OPTIMAL PARAMETER ESTIMATION FOR QUANTUM MARKOV CHAINS AND IID PURE STATE MODELS

ALFRED GODLEY

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

Supervised by

MĂDĂLIN GUŢĂ

University of Nottingham

September 2024

ABSTRACT

This thesis presents several optimal estimation strategies for quantum parameter estimation, mostly applied to quantum Markov chains (QMCs) - discrete-time quantum open systems. Chapter 4, based on [78], solves the problem of optimal estimation of a QMC using adaptive measurements. This culminates in an algorithm that determines the optimal measurement basis for the next step of the QMC by tracking an object known as the measurement filter. Crucially, this involves only local measurements on the environment of the QMC; a considerable simplification in comparison to other optimal strategies. Additionally, this estimation scheme utilizes a coherent absorber. This is a secondary system that we use to post-process the quantum Markov chain before measurement, which we include to purify the stationary state of the QMC.

Chapter 5, based on [77], then examines another measurement scheme - null measurements. A null measurement involves measuring the quantum system in a basis that contains the system's state. It has been claimed that this is an optimal measurement for parameter estimation, but we demonstrate that this is not the case through an independent identically distributed pure state model. This is due to an identifiability issue that occurs when one attempts the naive implementation the null measurement. We then present a solution to this problem, where we introduce some displacement into the null measurement. This removes the identification issue and provides a proper statistical foundation to the estimation scheme.

In chapter 6, based on [75], we apply this displaced null measurement to a QMC; this provides a second optimal estimation scheme for QMCs. In particular, the displaced null measurement is implemented through the coherent absorber and a simple fixed measurement basis on the environment of the QMC. Therefore, this estimation scheme can be implemented with considerably reduced measurement complexity. The coherent absorber plays an essential, but very different, role in this estimation scheme; it effectively reverses the evolution in each step of the QMC when an initial estimate θ_0 is equal to the true value of the parameters θ. The environment of the QMC can then be modelled as a vacuum state, but introducing some displacement results in 'excitation' patterns in this vacuum state. The chapter develops the mathematical theory of these patterns and the culminates in an optimal estimator that utilizes the observed pattern counts.

PUBLICATIONS

Some ideas and figures have appeared previously in the following publications:

[1] Federico Girotti, Alfred Godley, and Mădălin Guţă. "Estimating quantum Markov chains using coherent absorber post-processing and pattern counting estimator". arxiv:2408.00626.

[2] Federico Girotti, Alfred Godley, and Mădălin Guţă. "Optimal estimation of pure states with displaced-null measurements". In: *Journal of Physics A: Mathematical and Theoretical* 57 (24 June 2024), p. 245304.

[3] Alfred Godley and Mădălin Guţă. "Adaptive measurement filter: efficient strategy for optimal estimation of quantum Markov chains". In: *Quantum* 7 (2023), p. 973.

ACKNOWLEDGMENTS

First and foremost, I would like to thank my supervisor Mădălin Guţă for his continuous guidance throughout this PhD. It has been a pleasure to explore and develop your field of research with you over the past 4 years. In addition, you've supported me through an arduous and tumultuous period of my life with even-tempered advice and a commendably stoic attitude. I will miss our daily trips to Portland, espresso cups in hand.

Next, I would like to thank my partner, Katherine Hart. An extended project like this would never be possible without the support and encouragement of those closest to you; you have provided that in spades. There were times when this project felt overwhelming, but your belief in me never wavered, even when mine did. I look forward to entering this next stage of our lives together and I'm excited for the opportunities it brings.

I would also like to thank my parents, David and Helen, and siblings, Freddie, Greg, Dan, Josie and Rebecca. You all go above and beyond to make me feel included and supported whenever we are together; the peace of mind that this brings can never be overstated.

I would be remiss if I did not mention my father's siblings - Alan, Helen, Jane and Kipper - who all live around Nottingham. Thank you for the many trips to the pub; it's been a pleasure living close to you all. In particular, I would like to commemorate my uncle Alan, who passed away suddenly this year: you will be fondly missed!

Finally, I'd like to thank the friends I've made during my time in Nottingham, particularly George Bakewell-Smith, Cameron Bunney, Federico Girotti and Josh New. Your warmth and generosity has provided me with some of the happiest years of my life. And I still owe you dinner, Federico!

CONTENTS

1 Introduction 1

I Preliminary Theory

2 Quantum Information 8
- 2.1 Basic Quantum Mechanics 8
  - 2.1.1 States, Observables and Measurements 9
  - 2.1.2 Multipartite Systems 14
  - 2.1.3 Continuous Variable Systems 17
- 2.2 Quantum Channels 21
  - 2.2.1 Representations 22
  - 2.2.2 Common Properties 25
- 2.3 Open Quantum Systems and Quantum Markov Chains 30
  - 2.3.1 Quantum Markov Chains 30

3 Parameter Estimation 34
- 3.1 Classical Estimation 34
  - 3.1.1 Local Asymptotic Normality 43
- 3.2 Quantum Estimation 46
  - 3.2.1 Quantum Local Asymptotic Normality 51

II Results

4 Adaptive measurement filter: efficient strategy for optimal estimation of quantum Markov chains 55
- 4.1 Introduction 55
- 4.2 Optimal separable measurements 58
- 4.3 Discrete quantum Markov chains and the output QFI 60
- 4.4 Output post-processing using quantum coherent absorber 62
- 4.5 Adaptive measurement algorithm 63
  - 4.5.1 Derivation of the measurement filter 65
- 4.6 Fisher informations considerations 68
  - 4.6.1 Achievability of the QFI with adaptive output measurements 69
  - 4.6.2 Computing the classical Fisher information of the output 69
- 4.7 Numerical simulations 71
  - 4.7.1 Simplified Markov model for the first numerical investigation 71
  - 4.7.2 Simulation studies for the simplified model 73
  - 4.7.3 The second numerical investigation using the full adaptive protocol 75
- 4.8 Conclusions and Outlook 78
- 4.9 Proof of Proposition 1 79
- 4.10 Computation of finite time system-output QFI 81

5 Optimal estimation of pure states with displaced-null measurements 83

5.1 Introduction and main results 83
5.2 Achievability of the QCRB for pure states 89
5.3 Why the naive imp. of a null measurement does not work 93
5.3.1 Parameter localisation via a two step adaptive procedure 98
5.4 Displaced-null est. scheme for opt. est. of pure qubit states 100
5.4.1 The displaced-null measurement for one parameter qubit models 100
5.5 Displaced-null measurements in the asympt. Gaussian picture 101
5.5.1 Brief review of local asymptotic normality for pure qubit states 101
5.5.2 Asymptotic perspective on displaced-null measurements via local asymptotic normality 104
5.6 Multiparameter estimation for pure qudit states 105
5.6.1 Multiparameter estimation 105
5.6.2 Gaussian shift models and QLAN 106
5.6.3 Achieving the Holevo bound for pure qudit states via QLAN 110
5.6.4 Achieving the Holevo bound with displaced-null measurements 113
5.6.5 Estimating a completely unknown pure state with respect to the Bures distance 115
5.6.6 Achieving the QCRB with displaced-null measurements 116
5.7 Conclusions and outlook 118
5.8 Proof of Theorem 16 for weaker notions 119
5.9 Proof of Proposition 3 121
5.10 Proof of Theorem 17 123
5.11 Comparison with estimators developed in [120] 125
5.12 Proof of Proposition 4 127
5.13 Proof of Proposition 5 129
6 Estimating quantum Markov chains using coherent absorber post-processing and pattern counting estimator 130
6.1 Introduction 130
6.2 Quantum est. and the displaced null measurement technique 134
6.3 QMCs and post-processing using coherent absorbers 136
6.3.1 Quantum postprocessing with a coherent absorber 138
6.4 Translationally invariant modes in the output 140
6.5 Limit distribution of quadratures and number operators 142
6.6 Limit theorem for counting trajectories 145
6.7 Pattern counting estimator 147
6.8 Numerical experiments 149
6.9 Conclusions and Outlook 152
6.10 Proofs of Proposition 8 and Corollary 2 154
6.11 Proofs of Theorem 19 and Lemma 8 156
6.12 Proof of Corollary 3 165
6.13 Proof of Theorem 20 and Proposition 9 166
6.14 Proof of Corollary 4 176
6.15 Achievability of the QCRB under additional assumptions 178

III Appendix
A Numerical simulations Python code 183

Bibliography 189

LIST OF FIGURES

Figure 1.1 A quantum Markov chain: a series of input systems prepared in state $|\chi\rangle$ interact with the SOI once, then are measured. 3

Figure 2.1 The Bloch representation of the $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ state: (Green) The corresponding Bloch vector $\boldsymbol{r} = (1,0,0)$, (Blue) The point corresponding to the state $|+\rangle$. 13

Figure 4.1 Quantum input-output discrete-time dynamics with θ dependent unitary interaction U_θ. 56

Figure 4.2 Adaptive measurement filter: the output units undergo post-processing with a coherent quantum absorber, followed by applying an adaptive measurement computed with the algorithm. 57

Figure 4.3 Fisher informations as function of output length n: quantum Fisher information (blue), classical Fisher information for the adaptive measurement with/without system measurement (orange/green), classical Fisher information for a regular (non-adaptive) measurement (red). 74

Figure 4.4 Adaptive measurement trajectory with basis angle φ on left and φ plotted on the circle on the right. 75

Figure 4.5 Comparison of different Fisher informations of the output state as function of trajectory length n, obtained by averaging over $N = 10000$ trajectories, at $\lambda = 0.8, \varphi = \pi/4$. The quantum Fisher information (QFI) (blue) and the classical Fisher information (CFI) of adapted measurement (green) completely overlap due to optimality. The inverse MLE error of the adapted measurement (orange) approaches the QFI for large n. Similarly, for the standard measurement, the inverse MLE error (red) approaches the CFI (purple) for large n. 76

Figure 4.6 Histogram of the MLE distribution in the adaptive measurement scenario, with $n = 200, \theta = 0$ and 10000 samples, at $\lambda = 0.8, \varphi = \pi/4$. 77

Figure 4.7 Comparison of different Fisher informations of the output state as function of trajectory length n, in the fully adaptive numerical investigation, for two proportions of samples used in the preliminary stage: $q = 0.15$ (top panel) and $q = 0.25$ (bottom panel). The results are obtained by averaging over $N = 1000$ trajectories, at $\theta = 0.2$ and $\lambda = 0.8, \varphi = \pi/4$. We plot: the asymptotic QFI (purple line), observed Fisher information for adaptive measurement (orange line) and simple measurement (red line), inverse MSE of the MLE for adaptive measurement (blue line) and simple measurement (green line). 77

Figure 5.1 The figure illustrates the non-identifiability problem occurring with null measurement (first row) and how it is fixed by displaced-null measurement (second row). In the first column the red arc on the xz Bloch sphere circle (in blue) represents the set of parameters after localisation (confidence interval), the green disk represents the true parameter value $\theta = \theta_+$ and the blue disk (panel a) is the parameter θ_- which is indistinguishable from the true one, in the null basis. The black arrow represents the chosen measurement basis. The second column displays a plot of the single count probability as a function of the parameter: in the null measurement case such a function is not injective on the set of parameters determined after the localisation (panel b). The third column shows the phase space of a Gaussian model consisting of coherent states with unknown displacement along the Q axis: the red interval is the parameter space, the black dot corresponds to the number operator measured, the green disk to the true coherent state and the blue disk (panel c) is the coherent state which is indistinguishable from the true one in the null measurement case. The last column plots the intensity of the number operator as a function of the coherent state amplitude. 86

Figure 5.2 For a reasonable estimator $\tilde{\theta}_n$, the posterior distribution of θ is centred around $\tilde{\theta}_n$, and has width of order $n^{-(1-\epsilon)/2}$. The assumption amounts to the fact that the posterior has non-vanishing mass on either side of $\tilde{\theta}_n$ at distance larger than $n^{-(1-\epsilon/2)/2}$ which is much smaller that the typical standard deviation. 96

Figure 6.1 Basic elements of the pattern counting estimator. Panel a) A quantum Markov chain as a system interacting sequentially with the environment via a parameter dependent unitary U_θ. The first stage estimator $\tilde{\theta}_n$ is obtained by performing a standard sequential measurement on the output and equating empirical and expected counts. Panel b) Post-processing the output using a coherent absorber. When system and absorber parameters match, the output is identical to the input (vacuum) Panel c) After the first estimation stage the absorber is fixed at a value $\theta_{\text{abs}} = \tilde{\theta}_n - \delta_n$ where $\tilde{\theta}_n$ is the preliminary estimator and δ_n is the parameter shift required by the displaced-null measurement theory [76]. The output generated by the system and absorber dynamics with unitary $V_{\theta_{\text{abs}}} U_\theta$ is measured sequentially in the standard basis. Panel d) Given a measurement trajectory, excitation patterns are identified as binary sequences starting and ending with a 1 separated by long sequences of 0s. The final estimator is a correction to the preliminary estimator which depends only on the total number of patterns $\sum_\alpha N_{\alpha,n}$, the QFI f at $\tilde{\theta}_n$ and the displacement parameter τ_n. 131

Figure 6.2 (Blue) Counts histograms for patterns $\alpha = 1, 11, 101, 111$ from N=2000 trajectories. (Orange line) The Poisson distribution with intensity given by $|\mu_\alpha|^2 u^2$ matches well the empirical counts distribution, as prescribed by Corollary 4. 150

Figure 6.3 (Blue) Histogram of the final estimator $\hat{\theta}_n$ from N=1000 trajectories with no first stage estimation ($\theta = \tilde{\theta}_n$). The effective Fisher information (inverse of rescaled estimated variance) $F_{\text{eff}} \approx 13.8$ matches closely the QFI $f_\theta = 13.5$. (Orange line) For comparison we plot the density of the normal distribution with mean $\bar{\theta} = 0.2$ and variance $\sigma^2 = (n f_\theta)^{-1}$. 151

Figure 6.4 (Blue) Histogram of final estimator $\hat{\theta}_n$ from N=1070 trajectories; the effective Fisher information is $F_{\text{eff}} = 10.8$ compared to the QFI $f_\theta = 13.5$. (Orange line) The density of the normal distribution with mean $\bar{\theta} = 0.2$ and variance $\sigma^2 = (n f_\theta)^{-1}$. 152

ACRONYMS

CA	coherent absorber
CV	continuous variable
CFI	classical Fisher information
QFI	quantum Fisher information
CLT	central limit theorem
QCLT	quantum central limit theorem
CRB	Cramér-Rao bound
QCRB	quantum Cramér-Rao bound
DNM	displaced-null measurement
HL	Heisenberg limit
IID	independent identically distributed
I-O	input-output
LAN	local asymptotic normality
QLAN	quantum local asymptotic normality
LOCC	local operations and classical communication
MLE	maximum likelihood estimator
MSE	mean square error
ONB	orthonormal basis
POVM	positive operator-valued measure
QMC	quantum Markov chain
RV	random variable
SOI	system of interest
SLD	symmetric logarithmic derivative
SQL	standard quantum limit
TIM	translationally invariant mode
CPTP	completely positive trace preserving

1

INTRODUCTION

In the last few decades, the field of quantum mechanics has moved beyond theoretical physics into practical applications, leading to the emergence of quantum technologies. These technologies promise to revolutionize a wide range of fields from computation to communication to sensing. Among these fields, quantum parameter estimation is worthy of significant attention, having already contributed to several scientific advancements from gravitational wave detection [1, 2] to high precision magnetometry [23, 122]. It focuses on the use of quantum systems to estimate unknown parameters with high precision, often surpassing classical limits. This process lies at the intersection of several exciting fields including quantum information theory, statistics, and metrology.

This is a very fundamental problem as high-precision measurement forms the cornerstone of many modern scientific disciplines. Moreover, industries are increasingly reliant on high-precision measurements to develop and build modern technology. As a third and final justification of its importance, it also provides a fruitful avenue into better understanding quantum mechanics; by pushing the limits of precision and exploiting quantum effects we can better understand quantum mechanics. This may enable the discovery of new phenomena that could then lead to advancements in both fundamental physics and applied quantum technologies.

The core objective in quantum parameter estimation [25, 93, 95, 130] is to identify optimal strategies that maximize the information extracted from a quantum state $\rho_{\boldsymbol{\theta}}$, which we assume is known up to an unknown parameter $\boldsymbol{\theta}$. The difficulty lies in the probabilistic nature of measurements in quantum mechanics: the outcome of a measurement is essentially a classical random variable with probability distribution specified by the system's state and the choice of measurement. Therefore, our primary objective is to optimise this choice of measurement. We then have a classical statistical inference problem, where we need to design an optimal estimator $\hat{\boldsymbol{\theta}}$ that processes these outcomes.

The optimality of the estimation scheme is usually assessed through the Cramér-Rao bound (CRB) [44], which provide a lower bound on the covariance of an unbiased estimator. For a one-dimensional parameter this is

$$\mathrm{Var}(\hat{\theta}) \geq I_{\theta}^{-1}$$

when we have one copy of the quantum state, where I is the classical Fisher information (CFI). The CFI is a property of the probability distribution associated with the choice of measurement, so it depends both on this choice and the quantum state. The quantum Cramér-Rao bound (QCRB) [13, 25, 93, 95] then replaces

the CFI with the quantum Fisher information (QFI), so for a one-dimensional parameter we have

$$\mathrm{Var}(\hat{\theta}) \geq F_\theta^{-1},$$

where F is the QFI. This is an optimisation over all choices of measurement, so it only depends on the quantum state. In quantum parameter estimation, the goal is often to maximise the CFI by selecting an appropriate measurement, aiming to make the CFI as close as possible to the QFI. With an optimal estimator, the inequalities mentioned earlier can be transformed into equalities, resulting in the variance being bounded by $\mathrm{Var}(\hat{\theta}) = F_\theta^{-1}$. And together, the combination of an optimal measurement choice and an optimal estimator defines an optimal estimation strategy. From a quantum information perspective, this also conveniently isolates the quantum aspects (measurement choice) from the classical aspects (optimal estimator).

The nature of the optimal measurements is of significant interest to us; the QCRB is formulated in terms of the symmetric logarithmic derivatives (SLDs), which provide optimal observables for the estimation of each parameter. They are, however, generally very complex operators that involve a collective measurement on all subsystems of a multipartite quantum system. This can make them difficult to implement, so we are interested in finding optimal schemes that use local operations [182], where we measure each subsystem individually.

The optimal measurements often depend on the parameters of interest; a consequence of this is that these bounds are often only saturable in an asymptotic scenario [73, 92, 170], where we have n copies of the quantum state and we take the limit as $n \to \infty$. We can then use a vanishingly small proportion of the total copies to localize the parameters to a small region around an initial estimate θ_0, and design strategies that are optimal in this region.

In multi-dimensional parameter scenarios [4, 5, 49, 69, 158], the quantum Cramér-Rao bound (QCRB) is not generally attainable even in an asymptotic framework due to incompatibility in the optimal measurements for each parameter component. Fortunately, there are other bounds that are asymptotically achievable in multiparameter scenarios; the most common of which is the Holevo bound. Therefore, our goal shifts to strategies that saturate these alternate bounds instead.

Quantum parameter estimation is a broad church, covering a wide range of applications. For instance, quantum sensing [47, 121, 186] alone encompasses all uses of a quantum system to measure a physical quantity. This is done by first preparing a system in a known initial state, then encoding the parameters into the system via a mechanism that is known up to the parameters of interest. For instance, this mechanism could be the system's evolution and we're interested in some parameters of the Hamiltonian like the coupling constant of an interaction term. We then design an estimation strategy to extract this information as prescribed in quantum parameter estimation.

Quantum sensing includes fields such as gravitational wave detection [1, 2, 11, 168], quantum magnetometry [23, 28, 101, 159], quantum thermometry [46, 123, 145] and quantum clocks [26, 131, 156]. From a quantum information perspective, this field is usually called quantum metrology [49, 72, 73, 151, 160] and the focus is on designing strategies that are optimal in the sense outlined above,

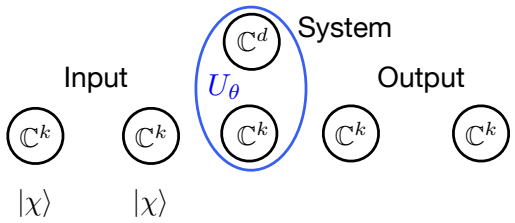

Figure 1.1: A quantum Markov chain: a series of input systems prepared in state $|\chi\rangle$ interact with the SOI once, then are measured.

whereas quantum sensing often focuses on strategies that are experimentally implementable.

In this thesis we develop optimal estimation schemes in two different settings: quantum Markov chains (QMCs) and independent identically distributed (IID) pure state models. The main setting of this thesis is QMCs [83, 84, 144] - a type of open quantum system involving a main system of interest (SOI) interacting with a series of initially uncorrelated environment states, cf. figure 1.1. Let $|\psi\rangle \in \mathbb{C}^d$ be the initial state of the main SOI, let $|\chi\rangle \in \mathbb{C}^k$ be the initial state of each environment state and let U be the unitary operation that describes the interaction. Then after n interactions the total state is

$$\rho^{\text{total}}(n) = U^{(n)} \dots U^{(1)} |\psi \otimes \chi^{\otimes n}\rangle \langle \psi \otimes \chi^{\otimes n}| U^{(1)*} \dots U^{(n)*},$$

where $U^{(i)}$ is the unitary applied to the SOI and ith environment state. If we fix an orthonormal basis (ONB) $\{|i\rangle : i = 1, \dots, k\}$ for the environment, then we can express the reduced state of the SOI as

$$\rho(n) = \sum_{i_1,\dots,i_n} K_{i_n} \dots K_{i_n} |\psi\rangle \langle\psi| K_{i_1}^* \dots K_{i_n}^* = T^n(|\psi\rangle \langle\psi|), \tag{1.1}$$

where $\{K_i = \langle i| U |\chi\rangle : i = 1, \dots k\}$ is a set of Kraus operators that satisfy normalisation condition $\sum_i K_i^* K_i = \mathbf{1}$, and $T(\cdot) = \sum_{i=1}^k K_i \cdot K_i^*$ is a quantum channel. The QMC consists of measuring the environment in this ONB, in which case, we obtain a sequence of outcomes $(i_1, \dots, i_n)$ with $i_j \in \{1, \dots, k\}$. Additionally, the state of the SOI jumps with each measurement, e.g.,

$$|0\rangle \langle 0| \to K_{i_1} |0\rangle |0\rangle K_{i_1}^* \to K_{i_2} K_{i_1} |0\rangle |0\rangle K_{i_1}^* K_{i_2}^* \to \dots,$$

where these states need to be properly normalised. Crucially, the probabilities of each possible jump $\rho \to K_i \rho K_i^*$ and the state after each jump depend only on the current state of the system.

To perform parameter estimation with a QMC, we simply let this unitary depend on some parameter of interest: $U = U_\theta$. Our aim is then to estimate this parameter using these measurements on the environment. This is a particularly exciting setting for parameter estimation as it allows a more fluid form of estimation: since we never measure the main system directly, we can simply keep measuring the environment to gain more information about the parameters. Additionally, this is a stark contrast to the standard parallel setting of quantum metrology, where we have to prepare n systems in advance. Instead, we simply prepare our system in some initial state, then repeatedly apply the quantum

channel whilst measuring the environment. Finally, this describes many common experimental setups such as a one-atom maser [91], quantum collision models [42] or a discretised optical cavity setups [10, 177].

Quantum channels [30, 33, 34, 175] provide one of the most general descriptions of quantum open systems and general quantum processes. We often utilize several properties of quantum channels [30, 54] in order to simplify the estimation problem. In particular, we often assume that the channel $T(\cdot)$ in equation (1.1) is primitive, which ensures that repeated applications of the quantum channel cause any initial state to converge to a fixed state ρ_* of maximal rank - the stationary state. We can then design our estimation protocol around this stationary regime. Since this stationary state is of maximal rank, this also ensures that the state's rank does not change at any point, which is often an important technicality.

In this thesis we investigate two alternate measurement strategies, both of which make use of an additional tool: a coherent absorber (CA). The CA [157] is another system that interacts with the environment after the main system of interest, where this interaction can be engineered. The total unitary in each step is then

$$W_\theta = V_{\theta_0} U_\theta,$$

where V_{θ_0} is the unitary for the interaction between an environment state and the CA, and $\theta_0 \in \Theta$ is a fixed value. In [78], we introduced the absorber in order to purify the system's stationary state at $\theta = \theta_0$, i.e.,

$$V_\theta U_\theta \left|\tilde{\psi}\right\rangle \otimes |\chi\rangle = \left|\tilde{\psi}\right\rangle \otimes |\chi\rangle ,$$

where $\left|\tilde{\psi}\right\rangle$ is the purified stationary state of the channel $T(\cdot)$. This ensures that in an asymptotic scenario we can guarantee that the stationary state is pure. However, the CA amounts to some post-processing of the state in general, so we can use this as a mechanism to alter the measurement implemented - a common feature in quantum information.

Our first optimal measurement scheme for a QMC adapts a local operations and classical communication (LOCC) measurement strategy developed in [182] to a QMC. This LOCC strategy utilized local, adaptive measurements on a pure multipartite quantum system to achieve the CRB. Our strategy then consists of an adaptive algorithm that determines each measurement inductively using the results of previous measurements. This involves tracking an object called the measurement filter through each estimation stage, which is updated after every measurement. It is then used to determine the next measurement basis. As mentioned in the previous paragraph, we introduced the CA to extend our measurement strategy from just pure state to mixed states (that we purify).

The second measurement setting, IID pure states, consists of n repeated copies of a pure state $|\psi_\theta\rangle$ from which we want to estimate the parameter $\theta \in \Theta \subseteq \mathbb{R}^k$. This simple setting was ideal for exploratory work involving displaced-null measurements (DNMs), motivated by the work of Yang et al. [177]. This switch away from QMCs allows us to develop the theory behind DNMs in a simplified setting. Regular null measurements consist of measuring in a basis that contains the system's state, i.e., for a pure state $|\psi_\theta\rangle$ we measure in a basis containing this state. This seems to be optimal as if we calculate the CFI of the state $|\psi_{\theta_0}\rangle$ in the limit $\theta_0 \to \theta$, then it is equal to the QFI. However, there is a procedural problem

with this: we have to specify how we generate θ_0 in the estimation scheme. And once we've specified how we select this initial estimate, the null measurement suffers from an identifiability problem locally around the initial estimate. This diminishes the actual corresponding CFI, so the regular null measurement fails to saturate the QCRB. Our second optimal estimation scheme is a solution to this: displaced-null measurements (DNMs). DNMs utilize the CA to implement a null measurement with $|\psi_{\theta_0+\tau}\rangle$, where τ is a suitable displacement. By carefully selecting this displacement we can eliminate the identification problem, allowing the estimation scheme to saturate the QCRB.

We then apply the displaced null measurement to a QMC. In this framework, the CA effectively reverses the evolution of the stationary state in each time step when $\theta_0 = \theta$. The output state of the environment, which we assume is 2-dimensional, is then in its stationary state. Therefore, we regard it as a "vacuum state"

$$|\Omega\rangle = |0\rangle^{\otimes n}$$

containing no information about the parameter. When we measure the environment in the standard basis $\{|0\rangle, |1\rangle\}$, a difference between θ_0 and θ causes "excitation patterns" in this vacuum state that start and end in a 1 (e.g.- $\alpha = 1$, 11, 101, etc.). In particular, we include some displacement τ to eliminate the identification problem found with a null measurement. We can calculate the expected rate of occurrence of each pattern, so we use this to design an optimal estimator based on the observed pattern counts.

This thesis consists of two parts. Part i functions as an introduction to the necessary topics of this thesis, so may be familiar to some readers. This starts with an introduction to the relevant quantum information theory in chapter 2, beginning with the basic topics of states, observables, operators, multipartite systems and continuous variable systems. It then introduces quantum channels in section 2.2, focusing on their mathematical theory and some important properties. Quantum channels are then related back to quantum open systems in section 2.3, with an introduction to quantum Markov chains included in section 2.3.1. Chapter 3 introduces classical and quantum estimation theory. This begins with the classical side in section 3.1, with a focus on asymptotic estimation theory and LAN. It then introduces their quantum analogues in section 3.2, culminating with QLAN in section 3.2.1.

With this background information out of the way, part ii presents the results of this PhD. Chapter 4 focuses on results from the paper [78]. This includes our first mention of a CA in section 4.4. The main result is presented in section 4.5, where we discuss the adaptive measurement scheme that we introduced above. The chapter then culminates with several numerical simulations in section 4.7, which demonstrate that our measurement scheme does indeed saturate the QCRB asymptotically.

In chapter 5, based on [77], we introduce the displaced null measurement scheme in the context of IID pure state models. We begin by presenting the necessary conditions to saturate the QCRB in Proposition 2. We then demonstrate why the regular null measurement fails to saturate the QCRB in section 5.3, culminating with Theorem 16. Section 5.4 introduces the displaced null measurement with Proposition 3 proving that this measurement scheme is optimal for one-

parameter qubit models. Section 5.5 explains the asymptotic perspective of the displaced null measurement through QLAN. Finally, section 5.6 generalizes all this to arbitrary d-dimensional qudits, demonstrating that the null measurement saturates the Holevo bound.

Finally, in chapter 6, based on [75], we apply the displaced null measurement scheme to a quantum Markov chain. To this end, in section 6.3 we discuss the role of the CA in this framework, mentioned above. In section 6.4 we formalize these excitation patterns in terms of translationally invariant modes (TIMs): essentially we can define "creation" $A^*_\alpha(n)$ and "annihilation" $A_\alpha(n)$ operators for the pattern, which map the vacuum state to a superposition of states with the corresponding pattern inserted in every available position, e.g., a state such as $A^*_1(n)\,|0\rangle^{\otimes n} = n^{-1/2}\sum_{i=1}^{n}|0\rangle^{\otimes i-1}\otimes|1\rangle\otimes|0\rangle^{n-i}$. Proposition 8 and Corollary 2 then demonstrate that these operators define the Fock basis of a continuous variable system in the asymptotic limit $n\to\infty$. The main result is Theorem 19, which relates the rate of occurrence of an excitation pattern back to its corresponding "quadratures"

$$Q_\alpha(n) = \frac{1}{\sqrt{2}}(A_\alpha(n) + A^*_\alpha(n)) \quad \text{and} \quad P_\alpha(n) = \frac{1}{\sqrt{2}i}(A_\alpha(n) - A^*_\alpha(n)).$$

We then utilize this in section 6.7 to design an optimal estimator based off counting the number of occurrences of each pattern in the output. In section 6.8 we then confirm that this estimator is optimal through numerical simulations.

Finally, appendix A discusses the Python code used to generate the numerical simulations, which can be found on GitHub at

`https://github.com/AGodley/ThesisCode`.

This mainly functions as a user's guide and an overview of the important files. There are several improvements that we would like to make to these files; mostly around restructuring the current code to increase usability. Therefore, we will include two folders: one as an archive of the current version of the code and one that we may edit in the future.

Part I

PRELIMINARY THEORY

In this part we introduce the necessary background quantum theory required to understand the main content of this thesis, which is found in part ii. Most of this content should be familiar to any reader with a background in quantum information, so it largely functions to introduce the notation used throughout this thesis. Towards the end of chapter 2, we also introduce quantum channels and quantum Markov chains in some depth - two topics that are considerably important throughout the rest of the thesis. In particular, we aim to elucidate some common properties of quantum channels that we utilize in part ii.

As this thesis focuses on quantum parameter estimation, it also has a large statistical element. Therefore, we spend considerable time introducing (classic/quantum) estimation theory in chapter 3. This focuses on defining the quantum estimation problem, where we want to estimate some parameters that are encoded into a quantum state. In particular, we highlight the conditions under which an estimation scheme can be considered optimal. These revolve around saturating fundamental bounds on the variance of an unbiased estimator. To this end, we also introduce several fundamental estimators and demonstrate how they achieve the relevant bounds. These sections ends with brief introductions to local asymptotic normality and its quantum equivalent. These are powerful results that aid in the design of optimal estimators.

2

QUANTUM INFORMATION

In this chapter we review the necessary concepts from quantum mechanics that are required to understand this thesis, starting with the basics in section 2.1. This section should be familiar to the reader and mostly functions as a space to introduce the notation used throughout this thesis. We begin with the common framework of quantum mechanics in sections 2.1.1 and 2.1.2, while highlighting aspects of particular importance to quantum information theory. We then discuss continuous variable systems in section 2.1.3 as they are necessary for chapter 3. Additionally, they describe the state of light in quantum optics [110]; a field that provides many of the most sophisticated experimental setups available. We then move onto quantum channels in section 2.2, which represent a general formalism for quantum open systems and noisy evolution. Essentially, a quantum channel can be regarded as a black box that maps valid states onto valid states in the Schrödinger picture, so they can be used to describe most processes in quantum mechanics. The section ends with a discussion of some common properties of quantum channels in section 2.2.2: *ergodicity*, *mixing*, *irreducibility*, and *primitivity*. We assume our quantum channels are primitive in chapters 5-7, so our primary aim for the chapter is to elucidate these properties. We then discuss quantum Markov chains in section 2.3, which are the consequence of measuring the environment of a discrete-time quantum open system. This results in a stochastic quantum trajectory, where the system's state jumps suddenly with each measurement of the environment. We then end the chapter by briefly discussing Lindbladian dynamics, which is the continuous time limit of a QMC.

2.1 BASIC QUANTUM MECHANICS

Quantum mechanics involves a mathematical framework that can be used to predict what is observed in physical situations. All possible predictions are calculated from the system's current state. In particular, the physical quantities that we can observe (energy, position, etc.) are described by mathematical operators known as observables; given a system's state and an observable, we can specify the possible outcomes of an experiment and their corresponding probabilities. These are the topics of this section, starting with states, observables and measurements.

2.1.1 *States, Observables and Measurements*

Without further ado, we introduce the basic objects of quantum mechanics [127, 146]. Associated with every isolated (closed) physical system is a complex Hilbert space $\mathcal{H}$ known as the *state space*. The system is then completely described by its *state*, which is a unit vector from the Hilbert space. We make use of Dirac notation, so a state is represented by a *ket*, say $|\psi\rangle \in \mathcal{H}$. We will typically restrict ourselves to finite-dimensional Hilbert spaces, $\mathcal{H} = \mathbb{C}^d$ for some finite $d < \infty$, so the theory will be presented in this context. However, the formalism can be extended to infinite-dimensional Hilbert spaces, which is relevant to section 2.1.3 where we discuss continuous variable systems.

Associated with every ket $|\phi\rangle$ is a *bra*, denoted $\langle\phi|$, which is a linear map that maps elements of the Hilbert space to complex numbers:

$$\begin{aligned} \langle\phi| : \mathcal{H} &\to \mathbb{C} \\ |\psi\rangle &\mapsto \langle\phi|\psi\rangle \end{aligned}$$

Therefore, $\langle\cdot|\cdot\rangle$ specifies an inner product which satisfies the following properties:

1. Conjugate symmetric: $\langle\phi|\psi\rangle = \overline{\langle\psi|\phi\rangle}$ for all $|\phi\rangle, |\psi\rangle \in \mathcal{H}$,
2. Linearity:
$$\langle\phi|a\psi_1 + b\psi_2\rangle = a\langle\phi|\psi_1\rangle + b\langle\phi|\psi_2\rangle$$
for all $a, b \in \mathbb{C}$ and $|\phi\rangle, |\psi_1\rangle, |\psi_2\rangle \in \mathcal{H}$
3. Positivity: $\langle\psi|\psi\rangle \geq 0$ for all $|\psi\rangle \in \mathcal{H}$,
4. Non-degeneracy: $\langle\psi|\psi\rangle = 0 \Leftrightarrow \psi = 0$,

where $\bar{\cdot}$ denotes complex conjugation.

Two vectors $|\psi\rangle, |\phi\rangle \in \mathcal{H}$ are orthogonal if $\langle\psi|\,\phi\rangle = 0$. An ONB for a Hilbert space $\mathcal{H} = \mathbb{C}^d$ is then a set of orthogonal unit vectors $\{|e_1\rangle, ..., |e_d\rangle \in \mathcal{H} : \langle e_i|e_j\rangle = \delta_{ij}, \langle e_i|e_i\rangle = 1\}$ that span the Hilbert space, i.e., we can express any state $|\psi\rangle \in \mathcal{H}$ as

$$|\psi\rangle = \sum_{i=1}^{d} a_i |e_i\rangle \tag{2.1}$$

for some $a_i \in \mathbb{C}$ that satisfy $\sum_{i=1}^{d} |a_i|^2 = 1$. Here we made use of the Kronecker delta δ_{ij}. The corresponding bra is then $\langle\psi| = \sum_{i=1}^{d} \overline{a_i} \langle e_i|$.

An example that is particularly relevant throughout quantum information is *qubits*, which represents the simplest non-trivial quantum system, and is the quantum equivalent of a classical bit taking binary values 0 or 1. Therefore, a qubit is a two-level system $\mathcal{H} = \mathbb{C}^2$ with standard ONB states $|0\rangle := \binom{1}{0}$ and $|1\rangle := \binom{0}{1}$. Since an arbitrary state is any unit vector from the Hilbert space, this allows for superpositions such as

$$|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix}.$$

Superpositions are closely related to entanglement [127], which is a very important concept in quantum information. Entanglement is introduced in section 2.1.2, where we introduce multipartite quantum systems.

The dual notion to a state is an *observable*, which is a self-adjoint (Hermitian) operator. An *operator* $A \in \mathcal{B}(\mathcal{H})$ is a bounded linear map on the Hilbert space

$$A : \mathcal{H} \to \mathcal{H},$$

and the *adjoint* of an operator $A \in \mathcal{B}(\mathcal{H})$ is another operator $A^* \in \mathcal{B}(\mathcal{H})$ such that

$$\langle \psi | A\phi \rangle = \langle A^*\psi | \phi \rangle$$

for all $|\psi\rangle, |\phi\rangle \in \mathcal{H}$. Therefore, an observable is an operator $A \in \mathcal{B}(\mathcal{H})$ that also satisfies $A = A^*$. Observables represent physical quantities (energy, position, etc.) that we can measure in an experiment. By fixing an ONB again, we can represent an operator as

$$A = \sum_{i,j=1}^{d} A_{ij} |e_i\rangle \langle e_j| ,$$

where $A_{ij} = \langle e_i | A | e_j \rangle$. The operator's adjoint is then the conjugate transpose of this: $A^* = \sum_{i,j=1}^{d} \bar{A}_{ji} |e_i\rangle \langle e_j|$.

Observables are the dual notion to states with respect to their expectation value:

$$\langle A \rangle_{|\psi\rangle} = \langle \psi | A | \psi \rangle = \mathrm{Tr}(A |\psi\rangle \langle \psi|)$$

Note: it is the operator $P_{|\psi\rangle} := |\psi\rangle \langle \psi|$ that is important here, not the state $|\psi\rangle$. This implies that states are only defined up to an arbitrary phase $e^{i\theta}$, $\theta \in \mathbb{R}$, since

$$P_{e^{i\theta}|\psi\rangle} = e^{i\theta} e^{-i\theta} |\psi\rangle \langle \psi| = P_{|\psi\rangle}.$$

Additionally, we will soon expand this notion to density matrices, which provide a more general description.

This operator $P_{|\psi\rangle}$ is an example of an orthogonal projection; it projects a vector $|\phi\rangle \in \mathcal{H}$ onto its overlap with the vector $|\psi\rangle \in \mathcal{H}$:

$$P_{|\psi\rangle} |\phi\rangle = \langle \psi | \phi \rangle \cdot |\psi\rangle .$$

A general orthogonal projection $P \in \mathcal{B}(\mathcal{H})$ is defined by the property $P^2 = P = P^*$, where this last equality demonstrates that it is clearly an observable. It can be expressed as $P = \sum_{i=1}^{r} |e_i\rangle \langle e_i|$ where $r \leq d$ and $\{|e_i\rangle : i = 1, ..., d\}$ forms an ONB for the Hilbert space, i.e., an orthogonal projection is diagonal with respect to an ONB. In particular, it projects a vector $|\psi\rangle$ onto the subspace spanned by a subset $\{|e_1\rangle, ..., |e_r\rangle\}$ of the full ONB:

$$P |\psi\rangle = \sum_{i=1}^{r} \sum_{j=1}^{d} |e_i\rangle \langle e_i| a_j |e_j\rangle = \sum_{i=1}^{r} a_i |e_i\rangle .$$

The number r specifies the dimension of the corresponding subspace, so the projection $P_{|\psi\rangle}$ was an example of a rank-one projection with only one term involving $|e_1\rangle = |\psi\rangle$. Two orthogonal projections P_i and P_j are mutually orthogonal if $P_i P_j = \delta_{ij} P_i$. In particular, a complete set of mutually orthogonal projections $\{P_i : i = 1, .., r\}$ for Hilbert space $\mathcal{H} = \mathbb{C}^d$ should satisfy $\sum_i P_i = \mathbf{1}$, where $\mathbf{1}$ is the identity operator, so that

$$\sum_i P_i |\psi\rangle = \mathbf{1} |\psi\rangle = |\psi\rangle ,$$

i.e., we can decompose a state into a sum of its projections $P_i |\psi\rangle$.

Any observable can be expressed in terms of orthogonal projections through the spectral decomposition:

Theorem 1 (Spectral decomposition)**.** *Let A be an observable on Hilbert space $\mathcal{H} = \mathbb{C}^d$. There exists a complete set of $r \leq d$ mutually orthogonal projections $\{P_i : i = 1, ..., r : \sum_{i=1}^{r} P_i = \mathbf{1}\}$ and real numbers $\{a_i \in \mathbb{R} : i = 1, .., r\}$ such that*

$$A = \sum_{i=1}^{r} a_i P_i. \tag{2.2}$$

The real numbers a_i are the eigenvalues of the observable. We expect d eigenvalues since the Hilbert space is d-dimensional, however, an eigenvalue may be degenerate, i.e., repeated. Therefore, the orthogonal projection P_i corresponds with the subspace spanned by a set of orthonormal vectors $\{|e_i\rangle\}$ that share the eigenvalue a_i. In total, we have d orthonormal eigenvectors as expected so we can equivalently write

$$A = \sum_{i=1}^{d} a_i |e_i\rangle \langle e_i| ,$$

for some ONB $\{|e_i\rangle : i = 1, ..., d\}$.

An important property of operators is *positivity*: an operator $A \in \mathcal{B}(\mathcal{H})$ is positive, denoted $A \geq 0$, if

$$\langle\psi| A |\psi\rangle \geq 0$$

for all states $|\psi\rangle \in \mathcal{H}$. Additionally, if $\langle\psi| A |\psi\rangle > 0$ we say A is strictly positive. Positivity implies that an operator is self-adjoint and has positive eigenvalues. Additionally, any positive operator $A \geq 0$ can be expressed as $A = B^* B$ for an operator $B \in \mathcal{B}(\mathcal{H})$.

We now have all the necessary elements to introduce *density operators*, which describe scenarios where there is some statistical uncertainty in the state preparation. For instance, consider a scenario where we prepare a system in state $|\psi_1\rangle \in \mathcal{H}$ with probability p or state $|\psi_2\rangle$ with probability $1 - p$. How would a third party model this situation if we do not tell them what state we prepared? The solution is to consider a mixture of the state's projections

$$\rho = p |\psi_1\rangle \langle\psi_1| + (1 - p) |\psi_2\rangle \langle\psi_2|$$

and this mixture is known as a density operator. We then replace the state's projection with this density operator, e.g., the expectation value of an observable $A \in \mathcal{B}(\mathcal{H})$ is $\langle A \rangle_\rho = \mathrm{Tr}(\rho A)$. A complete definition of density operators is the following:

Definition 1 (Density operators)**.** *A density operator $\rho \in \mathcal{B}(\mathcal{H})$ from a Hilbert space $\mathcal{H}$ is an operator that satisfies the following properties:*

1. *Normalization:* $\mathrm{Tr}(\rho) = 1$
2. *Positivity:* $\rho \geq 0$

We denote the space of density operators on the Hilbert space $\mathcal{H}$ as $S(\mathcal{H})$.

Each state $|\psi\rangle \in \mathcal{H}$ can be associated with a density operator $\rho \in S(\mathcal{H})$ via the rank-one projection $\rho = |\psi\rangle\langle\psi|$. However, not all density operators correspond with a rank-one projection: the space of density operators is convex, so a *mixture* of density operators $\rho_1, \rho_2 \in S(\mathcal{H})$ with form

$$\rho = p\rho_1 + (1-p)\rho_2, \quad 0 < p < 1,$$

is also a valid density operator. The extremal points of this convex set are precisely the rank-one projections, and their convex hull produces $S(\mathcal{H})$. The density operators that correspond with rank-one projections are known as the pure states as there is this direct association to a state and, as a projection, they satisfy the additional constraint $\rho^2 = \rho$.

The degree of mixedness can be quantified through the *purity* of a density operator $\rho \in S(\mathcal{H})$:

$$\gamma := \mathrm{Tr}(\rho^2).$$

This is a real number in the region $1/d \leq \gamma \leq 1$, where d is the dimension of the corresponding Hilbert space. For a pure state, we find that $\gamma = 1$ since $\mathrm{Tr}(\rho^2) = \mathrm{Tr}(\rho) = 1$. Density operators satisfying $\gamma = 1/d$ are known as maximally mixed. To see why, consider their spectral decomposition: we have

$$\rho = \sum_{i=1}^{d} p_i |e_i\rangle\langle e_i|,$$

for some $p_i \in \mathbb{R}$. Since $\mathrm{Tr}(\rho) = 1$, we know that $\sum_i p_i = 1$. Additionally, $p_i \geq 0$ since $\rho \geq 0$, so p_i represent valid probabilities. Minimising $\mathrm{Tr}(\rho^2) = \sum_{i=1}^{d} p_i^2$ under these constraints results in $p_i = 1/d$ as expected. Therefore, maximally mixed density operators correspond with an equal mixture of all states of an ONB.

Returning to the example of a qubit, states have a nice visual representation in terms of the Bloch sphere: the density matrix of any qubit can be expressed as

$$\rho = \frac{1}{2}(\mathbf{1} + \boldsymbol{r} \cdot \boldsymbol{\sigma}), \tag{2.3}$$

where $\boldsymbol{r} \in \mathbb{R}^3$ is the Bloch vector and $\boldsymbol{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$ is a vector of Pauli operators. This Bloch vector satisfies $|\boldsymbol{r}| \leq 1$ to ensure positivity. The state of a qubit can then be represented as the point given by its Bloch vector in a unit sphere - known as the Bloch sphere. Points on the surface of the sphere represent the pure states, whereas points within the sphere represent mixed states. This representation provides a very intuitive understanding of quantum qubit channels and operations as we can simply look at their effect on points from the Bloch sphere. An example point can be seen in figure 2.1.

We now discuss measurement in the context of measuring an observable; we will return to more general descriptions of measurement later. When we perform the measurement of an observable A with spectral decomposition as seen in equation (2.2) on a system in state ρ, the observed outcome is equal to one of its eigenvalues $a_i \in \mathbb{R}$, and occurs randomly with probability specified by Born's rule

$$p(i) = \mathrm{Tr}(\rho P_i),$$

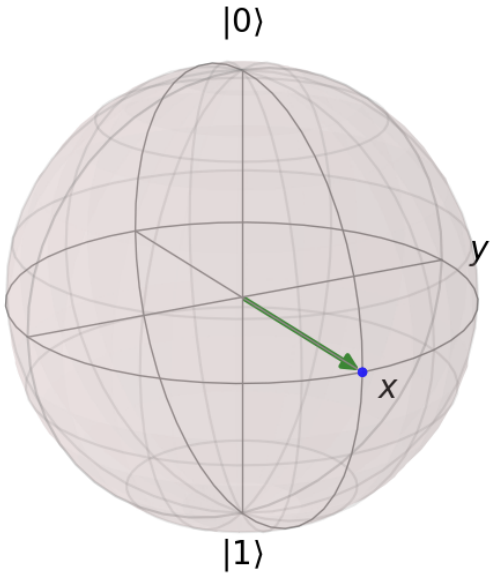

Figure 2.1: The Bloch representation of the $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ state: (Green) The corresponding Bloch vector $\boldsymbol{r} = (1, 0, 0)$, (Blue) The point corresponding to the state $|+\rangle$.

where P_i is the projection corresponding with eigenvalue a_i. Looking at the expectation value of an observable $A \in \mathcal{B}(\mathcal{H})$ again and using its spectral decomposition, we find

$$\langle A \rangle_\rho = \sum_{i=1}^{d} a_i \mathrm{Tr}(\rho P_i) = \sum_{i=1}^{d} a_i p(i)$$

so it is clearly the weighted average outcome.

The state directly after the measurement is

$$\rho' = \frac{P_i \rho P_i}{\mathrm{Tr}(\rho P_i)},$$

so the measurement projects the state onto the space spanned by the eigenvalue a_i that was observed. This is another significant diversion from classical mechanics; the act of measuring explicitly alters a system.

A consequence of this is that we cannot simultaneously measure two observables in general. Critically, since each measurement is disruptive, the order of measurements matters. For instance, consider measuring the Pauli operators on a qubit: if the qubit is initially in the σ_z eigenstate $|0\rangle$, then measuring σ_z results in observing outcome $+1$ with probability 1. Measuring σ_x then results in outcome $+1$ or -1 with equal probability. Meanwhile, reversing the measurement order results in either outcome $+1$ or -1 for the σ_x measurement. This measurement changes the system's state to an eigenvector of σ_x. The σ_z measurement then results in either outcome $+1$ or -1 with equal probability. Therefore, reversing the measurement order significantly altered the possible outcomes of the σ_z measurement. For two observables $A, B \in \mathcal{B}(\mathcal{H})$ to be simultaneously measurable, they must commute, i.e., $[A, B] := AB - BA = 0$. In which case, we can find a

set of eigenvectors that are simultaneous eigenvectors for each observable, i.e., $A|a,b\rangle = a|a,b\rangle$ and $B|a,b\rangle = b|a,b\rangle$.

In a physical setting, the time evolution of a closed quantum system is determined by solving the Schrödinger equation

$$\frac{\partial}{\partial t}|\psi(t)\rangle = -\frac{i}{\hbar}H(t)|\psi(t)\rangle,$$

where the *Hamiltonian* $H(t) \in \mathcal{B}(\mathcal{H})$ is a Hermitian operator related to the energy of the system. From a theoretical point of view this solution is specified by a unitary operator: $U \in \mathcal{B}(\mathcal{H})$ such that $U^*U = UU^* = \mathbf{1}$. For example, when H is fixed the solution to the Schrödinger equation is

$$|\psi(t)\rangle = U(t)|\psi(0)\rangle$$

where $U(t) = \exp(-\frac{i}{\hbar}Ht)$. From a mathematical point of view the objects of interest are these unitary operations, so we can avoid worrying about the underlying dynamics. Instead, we focus on the effects of unitary operations applied to quantum states.

Now that we've discussed the basics of states, measurement and dynamics, we move onto multipartite quantum systems.

2.1.2 *Multipartite Systems*

A system consisting of multiple subsystems is described using the tensor product $\otimes$ of their respective Hilbert spaces [127, 175]. The tensor product of two Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$ combines them into a larger Hilbert space $\mathcal{H}_{12} = \mathcal{H}_1 \otimes \mathcal{H}_2$. Let $\{|e_i\rangle : i = 1,...,d_1\}$ and $\{|f_j\rangle : i = 1,...,d_2\}$ be orthonormal bases for Hilbert spaces $\mathcal{H}_1 = \mathbb{C}^{d_1}$ and $\mathcal{H}_2 = \mathbb{C}^{d_2}$ respectively. Then $\mathcal{H}_{12}$ is the Hilbert space spanned by the ONB $\{|e_i\rangle \otimes |f_j\rangle : i = 1,..,d_1; j = 1,...,d_2\}$. This ONB has d_1d_2 elements, so the corresponding Hilbert space $\mathcal{H}_{12}$ clearly has dimension $\dim(\mathcal{H}_{12}) = d_1d_2$. Also, note: we often use the notation $|\psi \otimes \phi\rangle := |\psi\rangle \otimes |\phi\rangle$ to simplify expressions.

The inner product between two vectors $|\psi_1 \otimes \psi_2\rangle, |\phi_1 \otimes \phi_2\rangle \in \mathcal{H}_{12}$ is given by the inner products for each subspace multiplied together:

$$\langle \psi_1 \otimes \psi_2 | \phi_1 \otimes \phi_2 \rangle = \langle \psi_1 | \phi_1 \rangle \langle \psi_2 | \phi_2 \rangle.$$

Additionally, operators on a tensor product space can be expressed as linear combinations of tensor products of operators on the individual subsystems. If $A \in \mathcal{B}(\mathcal{H}_1)$ and $B \in \mathcal{B}(\mathcal{H}_2)$ then the tensor product $A \otimes B \in \mathcal{B}(\mathcal{H}_1 \otimes \mathcal{H}_2)$ has the following action on tensor product vectors:

$$A \otimes B|\psi\rangle \otimes |\phi\rangle = A|\psi\rangle \otimes B|\phi\rangle$$

By specifying both an operator and a state as a sums of product operators/states we can then apply generic operators to generic states.

Crucially, the existence of superpositions in a Hilbert space results in states that cannot be expressed as a simple tensor product, i.e., $|\Psi\rangle \neq |\psi\rangle \otimes |\phi\rangle$ for $|\psi\rangle \in \mathcal{H}_1, |\phi\rangle \in \mathcal{H}_2$. These states are called entangled states, whereas states

that can be expressed as a direct tensor product are known as product states. Entangled states exhibit correlations between the subsystems, which represent a uniquely quantum resource. In the next few paragraphs we demonstrate what we mean by correlations through a qubit-based example.

The standard basis in the space $\mathcal{H} = \mathbb{C}^2 \otimes \mathbb{C}^2$ of two qubits is

$$\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\},$$

where we use a different notation for simplicity, removing the tensor product symbol entirely. An example of an entangled state is $|\Phi^+\rangle := (|00\rangle + |11\rangle)/\sqrt{2}$, which is one of the four maximally entangled Bell states that also form an ONB for $\mathbb{C}^2 \otimes \mathbb{C}^2$

$$\begin{aligned}
\{|\Phi^+\rangle &:= \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle), \\
|\Phi^-\rangle &:= \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle), \\
|\Psi^+\rangle &:= \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle), \\
|\Psi^-\rangle &:= \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)\}
\end{aligned}$$

A characteristic feature of maximally entangled states is that they produce *perfectly correlated* outcomes when certain observables are measured on each qubit. For example, consider measuring the observable $\sigma_i \otimes \sigma_i$ with respect to the singlet state $|\Psi^-\rangle$, where σ_i is a Pauli operator. The structure of the state ensures that one always obtains opposite outcomes. This means that one can infer the outcome of the measurement on one side without performing the corresponding measurement, as long as the other side has been measured. This surprising fact is at the root of the famous EPR paradox [52] and the modern understanding of quantum non-locality as expressed by the Bell theory [9, 17].

An important result about bipartite states is the Schmidt decomposition.

Theorem 2 (Schmidt decomposition)**.** *Let $|\psi\rangle$ be a pure state on the Hilbert space $\mathcal{H}_{12} = \mathcal{H}_1 \otimes \mathcal{H}_2$. Then there exists an ONB $\{|e_i\rangle\}$ for $\mathcal{H}_1$ and an ONB $\{|f_i\rangle\}$ for $\mathcal{H}_2$ such that*

$$|\psi\rangle = \sum_{i=1}^{d} \sqrt{a_i}\, |e_i\rangle \otimes |f_i\rangle,$$

where $d = min(dim(\mathcal{H}_1), dim(\mathcal{H}_2))$ and a_i are non-negative real numbers satisfying $\sum_{i=1}^{d} a_i = 1$. These numbers are known as the Schmidt coefficients.

Note that while any vector can be decomposed in a product basis as

$$|\Psi\rangle = \sum_{ij} c_{ij}\, |e_i\rangle \otimes |f_j\rangle$$

the sum in the Schmidt decomposition runs over a single index and the Schmidt vectors $|e_i\rangle, |f_i\rangle$ are determined by the state $|\Psi\rangle$.

If we want to consider the state of one of the subsystems, we use the *partial trace*. For instance, let $\rho \in S(\mathcal{H}_{12})$, then the reduced state of the first subsystem is

$$\rho_1 = \mathrm{Tr}_2(\rho) = \sum_i \langle g_i^{(2)}| \rho |g_i^{(2)}\rangle , \tag{2.4}$$

where $\{|g_i^{(2)}\rangle\}$ is an ONB for $\mathcal{H}_2$. Similarly, the reduced state of the other subsystem is

$$\rho_2 = \mathrm{Tr}_1(\rho) = \sum_i \langle g_i^{(1)}| \rho |g_i^{(1)}\rangle , \tag{2.5}$$

where $\{|g_i^{(1)}\rangle\}$ is an ONB for $\mathcal{H}_1$. Additionally, this can be generalized to an arbitrary number of subsystems. The outcomes of measuring an observable on a subsystem can then be calculated using the subsystem's reduced state in the normal fashion. For instance, the expected value of an observable $A \in \mathcal{B}(\mathcal{H}_1)$ is $\langle A\rangle_{\rho_1} = \mathrm{Tr}(\rho_1 A)$.

