POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Anytime-Valid Quantum State Tomography via Confidence Sequences

Tesi di Laurea Magistrale in
Mathematical Engineering - Ingegneria Matematica

Author: **Aldo Cumitini**

Student ID: 247334
Advisor: Prof. Luca Barletta
Co-advisors: Prof. Osvaldo Simeone
Academic Year: 2025-26

Abstract

In this work, we address the problem of developing quantum state tomography (QST) methods that remain valid at any time during a sequence of measurements. Specifically, the aim is to provide a rigorous quantification of the uncertainty associated with the current state estimate as data are acquired incrementally. To this end, the proposed framework augments existing QST techniques by associating current point estimates of the state with confidence sets that are guaranteed to contain the true quantum state with a user-defined probability. The methodology is grounded in recent statistical advances in anytime-valid confidence sequences. Numerical results confirm the theoretical coverage properties of the proposed anytime-valid QST.

Keywords: Quantum state tomography, Bayesian inference, confidence sequences

Abstract in lingua italiana

In questo lavoro affrontiamo il problema di sviluppare metodi di tomografia quantistica (QST) che rimangano validi in qualunque istante durante una sequenza di misure. In particolare, l'obiettivo è fornire una quantificazione rigorosa dell'incertezza associata alla stima corrente dello stato mentre i dati vengono acquisiti in modo incrementale. A tal fine, il metodo proposto estende le tecniche esistenti di QST e, a partire dalle stime puntuali dello stato, costruisce insiemi di confidenza che contengono lo stato quantistico con una probabilità definita dall'utente. La metodologia si basa su recenti progressi statistici relativi alle sequenze di confidenza valide "in qualsiasi momento" (anytime-valid confidence sequences). I risultati numerici confermano le proprietà teoriche di copertura della QST proposta, valida in qualsiasi momento.

Parole chiave: Tomografia quantistica, inferenza Bayesiana, sequenze di confidenza

Contents

Acknowledgements

Scrivere una tesi è un percorso lungo, impegnativo e spesso faticoso. Nulla di tutto questo, infatti, sarebbe stato possibile senza l'aiuto di chi mi è stato accanto. Le persone che incontriamo e che scegliamo di avere vicino ci influenzano molto più di quanto immaginiamo: a volte basta un gesto, una parola o una presenza discreta per cambiare la direzione di una giornata, o persino di un intero cammino. Per questo desidero dedicare queste righe di riconoscenza a tutti coloro che mi hanno sostenuto.

Ringrazio innanzitutto i miei genitori, per il supporto e la fiducia che non mi sono mai mancati, anche nei momenti più difficili.

Desidero poi esprimere la mia sincera gratitudine al Prof. Luca Barletta per avermi dato l'opportunità di vivere un'esperienza straordinaria a Londra, per avermi guidato durante il percorso triennale e per averlo fatto ancora una volta con dedizione e attenzione in questa tesi magistrale. Ringrazio inoltre il Prof. Osvaldo Simeone, per avermi accolto al King's College London e per avermi seguito con grande disponibilità, cura e generosità, ben oltre quanto ci si possa aspettare.

Un grazie va anche a tutte le persone incontrate a Londra, al King's College London, con cui ho avuto la fortuna di stringere amicizie che porterò con me. Un pensiero speciale va a Kristian, sempre pronto ad aiutarmi quando si presentava una difficoltà burocratica o nei momenti di sconforto, e Brando, compagno di avventure con cui ho avuto la fortuna di condividere tante discussioni stimolanti e di conoscere la sua volpe londinese "domesticata".

Anche se nel tempo ormai lontani, desidero esprimere la mia profonda gratitudine a Gianluigi Liva e a tutto il team del DLR: senza quei momenti di confronto e crescita 3 anni fa non avrei mai scoperto il mondo del Quantum Computing, e questo lavoro non sarebbe mai nato.

Ringrazio inoltre tutti i miei amici stretti, Giovanni, Saverio, Stefano e Giorgio, e i colleghi dell'università, per i confronti, le risate e il sostegno lungo questo percorso. Un pensiero speciale va ai ragazzi dell'Alta Scuola Politecnica: insieme abbiamo condiviso

tante avventure, costruendo ricordi importanti e nuove amicizie.

Da ultimo, ma non per importanza, ringrazio Letizia, per avermi accompagnato in quest'ultimo anno di magistrale con una presenza costante, paziente e incoraggiante.

Questa tesi è stata un viaggio entusiasmante proprio perché vissuto e condiviso con tante persone. Porterò con me il valore di ogni incontro, di ogni consiglio e di ogni gesto ricevuto lungo la strada. A tutti voi, grazie.

"And so it turned out that only a life similar to the life of those around us, merging with it without a ripple, is genuine life, and that an unshared happiness is not happiness."

Boris Pasternak, Doctor Zhivago

1 | Introduction

Quantum state tomography (QST) is a fundamental procedure for characterizing quantum systems, enabling the estimation of an unknown quantum state from measurement outcomes. As quantum technologies mature and scale toward practical applications in computing, communication, and sensing, the ability to accurately estimate and validate quantum states becomes increasingly critical [1]. Traditional QST methods provide point estimates or credible regions based on accumulated data, but they lack rigorous guarantees in the presence of misspecified models or when inference is performed sequentially. The main objective of this work is to augment existing QST techniques by associating current state estimates with confidence sets that are guaranteed to contain the true quantum state with a user-defined probability *uniformly* over time.

QST has been extensively studied from both frequentist and Bayesian perspectives. Maximum likelihood estimation (MLE) for quantum states was explored in [2], providing consistent point estimates but no uncertainty quantification. The works [3] and [4] developed Bayesian methods using Markov chain Monte Carlo (MCMC) and sequential importance sampling (SIS) to construct posterior estimates and credible regions. The statistical validity of these techniques relies on the correct specification of the prior distribution. Recent advances have explored compressed sensing techniques [5], adaptive measurement strategies [4], and machine learning approaches [6–8]. There is also an important line of work on estimating specific properties of quantum states, rather than recovering the full state [9].

The problem of providing rigorous error bounds for QST has been addressed by several recent works. Notably, the paper [10] introduced confidence regions based on the likelihood ratio that offer rigorous frequentist coverage guarantees (see also [11] for a related theoretical study). While this method provides reliable error bounds, it is designed for a *fixed sample size*. That is, the confidence regions are theoretically valid only for a predetermined number of measurements decided before data collection begins. This limitation poses challenges in sequential or adaptive scenarios where the experimenter may wish to monitor the state estimate in real-time, make decisions based on intermediate results, or stop the experiment when a desired accuracy is reached.

In contrast, this thesis aims at developing *anytime-valid* confidence sets that maintain their coverage guarantees uniformly over all time steps, including at data-dependent stopping times. This property is critical for sequential decision-making. The proposed method, *anytime-valid quantum state tomography* (AV-QST), augments existing QST methods with rigorous, time-uniform confidence guarantees. AV-QST builds on *confidence sequences*, a recent advance in sequential inference [12].

A shorter version of this work has been submitted to *IEEE Signal Processing Letters*. The submission distills the core methodology and theoretical guarantees underlying AV-QST and reports a subset of the numerical results presented in this thesis. A preprint is available at `https://arxiv.org/abs/2601.20761`. The present document provides a more complete exposition, including extended background and discussion.

The remainder of the thesis is organized as follows. Chapter 2 introduces the background material and notation used throughout, including basic concepts from quantum information and the statistical tools underlying anytime-valid inference. Chapter 3 reviews the quantum state reconstruction problem and summarizes conventional approaches to QST, covering frequentist and Bayesian estimators as well as numerical methods used in practice. Chapter 4 presents the main contribution of this thesis by formalizing the notion of anytime validity and introducing the proposed AV-QST framework together with its theoretical guarantees. Chapter 5 reports numerical experiments that validate the theory and compare AV-QST with baseline methods in terms of coverage and set size. Finally, Chapter 6 concludes the thesis and outlines directions for future developments.

2 | Background and Preliminaries

In this chapter we present a brief recap of the preliminary concepts that will be useful for the rest of the thesis.

2.1. Notation

This thesis follows standard notation from quantum information, probability, and sequential inference. We collect here the main conventions used throughout.

Real and complex numbers are denoted by $\mathbb{R}$ and $\mathbb{C}$. Vectors are written in bold (e.g., $\boldsymbol{x}$) and matrices in uppercase (e.g., A). The identity matrix is denoted by I, with dimension clear from context. For a matrix A, $A^\dagger$ denotes the conjugate transpose, and $\mathrm{tr}(A)$ its trace. Tensor (Kronecker) products are written with $\otimes$.

The Hilbert space is denoted by $\mathcal{H}$, with $\dim(\mathcal{H}) = D$ (typically $D = 2^m$ for an m-qubit register). The set of density matrices on $\mathcal{H}$ is

$$\mathcal{D}(\mathcal{H}) \coloneqq \{\rho \in \mathbb{C}^{D\times D} : \ \rho \succeq 0, \ \mathrm{tr}(\rho) = 1\}.$$

The unknown true state is denoted by ρ^*, generic candidates by ρ, and estimators by $\hat{\rho}$ (with $\hat{\rho}_t$ for sequential estimators).

Pure states are represented using Dirac notation: $|\psi\rangle \in \mathcal{H}$ denotes a state vector, and $\langle\psi| \coloneqq |\psi\rangle^\dagger$ its conjugate transpose. The inner product between $|\phi\rangle$ and $|\psi\rangle$ is written as $\langle\phi\,|\,\psi\rangle = \langle\phi||\psi\rangle$. For an operator A on $\mathcal{H}$, the scalar $\langle\phi|A|\psi\rangle$ denotes the corresponding quadratic (or bilinear) form; in particular, $\langle\psi|A|\psi\rangle$ is the expectation of A in the pure state $|\psi\rangle$. The rank-one projector onto $|\psi\rangle$ is $|\psi\rangle\langle\psi|$, and we have the following identity $\langle\psi|A|\psi\rangle = \mathrm{tr}\big(A\,|\psi\rangle\langle\psi|\big)$.

We write $\propto$ to denote proportionality up to a normalization constant, and use standard Landau notation (e.g., $\mathcal{O}(\cdot)$) for computational complexity when needed.

2.2. Quantum Information Basics

This section reviews the basic notions of quantum information needed throughout the thesis; see [13, 14] for comprehensive introductions. We introduce qubits, describe their matrix representation, and summarize the measurement formalism used in quantum state tomography.

2.2.1. Qubits

A classical bit takes values in $\{0, 1\}$, and uncertainty about its value can be described by a Bernoulli distribution. In quantum mechanics, the analogous unit of information is a *qubit*, described by a two-dimensional complex Hilbert space $\mathcal{H} \cong \mathbb{C}^2$. Fixing the computational basis $\{|0\rangle, |1\rangle\}$, any pure qubit state can be written as

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle, \qquad \alpha, \beta \in \mathbb{C}, \quad |\alpha|^2 + |\beta|^2 = 1. \tag{2.1}$$

Vectors that differ by a global phase represent the same physical state, i.e., $|\psi\rangle$ and $e^{i\gamma}|\psi\rangle$ are equivalent for any $\gamma \in \mathbb{R}$.

Exploiting this phase invariance, any pure qubit state admits the polar parametrization

$$|\psi(\theta, \phi)\rangle = \cos\left(\frac{\theta}{2}\right)|0\rangle + e^{i\phi}\sin\left(\frac{\theta}{2}\right)|1\rangle, \qquad \theta \in [0, \pi], \ \phi \in [0, 2\pi). \tag{2.2}$$

Multi-qubit systems. An m-qubit register is described by the tensor-product Hilbert space

$$\mathcal{H}_m := (\mathbb{C}^2)^{\otimes m}, \qquad D := \dim(\mathcal{H}_m) = 2^m,$$

with computational basis $\{|x\rangle : x \in \{0, 1\}^m\}$, where $|x\rangle = |x_1\rangle \otimes \cdots \otimes |x_m\rangle$. A pure m-qubit state is a unit vector $|\psi\rangle \in \mathcal{H}_m$ and can be expanded as

$$|\psi\rangle = \sum_{x \in \{0,1\}^m} \alpha_x |x\rangle, \qquad \alpha_x \in \mathbb{C}, \quad \sum_x |\alpha_x|^2 = 1, \tag{2.3}$$

again defined only up to a global phase.

More general (possibly mixed) states are represented by *density matrices*. For a single qubit this is a 2×2 matrix, while for an m-qubit system it is a $D \times D$ matrix with $D = 2^m$, satisfying

$$\rho \succeq 0, \qquad \operatorname{tr}(\rho) = 1. \tag{2.4}$$

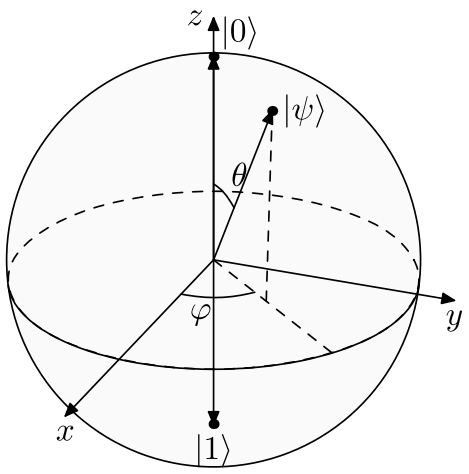


Figure 2.1: Bloch-sphere parametrization of a single-qubit pure state. The Bloch vector $\boldsymbol{r}$ forms a polar angle θ with the z axis and an azimuthal angle ϕ in the xy plane (measured from the x axis), corresponding to the state $|\psi(\theta,\phi)\rangle = \cos(\theta/2)|0\rangle + e^{i\phi}\sin(\theta/2)|1\rangle$.

Pure states correspond to rank-one projectors $\rho = |\psi\rangle\langle\psi|$, while mixed states can be written as convex combinations $\rho = \sum_j p_j|\psi_j\rangle\langle\psi_j|$ with $p_j \geq 0$ and $\sum_j p_j = 1$. A multi-qubit state is called *product* if it factorizes as $\rho = \rho_1 \otimes \cdots \otimes \rho_m$; otherwise it may exhibit *entanglement*, a key nonclassical feature of composite quantum systems.

Bloch-sphere representation (single qubit). Any single-qubit density matrix can be expressed using the Pauli matrices

$$\sigma_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma_y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad \sigma_z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix},$$

as

$$\rho = \frac{1}{2}\left(I + \boldsymbol{r}\cdot\boldsymbol{\sigma}\right) = \frac{1}{2}\left(I + r_x\sigma_x + r_y\sigma_y + r_z\sigma_z\right), \tag{2.5}$$

where $\boldsymbol{r} = (r_x, r_y, r_z) \in \mathbb{R}^3$ is the Bloch vector. The constraint $\rho \succeq 0$ is equivalent to $\|\boldsymbol{r}\| \leq 1$: pure states satisfy $\|\boldsymbol{r}\| = 1$ (surface of the Bloch sphere) and mixed states satisfy $\|\boldsymbol{r}\| < 1$ (interior). Figure 2.1 illustrates the geometric meaning of the angles (θ, ϕ) in (2.2) and the corresponding Bloch vector. For $|\psi(\theta,\phi)\rangle$ in (2.2), the corresponding Bloch vector is

$$\boldsymbol{r}(\theta,\phi) = \big(\sin\theta\cos\phi,\ \sin\theta\sin\phi,\ \cos\theta\big). \tag{2.6}$$

2.2.2. Measurements

A measurement maps a quantum state to a classical outcome. In the projective case, an orthonormal basis $\{|e_x\rangle\}$ defines projectors $|e_x\rangle\langle e_x|$ and yields outcomes distributed as

$$\Pr(x \mid \rho) = \operatorname{tr}(|e_x\rangle\langle e_x|\,\rho).$$

in accordance with Born's rule. Conditioned on observing x, the post-measurement state is projected onto the corresponding basis element (more generally, onto the associated projector), so the measurement irreversibly disturbs the state. As a consequence, a single projective measurement extracts only the information encoded in the chosen basis.

