NON-HOLONOMIC TOMOGRAPHY: A METHOD FOR ASSESSING VARIOUS STATE-PREPARATION AND MEASUREMENT CORRELATIONS

by

CHRISTOPHER S. JACKSON

A DISSERTATION

Presented to the Department of Physics
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

March 2017

DISSERTATION APPROVAL PAGE

Student: Christopher S. Jackson

Title: Non-Holonomic Tomography: A Method for Assessing Various State-Preparation and Measurement Correlations

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Physics by:

Michael Raymer	Chair
Steven J. van Enk	Advisor
Brian Smith	Member
Tim Cohen	Member
Xiaodi Wu	Institutional Representative

and

Scott L. Pratt	Dean of the Graduate School

Original approval signatures are on file with the University of Oregon Graduate School.

Degree awarded March 2017

DISSERTATION ABSTRACT

Christopher S. Jackson

Doctor of Philosophy

Department of Physics

March 2017

Title: Non-Holonomic Tomography: A Method for Assessing Various State-Preparation and Measurement Correlations

The following dissertation investigates a problem related to the practice of quantum tomography, where one usually estimates the parameters associated with either quantum states or measurements. In particular, the question answered is whether and how one could detect if states and measurements are correlated. A similar question answered is how one could detect state-state measurement-measurement correlations in multiqudit systems. The solution involves an analysis of certain matrix quantities called *partial determinants*. A partial determinants is an application of the Born rule that can be interpreted as tomography over a closed loop in the space of state and measurement settings. From this perspective, the notion of state and observable become *non-holonomic* — that is, state and observable parameters can be defined "locally" over each setting but not globally over all settings. As such, state and measurement parameters are not estimated because such estimated values do not exist in correlated systems. Rather the inability to estimate such values is quantified. Partial determinants are a measure of the amount of contradiction that would result from any claim of the existence of such estimated values, by propagating these estimates through a 'tomography loop' of data collected by various experiments. Such measures of contradiction are generally known as *holonomies*.

This dissertation includes previously published co-authored material.

CURRICULUM VITAE

NAME OF AUTHOR: Christopher S. Jackson

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:
University of Oregon, Eugene
University of Hawaii at Manoa, Honolulu

DEGREES AWARDED:
Doctor of Philosophy, Physics, 2017, University of Oregon
Bachelor of Science, Physics and Mathematics with distinction, 2011,
University of Hawaii at Manoa

AREAS OF SPECIAL INTEREST:
Quantum Information Theory
Condensed Matter Theory
Quantum Field Theory
Mathematical Physics
Geometry

PROFESSIONAL EXPERIENCE:

Graduate Teaching Assistant, University of Oregon, 2011-2017

Graduate Research Assistant, University of Oregon, 2015

GRANTS, AWARDS AND HONORS:

Science Literacy Teaching Fellowship, Quantum Mechanics for Everyone,
University of Oregon, 2015

Weiser Senior Teaching Assistant Award, Introduction to Physics with Calculus,
University of Oregon, 2014

Science Literacy Teaching Fellowship, Quantum Mechanics for Everyone,
University of Oregon, 2013

Department of Physics and Astronomy Achievement Scholarship,
University of Hawaii at Manoa, 2007-2010

Dean's List,
University of Hawaii at Manoa, 2006-2008

Certificate of Outstanding Achievement in Chemistry,
Hilo High School, 2005

PUBLICATIONS:

Christopher Jackson and S. J. van Enk, "Non-holonomic tomography II: Detecting correlations in multiqudit systems", arXiv:1702.06090 (2017.)

Christopher Jackson and S. J. van Enk, "Non-holonomic tomography I: The Born rule as a connection between experiments", arXiv:1702.00118 (2017.)

Christopher Jackson and S. J. van Enk, "Detecting correlated errors in state-preparation-and-measurement tomography", Physical Review A 92, 042312 (2015.)

ACKNOWLEDGEMENTS

To all of those who have helped me learn how to be myself.

May each of us find meaning and create something beautiful with this precious life.

To my mother, SC, and the community I've lived with and consider family.

To my professors and teachers:

especially Steven van Enk, Michael Raymer, John Toner, Davison Soper, Jim Isenberg, Arkady Vaintrob, John Learned, William Lampe, Robert Morse, Chester Vause, Luis Elias, Helen Nakamura, Suzanne Cama, and Harrilyn Chun-Akana.

To my friends and classmates:

Tyler Harvey, Dileep Reddy, Jonathan Mackrory, SP, George deCoster, Gabriel Barello, Jordan Pierce, Chris Newby, Kevin Croker, Sean Sanford, and many others.

Without any of you, a piece of my understanding of the nature of reality would be missing.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

The hallmark of quantum theory is a fundamental distinction [2] that is made between the *state* of a system and the *measurement* of that system, a distinction which is not present in classical theories. The paradigm of quantum theory is that these two distinct notions of state and measurement are not directly observable, but rather what is observed in the lab is understood to be the product of a relationship between the state and the measurement. Conceptually, one can imagine this with a proverb: Things (states) do not exist (become observed) without having properties (measurements) and properties do not exist without things to have them.

However, the quantum situation is still a little more subtle, because what is observed in the lab is fundamentally understood to be a discrete number of events (whether they're the positions and frequencies of light signals, vibrations of a thermal material, collisions in electric materials, excitations of Standard Model fields, etc.) These discrete events or quanta are considered to be generally random from event to event, and it is the distribution of these random events that is understood to be a product of the state and measurement, known as the Born rule. With this in mind, quantum theory suggests another paradigm: that the *actual* outcomes which appear in the lab are of a different nature than the possible[1] outcomes represented by the quantum state-measurement duality.

By now, this paradigm has evolved into the field of quantum information, a bonafide effort to understand how such "possible events" can create, send, and process information in ways that "actual events" fundamentally cannot.[2] Of particular interest to this thesis is the original motivation for non-holonomic tomography, from the effort to build a fault-tolerant quantum computer. To build a fault-tolerant computer, the error rate of a single operation or gate must be low enough so that the assembly of many such gates may actually result in a desired computation. [3, 4] The ability to perform such a computation undergoes a first-order transition as a function of the error rate (going to zero above a certain threshold under current error correction protocols.) [5–9] There are various

[1] This statement might be more traditionally phrased by saying quantum states can be in a superposition of possible outcomes.

[2] The more standard terms for "possible events" and "actual events" would be "quantum states" and "classical states", respectively.

estimates and predictions for what the value of the threshold is, but it is certainly important that measured error rates be below the threshold.

Currently, there are experiments which claim to be near the fault-tolerant threshold for a quantum computer. [10] The error rate for quantum gates could be quantified by the average gate fidelity, a measure of the distance between the target gate and the actual gate (where 1 corresponds to no error.) However, another issue has come up, which is that contributions to measured gate fidelities can be predominantly due to errors in the state-preparations and measurements (SPAM) used to probe such gates (rather than to errors in the gate.) [10–12] Considering there are such significant SPAM errors, this thesis addresses the concern of how to assess if such SPAM errors are correlated.

Such correlated SPAM errors can be quantified by a technique that we gave the name of "non-holonomic tomography", which can be considered a generalized application of the Born rule. Quantum (state) tomography originally referred to the estimation of an unknown quantum state using the projected information of the probability distributions resulting from a sufficient variety of known measurements. This concept quickly gave rise to the ideas of quantum detector tomography, where probabilities are used to determine the outcomes of unknown measurement devices from a sufficient variety of known states, and quantum process tomography, where probabilities are used to determine an unknown operation from a sufficient variety of known states and measurements. These applications of the Born rule may be summarized as the interpretation of probabilities as a map between state and measurement parameters.

To explain the main idea of non-holonomic tomography, let us consider a toy analog. Consider a device with various settings, a, each of which prepare a different signal on demand by the press of a button. Consider also a detector with various settings, i, each of which detect a particular property of the signal indicated by the blink of a light. Now suppose it is suspected that each setting of the preparation device actually produces the same signal, only that each setting produces the signal with varying probabilities of success, p_a. Suppose further that the detection device is expected to simply indicate the presence of the signal, only that each setting can register a signal with varying probabilities of success, w^i. We then imagine that p_a and w^i are actually unknown and that we are only able to change settings and record whether the light blinks or not.

FIGURE 1. On the left is a device which prepares various signals on demand depending on which button, a, is pressed. On the right is a device which blinks to indicate a signal with a certain property depending on which setting, i, a dial is turned to.

Let us indicate by ${f_a}^i$ the measured frequency with which the light blinks when the devices are set to (a, i). If one can assume that the performance of the devices and their settings are uncorrelated, then one can simply identify (after many runs of the experiment)

$$ {f_a}^i = p_a w^i. \tag{1.1} $$

Relaxing this assumption to allow for the possibility of correlations, one can make the more general identification that

$$ {f_a}^i = {\langle pw \rangle_a}^i \tag{1.2} $$

where we have introduced the notation ${\langle\rangle_a}^i$ to represent the average over the ensemble of trials for the pair of settings (a, i). The subtlety here is that the devices can still be represented by single parameters, p and w, only now these parameters are to be understood as random variables which fluctuate depending on the setting (a, i).

The presence of SPAM correlation is simply when the frequencies, ${\langle pw \rangle_a}^i$ (which are what we have access to) are such that

$$ {\langle pw \rangle_a}^i \neq \langle p \rangle_a \langle w \rangle^i. \tag{1.3} $$

It would seem that to identify such a circumstance one would have to measure $\langle p \rangle_a$ and $\langle w \rangle^i$ individually. However, such measurements would require devices which are already well characterized, unlike the devices we have. Nevertheless, one can still detect for such correlations in this setting. To illustrate this, one will only need to consider two settings per device, so let us denote them as p and q for the states and w and v for the measurements instead of $a = 1, 2$ and $i = 1, 2$.

To detect such correlations, one considers the settings (p, w), (q, w), (p, v), and (q, v) as individual experiments. Each individual experiment is *effectively uncorrelated* because we can always choose $\langle p \rangle$ and $\langle w \rangle$ such that $\langle pw \rangle = \langle p \rangle \langle w \rangle$. The freedom of that choice is a *gauge* degree of

freedom and is further a *local* one because each experiment has this property. However, if there are correlations there will not exist a choice for each of these gauges that is consistent with each other. The inability to make this choice can be quantified by what we call a partial determinant which has the interpretation of representing tomography being performed in a loop, as illustrated in Figure 2.

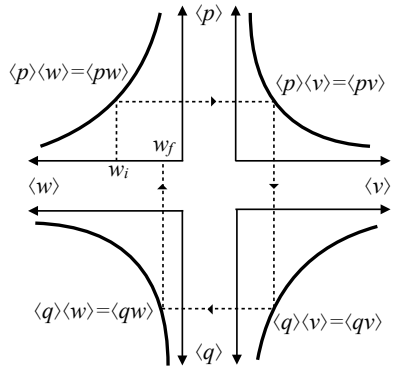


FIGURE 2. An illustration of the partial determinant as a holonomy. Each experiment (p, w) has a local gauge degree of freedom because it is effectively SPAM uncorrelated, $\langle p\rangle\langle w\rangle = \langle pw\rangle$. The data $\langle pw\rangle$ further provides a connection between adjacent gauge degrees of freedom by the assumption that they share independent settings. Such a connection defines a non-holonomic constraint when $w_f = \frac{\langle pv\rangle\langle qw\rangle}{\langle pw\rangle\langle qv\rangle} w_i \neq w_i$. A particular w_i fixes the gauge which can either represent an arbitrary choice or some external information. The PD $\Delta = \frac{w_f}{w_i}$ is gauge invariant.

The term "non-holonomic" originally comes from Lagrangian mechanics, in reference to certain constraints between coordinates and velocities that cannot be simplified to constraints between just coordinates. Such constraints are most abstractly represented by inexact one-forms which are the platform for thermodynamic theories. Non-holonomic constraints more generally arise in parallel transport as connections between tangent vectors or gauge parameters which have non-zero integrals or holonomies over closed loops.

In non-holonomic quantum tomography, where probabilities are estimated from unknown states and unknown measurements, there are gauge degrees of freedom associated with the fact that there is always a continuum of state-measurement parameters which are consistent with such estimated probabilities. These probabilities can more generally be interpreted as connections between these gauge parameters for experiments which share common device settings. In the presence of SPAM correlations, holonomies correspond to the contradictions which inevitably arise in any attempt to fix these gauge parameters in a way that is consistent among the various experiments one can perform.

This work was developed over the course of three publications[13–15], co-authored with Steven van Enk. These are, respectively, chapters III, IV, and V of this disseration. Chapter II is a partial review of the field of SPAM tomography and the fundamental perspectives behind it. Chapter III introduces the partial determinant and its application to detective correlated SPAM errors. Chapter IV presents a fully developed perspective behind the technique of using partial determinants, making precise analogies with other techniques throughout theoretical physics. Chapter V demonstrates how one can apply the technique of partial determinants to analyze correlated errors in multiqudit systems.

CHAPTER II

THE MODEL OF QUANTUM THEORY

By now, quantum theory has developed quite the reputation for being strange. All of this strangeness is ultimately due to a **fundamental distinction** [2] the theory makes between **three primary notions** — of *state*, *observables*, and *measured data* — which are related by an equation often referred to as the Born rule. The notion of *measured data* is an acknowledgement that physical nature is the articulation of a countable set of discrete events — i.e. quanta. Each event (or quantum) is further understood to be a random expression of only an aspect of a *state*, a notion suggesting the presence of a subtler circumstance which preceded the articulation (or measurement event.) Finally the state is conceived of through a holistic structure of alternative *observables*, where the notion is that each observable represents a set of possible outcomes which are distinct among themselves but are not distinct relative to the possible outcomes represented by other observables.

Although present in the original axioms, this fundamental distinction between these three notions often still remains conceptually obscured to both professionals and non-professionals. There are two reasons for this. The dominant reason is of course because these distinctions are subtle — so much so that they are irrelevant to the experiences and practices of most individuals. Nevertheless, for the individual who is concerned, there is the obstacle that these notions are almost always entirely represented with mathematics, requiring yet another layer of proficiency and integration. Colloquial references to the distinction of these notions exist, the most popular perhaps being the so called wave-particle duality. [16] This kind of language can be problematic in the absence of mathematics as it can suggest a notion of confused realism rather than a statement of this fundamental distinction between the three notions which are themselves the foundation of the model of quantum theory.

However, over the last several decades, this fundamental distinction is becoming distilled into a more direct and lucid picture of quantum phenomena. This philosophical "distillation" is certainly reflected in the emergence of quantum information as a field of research. As indicated by the name, quantum information owes many of its perspectives to concepts which were developed in information theory, computer science, communication theory, and statistics. Associated with quantum information are its many applications including quantum communication, quantum key

distribution, and quantum computation as well as theoretical perspectives in quantum foundations, quantum condensed matter, high energy physics, and possibly quantum gravity.

The work of this dissertation can be summarized as the beginning of a different way to use the Born rule by viewing it as simply a *constraint* between these three fundamental notions of state, observables, and measured data. What follows from this perspective is a technique for detecting correlations between state and observable parameters, even though state and observable parameters are themselves not directly measurable quantities. This technique centers around a mathematical object called a *partial determinant* which is a kind of holonomy or loop that measures contradictions. This holonomy in turn shows that the Born rule can be understood as measured data constraining state and observable parameters much in the same way that the average energy of a thermodynamic system constrains heat and work or the gauge fields of QED constrain the electron kinetic momentum and the photon vector potential.

The core of this work comes from three papers: one recently published [13] and two still under review [14, 15]. Chapter III presents what was originally called the *partial determinant* and applies it to a few state-observable correlated models. Chapter IV explains the theoretical interpretation of the partial determinant, including an explanation of the term *non-holonomic*, by demonstrating precise analogies with theoretical techniques from other fundamental physics models. Chapter V generalizes the technique to consider correlations in multi-partite systems. We begin this thesis with a review of the fundamentals of quantum theory. Such a review will serve as an opportunity to emphasize the subtleties within the standard language around the mathematics which will make explicit the subtleties that are key to an understanding of the non-holonomic perspective.

To do this, we will review quantum theory axiomatically in three iterations: the first being through the set of axioms most similar to their original form from the development of quantum mechanics, the second using the language which has become the current standard throughout quantum information, and the third emphasizing the important role of measurable data from the perspective of quantum tomography. From this third iteration, we will segue into a discussion about quantum estimation in practice and its associated perspectives and applications. Finally, we will discuss several potential applications of non-holonomic tomography.

Quantum Mechanics and Quantum Information

Original Axioms of Quantum Mechanics

The original axioms from the first half of the 20th century were designed with a very specific objective in mind: [17–19] to define a mathematical structure that reflected the quantum aspect of nature for the purpose of supporting mechanics.[1] All of the success in the abstraction of mechanical theories certainly served as a guiding principle in the birth of quantum mechanics. However, the representation of mechanics in a quantum theory should also be properly distinguished from the quantum setting itself. Mathematically, this is reflected in the fact that operators over Hilbert spaces can represent the *same* dynamical algebra as functions over (symplectic) phase spaces. As such, one must appreciate that the axioms are an attempt to describe the basic structure of a quantum setting, upon which dynamics can be supported. Put more simply, one must appreciate that what makes quantum mechanics "quantum" is much more fundamental than mechanics itself.

Historically, these axioms of quantum mechanics are also the culmination of an evolving correspondence principle [17–20] starting with Bohr (1913) and reaching the form to be presented here in 1927 at (or around) the famous Solvay conference. [21] Also important are several refinements in the formalism and interpretation of the theory [22–25] which continued to emerge since. To repeat ourselves one last time, there is actually nothing "mechanical" about these axioms. Nevertheless, it is convenient to use the word "mechanics" to emphasize the time period. To emphasize their fundamental notions, let us refer to these axioms as the axioms of *state*, *observables*, and *measured data*.

Axiom 1-State: States are represented as vectors in a complex Hilbert space.

The Hilbert property refers to the presence of an inner product which further defines the property of adjointness. This inner product represents a generalized form of logic and probability where orthogonality corresponds to mutual exclusivity. Adjointness represents a natural isomorphism between vectors and their dual space, the role of which will only become apparent

[1] Jumping ahead for a moment, we are referring to the fact that vectors in a (complex) Hilbert space automatically gives rise to the notion of unitarity through which one can represent dynamics because unitary transformations are generated by Hermitian operators which can represent physical observables. The treatment of physical observables as generators of continuous evolution owes its inspiration to classical Hamiltonian mechanics, where the kinematics generated by certain forces in real space are generalized to the canonical transformations generated by observables over phase space. From the perspective of a phase space, such continuous transformations correspond to equations of motion that are first order so that observables naturally inherit the Lie algebra of vector fields.

with the other two axioms. Although this is how the axiom is usually stated, there are actually two additional features one must understand:

The first is that in general one needs to further require that the Hilbert space, $\mathcal{H}$, be so called separable. Separability corresponds to the property that a Hilbert space have a countable basis which further implies that all physical wavefunctions are continuous (almost everywhere.) The separable condition is equivalent to the ability to estimate an unknown quantum state with arbitrary precision from the countable collection of measured quantum events. Nevertheless, we will only consider finite, d-dimensional Hilbert spaces, $\mathcal{H}_d$, in which case every basis is automatically countable. Further, the qualification that the vector space be Hilbert is redundant for finite-dimensional space, but "Hilbert" has colloquially become a handy word to use when referring to the "quantum part" among possibly other linear structures being considered.

The second feature is that we only distinguish vectors modulo the equivalence relation

$$z\,|\psi\rangle \sim |\psi\rangle\,, \tag{2.1}$$

where $z \in \mathbb{C}^*$ is a nonzero complex number and $|\psi\rangle \in \mathcal{H}$ is a vector — that is, vectors which are collinear in the complex sense represent the same state. So then the space of states is technically not the Hilbert space itself but what is called the complex projective space $\mathbb{C}P^{d-1} = \mathcal{H}_d/\sim$ — projective because every point represents a complex line — and one should think of vectors as actually representing points in this space. Half of this redundancy in representation is usually taken care of by requiring that every state be normalized — $\langle\psi|\psi\rangle = 1$ — while the "overall phase" remains implicitly understood to be arbitrary.[2] On the other hand, this phase is automatically absent in the representation of states in quantum information, as we will soon see.

Axiom 2-Observables: Observables correspond to Hermitian operators.

The Hermiticity of an operator is meaningful in light of the Spectral Theorem, which tells us that every Hermitian (or self-adjoint) operator, $A = A^\dagger$, has a complete set of orthonormal eigenvectors, $\{|k\rangle\}_{k=1}^d$, with real eigenvalues, $\{a_k\}_{k=1}^d$. In other words, every observable is of the

[2]Of course, one must make sure not to overextend this arbitrariness, particularly when $\mathcal{H}$ is part of a composite space or when considering evolution generated by observables with external parameters which vary cyclically.

form

$$A = \sum_k a_k |k\rangle\langle k| \tag{2.2}$$

where $\langle k|l\rangle = \delta_{kl}$ and $\sum_k |k\rangle\langle k| = 1_{d\times d}$. Physically, quantum observables represent the most basic type of measurement[3]: The set of possible outcomes is represented by the basis, while the eigenvalues are the numbers assigned to each outcome.[4] This brings us to the final axiom.

Axiom 3-Measured Data: Upon measuring an observable with respect to a given state, only one eigenvalue of the observable will occur randomly with a probability determined by the Born rule.

The Born rule is the usual "amplitude modulus squared":

$$P(k|\psi) = |\langle k|\psi\rangle|^2 \tag{2.3}$$

where $P(k|\psi)$ is the probability of an outcome k, corresponding to an eigenvector $\langle k|$ of an observable A, given a system in state ψ, corresponding to another vector $|\psi\rangle$. In particular, this means that the expectation value for the measurement of A is

$$\langle a_k\rangle = \sum_k a_k P(k|\psi) = \langle\psi|A|\psi\rangle\,. \tag{2.4}$$

A very important point on notation must be made here. The eigenvalues $\{a_k\}$ define a single random variable, k, distributed according to the state, ψ, represented mathematically by a vector, $|\psi\rangle$. The expectation value associated with this random variable is classically represented by angular brackets, e.g. $\langle f(k)\rangle$. (Indeed, this is what motivated the still popular Dirac bra-ket notation.) Often this same quantity is denoted by $\langle A\rangle$ to represent the computation on the right-hand side of equation (2.4). However, this double meaning of the angular brackets is inappropriate in the context of the work being presented here, and will lead to unnecessary confusion. The operator A, usually considered, is a fixed object governed by equations of motion, not a random variable. On the other

[3]Quantum observables can also represent generators of continuous evolution as mentioned earlier. It is here in the structure of quantum observables that mechanical algebras can be represented by the commutator. Thus it is worth repeating that such an introduction of mechanics is manifestly distinct from the notion of measurement.

[4]Indeed, the orthonormal basis is the really important part. In general, one could still use complex eigenvalues to represent physically meaningful quantities. In this case, all that needs to be done is to replace Hermitian operators with so called normal operators, defined by the property that $[A, A^\dagger] = 0$.

hand, we *will* want to consider random *operators*, in which case expressions like $\langle A \rangle$ are used where the angular brackets will continue to denote the expectation of an (operator) random variable.

Generalized Axioms in Quantum Information

Quantum mechanics makes explicit the fundamental notions of state and observables which gives rise to randomness in (ideal) measured data while maintaining generally non-local coherences. Although understood by their founders, these notions for a long time had an element of metaphysical reality to them which were historically interpreted as a relic of the incompleteness of the theory. A large reason for this is perhaps because there had also been much more technical refining left to do for the mathematics of mechanics itself — namely in group representations, coherent states, path integration, renormalization, geometry, and topology. That is to say those features which make quantum theory so distinct could not be broadly appreciated during the succession of developments in mathematical physics and its applications. Nevertheless, by now these basic notions have become of explicit importance, most notably since Bell's inequality (1964) for local hidden variable theories, Bennett and Brassard's protocol (1984) for quantum key distribution, and especially Shor's algorithm (1994) for factoring primes in polynomial time.

It is from these works that the fields of quantum communication, quantum computation, and most broadly quantum information have emerged. Of course, very much credit in these fields is due to the progress made in computer science, mathematical statistics, and communication theory. It is often not appreciated how much these other fields have paralleled the development of quantum physics throughout the 20th century. An important example is the path integral which was actually first invented by Norbert Wiener (1921) in the context of (classical) martingales [26].

From all of this combined history, quantum information takes on a slightly different appearance from quantum mechanics as originally formulated. The basic notions are in fact the same, it is just that the perspective has become much broader. As such, we shall now reiterate the original axioms with the more modern formalism[5] To show that the basic notions are in fact the same, we will further derive these axioms from the original ones. Considering that the original axioms are simply a special case of these, one can be assured that they are indeed equivalent. What this generalization does in

[5]Of course, this formalism was actually systematized almost in its entirety by von Neumann in 1927 with POVMs being first considered mathematically by Gelfand and Neumark in 1943 [27] and introduced to physics by Helstrom in 1976 [28].

effect is anticipate (quite elegantly) less idealized circumstances, where there may be coupling to an environment in which case the distributions of a particular system of interest can have fundamental entropy. These axioms can be thought of as the combination of the original quantum axioms with classical probability theory.

Axiom 1′-State: Any density operator represents a state.

A density operator is Hermitian with positive eigenvalues ($\rho > 0$) and unit trace ($\mathrm{Tr}\rho = 1$) which can represent a state. Sometimes it is convenient to consider states that are subnormalized, so that $0 < \mathrm{Tr}\rho \leq 1$ and the trace is interpreted as the probability that ρ was "prepared successfully", or simply not normalized at all, e.g. $\rho = e^{-\beta H}$ for an equilibrium system with Hamiltonian H and temperature β^{-1}. One way to think of these operators is to imagine states which can be represented by vectors in Hilbert space but are prepared according to a distribution, $P(\psi)$. We can then write expectation values in a different way:

$$\langle a_k \rangle = \sum_k a_k P(k) = \sum_k a_k \int d\psi \, P(k|\psi) P(\psi) = \mathrm{Tr}\rho A, \tag{2.5}$$

where

$$\rho = \int d\psi \, P(\psi) |\psi\rangle\langle\psi|. \tag{2.6}$$

Expressions such as equation (2.6) are referred to as *mixtures*, a term adopted from probability theory. This kind of addition is often considered classical, as opposed to the quantum superpositions within Hilbert space. States of the form $|\psi\rangle\langle\psi|$ have the special property that they cannot be written as mixtures of other states and are thus referred to as *pure*.

Importantly, since ρ is Hermitian[6], there is a special unique mixture consisting of its (orthogonal) eigenstates, $\{|r\rangle\}$:

$$\rho = \sum_r p_r |r\rangle\langle r|. \tag{2.7}$$

The collection $\{|r\rangle\langle r|\}$ can also be defined as the measurement which minimizes the entropy of the corresponding distribution for ρ.[7] When calculating a rate or cross-section, this mixture more often

[6] Although such operators could technically represent physical observables, it is important to understand that such objects play quite different roles as states.

[7] This value of the minimum entropy is the familiar von Neumann entropy, $-\mathrm{Tr}\rho \log \rho$.

appears under the guise of an "average over initial states." This kind of mixture also suggests a more quantum interpretation of these general mixed states as they can correspond to the entanglement of a pure composite state,

$$\rho = \mathrm{Tr}_B |\Psi\rangle\langle\Psi| \tag{2.8}$$

where, for some states $\{|r'\rangle\}$ in an unconsidered degree of freedom, B,

$$|\Psi\rangle = \sum_r \sqrt{p_r}\, |r\rangle_A \otimes |r'\rangle_B \,. \tag{2.9}$$

The partial trace operation in equation (2.8) may be more familiar as a "sum over final states."