With this in mind, we can relate the Schmidt decomposition back to the reduced states: for $\rho = |\psi\rangle\langle\psi|$ with $|\psi\rangle$ as defined in Theorem 2, we find

$$\rho_1 = \sum_i a_i |e_i\rangle\langle e_i|$$

and

$$\rho_2 = \sum_i a_i |f_i\rangle\langle f_i| .$$

Therefore, the Schmidt vectors $\{|e_i\rangle\}$ and $\{|f_i\rangle\}$ are the eigenvectors for the partial states and the eigenvalues are the corresponding Schmidt coefficients $\{a_i\}$. Overall, each reduced state is in its spectral decomposition.

A density operator $\rho \in S(\mathcal{H}_{12})$ is a product operator if

$$\rho = \rho_1 \otimes \rho_2$$

for (possibly mixed) density operators $\rho_1 \in S(\mathcal{H}_1), \rho_2 \in S(\mathcal{H}_2)$. A separable density operator can be expressed as a sum of product operators, i.e.,

$$\rho = \sum_i p_i \rho_i^{(1)} \otimes \rho_i^{(2)}$$

for $\rho_i^{(1)} \in S(\mathcal{H}_1), \rho_i^{(2)} \in S(\mathcal{H}_2), p_i \in \mathbb{R}$ such that $\sum_i p_i = 1$. And, finally, a density operator is entangled if it cannot be expressed as a convex combination in this manner.

For a pure state $|\Psi\rangle \in \mathcal{H}_{12}$ there is a single meaningful entanglement measure: the entropy of entanglement [139]:

$$\mathrm{Ent}(|\Psi\rangle\langle\Psi|) := S(\rho_1) = S(\rho_2),$$

where $S(\rho) := -\mathrm{Tr}(\rho\log(\rho))$ is the von-Neumann entropy, and ρ_1, ρ_2 are the reduced states of $|\Psi\rangle\langle\Psi|$. Using the Schmidt decomposition, we can see that

$$S(\rho) = -\sum_i a_i \log(a_i).$$

For mixed states there is a broad range of measures [139], none of which are clearly superior to the others. There is, however, some very rich theory surrounding them.

When applying an operator $A \in S(\mathcal{H}_{12})$ that is not a tensor product of two operators, $A \neq A_1 \otimes A_2$, to a product vector, one may obtain an entangled one. An example of this is the two qubits controlled-NOT (CNOT) gate:

$$\text{CNOT} = |0\rangle\langle 0| \otimes \mathbf{1} + |1\rangle\langle 1| \otimes \sigma_x,$$

which flips the second qubit (through application of σ_x) if the first qubit is in the $|1\rangle$ state. To see that this can create entanglement, consider its application to the product state $|+\rangle \otimes |0\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \otimes |0\rangle$:

$$\text{CNOT}\,|+\rangle \otimes |0\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle),$$

which is precisely one of the maximally entangled Bell state that we introduced earlier!

The final concept about bipartite quantum systems that we wish to present is *purification*: Let $\rho \in S(\mathcal{H})$ be a density operator with spectral decomposition

$$\rho = \sum_i p_i |i\rangle\langle i|.$$

We can define a pure state

$$|\psi\rangle := \sum_i \sqrt{p_i}\,|i\rangle \otimes |i\rangle$$

on the extended Hilbert space $\mathcal{H} \otimes \mathcal{H}$ such that

$$\rho = \text{Tr}_2(|\psi\rangle\langle\psi|).$$

This state is the purification of the mixed state ρ. So purification tells us that we can imagine any mixed state as a pure state on an extended space. This highlights the rich connection between quantum open systems and mixed states: we can imagine the universe as a closed system in which we get nice unitary dynamics involving pure states, but, in any physical experiment, we deal with a small part of this closed system. This results in open dynamics with mixed states instead.

2.1.3 *Continuous Variable Systems*

In this section we outline the fundamentals of continuous variable (CV) quantum systems [24, 110, 146]; the simplest of which is the simple harmonic oscillator. These systems are particularly important in quantum mechanics as many physical systems can be modelled as simple harmonics oscillators around their minimum energy. Additionally, the state of light as described in quantum optics is expressed in the same framework. Finally, they are particularly relevant to this thesis for two reasons:

1. The quantum Markov chain, introduced in section 2.3.1, can be seen as a discrete time version of a quantum optical system.

2. The quantum local asymptotic normality (Quantum local asymptotic normality (QLAN)) theory introduced in section 3.2.1 shows that n copies of a state $|\psi\rangle^{\otimes n}$ can be mapped into a continuous variable state in a statistical sense. This theory is utilized in chapters 5 and 6.

This motivates their introduction.

The Hilbert space for a one-mode CV system is given by $\mathcal{H} = L^2(\mathbb{R})$, where $L^2(\mathbb{R})$ is the space of square-integrable functions. The one-mode system can then be characterised in terms of two observables $Q, P \in \mathcal{B}(\mathcal{H})$, known as the *canonical coordinates*, which satisfy the commutation relation $[Q, P] = i$. Physically, these represent the position and momentum of a particle or an electromagnetic field, where we have rescaled the coordinates so that they are dimensionless for simplicity. An element from the Hilbert space then specifies a wavefunction $\psi(q) \in L^2(\mathbb{R})$ and these operators can be defined by their action on a wavefunction:

$$Q\psi(q) = q\psi(q), \quad P\psi(q) = -i\frac{d\psi}{dq}(q).$$

The probability of finding the system in some region $q_1 \leq q \leq q_2$ is then given by

$$\mathbb{P}_\psi^Q(q_1 \leq q \leq q_2) = \int_{q_1}^{q_2} |\psi(q)|^2 \, dq.$$

To make a similar statement about P we would have to consider the Fourier transform of the wavefunction $\tilde{\psi}(p) \in \mathcal{H}$ which swaps the roles of Q and P.

The canonical coordinates can be expressed in terms of the *creation* ($a^* \in \mathcal{B}(\mathcal{H})$) and *annihilation* ($a \in \mathcal{B}(\mathcal{H})$) operators as

$$Q = \frac{1}{\sqrt{2}}(a + a^*), \quad P = \frac{-i}{\sqrt{2}}(a - a^*), \tag{2.6}$$

which satisfy the commutation relation $[a, a^*] = \mathbf{1}$. These operators generate an ONB known as the Fock basis via the relations

$$a\,|n\rangle = \sqrt{n}\,|n-1\rangle\,, \quad a^*\,|n\rangle = \sqrt{n+1}\,|n+1\rangle\,,$$

where $n \in \mathbb{N}$ and we identify $|0\rangle$ with the ground state of the system. In quantum optics a Fock state $|n\rangle$ can be identified as the state with n photons (quanta of energy), so the creation/annihilation operators effectively add/remove a photon respectively. For general CV systems they add/remove *excitations* instead. The wavefunction corresponding to a Fock state $|n\rangle$ is

$$\psi_n(q) = \frac{H_n(q)e^{-q^2/2}}{(\sqrt{\pi}2^m m!)^{1/2}},$$

where H_n are the Hermite polynomials.

Note that the creation and annihilation operators are not observables, so cannot be measured. However, the *number operator* $N := a^*a$ is self-adjoint by construction and so can be directly measured. Additionally, its eigenvectors are the Fock states $N|n\rangle = n|n\rangle$. Given a state $|\psi\rangle = \sum_{n=0}^{\infty} c_n\,|n\rangle$ with $\sum_{n=0}^{\infty} |c_n|^2 = 1$, the probability of obtaining the outcome n (counting n photons in quantum optics) is $p(n) = |c_n|^2$ and the expectation value of the number operator is the average number of excitations in the system: $\langle N\rangle = \sum_{n=0}^{\infty} |c_n|^2 n$.

This defines the pure states of a CV system. We can then define density operators $\rho \in S(\mathcal{H})$ in a similar manner to in section 2.1.1 by considering statistical mixtures of pure density operators.

The most important states in quantum optics are the *coherent states*, which are the eigenstates of the annihilation operator:

$$a\,|\alpha\rangle = \alpha\,|\alpha\rangle\,, \qquad \alpha \in \mathbb{C}$$

and can be generated by applying the displacement operator $D(\alpha) := \exp(\alpha a^* - \bar{\alpha} a)$ to the vacuum state, i.e., $|\alpha\rangle = D(\alpha)\,|0\rangle$. In terms of the Fock basis, they can be expressed as

$$|\alpha\rangle = \exp(-\frac{|\alpha|^2}{2}) \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}}\,|n\rangle\,.$$

When we measure the Number operator with respect to a coherent state, we find the following:

1. The outcome n has distribution $\mathbb{P}_\alpha(n) = |\,\langle n|\alpha\rangle\,|^2 = e^{-|\alpha|^2}\frac{|\alpha|^2}{n!}$.
2. The expected value is $\langle N\rangle_\alpha = |\alpha|^2$.
3. The variance is equal to the expected value: $\mathrm{Var}(N) = |\alpha|^2$.

These are all the hallmarks of a Poisson distribution with intensity $|\alpha|^2$, so, in summary, photon counting with respect to a coherent state produces a Poisson distributed number of photons.

To see why these states are useful, consider measuring the canonical coordinates Q, P with respect to a coherent state. The expectation values of each coordinate are

$$\langle Q\rangle_{|\alpha\rangle} = \frac{(\alpha + \bar{\alpha})}{\sqrt{2}} = \sqrt{2}\mathrm{Re}(\alpha) \tag{2.7}$$

$$\langle P\rangle_{|\alpha\rangle} = -i\frac{(\alpha - \bar{\alpha})}{\sqrt{2}} = \sqrt{2}\mathrm{Im}(\alpha), \tag{2.8}$$

where $\mathrm{Re}(\cdot)$ and $\mathrm{Im}(\cdot)$ denote the real and imaginary parts. Therefore, measuring a canonical coordinate isolates the real or imaginary part of α. Additionally,

$$\begin{aligned}
\langle Q^2\rangle_{|\alpha\rangle} &= \frac{1}{2}\langle (a + a^*)^2\rangle_{|\alpha\rangle} \\
&= \frac{1}{2}\langle a^2 + (a^*)^2 + \mathbf{1} + 2a^*a\rangle_{|\alpha\rangle} \\
&= \frac{1}{2}(\alpha^2 + (\alpha)^2 + 1 + |\alpha|^2) \\
&= \frac{1}{2}(1 + 4\mathrm{Re}(\alpha)^2), \\
\langle P^2\rangle_{|\alpha\rangle} &= \frac{1}{2}(1 + 4\mathrm{Im}(\alpha^2)),
\end{aligned}$$

where we use the commutation relation $[a, a^*] = \mathbf{1}$ on line 2 and the calculation for $\langle P^2\rangle_{|\alpha\rangle}$ is similar to the first calculation. Therefore, the corresponding variances are

$$\mathrm{Var}_{|\alpha\rangle}(Q) = \langle Q^2\rangle_{|\alpha\rangle} - \langle Q\rangle^2_{|\alpha\rangle} = \frac{1}{2}, \tag{2.9}$$

$$\mathrm{Var}_{|\alpha\rangle}(P) = \langle P^2\rangle_{|\alpha\rangle} - \langle P\rangle^2_{|\alpha\rangle} = \frac{1}{2}. \tag{2.10}$$

The minimum variance of two observables A and B with respect to a state $|\psi\rangle$ is bounded by the Heisenberg uncertainty principle:

$$\mathrm{Var}_{|\alpha\rangle}(A)\mathrm{Var}_{|\alpha\rangle}(B) \geq \frac{1}{4}|\langle[A,B]\rangle_{|\psi\rangle}|^2.$$

In this case, $\mathrm{Var}_{|\alpha\rangle}(Q)\mathrm{Var}_{|\alpha\rangle}(P) \geq \frac{1}{4}$ when we substitute in their commutation relation. In fact, from the calculations above we know that this product saturates this inequality. Therefore, coherent states represent states with minimal uncertainty.

A convenient representation of CV states is in terms of their *Wigner function* which plays the role of a joint probability distribution of Q and P. Since these observables do not commute, they cannot be simultaneously measured. Therefore, the Wigner function is not a bona fide joint probability distribution. Indeed, it can take negative values in places, but shares other properties with a joint probability distribution. In order to define the Wigner function, it is easier to first define the characteristic function of a density operator $\rho \in S(L^2(\mathbb{R}))$:

$$\tilde{W}_\rho(u,v) \coloneqq \mathrm{Tr}(\rho \exp(-iuQ - ivP)).$$

This is the quantum Fourier transform of the density operator. The Wigner function of a state $\rho \in S(\mathcal{H})$ is then defined as the inverse Fourier transform of this:

$$W_\rho(q,p) \coloneqq \frac{1}{(2\pi)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\tilde{W}_\rho(u,v)\exp(iuq+ivp)dudv. \tag{2.11}$$

A property that this shares with a joint probability density is we can recover the marginal distributions for Q and P by integrating over the possible values of the other operator. For instance,

$$\langle q|\rho|q\rangle = \int_{-\infty}^{\infty} W_\rho(q,p)dp.$$

To define Gaussian states, we typically group the canonical coordinates together as $\boldsymbol{R} = (Q,P)$ so that we can express the covariance matrix as

$$\sigma_{kl}(\rho) = \frac{1}{2}\langle\{R_k,R_l\}\rangle_\rho - \langle R_k\rangle_\rho\langle R_l\rangle_\rho.$$

The uncertainty principle can then be encoded in the condition

$$\det(\sigma) \geq \frac{1}{4},$$

where $\det(\cdot)$ denotes the determinant. Additionally, we define the expectation value of $\boldsymbol{R}$ as $\langle\boldsymbol{R}\rangle_\rho = (\langle Q\rangle_\rho, \langle P\rangle_\rho)$. A state ρ is then *Gaussian* if its Wigner function is a Gaussian probability density, i.e., it is of the form

$$W_\rho(q,p) = \frac{1}{2\pi\sqrt{\det(\sigma(\rho))}}e^{-\frac{1}{2}(\boldsymbol{r}-\langle\boldsymbol{R}\rangle_\rho)^T\sigma^{-1}(\rho)(\boldsymbol{r}-\langle\boldsymbol{R}\rangle_\rho)}, \quad \boldsymbol{r}=(q,p).$$

Gaussian states include coherent states, but also squeezed states and thermal states. Squeezed states still saturate the uncertainty principle with $\det(\sigma) = \frac{1}{4}$, but they 'squeeze' the two variances $\mathrm{Var}(Q)$ and $\mathrm{Var}(P)$. This lowers one at

the cost of raising the other, so they are often more useful when we are more interested in one of the canonical coordinates over the other. Thermal states are generally mixed states that do not saturate the uncertainty relation, but their Wigner function still has a Gaussian shape around their mean.

If we want to consider a m-mode continuous variable system, then it generalises quite naturally: the system is described by canonical coordinates $Q_1, ..., Q_m, P_1, ..., P_m \in \mathcal{B}(\mathcal{H})$ that satisfy the commutation relations

$$[Q_i, P_j] = i\delta_{ij}, \quad [Q_i, Q_j] = [P_i, P_j] = 0.$$

Each mode has corresponding creation/annihilation operators satisfying

$$[a_i, a_j^*] = \delta_{ij}, \quad [a_i, a_j] = [a_i^*, a_j^*] = 0$$

that generates a Fock basis $\{|\boldsymbol{n}\rangle : \boldsymbol{n} = (n_1, ..., n_m) \in \mathbb{Z}^m\}$ via

$$\begin{aligned} a_i |\boldsymbol{n}\rangle &= \sqrt{n_i}\, |(n_1, ..., n_i + 1, ..., n_m))\rangle\,, \\ a_i^* |\boldsymbol{n}\rangle &= \sqrt{n_i + 1}\, |(n_1, ..., n_i + 1, ..., n_m)\rangle\,. \end{aligned}$$

Finally, we can define a m-mode coherent state $|\boldsymbol{\alpha}\rangle$ such that

$$a_i |\boldsymbol{\alpha}\rangle = \alpha_i |\boldsymbol{\alpha}\rangle\,,$$

which we can express as a tensor product of one-mode coherent states

$$|\boldsymbol{\alpha}\rangle = |\alpha_1\rangle \otimes ... \otimes |\alpha_m\rangle$$

to recover an expression in terms of the Fock basis of each mode.

A n-mode Gaussian state generalizes simply from the one-mode example: grouping together Qs and Ps as $\boldsymbol{R} = (Q_1, P_1, ..., Q_n, P_n)$, we can write

$$W(\boldsymbol{r}) = \frac{1}{(2\pi)^n \sqrt{\det(\boldsymbol{\sigma})}} e^{-\frac{1}{2}(\boldsymbol{r} - \langle \boldsymbol{R} \rangle_\rho)^T \boldsymbol{\sigma}^{-1}(\rho)(\boldsymbol{r} - \langle \boldsymbol{R} \rangle_\rho)}$$

with the constraint on the covariance $\det(\sigma) \geq (1/4)^n$

2.2 QUANTUM CHANNELS

In this section we introduce quantum channels [30, 54, 94, 103, 127, 175], highlighting the properties that are particularly relevant to this thesis. The first property that we need to discuss is *positivity*: a linear map $T : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H})$ is positive if $T(A) \geq 0$ for all $A \geq 0$.

With this in mind, we can proceed with the definition of a quantum channel:

Definition 2 (Quantum channels)**.** *A quantum channel is a linear map $T : \mathcal{B}(\mathcal{H}_1) \to \mathcal{B}(\mathcal{H}_2)$ that satisfies the following two conditions:*

1. *Trace preservation:* $\mathrm{Tr}(T(\rho)) = \mathrm{Tr}(\rho) = 1$ *for all* $\rho \in S(\mathcal{H}_1)$.
2. *Complete positivity: $T \otimes id_n$ is positive for all $n \in \mathbb{N}$, where id_n is the n-dimensional identity channel.*

A linear map that satisfies both these conditions is called completely positive trace preserving (CPTP).

We will assume that $\mathcal{H} = \mathcal{H}_1 = \mathcal{H}_2$ for simplicity. Additionally, We will denote the space of all quantum channels on Hilbert space $\mathcal{H}$ by $\mathcal{CPTP}(\mathcal{H})$.

Trace preservation and positivity ensure that if $\rho \in S(\mathcal{H})$ then $T(\rho) \in S(\mathcal{H})$. That is, it maps valid density operators onto valid density operators. However, it turns out that the map $T \in \mathcal{CPTP}(\mathcal{H}_1)$ is physical only if the stronger complete positivity condition holds. Indeed consider applying the channel T to the left side of the bipartite system $\mathcal{H} = \mathcal{H}_1 \otimes \mathbb{C}^n$, while no action is taken on the right side. Then the joint transformation is given by $T \otimes id_n$ which needs to be positive as well on physical grounds. In section 2.2.1 we discuss the Choi-Jamiolkowski correspondence which provides a simple way to check whether a map is a channel.

The standard example of a linear map that is positive but not completely positive is the qubit transposition with respect to the standard basis.

$$T(\cdot) = \sum_{i,j=0}^{1} \langle j| \cdot |i\rangle \, |i\rangle \langle j| \, .$$

One can verify that if A is positive then T(A) is also positive as is has the same eigenvalues as A.

However, consider applying the transpose to the first qubit of the Bell state $|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$. We then have

$$\rho = |\psi\rangle \langle\psi| = \frac{1}{2} \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} \rightarrow T \otimes \mathrm{id}_2(\rho) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

If we calculate the eigenvalues of this second matrix, we find one negative eigenvalue -1. Therefore, $T \otimes \mathrm{id}$ cannot be positive and so T is not completely positive.

With this initial definition out of the way, we move onto discussing the common representations of quantum channels.

2.2.1 *Representations*

The most common representation of a quantum channel is the well known *Kraus representation*:

Theorem 3 (Kraus representation)**.** *Let $T \in \mathcal{CPTP}(\mathcal{H})$ be a quantum channel. For some $m \leq d^2$, we can express T as*

$$T(\cdot) = \sum_{i=1}^{m} K_i \cdot K_i^*, \tag{2.12}$$

where the Kraus operators $\{K_i \in \mathcal{B}(\mathcal{H}) : i = 1, ..., m\}$ satisfy the condition $\sum_{i=1}^{m} K_i^* K_i = \mathbf{1}$.

The structure of the Kraus representation ensures complete positivity, while the condition on the Kraus operators ensures normalization as $\mathrm{Tr}(T(\rho)) = \sum_{i=1}^{m} \mathrm{Tr}(\rho K_i^* K_i) = \mathrm{Tr}(\rho) = 1$ for all $\rho \in S(\mathcal{H})$.

This is the representation of a quantum channel in the Schrödinger picture, with channels acting on density operators. However, we can also consider the Heisenberg picture, where the channel acts on operators $A \in \mathcal{B}(\mathcal{H})$ instead. The corresponding representation is then obvious from the properties of the trace

$$\mathrm{Tr}(T(\rho)A) = \sum_{i=1}^{m} \mathrm{Tr}(K_i \rho K_i^* A) = \sum_{i=1}^{m} \mathrm{Tr}(\rho K_i^* A K_i),$$

so the corresponding channel in the Heisenberg picture is given by $T_*(\cdot) = \sum_{i=1}^{r} K_i^* \cdot K_i$. Complete positivity is equivalent in both pictures, but the condition for trace preservation in the Heisenberg picture is $T_*(\mathbf{1}) = \mathbf{1}$. We denote the space of all channels on a Hilbert space $\mathcal{H}$ in the Heisenberg picture as $\mathcal{T}_*(\mathcal{H})$.

There is a useful correspondence between quantum channels and operators on a larger space, which is known as the *Choi-Jamiolkowski isomorphism*:

Theorem 4 (Choi-Jamiolkowski)**.** *The two equations*

$$\tau = (T \otimes id_d)(|\Omega\rangle\langle\Omega|), \quad \mathrm{Tr}(AT(B)) = d\mathrm{Tr}(\tau A \otimes B^T)$$

provide a one-to-one correspondence between linear maps $T : \mathcal{B}(\mathbb{C}^d) \to \mathcal{B}(\mathbb{C}^d)$ and operators $\tau \in \mathcal{B}(\mathbb{C}^d \otimes \mathbb{C}^d)$, for all $A, B \in \mathcal{B}(\mathbb{C}^d)$ and the d-dimensional maximally entangled state $|\Omega\rangle := \frac{1}{\sqrt{d}} \sum_{i=1}^{d} |ii\rangle$. The maps $T \mapsto \tau$ and $\tau \mapsto T$ defined by these equations are mutual inverses with the following correspondences:

1. *Complete positivity: T_* is CP iff $\tau \geq 0$*
2. *Trace preserving: $T_*(\mathbf{1}) = \mathbf{1}$ iff $Tr_1(\tau) = \mathbf{1}/d$*

In summary, as a linear map, a quantum channel can be represented as an operator acting on a larger space of dimension d^2, with this link provided by the Choi-Jamiolkowski isomorphism. Additionally, this provides a much easier method for checking whether a linear map T is completely positive; we just check that τ is positive. Furthermore, the minimal number of Kraus operators is simply the rank of this operator $r := \mathrm{rank}(\tau)$. This is known as the Kraus rank, and must satisfy $r \leq d^2$ since this is the dimension of the Choi state's Hilbert space. This minimal representation, however, is not unique. Additionally, in Theorem 3 we specified that there is some $m \leq d^2$ corresponding to a Kraus representation, but we can find a Kraus representation for any $m \geq r$ in general.

We have only included the correspondences that are relevant to quantum channels here; for a full list, we refer to the lecture notes by M. Wolf [175].

Quantum channels can be related back to quantum open systems via the *Stinespring representation*:

Theorem 5 (Stinespring representation)**.** *Let $T \in \mathcal{CPTP}(\mathbb{C}^d)$ be a quantum channel with Kraus rank $r \leq d^2$. Then for every $m \geq r$ there is an isometry $V : \mathbb{C}^d \to \mathbb{C}^d \otimes \mathbb{C}^m$ such that*

$$T(\rho) = Tr_E(V\rho V^*), \quad \forall \rho \in S(\mathbb{C}^d), \tag{2.13}$$

where the partial trace is taken over the "environment" $\mathbb{C}^m$. Alternatively, in the Heisenberg picture

$$T_*(A) = V^*(A \otimes \mathbf{1}_r)V, \quad \forall A \in \mathcal{B}(\mathbb{C}^d). \tag{2.14}$$

Note: the isometry performs the role of both embedding the system into a larger space and evolving the new system and environment with a unitary operator. For instance, let $m = r$, then we can write

$$V|\psi\rangle = U|\psi\rangle \otimes |\chi\rangle$$

for an environment state $|\chi\rangle \in \mathbb{C}^r$, unitary $U \in \mathcal{B}(\mathbb{C}^d \otimes \mathbb{C}^m)$ and all states $|\psi\rangle \in \mathbb{C}^d$. This isometry must still preserve the trace of a density operator, so it satisfies $V^*V = \mathbf{1}_d$.

The Kraus representation can then be recovered from the Stinespring representation by specifying an ONB $\{|i\rangle_E : i = 1, ..., r\}$ for the environment with

$$K_i = \langle i|_E U|\chi\rangle .$$

If we perform a projective measurement on the environment with respect to this ONB, we can specify it's effect on our system: let $\rho \in S(\mathcal{H})$ be the system's reduced state. Outcome i occurs with probability

$$p(i) = \mathrm{Tr}(\rho K_i^* K_i) \tag{2.15}$$

and the posterior state of the system is

$$\rho' = \frac{K_i \rho K_i^*}{\mathrm{Tr}(K_i \rho K_i^*)}. \tag{2.16}$$

In summary, a set of Kraus operators specify the effect on a system of a projective measurement on its environment. The outcome $i \in \{1, ..., r\}$, therefore, no longer corresponds with the eigenvalue of an observable. Instead, it specifies the Kraus operator that is applied to the system.

This is a very natural form of measurement for many realistic scenarios, where we do not directly observe a system. Instead we couple the system to another system that we refer to as the measurement device. This measurement device acts as the environment, so these two systems evolve together according to a unitary U. We then measure this measurement device to indirectly observe the main system.

In equation (2.15), we can replace the product of Kraus operators with a positive operator $M_i = K_i^* K_i$, where $\sum_i M_i = \mathbf{1}$ from the normalization condition of the Kraus operators. These operators still specify the probability of observing an index i through

$$p(i) = \mathrm{Tr}(\rho M_i).$$

However, we lose the connection to the Kraus operators, so we cannot specify the posterior state. This is not a problem in many scenarios, where measurement corresponds with our final action on a system. For instance, this is often the case in parameter estimation, where we use the outcomes of a final measurement to estimation some parameters. This idea generalizes to a positive operator-valued measure (POVM):

Definition 3 (Positive operator-valued measure). *Let $\mathcal{M} = \{M_i \in \mathcal{B}(\mathcal{H}) : i = 1, \ldots, m\}$ be a set of positive operators on Hilbert space $\mathcal{H}$ satisfying the normalization condition $\sum_{i=1}^m M_i = \mathbf{1}$. Then $\mathcal{M}$ specifies a final measurement on any state $\rho \in S(\mathcal{H})$ with outcomes $i \in \{1, \ldots, m\}$ and corresponding probabilities*

$$p(i) = \mathrm{Tr}(\rho M_i).$$

This provides the most general description of a projective measurement on an environment since we do not have to specify anything about the environment's structure or the projective measurement that we perform, we just specify the set of postive operators. The following theorem then provides the connection back to a projective measurement on some environment in manner to the Stinespring representation:

Theorem 6 (Naimark theorem)**.** *Let $\{M_1, ..., M_m\}$ be a POVM. There exists a Hilbert space $\mathcal{H}'$, a projective measurement $\{P_1, ..., P_m\}$ and an isometry $V : \mathcal{H} \to \mathcal{H}'$ such that*

$$M_i = V^* P_i V$$

for $i = 1, ..., m$.

That is, there exists a choice of embedding into an environment and a projective measurement on said environment that can realise any POVM.

2.2.2 *Common Properties*

The following properties [8, 30, 54] all concern the fixed points of quantum channels, which are related to the channel's eigenvalues. Therefore, it is prudent to first introduce these concepts in the context of quantum channels: the Choi-Jamiolkowski isomorphism demonstrates that a quantum channel $T : \mathcal{B}(\mathbb{C}^d) \to \mathcal{B}(\mathbb{C}^d)$ can be considered a linear map on a space of dimension d^2, where this linear map is the Choi state $\tau \in \mathcal{B}(\mathbb{C}^d \otimes \mathbb{C}^d)$. This is not the standard choice of linear map however: the operators form a vector space with respect to the Hilbert-Schmidt inner product:

Definition 4 (Hilbert-Schmidt inner product)**.** *For a Hilbert space $\mathcal{H} = \mathbb{C}^d$ and operators $A, B \in \mathcal{B}(\mathbb{C}^d)$, the associated inner product is the Hilbert-Schmidt inner product*

$$\langle A, B \rangle_{HS} := \mathrm{Tr}(A^* B),$$

where A^* denotes the adjoint (conjugate transpose) of A and $\mathrm{Tr}(\cdot)$ denotes the trace.

Provided this inner product, we can specify a orthonormal basis (ONB) on the vector space as a set of operators $\{E_i; i = 1, ..., d^2\}$ that satisfy $\langle E_i, E_j \rangle_{HS} = \delta_{ij}$. We can then vectorize the quantum channel as

$$\boldsymbol{T}_{ij} = \mathrm{Tr}(E_i^* T(E_j)). \tag{2.17}$$

The standard choice of ONB is $E_\alpha = |k\rangle \langle l|$ for $\alpha = (k, l)$ and $k, l = 1, ..., d$, i.e., a matrix of zeroes with element $\alpha = (k, l)$ set to 1. Substituting this ONB and a minimal Kraus representation $\{K_i; i = 1, ..., r\}$ into equation (2.17), we find

$$\begin{aligned} \boldsymbol{T}_{(k,l),(m,n)} &= \sum_{i=1}^{r} \langle k| K_i |m\rangle \langle n| K_i^* |l\rangle \\ &= \sum_{i=1}^{r} \langle k| K_i |m\rangle \langle l| \bar{K}_i |n\rangle \end{aligned}$$

where the $\bar{K}_i$ denotes entry-wise complex conjugation of K_i. This implies

$$\boldsymbol{T} = \sum_{i=1}^{r} K_i \otimes \bar{K}_i. \tag{2.18}$$

Having successfully vectorized the channel, its eigenvalues and eigenvectors can be found as the solution to the characteristic equation

$$\det(\lambda \mathbf{1} - \sum_{i=1}^{r} K_i \otimes \bar{K}_i) = 0,$$

where $\det(\cdot)$ denotes the determinant.

In the language of linear algebra, it has at most d^2 distinct eigenvalues $\lambda \in \mathbb{C}$ with corresponding eigenvectors $A \in \mathcal{B}(\mathcal{H})$ that satisfy

$$T(A) = \lambda A. \tag{2.19}$$

Since the channel acts on operators, these eigenvectors are actually operators $A \in \mathcal{B}(\mathbb{C}^d)$, not vectors $|\psi\rangle \in \mathcal{H}$ as would be the case for an operator or observable. Therefore, we will instead refer to them as eigenoperators to avoid confusion. Similarly, we use λ for the eigenvalues of a quantum channel in order to avoid confusing them with the eigenvalues of an operator as introduced in section 2.1.1.

What values can these eigenvalues take? The answer in linear algebra is given by Perron-Frobenius theory, so we are interested in its applications to the specific case of quantum channels. In particular there are 3 key results, which we group together for convenience:

Theorem 7 (Perron-Frobenius Results)**.** *For a quantum channel $T \in \mathcal{CPTP}(\mathcal{H})$ the following results hold:*

1. ***Density eigenoperator**: The channel has a density operator $\rho \in S(\mathcal{H})$ as an eigenoperator with corresponding eigenvalue $\lambda = 1$,*
2. ***Conjugate pairs**: If λ is an eigenvalue of channel $T \in \mathcal{CPTP}(\mathcal{H})$ with eigenoperator $A \in \mathcal{B}(\mathcal{H})$ then λ^* is also an eigenvalue with corresponding eigenoperator A^*,*
3. ***Unit circle**: The eigenvalues of a channel $T \in \mathcal{CPTP}(\mathcal{H})$ all satisfy $|\lambda| \leq 1$.*

The proof of 2. is obvious from the Kraus representation. Additionally, the proof of 3. is obvious from the Stinespring representation. The proof of 1., however, is not obvious and so is worth including [172]:

Proof. Let $T \in \mathcal{CPTP}(\mathcal{H})$ be our quantum channel. For every non-negative integer n, define a map $\phi_n \in \mathcal{CPTP}(\mathcal{H})$ as

$$\phi_n(A) = \frac{1}{2^n} \sum_{k=0}^{2^n-1} T^k(A)$$

for each $A \in \mathcal{B}(\mathcal{H})$, and define the set

$$\mathcal{C}_n = \{\phi_n(\rho) : \rho \in S(\mathcal{H})\}.$$

As T is linear, positive and trace-preserving, so is ϕ_n. It then follows that $\mathcal{C}_n$ is a compact and convex subset of $S(\mathcal{H})$. By this convexity of the set $\mathcal{C}_n$, it follows that

$$\phi_{n+1}(\rho) = \frac{1}{2}\phi_n(\rho) + \frac{1}{2}\phi_n(T^{2^n}(\rho)) \in \mathcal{C}_n$$

for every $\rho \in S(\mathcal{H})$. Therefore, $\mathcal{C}_{n+1} \subseteq \mathcal{C}_n$ for every n. As each $\mathcal{C}_n$ is compact and $\mathcal{C}_{n+1} \subseteq \mathcal{C}_n$ for all n, it follows that there must exist an element

$$\rho \in \mathcal{C}_0 \cap \mathcal{C}_1 \cap ... \tag{2.20}$$

contained in the intersection of all of these sets.

Now, fix any choice of ρ satisfying (2.20). For an arbitrary non-negative integer n, it holds that $\rho = \phi_n(\sigma)$ for some choice of $\sigma \in S(\mathcal{H})$, and therefore,

$$T(\rho) - \rho = T(\phi_n(\sigma)) - \phi_n(\sigma) = \frac{T^{2n}(\sigma) - \sigma}{2^n}.$$

As the trace distance between two density operators cannot exceed 2, it follows that

$$||T(\rho) - \rho||_1 \leq \frac{1}{2^{n-1}},$$

where $||\cdot||$ is the trace distance. This bound holds for every n, which implies $||T(\rho) - \rho||_1 = 0$, and, therefore, $T(\rho) = \rho$ as required. □

With these three results in mind, we define the first potential property of a quantum channel:

Definition 5 (Ergodicity)**.** *A quantum channel $T \in \mathcal{CPTP}(\mathcal{H})$ is ergodic if there exists a unique eigenoperator $\rho_* \in \mathcal{B}(\mathcal{H})$ with eigenvalue $\lambda = 1$.*

That is, there is only one eigenoperator with eigenvalue 1. The following theorem gives more insight into the properties of ergodic channels and could be used as an alternative definition of ergodicity.

Theorem 8. *A quantum channel $T \in \mathcal{CPTP}(\mathcal{H})$ is ergodic iff there exists a unique fixed state $\rho_* \in S(\mathcal{H})$ such that*

$$\lim_{N\to\infty} ||\Sigma_N(\rho) - \rho_*||_1 = 0, \quad \forall \rho \in S(\mathcal{H})$$

where

$$\Sigma_N(\rho) := \frac{1}{N+1}\sum_{n=0}^{N} T^n(\rho)$$

and $||A||_1 := \mathrm{Tr}(\sqrt{A^*A})$ is the trace norm.

This object $\Sigma_N(\cdot)$ is a convex combination of quantum channels, so it is also a valid quantum channel. The state $\Sigma_N(\rho)$ then represents the average state over repeated applications of the channel, starting with initial state $\rho \in S(\mathcal{H})$. This alternative definition tells us that the average state converges to the fixed point ρ_*. Accordingly, predictions about the average expectation value of an observable $A \in \mathcal{B}(\mathcal{H})$ in the limit $N \to \infty$ can be made from the fixed point alone since

$$\lim_{N\to\infty} \mathrm{Tr}(\Sigma_N(\rho)A) = \mathrm{Tr}(\rho_* A),$$

however, there is no guarantee that the actual state $T^N(\rho)$ converges to this fixed point. In fact, for an arbitrary initial state we expect some periodic behaviour that averages out to the fixed point. Of course, if we started in the fixed point we would have $T^n(\rho_*) = \rho_*$ for all $n \in \mathbb{N}$ and the state would never change.

An example of an ergodic channel in terms of qubits is the following flipping channel:

$$T(\cdot) = \langle 0| \cdot |0\rangle |1\rangle\langle 1| + \langle 1| \cdot |1\rangle |0\rangle\langle 0|.$$

This channel immediately kills all off-diagonal elements of a state $\rho \in S(\mathcal{H})$, but the diagonal elements keep flipping. In terms of the state's Bloch vector, the channel maps $\boldsymbol{r} = (r_x, r_y, r_z) \to (0, 0, -r_z)$ with repeated applications continuing to flip this r_z component. Therefore, the dynamics never settle down for an arbitrary initial state. Instead, the fixed point corresponds with Bloch vector $\boldsymbol{r_*} = (0, 0, 0)$ since there is no r_z to flip. This is the density operator $\rho_* = \frac{1}{2}(|0\rangle \langle 0| + |1\rangle \langle 1|)$. Additionally, any state with $r_z = 0$ will converge to this stationary state after a single application of the channel.

The next property then corresponds with the case where we get actual convergence to the fixed point:

Definition 6 (Mixing)**.** *A quantum channel $T \in \mathcal{CPTP}(\mathcal{H})$ is mixing if the fixed point $\rho_* \in S(\mathcal{H})$ is the only eigenoperator with eigenvalue λ such that*

$$|\lambda| = 1.$$

In analogue to the alternative definition of ergodicity presented in Theorem 8, we have

$$\lim_{n\to\infty} ||T^n(\rho) - \rho_*||_1 = 0$$

for all $\rho \in S(\mathcal{H})$, so we often refer to the fixed point as the stationary state.

An example of a mixing quantum channel is a probabilistic state replacement channel that replaces an initial state with a pure state $\rho = |\psi\rangle \langle\psi|$, $|\psi\rangle \in \mathcal{H}$, with some probability $p \in (0, 1)$

$$T(\cdot) = p\mathrm{Tr}(\cdot)\rho + (1 - p)\mathrm{id}(\cdot)$$

This density operator ρ is clearly a fixed point. Additionally, since $\lim_{n\to\infty} p^n = 0$, the initial state is guaranteed to converge to ρ. Another common example is a convex combination of an ergodic quantum channel T and the identity channel

$$T'(\cdot) = pT(\cdot) + (1 - p)\mathrm{id}(\cdot),$$

which one can show is mixing for all $p \in (0, 1)$ [30].

The final two properties are then special cases of ergodicity and mixing respectively, where we also insist that the fixed point is of *maximal rank*:

Definition 7 (Irreducibility)**.** *A quantum channel $T \in \mathcal{CPTP}(\mathcal{H})$ is irreducible if there is no non-trivial subspace of $\mathcal{H}$ that is left invariant by its action. That is, if $P \in \mathcal{B}(\mathcal{H})$ is an orthogonal projector such that*

$$T(PS(\mathcal{H})P) \subseteq PS(\mathcal{H})P$$

then $P \in \{0, \boldsymbol{1}\}$.

Irreducibility ensures there are no subspaces that 'trap' the action of the quantum channel, so repeated applications of the channel can explore the whole space. This is best understood through the following corollary, which relates irreducibility back to properties of the eigenvectors of the channel:

Corollary 1 (Irreducibility)**.** *An irreducible quantum channel $T \in \mathcal{CPTP}(\mathcal{H})$ is ergodic with fixed point $\rho_* \in S(\mathcal{H})$ and this fixed point is strictly positive (faithful) $\rho_* > 0$.*

So repeated applications of an irreducible channel may still result in some transient behaviour, but the average state will converge to the fixed point as before. Additionally, since there are no non-trivial subspaces, the fixed point must be of maximal rank. This is equivalent to stating that the fixed point is strictly positive, therefore, we can also specify the behaviour of the average state's rank in the asymptotic limit. The flipping channel that we presented as an example for ergodicity is also irreducible since its fixed point clearly has maximal rank.

Definition 8 (Primitivity)**.** *A quantum channel $T : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H})$ is primitive if it is mixing with stationary state $\rho_* \in S(\mathcal{H})$ and this stationary state is strictly positive (faithful) $\rho_* > 0$.*

In summary, a primitive quantum channel has guaranteed convergence to a stationary state ρ_* and this stationary state is strictly positive $\rho_* > 0$. This corresponds with the state being maximal rank, which is useful in proofs as we often rely on the state's rank not spontaneously changing at any point. With a primitive channel, we can guarantee the channel's rank does not change in the asymptotic limit.

An example of a primitive channel is the state replacement channel we used as an example for a mixing channel, where we also require that $\rho \geq 0$. Another related example is the depolarizing channel, which replaces the state of a qubit with the completely mixed state $\mathbf{1}/2$ with probability p:

$$T(\cdot) = p\mathrm{Tr}(\cdot)\frac{\mathbf{1}}{2} + (1-p)\mathrm{id}(\cdot). \tag{2.21}$$

In terms of the qubit's Bloch vector, this channel has the effect $\boldsymbol{r} = (r_x, r_y, r_z) \to (1-p)\cdot\boldsymbol{r}$, so repeated applications cause the Bloch vector to vanish. Additionally, we can easily specify the channels corresponding Kraus operators: consider the Bloch representation of an arbitrary qubit as introduced in equation (2.3). This implies that

$$\frac{\mathbf{1}}{2} = \frac{1}{4}\left[\rho + \sigma_x\rho\sigma_x + \sigma_y\rho\sigma_y + \sigma_z\rho\sigma_z\right]$$

for arbitrary $\rho \in S(\mathbb{C}^2)$. Utilizing this with equation (2.21), we find

$$\begin{aligned} T(\rho) &= \frac{p}{4}\left[\rho + \sigma_x\rho\sigma_x + \sigma_y\rho\sigma_y + \sigma_Z\rho\sigma_z\right] + (1-p)\rho \\ &= (1-\frac{3p}{4})\rho + \frac{p}{4}\left[\sigma_x\rho\sigma_x + \sigma_y\rho\sigma_y + \sigma_z\rho\sigma_z\right] \end{aligned}$$

from which we can identify

$$K_1 = \sqrt{1 - \frac{3p}{4}}\mathbf{1}, \qquad K_2 = \frac{\sqrt{p}}{2}\sigma_x$$
$$K_3 = \frac{\sqrt{p}}{2}\sigma_y, \qquad K_4 = \frac{\sqrt{p}}{2}\sigma_z.$$

Depolarization is a common form of noise in quantum systems, so this is a particularly important example of a primitive channel.

2.3 OPEN QUANTUM SYSTEMS AND QUANTUM MARKOV CHAINS

In this section we introduce the setting used throughout this thesis: quantum Markov chains (QMCs). These are discrete-time systems, so we then discuss how we can take their continuous time limit to produce the Lindbladian description of open quantum systems. This is followed by a brief description of some other open quantum systems.

2.3.1 *Quantum Markov Chains*

A quantum Markov chain (QMC) is the quantum analogue of a classical Markov chain. In a classical Markov chain, a system is described by a set of possible states and transition probabilities between each state. Since it relies on transition probabilities, this is a stochastic process, and the current state jumps from state to state in each step. For example, consider a simple model of the weather: let w_i be the current state of the weather on day i with possible states $W = \{'Sun', 'Rain'\}$. For our transition probabilities, let us assume that if it rained today, then there is a 20% chance it'll be sunny tomorrow, and if it was sunny then there is a 40% chance it'll rain tomorrow. What we observe is a sequence (w_i) such as $('Rain', 'Rain', 'Sun', 'Sun')$, and this is the classical Markov chain. Crucially, the probabilities of each possible transition depend on only the current state of the system; a stochastic process that satisfies this property is known as Markovian.

A QMC is a discrete-time input-output quantum system in which a succession of input systems (noise units) identically prepared in state $|\chi\rangle \in \mathbb{C}^k$ interact with another SOI $\mathbb{C}^d$. For clarity, we will refer to these input systems as input units. If $|\psi\rangle \in \mathbb{C}^d$ is the initial state of the SOI, then the joint state after n interactions is

$$|\Psi(n)\rangle = U^{(n)}...U^{(1)} |\psi\rangle \otimes |\chi\rangle^{\otimes n},$$

where $U^{(n)}$ is the unitary on $\mathbb{C}^d \otimes \mathbb{C}^k$ that describes the interaction between the SOI and the n-th input unit.

This can be expressed as the following by specifying a basis $\{|i_j\rangle ; i_j = 1, ..., k\}$ for each noise unit:

$$|\Psi(n)\rangle = \sum_{i_1,...,i_n} K_{i_n}...K_{i_1} |\psi\rangle \otimes |i_1 \otimes ... \otimes i_n\rangle$$

where $K_{i_j} = \langle i_j|U|\chi\rangle$ specifies a Kraus operator. In general this ONB $\{|i_j\rangle\}$ does not have to be the same for each input unit, however, if we use the same ONB, we

find that the reduced state of our SOI before measurement is given by a quantum channel repeatedly applied to the SOI's initial state:

$$\rho(n) = \sum_{i_1,\dots,i_n} K_{i_n}...K_{i_1} |\psi\rangle \langle\psi| K_{i_1}^*...K_{i_n}^* = T^n(|\psi\rangle \langle\psi|),$$

where T is the quantum channel corresponding to Kraus operators $K_i = \langle i| U |\chi\rangle$ with $i = 1, \dots, k$ and $T^n(\cdot)$ represents n repeated applications of the channel.

At this stage everything can be described in terms of pure states, however, we now want to measure the environment to indirectly observe the SOI. When we perform this measurement with respect to the ONB specified above, the (not-normalized) conditional state of our SOI given we observe $\boldsymbol{i} = (i_1, ..., i_n)$ is

$$\rho(n, \boldsymbol{i}) = K_{i_n}...K_{i_1} |\psi\rangle \langle\psi| K_{i_1}^*...K_{i_n}^*$$

Additionally, the probability of observing this sequence $\boldsymbol{i}$ is $\mathrm{Tr}(\rho(n, \boldsymbol{i}))$.

If we consider a single step, then we have probabilities $p(i_1) = \mathrm{Tr}(K_{i_1}\rho_0 K_{i_1}^*)$, $i_1 \in 1, ..., k$, and corresponding (not-normalized) posterior states

$$\rho(1, i_1) = K_{i_1}\rho_0 K_{i_1}^*$$

as we found in section 2.2.1. If we now consider an additional step, we find

$$p(i_1, i_2) = \mathrm{Tr}(K_{i_2}K_{i_1}\rho_0 K_{i_1}^* K_{i_2}^*) = \mathrm{Tr}(K_{i_2}\rho(1, i_1)K_{i_2}^*)$$

for $i_1, i_2 \in 1, ..., k$ and

$$\rho(2, i_1, i_2) = K_{i_2}K_{i_1}\rho_0 K_{i_1}^* K_{i_2}^* = K_{i_2}\rho(1, i_1)K_{i_2}^*$$

Similarly, $p(\boldsymbol{i}) = \mathrm{Tr}(K_{i_n}\rho(n-1, i_1, ..., i_{n-1})K_{i_n}^*)$ and

$$\rho(n, \boldsymbol{i}) = K_{i_n}\rho(n-1, i_1, ..., i_{n-1})K_{i_{n-1}}^*$$

Therefore, we can always calculate the probabilities and possible posterior states of the next step from the current state.

This is where we see the Markovian structure emerge, so we refer to this as a quantum Markov chain in analogy to the classical case. This setup is perfectly Markovian by construction since a new input unit is introduced in each step; this ensures that only the current state of the system influences the next state as we expect.

We now want to consider a QMCs continuous limit to demonstrate that we can recover the master equation for a Markovian quantum open system [27, 119, 144]. To this end, we take Kraus operators (up to the first order in Δt) of the form [177]

$$K_1 = \mathbf{1} - iH\Delta t - \frac{1}{2}\sum_i L_i^* L_i \Delta t, \tag{2.22}$$

$$K_i = L_i\sqrt{\Delta t} \quad i = 2, ..., r, \tag{2.23}$$

for a Hamiltonian $H \in \mathcal{B}(\mathcal{H})$ and (not necessarily Hermitian) jump operators $L_i \in \mathcal{B}(\mathcal{H})$. Since these Kraus operators are only specified up to the first order, they do not satisfy the normalization condition. We could, however, correct this by specifying the higher order terms and $\lim_{\Delta t\to 0}\sum_i K_i^* K_i = \mathbf{1}$ as expected.

Additionally, we could allow both the Hamiltonian and jump operators to depend on time, but we ignore this here for simplicity.

To derive a master equation for the evolution, consider the reduced state of the SOI after a single step:

$$\rho(t+\Delta t) = \sum_{i=1}^{r} K_i \rho(t) K_i^*.$$

Substituting in the Kraus operators from equations (2.22) and (2.23), one can show that

$$\partial_t \rho(t) = \lim_{\Delta t \to 0} \frac{\rho(t+\Delta t) - \rho(t)}{\Delta t} \tag{2.24}$$

$$= -i[H, \rho(t)] - \frac{1}{2} \sum_i \{L_i^* L_i, \rho(t)\} + L_i \rho(t) L_i^*, \tag{2.25}$$

which is the Gorini-Kossakowski-Sudarshan-Lindblad (GKSL) master equation [27, 119, 144] for quantum open systems. In particular, the operators L_i specify stochastic interactions between the system and its environment.

To elucidate this further, consider grouping together the terms of order Δt, we obtain

$$H_{\text{eff}} := H + \frac{i}{2} \sum_i L_i^* L_i.$$

This effective Hamiltonian describes the conditional evolution of the system given that no jumps has been emitted. This gives rise to a good interpretation of the master equation: the SOI evolves under this effective Hamiltonian and this evolution is interspersed with stochastic exchanges with the environment governed by the jump operators L_i. In particular, these terms $L_i^* L_i$ regulate energy transfer between the system and environment due to the stochastic jumps. The evolution of a pure state $|\psi\rangle \in \mathcal{H}$ given that jumps $(i_1, ..., i_n)$ were observed at times $(t_1, ..., t_n)$ is then

$$|\psi(t, \boldsymbol{i})\rangle = e^{-iH_{\text{eff}}(t-t_n)} L_{i_n} e^{-iH_{\text{eff}}(t_n - t_{n-1})} L_{i_{n-1}} ... L_{i_1} e^{-iH_{\text{eff}}(t_1)} |\psi\rangle, \quad t \geq t_n$$

up to normalization. And to find the reduced state of the SOI before measuring the environment, we'd have to consider the integral over all possible jumps occurring at all possible times:

$$\rho(t) = \sum_{k=0}^{\infty} \sum_{i_1,...,i_k} \int e^{-iH_{\text{eff}}(t-t_n)} L_{i_n} ... e^{-iH_{\text{eff}}(t_2 - t_1)} L_{i_1} e^{-iH_{\text{eff}} t_1} \rho_0 \tag{2.26}$$

$$\cdot e^{iH_{\text{eff}} t_1} L_{i_1}^* e^{iH_{\text{eff}}(t_2 - t_1)} ... L_{i_n}^* e^{iH_{\text{eff}}(t-t_n)} dt dt_1 ... dt_k, \tag{2.27}$$

where $\rho_0 = |\psi\rangle \langle\psi|$.

This is the most common formalism of open quantum systems, describing systems that are coupled to an environment under some reasonable assumptions [119]. In particular, we assume that three approximations hold:

1. Born approximation - the coupling between the system and environment is weak,

2. Markov approximation - correlations between the system and environment build up slowly and decay quickly ensuring the dynamics are Markovian for small time scales,

3. Rotating wave approximation - high frequency terms in the interaction picture can be neglected.

We refer to [119] for more details. Therefore, it describes many experimental settings well as they are designed with minimising noise in mind, often cooling down the apparatus, shielding it from external fields, etc. Alternative formalisms of quantum open systems all start with some unitary evolution between a system and an environment, then make different assumptions about the nature of this interaction and environment.

3

PARAMETER ESTIMATION

In this chapter we introduce the necessary theory from both classical and quantum estimation required for this thesis. As discussed in section 3.2, quantum estimation reduces to classical estimation after measurement, therefore, the brief review of classical estimation found in section 3.1 is helpful. This allows us to introduce the common ideas shared between both types of estimation - mainly estimators and the Cramér-Rao bound. In particular, we highlight some key statistics and how they fit into estimation theory. We then introduce asymptotic estimation schemes, which provide many tools that help us to construct optimal estimators. Section 3.1.1 gives a non-technical introduction to one of these tools: local asymptotic normality (LAN). LAN demonstrates that a model described by independent samples from a given distribution becomes statistically equivalent to one described by a single sample from a normal distribution in the asymptotic limit of large sample size. This is a fundamental result in asymptotic statistics and can be used in constructing optimal estimators. local asymptotic normality (LAN) can be proven through two different methods, so we focus on sketching out both these different proofs.

Section 3.2 introduces the basics of quantum parameter estimation. In quantum estimation we usually focus on the quantum aspects of the estimation problem, optimizing the choice of measurements and quantum states involved. Therefore, we focus on the quantum side of the problem. We then discuss some of the common frameworks for quantum estimation. Finally, we introduce the quantum analogue of LAN in section 3.2.1

3.1 CLASSICAL ESTIMATION

We start by introducing the following standard problem [3, 108, 151, 152, 170]. We are given a sample from a random variable X taking value in a measurable space (Ω, Σ), and we would like to infer a property of its distribution. In particular, we assume that the distribution is known up to some parameters $\boldsymbol{\theta} = (\theta_1, ..., \theta_k) \in \Theta \subseteq \mathbb{R}^k$; otherwise, the problem is intractable. That is,

$$X \sim \mathbb{P}_{\boldsymbol{\theta}} \in \mathcal{P},$$

where $\sim$ denotes that X is distributed according to distribution $\mathbb{P}_{\boldsymbol{\theta}}$, Θ is the parameter space, and $\mathcal{P} = \{\mathbb{P}_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta\}$ is known as the statistical model. Additionally, we assume the derivatives of $\mathbb{P}_{\boldsymbol{\theta}}$ with respect to the parameters exist so that the statistical manifold is smooth.

These are good assumptions in many physical situations, where the underlying distributions are well understood. For instance, the Poisson distribution characterizes 'rare' events well, such as the rate at which earthquakes occur or wait times in a queue. Additionally, in many situations the family of distributions is obvious from the process producing the random variable.

An illustrative example of this that we will refer back to often is a coin toss, which is clearly a Bernoulli trial $X \sim \mathcal{B}(q)$ with possible outcomes 'Heads' $H \equiv 1$ and 'Tails' $T \equiv 0$. The corresponding probabilities are then $p_q(1) := \mathbb{P}_q(X = 1) = q$ and $p_q(0) = 1 - q$ respectively, and the statistical model is $\mathcal{P} = \{\mathbb{P}_q : 0 \leq q \leq 1\}$.

In this section we will assume that the random variable (RV) has discrete outcomes for simplicity, but the theory can be extended to continuous random variables. In a similar manner to the coin flip, the probability of obtaining outcome $x \in \Omega$ in a general scenario is then $p_{\boldsymbol{\theta}}(x) := \mathbb{P}_{\boldsymbol{\theta}}(X = x)$. We can define the expected value of a real-valued function $f(X)$ as

$$\mathbb{E}(f(X)) := \sum_{x \in \Omega} p(x) f(x).$$

In particular the expected value of X is $\mathbb{E}(X)$ and the variance of X is $\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$.

If we have repeated access to the probability distribution, i.e., a random variable $\boldsymbol{X} = (X_1, ..., X_n) \sim \mathbb{P}_{\boldsymbol{\theta}}^n$ with

$$\mathbb{P}_{\boldsymbol{\theta}}^n(X_1 = x_1, ..., X_n = x_n) = \mathbb{P}_{\boldsymbol{\theta}}(X_1 = x_1) \cdots \mathbb{P}_{\boldsymbol{\theta}}(X_n = x_n),$$

then we can generate an IID sample $\mathbf{x} = (x_1, ..., x_n) \in \Omega^n$ from which we want to estimate the parameters. Our sample space Ω^n for the coin toss would be collections of individual flips, i.e., $\Omega^2 = \{HH, HT, TH, TT\}$.

From the random variable, we then calculate a *statistic* [3, 151]:

Definition 9 (Statistic)**.** *A statistic $Z(X)$ is a measurable function that maps outcomes x from the sample space Ω to values in another space Ω', i.e., $Z : \Omega \to \Omega'$*

If we are interested in using a statistic to estimate the parameters, then Ω' should be the same as the parameter space Θ, but this does not have to be the case in general. And this certainly is not the case outside of estimation; most statistics are used to gain some insight about the underlying data. For instance, if the data represents a population census, then we may want to calculate statistics such as the proportion of the population over the age of 60 in order to inform healthcare policy. Another example is hypothesis testing, where we would have $\Omega' = \{0, 1\}$ corresponding with whether we reject or approve some hypothesis.