2.2.3. POVMs

The most general measurement model considered in quantum information is a *positive operator-valued measure* (POVM). A POVM on a D-dimensional Hilbert space $\mathcal{H}$ is a collection of positive semidefinite operators

$$\mathcal{M} = \{\Pi_x\}_{x=1}^{M}, \qquad \Pi_x \succeq 0, \quad \sum_{x=1}^{M} \Pi_x = I. \tag{2.7}$$

Given $\rho \in \mathcal{D}(\mathcal{H})$, the outcome $X \in \{1, \ldots, M\}$ obeys

$$\mathbb{P}(X = x \mid \rho; \mathcal{M}) = \operatorname{tr}(\Pi_x \rho). \tag{2.8}$$

A POVM is *informationally complete* (IC) if the outcome probabilities $\{\operatorname{tr}(\Pi_x\rho)\}_{x=1}^{M}$ uniquely determine ρ, which is a key requirement for tomography.

For a single qubit, informational completeness requires at least enough outcomes to identify the three real degrees of freedom of a 2×2 density matrix (equivalently, the components of the Bloch vector), and the minimum possible number is $M = 4$. In this thesis we consider both a minimal informationally complete (MIC) design with four outcomes and an overcomplete, Pauli-basis design with six outcomes.

Example 1 (single-qubit MIC-POVM, $M = 4$). A minimal informationally complete POVM (MIC-POVM) [15] for a qubit has $M = 4$ outcomes. A standard choice is the tetrahedral symmetric IC (SIC) POVM, obtained from four pure states $\{|\psi_x\rangle\}_{x=0}^{3}$ whose Bloch vectors form a regular tetrahedron. Defining rank-one projectors $P_x = |\psi_x\rangle\langle\psi_x|$,

the POVM operators are

$$\Pi_x = \frac{1}{2} P_x = \frac{1}{2}|\psi_x\rangle\langle\psi_x|, \qquad x = 0, 1, 2, 3, \tag{2.9}$$

which satisfy $\sum_{x=0}^{3} \Pi_x = I$. One concrete parametrization consistent with the implementation used in this thesis is

$$\begin{aligned}
|\psi_0\rangle &= |0\rangle, \\
|\psi_1\rangle &= \frac{1}{\sqrt{3}}|0\rangle + \sqrt{\frac{2}{3}}|1\rangle, \\
|\psi_2\rangle &= \frac{1}{\sqrt{3}}|0\rangle + \sqrt{\frac{2}{3}} e^{i\frac{2\pi}{3}}|1\rangle, \\
|\psi_3\rangle &= \frac{1}{\sqrt{3}}|0\rangle + \sqrt{\frac{2}{3}} e^{i\frac{4\pi}{3}}|1\rangle.
\end{aligned} \tag{2.10}$$

Example 2 (single-qubit Pauli measurement, IC, $M = 6$). A common informationally complete design is obtained by combining the eigenbases of the Pauli observables $\sigma_z, \sigma_x, \sigma_y$. Let

$$|\pm\rangle = \frac{1}{\sqrt{2}}(|0\rangle \pm |1\rangle), \qquad |\pm i\rangle = \frac{1}{\sqrt{2}}(|0\rangle \pm i|1\rangle),$$

and consider the six projectors

$$P_0 = |0\rangle\langle 0|, \;\; P_1 = |1\rangle\langle 1|, \;\; P_+ = |+\rangle\langle +|, \;\; P_- = |-\rangle\langle -|, \;\; P_{+i} = |+i\rangle\langle +i|, \;\; P_{-i} = |-i\rangle\langle -i|.$$

A valid POVM can be formed by assigning equal weights,

$$\Pi_x = \frac{1}{3} P_x, \qquad x \in \{0, 1, +, -, +i, -i\}, \tag{2.11}$$

since $P_0 + P_1 = I$, $P_+ + P_- = I$, and $P_{+i} + P_{-i} = I$, hence $\sum_x \Pi_x = \frac{1}{3}(I + I + I) = I$. Operationally, this corresponds to measuring σ_z, σ_x, or σ_y uniformly at random and recording the resulting $\pm$ outcome.

2.3. Statistical Estimation and Inference

This section reviews a set of tools from sequential inference that will be used in Chapter 4 to construct anytime-valid uncertainty sets for quantum state tomography. A comprehensive treatment of e-values and their role in sequential inference can be found in [16].

2.3.1. E-Values, P-Values, and Tests

Let $\mathcal{P}$ denote a set of probability measures representing a null hypothesis (e.g., all data-generating distributions consistent with a given model). Throughout, expectations and probabilities under $\mathbb{P} \in \mathcal{P}$ are denoted by $\mathbb{E}_{\mathbb{P}}[\cdot]$ and $\mathbb{P}(\cdot)$.

An *e-value* (or *e-variable*) for $\mathcal{P}$ is a nonnegative random variable E satisfying

$$\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}[E] \leq 1. \tag{2.12}$$

Intuitively, large values of E provide evidence against the null $\mathcal{P}$, while condition (2.12) ensures that under the null, E cannot be large too often.

A *p-value* P for $\mathcal{P}$ is a random variable in $[0, 1]$ such that, for every $\alpha \in (0, 1)$,

$$\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}(P \leq \alpha) \leq \alpha. \tag{2.13}$$

Small values of P provide evidence against the null.

E-values and p-values play analogous roles (evidence against a null), but they behave differently under accumulation and sequential monitoring: unlike p-values, e-values are designed to combine naturally over time via a simple multiplication.

Finally, a *test* is a random variable $\phi \in [0, 1]$. A binary test takes values in $\{0, 1\}$, and its type-I error under $\mathbb{P} \in \mathcal{P}$ is $\mathbb{E}_{\mathbb{P}}[\phi]$. A test has level α for $\mathcal{P}$ if $\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}[\phi] \leq \alpha$.

2.3.2. E-Processes, Test Martingales, and Optional Stopping

Sequential data are modeled through a filtration $\{\mathcal{F}_t\}_{t\geq 0}$, where $\mathcal{F}_t$ represents the information available up to time t.

A nonnegative adapted process $\{M_t\}_{t\geq 0}$ with $M_0 = 1$ is called a *test supermartingale* for $\mathcal{P}$ if, for every $\mathbb{P} \in \mathcal{P}$,

$$\mathbb{E}_{\mathbb{P}}[M_t \mid \mathcal{F}_{t-1}] \leq M_{t-1} \quad \text{for all } t \geq 1, \qquad \text{and} \qquad \mathbb{E}_{\mathbb{P}}[M_0] \leq 1. \tag{2.14}$$

If equality holds in (2.14), $\{M_t\}$ is a *test martingale*. A sequence $\{E_t\}_{t\geq 0}$ of e-values that is adapted to $\{\mathcal{F}_t\}$ is called an *e-process* if for any stopping time $\tau \in \mathcal{T}$ and any $\mathbb{P} \in \mathcal{P}$, $\mathbb{E}_{\mathbb{P}}[E_\tau] \leq 1$.

A fundamental property is *optional stopping*: if τ is any stopping time with respect to

$\{\mathcal{F}_t\}$, then for every $\mathbb{P} \in \mathcal{P}$, $\mathbb{E}_{\mathbb{P}}[M_\tau] \leq 1$.

This makes test supermartingales robust to continuously monitoring the data and stopping based on intermediate results.

2.3.3. Markov's Inequality and Ville's Inequality

Markov's inequality provides a direct way to calibrate a single e-value into a valid tail bound. Let E be an e-variable for $\mathcal{P}$, then for any $\alpha \in (0, 1)$ and for all $\mathbb{P} \in \mathcal{P}$,

$$\mathbb{P}\left(E \geq \frac{1}{\alpha}\right) \leq \alpha. \tag{2.15}$$

Thus, thresholding an e-value at $1/\alpha$ yields a level-α test.

Ville's inequality is the time-uniform analogue of Markov's inequality for nonnegative test supermartingales (or e-processes). If $\{M_t\}_{t\geq 0}$ is a test supermartingale for $\mathcal{P}$, then for any $\alpha \in (0, 1)$,

$$\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left(\sup_{t\geq 0} M_t \geq \frac{1}{\alpha}\right) \leq \alpha. \tag{2.16}$$

Equivalently, under any $\mathbb{P} \in \mathcal{P}$, the probability that the process ever crosses the threshold $1/\alpha$ is at most α. A useful consequence of (2.16) is that the sequential test that rejects as soon as $M_t \geq 1/\alpha$ has level α, even if the stopping time is data-dependent.

2.3.4. Confidence Sequences via Inversion of E-Processes

Let $\vartheta = \vartheta(\mathbb{P}) \in \Theta$ be a parameter (or, more generally, a functional) of the data-generating distribution. In fixed-sample inference, a $(1-\alpha)$ confidence interval (or confidence set) is a random set $C(\alpha)$ satisfying

$$\mathbb{P}\big(\vartheta(\mathbb{P}) \in C(\alpha)\big) \geq 1 - \alpha \qquad \text{for all } \mathbb{P} \in \mathcal{P}. \tag{2.17}$$

In sequential settings, we define a $(1-\alpha)$ *confidence sequence* (CS): a sequence of random sets $\{C_t(\alpha)\}_{t\geq 0}$ with $C_t(\alpha)$ being $\mathcal{F}_t$-measurable and satisfying the time-uniform (or anytime-valid) coverage property

$$\mathbb{P}\Big(\forall t \geq 0,\ \vartheta(\mathbb{P}) \in C_t(\alpha)\Big) \geq 1 - \alpha \qquad \text{for all } \mathbb{P} \in \mathcal{P}. \tag{2.18}$$

As a direct corollary, for any stopping time τ, one has $\mathbb{P}(\vartheta(\mathbb{P}) \in C_\tau(\alpha)) \geq 1 - \alpha$, so reporting at a data-dependent stopping time remains valid.

A general and practically useful way to build confidence sequences is by *inverting* an e-process indexed by the parameter. Consider a family of e-processes $\{M_t(\theta)\}_{t\geq 0,\ \theta\in\Theta}$ such that the process $M_t(\theta)$ is a test supermartingale under the null hypothesis $\mathcal{P}_\theta := \{\mathbb{P}\in\mathcal{P} : \vartheta(\mathbb{P}) = \theta\}$.

Define the inverted set

$$C_t(\alpha) := \Big\{\theta\in\Theta : M_t(\theta) < \frac{1}{\alpha}\Big\}. \tag{2.19}$$

Then $\{C_t(\alpha)\}_{t\geq 0}$ is a $(1-\alpha)$ confidence sequence. Indeed, letting $\theta^* = \vartheta(\mathbb{P})$, the defining property of $M_t(\theta^*)$ and Ville's inequality (2.16) imply

$$\mathbb{P}\left(\exists t\geq 0 :\ M_t(\theta^*) \geq \frac{1}{\alpha}\right) \leq \alpha, \tag{2.20}$$

which is equivalent to (2.18) via the definition (2.19).

The construction (2.19) is the blueprint used later in this thesis (See Chapter 4). In AV-QST, the role of θ is played by the quantum state ρ, and the e-process is instantiated by a likelihood-ratio-type test martingale. Inverting this process yields a family of sets that forms an *anytime-valid* confidence sequence over the quantum state space.

3 | Quantum State Reconstruction

This chapter introduces the quantum state reconstruction problem and the main estimation and uncertainty-quantification tools used throughout the thesis. We begin by formalizing quantum state tomography as a statistical inference task based on Born's rule and a (possibly time-varying) measurement model. We then review standard tomography approaches from the frequentist and Bayesian perspectives, including maximum-likelihood estimation and Bayesian mean estimation, and discuss how uncertainty is typically summarized via credible sets and confidence regions. Finally, we describe the numerical procedures used in the experimental study.

3.1. Quantum State Tomography

Quantum state tomography is a set of experimental and computational techniques used to reconstruct the quantum state of a system from measurement data. Since a quantum state is disturbed by measurements and any single measurement yields only partial information, tomography proceeds by preparing many identical copies of the system and performing a carefully chosen collection of measurements to estimate the system's density matrix. In practice, quantum state tomography is formulated as a statistical inference problem: one seeks the physical state that best explains the observed outcomes.

3.2. Conventional Methods

We now summarize baseline frequentist and Bayesian QST methods.

3.2.1. Frequentist Quantum State Tomography

In the frequentist setting, the true quantum state $\rho^* \in \mathcal{D}(\mathcal{H})$ is assumed to be fixed but unknown, and the randomness in the data comes entirely from quantum measurement outcomes. For a D-dimensional system (e.g., $D = 2^m$ for an m-qubit register), the state is represented by a $D \times D$ complex-valued density matrix ρ that is positive semidefinite and has unit trace, i.e., $\rho \succeq 0$ and $\mathrm{tr}(\rho) = 1$.

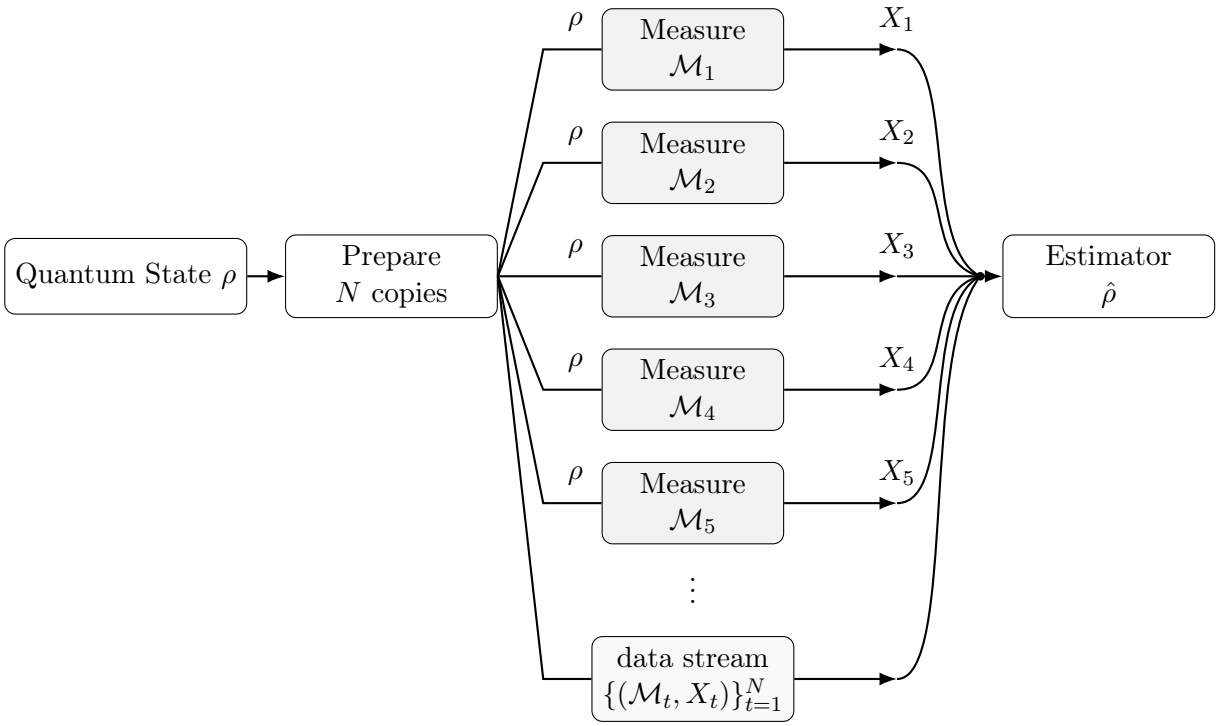


Figure 3.1: Schematic of QST: N identical copies of ρ are measured (possibly with varying POVMs $\mathcal{M}_t$), producing outcomes X_t that are aggregated to form an estimate $\hat{\rho}$.