Every state that is in the (tensor) product of two d-dimensional Hilbert spaces can be uniquely written in the form of equation (2.9) by the singular value decomposition theorem. The partial trace is a surjective map which defines what is referred to as the Hilbert-Schmidt bundle. Importantly, this is a method through which one can define a measure for the space of mixed state $d\rho$ as induced by the measure for pure states $d\Psi$.[8]

Axiom 2′-Observables: Any POVM represents a measurement.

A POVM or *positive operator-valued measure* is a set of positive operators $\{E_n\}$ such that $\sum_n E_n = 1$. One reason why a POVM can represent a measurement is because one can consider indirectly measuring a state by letting it interact with another system and measuring that system in the standard way. Let us suppose that the system to be measured (called a pointer) can be prepared in a pure state, $|0\rangle$ (such as the vacuum state of a mode.) Let U be the unitary operator representing the interaction between the pointer and the system of interest.[9] After the interaction, $\rho_A \otimes |0\rangle\langle 0|_B \longrightarrow U\rho_A \otimes |0\rangle\langle 0|_B\, U^\dagger$, the probability to find the pointer in a state $|n\rangle$ is

$$P(n) = \mathrm{Tr}_{AB}\Big(\big(U\rho_A \otimes |0\rangle\langle 0|_B\, U^\dagger\big)\big(1_A \otimes |n\rangle\langle n|_B\big)\Big) = \mathrm{Tr}_A(\rho_A K_n^\dagger K_n) \tag{2.10}$$

[8]This fibre bundle is also the origin of the strange formula for the popular state fidelity also known as the Bures distance, $\mathrm{Tr}\sqrt{\rho}\sqrt{\sigma} = \mathrm{Tr}\sqrt{\sqrt{\sigma}\rho\sqrt{\sigma}}$.

[9]Unitary evolutions, $U^\dagger = U^{-1}$, are uniquely defined as the continuous processes which preserve the Hilbert space's inner product. In particular, this means that outcomes which are mutually exclusive initially remain mutually exclusive under unitary evolution and so such evolutions are reversible. Further, this implies that such processes U are always generated by an observable A as determined by a Schrödinger equation, $U'(p) = -iA(p)U(p)$, where p is the classical/external/massive/slow parameter mechanically conjugate to the observable, A.

where the $K_n = \langle n| U |0\rangle$ are called Kraus operators and $\sum_n K_n^\dagger K_n = 1_A$ from unitarity. Specifically, this means each K_n can be *any* operator (not necessarily Hermitian or idempotent) so that $E_n = K_n^\dagger K_n$ can be any positive operator (with small enough eigenvalues.)

Further, this implies an intuitive notion that, in general, measurement outcomes can have certain amounts and types of ambiguity. To get a basic idea of this, consider the spectrum of an element $E = \sum_k e_k |k\rangle\langle k|$ in a two-outcome POVM, $\{E, \bar{E}\}$ (where $\bar{E} = 1 - E$.) One should observe that

$$\mathrm{Tr} E\bar{E} = \sum_k e_k(1 - e_k), \tag{2.11}$$

which is zero if and only if all the $e_k = 0, 1$ or equivalently if E is *idempotent*, $E^2 = E$. This has the interpretation that E and $\bar{E}$ may not correspond to mutually exclusive events. However for a two-outcome measurement, this kind of ambiguity can always be interpreted classically (because the commutator $[E, \bar{E}] = 0$) while for more outcomes these ambiguities become much subtler as different outcomes may not commute. POVMs where every element is idempotent are called *projective*. Further if the elements all have pairwise products equal to zero, then they are called *orthogonal projectors* and become equivalent to the usual measurements associated with an observable, including degeneracies. In particular, if each projector is of rank 1 (so that there are no degeneracies) the POVM is called a *von Neumann* measurement.

Axiom 3′-Measured Data: The Born rule is $p = \mathrm{Tr}\rho E$.

As is already evident from the previous discussion, the statistics of the measured data, the density operator of the state, and the POVM of the measurement are constrained by the Born rule: The probability p_n of an outcome n, represented by a POVM element E_n, given a state, represented by density operator ρ, is

$$p_n = \mathrm{Tr}\rho E_n. \tag{2.12}$$

It will be convenient to also refer to equations like (2.5) as the Born rule. What is really important to understand for the purpose of this thesis is that the trace is simply another inner product, called the Hilbert-Schmidt inner product, over the d^2-dimensional space of Hermitian operators over a d-dimensional Hilbert space. Most if not all of what is strange in quantum systems boils down to the simple fact that $d < d^2$ — that is, that the maximum number of unambiguous measurement

outcomes is strictly less than the number of degrees of freedom. This brings us to the topic of quantum tomography.

Quantum Tomography

The axioms of quantum theory were originally constructed to support mechanics. By the turn of the century, physicists found themselves with a handful of mechanical theories which were already refined to a high level of abstraction — from a particularly broad perspective, these theories would have been classified during the turn of the 20th century as continuum (fields), analytical (Lagrangian/Hamiltonian), and statistical mechanics. As such, the original intention of quantum theory has been to correctly *predict* the distributions of measurable quantum events, where states and observables are known from macroscopic observations and the appropriate application of the classical mechanicses. This is only one (albeit profound) application of the Born rule.

Unknown States: Part I

Although $d^2 > d$, it is still a finite number which means we can determine an unknown state to arbitrary precision by considering the statistics over various measurement outcomes. The prototypical example of this is the spin-1/2 ($d = 2$) system, similar in setup to the Stern-Gerlach experiment. In this case, we imagine silver ions which are being produced by an uncharacterized source, modeled by ρ, which may exhibit spin polarization. To begin determining the spin state, we might start by measuring the z-component represented by the Pauli observable $\sigma_z = |+\hat{z}\rangle\langle+\hat{z}| - |-\hat{z}\rangle\langle-\hat{z}|$ with corresponding POVM $\{|+\hat{z}\rangle\langle+\hat{z}|, |-\hat{z}\rangle\langle-\hat{z}|\}$. After the first quantum outcome, say it is "up", there is only one definite statement that we can make about the source: that it does not consistently prepare $|-\hat{z}\rangle$. After recording an arbitrarily large number of outcomes one has an estimate for the probability, p_z, of the "up" outcome, which according to the Born rule puts a constraint on the source, $2p_z - 1 = \mathrm{Tr}\rho\sigma_z$.

Unless we were so lucky to have estimated $p_z = 0, 1$ we do not have a prediction for the distribution of outcomes for any other (non-commuting) observable, e.g. σ_x. If we also estimate p_x, we have a second constraint, $2p_x - 1 = \mathrm{Tr}\rho\sigma_x$, as well as a prediction for all observables representing spin components within the zx-plane. However, there is still another observable degree of freedom for which we do not yet have a prediction, so we estimate, say p_y for σ_y. Finally we have an estimate

of the state[10] of the silver ion spins (conditioned on their detection) which we may write as:

$$\rho = \frac{1}{2}\sigma_o + (p_x - \frac{1}{2})\sigma_x + (p_y - \frac{1}{2})\sigma_y + (p_z - \frac{1}{2})\sigma_z \tag{2.13}$$

where σ_o is the identity and $\{\sigma_\mu\}_{\mu=0}^{d^2-1}$ is referred to as the Pauli (operator) basis.

Though not as typical, it will be very useful to consider the coefficient of σ_o as a degree of freedom as well, which has a simple interpretation: For the Stern-Gerlach example, this coefficient would be $\frac{m}{2M}$ where M is the mass lost by the source and m is the mass gained by the screens (while the $1/2$ is there to cancel the trace, $\mathrm{Tr}\sigma_o = 2$.) In other words, the state can be conditioned on some other indication that a state has been prepared which may not trigger a detection.[11]

This basic method of estimating a state, where a number of constraints are produced to uniquely determine the quantum degrees of freedom, is referred to as quantum (state) tomography. The word tomography is used because the measurement outcomes for each observable are only projected information about the state so that one must measure the state from various "angles" in operator space. This is also referred to as linear inversion because the Born rule gives us constraints that are linear in the state parameters. Quantum state tomography has become a standard tool in modern labs. [29–33] Other important forms of state estimation exist which we will discuss soon enough. However, let us first discuss some fundamental points-of-view that tomography has to offer.

Why Quantum?: Tomographic Axioms

Though the author has tried to supplement them with as much insight as he has, the original and generalized axioms are still very formal and can be rather opaque at explaining the principles that they represent. Of particular concern are two questions:

- Why do we use Hilbert spaces?
- Why do we use complex numbers?

Put another way, in what sense do we need to use the quantum theory that we have?

[10] We often simply say "perform *a* measurement on *the* state" when we should really say "perform sufficiently many repetitions of a measurement on the states being prepared by a source, resulting in the estimate of an average state."

[11] Indeed, this could be extended to a normalized state and a 3-outcome POVM over a larger Hilbert space; but this would introduce even more observable degrees of freedom which are beyond the capacity of the originally considered measurement devices.

Fortunately, there are a set of axioms thanks to Hardy [34] which are quite simple and provide an answer to these basic questions. They are what this author would call the most *tomographic* of the various new axioms out there. Other sets specifically rely instead on notions of information processes such as purification and distillation. On the other hand, this tomographic set of axioms is more attractive from a subjectivists perspective, where the model being constructed should clearly reflect the knowledge and judgements of the practitioner. [35]

What Hardy demonstrates is that a quantum theory is just the next simplest theory in a set of more general probability theories. This generalization in the notion of probability is due to a lucid, though subtle, distinction between two basic numbers:

- K: the minimum number of measurements required to determine a state, called the degree (as in the number of degrees of freedom)
- d: the maximum number of mutually exclusive measurement outcomes, called the dimension (foreshadowing the Hilbert space dimension.)

The axioms then proceed to declare the nature of these two numbers in relation to each other.

1-Probability: Observed frequencies have a limit *given* a state and a measurement.

This is the assertion that there exist mathematical representations, ρ and $\{E_k\}$, for the state and measurement, along with a map $f : (\rho, E_k) \mapsto p_k$ where the $\{p_k\}$ form a distribution which is the limit of the observed frequencies. By the definition of K, ρ can be represented as a list of K real numbers corresponding to the probabilities of some outcomes over possibly multiple measurements. Importantly, these lists are not generally distributions. Considering two such lists, it makes sense to mix states.

To any fixed outcome, E, it is useful to define a corresponding map, e, such that $e(\rho) = f(\rho, E)$. We insist that measurements respect state mixing so that the probability for a mixture of states is the same mixture of probabilities for each state — that is,

$$\lambda e(\rho) + (1 - \lambda)e(\sigma) = e\big(\lambda\rho + (1 - \lambda)\sigma\big) \tag{2.14}$$

for any $\lambda \in [0,1]$. This uniquely restricts all possible e to be linear functions over the K degrees of freedom. Further, it makes sense to mix outcomes between different measurements so that f must be bilinear.

2-Simplicity: K is a function of d that is the smallest consistent with the axioms.

Any function of the distributions over such d outcomes should be symmetric in their probabilities. In other words, "which outcome is which" is considered an arbitrary distinction that one could always change by permutation of their labels.[12] So the form of K is already quite restricted to be

$$K = \sum_{k=0}^{d} \chi_k \binom{d}{k} \tag{2.15}$$

where the χ_k are non-negative integers called the signature of the theory. One can think of these χ_k as enumerating the multiplicity of more general types of coherence.

3-Subsystems: Restricting to $c < d$ outcomes gives a system of dimension c.
4-Composite Systems: $K(d_A d_B) = K(d_A)K(d_B)$.

These two axioms are perhaps better considered as one axiom which we might call the Arithmetic axiom because it is a statement of how K should behave under the "addition" and "multiplication" of state spaces. The first simply asserts that K must be a strictly increasing function of d — in other words, every outcome is understood to have some "size." The second asserts that K should also reflect the multiplicity of considering multiple systems simultaneously. This restricts K to be an integer power[13] of d, $K = d^q$. It is a simple exercise to show that the only viable signatures are then

$$\chi = (1, 0, \ldots)\ , \quad (1, 2, 0, \ldots)\ , \quad (1, 6, 6, 0, \ldots)\ , \quad (1, 14, 36, 24, 0, \ldots)\ , \quad \text{etc.}$$

[12] This is actually a very crucial assumption which demonstrates how quantum mechanics might be modified. For example, one could consider a theory where energy plays a more fundamental role as the order of these outcomes. Perhaps quantum gravity will come from this sort of modification.

[13] This may be related to the existence of de Finetti theorems as we will discuss slightly in the next section.

If we were to stop here, then the simplicity axiom would tell us that $(1, 0, \ldots)$ is our theory, which simply corresponds to classical probability theory, where $K = d$ and there is actually no distinct notion of measurement.

5-Continuity: There is a reversible and continuous transformation between any two pure states.

Between this axiom and the unique declaration of our familiar quantum theory is the most significant portion of the work. Nevertheless, the logical steps that are left to be made are still simple. First, one needs to understand that the set of such transformations must form a Lie group as an extension of the permutation symmetry of d outcomes. This means that the space of pure states must be homogenous by the action of the Lie group. One must then show that the second simplest theory $(1, 2, 0, \ldots)$ can support such a continuous group action. Finally, it must be demonstrated that such a group is unique and in fact isomorphic to our beloved $\mathrm{SU}(d)$ in which case the space of pure states is the familiar $\mathbb{C}P^{d-1}$.[34]

Quantum State Estimation

Various techniques and perspectives in quantum state estimation have by now become standard practice.[36, 37] In addition to tomography, the most significant methods are Bayesian estimation[38], maximum-likelihood estimation (MLE), as well as hedged[39] MLE. These techniques are by no means unique to quantum systems, but the quantum distinction between K and d does introduce some extra issues of interpretation. Equally important is the need to understand when the estimation of a state can actually be a practice that is consistent in a larger scheme.[40–42] The theory of estimation and machine learning at the classical level is already quite subtle; so a quick review of classical estimation is in order before we can discuss how one may interpret estimating a quantum state.

Theories and Data

For any estimation scheme, one has a model or theory, θ, represented by the set of parameters $\theta = \{p_k\}$. The goal is to find values or distributions for these parameters that are in some sense the most consistent with some given data, D, represented by a set of events $D = \{e_t\}$, while also being the most predictive of further data. Of course, a theory is further understood to make mathematical predictions, specifically represented by the ability to calculate conditional probabilities, $P(D|\theta)$. Specifically defining good measures of a best estimate are beyond the scope of this work, but there are two such "best" estimates that are widely used which should be mentioned:

The most common method is the maximum likelihood estimate,

$$\theta_{ML}(D) = \arg\max_{\theta} P(D|\theta), \tag{2.16}$$

where θ is simply chosen to take the value which maximizes the likelihood function $P(D|\theta)$. A point to make here which will become important in quantum state estimation is that θ is always understood to be optimized over some "allowable" set of values. Although the interpretation of this estimate can be delicate (as will be explained) this is generally the easiest estimate one can calculate because it boils down to a convex optimization problem.

Perhaps the most ideal method is the Bayesian estimate,

$$\theta_B(D) = \int d\theta\, \theta P(\theta|D) = \frac{1}{P(D)} \int d\theta\, \theta P(D|\theta) P(\theta). \tag{2.17}$$

Here one calculates the mean value of the actual distribution $P(\theta|D)$, which by Bayes' rule is simply proportional to the likelihood, $P(D|\theta)$, times a prior, $P(\theta)d\theta$ (since we can treat $P(D) = \int d\theta\, P(D|\theta)P(\theta)$ as just a normalization constant.) While optimal by many theoretical standards, the distribution $P(\theta|D)$ (or at least a handful of its statistical moments) can be a useful thing to have in practice. However, a Bayesian estimate can be significantly more demanding to compute than that from MLE. Also an issue, as is the case in quantum state estimation, are the problems where the measure $d\theta$ is not uniquely motivated, which in turn makes a choice of the prior ambiguous or arbitrary.

A very important assumption one often also requires in order to calculate either of these estimates is that

$$P(D|\theta) = \prod_t P(e_t|\theta) \tag{2.18}$$

— that is, an assertion that the events are identically and independently distributed (or i.i.d.). Here the prototypical example is the estimation of the probability, p, for a coin flip, e, to turn up heads, $e = 1$. Under the i.i.d. assumption, the likelihood function for p is simply

$$P(\{e_k\}|p) = P(n, N|p) = p^n (1-p)^{N-n} \tag{2.19}$$

where N is the number of trials and $n = \sum_k e_k$ is the number of heads. It is a simple exercise to show that the two estimates are then

$$p_{ML} = \frac{n}{N} \qquad \text{and} \qquad p_B = \frac{n+1}{N+2} \tag{2.20}$$

where the Bayesian estimate has assumed a uniform prior, $P(p)dp = dp$. The Bayesian estimate in this case is also known as Laplace's rule, which has many nice properties.

There are two particular objections one should make here. The first is that MLE can result in unfortunately optimistic estimates as will be explained. With this in mind, under what circumstance are such estimates acceptable? Secondly, the i.i.d. assumption seems to be an unjustifiably strong

statement to make simply by fiat. In what sense can the i.i.d. assumption be verified or make operational sense? These issues are also signature of the subjectivist perspective (which the author enjoys) but their answers also have useful applications.

Statistical Fluctuations and Hedging

A maximum likelihood estimate can be especially inappropriate when considering a distribution with high purity. In the example of a biased coin, if an estimator witnesses $n = N$ heads, an estimate like $p = 1$ is inappropriate because it can be interpreted as the instruction that a gambler should bet all of their money on heads for the next flip. This is an especially striking claim for $N = 1$ but it is also bad for any N. Put simply, an estimate should always reflect the confidence of the estimator as determined by their experience.

This point becomes even more pertinent in quantum state estimation. Considering the estimation of a qubit, suppose a single measurement is made for σ_z and also for σ_x and say that the outcome for each of them was "up." MLE tells us that the estimated state should be a (pure) eigenstate of $\frac{1}{\sqrt{2}}(\sigma_z + \sigma_x)$! This same issue becomes apparent in tomography as well for suppose in addition σ_y is measured once with the outcome "up." Tomography tells us that if we were to measure $\frac{1}{\sqrt{3}}(\sigma_x + \sigma_y + \sigma_z)$, then we should expect the outcome "up" with probability $p = \sqrt{3}$! Such statistical fluctuations are of general concern in the estimation of high purity states because they invariably result in operators with negative eigenvalues or unfortunately low rank.[14]

In this regard, Bayesian techniques are certainly preferable for two reasons. The first is that one can simply set the prior equal to zero over all non-positive operators. The second is that, similar to the classical estimate in equation (2.20), estimates will always keep a safe distance away from the boundary of states (where operators have low rank.) However, it is worth repeating that Bayesian estimates are exceedingly more computation intensive, especially for high-dimensional states.

To enforce full rank estimates while keeping the computational simplicity of MLE, one can perform what is called *hedged* maximum likelihood estimation. The nicest kind of hedging is the so called "add β" technique, where one simply multiplies the likelihood by $\det \rho^\beta$ for some number β. This mimics the natural hedging in classical Bayesian estimation, specifically analogous to the use of Dirichlet-β priors, which retain their form under updating. In this classical case, one can even

[14]These are unacceptable properties if one takes seriously that one could have always chosen differently to measure any other observable, reflected by the roundness of the Bloch sphere which defines the boundary of positivity.

derive unique values for β depending on what the estimator considers optimal. Unfortunately, such values of β are not so clear for quantum estimation, but people often choose $\beta = 1/2$. In the current state of the art, a quantum estimator's uncertainty in this choice can only be slightly pacified by the observation that at least in the classical case an estimate will depend only weakly on β for large statistics.

Unknown States: Part II

The concern behind the i.i.d. assumption has been much more successfully resolved. We will discuss this issue from the subjective viewpoint, so probabilities are to be interpreted as a statement of the knowledge and experience of the estimator. As such, we must first resolve the interpretational issue of an unknown probability, sometimes referred to as the "man in the box." This is especially pressing for physics in general, where we imagine that the distributions we are measuring reflect consistent properties of a system with states that support the random events being recorded. Indeed, one can replace the notion of an unknown probability with an operational property thanks to the de Finetti representation theorem.

The Quantum de Finetti Theorem and a Tomographic Proof

The property de Finetti devised in the context of classical probability is called exchangeability which represents a symmetry between trial events and the ability to consistently incorporate new information. Specifically, let $P^{(n)}$ be a distribution over n trials consistent with the sequence of events $D = \{e_t\}$ — i.e. such that $P^{(n)}(e_1, \ldots, e_n) \neq 0$. We say that this distribution is *symmetric* if $P^{(n)}(e_{\pi(1)}, \ldots, e_{\pi(n)}) = P^{(n)}(e_1, \ldots, e_n)$ for any permutation, π. Every symmetric distribution can be represented equivalently by a function of the tallies for each outcome, $P^{(n)}(n_1, \ldots, n_d)$ where $\{1, \ldots, d\}$ enumerates the set of possible outcomes and n_i is the number of events in D with the value i. Further, a distribution is *extendible* if for any $N > n$, there is another symmetric distribution $P^{(N)}$ such that $P^{(n)}(e_1, \ldots, e_n) = \sum_{e_{n+1}} \cdots \sum_{e_N} P^{(N)}(e_1, \ldots, e_n, e_{n+1}, \ldots, e_N)$. A distribution is then called *exchangeable* if it is both symmetric and extendible.[15]

[15] A good example of a distribution that is symmetric but not exchangeable is the distribution from a GHZ state.

The de Finetti theorem asserts that every exchangeable distribution can be written uniquely in the form

$$P^{(n)}(e_1, \ldots, e_n) = \int_{\Delta_d} d\vec{p}\, P(\vec{p})\, p_{e_1} \cdots p_{e_n} \tag{2.21}$$

or equivalently

$$P^{(n)}(n_1, \ldots, n_d) = \int_{\Delta_d} d\vec{p}\, P(\vec{p})\, p_1^{n_1} \cdots p_d^{n_d} \tag{2.22}$$

for some common distribution P over the $(d-1)$-dimensional probability simplex, Δ_d. We call such expressions de Finetti representations. Also important to observe is that each $P^{(N)}$ is also exchangeable and thus has a de Finetti representation with the same integrated distribution, $P(\vec{p})$. The unique existence of this representation demonstrates that it is *as if* D consisted of i.i.d. events. So the i.i.d. assumption usually made by fiat can instead be interpreted as the explicit assertion of a symmetry in the model being considered which is further verified for the given set of trials to have the property of exchangeability. This is precisely the operational property we were looking for.

Of course, there is a quantum de Finetti theorem. The quantum property of exchangeability now refers to a density operator $\rho^{(n)}$ over the n-partite space of d-dimensional systems that is permutationally invariant and can always be written as a reduced state, $\rho^{(n)} = \mathrm{Tr}_{N-n}\rho^{(N)}$ for any N, where Tr_{N-n} refers unambiguously to a trace over any $N-n$ parts of the larger state. The quantum de Finetti representation theorem then states that every exchangeable state is a unique mixture of identical copies,

$$\rho^{(n)} = \int_{\mathcal{D}_d} d\rho\, P(\rho)\, \rho^{\otimes n} \tag{2.23}$$

where $\mathcal{D}_d$ is the space of (d^2-1)-dimensional density operators. An important corollary is that such states are never entangled.

Thanks to Caves et al. [43], there is a simple *tomographic* proof of the quantum de Finetti theorem assuming the classical de Finetti theorem that is worth sketching. One considers a tomographically complete POVM consisting of the products of the elements of a local d^2-outcome POVM. The classical de Finetti theorem tells us that we can write the distribution for this POVM as an integral over Δ_{d^2}. This in turn gives us an expression for the density operator as the mixture of identical operators. One must finally show that the support of this mixture is in fact only $\mathcal{D}_d \subset \Delta_{d^2}$.

An important note to make here is that not every probability theory has a de Finetti theorem. In particular, real and quaternionic quantum theories do not have such a theorem. The proof

sketched above alludes to the possibility that this has to do with the composite axiom of Section 2.2. In other words, the ability to perform complete tomography using only local measurements is a crucial step of the sketched proof. For (complex) quantum mechanics, this is ensured because $K_{\mathbb{C}}(d) = d^2$, where K is the number defined in section 2.2. On the other hand, one can easily count that for the real and quaternion theories[16]

$$K_{\mathbb{R}}(d) = \frac{1}{2}(d^2 + d) \qquad \text{and} \qquad K_{\mathbb{H}}(d) = 2d^2 - d \tag{2.24}$$

which do not satisfy the composite axiom but rather

$$K_{\mathbb{R}}(d_A d_B) > K_{\mathbb{R}}(d_A) K_{\mathbb{R}}(d_B) \qquad \text{and} \qquad K_{\mathbb{H}}(d_A d_B) < K_{\mathbb{H}}(d_A) K_{\mathbb{H}}(d_B). \tag{2.25}$$

So in the real theory, local tomography is never complete while in the quaternionic theory, products of states are not always states! This seems to correspond to the fact that not every exchangeable real state can be written as a de Finetti mixture while not every de Finetti mixture of quaternionic states is exchangeable![17]

A Global Representation Theorem of Symmetric States

There is a relaxed generalization of the quantum de Finetti theorem, thanks to Renato Renner [44], which has several useful implications. In this case, one drops any questions of exchangeability and only considers a symmetric N-partite state of d-dimensional systems and its n-partite reduced states. Indeed, the theorem establishes that such subsystems of larger symmetric states are approximately independent. Specifically any n-partite subsystem, $\rho^{(n)} = \mathrm{Tr}_{N-n}\rho^{(N)}$, of an N-partite symmetric system can be approximated by a state of the form

$$\tilde{\rho}^{(n)}_{n-m} = \int d\sigma \, P(\sigma) \, \sigma^{\otimes m} \otimes \tilde{\rho}^{(n-m)} \tag{2.26}$$

[16] Real and quaternionic quantum mechanics are the theories of signature $(1, 1, 0, \ldots)$ and $(1, 4, 0, \ldots)$ respectively.

[17] A fun question one might ask is whether the probability theory $(1, 6, 6, 0, \ldots)$, where $K(d) = d^3$, has a de Finetti theorem?

with an error given by the trace distance of

$$\varepsilon = \mathrm{Tr}|\rho^{(n)} - \tilde{\rho}_r^{(n)}| = 3(N-n)^d \exp\left(-r\frac{N-n}{N}\right) \tag{2.27}$$

for $r \ll n < N - \sqrt{N}$. Roughly speaking, every large N-partite symmetric system is approximatly independent so long as we can ignore a fraction $(N-n)/N$ of the system and tolerate interactions in r/n of the remaining system.

This is especially useful for the interpretation of many condensed matter theory results which are often based on a state that is assumed to have the i.i.d. property. In this case, Renner's theorem applies for two very fundamental reasons: The first is that exchange symmetry is already present when the system is modeled by a field as is often the treatment for many order parameters. The second is that such a notion of approximate independence is definitely good enough to justify the calculation of a macroscopic observable since a state like $\tilde{\rho}^{(r)}$ often corresponds to an interaction that is irrelevant under renormalization. So Renner's theorem explains essentially why and under what circumstances the many models based on an i.i.d. property are so effective.

In quantum state tomography (and quantum key distribution or QKD), Renner's theorem is also very useful. In these cases, the state $\rho^{(N)}$ is presumably made of N copies of an "unknown state" to be estimated (or an entangled-pair being shared between teleporters.) However, various systemic errors (or security attacks) could occur that would make the i.i.d. assumption generally false. Nevertheless, the i.i.d. property can be approximately restored according to Renner's theorem by simply applying a random permutation amoung the subsystems, thus enforcing symmetry in the composite system. The relaxed notion of independence is especially useful in QKD because the measurements of this state are followed by a distillation phase.

