N than standard quantum tomography because they have higher fit coefficient α. Besides, it can be seen that the initial sample that delivers the best infidelities for pure states is $N_0 = N/2$. On the other hand, the initial sample that delivers better infidelities for mixed states is $N_0 = N^{2/3}$. Weighted adaptive tomography delivers infidelities close to adaptive tomography except for quasi-pure states, despite the Weighted adaptive tomography minimizes another error function.

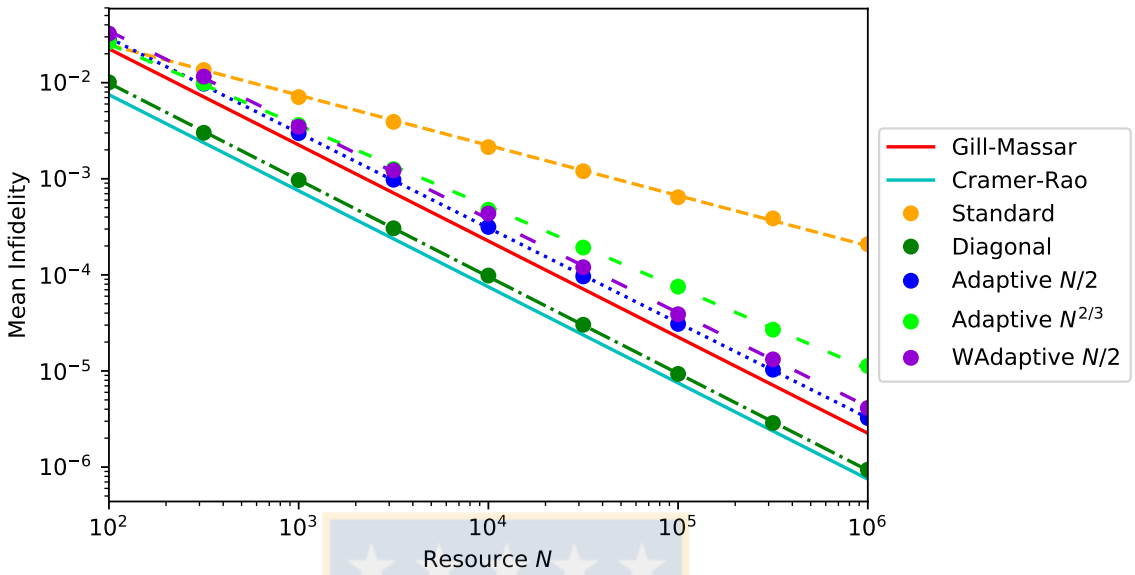

Figure 5.1: Comparison of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 1000 2-dimensional pure states. Source: Made by the author.

	α	β
Gill-Massar	1	2.25
Cramer-Rao	1	0.75
Standard	0.522668908263	0.274554132708
Diagonal	1.00730675152	1.02387919204
Adaptive $N/2$	0.988535035084	2.77264762811
Adaptive $N^{2/3}$	0.845224681067	1.23603792696
WAdaptive $N/2$	0.977353943227	3.10776158269

Table 5.1: Comparison of fit coefficients of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 1000 2-dimensional pure states. Source: Made by the author.

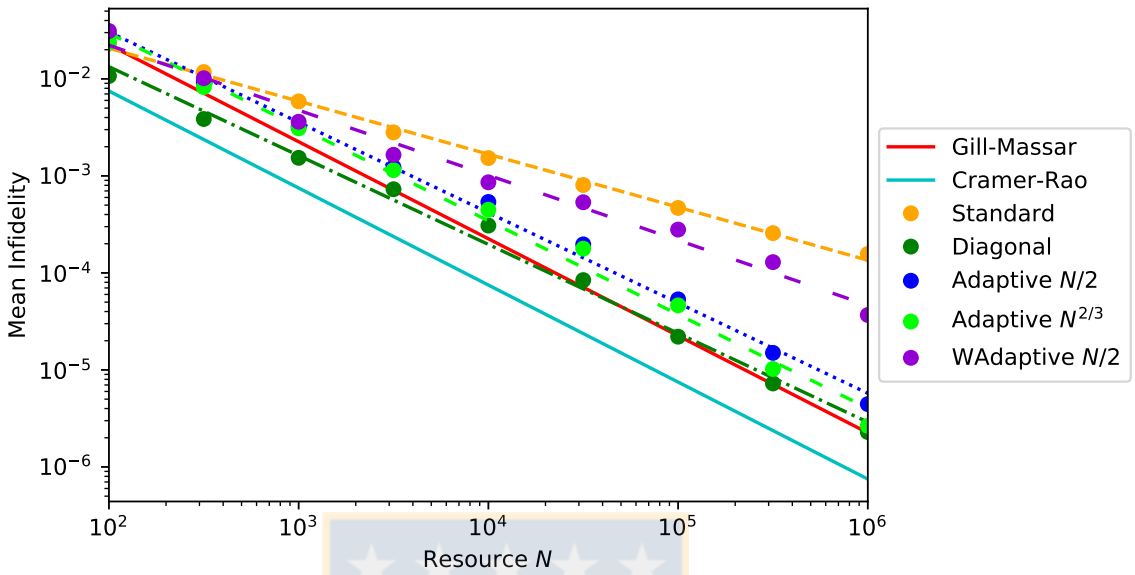

Figure 5.2: Comparison of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 1000 2-dimensional pure states with noise. Source: Made by the author.

	α	β
Gill-Massar	1	2.25
Cramer-Rao	1	0.75
Standard	0.546279818001	0.25649091079
Diagonal	0.916058230105	0.90762871724
Adaptive $N/2$	0.931322971963	2.22114366554
Adaptive $N^{2/3}$	0.966225512222	2.52922075262
WAdaptive $N/2$	0.670216628787	0.487350341131

Table 5.2: Comparison of fit coefficients of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 1000 2-dimensional pure states with noise. Source: Made by the author.

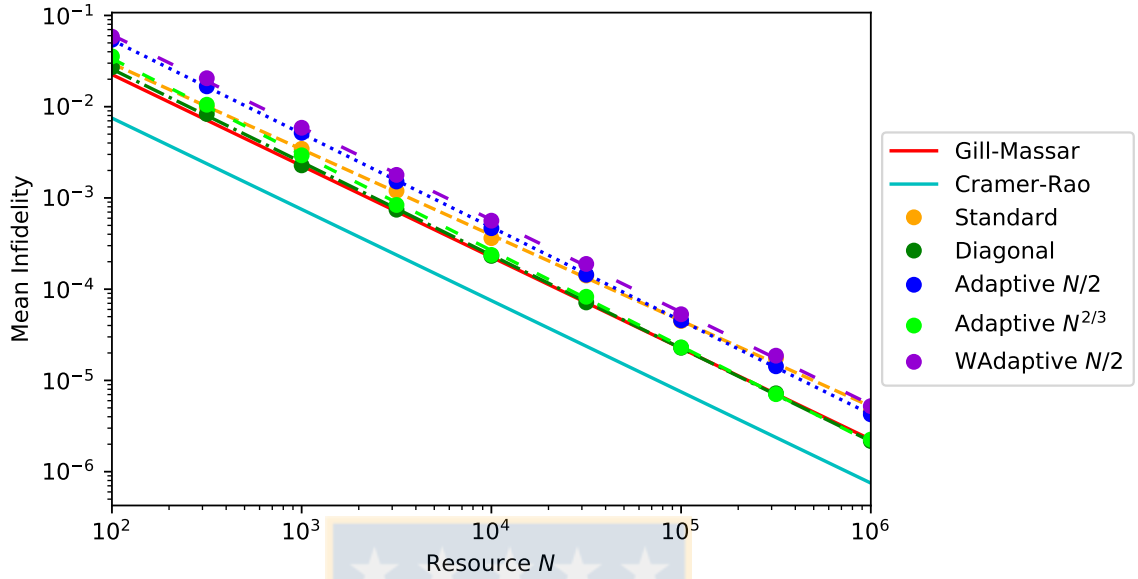

Figure 5.3: Comparison of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 1000 2-dimensional full rank states. Source: Made by the author.

	α	β
Gill-Massar	1	2.25
Cramer-Rao	1	0.75
Standard	0.939440415495	2.24594502968
Diagonal	1.01955462218	2.82365939907
Adaptive $N/2$	1.02533901125	6.04325441363
Adaptive $N^{2/3}$	1.05096054007	4.2472844212
WAdaptive $N/2$	1.01263809377	6.46776209741

Table 5.3: Comparison of fit coefficients of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 1000 2-dimensional states with rank 2. Source: Made by the author.

5.2 Foundations

The formulation of the qubit adaptive tomography require the use of quantities such as variance, fidelity and the quantum Fisher information matrix. In this section, we generalized these concepts to the case of a single qudit. Thereafter, these generalizations will be employed to extend adaptive tomography to the case of higher dimensional quantum systems.

5.2.1 Variance of Standard Tomography

Let us consider the standard tomography (4.25) of an unknown qudit. If we aim at the estimation of ρ with N preparations of the state, we employ $N/(d^2-1)$ different samples to obtain each parameter $S_i = \mathbb{E}(\sigma_i)$. This is equivalent to measuring each observable with probability $p(\sigma_j) = 1/(d^2-1)$, with $j = 1, ..., d^2-1$. The variance of measuring N times the parameters can be calculated using the fact that the conditional probabilities $\{p(i|\sigma_k)\}_{i=1,...,d}$ have, for a perfect detection process, multinomial noise, that is,

$$\mathrm{Cov}\big(\hat{p}(i|\sigma_k), \hat{p}(j|\sigma_k)\big) = \frac{d^2-1}{N} p(i|\sigma_k)\Big[\delta_{ij} - p(j|\sigma_k)\Big]. \tag{5.38}$$

Then, for S_{ij}^x and S_{ij}^y,

$$\begin{aligned}
\mathrm{Var}^{(N)}\big(\hat{S}_{ij}^{xy}\big) =& \mathrm{Var}\big(\hat{p}(+|\sigma_{ij}^{xy})\big) + \mathrm{Var}\big(\hat{p}(-|\sigma_{ij}^{xy})\big) + 2\mathrm{Cov}\big(\hat{p}(+|\sigma_{ij}^{xy}), \hat{p}(-|\sigma_{ij}^{xy})\big) && (5.39)\\
=& \frac{d^2-1}{N}\Big[p(+|\sigma_{ij}^{xy})\big[1-p(+|\sigma_{ij}^{xy})\big] + p(-|\sigma_{ij}^{xy})\big[1-p(-|\sigma_{ij}^{xy})\big] \\
& + 2p(+|\sigma_{ij}^{xy})p(-|\sigma_{ij}^{xy})\Big] && (5.40)\\
=& \frac{d^2-1}{N}\Big[p(+|\sigma_{ij}^{xy}) - p(+|\sigma_{ij}^{xy})^2 + p(-|\sigma_{ij}^{xy}) - p(-|\sigma_{ij}^{xy})^2 \\
& + 2p(+|\sigma_{ij}^{xy})p(-|\sigma_{ij}^{xy})\Big] && (5.41)\\
=& \frac{d^2-1}{N}\Big[p(+|\sigma_{ij}^{xy}) + p(-|\sigma_{ij}^{xy}) - \big[p(+|\sigma_{ij}^{xy}) - p(-|\sigma_{ij}^{xy})\big]^2\Big] && (5.42)\\
=& \frac{d^2-1}{N}\Big[p(+|\sigma_{ij}^{xy}) + p(-|\sigma_{ij}^{xy}) - (S_{ij}^{xy})^2\Big] && (5.43)\\
=& \frac{d^2-1}{N}\Big[p_i + p_j - (S_{ij}^{xy})^2\Big]. && (5.44)
\end{aligned}$$

The parameters S_k^z are a linear combination of $\{p_l\}_{l=1,\dots,d}$,

$$S_k^z = \sum_l \Gamma_{kl} p_l \qquad \Gamma_{kl} = \sqrt{\frac{2}{k(k+1)}}\Big[h_{k\geq l} - k\delta_{k+1,l}\Big], \tag{5.45}$$

where $h_{k\geq l}$ is the *step function* defined by

$$h_{k\geq l} = \begin{cases} 1, & k \geq l. \\ 0, & k < l. \end{cases} \tag{5.46}$$

Then, we can calculate their variances as

$$\begin{aligned}
\mathrm{Var}^{(N)}\big(\hat{S}_k^z\big) &= \sum_{m,n=1}^{d} \Gamma_{km}\Gamma_{kn}\mathrm{Cov}(\hat{p}_n, \hat{p}_m) && (5.47)\\
&= \sum_{n,m=1}^{d} \frac{2}{k(k+1)}\big(h_{k\geq m} - k\delta_{k+1,m}\big)\big(h_{k\geq n} - k\delta_{k+1,n}\big)\mathrm{Cov}(\hat{p}_n, \hat{p}_m) && (5.48)\\
&= \frac{2}{k(k+1)} \sum_{n,m=1}^{d} \mathrm{Cov}(\hat{p}_n, \hat{p}_m)\Big(h_{k\geq n}h_{k\geq m} - k\delta_{k+1,n}h_{k\geq m} \\
&\qquad - kh_{k\geq n}\delta_{k+1,m} + k^2\delta_{k+1,n}\delta_{k+1,m}\Big) && (5.49)\\
&= \frac{2}{k(k+1)}\Bigg[\sum_{n,m=1}^{k} \mathrm{Cov}(\hat{p}_n, \hat{p}_m) - k\sum_{m=1}^{k}\mathrm{Cov}(\hat{p}_{k+1}, \hat{p}_m) \\
&\qquad - k\sum_{n=1}^{k}\mathrm{Cov}(\hat{p}_n, \hat{p}_{k+1}) + k^2\mathrm{Cov}(\hat{p}_{k+1}, \hat{p}_{k+1})\Bigg] && (5.50)\\
&= \frac{2}{k(k+1)}\Bigg[\sum_{n,m=1}^{k} \mathrm{Cov}(\hat{p}_n, \hat{p}_m) - 2k\sum_{m=1}^{k}\mathrm{Cov}(\hat{p}_{k+1}, \hat{p}_m) \\
&\qquad + k^2\mathrm{Var}(\hat{p}_{k+1})\Bigg] && (5.51)\\
&= \frac{d^2-1}{N}\frac{2}{k(k+1)}\Bigg[\sum_{n=1}^{k} p_n - \sum_{n,m=1}^{k} p_n p_m - 2k\left(-p_{k+1}\sum_{m=1}^{k} p_m\right) \\
&\qquad + k^2\Big(p_{k+1} - (p_{k+1})^2\Big)\Bigg] && (5.52)\\
&= \frac{d^2-1}{N}\frac{2}{k(k+1)}\left[\sum_{n=1}^{k} p_n + k^2 p_{k+1} - \left(\sum_{n=1}^{k} p_n - kp_{k+1}\right)^2\right] && (5.53)
\end{aligned}$$

$$=\frac{d^2-1}{N}\left(\frac{2}{k(k+1)}\left[\sum_{n=1}^{k}p_n+k^2p_{k+1}\right]-\left(S_k^z\right)^2\right). \tag{5.54}$$

In standard tomography, each coefficient is estimated independently. Therefore, the covariance matrix is diagonal with elements

$$\mathrm{Var}^{(N)}\left(\hat{S}_{ij}^x\right)=\frac{d^2-1}{N}\left[p_i+p_j-\left(S_{ij}^x\right)^2\right], \tag{5.55}$$

$$\mathrm{Var}^{(N)}\left(\hat{S}_{ij}^y\right)=\frac{d^2-1}{N}\left[p_i+p_j-\left(S_{ij}^y\right)^2\right], \tag{5.56}$$

$$\mathrm{Var}^{(N)}\left(\hat{S}_k^z\right)=\frac{d^2-1}{N}\left[\frac{2}{k(k+1)}\left(\sum_{n=1}^{k}p_n+k^2p_{k+1}\right)-\left(S_k^z\right)^2\right]. \tag{5.57}$$

We see that the uncertainty in the estimation of the coefficients $\{S_i\}_{i=1,\ldots,d^2-1}$, described by the elements of the covariance matrix, also depends on these parameters, that is, the uncertainty in the estimation of the parameters defining ρ depends on ρ itself.

Since these parameters depend on the basis on which ρ is expressed, the uncertainty of the tomographic reconstruction depends on the basis on which it is performed [32,33]. In particular, if the state ρ has rank r, its parameters $\{S_i\}_{i=1,\ldots,d^2-1}$ on its basis of eigenvectors are

$$S_{ij}^x=0, \tag{5.58}$$

$$S_{ij}^y=0, \tag{5.59}$$

$$S_k^z=\begin{cases}\sqrt{\dfrac{2}{k(k+1)}}\left[\displaystyle\sum_{l=1}^{k}\lambda_l-k\lambda_{k+1}\right], & k<r,\\ \sqrt{\dfrac{2}{k(k+1)}}, & k\geq r,\end{cases} \tag{5.60}$$

$$\tag{5.61}$$

where $\{\lambda_k\}_{k=1,\ldots,r}$ are the eigenvalues of ρ. The variance of S_{ij}^x and S_{ij}^y becomes

$$\mathrm{Var}^{(N)}(\hat{S}_{ij}^x)=\mathrm{Var}^{(N)}(\hat{S}_{ij}^y)=\frac{d^2-1}{N}\left[\lambda_i+\lambda_j\right]. \tag{5.62}$$

For $k<r$, the variance of S_k^z is

$$\mathrm{Var}^{(N)}(\hat{S}_k^z)=\frac{d^2-1}{N}\left[\frac{2}{k(k+1)}\left(\sum_{n=1}^{k}\lambda_n+k^2\lambda_{k+1}\right)-\left(S_k^z\right)^2\right], \tag{5.63}$$

and for $k \geq r$ the variance becomes

$$\mathrm{Var}^{(N)}(\hat{S}_k^z) = \frac{d^2-1}{N}\left[\frac{2}{k(k+1)}\left(\sum_{n=1}^{k}\lambda_n + k^2\lambda_{k+1}\right) - \left(S_k^z\right)^2\right] \tag{5.64}$$

$$= \frac{1}{N}\left[\frac{2}{k(k+1)} - \frac{2}{k(k+1)}\right] \tag{5.65}$$

$$= 0. \tag{5.66}$$

Then, the variance in the estimates of $\{S_k^z\}_{k=r,\ldots,d-1}$ vanishes.

5.2.2 Approximation of the Fidelity

In this section, we calculate the second order approximation of the fidelity by resorting to the Fréchet derivative. The fidelity between two quantum states ρ and σ is defined as

$$F(\rho,\sigma) = \mathrm{Tr}\left(\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}}\right)^2. \tag{5.67}$$

The main advantage of resorting to the Fréchet derivative is that it does not require a particular parametrization of the space of density matrices, so this calculation is valid for any class of tomography.

Let us consider that the density matrix ρ has rank $r \leq d$ and that the state σ is an infinitesimal perturbation of ρ, that is,

$$\sigma = \rho + \Delta, \tag{5.68}$$

where Δ is a hermitian and traceless matrix such that $||\Delta|| = \sqrt{\mathrm{Tr}(\Delta^2)} \ll 1$. For a tomographic process, Δ is the error matrix. In the case of a standard tomography we have

$$\Delta = \sigma - \rho \tag{5.69}$$

$$= \frac{1}{d}\mathbb{I} + \frac{1}{2}\sum_{i=1}^{d^2-1}\hat{S}_i\sigma_i - \left(\frac{1}{d}\mathbb{I} + \frac{1}{2}\sum_{i=1}^{d^2-1}S_i\sigma_i\right) \tag{5.70}$$

$$= \frac{1}{2}\sum_{i=1}^{d^2-1}\left(\hat{S}_i - S_i\right)\sigma_i. \tag{5.71}$$

The norm of this matrix is similar to the uncertainty (2.113)

$$
\begin{aligned}
||\Delta|| =& \sqrt{\mathrm{Tr}(\Delta^2)} && (5.72)\\
=& \mathrm{Tr}\left(\left[\frac{1}{2}\sum_{i=1}^{d^2-1}(\hat{S}_i - S_i)\sigma_i\right]\left[\frac{1}{2}\sum_{j=1}^{d^2-1}(\hat{S}_j - S_j)\sigma_j\right]\right)^{1/2} && (5.73)\\
=& \left[\frac{1}{4}\sum_{i,j=1}^{d^2-1}(\hat{S}_i - S_i)(\hat{S}_j - S_j)\,\mathrm{Tr}\left(\sigma_i\sigma_j\right)\right]^{1/2} && (5.74)\\
=& \left[\frac{1}{2}\sum_{i=1}^{d^2-1}(\hat{S}_i - S_j)^2\right]^{1/2} && (5.75)\\
\sim& \mathcal{O}\left(1/\sqrt{N}\right). && (5.76)
\end{aligned}
$$

Without loss of generality, we can write ρ as a direct sum between the subspace with rank r and a $d-r$ null matrix Θ_{d-r},

$$
\rho = \rho_r \oplus \Theta_{d-r} = \begin{pmatrix} \rho_r & \Theta_{r,d-r} \\ \Theta_{d-r,r} & \Theta_{d-r} \end{pmatrix}. \tag{5.77}
$$

Writing

$$
\Delta = \begin{pmatrix} \Delta_r & \Delta_{r,d-r} \\ \Delta_{d-r,r} & \Delta_{d-r} \end{pmatrix}, \tag{5.78}
$$

where $\Delta_{d-r} \in \mathcal{P}(\mathcal{H})$ and $\mathrm{Tr}(\Delta_r) = -\mathrm{Tr}(\Delta_{d-r})$. Thus,

$$
\begin{aligned}
\sqrt{\rho}\Delta\sqrt{\rho} =& \begin{pmatrix} \sqrt{\rho_r} & \Theta_{r,d-r} \\ \Theta_{d-r,r} & \Theta_{d-r} \end{pmatrix}\begin{pmatrix} \Delta_r & \Delta_{r,d-r} \\ \Delta_{d-r,r} & \Delta_{d-r} \end{pmatrix}\begin{pmatrix} \sqrt{\rho_r} & \Theta_{r,d-r} \\ \Theta_{d-r,r} & \Theta_{d-r} \end{pmatrix} && (5.79)\\
=& \begin{pmatrix} \sqrt{\rho_r}\Delta_r & \sqrt{\rho_r}\Delta_{r,d-r} \\ \Theta_{d-r,r} & \Theta_{d-r} \end{pmatrix}\begin{pmatrix} \sqrt{\rho_r} & \Theta_{r,d-r} \\ \Theta_{d-r,r} & \Theta_{d-r} \end{pmatrix} && (5.80)\\
=& \begin{pmatrix} \sqrt{\rho_r}\Delta_r\sqrt{\rho_r} & \Theta_{r,d-r} \\ \Theta_{d-r,r} & \Theta_{d-r} \end{pmatrix} && (5.81)\\
=& \sqrt{\rho_r}\Delta_r\sqrt{\rho_r} \oplus \Theta_{d-r}. && (5.82)
\end{aligned}
$$

Thereby, the fidelity becomes

$$
\begin{aligned}
F(\rho, \rho+\Delta) =& \mathrm{Tr}\left(\sqrt{\rho^2 + \sqrt{\rho}\Delta\sqrt{\rho}}\right)^2 && (5.83)\\
=& \mathrm{Tr}\left(\sqrt{\rho_r^2 \oplus \Theta_{d-r} + \sqrt{\rho_r}\Delta_r\sqrt{\rho_r} \oplus \Theta_{d-r}}\right)^2 && (5.84)
\end{aligned}
$$

$$= \mathrm{Tr}\left(\sqrt{\rho_r^2 + \sqrt{\rho_r}\Delta_r\sqrt{\rho_r}}\right)^2 \tag{5.85}$$

$$= \mathrm{Tr}\left(\sqrt{\sqrt{\rho_r}\left(\rho_r + \Delta_r\right)\sqrt{\rho_r}}\right)^2. \tag{5.86}$$

Therefore, the fidelity between ρ and $\rho + \Delta$ only depends on the range of ρ. Let us note that $\rho_r + \Delta_r$ is not a quantum state, but this will not influence our arguments. Defining $f(A) := \sqrt{A}$, we can write the fidelity as

$$F(\rho, \rho + \Delta) = \mathrm{Tr}\left[f(\rho_r^2 + \sqrt{\rho_r}\Delta_r\sqrt{\rho_r})\right]^2. \tag{5.87}$$