A tomographic experiment probes ρ^* by measuring many identically prepared copies of the system. Each measurement is described by a POVM $\mathcal{M} = \{\Pi_x\}_{x=1}^M$, where every element satisfies $\Pi_x \succeq 0$ and the completeness relation $\sum_{x=1}^M \Pi_x = I$ holds, with I the $D \times D$ identity. If the system is in state ρ and we apply the POVM $\mathcal{M}$, then the probability of observing outcome x is given by Born's rule:

$$\mathbb{P}(x \mid \rho; \mathcal{M}) = \mathrm{tr}(\Pi_x \rho). \tag{3.1}$$

To model adaptive or time-sequential data acquisition, we consider discrete time steps $t = 1, 2, \ldots, T$. At each step, an experimenter chooses a POVM $\mathcal{M}_t = \{\Pi_{t,x}\}_{x=1}^M$ and records a single outcome $X_t \in \{1, \ldots, M\}$ drawn according to

$$\mathbb{P}(X_t = x \mid \rho; \mathcal{M}_t) = \mathrm{tr}(\Pi_{t,x} \rho). \tag{3.2}$$

Given the realized sequence of measurements and outcomes $\{(\mathcal{M}_t, X_t)\}_{t=1}^T$, the goal of quantum state tomography is to produce an estimate $\hat{\rho}$ that is as close as possible to the unknown ρ^* while respecting the physical constraints of a density matrix. Figure 3.1 provides a graphical summary of this sequential estimation procedure.

One of the earliest and most widely used frequentist estimators is maximum likelihood estimation (MLE) [2]. The key idea is simple: among all physically valid states, choose the one that makes the observed data most probable under the measurement model in (3.2). The likelihood of a candidate state ρ given the data up to time T is defined as

$$\mathcal{L}(\rho) = \prod_{t=1}^{T} \mathbb{P}(X_t \mid \rho; \mathcal{M}_t) = \prod_{t=1}^{T} \operatorname{tr}\big(\Pi_{t,X_t}\rho\big), \tag{3.3}$$

where Π_{t,X_t} denotes the POVM element corresponding to the outcome actually observed at time t. The MLE is then defined as the optimizer

$$\hat{\rho}^{\mathrm{MLE}} = \arg\max_{\rho\in\mathcal{D}(\mathcal{H})} \mathcal{L}(\rho). \tag{3.4}$$

In practice it is more convenient (and numerically stable) to maximize the log-likelihood, which turns the product in (3.3) into a sum:

$$\hat{\rho}^{\mathrm{MLE}} = \arg\max_{\rho\in\mathcal{D}(\mathcal{H})} \sum_{t=1}^{T} \log\left(\operatorname{tr}\big(\Pi_{t,X_t}\rho\big)\right). \tag{3.5}$$

Equivalently, $\hat{\rho}^{\mathrm{MLE}}$ minimizes the negative log-likelihood

$$\hat{\rho}^{\mathrm{MLE}} = \arg\min_{\rho\in\mathcal{D}(\mathcal{H})} -\sum_{t=1}^{T} \log\left(\operatorname{tr}\big(\Pi_{t,X_t}\rho\big)\right). \tag{3.6}$$

This formulation highlights why MLE is attractive in QST. The constraints $\rho \succeq 0$ and $\operatorname{tr}(\rho) = 1$, included in the definition of $\mathcal{D}(\mathcal{H})$, ensure the estimate is always a physical quantum state, unlike simple linear inversion which may yield an unphysical matrix [3]. However, MLE is not without problems. The most visible is that the estimate $\hat{\rho}^{\mathrm{MLE}}$ can be rank-deficient. If $|\psi\rangle$ is an eigenvector corresponding to a zero eigenvalue, then

$$\langle\psi|\hat{\rho}^{\mathrm{MLE}}|\psi\rangle = 0.$$

While such a result is not necessarily unphysical, it is often *implausible*: it assigns exactly zero probability to any outcome of the form $|\psi\rangle\langle\psi|$ whenever $\langle\psi|\hat{\rho}^{\mathrm{MLE}}|\psi\rangle = 0$. In other words, it asserts *complete certainty* that this outcome will never occur, a result that cannot be justified by any finite dataset. See [3] for further discussion. Moreover, the likelihood function in (3.3) can be challenging to optimize: it typically exhibits a complicated landscape, with sharp features and pronounced local maxima and minima.

In practice, a variety of numerical methods can be employed to locate the optimum (or, equivalently, the minimum of the negative log-likelihood in (3.6)).

As an illustration, Figs. 3.2 and 3.3 depict representative likelihood landscapes for a single-qubit state parameterized by the angular coordinates introduced in (2.2) under different measurement configurations. A possible strategy is multi-start optimization, for instance by initializing gradient-based methods from a grid of starting points (as in Figs 3.2 and 3.3). While this can improve the chances of locating the global maximum, it quickly becomes computationally expensive. Alternatively, one may use a smaller number of randomly chosen initializations, which typically improves efficiency but provides weaker assurance of global optimality. This randomized initialization approach is the one used throughout this work.

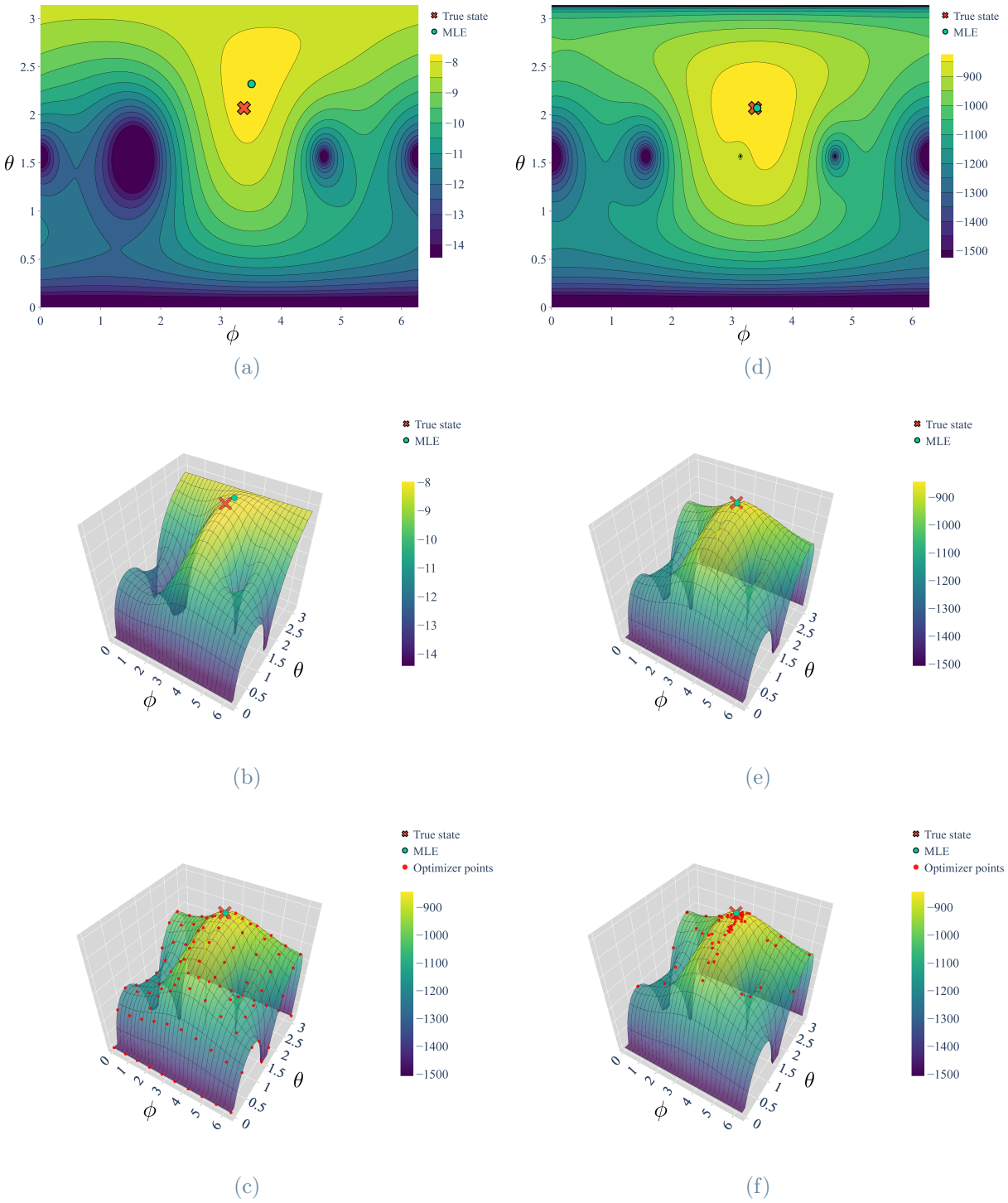


Figure 3.2: $\mathcal{L}(\rho)$ landscape for single-qubit MLE tomography in the (2.11) POVM configuration. Panels (a) and (d) show 2D contour plots of $\mathcal{L}(\rho)$ after $N = 5$ and $N = 500$ measurements, respectively, while panels (b) and (e) report the corresponding 3D surface visualizations. Panels (c) and (f) illustrate the optimization procedure: (c) shows the uniform grid of initial points at the first iteration of the gradient-descent routine, and (f) shows the locations of the starting points after the third iteration, highlighting how the iterates concentrate toward high-likelihood regions. The markers indicate the true state (red cross) and the resulting MLE (green circle).

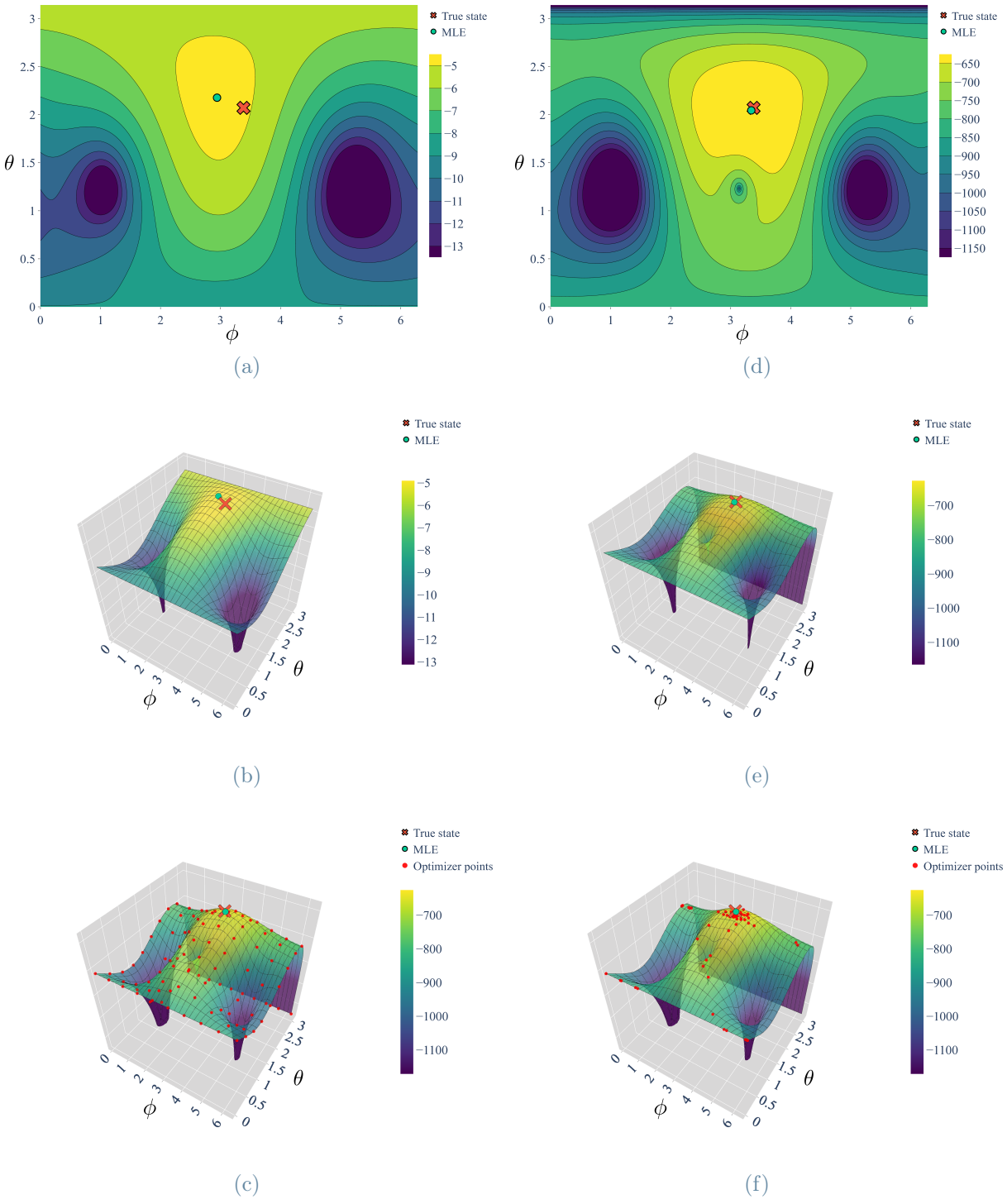


Figure 3.3: $\mathcal{L}(\rho)$ landscape for single-qubit MLE tomography in the (2.9) POVM configuration. Panels (a) and (d) show 2D contour plots of $\mathcal{L}(\rho)$ after $N = 5$ and $N = 500$ measurements, respectively, while panels (b) and (e) report the corresponding 3D surface visualizations. Panels (c) and (f) illustrate the optimization procedure: (c) shows the uniform grid of initial points at the first iteration of the gradient-descent routine, and (f) shows the locations of the starting points after the third iteration, highlighting how the iterates concentrate toward high-likelihood regions. The markers indicate the true state (red cross) and the resulting MLE (green circle).