Other Tomographies

Although there are more sophisticated techniques, from a practical point of view, linear inversion tomography represents a basic technique from which to conceptualize quantum information. Practically, tomography also serves as a great starting point from which to apply numerical techniques in MLE and Bayesian estimation. As well as with these other estimation schemes, quantum tomography is by no means restricted to state estimation.[18] One can also perform tomography on unknown observables or even unknown processes as we will now explain. For the rest of this thesis we will simply use the term tomography, implicitly referring to linear inversion.

"Standard" Tomographies

It will be useful at this point to rewrite the various linear constraints in tomography as simple matrix equations. Let $\{\rho_a\}_{a=1}^N$ be a set of states one can prepare and $\{\Sigma_i\}_{i=1}^M$ be a set of observables one can measure. It is convenient to refer to a and i as device settings and collections of a's and i's as experiments. The operator for each setting can be expanded as a linear combination

$$\rho_a = P_a{}^\mu \sigma_\mu \qquad \text{and} \qquad \Sigma^i = \sigma^\mu W_\mu{}^i \tag{2.28}$$

where $\{\sigma_\mu\}$ is *any*[19] basis for the d^2-dimensional space of Hermitian operators and $\{\sigma^\mu\}$ is the associated dual basis.[20] For every state-observable pair of settings, (a, i), one has access to the expectation value

$$S_a{}^i = \mathrm{Tr}\rho_a \Sigma^i = P_a{}^\mu W_\mu{}^i, \tag{2.29}$$

the set of which we may denote simply by the matrix equation

$$S = PW. \tag{2.30}$$

[18] ... although the PhySH labels available in APS Journals seem to suggest otherwise.

[19] One usually has in mind an orthogonal basis like the Pauli or Euclidean basis. However, it is important to understand that $\{\sigma_\mu\}$ could be any basis. Although this subtlety will make no difference in this section, it is the bane of self-consistent tomography and the key to non-holonomic tomography.

[20] Given a basis $\{\sigma_\mu\}$, the dual basis vectors $\{\sigma^\mu\}$ are the solution to the linear system of equations $(x^\mu, \sigma_\nu) = \delta^\mu_\nu$ for all ν, where $(,)$ denotes an inner product. A dual basis is similar to a reciprocal basis in solid state physics, where special momentum vectors are defined with respect to displacement vectors $\{a_i\}$ by $(k^i, a_j) = 2\pi\delta^i_j$ for all j.

The practice of state tomography is then equivalent to the ability to invert this equation,

$$P = SW^{-1}, \tag{2.31}$$

so that P can be estimated to arbitrary precision through the collection of data, S, and knowledge of the observables, W. Any use of equation (2.30) in this way, namely where parameters in P or W are assumed known, will be referred to as *standard* tomography.

Unknown Measurement Outcomes

An equation like (2.30) makes pretty obvious that we could also estimate W if we instead had previous knowedge of P since we can just as well write

$$W = P^{-1}S. \tag{2.32}$$

Such practice is referred to as detector tomography which can be quite useful when assessing a fabricated measurement device. Indeed, detector tomography and estimation have become a common practice as well. [45–47]

Estimating Unknown States from Known States

Similarly, one can also estimate states with unknown detectors if one has other known states. This is simply because we can write

$$S = PW \qquad \text{and} \qquad T = QW \tag{2.33}$$

where Q contains the set of state parameters which are known and T is the corresponding data. Simply substituting one into the other, one obtains

$$P = ST^{-1}Q. \tag{2.34}$$

This was only recently acknowledged [1] but in the context of maximum-likelihood techniques. Although this practice may seem trivial from a linear inversion perspective, equation (2.34) is conceptually only two steps away (pun intended) from non-holonomic tomography.

Unknown Operations

Finally, another form of tomography can be done if one knows both P and W. In particular, one can characterize an unknown process, $\rho \longrightarrow \Phi(\rho)$, because the process is also just a linear map so that

$$S_a{}^i = P_a{}^\nu \Phi_\nu{}^\mu W_\mu{}^i \tag{2.35}$$

which we can simply invert to give

$$\Phi = P^{-1} S W^{-1}. \tag{2.36}$$

This is similar in content to the estimation of a unitary "S-matrix" (which is to be compared to the operation Φ rather than the expectation value we are calling S) except that the dimensions are quadratically larger because these processes may include irreversible transformations such as decoherence, decay, and eavesdropping. Indeed, if Φ is to represent a physical operation, then one must further constrain Φ to be a *completely positive* map.[21]

Self-Consistent Tomographies and SPAM gauge

If P and W are not known, one can still collect the data, S. In such a case, however, it is no longer obvious what we can conclude from the measured data. At first glance, what should be apparent is that P and W can no longer be estimated uniquely because of the presence of so called gauge degrees of freedom. In particular, equation (2.30) is invariant under simultaneous linear transformations of the form

$$(P, W) \longrightarrow (PG^{-1}, GW). \tag{2.37}$$

These gauge degrees of freedom can be organized into particular classes described by the chain of subgroups

$$\mathrm{SU}(d) \quad < \quad \mathrm{SO}(d^2-1) \quad < \quad \mathbb{R}^{d^2-1} \rtimes \mathrm{GL}(d^2-1) \quad < \quad \mathrm{SL}(d^2). \tag{2.38}$$

$\mathrm{SU}(d)$ is the familiar unitary group on the Hilbert space, present in G as its adjoint representation and corresponding to the choice of an orthonormal basis. $\mathrm{SO}(d^2-1)$ is the smallest orthogonal group containing the unitary adjoint representations and is the symmetry group of the gram matrix technique in section 2.4. $\mathbb{R}^{d^2-1} \rtimes \mathrm{GL}(d^2-1)$ is the most general transformation when considering only

[21] This is simply the property that $\Phi \otimes 1_{d^2}(\rho)$ is positive for every positive ρ in a dilated space of states.

normalized states as will be described in chapter III. $\mathrm{SL}(d^2)$ is the most convenient set of general transformations to consider for the gate sets of section 2.4 where the "S" as in *s*pecial refers to the determinant being restricted to one.

Such gauge degrees of freedom have been considered early on under the name of unspeakable information. [48] Further, some consequences of this gauge have been identified relative to practicing state estimation when using imperfect detectors. [49] Very soon after this, it had become a definitive problem in process tomography after which so called self-consistent tomographies began to emerge. [10] It was then when SPAM (*s*tate *p*reparation *a*nd *m*easurement) errors and SPAM gauges got their name. The goal of self-consistent tomography is still to arrive at an estimate of the unknown quantum parameters which are unique modulo the relevant gauge degrees of freedom.

Gram Matrix Tomography

One sort of self-consistent SPAM tomography considers when the state-measurement Gram matrix[12],

$$\Gamma = \begin{bmatrix} P \\ W^{\mathsf{T}} \end{bmatrix} \begin{bmatrix} P^{\mathsf{T}} & W \end{bmatrix} = \begin{bmatrix} PP^{\mathsf{T}} & S \\ S^{\mathsf{T}} & W^{\mathsf{T}}W \end{bmatrix} \tag{2.39}$$

is so called completable[50]. As mentioned earlier, the Gram matrix still contains orthogonal gauge degrees of freedom, $(P, W) \longrightarrow (PR, R^{\mathsf{T}}W)$ where $R \in \mathrm{SO}(d^2)$. The off-diagonal blocks, S, are the accessible data while the diagonal blocks

$$(PP^{\mathsf{T}})_{ab} = \mathrm{Tr}\,\rho_a\rho_b \qquad \text{and} \qquad (W^{\mathsf{T}}W)^{ij} = \mathrm{Tr}\,\Sigma^i\Sigma^j \tag{2.40}$$

are what one trys to infer.

The problem of completability is related to the older problem of rigidity: Given $N + M$ point masses of which some pairs are connected by massless rigid rods, is the entire set of rods and masses rigid in a d^2 dimensional space? For example, consider 4 point masses $\{1, 2, 3, 4\}$ such that the following pairs are connected: $\{(1, 2), (2, 3), (3, 4), (4, 1)\}$ — that is, they are connected in a loop. This configuration is rigid in 1 dimension, but is wobbly in 2 dimensions. If we add the pair $(1, 3)$ to the set, then the configuration becomes rigid in 2 dimensions, but is still flappy in 3 dimensions. If we add another pair $(2, 4)$, then the configuration is a tetrahedron which is rigid in all dimensions.

So in gram-matrix estimation, one is concerned with when the state-measurement frame is "rigid" i.e. completable. Given only S, Γ is never completable. However, there are two kinds of

assumptions one can add which will make Γ completable for large enough M and N. (Remember, N and M were the number of state and measurement settings, respectively, considered in a tomography experiment.) One can assume that all the states are sufficiently pure so that $(P\bar{P})_{aa} = 1 - \epsilon$. On the other hand, one can assume that all of the measurements are sufficiently projective with known degeneracies, $\{d_i\}$ so that $(\bar{W}W)^{ii} = d_i \pm \epsilon$. For exactly what M and N is Γ completable depends on d in a non-trivial way. To give an example: for $d = 3$ with both the purity and projectivity assumptions, either ($M \geq 5$ and $N \geq 15$) or ($M \geq 6$ and $N \geq 9$) or ($M \geq 10$ and $N \geq 8$) will put Γ in the completable phase.

Gate Set Tomography

Another tomography to consider is estimating a set of unknown processes using unknown SPAM.[11] Here one considers many processes and so some new notation will be convenient:

$$\langle\langle \Sigma | \, \Phi \, | \rho \rangle\rangle = \mathrm{Tr}\, \Phi(\rho)\Sigma. \tag{2.41}$$

The scenario we imagine is that one can prepare a single unknown state, ρ, measure the outcome of a single unknown POVM, $\{E, \bar{E}\}$ which we represent by the observable $\Sigma = E - \bar{E}$, and perform a set of gates, $\{\Phi_\alpha\}$, which are to be characterized. The entire collection of the state, the measurement, and set of gates is called a *gate set*. Since the entire gate set is unknown, there is a gauge degree of freedom so that the transformation

$$\Big\{ |\rho\rangle\rangle \, , \langle\langle \Sigma | \, ; \{\Phi_\alpha\} \Big\} \longrightarrow \Big\{ G^{-1} |\rho\rangle\rangle \, , \langle\langle \Sigma | G \, ; \{ G^{-1} \Phi_\alpha G \} \Big\}. \tag{2.42}$$

for any $G \in SL(d^2)$ results in another gate set that is equally consistent with the measured data.

The first step is to cook up a set of operations to be used to generate tomographically complete sets of state-preparation and measurement:

$$\{F_\mu\}_{\mu=1}^{d^2} \Longrightarrow \{ \, |\rho_a\rangle\rangle = F_a \, |\rho\rangle\rangle \, , \langle\langle \Sigma_i | = \langle\langle \Sigma | \, F_i \}. \tag{2.43}$$

These can consist of single gates from the original set, sequences of such gates, or a combination of the two. Indeed, this is how different states are actually prepared in the lab. For example ρ could be the groundstate of an atom, Σ is something like the fluorescence, and the $\{F_\mu\}$ are the effective

operations generated by various laser pulses. One then proceeds to estimate the following matrix elements:

$$A_\mu = \langle\langle \Sigma | \, F_\mu \, | \rho \rangle\rangle = \langle\langle \Sigma | \rho_\mu \rangle\rangle = \langle\langle \Sigma_\mu | \rho \rangle\rangle$$

$$D_{ia} = \langle\langle \Sigma_i | \rho_a \rangle\rangle = \langle\langle \Sigma | \, F_i F_a \, | \rho \rangle\rangle \tag{2.44}$$

$$\Pi_{ia}(\alpha) = \langle\langle \Sigma_i | \, \Phi_\alpha \, | \rho_a \rangle\rangle = \langle\langle \Sigma | \, F_i \Phi_\alpha F_a \, | \rho \rangle\rangle \, ,$$

checking a posteriori that rank$D = d^2$ to ensure that the SPAM are tomographically complete.

Letting $\{ \, |\mu\rangle\rangle \, \}$ be an orthonormal basis one then defines two operators,

$$G = \sum_a |\rho_a\rangle\rangle\langle\langle a| = \sum_{\mu,a} P_a{}^\mu \, |\mu\rangle\rangle\langle\langle a| \qquad \text{and} \qquad H^* = \sum_i |i\rangle\rangle\langle\langle \Sigma_i| = \sum_{i,\mu} W_\mu{}^i \, |i\rangle\rangle\langle\langle \mu| \, . \tag{2.45}$$

Further, one can construct the following linear objects:

$$\begin{gathered} \tilde{1} = \textstyle\sum_{a,i} D_{ia} \, |i\rangle\rangle\langle\langle a| = H^* G \\ |\tilde{\rho}\rangle\rangle = \textstyle\sum_\mu A_\mu \, |\mu\rangle\rangle = H^* |\rho\rangle\rangle \\ \langle\langle \tilde{\Sigma}| = \textstyle\sum_\mu \langle\langle \mu| \, A_\mu = \langle\langle \Sigma| G \\ \tilde{\Phi}_\alpha = \textstyle\sum_{a,i} \Pi_{ia}(\alpha) \, |i\rangle\rangle\langle\langle a| = H^* \Phi_\alpha G \end{gathered} \tag{2.46}$$

which may be observed as related to the "actual" gate set, up to transformations involving G and H^*. The final step in gate set tomography is to use the inverse of $\tilde{1}$ to transform the above vectors and operators into a gate set that is equivalent to the "actual" gate set up to a gauge transformation:

$$\left\{ \quad \tilde{1}^{-1} \, |\tilde{\rho}\rangle\rangle = G^{-1} \, |\rho\rangle\rangle \qquad , \qquad \langle\langle \tilde{\Sigma}| = \langle\langle \Sigma| \, G \qquad ; \qquad \left\{ \tilde{1}^{-1} \tilde{\Phi}_\alpha = G^{-1} \Phi_\alpha G \right\} \quad \right\}. \tag{2.47}$$

In practice one usually has a target set of gates in mind, which have definite representations in an idealized basis. In such cases, the practice has been to follow up this tomography step with MLE optimization of the gate set over the gauge parameters.

Non-Holonomic Tomography and SPAM Correlations

Although self-consistent tomographies for the most part solve their estimation problem, they make one very crucial a priori assumption: that the devices used for state-preparation and measurement are uncorrelated. If you think about it, the idea of such correlations can seem quite puzzling — how would you know if $\langle PW \rangle \neq \langle P \rangle \langle W \rangle$ if neither $\langle P \rangle$ or $\langle W \rangle$ are available? Such is the fundamental problem for quantum systems with unknown SPAM.

An answer exists to this puzzle in the perspective of non-holonomic tomography. In non-holonomic tomography, states and observables are understood to be *effectively* uncorrelated for any minimally complete tomography experiment one can perform. However, this will lead to contradictions between multiple such tomography experiments when there are correlations. Such contradictions can be quantified by a matrix quantity called a partial determinant. Further, these contradictions suggest correlations are to be understood as a kind of global property due to the way that measurable data connects different tomography experiments. Such is the subject of this dissertation that we will explain in the following chapters.

CHAPTER III

DETECTING CORRELATED ERRORS IN SPAM TOMOGRAPHY

From Jackson, C. and van Enk, S. 2015. Detecting correlated errors in state-preparatoin-and-measurement tomography. *Physical Review A* 92(4):042312

Abstract

Whereas in standard quantum state tomography one estimates an unknown state by performing various measurements with known devices; and whereas in detector tomography one estimates the POVM elements of a measurement device by subjecting to it various known states, we consider here the case of SPAM (state preparation and measurement) tomography where neither the states nor the measurement device are assumed known. For d-dimensional systems measured by d-outcome detectors, we find there are at most $(d-1)(d^2-1)d^2$ "gauge" parameters that can never be determined by any such experiment. For the case $d = 2$ we find new gauge-invariant quantities that can be accessed experimentally and that can be used to debug SPAM. In particular, we identify parameters that allow one to detect correlations between SPAM errors. Moreover, standard quantum state tomography and detector tomography can be seen as fixing all gauge parameters by *fiat* in different ways.

Introduction

Quantum tomography has become an important tool for characterizing quantum devices [33, 37]. For example, in order to estimate the state ρ of quantum systems produced by some quantum source, we let our source create many copies of ρ, and subject each of those copies to different measurements. If we describe the whole set of measurements by POVM elements Π_k (such that $\sum_k \Pi_k = 1$), the experiment will produce estimates of probabilities

$$p_k = \mathrm{Tr}(\rho\Pi_k), \tag{3.1}$$

with $\sum_k p_k = 1$. If one has at least as many observed relative frequencies f_k (the number of times outcome k was observed divided by the total number of measurements) as there are parameters in ρ, then one may estimate ρ via a variety of different methods (from linear inversion to hedged

maximum likelihood estimation and Bayesian methods, see, e.g., [38–40]). Legitimately ending up with an estimate of a *single* density matrix ρ does require some assumptions about the experimental setup [42–44], namely, invariance under permutations of the copies. In particular, even if there is drift (for example, due to a slowly changing magnetic field or a slowly changing phase of the laser field one uses to prepare quantum systems), such that copy ρ_n differs slightly from the previous copy ρ_{n-1}, by performing the different measurements in random order one still estimates a single density matrix, namely, the average:

$$\langle \rho \rangle = \frac{1}{N_t} \sum_{n=1}^{N_t} \rho_n, \tag{3.2}$$

in an experiment with N_t copies of states in total.

In more recent times several experiments on detector tomography have been performed [31, 45–47]. Here one assumes that known states are fed into a measurement device, whose observed outcomes now tell us about the POVM elements that describe the detector. Detector tomography makes sense: A photodetector, for example, is not such a simple device theoretically. There are infinitely many modes of the electromagnetic field, and each mode is described by an infinite-dimensional Hilbert space. A simple candidate model for a photodetector can be formulated in terms of a quantum efficiency η and a dark count rate r, but these two parameters could drift over time and be field-mode dependent. In the case of detector tomography, we estimate probabilities of the form

$$p_m = \mathrm{Tr}(\rho_m \Pi), \tag{3.3}$$

for different known states ρ_m (where m runs from 1 to M), in order to estimate the POVM element Π, where the latter is really an average (under the proviso that one tested the measurement device with different states created in random order, to ensure symmetry under permutations of the measurements) over the course of the experiment of the different POVM elements describing the drifting detector. In complete analogy to (3.2) we may write then

$$\langle \Pi \rangle = \frac{1}{N_t} \sum_{n=1}^{N_t} \Pi_n, \tag{3.4}$$

if we performed this particular measurement N_t times in total.

The present paper focuses on two concerns: First, what exactly can we find out about measurements and states when we do not assume we know either one or the other? That is, what can we infer about the POVM elements Π_i describing different measurement settings and about different states ρ_m from estimates of probabilities

$$p_{m,i} = \mathrm{Tr}(\rho_m \Pi_i)? \tag{3.5}$$

Second, what if the fluctuations in state preparation and measurement are correlated? That is, what if the observed average relative frequencies are not determined by averaged states and averaged measurements? How can we infer there are correlations, i.e., that

$$\langle \mathrm{Tr}(\rho_m \Pi_i) \rangle \neq \mathrm{Tr}(\langle \rho \rangle_m \langle \Pi \rangle_i)? \tag{3.6}$$

Here the left-hand side is a quantity for which we obtain direct estimates from observed relative frequencies, but the two quantities $\overline{\rho}_m$ and $\overline{\Pi}_i$ on the right-hand side are not directly accessible individually and have to be inferred.

Such questions have become of interest in the context of debugging quantum devices that are meant to serve as fault tolerant quantum computers. The requirements on fault tolerance are quite stringent and more and more precise tools for analyzing tomography experiments have been developed very recently [10–12]. Correlations between errors are particularly bad for fault tolerance, hence our focus on detecting correlated SPAM errors.

We are going to analyze SPAM tomography for the case of a single qubit and two-outcome POVMs. Even for this simple case our results suggest novel experiments. Most results generalize easily to multiple qubits and/or higher-dimensional systems and measurements, and in particular to the combination of d dimensional systems and d-outcome POVMs.

SPAM Tomography

Let us first consider the simplest version of our problem: a single qubit state, ρ, and a single two-outcome qubit detector, with the two outcomes described by POVM elements $\{E, \bar{E}\}$ (read: E and NOT-E). From the conditions $\rho \geq 0, E \geq 0, \bar{E} \geq 0$, $\mathrm{Tr}\rho = 1$, and $E + \bar{E} = 1$, a general

parameterization can be written in terms of the Pauli matrices $\vec{\sigma} = (\sigma_x, \sigma_y, \sigma_z)^T$ as

$$\begin{aligned} \rho &= \frac{1}{2}\big(1 + \vec{p}\cdot\vec{\sigma}\big), \\ E &= \frac{1}{2}\Big((1+u)1 + \vec{w}\cdot\vec{\sigma}\Big), \\ \bar{E} &= \frac{1}{2}\Big((1-u)1 - \vec{w}\cdot\vec{\sigma}\Big), \end{aligned} \tag{3.7}$$

where positivity is ensured by the inequalities

$$\begin{aligned} |\vec{p}| &\leq 1, \\ |\vec{w}| + |u| &\leq 1. \end{aligned} \tag{3.8}$$

Altogether there are seven parameters. Three of these correspond to a choice of coordinate system (i.e., we have to specify what we mean by the components of spin in the x, y, z directions), or, equivalently, to a choice of basis for the 2D Hilbert space describing our qubit. The other four parameters can be thought of as state purity ($|\vec{p}|$), the detector's discrimination power ($|\vec{w}|$), detector bias (u), and the alignment (or fidelity) between detector and state ($\hat{p}\cdot\hat{w}$). We may conveniently represent the two-outcome POVM by a single observable:

$$\Sigma \equiv E - \bar{E} = u1 + \vec{w}\cdot\vec{\sigma}, \tag{3.9}$$

Knowledge of its expectation value S in the state ρ gives us a single constraint on these seven parameters:

$$S \equiv \mathrm{Tr}\rho\Sigma = \vec{p}\cdot\vec{w} + u. \tag{3.10}$$

Counting parameters

Now consider the case where our quantum source has M different settings, so that we may produce (at most) M different qubit states ρ_m. Similarly, our detector has N different settings, so there are $2N$ POVM elements described by the N different observables Σ_i.

Considering the number of undeterminable parameters, γ, a first guess might be that there are

$$\gamma = (\#\text{ unknowns}) - (\#\text{ constraints}) = 3M + 4N - NM \tag{3.11}$$

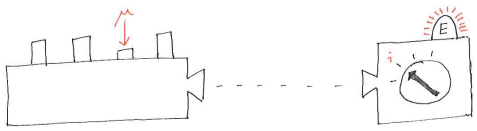


FIGURE 3. Performing two-outcome measurements with a variety of states and observables. The μ-th button of the quantum source is pressed, where μ runs from 1 to M, representing the preparation of state ρ_μ, while the dial of the measurement device is turned to the i-th setting (where i runs from 1 to N), symbolizing the observable Σ_i. The outcome of measurement i is either that the light blinks (E_i) or does not blink ($\bar{E}_i$).

of them (assuming we subject all M states to all N measurements). This kind of counting is correct so long as either $N < 4$ or $M < 3$ but otherwise it fails to take the following important observation into account: For $M = 4$, there are sufficiently many states to perform complete detector tomography on any number of detectors. Similarly, $N = 3$ would be a sufficient number of 2-outcome detectors to perform complete state tomography. Hence for any $N > 3$ and $M > 4$, there are just as many undeterminable parameters as when $N = 3$ and $M = 4$, because the addition of any states (or detectors) could reveal no more information about the detectors (or states) than the "original" 4 states (or 3 detectors) already yield. Therefore the number of undeterminable parameters is

$$\gamma = \begin{cases} 3M + 4N - NM & : \quad N < 4 \text{ or } M < 3 \\ 12 & : \quad N \geq 4 \text{ and } M \geq 3 \end{cases} \tag{3.12}$$

In order to find what these undeterminable parameters are we first introduce some convenient notation. First of all, the set of all constraints can be written succinctly as a matrix equation,

$$S = PW \tag{3.13}$$

where

$$P \equiv \begin{bmatrix} \cdots & \vec{p}_\mu & \cdots \\ \cdots & 1 & \cdots \end{bmatrix}^T \tag{3.14}$$

and

$$W \equiv \begin{bmatrix} \cdots & \vec{w}_i & \cdots \\ \cdots & u_i & \cdots \end{bmatrix} \tag{3.15}$$

are the $N \times 4$ and $4 \times M$ matrices of state and detector parameters, respectively, where we included a constant 1 (which stands for $\mathrm{Tr}(\rho_\mu)$) in the set of state parameters just to make the treatment of states and detectors more symmetric. The form of equation (3.13) makes obvious that the system of constraints has the symmetry:

$$P \longrightarrow PG^{-1}, \quad W \longrightarrow GW, \tag{3.16}$$

where G is of the form

$$G = \begin{bmatrix} H & \vec{0} \\ \vec{a}^T & 1 \end{bmatrix} \tag{3.17}$$

where H is any real 3×3 invertible matrix, $H \in \mathrm{GL}(3)$, and $\vec{a}$ is any real 3-vector, $\vec{a} \in \mathbb{R}^3$. The collection of all $G \in \mathbb{R}^3 \rtimes GL(3)$ forms the affine group.

Blame Gauges: A New Perspective on State/Detector Tomography

Some of the 12 parameters determining G should be familiar. Most important are perhaps the three parameters needed to describe matrices of the form

$$G = \begin{bmatrix} R & \mathbf{0} \\ \mathbf{0} & 1 \end{bmatrix}. \tag{3.18}$$

where R is in the adjoint representation of SU(2) (which here is just SO(3).) This space of symmetries is always present (also for higher-dimensional Hilbert spaces of size d where we would have the adjoint representation of SU(d)), and corresponds to the arbitrary choice of basis for the underlying Hilbert space. This sort of symmetry tends to be trivial, but becomes nontrivial (and interesting) in the context of quantum communication (the information to fix a reference frame was coined "unspeakable" by Asher Peres, for the reason that two distant parties cannot communicate the choice of coordinate system over the phone, unless they already share a reference frame).

Another simple undeterminable parameter follows by considering G of the form

$$G = \begin{bmatrix} g1_3 & \mathbf{0} \\ \mathbf{0} & 1 \end{bmatrix}, \tag{3.19}$$

where $g \in \mathbb{R}$. This kind of scaling transformation trades overall state purity for the detector's discrimination power. That is, imperfections may be blamed, within some limits, more on the preparation of the states than on the quality of our detectors or *vice versa*. Let us refer to parameters like this as *blame gauge* degrees of freedom.

We can characterize all 12 blame gauge parameters as follows. Every H (as appearing in (3.17)) has a polar decomposition $H = BR$ where R is real orthogonal and B is positive symmetric. This yields the following interpretation: We can always diagonalize B: the three eigenvalues are like the scale parameter g of (3.19). The basis in which B is diagonal (determined by three additional parameters) determine which components of $\vec{p}$ and $\vec{w}$ can be rescaled in concert.

The three parameters of R still correspond to the choice of coordinate system as in (3.18)

Finally, the last three parameters are featured in transformations of the form

$$G = \begin{bmatrix} 1_3 & \mathbf{0} \\ \vec{a}^T & 1 \end{bmatrix}, \tag{3.20}$$

where $\vec{a}$ is a real 3-vector. It is perhaps clearer what this transformation does when we write out the action on the system parameters explicitly:

$$\begin{cases} \vec{p}_\mu \longrightarrow \vec{p}_\mu - \vec{a} \\ \vec{w}_i \longrightarrow \vec{w}_i \\ u_i \longrightarrow u_i + \vec{a} \cdot \vec{w}_i. \end{cases} \tag{3.21}$$

We see that this exchanges state-detector alignment for detector bias, and thus is another type of blame gauge.