Now, we can approximate $F(\rho, \rho + \Delta)$ by performing the Taylor expansion of $f(A)$ around B,

$$f(A + B) = f(A) + \sum_{m=1}^{\infty} \frac{1}{m!} D^m F(A)([B]^m), \tag{5.88}$$

where $Df(A)(B)$ is the Directional Fréchet Derivative (2.67). In our case $A = \rho_r^2$ and $B = \sqrt{\rho_r}\Delta_r\sqrt{\rho_r}$. It should be noted that this expansion does not depend on any parameterization of density matrices. The first and second derivatives of f are (C)

$$\left[Df(A)(B)\right]_{ik} = \frac{B_{ik}}{\sqrt{a_i} + \sqrt{a_k}}, \tag{5.89}$$

$$\left[D^2 f(A)(B)(C)\right]_{ik} = -\frac{1}{\sqrt{a_i} + \sqrt{a_k}} \sum_{j=1}^{r} \left(\frac{B_{ij}}{\sqrt{a_i} + \sqrt{a_j}} \frac{C_{jk}}{\sqrt{a_j} + \sqrt{a_k}} + \frac{C_{ij}}{\sqrt{a_i} + \sqrt{a_j}} \frac{B_{jk}}{\sqrt{a_j} + \sqrt{a_k}} \right). \tag{5.90}$$

Thus, the approximation of the square root function up to second order is

$$\left[\sqrt{A + B}\right]_{ij} = \sqrt{a_i}\delta_{ij} + \frac{B_{ij}}{\sqrt{a_i} + \sqrt{a_j}} - \frac{1}{\sqrt{a_i} + \sqrt{a_j}} \sum_{k=1}^{r} \left(\frac{B_{ik}}{\sqrt{a_i} + \sqrt{a_k}} \frac{B_{kj}^*}{\sqrt{a_k} + \sqrt{a_j}} \right) + \mathcal{O}\left(||B||^3\right). \tag{5.91}$$

Taking the trace,

$$\mathrm{Tr}\left(\sqrt{A + B}\right) = \sum_{i=1}^{r} \left[\sqrt{A + B}\right]_{ii} \tag{5.92}$$

$$= \sum_{i=1}^{r} \left[\sqrt{a_i} + \frac{B_{ii}}{2\sqrt{a_i}} - \frac{1}{2\sqrt{a_i}} \sum_{k=1}^{r} \frac{|B_{ik}|^2}{(\sqrt{a_i}+\sqrt{a_k})^2} \right] + \mathcal{O}\big(||B||^3\big). \quad (5.93)$$

From $A = \rho_r^2$ and $B = \sqrt{\rho_r}\Delta_r\sqrt{\rho_r}$, we get

$$a_i = \lambda_i^2, \qquad B_{ij} = \sqrt{\lambda_i\lambda_j}\left\langle \lambda_i^r \middle| \Delta_r \middle| \lambda_j^r \right\rangle, \quad (5.94)$$

where $\{\lambda_i\}_{i=1,\dots,r}$ are eigenvalues of ρ_r with eigenvectors $\left\{\left|\lambda_j^r\right\rangle\right\}_{j=1,\dots,r}$. The matrix B can be rewritten in terms of the eigenvectors $\{|\lambda_j\rangle\}_{j=1,\dots,d}$ of ρ, where the first r vectors correspond to the non-null subspace of ρ,

$$B_{ij} = \sqrt{\lambda_i\lambda_j}\left\langle \lambda_i | \Delta | \lambda_j \right\rangle, \qquad i,j = 1,...,r. \quad (5.95)$$

Then,

$$\sqrt{F(\rho,\rho+\Delta)} = \sum_{i=1}^{r} \left[\lambda_i + \frac{\lambda_i \left\langle \lambda_i | \Delta | \lambda_i \right\rangle}{2\lambda_i} - \frac{1}{2\lambda_i} \sum_{k=1}^{r} \frac{\lambda_i\lambda_k}{(\lambda_i+\lambda_k)^2} |\left\langle \lambda_i | \Delta | \lambda_k \right\rangle|^2 \right] + \mathcal{O}\big(||\Delta||^3\big) \quad (5.96)$$

$$= \sum_{i=1}^{r} \left[\lambda_i + \frac{1}{2}\left\langle \lambda_i | \Delta_r | \lambda_i \right\rangle - \frac{1}{2} \sum_{k=1}^{r} \frac{\lambda_k}{(\lambda_i+\lambda_k)^2} |\left\langle \lambda_i | \Delta | \lambda_k \right\rangle|^2 \right] + \mathcal{O}\big(||\Delta||^3\big) \quad (5.97)$$

$$= 1 + \frac{1}{2}\sum_{i=1}^{r} \left\langle \lambda_i | \Delta | \lambda_i \right\rangle - \frac{1}{2} \sum_{i,k=1}^{r} \frac{\lambda_k}{(\lambda_i+\lambda_k)^2} |\left\langle \lambda_i | \Delta | \lambda_k \right\rangle|^2 + \mathcal{O}\big(||\Delta||^3\big). \quad (5.98)$$

The first order term can be rewritten using $\mathrm{Tr}(\Delta) = 0$,

$$\sum_{i=1}^{r} \left\langle \lambda_i | \Delta | \lambda_i \right\rangle = \mathrm{Tr}(\Delta) - \sum_{i=r+1}^{d} \left\langle \lambda_i | \Delta | \lambda_i \right\rangle = -\sum_{i=r+1}^{d} \left\langle \lambda_i | \Delta | \lambda_i \right\rangle. \quad (5.99)$$

The second order term can be rewritten because Δ is hermitian,

$$\sum_{i,k=1}^{r} \frac{\lambda_k}{(\lambda_i+\lambda_k)^2} |\left\langle \lambda_i | \Delta | \lambda_k \right\rangle|^2 \quad (5.100)$$

$$= \frac{1}{2}\sum_{i,k=1}^{r} \left[\frac{\lambda_k}{(\lambda_i+\lambda_k)^2} |\left\langle \lambda_i | \Delta | \lambda_k \right\rangle|^2 + \frac{\lambda_k}{(\lambda_i+\lambda_k)^2} |\left\langle \lambda_i | \Delta | \lambda_k \right\rangle|^2 \right] \quad (5.101)$$

$$= \frac{1}{2}\sum_{i,k=1}^{r} \left[\frac{\lambda_k}{(\lambda_i+\lambda_k)^2} |\left\langle \lambda_i | \Delta | \lambda_k \right\rangle|^2 + \frac{\lambda_i}{(\lambda_k+\lambda_i)^2} |\left\langle \lambda_k | \Delta^\dagger | \lambda_i \right\rangle|^2 \right] \quad (5.102)$$

$$= \frac{1}{2}\sum_{i,k=1}^{r}\left[\frac{\lambda_k}{(\lambda_i+\lambda_k)^2}|\langle\lambda_i|\Delta|\lambda_k\rangle|^2 + \frac{\lambda_i}{(\lambda_k+\lambda_i)^2}|\langle\lambda_i|\Delta|\lambda_k\rangle|^2\right] \tag{5.103}$$

$$= \frac{1}{2}\sum_{i,k=1}^{r}\frac{\lambda_k+\lambda_i}{(\lambda_i+\lambda_k)^2}|\langle\lambda_i|\Delta|\lambda_k\rangle|^2 \tag{5.104}$$

$$= \frac{1}{2}\sum_{i,k=1}^{r}\frac{|\langle\lambda_i|\Delta|\lambda_k\rangle|^2}{\lambda_i+\lambda_k}. \tag{5.105}$$

Thus, the fidelity up to second order becomes

$$\sqrt{F(\rho,\rho+\Delta)} = 1 - \frac{1}{2}\sum_{i=r+1}^{d}\langle\lambda_i|\Delta|\lambda_i\rangle - \frac{1}{4}\sum_{i,k=1}^{r}\frac{|\langle\lambda_i|\Delta|\lambda_k\rangle|^2}{\lambda_i+\lambda_k} + \mathcal{O}(||\Delta||^3), \tag{5.106}$$

and the infidelity between ρ and $\rho+\Delta$ is

$$\text{Inf}(\rho,\rho+\Delta) = 1 - F(\rho,\rho+\Delta) \tag{5.107}$$

$$= 1 - \left[1 - \frac{1}{2}\sum_{i=r+1}^{d}\langle\lambda_i|\Delta|\lambda_i\rangle - \frac{1}{4}\sum_{i,k=1}^{r}\frac{|\langle\lambda_i|\Delta|\lambda_k\rangle|^2}{\lambda_i+\lambda_k} + \mathcal{O}(||\Delta||^3)\right]^2 \tag{5.108}$$

$$= \sum_{i=r+1}^{d}\langle\lambda_i|\Delta|\lambda_i\rangle + \frac{1}{2}\sum_{i,k=1}^{r}\frac{|\langle\lambda_i|\Delta|\lambda_k\rangle|^2}{\lambda_i+\lambda_k} + \mathcal{O}(||\Delta||^2). \tag{5.109}$$

For a low rank state, the linear term in the expansion of the infidelity dominates,

$$\text{Inf}(\rho,\rho+\Delta) = \sum_{i=r+1}^{d}\langle\lambda_i|\Delta|\lambda_i\rangle + \mathcal{O}(||\Delta||^2). \tag{5.110}$$

Otherwise, for a full rank density matrix, the first order term vanishes because ρ does not have a null subspace. In this case, the approximation becomes up to second order [31]

$$\text{Inf}(\rho,\rho+\Delta) = \frac{1}{2}\sum_{i,k=1}^{d}\frac{|\langle\lambda_i|\Delta|\lambda_k\rangle|^2}{\lambda_i+\lambda_k} + \mathcal{O}(||\Delta||^3). \tag{5.111}$$

In summary, the infidelity is linear in $||\Delta||$ for low rank states and quadratic in $||\Delta||$ for full rank mixed states. This has as consequence that the tomography (5.76) of a low-rank state scales as $||\Delta|| \sim \mathcal{O}(1/\sqrt{N})$, while the tomography of a full-rank state scales $||\Delta||^2 \sim \mathcal{O}(1/N)$.

5.2.3 Bounds of Quantum Tomography

We now calculate the quantum Fisher Information matrix of a single qudit spanned by means of the Gell-Mann matrices (4.25), that is,

$$\rho = \frac{1}{d}\mathbb{I} + \frac{1}{2}\sum_{i=1}^{d^2-1} S_i\sigma_i. \tag{5.112}$$

First, we must obtain the symmetric logarithmic derivatives (3.88)

$$\frac{\partial\rho}{\partial S_i} = \frac{1}{2}\Big(\rho L_i + L_i\rho\Big) \tag{5.113}$$

$$\frac{1}{2}\sigma_i = \frac{1}{2}\Big(\rho L_i + L_i\rho\Big) \tag{5.114}$$

$$\sigma_i = \rho L_i + L_i\rho. \tag{5.115}$$

Projecting in the basis of eigenvectors $\{|\lambda_i\rangle\}_{i=1,\dots,d}$ of ρ,

$$\langle\lambda_j|\sigma_i|\lambda_k\rangle = \langle\lambda_j|\rho L_i|\lambda_k\rangle + \langle\lambda_j|L_i\rho|\lambda_k\rangle \tag{5.116}$$

$$\langle\lambda_j|\sigma_i|\lambda_k\rangle = \Big(\lambda_j + \lambda_k\Big)\langle\lambda_j|L_i|\lambda_k\rangle\,. \tag{5.117}$$

Supposing $\lambda_i \neq 0$ for all $i = 1, \dots, d$,

$$\langle\lambda_j|L_i|\lambda_k\rangle = \frac{1}{\lambda_j + \lambda_k}\langle\lambda_j|\sigma_i|\lambda_k\rangle\,. \tag{5.118}$$

Then, the quantum Fisher information matrix (3.106) is

$$J_{ij} = \frac{1}{2}\,\mathrm{Tr}\left(\rho\Big[L_iL_j + L_jL_i\Big]\right) \tag{5.119}$$

$$= \frac{1}{2}\sum_{k=1}^{d}\langle\lambda_k|\left(\rho\Big[L_iL_j + L_jL_i\Big]\right)|\lambda_k\rangle \tag{5.120}$$

$$= \frac{1}{2}\sum_{k=1}^{d}\lambda_k\langle\lambda_k|\Big[L_iL_j + L_jL_i\Big]|\lambda_k\rangle \tag{5.121}$$

$$= \frac{1}{2}\sum_{k,l=1}^{d}\lambda_k\Big[\langle\lambda_k|L_i|\lambda_l\rangle\langle\lambda_l|L_j|\lambda_k\rangle + \langle\lambda_k|L_j|\lambda_l\rangle\langle\lambda_l|L_i|\lambda_k\rangle\Big] \tag{5.122}$$

$$= \frac{1}{2}\sum_{k,l=1}^{d}\frac{\lambda_k}{(\lambda_k+\lambda_l)^2}\Big[\langle\lambda_k|\sigma_i|\lambda_l\rangle\langle\lambda_l|\sigma_j|\lambda_k\rangle + \langle\lambda_k|\sigma_j|\lambda_l\rangle\langle\lambda_l|\sigma_i|\lambda_k\rangle\Big] \tag{5.123}$$

$$= \frac{1}{2} \sum_{k,l=1}^{d} \left[\frac{\lambda_k}{(\lambda_k + \lambda_l)^2} \langle \lambda_k | \sigma_i | \lambda_l \rangle \langle \lambda_l | \sigma_j | \lambda_k \rangle + \frac{\lambda_l}{(\lambda_l + \lambda_k)^2} \langle \lambda_l | \sigma_j | \lambda_k \rangle \langle \lambda_k | \sigma_i | \lambda_l \rangle \right] \tag{5.124}$$

$$= \frac{1}{2} \sum_{k,l=1}^{d} \frac{\lambda_k + \lambda_l}{(\lambda_k + \lambda_l)^2} \langle \lambda_k | \sigma_i | \lambda_l \rangle \langle \lambda_l | \sigma_j | \lambda_k \rangle \tag{5.125}$$

$$= \frac{1}{2} \sum_{k,l=1}^{d} \frac{1}{\lambda_k + \lambda_l} \langle \lambda_k | \sigma_i | \lambda_l \rangle \langle \lambda_l | \sigma_j | \lambda_k \rangle . \tag{5.126}$$

Let us note that this is only true for full rank density matrices. Let us consider the infidelity for full rank density matrices (5.109),

$$\mathrm{Inf}(\rho, \rho + \Delta) = \frac{1}{2} \sum_{i,k=1}^{d} \frac{| \langle \lambda_i | \Delta | \lambda_k \rangle |^2}{\lambda_i + \lambda_k} + \mathcal{O}\big(||\Delta||^3\big). \tag{5.127}$$

For standard tomography, the error matrix is given by (5.71)

$$\Delta = \frac{1}{2} \sum_{i=1}^{d^2-1} \big(\hat{S}_i - S_i\big) \sigma_i. \tag{5.128}$$

Then,

$$\mathrm{Inf}(\rho, \rho + \Delta) = \frac{1}{2} \sum_{i,k=1}^{d} \frac{\langle \lambda_i | \Delta | \lambda_k \rangle \langle \lambda_k | \Delta | \lambda_i \rangle}{\lambda_i + \lambda_k} \tag{5.129}$$

$$= \frac{1}{8} \sum_{i,k=1}^{d} \sum_{l,m=1}^{d^2-1} \big(\hat{S}_l - S_l\big)\big(\hat{S}_m - S_m\big) \frac{\langle \lambda_i | \sigma_l | \lambda_k \rangle \langle \lambda_k | \sigma_m | \lambda_i \rangle}{\lambda_i + \lambda_k} \tag{5.130}$$

$$= \frac{1}{4} \sum_{l,m=1}^{d^2-1} \big(\hat{S}_l - S_l\big)\big(\hat{S}_m - S_m\big) \frac{1}{2} \sum_{i,k=1}^{d} \frac{\langle \lambda_i | \sigma_l | \lambda_k \rangle \langle \lambda_k | \sigma_m | \lambda_i \rangle}{\lambda_i + \lambda_k}. \tag{5.131}$$

Taking the expected value,

$$\mathbb{E}\Big[\mathrm{Inf}(\rho, \rho + \Delta)\Big] = \frac{1}{4} \sum_{l,m=1}^{d^2-1} \mathbb{E}\Big[\big(\hat{S}_l - S_l\big)\big(\hat{S}_m - S_m\big)\Big] \frac{1}{2} \sum_{i,k=1}^{d} \frac{\langle \lambda_i | \sigma_l | \lambda_k \rangle \langle \lambda_k | \sigma_m | \lambda_i \rangle}{\lambda_i + \lambda_k} \tag{5.132}$$

$$= \frac{1}{4} \mathrm{Tr}\Big(CJ\Big). \tag{5.133}$$

where C is the covariance matrix. The last expression relates infidelity to the covariance matrix, which allows us to find the quantum Cramer-Rao bound for infidelity. Then employing $C \geq J^{-1}/N$, where N is size of sample, we get

$$\mathbb{E}\Big[\mathrm{Inf}(\rho, \rho+\Delta)\Big] \geq \frac{1}{4N}\,\mathrm{Tr}\left(J^{-1}J\right) = \frac{1}{4N}\,\mathrm{Tr}(\mathbb{I}) = \frac{d^2-1}{4N}. \tag{5.134}$$

The lower bound $\mathrm{CR} = (d^2-1)/4N$ is the Cramer-Rao bound of the infidelity, which describes the optimal measurement among all measurement strategies (separable or collective). Therefore, with a sample N we can distinguish ρ from $\hat{\rho} = \rho + \Delta$ with an accuracy of $\mathcal{O}(1/N)$ for the infidelity.

Besides, replacing $W = J$ in (3.118) we obtain

$$\begin{aligned}
\mathbb{E}\Big[\mathrm{Inf}(\rho, \rho+\Delta)\Big] \geq & \frac{1}{4N}\frac{1}{d-1}\left(\mathrm{Tr}\left[\sqrt{\sqrt{J}J^{-1}\sqrt{J}}\right]\right)^2 && (5.135)\\
= & \frac{1}{4N}\frac{1}{d-1}\Big(\mathrm{Tr}\,\mathbb{I}\Big)^2 && (5.136)\\
= & \frac{1}{4N}\frac{1}{d-1}\big(d^2-1\big)^2 && (5.137)\\
= & \frac{1}{4N}\big(d^2-1\big)\big(d+1\big), && (5.138)
\end{aligned}$$

where $\mathrm{GM} = (d^2-1)(d+1)/4N$ is the Gill-Massar lower bound of the infidelity, which is the optimal bound for separable measurements. Clearly,

$$\mathbb{E}\Big[\mathrm{Inf}(\rho, \rho+\Delta)\Big] \geq \mathrm{GM} > \mathrm{CR}. \tag{5.139}$$

5.3 Qudit Adaptive Tomography

As we saw, the accuracy of quantum tomography depends on the state itself and the basis on which we perform the tomography. To minimize the infidelity

$$\mathrm{Inf}(\rho, \rho+\Delta) = \sum_{i=r+1}^{d} \langle\lambda_i|\Delta|\lambda_i\rangle + \frac{1}{2}\sum_{i,k=1}^{r} \frac{|\,\langle\lambda_i|\Delta|\lambda_k\rangle\,|^2}{\lambda_i+\lambda_k} + \mathcal{O}\big(||\Delta||^2\big), \tag{5.140}$$

we must accurately estimate the eigenvectors of small eigenvalues of ρ to cancel the first term. We will prove that this can be achieved by performing a standard

tomography on the basis of eigenvectors of ρ, that is, by measuring the following observables

$$\sigma_{ij}^{x} = |\lambda_i\rangle\langle\lambda_j| + |\lambda_j\rangle\langle\lambda_i|\,, \tag{5.141}$$

$$\sigma_{ij}^{y} = -\,\mathrm{i}\Big(\,|\lambda_i\rangle\langle\lambda_j| - |\lambda_j\rangle\langle\lambda_i|\,\Big), \tag{5.142}$$

$$\sigma_{k}^{z} = \sqrt{\frac{2}{k(k+1)}}\left(\sum_{j=1}^{k}|\lambda_j\rangle\langle\lambda_j| - k\,|\lambda_{k+1}\rangle\langle\lambda_{k+1}|\right). \tag{5.143}$$

The first term of the infidelity only depends on diagonal Gell-Mann matrices,

$$\sum_{i=r+1}^{d}\langle\lambda_i|\Delta|\lambda_i\rangle = \sum_{j=1}^{d^2-1}\sum_{i=r+1}^{d}\left(S_i - \hat{S}_i\right)\langle\lambda_i|\sigma_j|\lambda_i\rangle \tag{5.144}$$

$$= \sum_{j=1}^{d-1}\sum_{i=r+1}^{d}\left(S_i^z - \hat{S}_i^z\right)\langle\lambda_i|\sigma_j^z|\lambda_i\rangle \tag{5.145}$$

$$= \sum_{j=1}^{r-1}\sum_{i=r+1}^{d}\left(S_i^z - \hat{S}_i^z\right)\langle\lambda_i|\sigma_j^z|\lambda_i\rangle + \sum_{j=r}^{d-1}\sum_{i=r+1}^{d}\left(S_i^z - \hat{S}_i^z\right)\langle\lambda_i|\sigma_j^z|\lambda_i\rangle\,. \tag{5.146}$$

The first term at the right hand side vanishes. The diagonal matrices entering in this term are lineal combinations of projectors onto the eigenstates of ρ, that is the set $\{|\lambda_i\rangle_{i=1,\ldots,r}\}$,

$$\langle\lambda_i|\sigma_j^z|\lambda_i\rangle = \sqrt{\frac{2}{j(j+1)}}\,\langle\lambda_i|\left(\sum_{k=1}^{j}|\lambda_k\rangle\langle\lambda_k| - j\,|\lambda_{j+1}\rangle\langle\lambda_{j+1}|\right)|\lambda_i\rangle \tag{5.147}$$

$$= \sqrt{\frac{2}{j(j+1)}}\left(\sum_{k=1}^{j}\delta_{ik} - j\delta_{i,j+1}\right) \tag{5.148}$$

$$= 0. \tag{5.149}$$

The second term is also canceled because this contains the difference $S_i^z - \hat{S}_i^z$, which is of the order of magnitude of the standard deviation (2.113,5.76),

$$\left(S_i^z - \hat{S}_i^z\right)^2 \sim \begin{cases} \dfrac{1}{N}\left[\dfrac{2}{i(i+1)}\left(\displaystyle\sum_{n=1}^{i}\lambda_n + i^2\lambda_{i+1}\right) - (S_i^z)^2\right], & i < r. \\ 0, & i \geq r. \end{cases} \tag{5.150}$$

Thus, the first order term of infidelity is completely canceled by measuring in the base of eigenvectors,

$$\sum_{i=r+1}^{d} \langle\lambda_i|\Delta|\lambda_i\rangle \sim 0. \tag{5.151}$$

Therefore, this tomography improves the quality of the infidelity, reducing the distance between ρ and its estimation from $\mathcal{O}(||\Delta||)$ to $\mathcal{O}(||\Delta||^2)$ for all states,

$$\mathrm{Inf}(\rho, \rho+\Delta) = \frac{1}{2}\sum_{i,k=1}^{r} \frac{|\langle\lambda_i|\Delta|\lambda_k\rangle|^2}{\lambda_i+\lambda_k} + \mathcal{O}\big(||\Delta||^3\big). \tag{5.152}$$

Similarly to (5.133),

$$\mathbb{E}\Big[\mathrm{Inf}(\rho, \rho+\Delta)\Big] = \frac{1}{4}\sum_{l,m=1}^{d^2-1} \mathbb{E}\Big[\big(S_l - \hat{S}_l\big)\big(S_m - \hat{S}_m\big)\Big] \frac{1}{2}\sum_{i,k=1}^{r} \frac{\langle\lambda_i|\sigma_l|\lambda_k\rangle\ \langle\lambda_k|\sigma_m|\lambda_i\rangle}{\lambda_i+\lambda_k}. \tag{5.153}$$

Defining

$$\mathcal{J}_{ij} = \frac{1}{2}\sum_{k,l=1}^{r} \frac{1}{\lambda_k+\lambda_l} \langle\lambda_k|\sigma_i|\lambda_l\rangle\ \langle\lambda_l|\sigma_j|\lambda_k\rangle\,, \tag{5.154}$$

we can write the mean infidelity as

$$\mathbb{E}\Big[\mathrm{Inf}(\rho, \rho+\Delta)\Big] = \frac{1}{4}\,\mathrm{Tr}\Big(C\mathcal{J}\Big). \tag{5.155}$$

where $\mathcal{J}$ is the contribution of the non-null range of ρ to the quantum Fisher information matrix.