Some examples of statistics are the sample mean $\bar{\boldsymbol{X}} := 1/n \sum_{i=1}^n X_i$, the sample variance $1/n \sum_{i=1}^n (X_i - \bar{\boldsymbol{X}})^2$ or a random function like $Z(X_1, ..., X_n) = X_1 + X_2$, but these are not necessarily insightful. In particular, this last statistic does not even utilize all the data available. This leads us onto a key concept for statistics - *sufficiency* [3, 151]:

Definition 10 (Sufficiency)**.** *A statistic $Z(X)$ is sufficient for the parameters $\boldsymbol{\theta}$ if the conditional probability distribution given the value of the statistic*

$$\mathbb{P}_{\boldsymbol{\theta}}(X = x | Z(X) = z)$$

does not depend on the parameters.

And the following theorem [3] is very useful for proving when a statistic is sufficient:

Theorem 9 (Fisher-Neyman factorization)**.** *A statistic $Z(X)$ is sufficient for $\boldsymbol{\theta}$ iff for all $\boldsymbol{\theta} \in \Theta$,*

$$L_{\boldsymbol{\theta}}(x) = f(Z(x), \boldsymbol{\theta})g(x),$$

where $L_{\boldsymbol{\theta}}(x) := \mathbb{P}_{\boldsymbol{\theta}}(X = x)$ is the likelihood of sample $x \in \Omega$, the function $f(\cdot)$ depends on $\boldsymbol{\theta}$ and the statistic $Z(X)$, and the function $g(\cdot)$ does not depend on $\boldsymbol{\theta}$.

In particular, for IID random variables the likelihood is

$$L_{\boldsymbol{\theta}}(\boldsymbol{X}) := \prod_{i=1}^{n} \mathbb{P}_{\boldsymbol{\theta}}(X_i)$$

and for the coin toss we have

$$L_q(\boldsymbol{X}) = q^{\sum_i X_i}(1-q)^{n-\sum_i X_i} = q^{n\bar{\boldsymbol{X}}}(1-q)^{n(1-\bar{\boldsymbol{X}})},$$

so $\bar{\boldsymbol{X}}$ is a sufficient statistic for the coin toss by Theorem 9 with $f_q(\bar{\boldsymbol{X}}, \boldsymbol{X}) = q^{n\bar{\boldsymbol{X}}}(1-q)^{n(1-\bar{\boldsymbol{X}})}$ and $g(\boldsymbol{X}) = 1$.

For a random variable X and sufficient statistic Z, we can write

$$\mathbb{P}_{\boldsymbol{\theta}}(X = x) = \mathbb{P}(X = x | Z = Z(x))\mathbb{P}_{\boldsymbol{\theta}}(Z = Z(x)),$$

where $\mathbb{P}(X = x | Z(X) = Z(x))$ does not depend on $\boldsymbol{\theta}$ by the definition of a sufficient statistic. Therefore, all the dependence on the parameters is captured in the probability distribution of the sufficient statistic. In particular, if we know the value of the sufficient statistic, then we capture all available information about the unknown parameters. It is in this sense that a sufficient statistic is equivalent to the original data. Additionally, to recreate a sample from the sufficient statistic $Z(X) = z$, we could draw from the distribution $\mathbb{P}(X | Z(X) = z)$ which does not depend on $\boldsymbol{\theta}$ and so does not contain any information. The entire sample is always trivially a sufficient statistic, but one can also show that many other statistics are also sufficient.

This provides a natural progression onto estimators [3, 151, 170]:

Definition 11 (Estimators)**.** *An estimator $\hat{\boldsymbol{\theta}}$ is a statistic that maps samples $\boldsymbol{x}$ from the sample space Ω to values in the parameter space Θ. In addition, we call $\hat{\boldsymbol{\theta}}(\boldsymbol{x}) = (\hat{\theta}_1(\boldsymbol{x}), ..., \hat{\theta}_k(\boldsymbol{x}))$ an estimate of $\boldsymbol{\theta}$.*

Therefore, an estimator is simply a statistic that we would like to use to estimate the parameters $\boldsymbol{\theta}$. Now, this definition leaves a lot to be desired:

1. It does not characterise the performance of the estimator at all.
2. It does not describe how to construct 'good' estimators.

For instance, consider the coin toss again: an arbitrary estimator that assigns $\hat{q} = 1$ if $X_1 = 1$ and $\hat{q} = 0$ if $X_1 = 0$ is valid by this definition, but this clearly cannot be optimal for all possible q since the estimator will never predict the correct

value for $0 < q < 1$. It also only utilizes the first outcome, so any information about q contained in $X_2, ..., X_n$ is lost by this estimator. Therefore, we first need to establish what criteria makes an estimator 'good' in some sense, tackling this first problem before considering how to construct estimators. A natural starting point is to expect that 'good' statistics are sufficient so that they capture all the available information about the parameters. We also typically desire estimators that are *unbiased* [3, 151, 170]:

Definition 12 (Unbiased Estimators)**.** *An estimator $\hat{\boldsymbol{\theta}}$ for parameters $\boldsymbol{\theta} \in \Theta$ and sample space $\Omega \subseteq \mathbb{R}^k$ corresponding to a statistical model $\mathcal{P}$ is unbiased if*

$$\mathbb{E}(\hat{\boldsymbol{\theta}}) := \sum_{x \in \Omega} p_{\boldsymbol{\theta}}(X = x)\hat{\boldsymbol{\theta}}(x) = \boldsymbol{\theta}$$

where $\mathbb{E}(\cdot)$ denotes the expectation value.

An example of an unbiased estimator for the coin toss is the sample mean $\bar{\boldsymbol{X}}$ since $\mathbb{E}(\bar{\boldsymbol{X}}) = \frac{n}{n}q + \frac{n}{n}(1-q) = q$. However, unbiasedness does not characterise the estimator's performance in any meaningful way. The error can be characterised by the mean square error (MSE)

$$[\mathrm{MSE}(\hat{\boldsymbol{\theta}})]_{ij} := \mathbb{E}\left[(\hat{\theta}_i - \theta_i)(\hat{\theta}_j - \theta_j)\right],$$

which corresponds to the covariance matrix for an unbiased estimator:

$$[\mathrm{Cov}(\boldsymbol{\theta})]_{ij} := \mathbb{E}\left[(\hat{\theta}_i - \mathbb{E}(\theta_i))(\hat{\theta}_j - \mathbb{E}(\theta_j))\right].$$

The Cramér-Rao bound [49, 108, 151] then provides the optimal performance of an unbiased estimator through a lower bound on its covariance:

Theorem 10 (Cramér-Rao Bound)**.** *For a statistical model $\mathcal{P}$ with parameters $\boldsymbol{\theta} \in \Theta$ and IID sample of size n, the covariance of an unbiased estimator $\hat{\boldsymbol{\theta}}(\boldsymbol{x})$ can be bounded by*

$$Cov(\hat{\boldsymbol{\theta}}) \geq \frac{1}{n} I_{\boldsymbol{\theta}}^{-1},$$

where $(I_{\boldsymbol{\theta}})_{ij} := \sum_{\boldsymbol{x} \in \Omega^n} p_{\boldsymbol{\theta}}(\boldsymbol{x})\, \partial_{\theta_i} log p_{\boldsymbol{\theta}}(\boldsymbol{x})\, \partial_{\theta_j} log p_{\boldsymbol{\theta}}(\boldsymbol{x})$ is the CFI of $\mathcal{P}$.

This implies an estimator is optimal if its covariance is given by the inverse of the CFI matrix. Note: for a single parameter this is simply a bound on the variance of the estimator: $\mathrm{Var}(\hat{\theta}) \geq 1/nI_\theta$; additionally, we presented the CRB for an IID RV, but we could just as well formulate it for an arbitrary RV X with $[\mathrm{Cov}(\hat{\boldsymbol{\theta}})]_{ij} \geq \left(I_{\boldsymbol{\theta}}^{-1}\right)_{ij}$ instead. The formulation in terms of an IID RV simply highlights the scaling of the CFI with respect to the sample size n of the IID RV, i.e- $I_{\boldsymbol{\theta}} \propto n$. A consequence of this is that with a sufficiently large sample we can estimate the parameters up to any desired accuracy. Additionally, the CFI is a property of the probability distribution, therefore, it depends on the a priori unknown value of $\boldsymbol{\theta}$. This can make calculating the CFI difficult in practical applications.

The CRB is not generally saturable for a statistical model $\mathcal{P}$. There is, however, a special class of models for which the CRB can be saturated by a single RV: exponential models. These include many of the most common distributions

such as the normal distribution, the Poisson distribution, etc. The probability distribution of an exponential random variable for parameters $\boldsymbol{\theta} = (\theta_1, ..., \theta_k)$ has form

$$p_{\boldsymbol{\theta}}(\boldsymbol{x}) = h(\boldsymbol{x})\exp\left[\boldsymbol{\eta}(\boldsymbol{\theta}) \cdot \boldsymbol{T}(\boldsymbol{x}) - A(\boldsymbol{\theta})\right],$$

where $h(\boldsymbol{x})$ is a non-negative real-valued function, $A(\boldsymbol{\theta})$ is a real-valued function and $T(\boldsymbol{x}) = (T_1(\boldsymbol{x}), ..., T_k(\boldsymbol{x}))$, $\boldsymbol{\eta}(\boldsymbol{\theta}) = (\eta_1(\boldsymbol{\theta}), ..., \eta_k(\boldsymbol{\theta}))$ for real-valued functions $T_i(\boldsymbol{x}), \eta_i(\boldsymbol{\theta})$.

To see how they can saturate the CRB, consider a Gaussian model - one example of a exponential model. A random variable $\boldsymbol{X} = (X_1, ..., X_k)$ with sample space $\Omega = \mathbb{R}^k$ is Gaussian if its probability density $p := \frac{d\mathbb{P}}{d\mu}$ with respect to the Lebesgue measure μ has the form of a multidimensional normal distribution:

$$p(\boldsymbol{x}) = \frac{1}{(2\pi)^{k/2}}\det(\Gamma)^{-1/2}\exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\nu})^T\Gamma^{-1}(\boldsymbol{x}-\boldsymbol{\nu})\right],$$

where the parameters are its expected value $\boldsymbol{\nu} = \mathbb{E}(\boldsymbol{X})$ and its covariance $\Gamma = \mathrm{Cov}(\boldsymbol{X})$. In the one-dimensional case, this is simply

$$p(\boldsymbol{x}) = \frac{1}{\sqrt{2\pi\Gamma}}\exp\left[-\frac{1}{2\Gamma}(x-\nu)^2\right].$$

Note: this distribution is completely specified by two parameters, therefore, we will denote it by $N(\boldsymbol{\nu}, \Gamma)$. We consider a normal distribution with fixed Γ, so in the one-dimensional case we can easily identify

$$\begin{aligned}
h(x) &= \frac{1}{(2\pi)^{k/2}}\det(\Gamma)^{-1/2}\exp(-\frac{1}{2\Gamma}x^2),\\
A(\nu) &= \exp(\frac{1}{2\Gamma}\nu^2),\\
\eta(\nu) &= \nu,\\
T(x) &= -2x.
\end{aligned}$$

The Gaussian shift model is then

$$G = \{N(\boldsymbol{\nu}, \Gamma) : \boldsymbol{\nu} \in \mathbb{R}^k, \text{fixed } \Gamma\}.$$

Consider a scenario where we want to estimate $\boldsymbol{\nu}$ from a Gaussian RV: the data is automatically an unbiased estimator for itself since $\mathbb{E}(\boldsymbol{X}) = \boldsymbol{\nu}$ by definition. Additionally, if we set $\Gamma = I_{\nu}^{-1}$, the data is also an optimal estimator for itself since its variance is now specified by the inverse of the CFI.

For models that are not exponential, the CRB is generally unattainable. It is, however, attainable in the asymptotic limit of large IID sample, which we discuss after introducing a few important statistics. The first is the *score function* [170]: an alternative expression for the CFI is

$$(I_{\boldsymbol{\theta}})_{ij} := \mathbb{E}(\dot{\ell}_{\boldsymbol{\theta},i}\dot{\ell}_{\boldsymbol{\theta},j}), \tag{3.1}$$

where $\dot{\ell}_{\boldsymbol{\theta},i} := \frac{\partial}{\partial\theta_i}\ell_{\boldsymbol{\theta}}$ are the *score functions* and $\ell_{\boldsymbol{\theta}} := \log(p_{\boldsymbol{\theta}})$ is the log-likelihood. To start, let's assume that the there is only one parameter θ for simplicity. If the

probability density is correctly normalized, then we have $\sum_x p_\theta(x) = 1$. This implies that

$$\sum_x \dot{p}_\theta(x) = 0 \quad \text{and} \quad \sum_x \ddot{p}_\theta(x) = 0.$$

One can also show

$$\dot{\ell}_\theta = \frac{\dot{p}_\theta}{p_\theta}, \qquad \ddot{\ell}_\theta = \frac{\ddot{p}_\theta}{p_\theta} - \left(\frac{\dot{p}_\theta}{p_\theta}\right)^2$$

by directly differentiating the log-likelihood $\ell_\theta = \log(p_\theta)$. Putting these together, one can easily demonstrate that

$$\mathbb{E}(\dot{\ell}_\theta) = 0, \qquad \mathbb{E}(\ddot{\ell}_\theta^2) = -I(\theta). \tag{3.2}$$

And finally,

$$\mathbb{E}(\dot{\ell}_\theta^2) = \mathbb{E}\left(\frac{\ddot{p}_\theta}{p_\theta} - \ddot{\ell}_\theta\right) = I(\theta). \tag{3.3}$$

All together, we have shown that the score function has mean $\mathbb{E}(\dot{\ell}_\theta) = 0$ and variance $\mathrm{Var}(\dot{\ell}_\theta) = I(\theta)$. If we scale a RV X by a constant $a \in \mathbb{R}$, then the resultant RV aX has mean $\mathbb{E}(aX) = a\mathbb{E}(X)$ and variance $\mathrm{Var}(aX) = a^2\mathrm{Var}(X)$. Utilizing this we can construct an estimator

$$\hat{\theta}(X) = \theta + \frac{1}{I_\theta}\dot{\ell}_\theta(X), \tag{3.4}$$

which is unbiased and saturates the CRB by construction, i.e., it is an optimal estimator! This has little operational value since it requires the circular logic of knowing θ to estimate θ. Suppose that we know that θ is in a small neighbourhood of some fixed $\theta_0 \in \Theta$, then the estimator

$$\hat{\theta}(X) = \theta_0 + \frac{1}{I_{\theta_0}}\dot{\ell}_{\theta_0}(X)$$

is an optimal estimator for θ up to first order. Additionally, we can make this localisation argument rigorous in an asymptotic scenario.

The second statistic is the *log-likelihood ratio process* [170]; this requires the Radon-Nikodym derivative of one probability distribution with respect to another, so we will assume that both distributions have probability densities with respect to the same measure μ for all $\theta \in \Theta$ so that this is well defined.

Definition 13. *Let $\mathcal{P}$ be a statistical model for random variable X with parameters $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^k$. Then the log-likelihood ratio process is the map*

$$LLRP : \boldsymbol{\theta} \to log\left(\frac{d\,\mathbb{P}_{\boldsymbol{\theta}}}{d\,\mathbb{P}_{\boldsymbol{\theta}_0}}(X)\right) = log\left(\frac{p_{\boldsymbol{\theta}}}{p_{\boldsymbol{\theta}_0}}(X)\right), \quad \forall \boldsymbol{\theta} \in \Theta, \tag{3.5}$$

where $\boldsymbol{\theta}_0 \in \Theta$ is a fixed reference parameter. In particular the random variable

$$log\left(\frac{p_{\boldsymbol{\theta}}}{p_{\boldsymbol{\theta}_0}}(X)\right)$$

takes values in $\mathbb{R}^k$ for all $\boldsymbol{\theta} \in \Theta$.

This is a process as it defines a statistic for all possible values of $\boldsymbol{\theta}$, and these statistics all take values in a real space with the same dimension as $\boldsymbol{\theta}$. Additionally, we will show that this statistic is sufficient, so it captures all the information about the parameters. Therefore, it can be used to compare models. Moreover, two models are equivalent if their corresponding log-likelihood ratio processes are equal.

Let us denote the statistic as

$$\text{LLR}_{\boldsymbol{\theta}}(X) := \log\left(\frac{p_{\boldsymbol{\theta}}}{p_{\boldsymbol{\theta}_0}}(X)\right)$$

From the definition of sufficiency, we require that

$$\mathbb{P}(X = x | LLR_{\boldsymbol{\theta}}(X) = f(\tau))$$

does not depend on $\boldsymbol{\theta}$, where $f(\boldsymbol{\theta})$ is a real-valued function of $\boldsymbol{\theta} \in \Theta$. However, from the definition of the statistic, we know that if $LLR_{\boldsymbol{\theta}}(X) = f(\boldsymbol{\theta})$, then

$$p_{\boldsymbol{\theta}} = p_{\boldsymbol{\theta}_0} e^{f(\boldsymbol{\theta})}.$$

Therefore, if $LLR_{\boldsymbol{\theta}}(x) = f(\boldsymbol{\theta})$

$$\begin{aligned}
\mathbb{P}_{\boldsymbol{\theta}}(X = x | LLR_{\boldsymbol{\theta}}(X) = f(\boldsymbol{\theta})) &= \frac{\mathbb{P}_{\boldsymbol{\theta}}(X = x)}{\mathbb{P}_{\boldsymbol{\theta}}(LLR_{\boldsymbol{\theta}}(X) = f(\boldsymbol{\theta}))} \\
&= \frac{p_{\boldsymbol{\theta}_0}(x) e^{f(\boldsymbol{\theta})}}{\int_{x:LLR_{\boldsymbol{\theta}}(x)=f(\boldsymbol{\theta})} p_{\boldsymbol{\theta}_0}(x) e^{f(\boldsymbol{\theta})}} \\
&= \frac{p_{\boldsymbol{\theta}_0}(x)}{\int_{x:LLR_{\boldsymbol{\theta}}(x)=f(\boldsymbol{\theta})} p_{\boldsymbol{\theta}_0}(x)},
\end{aligned}$$

which clearly no longer depends on $\boldsymbol{\theta}$, so $LLR_{\boldsymbol{\theta}}(X)$ is a sufficient statistic. This holds for all $\boldsymbol{\theta} \in \Theta$, so the process provides a sufficient statistic for any $\boldsymbol{\theta} \in \Theta$.

For example, consider again the coin toss: since $\Omega = \{0, 1\}$ with $0/1$ corresponding with Tails/Heads, we can write

$$p_q(X) = q^X (1 - q)^{1-X}.$$

The corresponding log-likelihood ratio process is then

$$\log\left(\frac{p_q}{p_{q_0}}(X)\right) = X \log\left(\frac{q}{q_0}\right) + (1 - X) \log\left(\frac{1 - q}{1 - q_0}\right).$$

Additionally, for an IID sample of coin tosses, we have

$$\begin{aligned}
\log\left(\frac{p_q^n}{p_{q_0}^n}(\boldsymbol{X})\right) &= \sum_{i=1}^{n} \left[X_i \log\left(\frac{q}{q_0}\right) + (1 - X_i) \log\left(\frac{1 - q}{1 - q_0}\right) \right] \\
&= n\bar{\boldsymbol{X}} \log\left(\frac{q}{q_0}\right) + (1 - n\bar{\boldsymbol{X}}) \log\left(\frac{1 - q}{1 - q_0}\right).
\end{aligned}$$

This is a one-to-one function of a sufficient statistic, so it is clearly still a sufficient statistic as we expect.

We now proceed to asymptotic estimation: in asymptotic estimation theory we consider an IID sample $X_1, ..., X_n \sim \mathbb{P}_{\boldsymbol{\theta}}$ as introduced in the previous section, but

now we're interested in the limit where we let the sample size n tend to infinity. Therefore, we now denote the random variable as $\boldsymbol{X}_n \sim P_{\boldsymbol{\theta}}^n$, a sample as $\boldsymbol{x}_n$ and an estimator as $\hat{\boldsymbol{\theta}}_n(\boldsymbol{X}_n)$ so that the size n is explicit at every stage. Since the FI scales linearly with the sample size, we can in theory estimate the parameters up to arbitrary accuracy by going to a sufficiently large n while utilizing powerful asymptotic results. Our aim is to perform this estimation in an optimal manner as described in section 3.1, i.e., we want to attain the CRB.

The first of asymptotic result is that we can reduce the problem to a *local estimation problem*: since we have access to a very large amount of data in the asymptotic limit, we can simply use a vanishingly small proportion of the data in some sub-optimal initial estimation producing an estimate $\boldsymbol{\theta_0}$. We can then write

$$X_i \sim \mathbb{P}_{\boldsymbol{\theta_0}+\boldsymbol{h}}$$

for some small parameter $\boldsymbol{h} \in \mathbb{R}^r$ that we now want to estimate. In our work this involves using $n^{1-\epsilon}$ of the samples in the initial estimation for some small $\epsilon > 0$. Crucially, the proportion $n^{1-\epsilon}$ vanishes relative to n in the asymptotic limit, and, with some reasonable assumptions on the initial estimator (see section 5.3.1), it allows us to localize the estimation to a region $|\boldsymbol{h}| \leq n^{-1/2+\epsilon}$. Additionally, since the proportion vanishes in the asymptotic limit it also does not effect the estimation performance. We then define a *local parameter* $\boldsymbol{u} := \boldsymbol{h}\sqrt{n}$ for convenience so that $|\boldsymbol{u}| \leq n^{\epsilon}$. Our task is now to estimate $\boldsymbol{u}$ using IID samples from a random variable $X \sim \mathbb{P}_{\boldsymbol{\theta}_0+\boldsymbol{u}/\sqrt{n}}$.

This also aids us in the construction of unbiased estimators as we can relax the unbiased condition and instead look for *locally unbiased estimators* $\hat{\boldsymbol{\theta}}_n$, which, as the name suggests, are unbiased around a fixed parameter value $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ (the value found during our sub-optimal initial estimation). This corresponds with the following conditions often used in the quantum metrology literature:

$$\sum_{\mathbf{x}_n \in \Omega^n} p_{\boldsymbol{\theta}_0}(\mathbf{x}_n)\hat{\boldsymbol{\theta}}_n(\mathbf{x}_n) = \boldsymbol{\theta}_0, \qquad \sum_{\mathbf{x}_n \in \Omega^n} \nabla|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} p_{\boldsymbol{\theta}}(\mathbf{x}_n)\hat{\boldsymbol{\theta}}_n(\mathbf{x}_n)^T = \mathbf{1}, \tag{3.6}$$

where ∇ denotes the gradient with respect to the parameters $\boldsymbol{\theta}$.

One can show that many common estimators are locally unbiased in the asymptotic limit, such as the maximum likelihood estimator (MLE) or the method of moments estimators [170]. In particular, we make use of the MLE in chapter 4, so it is worth covering briefly; this estimator is given by the parameters $\boldsymbol{\theta}$ that maximise the likelihood $L_{\boldsymbol{\theta}}(\boldsymbol{x}_n) := \prod_{i=1}^n p_{\boldsymbol{\theta}}(x_i)$ for a sample $\mathbf{x}$:

$$\hat{\boldsymbol{\theta}}(\boldsymbol{x}_n) := \text{argmax}_{\boldsymbol{\theta}}\{L_{\boldsymbol{\theta}}(\boldsymbol{x}_n)\}.$$

This is a very natural estimator to consider as if $L_{\boldsymbol{\theta}_1}(\boldsymbol{x}_n) > L_{\boldsymbol{\theta}_2}(\boldsymbol{x}_n)$ for a sample $\boldsymbol{x}_n \in \Omega^n$, then the parameters $\boldsymbol{\theta}_1$ were more likely to produce the observed sample $\mathbf{x}_n$ than the parameters $\boldsymbol{\theta}_2$. Additionally, the natural logarithm $\log(\cdot)$ is a non-decreasing function, therefore, maximising the likelihood is equivalent to maximising the log-likelihood

$$\ell_{\boldsymbol{\theta}}(\boldsymbol{x}_n) := \log(L_{\boldsymbol{\theta}}(\boldsymbol{x}_n)) = \sum_{i=1}^n \log(p_{\boldsymbol{\theta}}(x_i))$$

that we introduced previously. This converts a product into a sum, so is often much easier to implement in practice.

A key result in asymptotic statistics is LAN [170], which shows that the local model with probability distribution $\mathbb{P}^n_{\boldsymbol{\theta}_0+\boldsymbol{u}/\sqrt{n}}$ converges to a Gaussian shift model G with normal distribution $N(\boldsymbol{u}, I^{-1}[\boldsymbol{\theta}_0])$ in the asymptotic limit under some reasonable assumptions [88, 170]. This allows us to describe the complicated behaviour of an IID sample through a single normally distributed random variable. To formulate this, we first have to introduce the concept of convergence of statistical models [170], which can be defined in several ways. Two of these require a notion of distance between pairs of samples $(\boldsymbol{x}, \boldsymbol{y})$ - say the Euclidean distance $d(x,y) = ||x-y|| = \sqrt{\sum_i x_i^2 - y_i^2}$:

Definition 14 (Convergence). *A sequence of random variables X_n converge* in distribution *to a random variable X, denoted $X_n \overset{D}{\rightsquigarrow} X$, if*

$$\mathbb{P}(X_n \leq x) \to \mathbb{P}(X \leq x)$$

as $n \to \infty$ for every x at which the distribution function $x \mapsto \mathbb{P}(X \leq x)$ is continuous.

They converge in probability *to X, denoted $X_n \overset{P}{\rightsquigarrow} X$, if*

$$\mathbb{P}(d(X_n, X) > \epsilon) \to 0$$

as $n \to \infty$ for all $\epsilon > 0$.

Finally, they converge almost surely *to X, denoted $X_n \overset{a.s.}{\rightsquigarrow} X$, if*

$$\mathbb{P}(\lim_{n\to\infty} d(X_n, X) = 0) = 1.$$

Convergence in distribution is the most general form of convergence as it only depends on the distribution functions, not the probability spaces of the random variables. The two other forms of convergence only make sense if the random variables are defined on the same probability space; they do, however, provide a hierarchy of convergence with almost sure convergence implying convergence in probability, which in turn implies weak convergence.

Two useful theorems [3, 170] that rely on these ideas of convergence are the *law of large numbers* and the *central limit theorem* (CLT):

Theorem 11 (Law of large numbers). *Let $\boldsymbol{X}_n = (X_1, ..., X_n)$ be an IID random variable with $X_i \sim \mathbb{P}$. If $\mathbb{E}(X_1) < \infty$, then*

$$\bar{\boldsymbol{X}}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \overset{a.s.}{\rightsquigarrow} \mathbb{E}(X_1)$$

as $n \to \infty$, that is, the sample mean converges to the true mean almost surely.

This is the *strong* law of large numbers as it relies on almost sure convergence, although, the weak law of large numbers that relies on convergence in distribution is often more useful.

Theorem 12 (Central limit theorem). *Let $\boldsymbol{X}_n = (X_1, ..., X_n)$ be an IID random variable with $X_i \sim \mathbb{P}$. If $\mathbb{E}(X_1) < \infty$ and $\mathbb{E}(X_1^2) < \infty$, then*

$$\sqrt{n}(\bar{\boldsymbol{X}}_n - \mathbb{E}(X_1)) \overset{D}{\rightsquigarrow} N(0, Cov(X_1)),$$

that is, the rescaled and shifted sample mean converges to a normally distributed random variable with mean 0 and variance given by the covariance of X_1.

This theorem provides additional insight into unbiased estimators as well: we can consider an estimator $\hat{\boldsymbol{\theta}}(X)$ for $X \sim \mathbb{P}_{\boldsymbol{\theta}}$ as a random variable itself. The estimator $(\hat{\boldsymbol{\theta}}(X_1), \ldots, \hat{\boldsymbol{\theta}}(X_n))$ corresponding to an IID random variable $\boldsymbol{X}_n = (X_1, \ldots, X_n)$, $X_i \sim \mathbb{P}_{\boldsymbol{\theta}}$ is then itself an IID random variable. For an unbiased estimator, theorem 12 then guarantees that the mean estimate

$$\bar{\hat{\boldsymbol{\theta}}}_n(\boldsymbol{X}_n) := \frac{1}{n}\sum_{i=1}^{n} \hat{\boldsymbol{\theta}}(X_i)$$

converges to a normally distributed random variable around $\boldsymbol{\theta}$. That is, the estimator $\bar{\hat{\boldsymbol{\theta}}}(\boldsymbol{X}_n)$ will concentrate around $\boldsymbol{\theta}$.

3.1.1 *Local Asymptotic Normality*

We now proceed with a non-technical statement of LAN for IID random variables:

> **Local asymptotic normality**: Let $\mathcal{P}_n = \{\mathbb{P}^n_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^k\}$ be a sequence of statistical models for corresponding IID random variables $\boldsymbol{X}_n = (X_1, ..., X_n)$ with $X_i \sim \mathbb{P}_{\boldsymbol{\theta}}$, $\boldsymbol{\theta} = \boldsymbol{\theta}_0 + \boldsymbol{u}/\sqrt{n}$ for fixed $\boldsymbol{\theta}_0$ and $\boldsymbol{u} \in \mathbb{R}^k$. Under reasonable smoothness assumptions about the parametrisation map $\boldsymbol{\theta} \mapsto p_{\boldsymbol{\theta}}$, one can show that the models
>
> $$\mathbb{P}^n_{\boldsymbol{\theta}_0+\boldsymbol{u}/\sqrt{n}} \quad \text{and} \quad N(\boldsymbol{u}, I_{\theta_0}^{-1}),$$
>
> where I_{θ_0} is the CFI of $\mathbb{P}_{\theta_0}$, are statistically equivalent in the limit $n \to \infty$.

In this thesis we will not include a formal proof of this statement. Instead, we will outline the two formulations and the methods used to prove them, and discuss the nature of the equivalence of the two models. It can be demonstrated in two ways - either through the log-likelihood ratio process that we introduced in section 3.1 or through randomizations (the classical analogue of a quantum channel). It is the second of these methods that can be more readily generalized to quantum statistics, so it is of more relevance to us. We will, however, discuss both methods in the following pages.

The first method focuses on the convergence of the *log-likelihood ratio process* [170]

$$\log\left(\frac{d\,\mathbb{P}^n_{\boldsymbol{\theta}_0+\boldsymbol{u}/\sqrt{n}}}{d\,\mathbb{P}^n_{\boldsymbol{\theta}_0}}(\boldsymbol{X}_n)\right) = \log\left(\Pi_{i=1}^n \frac{p_{\boldsymbol{\theta}_0+\boldsymbol{u}/\sqrt{n}}}{p_{\boldsymbol{\theta}_0}}(X_i)\right) \tag{3.7}$$

to the log-likelihood ratio process of the Gaussian shift model $N(\boldsymbol{u}, I^{-1}[\boldsymbol{\theta}_0])$ under the condition that the first and second derivatives of the log-likelihood exist. The statement also holds under a weaker condition called differentiability in quadratic mean [170].

For an intuitive idea of this convergence [170], we will assume the weaker condition that both $\dot{\ell}_{\boldsymbol{\theta},i}$ and $\ddot{\ell}_{\boldsymbol{\theta},i}$ exist. We will also assume that the parameter is one-dimensional for simplicity. Provided this, one uses a Taylor expansion to obtain

$$\log\left(\Pi_{i=1}^n \frac{p_{\theta_0+\boldsymbol{u}/\sqrt{n}}}{p_{\theta_0}}(X_i)\right) = \sum_{i=1}^{n}\left[\frac{u}{\sqrt{n}}\dot{\ell}_{\theta_0}(X_i) + \frac{1}{2}\frac{u^2}{n}\ddot{\ell}_{\theta_0}(X_i)\right] + \mathrm{Rem}_n,$$

where the remainder term Rem_n is negligible in the limit $n \to \infty$. Recalling equations (3.2) and (3.3), $\sum_{i=1}^{n} n^{-1}\ddot{\ell}_{\theta_0}(X_i)$ converges to $-I(\theta_0)$ by Theorem 11. Additionally, $n^{-1/2}\sum_{i=1}^{n}\dot{\ell}_{\theta_0}(x_i)$ converges to a normally distributed random variable $\Delta_{\theta_0} \sim N(0, I(\theta_0))$ by Theorem 12. Therefore, in the asymptotic limit, we have

$$\log\left(\frac{d\,\mathbb{P}^n_{\boldsymbol{\theta}_0+\boldsymbol{u}/\sqrt{n}}}{d\,\mathbb{P}^n_{\boldsymbol{\theta}_0}}(\boldsymbol{X}_n)\right) \overset{D}{\rightsquigarrow} u\Delta_{\theta_0} - u^2\frac{I(\theta_0)}{2}. \tag{3.8}$$

For a normally distributed random variable $X \sim N(u, I(\theta_0)^{-1})$, the corresponding probability density is

$$p_u(X) = \exp(\frac{-(X-u)^2 I(\theta_0)}{2}),$$

so the log-likelihood ratio process is

$$\begin{aligned}\log\left(\frac{p_u}{p_0}(X)\right) &= \log\left[\exp(\frac{-(X-u)^2 I(\theta_0) + X^2 I(\theta_0)}{2})\right] \\ &= uXI(\theta_0) - u^2\frac{I(\theta_0)}{2} \\ &= u\Delta_{\theta_0} - u^2\frac{I(\theta_0)}{2},\end{aligned} \tag{3.9}$$

where we use that $\text{Var(aX)} = a^2\text{Var(X)}$, $\mathbb{E}(aX) = a\mathbb{E}(X)$. Therefore, we have shown that equation 3.8 and equation 3.9 are identical in the asymptotic limit. Furthermore, since the ratio process is a sufficient statistic, the convergence implies both distributions capture all available information about the parameters and it is in this sense that the two models $(\mathbb{P}^n_{\theta_0=h/\sqrt{n}})$ and $N(u, I(\theta_0)^{-1})$ are statistically equivalent. Since this method is based on the convergence in distribution (weak) of the log-likelihood process, the statement is usually referred to as weak local asymptotic normality.

The second method for proving LAN is through randomizations [88]:

Definition 15 (Randomization)**.** *A randomization is a positive linear map*

$$R : L^1(\Omega_1, \Sigma_1, \mu_1) \to L^1(\Omega_2, \Sigma_2, \mu_2),$$

where $L^1(\Omega, \Sigma, \mu)$ is the space of absolutely integrable functions on Ω.

Informally, they represent maps that transform outcomes from one random variable X_1 with probability space $(\Omega_1, \Sigma_1, \mu_1)$ into outcomes from another random variable X_2 with probability space $(\Omega_2, \Sigma_2, \mu_2)$. This can be implemented by means of a statistic, or a randomised statistic where one makes use of an additional random variable with a given distribution. Indeed if $f : \Omega_1 \to \Omega_2$ is a statistic, then this induces a map $\tilde{R}_f : \mathbb{P}_1 \mapsto \mathbb{P}_2$ on the level of probability distributions. Assuming that $\mathbb{P}_i$ is absolutely continuous with respect to the measure μ_i, this in turn gives rise to the randomisation $R_f : p_1 \mapsto p_2$ where p_i is the probability density of $\mathbb{P}_i$ with respect to the measures μ_i.

Definition 16 (Statistical equivalence of models)**.** *Let $\mathcal{P}_1 = \{\mathbb{P}^{(1)}_\theta : \theta \in \Theta\}$ and $\mathcal{P}_2 = \{\mathbb{P}^{(2)}_\theta : \theta \in \Theta\}$ be statistical models where $\mathbb{P}^{(i)}_\theta$ are probability distributions*

on (Ω_i, Σ_i)*, and assume that* $\mathbb{P}_\theta^{(i)}$ *is absolutely continuous with respect to the measure* μ_i *with density* $p_\theta^{(i)}$*, for* $i = 1, 2$*. The models are called statistically equivalent if there exists randomizations* $R : L^1(\Omega_1, \Sigma_1, \mu_1) \to L^1(\Omega_2, \Sigma_2, \mu_2)$ *and* $S : L^1(\Omega_2, \Sigma_2, \mu_2) \to L^1(\Omega_1, \Sigma_1, \mu_1)$ *such that*

$$R(p_\theta^{(1)}) = p_\theta^{(2)}, \qquad S(p_\theta^{(2)}) = p_\theta^{(1)}$$

for all $\theta \in \Theta$*.*

To compare the two models in this context we use the deficiency, or, more accurately, the Le Cam distance:

Definition 17 (Deficiency and Le Cam distance)**.** *Let* $\mathcal{P}_1 = \{\mathbb{P}_\theta^{(1)} : \theta \in \Theta\}$ *and* $\mathcal{P}_2 = \{\mathbb{P}_\theta^{(2)} : \theta \in \Theta\}$ *be two statistical models, and assume that* $\mathbb{P}_\theta^{(i)}$ *are absolutely continuous with respect to reference distributions* $\mu^{(i)}$ *with densities* $p_\theta^{(i)}$*. The deficiency of* $\mathcal{P}_1$ *with respect to* $\mathcal{P}_2$ *is*

$$\delta(\mathcal{P}_1, \mathcal{P}_2) := \inf_R \sup_\theta ||R(p_\theta^{(1)}) - p_\theta^{(2)}||_1,$$

where the infimum is taken over all randomizations R*.*

The Le Cam distance between $\mathcal{P}_1$ *and* $\mathcal{P}_2$ *is simply*

$$\Delta(\mathcal{P}_1, \mathcal{P}_2) := max(\delta(\mathcal{P}_1, \mathcal{P}_2), \delta(\mathcal{P}_2, \mathcal{P}_1)).$$

An illustrative example is very useful here [89]: consider the coin toss again, but now let $X_i \sim \mathbb{P}_{0.5+u.n^{-1/2}}$. So we want to estimate some small bias u in the coin. LAN predicts that this is statistically equivalent to a random variable with normal distribution $N(u, I_{0.5}^{-1})$. We've already seen that the sample mean $\bar{\boldsymbol{X}}_n$ is a sufficient statistic, and, indeed, we know that $\sqrt{n}\bar{\boldsymbol{X}}_n$ converges to a normally distributed random variable $N(u, I_{0.5}^{-1})$ by Theorem 12. But how do we formalize this in terms of a randomization? For sample size n, the sample mean has corresponding sample space $\Omega = \{\frac{1}{n}, \frac{2}{n}, ..., \frac{n}{n}\}$. This only includes rational values between $\frac{1}{n}$ and 1, whereas the distribution that we want to map it onto is a continuous Gaussian. Therefore, the randomization $R(\cdot)$ takes this sufficient statistic and adds a Gaussian with small variance to obtain something with density with respect to the Lebesgue measure on $\mathbb{R}$. In this manner we can demonstrate convergence with respect to the Le Cam distance.

The strong version of LAN [170] states that the Le Cam distance between the sequence of IID statistical models $\mathcal{P}_n = \{\mathbb{P}^n_{\theta_0 + n^{-1/2}\mathbf{u}} : \|\mathbf{u}\| \leq n^\epsilon\}$ and the sequence of Gaussian shift models $G_n = \{N(u, I_{\theta_0}^{-1}) : \|\mathbf{u}\| \leq n^\epsilon\}$ converges to 0 in the limit of large n. The parameter ϵ is a small constant $\epsilon > 0$ that bounds the local parameter $\boldsymbol{u}$ for fixed n; a technical detail that is necessary for the proof. Since this holds uniformly over local parameters $\|\mathbf{u}\| \leq n^\epsilon$, this result can be used to establish asymptotically optimal estimation procedures and convergence rates; this involves a two step procedure where a small sample $n^{1-\epsilon}$ is used to localise the parameter, while the remaining data is mapped via an appropriate randomisation into a model that is approximately Gaussian, whose mean is then estimated in a simple way. The full procedure goes beyond the scope of the thesis but this philosophy will be applied for devising optimal estimation procedures for pure states in chapter 5.

3.2 QUANTUM ESTIMATION

In quantum parameter estimation [49, 130, 151, 152] we assume the parameter of interest $\boldsymbol{\theta}$ is encoded in a quantum state, that is we are given a system prepared in a state $\rho = \rho_{\boldsymbol{\theta}}$ depending on the unknown parameter $\boldsymbol{\theta}$. Our task is to measure this state to extract information about the parameter. In particular, the analogous setup to a classical statistical model is the quantum statistical model:

Definition 18 (Quantum statistical model)**.** *Given a quantum system described by a Hilbert space $\mathcal{H}$, a quantum statistical model $\mathcal{Q}$ is a family of states in $S(\mathcal{H})$ labelled by m real parameters $\boldsymbol{\theta} \in \Theta \subset \mathbf{R}^m$:*

$$\mathcal{Q} = \{\rho_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta\},$$

where the parametrization map $\boldsymbol{\theta} \to \rho_{\boldsymbol{\theta}}$ is injective.

The key aspect of the quantum estimation problem is the choice of the measurement. By performing a measurement with POVM $\mathcal{M} = \{M_i : i = 1, \ldots, k\}$, we obtain a random outcome $X \in \{1, \ldots, k\}$ with probability distribution $p_{\boldsymbol{\theta}}$

$$p_{\boldsymbol{\theta}}(i) := \mathrm{Tr}(\rho_{\boldsymbol{\theta}} M_i).$$

Therefore, once we have specified the measurement we are back in the realm of classical estimation. In particular, the choice of measurement has classical FI

$$I_{ij}(\rho_{\boldsymbol{\theta}}; \mathcal{M}) = \sum_m p_{\boldsymbol{\theta}}(m)\, \partial_{\theta_i} log p_{\boldsymbol{\theta}}(m)\, \partial_{\theta_j} log p_{\boldsymbol{\theta}}(m),$$

where we have explicitly highlighted that the FI depends on both the state $\rho_{\boldsymbol{\theta}}$ and the measurement $\mathcal{M}$. The Cramér-Rao bound then provides the ultimate limits on any unbiased estimator $\hat{\boldsymbol{\theta}}$ based off this measurement as introduced in section 3.1.

For pure state models we often exploit a state's global phase to simplify calculations. For instance, we often assume $\langle \psi | \dot{\psi} \rangle = 0$ in this thesis, where $|\dot{\psi}\rangle$ again denotes the derivative of $|\psi\rangle$ with respect to a parameter θ. This can always be achieved by replacing the state $|\psi\rangle$ with $|\tilde{\psi}\rangle = e^{i\varphi(\theta)} |\psi\rangle$, where $e^{i\varphi(\theta)}$ is a parameter dependant phase such that $\dot{\varphi} = i \langle \psi | \dot{\psi} \rangle$.

If we have access to multiple copies of the state, say $\rho_{\boldsymbol{\theta}}^{\otimes n}$, then we can derive the equivalent to an IID sample: the quantum statistical model is $\mathcal{Q}^n = \{\rho_{\boldsymbol{\theta}}^{\otimes n} : \boldsymbol{\theta} \in \Theta\}$, and the same POVM should be performed on each copy of the state so that the collective POVM is of the form

$$\mathcal{M}_n = \{M_{i_1,\ldots,i_n} := M_{i_1} \otimes \ldots \otimes M_{i_n} : (i_1, \ldots, i_n) \in \Omega^n\}.$$

We can then construct an unbiased estimator $\hat{\boldsymbol{\theta}}(\boldsymbol{i})$ for outcomes $\boldsymbol{i} := (i_1, \ldots, i_n)$ that is limited by the Cramér-Rao bound with FI $I(\rho_{\boldsymbol{\theta}}^{\otimes n}; \mathcal{M}_n) = nI(\rho_{\boldsymbol{\theta}}; \mathcal{M})$, so we achieve the same linear scaling as we expect for an IID sample.

There are, however, much more general measurements we could implement in this scenario. We could, for instance, perform a different POVM on each copy of the state; this is closely aligned with the measurement scheme that we use in chapter 4, where we also allow the results of previous measurement to

inform the next POVM. The most general measurement scheme though will also include highly entangled POVMs that cannot be expressed as a tensor product on individual copies. This means that they correspond to a collective measurement on multiple copies of the state $\rho_{\boldsymbol{\theta}}$ at once. Additionally, this is equally true for a system with multiple subsystems instead: the optimal POVM may involve collective measurements on multiple subsystems at once instead of measurements on each individual subsystem. Therefore, measurements may be highly entangled even when we have a single copy of the state - a sharp contrast to what is possible in classical estimation.

In order to identify the optimal measurement scheme we need to optimize over this more complex space of POVMs; the result is the quantum Cramér-Rao bound [49, 151]:

Theorem 13 (Quantum Cramér-Rao Bound)**.** *For a quantum statistical model $\mathcal{Q} = \{\rho_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta\}$ with $\Theta \subset \mathbb{R}^k$, the covariance of an unbiased estimator $\hat{\boldsymbol{\theta}}$ can be bounded by*

$$Cov(\hat{\boldsymbol{\theta}}) \geq F^{-1}(\rho_{\boldsymbol{\theta}}),$$

where

$$F_{ij}(\rho_{\boldsymbol{\theta}}) := \frac{1}{2} Tr(\rho_{\boldsymbol{\theta}} \{\mathcal{L}_{\theta_i}, \mathcal{L}_{\theta_j}\}) \tag{3.10}$$

is the quantum Fisher information (QFI) of $\mathcal{Q}_{\boldsymbol{\theta}}$ and the operators $\mathcal{L}_{\theta_i}$, $i = 1, ..., k$, are the symmetric logarithmic derivatives (SLDs) defined by the equations

$$\partial_{\theta_i} \rho_{\boldsymbol{\theta}} = \frac{1}{2} \{\rho_{\boldsymbol{\theta}}, \mathcal{L}_{\theta_i}\} \tag{3.11}$$

where $\{A, B\} = AB + BA$ is the anti-commutator.

A proof of the quantum Cramér-Rao bound for a single parameter can be seen in section 5.2, where we pay close attention to the necessary conditions for when it is attainable.

The SLD is the quantum analogue of the score function; it also provides an optimal estimator [55] for local estimation of a one-dimensional parameter $\theta \in \mathbb{R}$:

$$\hat{\theta}(X) = \hat{\theta}_0 + \frac{1}{F(\rho_{\theta_0})} X$$

where X is the outcome of measuring observable $\mathcal{L}_{\theta_0}$ and $F(\rho_{\theta_0})$ is the QFI of state ρ_{θ_0}. Since it depends on the parameters of interest, we are once again reliant on asymptotic techniques, where we first to localise the parameter around an initial estimate $\hat{\theta}_0$. Proving that this estimator is optimal then follows the exact same argument as we saw for equation (3.4).

For a pure state $\rho_{\boldsymbol{\theta}} = |\psi_{\boldsymbol{\theta}}\rangle \langle\psi_{\boldsymbol{\theta}}|$, calculating the SLDs is a simple affair: pure states satisfy $\rho^2 = \rho$, so, taking the derivative with respect to a parameter θ_i, we find

$$\partial_{\theta_i} \rho_{\boldsymbol{\theta}}^2 = (\partial_{\theta_i} \rho_{\boldsymbol{\theta}}) \rho_{\boldsymbol{\theta}} + \rho_{\boldsymbol{\theta}} (\partial_{\theta_i} \rho_{\boldsymbol{\theta}}).$$

Comparing this to equation 3.11, we can infer that $\mathcal{L}_i = 2(\partial_{\theta_i} \rho)$. However, calculating the SLD for arbitrary states is difficult and often relies on numerical methods.

There are a few problems with the SLDs: firstly, the SLDs are generally complex observables that contain entanglement between each subsystem of a multipartite system. For instance, if the system consists of n qubits, the SLD generally involves a collective measurement on all the qubits at once. While this may theoretically saturate the QCRB, it may not be practically implementable in an experimental quantum device as many modern devices still have severe limitations in the implementable operations. Or often the subsystems are separated by a considerable distance like in satellite-based optical interferometers, so we can only interact with them locally at each satellite. Secondly, for the multi-parameter scenarios each SLD is still the optimal observable to measure for each parameter, but these SLDs may not commute with each other. Therefore, we cannot necessarily simultaneously measure the SLDs; the consequence of this is that the QCRB may not be achievable in the multi-parameter scenario and so is not tight.

This first point motivates an interest in alternate measurement schemes [72, 73, 182] that still saturate the QCRB, but use a simplified class of measurements. Since we want to address issues around measurement complexity in multipartite systems, a natural choice is measuring each subsystem separately. The overall POVM can then be expressed as a tensor product of POVMs for each subsystem, say

$$\mathcal{M} = \{M_{i_1} \otimes M_{i_2}\}$$

for a two-partite quantum system. This involves only local operations performed on each subsystem, so seems to solve the issue with the SLDs. However, we can utilise some classical information to improve this: if we measure each subsystem individually, i.e., one at a time, then we can use the result of previous measurements to fine-tune the next choice of measurement. The POVM then has the form

$$M_{i_1} \otimes M_{i_1,i_2} \otimes M_{i_1,i_2,i_3} \otimes ... \otimes M_{\boldsymbol{i}_n},$$

where $\boldsymbol{i}_n = (i_1, ..., i_n)$. This extra step involves only some classical communication, so this new measurement strategy is referred to as a local operations and classical communication (LOCC) measurement scheme. In particular, it has been demonstrated [182] that this scheme can saturate the QCRB.

Since the QCRB is not achievable, one focuses instead on a *cost* - a type of MSE given by

$$\mathrm{Tr}(W\mathrm{Cov}(\hat{\boldsymbol{\theta}})),$$

where W is a positive weight matrix. This weight matrix effectively allows us to specify how important each parameter is to us. The cost is then a single number, so replacing the covariance matrix with costs converts matrix inequalities into scalar inequalities that can be optimized. The main bound that addresses the issues around achievability is the Holevo bound [49, 95]:

Theorem 14 (Holevo Bound). *For a quantum statistical model $\mathcal{Q} = \{\rho_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta\}$ with $\Theta \subset \mathbb{R}^k$, the cost $\mathcal{C}$ of a locally unbiased estimator $\hat{\boldsymbol{\theta}}$ can be bounded by*

$$\mathcal{C} = \mathrm{Tr}(W \cdot Cov(\hat{\boldsymbol{\theta}})) \geq \min_{\boldsymbol{X},V} \left(\mathrm{Tr}(WV) : V \geq Z[\boldsymbol{X}], \quad \mathrm{Tr}(\nabla\rho_{\boldsymbol{\theta}}\boldsymbol{X}^T = \boldsymbol{1})\right),$$

where $\boldsymbol{X} = [X_1, ..., X_k]^T$ is a vector of Hermitian operators $X_i \in \mathcal{B}(\mathcal{H})$, V is a $k \times k$ real matrix, $Z[\boldsymbol{X}] := \mathrm{Tr}(\rho_{\boldsymbol{\theta}}\boldsymbol{X}\boldsymbol{X}^T)$ is a $k \times k$ complex matrix and W is a real, symmetric cost matrix.

In general $Z[\boldsymbol{X}]$ is a complex matrix, but taking the trace of its product with a real matrix W results in some information being lost. Therefore, this final optimization over $V \geq Z[\boldsymbol{X}]$ is done to ensure that this information is captured as well.

This introduces the main field of quantum parameter estimation. Within this field there are many subfields that largely depend on the application of quantum parameter estimation and how the parameters are encoded into the quantum state. We will now introduce a selection of them that were prevalent throughout this thesis.

The main field relevant to this thesis is quantum metrology, in which we want to use quantum systems to measure physical quantities with high precision. To this end, we begin by preparing our system, which we will refer to as a probe, in some initial state ρ_0. The parameters we want to estimate are then encoded into the probe's state through the state's evolution described by unitary evolution or, more generally, by a quantum channel $T \in \mathcal{CPTP}(\mathcal{H})$. An example of this is a Hamitonian identification problem [129, 151], where the probe evolves due to a system's Hamiltonian $H(\theta)$ that is dependant on the parameter θ that we're interested in. We then have

$$\rho_\theta(t) = e^{-iH(\theta)t}\rho_0 e^{iH(\theta)t},$$

where t is some known fixed time. Once the parameters are encoded, we're in the realm of quantum parameter estimation. Therefore, we can use its techniques, optimising the estimator and measurement in order to saturate the QCRB or Holevo bound in the multiparameter case.

To see another advantage of quantum metrology, consider a scenario with multiple probes. If we initialise each probe in the same initial state ρ_0 and assume the operation encoding the parameters is a unitary operator $U_{\boldsymbol{\theta}}$, then the final state is

$$\rho_{\boldsymbol{\theta}}^{\text{IID}} = \rho_{\boldsymbol{\theta}}^{\otimes n}, \tag{3.12}$$

where $\rho_{\boldsymbol{\theta}} = U_{\boldsymbol{\theta}}\rho_0 U_{\boldsymbol{\theta}}^*$. This provides us with the quantum analogue of an IID sample since each probe is in the same state $\rho_{\boldsymbol{\theta}}$. We can, however, find better initial states than this: if each probe has Hilbert space $\mathcal{H}$, then the combined Hilbert space is $\mathcal{H}^{\otimes n}$. A general initial state is then any $\rho_0 \in \mathcal{S}(\mathcal{H}^{\otimes n})$, which allows for entanglement between each system. This provides another aspect that we can optimize over, enabling strategies that beat all available separable strategies. The corresponding final state is then

$$\rho_{\boldsymbol{\theta}}^{\text{gen}} = U_{\boldsymbol{\theta}}^{\otimes n}\rho_0 U_{\boldsymbol{\theta}}^{\otimes n*}$$

We are now back in the realm of quantum parameter estimation and can focus on optimizing the estimator and choice of measurement as normal.

Note: the setup described in the previous paragraph is known as a parallel setup since all the probes have the parameters encoded at the same time. There is, however, an equivalent sequential scheme [73, 184] that utilizes a single probe with some ancillas, n encoding steps and control operations between each encoding step.

One of the main goals in quantum metrology is to design strategies that achieve the so-called Heisenberg limit (HL) scaling of the Fisher information [41, 50,

72, 73]: $F(\rho_{\boldsymbol{\theta}}^{\text{gen}}) \propto n^2$. This represents a quadratic improvement in the Fisher information over an IID strategy, which is limited by the standard quantum limit (SQL): $F(\rho_{\boldsymbol{\theta}}^{\text{IID}}) \propto n$. Crucially, this requires inherently quantum resources in the probe state [72, 73], so highlights the advantage of using quantum systems.

For a simple demonstration of this HL scaling, consider a phase estimation problem where we want to estimate a phase $\phi \in \mathbb{R}$ that is encoded in a qubit using the unitary $U_\phi = \exp(-i\phi\sigma_z)$. If we initialise a qubit in the state $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$, then an IID scheme produces output state

$$|\Psi_\phi\rangle = \left(\frac{1}{\sqrt{2}}\right)^n (e^{-i\phi}|0\rangle + e^{i\phi}|1\rangle)^{\otimes n}.$$

For pure states, one can show that the QFI is

$$F(|\Psi_\phi\rangle\langle\Psi_\phi|) = 4\langle\dot{\Psi}_\phi|\dot{\Psi}_\phi\rangle$$

when $\langle\Psi_\phi|\dot{\Psi}_\phi\rangle = 0$. This holds in this case, so we have

$$F(|\Psi_\phi\rangle) = 4n.$$

This QFI has the SQL scaling, so any estimation scheme is restricted to SQL scaling. If we use a GHZ state $|GHZ\rangle = (|0\rangle^{\otimes n} + |1\rangle^{\otimes n})/\sqrt{2}$ as our initial state instead, we find

$$\tilde{\Psi}_\phi = \frac{1}{\sqrt{2}}\left(e^{-in\phi}|0\rangle^{\otimes n} + e^{in\phi}|1\rangle^{\otimes n}\right)$$

and this state has corresponding QFI

$$F(|\Psi_\phi\rangle) = 4n^2.$$

Therefore, a phase estimation scheme that utilizes a GHZ initial state can achieve the HL scaling instead.

Quantum metrology encompasses most experimental frameworks, so its applications are widespread. When the emphasis is more about measuring physical quantities using quantum systems and less focused on the estimation sensitivity, it is often referred to as quantum sensing [47]. One of the most active areas of metrology is quantum magnetometry [23, 28, 29, 101, 159], where we want to measure a magnetic field. Even within magnetometry, there are several competing devices in development from N-V centres in diamond [159] to ensembles of atoms [23] to optical setups [28]. Other fields of quantum metrology include quantum-enhanced imaging [4], thermometry [46, 123], clocks [47, 50] and gravitational wave detection [1, 40, 168], many of which have already been implemented in experiments [39, 101, 125, 168].

Unfortunately, the quantum resources that the HL relies upon are very susceptible to noise: when we introduce an effect such as decoherence to the state $\rho_{\boldsymbol{\theta}}$ the entanglement between different subsystems can be quickly destroyed, which, in turn, destroys the HL scaling of the CFI. Therefore, there is considerable interest in applying quantum error correction to a metrological setup [79, 107, 169, 181, 184], allowing this noise to be identified and corrected to restore the HL scaling.

The work in this thesis can be considered quantum metrology, however, the quantum Markov chain framework does not facilitate HL scaling. Since we

estimate parameters from the output of a single QMC, we do not have access to the necessary entanglement between multiple probes. We also don't assume that we can optimise the initial state since we rely on the primitivity of the QMC's corresponding channel. It does, however, model many practically implementable experimental setups from atom masers [91] to quantum collision models [42] to discrete-time optical detectors [10, 61, 171]. Therefore, the motivation of the estimation schemes that we present is that they are practically implementable. Part of this is that we don't need to initialize the probes in some complex initial state, which is a very difficult task in its own right. Additionally, a QMC does not require us to specify the available resources before the experiment. Instead, we can keep monitoring the environment until we achieve some desired tolerance level.