Having now completely classified the 12 gauge degrees of freedom in SPAM tomography, let us briefly revisit the more conventional types of (qubit) tomography. Having at our disposal $M = 4$ states assumed to be fully known allows for complete detector tomography precisely because we can write in that case $W = P^{-1}S$. Similarly, $N = 3$ known detectors allows us to perform complete quantum state tomography because we can write in that case $P = SW^{-1}$, where W^{-1} is to be interpreted as a right (pseudo-)inverse.

From the perspective of SPAM tomography, we see that the assumptions underlying state and detector tomography simply boil down to fixing the *blame gauge*. In state tomography on qubits

one fixes four parameters each of three fiducial measurements, and in detector tomography one fixes the four parameters each of three fiducial states.

Detecting Correlated Errors

Uncorrelated States and Detectors

A measurement yields averages of the form

$$S_{\mu i} = \langle \vec{p}_\mu \cdot \vec{w}_i + u_i \rangle, \tag{3.22}$$

which contains *two* types of averages: one quantum expectation value (in the state ρ_μ, denoted as usual by $\langle . \rangle$), and one average over the different runs of the experiment (indicated by the over line). The existence of a model (P, W) that satisfies Eq. (3.13) is precisely the statement that a system consists of *effectively uncorrelated* states and detectors. The point is, certainly for $M > 4$ and $N > 3$, not every possible set of data can be effectively uncorrelated simply because the number of data parameters is larger than the number of distinct uncorrelated model parameters whenever

$$MN > 3M + 4N - 12. \tag{3.23}$$

It is not hard to write down the number of constraints on the data that must exist for them to be effectively uncorrelated, namely, (# constraints) = (# data parameters) - (# distinct model parameters)= $(N-3)(M-4)$, as we will see in detail now.

Constraints for Uncorrelated Quantum Data: Partial Determinants

Suppose we have quantum data S that is effectively uncorrelated, i.e., described by a model (P, W). For the sake of presentation, let us consider first the case $N = M = 8$. Think of the data as partitioned into four 4×4 corners of an 8×8 matrix,

$$S \equiv \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix}, \tag{3.24}$$

such that

$$Q_{ab} = P_a W_b \tag{3.25}$$

where

$$P = \begin{bmatrix} P_1 \\ P_2 \end{bmatrix}; \; W = \begin{bmatrix} W_1 & W_2 \end{bmatrix}, \tag{3.26}$$

— that is,

$$P_1 = \begin{bmatrix} \vec{p}_1 & \cdots & \vec{p}_4 \\ 1 & \cdots & 1 \end{bmatrix}^T; \; P_2 = \begin{bmatrix} \vec{p}_5 & \cdots & \vec{p}_8 \\ 1 & \cdots & 1 \end{bmatrix}^T \tag{3.27}$$

are the top and bottom halves of P and

$$W_1 = \begin{bmatrix} \vec{w}_1 & \cdots & \vec{w}_4 \\ u_1 & \cdots & u_4 \end{bmatrix}; \; W_2 = \begin{bmatrix} \vec{w}_5 & \cdots & \vec{w}_8 \\ u_5 & \cdots & u_8 \end{bmatrix} \tag{3.28}$$

are the left and right halves of W.

These submatrices of the data, because of Eq. (3.25), have the property

$$Q_{11}^{-1} Q_{12} Q_{22}^{-1} Q_{21} = 1_4, \tag{3.29}$$

under the assumption that the two inverse matrices on the left-hand side exist. (For example, when state 1 and state 2 are the same, or measurements 1 and 3 are the same, Q_{11}^{-1} does not exist. Also, if one of the four vectors $\vec{p}_i$ for $i = 1, 2, 3, 4$ is a mixture of the other three, then Q_{11}^{-1} does not exist either.) We will refer to the left hand side of (3.29) as a *partial determinant* of S and write

$$\Delta(S) \equiv Q_{11}^{-1} Q_{12} Q_{22}^{-1} Q_{21}. \tag{3.30}$$

One should note that the ordering of the rows of P and columns of W was arbitrary and any such reordering would result in a different partial determinant. Note that a partial determinant can be defined for values of N and M as low as 5 The important conclusion is that the condition that $\Delta S = 1_4$ can be used to detect correlations: a violation of this equality (within error bars) shows that state preparation and measurement must have been correlated.

Nonsingular Partial Determinants

To simplify notation, instead of using Q_{ab}s, let us write the partitions of the data, S, simply as

$$S \equiv \begin{bmatrix} A & B \\ C & D \end{bmatrix} \tag{3.31}$$

in which case we have

$$\Delta(S) = A^{-1}BD^{-1}C. \tag{3.32}$$

An awkward feature of this quantity is that it may display singular behavior because of the presence of inverse matrices. A simple way to remedy this is to define a **nonsingular partial determinant**

$$\nabla(S) \equiv \det(D)A\big(\Delta - \mathbf{1}\big) = B\bar{D}^T C - (\det D)A \tag{3.33}$$

where $\bar{D}$ is the cofactor matrix of D so that, for example,

$$A^{-1} = \frac{1}{\det A}\bar{A}^T. \tag{3.34}$$

We chose to multiply by A simply to reduce computational complexity. This then gives rise to another (non-singular) condition on having uncorrelated SPAM. It has the disadvantage that satisfaction of the condition $\nabla(S) = 0$ may occur not because one's SPAM errors are uncorrelated, but because of having multiplied both sides of (3.29) with zero.

Estimates and Uncertainty for the Partial Determinant of Qubit Data

We consider here briefly error bars in the partial determinant, which are needed to decide whether or not one's data provides significant evidence for correlated SPAM errors.

The data is estimated to maximize the likelihood of MN binomial distributions:

$$\langle S\rangle_{\mu i} \equiv \langle S_{\mu i}\rangle = 2\frac{k_{\mu i}}{n_{\mu i}} - 1 \tag{3.35}$$

where $n_{\mu i}$ is the number of trials for the μ-th state and i-th detector settings and $k_{\mu i}$ is the number of positive outcomes. As such, the deviation of values from the actual mean is

$$(\delta S_{\mu i})^2 \equiv \langle {S_{\mu i}}^2 \rangle - \langle S_{\mu i} \rangle^2 = \frac{1 - {S_{\mu i}}^2}{n_{\mu i}}. \tag{3.36}$$

We can then propagate these uncertainties to the nonsingular partial determinant,

$$\nabla(S) = B\bar{D}^T C - (\det D)A. \tag{3.37}$$

Assuming independent Gaussian random variables, the relative coefficients for the qubit case are

$$\frac{\partial \nabla_{ab}}{\partial A_{cd}} = -\delta_{ac}\delta_{bd} \det D \tag{3.38}$$

$$\frac{\partial \nabla_{ab}}{\partial B_{cd}} = \delta_{ac}\varepsilon_{ijk}\varepsilon_{dlm} C_{ib} D_{jl} D_{km} \tag{3.39}$$

$$\frac{\partial \nabla_{ab}}{\partial C_{cd}} = \delta_{bd}\varepsilon_{cij}\varepsilon_{klm} B_{ak} D_{il} D_{jm} \tag{3.40}$$

$$\frac{\partial \nabla_{ab}}{\partial D_{cd}} = \varepsilon_{ijc}\varepsilon_{kld} \left(B_{ak} D_{jl} C_{ib} - \frac{1}{2} D_{ik} D_{jl} A_{ab} \right) \tag{3.41}$$

where δ & ε are the Kronecker & Levi-Civita symbols and repeated indices are to be summed over. If we choose $N = M = 8$, then the subdata can be chosen to be statistically independent and the error bar in the partial determinant is then simply

$$\begin{aligned}(\delta \nabla_{ab})^2 = \sum_{c,d} &\left(\frac{\partial \nabla_{ab}}{\partial A_{cd}} \delta A_{cd} \right)^2 + \left(\frac{\partial \nabla_{ab}}{\partial B_{cd}} \delta B_{cd} \right)^2 \\ &+ \left(\frac{\partial \nabla_{ab}}{\partial C_{cd}} \delta C_{cd} \right)^2 + \left(\frac{\partial \nabla_{ab}}{\partial D_{cd}} \delta D_{cd} \right)^2 .\end{aligned} \tag{3.42}$$

Otherwise, if a smaller number of states and detectors is considered, one must take into account that the entries are not all statistically independent when calculating error bars.

Examples of Correlated Models

A significant feature of these partial determinants is that they can be estimated directly from measured data. Stressing the point explicitly: "$\nabla(S) = 0$" may serve as an operational definition for

S to have no state-detector correlations. Certainly, if one finds that $\nabla(S) \neq 0$ with high confidence, then one has made a definitive statement that there are state-detector correlations. It is perhaps helpful if we illustrate this with some examples of correlated error models. For the sake of simplicity, let us restrict our attention to measurement devices whose biases may be chosen to all vanish.

Gaussian Model

One way to model correlation is to treat the Bloch vectors for all states and detectors as randomly fluctuating:

$$\begin{aligned} \vec{p}_\mu &= \vec{p}_{\mu o} + \vec{f}_\mu \\ \vec{w}_i &= \vec{w}_{io} + \vec{g}_i \end{aligned} \tag{3.43}$$

where the $\vec{p}_{\mu o}$ and $\vec{w}_{io}$ are constants and $\vec{f}_\mu$ and $\vec{g}_i$ are Gaussian noise with

$$\langle \vec{f}_\mu \rangle = \langle \vec{g}_i \rangle = 0. \tag{3.44}$$

Our data is then of the form:

$$S = P_o W_o + X \tag{3.45}$$

where

$$X_{\mu i} \equiv \langle \vec{f}_\mu \cdot \vec{g}_i \rangle \tag{3.46}$$

is the scalar noise correlation.

As a simple special case, consider $X = \chi 1_8$ where $\chi \ll 1$. We then have

$$S \equiv \begin{bmatrix} A & B \\ C & D \end{bmatrix} + \chi 1_8 = \begin{bmatrix} A + \chi 1_4 & B \\ C & D + \chi 1_4 \end{bmatrix} \tag{3.47}$$

where $A^{-1}BD^{-1}C = \Delta(P_o W_o) = 1_4$ so

$$\Delta = 1_4 - \chi(A^{-1} + C^{-1}D^{-1}C)$$

$$\Delta^{-1} = 1_4 + \chi(C^{-1}B^{-1}A + C^{-1}DB^{-1})$$

and

$$\nabla = -\chi(\det D)\Big(A(BC)^{-1}A - 1_4\Big). \tag{3.48}$$

As expected, the entries of ∇ are linear in the correlation, χ.

Causal Correlations

A second model considers data of the form

$$S_{\mu i} = \vec{p}_\mu \cdot \vec{w}_{i|\mu} \tag{3.49}$$

where the states, $\vec{p}_\mu$, are constants but each detector has a value conditioned on the state they are detecting. Correlations of this type are causal in the sense that the state preparation event is considered to be prior to the detection event and therefore the former cannot depend on the latter. More specifically, one could have, for example,

$$\vec{w}_{i|\mu} = \vec{w}_{io} + \chi\,\vec{p}_\mu \tag{3.50}$$

where $\vec{w}_{io}$ is fixed and $\chi \ll 1$ is a sort of scalar outcome susceptibility. In this case, the data is of the form

$$S = PW_o + Y \tag{3.51}$$

where

$$\begin{aligned} P &\equiv \begin{bmatrix} \cdots & \vec{p}_\mu & \cdots \\ \cdots & 1 & \cdots \end{bmatrix}^T, \\ W_o &\equiv \begin{bmatrix} \cdots & \vec{w}_{io} & \cdots \\ \cdots & 0 & \cdots \end{bmatrix}^T, \\ Y_{\mu i} &\equiv \chi|\vec{p}_\mu|^2. \end{aligned} \tag{3.52}$$

One can, obviously, construct additional correlated models containing larger numbers of adjustable parameters. Fitting actual data to the different models and applying model selection would allow

us to find the best description of the correlations present in our experiment (as explained further in Ref.in a similar context).

Conclusions

We considered a simple version of "gate-set tomography" [11], in which only state preparations and measurements (SPAM) are performed, and no gates in between. This simplification allowed us to fully analyze errors in SPAM and especially correlations between such errors. We found a simple test for detecting such correlated errors in terms of a "partial determinant." By considering the full set of gauge parameters (i.e., parameters not determinable by any (set of) SPAM experiment(s)), we found a novel interpretation of standard quantum state tomography and of the more recently developed detector topography: both can be seen as fixing the gauge parameters in their own particular way: the former by assuming fiducial measurements are fully known, the latter by assuming fiducial states are fully known.

Acknowledgments

SJvE was supported in part by LPS/ARO.

CHAPTER IV

THE BORN RULE AS A CONNECTION BETWEEN EXPERIMENTS

From Jackson, C. and van Enk, S. 2017. Non-holonomic tomography I: The Born rule as a connection between experiments. *arXiv preprint* arXiv:1702.00118

Abstract

In the context of quantum tomography, we recently introduced a quantity called a partial determinant [13]. PDs (partial determinants) are explicit functions of the collected data which are sensitive to the presence of state-preparation-and-measurment (SPAM) correlated errors. As such, PDs bypass the need to estimate state-preparation or measurement parameters individually. In the present work, we suggest a theoretical perspective for the PD. We show that the PD is a holonomy and that the notions of state, measurement, and tomography can be generalized to non-holonomic constraints. To illustrate and clarify these abstract concepts, direct analogies are made to parallel transport, thermodynamics, and gauge field theory. This paper is the first of a two part series where the second paper[15] is about scalable generalizations of the PD in multiqudit systems, with possible applications for debugging a quantum computer.

Introduction

In quantum computing, a recent problem has been learning how to estimate quantum gates while taking into account that there are small but significant errors in the states prepared and measurements made to probe such gates, so called SPAM errors [10]. Several works have come out to solve this, [10–12], all of which speak to the notion of a "self-consistent tomography." These works also make an important common assumption: that the uncontrolled fluctuations in the SPAM are not correlated. So in [13] the obvious question was asked: what if the states and measurements made were actually correlated with each other?

Even though this question can be asked for classical systems, this is an especially interesting question for quantum systems. Standard quantum theory tells us that reality articulates itself as discrete events. The probabilities of these events are further understood to be the product of *two* things: a state and a set of possible outcomes. More precisely, the Born rule in its modern form tells us that the distribution of these events is the inner product of a density operator and a POVM.

This is what makes a quantum theory distinct from a classical one as it allows for fundamental randomness because a state is no longer an outcome in itself: state and outcome become distinct notions. As distinct as these notions are, they are nevertheless inseparable because each quantum event measured is always and only the product of a state and a possible outcome — a fact which is especially apparent in tomography. Put another way, there is no quantum state defined operationally, independent of the resource of known possible outcomes and vice versa.

This brings up a fundamentally important point which is that the concept of states and observables as separate and independent is a subjective or man-made distinction, reflecting the model of standard quantum theory. In the presence of SPAM correlations, average state and average measurement parameters cannot be defined as statistically independent quantities, consistent with all possible state and measurement settings. However, one can still define average state and average measurement parameters *locally* over the space of device settings. A simple but subtle example of such locally defined quantities can be found in thermodynamics — the caloric, "Q", and potential energy, "W", represented by inexact heat and work forms which sum to changes in the energy, $dU = đQ + đW$, which is globally defined over the thermodynamic state space. A more standard example can be found in quantum electrodynamics — the electron kinetic momentum, $-iD_\mu$, and the photon vector potential, A_μ, which sum to the canonical momentum, $-i\partial_\mu = -iD_\mu + A_\mu$, globally defined over position space.

In order to illustrate these analogies explicitly, we will consider a toy analogy to quantum tomography with SPAM errors. This toy model replaces the state and measurement with single parameters, which can be correlated. We demonstrate precisely how the toy analog of the partial determinant from [13] has the same structure as $\oint đQ$ from thermodynamics or $\oint A \cdot dx$ from QED. Such "loop" integrals are generally called holonomies and the forms they integrate can be referred to as non-holonomic constraints. Finally, we translate these results to actual quantum tomography, completing the perspective of non-holonomic quantum tomography.

State-Preparation, Measurement, their Correlation, and Data

The Born Rule and Tomography

One could say that the Born rule was originally, since the 1920s, used exclusively to *predict* the distributions of events *from* states and observables. Standard textbook treatments will denote the

Born rule by $P(s|\psi) = |\langle s|\psi\rangle|^2$, thus introducing the notions of state and measurement outcome. Statistical observables[1] are then calculated from classical probability theory and typical expressions like

$$\langle s\rangle = \sum_s sP(s) = \sum_{s,\psi} sP(s|\psi)P(\psi) = \mathrm{Tr}\left(\sum_\psi P(\psi)|\psi\rangle\langle\psi|\right)\left(\sum_s s|s\rangle\langle s|\right) = \mathrm{Tr}\rho\Sigma = \langle\Sigma\rangle \qquad (4.1)$$

appear, introducing the notions of a classically mixed state and a quantum observable. Since distinct quantum systems can interact, the notion of an ancilla can be introduced and measurements can be generalized from an orthonormal basis to a positive operator valued measure (POVM).

In more recent years, the Born rule has found a different application in so called quantum state tomography [33, 37], where states are concluded from the distribution of measured events and various known POVMs. After this, it was quickly recognized that the Born rule could just as well be used for so called detector tomography [45, 46], where POVM elements are concluded from the distribution of events and known states. It had even been demonstrated that one could perform state tomography through unknown POVMs from other known states with a technique similar to applying the Born rule twice, bypassing the need to parameterize unknown POVMs [1].

Any application of the Born rule where *both* state preparation and measurement are unknown [10–12] we will henceforth refer to as SPAM tomography. The central feature which makes SPAM tomography distinct from other tomographies is the presence of gauge degrees of freedom. In this case, state and measurement parameters are explicitly inseparable because the Born rule cannot uniquely determine them individually from the statistics alone. Work has been done to recover unique estimates for individual state and measurement parameters [12] under further assumptions. Of course, such work also makes the implicit assumption that there are no correlated SPAM errors.

In the context of our work, where we *do* allow for correlated SPAM errors, a crucial point must be made concerning our use of the $\langle\rangle$ notation. On the leftmost side of Equation (4.1), $\langle\rangle$ refers to the expectation value of a random variable, s. On the rightmost side of Equation (4.1), $\langle\rangle$ loses this meaning as it does *not* refer to the expectation value of an operator, Σ, but rather an inner product of Σ with the density operator. In both cases, the *distribution of quantum events is completely*

[1]There is a slight conflict of language here as modern field theorists like to use "observable" to refer to cross-sections, lifetimes, etc. which we refer to as "statistical observables" as opposed to "quantum observables" which refer to operators in a theory and what we mean throughout this paper by "observable."

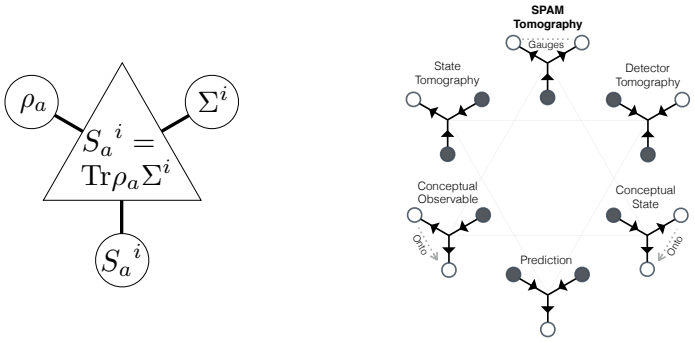


FIGURE 4. Schematic diagram for the various perspectives of the Born rule. Left: In the most general sense the Born rule is simply a constraint between states (ρ_a), observables (Σ^i), and data ($S_a{}^i$). Right: The six perspectives of the Born rule — states, observables, and data are represented spatially as in the left diagram and darker corners represent parameters that are fixed externally. "Prediction" fixes states and observables to conclude measured data. "State Tomography" uses fixed observables and data to conclude states. "Detector Tomography" uses fixed states and data to conclude observables. Dual to "State Tomography" is the "Conceptual State", where a fixed state is understood as a map from observables onto their 'expectation value'. Dual to "Detector Tomography" is the "Conceptual Observable", where a fixed observable is understood as a map from states onto their 'expectation value'. Finally, dual to "Prediction" is "SPAM Tomography", where state-observable relationships are concluded from fixed data.

attributed to the state and this assumption is perhaps further obscured by Dirac's bra-ket notation. For our purposes in SPAM tomography, we will not use $\langle\rangle$ in this way, beyond Equation (1). Rather, $\langle\rangle$ will refer to an expectation value where states and observables are themselves considered random variables. Specifically, if ρ is a density operator representing the state and E is a POVM element representing a possible outcome, then one must understand that

$$f = \langle \mathrm{Tr}\rho E\rangle \tag{4.2}$$

where f is an estimate of the probability (obtained from the frequency of the measured outcome) and $\langle\rangle$ is the average over the ensemble of trials. The measured frequency is SPAM correlated if

$$\langle \mathrm{Tr}\rho E\rangle \neq \mathrm{Tr}\langle\rho\rangle\langle E\rangle. \tag{4.3}$$

We will examine such correlations in a much simpler context in the next section but details and examples may also be found in [13].

A Toy Example

The problem of whether states and measurements are correlated is fundamentally interesting because states and observables are not individually accessible in principle by experiment alone. In other words, the quantities of the right hand side of Equation (4.3) cannot be measured without arbitrarily well characterized devices. Nevertheless, it was demonstrated in [13] that there is still a way to detect such correlations using properties of the data alone, bypassing the need to estimate state and observable parameters separately. The basic essence of that result can be illustrated by the following toy problem:

FIGURE 5. On the left is a device which prepares various signals on demand depending on which button, $a \in \{1, \ldots, N\}$, is pressed. On the right is a device which blinks to indicate a signal with a certain property depending on which setting, $i \in \{1, \ldots, M\}$, a dial is turned to.

Consider a device with various settings, a, each of which prepare a different signal on demand by the press of a button. Consider also a detector with various settings, i, each of which detect a particular property of the signal indicated by the blink of a light. Now suppose it is suspected that each setting of the preparation device actually produces the same signal, only that each setting produces the signal with varying probabilities of success, p_a. Suppose further that the detection device is expected to simply indicate the presence of the signal, only that each setting can register a signal with varying probabilities of success, w^i. We then imagine that p_a and w^i are actually unknown and that we are only able to change settings and record whether the light blinks or not.

Let us indicate by ${f_a}^i$ the measured frequency with which the light blinks when the devices are set to (a, i). If one can assume that the performance of the devices and their settings are uncorrelated, then one can simply identify (after many runs of the experiment)

$$ {f_a}^i = p_a w^i. \tag{4.4} $$

However, relaxing this assumption to allow for the possibility of correlations, one must be more careful about the quantities defined so to make the more general identification that

$$f_a{}^i = \langle pw \rangle_a{}^i \tag{4.5}$$

where we have introduced notation $\langle \rangle_a{}^i$ to represent the average over the ensemble of trials for the pair of settings (a, i). The subtlety here is that the devices can still be represented by single parameters, p and w, only now these parameters are to be understood as random variables which fluctuate depending on the setting (a, i).

The presence of SPAM correlation is simply when the frequencies, $\langle pw \rangle_a{}^i$ (which are what we have access to) are such that

$$\langle pw \rangle_a{}^i \neq \langle p \rangle_a \langle w \rangle^i. \tag{4.6}$$

It would seem that to identify such a circumstance one would have to measure $\langle p \rangle_a$ and $\langle w \rangle^i$ individually. However, such measurements would require devices which are already well characterized, unlike the devices we have. What we would like to do is detect if such correlations are present between our devices given only our humble, imperfect, uncalibrated devices.

Gauge Degrees of Freedom

In such a situation, one must acknowledge that there will always be so called gauge degrees of freedom. If one was given the promise that a pair of device parameters were in fact SPAM uncorrelated, then there would still be a one-parameter family of possible values for the average state parameter and average detector parameter. Specifically, for a possible pair of values $\big(\langle p \rangle, \langle w \rangle\big)$ such that $\langle pw \rangle = \langle p \rangle \langle w \rangle$, the pair $\big(g\langle p \rangle, g^{-1}\langle w \rangle\big)$ is just as possible.[2] If the devices are SPAM uncorrelated over a range of settings $a \in \{1, \ldots, N\}$ and $i \in \{1, \ldots, M\}$, then the set of possible average values continue to define exactly one gauge parameter. This is perhaps best illustrated by

[2]Of course, g must has a compact range so that the interpretation of $\big(g\langle p \rangle, g^{-1}\langle w \rangle\big)$ as a pair of probabilities still makes sense. However, this detail is not of concern for this paper.

observing $\langle pw\rangle_a{}^i = \langle p\rangle_a \langle w\rangle^i$ as a matrix equation,

$$\begin{bmatrix} \langle pw\rangle_1{}^1 & & \langle pw\rangle_1{}^M \\ & \ddots & \\ \langle pw\rangle_N{}^1 & & \langle pw\rangle_N{}^M \end{bmatrix} = \begin{bmatrix} \langle p\rangle_1 \\ \vdots \\ \langle p\rangle_N \end{bmatrix} \begin{bmatrix} \langle w\rangle^1 & \cdots & \langle w\rangle^M \end{bmatrix}, \tag{4.7}$$

so that if $\left(\left[\langle p\rangle_1 \cdots \langle p\rangle_N\right]^{\mathsf{T}}, \left[\langle w\rangle^1 \cdots \langle w\rangle^M\right] \right)$ is possible, then so is $\left(g \left[\langle p\rangle_1 \cdots \langle p\rangle_N\right]^{\mathsf{T}}, g^{-1} \left[\langle w\rangle^1 \cdots \langle w\rangle^M\right] \right)$.

To handle this gauge degree of freedom, it is useful to define the following notion: The collected data, $\langle pw\rangle_a{}^i$, for a pair of devices is *effectively* (SPAM) uncorrelated if Equation (4.7) exists — that is, if the experimentally accessible left-hand side can be expressed as in the right-hand side for some $[\langle p\rangle_a]^{\mathsf{T}}$ and $\left[\langle w\rangle^i\right]$. Considered as a matrix, $D = [\langle pw\rangle_a{}^i]$, one should recognize that this definition is equivalent to an upper bound on the rank, $\mathrm{rank}(D) \leq 1$. Such a bound on the rank can be further quantified by considering the determinant of every 2×2 submatrix of the data, so called (2×2) minors. Specifically, every such minor must be zero if the data is effectively uncorrelated. One should recognize that such conditions are properties of the data collected by just our humble devices alone.

Having mentioned some standard notions from linear algebra, there is an alternative set of notions which support the same analysis. These notions are also more geometric in their perspective, which one might have suspected to exist from the association of gauge. The technique which accompanies these notions further has an obvious tomographic interpretation. As a final statement of this prelude, the alternative technique we are referring to is also what generalizes to actual quantum tomography.

Partial Determinants

To demonstrate, we will need only to consider two settings per device, $N = M = 2$. For simplicity, let us denote the quantities $\langle pw\rangle_1{}^1$, $\langle pw\rangle_1{}^2$, $\langle pw\rangle_2{}^1$, and $\langle pw\rangle_2{}^2$ by simply $\langle pw\rangle$, $\langle pv\rangle$, $\langle qw\rangle$, and $\langle qv\rangle$, respectively and refer to them as *data*. Further, let us denote the settings $(1,1)$, $(1,2)$, $(2,1)$, and $(2,2)$ respectively as (p,w), (p,v), (q,w), and (q,v) and refer to them as *experiments*.