This tomography can not be done because the state ρ is unknown. Despite this, we can perform the following algorithm:

1. We first obtain an estimate of the basis of eigenvectors of ρ. This can be generated by means of a preliminary estimate $\hat{\rho}_0$ obtained by standard tomography using a fraction N_0 of the sample. It is not necessary that this estimation be with MLE, but rather the linear estimation (4.88) is enough, however the estimation with MLE should present better performance. In the worst case, the accuracy of this estimate is $\mathcal{O}(1/\sqrt{N_0})$. Let $\left\{\hat{\lambda}_i^0\right\}_{i=1,\ldots,d}$ and $\left\{|\hat{\lambda}_i^0\rangle\right\}_{i=1,\ldots,d}$ the

eigenvalues and eigenvectors of the estimate $\hat{\rho}_0$. We can write the estimated eigenvalues as $\hat{\lambda}_i^0 = \lambda_i + \delta_i$, where $\{\delta_i\}$ are the errors, and the estimated eigenvectors as a rotation of $\{|\lambda\rangle_i\}_{i=1,\dots,d}$ with the unitary matrix $U = e^{-\mathrm{i}\Omega_0}$, where Ω_0 is hermitian and traceless,

$$|\hat{\lambda}_i^0\rangle = U\,|\lambda_i\rangle\,. \tag{5.156}$$

Then, the estimated density matrix up to first order is

$$\hat{\rho}_0 = \sum_{i=1}^{d} \hat{\lambda}_i \left|\hat{\lambda}_i\right\rangle\!\left\langle\hat{\lambda}_i\right| \tag{5.157}$$

$$= \sum_{i=1}^{d} \hat{\lambda}_i U^\dagger\,|\lambda_i\rangle\langle\lambda_i|\,U \tag{5.158}$$

$$= \sum_{i=1}^{d} (\lambda_i + \delta_i)\big(|\lambda_i\rangle\langle\lambda_i| + \mathrm{i}\big[\Omega_0, |\lambda_i\rangle\langle\lambda_i|\,\big] + \mathcal{O}(||\Omega_0||^2)\big) \tag{5.159}$$

$$\begin{aligned}=&\rho + \sum_i \Big(\delta_i\,|\lambda_i\rangle\langle\lambda_i| + \mathrm{i}\lambda_i\big[\Omega_0, |\lambda_i\rangle\langle\lambda_i|\,\big]\Big)\\ &+ \sum_i \mathrm{i}\delta_i\big[\Omega_0, |\lambda_i\rangle\langle\lambda_i|\,\big] + \mathcal{O}(||\Omega_0||^2).\end{aligned} \tag{5.160}$$

Given that the error $\Delta_0 = \rho - \hat{\rho}_0$ in the estimated density matrix in the worst case scales as $\mathcal{O}(1/\sqrt{N_0})$ (5.76), we have that $\delta_i \sim \mathcal{O}(1/\sqrt{N})$ or $||\Omega_0|| \sim \mathcal{O}(1/\sqrt{N_0})$. Since only the eigenvectors will intervene in the protocol, we suppose that

$$||\Omega_0|| \sim \mathcal{O}\left(\frac{1}{\sqrt{N_0}}\right). \tag{5.161}$$

Thereby, using the Baker-Campbell-Hausdorff formula (2.61)

$$\left|\left\langle\lambda_i\middle|\hat{\lambda}_j^0\right\rangle\right|^2 = |\,\langle\lambda_i|U|\lambda_j\rangle|^2 \tag{5.162}$$

$$= \langle\lambda_j|U^\dagger\,|\lambda_i\rangle\langle\lambda_i|\,U|\lambda_j\rangle \tag{5.163}$$

$$\begin{aligned}= \langle\lambda_j|\Big(&|\lambda_i\rangle\langle\lambda_i| + \big[\mathrm{i}\Omega_0, |\lambda_i\rangle\langle\lambda_i|\,\big] + \frac{1}{2}\Big[\mathrm{i}\Omega_0, \big[\mathrm{i}\Omega_0, |\lambda_i\rangle\langle\lambda_i|\,\big]\Big]\Big)|\lambda_j\rangle\\ &+ \mathcal{O}(||\Omega_0||^3)\end{aligned} \tag{5.164}$$

$$= \Omega_{ij} + \mathrm{i}\,\langle\lambda_j|\big[\Omega_0, |\lambda_i\rangle\langle\lambda_i|\,\big]|\lambda_j\rangle - \frac{1}{2}\,\langle\lambda_j|\Big[\Omega_0, \big[\Omega_0, |\lambda_i\rangle\langle\lambda_i|\,\big]\Big]|\lambda_j\rangle$$

$$+\mathcal{O}(||\Omega_0||^3) \tag{5.165}$$
$$=\Omega_{ij}-\frac{1}{2}\langle\lambda_j|\Big[\Omega_0,\big[\Omega_0,|\lambda_i\rangle\langle\lambda_i|\,\big]\Big]|\lambda_j\rangle+\mathcal{O}(||\Omega_0||^3) \tag{5.166}$$
$$=\Omega_{ij}\Big(1-\langle\lambda_i|\Omega_0^2|\lambda_i\rangle\Big)+|\langle\lambda_j|\Omega_0|\lambda_i\rangle|^2+\mathcal{O}(||\Omega_0||^3) \tag{5.167}$$
$$=\Omega_{ij}-\mathcal{O}\big(||\Omega_0||^2\big). \tag{5.168}$$

Analogously, the probability of projecting the unknown state ρ onto the eigenstates of ρ_0 is given by

$$p_i^0=\Big\langle\hat{\lambda}_i^0\Big|\rho\Big|\hat{\lambda}_i^0\Big\rangle \tag{5.169}$$
$$=\langle\lambda_i|U^\dagger\rho U|\lambda_i\rangle \tag{5.170}$$
$$=\langle\lambda_i|\bigg(\rho+\Big[\mathrm{i}\Omega_0,\rho\Big]+\frac{1}{2}\Big[\mathrm{i}\Omega_0,\big[\mathrm{i}\Omega_0,\rho\big]\Big]\bigg)|\lambda_i\rangle+\mathcal{O}(||\Omega_0||^3) \tag{5.171}$$
$$=\lambda_i+\mathrm{i}\,\langle\lambda_i|\Big[\Omega_0,\rho\Big]|\lambda_i\rangle-\frac{1}{2}\langle\lambda_i|\Big[\Omega_0,\big[\Omega_0,\rho\big]\Big]|\lambda_i\rangle+\mathcal{O}(||\Omega_0||^3) \tag{5.172}$$
$$=\lambda_i-\frac{1}{2}\langle\lambda_i|\Big[\Omega_0,\big[\Omega_0,\rho\big]\Big]|\lambda_i\rangle+\mathcal{O}(||\Omega_0||^3) \tag{5.173}$$
$$=\lambda_i-\mathcal{O}\big(||\Omega_0||^2\big). \tag{5.174}$$

2. Thereafter, we perform a second standard tomography with the remaining sample $N-N_0$, writing the Gell-Mann operators on the basis of eigenvectors of $\hat{\rho}_0$. The basis of eigenvectors must be ordered decreasingly according to their respective eigenvalues. In this case, the linear term of the infidelity is not canceled (5.146).

   For the first sum of (5.146), we have

$$\Big\langle\hat{\lambda}_i^0\Big|\sigma_j^z\Big|\hat{\lambda}_i^0\Big\rangle=\sqrt{\frac{2}{j(j+1)}}\Big\langle\hat{\lambda}_i\Big|\left(\sum_{k=1}^{j}|\lambda_k^0\rangle\langle\lambda_k^0|-j\,|\lambda_{j+1}\rangle\langle\lambda_{j+1}|\right)\Big|\hat{\lambda}_i\Big\rangle \tag{5.175}$$
$$=\sqrt{\frac{2}{j(j+1)}}\left(\sum_{k=1}^{j}\Big|\Big\langle\hat{\lambda}_i^0\Big|\lambda_k\Big\rangle\Big|^2-j\Big|\Big\langle\hat{\lambda}_i^0\Big|\lambda_{j+1}\Big\rangle\Big|^2\right) \tag{5.176}$$
$$\sim\mathcal{O}\Big(||\Omega_0||^2\Big) \tag{5.177}$$
$$\sim\mathcal{O}\left(\frac{1}{N_0}\right). \tag{5.178}$$

For the second sum of (5.146), for $k \geq r$, we have

$$(S_k^z)^2 = \frac{2}{k(k+1)}\left(\sum_{j=1}^{k} p_j^0 - k p_{k+1}^0\right)^2 = \frac{2}{k(k+1)} + \mathcal{O}(||\Omega_0||^2). \tag{5.179}$$

Thereby

$$(S_k^z - \hat{S}_k^z)^2 \sim \frac{1}{N-N_0}\left[\frac{2}{k(k+1)}\left(\sum_{n=1}^{k}\lambda_n + k^2\lambda_{k+1}\right) - (S_k^z)^2\right] \tag{5.180}$$

$$\sim \frac{1}{N-N_0}\mathcal{O}\big(||\Omega_0||^2\big) \tag{5.181}$$

$$\sim \mathcal{O}\left(\frac{1}{(N-N_0)N_0}\right). \tag{5.182}$$

Therefore, the worst accuracy of the second estimate is

$$\sum_{i=r+1}^{d}\langle\lambda_i|\Delta|\lambda_i\rangle \sim \mathcal{O}\left(\frac{1}{\sqrt{N-N_0}N_0}\right) + \mathcal{O}\left(\frac{1}{\sqrt{(N-N_0)N_0}}\right) \tag{5.183}$$

$$\sim \mathcal{O}\left(\frac{1}{\sqrt{(N-N_0)N_0}}\right), \tag{5.184}$$

which is similar to the scaling of the second order term in the Taylor series. The choice of N_0 plays a key role in the performance of the two-stage adaptive quantum tomography. In case of qubit has been proposed to choose the preliminary ensemble as a fraction or a power of the total sample size N, that is $N_0 = N/a$ or $N_0 = N^b$ with $b \geq 2/3$. The first choice leads to an infidelity that scales as $\mathcal{O}(\sqrt{a(1-a)}/N)$, while with the second choice the infidelity scales as $\mathcal{O}(1/N^{(1+b)/2})$. For example, if we use $N_0 = N/2$ and $N_0 = N^{2/3}$ as the initial sample

$$N_0 = N/2 \Longrightarrow \sum_{i=r+1}^{d}\langle\lambda_i|\Delta|\lambda_i\rangle \sim \mathcal{O}\left(\frac{2}{N}\right), \tag{5.185}$$

$$N_0 = N^{2/3} \Longrightarrow \sum_{i=r+1}^{d}\langle\lambda_i|\Delta|\lambda_i\rangle \sim \mathcal{O}\left(\frac{1}{N^{5/6}}\right). \tag{5.186}$$

Finally, we must obtain the estimate by MLE. This optimization can be done with the measurements of both steps. However, for a large resource it is sufficient to use the measurements of the second step.

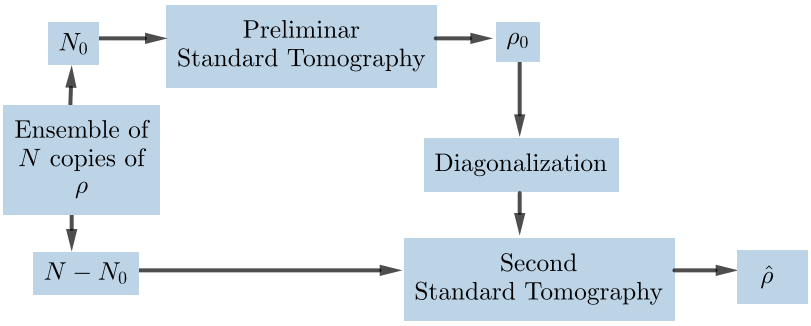

Figure 5.4: Diagram of the two-stages adaptive standard tomography. Source: Made by the author.

Moreover, we can define an adaptive quantum tomography which reduces another figure of merit, as the mean squared error $\mathrm{Tr}(WC)$. Let J_0 the quantum Fisher information matrix of ρ_0 on its basis of eigenvectors (D). Then, the classical Fisher information matrix becomes,

$$I_0 = (d-1)\sqrt{J_0}\frac{\sqrt{J_0^{-1/2}WJ_0^{-1/2}}}{\mathrm{Tr}\left(\sqrt{J_0^{-1/2}WJ_0^{-1/2}}\right)}\sqrt{J_0}. \tag{5.187}$$

Let suppose that I_0 has eigenvalues $\{a_i\}_{i=1,..,d^2-1}$ with eigenvectors $\{\boldsymbol{r}_i\}_{i=1,..,d^2-1}$. Then, the Gill-Massar bound can be attain measuring the observables $\left\{\boldsymbol{r}_i^{\mathrm{T}}\boldsymbol{\sigma}\right\}_{i=1,...,d^2-1}$ with probabilities

$$p_i = \frac{1}{d-1}\boldsymbol{r}_i^{\mathrm{T}}\left[I_0J_0^{-1}\right]\boldsymbol{r}_i = \frac{a_i}{d-1}\left[\boldsymbol{r}_i^{\mathrm{T}}J_0^{-1}\boldsymbol{r}_i\right], \tag{5.188}$$

where $\boldsymbol{\sigma} = \begin{pmatrix}\sigma_1 & \cdots & \sigma_{d^2-1}\end{pmatrix}^{\mathrm{T}}$. If $[J_0, W] = 0$, we obtain

$$I_0 = (d-1)\frac{\sqrt{WJ_0}}{\mathrm{Tr}\left(\sqrt{WJ_0^{-1}}\right)}. \tag{5.189}$$

Let us suppose $\{b_i\}_{i=1,...,d^2-1}$ and $\{w_i\}_{i=1,...,d^2-1}$ are the eigenvalues of J_0 and W, respectively. We get

$$p_i = \frac{\boldsymbol{r}_i^{\mathrm{T}}\sqrt{WJ_0^{-1}}\boldsymbol{r}_i}{\mathrm{Tr}\left[\sqrt{WJ_0^{-1}}\right]} = \frac{\sqrt{w_ib_i^{-1}}}{\sum_{i=1}^{d^2-1}\sqrt{w_ib_i^{-1}}}. \tag{5.190}$$

We can write these probabilities more explicitly considering that,

$$J_0 = \begin{pmatrix} J_0^x & \Theta & \Theta \\ \Theta & J_0^y & \Theta \\ \Theta & \Theta & J_0^z \end{pmatrix}, \qquad W = \begin{pmatrix} W^x & \Theta & \Theta \\ \Theta & W^y & \Theta \\ \Theta & \Theta & W^z \end{pmatrix}, \tag{5.191}$$

where (D)

$$[J_0^x]_{ij,i'j'} = \frac{1}{\lambda_i + \lambda_j}\delta_{ii'}\delta_{jj'}, \tag{5.192}$$

$$[J_0^y]_{ij,i'j'} = \frac{1}{\lambda_i + \lambda_j}\delta_{ii'}\delta_{jj'}, \tag{5.193}$$

$$[J_0^z]_{k,k'} = \frac{1}{2\sqrt{kk'(k+1)(k'+1)}}\left(\sum_{l=1}^{k}\frac{1}{\lambda_l} - \frac{k}{\lambda_{k+1}}h_{k'\geq k+1} + \frac{kk'}{\lambda_{k+1}}\delta_{kk'}\right). \tag{5.194}$$

Since J_0^x and J_0^y are diagonals, their elements are eigenvalues of J_0,

$$b_{ij}^x = b_{ij}^y = \frac{1}{\lambda_i + \lambda_j}. \tag{5.195}$$

On the other hand, J_0^z is not diagonal and it cannot be explicitly diagonalized. Let us suppose $\{\boldsymbol{r}_k^z\}_{k=1,\dots,d}$ are the eigenvectors of J_0^z with eigenvalues $\{b_k^z\}_{k=1,\dots,d-1}$. Then, the optimal measurements are $\left\{\sigma_{ij}^x, \sigma_{ij}^y, \boldsymbol{r}_k^{z\mathrm{T}}\boldsymbol{\sigma}^z\right\}$ with probabilities

$$p_{ij}^x = \frac{\sqrt{w_{ij}^x(\lambda_i+\lambda_j)}}{\sum_{1\leq i<j\leq d}\sqrt{\lambda_i+\lambda_j}\left[\sqrt{w_{ij}^x}+\sqrt{w_{ij}^y}\right] + \sum_{k=1}^{d-1} w_k^z[b_k^z]^{-1}}, \tag{5.196}$$

$$p_{ij}^y = \frac{\sqrt{w_{ij}^y(\lambda_i+\lambda_j)}}{\sum_{1\leq i<j\leq d}\sqrt{\lambda_i+\lambda_j}\left[\sqrt{w_{ij}^x}+\sqrt{w_{ij}^y}\right] + \sum_{k=1}^{d-1} w_k^z[b_k^z]^{-1}}, \tag{5.197}$$

$$p_k^z = \frac{\sqrt{w_{ij}^z[b_k^z]^{-1}}}{\sum_{1\leq i<j\leq d}\sqrt{\lambda_i+\lambda_j}\left[\sqrt{w_{ij}^x}+\sqrt{w_{ij}^y}\right] + \sum_{k=1}^{d-1} w_k^z[b_k^z]^{-1}}, \tag{5.198}$$

where $1 \leq i < j \leq d$ and $k = 1, ..., d-1$.

We will study some particular choices of W:

- If $W = \mathbb{I}$, the minimized function is the mean square error (2.154). The optimal observables are $\left\{\sigma_{ij}^x, \sigma_{ij}^y, \boldsymbol{r}_k^{z\mathrm{T}}\boldsymbol{\sigma}^z\right\}$ with probabilities

$$p_{ij}^x = p_{ij}^y = \frac{\sqrt{\lambda_i+\lambda_j}}{2\sum_{1\leq i<j\leq d}\sqrt{\lambda_i+\lambda_j} + \sum_{k=1}^{d-1}[b_k^z]^{-1}}, \tag{5.199}$$

$$p_k^z = \frac{\sqrt{[b_k^z]^{-1}}}{2\sum_{1\leq i<j\leq d}\sqrt{\lambda_i+\lambda_j}+\sum_{k=1}^{d-1}[b_k^z]^{-1}}. \tag{5.200}$$

Despite this technique decreases the mean squared error and does not decrease the infidelity, we will study its effects on infidelity.

- If $W = J$, the minimized function is the second order infidelity (5.133),

$$\mathrm{Tr}(JC) = 4\mathrm{Inf}(\rho,\hat{\rho}). \tag{5.201}$$

The optimal observables are $\left\{\sigma_{ij}^x, \sigma_{ij}^y, \sigma_k^z\right\}$ with probabilities

$$p_i = \frac{1}{d^2-1}. \tag{5.202}$$

5.4 Simulations

We perform computational simulations of adaptive quantum tomography in order to study its accuracy. We compute and compare the mean infidelity between a state and its estimate by standard tomography, diagonal tomography, adaptive tomography with $N_0 = N/2$ and $N_0 = N^{2/3}$, and weighted adaptive tomography $W = \mathbb{I}$ with $N_0 = N/2$. The measurement outcomes have been simulated by multinomial distribution (4.79), obtaining estimated probabilities $\left\{\hat{p}_i^j\right\}_{i=1,\ldots,d}$ when the matrix σ_k has been measurement. Thereby, the estimated expected value of the Gell-Mann matrix σ_k is

$$\hat{S}_k = \sum_{i=1}^{d}\lambda_i^k\hat{p}_i^k, \tag{5.203}$$

where $\{\lambda_i^k\}_{i=1,\ldots,d}$ are the eigenvalues of σ_k (4.21,4.22,4.23). The estimation by maximum likelihood was only performed in the second step of the tomographies. The mean infidelity on the state space is obtained by (E)

$$\int_{\mathcal{H}}\mathrm{Inf}(\rho,\hat{\rho})\mathrm{d}\mu(\rho) \approx \frac{1}{M}\sum_{i=1}^{M}\mathrm{Inf}(\rho_i,\hat{\rho}_i), \tag{5.204}$$

where $\{\rho_i\}_{i=1,\ldots,M}$ are M uniformly distributed random states. We analyze the mean infidelity of states in different dimensions and with different ranks. Besides, we fit the function $\mathrm{Inf}(N,\alpha,\beta) = \beta/N^\alpha$ to the data by least squares.

5.4.1 Pure states

Figures (5.5) to (5.10) compare the mean infidelity with respect to the sample size N for pure states in different dimensions. The dots are the simulated tomographies and the dot-dashed lines are the best fit by least squares. Tables (5.4) to (5.9) show the fit coefficients. It can be seen that with the same amount of resources adaptive tomography and weighted adaptive tomography achieve better mean infidelity than standard quantum tomography. Moreover, adaptive tomography and weighted adaptive tomography scale better with the sample size N than standard quantum tomography because they have a higher fit coefficient α. Adaptive tomography with $N_0 = N/2$ achieves the best infidelity for pure states in all dimensions. Diagonal tomography delivers mean infidelity equal to or less than the Cramer-Rao bound, because the Gill-Massar bound for the estimation of a pure state (which only has $2[d-1]$ independent real parameters) is

$$\mathrm{GM}_{pure} = \frac{d-1}{N} \leq \mathrm{CR} \leq \mathrm{GM}. \tag{5.205}$$

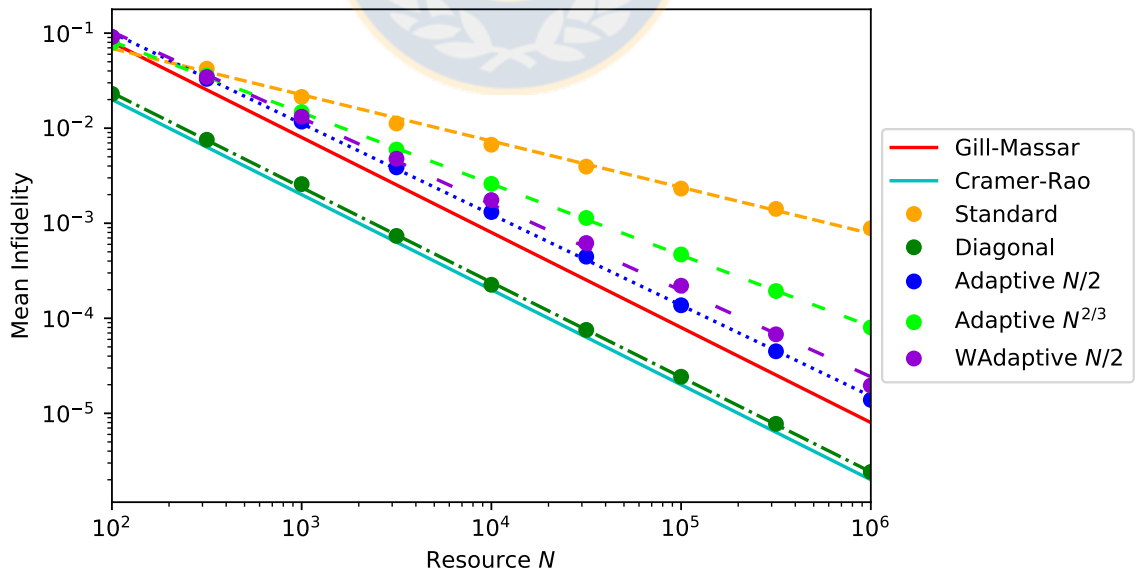

Figure 5.5: Comparison of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 1000 3-dimensional pure states. Source: Made by the author.

	α	β
Gill-Massar	1	8.0
Cramer-Rao	1	2.0
Standard	0.486840610328	0.648924437556
Diagonal	0.997295586425	2.33049882612
Adaptive $N/2$	0.955865495569	8.2410556636
Adaptive $N^{2/3}$	0.749563017847	2.58419396358
WAdaptive $N/2$	0.907667688012	6.76682644859

Table 5.4: Comparison of fit coefficients of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 1000 3-dimensional pure states. Source: Made by the author.