Overall, the breadth of applications discussed here highlights the importance of quantum parameter estimation. Measurement - or, more accurately, estimation - is a very fundamental problem, so improving its efficiency causes many knock-on improvements elsewhere. In particular, these improvements may allow us to measure phenomena that were undetectable until now such as gravitational wave detection [1, 40, 168].

3.2.1 *Quantum Local Asymptotic Normality*

In this section we introduce the quantum analogue of local asymptotic normality (LAN). As we saw in section 3.1.1, LAN demonstrates that a sequence of local statistical models, $\mathcal{P}_n = \{\mathbb{P}^n_{\boldsymbol{\theta}_0+\boldsymbol{u}/\sqrt{n}} : \boldsymbol{u} \in \mathbb{R}^k, \text{fixed } \boldsymbol{\theta}_0 \in \mathbb{R}^k\}$, converges to a Gaussian shift model $G = \{N(\boldsymbol{u}, I^{-1}_{\boldsymbol{\theta}_0}) : \boldsymbol{u} \in \mathbb{R}^k\}$ in the limit $n \to \infty$. And this convergence could be proven through two different methods: strong convergence via randomizations or the weak convergence of the log-likelihood ratio process. Its analogue, quantum local asymptotic normality (QLAN), demonstrates that a sequence of local IID quantum statistical models

$$\mathcal{Q}_n = \{\rho^{\otimes n}_{\boldsymbol{\theta}_0+\boldsymbol{u}/\sqrt{n}} : \text{fixed } \boldsymbol{\theta}_0 \in \Theta \subseteq \mathbb{R}^{2k}, \boldsymbol{u} \in \mathbb{R}^{2k}\},$$

converges to a continuous variable Gaussian state model $\mathcal{G} = \{\phi_u : u \in \mathbb{R}^k\}$, where we have highlighted that the Gaussian state ϕ_u has k-modes for a $2k$-dimensional parameter $\boldsymbol{u} \in \mathbb{R}^{2k}$.

There is a general statement of QLAN for arbitrary finite-dimensional mixed states [88, 89, 102], but there is also a similar statement specific to quantum Markov chains (QMCs) [37, 83, 84, 86]. We will focus on IID pure state models for simplicity, which ensure we can present the quantum analogue of both weak and strong convergence. Weak convergence cannot be generalised for arbitrary mixed states, so focusing on IID pure states allows us to draw more parallels the theory of classical LAN, which we introduced in section 3.1.1.

For example, consider an pure qubit model [83] with

$$|\psi_{\boldsymbol{\theta}}\rangle := \exp\left(\frac{i}{\sqrt{2}}(\theta_2\sigma_x - \theta_1\sigma_y)\right)|0\rangle, \quad \boldsymbol{\theta} = (\theta_1, \theta_2) \in \mathbb{R}^2.$$

In terms of a qubit's Bloch sphere, the $|0\rangle$ state corresponds with the sphere's north pole, i.e., an extremal point on the σ_z-axis with Bloch vector $\boldsymbol{r} = (0, 0, 1)$.

This state corresponds with a rotation of this point around the σ_x and σ_y axes by angles $\theta_1/\sqrt{2}$ and $\theta_2/\sqrt{2}$ respectively, where this factor of $\sqrt{2}$ has been chosen for later convenience. We consider a local model again, so $\boldsymbol{\theta} = \boldsymbol{\theta}_0 + \boldsymbol{u} \cdot n^{-1/2}$ for local parameter $\boldsymbol{u} \in \mathbb{R}^2$ that we want to estimate. We fix $\boldsymbol{\theta} = (0,0)$ so that $\boldsymbol{u} = (0,0)$ corresponds with the $|0\rangle$ state. This could be achieved by reversing the rotation due to the known parameter $\boldsymbol{\theta}_0$, so this results in no loss of generality. The corresponding IID pure state model is then

$$\mathcal{Q}_n \coloneqq \{|\psi_{\boldsymbol{u}}^n\rangle = |\psi_{\boldsymbol{u}/\sqrt{n}}\rangle^{\otimes n} : \boldsymbol{u} \in \mathbb{R}^2\}$$

And the corresponding Gaussian state model is

$$\mathcal{G} = \{|u_1 + iu_2\rangle : \boldsymbol{u} \in \mathbb{R}^2\},$$

i.e., a one-mode coherent state $|\alpha\rangle$ with $u_1 = \mathrm{Re}(\alpha)$ and $u_2 = \mathrm{Im}(\alpha)$.

To see this convergence, consider the collective spin observables

$$S_x(n) \coloneqq \frac{1}{\sqrt{2n}} \sum_{i=1}^{n} \sigma_x^{(i)}$$

and

$$S_y(n) \coloneqq \frac{1}{\sqrt{2n}} \sum_{i=1}^{n} \sigma_y^{(i)}$$

where $\sigma_j^{(i)}$ is the Pauli operator applied to the ith qubit with the identity operator applied elsewhere.

One can show through the quantum equivalent of the central limit theorem [135] that the distribution of these observables converge in distribution to normal distributions

$$S_x(n) \overset{D}{\rightsquigarrow} N(u_1, \frac{1}{2}),$$
$$S_y(n) \overset{D}{\rightsquigarrow} N(u_2, \frac{1}{2}).$$

This is the expected distribution of canonical coordinates Q and P with respect to a coherent state $|u_1 + iu_2\rangle$, so this effectively shows that $\mathcal{Q}_n$ can be approximated by the coherent state model $\mathcal{G}$ as expected.

Another way to capture this convergence between the two models $|\psi_{\boldsymbol{u}}^n\rangle$ and $|\boldsymbol{u}\rangle$ is through an inner product:

$$\lim_{n\to\infty} \langle \psi_{\boldsymbol{u}}^n | \psi_{\boldsymbol{v}}^n \rangle = \langle \boldsymbol{u} | \boldsymbol{v} \rangle .$$

This is equivalent to the weak convergence in classical LAN. However, it can only be defined for pure states, so it is less general than in LAN.

To define the stronger form of equivalence, we have to introduce the quantum equivalent of the Le Cam distance, which replaces randomizations with quantum channels:

Definition 19 (Quantum Le Cam distance)**.** *Let* $\mathcal{Q}_1 = \{\rho_{\boldsymbol{\theta}}^{(1)} \in \mathcal{H}_1 : \boldsymbol{\theta} \in \Theta\}$ *and* $\mathcal{Q}_2 = \{\rho_{\boldsymbol{\theta}}^{(2)} \in \mathcal{H}_2 : \boldsymbol{\theta} \in \Theta\}$ *be two quantum statistical models for* $\boldsymbol{\theta} \in \Theta$*. The deficiency of* $\mathcal{Q}_1$ *with respect to* $\mathcal{Q}_2$ *is defined as*

$$\delta(\mathcal{Q}_1, \mathcal{Q}_2) = \inf_T \sup_{\boldsymbol{\theta}\in\Theta} ||\rho_{\boldsymbol{\theta}}^{(1)} - T(\rho_{\boldsymbol{\theta}}^{(2)})||_1,$$

where the infimum is taken over all quantum channels $T : \mathcal{H}_2 \to \mathcal{H}_1$. The Le Cam distance between $\mathcal{Q}_1$ and $\mathcal{Q}_2$ is then

$$\Delta(\mathcal{Q}_1, \mathcal{Q}_2) := max(\delta(\mathcal{Q}_1, \mathcal{Q}_2), \delta(\mathcal{Q}_2, \mathcal{Q}_1)).$$

The strong theory of QLAN then demonstrates that there exists two series of quantum channels T_n and S_n such that

$$\lim_{n\to\infty} \sup_{||\boldsymbol{u}||\leq C} ||\phi_{\boldsymbol{u}} - T_n(\rho_{\boldsymbol{u}}^n)||_1 \overset{D}{\rightsquigarrow} 0, \tag{3.13}$$

$$\lim_{n\to\infty} \sup_{||\boldsymbol{u}||\leq C} ||S_n(\phi_{\boldsymbol{u}}) - \rho_{\boldsymbol{u}}^n||_1 \overset{D}{\rightsquigarrow} 0, \tag{3.14}$$

where $||\cdot||_1$ is once again the trace norm and C is a suitable limit on $\boldsymbol{u}$, which we often take as n^ϵ for a small $\epsilon > 0$. That is, there exists channels that can map the Gaussian state into the IID quantum statistical model for all suitable $\boldsymbol{u}$ and vice versa.

Part II

RESULTS

In this part we present the main results of this thesis. This comprises of three optimal estimation schemes, two of which apply to quantum Markov chains (QMC). The first, see Chapter 4, comprises of an adaptive measurement scheme applied to a QMC. This scheme utilizes local measurements on the QMC's output that we tailor based on the results of previous measurements. In particular, we demonstrate that this scheme is optimal through numerical simulations. Additionally, this chapter introduces the absorber for the first time, which we use to post-process the QMC's output before measurement. This post-processing is key to both this estimation scheme and the scheme presented in chapter 6.

Chapter 5 focuses on a different measurement scheme - displaced null measurements. A null measurement refers to measuring a system using an orthonormal basis containing the system's state itself. Operationally, we cannot implement a true null measurement since it requires knowledge of the state's parameter θ beforehand. Instead, we have to measure an approximate null measurement with θ_0 close to θ. This is expected to be optimal, but in this chapter we demonstrate that this is not the case using independent identically prepared pure state models. In particular, the method through which one specifies the initial estimate leads to an identification problem when we perform the null measurement. We then present the displaced null measurement - our solution to this problem. This involves displacing the parameter $\theta_0 \to \theta_0 + \tau$ used in the null measurement to remove this identification problem.

Chapter 6 applies this null measurement scheme to a QMC. In particular, the absorber is used here to implement the displaced null measurement. We demonstrate that the expected output is a vacuum state when $\theta_0 = \theta$. A difference in these parameters $|\theta_0 - \theta| > 0$ then results in binary excitation patterns in the output of the QMC. We develop the theory behind these excitation patterns and demonstrate that we can calculate the rate of each pattern in the asymptotic limit. This is then used to design an optimal estimator, which relies on the counts of each pattern.

4

ADAPTIVE MEASUREMENT FILTER: EFFICIENT STRATEGY FOR OPTIMAL ESTIMATION OF QUANTUM MARKOV CHAINS

4.1 INTRODUCTION

The quantum input-output formalism is an effective framework for describing the evolution, monitoring and control of Markovian quantum open systems [43, 66, 174]. In this setting, the interaction with the environment is modelled by coupling the system of interest with a quantum transmission line (channel) represented by a Gaussian bosonic field. The output field carries information about the open system's dynamics which can be accessed by performing continuous-time measurements, and the corresponding conditional system evolution is described in terms of stochastic Schrödinger or filtering equations [16, 22, 35, 45, 173].

While these theories are key to quantum engineering applications, they rely on the precise knowledge of the system's dynamical parameters (e.g.- Hamiltonian of field coupling), which are often uncertain, or completely unknown, and therefore need to be estimated from measurement data. The input-output (I-O) formalism is ideally suited for this statistical inference task, and more generally for implementing online system identification methods [116]. Unlike direct measurement techniques which require repeated system re-preparations and fast control operations [18, 48, 50, 53, 97, 155, 184], the parameters can be estimated continuously from the output measurement trajectory, even if the system is not directly accessible or it is involved in an information processing task. The first investigation in parameter estimation for continuously-observed quantum systems considered the estimation of the Rabi frequency of an atom in a cavity mode, while a photon counting measurement is performed on the cavity output [117]. Subsequent works have addressed a variety of related problems including the dependence on measurement choice [62], adaptive estimation [20, 140] filtering with uncertain parameters [142], particle filters for estimation [38, 154] achieving Heisenberg scaling [6, 7, 118], sensing with error correction [138], Bayesian estimation [64, 105, 126, 143, 180], quantum smoothing [82, 161, 165, 166], waveform estimation [163, 167] estimation of linear systems [68, 85, 112, 113], classical and quantum Fisher informations of the output channel [37, 64, 65, 68, 83, 84, 86, 104].

An upshot of these studies is that standard measurements such as counting, homodyne or heterodyne generally do no achieve the ultimate limit given by the quantum Cramér-Rao bound [25, 93, 95], while the optimal measurement prescribed by the symmetric logarithmic derivative requires collective operations on the output state. In this work we make a first step towards addressing the key

issue of devising *realistic and statistically effective* measurement strategies within the framework of the I-O theory. By realistic we mean procedures which involve sequential continuous-time measurements (as opposed to general non-separable measurements on the output state), possibly combined with more advanced but theoretically well understood operations such as series connections and feedback [81].

For conceptual clarity we focus primarily on discrete-time dynamics, but we will indicate how the techniques may be extended to continuous-time. In the discrete-time setting, the I-O dynamics consists of a d-dimensional system of interest interacting sequentially with a chain of k-dimensional 'noise' input units, which are identically and independently prepared in a state $|\chi\rangle$, cf. Figure 4.1. We assume that the interaction unitary U_θ acting on $\mathbb{C}^d \otimes \mathbb{C}^k$ depends on a

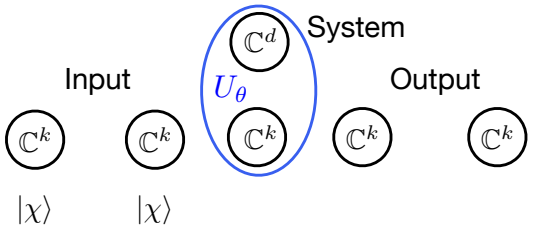

Figure 4.1: Quantum input-output discrete-time dynamics with θ dependent unitary interaction U_θ.

parameter $\theta \in \mathbb{R}$, which we would like to estimate by measuring the output after n evolution steps. In principle, this can be done by applying the adaptive, separable measurement scheme developed in [183] to the joint pure system-output state; indeed this has been shown to attain the quantum Cramér-Rao bound (QCRB). However, while theoretically applicable, the algorithm involves manipulating multi-partite operators, making it unsuitable for processing output states with a large number n of noise units. In addition, it is not clear how the algorithm can be applied to continuous-time dynamics.

Our main contribution is to eliminate these drawbacks by devising a scheme which exploits the intrinsic Markovian structure of the problem. Concretely, we propose an algorithm which finds optimal measurement bases for each of the output units by only performing computations on the space of a doubled-up system and a noise unit, i.e., $\mathbb{C}^d \otimes \mathbb{C}^d \otimes \mathbb{C}^k$. Our algorithm has a similar structure to that of the quantum state filter describing the system's conditional evolution, and can be run in real-time without having to specify the time length n in advance.

While our general algorithm requires measurements on both the output and system in order to achieve finite sample optimality, in Proposition 1 we prove that by measuring only the output we incur a loss of Fisher information which is bounded by a constant, independent of the time n. Since the quantum Fisher information scales linearly in time, this implies that the output measurement is optimal in the leading contribution to the quantum Fisher information (QFI).

We now describe our scheme in more detail. In the first stage of the protocol we use a small proportion of the output units (of sample size $\tilde{n} \approx n^{1-\epsilon}$ with small ϵ) in order to compute a preliminary 'rough estimator' θ_0 of the true parameter θ, by performing a standard sequential measurement. This step is necessary in *any* quantum estimation problem in which the optimal measurement depends on

the unknown parameter [71]. In particular, this means that strictly speaking one can only attain optimality in the limit of large sample sizes, as $\theta - \theta_0$ decays as $n^{-(1-\epsilon)/2}$ thanks to the preliminary estimation stage. In the second (main) stage

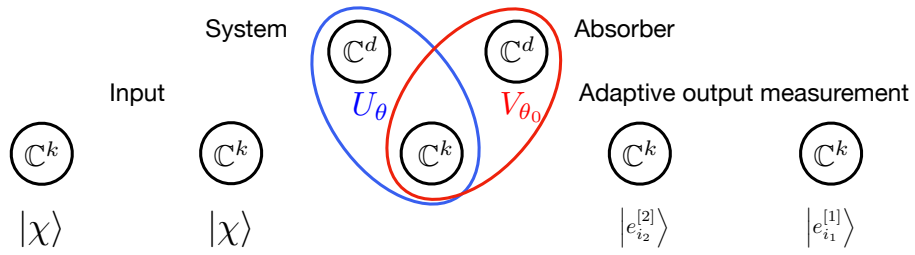

Figure 4.2: Adaptive measurement filter: the output units undergo post-processing with a coherent quantum absorber, followed by applying an adaptive measurement computed with the algorithm.

of the protocol we use θ_0 to design a sequential measurement which achieves the output QFI at $\theta = \theta_0$. Since the first stage insures that $\theta - \theta_0$ decays and the QFI is continuous with respect to θ, we find that the overall scheme is asymptotically optimal at any parameter value θ.

The second stage is illustrated in Figure 4.2: each output unit undergoes a physical transformation (which we call 'quantum post-processing') followed by an adaptive projective measurement whose basis is computed according to the 'measurement filter' algorithm described below. More specifically, after interacting with the system, the post-processing consists in applying a unitary V_{θ_0} to the output noise unit together with an additional ancilla of the same size d as the system. The system and ancilla can be regarded as a single open system (denotes 's+a') of dimension $D = d^2$ coupled to the noise units via the unitary $W_\theta = V_{\theta_0} U_\theta$. The unitary V_{θ_0} is chosen such that s+a has a *pure stationary state* $|\psi\rangle \in \mathbb{C}^D$ at $\theta = \theta_0$, and the output state is identical to that of the input. This is a discrete-time analogue of the notion of *coherent quantum absorber* introduced in [157], and it insures that the 'reference' output state at $\theta = \theta_0$ is the same as the product input state (the 'vacuum'), while deviations from θ_0 produce 'excitations' in the output. After the interaction with the ancilla (absorber), the noise unit is measured in a basis determined by a simple iterative algorithm detailed in section 4.5.

The iterative step consists of using the current value of a certain 'filter operator' on system+absorber to determine the next measurement basis, and then using the measurement outcome to update the filter operator. This simplification relies on the fact that the output is uncorrelated from system (and absorber), which is not the case in the original dynamical setup of Figure 4.1.

In section 4.7 we describe the results of two numerical investigations testing our theoretical results. The first investigation focuses on a simplified model where the system plus absorber are represented by a two-dimensional system with a pure stationary state. While this sidesteps the preliminary estimation stage of the protocol, it allows us to specifically test the key features of the adaptive measurement algorithm with a reasonably large trajectory length and a high number of repetitions. For this model, we can explicitly compute the system-output QFI (cf. Lemma 4), while the classical Fisher information of any output measurement

strategy can be estimated by sampling techniques. The results confirm that the adaptive measurement attains the QFI when the system is measured at the end, while the output-only strategy is only worse by a constant independent of trajectory length. On the other hand, simple measurements (same fixed basis for each unit) perform strictly worse even when the measurement basis is optimised. While the improvement here is not dramatic, our preliminary investigations indicate that the gap increases significantly with the system dimension, depending on the chosen model. We further test the performance of the maximum likelihood estimator and find that its mean square error approaches the inverse of the classical Fisher information in the long-time limit, which agrees broadly with the Cramér-Rao bound. The second numerical investigation implements the full two-stage adaptive measurement algorithm including the use of the coherent absorber.

Finally, we note that our scheme can be extended to continuous-time dynamics by using standard time-discretisation techniques [111, 128]. Although we do not treat this in detail here, we comment on this extension at the end of the chapter.

The chapter is organised as follows. In section 4.2 we briefly review the adaptive algorithm for optimal separable measurements developed in [183]. Section 4.3 introduces the Markov dynamics setting and reviews a key result on the asymptotic QFI of the output. Section 4.4 explains how the use of 'post-processing' by quantum absorber reduces the general estimation problem to one concerning a system with a pure stationary state. This is then used in section 4.5, which details the adaptive measurement procedure including the key 'measurement filter' algorithm. In section 4.6.2 we show that the proposed adaptive output measurement achieves the optimal QFI rate even if the system is not measured. We also devise a scheme to estimate the classical Fisher information of the measurement process by sampling over trajectories. Section 4.7 presents simulation results using an elaboration of an amplitude decay qubit model.

4.2 OPTIMAL SEPARABLE MEASUREMENTS

In this section we review a result by Zhou, Zou and Jiang [183] concerning optimal parameter estimation for multipartite pure states, using separable measurements (local measurements and classical communication). Their method will then be applied to the problem of estimating parameters of discrete time quantum input-output systems. By exploiting the Markovian nature of the dynamics, we will show that the algorithm can be recast in a simpler procedure akin to that of a quantum state filter.

Consider a one parameter quantum statistical model $\{\rho_\theta : \theta \in \mathbb{R}\}$ where ρ_θ is a state on a Hilbert space $\mathcal{H}$ which depends smoothly on the unknown parameter θ. To estimate θ we perform a measurement on the state ρ_θ and compute an estimator $\hat{\theta}$ based on the measurement outcome. According to the QCRB [25, 93, 95], the variance of any unbiased estimator $\hat{\theta}$ is lower bounded as

$$\mathrm{V}ar(\hat{\theta}) = \mathbb{E}\left[(\hat{\theta} - \theta)^2\right] \geq F_\theta^{-1}$$

where F_θ is the quantum Fisher information defined as $F(\theta) = \mathrm{Tr}(\rho_\theta \mathcal{L}_\theta^2)$, and $\mathcal{L}_\theta$ is the symmetric logarithmic derivative satisfying $\frac{d}{d\theta}\rho_\theta = \frac{1}{2}\{\mathcal{L}_\theta, \rho_\theta\}$. In general, for

any given parameter value θ_0, the QCRB is saturated[1] by measuring the symmetric logarithmic derivative (SLD) $\mathcal{L}_{\theta_0}$ and constructing a locally unbiased estimator $\hat{\theta} = \theta_0 + X/F(\theta_0)$ where X is the measurement outcome.

While for full rank states the optimal measurement is essentially unique, for rank deficient states this is not the case and a necessary and sufficient condition for a measurement to saturate the QCRB has been derived in [25]. This has practical relevance for multipartite systems where the measurement of the SLD may not be easy to implement. Motivated by this limitation, the saturability condition has been further investigated in [183] where it is shown that the QCRB for pure states of multipartite systems is achievable using separate measurements constructed in an adaptive fashion which we now proceed to describe.

Consider the pure state model $\rho_\theta = |\psi_\theta\rangle\langle\psi_\theta|$ with $|\psi_\theta\rangle \in \mathcal{H}$. We denote $|\dot{\psi}_\theta\rangle = \frac{d}{d\theta}|\psi_\theta\rangle$ and assume that $\langle\psi_\theta|\dot{\psi}_\theta\rangle = 0$. This can generally be arranged by choosing the (unphysical) phase of $|\psi_\theta\rangle$ to have an appropriate dependence on θ. In particular, in this case we have $|\psi_\theta^\perp\rangle := (1 - |\psi_\theta\rangle\langle\psi_\theta|)|\dot{\psi}_\theta\rangle = |\dot{\psi}_\theta\rangle$. Under this assumption the QFI is given by

$$F_\theta = 4\|\dot{\psi}_\theta\|^2. \tag{4.1}$$

Further, we define the operator M which will play a key role in the analysis

$$M = |\psi_\theta\rangle\langle\dot{\psi}_\theta| - |\dot{\psi}_\theta\rangle\langle\psi_\theta|. \tag{4.2}$$

The authors of [183] note that if a projective rank-one measurement $\{E_i = |e_i\rangle\langle e_i|\}$ satisfies the conditions

$$\langle e_i|M|e_i\rangle = 0, \qquad \text{and} \qquad p_\theta(i) = |\langle e_i|\psi_\theta\rangle|^2 = 1/k, \qquad k = dim(\mathcal{H}) \tag{4.3}$$

then it fulfils the general criteria of [25] and therefore it achieves the QCRB. In fact, the second condition can be relaxed to $p(i) \neq 0$ for all i, but we will stick to the chosen expression for concreteness. The achievability can be understood as follows. The conditions $\langle e_i|M|e_i\rangle = 0$ implies that $\langle e_i|\psi_\theta\rangle\langle\dot{\psi}_\theta|e_i\rangle$ is real, so that the phase of the basis vectors $|e_i\rangle$ can be chosen such that both $\langle e_i|\psi_\theta\rangle$ and $\langle e_i|\dot{\psi}_\theta\rangle$ are real for all i. Together with the condition $p_\theta(i) \neq 0$, this means that in the first order of approximation, the quantum model is described by vectors with real coefficients with respect to the measurement basis. In this case the classical and quantum informations coincide

$$\begin{aligned} I_\theta &= \sum_i p_\theta(i)\left(\frac{d\log p_\theta(i)}{d\theta}\right)^2 = 4\sum_i \frac{(\mathrm{Re}\langle e_i|\psi_\theta\rangle\langle\dot{\psi}_\theta|e_i\rangle)^2}{\langle e_i|\psi_\theta\rangle\langle\psi_\theta|e_i\rangle} \\ &= 4\sum_i \frac{|\langle e_i|\psi_\theta\rangle\langle\dot{\psi}_\theta|e_i\rangle|^2}{\langle e_i|\psi_\theta\rangle\langle\psi_\theta|e_i\rangle} = 4\sum_i |\langle e_i|\dot{\psi}_\theta\rangle|^2 = 4\|\dot{\psi}_\theta\|^2 = F_\theta. \end{aligned}$$

We now assume that we deal with a multipartite system such that $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \cdots \otimes \mathcal{H}_n$, with $dim(\mathcal{H}_i) = k_i$, and follow [183] to show that a separable measurement satisfying the above conditions can be constructed by using the

1 This achievability argument can be made rigorous in an asymptotic setting where the experimenter has an ensemble of n independent, identically prepared systems and employs an adaptive procedure for 'locating' the parameter [49, 71].

algorithm outlined below. For any set of indices $A \subset \{1,\dots,n\}$ we denote its complement by A^c and by $\rho_A = \mathrm{Tr}_{A^c}(\rho_\theta)$ the partial state of the sub-systems with indices in A. For $m > 1$ we denote by $\underline{m}$ the set $\{1,\dots,m\}$. Similarly, we denote $M_A = \mathrm{Tr}_{A^c}(M)$ and for single sub-systems ($A = \{i\}$) we use the notation ρ_i and M_i.

We measure the sub-systems sequentially, such that each individual measurement basis depends on the outcomes of the previous measurements, as follows. In the first step, the measurement basis $\{|e_i^{[1]}\rangle\}_{i=1}^{k_1}$ of system $\mathcal{H}_1$ is chosen such that

$$\left\langle e_i^{[1]} \right| M_1 \left| e_i^{[1]} \right\rangle = 0 \qquad \text{and} \qquad p_1(i) = \left\langle e_i^{[1]} \right| \rho_1 \left| e_i^{[1]} \right\rangle = \frac{1}{k_1}.$$

The existence of such a basis can be established by induction with respect to dimension, cf. proof of Lemma 1 in [183]. The concrete construction in two dimensions is described in section 4.7.

After this, the following procedure is applied sequentially to determine the measurement basis for system $j+1$ with $j = 1,\dots,n-1$: given the outcomes $i_{\underline{j}} := \{i_1,\dots,i_j\}$ of the first j measurements, we choose the basis $\left\{|e_i^{[j+1]}\rangle\right\}_{i=1}^{k_{j+1}}$ in $\mathcal{H}_{j+1}$ such that

$$\left\langle e_i^{[j+1]} \right| M_{j+1}(i_{\underline{j}}) \left| e_i^{[j+1]} \right\rangle = 0$$

and

$$p_{j+1}(i_{\underline{j}}, i) = \left\langle e_i^{[j+1]} \right| \rho_{j+1}(i_{\underline{j}}) \left| e_i^{[j+1]} \right\rangle = \frac{1}{k_1 \cdot k_2 \cdot \dots \cdot k_{j+1}}$$

for all $i = 1,\dots,k_{j+1}$, where

$$M_{j+1}(i_{\underline{j}}) = \left\langle e_{i_{\underline{j}}}^{[j]} \right| M_{\underline{j+1}} \left| e_{i_{\underline{j}}}^{[j]} \right\rangle, \quad \rho_{j+1}(i_{\underline{j}}) = \left\langle e_{i_{\underline{j}}}^{[j]} \right| \rho_{\underline{j+1}} \left| e_{i_{\underline{j}}}^{[j]} \right\rangle,$$

and

$$\left| e_{i_{\underline{j}}}^{[j]} \right\rangle = \left| e_{i_1}^{[1]} \right\rangle \otimes \dots \otimes \left| e_{i_j}^{[j]} \right\rangle$$

Note that the second condition means that for each j the outcome i_j is independent of the others and has equal probabilities $1/k_j$.

After n steps we have defined in an adpative fashion a product measurement basis

$$\left| e_{i_{\underline{n}}}^{[\underline{n}]} \right\rangle = \left| e_{i_1}^{[1]} \right\rangle \otimes \dots \otimes \left| e_{i_n}^{[n]} \right\rangle$$

and one can verify that such a measurement satisfies the general condition (4.3).

4.3 DISCRETE QUANTUM MARKOV CHAINS AND THE OUTPUT QFI

In the input-output formalism the dynamics of a discrete-time quantum open system $\mathcal{H}_s \cong \mathbb{C}^d$ is modeled by successive unitary interactions with independent 'noise units', identically prepared in a state $|\chi\rangle \in \mathcal{H}_u \cong \mathbb{C}^k$. This can be pictured as a conveyor belt where the incoming 'noise units' constitute the input, while the outgoing 'noise units' make up the output of the process, cf. Figure 4.1. If $|\phi\rangle \in \mathcal{H}_s$ is the initial state of the system, and U is the unitary on $\mathcal{H}_s \otimes \mathcal{H}_u$

describing the interaction between system and a noise unit, then the state of the system and output after n time units is

$$|\Psi(n)\rangle = U(n)|\phi \otimes \chi^{\otimes n}\rangle = U^{(n)} \cdot \cdots \cdot U^{(2)} \cdot U^{(1)}|\phi \otimes \chi^{\otimes n}\rangle \in \mathcal{H}_s \otimes \mathcal{H}_u^{\otimes n} \quad (4.4)$$

where $U^{(i)}$ is the unitary acting on the system and the i-th noise unit. From equation (4.4) we find that the reduced state of the system at time n is given by

$$\rho(n) := \mathrm{Tr}_{out}(|\Psi(n)\rangle\langle\Psi(n)|) = T^n(\rho_{in}), \qquad \rho_{in} = |\phi\rangle\langle\phi|,$$

where the partial trace is taken over the output (noise units), and $T : \mathcal{T}_1(\mathcal{H}_s) \to \mathcal{T}_1(\mathcal{H}_s)$ is the Markov transition operator

$$T : \rho \mapsto \mathrm{Tr}_u(U(\rho \otimes \tau)U^*), \qquad \tau := |\chi\rangle\langle\chi|.$$

Fixing an orthonormal basis $\{|1\rangle, \ldots, |k\rangle\}$ in $\mathcal{H}_u$, we can express the system-output state as a matrix product state

$$|\Psi(n)\rangle = \sum_{i_1,\ldots,i_n=1}^{k} K_{i_n} \ldots K_{i_1}|\phi\rangle \otimes |i_1\rangle \otimes \cdots \otimes |i_n\rangle \quad (4.5)$$

where $K_i = \langle i|U|\chi\rangle$ are the Kraus operators of T, so that $T(\rho) = \sum_i K_i \rho K_i^*$.

Now, let us assume that the dynamics depends on a parameter $\theta \in \mathbb{R}$ which we would like to estimate, so that $U = U_\theta$ and $|\Psi(n)\rangle = |\Psi_\theta(n)\rangle$. In the input-output formalism it is usually assumed that the experimenter can measure the output (noise units after the interaction) but may not have access to the system. In this case the relevant quantum statistical model is that of the (mixed) output state given by

$$\rho_\theta^{out}(n) = \mathrm{Tr}_s(|\Psi_\theta(n)\rangle\langle\Psi_\theta(n)|).$$

The problem of estimating θ in this formulation has been investigated in both the discrete time [83, 86] and the continuous time [37, 65, 84] settings. For our purposes, we summarise here the relevant results of [83]. We will assume that the Markov chain is *primitive*, i.e., the transition operator T has a unique full-rank steady state ρ_{ss} (i.e.- $T(\rho_{ss}) = \rho_{ss}$) , and is aperiodic (i.e.- the only eigenvalue of T on the unit circle is 1). In particular, for any initial state ρ_{in}, the system converges to the stationary state $T^n(\rho_{in}) \to \rho_{ss}$ in the large n limit. Therefore, for the asymptotic analysis we can assume that the dynamics is in the *stationary regime* and focus on the large time properties of the stationary output state. The following theorem shows that the output QFI scales linearly with time and provides an explicit expression of the rate.

Theorem 15. *Consider a primitive discrete time Markov chain as described above, whose unitary depends smoothly on a one-dimensional parameter θ, so that $U = U_\theta$. The quantum Fisher information $F_\theta(n)$ of the output state $\rho_\theta^{out}(n)$ scales linearly with n and its rate is equal to*

$$\lim_{n\to\infty} \frac{1}{n} F_\theta(n) = f_\theta = 4 \sum_{i=1}^{k} \left[Tr\left[\rho_{ss} \dot{K}_i^* \dot{K}_i\right] + 2Tr\left[Im(K_i \rho_{ss} \dot{K}_i^*) \cdot \mathcal{R}(Im \sum_j \dot{K}_j^* K_j) \right] \right] \quad (4.6)$$

where $\mathcal{R}$ is the Moore-Penrose inverse of $Id - T_\theta$.

Following standard quantum Cramér-Rao theory [25, 93, 95], the theorem implies that the variance of any (unbiased) output-based estimator is bounded from below by $n^{-1}/f(\theta)$ for large n. A more in depth analysis [83] shows that the output model satisfies the property of *local asymptotic normality* which pertains to a certain quantum Gaussian approximation of the output state and implies that there exists an estimator $\hat{\theta}_n$ which achieves the Cramér-Rao bound (CRB) asymptotically and has normally distributed errors:

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \longrightarrow N(0, f(\theta)^{-1})$$

where the convergence is in distribution to a normal variable with variance $f(\theta)^{-1}$. Below, we will make use of an extension of Theorem 15 which shows that the same result holds for rank-deficient stationary states of ergodic chains, and in particular for pure states [106].

Having identified the output QFI rate, we would like to investigate measurement schemes which can provide good accuracy for estimating the parameter θ. As noted before, the QCRB can be achieved by measuring the SLD of the statistical model. However, the SLD of the output state is generally a complicated operator whose measurement requires collective operations on the noise units. On the other hand, one can consider separate measurements of the same observable on the different noise units and system. The average statistic may provide an efficient estimator and its (asymptotic) Fisher information can be computed explicitly [37]. However, such measurements are in general not optimal. Here would would like to ask the more fundamental question: is it possible to achieve the QCRB using simpler 'local' manipulation of the output units which involve operations on single, rather than multiple units.

4.4 OUTPUT POST-PROCESSING USING QUANTUM COHERENT ABSORBER

In this section we introduce a key tool which will allow us to recast the estimation problem for a general primitive Markov chain (which typically has a mixed stationary state) into one concerning a Markov chain with a 'doubled-up' system having a *pure stationary state*. The construction is a discrete-time adaptation on the concept of *coherent quantum absorber* introduced in [157] for continuous-time dynamics. In section 4.5 we then show how the absorber can be used to compute an adaptive, separable measurement in a simple recursive algorithm.

Consider the input-output system of section 4.3 characterised by a unitary U on $\mathcal{H}_s \otimes \mathcal{H}_u$. We now modify the setup as illustrated in Figure 4.2 by inserting an additional physical d-dimensional system $\mathcal{H}_a$ called absorber which interacts with each of the noise units via a fixed unitary V on $\mathcal{H}_a \otimes \mathcal{H}_u$, applied immediately after U. This can be seen as a type of quantum post-processing of the output prior to the measurement. The original system and the absorber can be considered a single open system with space $\mathcal{H}_s \otimes \mathcal{H}_a$ which interacts with the same conveyor belt of noise units via the unitary $W_\theta = V \cdot U$ where V, U are now understood as the ampliations of the unitaries to the tensor product $\mathcal{H}_s \otimes \mathcal{H}_a \otimes \mathcal{H}_u$. The following lemma shows that for certain choices of V the auxiliary system forms a pure stationary state together with the original one, and the noise units pass unperturbed from input to output. This explains the 'absorber' terminology,

which was originally introduced in the context of continuous-time input-output dynamics [157].

Lemma 1. *Any given primitive quantum Markov chain with unitary U can be extended to a quantum Markov chain including an absorber with unitary V, such that the doubled-up system has a pure stationary state $|\tilde{\psi}\rangle \in \mathcal{H}_s \otimes \mathcal{H}_a$ and*

$$W : |\tilde{\psi}\rangle \otimes |\chi\rangle \mapsto |\tilde{\psi}\rangle \otimes |\chi\rangle, \qquad W = VU.$$

In particular, if the initial state of the doubled-up system is $|\tilde{\psi}\rangle$, then the n-steps output state of the doubled-up system is identical to the input state $|\chi\rangle^{\otimes n}$.

Proof. Let $\rho_{ss} = \sum_i \lambda_i |f_i\rangle\langle f_i|$ be the spectral decomposition of the stationary state of the original system with unitary U. We construct the purification

$$|\tilde{\psi}\rangle = \sum \sqrt{\lambda_i} |f_i\rangle \otimes |f_i\rangle \in \mathcal{H}_s \otimes \mathcal{H}_a$$

which will play the role of stationary state of the extended system. Let $|\phi\rangle := U|\tilde{\psi} \otimes \chi\rangle \in \mathcal{H}_s \otimes \mathcal{H}_a \otimes \mathcal{H}_u$ be the state after applying U. We therefore look for unitary V on $\mathcal{H}_a \otimes \mathcal{H}_u$ (ampliated by identity on $\mathcal{H}_s$) such that V reverts the action of U

$$V : |\phi\rangle \mapsto |\tilde{\psi} \otimes \chi\rangle.$$

Since $\mathrm{Tr}_a(|\tilde{\psi}\rangle\langle\tilde{\psi}|) = \rho_{ss}$ this means that the reduced state of the system after applying U is still the stationary state, so that

$$|\phi\rangle = \sum_i \sqrt{\lambda_i} |f_i\rangle \otimes |g_i\rangle$$

where $|g_i\rangle$ are mutually orthogonal unit vectors in $\mathcal{H}_a \otimes \mathcal{H}_u$. We now choose a unitary V such that $V|g_i\rangle = |f_i \otimes \chi\rangle$, for all i, which is always possible due to orthogonality. □

4.5 ADAPTIVE MEASUREMENT ALGORITHM

In this section we describe our adaptive output measurement protocol for estimating an unknown one-dimensional dynamical parameter θ of a discrete time quantum Markov chain with unitary U_θ, as described in section 4.3. The protocol has two stages. In the first stage we use a small proportion (e.g.- $\tilde{n} = n^{1-\epsilon}$, with $0 < \epsilon \ll 1$) of the output units in order to compute a preliminary 'rough estimator' θ_0 of the true parameter θ by performing a standard sequential measurement. This step is necessary in any quantum estimation problem in which the optimal measurement depends on the unknown parameter [49, 71], and will inform the second stage of the protocol. In the second stage we use θ_0 to design a *optimal* sequential measurement for $\theta = \theta_0$, i.e., one that achieves the output QFI at $\theta = \theta_0$. Since $\delta\theta = \theta - \theta_0 = O(n^{-1/2+\epsilon})$ [37], this implies that the procedure is asymptotically optimal for *any* parameter value θ, in that the classical Fisher information has the same linear scaling as the QFI $F_\theta(n)$. This stage has two key ingredients (cf. Figure 4.2): a quantum 'post-processing' operation on output units immediately after interacting with the system, followed by an adaptive projective measurement whose basis is computed according to a 'measurement filter' algorithm inspired by [183]. We now describe the two steps in detail.

Quantum post-processing. After the interaction U_θ with the system, the output units interact sequentially with a d-dimensional ancillary system, cf. Figure 4.2. The interaction unitary V_{θ_0} is chosen such that the ancillary system is a coherent quantum absorber for $\theta = \theta_0$, see section 4.4 and Lemma 1 for construction. This means that the system plus absorber (s+a) can be regarded as a single $D = d^2$ dimensional open system with associated unitary $W_\theta = V_{\theta_0} U_\theta$, which has a pure stationary state at $\theta = \theta_0$ denoted $|\psi\rangle \in \mathbb{C}^D$, and whose output state is identical to the input. The general estimation problem for U_θ has been reduced to a special one for a doubled-up system with unitary W_θ which features a pure stationary state at $\theta = \theta_0$.

Remark 1. *Since the absorber transformation V_{θ_0} does not depend on θ and is applied after U_θ, the overall effect over an n steps interval is to rotate the absorber plus output state by a fixed unitary $V_{\theta_0}^{(n)} \dots V_{\theta_0}^{(1)}$. This means that the total QFI does not change by introducing the absorber.*

Adaptive measurement algorithm. We will assume for simplicity that the initial state of $s + a$ is the stationary state $|\psi\rangle$ such that the full (s+a)-output state is $|\Psi_\theta(n)\rangle = W_\theta(n)|\psi \otimes \chi^{\otimes n}\rangle$ as defined in (4.4). One could obtain similar results for different initial states by simply waiting long enought for the system and absorber to converge to the stationary state $|\psi\rangle$. In principle we could now apply the algorithm described in section 4.2 to construct and adaptive measurement whose classical Fisher information is equal to the system-output QFI. In fact we could have done this without using the absorber. However, this procedure has some drawbacks. Indeed, in order to compute the measurement bases one needs to work with large dimensional spaces which becomes unfeasible in an asymptotic setting. Secondly, it is not clear a priori whether the output units can be measured immediately after the interaction with the system, and whether the measurements depend on the length of the output (sample size). In addition, the procedure requires a final measurement on the system, which may be impractical in the context of input-output dynamics. We will show that all these issues can be addressed by taking into account the Markovian structure of our model, and exploiting the pure stationary state property. Let us denote $W = W_{\theta_0}$, $\dot{W} = \frac{dW}{d\theta}\Big|_{\theta_0}$ and

$$\begin{aligned} A_1 = M^{(1)} &= \dot{W}|\psi \otimes \chi\rangle\langle\psi \otimes \chi|W^* - W|\psi \otimes \chi\rangle\langle\psi \otimes \chi|\dot{W}^* \\ &= \dot{W}P_{\psi\otimes\chi} - P_{\psi\otimes\chi}\dot{W}^*, \end{aligned}$$

and

$$B_1 = \mathrm{Tr}_{s+a} A_1 = KP_\chi - P_\chi K^*, \qquad \text{with } K = \langle\psi|\dot{W}|\psi\rangle.$$

In section 4.5.1 we show that the adaptive measurement of [183] reduces to the following iterative algorithm which is conceptually similar to quantum state filtering [16, 22], and involves individual measurements on the output units immediately after the interaction with the system, and computations with operators on $\mathbb{C}^D \otimes \mathbb{C}^k$ at each step.

Initialisation step (j=1). The first measurement basis $\left\{|e_i^{[1]}\rangle\right\}$ in $\mathbb{C}^k$ is chosen such that the following conditions are fulfilled:

$$\left\langle e_i^{[1]}\right| B_1 \left|e_i^{[1]}\right\rangle = 0, \qquad \text{and} \qquad \left|\langle e_i^{[1]}|\chi\rangle\right|^2 = \frac{1}{k}, \qquad \text{for all } i = 1\dots,k.$$

The first noise unit is measured in this basis and the outcome $X_1 = i_1$ is obtained. The filter at time $j = 1$ is defined as the (trace zero) s+a operator

$$\Pi_1 = \left\langle e^{[1]}_{i_1} \right| A_1 \left| e^{[1]}_{i_1} \right\rangle .$$

Iterative step. The following step is iterated for $j = 2, \ldots, n$. Given the filter operator Π_{j-1} of the previous step, we define

$$A_j = \frac{1}{D^{j-1}} A_1 + W \left(\Pi_{j-1} \otimes P_\chi \right) W^*, \qquad B_j = \mathrm{Tr}_{s+a} A_j$$

The j-th measurement basis $\left\{ |e^{[j]}_i\rangle \right\}$ is chosen to fulfill the conditions

$$\left\langle e^{[j]}_i \right| B_j \left| e^{[j]}_i \right\rangle = 0, \qquad \text{and} \qquad \left| \langle e^{[j]}_i | \chi \rangle \right|^2 = \frac{1}{k}, \qquad \text{for all } i = 1, \ldots k. \tag{4.7}$$

We measure the j-th noise unit in the basis $\left\{ |e^{[j]}_i\rangle \right\}$ and obtain the result $X_j = i_j$. The filter at time j is updated to

$$\Pi_j = \left\langle e^{[j]}_{i_j} \right| A_j \left| e^{[j]}_{i_j} \right\rangle .$$

Final s+a measurement. This is an optional step which involves a final joint measurement on system and absorber. The basis $\left\{ |e^{[s+a]}_i\rangle \right\}_{i=1}^{D}$ is determined by the following conditions

$$\left\langle e^{[s+a]}_i \right| \Pi_n \left| e^{[s+a]}_i \right\rangle = 0, \qquad p(i|i_1, \ldots, i_n) = \left| \langle e^{[s+a]}_i | \psi \rangle \right|^2 = 1/D.$$

The system and absorber are measured in this basis and the outcome $X = i_0$ is obtained.

The output measurement record $\{i_1, \ldots, i_n, i_0\}$ is collected and used for estimating the parameter θ. The likelihood function is given by

$$p_\theta(i_1, \ldots i_n, i_0) = \left| \left\langle e^{[s+a]}_{i_0} \otimes e^{[1]}_{i_1} \otimes \cdots \otimes e^{[n]}_{i_n} \mid \Psi_\theta(n) \right\rangle \right|^2 . \tag{4.8}$$

For later use, we denote by $p_\theta(i_1, \ldots i_n)$ the marginal distribution of the output measurement record only.

4.5.1 *Derivation of the measurement filter*

In this section we provide the derivation of the measurement filter: we start by applying the algorithm [183] to our estimation problem. The natural setup is to measure the noise units in the order in which they emerge in the output, followed by a final measurement on the system+absorber. For simplicity we refer to the latter as the system. We start by defining

$$M^{(n)} := |\Psi(n)\rangle \langle \dot{\Psi}(n)| - |\dot{\Psi}(n)\rangle \langle \Psi(n)| , \tag{4.9}$$

where the index n keeps track of the output length. In this section we will use the label 0 for the system, and $1, \ldots, n$ for the noise units of the output. The algorithm

prescribes measurement bases $\left\{|e_i^{[l]}\rangle\right\}_{i=1}^{k}$ for each of the output units $l = 1, \ldots, n$ in an adaptive, sequential fashion. The first basis satisfies the equations

$$\left\langle e_i^{[1]} \right| M_1^{(n)} \left| e_i^{[1]} \right\rangle = 0 \qquad \text{and} \qquad \left| \left\langle e_i^{[1]} \right| \chi \right\rangle \Big|^2 = \frac{1}{k}, \qquad \text{for all } i.$$

where $M_1^{(n)} = \mathrm{Tr}_{0,2,\ldots,n} M^{(n)}$. The second equality follows from the fact that at θ_0 we have $|\Psi(n)\rangle = |\psi \otimes \chi^{\otimes n}\rangle$. The next basis depends on the outcome i_1 of the first measurement and satisfies the constraints

$$\left\langle e_i^{[2]} \right| M_2^{(n)}(i_1) \left| e_i^{[2]} \right\rangle = 0 \qquad \text{and} \qquad \left| \left\langle e_i^{[2]} \right| \chi \right\rangle \Big|^2 = \frac{1}{k}, \qquad \text{for all } i$$

where

$$M_2^{(n)}(i_1) = \mathrm{Tr}_{0,3,\ldots,n} \left\langle e_{i_1}^{[1]} \right| M^{(n)} \left| e_{i_1}^{[1]} \right\rangle .$$

Assuming the first $j < n$ units have been measured and a measurement record $i_{\underline{j}} := \{i_1, \ldots, i_j\}$ has been obtained, we denote

$$M^{(n)}(i_{\underline{j}}) := \left\langle e_{i_1}^{[1]} \otimes e_{i_2}^{[2]} \otimes \cdots \otimes e_{i_j}^{[j]} \right| M^{(n)} \left| e_{i_1}^{[1]} \otimes e_{i_2}^{[2]} \otimes \cdots \otimes e_{i_j}^{[j]} \right\rangle ,$$
$$M_{j+1}^{(n)}(i_{\underline{j}}) := \mathrm{Tr}_{0,j+2,\ldots,n} M^{(n)}(i_{\underline{j}}).$$

The measurement basis $\left\{ \left| e_i^{[j+1]}(i_{\underline{j}}) \right\rangle \right\}$ for unit $j+1$ is then obtained by solving the constraints

$$\left\langle e_i^{[j+1]}(i_{\underline{j}}) \right| M_{j+1}^{(n)}(i_{\underline{j}}) \left| e_i^{[j+1]}(i_{\underline{j}}) \right\rangle = 0 \quad \text{and} \quad \left| \left\langle e_i^{[j+1]}(i_{\underline{j}}) \right| \chi \right\rangle \Big|^2 = \frac{1}{k}, \qquad \text{for all } i.$$

The last step consists of measuring the system using the same procedure as for the output units.

A priori, the procedure depends on the size n of the output. The following lemma shows that the optimal bases obtained for different output sizes coincide.

Lemma 2. *Let j, n be two output lengths with $j < n$ and consider applying the above procedure to the corresponding states $|\Psi^{(j)}\rangle$ and respectively $|\Psi^{(n)}\rangle$. The optimal measurement bases $\{|e_i^{[l]}\rangle\}$ for the units $l = 1, \ldots, j$ satisfy the same constraints and can be chosen to be the same. In addition we have*

$$M_j^{(n)}(i_{\underline{j-1}}) = M_j^{(j)}(i_{\underline{j-1}}).$$

Proof. Recall that $|\Psi(n)\rangle = W^{(n)} \ldots W^{(1)} |\psi \otimes \chi^{\otimes n}\rangle$, and let us denote

$$P(n) := \left|\psi \otimes \chi^{\otimes n}\right\rangle \left\langle \psi \otimes \chi^{\otimes n}\right| = |\Psi_{\theta_0}(n)\rangle\langle\Psi_{\theta_0}(n)|.$$

Then

$$\begin{aligned} |\dot{\Psi}(n)\rangle &= \sum_{i=1}^{n} W^{(n)} \ldots \dot{W}^{(i)} \ldots W^{(1)} |\psi \otimes \chi^{\otimes n}\rangle \\ &= \sum_{i=1}^{n} W^{(n)} \ldots \dot{W}^{(i)} |\psi \otimes \chi^{\otimes n}\rangle \end{aligned}$$

where we used the fact that W leaves $|\psi \otimes \chi\rangle$ invariant.

We first show that the matrix $M_1^{(n)}$ does not depend on n, and therefore the first measurement basis does not depend on the length of the output.

$$
\begin{aligned}
M_1^{(n)} &= \mathrm{Tr}_{0,2,\ldots,n}\left[M^{(n)}\right] \qquad (4.10)\\
&= \sum_i \mathrm{Tr}_{0,2,\ldots,n}\Big[\, W^{(n)}\cdots W^{(i)}\cdots W^{(1)}P(n).\\
&\qquad\qquad .W^{(1)*}\cdots \dot{W}^{(i)*}\cdots W^{(n)*} - c.c\,\Big]\\
&= \sum_i \mathrm{Tr}_{0,2,\ldots,n}\Big[\, W^{(n)}\cdots W^{(i)}P(n)\dot{W}^{(i)*}\cdots W^{(n)*} - c.c\,\Big]\\
&= \sum_i \mathrm{Tr}_{0,2,\ldots,n}\Big[\, W^{(i)}P(n)\dot{W}^{(i)*} - c.c\,\Big]\\
&= \mathrm{Tr}_0\left[W^{(1)}P(1)\dot{W}^{(1)*} - c.c\right] + \sum_{i=2}\mathrm{Tr}_{0,2,\ldots,n}\left[W^{(i)}P(n)\dot{W}^{(i)*} - c.c\right]\\
&= \mathrm{Tr}_0\left[M^{(1)}\right] = M_1^{(1)}, \qquad (4.11)
\end{aligned}
$$

where $c.c$, denotes the adjoint term. In the third equality we used

$$\mathrm{Tr}_{0,2}[W_{0,2}A_{0,1,2}W_{0,2}^*] = \mathrm{Tr}_{0,2}[A_{0,1,2}]$$

where $W_{0,2}$ is a unitary acting on subsystems $0,2$ of a tripartite system, and $A_{0,1,2}$ acts on subsystems $0,1,2$. In the last equality we used our assumption

$$\langle\psi\otimes\chi|\, W^*\dot{W}\,|\psi\otimes\chi\rangle = 0.$$

Let $\left\{\left|e_i^{[1]}\right\rangle\right\}$ be the measurement basis determined from $M_1 = M_1^{(n)} = M_1^{(j)}$. We now show that, conditional on the outcome i_1 of this measurement, the second basis $\left\{\left|e_i^{[2]}\right\rangle\right\}$ does not depend on the length of the output, which can be take to be equal to 2.

$$
\begin{aligned}
M_2^{(n)}(i_1) &= \mathrm{Tr}_{0,3,\ldots,n}\left[M^{(n)}(i_1)\right]\\
&= \sum_i \mathrm{Tr}_{0,3,\ldots,n}\left[\left\langle e_{i_1}^{[1]}\right| W^{(n)}\cdots W^{(i)}\cdots W^{(1)}P(n).\right.\\
&\qquad\qquad \left.\left.\cdots W^{(1)*}\cdots \dot{W}^{(i)*}\cdots W^{(n)*}\left|e_{i_1}^{[1]}\right\rangle - c.c\right]\right.\\
&= \mathrm{Tr}_{0,3,\ldots,n}\left[\left\langle e_{i_1}^{[1]}\right| W^{(n)}\cdots W^{(1)}P(n)\dot{W}^{(1)*}\cdots W^{(n)*}\left|e_{i_1}^{[1]}\right\rangle - c.c\right]\\
&\quad + \mathrm{Tr}_{0,3,\ldots,n}\left[\left\langle e_{i_1}^{[1]}\right| W^{(n)}\cdots W^{(2)}W^{(1)}P(n).\right.\\
&\qquad\qquad \left.\left.\cdots W^{(1)*}\dot{W}^{(2)*}\cdots W^{(n)*}\left|e_{i_1}^{[1]}\right\rangle - c.c\right]\right.\\
&\qquad + \sum_{i=3}^{n}\mathrm{Tr}_{0,3,\ldots,n}\left[\left\langle e_{i_1}^{[1]}\right| W^{(n)}\cdots W^{(i)}\cdots W^{(1)}P(n).\right.\\
&\qquad\qquad \left.\left.\cdots W^{(1)*}\cdots \dot{W}^{(i)*}\cdots W^{(n)*}\left|e_{i_1}^{[1]}\right\rangle - c.c\right]\right.\\
&= \mathrm{Tr}_0\left[W^{(2)}\left\langle e_{i_1}^{[1]}\right| W^{(1)}P(2)\dot{W}^{(1)*}\left|e_{i_1}^{[1]}\right\rangle W^{(2)*} - c.c\right]\\
&\quad + \frac{1}{k}\mathrm{Tr}_0\left[W^{(1)}P(1)\dot{W}^{(1)*} - c.c\right]\\
&= \mathrm{Tr}_0\left[W^{(2)}\left(\left\langle e_{i_1}^{[1]}\right| M^{(1)}\left|e_{i_1}^{[1]}\right\rangle\otimes P_\chi\right)W^{(2)*}\right] + \frac{1}{k}\mathrm{Tr}_0\left[M^{(1)}\right] = M_2^{(2)}(i_1).
\end{aligned}
$$

The equalities follow in the same way as in (4.11), and in addition we used $\left|\langle e_{i_1}^{[1]}|\chi\rangle\right|^2 = \frac{1}{k}$ in the third equality.

Using the same techniques as above we obtain the general statement

$$\begin{aligned} M_j^{(n)}(\underline{i_{j-1}}) &= \mathrm{Tr}_{0,j+1,\dots,n}\left[M^{(n)}(\underline{i_{j-1}})\right] \qquad (4.12)\\ &= \frac{1}{k^{j-1}}\mathrm{Tr}_0\left[M^{(1)}\right] \\ &\quad + \mathrm{Tr}_0\left[W^{(j)}\left(\langle e_{\underline{i_{j-1}}}^{[j-1]}|\, M^{(j-1)}(\underline{i_{j-2}})\, |e_{\underline{i_{j-1}}}^{[j-1]}\rangle \otimes P_\chi\right) W^{(j)*}\right] \\ &= \mathrm{Tr}_0\left[M^{(j)}(\underline{i_{j-1}})\right] = M_j^{(j)}(\underline{i_{j-1}}). \qquad (4.13) \end{aligned}$$

□

Next, we show that we can express $M^{(j)}(\underline{i_{j-1}})$ in terms of $M^{(1)}$, $M^{(j-1)}(\underline{i_{j-2}})$ and $|e_{j-1}^{[j-1]}\rangle$. Indeed by writing $W(j) = W^{(j)}W(j-1)$ we have

$$\begin{aligned} M^{(j)}(\underline{i_{j-1}}) &:= \left\langle e_{\underline{i_{j-1}}}^{[j-1]}\right| M^{(j)} \left|e_{\underline{i_{j-1}}}^{[j-1]}\right\rangle \\ &= \left\langle e_{\underline{i_{j-1}}}^{[j-1]}\right| W^{(j)}W(j-1)P(j)W^*(j-1)\dot{W}^{(j)*} \left|e_{\underline{i_{j-1}}}^{[j-1]}\right\rangle - c.c \\ &\quad + \left\langle e_{\underline{i_{j-1}}}^{[j-1]}\right| W^{(j)}W(j-1)P(j)\dot{W}(j-1)^*W^{(j)*} \left|e_{\underline{i_{j-1}}}^{[j-1]}\right\rangle - c.c \\ &= \frac{1}{k^{j-1}}M^{(1)} + W^{(j)}\left(\left\langle e_{i_{j-1}}^{[j-1]}\right| M^{(j-1)}(\underline{i_{j-2}}) \left|e_{i_{j-1}}^{[j-1]}\right\rangle \otimes P_\chi\right) W^{(j)*} \qquad (4.14) \end{aligned}$$

This can then be used to determine the next measurement basis, producing the iterative procedure described in section 4.5, which consists in updating the 'filter' that determines the optimal basis at each time step using the last measurement outcome $|e_{i_j}^{[j]}\rangle$.