If one considers only the measured quantity $\langle pw\rangle$, then such a datum is always effectively uncorrelated and should thus be associated with a gauge degree of freedom. This is true as well for the other data, $\langle pv\rangle$, $\langle qw\rangle$, and $\langle qv\rangle$ considered individually. Considering these data individually means that they have the property of being (effectively) uncorrelated *locally*, in which case each

experiment should be understood to correspond to a local gauge degree of freedom. Explicitly, "local" is relative to the space of experimental settings which here consists only of 4 points (though we will consider the continuous case soon enough in the next section.)

Each gauge degree of freedom is arbitrary in the sense that they cannot be defined without the resource of better calibrated devices.[3] However, these gauge degrees of freedom are still related to each other because the experiments can share common settings. For example, let us parameterize the gauge of the experiment (p, w) with $\langle w\rangle$ and the gauge of (p, v) with $\langle v\rangle$. Since these two experiments share the setting p, their corresponding gauge degrees of freedom are related by the data because $\langle v\rangle = \frac{\langle pv\rangle}{\langle pw\rangle}\langle w\rangle$. In other words, the data can be interpreted as a *connection* between the gauge of each experiment. The connection is itself not uniquely determined by the data, but this is only because it is intimately related to the gauge — e.g. if we had instead parameterized the the gauge of experiment (p, w) with $\langle p\rangle$, then the above connection would have been rather $\langle v\rangle = \langle pv\rangle/\langle p\rangle$.

The gauge of each experiment (p, w) represents the fact that the corresponding data $\langle pw\rangle$ is locally (effectively) uncorrelated. Nevertheless, it may still be the case that the data of all four experiments is not (effectively) uncorrelated *globally* so that one may not be able to write

$$D = \begin{bmatrix} \langle pw\rangle & \langle pv\rangle \\ \langle qw\rangle & \langle qv\rangle \end{bmatrix} = \begin{bmatrix} \langle p\rangle \\ \langle q\rangle \end{bmatrix} \begin{bmatrix} \langle w\rangle & \langle v\rangle \end{bmatrix} \tag{4.8}$$

simultaneously. As observed earlier, such data is globally (effectively) uncorrelated if and only if $\det D = 0$. Assuming $\langle pw\rangle\langle qv\rangle \neq 0$, the $\det D = 0$ condition is equivalent to

$$\Delta(D) \equiv \frac{\langle pv\rangle\langle qw\rangle}{\langle pw\rangle\langle qv\rangle} = 1 \tag{4.9}$$

and it is this quantity which generalizes to the full quantum problem.[13] Since Δ is only a function of data, it is manifestly gauge invariant. Δ is called a *partial determinant* because of the analogy to the above problem and because it is not generally a single number, but rather a matrix of reduced size ($d^2 \times d^2$ for d-dimensional Hilbert spaces.)

Restating the (toy) result,

$$D \text{ is globally (effectively) uncorrelated if and only if } \Delta(D) = 1,$$

[3] One can argue that such devices do not exist other than by assumption!

The reader may be familiar with a proof of this using the language of standard linear algebra[4] (considering D as an operator and considering its null space, etc.) However to emphasize the perspective, we include here a more tomographic proof: The "only if" can be proved by simple substitution. For the "if" direction, one first remembers that they can always choose $\langle p\rangle$ and $\langle w\rangle$ such that $\langle pw\rangle = \langle p\rangle\langle w\rangle$. Having chosen $\langle p\rangle$ and $\langle w\rangle$, one may then fix $\langle q\rangle = \langle qw\rangle/\langle w\rangle$ and $\langle v\rangle = \langle pv\rangle/\langle p\rangle$. Notice that this fixing of $\langle q\rangle$ and $\langle v\rangle$ is analogous to state and detector tomography. Finally, if $\Delta(D) = \langle pv\rangle\langle qw\rangle\langle pw\rangle\langle qv\rangle = 1$ then $\langle qv\rangle = \langle pv\rangle\langle qw\rangle/\langle pw\rangle = \langle q\rangle\langle w\rangle$, which finishes the proof.

Summarizing, we have developed a perspective for analyzing toy data which parallels a perspective for analyzing quantum data: Considering the settings (p, w), (q, w), (p, v), and (q, v) as individual experiments, these settings act as *coordinates* for the space of experiments so that one can say, for example, experiments (p, w) and (q, w) are displaced from each other by keeping the measurement setting constant. Further, each individual experiment is *effectively uncorrelated* because we can always choose $\langle p\rangle$ and $\langle w\rangle$ such that $\langle pw\rangle = \langle p\rangle\langle w\rangle$. The freedom of that choice is a *gauge* degree of freedom and is further a *local* one because each experiment has this property. Finally, there is a *connection* between the gauges of each experiment because we can write equations like $\langle p\rangle = \langle pw\rangle/\langle w\rangle$ — that is, a choice of $\langle w\rangle$ fixes the gauge of experiment (p, w) which consequently fixes the gauge of experiment (p, v). With this connection, the partial determinant has the interpretation of performing tomography in a loop, with a value which measures a contradiction (see Figure 6), reflecting the presence of SPAM correlation. Cast in this language, we have demonstrated that a PD is a *holonomy*. We shall proceed to explain this further. At last, it is the tomographic interpretation of this holonomy which is why we refer to any analysis with PDs *non-holonomic tomography*.

Holonomy

Holonomy is a concept which has become quite ubiquitous in modern physics and mathematics. Applications range from geometric phases to Yang-Mills Lagrangians, all of which share the notion of a non-holonomic constraint. Perhaps the simplest physical examples of non-holonomic constraint are the thermodynamic concepts of heat and work, although thermodynamics is typically not considered in this way. The simplest mathematical example is probably parallel transport through a sphere,

[4]Remember that this result is equivalent to the commonly known property that D is rank ($\leq$) 1 if and only if $\mathrm{Det}D = 0$.

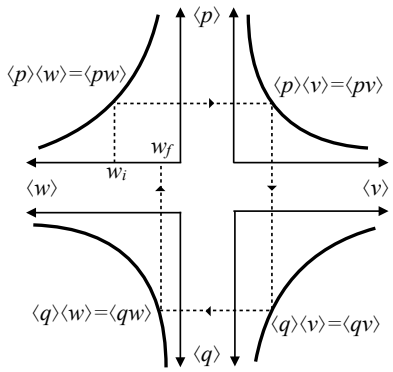


FIGURE 6. Illustration of the PD as a Holonomy: Each experiment (p, w) has a local gauge degree of freedom because it is effectively SPAM uncorrelated, $\langle p\rangle\langle w\rangle = \langle pw\rangle$. The data $\langle pw\rangle$ further provides a connection between adjacent gauge degrees of freedom by the assumption that they share independent settings. Such a connection defines a non-holonomic constraint when $w_f = \frac{\langle pv\rangle\langle qw\rangle}{\langle pw\rangle\langle qv\rangle} w_i \neq w_i$. A particular w_i fixes the gauge which can either represent an arbitrary choice or some external information. The PD $\Delta = \frac{w_f}{w_i}$ is gauge invariant.

where a tangent vector will turn with an angle proportional to the solid-angle subtended by the loop traversed (Figure 7.)

Characteristic of these non-holonomic systems are local degrees of freedom (such as heat or angle) whose differential can be integrated over contours defined within certain dimensions (such as the thermodynamic state or the point on a sphere.) However, these integrals will have non-zero values over *closed* contours, reflecting that these local degrees of freedom cannot be globally defined as additional dimensions like the ones which defined the contour. Such integrals are called holonomies and their non-zero values may be interpreted as a measure of contradiction or inability to integrate the local degree of freedom to a global coordinate.

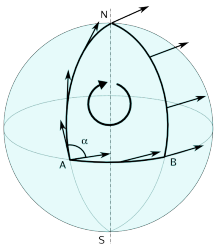


FIGURE 7. Probably the most familiar example of holonomy is the parallel transport of a tangent vector on the sphere.

The technical notion of heat as a holonomy is not standard and so an elaboration is in order. This will allow us to draw an analogy from which the perspective of non-holonomic tomography

will be more explicit. Using the language of gauges in such a non-standard way, it will also be appropriate to relate these notions to their more familiar application in gauge field theory. After having established theses connections (no pun intended) we will then rewrite non-holonomic tomography in this field theoretic language. For completeness, we include a section on the actual quantum analogue of the toy problem to make all the respective technical aspects clear.

Analogy: Thermodynamics

For a thermodynamic system such as an ideal piston, the notion of an adiabatic process can be defined but cannot be extended to a notion of heat as a quantity. This is because heat can be transferred (into other forms of energy) over closed loops in state space (see Figure 8.) This transfer of heat is the holonomy and the integrals $\int_\gamma đQ$ are the connection. Put another way, the connection $\int_\gamma đQ$ can be thought of as a change in some quantity (like caloric), ΔQ, but only locally because one can have nonzero changes in the heat upon a *return* to the same state.

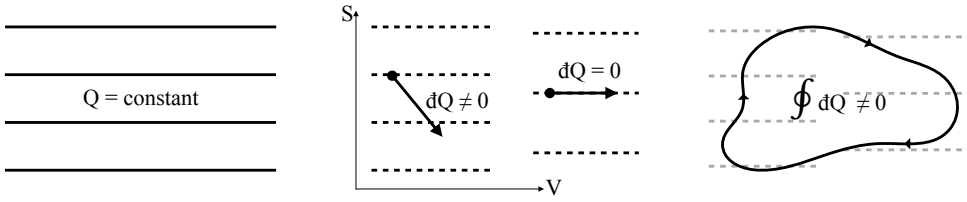


FIGURE 8. Left: Holonomic constraints can be written globally and therefore used as coordinates. Middle: Non-holonomic constraints are only local and cannot define coordinates. The dashed lines are supposed to convey that a notion of "transverse" is still present but the distance between the layers of constraint can be correlated with coordinates along the layers. Right: Non-holonomic constraints thus give rise to holonomies or non-zero integrals over closed contours.

However, the notions of energy and entropy do exist as globally defined state variables and heat can be thought of as the energetic response generated by changes in entropy,

$$đQ = TdS. \tag{4.10}$$

The coefficient of response is the temperature which can depend on other degrees of freedom within the state space, such as volume:

$$T(S,V) = \left.\frac{\partial U}{\partial S}\right|_V \tag{4.11}$$

This extra dependence on other degrees of freedom is what makes $đQ$ non-holonomic, non-integrable, or inexact (words which are synonymous in this context.) For such a temperature that depends on

volume, one could say that the energy transfer generated by a fixed displacement in entropy is *correlated* with the volume.

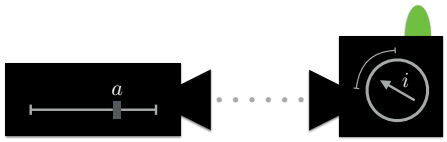


FIGURE 9. Our state and measurement devices, now with continuous settings!

Similarly, as in Figure 6, we know what it means to keep the "state device setting" constant so that we may coordinate (p, w) & (p, v) or (q, w) & (q, v) as being in the same layer. We even have the notion of an "average state parameter change" generated by an "iso-measurement-ic" process because we can write

$$\langle q\rangle = \frac{\langle qw\rangle}{\langle pw\rangle}\langle p\rangle \qquad \text{or} \qquad \langle q\rangle = \frac{\langle qv\rangle}{\langle pv\rangle}\langle p\rangle. \tag{4.12}$$

Further, such an "average state parameter change" may not be holonomic because one could have

$$\frac{\langle qw\rangle}{\langle pw\rangle} \neq \frac{\langle qv\rangle}{\langle pv\rangle} \tag{4.13}$$

so that the response in the "average state parameter" with respect to changes in the "state device setting" is a function of "measurement device setting." Importantly, the isomorphism from the ideal piston to SPAM tomography is algebraically exponential — that is, for example,

$$\frac{\langle qw\rangle}{\langle pw\rangle} \sim \exp\int đQ. \tag{4.14}$$

Ideal Piston	Toy SPAM Tomography
State Space (S, V)	Device Setting Space (a, i)
Entropy, S	"State Device Setting", a
Volume, V	"Measurement Device Setting", i
Energy, $U(S, V)$	Data, $\log \langle pw\rangle_a{}^i$
Temperature, $T = \frac{\partial U}{\partial S}\big\vert_V$	Response, $\chi = \frac{\partial}{\partial a}\big\vert_i \log\langle pw\rangle$
Pressure, $P = -\frac{\partial U}{\partial V}\big\vert_S$	Response, $\xi = -\frac{\partial}{\partial i}\big\vert_a \log\langle pw\rangle$
Heat, $đQ = TdS$	Average State Parameter Change, $đ\log\langle p\rangle = \chi da$
Adiabatic/Isentropic	"Iso-state-ic"
Work, $đW = -PdV$	Average Measurement Parameter Change, $đ\log\langle w\rangle = -\xi di$
Isochoric	"Iso-measurement-ic"

TABLE 1. A table to help with the corresponding terms in the Piston-SPAM analogy.

Indeed, this analogy can be made even more exact (see Table 1 and Figures 9 and 10.) Returning to our toy devices, suppose instead that the state and observable settings could be dialed continuously and call these external parameters a & i respectively. Assuming that a & i are the only controls, then the data $\langle pw \rangle$ is a well defined function over the space of (a, i). We can also define responses in the data with respect to these parameters:

$$\chi = \frac{\partial}{\partial a}\Big|_i \log\langle pw\rangle \qquad \text{and} \qquad \xi = -\frac{\partial}{\partial i}\Big|_a \log\langle pw\rangle. \tag{4.15}$$

These responses provide equations of state which we may then attribute to notions of non-holonomic average state parameter & average measurement parameter changes,

$$đ \log\langle p\rangle = \chi(a,i)da \qquad \text{and} \qquad đ \log\langle w\rangle = -\,\xi(a,i)di, \tag{4.16}$$

which are related to the original data:

$$d \log\langle pw\rangle = đ \log\langle p\rangle + đ \log\langle w\rangle. \tag{4.17}$$

The exponential maps between the finite and the infinitesimal processes may now be written explicitly:

$$\frac{\langle qw\rangle}{\langle pw\rangle} = \exp\left(\int_p^q \chi(a,w)da\right) \qquad \text{and} \qquad \frac{\langle pv\rangle}{\langle pw\rangle} = \exp\left(\int_w^v \xi(p,i)di\right). \tag{4.18}$$

Finally, we have for the partial determinant

$$\Delta = \frac{\langle pv\rangle\langle qw\rangle}{\langle pw\rangle\langle qv\rangle} = \exp\left(\oint đ \log\langle p\rangle\right) = \exp\left(-\oint đ \log\langle w\rangle\right) = \exp\left(\iint \Gamma\, da di\right) \tag{4.19}$$

where the integrals are counterclockwise in Figure 10 and

$$\Gamma = \frac{\partial \chi}{\partial i} = -\frac{\partial \xi}{\partial a} = -\frac{\partial^2 \log\langle pw\rangle}{\partial a \partial i} \tag{4.20}$$

is a kind of correlation density.

When considering this treatment for the response of quantum data to continuous device settings, p and w become $d^2 \times d^2$ matrix quantities, P and W such that $D \stackrel{?}{=} PW$, representing minimally complete tomography experiments for a d-dimensional Hilbert space, as will be explained

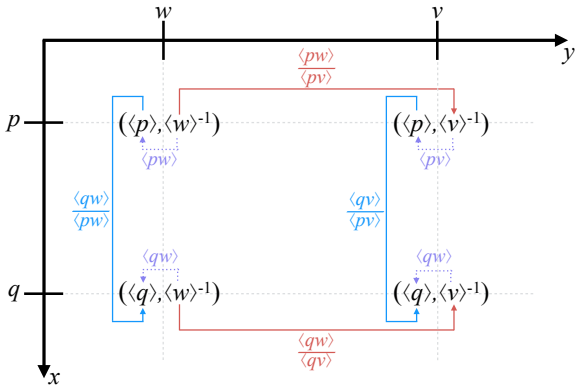


FIGURE 10. An "S-V" diagram for toy non-holonomic tomography. Ratios between horizontally adjacent data can be interpreted as"iso-states-ic" processes and vertical ratios as "iso-observables-ic". These processes are non-holonomic and so demote the notions of "average state" and "average observable" from physical coordinates to a gauge degree of freedom.

in section 4.4. As such, the inexact forms in Equation (4.17) should be replaced with the forms $(d\langle P\rangle)\langle P\rangle^{-1}$ or $\langle W\rangle^{-1}d\langle W\rangle$. These forms may be recognized as Maurer-Cartan forms for the Lie Group $GL(d^2)$ or also the GL-equivalent of Mead-Berry Potentials.

Analogy: Interactions of a Single Quantum with a Gauge Field

Perhaps the most effective (no pun intended) place to start here is with the gauge interaction of a single electron in an external electromagnetic field.[5] The wavefunction can be written as a path integral,

$$\Psi[\gamma_1,\gamma_0;A]=\int\mathcal{D}\gamma\, e^{iq\int_\gamma dx\cdot A}e^{iS_o[\gamma]}, \tag{4.21}$$

where A is the 4-vector potential, S_o is the action for the electron in no field, and the integral is over all paths with initial and final spacetime events γ_0 & γ_1, respectively. We will not be interested in the spacetime dependence here and will thus denote the wavefunction as just $\Psi[A]$. On the other hand, the field dependence has a gauge degree of freedom represented by the (projective) symmetry,

$$\Psi[A+\partial\zeta]=e^{iq(\zeta(\gamma_1)-\zeta(\gamma_0))}\Psi[A] \tag{4.22}$$

[5]Indeed, we could just as well have a discussion about general partition functions in statistical mechanics. Their dependence on reservoir parameters can be probed with the mode of the ensemble distribution. The conclusions of such a discussion would have the same essence as the previous section with only the advantage of a technically broader perspective. The logic would exactly parallel the following discussion so we will not go further than to simply acknowledge its existence.

which (according to the Born rule) leaves the transition rate between events at γ_0 & γ_1 invariant. The potential A is also called a *connection* because it fixes the phase of the wavefunction at γ_1 relative to the phase at γ_0.

The dominant contribution to the wavefunction is from the path satisfying the classical equation of motion, $\frac{\delta S_o}{\delta \gamma} = qF\dot{\gamma}$. If S_o is the free particle action, then the equation of motion is just the Lorentz force law. However, we could just as well incorporate external interactions into S_o which overpower the Lorentz force and fix γ arbitrarily to $\frac{\delta S_o}{\delta \gamma} = 0$. In which case we can write the wavefunction with a classical approximation,

$$\Psi[\gamma, A] \propto e^{iq\int_\gamma dx \cdot A} \tag{4.23}$$

where it is understood now that γ can be fixed arbitrarily. We do not bother with the normalization constant or the external phase here because we wish only to illustrate the dependence of the wavefunction on A which we can now imagine is being probed through γ, which can be externally controlled.

The quantity

$$W_\gamma = e^{iq\int_\gamma dx \cdot A} \tag{4.24}$$

is called a Wilson line. Also important is the Wilson loop

$$W_\gamma = \mathrm{Tr}\left(e^{iq\oint_\gamma dx \cdot A}\right) \tag{4.25}$$

where a trace has been introduced to include non-abelian gauge fields where there are several As, one for each generator of the gauge group. The general wavefunction, Equation (4.21), is often referred to as the "quantum expectation value" of the Wilson loop in this context. Normally, the application of the Wilson loop is to determine the dynamics of γ from a theory of the gauge field. However, our purpose for the Wilson loop is to represent how the gauge field could be probed by an externally fixed γ. (See Figure 4)

When we consider partial determinants in section 4.4, the analogous quantity will be just the closed Wilson line, i.e. a Wilson loop without the trace. Aside from the difference between a single number and a matrix, an important distinction is that closed Wilson line actually depend on the initial/final point from which γ is drawn, while Wilson loops do not. However, the dependence is

simple and only such that the closed Wilson line is gauge covariant instead of invariant

$$W_\gamma \longrightarrow U(\gamma_1) W_\gamma U^{-1}(\gamma_0) \tag{4.26}$$

where $\gamma_1 = \gamma_0$ for a closed contour. Although this does not have any significance in gauge field theories, it is significant for a theory of SPAM correlations.

Analogous to a Wilson line, one can define a tomography line:

$$\Delta\big(\gamma, \tau\big) = \exp \int_\gamma \tau. \tag{4.27}$$

which represents a specific type tomography, where the gauge parameter of experiment γ_1 is concluded from the gauge parameter of experiment γ_0 through the data, represented by the connection τ, along changes in the device parameters, represented by the contour γ. The tomographic connection, τ, is not uniquely determined by the data but is nonetheless intimately related to the interpretation of the gauge at each experiment along γ. Formally this is represented by the tomography lines being equivalent by a local gauge transformation

$$\Delta\big(\gamma, \tau + dg\big) = e^{g(\gamma_1) - g(\gamma_0)} \Delta\big(\gamma, \tau\big) = e^{g(\gamma_1)} \Delta\big(\gamma, \tau\big) e^{-g(\gamma_0)} \tag{4.28}$$

where the effect of the transformation is only to relabel the initial and final gauge parameters.

Returning to our toy devices, suppose that the gauge at each experiment is represented by an average state parameter (one could call this fixing the *state* gauge.) Then for $da = 0$, Δ would be the identity, while along the a-direction

$$\Delta\big(\gamma, đ \log\langle p \rangle\big) = \exp\left(\int_\gamma \chi da\right) \tag{4.29}$$

would represent iso-measurement-ic tomography. Similarly, if the gauge at each experiment is represented by an average measurement parameter (let's call this *measurement* gauge), then

$$\Delta\big(\gamma, -đ \log\langle w \rangle\big) = \exp\left(\int_\gamma \xi di\right) \tag{4.30}$$

would represent iso-state-ic tomography. Most importantly, these tomographies are equivalent to each other modulo a local gauge transformation:

$$\Delta\big(\gamma, -đ\log\langle w\rangle\big) = \Delta\big(\gamma, đ\log\langle p\rangle - d\log\langle pw\rangle\big) \tag{4.31}$$

$$= \exp\left(-\int_\gamma d\log\langle pw\rangle\right)\Delta\big(\gamma, đ\log\langle p\rangle\big) \tag{4.32}$$

$$= \frac{\langle pw\rangle(\gamma_0)}{\langle pw\rangle(\gamma_1)}\Delta\big(\gamma, đ\log\langle p\rangle\big). \tag{4.33}$$

In the electromagnetism analogy, these are the equivalent of Landau gauges (see Figure 13.)

Non-Holonomic Quantum Tomography and Non-Abelian Lattice Gauge

Having hopefully made the perspective of non-holonomic tomography clear through these analogies for the toy problem, some discussion about the actual quantum problem is due.[13] The quantum problem is the same as the toy problem except that we assume the state and measurement devices are parameterized by Hermitian operators (a density operator and a POVM element, respectively) over a d-dimensional Hilbert space. In particular, this means that the devices are to be modeled by d^2 random variables each. If all the device parameters were uncorrelated, then one could write these operators as

$$\rho_a = \frac{1}{d}{p_a}^\mu\sigma_\mu \qquad \text{and} \qquad E^i = \sigma^\mu {w_\mu}^i \tag{4.34}$$

where the $\{\sigma_\mu\}_{\mu=0}^{d^2-1}$ is some operator basis of Hermitian operators, $\{\sigma^\mu\}$ is its reciprocal basis, and a sum over repeated indices is implied. If $\sigma_0 = 1$ and the other σ_μ are traceless, then ${p_a}^0$ and ${w_0}^i$ are identical to the single device parameters of the toy problem.

The measured frequencies, a.k.a. "the data", are now given by

$${f_a}^i = \langle \mathrm{Tr}\rho E\rangle_a{}^i = \langle p^\mu w_\mu\rangle_a{}^i. \tag{4.35}$$

To be effectively uncorrelated in this case means that the data can be decomposed into the form

$$F = \begin{bmatrix} \langle p^\mu w_\mu \rangle_1{}^1 & & \langle p^\mu w_\mu \rangle_1{}^M \\ & \ddots & \\ \langle p^\mu w_\mu \rangle_N{}^1 & & \langle p^\mu w_\mu \rangle_N{}^M \end{bmatrix} = \begin{bmatrix} \langle p^0 \rangle_1 & \cdots & \langle p^{d^2-1} \rangle_1 \\ & \vdots & \\ \langle p^0 \rangle_N & \cdots & \langle p^{d^2-1} \rangle_N \end{bmatrix} \begin{bmatrix} \langle w_0 \rangle^1 & & \langle w_{d^2-1} \rangle^M \\ \vdots & \cdots & \vdots \\ \langle w_0 \rangle^1 & & \langle w_{d^2-1} \rangle^M \end{bmatrix}, \tag{4.36}$$

which is equivalent to saying that the rank is bounded above by $\text{rank}(D) \leq d^2$. To define a partial determinant, the simplest way is to consider $M = N = 2d^2$ and partition the data into 4 $d^2 \times d^2$ corners,

$$F = \begin{bmatrix} A & B \\ C & D \end{bmatrix}. \tag{4.37}$$

The partial determinant is

$$\Delta(F) = D^{-1}CB^{-1}A \tag{4.38}$$

which is significant because of the result

$$\text{F is globally uncorrelated if and only if } \Delta(F) = 1.$$

Specifically, Δ parameterizes d^4 degrees of correlation. However, because of gauge covariance (Equation 4.26,) only d^2 of these are gauge invariant parameters.

In the quantum case, it becomes important to pay attention to the arrangement of the settings when the data is considered in the form of Equation (22) so let us define indices:

$$F = \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} D_0{}^0 & D_0{}^1 \\ D_1{}^0 & D_1{}^1 \end{bmatrix} \tag{4.39}$$

where the matrix elements of these corners are

$$(D_a{}^i)_\alpha{}^\iota = f_{ad^2+\alpha+1}{}^{id^2+\iota+1}. \tag{4.40}$$

The *corners* are coordinated by (a, i) and understood to be 2×2 minimally complete tomography experiments we call a *square*. Each minimally complete tomography experiment consists of d^2 states enumerated by α and d^2 measurements enumerated by ι and is further associated with d^4 gauge

degrees of freedom reflecting the fact that the data of each corner is locally (effectively) uncorrelated,

$$(D_a{}^i)_\alpha{}^\iota = (P_a)_\alpha{}^\mu (W^i)_\mu{}^\iota = (P_a)_\alpha{}^\mu G_\mu{}^\lambda G^{-1}{}_\lambda^\nu (W^i)_\nu{}^\iota. \tag{4.41}$$

The corners are understood to be displaced from each other through changes in (a, i) and it is useful to think of these indices as pairs of points on a continuum (see Figure 11.) As such, the data matrix is conceptually reorganized as a square which has gauges at each corner (experiment) which are connected to each other by the edges over which the data define a connection.

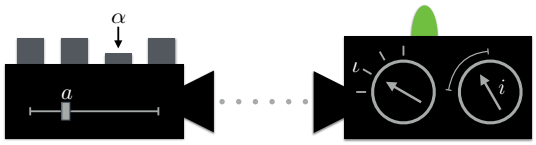


FIGURE 11. The d^2 buttons enumerate a (detector) tomographically complete frame of states. The d^2 notches enumerate a (state) tomographically complete frame of observables. The continuous slider and continuous dial are the square coordinates which displace settings.