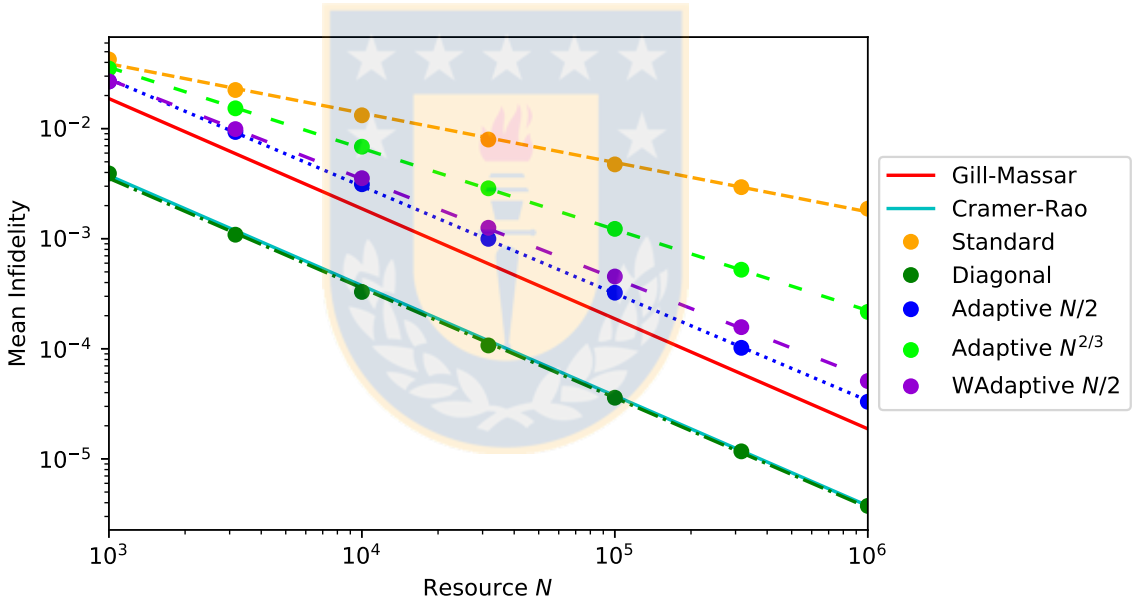

Figure 5.6: Comparison of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 1000 4-dimensional pure states. Source: Made by the author.

	α	β
Gill-Massar	1	18.75
Cramer-Rao	1	3.75
Standard	0.448386208637	0.8606839038
Diagonal	0.996933431113	3.46605333551
Adaptive $N/2$	0.974034326237	23.6266047634
Adaptive $N^{2/3}$	0.736461721101	5.84541370537
WAdaptive $N/2$	0.904034355563	14.4229945003

Table 5.5: Comparison of fit coefficients of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 1000 4-dimensional pure states. Source: Made by the author.

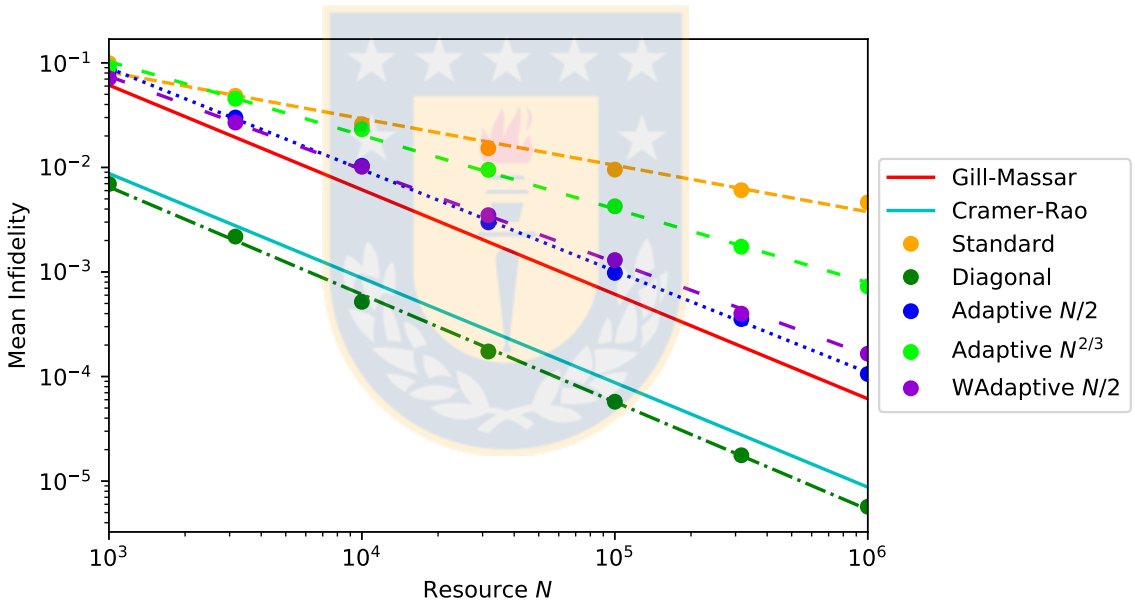

Figure 5.7: Comparison of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 100 6-dimensional pure states. Source: Made by the author.

	α	β
Gill-Massar	1	61.25
Cramer-Rao	1	8.75
Standard	0.445524616668	1.77256685365
Diagonal	1.02862952388	7.93763170438
Adaptive $N/2$	0.970478935162	72.667504549
Adaptive $N^{2/3}$	0.704983162844	13.4376970486
WAdaptive $N/2$	0.88929748672	34.6595173618

Table 5.6: Comparison of fit coefficients of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 100 6-dimensional pure states. Source: Made by the author.

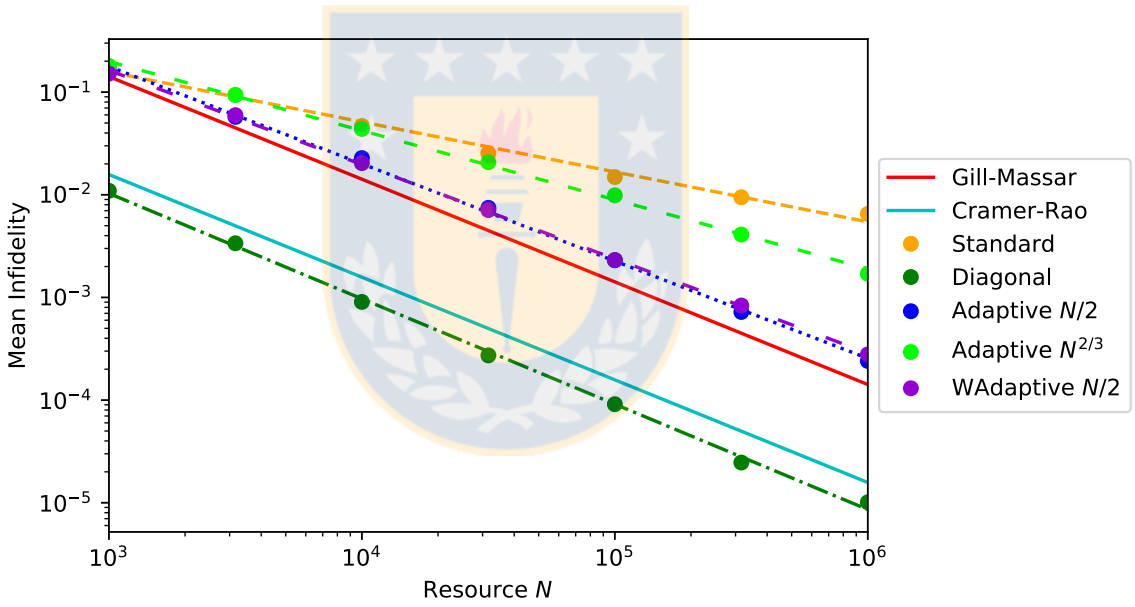

Figure 5.8: Comparison of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 100 8-dimensional pure states. Source: Made by the author.

	α	β
Gill-Massar	1	141.75
Cramer-Rao	1	15.75
Standard	0.487009852719	4.5493865281
Diagonal	1.0263897093	12.3663703431
Adaptive $N/2$	0.947713881087	123.795827206
Adaptive $N^{2/3}$	0.673339697624	20.819804269
WAdaptive $N/2$	0.918303133516	92.8045975921

Table 5.7: Comparison of fit coefficients of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 100 8-dimensional pure states. Source: Made by the author.

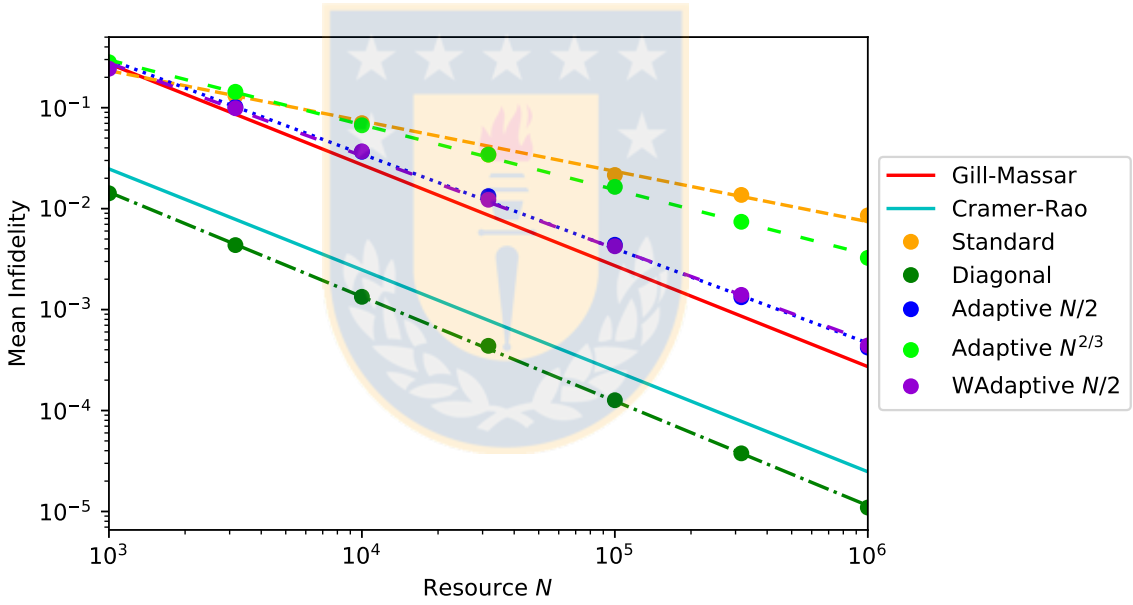

Figure 5.9: Comparison of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 100 10-dimensional pure states. Source: Made by the author.

	α	β
Gill-Massar	1	272.25
Cramer-Rao	1	24.75
Standard	0.498608354403	7.29716124083
Diagonal	1.03492218404	18.5328804977
Adaptive $N/2$	0.93630391787	193.445155351
Adaptive $N^{2/3}$	0.642792892212	25.2454127239
WAdaptive $N/2$	0.918908650875	158.452446585

Table 5.8: Comparison of fit coefficients of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 100 10-dimensional pure states. Source: Made by the author.

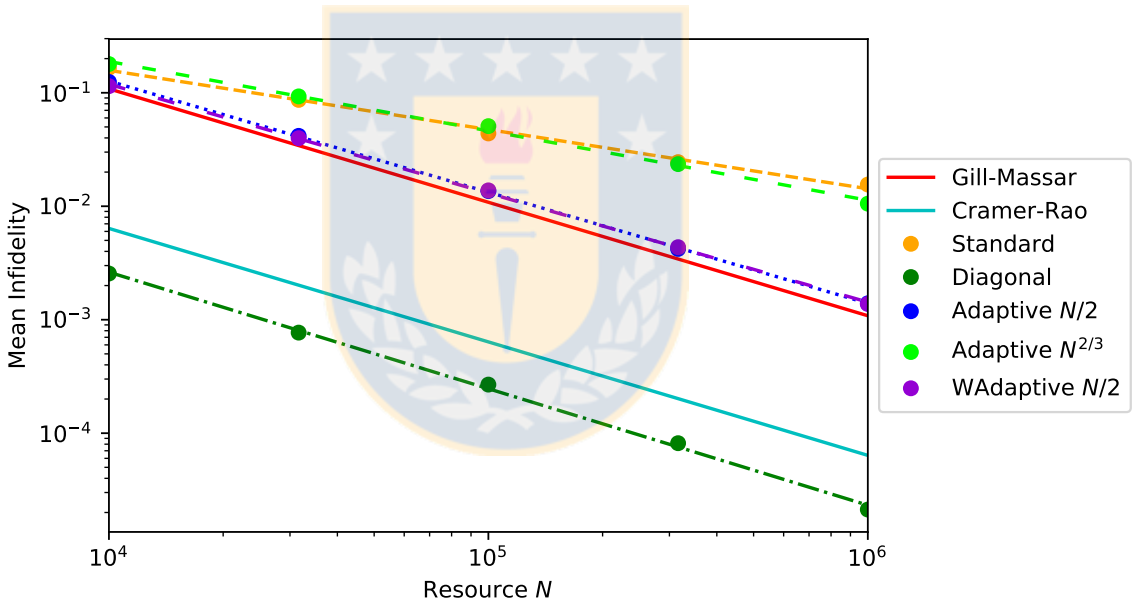

Figure 5.10: Comparison of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 100 16-dimensional pure states. Source: Made by the author.

	α	β
Gill-Massar	1	1083.75
Cramer-Rao	1	63.75
Standard	0.521728796368	19.2589060538
Diagonal	1.02687772882	33.6100975019
Adaptive $N/2$	0.979562162464	1050.388139
Adaptive $N^{2/3}$	0.6115039318	52.6933278002
WAdaptive $N/2$	0.960106416413	823.5077895

Table 5.9: Comparison of fit coefficients of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 100 16-dimensional pure states. Source: Made by the author.

5.4.2 Low rank states

Figures (5.11) to (5.15) compare the mean infidelity with respect to the sample size N for low rank states in different dimensions. The dots are the simulated tomographies and the dot-dashed lines are the best fit by least squares. Tables (5.10) to (5.14) show the fit coefficients. It can be seen that with the same amount of resources adaptive tomography and weighted adaptive tomography achieve lower mean infidelity than standard quantum tomography. Moreover, adaptive tomography and weighted adaptive tomography scale better with the sample size N than standard quantum tomography because they have higher fit coefficient α. Nevertheless, this fit coefficient is lower than the fit coefficient for pure states, that is the protocol reduces its performance by increasing the rank of the state. Adaptive tomography with $N_0 = N/2$ achieves the best mean infidelity for low rank states in all dimensions. Adaptive tomography does not reach the Gill-Masar bound for low rank states.

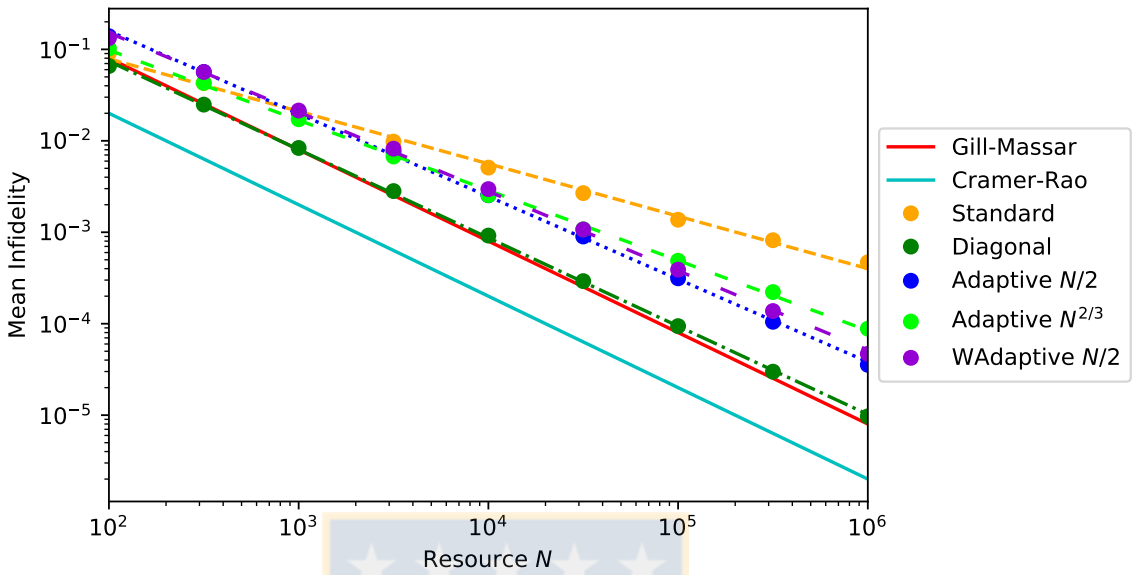

Figure 5.11: Comparison of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 1000 3-dimensional states with rank 2. Source: Made by the author.

	α	β
Gill-Massar	1	8.0
Cramer-Rao	1	2.0
Standard	0.5735058611	1.1044278579
Diagonal	0.965589328922	6.3288165699
Adaptive $N/2$	0.904625777431	10.2298582157
Adaptive $N^{2/3}$	0.76693327354	3.3590855469
WAdaptive $N/2$	0.867208473038	8.23804336571

Table 5.10: Comparison of fit coefficients of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 1000 3-dimensional states with rank 2. Source: Made by the author.

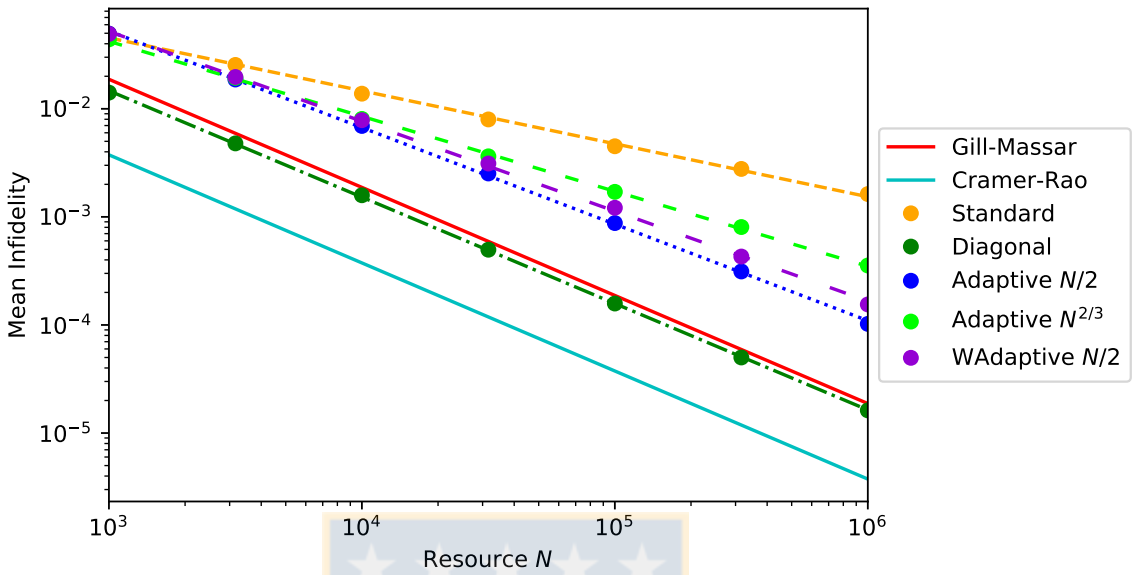

Figure 5.12: Comparison of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 1000 4-dimensional states with rank 2. Source: Made by the author.

	α	β
Gill-Massar	1	18.75
Cramer-Rao	1	3.75
Standard	0.490159011952	1.34052701646
Diagonal	0.984495791835	13.2059544052
Adaptive $N/2$	0.893368706158	25.1088552827
Adaptive $N^{2/3}$	0.693821250691	5.07688005339
WAdaptive $N/2$	0.831483384519	16.2403694326

Table 5.11: Comparison of fit coefficients of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 1000 4-dimensional states with rank 2. Source: Made by the author.

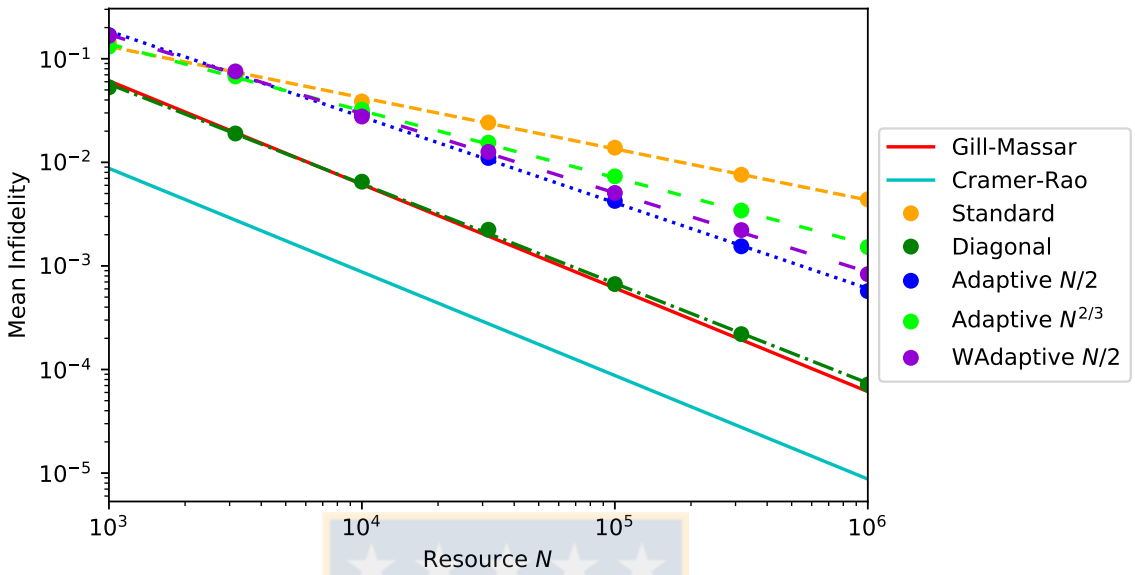

Figure 5.13: Comparison of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 100 6-dimensional states with rank 3. Source: Made by the author.

	α	β
Gill-Massar	1	61.25
Cramer-Rao	1	8.75
Standard	0.494166775132	3.98878822666
Diagonal	0.962419872982	43.9666089538
Adaptive $N/2$	0.828636047109	56.7238428081
Adaptive $N^{2/3}$	0.645557303655	12.0368688206
WAdaptive $N/2$	0.764583373286	33.7432853016

Table 5.12: Comparison of fit coefficients of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 100 6-dimensional states with rank 3. Source: Made by the author.

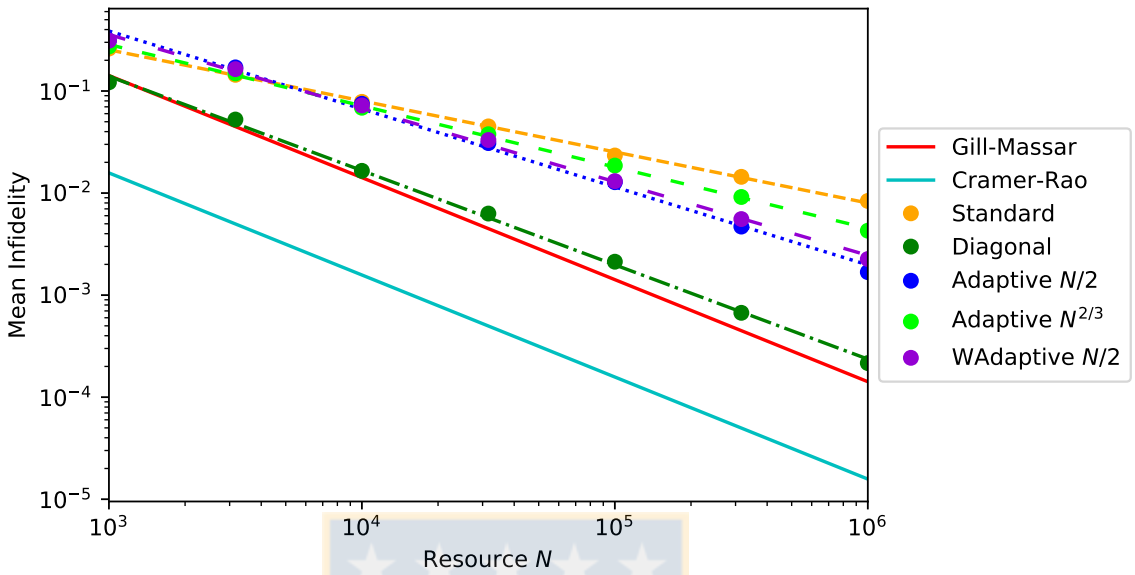

Figure 5.14: Comparison of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 100 8-dimensional states with rank 4. Source: Made by the author.

	α	β
Gill-Massar	1	141.75
Cramer-Rao	1	15.75
Standard	0.500377450872	8.03219811731
Diagonal	0.923805528625	82.3712503188
Adaptive $N/2$	0.763999993169	75.8414101667
Adaptive $N^{2/3}$	0.600350756358	18.0493252385
WAdaptive $N/2$	0.721262563098	52.1603730077

Table 5.13: Comparison of fit coefficients of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 100 8-dimensional states with rank 4. Source: Made by the author.