Let $A_1 = M^{(1)}$ and $A_j := M^{(j)}(\underline{i_{j-1}})$ for $j > 1$, and denote

$$\Pi_j := \left\langle e_{i_j}^{[j]}\right| A_j \left|e_{i_j}^{[j]}\right\rangle = \left\langle e_{i_j}^{[j]}\right| M^{(j)}(\underline{i_{j-1}}) \left|e_{i_j}^{[j]}\right\rangle$$

Then equation (4.14) can be written as

$$A_j = \frac{1}{k^{j-1}}A_1 + U^{(j)}\left(\Pi_{j-1}\otimes P_\chi\right)U^{(j)*}$$

The optimal measurement is obtained by applying the conditions to the operator $B_j := \mathrm{Tr}_0 A_j$.

4.6 FISHER INFORMATIONS CONSIDERATIONS

In this section we investigate the relationship between the classical Fisher information of the output adaptive measurement process and the system-output QFI. We prove that both scale with the same rate and the latter may be larger than the former by at most a constant, independent of time.

We also provide an expression of the CFI of sequential (adaptive or standard) output measurements, which is amenable to estimation by sampling. This tool will be used to confirm the optimality of our adaptive algorithm in numerical simulations.

4.6.1 *Achievability of the QFI with adaptive output measurements*

The adaptive measurement scheme described in section 4.5 insures the CFI of the full measurement (output and s+a) is equal to the QFI of the full pure state model

$$I_{\theta_0}^{(\mathrm{s}+a+o)}(n) = F_{\theta_0}^{(\mathrm{s}+a+o)}(n) = F_{\theta_0}^{(\mathrm{s}+o)}(n) \tag{4.15}$$

where the last equality follows from the fact that the absorber acts as an additional rotation which does not change the QFI, cf. Remark 1. However, in certain physical implementations the system may not be accessible for measurements, so the more interesting scenario is that in which only the output state is measured. In this case the CFI will generally be strictly smaller that the QFI, and the question is whether by measuring only the output we incur a significant loss of information. In proposition 1 we show that this is not the case: the difference between the QFI and the output CFI is bounded by a constant, so for large times the loss of information is negligible compared to both QFI and output CFI, which scale linearly with time. In section 4.7 we will illustrate the result on a specific model.

Proposition 1. *Consider the setup described in section 4.5, and let $F(n)$ be the system-absober-output QFI, and $I^{(o)}(n)$ be the output CFI for the optimal adaptive measurement, at $\theta = \theta_0$. Then $F(n) - I^{(o)}(n) < c$ for all n where c is a constant depending only on the model U_θ. Consequently,*

$$\lim_{n\to\infty} \frac{1}{n} I^{(o)}(n) = \lim_{n\to\infty} \frac{1}{n} F(n) = f > 0$$

where $f = f_{\theta_0}$ is the QFI rate (4.6).

The proof of Proposition 1 can be found in section 4.9.

4.6.2 *Computing the classical Fisher information of the output*

The *classical Fisher information* of the output measurement process at θ_0 is

$$I^{(o)}(n) = \mathbb{E}_{\theta_0}\left(\frac{d\log p_\theta}{d\theta}\right)^2 = \sum_{i_1,\ldots,i_n} p_{\theta_0}(i_1,\ldots i_n)^{-1}\left(\frac{dp_\theta(i_1,\ldots,i_n)}{d\theta}\bigg|_{\theta_0}\right)^2$$

where the sum runs over indices such that $p_\theta(i_1,\ldots,i_n) > 0$. In general

$$I^{(o)}(n) \leq F^{(o)}(n) \leq F^{(\mathrm{s}+o)}(n)$$

where the successive upper bounds are output and system-output QFIs respectively.

In our simulation study we will be interested to study to what extent these bounds are saturated in the adaptive and non-adaptive scenarios, and in particular, to verify the prediction of Proposition 1. Since the classical Fisher information is difficult to compute for long trajectories, we will recast it as an expectation which can be estimated by sampling measurement trajectories. In Lemma 3 below, we will use the fact that at $\theta = \theta_0$ the vector $|\psi\rangle$ is the stationary state, and therefore

$$K_i^{[j]}|\psi\rangle = c_i^{[j]}|\psi\rangle \tag{4.16}$$

for *any* Kraus decomposition $K_i^{[j]} = \langle e_j^{[i]}|W|\chi\rangle$ (for simplicity we use the same notation for s+a Kraus operators as in section 4.3).

Lemma 3. *Consider the setup described in section 4.5. The output CFI at $\theta = \theta_0$ is given by*

$$I^{(o)}(n) = \mathbb{E}_{\theta_0}(f^2) = \sum_{i_1,\dots,i_n} p_{\theta_0}(i_1,\dots,i_n) f^2(i_1,\dots,i_n) \tag{4.17}$$

where f is the function

$$f(i_1,\dots,i_n) = 2Re \sum_{j=1}^{n} \frac{\langle\psi|K_{i_n}^{[n]}\dots K_{i_{j+1}}^{[j+1]}\dot{K}_{i_j}^{[j]}|\psi\rangle}{c_{i_j}^{[j]}\dots c_{i_n}^{[n]}}$$

and the constants $c_{i_j}^{[j]}$ are defined by equation 4.16. In particular, $I^{(o)}(n)$ can be estimated by computing the empirical average of f^2 over sampled trajectories.

Proof. The CFI of any output measurement can be computed explicitly by writing

$$\begin{aligned}
\frac{d}{d\theta}p_\theta(i_1,\dots i_n)\Big|_{\theta_0} &= \frac{d}{d\theta}\left\|K_{i_n}^{[n]}\cdots K_{i_1}^{[1]}\psi\right\|^2\Big|_{\theta_0} \\
&= 2\mathrm{Re}\sum_{j=1}^{n}\langle\psi|K_{i_1}^{[1]*}\cdots K_{i_n}^{[n]*}K_{i_n}^{[n]}\cdots\dot{K}_{i_j}^{[j]}\cdots K_{i_1}^{[1]}|\psi\rangle \\
&= \left|c_{i_1}^{[1]}\cdots c_{i_n}^{[n]}\right|^2\cdot 2\mathrm{Re}\sum_{j=1}^{n}\frac{\langle\psi|K_{i_n}^{[n]}\dots K_{i_{j+1}}^{[j+1]}\dot{K}_{i_j}^{[j]}|\psi\rangle}{c_{i_j}^{[j]}\dots c_{i_n}^{[n]}}
\end{aligned}$$

Therefore

$$\begin{aligned}
I_{\theta_0}^{(out)}(n) &= \sum_{i_1,\dots,i_n}\left|c_{i_1}^{[1]}\cdots c_{i_n}^{[n]}\right|^2\left(2\mathrm{Re}\sum_{j=1}^{n}\frac{\langle\psi|K_{i_n}^{[n]}\dots K_{i_{j+1}}^{[j+1]}\dot{K}_{i_j}^{[j]}|\psi\rangle}{c_{i_j}^{[j]}\dots c_{i_n}^{[n]}}\right)^2 \\
&= \sum_{i_1,\dots,i_n} p_{\theta_0}(i_1,\dots,i_n)f^2(i_1,\dots,i_n) = \mathbb{E}_{\theta_0}(f^2)
\end{aligned} \tag{4.18}$$

where f is the function

$$f(i_1,\dots,i_n) = 2\mathrm{Re}\sum_{j=1}^{n}\frac{\langle\psi|K_{i_n}^{[n]}\dots K_{i_{j+1}}^{[j+1]}\dot{K}_{i_j}^{[j]}|\psi\rangle}{c_{i_j}^{[j]}\dots c_{i_n}^{[n]}}$$

□

We now consider the case where a (projective) measurement $\{P_i^{(s+a)}\}$ is performed on s+a, after obtaining the output measurement record $(i_1,\dots,i_n)$. We denote the additional outcome by i_0. The CFI of the full process is

$$I^{(s+a+o)}(n) = \mathbb{E}_{\theta_0}\left(\frac{d\log p_\theta}{d\theta}\Big|_{\theta_0}\right)^2$$

where

$$p_\theta(i_1,\dots,i_n,i_0) = \left\|P_{i_0}^{(s+a)}K_{i_n}^{[n]}\dots K_{i_1}^{[1]}\psi\right\|^2$$

is the likelihood function of a trajectory augmented by the system measurement outcome i_0. The relevant upper bound in this case is

$$I^{(s+a+o)}(n) \le F^{(s+a+o)}(n) = F^{(s+o)}(n). \tag{4.19}$$

A similar computation to that of Lemma 3 gives the system-output classical Fisher information

$$I^{(s+a+o)}(n) = \mathbb{E}_{\theta_0}(\tilde{f}^2)$$

where $\tilde{f}$ is the function

$$\tilde{f}(i_1, \ldots, i_n, i_0) \quad = \quad 2\,\mathrm{Re} \sum_{j=1}^{n} \frac{\langle \psi | P_{i_0}^{(s)} K_{i_n}^{[n]} \ldots K_{i_{j+1}}^{[j+1]} \dot{K}_{i_j}^{[j]} | \psi \rangle}{c_{i_j}^{[j]} \ldots c_{i_n}^{[n]}}$$

For fixed (non-adaptive) measurements, the bound (4.19) is generally not saturated except for special models (e.g.- if the state coefficients in the measurement basis are real for all θ). In contrast, the system-absorber-output classical Fisher information for adaptive measurements is equal to the QFI thanks to the optimality of the adaptive measurement procedure (4.15). This is confirmed by our simulation study which also investigates the performance of the fixed measurement scenario.

4.7 NUMERICAL SIMULATIONS

We now test the key properties of the adaptive measurement scheme developed in section 4.5, in two separate numerical investigations.

The first investigation described in subsections 4.7.1 and 4.7.2 employs a simplified Markov model which bypasses stage-one of the scheme (computing a rough estimator θ_0) and simulates data at $\theta = \theta_0$. This allows us to directly study the performance of the algorithm itself (stage two), rather than that of the combination of the two stages. The second simplification of this investigation is that we choose a system which has a pure stationary state at θ_0; this means that no absorber is required, so the system can be seen as a surrogate for the system+absorber in the general scheme. The reason for this is mainly practical, as it allows us to use a two-dimensional system while system+absorber would have dimension at least four.

The second numerical investigation consists of a full simulation study including the use of the coherent absorber and the two stage estimation procedure, and its results are presented in subsection 4.7.3. Here we can see the overall performance of the estimation method, but it is harder to estimate the Fisher information of the measurement process and to separate the contribution of the two stages in the overall estimation error.

4.7.1 *Simplified Markov model for the first numerical investigation*

We consider a dynamical model consisting of a two-dimensional system coupled via a unitary U_θ to two dimensional noise units in state $|\chi\rangle = |0\rangle$. The input state and the unitary are designed such that the stationary state at $\theta_0 = 0$ is $|\psi\rangle = |0\rangle$. Since the input is prepared in a fixed state, we only need to define the action of

U_θ on the basis vectors $|0\rangle \otimes |0\rangle$ and $|1\rangle \otimes |0\rangle$. The following choice has unknown parameter θ and two known parameters λ and ϕ

$$\begin{aligned} U_\theta : |00\rangle &\longrightarrow \cos(\theta)\sqrt{1-\theta^2}\,|00\rangle + i\sin(\theta)\sqrt{1-\theta^2}\,|10\rangle + \theta\,|11\rangle\,, \\ U_\theta : |10\rangle &\longrightarrow i\sin(\theta)\sqrt{1-\lambda}\,|00\rangle + \cos(\theta)\sqrt{1-\lambda}\,|10\rangle + \sqrt{\lambda}e^{i\phi}\,|01\rangle\,. \end{aligned} \tag{4.20}$$

A non-zero value of the phase parameter ϕ ensures that the system-output state does not have real coefficients in the standard basis, in which case the standard basis measurement would be optimal. Note that at θ_0 the dynamics provides a simple model for 'photon-decay' with decay parameter λ.

We compare two measurements scenarios. In the first, non-adaptive scenario, the noise units are measured in a fixed orthonormal basis

$$\{|f_0\rangle = (|0\rangle + |1\rangle)/\sqrt{2}, |f_1\rangle = (|0\rangle - |1\rangle)/\sqrt{2}\}$$

while in the second scenario they are measured adaptively following the algorithm described in section 4.5. In both cases, given the measurement record $\{i_1, \ldots, i_n\} \in \{0,1\}^n$, the conditional state of the system is

$$|\psi_n(i_1, \ldots i_n)\rangle = \frac{K_{i_n}^{[n]} \ldots K_{i_1}^{[1]} |\psi\rangle}{\|K_{i_n}^{[n]} \ldots K_{i_1}^{[1]} \psi\|} \tag{4.21}$$

where the Kraus operators are $\{K_0 = \langle f_0|U|\chi\rangle, K_1 = \langle f_1|U|\chi\rangle\}$ in the first scenario, and $K_i^{[j]} = \langle e_i^{[j]}|U|\chi\rangle$ in the second one. The likelihood of the output measurement trajectory is

$$p_\theta(i_1, \ldots i_n) = \|K_{i_n}^{[n]} \ldots K_{i_1}^{[1]} \psi\|^2. \tag{4.22}$$

Recall that the optimal measurement basis needs to satisfy the conditions (4.7) in section 4.5. For two dimensional noise units the computation reduces to the following scheme. We express the traceless, anti-Hermitian matrix B_j defined in section 4.5 as $B_j = i\boldsymbol{r}^{[j]} \cdot \boldsymbol{\sigma}$ where $\boldsymbol{r}^{[j]} = (r_x^{[j]}, r_y^{[j]}, 0)$ is its Bloch vector, and similarly, we let $\pm\boldsymbol{s}^{[j]}$ be the Bloch vectors of the basis vectors $\{|e_1^{[j]}\rangle, |e_2^{[j]}\rangle\}$ satisfying the conditions (4.7). Then one finds that the conditions (4.7) are satisfied if $\boldsymbol{s}^{[j]}$ is taken to be $\boldsymbol{s}^{[j]} = (r_y^{[j]}, -r_x^{[j]}, 0)$.

We study both the scenario where system of interest is measured after obtaining the output trajectory, as well as the one where only the output measurement is considered. In the former case, the algorithm guarantees that the classical Fisher information of the full measurement record is equal to the QFI of the system-output state. This claim is verified numerically by comparing the estimated classical Fisher information computed using the method described in section 4.6.2 with the *quantum* Fisher information of the system-output state. In the latter scenario, Proposition 1 insures that the loss of information compared to a 'full measurement' is bounded by a constant which does not depend on time.

As shown in Theorem 15, the system-output QFI scales linearly with time, i.e.,

$$F_\theta^{(s+o)}(n) = nf_\theta + o(n)$$

where the QFI *rate* is given by the equation (4.6) and the $o(n)$ term depends on the specific system parameters. Applying this to our model we obtain

$$f_{\theta_0} = \frac{8}{1-\sqrt{1-\lambda}}. \tag{4.23}$$

However, it turns out that for this specific model and parameter value, the QFI can be computed explicitly for *any* fixed n. The proofs of (4.23) and of the following lemma can be found in section 4.10.

Lemma 4. *The system-output QFI at $\theta = \theta_0$ is given by the formula*

$$F_\theta^{(s+o)}(n) \quad = \quad = \frac{8n}{1-a} + 4\left[\frac{2(a^2-a^n)}{(1-a)^2} - 2\frac{b^2(a-a^n)}{(1-a)^3} + \frac{2(a^2-a^{2n})}{b^2}\right] \tag{4.24}$$

where $a = \sqrt{1-\lambda}$ and $b = \sqrt{\lambda}$. In particular, the leading term in n is given by (4.23) while the remaining terms are bounded.

4.7.2 *Simulation studies for the simplified model*

We now present the results of our first numerical investigation consisting of 3 simulation studies using the simplified Markov model described above.

The first simulation study focuses on the comparison between the different notions of Fisher information: the system-output QFI, the CFI of the output trajectory in the non-adapted and adapted scenarios, and the CFI of the system-output measurement process in the adapted measurement scenario. The QFI is computed using the formula in Lemma 4 while the CFIs are estimated by sampling using the expression in Lemma 3.

The results are illustrated in Figure 4.3 where the different informations are plotted as a function of time (trajectory length) n, for $\lambda = 0.8, \phi = \pi/4$. The simulation confirms the fact that the adaptive algorithm achieves the QFI when the system is measured together with the output, while the CFI of the output trajectory (without measuring the system) provides a close approximation which only differs by a constant factor. In contrast, the CFI of the standard measurement has a smaller rate of increase; additional numerical work shows that the CFI rate can be improved by optimising the basis of the standard measurement but it does not achieve the QFI.

The second simulation study focuses on the 'measurement trajectory', i.e., the sequence of measurement settings produced in the adaptive measurement scenario. Since all measurement bases consist of vectors in the equatorial plane on the Bloch sphere, one can parametrise each basis by the polar coordinate φ of the basis vector $(|0\rangle \pm e^{i\varphi}|1\rangle)/\sqrt{2}$ for which φ belongs to a specified interval of length π. This is illustrated in the left panel of Figure 4.4. This parametrisation has the disadvantage that it does not reproduce the topology of the space of measurements which is that of a circle, leading to some jumps in measurement angles appearing to be larger than the actual 'distance' between measurement bases. To remedy this, in the right panel of Figure 4.4 we plot 2φ on circles of radius increasing linearly with time. We note that the initial steps of the trajectory show large variations which then 'stabilise' around a certain range of values. This

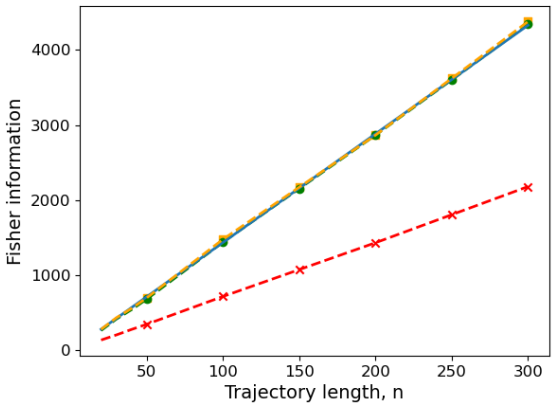

Figure 4.3: Fisher informations as function of output length n: quantum Fisher information (blue), classical Fisher information for the adaptive measurement with/without system measurement (orange/green), classical Fisher information for a regular (non-adaptive) measurement (red).

has to do with the fact that the initial angle can be chosen arbitrarily in this model as $B_1 = 0$, cf. section 4.5. Understanding the nature of this stochastic process remains an interesting topic of future research.

The third simulation study concerns the performance of the maximum likelihood estimator (MLE) in an adapted and non-adapted output measurement scenarios. The MLE is defined by

$$\hat{\theta}_n := \arg\max_{\tau} \; p_\tau(i_1, \dots i_n).$$

where the likelihood $p_\tau(i_1, \dots i_n)$ is computed as in equation (4.22). In numerics, we maximise the log-likelihood function which can be computed as the sum

$$\log p_\tau(i_i, \dots i_n) = \sum_{j=1}^{n} \log \|K_{i_j}^{[j]} \psi(i_1, \dots i_{j-1})\|^2$$

where $\psi(i_1, \dots i_{j-1})$ is the system's state conditional on the output trajectory (filter), cf. equation (4.21). The MLE accuracy is quantified by the mean square error (MSE) $E(n) = \mathbb{E}_\theta(\hat{\theta}_n - \theta)^2$, which is estimated empirically by averaging over a number of simulations runs to obtain $\hat{E}(n)$. To verify that the MSE scales as n^{-1} we plot the inverse empirical error $\hat{E}^{-1}(n)$ as a function of time. Figure 4.5 shows the inverse error for the adaptive and non-adaptive measurements together with the QFI (which is equal to the CFI of the adaptive measurement) and the CFI of the non-adaptive measurement. We note that for small values of n the inverse error is significantly lower that the corresponding Fisher information, but it approaches the latter for larger values of n. This suggests that the MLE achieves the Cramér-Rao bound asymptotically, which is not surprising since the MLE is known to be asymptotically optimal for independent samples as well as for certain classes of hidden Markov chains [21, 108]. However, proving its optimality for the adaptive measurement process remains an open problem. In addition to the mean square error, we looked at the distribution of the MLE.

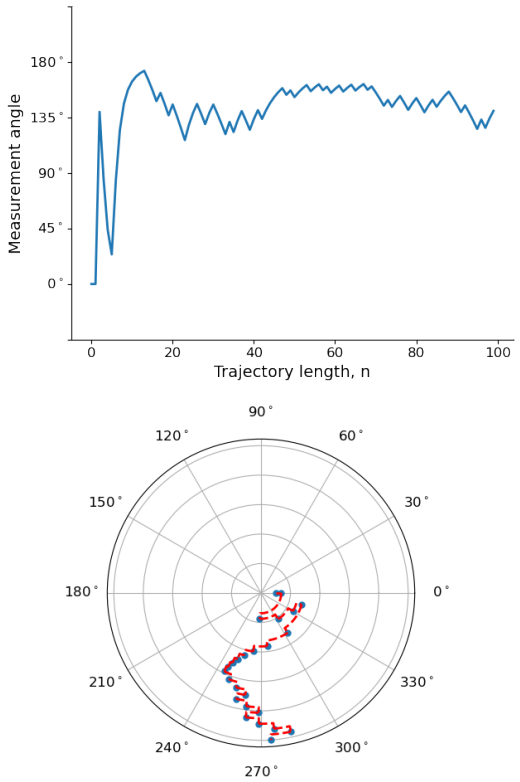

Figure 4.4: Adaptive measurement trajectory with basis angle φ on left and φ plotted on the circle on the right.

Figure 4.6 shows a histogram of the MLE based on $N = 10000$ simulations with $n = 200$, and it indicates that the MLE is approximately normally distributed. Based on this, it is reasonable to conjecture that the MLE is an *efficient estimator* [21, 108] (i.e.- it has asymptotically normal distribution with variance equal to the inverse of the Fisher information).

4.7.3 *The second numerical investigation using the full adaptive protocol*

In this section we present the results of the second numerical investigation which implements the full estimation scheme proposed in this chapter, including the preliminary estimation stage and the use of the coherent absorber. We use the same input-output model as described by the unitary in equation (4.20), but the data will be simulated at a true value of $\theta = 0.2$ instead of $\theta = 0$. While at $\theta = 0$ the system has a pure stationary state and the coherent absorber is not needed, away from this value the stationary state is mixed and we will apply the full protocol described in section 4.4.

In the first stage of the adaptive estimation procedure we use a fixed proportion q of the total sample size n to obtain a preliminary estimator θ_0 of θ by measuring each output unit in the standard basis. The parameter is estimated using the maximum likelihood method. In the second stage we apply the adaptive scheme with an absorber 'tuned' to the parameter value θ_0 (cf. section 4.4). The system and absorber are prepared in the pure stationary state at θ_0, the dynamics is run for time $(1 - q)n$ and the output is measured according to the adaptive

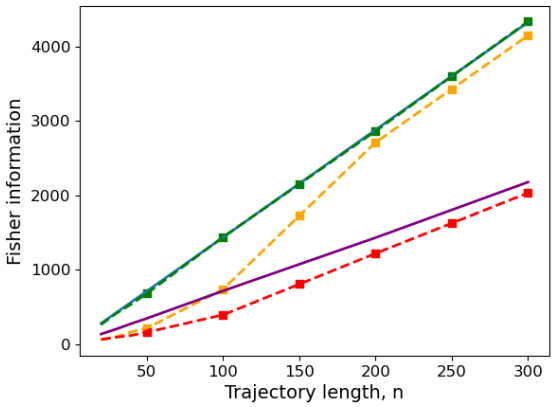

Figure 4.5: Comparison of different Fisher informations of the output state as function of trajectory length n, obtained by averaging over $N = 10000$ trajectories, at $\lambda = 0.8, \varphi = \pi/4$. The quantum Fisher information (QFI) (blue) and the classical Fisher information (CFI) of adapted measurement (green) completely overlap due to optimality. The inverse MLE error of the adapted measurement (orange) approaches the QFI for large n. Similarly, for the standard measurement, the inverse MLE error (red) approaches the CFI (purple) for large n.

measurement filter algorithm. The maximum likelihood estimator for the stage one *and* two data is then computed.

For comparison, we also run a non-adaptive scheme where the output is measured in the standard basis for the whole duration n, without using an absorber. The maximum likelihood estimator is again computed from the measurement data. In both experiments, the mean square error of the MLE is estimated by averaging over 1000 repetitions.

Unlike the setup of the previous numerical study, the estimation of the classical Fisher information of the adaptive measurement process was too costly and is not included in the study. As a proxy for the classical Fisher information we plot the average value of the *observed Fisher information* for each of the two simulations. For a given measurement run $(i_1, \dots i_n)$, the observed Fisher information is defined as the second derivative of the log-likelihood function evaluated at the maximum likelihood estimator $\hat{\theta}_n$:

$$I_{obs}(i_1, \dots i_n) = -\left.\frac{d^2 \log p_\theta(i_1, \dots i_n)}{d\theta^2}\right|_{\theta=\hat{\theta}_n}$$

While a full theoretical justification of I_{obs} goes beyond the scope of this chapter, we note that the use of the observed Fisher information for independent identically distributed data is well grounded in statistical methodology and is closely relate to the asymptotic normality property [51].

Figure 4.7 shows the results of the numerical experiments, for two values $q = 0.15$ and $q = 0.25$ of the proportion of samples used in the preliminary estimation stage. As before, we plot the inverse of the estimated mean square error for the simple measurement (green line) and the adaptive measurement (blue line). In addition we plot the leading contribution to the quantum Fisher information $nf(\theta)$ (purple line) computed using the results in Theorem 15, which

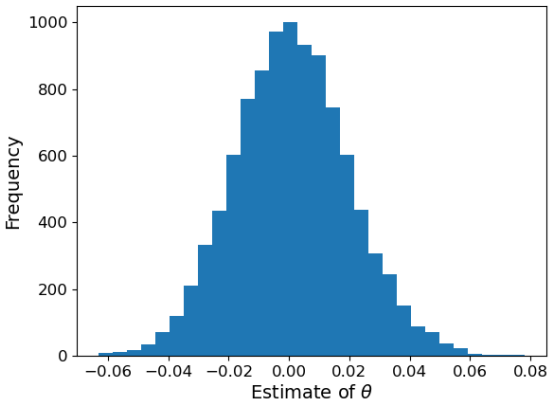

Figure 4.6: Histogram of the MLE distribution in the adaptive measurement scenario, with $n = 200, \theta = 0$ and 10000 samples, at $\lambda = 0.8, \varphi = \pi/4$.

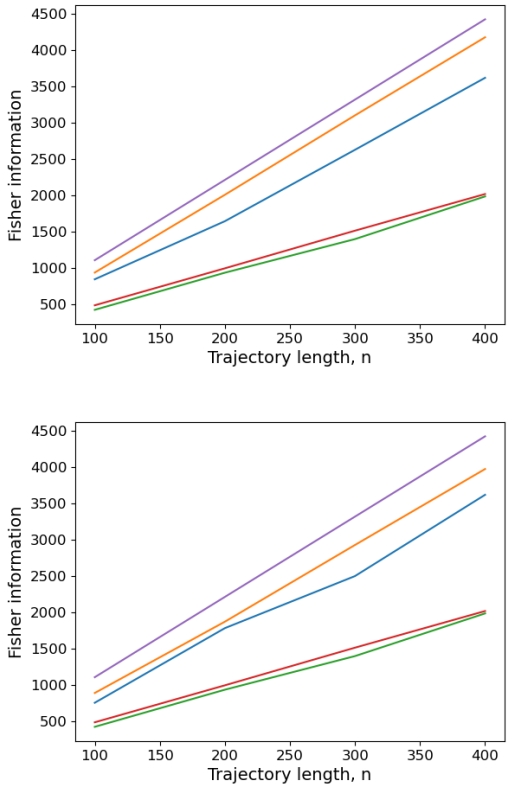

Figure 4.7: Comparison of different Fisher informations of the output state as function of trajectory length n, in the fully adaptive numerical investigation, for two proportions of samples used in the preliminary stage: $q = 0.15$ (top panel) and $q = 0.25$ (bottom panel). The results are obtained by averaging over $N = 1000$ trajectories, at $\theta = 0.2$ and $\lambda = 0.8, \varphi = \pi/4$. We plot: the asymptotic QFI (purple line), observed Fisher information for adaptive measurement (orange line) and simple measurement (red line), inverse MSE of the MLE for adaptive measurement (blue line) and simple measurement (green line).

provides the asymptotic slope of the actual output quantum Fisher information. The average observed Fisher information is plotted for both the simple (red line) and adaptive (orange line) measurement setups. We note a good agreement between the inverse mean square error of the MLEs and the observed Fisher informations. For the adaptive measurements, the observed Fisher information also shows the same slope as the asymptotic Fisher information, as expected. We also note that in the adaptive measurement, the mean square error does not quite achieve the (observed) quantum Fisher information for the range of times we considered, although it gets closer to it in the case $q = 0.25$ (right panel). We speculate that this may be related to a number of factors such as the choice of q for different sample sizes, the details of the actual measurement procedure implemented in practice (in contrast to the theoretical prescription) and even the implementation of the MLE. For instance, in practice it may be beneficial to implement several adaptive estimation stages where the preliminary estimator is gradually improved and used in tuning the absorber in the next stage. All these remain interesting questions which are worth investigating in more detail and on a case by case basis. However, these are somewhat separate issues from that of designing adaptive measurements that achieve the QFI, which was the main focus of this work.

4.8 CONCLUSIONS AND OUTLOOK

In this chapter we developed an efficient iterative algorithm for optimal estimation of dynamical parameters of a discrete-time quantum Markov chain, using adaptive sequential measurements on the output. The algorithm builds on the general measurement scheme of [183] which achieves the quantum Fisher information for pure state models of multipartite systems with one dimensional unknown parameters. However, unlike the scheme of [183] which requires manipulations involving the full multipartite state, the proposed algorithm only involves computations on $d^2 \cdot k$-dimensional systems where d and k are the dimensions of the open system and noise unit respectively. Therefore, the method can be readily applied to Markov parameter estimation for large output sizes. The algorithm exploits the Markovian structure of the dynamics to sequentially compute optimal measurement bases in terms of a single time-dependent 'measurement filter' operator, which is updated in a way that is reminiscent of a state filter. One of the key ingredients of the proposed scheme is the use of a coherent quantum absorber [157] which reduces the estimation problem to one concerning a system with a pure stationary state. We considered both output and output-system measurements scenarios and we showed that while the former achieves the full quantum Fisher information, the latter is short of this by just a fixed constant, and in particular both have the same scaling with time. Our theoretical results are confirmed by numerical simulations using a simplified model related to the amplitude decay channel. We also presented results from 'full simulation' studies involving the use of the coherent absorber.

Our discrete-time procedure raises interesting questions about the possibility to design realistic optimal sequential measurements in continuous-time dynamics. In principle the scheme can be applied to continuous time by using

time-discretisation techniques [111, 128]. Although we did not treat this in detail here, we can readily make several observations in the special case of a single Bosonic I-O channel. For small enough time intervals δt, the field can be approximated by a noise unit $\mathbb{C}^2$ with basis vectors representing the vacuum and a one-photon state respectively. The proposed measurements consist of projections whose corresponding Bloch vectors are in the equatorial plane; loosely speaking, this corresponds to an adaptive homodyne measurement with a time-dependent angle. Our preliminary investigations indicate that the behaviour of the time-dependent angle ranges from deterministic evolution to a noisy stochastic process which does not appear to be of diffusive type. This raises the question whether the remaining freedom in choosing the measurement bases can be used to improve the adaptive algorithm and produce a more regular measurement process. Indeed, the second defining condition for the measurement vectors can be relaxed, allowing for more general classes of optimal measurements, which may be more suitable for continuous-time detection. This will be the topic of a future investigation.

Another important open question concerns the extension to chains with mixed input states, or more multiple inputs, of which only some are observed. We speculate that for small departures from the current scheme, the algorithm will be quasi-optimal for some time interval but will be sub-optimal in the long time limit. In this case, restarting the evolution and measurement filter at regular intervals may be more efficient.

From a theoretical viewpoint, it is important to understand the mathematical properties of the stochastic processes introduced here, the adaptive measurement and the measurement trajectory. Finally, it is intriguing to consider to what extent the proposed method can adapted to multi-parameter estimation and to general (time-dependent) matrix product states, as opposed to stationary output states of Markov processes.

4.9 PROOF OF PROPOSITION 1

We start by using a general Fisher information identity for bipartite systems. Consider a generic pure state model $|\psi_\theta\rangle \in \mathcal{H}_s \otimes \mathcal{H}_o$ and let $\{|e_i^o\rangle\}$ and $\{|e_j^s\rangle\}$ be optimal bases in the 'output' and 'system' subsystems, for estimating θ at a particular value θ_0, as prescribed by [183]. Assume we perform the 'output' measurement and let X denote the outcome whose distribution is

$$\mathbb{P}_\theta(X = i) = \langle\psi_\theta|\mathbf{1} \otimes P_i|\psi_\theta\rangle, \qquad P_i = |e_i^o\rangle\langle e_i^o|.$$

The conditional state of the 'system' given $X = i$ is

$$|\psi_\theta(i)\rangle = \frac{\mathbf{1} \otimes P_i|\psi_\theta\rangle}{\|\mathbf{1} \otimes P_i\psi_\theta\|}$$

and this state contains the 'remaining' information about θ. The following inequality bounds the total available information as

$$I_\theta(X) + \mathbb{E}_X F(\psi_\theta(X)) \leq F_\theta$$

where the first term on the left side is the classical Fisher information of the outcome distribution $\mathbb{P}_\theta$ and the second is the expected QFI of the conditional

state $|\phi_\theta^i\rangle$. Consider now the second measurement $|e_j^s\rangle$ on the system and let Y be its outcome. Then

$$F_\theta = I_\theta(X) + \mathbb{E}_X I_\theta(Y|X) \leq I_\theta(X) + \mathbb{E}_X F(\psi_\theta(X)) \leq F_\theta$$

where the first equality is due to measurement optimality, while the second is the inequality between classical and quantum information. This implies that

$$F_\theta - I_\theta(X) = \mathbb{E}_X F_\theta(\psi_\theta(X)). \tag{4.25}$$

We now consider the Markov setup in which the system+absorber play the role of 'system' while the n noise units are the 'output'. We assume that the output and the system+absorber are measured according to the optimal scheme presented in section 4.5. The joint state is given by

$$|\Psi_\theta(n)\rangle = W_\theta^{(n)} \cdot \cdots \cdot W_\theta^{(1)} |\psi \otimes \chi^{\otimes n}\rangle$$

and the conditional states are

$$|\psi_\theta(i_1, \ldots, i_n)\rangle = \frac{K_{\theta,i_n}^{[n]} \ldots K_{\theta,i_1}^{[1]} |\psi\rangle}{\sqrt{p_\theta(i_1, \ldots i_n)}}.$$

We will show that at $\theta = \theta_0$ the left side of (4.25) is bounded by a constant which does not depend on n. For simplicity, whenever possible we will use the compact notations such as

$$K_{\mathbf{i}} := K_{\theta,i_n}^{[n]} \ldots K_{\theta,i_1}^{[1]}, \qquad \text{and} \qquad |e_{\mathbf{i}}\rangle = |e_{i_n}^{[n]} \otimes \cdots \otimes e_{i_1}^{[1]}\rangle.$$

Recall that by design the following condition holds $W_{\theta_0}|\psi \otimes \chi\rangle = |\psi \otimes \chi\rangle$, which implies $|\Psi_{\theta_0}(n)\rangle = |\psi \otimes \chi^{\otimes n}\rangle$ and also $K_{i_j}^{[j]}|\psi\rangle = c_{i_j}^{[j]}|\psi\rangle$ for some constants $c_{i_j}^{[j]}$.

Recall that for a pure state model $|\psi_\theta\rangle$ the QFI is given by

$$F_\theta = 4\left(\|\dot\psi_\theta\|^2 - |\langle\dot\psi_\theta|\psi_\theta\rangle|^2\right) = 4\|\psi_\theta^\perp\|^2, \qquad |\psi_\theta^\perp\rangle = |\dot\psi_\theta\rangle - P_{\psi_\theta}|\dot\psi_\theta\rangle.$$

Therefore, the expected system QFI on the left side of (4.25) is given by

$$F_s(\theta_0) = 4\sum_{\mathbf{i}} p(\mathbf{i}) \|\psi_{\theta_0}^\perp(\mathbf{i})\|^2 \tag{4.26}$$

where

$$|\psi_{\theta_0}^\perp(\mathbf{i})\rangle = |\dot\psi_{\theta_0}(\mathbf{i})\rangle - P_\psi|\dot\psi_{\theta_0}(\mathbf{i})\rangle$$

For simplicity we now drop the subscript θ_0 and we have

$$|\dot\psi(\mathbf{i})\rangle = \frac{\dot K_{\mathbf{i}}|\psi\rangle}{\sqrt{p(\mathbf{i})}} - \frac{1}{2}\frac{K_{\mathbf{i}}|\psi\rangle}{p^{3/2}(\mathbf{i})}\dot p(\mathbf{i}) = \frac{\dot K_{\mathbf{i}}|\psi\rangle}{\sqrt{p(\mathbf{i})}} - \frac{1}{2}\frac{c(\mathbf{i})\dot p(\mathbf{i})}{p^{3/2}(\mathbf{i})}|\psi\rangle$$

and therefore

$$|\psi^\perp(\mathbf{i})\rangle = \frac{1}{\sqrt{p(\mathbf{i})}}\left(I - P_\psi\right)\dot K_{\mathbf{i}}|\psi\rangle.$$

Equation (4.26) becomes

$$F_s(\theta_0) = 4\sum_{\mathbf{i}} \left\| P_\psi^\perp \dot K_{\mathbf{i}} \psi \right\|^2 = 4\left\| (P_\psi^\perp \otimes I)\dot\Psi(n) \right\|^2.$$

Now

$$|\Psi(n)\rangle = \sum_{j=1}^{n} W^{(n)} \dots W^{(j+1)} \dot{W}^{(j)} W^{(j-1)} \dots W^{(1)} |\psi \otimes \chi^{\otimes n}\rangle$$
$$= \sum_{j=1}^{n} W^{(n)} \dots W^{(j+1)} \dot{W}^{(j)} |\psi \otimes \chi^{\otimes n}\rangle$$

and by triangle inequality

$$F_s(\theta_0) \leq 4 \left(\sum_{j=1}^{n} \left\| (P_\psi^\perp \otimes I) W^{(n)} \dots W^{(j+1)} \dot{W}^{(j)} \psi \otimes \chi^{\otimes n} \right\| \right)^2$$

Let $|\dot{\Psi}(1)\rangle := \dot{W}|\psi \otimes \chi\rangle$ and let $\tau := \mathrm{Tr}_1(\dot{\Psi}(1)\rangle\langle\dot{\Psi}(1)|)$. Then

$$\left\| (P_\psi^\perp \otimes I) W^{(n)} \dots W^{(j+1)} \dot{W}^{(j)} \psi \otimes \chi^{\otimes n} \right\|^2$$
$$= \mathrm{Tr}_{0,1,\dots n-j} \left((P_\psi^\perp \otimes I) W^{(n-j)} \dots W^{(1)} \tau W^{(1)*} \dots W^{(n-j)*} \right)$$
$$= \mathrm{Tr}_0 (P_\psi^\perp T^{n-j}(\tau))$$

where in the last equality we have used the definition of the transition operator T of the system+absorber. Assuming that T is ergodic we have

$$T^n(\tau) \to P_\psi$$

exponentially fast with n so that

$$\mathrm{Tr}_0 (P_\psi^\perp T^{n-j}(\tau)) \leq a^{2(n-j)}$$

for some $a < 1$. Therefore

$$F_s(\theta_0) \leq 4 \left(\sum_{j=1}^{n} a^{n-j} \right)^2 \leq 4 \frac{1}{(1-a)^2}$$

Note that if the spectral gap of T becomes small then the convergence to stationarity is slower and the upper bound increases. □

4.10 COMPUTATION OF FINITE TIME SYSTEM-OUTPUT QFI

From (4.5) and (4.1) we have

$$\begin{aligned} F_\theta^{(s+o)}(n) &= 4 \|\dot{\Psi}_\theta(n)\|^2 \\ &= 4 \sum_{i_1,\dots i_n} \left\| \sum_{j=1}^{n} K_{i_n} \dots \dot{K}_{i_j} \dots K_{i_1} \psi \right\|^2 \end{aligned} \tag{4.27}$$

where K_i are the (fixed) Kraus operators with respect to the standard basis, and $|\psi\rangle = |0\rangle$. Our specific model has the feature that both K_i and $\dot{K}_i$ map the basis vectors into each other:

$$K_0|0\rangle = |0\rangle \quad \dot{K}_0|0\rangle = i|1\rangle \quad K_0|1\rangle = \sqrt{1-\lambda}|1\rangle \quad \dot{K}_0|1\rangle = i\sqrt{1-\lambda}|0\rangle$$
$$K_1|0\rangle = 0 \quad \dot{K}_1|0\rangle = |1\rangle \quad K_1|1\rangle = \sqrt{\lambda} e^{i\phi}|0\rangle \quad \dot{K}_1|1\rangle = 0$$

This allows to compute the QFI explicitly by noting the terms in the sum (4.27) with more that two indices equal to 1 have zero contribution. The remaining terms can be computed as follows. The term with only zero indices is

$$F^{(0)} = 4\left\|\sum_{j=1}^{n} K_0 \dots \dot{K}_0 \dots K_0|0\rangle\right\|^2 = 4\left|\sum_{j=1}^{n} i a^{n-j}\right|^2 = 4\left(\frac{1-a^n}{1-a}\right)^2 \tag{4.28}$$

where $a = \sqrt{1-\lambda}$.

Consider now a sequence $(0, \dots, , 0, 1, 0, \dots 0)$ with a single one on position l. Since $K_0 \dots \dot{K}_0 \dots K_1 \dots K_0|0\rangle = 0$ the only contributing terms will be those with derivative on the first $(l-1)$ K_0s or on K_1. This gives

$$\begin{aligned} F^{(1)} &= 4\sum_{l=2}^{n}\left\|i\sum_{r=1}^{l-1} e^{i\phi}a^{l-1-r}b|0\rangle + a^{n-l}|1\rangle\right\|^2 + 4\|a^{n-1}|1\rangle\|^2 \\ &= 4\sum_{l=2}^{n}\left(b^2\frac{1-a^{l-1}}{1-a} + a^{2(n-l)}\right) + 4a^{2(n-1)} \\ &= 4(n-1)\frac{b^2}{(1-a)^2} + 4\frac{a^2-a^{2n}}{(1-a)^2} \\ &\qquad - 8\frac{b^2(a-a^n)}{(1-a)^3} + 4\frac{1-a^{2(n-1)}}{b^2} + 4a^{2(n-1)} \end{aligned} \tag{4.29}$$

where $b = \sqrt{\lambda}$.

Finally, consider the sequences of the type $(0, \dots 0, 1, 0, \dots 0, 1, 0 \dots 0)$ with 1s on positions $1 \le i < k \le n$. In this case the nonzero contributions come from terms where the derivative is on positions $j = i$. The Fisher contribution is

$$\begin{aligned} F^{(2)} &= 4\sum_{1\le i<k\le n}\|e^{i\phi}\sqrt{\lambda}(1-\lambda)^{(k-i-1)/2}|0\rangle\|^2 \\ &= 4\sum_{1\le i<k\le n} b^2 a^{2(k-i-1)} = 4(n-1) - 4\frac{a^2-a^{2n}}{b^2}. \end{aligned} \tag{4.30}$$

Adding together the contributions (4.28), (4.29) and (4.30) we obtain the total QFI

$$\begin{aligned} F_\theta^{(s+o)}(n) &= F^{(0)} + F^{(1)} + F^{(2)} \\ &= \frac{8n}{1-a} \\ &\qquad + 4\left[\left(\frac{1-a^n}{1-a}\right)^2 - \frac{b^2}{(1-a)^2} + \frac{a^2-a^{2n}}{(1-a)^2}\right. \\ &\qquad\qquad \left. -2\frac{b^2(a-a^n)}{(1-a)^3} + \frac{1-a^{2(n-1)}}{b^2} + a^{2(n-1)}\right] \\ &\qquad - 4\left[1 + \frac{a^2-a^{2n}}{b^2}\right] \end{aligned} \tag{4.31}$$

where the leading term is consistent with the QFI rate formula (4.23).

5

OPTIMAL ESTIMATION OF PURE STATES WITH DISPLACED-NULL MEASUREMENTS

5.1 INTRODUCTION AND MAIN RESULTS

As we've seen in the previous chapters, a common feature of many quantum estimation problems is that 'optimal' measurements depend on the unknown parameter, so they can only be implemented approximately, and the optimality is at best achieved in the limit of large 'sample size'. This raises the question of how to interpret theoretical results such as the QCRB [15, 25, 93, 95, 124, 179] and how to design adaptive measurement strategies which attain the optimal statistical errors in the asymptotic limit. When multiple copies of the state are available, the standard strategy is to use a sub-sample to compute a rough estimator and then apply the optimal measurement corresponding to the estimated value. Indeed this works well for the case of the symmetric logarithmic derivative [71], an operator which saturates the quantum Cramér-Rao bound (QCRB) for one-dimensional parameters. However, the QCRB fails to predict the correct *attainable* error for quantum metrology models which consist of correlated states and exhibit Heisenberg (quadratic) scaling for the mean square error [80]. This is due to the fact that in order to saturate the QCRB one needs to know the parameter to a precision comparable to what one ultimately hopes to achieve.

In this chapter we uncover a somewhat complementary phenomenon, where the usual adaptive strategy fails *precisely* because it is applied to a 'good' guess of the true parameter value. This happens in the standard multi-copy setting when estimating a pure state by means of 'null measurements', where the experimenter aims to measure in a basis that contains the unknown state. While this can only be implemented approximately, the technique is known to exhibit certain Fisher-optimality properties [115, 136, 137] and has the intuitive appeal of 'locking' onto the correct value as outcomes corresponding to other measurement vectors become more and more unlikely.

In Theorem 16, which is our first main result, we show that the standard adaptive strategy in which the parameter is first estimated on a sub-sample and then the null-measurement for this rough value is applied to the rest of the ensemble, fails to saturate the QCRB, and indeed does not attain the standard rate of precision. Our result shows the importance of accompanying mathematical properties with clear operational procedures that allow us to draw statistical conclusions; this provides another example of the limitations of the 'local' estimation approach based on the quantum Cramér-Rao bound [164]. Indeed the reason behind the failure of the standard adaptive strategy is the fact that null-measurements suffer

from non-identifiability issues when the true parameter and the rough preliminary estimator are too close to each other, i.e., when the latter is a reasonable estimator of the former.

Fortunately, it turns out that the issue can be resolved by deliberately shifting the measurement reference parameter away from the estimated value by a vanishingly small but sufficiently large amount to resolve the non-identifiabilty issue. Using this insight we devise a novel adaptive measurement strategy which achieves the Holevo bound for arbitrary multi-parameter models, asymptotically with the sample size. This second main result is described in Theorem 17. In particular our method can be used to achieve the quantum Cramér-Rao bound for models where this is achievable, which was the original theme of [115, 136, 137]. The validity of the displaced-null strategy goes beyond the setting of the estimation with independent copies and has already been employed for optimal estimation of dynamical parameters of open quantum systems by counting measurements [177]. The extension of our present results to the setting of quantum Markov chains will be presented in a forthcoming publication [75]. In the rest of this section we give a brief review of the main results of this chapter.

The quantum Cramér-Rao bound and the symmetric logarithmic derivative

As we saw in Chapter 3, the quantum estimation problem is formulated as follows: given a quantum system prepared in a state ρ_θ which depends on an unknown (finite dimensional) parameter $\theta \in \Theta$, one would like to estimate θ by performing a measurement M and constructing an estimator $\hat{\theta} = \hat{\theta}(X)$ based on the (stochastic) outcome X. The Cramér-Rao bound [170, 185] shows that for a given measurement M, the covariance of any unbiased estimator is lower bounded as $\mathrm{Cov}(\hat{\theta}) \geq I_M^{-1}(\theta)$ where $I_M(\theta)$ is the classical Fisher information (CFI) of the measurement outcome.

Optimising over all possible measurements, the CFI $I_M(\theta)$ is upper bounded by the quantum Fisher information (QFI) $F(\theta)$ - an intrinsic property of the quantum statistical model $\{\rho_\theta\}_{\theta\in\Theta}$. By combining the two bounds we obtain the celebrated quantum Cramér-Rao bound (QCRB) [15, 25, 93, 95, 124, 179] $\mathrm{Cov}(\hat{\theta}) \geq F^{-1}(\theta)$. For one dimensional parameters the QFI can be (formally) achieved by measuring an observable $\mathcal{L}_\theta$ called the symmetric logarithmic derivative (SLD), defined as the solution of the Lyapunov equation $\frac{d\rho_\theta}{d\theta} = \frac{1}{2}(\rho_\theta \mathcal{L}_\theta + \mathcal{L}_\theta \rho_\theta)$. However, since the SLD depends on the unknown parameter θ, this measurement cannot be performed without its prior knowledge, and the formal achievability is unclear without further operational specifications.

Fortunately, this apparent circularity issue can be solved in the context of asymptotic estimation [92]. In most practical applications one does not measure a single system but deals with (large) ensembles of identically prepared systems, or multi-partite correlated states as in quantum enhanced metrology [129, 184] and continuous time estimation of Markov dynamics [63, 83, 84, 177]. Here one considers issues such as the scaling of errors with sample size, collective versus separable measurements, and whether one needs fixed or adaptive measurements. In particular, in the case of *one-dimensional* models, the QCRB can be achieved *asymptotically* with respect to the size n of an ensemble of independent identically prepared systems, by using a two steps *adaptive* measurement strategy [71]. In

the first step, a preliminary 'rough' estimator $\tilde{\theta}_n$ is computed by measuring a sub-ensemble of $\tilde{n} = o(n)$ systems, after which the SLD for parameter value $\tilde{\theta}_n$ (our best guess at the optimal observable $\mathcal{L}_\theta$) is measured on each of the remaining systems. In the limit of large sample size n, the preliminary estimator $\tilde{\theta}_n$ approaches θ and the two step procedure achieves the QCRB in the sense that the mean square error (MSE) of the final estimator scales as $(nF(\theta))^{-1}$.

By implicitly invoking the above adaptive measurement argument, the quantum estimation literature has largely focused on computing or estimating the QFI of specific models, or designing input states which maximise the QFI in quantum metrology settings. However, as shown in [80], the adaptive argument breaks down for models exhibiting quadratic (or Heisenberg) scaling of the QFI where the *achievable* MSE is larger by a constant factor compared to the QCRB prediction, even asymptotically. In this work we show that similar care needs to be taken even when considering standard estimation problems involving ensembles of independent quantum systems and standard error scaling.

Null measurements and their standard adaptive implementation

Specifically, we revisit the problem of estimating a parameter of a *pure state model* $\{|\psi_\theta\rangle\}_{\theta\in\Theta}$ and analyse a measurement strategy [115, 136, 137], which we broadly refer to as *null measurement*. The premise of the null measurement is the observation that if one measures $|\psi_\theta\rangle$ in an orthonormal basis $\mathcal{B}(\theta) := \{|v_1\rangle, \ldots, |v_d\rangle\}$ such that $|v_1\rangle = |\psi_\theta\rangle$ then the only possible outcome is $X = 1$ and all other outcomes have probability zero. Since θ is unknown, in practice one would measure in a basis $\mathcal{B}(\tilde{\theta})$ corresponding to an approximate value $\tilde{\theta}$ of the true parameter θ, and exploit the occurrence of low probability outcomes $X \neq 1$ in order to estimate the deviation of θ from $\tilde{\theta}$. This intuition is supported by the following property which is a specialisation to one-dimensional parameters of a more general result derived in [115, 136, 137]: as $\tilde{\theta}$ approaches θ, the classical Fisher information $I_{\tilde{\theta}}(\theta)$ associated with $\mathcal{B}(\tilde{\theta})$ converges to the QFI $F(\theta)$. This implies that null measurements can achieve MSE rates scaling as n^{-1} with constants that are *arbitrarily close* to $F^{-1}(\theta)$, by simply measuring all n systems of an ensemble in a basis $\mathcal{B}(\tilde{\theta})$ with a *fixed* $\tilde{\theta}$ that is close to θ:

$$n\mathbb{E}_\theta[(\hat{\theta}_n - \theta)^2] \to I_{\tilde{\theta}}^{-1}(\theta) \approx F^{-1}(\theta).$$

Do null measurements actually achieve the QCRB (asymptotically) or just 'come close' to it? In absence of a detailed multi-copy operational interpretation in [115, 136, 137], the most natural strategy is to apply the same two step adaptive procedure which worked well in the case of the SLD measurement. A preliminary estimator $\tilde{\theta}_n$ is first computed by measuring $\tilde{n}$ systems and the rest of the ensemble is subsequently measured in the basis $\mathcal{B}(\tilde{\theta}_n)$. Since $I_{\tilde{\theta}_n}(\theta)$ converges to $F(\theta)$ as $\tilde{\theta}_n$ approaches θ, it would appear that the QCRB is achieved asymptotically. One of our main results is to show that this adaptive procedure actually *fails* to achieve the QCRB even in the simple qubit model

$$|\psi_\theta\rangle = \cos\theta|0\rangle + \sin\theta|1\rangle, \tag{5.1}$$

thus providing another example where caution is needed when using arguments based on Fisher information, see [164] for other examples.

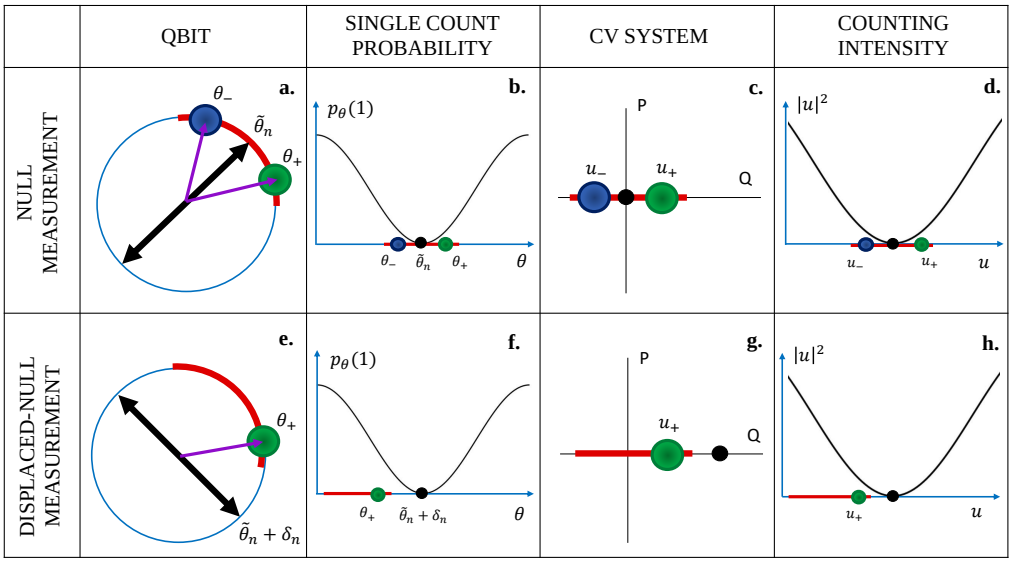

Figure 5.1: The figure illustrates the non-identifiability problem occurring with null measurement (first row) and how it is fixed by displaced-null measurement (second row). In the first column the red arc on the xz Bloch sphere circle (in blue) represents the set of parameters after localisation (confidence interval), the green disk represents the true parameter value $\theta = \theta_+$ and the blue disk (panel a) is the parameter θ_- which is indistinguishable from the true one, in the null basis. The black arrow represents the chosen measurement basis. The second column displays a plot of the single count probability as a function of the parameter: in the null measurement case such a function is not injective on the set of parameters determined after the localisation (panel b). The third column shows the phase space of a Gaussian model consisting of coherent states with unknown displacement along the Q axis: the red interval is the parameter space, the black dot corresponds to the number operator measured, the green disk to the true coherent state and the blue disk (panel c) is the coherent state which is indistinguishable from the true one in the null measurement case. The last column plots the intensity of the number operator as a function of the coherent state amplitude.