For simplicity, each minimally complete tomography experiment will henceforth be referred to as just an experiment. For each experiment, the Born rule, $A = PW$, can be thought of as a connection between gauge parameters, e.g. $P = AW^{-1}$ or $W^{-1} \xrightarrow{A} P$. In other words, the data from experiments can be interpreted as defining maps. For multiple experiments sharing devices, there are degrees of choice as to how one can represent the gauge degrees of freedom for each pair of devices. These choices simultaneously correspond to the choices of how to embed the data in the maps between these experiments. Let us go over a few particularly meaningful examples.

A couple of gauges that should be familiar are what we would like to call standard gauges (Figure 12.) Every arrow represents a constraint which may be interpreted as a tomography — e.g. in the right diagram of Figure 12, $P \xrightarrow{A^{-1}} W^{-1}$ represents the equation $W^{-1} = A^{-1}P$ which may be interpreted as a detector tomography. This gauge is in fact the gauge used in the tomographic proof of section 5.7. Also important are what we call tomographies in "Landau" gauge (see Figure 13) which have actually appeared (sections 5.7 and 4.4.) The reader is encouraged to stare at these 4 gauges and try to see how they are each an equivalent representation of the same organization of information as Figure 10.

FIGURE 12. Tomography in "Standard" Gauge. We call them standard gauges because, considering for instance the left connection: The measurement parameters of the top-left experiment are imagined to be fixed in which case the data from this experiment can be interpreted as a standard state tomography on the top-right experiment, and from the top-right the connection does standard detector tomography on the bottom-right, etc. The choice of representing the top-left experiment's gauge by its measurement device parameters, the top-right experiment's gauge by its state device paramters, etc. uniquely defines how the data is to be organized as a connection in between these experiment's gauge parameters.

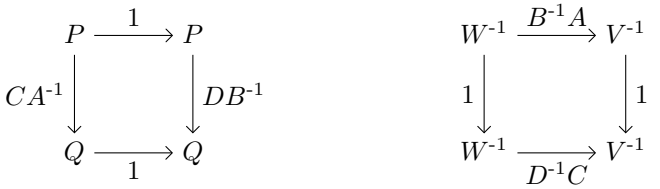


FIGURE 13. Tomography in "Landau" Gauge. Left: iso-measurement-ic tomography, the arrangement of quantum data in state gauge, Equation (4.29). Right: iso-state-ic tomography, the arrangement of quantum data in measurement gauge, Equation (4.30). These are called Landau because they keep gauge parameters in either the state or measurement direction constant just like the vector potential for a 2-d surface in the x- or y-direction can be chosen to be zero . The left gauge is a tomography where data from two experiments (either A and C or B and D) with a common measurement device is used to infer an unknown state device (Q) from a "known" state device (P.) This kind of tomography has been thought of before and already put into practice [1] (instead using a maximum likelihood method to estimate parameters rather than linear inversion, which we are considering.) As far as the authors are aware, the right gauge is a tomography yet unperformed.

All of these gauges are formally related to each other by local gauge transformations. As such, an explanation of gauge transformations on a lattice is in order (see Figure 14.) Instead of considering only a square of experiments, it is conceptually more useful to think about a *lattice* of experiments sharing devices. Something to notice is that g is not exactly the G in Equation 4.41, but rather $g\Gamma = \Gamma G$ or $g = \Gamma G \Gamma^{-1}$.

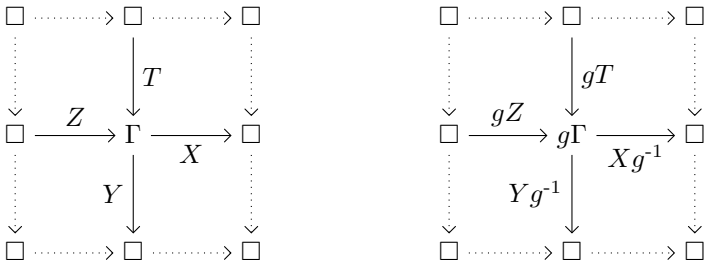


FIGURE 14. Local Gauge Transformations: The vertical direction represents displacement in state a and the horizontal direction represents displacement in measurement i. At each vertex (experiment) is a $d^2 \times d^2$ matrix of gauge parameters, Γ. At each adjacent edge (connection to the adjacent experiment) is a component of the connection, $X = \tau_i(a, i)$, $Y = \tau_a(a, i)$, $Z = \tau_a(a-1, i)$, $T = \tau_i(a, i-1)$ (see Equation 4.27.) The distance between lattice sites is defined by distances along continuous device settings (see Figure 11.) The right lattice is a local gauge transformation, g, of the left lattice at the center vertex. These transformations leave the constraints represented by each connection invariant.

Having re-expressed non-holonomic tomography for quantum systems, some distinctions are in order. First, as already mentioned one should not forget that unlike in the toy model, the gauges of quantum tomography are non-abelian — particularly, the gauge does not generally commute with the connection — which results in a covariance (see Equation 4.26) of closed-line tomographies on the gauge at the initial/terminal experiment. Second, the gauge groups, $\mathrm{GL}(d^2, \mathbb{R})$, we are concerned with are actually not compact like the unitary groups of Yang-Mills theories.[6] Third, one could imagine having d^2 continuous settings per device, in which case the gauge group becomes a tangent space, where the frame, P, and coframe, W^{-1}, are then like vierbein. Fourth, an experimentalist may not have any "sliders" but rather just have $2d^2$ "buttons" per device in which case a metric for the distance between experiments is obscured. Finally, in the "only buttons" scenario, localizing settings to corners of a square becomes arbitrary — i.e. whether settings $\{1, 2, 3, 4\}$ are to appear in the first corner or $\{2, 6, 4, 7\}$ is arbitrary.

Conclusion and Discussion

In this work, we considered non-holonomic quantum tomography as a perspective for the method of partial determinants[13]. Partial determinants are matrix quantities which analyze quantum data to detect and quantify SPAM correlations, without estimating average state-preparation or measurement parameters. We particularly focused on a toy model to illustrate that

[6].. ignoring positivity constraints.

the partial determinant is in fact a holonomy, showing that one can formalize SPAM tomographies in direct analogy to thermodynamic theories and gauge field theories. A SPAM tomography is then non-holonomic if the partial determinants (i.e. tomographic holonomies) have nontrivial values, which can be interpreted as correlations between state and measurement parameters.

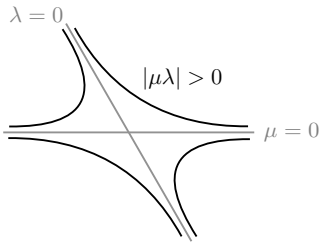


FIGURE 15. Using a determinant to define the distance of a rank 2 matrix from the space of rank 1 matrices can be a subtle point. If λ and μ are the singular values of a matrix M, then $|\mathrm{Det}M| = \lambda\mu$ is a type of distance from the axes (which are rank 1), modulo area preserving transformations. The axes are drawn askew to emphasize that there is no notion of metric distance.

From a practical perspective, the matrix elements of a PD can be used to *detect* amounts of SPAM correlation. However, the way in which a PD measures distances away from a correlated model can be a little subtle because these distances are not a metric, in the standard mathematical sense. The subtlety simply reduces to the fact that the determinant of a matrix alone does not actually tell you how large its smallest singular value is (see Figure 15.) Rather than think of distances away from the space of uncorrelated data, one must think in terms of inherited notions of distance from continuous device settings. The equations such as (4.20) can quantitatively measure correlations relative to areas in setting space.

A broader observation should also be made about device parameters and gauge dimensions. Importantly, one should notice that the only property which distinguished the toy problem from the quantum problem was a mere "speculation" about the number of degrees of freedom which parameterize the devices. In the most general scheme, an $r \times r$ PD is a test of the ability to model the data by uncorrelated r-dimensional state and measurement vectors. For quantum probabilities, one has further interpretations for the $r = d^2$ dimensions reflecting that the state and measurement vectors are also operators on a d-dimensional vector space. As an example of a more general application of PDs, one could consider 2×2 PDs for an uncorrelated qubit system. Such a PD would generally take a value different from the identity which can be interpreted as a measure of the inability to model the data by uncorrelated *classical bit* state and measurement parameters.

Acknowledgements

This research was supported in part by Perimeter Institute for Theoretical Physics. Research at Perimeter Institute is supported by the Government of Canada through the Department of Innovation, Science and Economic Development and by the Province of Ontario through the Ministry of Research, Innovation and Science. C. Jackson would like to particularly thank R. Spekkens for a useful conversation which helped to make section 4.3 more concrete. S.J. van Enk was supported in part by ARO/LPS under Contract No. W911NF-14-C-0048.

CHAPTER V

DETECTING CORRELATIONS IN MULTIQUDIT SYSTEMS

From Jackson, C. and van Enk, S. 2017. Non-holonomic tomography II: Detecting correlations in multiqudit systems. *arXiv preprint* arXiv:1702.06090

Abstract

In the context of quantum tomography, quantities called partial determinants[13] were recently introduced. PDs (partial determinants) are explicit functions of the collected data which are sensitive to the presence of state-preparation-and-measurement (SPAM) correlations. In this paper, we demonstrate further applications of the PD and its generalizations. In particular we construct methods for detecting various types of SPAM correlation in multiqudit systems — e.g. measurement-measurement correlations. The relationship between the PDs of each method and the correlations they are sensitive to is topological. We give a complete classification scheme for all such methods but focus on the explicit details of only the most scalable methods, for which the number of settings scales as $\mathcal{O}(d^4)$. This paper is the second of a two part series where the first paper[14] is about a theoretical perspective for the PD, particularly its interpretation as a holonomy.

Introduction

In any quantum tomography experiment, one has the ability to perform various state preparations and measurements. We may abstractly represent these abilities by devices with various settings (Figure 16.) In standard quantum tomographies such as state, detector, or process tomography, it is assumed, respectively, that either the measurement device, the state device, or both are already characterized and may thus provide a resource to determine the parameters associated with the yet uncharacterized devices. Fundamental to the practice of these tomographies is a much subtler assumption: that the performance of each device is independent of the use and history of every other device.

A problem in recent years has been the issue of estimating quantum gates while taking into account that there are small but significant errors in the states prepared and measurements made to probe the gates, so called SPAM errors [10]. Any practice which takes into account SPAM errors will be generically referred to as SPAM tomography. Several works have come out in SPAM tomography

particular to the task of making estimates in spite of such conditions [10–12], all of which speak to the notion of a "self-consistent tomography." However, these works consistently assume by fiat that the SPAM errors are uncorrelated.

In [13] it was demonstrated that one can test for the presence of correlated SPAM errors using so called partial determinants (PDs) which bypass any need to estimate state or measurement parameters individually. The logic behind the PD is simple, uncorrelated SPAM corresponds to a particular ability to factorize the estimated frequencies into a product of state and measurement parameters. Such a factorization always exists for small enough numbers of settings but does not exist for larger numbers of settings if there are correlations. Thus, the notion of parameter independence can be viewed as either a local or a global property. PDs are then a measure of the contradiction that results from requiring that multiple sets of locally uncorrelated settings be consistent with each other. In other words, SPAM correlations correspond to holonomies (or measures of global contradiction) in overcomplete tomography experiments (hence the title, "non-holonomic tomography.") Further details on this perspective may be found in [14].

For multiqudit systems, the notions of product state and product measurement introduce further kinds of factorizability in estimated frequencies. Particularly, we will focus on systems where there is a single device associated with the preparation of multiqudit states and a measurement device for each qudit separately (but not necessarily independently) — i.e. systems where we expect outcome probabilities to factor into the form $\mathrm{Tr}\rho(E\otimes\cdots\otimes E)$. Sure enough, PDs can be generalized to measure a degree to which such factorizations do not exist. Thus, these generalized PDs can serve as tests for the presence of various state-state correlations, measurement-measurement correlations, as well as mixed SPAM correlations. Further, such generalized PDs can be much more scalable than the original PD — $\mathcal{O}(d^4)$ settings versus $\mathcal{O}(d^{4m})$ settings where m is the number of qudits. The main portion of this work will demonstrate how to classify the various PDs one could consider.

FIGURE 16. On the left is a device which prepares various signals on demand depending on which button, $a \in \{1,\ldots,N\}$, is pressed. On the right is a device which blinks to indicate a signal with a certain property depending on which setting, $i \in \{1,\ldots,M\}$, a dial is turned to. For each pair of settings (a,i), if ${n_a}^i$ is the number of times the light blinks and ${N_a}^i$ is the number of times the button is pressed, then the estimated frequency is just ${f_a}^i = {n_a}^i/{N_a}^i$.

The most basic aspect of non-holonomic tomography relies on the notion of an effectively uncorrelated system. With this notion, one emphasizes the perspective that, although one is not able to measure individual device parameters, correlation is simply the inability to define parameters that are organized according to a particular model. Similar forms of analysis have come up in the context of matrix product states[51–53], a way of representing various kinds of many-body quantum states that is particularly elegant for calculating correlation functions. Similar analyses can also be found in the more abstract context of (generalized) Baysian networks[54–56] where the presence of hidden variables with in a model or causal structure result in a rich set of testable constraints on the probabilities associated with the observed variables. (Bell inequalities are an example of this.) Also from a fundamental perspective, similar analyses may be found in works on general probabilistic theories [34, 57] where much attention is spent on the correspondence between operational descriptions of systems and the mathematical calculations that represent them.

Tomography: States, Observables, and Data

The Born Rule Revisited

For every quantum experiment, quantum events are counted and the frequency of each outcome is understood to estimate the product of a state and a POVM element (Figure 16.) This is the famous Born Rule, usually denoted

$$f_a{}^i = \mathrm{Tr}\rho_a E^i \tag{5.1}$$

where ρ_a is the density operator for the state prepared according to a, E^i is the POVM element for an outcome of the measurement made according to i, and $f_a{}^i$ is the estimated frequency. However, we wish to consider the situation where the state-preparations and measurements (SPAM) behind these estimated frequencies actually fluctuate. In such a case, one must modify the Born Rule to read

$$f_a{}^i = \langle \mathrm{Tr}\rho E \rangle_a{}^i \tag{5.2}$$

where $\langle\rangle_a{}^i$ denotes the average over the ensemble of trial runs of the devices set to a and i — that is, ρ and E are now to be considered (positive operator-valued) random variables, distributed according to the setting (a, i).

It is useful to more generally consider estimates of any statistical observable, ${S_a}^i$ such that

$$ {S_a}^i = \langle \mathrm{Tr}\rho\Sigma \rangle_a{}^i, \tag{5.3} $$

where Σ is a Hermitian (not necessarily positive) operator-valued random variable representing the corresponding quantum observable. The setting i still represents a measurement, but can be more generally associated with a specific linear combination of outcomes which may be useful to consider — e.g. $\Sigma^i = |+_i\rangle\langle+_i| - |-_i\rangle\langle-_i|$ where $|\pm_i\rangle$ are eigenstates of spin in the i-direction. Any such ${S_a}^i$ will be referred to as quantum *data*, calculated as the same linear combinations of measured frequencies as the observables they correspond to — that is, ${S_a}^i = {f_a}^k {c_k}^i$ just as $\Sigma^i = E^k {c_k}^i$ for whatever ${c_k}^i$ are useful. More traditional language would refer to ${S_a}^i$ as a "quantum expectation value" of the observable i given state a. However, for the purposes of this paper one should refrain from such language as it is crucial to focus instead on states and observables themselves as the random variables, rather than the actual result or "blinking of the light" for each trial.[1]

The Partial Determinant: A Test for Correlated SPAM errors

In standard state, detector, and process tomographies, an experimentalist can ignore the ensemble average because they are (respectively) able to control either the measurements, the state preparations, or both. However, if one is doing SPAM tomography, where neither the state preparations nor measurements are assumed to be controlled, then the ensemble average suggests the possibility that SPAM errors are correlated — i.e.

$$ \langle \mathrm{Tr}\rho\Sigma \rangle_a{}^i \neq \mathrm{Tr}\langle\rho\rangle_a \langle\Sigma\rangle^i. \tag{5.4} $$

From the perspective of doing any of the standard tomographies, this is an awkward statement indeed because one does not have the resources necessary to access quantities such as $\langle\rho\rangle_a$ or $\langle\Sigma\rangle^i$ individually.[2] One may thus be tempted to conclude that correlations (or lack thereof) cannot be determined without access to the individual expectation values. However, this is not the case.

[1] The author is even inclined to suggest that states and outcomes should fundamentally be thought of as on an equal footing.

[2] One could parse correlations into two separate kinds of independence: The first kind being when $\langle \mathrm{Tr}\rho\Sigma \rangle_a{}^i = \mathrm{Tr}\langle\rho\rangle_a{}^i \langle\Sigma\rangle_a{}^i$. The second kind being when $\langle\rho\rangle_a{}^i = \langle\rho\rangle_a$ and $\langle\Sigma\rangle_a{}^i = \langle\Sigma\rangle^i$

Correlations such as Equation (5.4) can be determined without individual expectation values because equations like $\langle \mathrm{Tr}\rho\Sigma\rangle_a{}^i = \mathrm{Tr}\langle\rho\rangle_a\langle\Sigma\rangle^i$ express a very special factorizability of the data, Equation (5.3). We thus proceed with the following operational definition: we say that data $S_a{}^i$ is *effectively (SPAM) uncorrelated* when we can express it as a simple matrix equation:

$$S_a{}^i = P_a{}^\mu W_\mu{}^i. \tag{5.5}$$

The rows of P and columns of W (when they exist) represent the states and observables, $\rho_a = P_a{}^\mu \sigma_\mu$ and $\Sigma^i = \sigma^\mu W_\mu{}^i$, where $\{\sigma_\mu\}$ is some operator basis and $\{\sigma^\mu\}$ is the corresponding dual basis. Repeated indices are to be summed over. We relax the requirement that the rows of P correspond to positive operators.

Three important observations should be made about this definition. First, Equation (5.5) requires that the sum on μ be over $\leq n^2$ operators for an n-dimensional Hilbert space. (Indeed, the notion of effectively uncorrelated is always relative to an assumed dimension, n.) Second, one can always write an expression like Equation (5.5) (with the sum on μ being over $\leq n^2$) so long as the number of state settings, N, and detector settings, M, are both $\leq n^2$. Third, when P and W exist, they are in general not unique because one could just as well use PG and $G^{-1}W$ where G is an $n^2 \times n^2$ real invertible matrix. The components of G are gauge degrees of freedom which have been referred to as *SPAM gauge*[10–12] or *blame gauge*[13].

The combination of these observations suggest the following generic protocol for quantifying correlations: First, perform SPAM tomography with $N > d^2$ and $M > d^2$. Such SPAM tomography may be referred to as overcomplete because such numbers of setttings would correspond to overcomplete standard tomographies, if only the appropriate devices were controlled and well characterized. Second, consider $d^2 \times d^2$ submatrices of the data, which can be thought of as corresponding to minimally complete tomographies. Each such submatrix is effectively uncorrelated and thus may be associated with a "local" gauge degree of freedom. Finally, check whether the states and observables for each minimally complete submatrix can be chosen so that such choices among all submatrices are consistent with each other. It turns out that the amount of inconsistency can be quantified rather elegantly by what has been called a partial determinant.[13]

Such protocols can be understood as an organization of the data into a *fiber bundle*. Fiber bundles are mathematical structures which support the notions of *connection* and *holonomy* which

are rather ubiquitous concepts. In [14], it is demonstrated how tomography and partial determinants can be interpreted as connections and holonomies, respectively. However, for the sake of those who are interested exclusively in potential applications, an effort has been made to avoid such language in this current paper. Nevertheless, occasional references will be made and terms will be defined which allude to this perspective. It is from these perspectives that the title "Non-holonomic tomography" is fully justified as a name for partial determinants.

Forgetting this alternative perspective and using standard linear algebraic considerations, instead one can make the observation that the definition for effective independence is equivalent to the statement that the data must be such that $\text{rank}(S) \leq n^2$. For example, consider devices whose numbers of settings are $M = N = 2n^2$ and organize the data as

$$S = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \tag{5.6}$$

where each corner is an $n^2 \times n^2$ matrix. One can define a partial determinant (or PD) for this arrangement of the data,

$$\Delta(S) = A^{-1}BD^{-1}C. \tag{5.7}$$

The PD has the property that it is equal to the identity matrix if and only if the data is effectively SPAM uncorrelated.[13] The proof of this is simple if one observes that $\text{rank}(S) \leq n^2$ if and only if there exist $n^2 \times n^2$ matrices P_1, P_2, W_1, and W_2 such that

$$S = \begin{bmatrix} P_1 \\ P_2 \end{bmatrix} \begin{bmatrix} W_1 & W_2 \end{bmatrix}. \tag{5.8}$$

Multiqudit Correlations: SPAMs and Non-Localities

In this paper we consider extensions of the general notion of a PD to multiqudit systems ($n = d^m$ for m qu*d*its) with a concentration on PD constructions with a number of device settings which scales to lowest order, $\mathcal{O}(d^4)$. Specifically, we will focus on systems where the preparation of a multiqudit state can be represented by a single device and the measurement of each qudit can be represented separately by separate (but not necessarily independent) devices.

Uncorrelated measurements between different qudits will be referred to as *local* measurements.[3] If all measurements are effectively local and SPAM uncorrelated, then any data collected for m qudits can be factored into the form

$$S_a^{ijk\cdots} = \mathrm{Tr}\big(\rho_a\ \Sigma_1^i \otimes \Sigma_2^j \otimes \Sigma_3^k \cdots\big) = R_a^{\lambda\mu\nu\cdots} W_{1\lambda}{}^i W_{2\mu}{}^j W_{3\nu}{}^k \cdots \tag{5.9}$$

where $\rho_a = R_a^{\mu\nu\cdots}\sigma_\mu \otimes \sigma_\nu \otimes \cdots$ and $\Sigma_q^i = W_{q\mu}^i \sigma^\mu$ for each qudit $q \in \{1, \ldots, m\}$, for some operator bases $\{\sigma_\mu\}$ and dual bases $\{\sigma^\mu\}$ and a sum over repeated greek indices (from 1 to d^2) is always implied.[4]

Two kinds of indices can be distinguished in these expressions. There are those indices which correspond to device *settings* (a, i, j, ...) which are associated with degrees of freedom which can be controlled. Such indices may be referred to as "external" because they correspond to degrees of freedom outside of the quantum system being probed. Then there are those indices which correspond to device *parameters* (μ, ν, ...) and represent the model — the Born rule with d-dimensional Hilbert spaces. These indices may be referred to as "internal" because they are always summed over and thus are accompanied with gauge degrees of freedom.

For such multiqudit systems, there are now multiple kinds of correlation one can have. We refer to correlations between states and measurements on qudit q as SPAM_q correlations. Further, let us refer to correlations between measurements on qudit q and measurements on qudit p as (q,p)-nonlocalities. Such correlations are to be understood with the notions of *effectively* SPAM_q uncorrelated data and *effectively* (q,p)-local data. Then one may proceed to categorize the various ways a partial determinant may be constructed to test if a system is SPAM_q correlated or (q,p)-nonlocal. For simplicity, we are only considering 2-point correlations in this paper (see Conclusions and Discussion.)

Local Measurements of Two Qudits

For two qudits, a.k.a. a qud^{2}it, the data is an object with 3 (external) indices, 1 for state preparations and 2 for the measurements on each qudit. If there are no correlations, then we may

[3]This second meaning of the word "local" should not be too confusing as it will be clear from context whether we are considering individual qudit observables or small numbers of state and measurement settings.

[4]Technically, we should write $(\sigma_q)_\mu$ to emphasize that the qudit measurements do not necessarily share a reference frame, but we will not write this here for the sake of reducing index clutter.

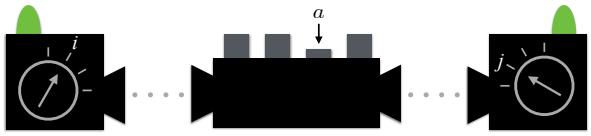


FIGURE 17. A two qudit experiment where there is a single device which prepares qud^{2}it states and two devices which make qudit measurements. We would like to know if the data can be modeled by equation (5.10) — that is, does the factorization in Equation (5.10) exist for the accessible S_a^{ij} of this experiment?

write

$$S_a^{ij} = R_a^{\mu\nu} V_\mu{}^i W_\nu{}^j. \tag{5.10}$$

One can consider this as a matrix equation in the most obvious way:

$$S_a{}^I = R_a{}^M X_M{}^I \tag{5.11}$$

where $M = (\mu, \nu)$, $I = (i, j)$, and $X_M{}^I = V_\mu{}^i W_\nu{}^j$, treating the two qudit measurements as one qud^{2}it measurement. This separation of parameters suggest the original protocol [13] for detecting what will now be called *generic* SPAM correlations, constructing a partial determinant for $n = d^2$.

One can also consider equation (5.10) as a matrix equation in another way:

$$S_A{}^j = P_A{}^\nu W_\nu{}^j \tag{5.12}$$

where $A = (a, i)$ and $P_A{}^\nu = R_a^{\mu\nu} V_\mu{}^i$. One can interpret this separation as the measurement settings of one qudit being used to effectively prepare states for the other qudit. In this case, we know that if these effective-states, set by A, are uncorrelated with the other qudit measurements, set by j, (so that we may write Equation (5.12)) then a smaller ($n = d$) PD of $S_A{}^j$ must be the identity. We should stress that taking the inverse of a matrix like $S_A{}^j$ is very different from taking the inverse of a matrix like $S_a{}^I$ even if they consist of the same entries, only organized differently.

One sees that there are already two distinct ways to be effectively uncorrelated: The first is to be SPAM uncorrelated in the generic sense, such that Equation (5.11) exists. In this case, the rank of $S_a{}^I$ must be $\leq d^4$, particularly for $> d^4$ state settings, a, and $> d^4$ measurement settings, I. The second is to be uncorrelated such that Equation (5.12) exists. In this case, the rank of $S_A{}^j$ must

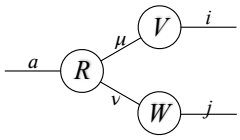


FIGURE 18. Diagrammatic representation of effectively completely uncorrelated data, Equation (5.10). Each internal line represents a sum over d^2 operators while each external line represents a setting.

be $\leq d^2$, particularly for $> d^2$ effective state settings, A, and $> d^2$ measurement settings, j, for the second qudit.

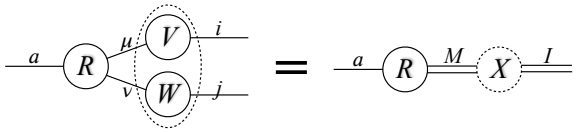


FIGURE 19. Diagrammatic representation of Equation (5.11), a weaker form of effective independence. The right-hand matrix has a rank bounded by d^4 because the dotted separation cuts two internal lines. Double lines represent product indices.

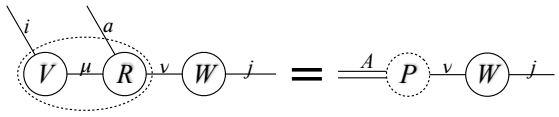


FIGURE 20. Diagrammatic representation of Equation (5.12), another kind of weaker effective independence. The rank of the right-hand matrix is bounded even lower by d^2 because the dotted separation cuts only one internal line.

Equations (5.11) and (5.12) represent weaker forms of effective independence than Equation (5.10). One should think of them as potential factorizations of the data which may or may not exist. It is helpful to represent Equations (5.10) through (5.12) diagrammatically, as in Figures 18 thru 20. Being able to factorize the data as in Equation (5.10) means that the system can be considered completely uncorrelated. Being able to factorize the data as in Equation (5.11) means that the data is effectively SPAM_1 and SPAM_2 uncorrelated. (Recall the terminology from the end of section 5.3) Being able to factorize the data as in Equation (5.12) means that the data has effective SPAM_2 independence and (1,2)-locality. Similarly, there is another factorization that results from permuting the qudits whose existence would mean the system is effectively SPAM_1 uncorrelated and (1,2)-local.