3.2.2. Bayesian Quantum State Tomography

Bayesian quantum state tomography offers an alternative to MLE by treating the unknown state ρ as a random variable and updating our information about it using Bayes' rule. Starting from a prior distribution $\pi_0(\rho)$ over quantum states and the likelihood of the observed data D, one obtains a posterior distribution $\pi(\rho \mid \mathrm{D})$ that quantifies which states remain plausible after the experiment. Rather than selecting a single *best-fit* state, the Bayesian approach accounts for *all* states compatible with the data, weighted by their posterior probability. A natural point estimator is the Bayesian mean estimate (BME),

$$\hat{\rho}_{\mathrm{BME}} = \mathbb{E}[\rho \mid \mathrm{D}],$$

which is automatically physical (positive semidefinite with unit trace) and, unlike MLE, does not assign exactly zero probability to outcomes on the basis of finite data. Moreover, the posterior provides uncertainty quantification "for free": credible regions can be extracted directly from $\pi(\rho \mid \mathrm{D})$. Finally, Bayesian estimators such as the posterior mean admit a clear decision-theoretic justification: the posterior mean minimizes the expected loss, making it optimal on average [3].

We now briefly outline the Bayesian update used in quantum state tomography; see, e.g., [3, 4] for further details. Like frequentist QST, Bayesian QST produces a point estimate of the underlying state, but it also quantifies uncertainty via a *credible set*. As we previously said, the starting point is the choice of a prior $\pi_0(\rho)$. For pure states, a common "uninformative" choice is the Haar measure (uniform over the unit sphere in Hilbert space), while for mixed states one often adopts an induced prior, discussed in detail in Sec. 3.3.3 and [3].

Considering the same model used for MLE, suppose that by time t we have collected outcomes $\{X_{t'}\}_{t'=1}^{t}$ from measurements $\{\mathcal{M}_{t'}\}_{t'=1}^{t}$, where the probability of observing outcome X_t given ρ and measurement $\mathcal{M}_t$ is given by Born rule

$$\mathbb{P}(X_t \mid \rho; \mathcal{M}_t) = \mathrm{tr}(\Pi_{t,X_t}\rho),$$

with Π_{t,X_t} the POVM element associated with X_t. Let $\mathcal{L}_t(\rho)$ denote the likelihood of the data up to time t (e.g., $\mathcal{L}_t(\rho) = \prod_{t'=1}^{t} \mathrm{tr}(\Pi_{t',X_{t'}}\rho)$ for conditionally independent samples). The posterior at time t is then

$$\pi_t(\rho) \propto \pi_0(\rho)\, \mathcal{L}_t(\rho). \tag{3.7}$$

Equivalently, the posterior can be updated recursively by incorporating the new observation X_t:

$$\pi_t(\rho) \propto \pi_{t-1}(\rho)\,\mathrm{tr}(\Pi_{t,X_t}\rho). \tag{3.8}$$

In this case the BME for the dataset available at time t results in the computation of the integral,

$$\hat{\rho}_t^{\text{B-QST}} = \int \rho\,\pi_t(\rho)\,d\rho. \tag{3.9}$$

In general, the integral in (3.9) is not available in closed form and must be approximated numerically. The numerical procedures adopted in this thesis follow [3] and are described in Sec. 3.3.

Beyond the point estimate, Bayesian QST augments $\hat{\rho}_t^{\text{B-QST}}$ with a credible set C_t [3] targeting coverage level $1-\alpha$, i.e.,

$$\mathbb{P}\left(\rho^* \in C_t \,\middle|\, \{X_{t'}\}_{t'=1}^{t}\right) \geq 1-\alpha, \tag{3.10}$$

where the probability is Bayesian (computed under the prior π_0 for a random $\rho^* \sim \pi_0$), rather than a frequentist guarantee for a fixed true state.

3.3. Numerical Evaluation

In practice, computing the BME reduces to evaluating the posterior expectation (3.9), i.e., an integral over the high-dimensional set of density matrices. Whether this integral admits a closed-form expression is generally unknown, and even when it does, it is rarely available in a form that is useful for tomography. Consequently, one typically resorts to numerical approximations. The procedures described in this section come from [3], they are not optimal but are a proof of principle for BME.

A key simplification for the computation of the integral is that, for a fixed dataset, the likelihood is straightforward to evaluate. Given a sequence of measurement outcomes $\{(\mathcal{M}_t, X_t)\}_{t=1}^{T}$, we recall the likelihood of a candidate state ρ is

$$\mathcal{L}(\rho) = \prod_{t=1}^{T} \mathbb{P}(X_t \mid \rho; \mathcal{M}_t) = \prod_{t=1}^{T} \mathrm{tr}\big(\Pi_{t,X_t}\rho\big), \tag{3.11}$$

where Π_{t,X_t} denotes the POVM element corresponding to the observed outcome at time t. When the data are summarized as outcome counts for each measurement setting, i.e.,

for a fixed POVM $\mathcal{M} = \{\Pi_x\}_{x=1}^{M}$ outcome x is observed n_x times, with $\sum_x n_x = N$, the likelihood can be written compactly as

$$\mathcal{L}(\rho) = \prod_{x=1}^{M} \mathrm{tr}(\Pi_x \rho)^{n_x}, \tag{3.12}$$

which can be evaluated in $\mathcal{O}(M)$ time once the Born probabilities $\mathrm{tr}(\Pi_x\rho)$ are known.

3.3.1. Monte Carlo Integration

In the absence of an analytic expression for the posterior mean (3.9), the required integrals can be approximated using Monte Carlo techniques. A direct approach based on independent random samples of ρ is typically ineffective in tomography, because the posterior density $\pi(\rho) \propto \pi_0(\rho)\mathcal{L}(\rho)$ is often highly concentrated in a small region of the high-dimensional state space. As a result, naive sampling spends most of its time proposing states with negligible posterior weight and converges very slowly.

The Metropolis algorithm [17], a member of the Markov chain Monte Carlo (MCMC) family, is designed precisely for such sharply peaked targets. In particular, the variant Metropolis–Hastings (MH) algorithm [18, 19] is often used for Bayesian estimation: it constructs a Markov chain $\{\rho^{(k)}\}_{k\geq 0}$ whose stationary distribution is the desired posterior (or, equivalently, a density proportional to $\pi_0(\rho)\mathcal{L}(\rho)$).

3.3.2. Metropolis-Hastings Sampling

It is now given a generic version of the MH algorithm for illustrative purposes and then it will be applied to the specific case under exam.

It is often convenient to view Bayesian estimation as the computation of expectations with respect to a *target* probability density. Let π be a probability density on a state space $\mathcal{X}$, and assume that it is only known up to an unknown normalizing constant, i.e.,

$$\pi(x) \propto \tilde{\pi}(x), \qquad x \in \mathcal{X}.$$

The Metropolis–Hastings (MH) algorithm provides a generic way to construct an ergodic and stationary Markov chain

$$X^{(0)}, X^{(1)}, \ldots, X^{(t)}, \ldots$$

whose stationary distribution is π, which means that if $X^{(t)} \sim \pi(x)$ then $X^{(t+1)} \sim \pi(x)$,

thus the Markov Chain converges in distribution to π. After the chain has reached stationarity (approximately), the iterates can be regarded as dependent samples from π, and expectations under π can be approximated by empirical averages. Concretely, for any integrable test function $h : \mathcal{X} \to \mathbb{R}$,

$$\mathcal{J}(h) \coloneqq \mathbb{E}_\pi[h(X)] \quad \text{can be estimated by} \quad \hat{\mathcal{J}}_T(h) \coloneqq \frac{1}{T}\sum_{t=1}^{T} h\big(X^{(t)}\big), \tag{3.13}$$

and, under standard ergodicity conditions, $\hat{\mathcal{J}}_T(h)$ converges to $\mathcal{J}(h)$ as $T \to \infty$. Hence, formula (3.9) , which is just $\mathbb{E}_{\pi_t}[\rho]$, can be estimated by means of MH algorithm. In practice, the initial iterates of the Markov Chain returned by the algorithm are typically influenced by the starting point and are usually discarded as *burn-in*.

It is now given an outline of a general MH algorithm. It is fixed a proposal (transition) density $q(x' \mid x)$ on $\mathcal{X}$. Given the current state $X^{(t)} = x$, MH proceeds as follows:

1. **Propose:** draw a candidate $x' \sim q(\,\cdot \mid x)$.
2. **Accept/reject:** compute the acceptance probability

$$\alpha(x, x') = \min\left\{1,\ \frac{\tilde{\pi}(x')\, q(x \mid x')}{\tilde{\pi}(x)\, q(x' \mid x)}\right\}. \tag{3.14}$$

   Draw $u \sim \mathrm{Unif}[0,1]$. If $u \le \alpha(x, x')$, accept and set $X^{(t+1)} = x'$; otherwise reject and set $X^{(t+1)} = x$.

When the proposal is *symmetric*, i.e. $q(x' \mid x) = q(x \mid x')$, the acceptance probability simplifies to

$$\alpha(x, x') = \min\left\{1,\ \frac{\tilde{\pi}(x')}{\tilde{\pi}(x)}\right\}. \tag{3.15}$$

Intuitively, the algorithm attempts local moves throughout $\mathcal{X}$, but it preferentially spends time in regions where $\tilde{\pi}$ (and hence π) is large, which is precisely what is needed when the target density is sharply concentrated. It is given here a pseudocode for illustrative purposes, with burn-in condition included:

Algorithm 3.1 Metropolis–Hastings (generic)

Require: Unnormalized target density $\tilde{\pi}(x)$ on $\mathcal{X}$ (so that $\pi(x) \propto \tilde{\pi}(x)$); proposal density $q(x' \mid x)$; initial state $x^{(0)} \in \mathcal{X}$; number of iterations T; burn-in length $B < T$.

for $t = 0$ **to** $T - 1$ **do**
  Propose: sample $x' \sim q(\cdot \mid x^{(t)})$.
  Compute acceptance probability:

$$\alpha \leftarrow \min\left\{1,\ \frac{\tilde{\pi}(x')\, q(x^{(t)} \mid x')}{\tilde{\pi}(x^{(t)})\, q(x' \mid x^{(t)})}\right\}.$$

  Sample $u \sim \mathrm{Unif}[0, 1]$.
  if $u \leq \alpha$ **then**
    Accept: $x^{(t+1)} \leftarrow x'$.
  else
    Reject: $x^{(t+1)} \leftarrow x^{(t)}$.
  end if
  if $t + 1 > B$ **then**
    Store sample: add $x^{(t+1)}$ to the output sample set.
  end if
end for

In the case under exam we want to compute the average value of ρ over the integration measure $\pi_0(\rho)\,\mathcal{L}(\rho)d\rho$. By following the analysis in [3] we need a proposal rule J (playing the role of the proposal distribution $q(\cdot \mid x)$ in Algorithm 3.1) for jumping from any ρ to a nearby $\rho' = J(\rho)$, which will be discussed in the next section. Defining the unnormalized target density as $\tilde{\pi}(\rho) = \pi_0(\rho)\,\mathcal{L}_t(\rho)$ we arrive at the following MH procedure:

Algorithm 3.2 Metropolis–Hastings (likelihood)

for $t = 0$ **to** $T - 1$ **do**
  Propose: $\rho' = J(\rho^{(t)})$.
  Compute acceptance probability:

$$\alpha \leftarrow \min\left\{1,\ \frac{\pi_0(\rho')\,\mathcal{L}_t(\rho')}{\pi_0(\rho^{(t)})\,\mathcal{L}_t(\rho^{(t)})}\right\}.$$

  Sample $u \sim \mathrm{Unif}[0,1]$.
  if $u \le \alpha$ **then**
    Accept: $\rho^{(t+1)} \leftarrow \rho'$.
  else
    Reject: $\rho^{(t+1)} \leftarrow \rho^{(t)}$.
  end if
  if $t + 1 > B$ **then**
    Store sample: add $\rho^{(t+1)}$ to the output sample set.
  end if
end for

3.3.3. Induced Measures on Density Matrices

To apply Metropolis–Hastings to quantum states, the crucial ingredient is the *proposal rule* J, which must generate nearby candidates while remaining consistent with the reference measure underlying the prior [3]. In general, measures that can be regarded as "uniform" are characterized by invariance under a symmetry group. For pure states on a d-dimensional Hilbert space, the natural symmetry group is $\mathrm{SU}(d)$, the *special unitary group* consisting of all $d \times d$ unitary matrices with determinant one, and the corresponding invariant distribution is the so-called Haar measure. Extending such a notion of uniformity to *mixed* states is more subtle: unlike pure states, density matrices have eigenvalues that do not enjoy an obvious symmetry, and there is no single canonical choice of a uniform measure on $\mathcal{D}(\mathcal{H})$.

A useful and widely adopted family of priors is given by the *induced measures*. The idea is to generate a mixed state by embedding the system into a larger Hilbert space $\mathcal{H} \otimes \mathcal{K}$ with $\dim(\mathcal{H}) = d$ and $\dim(\mathcal{K}) = k$, drawing a Haar-random *pure* state $|\psi\rangle \in \mathcal{H} \otimes \mathcal{K}$, and then tracing out the auxiliary subsystem:

$$\rho = \mathrm{tr}_{\mathcal{K}}\left(|\psi\rangle\langle\psi|\right). \tag{3.16}$$

Varying k produces a one-parameter family of measures on $\mathcal{D}(\mathcal{H})$, which we denote by $d_k\rho$. Two limiting cases are particularly important: when $k = 1$, the construction reduces to the Haar measure on pure states; when $k = d$, one obtains the Hilbert–Schmidt measure (equivalently, the Lebesgue measure on the real vector space of Hermitian $d \times d$ matrices restricted to $\mathcal{D}(\mathcal{H})$).

This induced-measure construction suggests an efficient way to implement MH proposals on mixed states: instead of proposing ρ directly, we maintain a representative pure state $|\psi\rangle$ in the enlarged space $\mathcal{H} \otimes \mathcal{K}$ and map it to a density matrix via (3.16). The auxiliary degrees of freedom act as internal "hidden variables" that make it straightforward to respect the chosen measure $d_k\rho$ while proposing local moves.

A local random-walk proposal. A naive way to sample Haar-random pure states would be to draw an independent Haar-random unitary at each step, but this produces non-local jumps, hence making the use of MH impossible, and is computationally expensive. Instead, we use a *local* random walk on $\mathrm{SU}(d \times k)$ generated by simple two-dimensional rotations, which is cheap to apply and mixes toward the Haar measure.