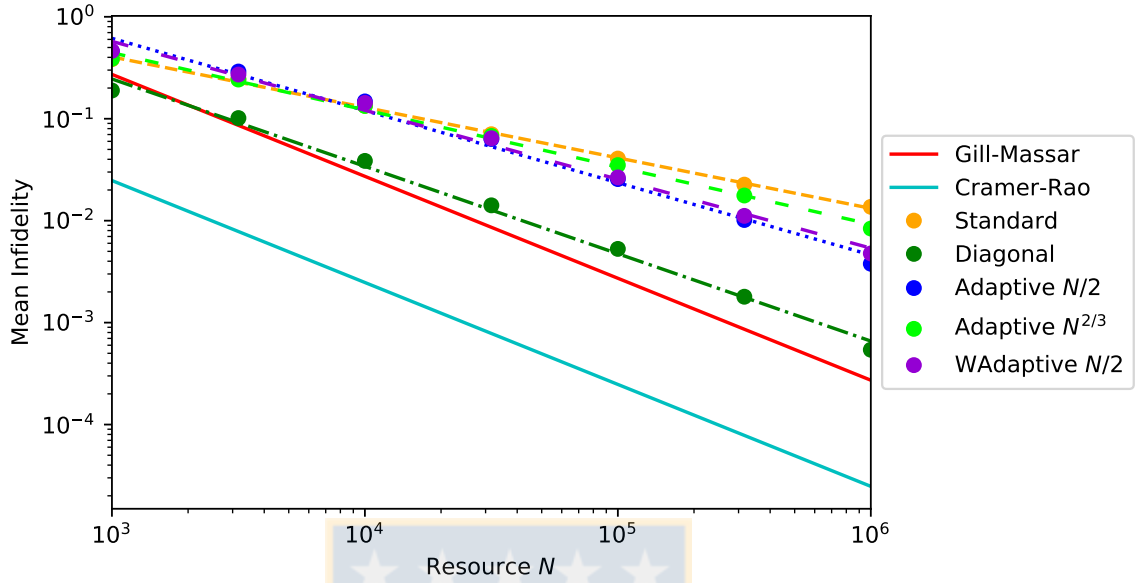

Figure 5.15: Comparison of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 100 10-dimensional states with rank 5. Source: Made by the author.

	α	β
Gill-Massar	1	272.25
Cramer-Rao	1	24.75
Standard	0.495213050342	12.3665995864
Diagonal	0.857016622493	91.4237327082
Adaptive $N/2$	0.708378540134	82.0456789328
Adaptive $N^{2/3}$	0.560884394015	21.3959430662
WAdaptive $N/2$	0.676096905836	61.2521742538

Table 5.14: Comparison of fit coefficients of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 100 10-dimensional states with rank 5. Source: Made by the author.

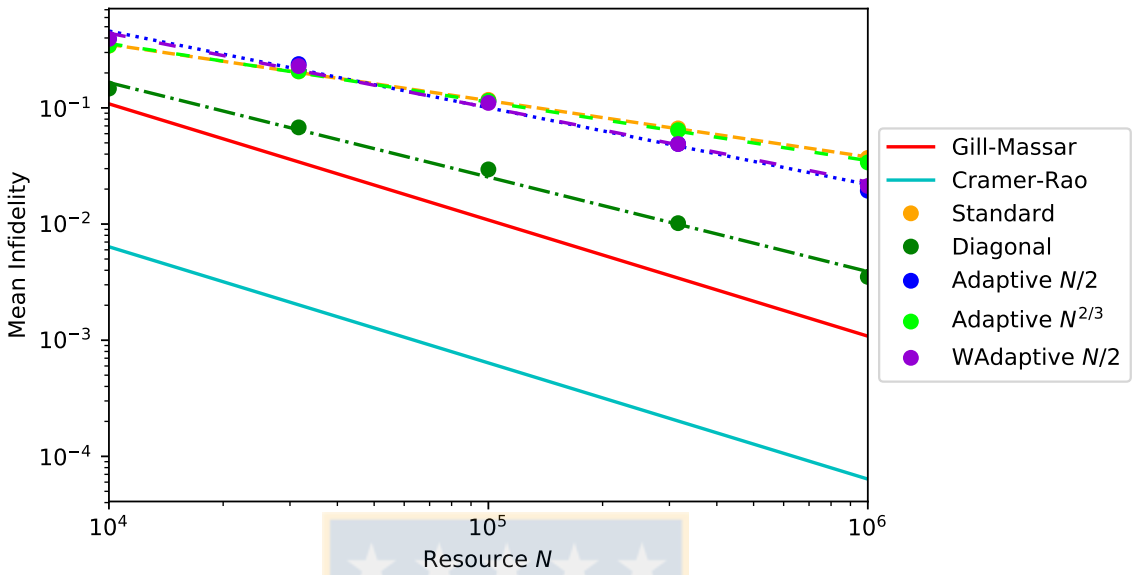

Figure 5.16: Comparison of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 100 16-dimensional states with rank 8. Source: Made by the author.

	α	β
Gill-Massar	1	1083.75
Cramer-Rao	1	63.75
Standard	0.48469739021	30.6180207697
Diagonal	0.813359936615	296.013394515
Adaptive $N/2$	0.660504196059	201.326947904
Adaptive $N^{2/3}$	0.504061815008	37.2065879176
WAdaptive $N/2$	0.640672964757	160.711830893

Table 5.15: Comparison of fit coefficients of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 1000 16-dimensional states with rank 8. Source: Made by the author.

5.4.3 Full rank states

Figures (5.17) to (5.21) compare the mean infidelity with respect to the sample size N for full rank states in different dimensions. The dots are the simulated tomographies and the dot-dashed lines are the best fit by least squares. Tables (5.16) to (5.20) show the fit coefficients. It can be seen that with the same amount of re-

sources adaptive tomography $N_0 = N/2$ and weighted adaptive tomography achieve similar mean infidelity than standard quantum tomography. Adaptive tomography $N_0 = N^{2/3}$ achieve lower mean infidelities than standard quantum tomography. Adaptive tomography does not reach the Gill-Massar bound for full rank states.

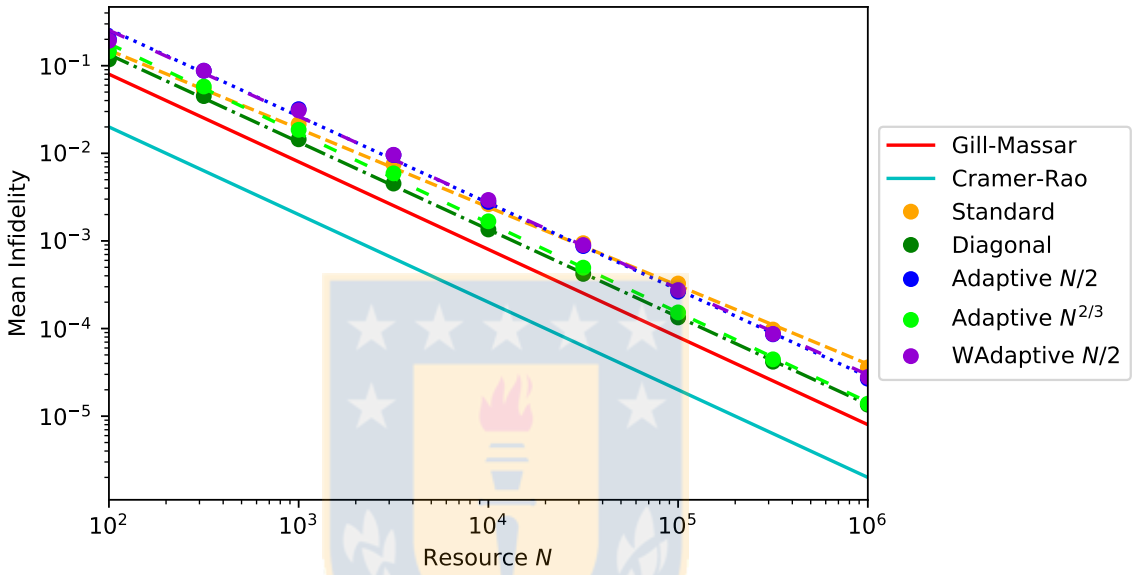

Figure 5.17: Comparison of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 1000 3-dimensional full rank states. Source: Made by the author.

	α	β
Gill-Massar	1	8.0
Cramer-Rao	1	2.0
Standard	0.895737828827	9.30606562924
Diagonal	0.99844569546	13.3146866104
Adaptive $N/2$	0.990872733639	24.9249150176
Adaptive $N^{2/3}$	1.02255040469	19.8135798498
WAdaptive $N/2$	0.983177160201	23.5067587495

Table 5.16: Comparison of fit coefficients of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 1000 3-dimensional full rank states. Source: Made by the author.

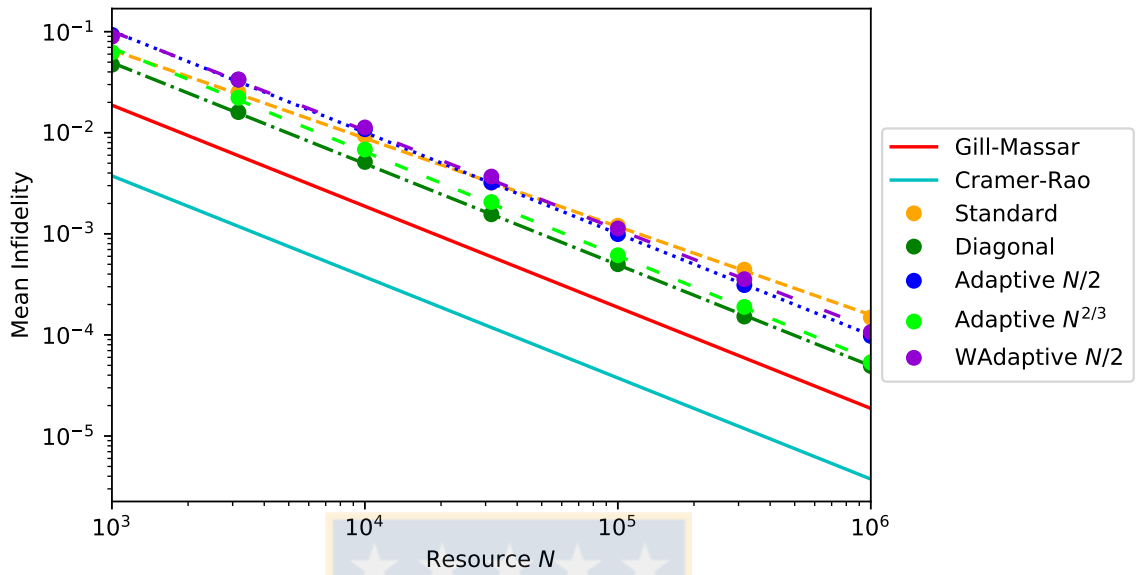

Figure 5.18: Comparison of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 1000 4-dimensional full rank states. Source: Made by the author.

	α	β
Gill-Massar	1	18.75
Cramer-Rao	1	3.75
Standard	0.874049329003	27.681258897
Diagonal	0.999840098803	49.352890542
Adaptive $N/2$	1.00269203039	103.489236591
Adaptive $N^{2/3}$	1.02730740327	83.3834150242
WAdaptive $N/2$	0.979969804744	87.6667655073

Table 5.17: Comparison of fit coefficients of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 1000 4-dimensional full rank states. Source: Made by the author.

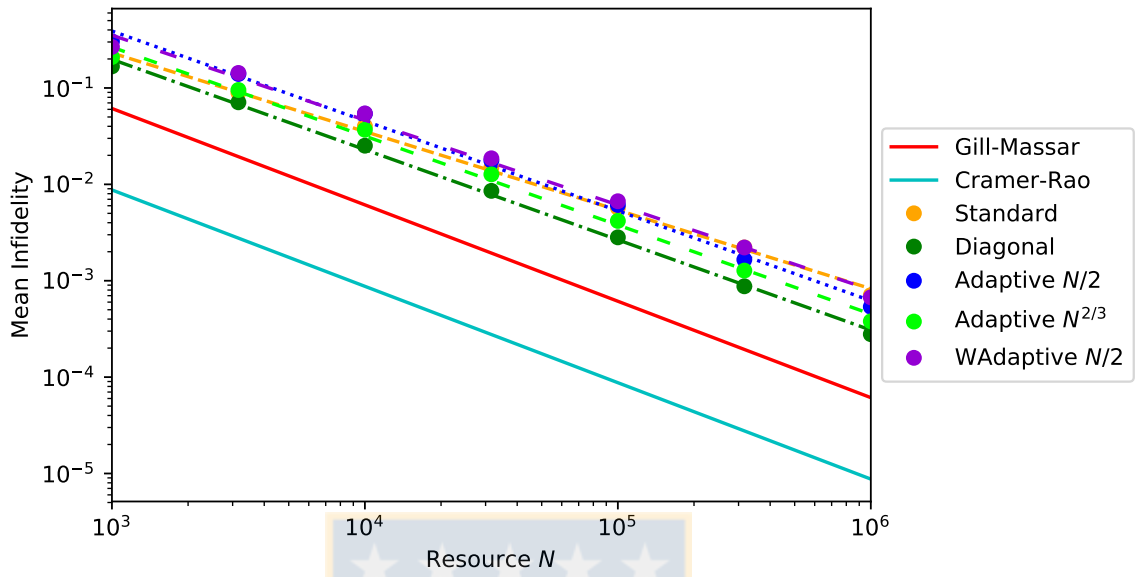

Figure 5.19: Comparison of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 100 6-dimensional full rank states. Source: Made by the author.

	α	β
Gill-Massar	1	61.25
Cramer-Rao	1	8.75
Standard	0.815134790472	64.3366062038
Diagonal	0.93617018919	127.409248003
Adaptive $N/2$	0.932765720494	245.03533196
Adaptive $N^{2/3}$	0.922941749331	155.81935465
WAdaptive $N/2$	0.880612280843	154.608851485

Table 5.18: Comparison of fit coefficients of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 100 6-dimensional full rank states. Source: Made by the author.

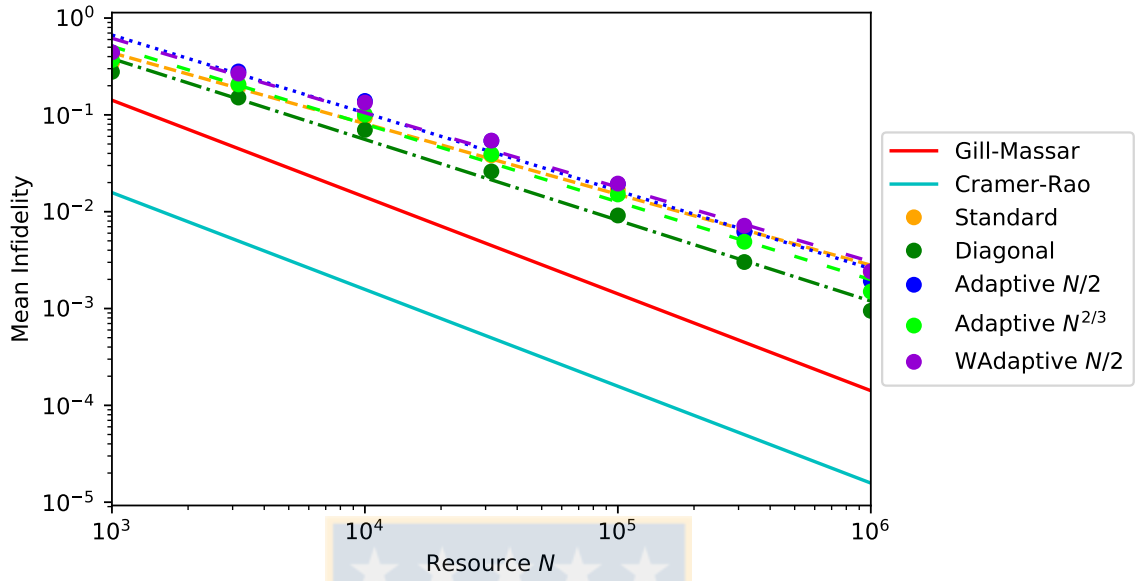

Figure 5.20: Comparison of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 100 8-dimensional full rank states. Source: Made by the author.

	α	β
Gill-Massar	1	141.75
Cramer-Rao	1	15.75
Standard	0.731163938682	68.3112146739
Diagonal	0.834762307376	121.71674307
Adaptive $N/2$	0.804268517365	172.798273961
Adaptive $N^{2/3}$	0.803363580332	131.237330839
WAdaptive $N/2$	0.768960257541	125.288470459

Table 5.19: Comparison of fit coefficients of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 100 8-dimensional full rank states. Source: Made by the author.

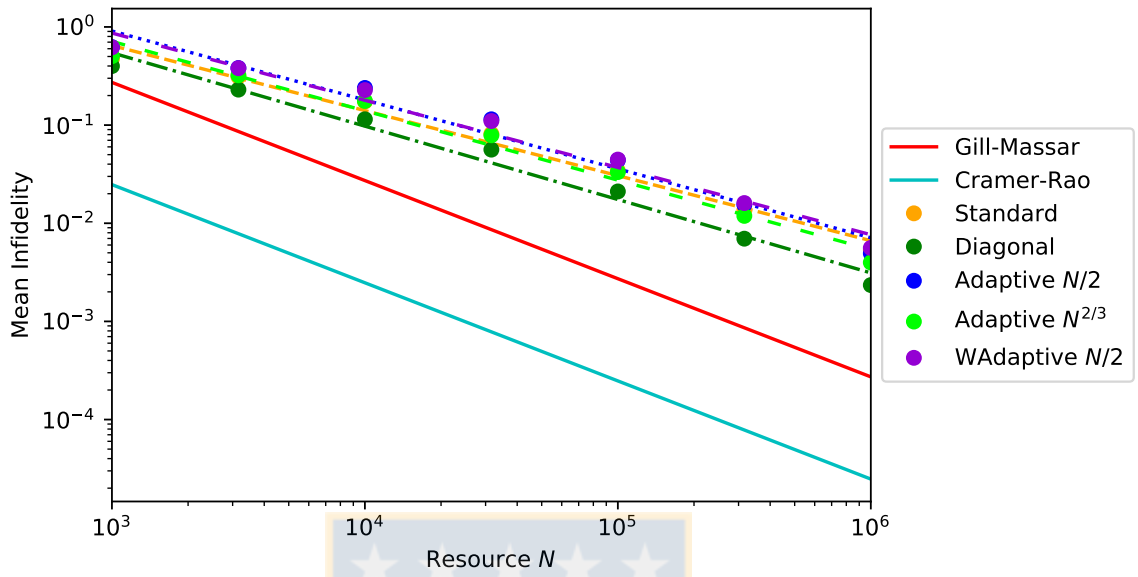

Figure 5.21: Comparison of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 100 10-dimensional full rank states. Source: Made by the author.

	α	β
Gill-Massar	1	272.25
Cramer-Rao	1	24.75
Standard	0.662719194324	62.9094012451
Diagonal	0.747587012886	95.4349333021
Adaptive $N/2$	0.702127721825	116.099265098
Adaptive $N^{2/3}$	0.706770996405	93.9709275126
WAdaptive $N/2$	0.684711846901	98.0978592067

Table 5.20: Comparison of fit coefficients of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 100 10-dimensional full rank states. Source: Made by the author.

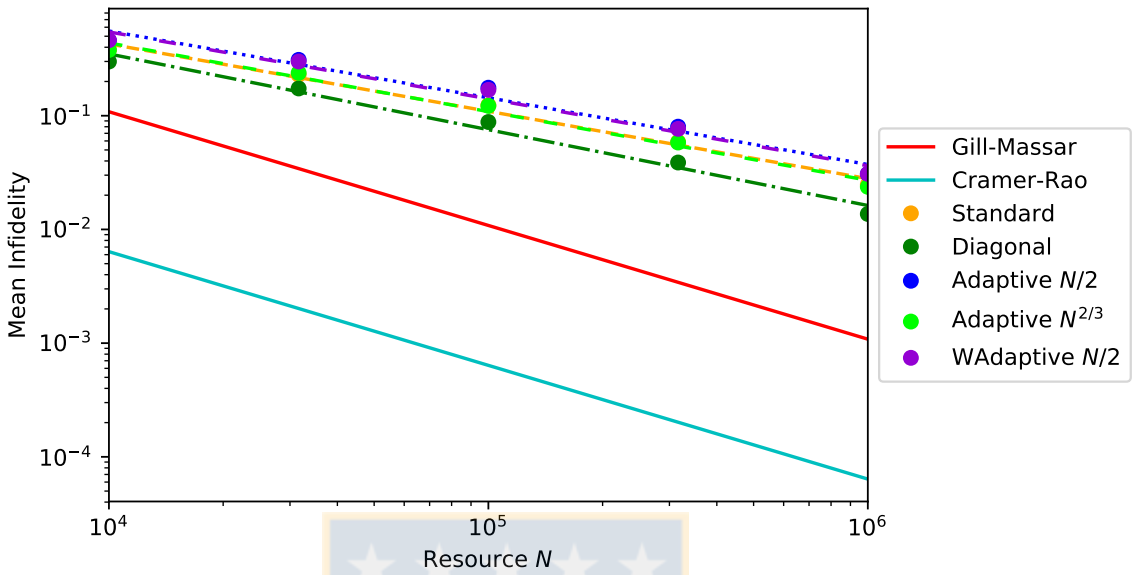

Figure 5.22: Comparison of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 100 16-dimensional full rank states. Source: Made by the author.

	α	β
Gill-Massar	1	1083.75
Cramer-Rao	1	63.75
Standard	0.59238121467	100.176478771
Diagonal	0.666535189765	162.087937747
Adaptive $N/2$	0.586882311134	123.700912096
Adaptive $N^{2/3}$	0.603959490009	113.761293979
WAdaptive $N/2$	0.588604619127	123.391942412

Table 5.21: Comparison of fit coefficients of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 1000 16-dimensional full rank states. Source: Made by the author.

5.4.4 Pure States with Noise

Figures (5.23) to (5.34) compare the mean infidelity with respect to the sample size N for pure states with noise in different dimensions. That is

$$\rho_i = \lambda \left|\psi_i\right\rangle\!\left\langle\psi_i\right| + \frac{1-\lambda}{d}\mathbb{I}, \qquad i = 1, ..., M. \tag{5.206}$$

where we have chosen $\lambda = 0.999$ and $\lambda = 0.99$. The purity of these states is,

$$\mathrm{Tr}\left(\rho_i^2\right) = \mathrm{Tr}\left[\left(\lambda\left|\psi_i\right\rangle\!\left\langle\psi_i\right| + \frac{1-\lambda}{d}\mathbb{I}\right)\left(\lambda\left|\psi_i\right\rangle\!\left\langle\psi_i\right| + \frac{1-\lambda}{d}\mathbb{I}\right)\right] \tag{5.207}$$

$$= \lambda^2 + 2\lambda\frac{1-\lambda}{d} + \left(\frac{1-\lambda}{d}\right)^2 d \tag{5.208}$$

$$= \lambda^2 + \frac{2\lambda - 2\lambda^2}{d} + \frac{1 - 2\lambda + \lambda^2}{d} \tag{5.209}$$

$$= \left(1 - \frac{1}{d}\right)\lambda^2 + \frac{1}{d}. \tag{5.210}$$

For $\lambda = 0.999$ and $\lambda = 0.99$, the purities of those states is near to 99.8% and 98%, respectively. This class of states is important due to the fact that it is usually employed to model errors in certain experiments. The dots are the simulated tomographies and the dot-dashed lines are the best fit by least squares. Tables (5.22) to (5.33) show the fit coefficients. It can be seen that with the same amount of resources adaptive tomography and weighted adaptive tomography achieve lower mean infidelity than standard quantum tomography. Moreover, adaptive tomography and weighted adaptive tomography scale better with the sample size N than standard quantum tomography because they have higher fit coefficient α. Adaptive tomography with $N_0 = N^{2/3}$ delivers the best infidelity for low dimensions but reduces its performance in high dimensions, where $N_0 = N/2$ lends to better mean infidelity. This is because, despite the purity decrease with de dimension, the unknown state has several eigenvalues near to zero. Adaptive tomography $N_0 = N/2$ and weighted adaptive tomography $N_0 = N/2$ achieve similar mean infidelity. The unusual behavior of diagonal tomography is due to that for low resource small eigenvalues are not distinguished, so the achieved mean infidelity is similar to a pure state. For a large enough sample, the mean infidelity achieved by diagonal tomography is similar to a full rank state.