More precisely, we show that if the preliminary estimator $\tilde{\theta}_n$ is reasonably good (cf. section 5.3 for precise formulation), any final estimator $\hat{\theta}_n$ computed from the outcomes of the null measurement $\mathcal{B}(\tilde{\theta}_n)$ is not only suboptimal but does not even achieve the standard n^{-1} estimation MSE rate. The reason for the radically different behaviors of the SLD and null meaurement settings is that the latter suffers from a *non-identifiability* problem when the parameter $\tilde{\theta}$ (which determines the null basis) is close to θ. Indeed, since at $\tilde{\theta} = \theta$ the null measurement has a deterministic outcome, for $\tilde{\theta} \approx \theta$ the outcome probabilities are quadratic in $\epsilon = \theta - \tilde{\theta}$ and therefore, the parameters $\theta_\pm = \tilde{\theta} \pm \epsilon$ cannot be distinguished (at least in second order). If $\tilde{\theta}_n$ is a reasonably good estimator, then $\epsilon_n = |\theta - \tilde{\theta}_n|$ is of the order $\tilde{n}^{-1/2}$, so the error in estimating θ is at least of the order of the distance $|\theta_+ - \theta_-|$ between the two undistinguishable candidate parameters $\theta_\pm = \tilde{\theta}_n \pm \epsilon_n$, which scales as $\tilde{n}^{-1/2}$ instead of $n^{-1/2}$. Since $\tilde{n} = o(n)$ the mean square error decreases slower that the standard rate n^{-1}. This argument is illustrated in Figure 5.1a. for the simple case of the qubit rotation model (5.1) which is discussed in detail in section 5.3.

Asymptotic optimality of displaced-null measurements

Fortunately, the above explanation offers an intuitive solution to the non-identifiability problem. Assuming that the preliminary estimator $\tilde{\theta}_n$ satisfies standard concentration properties (e.g.- asymptotic normality), one finds that θ belongs (with high probability) to a confidence interval I_n centered at $\tilde{\theta}_n$, whose length is slightly larger than the estimation uncertainty $\tilde{n}^{-1/2}$. Therefore by displacing $\tilde{\theta}_n$ by a (vanishingly small) amount $\delta_n > 0$ that is larger than this uncertainty, we can make sure that I_n lies at the left side of $\theta'_n := \tilde{\theta}_n + \delta_n$ and therefore measuring in the basis $\mathcal{B}(\theta'_n)$ circumvents the non-identifiability issue. This is illustrated in panels e. and f. of Figure 5.1.

The main aim of the chapter is to investigate this method which we call a *displaced-null* measurement strategy and derive asymptotic optimality results for the resulting estimators. In section 5.4.1 we show that the displaced-null measurement achieves the QCRB in the one-parameter qubit model for which the standard adaptive procedure failed; the corresponding second stage estimator is a simple average of measurement outcomes and satisfies asymptotic normality, thus allowing practitioners to define asymptotic confidence intervals.

In section 5.6 we extend the null-measurement strategy to *multi-parameter* models of pure qudit states. In this case, the QCRB is typically not attainable even asymptotically due to the incompatibility of optimal measurements corresponding to different parameter components. However, we show that the Holevo bound [95] *can* be achieved asymptotically. We first consider the task of estimating a completely unknown pure state with respect to the Bures (fidelity) distance. In this case we show that the Holevo bound can be achieved by using two separate displaced-null measurements, for the real and imaginary parts of the state coefficients with respect to a basis containing $|\psi_{\theta'_n}\rangle$ as a vector. The second task is to estimate a general m-dimensional model with respect to an arbitrary locally quadratic distance on the parameter space. Here we show that the Holevo bound is achievable by applying displaced-null measurements on copies of the systems coupled with an ancilla in a fixed state. The proof relies

on the intuition gained from quantum local asymptotic normality theory and its use in establishing the achievability of the Holevo bound [49, 70] by mapping the ensemble onto a continuous variables system. However, unlike the latter, the displaced-null technique only involves separate projective measurements on system-ancilla pairs.

Finally, in section 5.6.6 we show that for multiparameter models where the QCRB is achievable, this can be done using displaced-null measurements. This puts related results of [115, 136, 137] on a firm operational basis.

Local asymptotic normality perspective

The theory of quantum local asymptotic normality (QLAN) [70, 88, 89, 102] offers an alternative perspective on the displaced-null measurements strategy outlined above. In broad terms, QLAN is a statistical tool that allows us to approximate the independent identically distributed (IID) model describing the joint state of an ensemble of systems, by a single continuous variables Gaussian state whose mean encodes information about the unknown parameter (cf. sections 5.5.1 and 5.6.2 for more details). By applying this approximation, the null measurement problem discussed earlier can be cast into a Gaussian version formulated as follows. Suppose we are given a one-mode continuous variables system prepared in a coherent state $|u\rangle$ with unknown displacement $u \in \mathbb{R}$ along the Q axis, and assume that $|u| \leq a_n$ for some bound a_n which diverges with n. At $u = 0$, the system state is the vacuum, and the measurement of the number operator N is a null measurement (see Figure 5.1c.). However, for a given $u \neq 0$ the number operator has Poisson distribution with intensity $|u|^2$, and therefore cannot distinguish between parameters $u_\pm := \pm u$, cf. Figure 5.1d. This means that any estimator will have large MSEs of order a_n^2 for large values of u. In contrast, measuring the quadrature Q produces (optimal) estimators with fixed MSE given by the vacuum fluctuations. However, the non-identifiability problem of the counting measurement can be lifted by displacing the coherent state along the Q axis by an amount $\Delta_n > a_n$ and then measuring N. Equivalently, one can measure the corresponding displaced number operator on the original coherent state as illustrated in panels g. and h. of Figure 5.1. In this case the intensity $(u - \Delta_n)^2$ is in one-to-one correspondence with u so the parameter is identifiable. Moreover, for large n, the counting measurement can be linearised and becomes equivalent to measuring the quadrature Q, a well known fact from homodyne detection [109].

QLAN shows that the Gaussian problem discussed above is the asymptotic version of the one-parameter qubit rotation model (5.1) which we used earlier to illustrate the concept of approximate and displaced null measurements. The coherent state $|u\rangle$ corresponds to all qubits in the state $|\psi_{u/\sqrt{n}}\rangle$ (assuming for simplicity that $\tilde{\theta}_n = 0$ and writing $\theta = u/\sqrt{n}$). The number operator corresponds to measuring in the standard basis, which is an exact null measurement at $u = 0$. On the other hand, the displaced number operator corresponds to measuring in the rotated basis with angle $\delta_n = n^{-1/2}\Delta_n$.

The same Gaussian correspondence is used in section 5.6 for more general problems involving multiparameter estimation for pure qudit state models and establishing the achievability of the Holevo bound, cf. Theorem 17. The general

strategy is to translate the IID problem into a Gaussian one, solve the latter by using displaced number operators in a specific mode decomposition and then translate this into qudit measurement with respect to specific rotated bases.

This chapter is organised as follows. Section 5.2 reviews the QCRB and the conditions for its achievability. In section 5.3 we show that null measurements based at reasonable preliminary estimators fail to achieve the standard error scaling. In section 5.4.1 we introduce the idea of displaced-null measurement and prove its optimality in the paradigmatic case of a one-parameter qubit model. In section 5.6 we treat the general case of d dimensional systems and show how the Holevo bound is achieved on general models, and deal with the case where the multi-parameter QCRB is achievable.

5.2 ACHIEVABILITY OF THE QUANTUM CRAMÉR-RAO BOUND FOR PURE STATES

In this section we review the quantum Cramér-Rao bound (QCRB) and the conditions for its achievability in the case of models with *one-dimensional* parameters, which will be relevant for the first part of the chapter.

The estimation of multidimensional models and the corresponding Holevo bound is discussed in section 5.6.

Consider a quantum statistical model given by a family of d-dimensional density matrices ρ_θ which depend smoothly on an unknown parameter $\theta \in \mathbb{R}$. Let $\mathcal{M}$ be a measurement on $\mathbb{C}^d$ with positive operator-valued measure (POVM) elements $\{M_0, \dots, M_p\}$. By measuring ρ_θ we obtain an outcome $X \in \{0, \dots, p\}$ with probabilities

$$p_\theta(X = i) = p_\theta(i) = \mathrm{Tr}(M_i \rho_\theta), \qquad i = 0, \dots, p.$$

The classical Cramér-Rao bound states that the variance of any *unbiased* estimator $\hat{\theta} = \hat{\theta}(X)$ of θ is lower bounded as

$$\mathrm{Var}(\hat{\theta}) := \mathbb{E}_\theta[(\hat{\theta} - \theta)^2] \geq I_{\mathcal{M}}(\theta)^{-1} \tag{5.2}$$

where $I_{\mathcal{M}}(\theta)$ is the classical Fisher information (CFI)

$$I_{\mathcal{M}}(\theta) = \mathbb{E}_\theta \left[\left(\frac{d \log p_\theta}{d\theta} \right)^2 \right] = \sum_{i : p_\theta(i) > 0} p_\theta^{-1}(i) \left(\frac{d p_\theta(i)}{d\theta} \right)^2 . \tag{5.3}$$

The CFI associated to any measurement is upper bounded by the quantum Fisher information (QFI) [25, 26]

$$I_{\mathcal{M}}(\theta) \leq F(\theta) \tag{5.4}$$

where $F(\theta) = \mathrm{Tr}(\rho_\theta \mathcal{L}_\theta^2)$ and $\mathcal{L}_\theta$ is the symmetric logarithmic derivative (SLD) defined by the Lyapunov equation

$$\frac{d\rho_\theta}{d\theta} = \frac{1}{2}(\mathcal{L}_\theta \rho_\theta + \rho_\theta \mathcal{L}_\theta).$$

By putting together (5.2) and (5.4) we obtain the quantum Cramér-Rao bound (QCRB) [93, 95]

$$\mathrm{Var}(\hat{\theta}) := \mathbb{E}_\theta[(\hat{\theta} - \theta)^2] \geq F(\theta)^{-1}. \tag{5.5}$$

which sets a fundamental limit to the estimation precision. A similar bound on the covariance matrix of an unbiased estimator holds for multidimensional models, cf. section 5.6.

An important question is which measurements saturate the bound (5.4), and what is the statistical interpretation of the corresponding QCRB (5.5). For completeness, we state the exact conditions in the following Proposition whose formulation is adapted from [184].

Proposition 2. *Let ρ_θ be a one-dimensional quantum statistical model and let $\mathcal{M} := \{M_0, \dots, M_p\}$ be a measurement with probabilities $p_\theta(i) := Tr(\rho_\theta M_i)$. Then $\mathcal{M}$ achieves the bound (5.4) if and only if the following conditions hold:*

1) if $p_\theta(i) > 0$ there exists $\lambda_i \in \mathbb{R}$ such that

$$M_i^{1/2}\rho_\theta^{1/2} = \lambda_i M_i^{1/2}\mathcal{L}_\theta\rho_\theta^{1/2} \tag{5.6}$$

2) if $p_\theta(i) = 0$ for some i then $Tr(M_i\mathcal{L}_\theta\rho_\theta\mathcal{L}_\theta) = 0$.

Proof. Here we present a proof of the quantum Cramér-Rao bound, paying close attention to the necessary and sufficient conditions for a quantum Cramér-Rao bound saturating POVM. This is based on the proof by Zhou et al [184], and the aim is to highlight the features that allow this POVM to saturate the bound, producing Proposition 2. The classical Fisher information corresponding to the measurement $\mathcal{M} = \{M_i\}$ is given by

$$\begin{aligned} I_{\mathcal{M}}(\theta) &= \sum_{i:\mathrm{Tr}(M_i\rho_\theta)\neq 0} \frac{(\mathrm{Tr}(M_i\partial_\theta\rho_\theta))^2}{\mathrm{Tr}(M_i\rho_{\boldsymbol{\theta}})} \\ &= \sum_{i:\mathrm{Tr}(M_i\rho_\theta)\neq 0} \frac{(\mathrm{Re}[\mathrm{Tr}(M_i\mathcal{L}_\theta\rho_{\boldsymbol{\theta}})])^2}{\mathrm{Tr}(M_i\rho_{\boldsymbol{\theta}})}, \end{aligned}$$

where we have used the Lyapunov equation and identified that the resulting term corresponds to the above real component. Clearly

$$\begin{aligned} &\sum_{i:\mathrm{Tr}(M_i\rho_\theta)\neq 0} \frac{(\mathrm{Re}[\mathrm{Tr}(M_i\mathcal{L}_\theta\rho_{\boldsymbol{\theta}})])^2}{\mathrm{Tr}(M_i\rho_{\boldsymbol{\theta}})} \\ &\qquad \leq \sum_{i:\mathrm{Tr}(M_i\rho_\theta)\neq 0} \frac{|\mathrm{Tr}(M_i\mathcal{L}_\theta\rho_{\boldsymbol{\theta}})|^2}{\mathrm{Tr}(M_i\rho_{\boldsymbol{\theta}})}, \end{aligned}$$

where we have equality when $\mathrm{Im}[\mathrm{Tr}(M_i\mathcal{L}_\theta\rho_{\boldsymbol{\theta}})] = 0$. We now use the Cauchy-Schwarz inequality to cancel the denominator, identifying the terms $M_i^{\frac{1}{2}}\rho_{\boldsymbol{\theta}}^{\frac{1}{2}}$ and $M_i^{\frac{1}{2}}\mathcal{L}_\theta\rho_{\boldsymbol{\theta}}^{\frac{1}{2}}$ within the expression above, finding

$$\sum_{i:\mathrm{Tr}(M_i\rho_\theta)\neq 0} \frac{|\mathrm{Tr}(M_i\mathcal{L}_\theta\rho_{\boldsymbol{\theta}})|^2}{\mathrm{Tr}(M_i\rho_{\boldsymbol{\theta}})} \leq \sum_{i:\mathrm{Tr}(M_i\rho_\theta)\neq 0} \mathrm{Tr}(M_i\mathcal{L}_\theta\rho_{\boldsymbol{\theta}}\mathcal{L}_\theta). \tag{5.7}$$

The equality holds if

$$M_i^{\frac{1}{2}}\rho_{\boldsymbol{\theta}}^{\frac{1}{2}} = \lambda_i M_i^{\frac{1}{2}}\mathcal{L}_\theta\rho_{\boldsymbol{\theta}}^{\frac{1}{2}} \tag{5.8}$$

for all i such that $\mathrm{Tr}(M_i\rho_\theta) \neq 0$ for some $\lambda_i \in \mathbb{C}$. Finally, since M_i represents a POVM we have

$$\sum_{i:\mathrm{Tr}(M_i\rho_\theta)\neq 0} \mathrm{Tr}(M_i\mathcal{L}_\theta\rho_{\boldsymbol{\theta}}\mathcal{L}_\theta) \leq \mathrm{Tr}(\mathcal{L}_\theta^2\rho_{\boldsymbol{\theta}}) \equiv F(\theta). \tag{5.9}$$

where $F(\theta)$ is the quantum Fisher information. Equality is achieved when

$$\mathrm{Tr}(M_i\mathcal{L}_\theta\rho_{\boldsymbol{\theta}}\mathcal{L}_\theta) = 0$$

for all i such that $\mathrm{Tr}(M_i\rho_{\boldsymbol{\theta}}) = 0$. Note: this ensures that all the information on the parameter is contained in measurable outcomes with non-zero probabilities.

In summary, to achieve the QFI the POVM $\{M_i\}$ needs to satisfy the 3 following conditions:

1. If $\mathrm{Tr}(M_i\rho_{\boldsymbol{\theta}}) > 0$ then $\mathrm{Im}[\mathrm{Tr}(M_i\mathcal{L}_\theta\rho_{\boldsymbol{\theta}})] = 0$,
2. If $\mathrm{Tr}(M_i\rho_{\boldsymbol{\theta}}) > 0$ then $M_i^{\frac{1}{2}}\rho_{\boldsymbol{\theta}}^{\frac{1}{2}} = \lambda_i M_i^{\frac{1}{2}}\mathcal{L}_\theta\rho_{\boldsymbol{\theta}}^{\frac{1}{2}}$,$\lambda_i \in \mathbb{C}$,
3. If $\mathrm{Tr}(M_i\rho_{\boldsymbol{\theta}}) = 0$ then $\mathrm{Tr}(M_i\mathcal{L}_\theta\rho_{\boldsymbol{\theta}}\mathcal{L}_\theta) = 0$.

We can combine conditions 1. and 2. into the following condition:

4. If $\mathrm{Tr}(M_i\rho_{\boldsymbol{\theta}}) > 0$ then $M_i^{\frac{1}{2}}\rho_{\boldsymbol{\theta}}^{\frac{1}{2}} = \lambda_i M_i^{\frac{1}{2}}\mathcal{L}_\theta\rho_{\boldsymbol{\theta}}^{\frac{1}{2}}$, $\lambda_i \in \mathbb{R}$.

□

One can check that the conditions in Proposition 2 are satisfied, and hence the bound (5.4) is saturated, if $\mathcal{M}$ is the measurement of the observable $\mathcal{L}_\theta$. However, in general this observable depends on the unknown parameter, so achieving the QFI does not have an immediate statistical interpretation. Nevertheless, one can provide a meaningful operational interpretation in the scenario in which a large number n of copies of ρ_θ is available. In this case one can apply the adaptive scheme presented in the introduction: using a (small) sub-sample to obtain a 'rough' preliminary estimator $\tilde{\theta}$ of θ and then measuring $\mathcal{L}_{\tilde{\theta}}$ on the remaining copies. This adaptive procedure provides estimators $\hat{\theta}_n$ which achieve the Cramér-Rao bound asymptotically in the sense that (see, e.g., [70, 71])

$$n\mathbb{E}_\theta[(\hat{\theta}_n - \theta)^2] \to F^{-1}(\theta).$$

Pure state models. While for full rank states ($\rho_\theta > 0$) the second condition in Proposition 2 is irrelevant, this is not the case for rank deficient states, and in particular for pure state models.

Indeed let us assume that the model consists of pure states $\rho_\theta = |\psi_\theta\rangle\langle\psi_\theta|$ and let us choose the phase dependence of the vector state such that $\langle\dot{\psi}_\theta|\psi_\theta\rangle = 0$ (alternatively, one can use $|\psi_\theta^\perp\rangle := |\dot{\psi}_\theta\rangle - \langle\psi_\theta|\dot{\psi}_\theta\rangle|\psi_\theta\rangle$ instead of $|\dot{\psi}_\theta\rangle$ in the equations below). Then

$$\mathcal{L}_\theta = 2(|\dot{\psi}_\theta\rangle\langle\psi_\theta| + |\psi_\theta\rangle\langle\dot{\psi}_\theta|), \quad \text{and} \quad F(\theta) = 4\|\dot{\psi}_\theta\|^2.$$

Let $\mathcal{M}$ to be a projective measurement with $M_i = |v_i\rangle\langle v_i|$ where

$$\mathcal{B} := \{|v_0\rangle, \ldots, |v_{d-1}\rangle\}$$

is an orthonormal basis (ONB). Without loss of generality we can choose the phase factors such that $\langle v_i|\psi_\theta\rangle \in \mathbb{R}$ at the particular value of interest θ. Equation (5.6) in Proposition 2 becomes $\langle v_i|\dot{\psi}_\theta\rangle \in \mathbb{R}$, i.e., in the first order, the statistical model is in the real span of the basis vectors. Condition 2 requires that if $\langle v_i|\psi_\theta\rangle = 0$ then $\langle v_i|\dot{\psi}_\theta\rangle = 0$. Intuitively, this implies that, in the first order, the model is restricted to the real subspace spanned by the basis vectors with positive probabilities. For example if

$$|\psi_\theta\rangle := \cos\theta|0\rangle + \sin\theta|1\rangle \in \mathbb{C}^2, \tag{5.10}$$

then any measurement with respect to an ONB consisting of superpositions of $|0\rangle$ and $|1\rangle$ with *nonzero real* coefficients achieves the QCRB at $\theta = 0$, and no other measurement does so. This model will be discussed in detail in sections 5.3 and 5.4.

Null measurements. We now formally introduce the concept of a *null measurement* which will be the focus of our investigation. The general idea is to choose a measurement basis such that one of its vectors is equal or close to the unknown state. In this case, the corresponding outcome has probability close to one while the occurrence of other outcomes can serve as a 'signal' about the deviation from the true state. Let us consider first an *exact null measurement*, i.e., one in which the measurement basis $\mathcal{B} = \mathcal{B}(\theta)$ is chosen such that $|v_0\rangle = |\psi_\theta\rangle$, e.g., in the example in equation (5.10) the null measurement at $\theta = 0$ is determined by the standard basis. Such a measurement does not satisfy the conditions for achieving the QCRB. Indeed, we have $p_\theta(i) = \delta_{0,i}$ and condition 2 implies $\langle v_i|\dot{\psi}_\theta\rangle = 0$ for all $i = 1, \ldots, d-1$. Since $\mathcal{B}$ is an ONB, we then expect $\langle v_0|\dot{\psi}_\theta\rangle \neq 0$. This creates a contraction as we already assumed that $\langle\dot{\psi}_\theta|\psi_\theta\rangle = 0$. Therefore, the exact null measurement cannot achieve the QFI. In fact, the exact null measurement has zero CFI, which implies that there exists no (locally) unbiased estimator. Indeed, since probabilities belong to $[0, 1]$, and $p_\theta(i)$ is either 0 or 1 for a null measurement, all first derivatives at θ are zero so the CFI (5.3) is equal to zero, i.e., $I_{\mathcal{B}(\theta)}(\theta) = 0$.

One can rightly argue that the exact null measurement as defined above is not an operationally useful concept and cannot be implemented experimentally as it requires the exact knowledge of the unknown parameter. However, in a multi-copy setting the measurement *can* incorporate information about the parameter, as this can be obtained by measuring a sub-ensemble of systems in a preliminary estimation step, similarly to the SLD case. It is therefore meaningful to consider *approximate null* measurements, which satisfy the null property at $\tilde{\theta} \approx \theta$, i.e., we measure in a basis $\mathcal{B}(\tilde{\theta}) = \{|v_0^{\tilde{\theta}}\rangle, \ldots, |v_{d-1}^{\tilde{\theta}}\rangle\}$ with $|v_0^{\tilde{\theta}}\rangle = |\psi_{\tilde{\theta}}\rangle$. Interestingly, while the exact null measurement has zero CFI, an approximate null measurement $\mathcal{B}(\tilde{\theta})$ 'almost achieves' the QCRB in the sense that the corresponding classical Fisher information $I_{\mathcal{B}(\tilde{\theta})}(\theta)$ converges to $F(\theta)$ as $\tilde{\theta}$ approaches θ [115, 136, 137]. This means that by using an approximate null measurement we can achieve asymptotic error rates arbitrarily close (but not equal) to the QCRB, by measuring in a basis $\mathcal{B}(\tilde{\theta})$ with a fixed $\tilde{\theta}$ close to θ.

The question is then, is it possible to achieve the QCRB asymptotically with respect to the sample size by employing null measurements determined by an *estimated* parameter value, as in the case of the SLD measurement? References [115, 136, 137] do not address this question, aside from the above Fisher information convergence argument.

To answer the question we allow for measurements which have the null property at parameter values determined by *reasonable* preliminary estimators based on measuring a sub-sample of a large ensemble of identically prepared systems (cf. section 5.3 for precise definition). We investigate such measurement strategies and show that the natural two step implementation – use the rough estimator as a vector in the second step measurement basis – *fails* to achieve the standard rate $n^{-1/2}$ on simple qubit models. We will see that this is closely related to the fact that the CFI of the exact null measurement is zero, unlike the SLD case.

Nevertheless, in section 5.4.1 we show that a modified strategy which we call a *displaced-null measurement* does achieve asymptotic optimality in the simple qubit model discussed above. This scheme is then extended to general multidimensional qudit models in section 5.6 and shown to achieve the Holevo bound for general multiparameter models.

5.3 WHY THE NAIVE IMPLEMENTATION OF A NULL MEASUREMENT DOES NOT WORK

In this section we analyse the *null measurement* scheme described in section 5.2, for the case of a simple one-parameter qubit rotation model. The main result is Theorem 16 which shows that the naive/natural implementation of the null-fails to achieve the QCRB.

Let

$$|\psi_\theta\rangle = e^{-i\theta\sigma_y}|0\rangle = \cos(\theta)|0\rangle + \sin(\theta)|1\rangle \tag{5.11}$$

be a one-parameter family of pure states which describes a circle in the xz plane of the Bloch sphere. To simplify some of the arguments below we will assume that θ is known to be in the open interval $\Theta = (-\pi/8, \pi/8)$, but the analysis can be extended to completely unknown θ. The quantum Fisher information is

$$F(\theta) = 4\mathrm{Var}(\sigma_y) = 4\langle\psi_\theta|\sigma_y^2|\psi_\theta\rangle - 4\langle\psi_\theta|\sigma_y|\psi_\theta\rangle^2 = 4.$$

We now consider the specific value $\theta = 0$, so $|\psi_0\rangle = |0\rangle$ and $|\dot{\psi}_0\rangle = |1\rangle$. According to Proposition 2 any measurement with respect to a basis consisting of real combinations of $|0\rangle$ and $|1\rangle$ achieves the QCRB, with the exception of the basis $\{|0\rangle, |1\rangle\}$ itself. Indeed, let

$$|v_0^\tau\rangle = \exp(-i\tau\sigma_y)|0\rangle, \quad |v_1^\tau\rangle = \exp(-i\tau\sigma_y)|1\rangle \tag{5.12}$$

be such a basis ($\tau \neq 0$) , then the probability distribution is

$$p_\theta(0) = \cos^2(\theta - \tau), \qquad p_\theta(1) = \sin^2(\theta - \tau)$$

and the classical Fisher information is

$$I_\tau(\theta = 0) = \mathbb{E}_{\theta=0}\left[\left(\frac{d\log p_\theta}{d\theta}\right)^2\right] = 4.$$

However, at $\tau = 0$ we have $I_0(\theta = 0) = 0$ in agreement with the general fact that exact null measurements have zero CFI. This reveals a curious singularity in the space of optimal measurements, and our goal is to understand to what extent this is a mathematical artefact or it has a deeper statistical significance.

To start, we note that the failure of the standard basis measurement can also be understood as a consequence of parameter *non-identifiability* around the parameter value 0. Indeed, for $\tau = 0$ we have $p_\theta(i) = p_{-\theta}(i)$ so this measurement cannot distinguish θ from $-\theta$. A similar issue exists for $\tau \neq 0$, if θ is assumed to be completely unknown, or in an interval containing τ, cf. Figure 5.1. On the other hand, if θ is *known* to belong to an interval I and τ is outside this interval, then the parameter *is* identifiable and the standard asymptotic theory applies. For instance, measuring σ_x leads to an identifiable statistical model for our quantum qubit model.

Consider now the following two step procedure, which arguably is the most natural way of implementing approximate-null measurements. A sub-ensemble of $\tilde{n}$ systems is used to compute a preliminary estimator $\tilde{\theta}_n$, and subsequently the remaining samples are measured in the null-basis at angle $\tau = \tilde{\theta}_n$. For concreteness we assume that $\tilde{n} = n^{1-\epsilon}$ for some small constant $\epsilon > 0$, but our results hold more generally for $\tilde{n} = o(n)$ and $\tilde{n} \to \infty$ with n.

To formulate our theoretical result, we use the language of Bayesian statistics which we temporarily adopt for this purpose. We consider that the unknown parameter θ is random and is drawn from the uniform *prior distribution* $\pi(d\theta) = \frac{4}{\pi} d\theta$ over the parameter space Θ. Adopting a Bayesian notation we let $p(d\tilde{\theta}_n|\theta) := p_\theta(d\tilde{\theta}_n)$ be the distribution of $\tilde{\theta}_n$ given θ. The joint distribution of $(\theta, \tilde{\theta}_n)$ is then

$$p(d\theta, d\tilde{\theta}_n) = \pi(d\theta)p(d\tilde{\theta}_n|\theta) = p(d\tilde{\theta})\pi(d\theta|\tilde{\theta}_n)$$

where $\pi(d\theta|\tilde{\theta}_n)$ is the *posterior distribution* of θ given $\tilde{\theta}_n$.

Reasonable estimator hypothesis: we assume that $\tilde{\theta}_n$ is a *reasonable estimator* in the sense that the following conditions are satisfied for every $n \geq 1$:

1. $\pi(d\theta|\tilde{\theta}_n)$ has a density $\pi(\theta|\tilde{\theta}_n)$ with respect to the Lebesgue measure;
2. For each n there exist a set $A_n \subseteq \Theta$ such that $\mathbb{P}(\tilde{\theta}_n \in A_n) > c$ for some constant $c > 0$, and the following condition holds: for each $\tilde{\theta}_n \in A_n$, the positive symmetric function

$$g_{n,\tilde{\theta}_n}(r) := \min\{\pi(\tilde{\theta}_n + r|\tilde{\theta}_n), \pi(\tilde{\theta}_n - r|\tilde{\theta}_n)\}$$

satisfies

$$\int_{r \geq \tau_n} g_{n,\tilde{\theta}_n}(r)dr \geq C \tag{5.13}$$

where $\tau_n := n^{-1/2+\epsilon/4}$ and $C > 0$ is a constant independent on n and $\tilde{\theta}_n$.

Condition 2. means that the posterior distribution has significant mass on *both* sides of the preliminary estimator $\tilde{\theta}_n$, at a distance which is larger than $n^{-1/2+\epsilon/4}$, as illustrated in Figure 5.2. Since standard estimators such as maximum likelihood have asymptotically normal posterior distribution with standard deviation $\tilde{n}^{-1/2} = n^{-1/2+\epsilon/2} \gg n^{-(1-\epsilon/2)/2}$, condition 2. is expected to hold quite generally, hence the name reasonable estimator. The following lemma shows that the natural estimator in our model is indeed reasonable.

Lemma 5. *Consider the measurement of σ_x on a sub-ensemble of $\tilde{n} = n^{1-\epsilon}$ systems, and let $\tilde{\theta}_n$ be the maximum likelihood estimator. Then $\tilde{\theta}_n$ is a reasonable estimator.*

Proof. The outcomes $X_1, \dots, X_{\tilde{n}}$ have probabilities

$$p_\theta(0) = \cos^2(\theta - \pi/4), \quad p_\theta(1) = \sin^2(\theta - \pi/4).$$

The maximum likelihood estimator is $\tilde{\theta}_n = \pi/4 + f^{-1}(\bar{X}_n)$ where $\bar{X}_n = \frac{1}{\tilde{n}} \sum_{i=1}^{\tilde{n}} X_i$ and $f(x) = \sin^2(x)$ (which is invertible on $(-\pi/2, 0)$).

For every $k = 0, \dots, \tilde{n}$ and $\theta \in [-\pi/8, \pi/8]$, the density of the posterior at time n is given by

$$\pi(\theta | \tilde{\theta}_n = \pi/4 + f^{-1}(k/\tilde{n})) = \frac{\sin^{2k}(\theta - \pi/4) \cos^{2(\tilde{n}-k)}(\theta - \pi/4)}{\int_{-\frac{\pi}{8}}^{\frac{\pi}{8}} \sin^{2k}(\zeta - \pi/4) \cos^{2(\tilde{n}-k)}(\zeta - \pi/4) d\zeta} \tag{5.14}$$

and the unconditional distribution of $\tilde{\theta}_n$ is given by

$$\mathbb{P}(\tilde{\theta}_n = \pi/4 + f^{-1}(k/\tilde{n})) = \frac{4}{\pi} \binom{\tilde{n}}{k} \int_{-\frac{\pi}{8}}^{\frac{\pi}{8}} \sin^{2k}(\theta - \pi/4) \cos^{2(\tilde{n}-k)}(\theta - \pi/4) d\theta.$$

Consistency of $\tilde{\theta}_n$ and the dominated convergence theorem imply that for every Borel set

$$\lim_{n \to +\infty} \mathbb{P}(\tilde{\theta}_n \in A) = \frac{4}{\pi} \int_{-\frac{\pi}{8}}^{\frac{\pi}{8}} \chi_A(\theta) d\theta.$$

This allows us to consider for instance $A_n := (-\pi/8, \pi/8)$. Moreover, we can rewrite equation (5.14) as

$$\frac{e^{-\tilde{n} H(\sin^2(\tilde{\theta}_n - \pi/4), \sin^2(\theta - \pi/4))}}{\int_{-\frac{\pi}{8}}^{\frac{\pi}{8}} e^{-\tilde{n} H(\sin^2(\tilde{\theta}_n - \pi/4), \sin^2(\zeta - \pi/4))} d\zeta}$$

where $H(p, q) = -p \log(q) - (1 - p) \log(1 - q)$. Notice that if $\tilde{\theta}_n \in (-\pi/8, \pi/8)$, then $H(\sin^2(\tilde{\theta}_n - \pi/4), \sin^2(\theta - \pi/4))$ admits a unique minimum in $[-\pi/8, \pi/8]$ at $\theta = \tilde{\theta}_n$ where it vanishes and where the value of the second derivative is equal to 4. Therefore, for n big enough (uniformly in $\tilde{\theta}_n \in (-\pi/8, \pi/8)$) one has that

$$\begin{aligned} &\int_{\tilde{\theta}_n + \tau_n}^{\tilde{\theta}_n + \sqrt{\tilde{n}}} e^{-\tilde{n} H(\sin^2(\tilde{\theta}_n - \pi/4), \sin^2(\theta - \pi/4))} d\theta \\ &\geq \int_{\tilde{\theta}_n + \tau_n}^{\tilde{\theta}_n + \tilde{n}^{-1/2}} e^{-\frac{5\tilde{n}(\theta - \tilde{\theta}_n)^2}{2}} d\theta \\ &= \frac{1}{\sqrt{\tilde{n}}} \int_{n^{-\epsilon/4}}^{1} e^{-\frac{5\theta^2}{2}} d\theta \end{aligned}$$

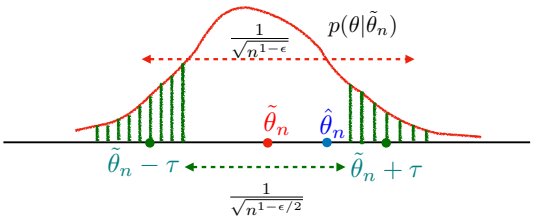

Figure 5.2: For a reasonable estimator $\tilde{\theta}_n$, the posterior distribution of θ is centred around $\tilde{\theta}_n$, and has width of order $n^{-(1-\epsilon)/2}$. The assumption amounts to the fact that the posterior has non-vanishing mass on either side of $\tilde{\theta}_n$ at distance larger than $n^{-(1-\epsilon/2)/2}$ which is much smaller that the typical standard deviation.

and analogously for the integral in the interval $[\tilde{\theta}_n - \sqrt{\tilde{n}}, \tilde{\theta}_n - \tau_n]$. Moreover

$$\begin{aligned}
&\int_{-\frac{\pi}{8}}^{\frac{\pi}{8}} e^{-\tilde{n}H(\sin^2(\tilde{\theta}_n-\pi/4),\sin^2(\zeta-\pi/4))}d\zeta \\
&= \int_{|\zeta-\tilde{\theta}_n|\geq n^{-1/2+\epsilon}} e^{-\tilde{n}H(\sin^2(\tilde{\theta}_n-\pi/4),\sin^2(\zeta-\pi/4))}d\zeta \\
&+ \int_{|\zeta-\tilde{\theta}_n|\leq n^{-1/2+\epsilon}} e^{-\tilde{n}H(\sin^2(\tilde{\theta}_n-\pi/4),\sin^2(\zeta-\pi/4))}d\zeta \\
&\leq e^{-3n^\epsilon/2} + \int_{\tilde{\theta}_n-n^{-1/2+\epsilon}}^{\tilde{\theta}_n+n^{-1/2+\epsilon}} e^{-\frac{3\tilde{n}(\zeta-\tilde{\theta}_n)^2}{2}}d\zeta \\
&= e^{-3n^\epsilon/2} + \frac{1}{\sqrt{\tilde{n}}}\int_{-n^{-\epsilon/2}}^{n^{-\epsilon/2}} e^{-\frac{3\zeta^2}{2}}d\zeta
\end{aligned}$$

and we are done, since we just proved that for n big enough (uniformly in $\tilde{\theta}_n \in (-\pi/8, \pi/8)$), one has

$$\begin{aligned}
&\mathbb{P}(\theta \geq \tilde{\theta}_n \pm \tau_n) \\
&\geq \frac{\int_{n^{-\epsilon/4}}^{1} e^{-\frac{5\theta^2}{2}}d\theta}{\sqrt{\tilde{n}}e^{-3n^\epsilon/2} + \int_{-n^{-\epsilon/2}}^{n^{-\epsilon/2}} e^{-\frac{3\zeta^2}{2}}d\zeta} \geq c > 0
\end{aligned}$$

for some c independent on n and $\tilde{\theta}_n$. This method can be extended to a wide class of estimators, since it essentially relies on assumptions which are quite standard in usual statistical problems. □

The next theorem is the main result of this section and shows that if a reasonable (preliminary) estimator is used as reference for a null measurement on the remaining samples, the MSE of the final estimator cannot achieve the QCRB, indeed it cannot even achieve standard scaling.

Theorem 16. *Assume that $\tilde{\theta}_n$ is a reasonable estimator as defined above, obtained by measuring a sub-ensemble of size $\tilde{n} := n^{1-\epsilon}$. Let $\hat{\theta}_n$ be an estimator of θ based on measuring the remaining $n - n^{1-\epsilon}$ sub-ensemble in the basis corresponding to angle $\tilde{\theta}_n$. Then*

$$\lim_{n\to\infty} nR_\pi(\hat{\theta}_n) = \infty$$

where

$$R_\pi(\hat{\theta}_n) = \int_\Theta \pi(d\theta)\mathbb{E}_\theta[(\hat{\theta}_n - \theta)^2]$$

is the average mean square error risk.

Proof. Taking into account the two step procedure we write

$$\begin{aligned}\mathbb{E}_\theta[(\hat{\theta}_n - \theta)^2] &= \int_{\mathbb{R}} p(d\hat{\theta}_n|\theta)(\hat{\theta}_n - \theta)^2 \\ &= \int_{\mathbb{R}^2} p(d\tilde{\theta}_n|\theta)p(d\hat{\theta}_n|\theta,\tilde{\theta}_n)(\hat{\theta}_n - \theta)^2\end{aligned}$$

where $p(d\tilde{\theta}_n|\theta)$ is the distribution of the preliminary estimator at θ and $p(d\hat{\theta}_n|\theta,\tilde{\theta}_n)$ is the distribution of the final estimator given θ and $\tilde{\theta}_n$. Since the final estimator is obtained by measuring at angle $\tilde{\theta}_n$, its distribution depends only on $r = |\tilde{\theta}_n - \theta|$, so $p(d\hat{\theta}_n|\tilde{\theta}_n,\theta) = p_r(d\hat{\theta}_n)$.

The Bayesian risk of the final estimator $\hat{\theta}_n$ is

$$\begin{aligned}R_\pi(\hat{\theta}_n) &= \mathbb{E}[(\hat{\theta}_n - \theta)^2] \\ &= \int_{\Theta\times\mathbb{R}^2} \pi(d\theta)p(d\tilde{\theta}_n|\theta)p(d\hat{\theta}_n|\theta,\tilde{\theta}_n)(\hat{\theta}_n - \theta)^2 \\ &= \int_{\mathbb{R}} p(d\tilde{\theta}_n)\int_{\Theta\times\mathbb{R}} \pi(d\theta|\tilde{\theta}_n)p(d\hat{\theta}_n|\theta,\tilde{\theta}_n)(\hat{\theta}_n - \theta)^2\end{aligned}$$

We have

$$\begin{aligned}&\int_\Theta\int_{\mathbb{R}} \pi(d\theta|\tilde{\theta}_n)p(d\hat{\theta}_n|\theta,\tilde{\theta}_n)(\hat{\theta}_n - \theta)^2 \\ &= \int_{r\geq 0} dr \int_{\mathbb{R}} p_r(d\hat{\theta}_n) \times \\ &[\pi(\tilde{\theta}_n + r|\tilde{\theta}_n)(\hat{\theta}_n - \tilde{\theta}_n - r)^2 + \\ &\pi(\tilde{\theta}_n - r|\tilde{\theta}_n)(\hat{\theta}_n - \tilde{\theta}_n + r)^2]\end{aligned}$$

By assumption, $\pi(\tilde{\theta}_n \pm r|\tilde{\theta}_n) \geq g_{n,\tilde{\theta}_n}(r)$ and since

$$(\hat{\theta}_n - \tilde{\theta}_n - r)^2 + (\hat{\theta}_n - \tilde{\theta}_n + r)^2 \geq 2r^2$$

we get that for every $\tilde{\theta}_n \in A_n$

$$\begin{aligned}&\int_{\Theta\times\mathbb{R}} \pi(d\theta|\tilde{\theta}_n)p(d\hat{\theta}_n|\theta,\tilde{\theta}_n)(\hat{\theta}_n - \theta)^2 \\ &\geq \int_{|r|\geq n^{-(1/2-\epsilon/4)}} g_{n,\tilde{\theta}_n}(r)2r^2dr \int_{\mathbb{R}} p_r(d\hat{\theta}_n) \\ &\geq 2Cn^{-1+\epsilon/2},\end{aligned}$$

where the last inequality follows from point 2. in the definition of the reasonable estimator. Finally, since $\mathbb{P}(\tilde{\theta}_n \in A_n) \geq c$, we get

$$R_\pi(\hat{\theta}_n) \geq 2cCn^{-1+\epsilon/2}$$

which implies the result. □

The fact that a reasonable estimator has a 'balanced' posterior was key in obtaining the negative result in Theorem 16. This encodes the fact that the null measurement cannot distinguish between possible parameter values $\theta = \tilde{\theta}_n + \tau_n$ and $\theta = \tilde{\theta}_n - \tau_n$ leading to errors that are larger than $n^{-1/2}$. In section 5.4 we show how we can go around this problem by deliberately choosing the reference parameter of the null measurement to be displaced away from a reasonable estimator $\tilde{\theta}_n$ by an amount δ_n that is large enough to insure identifiability, but small enough to still be in a shrinking neighbourhood of θ.

In the proof of Theorem 16 we made use of the fact that, for the statistical model defined in equation (5.11), the law of the measurement in the basis containing $|\psi_{\tilde{\theta}_n}\rangle$ could not distinguish between $\tilde{\theta}_n \pm r$. Although for general pure state models this might not be the case, in section 5.8 we show that under some mild additional assumptions, the result of Theorem 16 extends to weaker notions of non-identifiability.

5.3.1 *Parameter localisation via a two step adaptive procedure*

Here we discuss in more detail the general *parameter localisation* principle to which we refer repeatedly in this chapter. The principle is formulated for one-dimensional models, but can be extended staightforwardly to multidimensional ones.

Suppose we are given a large number n of independent, identically prepared systems in the state ρ_θ, depending smoothly on a parameter θ which lies in an open set $\Theta \subset \mathbb{R}$. To avoid pathological cases we assume that $F(\theta) > f > 0$ for all $\theta \in \Theta$. Even though the set Θ is a priori 'large', we can 'localise' the parameter and subsequently perform measurements adapted to the parameter value, by using the following two step procedure.

Consider a measurement $\mathcal{M}$ such that θ is *identifiable*, i.e., no two different parameters produce outcomes with identical probability distributions. In the first step we apply $\mathcal{M}$ to each system belonging to a vanishingly small proportion $\tilde{n} = o(n)$ of the samples, with $\tilde{n}$ growing with n. For concreteness we assume that $\tilde{n} = n^{1-\epsilon} \ll n$, with $\epsilon > 0$ a small number, but the arguments hold for a generic choice. Using the data obtained from measuring this sub-ensemble we construct a preliminary rough estimator $\tilde{\theta}_n$ of θ.

Naturally, one would like the estimator $\tilde{\theta}_n$ to be 'pretty good' (given the used sample size), but not necessarily optimal. There are several properties that could embody this requirement; for example one could require that the mean square error (MSE) scales at the standard rate $\tilde{n}^{-1} = n^{-1+\epsilon}$ and $\tilde{\theta}_n$ is asymptotically normal, i.e., it concentrates around θ at rate $\tilde{n}^{-1/2}$ with

$$\sqrt{\tilde{n}}(\tilde{\theta}_n - \theta) \longrightarrow N(0, V_\theta)$$

where the convergence to the normal distribution holds in law as $n \to \infty$ and the variance satisfies $V_\theta \geq F(\theta)^{-1}$. Standard estimators such as maximum likelihood typically satisfy this property [185]. In particular, this implies that the (confidence) interval

$$I_n = (\tilde{\theta}_n - n^{-1/2+\epsilon},\ \tilde{\theta}_n + n^{-1/2+\epsilon})$$

contains θ with probability converging to one exponentially fast. This follows from the fact that the ratio between the size of I_n and the standard deviation of $\tilde{\theta}_n - \theta$ diverges as $|I_n|/n^{(-1+\epsilon)/2} = n^{\epsilon/2}$.

On the other hand, if one adopts a Bayesian viewpoint and assumes the existence of a prior distribution on Θ with density $\pi(\theta)$, then it is natural to require the asymptotic normality of the posterior density

$$p(\theta|\tilde{\theta}_n) := \frac{\pi(\theta)p(\tilde{\theta}_n|\theta)}{p(\tilde{\theta}_n)}$$

where $p(\tilde{\theta}_n|\theta) = p_\theta(\tilde{\theta}_n)$ and $p(\tilde{\theta}_n) = \int p(\tilde{\theta}_n|\theta)\pi(d\theta)$. Intuitively this follows from the asymptotic normality of $\tilde{\theta}_n$. Indeed if the prior $\pi(\theta)$ and the variance V_θ are sufficiently regular with respect to θ and V_θ is bounded away from zero, then $p(\theta|\tilde{\theta}_n) \propto \exp(-\tilde{n}(\tilde{\theta}_n - \theta)^2/2V_\theta)$ concentrates around $\tilde{\theta}_n$ with approximately normal distribution. For more details on asymptotic normality theory we refer to [170, 178]. For our purposes, it will suffice to assume that $\tilde{\theta}_n$ is a 'reasonable' estimator in the sense that the posterior distribution is 'balanced' with respect to $\tilde{\theta}_n$ in a sense that is precisely defined in section 5.3. In particular, this means that we exclude 'dishonest' estimators for which the mass of the posterior distribution lies largely on one side of the estimator. For instance, taking a reasonable estimator $\tilde{\theta}_n$ and adding a constant δ_n such that $\delta_n/\tilde{n}^{-1/2} \to \infty$ for large n would be an example of a dishonest estimator. As we will see later this distinction becomes important since the preliminary estimator enters the definition of the second stage estimator, and the performance of null measurements based on reasonable/dishonest estimators is radically different.

Adaptive step. We pass now to the second step of the estimation procedure in which one measures the remaining $n' = n - \tilde{n}$ systems, taking into account the information provided by the first step. We distinguish two measurement strategies, the SLD measurement and the approximate null measurement.

For one-dimensional parameters, an *optimal* procedure is to measure the SLD $\mathcal{L}_{\tilde{\theta}_n}$ separately on each system and then construct the (final) estimator $\hat{\theta}_n = \tilde{\theta}_n + \bar{X}_n/F(\tilde{\theta}_n)$ where $\bar{X}_n = \frac{1}{n'}\sum_{i=1}^{n'} X_i$ is the average of the measurement outcomes. Assuming that the preliminary estimator is consistent (i.e.- $\tilde{\theta}_n \to \theta$ for large n), we obtain that $\hat{\theta}_n$ achieves the multicopy (asymptotic) version of the QCRB in the sense that

$$\lim_{n\to\infty} n\mathbb{E}_\theta[(\hat{\theta}_n - \theta)^2] = F(\theta)^{-1}.$$

Moreover, $\hat{\theta}_n$ is asymptotically normal

$$\sqrt{n}(\hat{\theta}_n - \theta) \longrightarrow N(0, F(\theta)^{-1})$$

thus providing us with simple asymptotic confidence intervals.

Let us consider now the case of null measurements. In section 5.2 we showed that if we measure in a basis $\mathcal{B}(\theta) = \{|v_1\rangle, \ldots, |v_d\rangle\}$ such that $|v_1\rangle = |\psi_\theta\rangle$ then the classical Fisher information is zero. However, at $\tilde{\theta} \approx \theta$ ($\tilde{\theta} \neq \theta$) the approximate null measurement with respect to a basis $\mathcal{B}(\tilde{\theta})$ has classical Fisher information $I_{\mathcal{B}(\tilde{\theta})} \approx F(\theta)$. As anticipated in section 5.2, the adaptive strategy used for the SLD measurement does not work in the case of null measurements when the initial estimator is reasonable. Proving this will be the subject of section 5.3.

Finally, let us briefly consider the case of multidimensional parameter models. In this setting, separate measurements may not be optimal in the second step due to non-commutativity of the SLD operators for different parameter components. However, using the information contained in $\tilde{\theta}_n$, we can devise collective measurements procedures which are asymptotically optimal in the sense of achieving the Holevo bound [95]. This can be understood by employing the local asymptotic normality (LAN) theory [88, 89, 102], which we briefly recall in section 5.6.2.

Note that since step one uses a vanishing proportion of the samples, the asymptotic result remains the same if we assume that n samples are available in the second step. Therefore, in order to simplify notation, in what follows we will replace $n' = n - \tilde{n}$ by n.

5.4 DISPLACED-NULL ESTIMATION SCHEME FOR OPTIMAL ESTIMATION OF PURE QUBIT STATES

In section 5.3 we showed that a null measurement that uses a reasonable preliminary estimator as reference parameter is sub-optimal. We will now show that one can achieve the asymptotic version of the QCRB (5.5) by employing a null measurement at a reference parameter that is *deliberately shifted* away from the reasonable estimator by a certain amount. We will call these *displaced-null* measurements.

5.4.1 *The displaced-null measurement for one parameter qubit models*

We consider the one parameter model $|\psi_\theta\rangle$ defined in equation (5.11) and assume that we are given n identical copies of $|\psi_\theta\rangle$. We apply the usual two step adaptive procedure: in the first step we use a vanishingly small proportion of the samples containing $\tilde{n} = n^{1-\epsilon}$ copies (where $\epsilon > 0$ is a small parameter) to perform a preliminary (non-optimal) estimation producing a reasonable estimator $\tilde{\theta}_n$. For concreteness we assume that $\tilde{\theta}_n$ is the estimator described in Lemma 5. Using Hoeffding's bound we find that $\tilde{\theta}_n$ satisfies the concentration bound

$$\mathbb{P}_\theta(|\tilde{\theta}_n - \theta| > n^{-1/2+\epsilon}) \leq Ce^{-n^\epsilon r} \tag{5.15}$$

for some constants $C, r > 0$. This means that with high probability, θ belongs to the confidence interval $I_n = (\tilde{\theta}_n - n^{-1/2+\epsilon},\ \tilde{\theta}_n + n^{-1/2+\epsilon})$ whose size shrinks at a slightly slower rate than $n^{-1/2}$.

In the second step we would like to measure all remaining qubits in a basis which contains a vector that is close to $|\psi_\theta\rangle$. However, as argued in section 5.3, the null measurement basis $\{|v_0^{\tilde{\theta}_n}\rangle, |v_1^{\tilde{\theta}_n}\rangle\}$ satisfying $|v_0^{\tilde{\theta}_n}\rangle = |\psi_{\tilde{\theta}_n}\rangle$ is suboptimal. More generally, for any angle $\tau \in I_n$, the basis defined by equation (5.12) suffers an identifiability problem as illustrated in panels a. and b. of Figure 5.1. For this reason, in the second step we choose the reference value

$$\theta'_n := \tilde{\theta}_n + \delta_n, \quad \delta_n := n^{-1/2+3\epsilon},$$

such that θ'_n is well outside I_n but nevertheless, $\theta'_n \to \theta$ for large n (assuming $\epsilon < 1/6$). The 3ϵ factor in the exponent is chosen such that the result of Proposition

3 below holds, but any factor larger than 2ϵ suffices. We measure all remaining samples in the basis $\{|v_0^{\theta'_n}\rangle, |v_1^{\theta'_n}\rangle\}$ (cf eq. (5.12)) to obtain outcomes $X_1, \dots, X_n \in \{0,1\}$ with probability distribution

$$P_\theta^{(n)} = (1 - p_\theta^{(n)}, p_\theta^{(n)}), \quad p_\theta^{(n)} = \sin^2(\theta - \theta'_n).$$

Proposition 3. *Assume that Θ is bounded and $\epsilon < 1/10$ is fixed, and let $\tilde{\theta}_n$ be the preliminary estimator based on $\tilde{n} = n^{1-\epsilon}$ samples.*

Let $\hat{\theta}_n$ be the estimator

$$\hat{\theta}_n := \tilde{\theta}_n + \frac{n^{-1/2+3\epsilon}}{2} - \frac{n^{1/2-3\epsilon}}{2}\hat{p}_n$$

where $\hat{p}_n$ is the empirical estimator of $p_\theta^{(n)}$, i.e.,

$$\hat{p}_n = \frac{|\{i : X_i = 1,\ i = 1, \dots, n\}|}{n}. \tag{5.16}$$

Then $\hat{\theta}_n$ is asymptotically optimal in the sense that

$$\lim_{n\to\infty} n\mathbb{E}_\theta[(\hat{\theta}_n - \theta)^2] = F^{-1}(\theta) = \frac{1}{4}.$$

Moreover, $\hat{\theta}_n$ is asymptotically normal, i.e.,

$$\sqrt{n}(\hat{\theta}_n - \theta) \to N\left(0, \frac{1}{4}\right)$$

where the convergence holds in distribution.

The proof of Proposition 3 can be found in section 5.9. Note that we chose to identify n and $n' = n - n^{1-\epsilon}$ in order to simplify the notation and the proofs, but it is immediate to adapt the reasoning in order to deal with this technicality. We also remark that the assumption $\epsilon < 1/10$ is not essential and could be removed at the price of using more involved analysis of the concentration properties of $\tilde{\theta}_n$ and the definition of the displacement parameter δ_n.

5.5 DISPLACED-NULL MEASUREMENTS IN THE ASYMPTOTIC GAUSSIAN PICTURE

In this section we cast the null-measurement problem into a companion Gaussian estimation problem which arises in the limit of large sample sizes. The Gaussian approximation is described by the theory of quantum local asymptotic normality (QLAN) developed in [70, 88, 89, 102]. For reader's convenience we review the special case of pure qubit states in section 5.5.1.

5.5.1 *Brief review of local asymptotic normality for pure qubit states*

The QLAN theory is closely related to the quantum central limit theorem (QCLT) and shows that for large n the statistical model describing ensembles of n identically prepared qubits can be approximated (locally in the parameter space) by a

single coherent state of a one-mode continuous variable (CV) system, whose mean encodes the unknown qubit rotation angle. We refer to [87, 89] for mathematical details and focus here on the intuitive correspondence between qubit ensembles and the CV mode.

We start with a completely unknown pure qubit state described by a one-dimensional projection $P = |\psi\rangle\langle\psi|$. In the first step we measure a sub-sample of $\tilde{n} = n^{1-\epsilon}$ systems and obtain a preliminary estimator $\tilde{P}_n = |\tilde{\psi}_n\rangle\langle\tilde{\psi}_n|$. We assume that $\tilde{P}_n$ satisfies a concentration bound similar to the one in equation (5.15) so that P lies within a ball of size $n^{-1/2+\epsilon}$ around $\tilde{P}_n$ with high probability. For more about the localisation procedure we refer to section 5.3.1.

We now choose the ONB $\{|0\rangle, |1\rangle\}$ such that $|0\rangle := |\tilde{\psi}_n\rangle$.

Thanks to parameter localisation we can focus our attention on 'small' rotations around $|0\rangle$ whose magnitude is of the order $n^{-1/2+\epsilon}$ where n is the sample size and $\epsilon > 0$ is small. We parametrise such states as

$$|\psi_{\boldsymbol{u}/\sqrt{n}}\rangle := U\left(\frac{\boldsymbol{u}}{\sqrt{n}}\right)|0\rangle = e^{-i(u_1\sigma_y - u_2\sigma_x)/\sqrt{n}}|0\rangle,$$

where $\boldsymbol{u} = (u_1, u_2)$ is a two-dimensional local parameter of magnitude $|\boldsymbol{u}| < n^{\epsilon}$. The joint state of the ensemble of n identically prepared qubits is then

$$|\Psi_{\boldsymbol{u}}^n\rangle = |\psi_{\boldsymbol{u}/\sqrt{n}}\rangle^{\otimes n}.$$

We now describe the *Gaussian shift model* which approximates the IID qubit model in the large sample size limit. A one mode CV system is specified by canonical coordinates Q, P satisfying $[Q, P] = i\mathbb{1}$. These act on a Hilbert space $\mathcal{H}$ with a orthonormal Fock basis $\{|k\rangle : k \geq 0\}$, such that $a|k\rangle = \sqrt{k}|k-1\rangle$, where a is the annihilation operator $a = (Q + iP)/\sqrt{2}$. The coherent states are defined as

$$|z\rangle := e^{-|z|^2/2}\sum_{k=0}^{\infty}\frac{z^k}{\sqrt{k!}}|k\rangle, \quad z \in \mathbb{C}$$

and satisfy $\langle z|a|z\rangle = z$. In the coherent state $|z\rangle$, the canonical coordinates Q, P have normal distributions $N\left(\sqrt{2}\mathrm{Re}(z), \frac{1}{2}\right)$ and $N\left(\sqrt{2}\mathrm{Im}(z), \frac{1}{2}\right)$, respectively. In addition, the number operator $N := a^*a$ has Poisson distribution with intensity $|z|^2$.