Let $D_{ext} = d \times k$ and consider a pure state $|\psi\rangle \in \mathbb{C}^{D_{ext}}$. One MH proposal step consists of:

1. **Choose a two-dimensional subspace:** sample indices $i, j \in \{0, \ldots, D_{ext} - 1\}$ and define a Hermitian generator H_{ij} acting nontrivially only on $\mathrm{span}\{|i\rangle, |j\rangle\}$ (and as the identity elsewhere). A convenient choice is to use Pauli-type generators on this subspace.
2. **Choose a step size:** draw δ from a symmetric distribution with mean 0 and variance Δ^2 (e.g. $\delta \sim \mathcal{N}(0, \Delta^2)$).
3. **Apply the update:** set

$$|\psi'\rangle = J(|\psi\rangle) := e^{i\delta H_{ij}}|\psi\rangle. \tag{3.17}$$

Finally, map the proposed pure state to a mixed-state proposal via

$$\rho' = \mathrm{tr}_{\mathcal{K}}\left(|\psi'\rangle\langle\psi'|\right). \tag{3.18}$$

Because the unitary in (3.17) acts only on a two-dimensional subspace, each proposal is fast to compute and yields a *local* move in state space.

Choosing the step size. The scale Δ controls the trade-off between exploration and acceptance. If Δ is too large, proposals move too far and are often rejected; if Δ is too small, almost all proposals are accepted but the chain explores the posterior slowly. Since the optimal value depends on the unknown posterior concentration (hence on $\mathcal{L}(\rho)$), it is typically tuned adaptively using the observed acceptance rate. A common heuristic is to aim for an acceptance rate on the order of 60%, adjusting Δ gradually based on a sliding window of recent iterations.

The final algorithm which integrates the proposal rule J described previously then becomes:

Algorithm 3.3 Metropolis–Hastings with local Haar-walk proposal J (likelihood target)

Require: Prior density $\pi_0(\rho)$; likelihood $\mathcal{L}(\rho)$; dimensions d, k; steps T; burn-in B; step-size Δ.

Initialize a joint pure state $|\psi^{(0)}\rangle \in \mathbb{C}^{dk}$.
Set $\rho^{(0)} \leftarrow \mathrm{tr}_{\mathcal{K}}\big(|\psi^{(0)}\rangle\langle\psi^{(0)}|\big)$.
for $t = 0$ **to** $T - 1$ **do**
  Proposal J**:** starting from $|\psi^{(t)}\rangle$,
    Sample indices $i, j \in \{0, \ldots, d \times k - 1\}$ uniformly.
    Construct the Hermitian generator H_{ij}.
    Sample a step $\delta \sim \mathcal{N}(0, \Delta^2)$.
    Set $U \leftarrow e^{i\delta H_{ij}}$ and $|\psi'\rangle \leftarrow U|\psi^{(t)}\rangle$.
    Map to a mixed-state proposal $\rho' \leftarrow \mathrm{tr}_{\mathcal{K}}\big(|\psi'\rangle\langle\psi'|\big)$.
  Acceptance probability:

$$\alpha \leftarrow \min\left\{1,\ \frac{\pi_0(\rho')\,\mathcal{L}(\rho')}{\pi_0(\rho^{(t)})\,\mathcal{L}(\rho^{(t)})}\right\}.$$

  Sample $u \sim \mathrm{Unif}[0, 1]$.
  if $u \leq \alpha$ **then**
    Accept: $|\psi^{(t+1)}\rangle \leftarrow |\psi'\rangle, \quad \rho^{(t+1)} \leftarrow \rho'$.
  else
    Reject: $|\psi^{(t+1)}\rangle \leftarrow |\psi^{(t)}\rangle, \quad \rho^{(t+1)} \leftarrow \rho^{(t)}$.
  end if
  if $t + 1 > B$ **then**
    Store sample: add $\rho^{(t+1)}$ to the output chain.
  end if
end for

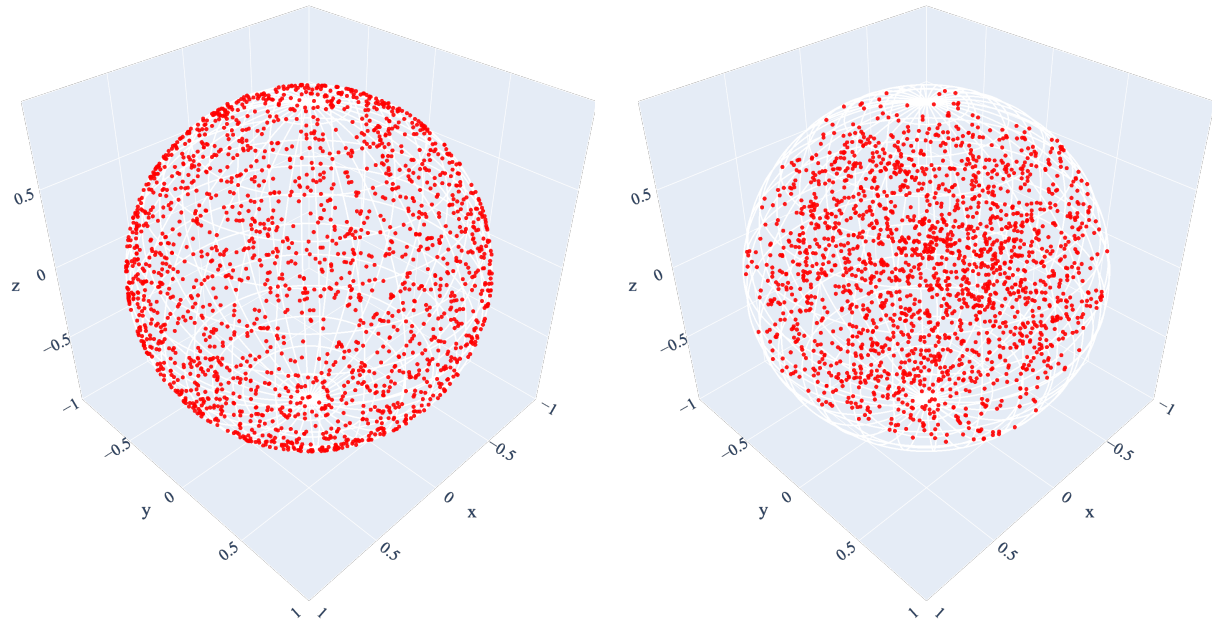


Figure 3.4: Metropolis–Hastings sampling (red dots) from induced priors on single-qubit states, visualized on the Bloch sphere. Panel (a) shows samples drawn with $k = 1$, which corresponds to the Haar-uniform distribution over pure states (points concentrated on the surface). Panel (b) shows samples drawn with $k = 2$, yielding an induced distribution over mixed states (points spread in the interior).

To illustrate the behavior of Algorithm 3.3, Fig. 3.4 shows representative Markov-chain samples (after burn-in) on the Bloch sphere for two choices of induced prior. In the case $k = 1$, the induced measure reduces to the Haar-uniform distribution over pure states, and the chain concentrates on the surface of the sphere, as shown in Fig. 3.4(a). For $k = 2$, the construction yields the Hilbert–Schmidt induced prior over single-qubit mixed states, and the samples spread throughout the Bloch ball, as in Fig. 3.4(b).

3.4. Uncertainty Sets

Beyond producing a point estimate $\hat{\rho}$, a central objective of quantum state tomography is to *quantify the remaining uncertainty* about the unknown state. In practice, an experimentalist typically wants more than a single reconstructed density matrix: they want an assessment of which other states are still compatible with the observed data, and how reliable derived predictions should be. This motivates the construction of *uncertainty sets*, subsets of the state space $\mathcal{D}(\mathcal{H})$ that summarize plausible candidates for the true state.

In this section we review two complementary approaches. The first is Bayesian: starting from a posterior distribution, one can report *credible sets* that contain a prescribed fraction of posterior mass, thereby providing a natural probabilistic summary of uncertainty [3]. The second is frequentist: one constructs *confidence regions* with guaranteed coverage, meaning they contain the true (fixed) state ρ^* with probability at least $1-\alpha$ under repeated sampling [10]. Such confidence regions yield rigorous error bars and represent a state-of-the-art approach to uncertainty quantification in QST; see also [11] for a related construction.

3.4.1. Bayesian Credible Sets

In Bayesian QST, uncertainty is summarized by the posterior distribution $\pi_t(\rho)$ over the state space $\mathcal{D}(\mathcal{H})$ given the data observed up to time t. A *credible set* is then a subset $C_t \subseteq \mathcal{D}(\mathcal{H})$ that contains a prescribed fraction of posterior mass. For a target level $1-\alpha \in (0,1)$, this is typically expressed as

$$\mathbb{P}\left(\rho \in C_t \,\middle|\, \{X_{t'}\}_{t'=1}^{t}\right) \geq 1-\alpha, \tag{3.19}$$

where the probability is computed under the posterior $\pi_t(\rho)$. It is important to emphasize that (3.19) is a *Bayesian* statement: it quantifies uncertainty for a fixed dataset through the posterior induced by the chosen prior $\pi_0(\rho)$. In particular, it does not by itself guarantee frequentist coverage for a fixed, unknown true state ρ^*, and its calibration can be sensitive to prior misspecification.

A canonical choice to calculate the set C_t in (3.19) is the *highest-posterior-density* (HPD) credible region, defined as the set of states whose posterior density exceeds a threshold chosen so that the region carries posterior mass $1-\alpha$. While conceptually appealing, computing exact HPD regions in $\mathcal{D}(\mathcal{H})$ is typically intractable, and in practice one relies on approximations (e.g. particle representations or local Gaussian approximations).

Gaussian credible ellipsoids from posterior moments. A convenient and widely used approximation is to summarize the posterior locally by its mean and covariance, and to define a credible region as an ellipsoid in a suitable coordinate representation [3]. For a single-qubit state ($D=2$), one may represent ρ by its Bloch vector coordinates

$$\mathrm{vec}(\rho) = \begin{bmatrix} \mathrm{tr}(\rho\sigma_x) \\ \mathrm{tr}(\rho\sigma_y) \\ \mathrm{tr}(\rho\sigma_z) \end{bmatrix} \in \mathbb{R}^3, \tag{3.20}$$

where $\{\sigma_x, \sigma_y, \sigma_z\}$ are the Pauli matrices. Given a posterior $\pi_t(\rho)$ and a point estimator $\hat{\rho}_t^{\text{B-QST}}$ (typically the posterior mean), the posterior covariance of these coordinates can be written as

$$\Sigma_t^{\text{B-QST}} = \int \text{vec}(\rho)\,\text{vec}(\rho)^\top \pi_t(\rho)\, d\rho - \text{vec}(\hat{\rho}_t^{\text{B-QST}})\,\text{vec}(\hat{\rho}_t^{\text{B-QST}})^\top. \tag{3.21}$$

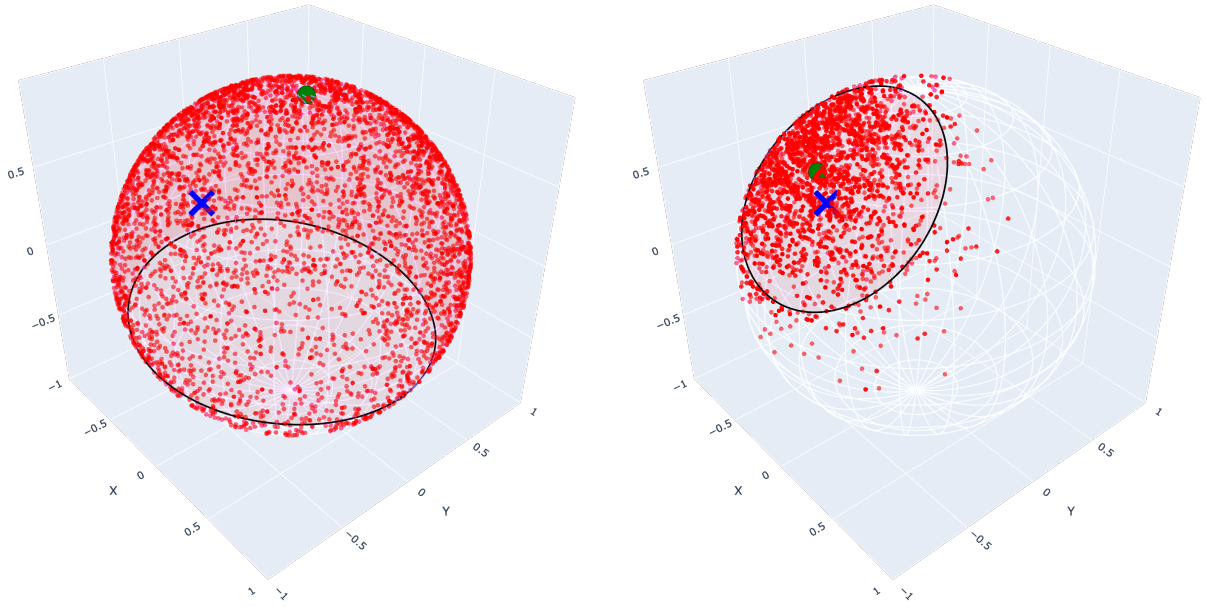
Using this covariance [3], an approximate $(1-\alpha)$ credible set can be defined as the ellipsoidal region

$$C_t^{\text{B-QST}} = \left\{ \rho \in \mathcal{D}(\mathcal{H}) : \big(\text{vec}(\rho) - \text{vec}(\hat{\rho}_t^{\text{B-QST}})\big)^\top \big(\Sigma_t^{\text{B-QST}}\big)^{-1} \big(\text{vec}(\rho) - \text{vec}(\hat{\rho}_t^{\text{B-QST}})\big) \le c_t \right\}, \tag{3.22}$$

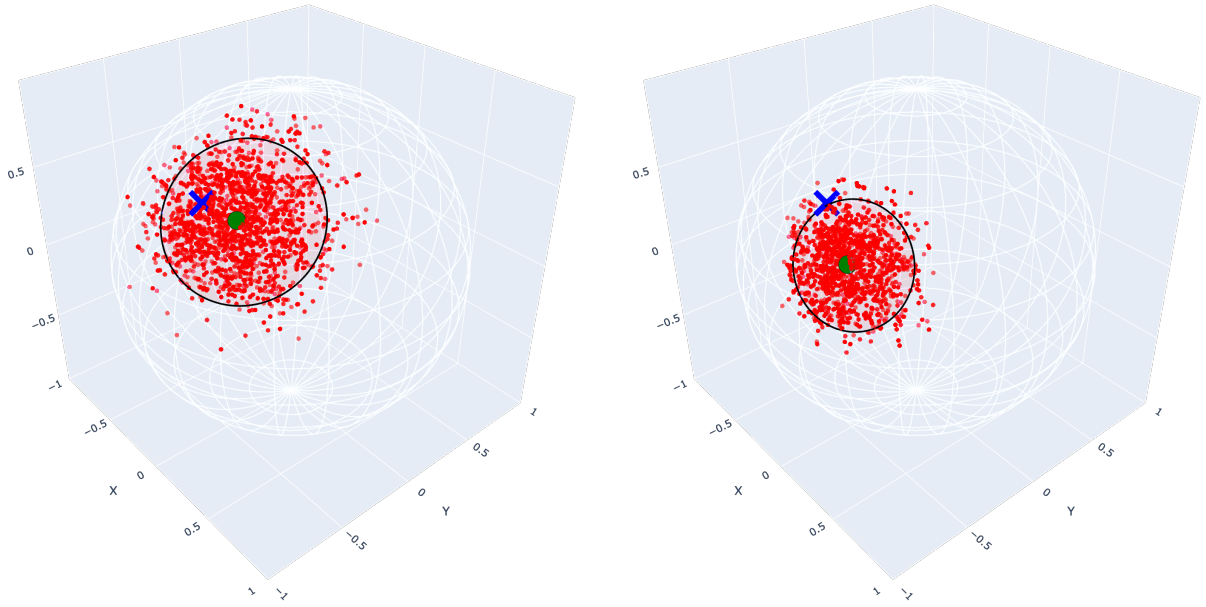
where c_t is a threshold chosen to achieve the desired posterior mass. In a particle approximation of the posterior (e.g. MCMC or SIS), a practical choice, adopted throughout this work, is to select c_t so that approximately a fraction $1-\alpha$ of the posterior samples lies inside the set defined in (3.22).