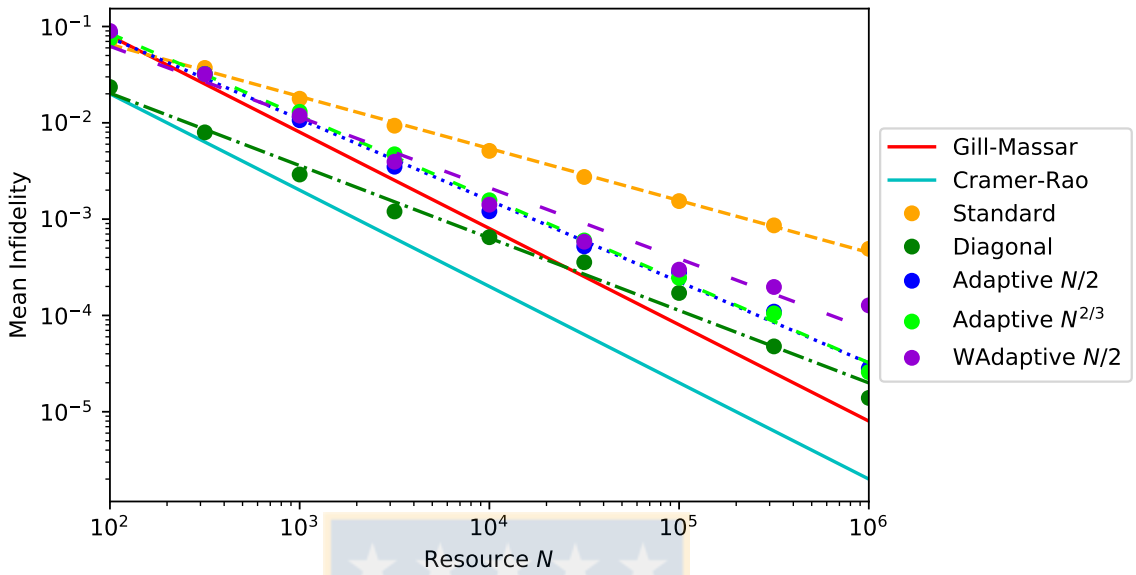

Figure 5.23: Comparison of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 1000 3-dimensional pure states with noise $\lambda = 0.999$. Source: Made by the author.

	α	β
Gill-Massar	1	8.0
Cramer-Rao	1	2.0
Standard	0.540221028399	0.784448215824
Diagonal	0.752108320951	0.649085586915
Adaptive $N/2$	0.844502901297	3.73755966892
Adaptive $N^{2/3}$	0.854791892538	4.34028742453
WAdaptive $N/2$	0.735244597942	1.84333061196

Table 5.22: Comparison of fit coefficients of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 1000 3-dimensional pure states with noise $\lambda = 0.999$. Source: Made by the author.

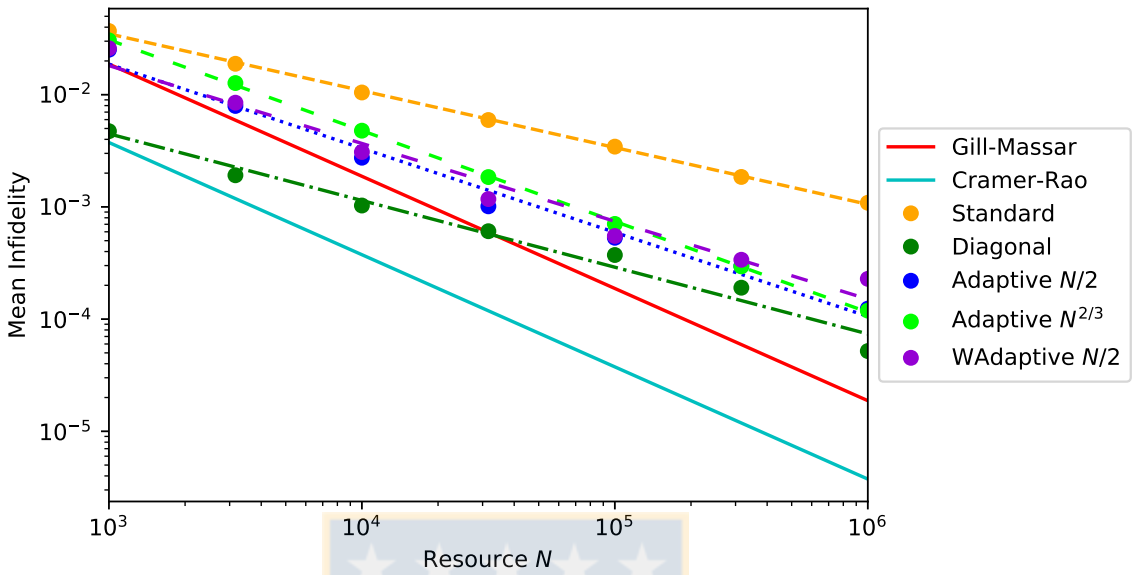

Figure 5.24: Comparison of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 1000 4-dimensional pure states with noise $\lambda = 0.999$. Source: Made by the author.

	α	β
Gill-Massar	1	18.75
Cramer-Rao	1	3.75
Standard	0.506600622327	1.15465550647
Diagonal	0.594617125513	0.272645581637
Adaptive $N/2$	0.749321456463	3.31198111333
Adaptive $N^{2/3}$	0.809036870001	8.24295347788
WAdaptive $N/2$	0.693853919176	2.19293300341

Table 5.23: Comparison of fit coefficients of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 1000 4-dimensional pure states with noise $\lambda = 0.999$. Source: Made by the author.

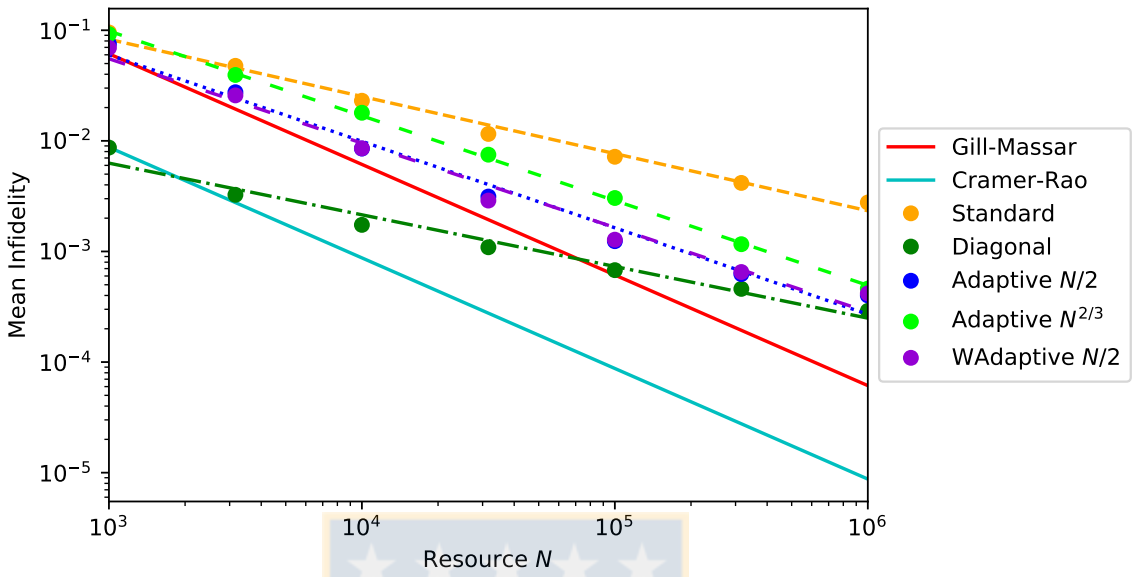

Figure 5.25: Comparison of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 100 6-dimensional pure states with noise $\lambda = 0.999$. Source: Made by the author.

	α	β
Gill-Massar	1	61.25
Cramer-Rao	1	8.75
Standard	0.516797507705	2.94261402836
Diagonal	0.467189789879	0.158564714912
Adaptive $N/2$	0.782064781967	13.3079856921
Adaptive $N^{2/3}$	0.766570859865	19.6234857519
WAdaptive $N/2$	0.762897162451	10.7414705412

Table 5.24: Comparison of fit coefficients of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 100 6-dimensional pure states with noise $\lambda = 0.999$. Source: Made by the author.

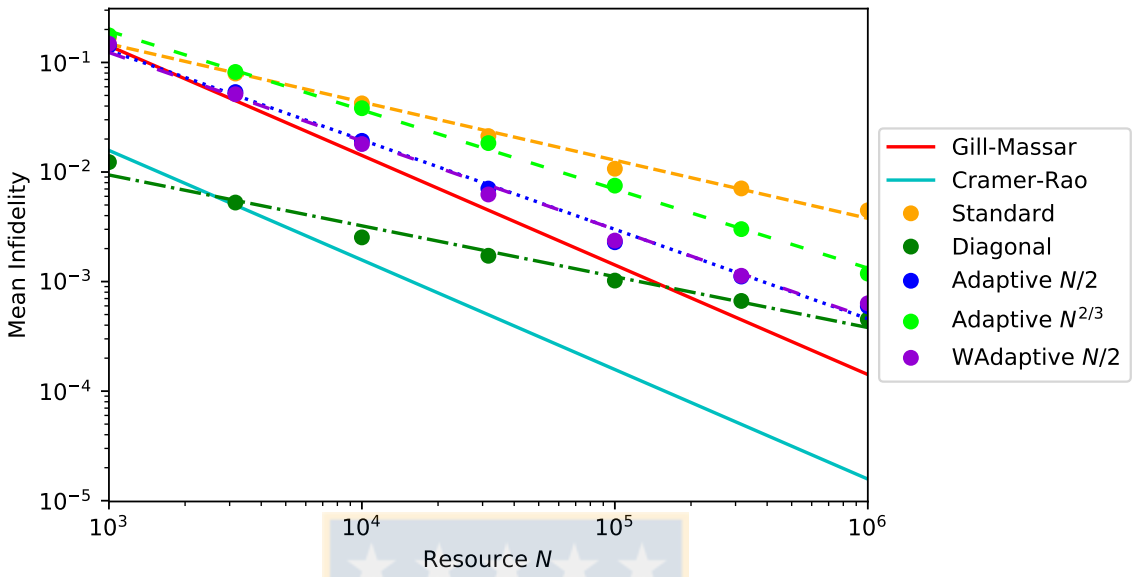

Figure 5.26: Comparison of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 100 8-dimensional pure states with noise $\lambda = 0.999$. Source: Made by the author.

	α	β
Gill-Massar	1	141.75
Cramer-Rao	1	15.75
Standard	0.530843722277	5.7824827963
Diagonal	0.464155852402	0.232010687621
Adaptive $N/2$	0.816341671561	36.1950617003
Adaptive $N^{2/3}$	0.72138055061	28.2573262078
WAdaptive $N/2$	0.807348785568	32.5844828634

Table 5.25: Comparison of fit coefficients of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 100 8-dimensional pure states with noise $\lambda = 0.999$. Source: Made by the author.

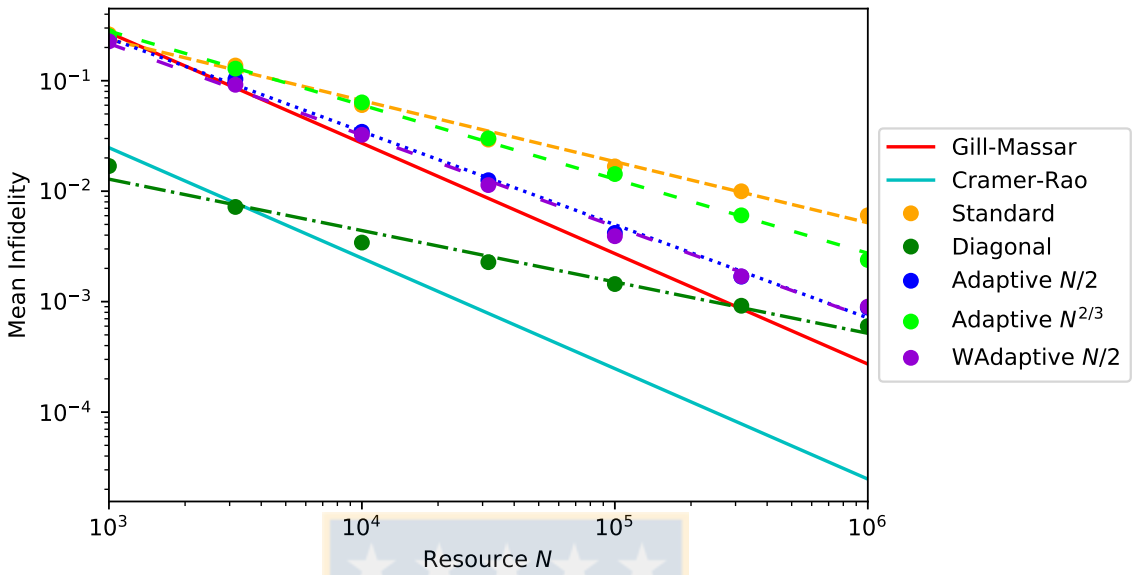

Figure 5.27: Comparison of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 100 10-dimensional pure states with noise $\lambda = 0.999$. Source: Made by the author.

	α	β
Gill-Massar	1	272.25
Cramer-Rao	1	24.75
Standard	0.553350380036	10.8172157727
Diagonal	0.465216319499	0.319513599256
Adaptive $N/2$	0.844056495002	82.6329174255
Adaptive $N^{2/3}$	0.67051003048	28.9403148461
WAdaptive $N/2$	0.828021608827	66.16230481

Table 5.26: Comparison of fit coefficients of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 100 10-dimensional pure states with noise $\lambda = 0.999$. Source: Made by the author.

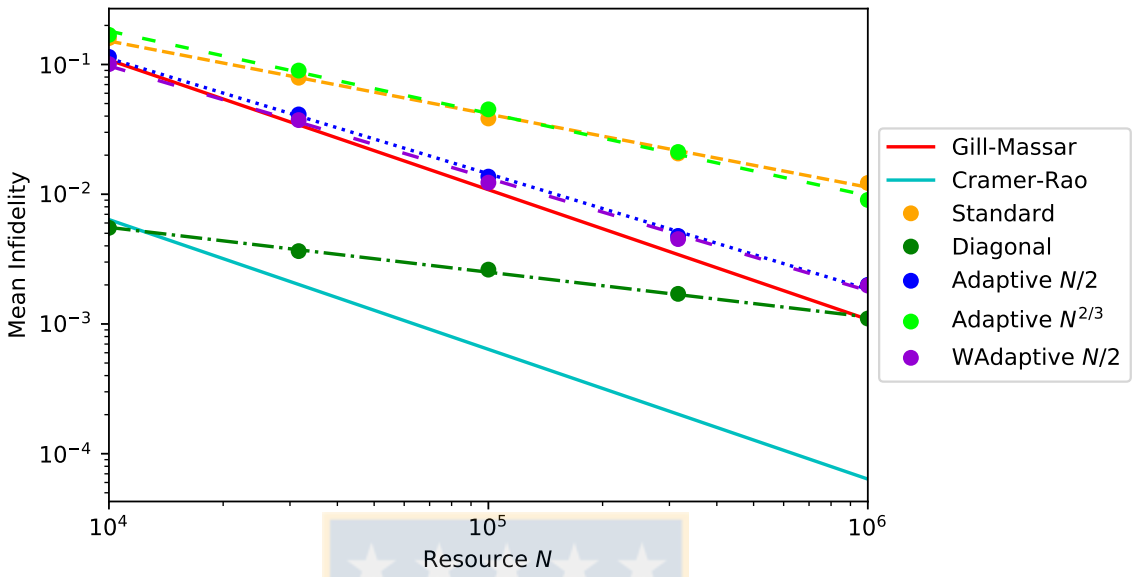

Figure 5.28: Comparison of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 100 16-dimensional pure states with noise $\lambda = 0.999$. Source: Made by the author.

	α	β
Gill-Massar	1	1083.75
Cramer-Rao	1	63.75
Standard	0.563494051051	27.2124622128
Diagonal	0.345544918203	0.133907026198
Adaptive $N/2$	0.889996968123	405.194498184
Adaptive $N^{2/3}$	0.634482960243	62.4984663451
WAdaptive $N/2$	0.866205686123	284.400555289

Table 5.27: Comparison of fit coefficients of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 100 16-dimensional pure states with noise $\lambda = 0.999$. Source: Made by the author.

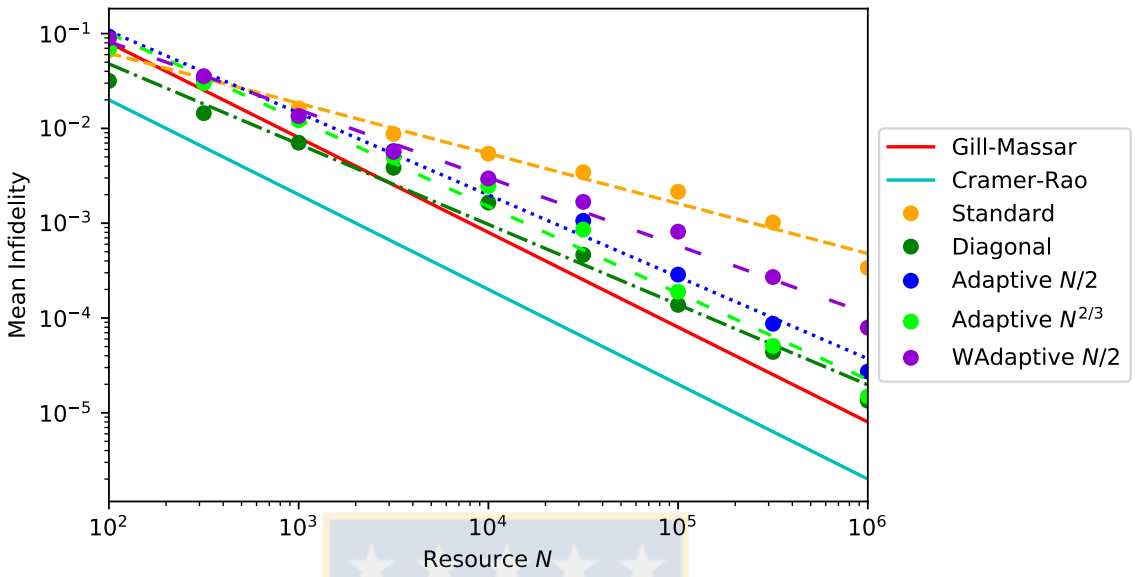

Figure 5.29: Comparison of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 100 3-dimensional pure states with noise $\lambda = 0.99$. Source: Made by the author.

	α	β
Gill-Massar	1	8.0
Cramer-Rao	1	2.0
Standard	0.528425655925	0.70829593936
Diagonal	0.846353175779	2.3592544272
Adaptive $N/2$	0.864305458869	5.69665745524
Adaptive $N^{2/3}$	0.91322822953	6.75067544367
WAdaptive $N/2$	0.717420015388	2.2376740063

Table 5.28: Comparison of fit coefficients of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 100 3-dimensional pure states with noise $\lambda = 0.99$. Source: Made by the author.

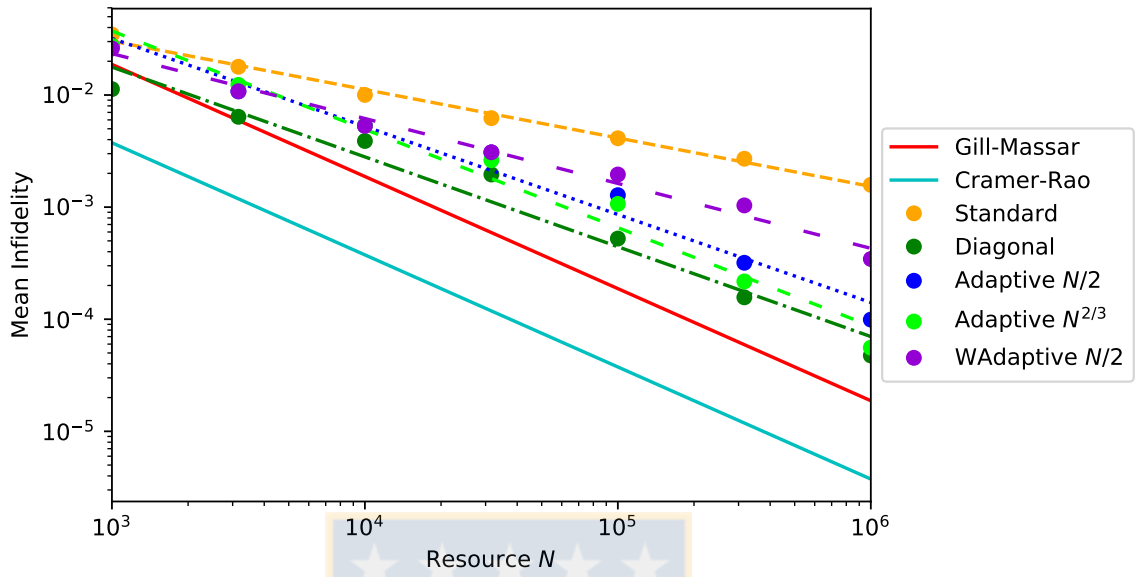

Figure 5.30: Comparison of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 100 4-dimensional pure states with noise $\lambda = 0.99$. Source: Made by the author.

	α	β
Gill-Massar	1	18.75
Cramer-Rao	1	3.75
Standard	0.432147979285	0.599381533441
Diagonal	0.801343622837	4.50162895128
Adaptive $N/2$	0.78524726407	7.25601432931
Adaptive $N^{2/3}$	0.877138930205	15.9366968172
WAdaptive $N/2$	0.578957257659	1.27717646592

Table 5.29: Comparison of fit coefficients of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 100 4-dimensional pure states with noise $\lambda = 0.99$. Source: Made by the author.

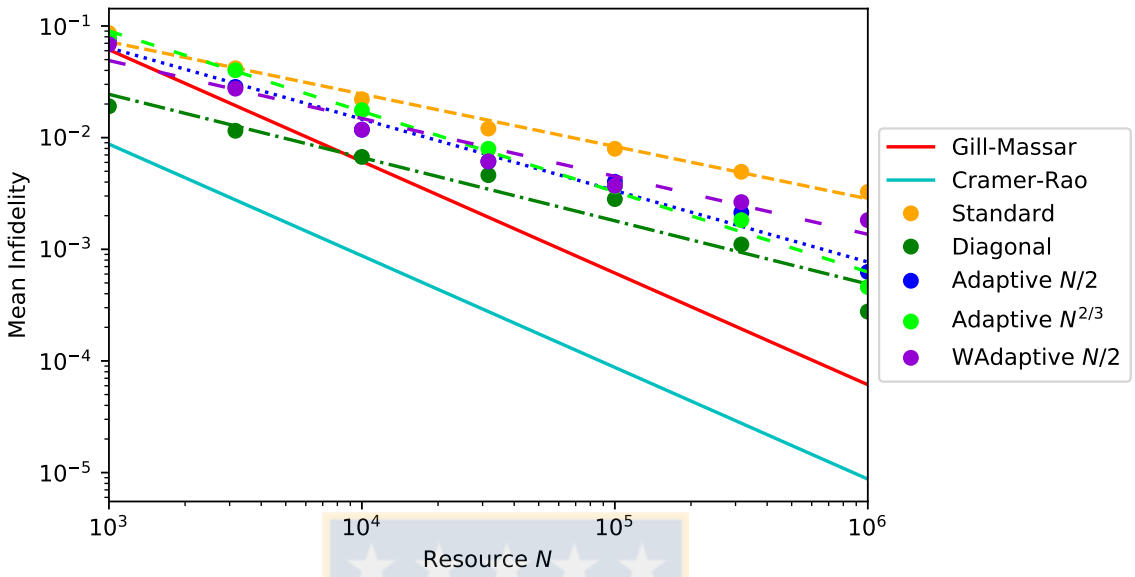

Figure 5.31: Comparison of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 100 6-dimensional pure states with noise $\lambda = 0.99$. Source: Made by the author.

	α	β
Gill-Massar	1	61.25
Cramer-Rao	1	8.75
Standard	0.468978203196	1.8454848187
Diagonal	0.566441530146	1.22291751384
Adaptive $N/2$	0.639690378102	5.29884666942
Adaptive $N^{2/3}$	0.719327852801	12.9680522034
WAdaptive $N/2$	0.518698314956	1.76457614137

Table 5.30: Comparison of fit coefficients of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 100 6-dimensional pure states with noise $\lambda = 0.99$. Source: Made by the author.