Classifying Different PDs

PDs which test for generic SPAM correlations are relatively straight forward as there are only 2 main variations on their construction. In contrast, there are 11 distinct PDs one can consider relative to factoring the data as in Equation (5.12). These PDs differ in their construction by the number of settings used for a and j (or i) and by how these settings are organized. These different constructions are sensitive to different kinds of correlation. Specifically, each PD will be equal to the identity when the system is effectively uncorrelated in a corresponding way. In order to organize the description of these various constructions, we must establish a few definitions and some notation.

The procedure for constructing a PD can be summarized in two steps. The basic goal is to organize the data so that it is of the same form as Equation (5.6). In addition to the original PD construction, rows and columns may now be products of multiple settings. The first step is then to organize the settings so to construct a *corner* template. The second step is to "displace" four instances of that corner which can then be connected in a *loop*, as in Equation (5.7). This constructed matrix of four corners shall be called a *square*.

Generic SPAM Correlations

For detecting generic SPAM correlations, such that Equation (5.11) does not exist, we denote the various numbers of experimental settings by $[N : M_1, M_2]$, where N is the number of state settings (the range of a) and M_q is the number of local measurement settings for qudit$_q$ (the range of i or j.) The colon can be thought of as representing the dotted separation of Figure 19. Settings to the left of the colon are to be organized as a row index while settings to the right are to be columns.

To calculate a partial determinant in this case, one needs to consider corners that are $d^4 \times d^4$ which further requires $N = d^4$ and $M_1 = M_2 = d^2$. (Recall that this is because the data must have rank $\leq d^4$ if Equation (5.11) exists.) A square can then be assembled from such a corner in two ways, which we denote simply by multiplying the appropriate setting number by 2:

$$[2d^4 : 2d^2, d^2] \qquad \text{and} \qquad [2d^4 : d^2, 2d^2]. \tag{5.13}$$

Thus we have two kinds of generic PD.

Importantly, we use a '2' in our bracket notation as if to suggest implementing twice as many settings, as done originally. However, one could just as well make a square from any number of rows

and columns, each $> d^4$. In Appendix 5.7, we demonstrate how to construct an $r \times r$ PD for an $(r+1)\times(r+1)$ matrix. Nevertheless, we will always write '2's in our bracket notation for simplicity. To summarize, factors of d in this square bracket notation represent an organization template for the settings in each corner, while '2's represent which device settings one changes when going from one corner to the next. The nature of these representations should become much clearer in the following, more intricate factorization problem.

Nonlocalities and SPAM$_q$ Correlations

For detecting correlations such that Equation (5.12) does not exist, we denote the numbers of settings by $[N; L : M]$. We now make a distinction between L, the number of observable settings for qudit$_1$, used to effectively prepare states, and M, the number of observable settings for qudit$_2$, used to measure them. Again, we can interpret the colon as the dotted separation in Figure 20, between effective state preparations and measurements of qudit$_2$. A semicolon after the first argument is just to distinguish the first argument as the number of (joint) state preparations. Of course, there are actually 2 distinct schemes of type $[N; L : M]$ depending on which qudit we consider part of the effective state preparation. We denote the other by $\pi[N; L : M]$ where π means 'permute the two qudits.'

Corners	Squares	
$[N; L : M]$	$[2N; L : 2M]$	$[N; 2L : 2M]$
$[d^2; 1 : d^2]$	$[2d^2; 1 : 2d^2]$	$[d^2; 2 : 2d^2]$
$[d; d : d^2]$	$[2d; d : 2d^2]$	$[d; 2d : 2d^2]$
$[1; d^2 : d^2]$	$[2; d^2 : 2d^2]$	$[1; 2d^2 : 2d^2]$

(5.14)

TABLE 2. Each row is a way to make a corner while each column is a way to make a square.

Corners and squares can now be made in several ways. Corners must be $d^2 \times d^2$ (because the data must have rank $\leq d^2$ if Equation (5.12) exists.) There are 3 ways one can do this because we must take $M = d^2$ while there are 3 different ways to make d^2 effective states, $[N; L] = [d^2; 1]$, $[d; d]$, and $[1; d^2]$, (restricting ourselves to nice multiples.) Then there are 2 ways each to make a square, $[2N; L : 2M]$ or $[N; 2L : 2M]$, (ignoring that one could mix corner types in a single PD.) (See Table 2.) Having picked one of the 2 qudits, there are almost $2 \times 6 = 12$ PDs, except that $\pi[1; 2d^2 : 2d^2] = [1; 2d^2 : 2d^2]$ is actually a symmetric construction. So there are $12 - 1 = 11$ in total PDs of the type $[N; L : M]$. To make the construction of these PDs as clear as possible, Figures 22 and 21 are given to go over each of them individually.

In Figures 21, it is important to recall the distinction between device settings and a device parameters. Device settings are the external controls that are available in an experiment, while device parameters are the model dependent numbers used to describe the behavior of the experiment. Changes in settings can be understood as generating changes in parameters, but only in a local sense (from corner to corner) which one might not be able to integrate to a global correspondence (because there could be correlations.) In other words, a constraint such as "keeping state parameters fixed" can still be operationally defined but will in general be a non-holonomic constraint. Similar distinctions are represented mathematically in other physical theories as well, a discussion of which may be found in [14].

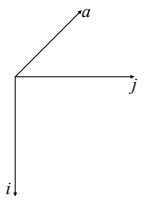
In Figures 22, Corners have been given qualitative names for how they "fill" the space of settings as represented in Figures 21. Solid lines represent a range of d^2, dashed lines have range d, and amputated lines are single valued. A vertex joining one solid line with two dashed lines represents the delta function

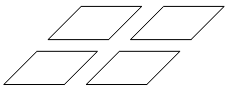
$$\delta_A^{ab} = \begin{cases} 1 & A = ad + b \\ 0 & \text{otherwise} \end{cases} \tag{5.15}$$

where $A \in \{0, 1, \ldots, d^2 - 1\}$ is the solid line and $a, b \in \{0, 1, \ldots, d - 1\}$ are the dashed lines. Dotted lines with small circular endpoints represent the settings used to displace the corners of a square, i.e. the '2's in square bracket notation. Squares have been further labelled based on how they are oriented in the setting dimensions as represented by the placement of '2's in bracket notation as well as in Figures 21.

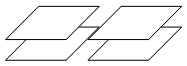
Further in Figures 22, the backbone (Figure 18) of each diagram represents the hypothesis that the data is effectively completely uncorrelated. However, once a corner is assembled, one can then see from the diagram how this hypothesis may be relaxed to weaker types of independence that would still give the PD a trivial value. Digrammatically, this corresponds to the property that the minimum number of lines one must cut in order to detach the external solid lines corresponds exactly to the upper bound in the rank. Moreover, the displacing lines or '2's can empirically suggest different models of correlation for nontrivial values in the corresponding PD. For example, a nontrivial value for $[2d^2; 1 : 2d^2]$ suggests SPAM_2 correlations while $[d^2; 2 : 2d^2]$ suggests (1,2)-nonlocalities.

To summarize, a square bracket notation has been introduced to represent different PDs one can construct for two-qudit (or qud^2it) systems. Each PD will have a trivial value, $A^{-1}BD^{-1}C = 1_{d^2 \times d^2}$,

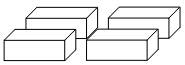




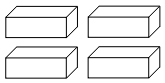
(a) $[\mathbf{2d^2}; \mathbf{1} : \mathbf{2d^2}]$ Using $2d^2$ states, 1 observable for qudit_1, and the usual $2d^2$ observables for qudit_2. If the qudit_1 observable is the identity, this is simply SPAM_2 tomography.



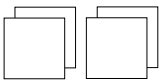
(b) $[\mathbf{d^2}; \mathbf{2} : \mathbf{2d^2}]$ Using d^2 states, 2 observables for qudit_1, and the usual $2d^2$ observables for qudit_2. Another PD can be constructed by permuting the two qudits, $\pi[d^2; 2 : 2d^2]$.



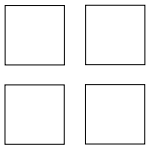
(c) $[\mathbf{2d}; \mathbf{d} : \mathbf{2d^2}]$ Using $2d$ states and d qudit_1 observables (such as a d-outcome POVM.) Each stick of butter represents a square matrix that has been rolled up or folded.



(d) $[\mathbf{d}; \mathbf{2d} : \mathbf{2d^2}]$ Using d states and $2d$ qudit_1 observables (such as two d-outcome POVMs.) Another PD can be constructed by using vertical sticks of butter, $\pi[d; 2d : 2d^2]$.



(e) $[\mathbf{2}; \mathbf{d^2} : \mathbf{2d^2}]$ Using two states and d^2 qudit_1 observables. Another protocol exists by permuting measurement locations.



(f) $[\mathbf{1}; \mathbf{2d^2} : \mathbf{2d^2}]$ Using one state and $2d^2$ and qudit_1 observables . This particular protocol is symmetric under permuting qudits, $\pi[1; 2d^2 : 2d^2] = [1; 2d^2 : 2d^2]$.

FIGURE 21. At the top is a coordinate system for the entries of the data ${S_a}^{ij}$ where a is a state setting, i is a qudit_1 measurement setting, and j is a qudit_2 measurement setting. Regions covered by the various shapes represent collected data. Each subfigure can be associated with a different measurement protocol one can consider which may further suggest different models of correlation. Each shape corresponds to a corner template while the arrangement of the 4 copies correspond to a square one can translate through. The permuted versions of each PD corresponding to the above subfigures are all distinct, except for the PD of subfigure 21f.

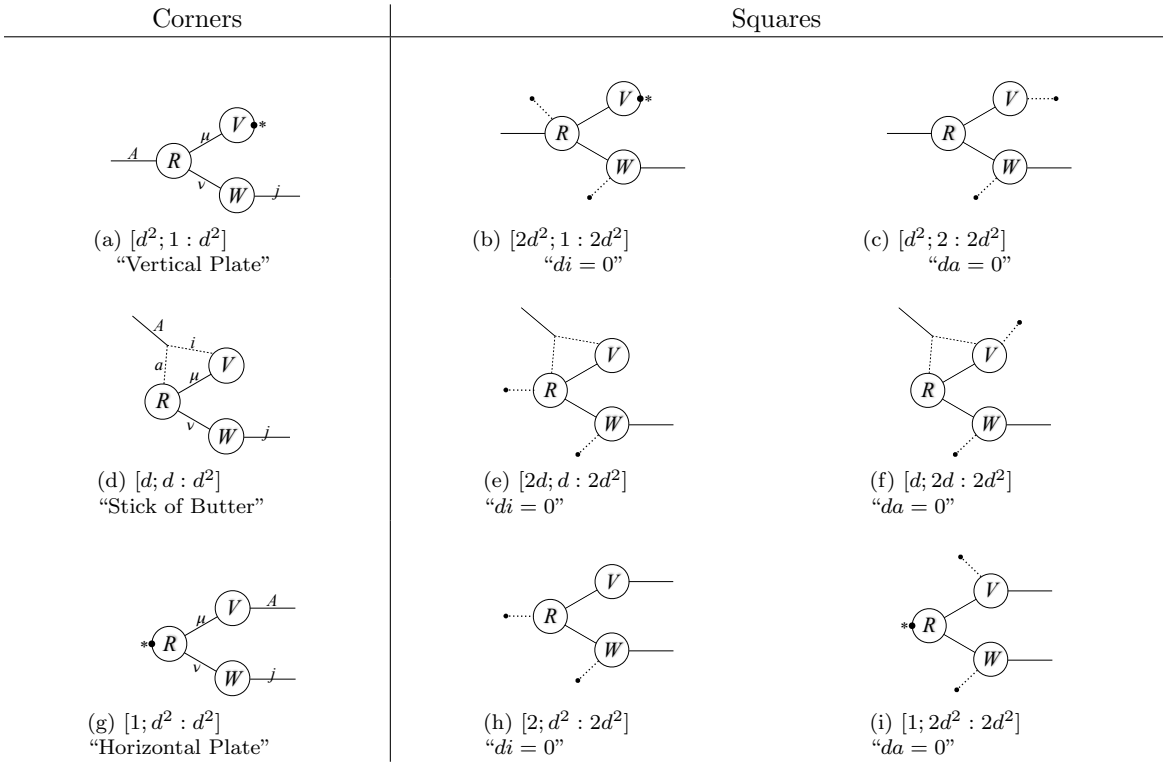


FIGURE 22. Diagrammatic representations of PD constructions as arranged in Table 2. Circles represent device parameters. External lines represent experimental settings. Internal lines represent a sum over the number of independent model parameters. The backbone of each diagram, Figure 18, represents the hypothesis that the data (of an experiment like Figure 17) can be modeled by Equation (5.10). Solid lines represent a range of d^2, dashed lines have range d, and amputated lines are single valued. A vertex joining one solid line with two dashed lines represents the delta function, Equation (5.15). Dotted lines with small circular endpoints correspond to a '2', a setting used to displace or distinguish corners.

if the system is effectively uncorrelated in that corresponding way. The types of correlation which violate these PDs should be clear from the topology of their effective backbone (see Figures 19 and 20.) The first is that $[N : M_1, M_2]$ PDs are trivial if $\langle \mathrm{Tr}\rho\,(\Sigma \otimes \Sigma)\rangle = \mathrm{Tr}\langle\rho\,\rangle\langle\Sigma \otimes \Sigma\rangle$ and are thus not sensitive to (12-)nonlocalities. The second is that $[N; L : M]$ PDs are trivial if $\langle \mathrm{Tr}\rho\,(\Sigma\otimes\Sigma)\rangle = \mathrm{Tr}\langle\rho\,\Sigma\rangle \otimes \langle\Sigma\rangle$ so are not sensitive to SPAM_1 correlations. Similarly $\pi[N; L : M]$ are insensitive to SPAM_2 correlations.

More than Two Qudits

Increasing the number of qudits, $m > 2$, there are many more variations in the kinds of corners and squares we can construct and so there are many more different types of experiments one can

do to detect many more different types of correlation. One fruitful way of classifying PDs (and the corresponding experiments) is by the matrix *rank* that the corresponding square should have in the absence of correlations. In particular, for qudit measurements on qud^{m}it there are m types of PDs corresponding to m different ranks, $\text{rank}(M) = d^{2k}$ for $k = 1, \ldots, m$. (See Figures 19, 20, and 23 and Tables 3 and 4.) Remember that the rank also determines how the number of experimental settings scales, namely, as "pairs of settings" $= \text{rank}(M)^2 = d^{4k}$.

Following our previous notation, these classes will be denoted with square brackets by

$$\Delta_k = [N; L_1, \ldots, L_{m-k} : M_1, \ldots, M_k] \tag{5.16}$$

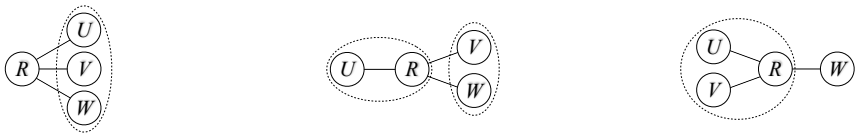
The generic PD corresponds to $k = m$ which has only 1 corner type (because all the measurement devices are to the right of the colon) and m square types (because there are m devices to the right of the colon which can be used for displacement) as in section 5.4. Those PDs which demand the least number of experimental settings correspond to $k = 1$ for which there are 4 kinds of corner, 10 kinds of square, and $\frac{1}{2}m(7m^2 - 12m + 7)$ permutational variants, as will be explained. All of the main variations in $k = 1$ are present for $m = 3$, so we will start there. We will also briefly include $k = 2$ for $m = 3$ qudits to make the construction of the more general PDs clear.

Three Qudits

For three qudits, the data has $1 + 3 = 4$ indices or device settings. If the data is completely uncorrelated, then we may write

$$S_a{}^{ijk} = R_a^{\lambda\mu\nu} U_\lambda{}^i V_\mu{}^j W_\nu{}^k. \tag{5.17}$$

Such data can be organized into a matrix in 3 basic ways as represented in Figure 23. These 3 ways further represent separate classes of PD one can construct, each of which are sensitive to different correlations. Generic PDs, $[N : M, M, M]$, are insensitive to all (p, q)-nonlocalities. The "$k = 2$" PDs, $[N; L : M, M]$, are insensitive to $(2, 3)$-nonlocality and SPAM_1 correlation, but are sensitive to $(1, 2)$-nonlocality, $(1, 3)$-nonlocality, SPAM_2 correlation, and SPAM_3 correlation. The most scalable PDs, $[N; L, L : M]$, are insensitive to $(1, 2)$-nonlocalities, SPAM_1, and SPAM_2 correlations. but are sensitive to $(1, 3)$-nonlocality, $(2, 3)$-nonlocality, and SPAM_3 correlation.



(a) $[N : M_1, M_2, M_3]$
Uncorrelated $\leftrightarrow$ rank $\leq d^6$

(b) $[N; L : M_1, M_2]$
Uncorrelated $\leftrightarrow$ rank $\leq d^4$

(c) $[N; L_1, L_2 : M]$
Uncorrelated $\leftrightarrow$ rank $\leq d^2$

FIGURE 23. Each circular vertex also has an implied external index attached to it like in Figures 19 & 20. Each line cut by the dotted separation represents a sum over d^2 degrees of freedom. These factors determine the upper bound on the rank of the data respectively organized.

Of course, one can permute the qudits to make similar statements. Perhaps the best way to denote each of these is by $\pi[N; \ldots M]$ where now π could denote any permutation of 3 elements. Further, we may denote each π the most succinctly with cyclic notation. For example $(123)[N; L : M, M]$ PDs are insensitive to 31-nonlocalities and SPAM_2 correlation. This notion is important for discussing PD symmetries which brings us to the discussion on the ways one can construct corners and squares.

Generic PDs, $[N : M, M, M]$, have no variability in corner types and only 1 basic kind of square, 3 considering which qubit you choose to displace the measurement dimension. These can be represented in permutation notation as Δ, $(12)\Delta$, and $(13)\Delta$ where $\Delta = [2d^6 : 2d^2, d^2, d^2]$. The permutations $\{1, 12, 13\}$ represent the coset for the subgroup $\{1, 23\}$ corresponding to the symmetry $(23)\Delta = \Delta$.

Corners	Squares	
$[N; L : M_1, M_2]$	$[2N; L : 2M_1, M_2]$	$[N; 2L : 2M_1, M_2]$
$[d^4; 1 : d^2, d^2]$	$[2d^4; 1 : 2d^2, d^2]$	$[d^4; 2 : 2d^2, d^2]$
$[d^3; d : d^2, d^2]$	$[2d^3; d : 2d^2, d^2]$	$[d^3; 2d : 2d^2, d^2]$
$[d^2; d^2 : d^2, d^2]$	$[2d^2; d^2 : 2d^2, d^2]$	$[d^2; 2d^2 : 2d^2, d^2]$

(5.18)

TABLE 3. $[N; L : M, M]$ PDs require $\mathcal{O}(d^8)$ settings. The number of settings is determined by the expected rank of these matrices, d^4, for a completely uncorrelated model. See Figure 23.

For $k = 2$ PDs, $[N; L : M, M]$, we have 3 kinds of corner and 6 kinds of square (see Table 3 and compare to Table 2.) We even continue to have the symmetry $(12)\Delta = \Delta$ for $\Delta = [d^2; 2d^2 : 2d^2, d^2]$. Except now, a square can be displaced in 3 measurement dimensions. This gives a total of $3 \times 11 = 33$ partial determinants, 6 per square except for the one with a symmetry (which only only gives $6/2 = 3$.) For m qubits, this would be $11\binom{m}{2}$ PDs.

Qudits given a "1" in square bracket notation can be considered a trace over that qudit, i.e. choose the identity observable. Most practical instances will consider qubits, $d = 2$, in which case

it is important to remember a "d" and a "2" are still different in that a d refers to settings used to make a single kind of corner while a 2 is used to displace different corners in a square. Diagrams could be drawn as before to represent corners and squares where we would go on to interpret what trivial values for such PDs mean.

Corners	Squares		
$[N; L_1, L_2 : M]$	$[2N; L_1, L_2 : 2M]$	$[N; 2L_1, L_2 : 2M]$	$[N; L_1, 2L_2 : 2M]$
$[d^2; 1, 1 : d^2]$	$[2d^2; 1, 1 : 2d^2]$	$[d^2; 2, 1 : 2d^2]$	
$[d; d, 1 : d^2]$	$[2d; d, 1 : 2d^2]$	$[d; 2d, 1 : 2d^2]$	$[d; d, 2 : 2d^2]$
$[1; d^2, 1 : d^2]$	$[2; d^2, 1 : 2d^2]$	$[1; 2d^2, 1 : 2d^2]^*$	$[1; d^2, 2 : 2d^2]$
$[1; d, d : d^2]$	$[2; d, d : 2d^2]^{**}$	$[1; 2d, d : 2d^2]$	

(5.19)

TABLE 4. $[N; L, L : M]$ PDs require $\mathcal{O}(d^4)$ settings. The number of settings is determined by the expected rank of these matrices, d^2, for a completely uncorrelated model. See Figure 23.

Finally for the most scalable PDs, $[N; L, L : M]$, we have 4 types of corner and 10 types of square (see Table 4.) Entries kept blank are simply because they are equivalent by permutation with the entry to the left in the table. Entries marked with an asterisk have a symmetry. All together, these make 51 PDs which we will explain the combinatorics for in the next section on general m. Illustrating the corners diagrammatically as in Figures 24, we can interpret the meaning of a non-trivial value for each corresponding PD and we will refer to them by there subfigures:

- Rank d^2 PDs displaced from Δ_a, Δ_b, and Δ_d have a trivial value
  if and only if $\langle RUVW\rangle = \langle RUV\rangle\langle W\rangle$.

- On the other hand, rank d^2 PDs displaced from Δ_c will have a trivial value
  if either $\langle RUVW\rangle = \langle RUV\rangle\langle W\rangle$ *or* $\langle RUVW\rangle = \langle U\rangle\langle RVW\rangle$.

- Further, if $\langle RUVW\rangle \neq \langle RUV\rangle\langle W\rangle$ but $\langle RUVW\rangle = \langle U\rangle\langle RVW\rangle$,
  then a rank d^2 PD displaced from Δ_b will have a nontrivial value,
  but one of rank d^3 will be trivial.

- Finally, if $\langle RUVW\rangle \neq \langle RUV\rangle\langle W\rangle$ but a PD from Δ_d but of rank d^3 had a trivial value
  then either $\langle RUVW\rangle = \langle U\rangle\langle RVW\rangle$ *or* $\langle RUVW\rangle = \langle V\rangle\langle RUW\rangle$

The Most Scalable m Qudit PDs

The PDs of $k = 1$ for $m > 3$ qudits are essentially no different from $m = 3$ because after 3 qudits are chosen, the remaining are fixed to 1 observable or just traced out. For completeness, let

(a) $[d^2; 1, 1 : d^2]$ (b) $[d; d, 1 : d^2]$ (c) $[1; d^2, 1 : d^2]$ (d) $[1; d, d : d^2]$

FIGURE 24. Diagrams for $[N; L_1, L_2 : M]$ corners, the rows of Table 4. These are the most scalable because the external lines can be detached from each other by cutting a single internal line, representing the upper bound on the rank of d^2 if such a factorization exists. Further, corners (b) and (d) suggest meaningful PDs of rank d^3 because they can factor by cutting $1\frac{1}{2}$ lines.

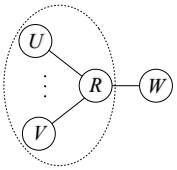


FIGURE 25. There are $\frac{1}{2}m(7m^2 - 12m + 7)$ of the most scalable m qudit PDs, which require $\mathcal{O}(d^4)$ settings.

us write the completely uncorrelated data of $m + 1$ indices,

$$S_a{}^{i\ldots jk} = R_a^{\lambda\ldots\mu\nu} U_\lambda{}^i \cdots V_\mu{}^j W_\nu{}^k, \tag{5.20}$$

and represent the $k = 1$ class of PDs diagrammatically in Figure 25. Also, we include Tables 5 which are just like the last section except with a bunch of ellipses to denote '1's for the remaining qudit measurement settings. We also include the combinatorics for the permuted variations of each PD. Trivial values are also interpreted just the same as for 3 qudits.

Conclusions and Discussion

In this paper, we considered non-holonomic tomography and its application to multiqudit systems. Non-holonomic tomography is the use of partial determinants (PDs) to analyze quantum data for detecting various kinds of correlation in SPAM tomography, where both state and measurement devices have errors. We demonstrated that there are a multitude of PDs one can consider which are sensitive in different ways to the various correlations that can occur. Further, we were able to describe these sensitivities based on the topology of the factorization associated with corresponding notions of an effectively uncorrelated system. For single qudit measurements on a qud^{m}it state, there are m major classes of PD corresponding to matrix rank,

Corners	Squares		
$[N; L_1, L_2, ... : M]$	$[2N; L_1, L_2, ... : 2M]$	$[N; 2L_1, L_2, ... : 2M]$	$[N; L_1, 2L_2, ... : 2M]$
$[d^2; 1, 1, ... : d^2]$	$[2d^2; 1, 1, ... : 2d^2]$	$[d^2; 2, 1, ... : 2d^2]$	
$[d; d, 1, ... : d^2]$	$[2d; d, 1, ... : 2d^2]$	$[d; 2d, 1, ... : 2d^2]$	$[d; d, 2, ... : 2d^2]$
$[1; d^2, 1, ... : d^2]$	$[2; d^2, 1, ... : 2d^2]$	$[1; 2d^2, 1, ... : 2d^2]^*$	$[1; d^2, 2, ... : 2d^2]$
$[1; d, d, ... : d^2]$	$[2; d, d, ... : 2d^2]^{**}$	$[1; 2d, d, ... : 2d^2]$	

(5.21)

Corners	Squares		
$[N; L_1, L_2, ... : M]$	$[2N; L_1, L_2, ... : 2M]$	$[N; 2L_1, L_2, ... : 2M]$	$[N; L_1, 2L_2, ... : 2M]$
$[d^2; 1, 1, ... : d^2]$	m	$m(m-1)$	
$[d; d, 1, ... : d^2]$	$m(m-1)$	$m(m-1)$	$m(m-1)(m-2)$
$[1; d^2, 1, ... : d^2]$	$m(m-1)$	$m(m-1)/2$	$m(m-1)(m-2)$
$[1; d, d, ... : d^2]$	$m(m-1)(m-2)/2$	$m(m-1)(m-2)$	

(5.22)

TABLE 5. The most scalable PDs, requiring only $\mathcal{O}(d^4)$ device settings, are just like those for 3 qudits (Table 4) except that there are more of them by qudit permutation. The combinatorics for the distinct permutations of each PD are given in the second table, with a total of $\frac{1}{2}m(7m^2 - 12m + 7)$.

rank $= d^{2k}$ for $k = 1, \ldots, m$. These ranks in turn determine how many device settings are needed ($\mathcal{O}(\text{rank}^2) = \mathcal{O}(d^{4k})$) to experimentally determine the PD. Finally, we enumerated the class of PDs which require the least number of experimental settings, $k = 1$, for any number of qudits. Figure 26 is provided as a logical sketch for the technique of non-holonomic tomography.

$$\begin{array}{ccccc} \text{(Uncorrelated) Model} & \xrightarrow{\text{“Topology”}} & \text{Rank} \leq r & \longleftrightarrow & 1_{r\times r} \\ \text{(Data) Tensor} & \xrightarrow{\text{Experimental Protocol}} & \text{Matrix} & \xrightarrow{\text{rank } r \text{ PD}} & \Delta_{r\times r} \end{array} \tag{5.23}$$

FIGURE 26. A summary of the logic behind non-holonomic tomography: The data collected from a quantum experiment is a tensor, where each index is associated with each (state preparation or measurement) device. This tensor can be organized as a matrix or “square” in various ways and partial determinants can be calculated for these matrices. Uncorrelated devices correspond to a specific factorization model of the data. The “topology” of this factorization model then sets upper bounds on the rank of any matrix organized from the data. The rank of these matrices are equal to their upper bound if and only if the partial determinant of their correspondingly sized PD is equal to the identity.