Fig. 3.5 illustrates the evolution of the posterior samples and the corresponding credible regions, constructed via (3.22), as additional measurement data are collected.

While credible sets may be intuitive and easy to compute from posterior samples, they come with important caveats. First, their interpretation depends on the prior: if $\pi_0(\rho)$ is misspecified, the resulting posterior mass statements may be poorly calibrated. Second, ellipsoidal regions such as (3.22) rely on a (local) Gaussian approximation of the posterior, which may be inaccurate. Finally, credible sets are typically constructed *separately* for each fixed time t, and therefore do not automatically satisfy the stronger *time-uniform* (anytime-valid) requirement introduced in Chapter 4.



(a) $t = 5$ (b) $t = 20$



(c) $t = 50$ (d) $t = 100$

Figure 3.5: Evolution of Bayesian uncertainty for a single-qubit state as data accumulate. Panels (a)–(d) show posterior samples on the Bloch sphere at times $t \in \{5, 20, 50, 100\}$: red points are MH samples drawn from the posterior $\pi_t(\rho)$, the green marker is the BME $\hat{\rho}_t^{\text{B-QST}}$, and the blue cross denotes the true state ρ^*. The black curve delineates the 90% credible region computed from (3.22). As t increases, the posterior contracts and the credible region shrinks; notably, at $t = 100$ the resulting 90% region fails to include the true state.

3.4.2. Confidence Regions

Bayesian credible sets provide an intuitive description of posterior uncertainty, but their interpretation depends on the chosen prior and they do not generally come with a *frequentist* coverage guarantee for a fixed, unknown true state. In contrast, *confidence regions* are designed to satisfy a rigorous coverage property: for a prescribed level $1-\alpha$, the reported set $\hat{R}_\alpha(\mathcal{D}) \subseteq \mathcal{D}(\mathcal{H})$ must contain the true state ρ^* with probability at least $1-\alpha$ under repeated sampling of the data $\mathcal{D}$ generated from ρ^*.

A particularly effective construction for quantum tomography is given by the *likelihood-ratio based* (LR) confidence regions introduced in [10], which we will refer to as LR-QST here. These regions generalize classical error bars to the quantum setting and have appealing features: they achieve guaranteed, user-specified coverage and they are straightforward to define from the likelihood function. A key point for later comparison is that LR-QST is an *offline* procedure: it assumes that the entire dataset $\mathcal{D}$ has already been collected before constructing the region $\hat{R}_\alpha(\mathcal{D})$. As a consequence, the resulting guarantee is tied to a *single, pre-specified* sample size and does not, in general, satisfy the *time-uniform* (anytime-valid) requirement introduced in Chapter 4.

Given observed data $\mathcal{D}$, define the likelihood of a candidate state ρ as

$$\mathcal{L}(\rho) = \mathbb{P}(\mathcal{D} \mid \rho),$$

which in our measurement model takes the product form in (3.3) (or the count form in (3.12)). Let

$$\hat{\rho}^{\mathrm{MLE}} = \arg\max_{\rho \in \mathcal{D}(\mathcal{H})} \mathcal{L}(\rho)$$

be the maximum-likelihood estimate. The (log) likelihood-ratio statistic is then defined as

$$\lambda(\rho) := -2\log\left(\frac{\mathcal{L}(\rho)}{\mathcal{L}(\hat{\rho}^{\mathrm{MLE}})}\right) = 2\Big(\log\mathcal{L}(\hat{\rho}^{\mathrm{MLE}}) - \log\mathcal{L}(\rho)\Big), \tag{3.23}$$

so that small values of $\lambda(\rho)$ correspond to states whose likelihood is close to the maximum.

Likelihood-ratio confidence region. For a target confidence level $1-\alpha$, the LR confidence region in [10] is defined as the set

$$\hat{R}_\alpha(\mathcal{D}) := \big\{\rho \in \mathcal{D}(\mathcal{H}) : \lambda(\rho) \le \lambda_\alpha\big\}, \tag{3.24}$$

where the threshold λ_α is chosen to guarantee frequentist coverage. Increasing λ_α enlarges the region and increases its coverage probability, but decreases its *power* (the ability to

rule out incorrect states). Thus λ_α should be taken as small as possible while still ensuring the desired coverage.

Choosing the threshold. For a generic region construction, the required threshold can depend significantly on the unknown true state ρ^*. A key empirical and theoretical insight in [10] is that the distribution of the LR statistic $\lambda(\rho^*)$ is, in many tomography regimes, approximately *independent* of ρ^*, which motivates replacing a potentially state-dependent threshold by a constant λ_α that satisfies a uniform bound

$$\lambda_\alpha \geq \lambda_\alpha(\rho) \qquad \text{for all } \rho \in \mathcal{D}(\mathcal{H}). \tag{3.25}$$

In principle, one can define the complementary cumulative distribution function (CCDF)

$$F(\lambda \mid \rho) \coloneqq \mathbb{P}\big(\lambda(\rho) > \lambda \mid \rho\big), \tag{3.26}$$

and choose λ_α so that $\sup_\rho F(\lambda_\alpha \mid \rho) \leq \alpha$. Computing $F(\cdot \mid \rho)$ exactly is generally difficult. The work [10] assumes a Gaussian approximation of the data, under which $\lambda(\rho)$ behaves like a χ^2_k random variable. In that case one would set λ_α from

$$F_{\chi^2_k}(\lambda_\alpha) = \frac{\gamma\left(k/2, \lambda_\alpha/2\right)}{\Gamma(k/2)}, \tag{3.27}$$

where $\Gamma(\cdot)$ is the Gamma function and $\gamma(\cdot,\cdot)$ denotes the lower incomplete Gamma function. For large thresholds, one may further use the asymptotic form

$$F_{\chi^2_k}(\lambda_\alpha) \approx \frac{1}{\left(\frac{k}{2}-1\right)!}\left(\frac{\lambda_\alpha}{2}\right)^{k/2-1} e^{-\lambda_\alpha/2}, \qquad \lambda_\alpha \to \infty. \tag{3.28}$$

However, tomographic data are multinomial rather than Gaussian, and this approximation can be overly optimistic, yielding coverage that is correct only on average and potentially much smaller for some states. For this reason, [10] derives a practical *upper bound* that guarantees coverage, at the price of larger regions. In particular, they obtain the bound

$$\begin{aligned} F(\lambda_\alpha) &\leq F_{\chi^2_k}(\lambda_\alpha) + e^{-\lambda_\alpha/2}\left[\left(1+\frac{\sqrt{3e\,\lambda_\alpha}}{\pi}\right)^k - \left(\frac{\sqrt{3e\,\lambda_\alpha}}{\pi}\right)^k\right] \\ &\longrightarrow \quad F_{\chi^2_k}(\lambda_\alpha) + e^{-\lambda_\alpha/2}\, k\, \lambda_\alpha^{(k-1)/2}, \qquad \text{as } \lambda_\alpha \to \infty. \end{aligned} \tag{3.29}$$

In order to calculate the values λ_α one has to solve numerically (3.29).

4 | Problem Formulation

This thesis concerns quantum state tomography (QST) in *sequential* experimental regimes, where measurement data arrive over time and one may wish to inspect, report, or stop the experiment at arbitrary times. Standard tomographic pipelines are typically designed for an *offline* setting: a fixed dataset is collected first, and only then a point estimate $\hat{\rho}$ (and sometimes an uncertainty quantification) is produced. In contrast, many practical scenarios—including adaptive tomography, real-time calibration, and experiments with data-dependent stopping rules—require estimates that are meaningful *at any time* during data acquisition.

Classical frequentist QST procedures commonly output only a point estimate $\hat{\rho}_t$ at time t (see Sec. 3.2.1). Bayesian approaches instead return a posterior distribution and may summarize uncertainty by a credible set $C_t \subseteq \mathcal{D}(\mathcal{H})$ (Sec. 3.2.2). While such credible sets can be informative, their interpretation depends on the correctness of the chosen prior and, in general, they do not provide a frequentist guarantee for a fixed but unknown true state ρ^*.

Motivated by these limitations, we study how to complement either frequentist or Bayesian point estimates with *uncertainty sets* that remain statistically valid throughout the entire measurement sequence. Concretely, we seek a sequence of sets $\{C_t\}_{t\geq 1}$, with $C_t \subseteq \mathcal{D}(\mathcal{H})$, such that at any time t—including random, data-dependent stopping times—the reported set contains the true state with a user-chosen confidence level. This is precisely the notion of *anytime validity* (See Sec. 2.3.4 and [16]).

4.1. Anytime Validity

Let $\rho^* \in \mathcal{D}(\mathcal{H})$ denote the (fixed, unknown) true quantum state and let $\mathcal{F}_t$ be the filtration generated by the observations up to time t (i.e., all measurement choices and outcomes observed so far). An *anytime-valid* (or *time-uniform*) confidence sequence is a sequence of

random sets $\{C_t\}_{t\geq 1}$, where each C_t is $\mathcal{F}_t$-measurable, satisfying the coverage requirement

$$\mathbb{P}\left(\forall t \geq 1,\ \rho^* \in C_t\right) \geq 1-\alpha, \tag{4.1}$$

for a user-specified $\alpha \in (0,1)$.

The strength of (4.1) is that it controls *all times simultaneously*. In particular, it implies valid inference at any deterministic time t, and it also implies validity under optional stopping: for any (possibly random) stopping time τ with respect to the filtration $\{\mathcal{F}_t\}$,

$$\mathbb{P}\left(\rho^* \in C_\tau\right) \geq 1-\alpha. \tag{4.2}$$

Thus, an experimenter may monitor the evolving estimate, stop when a desired precision is reached, and still retain the same confidence level $1-\alpha$ without any post-hoc correction.

It is important to stress that (4.1) is a *frequentist* guarantee: ρ^* is fixed, and the probability is taken over the randomness of the observed measurement outcomes (and any randomized measurement choices), not over a prior on ρ.

4.2. Anytime-Valid Quantum State Tomography

We now introduce *anytime-valid quantum state tomography* (AV-QST), a framework for *online* uncertainty quantification in QST. The central goal is to augment *any* sequential point estimator $\{\hat{\rho}_t\}_{t\geq 1}$, frequentist or Bayesian, with a sequence of confidence sets $\{C_t\}_{t\geq 1}$ that satisfies the anytime-valid coverage requirement (4.1). Our construction is an instance of the general recipe from Sec. 2.3.4: build an e-process (equivalently, a nonnegative test martingale) indexed by candidate states, and invert it to obtain a confidence sequence. The key technical tool ensuring time-uniform validity is Ville's inequality (Sec. 2.3.3); see [16].

4.2.1. Confidence sequences via Likelihood-Ratios

At each time t, AV-QST assumes access to a point predictor from the previous step, $\hat{\rho}_{t-1}$ (for instance, an MLE, a BME, or any other estimator). We initialize the predictor with a fixed random state $\hat{\rho}_0$. Given the data collected up to time t, namely $\{(\mathcal{M}_{t'}, X_{t'})\}_{t'=1}^{t}$, we define a *likelihood-ratio process* indexed by candidate states $\rho \in \mathcal{D}(\mathcal{H})$ as

$$R_t(\rho) := \prod_{t'=1}^{t} \frac{\operatorname{tr}\left(\Pi_{t',X_{t'}}\,\hat{\rho}_{t'-1}\right)}{\operatorname{tr}\left(\Pi_{t',X_{t'}}\,\rho\right)}. \tag{4.3}$$

Intuitively, $R_t(\rho)$ compares how well the one-step-behind predictors $\hat{\rho}_{t'-1}$ explain the observed outcomes relative to the fixed candidate ρ. Smaller values of $R_t(\rho)$ indicate that, along the realized data path, the candidate ρ is more plausible (i.e., it is not strongly disfavored) compared to the running point estimates.

Using this statistic, AV-QST outputs an anytime-valid uncertainty set at time t defined by:

$$C_t^{\text{AV-QST}}(\alpha) := \left\{\rho \in \mathcal{D}(\mathcal{H}) : R_t(\rho) \leq \frac{1}{\alpha}\right\}, \tag{4.4}$$

where $\alpha \in (0, 1)$ is the user-chosen error level. This is equivalent to the inversion rule (2.19) with the e-process $M_t(\rho) = R_t(\rho)$ (Sec. 2.3.4). The set $C_t^{\text{AV-QST}}(\alpha)$ contains exactly those states whose likelihood ratio is not too large; equivalently, there is not overwhelming evidence against them at confidence level $1 - \alpha$.

The related state of the art construction [10] seen in Sec. 3.4.2, called in this work as LR-QST, returns a *single* confidence set only at a fixed final sample size T:

$$C_T^{\text{LR-QST}}(\alpha) := \left\{\rho \in \mathcal{D}(\mathcal{H}) : R_T^{\text{LR-QST}}(\rho) \leq \lambda_\alpha\right\}, \tag{4.5}$$

where the batch likelihood ratio is

$$R_T^{\text{LR-QST}}(\rho) := -2 \log \left(\frac{\mathcal{L}_T(\rho)}{\mathcal{L}_T\left(\hat{\rho}_T^{\text{MLE}}\right)}\right) \tag{4.6}$$

and the threshold λ_α is obtained by solving the implicit calibration equation (3.29). Unlike LR-QST, AV-QST is designed to provide valid uncertainty sets *simultaneously for all times* and it naturally accommodates stronger point predictors than the MLE, such as the BME.