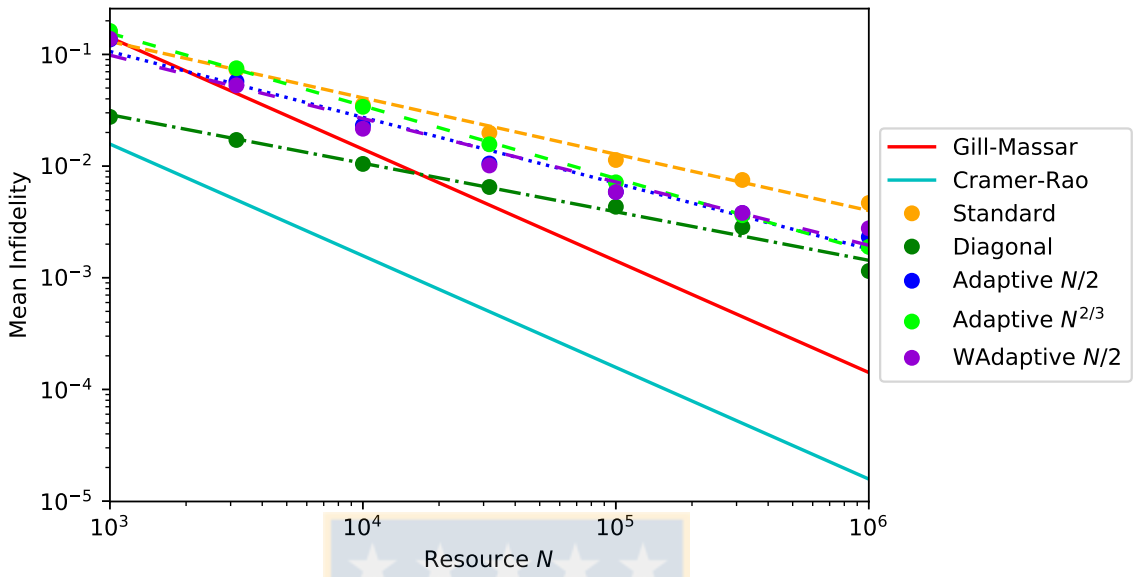

Figure 5.32: Comparison of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 100 8-dimensional pure states with noise $\lambda = 0.99$. Source: Made by the author.

	α	β
Gill-Massar	1	141.75
Cramer-Rao	1	15.75
Standard	0.504690504702	4.26365382741
Diagonal	0.434589318888	0.580526235074
Adaptive $N/2$	0.589682460744	6.24512411671
Adaptive $N^{2/3}$	0.650609782925	13.8815159729
WAdaptive $N/2$	0.56740939244	4.95852960559

Table 5.31: Comparison of fit coefficients of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 100 8-dimensional pure states with noise $\lambda = 0.99$. Source: Made by the author.

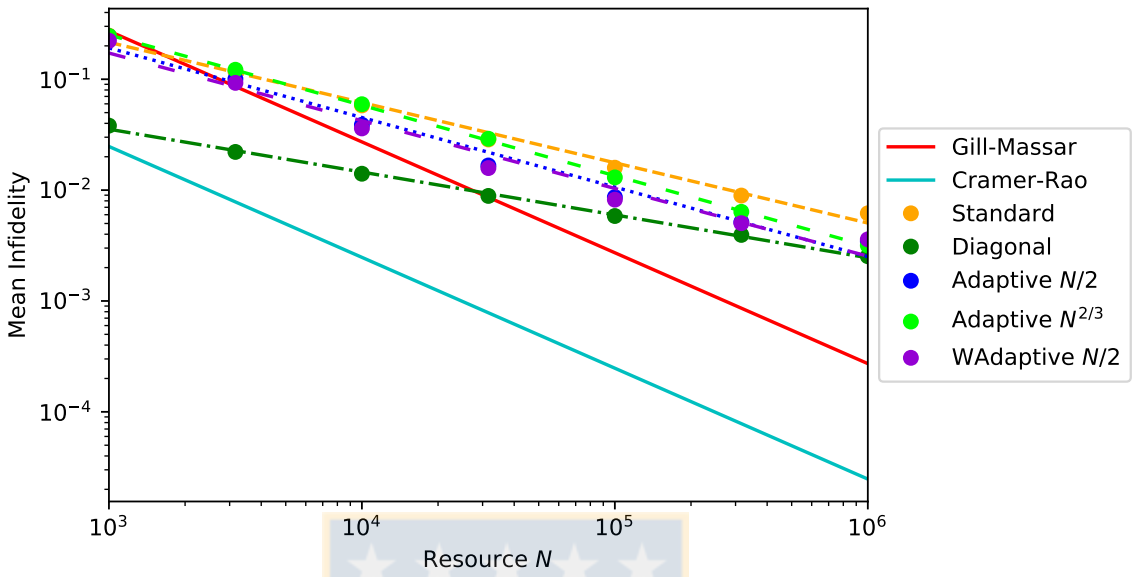

Figure 5.33: Comparison of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 100 10-dimensional pure states with noise $\lambda = 0.99$. Source: Made by the author.

	α	β
Gill-Massar	1	272.25
Cramer-Rao	1	24.75
Standard	0.543474571266	9.20315039447
Diagonal	0.386584133611	0.511949191381
Adaptive $N/2$	0.627594623655	14.6317704245
Adaptive $N^{2/3}$	0.635005219492	20.1848087699
WAdaptive $N/2$	0.610657397526	11.6999022371

Table 5.32: Comparison of fit coefficients of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 100 10-dimensional pure states with noise $\lambda = 0.99$. Source: Made by the author.

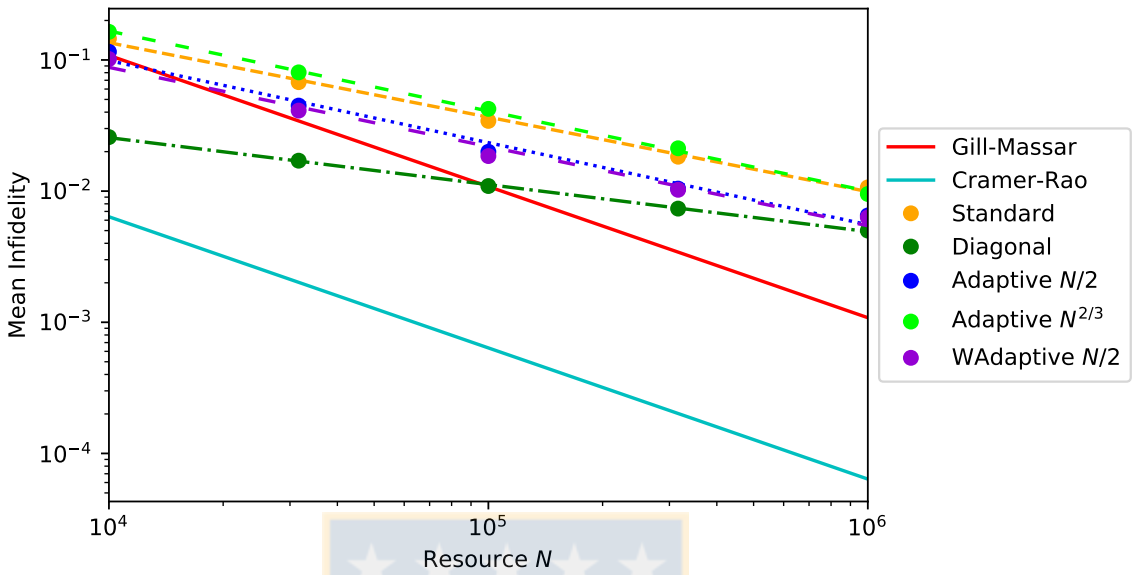

Figure 5.34: Comparison of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 100 16-dimensional pure states with noise $\lambda = 0.99$. Source: Made by the author.

	α	β
Gill-Massar	1	1083.75
Cramer-Rao	1	63.75
Standard	0.567622294413	25.2422062666
Diagonal	0.358187207654	0.692795186052
Adaptive $N/2$	0.625732423436	31.4910865197
Adaptive $N^{2/3}$	0.610296783383	45.90060537
WAdaptive $N/2$	0.603786790486	22.8357306162

Table 5.33: Comparison of fit coefficients of mean infidelity obtained from the simulation of the standard tomography and adaptive tomography of 100 16-dimensional pure states with noise $\lambda = 0.99$. Source: Made by the author.

5.4.5 Summary of the Simulations

From the simulations, it can be seen that with the same amount of resources, at least one adaptive tomographic method leads to better mean infidelity than standard tomography. For low rank states, adaptive tomography with $N_0 = N/2$ delivers the best mean infidelity. On the other hand, for mixed states, adaptive tomography with $N_0 = N^{2/3}$ leads to the best mean infidelity. In the case of pure states, this method provides mean infidelities close to the Gill-Massar bound. In the case of states with higher rank, this does not hold. For pure states with noise, the best technique is the adaptive tomography with $N_0 = N^{2/3}$, but from a sufficiently high dimension the best technique is the adaptive tomography with $N_0 = N/2$. Weighted adaptive tomography with $N_0 = N/2$ delivers mean infidelity close to adaptive tomography with $N_0 = N/2$, despite it does not minimize the infidelity. Therefore, the optimal value of the preliminary ensemble N_0 to optimize the two-stage adaptive standard tomography depends on the rank, dimension and the small eigenvalues of the state to be reconstructed, so in general, this is unknown. We recommend using adaptive tomography with the choice $N_0 = N/2$ for the reconstruction of an unknown state, because this method delivers in the best case (low rank states) a much better mean infidelity than standard tomography and a similar mean infidelity in the worst case (full rank states).

Chapter 6

Conclusion

We have generalized the two-stage adaptive tomographic method for a two-dimensional system (qubit) to the case of a d-dimensional system (qudit), with d arbitrary. In the first stage, a low precision estimate ρ_0 is obtained by standard tomography on a fraction N_0 of the total number N of identically prepared copies of the unknown state to be reconstructed. Later, we adapt the measurement base of the Generalized Gell-Mann matrices to the basis of eigenvectors of ρ_0 and perform a second higher accuracy standard tomography on the remaining ensemble of size $N-N_0$.

The two-stage adaptive tomographic protocol is based on the Taylor series of the infidelity and on the uncertainty of measurements in Gell-Mann matrices. We have obtained the series expansion of the infidelity by means of the Fréchet derivative, which has the advantage that it does not require a parameterization of the states. Thereby, this expansion is valid for any type of tomography. The variance in the measurements of the Gell-Mann matrices has been obtained by propagating the uncertainty of the measurement outcomes, which are in a multinomial distribution. The uncertainty in the estimation of the parameters defining the unknown state depends on these parameters.

The two-stage adaptive tomographic method improves the average infidelity between an unknown state of any rank and its estimate from $\mathcal{O}(1/\sqrt{N})$ to $\mathcal{O}(1/N)$. Thereby, this protocol delivers a much better infidelity than standard tomography

for equal resource N, with N large enough. In the case of pure unknown states, this method provides mean infidelities close to the Gill-Massar bound. Thereby, this protocol is near to the optimal estimation strategy for pure states by separable measurements. In the case of states with a higher rank, this does not hold.

We have also formulated a two-step adaptive tomography that reduces the weighted mean square error in order to attain the Gill-Massar bound. We have studied its effect on infidelity, although it minimizes another figure of merit. The weighted adaptive tomographic method delivers mean infidelities marginally worse than adaptive tomography for all types of states. It is necessary to study in more detail the effect of this algorithm on the weighted mean square error by numerical simulations.

We have seen that the performance of the two-stage adaptive tomographic method depends on the particular distribution of the ensemble among the two stages of the tomography. In general, the optimal value of N_0 is unknown. However, detailed simulations point out that the sample size of the preliminary estimation that leads to better mean infidelity for low rank states is $N_0 = N/2$. Instead, the sample size that delivers the best mean infidelity for full rank states is $N_0 = N^{2/3}$. We recommend using adaptive tomography with the choice $N_0 = N/2$ for the reconstruction of an unknown state because this method delivers in the best case a much better mean infidelity than standard tomography and a similar mean infidelity in the worst case.

Adaptive tomography requires measurements in the standard tomographic bases, which exist in any dimension. Therefore, this protocol can be applied to any dimension, which is an advantage over mutually unbiased bases based tomography. This is because the latter exists in dimensions that correspond to integer powers of prime numbers, otherwise its existence is uncertain. Besides, two-stage adaptive tomography has an experimental advantage over tomography by SIC-POVM, because the bases to be measured are easier to implement than a POVM on high dimensions. Moreover, computational expense of this two stages adaptive tomography method

is lesser than a self-learning adaptive tomography. However, this method also has several disadvantages. In the first place, this method is difficult to implement in multipartite systems, because it requires measurements on entangled bases. Another disadvantage of this method is that it requires twice the number of measurements as standard tomography, that is $2d(d^2-1)$ measurement outcomes. Besides, the computational expense is higher than that of standard tomography, because it needs to diagonalize the preliminary density matrix and the post-processing is with twice the data.

Adaptive two-step tomography improves the accuracy of the mean infidelity. So far, our analysis considers a source of error of finite character of the ensemble to be measurement. The realization of simulations of this algorithm considering experimental noises is pending. For example, it is observed that the detection of photons has Poisson distribution instead multinomial distribution. These simulations would allow us to find ranges where the improvement in accuracy of the infidelity provided by this method is not overshadowed by experimental noise. Besides, these simulations would be useful for experimental realization of the two-stage adaptive tomographic method.

The two-stage adaptive tomographic method improves the precision of the reconstruction of an unknown state, which has positive implications in applications of Quantum Mechanics, such as Quantum Information, Quantum Computation and Quantum Metrology. This method can contribute to the formulation of new adaptive tomographic methods in high dimensions. The natural extension of this method is to replace the standard tomographies by mutually unbiased bases based tomographies, which has already been studied in the context of qubits. This protocol reduces the number of outcomes from $2d(d^2-1)$ to $2d(d+1)$ on prime power dimensional spaces.

Appendix A

Products between Gell-Mann matrices

In this appendix we proof the product between the Gell-Mann matrices,

$$\mathrm{Tr}(\sigma_i\sigma_j) = 2\delta_{ij}. \tag{A.1}$$

- If $j = l$ it is impossible that $i = m$,

$$\begin{aligned}
\mathrm{Tr}\big(\sigma^x_{ij}\sigma^x_{lm}\big) &= \mathrm{Tr}\Big[\Big(\,|i\rangle\langle j| + |j\rangle\langle i|\,\Big)\Big(\,|l\rangle\langle m| + |m\rangle\langle l|\,\Big)\Big], && \text{(A.2)}\\
&= \mathrm{Tr}\Big[\delta_{jl}\,|i\rangle\langle m| + \delta_{jm}\,|i\rangle\langle l| + \delta_{il}\,|j\rangle\langle m| + \delta_{im}\,|j\rangle\langle l|\,\Big], && \text{(A.3)}\\
&= \delta_{jl}\delta_{im} + \delta_{jm}\delta_{il} + \delta_{il}\delta_{jm} + \delta_{im}\delta_{jl}, && \text{(A.4)}\\
&= 2\delta_{il}\delta_{jm}. && \text{(A.5)}
\end{aligned}$$

$$\begin{aligned}
\mathrm{Tr}\big(\sigma^y_{ij}\sigma^y_{lm}\big) &= \mathrm{i}^2\,\mathrm{Tr}\Big[\Big(\,|i\rangle\langle j| - |j\rangle\langle i|\,\Big)\Big(\,|l\rangle\langle m| - |m\rangle\langle l|\,\Big)\Big] && \text{(A.6)}\\
&= -\,\mathrm{Tr}\Big[\delta_{jl}\,|i\rangle\langle m| - \delta_{jm}\,|i\rangle\langle l| - \delta_{il}\,|j\rangle\langle m| + \delta_{im}\,|j\rangle\langle l|\,\Big] && \text{(A.7)}\\
&= -\delta_{jl}\delta_{im} + \delta_{jm}\delta_{il} + \delta_{il}\delta_{jm} - \delta_{im}\delta_{jl} && \text{(A.8)}\\
&= 2\delta_{il}\delta_{jm}. && \text{(A.9)}
\end{aligned}$$

$$\begin{aligned}
\mathrm{Tr}\big(\sigma^x_{ij}\sigma^y_{ml}\big) &= -\mathrm{i}\,\mathrm{Tr}\Big[\Big(\,|i\rangle\langle j| + |j\rangle\langle i|\,\Big)\Big(\,|l\rangle\langle m| - |m\rangle\langle l|\,\Big)\Big] && \text{(A.10)}\\
&= -\mathrm{i}\,\mathrm{Tr}\Big[\delta_{jl}\,|i\rangle\langle m| - \delta_{jm}\,|i\rangle\langle l| + \delta_{il}\,|j\rangle\langle m| - \delta_{im}\,|j\rangle\langle l|\,\Big] && \text{(A.11)}\\
&= -\mathrm{i}\Big(\delta_{jl}\delta_{im} - \delta_{jm}\delta_{il} + \delta_{il}\delta_{jm} - \delta_{im}\delta_{jl}\Big) && \text{(A.12)}\\
&= 0. && \text{(A.13)}
\end{aligned}$$

- Let us suppose $k \leq m$,

$$\mathrm{Tr}(\sigma_k^z \sigma_m^z) = \sqrt{\frac{4}{k(k+1)m(m+1)}}\,\mathrm{Tr}\left[\left(\sum_{l=1}^{k} |l\rangle\langle l| - k\,|k+1\rangle\langle k+1|\right)\right. \\ \left.\times\left(\sum_{n=1}^{m} |n\rangle\langle n| - m\,|m+1\rangle\langle m+1|\right)\right] \tag{A.14}$$

$$= \frac{2}{\sqrt{k(k+1)m(m+1)}}\,\mathrm{Tr}\left[\sum_{l,n=1}^{k,m} \delta_{ln}\,|l\rangle\langle n| \right. \\ - k\left(\sum_{n=1}^{m+1} \delta_{k+1,n}\,|k+1\rangle\langle n| - \delta_{k+1,m+1}\,|k+1\rangle\langle m+1|\right) \\ \left. + mk\delta_{k+1,m+1}\,|k+1\rangle\langle m+1|\right] \tag{A.15}$$

$$= \frac{2}{\sqrt{k(k+1)m(m+1)}}\Big[k - k(1-\delta_{km}) + k^2\delta_{mk}\Big] \tag{A.16}$$

$$= \frac{2k(k+1)\delta_{mk}}{\sqrt{k(k+1)m(m+1)}} \tag{A.17}$$

$$= 2\delta_{km} \tag{A.18}$$

- Because $i \neq j$,

$$\mathrm{Tr}\big(\sigma_k^z \sigma_{ij}^x\big) = \sqrt{\frac{2}{k(k+1)}}\,\mathrm{Tr}\left(\left[\sum_{l=1}^{k} |l\rangle\langle l| - k\,|k+1\rangle\langle k+1|\right]\Big[\,|i\rangle\langle j| + |j\rangle\langle i|\,\Big]\right) \tag{A.19}$$

$$= \sqrt{\frac{2}{k(k+1)}}\,\mathrm{Tr}\left(\sum_{l=1}^{k} \delta_{li}\,|l\rangle\langle j| + \sum_{l=1}^{k} \delta_{lj}\,|l\rangle\langle i| \right. \\ \left. - k\delta_{k+1,i}\,|k+1\rangle\langle j| - k\delta_{k+1,j}\,|k+1\rangle\langle i|\right) \tag{A.20}$$

$$= \sqrt{\frac{2}{k(k+1)}}\Big[\delta_{ij} + \delta_{ji} - k\delta_{k+1,i}\delta_{k+1,j} - k\delta_{k+1,j}\delta_{k+1,i}\Big] \tag{A.21}$$

$$= 0 \tag{A.22}$$

$$\mathrm{Tr}\big(\sigma_k^z \sigma_{ij}^y\big) = -\mathrm{i}\sqrt{\frac{2}{k(k+1)}}\,\mathrm{Tr}\left(\left[\sum_{l=1}^{k} |l\rangle\langle l| - k\,|k+1\rangle\langle k+1|\right]\Big[\,|i\rangle\langle j| - |j\rangle\langle i|\,\Big]\right) \tag{A.23}$$

$$= -\mathrm{i}\sqrt{\frac{2}{k(k+1)}}\,\mathrm{Tr}\left(\sum_{l=1}^{k}\delta_{li}\,|l\rangle\langle j| - \sum_{l=1}^{k}\delta_{lj}\,|l\rangle\langle i| - k\delta_{k+1,i}\,|k+1\rangle\langle j| + k\delta_{k+1,j}\,|k+1\rangle\langle i|\right) \tag{A.24}$$

$$= -\mathrm{i}\sqrt{\frac{2}{k(k+1)}}\left[\delta_{ij} - \delta_{ji} - k\delta_{k+1,i}\delta_{k+1,j} + k\delta_{k+1,j}\delta_{k+1,i}\right] \tag{A.25}$$

$$= 0. \tag{A.26}$$

Appendix B

Generalized Bloch Representation

Let us consider the representation of qudits by Gell-Mann matrices,

$$\rho = \frac{1}{d}\mathbb{I} + \frac{1}{2}\sum_{i=1}^{d^2-1} S_i \sigma_i. \tag{B.1}$$

The purity of ρ is

$$\begin{aligned}
\mathrm{Tr}\big(\rho^2\big) &= \frac{1}{d} + \frac{1}{4}\sum_{i,j=1}^{d^2-1} S_i S_j \,\mathrm{Tr}(\sigma_i \sigma_j) && \text{(B.2)}\\
&= \frac{1}{d} + \frac{1}{4}\sum_{\substack{1\le i<j\le d\\ 1\le l<m\le d}} S_{ij}^x S_{lm}^x \,\mathrm{Tr}\big(\sigma_{ij}^x \sigma_{lm}^x\big) + \frac{1}{4}\sum_{\substack{1\le i<j\le d\\ 1\le l<m\le d}} S_{ij}^y S_{lm}^y \,\mathrm{Tr}\big(\sigma_{ij}^y \sigma_{lm}^y\big)\\
&\quad + \frac{1}{4}\sum_{i,j=1}^{d-1} S_i^z S_j^z \,\mathrm{Tr}\big(\sigma_i^z \sigma_j^z\big) && \text{(B.3)}\\
&= \frac{1}{d} + \frac{1}{4}\sum_{\substack{1\le i<j\le d\\ 1\le l<m\le d}} S_{ij}^x S_{lm}^x 2\delta_{il}\delta_{jm} + \frac{1}{4}\sum_{\substack{1\le i<j\le d\\ 1\le l<m\le d}} S_{ij}^y S_{lm}^y 2\delta_{il}\delta_{jm}\\
&\quad + \frac{1}{4}\sum_{i,j=1}^{d-1} S_i^z S_j^z 2\delta_{ij} && \text{(B.4)}\\
&= \frac{1}{d} + \frac{1}{2}\sum_{1\le i<j\le d} (S_{ij}^x)^2 + \frac{1}{2}\sum_{1\le i<j\le d} (S_{ij}^y)^2 + \frac{1}{2}\sum_{i=1}^{d-1} (S_i^z)^2 && \text{(B.5)}\\
&= \frac{1}{d} + \frac{1}{2}\sum_{i=1}^{d^2-1} S_i^2. && \text{(B.6)}
\end{aligned}$$

Then, using $1/d \leq \mathrm{Tr}(\rho^2) \leq 1$, we get

$$0 \leq \sum_{i=1}^{d^2-1} S_i^2 \leq 2\frac{d-1}{d}, \tag{B.7}$$

where the upper bound represents the pure states and the lower bound represents the maximally mixed state. The last inequality told us that the states are on the inside of a d^2-1 dimensional sphere with radius $r_d = \sqrt{2(d-1)/d}$, which is analogous to the Bloch sphere, but in higher dimensions. But since $-1 \leq S_{ij}^x \leq 1$, $-1 \leq S_{ij}^y \leq 1$ and $-k\sqrt{2/k(k+1)} \leq S_k^z \leq \sqrt{2/k(k+1)}$, actually the states are only a section of this sphere.