PDs have been classified by the types of experimental protocols or “squares” one can consider. However, for each square there are still more PDs corresponding to the order in which the settings are actually put into a matrix. If one considers data from $2d^2 \times 2d^2$ distinct settings, then for a fixed type of square there are actually $[(2d^2)!]^2$ different PDs by permutation of rows and columns. On the other hand, these PDs are certainly not distinct quantities. Some permutations result in PDs which are obviously equivalent, up to familiar transformations, while others result in more obscure equivalences. Analyzing these various permutationally equivalent PDs may help to determine which more precisely which settings have correlation.

The first group of permutations which give obviously equivalent PDs correspond to the ways one can traverse the corners of a square — 4 starting points times 2 directions. If the corners are originally A, B, C, and D as in Equation (5.6), then the 8 PDs that result are

$$\begin{array}{cccc} A^{-1}BD^{-1}C & BD^{-1}CA^{-1} & D^{-1}CA^{-1}B & CA^{-1}BD^{-1} \\ C^{-1}DB^{-1}A & DB^{-1}AC^{-1} & B^{-1}AC^{-1}D & AC^{-1}DB^{-1}. \end{array} \tag{5.24}$$

PDs in the same row are related by cyclic permutation while those in the same column are inverses of each other. Each of these PDs are equivalent to each other up to inverse and conjugation — e.g. $(B^{-1}AC^{-1}D) = (B^{-1}A)\,(A^{-1}BD^{-1}C)^{-1}(B^{-1}A)^{-1}$. The second group corresponds to those permutations that keep settings within their respective corners:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \longrightarrow \begin{bmatrix} \pi_{SP1} & 0 \\ 0 & \pi_{SP2} \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} \pi_{M1} & 0 \\ 0 & \pi_{M2} \end{bmatrix}^{-1} \tag{5.25}$$

where all the πs are $d^2 \times d^2$ permutation matrices. There are $(d^2!)^4$ such elements. These PDs are equivalent to each other up to conjugation since

$$\Delta\left(\begin{bmatrix} \pi_{SP1} & 0 \\ 0 & \pi_{SP2} \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} \pi_{M1} & 0 \\ 0 & \pi_{M2} \end{bmatrix}^{\mathsf{T}}\right) = \pi_{M1}^{-1}\Delta\left(\begin{bmatrix} A & B \\ C & D \end{bmatrix}\right)\pi_{M1} \tag{5.26}$$

where Δ is the standard PD defined by Equations (5.6) and (5.7).[5] Permutations beyond these two groups "delocalize" settings across corners (experiments) and thus give PDs which are equivalent but in a much less obvious way.

Another important comment is that the links between corners, as considered in this paper, have no immediate sense of distance. This is a consequence of the gauge degrees of freedom. We don't a priori have the ability to say how different states in experiment A are from states in experiment B, even if they share the same state settings. However, a notion of distance can be introduced if the

[5]If one considers data from $(d^2+1) \times (d^2+1)$ settings (as described in Appendix 5.7,) then there are $\binom{d^2+1}{2}^2$ distinct PDs, having already divided out the aforementioned equivalences. This is because one must choose the d^2-1 rows and d^2-1 columns of the data that will be common to each corner. The remaining 2 rows and 2 columns are what actually displace the corners.

devices are also equipped with *continuous* settings. A discussion of this technique may be found in [14], Sections III.A, III.C, and IV.

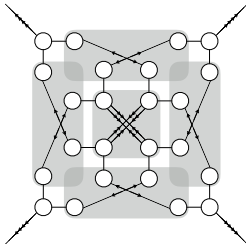


FIGURE 27. A PD (for a two qubit system) that has the topology of a 2-dimensional surface, testing for fundamentally 3-point correlations. Eight copies of Figure 18 can be found along with 4-lined vertices corresponding to contractions with antisymmetric symbols, $\varepsilon_{\mu\nu\rho\sigma}$, which appear in expressions for matrix inverses.

From a mathematical perspective, it is intriguing that there is this relationship between matrix rank and holonomy. These holonomies can in fact be generalized to higher-dimensional quantities (like surfaces etc., rather than just loops) which can test for more general tensor ranks. Such tests can be interpreted as measures of higher n-point correlations between devices. Their construction is relatively simple and requires just one observation: that matrix inverses just consist of several contractions with antisymmetric tensors (or Levi-Civita epsilon symbols.) The first author expects to soon publish full details on the method for constructing such quantities (see Figure 27.)

Acknowledgements

S.J.v.E. was supported in part by ARO/LPS under Contract No. W911NF-14-C-0048.

An $\mathbf{r} \times \mathbf{r}$ PD for an $(\mathbf{r}+\mathbf{1}) \times (\mathbf{r}+\mathbf{1})$ Matrix

In this section, we show how to use PDs to test if an $(r+1) \times (r+1)$ matrix has rank r. Of course, this test should be equivalent to checking if a regular determinant is zero. However, this construction also generalizes to check if any $s \times s$ matrix has rank $r < s$. These tests can be associated with experimental protocols which require fewer device settings than the $2r \times 2r$ protocol. In fact, the $(r+1) \times (r+1)$ protocol has already been applied.[58]

Suppose we have an $(r+1)\times(r+1)$ matrix, S, which we suspect has rank $\leq r$. We can calculate an $r\times r$ PD by generating a $2r\times 2r$ matrix, $\tilde{S}$, partitioning S as follows

$$S=\left[\begin{array}{c|ccc|c} * & * & * & * & * \\ \hline * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ \hline * & * & * & * & * \end{array}\right]=\left[\begin{array}{ccc} a & \vec{\beta}^{\mathsf{T}} & b \\ \vec{\alpha} & \mathbf{M} & \vec{\delta} \\ c & \vec{\gamma}^{\mathsf{T}} & d \end{array}\right]\longrightarrow \tilde{S}=\left[\begin{array}{cc|cc} a & \vec{\beta}^{\mathsf{T}} & \vec{\beta}^{\mathsf{T}} & b \\ \vec{\alpha} & \mathbf{M} & \mathbf{M} & \vec{\delta} \\ \hline \vec{\alpha} & \mathbf{M} & \mathbf{M} & \vec{\delta} \\ c & \vec{\gamma}^{\mathsf{T}} & \vec{\gamma}^{\mathsf{T}} & d \end{array}\right]\equiv\left[\begin{array}{cc} A & B \\ C & D \end{array}\right].$$

It is useful to define the following matrices:

$$\tilde{\alpha}=\left[\begin{array}{cc} 1 & \vec{0}^{\mathsf{T}} \\ -\mathbf{M}^{-1}\vec{\alpha} & \mathbf{1} \end{array}\right]\qquad \tilde{\beta}=\left[\begin{array}{cc} 1 & -\vec{\beta}^{\mathsf{T}}\mathbf{M}^{-1} \\ \vec{0} & \mathbf{1} \end{array}\right]$$

$$\tilde{\gamma}=\left[\begin{array}{cc} \mathbf{1} & \vec{0} \\ -\vec{\gamma}^{\mathsf{T}}\mathbf{M}^{-1} & 1 \end{array}\right]\qquad \tilde{\delta}=\left[\begin{array}{cc} \mathbf{1} & -\mathbf{M}^{-1}\vec{\delta} \\ \vec{0}^{\mathsf{T}} & 1 \end{array}\right]$$

(which one may note are representations of $(r-1)$-dimensional translation groups.) These matrices allow us to partially diagonalize each corner:

$$\tilde{\beta}A\tilde{\alpha}=\left[\begin{array}{cc} A/M & \vec{0}^{\mathsf{T}} \\ \vec{0} & \mathbf{M} \end{array}\right]\qquad \tilde{\beta}B\tilde{\delta}=\left[\begin{array}{cc} \vec{0}^{\mathsf{T}} & B/M \\ \mathbf{M} & \vec{0} \end{array}\right]$$

$$\tilde{\gamma}C\tilde{\alpha}=\left[\begin{array}{cc} \vec{0} & \mathbf{M} \\ C/M & \vec{0}^{\mathsf{T}} \end{array}\right]\qquad \tilde{\gamma}D\tilde{\delta}=\left[\begin{array}{cc} \mathbf{M} & \vec{0} \\ \vec{0}^{\mathsf{T}} & D/M \end{array}\right]$$

where we denote the Schur complements by

$$A/M=a-\vec{\beta}^{T}M^{-1}\vec{\alpha}\qquad B/M=b-\vec{\beta}^{T}M^{-1}\vec{\delta}$$
$$C/M=c-\vec{\gamma}^{T}M^{-1}\vec{\alpha}\qquad D/M=d-\vec{\gamma}^{T}M^{-1}\vec{\delta}.$$

All of the various partial determinants can thus be simplified:

$$A^{-1}BD^{-1}C = 1 + (x-1)\tilde{\alpha} \qquad C^{-1}DB^{-1}A = \frac{1}{x^2}\left[1 + (x-1)\tilde{\alpha}^{-1}\right]$$

$$BD^{-1}CA^{-1} = 1 + (x-1)\tilde{\beta}^{-1} \qquad AC^{-1}DB^{-1} = \frac{1}{x^2}\left[1 + (x-1)\tilde{\beta}\right]$$

$$D^{-1}CA^{-1}B = 1 + (x-1)\tilde{\delta} \qquad B^{-1}AC^{-1}D = \frac{1}{x^2}\left[1 + (x-1)\tilde{\delta}^{-1}\right]$$

$$CA^{-1}BD^{-1} = 1 + (x-1)\tilde{\gamma}^{-1} \qquad DB^{-1}AC^{-1} = \frac{1}{x^2}\left[1 + (x-1)\tilde{\gamma}\right]$$

where

$$x = \frac{(B/M)(C/M)}{(A/M)(D/M)} = \frac{\det B \det C}{\det A \det D}.$$

One can see that each of these PDs is equal to the identity if and only if $x = 1$. This condition on x must be equivalent to $\mathrm{Det}S = 0$ (given the existence of M^{-1}.)

CHAPTER VI

CONCLUSION

Quantum tomography has become an important tool for characterizing quantum devices. [31, 33, 37, 45–47] However, ending up with a legitimate estimate of a *single* density operator, POVM element, or CP map requires some important assumptions about the experimental setup. [42–44] We considered tomography in the setting where one only has imperfect and uncharacterized state-preparation and measurement devices. Particularly, we asked how one could determine if the fluctuations in such state-preparation and measurement devices were correlated, such that $\langle \mathrm{Tr}\rho E\rangle_a{}^i \neq \mathrm{Tr}\langle\rho\rangle_a\langle E\rangle^i$. Such a determination is fundamentally interesting because one does not have access to the individual expectation values, $\langle\rho\rangle_a$ or $\langle E\rangle^i$, but rather only has access to the data or measured frequencies, $\langle \mathrm{Tr}\rho E\rangle_a{}^i$.

We have demonstrated that one can test for the presence of various correlated SPAM errors using so called partial determinants (PDs) which bypass any need to estimate state or measurement parameters individually. The logic behind the PD is simple, uncorrelated SPAM corresponds to a particular ability to factorize the estimated frequencies into a product of state and measurement parameters. Such factorizations always exist for small enough numbers of settings but do not exist for larger numbers of settings if there are correlations. Thus, the notion of parameter independence can be viewed as a local property which may not be "extendable" (i.e. integrable) to a global property. PDs are then a measure of the contradiction that results from requiring that multiple sets of locally uncorrelated settings be consistent with each other. In other words, SPAM correlations correspond to holonomies (or measures of global contradiction) in overcomplete tomography experiments (hence the title, "non-holonomic tomography.")

Non-holonomic tomography quantifies the extent to which a (data) matrix does not have low enough rank to be modeled by uncorrelated devices. This problem of quantifying matrix rank can be treated just as well with two other methods: One could do a *singular value decomposition* (SVD) of the data or one could calculate all *minors* of the appropriate rank. Nevertheless, PDs have the important advantage of an immediate operational and theoretical interpretation, allowing for strong analogies with other techniques in physics and mathematics, thus promoting deeper perspectives.

Non-holonomic tomography has potential applications for quantum computers as well as quantum key distribution. Particularly relevant in quantum information processing is the need for fault-tolerant quantum computation. [3, 4] Various thresholds for fault-tolerance as a function of gate fidelity exist. [5–9] However, all of these thresholds rely on error-correction models that only work under the assumption that gate errors are uncorrelated (so that the probability of n errors scales as ϵ^n for some error rate ϵ.) Further, the dominant contribution to some experimental measurements of gate fidelity is SPAM error. Similarly, there could be analogous correlations of the measurement-measurement type which could prove problematic in measurement based computation. [59–64]

In quantum key distribution (QKD,) several problems have appeared concerning security relative to the inability to trust devices which are used to perform a protocol. Device independent QKD (diQKD) is a protocol where Alice and Bob have measurement devices with settings that correspond to unknown local measurements on an unknown two-qubit state. In such a protocol, security of their keys can be insured by the violation of a Bell inequality. [65–68] However, security is only insured under the assumption that the two measurement devices along with the state preparation device are all uncorrelated with each other — that is, Bell inequalities can certainly be violated by simply correlating the device settings themselves. [69] Similar device correlations can also ruin security in so called measurement-device independent QKD (mdiQKD) where Alice and Bob send local states to an unknown two-qubit measurement, for which security of their keys can be insured by certain outcome selection-rules. [70] Performing non-holonomic tomography in parallel with these protocols, one could in principle be able to detect if such correlations are present. Indeed, non-holonomic tomography could be performed in some already existing protocols which employ so called decoy state. [71]

PDs may also have relevance to problems which inherently have correlations. For example, in state discrimination protocols, so called adaptive measurement techniques are used. [72–74] In state discrimination, states chosen from a known finite set are measured with the task of determining which states had been chosen. Adaptive measurements choose what measurements to perform based on the outcomes due to previous states and measurements. Such a system obviously has SPAM correlation, which may be represented by the value of a partial determinant.

REFERENCES CITED

[1] Merlin Cooper, Michał Karpiński, and Brian J Smith. Local mapping of detector response for reliable quantum state estimation. *Nature communications*, 5, 2014.

[2] Leonard Susskind and Art Friedman. *Quantum mechanics: the theoretical minimum*, volume 2. Basic Books, 2015.

[3] Dorit Aharonov and Michael Ben-Or. Fault-tolerant quantum computation with constant error. In *Proceedings of the twenty-ninth annual ACM symposium on Theory of computing*, pages 176–188. ACM, 1997.

[4] Dorit Aharonov and Michael Ben-Or. Fault-tolerant quantum computation with constant error rate. *SIAM Journal on Computing*, 38(4):1207–1282, 2008.

[5] P Aleferis, D Gottesman, and J Preskill. Quantum accuracy threshold for concatenated distance-3 code. *Quant. Inf. Comput*, 6:97–165, 2006.

[6] Panos Aliferis and John Preskill. Fibonacci scheme for fault-tolerant quantum computation. *Physical Review A*, 79(1):012332, 2009.

[7] Emanuel Knill. Quantum computing with realistically noisy devices. *Nature*, 434(7029):39–44, 2005.

[8] Robert Raussendorf and Jim Harrington. Fault-tolerant quantum computation with high threshold in two dimensions. *Physical review letters*, 98(19):190504, 2007.

[9] David S Wang, Austin G Fowler, and Lloyd CL Hollenberg. Surface code quantum computing with error rates over 1%. *Physical Review A*, 83(2):020302, 2011.

[10] Seth T Merkel, Jay M Gambetta, John A Smolin, Stefano Poletto, Antonio D Córcoles, Blake R Johnson, Colm A Ryan, and Matthias Steffen. Self-consistent quantum process tomography. *Phys. Rev. A*, 87(6):062119, 2013.

[11] Robin Blume-Kohout, John King Gamble, Erik Nielsen, Jonathan Mizrahi, Jonathan D Sterk, and Peter Maunz. Robust, self-consistent, closed-form tomography of quantum logic gates on a trapped ion qubit. *arXiv preprint arXiv:1310.4492*, 2013.

[12] Cyril Stark. Self-consistent tomography of the state-measurement gram matrix. *Phys. Rev. A*, 89(5):052109, 2014.

[13] Christopher Jackson and SJ van Enk. Detecting correlated errors in state-preparation-and-measurement tomography. *Physical Review A*, 92(4):042312, 2015.

[14] Christopher Jackson and Steven van Enk. Non-holonomic tomography i: The born rule as a connection between experiments. *arXiv preprint arXiv:1702.00118*, 2017.

[15] Christopher Jackson and Steven van Enk. Non-holonomic tomography ii: Detecting correlations in multiqudit systems. *arXiv preprint arXiv:1702.06090*, 2017.

[16] Claude Cohen-Tannoudji, Jacques Dupont-Roc, and Gilbert Grynberg. *Photons and Atoms: Introduction to Quantum Electrodynamics*. Wiley-VCH, 2004.

[17] W Heisenberg. Heisenberg nobel lecture. 1933.

[18] Erwin Schrödinger. The fundamental idea of wave mechanics. *Nobel lecture*.

[19] P.A.M. Dirac et al. principles of quantum mechanics. 1930.

[20] Niels Bohr. The structure of the atom. *Nobel lecture*, 11:1922, 1922.

[21] Guido Bacciagaluppi and Antony Valentini. Quantum theory at the crossroads: Reconsidering the 1927 solvay conference. *arXiv preprint quant-ph/0609184*, 2006.

[22] John Von Neumann. *Mathematical foundations of quantum mechanics*. Number 2. Princeton university press, 1955.

[23] P.A.M. Dirac. "the lagrangian in quantum mechanics." this article was first published in physikalische zeitschrift der sowjetunion, band 3, heft 1 (1933), pp. 64–72. 1933.

[24] RP Feynman and AR Hibbs. Quantum mechanics and path integrals. 1965.

[25] Max Born. The statistical interpretation of quantum mechanics. *Nobel Lecture*, 11:1942–1962, 1954.

[26] M Chaichian and A Demichev. Path integrals in physics, vol. 1: Stochastic processes and quantum mechanics. *IOP, Bristol, UK*, 2001.

[27] Israel Gelfand and Mark Neumark. On the imbedding of normed rings into the ring of operators in hilbert space. , 12(2):197–217, 1943.

[28] Edward Brian Davies. Quantum theory of open systems. 1976.

[29] K Vogel and H Risken. Determination of quasiprobability distributions in terms of probability distributions for the rotated quadrature phase. *Physical Review A*, 40(5):2847, 1989.

[30] DT Smithey, M Beck, Michael G Raymer, and A Faridani. Measurement of the wigner distribution and the density matrix of a light mode using optical homodyne tomography: Application to squeezed states and the vacuum. *Physical review letters*, 70(9):1244, 1993.

[31] Jelmer J Renema, Giulia Frucci, Michiel JA de Dood, Richard Gill, Andrea Fiore, and Martin P van Exter. Tomography and state reconstruction with superconducting single-photon detectors. *Phys. Rev. A*, 86(6):062113, 2012.

[32] Joseph B Altepeter, Evan R Jeffrey, and Paul G Kwiat. Photonic state tomography. *Advances in Atomic, Molecular, and Optical Physics*, 52:105–159, 2005.

[33] Alexander I Lvovsky and Michael G Raymer. Continuous-variable optical quantum-state tomography. *Rev. Mod. Phys.*, 81(1):299, 2009.

[34] Lucien Hardy. Quantum theory from five reasonable axioms. *arXiv preprint quant-ph/0101012*, 2001.

[35] Edwin T Jaynes. Information theory and statistical mechanics. *Physical review*, 106(4):620, 1957.

[36] Ulf Leonhardt. *Measuring the quantum state of light*, volume 22. Cambridge university press, 1997.

[37] Matteo Paris and Jaroslav Rehacek. *Quantum state estimation*, volume 649. Springer Science & Business Media, 2004.

[38] Robin Blume-Kohout. Optimal, reliable estimation of quantum states. *New. J. Phys.*, 12(4):043034, 2010.

[39] Robin Blume-Kohout. Hedged maximum likelihood quantum state estimation. *Phys. Rev. Lett.*, 105(20):200504, 2010.

[40] Christian Schwemmer, Lukas Knips, Daniel Richart, Harald Weinfurter, Tobias Moroder, Matthias Kleinmann, and Otfried Gühne. Systematic errors in current quantum state tomography tools. *Phys. Rev. Lett.*, 114(8):080403, 2015.

[41] JOS Yin and SJ van Enk. Information criteria for efficient quantum state estimation. *Physical Review A*, 83(6):062110, 2011.

[42] Steven J van Enk and Robin Blume-Kohout. When quantum tomography goes wrong: drift of quantum sources and other errors. *New J. Phys.*, 15(2):025024, 2013.

[43] Carlton M Caves, Christopher A Fuchs, and Rüdiger Schack. Unknown quantum states: the quantum de finetti representation. *J. Math. Phys.*, 43(9):4537–4559, 2002.

[44] Renato Renner. Symmetry of large physical systems implies independence of subsystems. *Nature Physics*, 3(9):645–649, 2007.

[45] JS Lundeen, A Feito, H Coldenstrodt-Ronge, KL Pregnell, Ch Silberhorn, TC Ralph, J Eisert, MB Plenio, and IA Walmsley. Tomography of quantum detectors. *Nature Physics*, 5(1):27–30, 2009.

[46] A Feito, JS Lundeen, H Coldenstrodt-Ronge, J Eisert, MB Plenio, and IA Walmsley. Measuring measurement: theory and practice. *New Journal of Physics*, 11(9):093038, 2009.

[47] JJ Renema, R Gaudio, Q Wang, Z Zhou, A Gaggero, F Mattioli, R Leoni, D Sahin, MJA De Dood, A Fiore, et al. Experimental test of theories of the detection mechanism in a nanowire superconducting single photon detector. *Phys. Rev. Lett.*, 112(11):117604, 2014.

[48] Asher Peres and Petra F Scudo. Unspeakable quantum information. *arXiv preprint quant-ph/0201017*, 2002.

[49] Denis Rosset, Raphael Ferretti-Schöbitz, Jean-Daniel Bancal, Nicolas Gisin, and Yeong-Cherng Liang. Imperfect measurement settings: Implications for quantum state tomography and entanglement witnesses. *Physical Review A*, 86(6):062325, 2012.

[50] Cyril Stark. Global completability with applications to self-consistent quantum tomography. *arXiv preprint arXiv:1209.6499*, 2012.

[51] David Perez-Garcia, Frank Verstraete, Michael M Wolf, and J Ignacio Cirac. Matrix product state representations. *arXiv preprint quant-ph/0608197*, 2006.

[52] C Schön, K Hammerer, Michael M Wolf, J Ignacio Cirac, and E Solano. Sequential generation of matrix-product states in cavity qed. *Physical Review A*, 75(3):032311, 2007.

[53] Gregory M Crosswhite and Dave Bacon. Finite automata for caching in matrix product algorithms. *Physical Review A*, 78(1):012356, 2008.

[54] Dan Geiger, David Heckerman, Henry King, and Christopher Meek. Stratified exponential families: graphical models and model selection. *Annals of statistics*, pages 505–529, 2001.

[55] Luis David Garcia, Michael Stillman, and Bernd Sturmfels. Algebraic geometry of bayesian networks. *Journal of Symbolic Computation*, 39(3-4):331–355, 2005.

[56] Joe Henson, Raymond Lal, and Matthew F Pusey. Theory-independent limits on correlations from generalized bayesian networks. *New Journal of Physics*, 16(11):113043, 2014.

[57] Lucien Hardy. A formalism-local framework for general probabilistic theories, including quantum theory. *Mathematical Structures in Computer Science*, 23(02):399–440, 2013.

[58] AF McCormick, SJ van Enk, and M Beck. Experimental demonstration of state-preparation-and-measurement tomography. *arXiv preprint arXiv:1701.06498*, 2017.

[59] Robert Raussendorf and Hans J Briegel. Quantum computing via measurements only. *arXiv preprint quant-ph/0010033*, 2000.

[60] M Nielsen. Universal quantum computation using only projective measurement, quantum memory, and preparation of the— 0¿ state, 2001. *arXiv preprint quant-ph/0108020*.

[61] Daniel Gottesman and Isaac L Chuang. Demonstrating the viability of universal quantum computation using teleportation and single-qubit operations. *Nature*, 402(6760):390–393, 1999.

[62] Philippe Jorrand and Simon Perdrix. Unifying quantum computation with projective measurements only and one-way quantum computation. In *Moscow, Russia*, pages 44–51. International Society for Optics and Photonics, 2005.

[63] Robert Raussendorf, Jim Harrington, and Kovid Goyal. Topological fault-tolerance in cluster state quantum computation. *New Journal of Physics*, 9(6):199, 2007.

[64] An-Ning Zhang, Chao-Yang Lu, Xiao-Qi Zhou, Yu-Ao Chen, Zhi Zhao, Tao Yang, and Jian-Wei Pan. Experimental construction of optical multiqubit cluster states from bell states. *Physical Review A*, 73(2):022330, 2006.

[65] Antonio Acín, Nicolas Brunner, Nicolas Gisin, Serge Massar, Stefano Pironio, and Valerio Scarani. Device-independent security of quantum cryptography against collective attacks. *Physical Review Letters*, 98(23):230501, 2007.

[66] Ben W Reichardt, Falk Unger, and Umesh Vazirani. Classical command of quantum systems. *Nature*, 496(7446):456–460, 2013.

[67] Nicolas Brunner, Daniel Cavalcanti, Stefano Pironio, Valerio Scarani, and Stephanie Wehner. Bell nonlocality. *Reviews of Modern Physics*, 86(2):419, 2014.

[68] Valerio Scarani, Helle Bechmann-Pasquinucci, Nicolas J Cerf, Miloslav Dušek, Norbert Lütkenhaus, and Momtchil Peev. The security of practical quantum key distribution. *Reviews of modern physics*, 81(3):1301, 2009.

[69] Ilja Gerhardt, Qin Liu, Antía Lamas-Linares, Johannes Skaar, Valerio Scarani, Vadim Makarov, and Christian Kurtsiefer. Experimentally faking the violation of bells inequalities. *Physical Review Letters*, 107(17):170404, 2011.

[70] Hoi-Kwong Lo, Marcos Curty, and Bing Qi. Measurement-device-independent quantum key distribution. *Physical review letters*, 108(13):130503, 2012.

[71] Hoi-Kwong Lo, Xiongfeng Ma, and Kai Chen. Decoy state quantum key distribution. *Physical review letters*, 94(23):230504, 2005.

[72] FE Becerra, J Fan, G Baumgartner, J Goldhar, JT Kosloski, and A Migdall. Experimental demonstration of a receiver beating the standard quantum limit for multiple nonorthogonal state discrimination. *Nature Photonics*, 7(2):147–152, 2013.

[73] FE Becerra, J Fan, and A Migdall. Implementation of generalized quantum measurements for unambiguous discrimination of multiple non-orthogonal coherent states. *Nature communications*, 4, 2013.

[74] FE Becerra, J Fan, and A Migdall. Photon number resolution enables quantum receiver for realistic coherent optical communications. *Nature Photonics*, 9(1):48–53, 2015.