4.2.2. Theoretical Guarantees

The defining feature of AV-QST is that the sequence $\{C_t^{\text{AV-QST}}(\alpha)\}_{t\geq 1}$ is anytime-valid in the frequentist sense. We have the following

Proposition 4.1 (Anytime validity of AV-QST). *For any (possibly data-dependent) sequence of point estimates $\{\hat{\rho}_t\}_{t\geq 1}$, the uncertainty sets defined in* (4.4) *satisfy*

$$\mathbb{P}\left(\forall t \geq 1,\ \rho^* \in C_t^{\text{AV-QST}}(\alpha)\right) \geq 1 - \alpha. \tag{4.7}$$

Proof. By the definition of $C_t^{\text{AV-QST}}(\alpha)$ in (4.4),

$$\Big\{\forall t \geq 1,\ \rho^* \in C_t^{\text{AV-QST}}(\alpha)\Big\} = \Big\{\forall t \geq 1,\ R_t(\rho^*) \leq \frac{1}{\alpha}\Big\}. \tag{4.8}$$

It remains to show that $R_t(\rho^*)$ is a nonnegative test martingale. Define $R_0(\rho^*) \coloneqq 1$ and write (4.3) recursively as

$$R_t(\rho^*) = \frac{\operatorname{tr}\big(\Pi_{t,X_t}\,\hat{\rho}_{t-1}\big)}{\operatorname{tr}\big(\Pi_{t,X_t}\,\rho^*\big)}\, R_{t-1}(\rho^*). \tag{4.9}$$

Let $\mathcal{F}_t = \sigma\big((\mathcal{M}_1, X_1), \ldots, (\mathcal{M}_t, X_t)\big)$ be the natural filtration. Taking conditional expectation given $\mathcal{F}_{t-1}$ and using Born's rule under the true state ρ^* yields

$$\begin{aligned}
\mathbb{E}\left[R_t(\rho^*) \mid \mathcal{F}_{t-1}\right] &= R_{t-1}(\rho^*)\, \mathbb{E}\left[\frac{\operatorname{tr}\big(\Pi_{t,X_t}\,\hat{\rho}_{t-1}\big)}{\operatorname{tr}\big(\Pi_{t,X_t}\,\rho^*\big)} \,\middle|\, \mathcal{F}_{t-1}\right] \\
&= R_{t-1}(\rho^*) \sum_{x=1}^{M} \operatorname{tr}\big(\Pi_{t,x}\rho^*\big)\, \frac{\operatorname{tr}\big(\Pi_{t,x}\,\hat{\rho}_{t-1}\big)}{\operatorname{tr}\big(\Pi_{t,x}\,\rho^*\big)} \\
&= R_{t-1}(\rho^*) \sum_{x=1}^{M} \operatorname{tr}\big(\Pi_{t,x}\,\hat{\rho}_{t-1}\big) = R_{t-1}(\rho^*),
\end{aligned} \tag{4.10}$$

where we used $\sum_{x=1}^{M} \Pi_{t,x} = I$ and $\operatorname{tr}(\hat{\rho}_{t-1}) = 1$ in the last step. Moreover, we have

$$\mathbb{E}[R_t(\rho^*)] = \mathbb{E}[\mathbb{E}(R_t(\rho^*) \mid \mathcal{F}_{t-1})] = \mathbb{E}[R_{t-1}(\rho^*)] = \mathbb{E}[R_0(\rho^*)] = 1. \tag{4.11}$$

Hence $\{R_t(\rho^*)\}_{t\geq 1}$ is a nonnegative martingale with $\mathbb{E}[R_t(\rho^*)] = 1$ for all t.

Finally, Ville's inequality for nonnegative martingales implies

$$\mathbb{P}\left(\sup_{t\geq 1} R_t(\rho^*) \geq \frac{1}{\alpha}\right) \leq \alpha, \tag{4.12}$$

equivalently,

$$\mathbb{P}\left(\forall t \geq 1,\ R_t(\rho^*) \leq \frac{1}{\alpha}\right) \geq 1 - \alpha, \tag{4.13}$$

which is exactly (4.7). □

Proposition 4.1 shows that the AV-QST sets provide a *time-uniform* frequentist guarantee: at any time t, and even under data-dependent stopping, the reported uncertainty set maintains confidence level $1 - \alpha$ for containing the fixed, unknown true state ρ^*.

5 | Experimental Evaluation

This chapter assesses the practical performance of the proposed anytime-valid quantum state tomography framework through numerical simulations. In particular, we empirically compare AV-QST against representative baseline approaches from the literature in terms of (i) *coverage*—whether the reported confidence region contains the true state—and (ii) *informativeness*—how quickly the confidence region shrinks as more measurement data are collected.

5.1. Numerical Results

In this section, we present numerical results for simulated tomography experiments of the proposed AV-QST procedure, and compare it with the Bayesian credible-region approach B-QST [3] and the likelihood-ratio based procedure LR-QST [10]. All simulations were implemented in Python and executed on an M3 Max MacBook Pro. Throughout, we perform local measurements using the MIC-POVM presented in (2.9) separately on each qubit (see [15] and Sec. 2.2.3). For each Monte Carlo run, we generate a random pure quantum state ρ^* according to the Haar-uniform distribution (Sec. 3.3.3). In AV-QST, the point estimate $\hat{\rho}_t$ is taken to be the MLE. For LR-QST, we evaluate the confidence set in (4.5) separately at each time t (in particular, at $t = T$ for the reported comparisons).

A representative visualization of the confidence regions produced by AV-QST and B-QST is shown in Fig. 5.1. In this example, the true state ρ^* is not contained in the B-QST region at time $t = 22$, illustrating the key limitation of B-QST in the sequential setting: the lack of an anytime-validity guarantee. In contrast, AV-QST is designed to maintain the desired coverage uniformly over time.

We next summarize aggregate results over repeated Monte Carlo runs for the one, two, three, and four-qubit settings, with Hilbert-space dimensions $D = 2, 4, 8$, and 16, respectively. The one and two-qubit results are reported in Fig. 5.2, while the three and four-qubit results are reported in Fig. 5.3. Fixing the target miscoverage level to $\alpha = 0.1$, panels (a) and (d) display the empirical miscoverage rate as functions of the time index

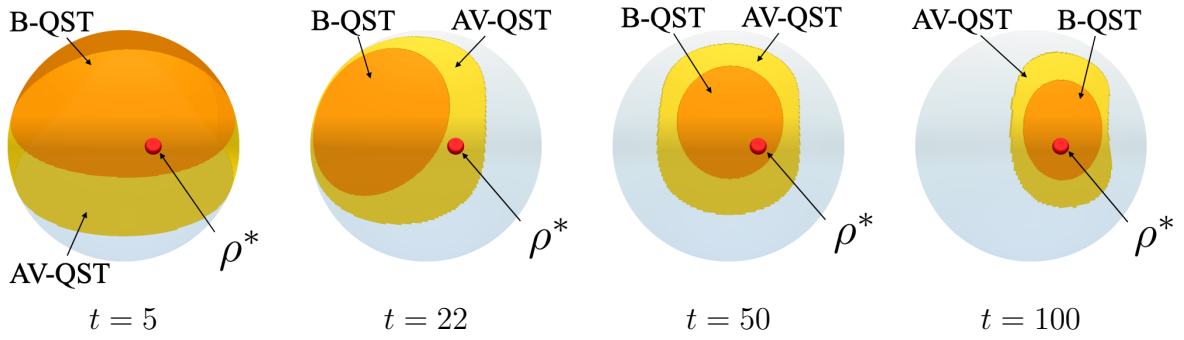


Figure 5.1: Confidence regions obtained as measurements are carried out on copies of a true state ρ^* over discrete time $t = 1, 2, \ldots$ using Bayesian credible regions (B-QST) [3] and the proposed anytime-valid QST (AV-QST). Bayesian QST (B-QST) does not provide time-uniform frequentist coverage guarantees (highlighted at $t = 22$); while the proposed AV-QST ensures that the confidence region includes the true state ρ^* (red dot) with probability no smaller than $1 - \alpha$ *uniformly* over *all* time steps t.

t, while panels (b) and (e) show the corresponding normalized set sizes. The empirical miscoverage rate at time t is estimated by computing the fraction of runs in which the true state ρ^* is not contained in the confidence set up to time t. The normalized set size is estimated by approximating the fraction of the space of density matrices covered by the confidence set, using a sampling-based procedure (Sec. 3.3.1).

Across all system sizes, the confidence regions produced by all methods shrink as t increases and more measurement data are collected, reflecting the increasing information provided by the measurement record. The three procedures, however, exhibit different behavior in terms of coverage. The empirical miscoverage rate of AV-QST remains below the α specified in (4.7) at all times, consistent with the anytime-valid design objective. By contrast, the empirical miscoverage rate of the B-QST credible region increases substantially above the target level, highlighting that it does not provide valid sequential coverage when the confidence region is monitored over time.

For LR-QST, we observe a dimension-dependent conservativeness. In the one-qubit case (panels (a)–(c) of Fig. 5.2), LR-QST behaves similarly to AV-QST in terms of both miscoverage and normalized set size, and we do not observe a pronounced practical difference between the two methods. As the number of qubits grows (panels (d)–(f) of Fig. 5.2 and Fig. 5.3), however, LR-QST becomes increasingly conservative: it almost never miscovers the true state, but this is achieved by producing confidence regions that are substantially larger than those of AV-QST. This widening gap is consistent with the fact that the

LR-QST upper bound depends on the dimension of the system. Consequently, LR-QST becomes markedly less informative than AV-QST in the three and four-qubit settings (Fig. 5.3).

Finally, panels (c) and (f) of Fig. 5.2 (and the corresponding bottom-row panels of Fig. 5.3) report the empirical miscoverage rate as a function of the target miscoverage α, evaluated at $t = 100$. These results reinforce the same qualitative picture: B-QST regions are consistently smaller than the corresponding AV-QST confidence sets, but fail to achieve the desired miscoverage level, whereas LR-QST regions are consistently larger than those of AV-QST, with the degree of over-conservativeness increasing with the number of qubits. Overall, the numerical experiments support the conclusion that AV-QST maintains the required sequential coverage while producing confidence regions that remain informative as the system dimension grows.

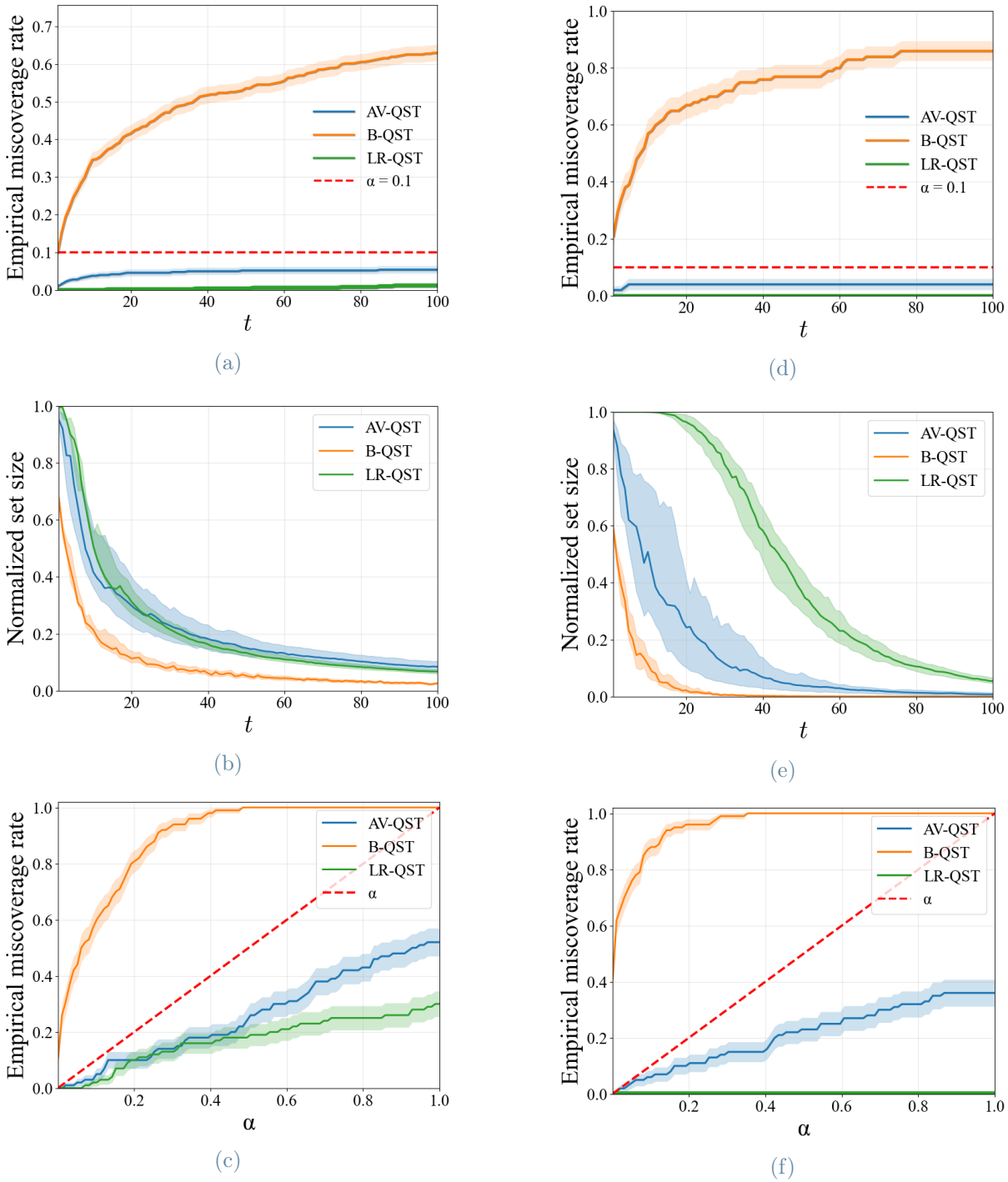


Figure 5.2: Empirical miscoverage and normalized set size for B-QST [3], LR-QST [10], and AV-QST (this thesis). Panels (a)–(c) correspond to the one-qubit setting and panels (d)–(f) to the two-qubit setting. Within each column: top row shows miscoverage rate vs. time t, middle row shows normalized set size vs. time t, and bottom row shows miscoverage rate vs. target α. Lines represent median values and the shaded areas correspond to the interquartile range (25th–75th percentiles).