We now outline two approaches to QLAN embodying different ways to express the closeness of the multiqubits model $\{|\Psi_{\boldsymbol{u}}^n\rangle : |\boldsymbol{u}| \leq n^{\epsilon}\}$ to the quantum Gaussian shift model $\{|u_1 + iu_2\rangle : |\boldsymbol{u}| \leq n^{\epsilon}\}$. By applying the QCLT [135], one shows that the collective spin in the 'transverse' directions x and y have asymptotically normal distributions

$$\begin{aligned}\frac{1}{\sqrt{2n}}S_x(n) := \frac{1}{\sqrt{2n}}\sum_{i=1}^{n}\sigma_x^{(i)} &\rightarrow N\left(\sqrt{2}u_1, \frac{1}{2}\right)\\ \frac{1}{\sqrt{2n}}S_y(n) := \frac{1}{\sqrt{2n}}\sum_{i=1}^{n}\sigma_y^{(i)} &\rightarrow N\left(\sqrt{2}u_2, \frac{1}{2}\right)\end{aligned}$$

where the arrows represent convergence in distribution with respect to $|\Psi_u^n\rangle$. In fact the convergence holds for the whole 'joint distribution' which we write symbolically as

$$\left(\frac{1}{\sqrt{2n}}S_x(n), \frac{1}{\sqrt{2n}}S_y(n) \,:\, |\Psi_{\boldsymbol{u}}^n\rangle\right) \rightarrow (Q, P \,:\, |u_1 + iu_2\rangle)\,.$$

So, in what concerns the collective spin observables, the joint qubit state converges to a coherent state whose displacement is linear with respect to the local rotation parameters.

An alternative way to formulate the convergence to the Gaussian model is to show that the two models can be mapped into each other by means of physical operations (quantum channels) with asymptotically vanishing error, uniformly over all local parameters $|\boldsymbol{u}| \leq n^{\epsilon}$. Consider the isometric embedding of the symmetric subspace $\mathcal{S}_n = (\mathbb{C}^2)^{\otimes_s^n}$ of the tensor product $(\mathbb{C}^2)^{\otimes n}$ into the Fock space

$$\begin{aligned} V_n := \mathcal{S}_n &\rightarrow \mathcal{H} \\ |k,n\rangle &\mapsto |k\rangle \end{aligned}$$

where $|k,n\rangle$ is the normalised projection of the vector $|1\rangle^{\otimes k} \otimes |0\rangle^{\otimes n-k}$ onto $\mathcal{S}_n$. The following limits hold [89]

$$\begin{aligned} \lim_{n\to\infty} \sup_{|\boldsymbol{u}|\leq n^{1/2-\eta}} \| V_n |\Psi_{\boldsymbol{u}}^n\rangle - |u_1 + iu_2\rangle \| = 0, \\ \lim_{n\to\infty} \sup_{|\boldsymbol{u}|\leq n^{1/2-\eta}} \| |\Psi_{\boldsymbol{u}}^n\rangle - V_n^* |u_1 + iu_2\rangle \| = 0. \end{aligned}$$

where $\eta > 0$ is an arbitrary fixed parameter. In particular, for $\eta < 1/2 - \epsilon$ the supremum is taken over regions that contain all $|\boldsymbol{u}| < n^{\epsilon}$, which means that the Gaussian approximation holds uniformly over all values of the local parameter arising from the preliminary estimation step.

We now move to describe the relationship between qubit rotations and Gaussian displacements in the QLAN approximation. Let $U^n(\boldsymbol{\Delta}) := U(n^{-1/2}\boldsymbol{\Delta})^{\otimes n}$ be a qubit rotation by small angles $\boldsymbol{\delta} := n^{-1/2}\boldsymbol{\Delta}$ and let $D(\boldsymbol{\Delta}) = \exp(-i\sqrt{2}(\Delta_1 P - \Delta_2 Q))$ be the corresponding displacement operator. Then the following commutative diagram shows how QLAN translates (small) rotations into displacements (asymptotically with n and uniformly over local parameters)

$$\begin{array}{ccc} |\Psi_{\boldsymbol{u}}^n\rangle & \xrightarrow{V_n} & |u_1 + iu_2\rangle \\ \Big\downarrow {\scriptstyle U^n(-\boldsymbol{\Delta})} & & \Big\downarrow {\scriptstyle D(-\boldsymbol{\Delta})} \\ |\Psi_{\boldsymbol{u}-\boldsymbol{\Delta}}^n\rangle & \xrightarrow{V_n} & |u_1 - \Delta_1 + i(u_2 - \Delta_2)\rangle \end{array}$$

Notice that also the vertical arrow on the left of the diagram is true asymptotically with n and has to be intended as $\lim_{n\to+\infty} \| U^n(-\boldsymbol{\Delta})|\Psi_{\boldsymbol{u}}^n\rangle - |\Psi_{\boldsymbol{u}-\boldsymbol{\Delta}}^n\rangle \| = 0$.

Finally, we note that while the transverse spin components S_x, S_y converge to the canonical coordinates of the CV mode, the collective operator related to the total spin in direction z becomes the number operator N. Indeed if $E_n := (n\mathbb{1} - S_z)/2$ then $E_n|k,n\rangle = k|k,n\rangle$ so $E_n = V_n^* N V_n$. This correspondence can be extended to small rotations of such operators. Consider the collective operator

$$N_{\boldsymbol{\Delta}}^n := U^n(\boldsymbol{\Delta})(n\mathbb{1} - S_z)U^n(-\boldsymbol{\Delta})$$

which corresponds to measuring individual qubits in the basis

$$|v_0^{\delta}\rangle = U(\boldsymbol{\delta})|0\rangle, \quad |v_1^{\delta}\rangle = U(\boldsymbol{\delta})|1\rangle$$

and adding the resulting $\{0,1\}$ outcomes. In the limit Gaussian model, this corresponds to measuring the displaced number operator $N_{\boldsymbol{\Delta}} = D(\boldsymbol{\Delta})ND(-\boldsymbol{\Delta})$. More precisely, the binomial distribution $p^{(n)}_{\boldsymbol{u},\boldsymbol{\Delta}}$ of $N^n_{\boldsymbol{\Delta}}$ computed in the state $|\Psi^n_{\boldsymbol{\Delta}}\rangle$ converges to the Poisson distribution of $N_{\boldsymbol{\Delta}}$ with respect to the state $|u_1 + iu_2\rangle$

$$\lim_{n\to\infty} p^{(n)}_{\boldsymbol{u},\boldsymbol{\Delta}}(k) = e^{-\|\boldsymbol{u}-\boldsymbol{\Delta}\|^2} \frac{\|\boldsymbol{u}-\boldsymbol{\Delta}\|^{2k}}{k!}, \qquad k \geq 0.$$

5.5.2 *Asymptotic perspective on displaced-null measurements via local asymptotic normality*

We now offer a complementary picture of the displaced-null measurement schemes outlined in section 5.4.1, using the QLAN theory of section 5.5.1. In the Gaussian limit, the qubits ensemble is replaced by a single coherent state while the qubit null measurement becomes the number operator measurement. The Gaussian picture will illustrate why the null measurement does not work and how this problem can be overcome by using the displaced null strategy.

Consider first the one dimensional model given by equation (5.11), and let us assume for simplicity that the preliminary estimator takes the value $\tilde{\theta}_n = 0$. The general case can be reduced to this by a rotation of the block sphere.

We write θ in terms of the local parameter u as $\theta = \tilde{\theta}_n + u/\sqrt{n} = u/\sqrt{n}$ with $|u| \leq n^{\epsilon}$. By employing QLAN we map the IID model $|\Psi^n_u\rangle$ (approximately) into the limit coherent state model $|u\rangle$. At $\tilde{\theta}_n = 0$ the null measurement for an individual qubit is that of σ_z (standard basis). On the ensemble level this translates into measuring the collective spin observable S_z, which converges to the number operator N in the limit model, cf. section 5.5.1. Indeed, at $u = 0$ the coherent state is the vacuum which is an eigenstate of N.

As in the qubit case, the number measurement suffers from the non-identifiabilty issue since both $|\pm u\rangle$ states produce the same Poisson distribution (see panels c. and d. in Figure 5.1).

We now interpret the displaced-null measurement in the QLAN picture. Recall that if we measure each qubit in the rotated basis

$$|v^{\delta_n}_0\rangle = U((\delta_n, 0))|0\rangle, \quad |v^{\delta_n}_1\rangle = U((\delta_n, 0))|1\rangle,$$

then the non-identifiability is lifted and the parameter can be estimated optimally. The collective spin in this rotated basis is

$$N^n_{(\Delta_n,0)} := U^n((\Delta_n, 0))(\mathbb{1} - S_z)U^n((-\Delta_n, 0)).$$

where $\Delta_n = n^{1/2}\delta_n = n^{3\epsilon}$ and by the QLAN correspondence it maps to the displaced number operator

$$N_{(\Delta_n,0)} = D((\Delta_n, 0))ND((-\Delta_n, 0)).$$

In this case the distribution with respect to the state $|u\rangle$ is Poisson$(|\Delta_n - u|^2)$, and since $\Delta_n = n^{3\epsilon} \gg |u|$, the model is identifiable, i.e., the correspondence the intensity $|\Delta_n - u|^2$ and u is one-to-one (see panels g. and h. in Figure 5.1). Moreover, for large n the measurement provides an optimal estimator of u. Indeed by writing

$$N_{(\Delta_n,0)} = (a - \Delta_n \mathbf{1})^*(a - \Delta_n \mathbf{1}) = a^*a - \Delta_n(a + a^*) + \Delta_n^2 \mathbf{1} \tag{5.17}$$

and noting that the term a^*a is $O(n^{2\epsilon})$ (for $|u| \leq n^{\epsilon}$) we get

$$\frac{1}{2}\Delta_n - \frac{1}{2\Delta_n} N_{(\Delta_n,0)} = \frac{Q}{\sqrt{2}} + o(1) \tag{5.18}$$

where we recover the well known fact that quadrature (homodyne) measurement can be implemented by displacement and counting. By measuring the operator on the lefthand side of (5.18) we obtain an asymptotically optimal estimator of u, which corresponds to the qubit estimator constructed in section 5.4.1.

5.6 MULTIPARAMETER ESTIMATION FOR PURE QUDIT STATES

In this section we discuss the general case of a multidimensional statistical model for a d-dimensional quantum system (qudit).

The first two subsections review the theory of multiparameter estimation and how QLAN is used to establish the asymptotic achievability of the Holevo bound. This circle of ideas will be helpful in understanding the results in the following sections which deal with displaced-null estimation of qudit models. In particular we show that displaced-null measurements achieve the following:

1. The Holevo bound for completely unknown pure state models where the figure of merit is given by the Bures distance (Proposition 4);
2. The quantum Cramér-Rao bound in statistical models where the parameters can be estimated simultaneously (Proposition 5), providing an operational implementation for the results in [115, 136, 137];
3. The Holevo bound for completely general pure state models and figures of merit (Theorem 17).

Since the two stage strategy is discussed in detail in section 5.3.1, we do not give a detailed account of the preliminary stage and assume that the parameter has been localised in a neighbourhood of size $n^{-1/2+\epsilon}$ around a preliminary estimator with probability that converges to 1 exponentially fast in n.

5.6.1 *Multiparameter estimation*

Let us consider the problem of estimating the parameter $\boldsymbol{\theta}$ belonging to an open set $\Theta \subseteq \mathbb{R}^m$ given the corresponding family of states $\rho_{\boldsymbol{\theta}}$ of a d-dimensional quantum system. Given a measurement with POVM $\mathcal{M} := \{M_0, \ldots, M_p\}$, the CFI matrix is given by

$$I_{\mathcal{M}}(\boldsymbol{\theta})_{ij} = \mathbb{E}_{\theta}\left[\frac{\partial \log p_{\boldsymbol{\theta}}}{\partial \theta_i}\frac{\partial \log p_{\boldsymbol{\theta}}}{\partial \theta_j}\right].$$

The QFI matrix is $F(\theta)_{ij} = \frac{1}{2}\mathrm{Tr}(\rho_{\boldsymbol{\theta}}\{\mathcal{L}^i_{\boldsymbol{\theta}}, \mathcal{L}^j_{\boldsymbol{\theta}}\})$ where $\mathcal{L}^j_{\boldsymbol{\theta}}$ are the SLDs satisfying $\partial_i \rho_{\boldsymbol{\theta}} = \frac{1}{2}\{\mathcal{L}^j_{\boldsymbol{\theta}}, \rho_{\boldsymbol{\theta}}\}$ and $\{\cdot,\cdot\}$ denotes the anti-commutator again. If $\hat{\boldsymbol{\theta}}$ is an unbiased estimator then the multidimensional QCRB states that its covariance matrix is lower bounded as

$$\mathrm{Cov}_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}) := \mathbb{E}_{\boldsymbol{\theta}}[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T] \geq I_{\mathcal{M}}(\boldsymbol{\theta})^{-1} \geq F(\boldsymbol{\theta})^{-1}. \tag{5.19}$$

In general, the second lower bound is not achievable even asymptotically. Roughly, this is due to the fact that the optimal measurements for estimating the different components of $\boldsymbol{\theta}$ are incompatible with each other. The precise condition for the achievability of the QCRB is [49, 141]

$$\mathrm{Tr}(\rho_{\boldsymbol{\theta}}[\mathcal{L}^i_{\boldsymbol{\theta}}, \mathcal{L}^j_{\boldsymbol{\theta}}]) = 0, \quad i,j = 1,\dots,m. \tag{5.20}$$

which in the case of a pure statistical model $|\psi_{\boldsymbol{\theta}}\rangle$ becomes

$$\mathrm{Im}(\langle \partial_{\theta_i}\psi | \partial_{\theta_j}\psi\rangle) = 0, \quad i,j = 1,\dots,m. \tag{5.21}$$

When the QCRB is not achievable, one may look for measurements that optimise a specific figure of merit. The simplest example is that of a quadratic form with positive weight matrix W

$$R_W(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}) = \mathbb{E}_\theta[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T W (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})]$$

This choice is not as restrictive as it may seem since many interesting loss functions have a local quadratic approximation which determines the leading term of the asymptotic risk. A straightforward lower bound on R_W can be obtained by taking the trace with W in (5.19) but this bound is not achievable either. A better one is the Holevo bound [95]

$$\begin{aligned} & \mathrm{Tr}(W\mathrm{Cov}_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}})) \geq \mathcal{H}^W(\boldsymbol{\theta}) \\ := \quad & \min_{\mathbf{X}} \mathrm{Tr}(\mathrm{Re}(Z(\mathbf{X}))W) + \mathrm{Tr}\left|W^{1/2}\mathrm{Im}(Z(\mathbf{X}))W^{1/2}\right| \end{aligned} \tag{5.22}$$

where the minimum runs over m-tuples of selfadjoint operators $\boldsymbol{X} = (X_1,\dots,X_m)^T$ acting on the system, which satisfy the constraints $\mathrm{Tr}(\nabla_{\boldsymbol{\theta}}\rho_{\boldsymbol{\theta}}\boldsymbol{X}^T) = \mathbf{1}$, and $Z(\mathbf{X})$ is the $m \times m$ complex matrix with entries $Z(\mathbf{X})_{ij} = \mathrm{Tr}(\rho_{\boldsymbol{\theta}} X_i X_j)$. Unlike the multidimensional QCRB, the Holevo bound is achievable asymptotically in the IID scenario [49, 70]. In the next two section we will give an intuitive explanation based on the QLAN theory.

5.6.2 *Gaussian shift models and QLAN*

Quantum Gaussian shift models play a fundamental role in quantum estimation theory [95]. Such models are fairly tractable in that the Holevo bound is achievable with simple linear measurements. More importantly, Gaussian shift models arise as limits of IID models in the QLAN theory, which offers a recipe for constructing estimators which achieve the Holevo bound asymptotically in the IID setting.

For the purposes of this work, the asymptotic Gaussian limit offers a clean intuition about the working of the proposed estimators, but is not explicitly used in deriving the mathematical results. We therefore keep the presentation on a intuitive level and refer to the papers [32, 49, 70] for more details.

In this subsection we recall the essentials of multiparameter estimation in a pure quantum Gaussian shift model and of QLAN theory for pure states of finite dimensional quantum systems, extending what we already presented in the case of qubits in Section 5.5.

5.6.2.1 *Achieving the Holevo bound in a pure Gaussian shift model*

Consider a CV system consisting of $(d-1)$ modes. The corresponding Hilbert space $\mathcal{H}$ is the multimode Fock space which will be identified with the tensor product of $d-1$ copies of the single mode spaces, with ONB given by the Fock vectors $|\mathbf{k}\rangle := |k_1\rangle \otimes \cdots \otimes |k_{d-1}\rangle$, with $\mathbf{k} = (k_1, \ldots, k_{d-1}) \in \mathbb{N}^{d-1}$. The creation/annihilation operators, canonical coordinates and number operator of the individual modes are denoted a_i^*, a_i, $Q_i = (a_i + a_i^*)/\sqrt{2}$, $P_i = (a_i - a_i^*)/(\sqrt{2}i)$ and $N_i = a_i^* a_i$ for $i = 1, \ldots, d-1$.

We denote by $|\boldsymbol{z}\rangle = |z_1\rangle \otimes \cdots \otimes |z_{d-1}\rangle$ the multimode coherent states with $\boldsymbol{z} = (z_1, \ldots, z_{d-1}) \in \mathbb{C}^{d-1}$, so that Q_i and P_i have normal distribution with variance $1/2$ and mean $\sqrt{2}\mathrm{Re}(z_i)$ and $\sqrt{2}\mathrm{Im}(z_i)$, respectively, while N_i have Poisson distributions with intensities $|z_i|^2$. We denote by $\mathbf{R} := (Q_1, \ldots, Q_{d-1}, P_1, \ldots, P_{d-1})^T$ the vector of canonical coordinates which satisfy commutation relations $[R_i, R_j] = i\Omega_{ij}$ where Ω is the $2(d-1) \times 2(d-1)$ symplectic matrix

$$\Omega = \begin{pmatrix} \mathbf{0} & \mathbf{1} \\ -\mathbf{1} & \mathbf{0} \end{pmatrix}.$$

Let $\boldsymbol{u} \in \mathbb{R}^m$ be an unknown parameter and let $\mathcal{G}$ be the *quantum Gaussian shift model*

$$\mathcal{G} := \{|C\boldsymbol{u}\rangle : \boldsymbol{u} \in \mathbb{R}^m\}$$

where $C : \mathbb{R}^m \to \mathbb{C}^{d-1}$ is a linear map. The goal is to estimate $\boldsymbol{u}$ optimally for a given figure of merit.

Denoting the entries of C as $C_{k,j} = c_{kj}^q + ic_{kj}^p$ for $k = 1, \ldots, d-1$ and $j = 1, \ldots, m$, we call D the real $2(d-1) \times m$ matrix with elements $D_{k,j} = \sqrt{2}c_{kj}^q$, $D_{k+(d-1),j} = \sqrt{2}c_{kj}^p$ with $k = 1, \ldots d-1$; notice that $\mathbb{E}_{\boldsymbol{u}}[\mathbf{R}] = D\boldsymbol{u}$. We remark that $\boldsymbol{u}$ is identifiable if and only if D has rank equal to m. The quantum Fisher information matrix is independent of $\boldsymbol{u}$ and is given by $F = 2D^T D > 0$.

Let us first consider the case when the QCRB is achievable (in which case it leads to the Holevo bound by tracing with W). Condition (5.21) amounts to C^*C being a *real* matrix which is equivalent to $D^T \Omega D = 0$ and the fact that the generators of the Gaussian shift model $\mathcal{G}$

$$S_j = \sum_{k=1}^{d-1} c_{kj}^q P_k - c_{kj}^p Q_k = (D^T \Omega \mathbf{R})_j, \quad j = 1, \ldots m$$

commute with each other.

An optimal unbiased measurement consists of simultaneously measuring the commuting operators $\boldsymbol{Z} = \Sigma^{-1} D^T \mathbf{R}$, where $\Sigma := D^T D = F/2$. Indeed

- $[\mathbf{Z}, \mathbf{Z}^T] = \Sigma^{-1} D^T \Omega D \Sigma^{-1} = 0$ (commutativity),
- $\mathbb{E}_{\boldsymbol{u}}[\mathbf{Z}] = \Sigma^{-1} D^T D\boldsymbol{u} = \boldsymbol{u}$ (unbiasedeness),
- $\mathrm{Cov}_{\boldsymbol{u}}(\mathbf{Z}) = \Sigma^{-1}/2$ (achieves the QCRB).

Consider now the case when the QCRB is not achievable. For a given positive weight matrix W, the corresponding Holevo bound is given by

$$\begin{aligned} &\mathrm{Tr}(\mathrm{Cov}_{\boldsymbol{u}}(\hat{\boldsymbol{u}})W) \geq \mathcal{H}^W(\mathcal{G}) \\ &:= \min_B \frac{1}{2}\left(\mathrm{Tr}(WBB^T) + \mathrm{Tr}(|\sqrt{W}B\Omega B^T\sqrt{W}|)\right), \end{aligned} \tag{5.23}$$

where $\hat{\boldsymbol{u}}$ is an unbiased estimator and the minimum is taken over real $m \times 2(d-1)$ matrices B such that $BD = \mathbf{1}$. The Holevo bound can be saturated by coupling the system with another ancillary $(d-1)$-dimensional CV system in the vacuum state and with position and momentum vector that we denote by $\mathbf{R}' = (Q'_1, \ldots, Q'_{d-1}, P'_1, \ldots, P'_{d-1})^T$. In order to estimate $\boldsymbol{u}$, we consider a vector of quadratures of the form $\boldsymbol{Z} = B\mathbf{R} + B'\mathbf{R}'$ for B, B' real $m \times (d-1)$ matrices and we require that $\boldsymbol{Z}$ is unbiased and belongs to a commutative family:

- $B'\Omega B'^T = -B\Omega B^T$ (commutativity of the Z_i's),
- $\langle C\boldsymbol{u} \otimes \mathbf{0}|\mathbf{Z}|C\boldsymbol{u} \otimes \mathbf{0}\rangle = \boldsymbol{u} \Leftrightarrow BD = \mathbf{1}$ (unbiasedeness).

The corresponding risk is

$$R_{\mathbf{Z}} = \frac{1}{2}\Big(\mathrm{Tr}(WBB^T) + \mathrm{Tr}(WB'B'^T)\Big)$$

and by minimizing over B and B' one obtains the expression of the Holevo bound in Equation (5.23). Therefore, given a minimiser $(B^\star, B'^\star)$, the corresponding vector of quadratures $\boldsymbol{Z}^\star$ is an optimal estimator for any $\boldsymbol{u}$.

To summarise, in the pure Gaussian shift model there always exists a set of commuting quadratures $\mathbf{Z}^\star$ of a doubled up system that achieves the Holevo bound; in the case when the QCRB is achievable, one does not need an ancilla.

For the discussion in section 5.6.4 it is useful to consider the following implementation of the optimal measurement. Let $(\tilde{Q}_1, \ldots, \tilde{Q}_{2(d-1)}, \tilde{P}_1, \ldots, \tilde{P}_{2(d-1)})$ be a choice of vacuum modes of the doubled-up CV system such that $\mathbf{Z}^\star = T\tilde{\mathbf{Q}}$ where $\tilde{\mathbf{Q}} = (\tilde{Q}_1, \ldots, \tilde{Q}_m)^T$ for some $m \times m$ invertible matrix T with real entries. Up to classical post-processing, measuring $Z_1, \ldots, Z_m$ is equivalent to measuring $\tilde{Q}_1, \ldots, \tilde{Q}_m$. If we denote the outcomes of the latter by $\tilde{\mathbf{q}} := (\tilde{q}_1, \ldots \tilde{q}_m)$ then an optimal unbiased estimator of $\mathbf{u}$ is given by $\hat{\mathbf{u}} = T\tilde{\mathbf{q}}$.

5.6.2.2 *QLAN for IID pure qudit models*

The idea of QLAN is that the states in a shrinking neigbourhood of a fixed state can be approximated by a Gaussian shift model. In the next section we will show how this can be used as an estimation tool, but here we describe the general structure of QLAN for *pure* qudit states.

We choose the centre of the neighbourhood to be the first vector of an ONB $\{|0\rangle, \ldots, |d-1\rangle\}$, and parametrise the local neighborhood of states around $|0\rangle$ as

$$|\psi_{\boldsymbol{u}/\sqrt{n}}\rangle = \exp\left(-i\sum_{k=1}^{d-1}(u_1^k\sigma_y^k - u_2^k\sigma_x^k)/\sqrt{n}\right)|0\rangle \tag{5.24}$$

for $\boldsymbol{u} = (\boldsymbol{u}_1, \boldsymbol{u}_2) \in \mathbb{R}^{2(d-1)}$, $\|\boldsymbol{u}\| \leq n^\epsilon$, $\sigma_y^k = i\,|k\rangle\langle 0| - i\,|0\rangle\langle k|$ and $\sigma_x^k = |k\rangle\langle 0| + |0\rangle\langle k|$. As in the qubit case, the appropriately rescaled collective variables converge to position, momentum and number operators in 'joint distribution' with respect to $|\Psi_{\boldsymbol{u}}^n\rangle := |\psi_{\boldsymbol{u}/\sqrt{n}}\rangle^{\otimes n}$

$$\left(\frac{1}{\sqrt{2n}}S_x^k(n), \frac{1}{\sqrt{2n}}S_y^k(n), n\mathbb{1} - S_z^k(n) \;:\; |\Psi_{\boldsymbol{u}}^n\rangle\right) \\ \to \left(Q_k, P_k, N_k \;:\; |\mathbf{z} = \boldsymbol{u}_1 + i\boldsymbol{u}_2\rangle\right),$$

where $S_\alpha^k(n) = \sum_{l=1}^n (\sigma_\alpha^k)^{(l)}$ for $\alpha \in \{x, y, z\}$. More generally, we have a (real) linear map between the orthogonal complement of $|0\rangle$ and Gaussian quadratures: for every vector $|v\rangle = \sum_{k=1}^{d-1}(v_x^k + iv_y^k)\,|k\rangle$ we construct the corresponding Pauli operator $\sigma(v) = |v\rangle\langle 0| + |0\rangle\langle v|$ and the following central limit theorem (CLT) holds

$$\left(\frac{1}{\sqrt{2n}} S_v(n) : |\Psi_u^n\rangle\right) \to (X(v) : |\mathbf{z} = \boldsymbol{u}_1 + i\boldsymbol{u}_2\rangle) \tag{5.25}$$

where $S_v(n) := \sum_{l=1}^n \sigma(v)^{(l)}$ and $X(v) := \sum_{k=1}^n v_x^k Q_k + v_y^k P_k$.

In addition to the QCLT, the following strong QLAN statement holds: the statistical model $\{|\Psi_{\boldsymbol{u}}^n\rangle\}$ can be approximated by a pure Gaussian shift model in the sense that

$$\lim_{n\to\infty} \sup_{|\boldsymbol{u}|\le n^{1/2-\eta}} \| V_n |\Psi_{\boldsymbol{u}}^n\rangle - |\boldsymbol{u}_1 + i\boldsymbol{u}_2\rangle \| = 0, \tag{5.26}$$

$$\lim_{n\to\infty} \sup_{|\boldsymbol{u}|\le n^{1/2-\eta}} \| |\Psi_{\boldsymbol{u}}^n\rangle - V_n^* |\boldsymbol{u}_1 + i\boldsymbol{u}_2\rangle \| = 0 \tag{5.27}$$

for any fixed $0 < \eta < 1/2$. V_n is the isometric embedding of the symmetric subspace $\mathcal{S}_d^{(n)} := \left(\mathbb{C}^d\right)^{\otimes_s n}$ into a $(d-1)$-mode Fock space $\mathcal{H}$ (cf. previous section) characterised by

$$\begin{aligned} V_n : \mathcal{S}_d^{(n)} &\to \mathcal{H} \\ |\mathbf{k}; n\rangle &\mapsto |\mathbf{k}\rangle \end{aligned} \tag{5.28}$$

where $|\mathbf{k}; n\rangle$ denotes the normalised vector obtained by symmetrising

$$|1\rangle^{\otimes k_1} \otimes \cdots \otimes |d-1\rangle^{\otimes k_{d-1}} \otimes |0\rangle^{\otimes (n-(k_1+\cdots+k_{d-1}))}.$$

As in the qubits case, the Gaussian approximation maps small rotations into displacements of the coherent states. Consider collective qubit rotations by small angles $\delta := n^{-1/2}\boldsymbol{\Delta}$

$$U^n(\boldsymbol{\Delta}) := \left(\exp\left(-i\sum_{k=1}^{d-1}(n^{-1/2}\Delta_1^k\sigma_y^k - n^{-1/2}\Delta_2^k\sigma_x^k)\right)\right)^{\otimes n}$$

and the corresponding displacement operators

$$D(\boldsymbol{\Delta}) = \exp\left(-i\sum_{k=1}^{d-1}(\Delta_1^k P_k - \Delta_2^k Q_k)\right).$$

The diagram below conveys the asymptotic correspondence between rotations and displacements, where the arrows should be interpreted in the same way as the strong convergence equations (5.26) and (5.27)

$$\begin{array}{ccc} |\Psi_{\boldsymbol{u}}^n\rangle & \xrightarrow{V_n} & |\boldsymbol{u}_1 + i\boldsymbol{u}_2\rangle \\ \downarrow U^n(-\boldsymbol{\Delta}) & & \downarrow D(-\boldsymbol{\Delta}) \\ |\Psi_{\boldsymbol{u}-\boldsymbol{\Delta}}^n\rangle & \xrightarrow{V_n} & |\boldsymbol{u}_1 - \boldsymbol{\Delta}_1 + i(\boldsymbol{u}_2 - \boldsymbol{\Delta}_2)\rangle \end{array}$$

A similar correspondence holds for measurements with respect to rotated bases and displaced number operators

$$\begin{array}{ccc} |\Psi_{\boldsymbol{u}}^n\rangle & \xrightarrow{V_n} & |\boldsymbol{u}_1 + i\boldsymbol{u}_2\rangle \\ \downarrow N_{\boldsymbol{\Delta}}^i(n) & & \downarrow N_{\boldsymbol{\Delta}}^i \\ p^n(\boldsymbol{u}, \boldsymbol{\Delta}) & \xrightarrow{V_n} & \text{Poisson}(\|\boldsymbol{u}_1 - \boldsymbol{\Delta}_1 + i(\boldsymbol{u}_2 - \boldsymbol{\Delta}_2)\|^2) \end{array}$$

More precisely, suppose we measure the commuting family of operators $\{N_{\boldsymbol{\Delta}}^i(n), i = 1, \ldots, d-1\}$ given by

$$N_{\boldsymbol{\Delta}}^i(n) := U^n(-\boldsymbol{\Delta})(n\mathbb{1} - S_z^i(n))U^n(\boldsymbol{\Delta}) \quad i = 1, \ldots, d-1,$$

which amounts to measuring individual qudits in the basis

$$|v_i^\delta\rangle = U(\delta)|i\rangle \quad i = 0, \ldots, d-1$$

and collecting the total counts for individual outcomes in $\{0, \ldots, d-1\}$. In the Gaussian model this corresponds to measuring the displaced number operators $N_{\boldsymbol{\Delta}}^i = D(-\boldsymbol{\Delta})N^i D(\boldsymbol{\Delta})$, and by QLAN, the multinomial distribution $p^n(u, \boldsymbol{\Delta})$ of $N_{\boldsymbol{\Delta}}^i(n)$ converges to the law of the vector of Poisson random variables obtained by measuring $N_{\boldsymbol{\Delta}}^i$ with respect to the state $|\boldsymbol{u}_1 + i\boldsymbol{u}_2\rangle$.

5.6.3 *Achieving the Holevo bound for pure qudit states via QLAN*

We will now treat a general pure states statistical model and show how one can use QLAN to achieve the Holevo bound (5.22) asymptotically with the sample size. Let $|\psi_{\boldsymbol{\theta}}\rangle$ be a statistical model where $\boldsymbol{\theta} = (\theta^j)_{j=1}^m$ belongs to some open set $\Theta \subset \mathbb{R}^m$ with $m \leq 2(d-1)$ and the parameter is assumed to be identifiable. Given an ensemble of n copies of the unknown state, we would like to devise a measurement strategy and estimation procedure which attains the smallest average error (risk), asymptotically with n. For mixed states, a general solution has been discussed in [49] where it is shown how the Holevo bound can be achieved asymptotically using the QLAN machinery. Here we adapt this method to the case of pure state models.

In brief, the procedure involves three steps. We first use $\tilde{n} = n^{1-\epsilon}$ samples to produce a preliminary estimator $\tilde{\boldsymbol{\theta}}_n$ and write $\boldsymbol{\theta} = \tilde{\boldsymbol{\theta}}_n + \boldsymbol{u}/\sqrt{n}$ where $\boldsymbol{u}$ is the local parameter satisfying $\|\boldsymbol{u}\| \leq n^\epsilon$ (with high probability). We chooose an ONB $\{|0\rangle, \ldots |d-1\rangle\}$ such that $|\psi_{\tilde{\boldsymbol{\theta}}_n}\rangle = |0\rangle$ and use the QLAN isometry V_n (cf. equation 5.28) to map the remaining qubits $|\Psi_{\boldsymbol{u}}^n\rangle := |\psi_{\tilde{\boldsymbol{\theta}}_n + \boldsymbol{u}/\sqrt{n}}\rangle^{\otimes n}$ approximately into the Gaussian state $|C\boldsymbol{u}\rangle$. We then use the method described in section 5.6.2.1 to estimate the unknown parameter and achieve the Holevo bound.

We start by expressing the local states as small rotations around $|0\rangle$

$$\begin{aligned} &|\psi_{\tilde{\boldsymbol{\theta}}_n + \boldsymbol{u}/\sqrt{n}}\rangle \\ &= \exp\left(-i\sum_{k=1}^{d-1}\left(f_k^q\left(\frac{\boldsymbol{u}}{\sqrt{n}}\right)\sigma_y^k - f_k^p\left(\frac{\boldsymbol{u}}{\sqrt{n}}\right)\sigma_x^k\right)\right)|0\rangle \end{aligned} \tag{5.29}$$

where f_k^q and f_k^p are real functions and σ_y^k and σ_x^k are the Pauli matrices of equation (5.24). We now 'linearise' the generators of the rotations and define

$$|\tilde{\psi}_{\boldsymbol{u}/\sqrt{n}}\rangle := \exp\left(-i\sum_{j=1}^{m} u_j S_j/\sqrt{n}\right)|0\rangle \tag{5.30}$$

where

$$S_j = \sum_{k=1}^{d-1}(c_{kj}^q \sigma_y^k - c_{kj}^p \sigma_x^k), \quad c_{kj}^{q,p} = \partial_j f_k^{q,p}(\boldsymbol{u})\big|_{\boldsymbol{u}=\boldsymbol{0}}.$$

We denote the ensemble state of the linearised model $|\widetilde{\Psi}_{\boldsymbol{u}}^n\rangle := |\tilde{\psi}_{\boldsymbol{u}/\sqrt{n}}\rangle^{\otimes n}$. The following lemma shows that the original and the 'linearised' models are locally undistinguishable in the asymptotic limit.

Lemma 6. *With the above notations if $\epsilon < 1/6$ one has*

$$\lim_{n\to\infty} \sup_{\|\boldsymbol{u}\|\leq n^\epsilon} \||\Psi_{\boldsymbol{u}}^n\rangle\langle\Psi_{\boldsymbol{u}}^n| - |\widetilde{\Psi}_{\boldsymbol{u}}^n\rangle\langle\widetilde{\Psi}_{\boldsymbol{u}}^n|\|_1 = 0$$

where $\|\cdot\|_1$ denotes the trace distance.

Proof. In the present section we want to show that the statistical models $|\Psi_{\boldsymbol{u}}^n\rangle$ and $|\tilde{\Psi}_{\boldsymbol{u}}^n\rangle$ (which are the ensemble states corresponding to the models in Equations (5.29) and (5.30)) become equivalent in Le Cam distance when the neighbourhood of parameters considered shrinks around $\tilde{\boldsymbol{\theta}}_n$. Let us first compute the overlaps between states corresponding to the same local parameter $\boldsymbol{u}$ in the single copy scenario: expanding the unitary rotations one obtains

$$\begin{aligned}
&\langle\tilde{\psi}_{\boldsymbol{u}/\sqrt{n}}|\psi_{\tilde{\boldsymbol{\theta}}_n+\boldsymbol{u}/\sqrt{n}}\rangle = \left\langle 0\middle|\left(\mathbf{1}+i\frac{S(\boldsymbol{u})}{\sqrt{n}}-\frac{S(\boldsymbol{u})^2}{2n}+o\left(\frac{1}{n}\right)\right)\cdot\right.\\
&\left.\cdot\left(\mathbf{1}+i\frac{S(\boldsymbol{u})}{\sqrt{n}}-i\frac{T(\boldsymbol{u})}{\sqrt{n}}-\frac{S(\boldsymbol{u})^2}{2n}+o\left(\frac{1}{n}\right)\right)|0\right\rangle\\
&= 1 - i\langle 0|T(\boldsymbol{u})0\rangle/\sqrt{n} + o(1/n) = 1 + o(1/n),
\end{aligned}$$

where

$$\begin{aligned}
S(\boldsymbol{u}) &= \sum_{j=1}^{m} u_j S_j,\\
T(\boldsymbol{u}) &= \sum_{i,j=1} u_i u_j \sum_{k=1}^{d-1} \partial_{ij} f_k^q(0)\sigma_y^k - \partial_{ij} f_k^p(0)\sigma_x^k.
\end{aligned}$$

Notice that the last equality in the computation of the overlap is true because $T(\boldsymbol{u})$ has zero expectation in $|0\rangle$. We remark that the error remains of the order of $o(1/n)$ and is uniform in $\boldsymbol{u}$ if $\|\boldsymbol{u}\| \leq n^\epsilon$ with $\epsilon < 1/6$. Now we can conclude easily using the expressions of the trace norm between two pure states in terms of the overlap of two representative vectors, and noticing that

$$\langle\tilde{\Psi}_{\boldsymbol{u}}^n|\Psi_{\boldsymbol{u}}^n\rangle = \langle\tilde{\psi}_{\boldsymbol{u}/\sqrt{n}}|\psi_{\tilde{\boldsymbol{\theta}}_n+\boldsymbol{u}/\sqrt{n}}\rangle^n = (1+o(1/n))^n \to 1$$

uniformly in $\|\boldsymbol{u}\| \leq n^\epsilon$.

□

Thanks to such uniform approximation results, one can replace the original model with the linearised one without affecting the asymptotic estimation analysis. We denote the latter by

$$\mathcal{Q}_n := \{|\widetilde{\Psi}^n_{\boldsymbol{u}}\rangle : \boldsymbol{u} \in \mathbb{R}^m, \|\boldsymbol{u}\| \leq n^\epsilon\}.$$

Let us now consider the second ingredient of the estimation problem, the risk (figure of merit). We fix a loss function $L : \Theta \times \Theta \to \mathbb{R}_+$, so that the risk of an estimator $\hat{\boldsymbol{\theta}}_n$ at $\boldsymbol{\theta}$ is $R(\hat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{\theta}}[L(\hat{\boldsymbol{\theta}}_n, \boldsymbol{\theta})]$. We assume that the loss function is locally quadratic around any point and in particular

$$L(\tilde{\boldsymbol{\theta}}_n + \boldsymbol{u}, \tilde{\boldsymbol{\theta}}_n + \boldsymbol{v}) \approx \sum_{i,j=1}^{m} w_{ij}(\tilde{\boldsymbol{\theta}}_n)(u_i - v_i)(u_j - v_j)$$

for a strictly positive weight matrix function $\boldsymbol{\theta}' \mapsto W(\boldsymbol{\theta}') = (w_{ij}(\boldsymbol{\theta}'))$ (which we assume to be continuous in $\boldsymbol{\theta}'$). In asymptotics, $\tilde{\boldsymbol{\theta}}_n \to \boldsymbol{\theta}$ and the loss function can be replaced by its quadratic approximation at the true parameter $\boldsymbol{\theta}$ without affecting the leading contribution to the estimation risk. We denote $W := W(\boldsymbol{\theta})$.

Returning to the original estimation problem, we now show how QLAN can be used to construct an estimator which achieves the Holevo bound asymptotically.

We couple each system with a d-dimensional ancillary system in state $|0'\rangle$ and fix an ONB for the ancilla $\mathcal{B}' = \{|0'\rangle, \ldots, |d-1'\rangle\}$. The extended IID statistical model is $|\Psi^n_{\boldsymbol{u}}\rangle \otimes |0'\rangle^{\otimes n}$. By quantum LAN, the joint ensemble can be approximated by a pure Gaussian shift model coupled with an ancillary $(d-1)$-modes CV system prepared in the vacuum: $|C\boldsymbol{u}\rangle \otimes |\boldsymbol{0}\rangle$ where C is the $(d-1) \times m$ complex matrix with entries $C_{kj} = c^q_{kj} + ic^p_{kj}$; more precisely we map the two qudit ensembles into their Fock spaces by means of a tensor of isometries as in equation (5.28) and we consider the $2(d-1)$ modes which correspond to the linear space $\mathcal{L} := \mathrm{Lin}\{|0\rangle \otimes |i'\rangle, |i\rangle \otimes |0'\rangle : i = 1, \ldots d-1\}$ (which contains $\{|\psi_{\boldsymbol{\theta}}\rangle \otimes |0'\rangle\}_{\boldsymbol{\theta}\in\Theta}$). Alternatively, one can map the original ensemble to the CV space and *then* add a second CV system in the vacuum state. The reason we chose to add an ancillary ensemble at the beginning is because this same setup will be used in the next section in the context of displaced-null measurements.

We now apply the optimal measurement for the Gaussian shift model $|C\boldsymbol{u}\rangle$ with weight matrix W, as described in section 5.6.2.1. This involves measuring commuting quadratures of the doubled up CV system, such that the resulting estimator $\hat{\boldsymbol{u}}_n$ achieves the Gaussian Holevo bound (5.23) in the limit of large n.

Thanks to the parameter localisation and LAN, the asymptotic (rescaled) risk of the corresponding 'global' estimator $\hat{\boldsymbol{\theta}}_n = \tilde{\boldsymbol{\theta}}_n + \hat{\boldsymbol{u}}_n/\sqrt{n}$ satisfies

$$\lim_{n\to\infty} nR(\hat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}) = \mathcal{H}^W(\mathcal{G}).$$

Finally we note that the expressions of the Holevo bound (5.22) in IID model $|\psi_{\boldsymbol{\theta}}\rangle$ with loss function L, and the corresponding Gaussian shift model $|C\boldsymbol{u}\rangle$ with weight matrix W coincide: $\mathcal{H}^W(\boldsymbol{\theta}) = \mathcal{H}^W(\mathcal{G})$. Indeed, since $\rho_{\boldsymbol{\theta}} = |\psi_{\boldsymbol{\theta}}\rangle\langle\psi_{\boldsymbol{\theta}}|$ is a pure state, the minimisation in (5.22) can be restricted to operators $\boldsymbol{X} = (X_1, \ldots, X_m)$ such that $PX_iP = P^\perp X_i P^\perp = 0$ where $P = \rho_{\boldsymbol{\theta}}, P^\perp = \mathbf{1} - P$. In this case the two Holevo bounds coincide after making the identification $B_{j,k} = \sqrt{2}\mathrm{Re}\langle k|X_j|0\rangle, B_{j+d-1,k} = \sqrt{2}\mathrm{Im}\langle k|X_j|0\rangle$.

5.6.4 *Achieving the Holevo bound with displaced-null measurements*

In this section we show how displaced-null measurements offer an alternative strategy to the one presented in the previous section, for optimal estimation in a general finite dimensional pure statistical model $|\psi_{\boldsymbol{\theta}}\rangle$ with $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^m$. As before, we assume that the risk function $L : \Theta \times \Theta \to \mathbb{R}_+$ has a continuous quadratic local approximation given by the matrix valued function $W(\boldsymbol{\theta})$.

The first steps are the same as in the estimation procedure in section 5.6.3: we use $\tilde{n} = n^{1-\epsilon}$ samples to produce a preliminary estimator $\tilde{\boldsymbol{\theta}}_n$ and we write $\boldsymbol{\theta} = \tilde{\boldsymbol{\theta}}_n + \boldsymbol{u}/\sqrt{n}$ where $\boldsymbol{u}$ is the local parameter such that $\|\boldsymbol{u}\| \leq n^{\epsilon}$ with high probability. We choose an ONB $\mathcal{B} = \{|0\rangle, \ldots, |d-1\rangle\}$ such that $|0\rangle := |\psi_{\tilde{\boldsymbol{\theta}}_n}\rangle$ and apply Lemma 6 to approximate the local model as in equation (5.30). We couple each system with an ancillary qudit in state $|0'\rangle$. By QLAN, the joint model is approximated by the Gaussian shift model consisting of coherent states $|C\boldsymbol{u}\rangle \otimes |0\rangle$ of a $2(d-1)$-modes CV system.

As detailed in section 5.6.2.1, the Holevo bound for the Gaussian shift can be attained by measuring a certain set of canonical coordinates $\widetilde{Q} := (\widetilde{Q}_1, \ldots, \widetilde{Q}_m)$ of the doubled-up systems. In turn, this provides an asymptotically optimal measurement for the IID qudit model as explained in section 5.6.3. Instead of measuring these quadratures, here we adopt the displaced-null measurements philosophy used in section 5.4, which achieves the same asymptotic risk. This means that one measures the commuting set of displaced number operators $\tilde{N}^j_{\boldsymbol{\Delta}_n} = D(-\boldsymbol{\Delta}_n)\widetilde{N}^j D(\boldsymbol{\Delta}_n)$ where $\widetilde{N}^j = \widetilde{a}^*_j \widetilde{a}_j$ is the number operator corresponding to the mode $(\widetilde{Q}_j, \widetilde{P}_j)$ and

$$D(\boldsymbol{\Delta}_n) = \exp\left(-i\Delta_n \sum_{k=1}^{m} \tilde{P}_k\right), \quad \Delta_n = \sqrt{n}\delta_n = n^{3\epsilon}.$$

We note that

$$\tilde{N}^j_{\boldsymbol{\Delta}_n} = (\widetilde{a}_j - n^{3\epsilon}\mathbf{1})^*(\widetilde{a}_j - n^{3\epsilon}\mathbf{1}) = n^{6\epsilon}\mathbf{1} - \sqrt{2}\widetilde{Q}_j n^{3\epsilon} + \widetilde{N}^j,$$

so for large n, measuring $\tilde{N}^j_{\boldsymbol{\Delta}_n}$ is equivalent to measuring $\widetilde{Q}_j$. We recall that by measuring $\mathbf{Z}^\star := T\widetilde{\mathbf{Q}}$ we obtain an optimal unbiased estimator of $\mathbf{u}$, where T is the invertible matrix defined at the end of section 5.6.2.1. Therefore, using the above equation we can construct an (asymptotically) optimal estimator given by the outcomes of the following set of commuting operators

$$\sum_{k=1}^{m} T_{jk}\left(\frac{n^{3\epsilon}}{\sqrt{2}}\mathbf{1} - \frac{n^{-3\epsilon}}{\sqrt{2}}\tilde{N}^k_{\boldsymbol{\Delta}_n}\right) \approx Z_i$$

We are now ready to translate the above CV measurement into its corresponding projective qudit measurement using the correspondence between displaced number operators measurements and rotated bases, described in section 5.6.2.2.

Using the general CLT map (5.25), we identify vectors $\{|\tilde{1}\rangle, \ldots, |\widetilde{m}\rangle\}$ in the orthogonal complement of $|\tilde{0}\rangle = |0\rangle \otimes |0'\rangle$ such that their corresponding limit quadratures are $X(|\tilde{k}\rangle) = \tilde{Q}_k$, for $k = 1, \ldots, m$. By virtue of the CLT the vectors $|\tilde{0}\rangle, |\tilde{1}\rangle, \ldots, |\widetilde{m}\rangle$ are normalised and orthogonal to each other, so we can complete

the set to an ONB $\tilde{\mathcal{B}} := \{|\tilde{0}\rangle, \ldots \widetilde{|d^2-1\rangle}\}$ of $\mathbb{C}^d \otimes \mathbb{C}^d$ where the remaining vectors are chosen arbitrarily. Now let $\tilde{\mathcal{B}}_n$ be the rotated basis

$$|v_j^{\delta_n}\rangle = U(\delta_n)|\tilde{j}\rangle = \exp\left(-i\delta_n \sum_{k=1}^{m} \sigma(i\tilde{k})\right)|\tilde{j}\rangle$$

for $\delta_n = n^{-1/2+3\epsilon}$ and $\sigma(i\tilde{k}) := -i|\tilde{0}\rangle\langle\tilde{k}| + i|\tilde{k}\rangle\langle\tilde{0}|$. Note that $\tilde{\mathcal{B}}_n$ is a small rotation of the basis $\mathcal{B}$ which contains the reference state $|\tilde{0}\rangle = |0\rangle \otimes |0'\rangle$, so the corresponding measurement is of the displaced-null type.

We measure each of the qudits in the basis $\tilde{\mathcal{B}}_n$ and obtain IID outcomes $X_1, \ldots, X_n$ taking values in $\{0, \ldots, d^2-1\}$, and let $p_{\boldsymbol{u}}^{(n)}$ be their distribution:

$$p_{\boldsymbol{u}}^{(n)}(j) = |\langle \psi_{\boldsymbol{u}/\sqrt{n}} \otimes 0' | v_j^{\delta_n}\rangle|^2, \quad j = 0, \ldots, d^2-1.$$

The following theorem is one of the main results of the chapter and shows that the Holevo bound can be attained by using displaced-null measurements.

Theorem 17. *Assume we are given n samples of the qudit state $|\psi_{\boldsymbol{\theta}}\rangle$ where $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^m$ is unknown. We further assume that Θ is bounded and $\epsilon < 1/10$. Using $\tilde{n} = n^{1-\epsilon}$ samples, we compute a preliminary estimator $\tilde{\boldsymbol{\theta}}_n$, and we measure the rest of the systems in the ONB $\tilde{\mathcal{B}}_n$, as defined above. Let*

$$\hat{\boldsymbol{\theta}}_n := \tilde{\boldsymbol{\theta}}_n + \hat{\boldsymbol{u}}_n/\sqrt{n}$$

be the estimator with

$$\hat{u}_n^j = \sum_{k=1}^{m} T_{jk}\left(\frac{n^{3\epsilon}}{\sqrt{2}} - \frac{n^{1-3\epsilon}}{\sqrt{2}}\hat{p}_n(k)\right), \quad j = 1, \ldots, m$$

where $\hat{p}_n(j)$ is the empirical estimator of $p_{\boldsymbol{u}}^{(n)}(j)$, i.e.,

$$\hat{p}_n(j) = \frac{|\{i : X_i = j,\ i = 1, \ldots, n\}|}{n},$$

for $j = 1, \ldots, m$.

Then $\hat{\boldsymbol{\theta}}_n$ is asymptotically optimal in the sense that for every $\boldsymbol{\theta} \in \Theta$

$$\lim_{n\to\infty} nR_n(\hat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}) = \mathcal{H}^{W(\boldsymbol{\theta})}(\boldsymbol{\theta})$$

Moreover, $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})$ converges in law to a centered normal random variable with covariance given by $TT^T/2$.

The proof of Theorem 17 can be found in section 5.10.

Our measurement has been obtained by modifying the optimal linear measurement for the limiting Gaussian shift to displaced counting one, and translating this to a qudit and ancilla measurement with repect to a displaced-null basis. Interestingly, this resulting measurement is closely connected to the optimal measurement described in [120]. The connection is discussed in section 5.11.

5.6.5 *Estimating a completely unknown pure state with respect to the Bures distance*

In this section we consider the problem of estimating a completely unknown pure qudit state, when the loss function (figure of merit) is defined as the squared Bures distance

$$d_b^2(|\psi\rangle\langle\psi|, |\phi\rangle\langle\phi|) = 2(1 - |\langle\psi|\phi\rangle|).$$

In this particular case, we will show that one can asymptotically achieve the Holevo bound using displaced-null measurement without the need of any ancillary system.

We parametrise a neighbourhood of the preliminary estimator $|0\rangle := |\tilde{\psi}_n\rangle$ as

$$|\psi_{\boldsymbol{u}/\sqrt{n}}\rangle = \exp\left(-i\sum_{k=1}^{d-1}(u_1^k\sigma_y^k - u_2^k\sigma_x^k)/\sqrt{n}\right)|0\rangle$$

where $\boldsymbol{u} = (u_1^1, u_2^1, \ldots, u_1^{d-1}, u_2^{d-1}) \in \mathbb{R}^{2(d-1)}$ satisfies $\|\boldsymbol{u}\| \leq n^{\epsilon}$ with high probability.

For small deviations from $|0\rangle$ the Bures distance has the quadratic approximation

$$d_b^2\left(|\psi_{\frac{\boldsymbol{u}}{\sqrt{n}}}\rangle\langle\psi_{\frac{\boldsymbol{u}}{\sqrt{n}}}|, |\psi_{\frac{\boldsymbol{u}'}{\sqrt{n}}}\rangle\langle\psi_{\frac{\boldsymbol{u}'}{\sqrt{n}}}|\right) = \frac{1}{n}\|\boldsymbol{u} - \boldsymbol{u}'\|^2 + o(n^{-1+2\epsilon})$$

which determines the optimal measurement and error rate in the asymptotic regime.

The Gaussian approximation consists in the model $|\boldsymbol{u}_1 + i\boldsymbol{u}_2\rangle$ and the optimal measurement with respect to the identity cost matrix would be to measure the Q_k's and P_k's. In order to estimate $\boldsymbol{u}$, instead of using an ancilla, we split the ensemble of n qudits in two equal sub-ensembles and perform separate 'displaced-null' measurements on each of them in the following bases which are obtained by rotating $\{|0\rangle, \ldots |d-1\rangle\}$ by (small) angles of size $\delta_n = n^{-1/2+3\epsilon}$

$$|v_j^{\delta_n}\rangle = U_1(\delta_n)|j\rangle = \exp\left(-i\delta_n\sum_{k=1}^{d-1}\sigma_y^k\right)|j\rangle \tag{5.31}$$

$$|w_j^{\delta_n}\rangle = U_2(\delta_n)|j\rangle = \exp\left(i\delta_n\sum_{k=1}^{d-1}\sigma_x^k\right)|j\rangle. \tag{5.32}$$

Therefore in the asymptotic picture, the proposed measurements are effectively joint measurements of $\{Q_i, i = 1, \ldots d-1\}$ and respectively $\{P_i, i = 1, \ldots d-1\}$ which are known to be optimal measurements for the local parameter $\boldsymbol{u}$ in the Gaussian shift model when performed on two separate copies of $|(\boldsymbol{u}_1 + i\boldsymbol{u}_2)/\sqrt{2}\rangle$ obtained from the original state by using a beamsplitter.

Let $X_1, \ldots, X_{n/2}$ and $Y_1, \ldots, Y_{n/2}$ be the independent outcomes of the two types of measurements, taking values in $\{0, \ldots, d-1\}$, and let $p_{\boldsymbol{u}}^{(n)}$ and $q_{\boldsymbol{u}}^{(n)}$ be their respective distributions

$$p_{\boldsymbol{u}}^{(n)}(j) = |\langle\psi_{\boldsymbol{u}/\sqrt{n}}|v_j^{\delta_n}\rangle|^2, \quad q_{\boldsymbol{u}}^{(n)}(j) = |\langle\psi_{\boldsymbol{u}/\sqrt{n}}|w_j^{\delta_n}\rangle|^2. \tag{5.33}$$

Proposition 4. *Assume $\epsilon < 1/10$ and let*

$$|\hat{\psi}\rangle := |\psi_{\hat{\boldsymbol{u}}/\sqrt{n}}\rangle$$

be the state estimator with local parameter $\hat{\boldsymbol{u}}_n$ defined as

$$\begin{aligned}\hat{u}_1^j &= \frac{n^{3\epsilon}}{2} - \frac{n^{1-3\epsilon}}{2}\hat{p}_n(j),\\ \hat{u}_2^j &= \frac{n^{3\epsilon}}{2} - \frac{n^{1-3\epsilon}}{2}\hat{q}_n(j), \quad j = 1,\dots,d-1,\end{aligned}$$

where $\hat{p}_n$, $\hat{q}_n$ are the empirical estimator of $p_{\boldsymbol{u}}^{(n)}$ and $q_{\boldsymbol{u}}^{(n)}$, respectively, i.e.,

$$\begin{aligned}\hat{p}_n(j) &= \frac{|\{i : X_i = j,\ i = 1,\dots,n/2\}|}{n/2},\\ \hat{q}_n(j) &= \frac{|\{i : Y_i = j,\ i = 1,\dots,n/2\}|}{n/2},\end{aligned}$$

for $j = 1,\dots,d-1$.