Appendix C

Derivatives of Square Root Function

Here we calculate the first and second derivatives of square root function $f(A) = \sqrt{A}$ by Freshet derivative. They can be obtained from the product rule

$$Df^2(A)(B) = D[f \cdot f](A)(B) = Df(A)(B) \cdot f(A) + f(A) \cdot Df(A)(B). \tag{C.1}$$

Since $Df^2(A)(B) = B$, we obtain that

$$Df(A)(B) \cdot \sqrt{A} + \sqrt{A} \cdot Df(A)(B) = B. \tag{C.2}$$

Let us now consider $\{|a_i\rangle\}_{i=1,\dots,r}$ the basis of eigenvectors of A and $\{a_i\}_{i=1,\dots,r}$ the corresponding eigenvalues of A, where r is the dimension of A, which in this case is equal to rank of ρ. Expanding on this basis, we have

$$\sum_{j=1}^{r} \Big(\big[Df(A)(B)\big]_{ij} \sqrt{a_j} \delta_{jk} + \sqrt{a_i} \delta_{ij} \big[Df(A)(B)\big]_{jk} \Big) = B_{ik} \tag{C.3}$$

$$\sqrt{a_k} \big[Df(A)(B)\big]_{ik} + \sqrt{a_i} \big[Df(A)(B)\big]_{ik} = B_{ik} \tag{C.4}$$

$$(\sqrt{a_i} + \sqrt{a_k}) \big[Df(A)(B)\big]_{ik} = B_{ik}. \tag{C.5}$$

Then,

$$\big[Df(A)(B)\big]_{ik} = \frac{B_{ik}}{\sqrt{a_i} + \sqrt{a_k}}. \tag{C.6}$$

Since $A = \rho_r^2$ is full rank, there is no indetermination in the derivative. For the second derivative we use the product rule again,

$$D^2 f^2(A)(B)(C) = D\big[Df^2(A)(B)\big](C) \tag{C.7}$$

$$= D\big[Df(A)(B)\cdot f(A) + f(A)\cdot Df(A)(B)\big](C) \tag{C.8}$$

$$= D^2f(A)(B)(C)\cdot f(A) + Df(A)(B)\cdot Df(A)(C) \tag{C.9}$$

$$+ Df(A)(C)\cdot Df(A)(B) + f(A)\cdot D^2f(A)(B)(C) \tag{C.10}$$

$$= D^2f(A)(B)(C)\cdot\sqrt{A} + Df(A)(B)\cdot Df(A)(C) \tag{C.11}$$

$$+ Df(A)(C)\cdot Df(A)(B) + \sqrt{A}\cdot D^2f(A)(B)(C). \tag{C.12}$$

Using $D^2f^2(A)(B)(C) = 0$ and writing per component,

$$0 = \sum_{j=1}^{r}\left(\big[D^2f(A)(B)(C)\big]_{ij}\sqrt{a_j}\delta_{jk} + \frac{B_{ij}}{\sqrt{a_i}+\sqrt{a_j}}\frac{C_{jk}}{\sqrt{a_j}+\sqrt{a_k}}\right.$$

$$\left. + \frac{C_{ij}}{\sqrt{a_i}+\sqrt{a_j}}\frac{B_{jk}}{\sqrt{a_j}+\sqrt{a_k}} + \sqrt{a_i}\delta_{ij}\big[D^2f(A)(B)(C)\big]_{jk}\right) \tag{C.13}$$

$$= \sqrt{a_k}\big[D^2f(A)(B)(C)\big]_{ik} + \sqrt{a_i}\big[D^2f(A)(B)(C)\big]_{ik}$$

$$+ \sum_{j=1}^{r}\left(\frac{B_{ij}}{\sqrt{a_i}+\sqrt{a_j}}\frac{C_{jk}}{\sqrt{a_j}+\sqrt{a_k}} + \frac{C_{ij}}{\sqrt{a_i}+\sqrt{a_j}}\frac{B_{jk}}{\sqrt{a_j}+\sqrt{a_k}}\right) \tag{C.14}$$

$$= (\sqrt{a_i}+\sqrt{a_k})\big[D^2f(A)(B)(C)\big]_{ik}$$

$$+ \sum_{j=1}^{r}\left(\frac{B_{ij}}{\sqrt{a_i}+\sqrt{a_j}}\frac{C_{jk}}{\sqrt{a_j}+\sqrt{a_k}} + \frac{C_{ij}}{\sqrt{a_i}+\sqrt{a_j}}\frac{B_{jk}}{\sqrt{a_j}+\sqrt{a_k}}\right). \tag{C.15}$$

Then, the second derivative is

$$\big[D^2f(A)(B)(C)\big]_{ik} = -\frac{1}{\sqrt{a_i}+\sqrt{a_k}}\sum_{j=1}^{r}\left(\frac{B_{ij}}{\sqrt{a_i}+\sqrt{a_j}}\frac{C_{jk}}{\sqrt{a_j}+\sqrt{a_k}}\right.$$

$$\left. + \frac{C_{ij}}{\sqrt{a_i}+\sqrt{a_j}}\frac{B_{jk}}{\sqrt{a_j}+\sqrt{a_k}}\right). \tag{C.16}$$

Appendix D

Quantum Fisher Information Matrix of Adaptive Tomography

Let J the quantum Fisher information matrix of ρ. We will obtain this matrix by getting its blocks,

$$J = \begin{pmatrix} J^x & J^{xy} & J^{xz} \\ J^{yx} & J^y & J^{yz} \\ J^{zx} & J^{zy} & J^z \end{pmatrix}. \tag{D.1}$$

Let us suppose $i < j$ and $i' < j'$,

$$[J^x]_{ij,i'j'} = \frac{1}{2} \sum_{k,l=1}^{d} \frac{1}{\lambda_k + \lambda_l} \langle \lambda_k | \sigma^x_{ij} | \lambda_l \rangle \ \langle \lambda_l | \sigma^x_{i'j'} | \lambda_k \rangle \tag{D.2}$$

$$= \frac{1}{2} \sum_{k,l=1}^{d} \frac{1}{\lambda_k + \lambda_l} (\delta_{ik}\delta_{jl} + \delta_{jk}\delta_{il})(\delta_{i'k}\delta_{j'l} + \delta_{j'k}\delta_{i'l}) \tag{D.3}$$

$$= \frac{1}{\lambda_i + \lambda_j} (\delta_{ii'}\delta_{jj'} + \delta_{ij'}\delta_{ji'}) \tag{D.4}$$

$$= \frac{1}{\lambda_i + \lambda_j} \delta_{ii'}\delta_{jj'}. \tag{D.5}$$

$$[J^y]_{ij,i'j'} = \frac{1}{2} \sum_{k,l=1}^{d} \frac{1}{\lambda_k + \lambda_l} \langle \lambda_k | \sigma^x_{ij} | \lambda_l \rangle \ \langle \lambda_l | \sigma^x_{i'j'} | \lambda_k \rangle \tag{D.6}$$

$$= \frac{1}{2} \sum_{k,l=1}^{d} \frac{-\mathrm{i}^2}{\lambda_k + \lambda_l} (\delta_{ik}\delta_{jl} - \delta_{jk}\delta_{il})(\delta_{i'k}\delta_{j'l} - \delta_{j'k}\delta_{i'l}) \tag{D.7}$$

$$= \frac{1}{\lambda_i + \lambda_j} (\delta_{ii'}\delta_{jj'} - \delta_{ij'}\delta_{ji'}) \tag{D.8}$$

$$= \frac{1}{\lambda_i + \lambda_j} \delta_{ii'} \delta_{jj'}. \tag{D.9}$$

$$[J^{xy}]_{ij,i'j'} = \frac{1}{2} \sum_{k,l=1}^{d} \frac{1}{\lambda_k + \lambda_l} \langle \lambda_k | \sigma^x_{ij} | \lambda_l \rangle \ \langle \lambda_l | \sigma^y_{i'j'} | \lambda_k \rangle \tag{D.10}$$

$$= \frac{\mathrm{i}}{2} \sum_{k,l=1}^{d} \frac{1}{\lambda_k + \lambda_l} (\delta_{ki} \delta_{jl} + \delta_{kj} \delta_{il})(\delta_{ki'} \delta_{j'l} - \delta_{kj'} \delta_{i'l}) \tag{D.11}$$

$$= \frac{\mathrm{i}}{2} \frac{1}{(\lambda_i + \lambda_j)} (\delta_{ii'} \delta_{jj'} - \delta_{ij'} \delta_{ji'} + \delta_{i'j} \delta_{ij'} - \delta_{jj'} \delta_{ii'}) \tag{D.12}$$

$$= 0. \tag{D.13}$$

$$[J^{xz}]_{ij,k} = \frac{1}{2} \sum_{l,m=1}^{d} \frac{1}{\lambda_l + \lambda_m} \langle \lambda_l | \sigma^x_{ij} | \lambda_m \rangle \ \langle \lambda_m | \sigma^z_{k} | \lambda_l \rangle \tag{D.14}$$

$$= \frac{1}{2} \sum_{l,m=1}^{d} \frac{1}{\lambda_l + \lambda_m} (\delta_{li} \delta_{jm} + \delta_{lj} \delta_{im}) \langle \lambda_m | \sigma^z_{k} | \lambda_l \rangle \tag{D.15}$$

$$= \frac{1}{2} \frac{1}{\lambda_i + \lambda_j} \Big(\langle \lambda_i | \sigma^z_{k} | \lambda_j \rangle + \langle \lambda_j | \sigma^z_{k} | \lambda_i \rangle \Big) \tag{D.16}$$

$$= 0. \tag{D.17}$$

$$[J^{yz}]_{ij,k} = \frac{1}{2} \sum_{l,m=1}^{d} \frac{1}{\lambda_l + \lambda_m} \langle \lambda_l | \sigma^y_{ij} | \lambda_m \rangle \ \langle \lambda_m | \sigma^z_{k} | \lambda_l \rangle \tag{D.18}$$

$$= \frac{1}{2} \sum_{l,m=1}^{d} \frac{-\mathrm{i}}{\lambda_l + \lambda_m} (\delta_{li} \delta_{jm} - \delta_{lj} \delta_{im}) \langle \lambda_m | \sigma^z_{k} | \lambda_l \rangle \tag{D.19}$$

$$= \frac{1}{2} \frac{-\mathrm{i}}{\lambda_i + \lambda_j} \Big(\langle \lambda_i | \sigma^z_{k} | \lambda_j \rangle - \langle \lambda_j | \sigma^z_{k} | \lambda_i \rangle \Big) \tag{D.20}$$

$$= 0. \tag{D.21}$$

Let us suppose $k \leq k'$,

$$[J^{z}]_{k,k'} = \frac{1}{2} \sum_{l,m=1}^{d} \frac{1}{\lambda_l + \lambda_m} \langle \lambda_l | \sigma^z_{k} | \lambda_m \rangle \ \langle \lambda_m | \sigma^z_{k'} | \lambda_l \rangle \tag{D.22}$$

$$= \frac{1}{2} \sum_{l=1}^{d} \frac{1}{\lambda_l} \frac{1}{\sqrt{kk'(k+1)(k'+1)}} \Big(h_{k \geq l} - k \delta_{k+1,l} \Big) \Big(h_{k' \geq l} - k' \delta_{k'+1,l} \Big) \tag{D.23}$$

$$=\frac{1}{2\sqrt{kk'(k+1)(k'+1)}}\left(\sum_{l=1}^{k}\frac{1}{\lambda_l}-\frac{k}{\lambda_{k+1}}h_{k'\geq k+1}+\frac{kk'}{\lambda_{k+1}}\delta_{kk'}\right). \tag{D.24}$$

Thereby, the quantum Fisher matrix is diagonal except in the block J^z,

$$J=\begin{pmatrix} J^x & \Theta & \Theta \\ \Theta & J^y & \Theta \\ \Theta & \Theta & J^z \end{pmatrix}. \tag{D.25}$$

Appendix E

Integration on the State Space

In quantum information, it is common to find the need to perform integrals over quantum states,

$$\Gamma = \int_{\mathcal{H}} F(\rho) \mathrm{d}\mu(\rho). \tag{E.1}$$

An example can be the mean of some function over the space of Hilbert. From the numerical point of view, these integrals are solved by the method of Monte Carlo. This consists in estimating this integral using the average of the function evaluated in N uniformly distributed random states $\{\rho_i\}_{i=1,\ldots,N}$,

$$\Gamma_N = \frac{1}{N} \sum_{i=1}^{N} F(\rho_i). \tag{E.2}$$

The exact integral is obtained at the infinite limit of the estimate,

$$\Gamma_N = \lim_{N \to \infty} \Gamma_N. \tag{E.3}$$

The states are generated by *uniformly Haar measure* [74], that is a unitary invariant measure,

$$\mathrm{d}\mu(\rho) = \mathrm{d}\mu(U^\dagger \rho U). \tag{E.4}$$

In order to do this, we us a matrix Z, whose z_{ij} elements are randomly chosen according to the complex normal distribution,

$$p(z_{ij}) = \frac{1}{\pi} e^{-|z_{ij}|^2}. \tag{E.5}$$

Then, the probability density of Z is

$$f(Z) = \prod_{i,j=1}^{N} p(z_{ij}) = \frac{1}{\pi^{N^2}} e^{-\sum_{i,j=1}^{N} |z_{ij}|^2} = \frac{1}{\pi^{N^2}} e^{-\operatorname{Tr}\left(Z^\dagger Z\right)}. \tag{E.6}$$

This density probability is invariant under $Z \to UZ$, because the trace is cyclic,

$$f(UZ) = \frac{1}{\pi^{N^2}} e^{-\operatorname{Tr}\left(U^\dagger Z^\dagger Z U\right)} = \frac{1}{\pi^{N^2}} e^{-\operatorname{Tr}\left(Z^\dagger Z\right)} = f(Z). \tag{E.7}$$

We can decompose the matrix Z into a unitary matrix Q and an upper triangular matrix R,

$$Z = QR. \tag{E.8}$$

This is called the *QR decomposition* of Z. The columns of the matrix Q are the vectors of the base resulting from applying the Gram-Schmidt procedure (2.24) to the columns of the matrix Z. The probability density of Q is

$$f(Q) = \frac{1}{\pi^{N^2}} e^{-\operatorname{Tr}\left((QR)^\dagger QR\right)} = \frac{1}{\pi^{N^2}} e^{-\operatorname{Tr}\left(R^\dagger Q^\dagger QR\right)} = \frac{1}{\pi^{N^2}} e^{-\operatorname{Tr}\left(R^\dagger R\right)}. \tag{E.9}$$

The matrix Q has uniformly probability density and unitary invariant because this does not depend on Q. Therefore, this method generates a unitary matrix Q uniformly distributed by Haar measure.

Thereby, the following density matrix is uniformly distributed by Haar measure,

$$\rho = QLQ^\dagger, \tag{E.10}$$

where L is a diagonal matrix with the eigenvalues of ρ. The way to select L defines different random density matrices. The probability measure of ρ has a product form [75]

$$\mathrm{d}\mu(\rho) = \mathrm{d}\mu(L) \times \mathrm{d}\mu(Q). \tag{E.11}$$

In particular, we have chosen $L = RR^\dagger / \operatorname{Tr}\left(RR^\dagger\right)$,

$$\rho = Q \frac{RR^\dagger}{\operatorname{Tr}(RR^\dagger)} Q^\dagger = \frac{QR(QR)^\dagger}{\operatorname{Tr}\left(QR(QR)^\dagger\right)} = \frac{ZZ^\dagger}{\operatorname{Tr}(ZZ^\dagger)}. \tag{E.12}$$

Bibliography

[1] Č. Brukner, M. Żukowski y A. Zeilinger, Phys. Rev. Lett. **89**, 19 (2002).

[2] G. M. Nikolopoulos, K. S. Ranade y G. Alber, Phys. Rev. A **73**, 3 (2006).

[3] N. J. Cerf, M. Bourennane, A. Karlsson y N. Gisin, Phys. Rev. Lett. **88**, 12 (2001).

[4] H. S. Tonchev y N. V. Vitanov, Phys. Rev. A **94**, 4 (2016).

[5] M.-X. Luo, X.-B. Chen, Y.-X. Yang y X. Wang, Sci. Rep. **4**, 4044 (2014).

[6] D. P. O'Leary, G. K. Brennen y S. S. Bullock, Phys. Rev. A **74**, 3 (2006).

[7] D. Collins, S. Popescu, D. Collins, S. Popescu, N. Gisin, N. Linden y S. Massar, Phys. Rev. Lett. **88**, 4 (2002).

[8] D. Kaszlikowski, P. Gnaciński, M. Zukowski, W. Miklaszewski y A. Zeilinger, Phys. Rev. Lett. **85**, 21 (2000).

[9] R. T. Thew, K. Nemoto, A. G. White and W. J. Munro, Phys. Rev. A **66**, 1 (2002).

[10] K. Vogel y H. Risken, Phys. Rev. A **40**, 5 (sep 1989).

[11] D. T. Smithey, M. Beck, M. G. Raymer and A. Faridani, D. T. Smithey, M. Beck, M. G. Raymer y A. Faridani, Phys. Rev. Lett. **70**, 9 (mar 1993).

[12] C. R. Müller, B. Stoklasa, C. Peuntinger, C. Gabriel, J. Řeháček, Z. Hradil, A. B. Klimov, G. Leuchs, C. Marquardt y L. L. Sánchez-Soto, New J. Phys. **14** (2012).

[13] M. Guţă, T. Kypraios y I. Dryden, New J. Phys. **14**, 10 (2012).

[14] H. Häffner, W. Hänsel, C. F. Roos, J. Benhelm, D. Chek-al kar, M. Chwalla, T. Körber, U. D. Rapol, M. Riebe, P. O. Schmidt, C. Becher, O. Gühne, W. Dür y R. Blatt, Nature **438**, 7068 (2005).

[15] S. Wallentowitz, B. Seifert y S. Godoy, New J. Phys. **14** (2012).

[16] C Sayrin, I Dotsenko, S Gleyzes, M Brune, J M Raimond and S Haroche, New J. Phys. **14** (2012).

[17] S. L. Christensen, J. B. Béguin, H. L. Sørensen, E. Bookjans, D. Oblak, J. H. Müller, J. Appel y E. S. Polzik, New J. Phys. **15** (2013).

[18] G. Mauro D'Ariano, Martina De Laurentis, Matteo G. A. Paris, Alberto Porzio and Salvatore Solimeno, J. Opt. B Quantum Semiclass. Opt **4** (2002).

[19] J. S. Lundeen, A. Feito, H. Coldenstrodt-Ronge, K. L. Pregnell, C. Silberhorn, T. C. Ralph, J. Eisert, M. B. Plenio y I. A. Walmsley, Nat. Phys. **5**, 1 (2009).

[20] T. D. Ladd, Nature **464** (2010).

[21] R. Blume-Kohout, New J. Phys. **12** (2010).

[22] A. I. Lvovsky and M. G. Raymert, Rev. Mod. Phys. **81** (2009).

[23] V. Giovannetti, S. Lloyd, L. MacCone y V. Giovannetti, S. Lloyd and L. Maccone, Nat. Photonics **5**, 4 (2011).

[24] S. Lloyd, Science (80-.). **273** (1996).

[25] K. Banaszek, M. Cramer y D. Gross, New J. Phys. **15** (2013).

[26] C. W. Helstrom, Phys. Lett. A **25**, 2 (1967).

[27] A. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (Edizioni della Normale, 2011), primera edición.

[28] E. Bagan, M. A. Ballester, R. D. Gill, R. Muñoz-Tapia y O. Romero-Isart, Phys. Rev. Lett. **97**, 13 (sep 2006).

[29] R. D. Gill y S. Massar, Phys. Rev. A **61**, 4 (2000).

[30] M. A. Nielsen and I.L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, England, 2010).

[31] M. Hübner, Phys. Lett. A **163**, 4 (1992).

[32] D. H. Mahler, L. A. Rozema, A. Darabi, C. Ferrie, R. Blume-Kohout y A. M. Steinberg, Phys. Rev. Lett. **111**, 18 (oct 2013).

[33] D. Mahler, L. Rozema, A. Darabi, J. Combes, C. Ferrie, R. Blume-Kohout y A. M. Steinberg, en *Front. Opt. 2012/Laser Sci. XXVIII* (Optical Society of America, 2012), pág. FTh4B.3.

[34] S. Straupe, JETP Lett. **104**, 7 (2016).

[35] Z. Hou, H. Zhu, G. Xiang, C.-F. Li y G.-C. Guo, npj Quantum Inf. **2**, 16001 (2016).

[36] D. G. Fischer, S. H. Kienle y M. Freyberger, Phys. Rev. A **61**, 3 (2000).

[37] A. Kalev y I. Hen, New J. Phys. **17**, 9 (2015).

[38] T. Hannemann, D. Reiss, C. Balzer, W. Neuhauser, P. E. Toschek y C. Wunderlich, Phys. Rev. A - At. Mol. Opt. Phys. **65**, 5 (2002).

[39] C. Granade, C. Ferrie y S. T. Flammia, New J. Phys. **19**, 11 (2017).

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

[41] C. Radhakrishna Rao, Bull. Calcutta Math. Soc. **37** (1945).

[42] H. Cramér, *Mathematical Methods of Statistics.*, PMS-9 / Princeton Landmarks in Mathematics (Princeton University Press, 1946).

[43] Vlatko Vedral, *Introduction to Quantum Information Science* (Oxford Graduate Texts, 2006).

[44] S. Pirandola, J. Eisert, C. Weedbrook, A. Furusawa y S. Braunstein, Nat. Photonics **9** (2015).

[45] A. K. Ekert, Phys. Rev. Lett. **67**, 6 (aug 1991).

[46] E. S. Gómez, S. Gómez, P. González, G. Cañas, J. F. Barra, A. Delgado, G. B. Xavier, A. Cabello, M. Kleinmann, T. Vértesi y G. Lima, Phys. Rev. Lett. **117**, 26 (2016).

[47] O. Gühne y G. Tóth, Phys. Rep. **474**, 1 (2009).

[48] J. Szangolies, H. Kampermann y D. Bruß, New J. Phys. **17**, 11 (2015).

[49] S. M. Barnett y S. Croke, Adv. Opt. Photon. **1**, 2 (apr 2009).

[50] S. Massar, Phys. Rev. A **62**, 4 (2000).

[51] Z. Huangjun, *Quantum State Estimation and Symmetric Informationally Complete POMs*, Tesis Doctoral (2012).

[52] M. Grassl, Elec. Notes Discret. Math **20**, 151 (2005).

[53] S. N. Filippov and V. I. Man'ko, Phys. Scr. **2011**, T143 (2011).

[54] N. K. Langford, New J. Phys. **15**, 3 (2013).

[55] W. K. Wooters and W. H. Zurek, Nature **299** (1982).

[56] D. F. V. James, P. G. Kwiat, W. J. Munro, A. G. White y D. F. V. James, P. G. Kwiat, W. J. Munro and A. G. White, Phys. Rev. A **64**, 5 (oct 2001).

[57] I. D. Ivanović, J. Phys. A Math. Gen. **14**, 3241 (1981).

[58] A. Klappenecker and M. Rötteler, en *Finite Fields Appl. 7th Int. Conf. Fq7, Toulouse, Fr. May 5-9, 2003. Revis. Pap.*, editado por A. P. G. L. Mullen y

H. .Stichtenoth (Springer Berlin Heidelberg, Berlin, Heidelberg, 2004), págs. 137–144.

[59] D. M. Appleby, J. Math. Phys. **46**, 52107 (2005).

[60] J. M. Renes, R. Blume-Kohout, A. J. Scott and C. M. Caves, J. Math. Phys. **45**, 2171 (2004).

[61] G. Zauner, *Quantendesigns: Grundzüge einer nichtkommutativen Designtheorie*, Tesis Doctoral, Universität Wien (1999).

[62] Z. Hradil, Phys. Rev. A **55**, 3 (mar 1997).

[63] M. S. Kaznady y D. F. James, Phys. Rev. A **79**, 2 (2009).

[64] J. Shang, Z. Zhang y H. K. Ng, Phys. Rev. A **95**, 6 (2017).

[65] J. Řeháček, B. G. Englert y D. Kaszlikowski, Phys. Rev. A **70**, 5 A (2004).

[66] C. Ferrie, Phys. Rev. Lett. **113**, 19 (2014).

[67] D. Goyeneche, G. Cañas, S. Etcheverry, E. S. Gómez, G. B. Xavier, G. Lima y A. Delgado, Phys. Rev. Lett. **115**, 9 (2015).

[68] X. Ma, T. Jackson, H. Zhou, J. Chen, D. Lu, M. D. Mazurek, K. A. Fisher, X. Peng, D. Kribs, K. J. Resch, Z. Ji, B. Zeng y R. Laflamme, Phys. Rev. A **93**, 3 (2016).

[69] Robert J. Chapman, Christopher Ferrie and Alberto Peruzzo, Phys. Rev. Lett **117**, 40402 (2016).

[70] K. S. Kravtsov, S. S. Straupe, I. V. Radchenko, N. M. T. Houlsby, F. Huszár y S. P. Kulik, Phys. Rev. A **87**, 6 (2013).

[71] N. Bent, H. Qassim, A. A. Tahir, D. Sych, G. Leuchs, L. L. Sánchez-Soto, E. Karimi y R. W. Boyd, Phys. Rev. X **5**, 4 (2015).

[72] R. Okamoto, M. Iefuji, S. Oyama, K. Yamagata, H. Imai, A. Fujiwara y S. Takeuchi, Phys. Rev. Lett. **109**, 13 (2012).

[73] E. Bolduc, G. C. Knee, E. M. Gauger y J. Leach, npj Quantum Inf. **3**, 1 (2017).

[74] K. Zyczkowski, K. A. Penson, I. Nechita y B. Collins, J. Math. Phys. **52**, 6 (2011).

[75] H.-U. Sommers y K. Zyczkowski, J. Phys. A Math. Gen **37**, 8457 (2004).