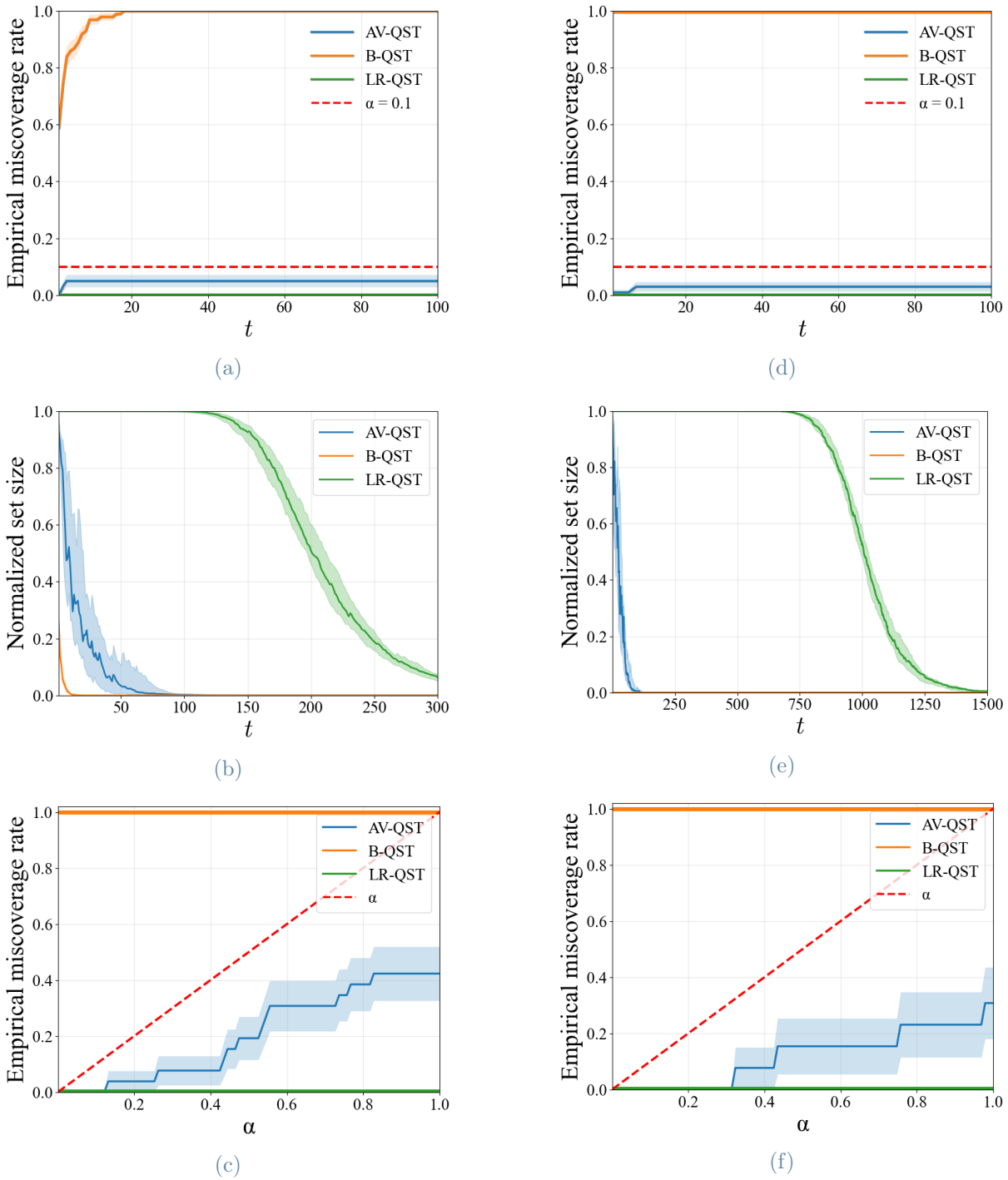


Figure 5.3: Empirical miscoverage and normalized set size for B-QST [3], LR-QST [10], and AV-QST (this thesis). Panels (a)–(c) correspond to the three-qubit setting and panels (d)–(f) to the four-qubit setting. Within each column: top row shows miscoverage rate vs. time t, middle row shows normalized set size vs. time t, and bottom row shows miscoverage rate vs. target α. Lines represent median values and the shaded areas correspond to the interquartile range (25th–75th percentiles).

6 | Conclusions and Future Developments

This chapter concludes the thesis by summarizing the main contributions and findings, and by discussing directions for future research. In particular, we outline extensions of the proposed anytime-valid tomography framework aimed at broadening its applicability and improving its efficiency in realistic experimental scenarios.

6.1. Conclusions

This thesis has developed and evaluated anytime-valid confidence sets for QST. The proposed approach, termed AV-QST, provides a principled method for sequentially reporting confidence regions that satisfy a uniform-in-time coverage requirement, addressing a key limitation of standard non-sequential or fixed-time methods. The theoretical properties of AV-QST (stated in (4.7)) were validated by the numerical simulations reported in Chapter 5 (see Figs. 5.2 and 5.3), which demonstrate practical benefits both in terms of maintaining the desired coverage level and in producing confidence regions of competitive size compared to existing alternatives, namely [10].

More broadly, this work contributes to a growing line of research focused on bringing modern statistical methodology, especially tools designed for sequential and adaptive settings, to quantum information processing [20, 21]. The numerical evidence presented in this thesis supports the central conclusion that anytime-valid inference can provide a favorable balance between rigorous sequential coverage and practical informativeness, making it well suited for real-time QST workflows.

6.2. Future Developments

An important extension of the current work is the development of statistically valid change-detection mechanisms that can detect shifts in the underlying quantum state during an experiment while retaining rigorous guarantees [22, 23]. Such tools would

be valuable in non-stationary settings, where maintaining confidence statements without accounting for changes can lead to misleading inferences. Integrating change detection with anytime-valid confidence sets could enable reliable online monitoring: raising alarms when changes are detected while continuing to provide valid uncertainty quantification before and after change points.

A second promising direction is the design of *adaptive* measurement strategies supported by the proposed framework. Because AV-QST produces valid confidence sets at every time step, it is naturally compatible with tomography protocols in which the measurement settings are updated online as data arrive. The central objective is to accelerate uncertainty reduction—that is, to obtain smaller confidence regions with fewer measurement shots—while retaining rigorous statistical validity.

Concretely, one may couple AV-QST with an experiment-design rule that selects the next measurement configuration by optimizing an information or uncertainty criterion computed from the current estimate and confidence region. For instance, the next POVM could be chosen to minimize the expected volume of the subsequent confidence set, to maximize an information gain proxy. Related ideas have been explored in adaptive tomography via entropy-based criteria [4].

Finally, the numerical investigations carried out in this thesis can be substantially broadened. In addition to the pure-state setting, the same simulation pipeline can be extended to *mixed* states by adopting appropriate priors and estimators, and by tracking the resulting confidence sequences over the full state space. Moreover, the proposed framework is agnostic to the specific point predictor: beyond the MLE, one may use the BME, hybrid predictors that combine MLE and BME, or any other estimator (or data-dependent combination of estimators) that yields sequential predictions. Systematically comparing these choices—in terms of computational cost, empirical tightness of the regions, and robustness across measurement configurations—is a natural next step.

Bibliography

[1] Xiao-Dong Yu, Jiangwei Shang, and Otfried Gühne. Statistical methods for quantum state verification and fidelity estimation. *Advanced Quantum Technologies*, 5(5): 2100126, 2022.

[2] Z. Hradil. Quantum-state estimation. *Phys. Rev. A*, 55:R1561–R1564, Mar 1997. doi: 10.1103/PhysRevA.55.R1561. URL `https://link.aps.org/doi/10.1103/PhysRevA.55.R1561`.

[3] Robin Blume-Kohout. Optimal, reliable estimation of quantum states. *New Journal of Physics*, 12(4):043034, April 2010. ISSN 1367-2630. doi: 10.1088/1367-2630/12/4/043034. URL `http://dx.doi.org/10.1088/1367-2630/12/4/043034`.

[4] Ferenc Huszár and Neil MT Houlsby. Adaptive Bayesian quantum tomography. *Physical Review A*, 85(5):052120, 2012.

[5] David Gross, Yi-Kai Liu, Steven T Flammia, Stephen Becker, and Jens Eisert. Quantum state tomography via compressed sensing. *Physical Review Letters*, 105(15): 150401, 2010.

[6] Giacomo Torlai, Guglielmo Mazzola, Juan Carrasquilla, Matthias Troyer, Roger Melko, and Giuseppe Carleo. Neural-network quantum state tomography. *Nature Physics*, 14(5):447–450, 2018.

[7] Juan Carrasquilla, Giacomo Torlai, Roger G Melko, and Leandro Aolita. Reconstructing quantum states with generative models. *Nature Machine Intelligence*, 1(3): 155–161, 2019.

[8] Duoduo Xue, Wenrui Dai, Ziyang Zheng, Chenglin Li, Junni Zou, and Hongkai Xiong. Generalized quantum state tomography with hybrid denoising priors. *Trans. Sig. Proc.*, 73:1532–1548, January 2025. ISSN 1053-587X. doi: 10.1109/TSP.2025.3546655. URL `https://doi.org/10.1109/TSP.2025.3546655`.

[9] Andreas Elben, Steven T. Flammia, Hsin-Yuan Huang, Richard Kueng, John Preskill, Benoît Vermersch, and Peter Zoller. The randomized measurement toolbox.

Nature Reviews Physics, 5(1):9–24, December 2022. ISSN 2522-5820. doi: 10.1038/s42254-022-00535-2. URL http://dx.doi.org/10.1038/s42254-022-00535-2.

[10] Robin Blume-Kohout. Robust error bars for quantum tomography. *arXiv preprint arXiv:1202.5270*, 2012.

[11] Matthias Christandl and Renato Renner. Reliable quantum state tomography. *Phys. Rev. Lett.*, 109:120403, Sep 2012. doi: 10.1103/PhysRevLett.109.120403. URL https://link.aps.org/doi/10.1103/PhysRevLett.109.120403.

[12] Steven R Howard, Aaditya Ramdas, Jon McAuliffe, and Jasjeet Sekhon. Time-uniform, nonparametric, nonasymptotic confidence sequences. *The Annals of Statistics*, 49(2):1055–1085, 2021.

[13] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, 2010.

[14] Osvaldo Simeone. An introduction to quantum machine learning for engineers, 2022. URL https://arxiv.org/abs/2205.09510.

[15] Joseph M. Renes, Robin Blume-Kohout, A. J. Scott, and Carlton M. Caves. Symmetric informationally complete quantum measurements. *Journal of Mathematical Physics*, 45(6):2171–2180, June 2004. ISSN 1089-7658. doi: 10.1063/1.1737053.

[16] Aaditya Ramdas and Ruodu Wang. Hypothesis testing with e-values. *Foundations and Trends® in Statistics*, 1(1–2):1–390, 2025. ISSN 2978-4220. doi: 10.1561/3600000002.

[17] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 06 1953. ISSN 0021-9606. doi: 10.1063/1.1699114. URL https://doi.org/10.1063/1.1699114.

[18] W. K. Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109, 04 1970. ISSN 0006-3444. doi: 10.1093/biomet/57.1.97. URL https://doi.org/10.1093/biomet/57.1.97.

[19] Siddhartha Chib and Edward Greenberg. Understanding the metropolis-hastings algorithm. *The American Statistician*, 49(4):327–335, 1995. doi: 10.1080/00031305.1995.10476177. URL https://www.tandfonline.com/doi/abs/10.1080/00031305.1995.10476177.

[20] Sangwoo Park and Osvaldo Simeone. Quantum conformal prediction for reliable un-

certainty quantification in quantum machine learning. *IEEE Transactions on Quantum Engineering*, 5:1–24, 2023.

[21] Matteo Zecchin, Osvaldo Simeone, and Aaditya Ramdas. Quantum sequential universal hypothesis testing. *arXiv preprint arXiv:2508.21594*, 2025.

[22] Shubhanshu Shekhar and Aaditya Ramdas. Reducing sequential change detection to sequential estimation, 2023. URL `https://arxiv.org/abs/2309.09111`.

[23] Shubhanshu Shekhar and Aaditya Ramdas. Sequential changepoint detection via backward confidence sequences. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett, editors, *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 30908–30930. PMLR, 23–29 Jul 2023. URL `https://proceedings.mlr.press/v202/shekhar23a.html`.

AI Use Disclosure

In compliance with university guidelines, AI-based tools were used to a limited extent in preparing this thesis and only for assistance with writing and presentation, including reformulating sentences for clarity, converting existing material (e.g., figures or handwritten notes) into LaTeX expressions, and improving the aesthetic appearance of plots. All scientific content, analyses, implementations, interpretations, and conclusions are entirely the result of the author's own work. AI tools did not contribute to the development of original ideas, the design of the research methodology, the production of results, or the derivation of any research outcomes.

List of Figures

List of Symbols

Symbol	Description
$\mathbb{R}, \mathbb{C}$	Real and complex numbers
$\mathbb{P}$	Generic probability measure
$\mathbb{E}$	Expectation operator
I	Identity operator (dimension clear from context)
$\mathrm{tr}(\cdot)$	Trace
$\otimes$	Tensor (Kronecker) product
$(\cdot)^\dagger$	Conjugate transpose
$\mathcal{O}(\cdot)$	Landau notation (order of growth)
$\propto$	Proportionality up to normalization
$\mathcal{H}$	Hilbert space of the system
$\mathcal{K}$	Auxiliary Hilbert space used in the induced-measure construction
D	Dimension of $\mathcal{H}$
d	Dimension of the system Hilbert space in the induced-measure section
$\mathcal{D}(\mathcal{H})$	Set of density matrices on $\mathcal{H}$
ρ	Generic quantum state (density matrix)
ρ^*	True (unknown) quantum state
$\hat{\rho}$	Generic state estimator
$\hat{\rho}_t$	Sequential estimate at time t
$\hat{\rho}^{\mathrm{MLE}}$	Maximum-likelihood estimate (MLE)
$\hat{\rho}_t^{\text{B-QST}}$	Bayesian mean estimate (BME) at time t
$\lvert\psi\rangle, \langle\psi\rvert$	State vector and its dual
$\langle\phi \,\vert\, \psi\rangle$	Inner product
$\lvert\psi\rangle\langle\psi\rvert$	Rank-one projector
$\sigma_x, \sigma_y, \sigma_z$	Pauli matrices
$\boldsymbol{r}$	Bloch vector
r_x, r_y, r_z	Components of the Bloch vector
θ, ϕ	Bloch-sphere angles

Symbol	Description
$\mathrm{vec}(\rho)$	Bloch-vector coordinates of ρ
$\mathcal{M}$	POVM (measurement)
$\mathcal{M}_t$	POVM used at time t
M	Number of outcomes of a POVM
Π_x	POVM element for outcome x
$\Pi_{t,x}$	POVM element for outcome x at time t
X_t	Measurement outcome at time t
N	Number of measurement shots (when used as total count)
n_x	Count of outcome x
$\mathcal{L}(\rho)$	Likelihood (full dataset)
$\mathcal{L}_t(\rho)$	Likelihood up to time t
$\log \mathcal{L}_t(\rho)$	Log-likelihood up to time t
$\pi_0(\rho)$	Prior density over states
$\pi_t(\rho)$	Posterior density at time t
C_t	Generic uncertainty set at time t
$C_t^{\text{B-QST}}$	Bayesian credible set at time t
$\Sigma_t^{\text{B-QST}}$	Posterior covariance
c_t	Credible-ellipsoid threshold
ϑ	Generic parameter / functional of interest
Θ	Parameter space for ϑ
$\mathcal{P}$	Null hypothesis / family of probability measures
$\lambda(\rho)$	Likelihood-ratio statistic
λ_α	LR-QST threshold at level $1-\alpha$
$\hat{R}_\alpha(\mathcal{D})$	LR-QST confidence region
$F(\lambda \mid \rho)$	CCDF used for LR-QST threshold calibration
$F_{\chi^2_k}(\cdot)$	Chi-square CDF with k degrees of freedom
k	Degrees of freedom / induced-prior ancilla dimension (context-dependent)
α	Miscoverage level
$1-\alpha$	Target coverage level
$\mathcal{F}_t$	Filtration (information up to time t)
τ	(Data-dependent) stopping time
M_t	Test martingale / test supermartingale at time t
E_t	E-process / sequential e-value at time t
$R_t(\rho)$	AV-QST likelihood-ratio process

Symbol	Description
$C_t^{\text{AV-QST}}(\alpha)$	Anytime-valid confidence set at time t
$\tilde{\pi}$	Unnormalized target density in Metropolis–Hastings
$\mathcal{X}$	State space of the generic Metropolis–Hastings chain
J	Proposal rule (quantum-state proposal)
$q(\cdot \mid \cdot)$	Proposal density (generic MH notation)
T	Number of MH iterations / time horizon (context-dependent)
B	Burn-in length
Δ	MH step-size parameter
$\alpha(x, x')$	MH acceptance probability