Then under $\mathbb{P}_{\boldsymbol{u}}$, $\sqrt{n}(\hat{\boldsymbol{u}}_n - \boldsymbol{u})$ is asymptotically distributed as a centered Gaussian random vector with covariance $\mathbf{1}/2$ and $|\hat{\psi}_n\rangle$ is asymptotically optimal in the sense that it achieves the Holevo bound:

$$\lim_{n\to\infty} n\mathbb{E}_{|\psi\rangle}[d_b^2(|\psi\rangle\langle\psi|, |\hat{\psi}_n\rangle\langle\hat{\psi}_n|)] = d-1.$$

The proof of Proposition 4 can be found in section 5.12.

5.6.6 *Achieving the QCRB with displaced-null measurements*

We now consider quantum statistical models for which the QCRB is (asymptotically) achievable. In contrast to models discussed in sections 5.6.4 and 5.6.5, in this case all parameter components can be estimated simultaneously at maximum precision. We will provide a class a displaced-null measurements which achieve the QCRB asymptotically.

Let us consider the statistical model $\{|\psi_{\boldsymbol{\theta}}\rangle\}$, $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^m$ with $m \leq 2(d-1)$ and assume that the parameter is identifiable and that the QCRB is achievable for all $\theta \in \Theta$. This is equivalent to condition (5.21) for all $\boldsymbol{\theta} \in \Theta$. The QFI is given by

$$F(\boldsymbol{\theta})_{ij} = 4\langle\partial_i\psi_{\boldsymbol{\theta}}|\partial_j\psi_{\boldsymbol{\theta}}\rangle - 4\langle\psi_{\boldsymbol{\theta}}|\partial_j\psi_{\boldsymbol{\theta}}\rangle\langle\partial_i\psi_{\boldsymbol{\theta}}|\psi_{\boldsymbol{\theta}}\rangle,$$

for $i,j = 1,\dots,m$. Let $|0\rangle := |\psi_{\tilde{\boldsymbol{\theta}}_n}\rangle$ be the preliminary estimator. We write $\boldsymbol{\theta} = \tilde{\boldsymbol{\theta}}_n + \boldsymbol{u}/\sqrt{n}$ with $\boldsymbol{u}$ the local parameter satisfying $\|\boldsymbol{u}\| \leq n^{\epsilon}$ with high probability. We assume that the phase of $|\psi_\theta\rangle$ has been chosen such that $\langle\dot{\psi}_i|0\rangle = 0$ for all i, and denote $\dot{\psi}_i := \partial_i\psi_{\tilde{\boldsymbol{\theta}}_n}$.

We now describe a class of measurements that will be shown to achieve the QCRB asymptotically. We choose an orthonormal basis $\mathcal{B} := \{|0\rangle, |1\rangle, \dots, |d-1\rangle\}$ whose first vector is $|0\rangle$ and the other vectors satisfy

$$c_{ki} := \langle k|\dot{\psi}_i\rangle \in \mathbb{R}, \qquad i = 1,\dots,m, \quad k = 1,\dots,d-1. \tag{5.34}$$

This condition is similar to equation (7) in [137], but unlike this reference we do not impose additional conditions for the case when $\langle k|\dot{\psi}_i\rangle = 0$ for all $i = 1,\dots,m$. If we assume that the parameter $\boldsymbol{\theta}$ is identifiable, then the matrix $C = (c_{ki})$ needs to have rank m.

We will further rotate $\mathcal{B}$ with a unitary $U = \exp(-i\delta_n G)$ where $\delta_n = n^{-1/2+3\epsilon}$ and

$$G = \sum_{k=1}^{d-1} g_k \sigma_y^k, \quad \sigma_y^k = -i\,|0\rangle\langle k| + i\,|k\rangle\langle 0|$$

where $g_k \neq 0$ are arbitrary real coefficients. We obtain the ONB $\{|v_0^{\delta_n}\rangle, \dots |v_{d-1}^{\delta_n}\rangle\}$ with

$$|v_k^{\delta_n}\rangle = U|k\rangle, \qquad k = 0, \dots d-1.$$

We measure all the systems in the basis $\tilde{\mathcal{B}}$ and obtain IID outcomes $X_1, \dots, X_n \in \{0, \dots, d-1\}$ and denote by $\hat{p}_n$ the corresponding empirical frequency. We denote by $T = (T_{ij})$ the $m \times (d-1)$ matrix defined as

$$T = (C^T C)^{-1} C^T.$$

Proposition 5. *Assume that Θ be bounded and $\epsilon < 1/10$. Let $\hat{\boldsymbol{\theta}}_n = \tilde{\boldsymbol{\theta}}_n + \hat{\boldsymbol{u}}_n/\sqrt{n}$ be the estimator determined by*

$$\hat{\boldsymbol{u}}_n^j = \sum_{k=1}^{d-1} T_{jk} \left(\frac{g_k n^{3\epsilon}}{2} - \frac{n^{1-3\epsilon}}{2g_k} \hat{p}_n(k) \right).$$

Then $\hat{\boldsymbol{\theta}}_n$ achieves the QCRB, i.e.,

$$\lim_{n\to\infty} n\mathbb{E}_{\boldsymbol{\theta}}[(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})^T] = F(\boldsymbol{\theta})^{-1}.$$

The proof of Proposition 5 can be found in section 5.13.

We now give a QLAN interpretation of the above construction. The fact that c_{ki} are real implies that the linearisation of the model around the preliminary estimation is given by

$$|\tilde{\psi}_{\boldsymbol{u}/\sqrt{n}}\rangle = \exp\left(-i \sum_{j=1}^{m} u_j S_j/\sqrt{n} \right) |0\rangle$$

with

$$S_j = \sum_{k=1}^{d-1} c_{kj} \sigma_y^k, \qquad c_{kj} = \langle k|\dot{\psi}_j\rangle.$$

By QLAN, the corresponding Gaussian model consists of coherent states $|C\boldsymbol{u}\rangle$ of a $(d-1)$-modes CV system where $C : \mathbb{R}^m \to \mathbb{C}^{d-1}$ is given by the *real* coefficients $c_{kj} = \langle k|\dot{\psi}_j\rangle$. This means that each of the $(d-1)$ modes is in a coherent state whose displacement is along the Q axis, so $\langle C\boldsymbol{u}|P_k|C\boldsymbol{u}\rangle = 0$ for all k, while

$$q_k := \langle C\boldsymbol{u}|Q_k|C\boldsymbol{u}\rangle = \sqrt{2}\sum_{j=1}^{m} c_{kj} u_j.$$

As we mentioned in Section 5.6.2.1, the QCRB is achievable for the limit model too and the simultaneous measurement of all Q_k is optimal. This is asymptotically obtained by the counting in the rotated basis.

5.7 CONCLUSIONS AND OUTLOOK

In this chapter we showed that the framework of displaced-null measurements provides a general scheme for optimal estimation of unknown parameters $\boldsymbol{\theta} \in \mathbb{R}^m$ of pure states models $|\psi_{\boldsymbol{\theta}}\rangle \in \mathbb{C}^d$. In particular, displaced-null measurements achieve the quantum Cramér-Rao bound (QCRB) for models in which the bound is achievable, and the Holevo bound for general qudit models.

Our method is related to previous works [115, 136, 137] that deal with the achivebility of the QCRB for pure state models $|\psi_{\boldsymbol{\theta}}\rangle$. These works exhibit a class of parameter-dependent orthonormal bases $\mathcal{B}(\tilde{\boldsymbol{\theta}})$ whose associated classical Fisher information $I_{\tilde{\boldsymbol{\theta}}}(\boldsymbol{\theta})$ converges to the quantum Fisher information $F(\boldsymbol{\theta})$ of $|\psi_{\boldsymbol{\theta}}\rangle$ as $\tilde{\boldsymbol{\theta}}$ approaches the true unknown state parameter $\boldsymbol{\theta}$. The measurement basis $\mathcal{B}(\tilde{\boldsymbol{\theta}})$ has the special feature that it contains the state $|\psi_{\tilde{\boldsymbol{\theta}}}\rangle$ as one of its elements, so that at $\tilde{\boldsymbol{\theta}} = \boldsymbol{\theta}$ the measurement has only one outcome, while for $\tilde{\boldsymbol{\theta}} \approx \boldsymbol{\theta}$ the occurrence of other outcomes can be interpreted as signaling the deviation from the reference value $\tilde{\boldsymbol{\theta}}$. With this in mind we called such measurements, null measurements.

However, the references [115, 136, 137] do not provide an explicit operational implementation of a strategy that achieves the QCRB. The naive solution would be to choose the reference parameter as a preliminary estimator $\tilde{\boldsymbol{\theta}}_n$ obtained by measuring a sub-sample of $\tilde{n} \ll n$ systems, and to apply the approximate null measurement $\mathcal{B}_{\tilde{\boldsymbol{\theta}}_n}$ to the rest of the systems. Surprisingly, it turned out that this adaptive strategy fails to achieve the QCRB, and indeed does not even reach the standard n^{-1} scaling of precision, when the preliminary estimator satisfies certain natural assumptions. This is due to the fact that $\tilde{\boldsymbol{\theta}}_n$ lies in the interior of a confidence interval of $\boldsymbol{\theta}$ and the measurement cannot distinguish positive and negative deviations from the reference since probabilities depend on the square of the deviations. This is an important finding which shows the pitfalls of drawing statistical conclusions based solely on Fisher information arguments.

To avoid this issue, we proposed to displace the preliminary estimator by a small amount $\boldsymbol{\delta}_n$ which is however sufficiently large to ensure that the new reference parameter $\tilde{\boldsymbol{\theta}}_n + \boldsymbol{\delta}_n$ is outside the confidence interval of $\boldsymbol{\theta}$. Building on this idea we showed the achievability of the QCRB in the setting of [115, 136, 137]. Furthermore, for general pure state models and locally quadratic loss functions, we devised displaced-null measurements which achieve the Holevo bound asymptotically for arbitrary qudit models.

The theory of quantum local asymptotic normality (QLAN) has played an important role in our investigations. The QLAN machinery translates the multi-copy estimation problem into one about estimating the mean of a multi-mode coherent state. In the latter case, counting measurements are paradigmatic example of null-measurements, while appropriately displacing the number operators provides the basis for displaced-null measurements. Using the QLAN correspondence, this translates into a simple prescription for rotating a basis containing the preliminary estimator $|\psi_{\tilde{\boldsymbol{\theta}}_n}\rangle$ into that of the displaced-null measurement. Interestingly, the obtained measurement turned out to be closely related to the parameter-dependent measurements proposed by Matsumoto in [120], and our approach offers an alternative asymptotic perspective on this work.

An exciting area of applications for displaced-null measurements is that of optimal estimation of dynamical parameters of open systems [37, 56, 62, 65, 83,

84, 86, 98]. Recent works [78, 177] have shown out that quantum post-processing by means of coherent absorbers allows for optimal estimation of such parameters. In particular [177] pointed out that a basic measurement such as photon counting constitutes a null-measurement, thus opening the route for devising optimal measurements for multidimensional estimation of Markov dynamics. An asymptotic analysis of displaced null measurements in this context will be the subject of a forthcoming publication [75].

Another area of future interest is to extend the method to models consisting of mixed states. While this may not work for arbitrary mixed states, the ideas presented here may be useful for models consisting of states with a high degree of purity which is the relevant setup in many quantum technology applications. Another important extension is towards refining the methodology for optimal estimation in the finite sample rather than asymptotic regime. Finally, we would like to better understand how displaced-null measurements can be used in the context of quantum metrology and interferometry [90, 148].

5.8 PROOF OF THEOREM 16 FOR WEAKER NOTIONS OF UNIDENTIFIABILITY

As we already mentioned, in the proof of Theorem 16 we made use of the fact that for the statistical model defined in equation (5.11), the law of the measurements in the basis containing $|\psi_{\tilde{\theta}_n}\rangle$ could not distinguish between $\tilde{\theta}_n \pm r$.

In the qubit case, we can still prove Theorem 16 for a wider class of one-parameter models under two additional assumptions. The first one is asking that unidentifiable parameters concentrate around the preliminary estimate at the same speed on both sides; more precisely, let us consider a general (smooth) one parameter model $|\psi_{\tilde{\theta}_n+r}\rangle$ for $r \in (-a, a)$; the corresponding probabilities describing the measurement in the $\tilde{\theta}_n$-null-basis are given by

$$p_r(1) = |\langle \psi_{\tilde{\theta}_n} | \psi_{\tilde{\theta}_n+r} \rangle|^2 = 1 - p_r(0).$$

In general, there is no reason why $p_r = p_{-r}$, however at $r = 0$ the function $p_r(1)$ has a global minimum and we can pick a neighborhood $(-a', b')$ such that

1. $p_r(1)$ is invertible on $(-a', 0]$ and $[0, b')$,
2. $p_r(1)$ maps both $(-a', 0]$ and $[0, b')$ onto the same interval.

A priori, the neighborhood depends on $\tilde{\theta}_n$, but if the preliminary estimator takes value in a compact set, we can find a nonempty neighborhood $(-a', b')$ that works for every value of $\tilde{\theta}_n$. If we denote by $r'(r)$ the unique value in $(-a', 0]$ such that $p_{r'(r)}(1) = p_r(1)$ for $r \in [0, b')$, we require

$$\begin{aligned} &r \mapsto r'(r) \text{ and its inverse to be Lipschitz with a} \\ &\text{Lipschitz constant } L \text{ that is uniform in } \tilde{\theta}_n. \end{aligned} \tag{5.35}$$

The second assumption consists in replacing the condition in equation 5.13 with

$$\min \left\{ \int_{\tau_n}^{\tau_n'} g_{n,\tilde{\theta}_n}(r) dr, \int_{\tau_n/L}^{L\tau_n'} g_{n,\tilde{\theta}_n}(r) dr \right\} \geq C \tag{5.36}$$

where $\tau_n := n^{-(1-\epsilon+\alpha)/2}$, $\tau'_n := n^{-(1-\epsilon-\beta)/2}$ for some α, β such that $\tau'_n = o(1)$, and $C > 0$ is a constant independent on n. The additional requirement is that the posterior measure concentrates around the preliminary estimator. Under assumptions (5.35) and (5.36), the proof of Theorem 16 can be adapted quite straightforwardly. This reparametrisation trick, however, cannot be repeated in the qudit case.

More generally, if instead of conditions 5.13 and 5.36, we assume that

$$\int_{\tau_n}^{\tau'_n} g_{n,\tilde{\theta}_n}(r)dr \geq C \tag{5.37}$$

where τ_n, τ'_n and $C > 0$ satisfy the same conditions above and, moreover, we require the preliminary estimator to be enough accurate, i.e., $\epsilon < 1/3 - (2\alpha + 5\beta)/3$, we can prove Theorem 16 without any assumption on the statistical model. We will make use of the fact that the conditional law of the measurements in the $\tilde{\theta}_n$-null basis conditional to the parameter $\theta = \tilde{\theta}_n + r$ does not distinguish between $\theta = \tilde{\theta}_n \pm r$ locally (which is the condition equivalent to have zero Fisher information), i.e.,

$$\dot{p}_0(0) = \dot{p}_0(1) = 0,$$

where the derivative is taken with respect to r.

Proposition 6. *Consider any one-parameter qudit model $\{|\psi_\theta\rangle\}$ and assume that $\tilde{\theta}_n$ is a reasonably good estimator satisfying condition (5.37), obtained by measuring a sub-ensemble of size $n^{1-\epsilon}$ with $\epsilon < 1/3 - (2\alpha + 5\beta)/3$. Let $\hat{\theta}_n$ be an estimator of θ based on measuring the remaining $n - n^{1-\epsilon}$ sub-ensemble in a basis containing $|\psi_{\tilde{\theta}_n}\rangle$. Then*

$$\lim_{n\to\infty} nR_\pi(\hat{\theta}_n) = \infty.$$

Proof. First notice that

$$\begin{aligned}\mathbb{E}[(\hat{\theta}_n - \theta)^2] \geq\ & \mathbb{P}(|\hat{\theta}_n - \theta| > \tau'_n)\tau'^2_n \\ & + \mathbb{E}\left[\chi_{|\hat{\theta}_n-\theta|\leq\tau'_n}(\hat{\theta}_n - \theta)^2\right].\end{aligned}$$

Moreover we can write

$$\begin{aligned}&\int_{|\hat{\theta}_n-\theta|\leq\tau'_n} \pi(d\theta|\tilde{\theta}_n)p(d\hat{\theta}_n|\theta,\tilde{\theta}_n)(\hat{\theta}_n - \theta)^2 \geq \\ &\int_{\substack{\tau_n\leq|\theta-\tilde{\theta}_n|\leq\tau'_n \\ |\hat{\theta}_n-\theta|\leq\tau'_n}} \pi(d\theta|\tilde{\theta}_n)p(d\hat{\theta}_n|\theta,\tilde{\theta}_n)(\hat{\theta}_n - \theta)^2 \geq \\ &\int_{\substack{\tau_n\leq r\leq\tau'_n \\ |\hat{\theta}_n-\theta|\leq\tau'_n}} g_{n,\tilde{\theta}_n}(r)(p_r(d\hat{\theta}_n)(\hat{\theta}_n - \tilde{\theta}_n - r)^2 + \\ &p_{-r}(d\hat{\theta}_n)(\hat{\theta}_n - \tilde{\theta}_n + r)^2) = \\ &\int_{\substack{\tau_n\leq r\leq\tau'_n \\ |\hat{\theta}_n-\theta|\leq\tau'_n}} g_{n,\tilde{\theta}_n}(r)(p_r(d\hat{\theta}_n) + p_{-r}(d\hat{\theta}_n))((\hat{\theta}_n - \tilde{\theta}_n)^2 + r^2)) + \\ &- 2\int_{\substack{\tau_n\leq r\leq\tau'_n \\ |\hat{\theta}_n-\theta|\leq\tau'_n}} g_{n,\tilde{\theta}_n}(r)(p_r(d\hat{\theta}_n) - p_{-r}(d\hat{\theta}_n))(\hat{\theta}_n - \tilde{\theta}_n)r\end{aligned}$$

The first addend can be lower bounded by

$$\int_{\substack{\tau_n \le r \le \tau_n' \\ |\hat{\theta}_n - \theta| \le \tau_n'}} g_{n,\tilde{\theta}_n}(r)(p_r(d\hat{\theta}_n) + p_{-r}(d\hat{\theta}_n))\tau_n^2.$$

The second addend will be negligible in the final analysis. Indeed, we have that

$$\begin{aligned}
&\sum_{x \in \{0,1\}^n} |\Pi_{k=1}^n p_r(x_k) - \Pi_{k=1}^n p_{-r}(x_k)| \le \\
&\sum_{l=1}^n \sum_{x \in \{0,1\}^n} p_{-r}(x_1) \cdots p_{-r}(x_{l-1}) |p_r(x_l) - p_{-r}(x_l)| \cdot \\
&\cdot q_r(x_{l+1}) \cdots q_r(x_n) = \\
&n \sum_{x=0,1} |p_r(x_l) - p_{-r}(x_l)| \lesssim n\tau_n'^3,
\end{aligned}$$

where in the last inequality we used that

$$p_0(x) - p_0(x) = \dot{p}_0(x) + \dot{p}_0(x) = \ddot{p}_0(x) - \ddot{p}_0(x) = 0$$

for $x = 0, 1$ and the symbol $\lesssim$ means that the left hand side is less or equal than a constant times the right hand side. Therefore, the second term can be upper bounded by a constant times $n\tau_n'^5 = n^{-3/2+5(\epsilon+\beta)/2}$, which is $o(\tau_n^2)$ if $\epsilon < 1/3 - (2\alpha + 5\beta)/3$. To sum up, one has that

$$\begin{aligned}
&\mathbb{E}[(\hat{\theta}_n - \theta)^2] \ge \mathbb{P}(|\hat{\theta}_n - \theta| > \tau_n')\tau_n'^2 + \\
&\int_{\substack{\tau_n \le r \le \tau_n' \\ |\hat{\theta}_n - \theta| \le \tau_n'}} \pi(d\tilde{\theta}_n) g_{n,\tilde{\theta}_n}(r)(p_r(d\hat{\theta}_n) + p_{-r}(d\hat{\theta}_n))\tau_n^2 + o(\tau_n^2) \ge \\
&\int_{\tau_n \le r \le \tau_n'} \pi(d\tilde{\theta}_n) g_{n,\tilde{\theta}_n}(r)(p_r(d\hat{\theta}_n) + p_{-r}(d\hat{\theta}_n))\tau_n^2 + o(\tau_n^2) \ge \\
&cC\tau_n^2 + o(\tau_n^2).
\end{aligned}$$

□

We remark that, if all the derivatives of p_r up to the $2s-1$-th order for some $s \ge 1$ vanish at 0, then we get that

$$\sum_{x=0,1} |p_r(x) - p_{-r}(x)| \lesssim \tau_n'^{2s+1}$$

and we obtain the statement under the assumption that $\epsilon < (2s-1)/(2s+1) - (2\alpha + (2s+3)\beta)/(2s+1)$. Notice that in general we can pick α and β arbitrarily small, hence the restriction on ϵ effectively becomes $\epsilon < (2s-1)/(2s+1)$, which does not preclude any value in the limit $s \to +\infty$.

5.9 PROOF OF PROPOSITION 3 ON OPTIMALITY OF DISPLACED NULL MEASUREMENTS

Since this measurement setting depends on n, we need to look in more detail at the asymptotic behaviour of the estimation problem.

We start by assuming that $\theta \in I_n$ and at the end of the proof we treat the case $\theta \notin I_n$ by employing the concentration bound in equation (5.15).

Since $\theta \in I_n$, we can write $\theta = \tilde{\theta}_n + u/n^{1/2}$ with local parameter u satisfying $|u| \leq n^{\epsilon}$. Then

$$\theta - \theta'_n = un^{-1/2} - n^{-1/2+3\epsilon} = O(n^{-1/2+3\epsilon})$$

and

$$\begin{aligned} p_\theta^{(n)} &= \sin^2(\theta - \theta'_n) = (\theta - \theta'_n)^2 + O(n^{-2+12\epsilon}) \\ &= n^{-1}(u - n^{3\epsilon})^2 + O(n^{-2+12\epsilon}). \end{aligned}$$

From this we get that

$$u = \frac{n^{3\epsilon}}{2} - \frac{n^{1-3\epsilon}}{2} p_\theta^{(n)} + O(n^{-\epsilon}) \tag{5.38}$$

where the u^2 term is negligible compared to $un^{3\epsilon}$ and the remainder is $O(n^{-\epsilon})$ for $\epsilon < 1/10$.

The probability $p_\theta^{(n)}$ can be estimated by the empirical frequency (5.16) whose distribution is the binomial $\text{Bin}(p_\theta^{(n)}, n)$. Taking into account that $\theta = \tilde{\theta}_n + u/n^{1/2}$ and using (5.38) we define the estimator

$$\hat{\theta}_n = \tilde{\theta}_n + \frac{n^{-1/2+3\epsilon}}{2} - \frac{n^{1/2-3\epsilon}}{2} \hat{p}_n \tag{5.39}$$

with $\hat{p}_n$ as in (5.16). Now from (5.38) we get

$$\sqrt{n}(\hat{\theta}_n - \theta) = \frac{n^{1-3\epsilon}}{2}(p_\theta^{(n)} - \hat{p}_n) + O(n^{-\epsilon}).$$

Conditional to a certain value of $\tilde{\theta}_n$, $\hat{p}_n$ has a binomial distribution with parameters $p_\theta^{(n)}$ and the term $O(n^{-\epsilon})$ is deterministic, hence

$$n\mathbb{E}[(\hat{\theta}_n - \theta)^2 | \tilde{\theta}_n] = \frac{n^{2-6\epsilon}}{4} p_\theta^{(n)}(1 - p_\theta^{(n)}) = \frac{1}{4} + o(1).$$

In order to study the convergence in law of $\sqrt{n}(\hat{\theta}_n - \theta)$, one can consider the conditional characteristic function of $n^{1-3\epsilon}(p_\theta^{(n)} - \hat{p}_n)/2$ instead (conditional to $\tilde{\theta}_n$, they only differ by a deterministic vanishing quantity):

$$\begin{aligned} &\mathbb{E}_\theta \left[\exp(ian^{1-3\epsilon}(\hat{p}_n - p_\theta^{(n)})/2) | \tilde{\theta}_n \right] \\ &= \mathbb{E}_\theta \left[\exp\left(ian^{-3\epsilon} \sum_{i=1}^{n} (X_i - p_\theta^{(n)})/2 \right) \middle| \tilde{\theta}_n \right] \\ &= \mathbb{E}_\theta \left[\exp(ian^{-3\epsilon}(X_1 - p_\theta^{(n)})/2 | \tilde{\theta}_n \right]^n \\ &= \left(1 - \frac{a^2}{8} n^{-6\epsilon} p_\theta^{(n)}(1 - p_\theta^{(n)}) + o(n^{-1}) \right)^n \\ &= \left(1 - \frac{a^2}{8n} + o(n^{-1}) \right)^n = e^{-a^2/8} + o(1). \end{aligned}$$

Notice that for every $a \in \mathbb{R}$

$$\begin{aligned} \mathbb{E}_\theta[e^{ia\sqrt{n}(\hat{\theta}_n - \theta)}] = e^{\frac{-a^2}{8}} &\mathbb{P}(\theta \in I_n) \\ &+ \int_{\theta \in I_n} p(d\tilde{\theta}_n | \theta)(\mathbb{E}_\theta[e^{ia\sqrt{n}(\hat{\theta}_n - \theta)} | \tilde{\theta}_n] - e^{\frac{a^2}{8}}) \\ &+ \int_{\theta \notin I_n} p(d\tilde{\theta}_n | \theta) \mathbb{E}_\theta[e^{ia\sqrt{n}(\hat{\theta}_n - \theta)} | \tilde{\theta}_n]. \end{aligned}$$

Since $\mathbb{P}_\theta(\theta \notin I_n)$ goes to zero, the first term goes to $e^{\frac{-a^2}{8}}$ and the third one vanishes. The second term vanishes because $|(\mathbb{E}_\theta[e^{ia\sqrt{n}(\hat{\theta}_n-\theta)}|\tilde{\theta}_n] - e^{\frac{a^2}{8}})|\chi_{\{\theta\in I_n\}}$ can be upper bounded uniformly in $\tilde{\theta}_n$ by a sequence converging to 0. Therefore we obtain the convergence of $\sqrt{n}(\hat{\theta}_n - \theta)$ to the normal, in distribution. For the convergece of the rescaled MSE we note that since $\theta, \tilde{\theta}_n$ are bounded, equation (5.39) shows that the square error $n(\hat{\theta}_n - \theta)^2$ does not grow more than n^2; using the fact that $\mathbb{P}_\theta(\theta \notin I_n)$ decays exponentially fast, one can remove the conditioning also in the convergence of the MSE.

□

5.10 PROOF OF THEOREM 17

We first assume that $\boldsymbol{\theta} \in I_n$ where

$$I_n = \{\boldsymbol{\theta} \in \mathbb{R}^m : \|\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}_n\| \leq n^{-1/2+\epsilon}\}.$$

Recall that $|0\rangle = |\psi_{\tilde{\theta}_n}\rangle$ is the preliminary estimator, and we denote $|\tilde{0}\rangle := |0\rangle \otimes |0'\rangle$ the first basis vector of an ONB $\tilde{B} := \{|\tilde{0}\rangle, \ldots, |\widetilde{d^2-1}\rangle\}$ in $\mathbb{C}^d \otimes \mathbb{C}^d$ which is chosen such that $|\tilde{1}\rangle, \ldots, |\tilde{m}\rangle$ are vectors corresponding to the canonical variables $\tilde{Q}_1, \ldots, \tilde{Q}_m$ which span the elements of the optimal unbiased set of observables $\mathbf{Z}^*$. Without loss of generality we can assume that $\{|\tilde{1}\rangle, \ldots, |\widetilde{2d-1}\rangle\}$ form an ONB of the subspace $\mathcal{L} := \mathrm{Lin}\{|0\rangle \otimes |i'\rangle, |i\rangle \otimes |0'\rangle : i = 1, \ldots d-1\}$. The local state (of system and ancilla) can be written as

$$|\tilde{\psi}_{\tilde{\boldsymbol{\theta}}_n + \boldsymbol{u}/\sqrt{n}}\rangle = e^{-i\sum_{k=1}^{2(d-1)}\left(\tilde{f}_1^k\left(\frac{\boldsymbol{u}}{\sqrt{n}}\right)\tilde{\sigma}_y^k - \tilde{f}_2^k\left(\frac{\boldsymbol{u}}{\sqrt{n}}\right)\tilde{\sigma}_x^k\right)}|\tilde{0}\rangle.$$

where $\tilde{f}_{1,2}^k$ are smooth real valued functions and $\tilde{\sigma}_{x,y}^k$ are the Pauli operators for the vectors in the basis $\tilde{\mathcal{B}}$.

From the definition of the basis $\tilde{\mathcal{B}}$, the subspace $\mathcal{L}$ and of the matrix T defined at the end of section 5.6.2.1 we have

$$(T^{-1})_{kj} = \sqrt{2}\partial_j \tilde{f}_1^k(0) \quad \text{for } j = 1, \ldots, m.$$

In particular, we note that

$$\frac{1}{2}\mathrm{Tr}(()W(\tilde{\boldsymbol{\theta}}_n)TT^T) = \mathcal{H}^{W(\tilde{\boldsymbol{\theta}}_n)}(\tilde{\boldsymbol{\theta}}_n). \tag{5.40}$$

Expanding the unitary rotation, one has

$$\begin{aligned}|\tilde{\psi}_{\tilde{\boldsymbol{\theta}}+\boldsymbol{u}/\sqrt{n}}\rangle = |\tilde{0}\rangle + \frac{1}{\sqrt{2}}\sum_{k=1}^m\left(\sum_{j=1}^m T_{kj}^{-1}\frac{u_j}{\sqrt{n}}\right)|\tilde{k}\rangle + \\ i\sum_{k=1}^m\left(\sum_{j=1}^m \partial_j f_2^k(0)\frac{u_j}{\sqrt{n}}\right)|\tilde{k}\rangle + \\ + O(n^{-1+2\epsilon}).\end{aligned} \tag{5.41}$$

The Taylor expansion for the vectors in the rotated basis is

$$|v_j^{\delta_n}\rangle = \exp\left(-i\delta_n\left(\sum_{k=1}^m \tilde{\sigma}_y^k\right)\right)|\tilde{j}\rangle \tag{5.42}$$
$$= \begin{cases} |\tilde{j}\rangle - \delta_n|\tilde{0}\rangle + O(n^{-1+6\epsilon}) & \text{if } j = 1,\dots,m \\ |\tilde{j}\rangle & \text{otherwise} \end{cases}.$$

Therefore one obtain the following expression for the outcome probability measure

$$p_{\boldsymbol{u}}^{(n)}(k) = \frac{1}{n}\left(\sum_{j=1}^m \frac{T_{kj}^{-1}}{\sqrt{2}}u_j - n^{3\epsilon}\right)^2 + \frac{1}{n}\left(\sum_{j=1}^m \partial_j f_2^k(0)u_j\right)^2 + O(n^{-3/2+9\epsilon})$$

if $k = 1,\dots,m$ and $p_{\boldsymbol{u}}^{(n)}(k) = O(n^{-1+2\epsilon})$ otherwise. Using the fact that $\|\boldsymbol{u}\| \leq n^{\epsilon}$ one can neglect the quadratic terms in $\boldsymbol{u}$ and write

$$u_j = \sum_{k=1}^m T_{jk}\left(\frac{n^{3\epsilon}}{\sqrt{2}} - \frac{n^{1-3\epsilon}}{\sqrt{2}}p_{\boldsymbol{u}}^{(n)}(k)\right) + O(n^{-\epsilon}).$$

Moreover, from explicit computations one can see that for every $j \neq k = 1,\dots,m$

$$\frac{n^{2-6\epsilon}}{2}\mathbb{E}_{\boldsymbol{\theta}}[(\hat{p}_n(j) - p_{\boldsymbol{u}}^{(n)}(j))^2|\tilde{\boldsymbol{\theta}}_n] =$$
$$\frac{n^{1-6\epsilon}}{2}p_{\boldsymbol{u}}^{(n)}(j)(1 - p_{\boldsymbol{u}}^{(n)}(j)) = \frac{1}{2} + o(1),$$

and

$$\frac{n^{2-6\epsilon}}{2}\mathbb{E}_{\boldsymbol{\theta}}[(\hat{p}_n(j) - p_{\boldsymbol{u}}^{(n)}(j))(\hat{p}_n(k) - p_{\boldsymbol{u}}^{(n)}(k))|\tilde{\boldsymbol{\theta}}_n] =$$
$$-\frac{n^{1-6\epsilon}}{2}p_{\boldsymbol{u}}^{(n)}(j)p_{\boldsymbol{u}}^{(n)}(k) = 0 + o(1).$$

Therefore

$$\begin{aligned} &n\mathbb{E}_{\boldsymbol{\theta}}[L(\boldsymbol{\theta},\hat{\boldsymbol{\theta}}_n)^2|\tilde{\boldsymbol{\theta}}_n] \\ &= \mathbb{E}_{\boldsymbol{\theta}}[\mathrm{Tr}((\hat{\boldsymbol{u}}_n - \boldsymbol{u})^T W(\tilde{\boldsymbol{\theta}}_n)(\hat{\boldsymbol{u}}_n - \boldsymbol{u}))|\tilde{\boldsymbol{\theta}}_n] + o(1) \\ &= \frac{n^{2-6\epsilon}}{2}\sum_{j,k=1}^m (T^T W(\tilde{\boldsymbol{\theta}}_n)T)_{kj} \cdot \\ &\cdot \mathbb{E}_{\boldsymbol{\theta}}[(\hat{p}_n(j) - p_{\boldsymbol{u}}^{(n)}(j))(\hat{p}_n(k) - p_{\boldsymbol{u}}^{(n)}(k))|\tilde{\boldsymbol{\theta}}_n]) \\ &+ o(1) \\ &= \frac{1}{2}\mathrm{Tr}(W(\tilde{\boldsymbol{\theta}}_n)TT^T) = \mathcal{H}^{W(\tilde{\boldsymbol{\theta}}_n)}(\tilde{\boldsymbol{\theta}}_n) + o(1). \end{aligned}$$

In order to derive the asymptotic normality result for $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$, we first consider the characteristic function of $n^{1-3\epsilon}(\hat{p}_n - p_{\boldsymbol{\theta}}^{(n)})/\sqrt{2}$: for every $\boldsymbol{a} \in \mathbb{R}^{d-1}$, one has

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{\theta}}\left[\exp(in^{1-3\epsilon}\boldsymbol{a}\cdot(\hat{p}_n - p_{\boldsymbol{\theta}}^{(n)})/\sqrt{2})|\tilde{\boldsymbol{\theta}}_n\right] \\
&= \mathbb{E}_{\boldsymbol{\theta}}\left[\exp\left(in^{-3\epsilon}\sum_{i=1}^{n}\frac{\boldsymbol{a}}{\sqrt{2}}\cdot(X_i - p_{\boldsymbol{\theta}}^{(n)})\right)\middle|\tilde{\boldsymbol{\theta}}_n\right] \\
&= \mathbb{E}_{\boldsymbol{\theta}}\left[\exp\left(in^{-3\epsilon}\frac{\boldsymbol{a}}{\sqrt{2}}\cdot(X_1 - p_{\boldsymbol{\theta}}^{(n)})\right)\middle|\tilde{\boldsymbol{\theta}}_n\right]^n \\
&= \left(1 - \frac{\|\boldsymbol{a}\|^2}{4}n^{-6\epsilon}p_{\boldsymbol{\theta}}^{(n)}(1 - p_{\boldsymbol{\theta}}^{(n)}) + o(n^{-1})\right)^n \\
&= \left(1 - \frac{\|\boldsymbol{a}\|^2}{4n} + o(n^{-1})\right)^n = e^{-\|\boldsymbol{a}\|^2/4} + o(1).
\end{aligned}$$

Indeed, using that $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = n^{1-3\epsilon}T(\hat{p}_n - p_{\boldsymbol{\theta}}^{(n)})/\sqrt{2}$, one has that for every $\boldsymbol{a} \in \mathbb{R}^m$

$$\mathbb{E}_{\boldsymbol{\theta}}[\exp(i\sqrt{n}\boldsymbol{a}\cdot(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}))|\tilde{\boldsymbol{\theta}}_n] = e^{-\frac{\boldsymbol{a}^T\cdot TT^T\cdot\boldsymbol{a}}{4}} + o(1).$$

We can now remove the conditioning with respect to the preliminary estimate and take the limit for $n \to +\infty$ (we will only show the computations for the risk, but they are the same in the case of the characteristic function):

$$\begin{aligned}
&n\mathbb{E}_{\boldsymbol{\theta}}[L(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}_n)^2] = \mathcal{H}^{W(\boldsymbol{\theta})}(\boldsymbol{\theta})\mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta} \in I_n) \\
&+ \int_{\theta\in I_n} p(d\tilde{\theta}_n|\theta)(\mathcal{H}^{W(\tilde{\boldsymbol{\theta}}_n)}(\tilde{\boldsymbol{\theta}}_n) - \mathcal{H}^{W(\boldsymbol{\theta})}(\boldsymbol{\theta})) \\
&+ \int_{\theta\in I_n} p(d\tilde{\theta}_n|\theta)(n\mathbb{E}_{\boldsymbol{\theta}}[d(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}_n)^2|\tilde{\boldsymbol{\theta}}_n] - \mathcal{H}^{W(\tilde{\boldsymbol{\theta}}_n)}(\tilde{\boldsymbol{\theta}}_n)) \\
&+ \int_{\theta\notin I_n} p(d\tilde{\theta}_n|\theta)n\mathbb{E}_{\boldsymbol{\theta}}[d(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}_n)^2|\tilde{\boldsymbol{\theta}}_n].
\end{aligned}$$

The first term in the sum tends to $\mathcal{H}^{W(\boldsymbol{\theta})}$, while all the other ones tend to 0 because of the continuity of $\mathcal{H}^{W(\boldsymbol{\theta})}$, the fact that $n\mathbb{E}_{\boldsymbol{\theta}}[L(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}_n)^2|\tilde{\boldsymbol{\theta}}_n] - \mathcal{H}^{W(\tilde{\boldsymbol{\theta}}_n)}(\tilde{\boldsymbol{\theta}}_n)$ is uniformly bounded by a vanishing sequence on I_n and that the last term can be upper bounded by a constant times $n\mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta} \notin I_n)$. The same reasoning shows unconditional asymptotic normality.

□

5.11 COMPARISON WITH ESTIMATORS DEVELOPED IN [120]

In this section we elucidate the connection between the measurement strategy that we propose and the optimal measurement for pure statistical models pointed out in [120]. Theorems 1 and 2 in [120] show that for every parameter value $\boldsymbol{\theta}$, there exists a measurement basis that allows to attain the Holevo bound at $\boldsymbol{\theta}$ *in one shot*: given the optimal estimator $\mathbf{Z}$ of the limit Gaussian model at $\boldsymbol{\theta}$, one needs to consider the corresponding vectors $|z_1\rangle, \ldots, |z_m\rangle$ via QCLT (see Eq. (5.25)) and pick any orthonormal basis $\{|b_k\rangle\}_{k=0}^m$ of $\mathrm{span}_{\mathbb{R}}\{|\psi_{\boldsymbol{\theta}}\rangle, |z_1\rangle, \ldots, |z_m\rangle\}$ such that $\langle b_k|\psi_{\boldsymbol{\theta}}\rangle \neq 0$ for every k. The measurement in any orthonormal basis

containing $|b_k\rangle_{k=0}^m$ is optimal and the estimator achieving the Holevo bound is given by

$$\hat{\hat{\theta}}^i = \frac{\langle b_k|z_i\rangle}{\sqrt{2}\langle b_k|\psi_{\boldsymbol{\theta}}\rangle} + \theta^i$$

if k is observed for $k = 0, \dots, m$ and 0 otherwise.

As in the case of the SLD, such an optimal measurement depends on the true parameter; in order to come up with a concrete estimation strategy, one needs a two step procedure. After producing a preliminary estimate $\tilde{\boldsymbol{\theta}}_n$ of the parameter, one would then choose an orthonormal basis $\{b_k\}_{k=0}^m$ of $\mathrm{span}_{\mathbb{R}}\{|\psi_{\tilde{\theta}_n}\rangle, |z_1\rangle, \dots, |z_m\rangle\}$ such that $\langle b_k|\psi_{\tilde{\theta}_n}\rangle \neq 0$ for every k and measure in any orthonormal basis containing $\{b_k\}_{k=0}^m$. The final estimator would be given by

$$\hat{\hat{\theta}}_n^i = \sum_{k=0}^m \frac{\langle b_k|z_i\rangle}{\sqrt{2}\langle b_k|\psi_{\tilde{\theta}_n}\rangle}\hat{p}_n(k) + \tilde{\theta}_n^i, \tag{5.43}$$

where $\hat{p}_n(k)$ is the empirical probability of observing k.

However, Theorem 16 shows that if $\{b_k\}_{k=0}^m$ is too close to be a null-basis, such a strategy does not even achieve a standard scaling due to identifiability problems. The basis $\{|v_k^{\delta_n}\rangle\}_{k=0}^m$ that we propose satisfies the assumptions above for being optimal at $\tilde{\boldsymbol{\theta}}_n$ and ensures an asymptotically optimal estimation precision; moreover, in this case $\hat{\hat{\boldsymbol{\theta}}}_n$ and $\hat{\boldsymbol{\theta}}_n$ are equivalent in the following sense.

Proposition 7. *Let $\hat{\hat{\boldsymbol{\theta}}}_n$ and $\hat{\boldsymbol{\theta}}$ the estimators defined in Eq. (5.43) and Theorem 17, respectively. Then the following holds true:*

$$\lim_{n\to+\infty} n\mathbb{E}_{\boldsymbol{\theta}}[(\hat{\hat{\boldsymbol{\theta}}}_n - \hat{\boldsymbol{\theta}}_n)^2] = 0.$$

Proof. First we condition on $\tilde{\boldsymbol{\theta}}_n \in I_n$, where

$$I_n = \{\boldsymbol{\theta} \in \mathbb{R}^m : \|\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}_n\| \leq n^{-1/2+\epsilon)}\}.$$

Using Eq. (5.42) and $|z_i\rangle = \sum_{k=1}^m T_{ik}|\tilde{k}\rangle$, one has that

$$\langle b_k|z_i\rangle = T_{ik} - \frac{\delta_n^2}{2}\sum_{j=1}^m T_{ij} + O(n^{-3/2+9\epsilon}),$$
$$\langle b_k|\psi_{\tilde{\boldsymbol{\theta}}}\rangle = -\delta_n + O(n^{-3/2+9\epsilon})$$

for $k = 1, \dots, m$ and

$$\langle b_0|z_i\rangle = \delta_n\sum_{j=1}^m T_{ij} + O(n^{-3/2+9\epsilon}),$$
$$\langle b_0|\psi_{\tilde{\boldsymbol{\theta}}}\rangle = 1 + O(n^{-1+6\epsilon}).$$

Therefore we can write

$$\begin{aligned}\hat{\tilde{\theta}}_n^i - \tilde{\theta}_n^i &= \sum_{k=1}^m \frac{\langle b_k|z_i\rangle}{\sqrt{2}\langle b_k|\psi_{\tilde{\theta}_n}\rangle}\hat{p}_n(k) \\ &= \frac{n^{-1/2+3\epsilon}}{\sqrt{2}}\sum_{k=1}^m T_{ik}\hat{p}_n(0) - \frac{n^{1/2-3\epsilon}}{\sqrt{2}}\sum_{k=1}^m T_{ik}\hat{p}_n(k) \\ &+ \sum_{j=1}^m \frac{n^{-1/2+3\epsilon}}{4} T_{ij}\sum_{k=1}^m \hat{p}_n(k) + R \\ &= \hat{\tilde{\theta}}_n^i - \tilde{\theta}_n^i + R \\ &+ \frac{n^{-1/2+3\epsilon}}{\sqrt{2}}\sum_{k=1}^m T_{ik}(\hat{p}_n(0) - 1) \quad (I) \\ &+ \sum_{j=1}^m \frac{n^{-1/2+3\epsilon}}{4} T_{ij}\sum_{k=1}^m \hat{p}_n(k) \quad (II),\end{aligned}$$

where R is a random variable whose standard deviation is $O(n^{-3/2+9\epsilon})$. Moreover, both (I) and (II) are negligible too: indeed, for every $k = 0, \dots, m$

$$\mathbb{E}_{\boldsymbol{\theta}}[(\hat{p}_n(k) - p_{\boldsymbol{\theta}}^{(n)}(k))^2|\tilde{\boldsymbol{\theta}}_n] = o(1/n)$$

and

$$p_{\boldsymbol{\theta}}^{(n)}(k) = \delta_{0k} + O(n^{-1+6\epsilon}).$$

The statement follows removing the conditioning can be shown with the same technique as in the proof of Theorem 17. □

5.12 PROOF OF PROPOSITION 4

We denote by I_n the set of states

$$\{|\psi\rangle : d_b(|\psi\rangle\langle\psi|, |\tilde{\psi}_n\rangle, \langle\tilde{\psi}_n|) \leq n^{(1-\epsilon)/2}\}$$

and we assume that $|\psi\rangle$ belongs to I_n (the converse can be dealt with as in the proof of Theorem 17). Therefore, we can write

$$|\psi\rangle = \exp\left(-i\sum_{k=1}^{d-1}(u_1^k\sigma_y^k - u_2^k\sigma_x^k)/\sqrt{n}\right)|\tilde{\psi}_n\rangle$$

for some $\boldsymbol{u} = (u_1^1, u_2^1, \dots, u_1^{d-1}, u_2^{d-1}) \in \mathbb{R}^{2(d-1)}$ that satisfies $\|\boldsymbol{u}\| = O(n^\epsilon)$. Notice that for $j = 1, \dots, d$ one has

$$\begin{aligned}p_{\boldsymbol{u}}^{(n)}(j) &= |\langle\psi_{\boldsymbol{u}/\sqrt{n}}|v_j^{\delta_n}\rangle|^2 \\ &= \left|\left\langle j\left|\exp\left(i\delta_n\sum_{k=1}\sigma_y^k\right)\cdot\right.\right. \\ &\cdot \left.\left.\exp\left(-i\sum_{k=1}^{d-1}(u_1^k\sigma_y^k - u_2^k\sigma_x^k)/\sqrt{n}\right)\right|\tilde{\psi}_n\right\rangle\right|^2 \\ &= (u_1^j/\sqrt{n} - \delta_n)^2 + (u_2^j/\sqrt{n})^2 + O(n^{-2+12\epsilon}),\end{aligned}$$

where the last equality is obtained expanding the matrix exponential. Analogously one obtains

$$q_{\boldsymbol{u}}^{(n)}(j) = (u_1^j/\sqrt{n})^2 + (u_2^j/\sqrt{n} - \delta_n)^2 + O(n^{-2+12\epsilon})$$

for $j = 1, \dots, d-1$. This implies that

$$u_1^j = \frac{n^{3\epsilon}}{2} - \frac{n^{1-3\epsilon}}{2} p_{\boldsymbol{u}}^{(n)}(j) + O(n^{-\epsilon})$$

and

$$u_2^j = \frac{n^{3\epsilon}}{2} - \frac{n^{1-3\epsilon}}{2} q_{\boldsymbol{u}}^{(n)}(j) + O(n^{-\epsilon}).$$

Moreover, for every $j \neq k = 1, \dots, d-1$ one has

$$\frac{n^{2-6\epsilon}}{4} \mathbb{E}_{|\psi\rangle}[(\hat{p}_n(j) - p_{\boldsymbol{u}}^{(n)}(j))^2 | \, |\tilde{\psi}_n\rangle] = \frac{n^{1-6\epsilon}}{2} p_{\boldsymbol{u}}^{(n)}(j)(1 - p_{\boldsymbol{u}}^{(n)}(j)) = \frac{1}{2} + o(1),$$

and

$$\frac{n^{2-6\epsilon}}{4} \mathbb{E}_{|\psi\rangle}[(\hat{p}_n(j) - p_{\boldsymbol{u}}^{(n)}(j))(\hat{p}_n(k) - p_{\boldsymbol{u}}^{(n)}(k)) | \, |\tilde{\psi}_n\rangle] = -\frac{n^{1-6\epsilon}}{2} p_{\boldsymbol{u}}^{(n)}(j) p_{\boldsymbol{u}}^{(n)}(k) = 0 + o(1).$$

Another consequence is that for $\boldsymbol{a} \in \mathbb{R}^{d-1}$ one has

$$\begin{aligned}
&\mathbb{E}_{|\psi\rangle}\left[\exp(in^{1-3\epsilon}\boldsymbol{a} \cdot (\hat{p}_n - p_{\boldsymbol{u}}^{(n)})/2) | \, |\tilde{\psi}_n\rangle\right] \\
&= \mathbb{E}_{|\psi\rangle}\left[\exp\left(in^{-3\epsilon}\sum_{i=1}^{n/2} \boldsymbol{a} \cdot (X_i - p_{\boldsymbol{u}}^{(n)})\right) \middle| \, |\tilde{\psi}_n\rangle\right] \\
&= \mathbb{E}_{|\psi\rangle}\left[\exp(in^{-3\epsilon}\boldsymbol{a} \cdot (X_1 - p_{\boldsymbol{u}}^{(n)}) | \, |\tilde{\psi}_n\rangle\right]^{n/2} \\
&= \left(1 - \frac{\|\boldsymbol{a}\|^2}{2} n^{-6\epsilon} p_{\boldsymbol{u}}^{(n)}(1 - p_{\boldsymbol{u}}^{(n)}) + o(n^{-1})\right)^{n/2} \\
&= \left(1 - \frac{\|\boldsymbol{a}\|^2}{2n} + o(n^{-1})\right)^{n/2} = e^{-\|\boldsymbol{a}\|^2/4} + o(1).
\end{aligned}$$

The same can be proved for the other batch. Notice that

$$nd_b(|\psi\rangle\langle\psi|, |\hat{\psi}_n\rangle\langle\hat{\psi}_n|)^2 = \|\hat{\boldsymbol{u}} - \boldsymbol{u}\|^2 + o(1) = \frac{n^{2-6\epsilon}}{4} \sum_{j=1}^{d-1} (\hat{p}_n(j) - p_{\boldsymbol{u}}^{(n)}(j))^2 + (\hat{q}_n(j) - q_{\boldsymbol{u}}^{(n)}(j))^2 + o(1).$$

Therefore, if $|\psi\rangle \in I_n$

$$\mathbb{E}_{|\psi\rangle}[d_b(|\psi\rangle\langle\psi|, |\hat{\psi}_n\rangle\langle\hat{\psi}_n|)^2 | \, |\tilde{\psi}_n\rangle] = d - 1 + o(1)$$

and for every $\boldsymbol{a} \in \mathbb{R}^{2(d-1)}$

$$\mathbb{E}_{|\psi\rangle}[e^{i\boldsymbol{a}\cdot(\hat{\boldsymbol{u}}-\boldsymbol{u})} | \, |\tilde{\psi}_n\rangle] = e^{-\|\boldsymbol{a}\|^2/4} + o(1).$$

The rest of the proof is similar to the one of Proposition 3.

5.13 PROOF OF PROPOSITION 5

In order to avoid confusion and conversely to the main text, in this proof we stress the dependence of C and B on the preliminary estimator $\tilde{\boldsymbol{\theta}}_n$.

As usual, we assume that $\boldsymbol{\theta} \in I_n$, where

$$I_n = \{\boldsymbol{\theta} \in \mathbb{R}^m : \|\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}_n\| \leq n^{-1/2+\epsilon)}\};$$

then one can write $\boldsymbol{\theta} = \tilde{\boldsymbol{\theta}}_n + \boldsymbol{u}/\sqrt{n}$ for some $\boldsymbol{u}$ such that $\|u\| \leq n^{\epsilon}$ and the probability law of the X_j's is given by

$$\begin{aligned} p_{\boldsymbol{u}}^{(n)}(k) &= \left(\sum_{j=1}^{m} c_{kj}(\tilde{\boldsymbol{\theta}}_n)\theta^j - g_k\delta_n\right)^2 + O(n^{-3+9\epsilon}) \\ &= g_k^2 n^{-1+6\epsilon} - 2n^{-1+3\epsilon}\left(\sum_{j=1}^{m} c_{kj}(\tilde{\boldsymbol{\theta}}_n)u^j\right) g_k + O(n^{-1+2\epsilon}) \end{aligned}$$

for $k = 1, \ldots, d-1$. Equivalently, using that $B(\tilde{\boldsymbol{\theta}}_n)C(\tilde{\boldsymbol{\theta}}_n) = \mathbf{1}$, we can write

$$u^j = \sum_{k=1}^{d-1} b_{jk}(\tilde{\boldsymbol{\theta}}_n)\left(\frac{g_k n^{3\epsilon}}{2} - \frac{n^{1-3\epsilon}}{2g_k}p^{(n)}(k)\right) + O(n^{-\epsilon}).$$

Therefore

$$\begin{aligned} &n\mathbb{E}_{\boldsymbol{\theta}}[(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})^T(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})|\tilde{\boldsymbol{\theta}}_n] = \mathbb{E}_{\boldsymbol{\theta}}[(\hat{\boldsymbol{u}}_n - \boldsymbol{u})^T(\hat{\boldsymbol{u}}_n - \boldsymbol{u})|\tilde{\boldsymbol{\theta}}_n] \\ &= \frac{n^{2-6\epsilon}}{4}BG\mathbb{E}_{\boldsymbol{\theta}}[(\hat{p}_n - p_{\boldsymbol{u}}^{(n)})(\hat{p}_n - p_{\boldsymbol{u}}^{(n)})^T|\tilde{\boldsymbol{\theta}}_n]GB^T + o(1), \end{aligned}$$

where G is the diagonal matrix with entries given by $(1/g_k)_{k=1}^{d-1}$. Explicit computations show that

$$\frac{n^{2-6\epsilon}}{4}G\mathbb{E}_{\boldsymbol{\theta}}[(\hat{p}_n - p_{\boldsymbol{u}}^{(n)})(\hat{p}_n - p_{\boldsymbol{u}}^{(n)})^T|\tilde{\boldsymbol{\theta}}_n]G = \mathbf{1}/4 + o(1).$$

Therefore

$$n\mathbb{E}_{\boldsymbol{\theta}}[(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})^T|\tilde{\boldsymbol{\theta}}_n] = B(\tilde{\boldsymbol{\theta}}_n)B(\tilde{\boldsymbol{\theta}}_n)^T/4 + o(1).$$

Notice that $B(\tilde{\boldsymbol{\theta}}_n)B(\tilde{\boldsymbol{\theta}}_n)^T/4 = F(\tilde{\boldsymbol{\theta}}_n)^{-1}$: indeed, using the explicit expression of $B(\tilde{\boldsymbol{\theta}}_n)$

$$B(\tilde{\boldsymbol{\theta}}_n)B(\tilde{\boldsymbol{\theta}}_n)^T = (C(\tilde{\boldsymbol{\theta}}_n)^T C(\tilde{\boldsymbol{\theta}}_n))^{-1} = 4F(\tilde{\boldsymbol{\theta}}_n).$$

The rest of the proof is the same as in the one of Theorem 17 and uses the continuity of $F(\tilde{\boldsymbol{\theta}}_n)$.

6

ESTIMATING QUANTUM MARKOV CHAINS USING COHERENT ABSORBER POST-PROCESSING AND PATTERN COUNTING ESTIMATOR

6.1 INTRODUCTION

In this chapter we apply the displaced-null measurement (DNM) strategy presented in the previous chapter to a quantum Markov chain (QMC), cf. Chapter 2 and Figure 6.1 a). As discussed, this is an alternative to the usual approach that measures the symmetric logarithmic derivative (SLD). For one-dimensional parameters, this SLD measurement is optimal in the sense that it attains the quantum Cramér-Rao bound (QCRB) in the limit of many copies. However, when dealing with complex models involving correlated states of many-body systems, it may be hard to compute and implement in practice, and in the case of pure states (and more generally, rank deficient states) such a measurement is highly non-unique. Therefore, it is particularly important to devise *realistic* measurement schemes which allow the estimation of unknown parameters with close to optimal precision, by means of *computationally efficient* estimators.

The QMC setup used here is similar to Haroche's photon-box one-atom maser [91] and to that used in quantum collision models [42], and provides a physical mechanism for generating versatile many-body states such as matrix product states [132, 149] and finitely correlated states [57, 58]. By discretising time, QMCs can be used to model continuous-time dynamics of a Markovian open system coupled with Bosonic input-output channels [10, 61, 171].

For clarity, it is useful to distinguish between two mainstream approaches to parameter estimation in quantum open systems. In the setting of [18, 48, 79, 97, 150, 155, 184], the quantum system undergoes a noisy evolution depending on an unknown parameter, and the experimenter tries to extract information about the parameter by repeatedly applying instantaneous direct measurements and control operations while the system is evolving. In contrast, the setting adopted in this chapter is commonly used in quantum optics and input-output theory [43, 66, 81, 174] where the experimenter does not have direct access to the system but can measure the output field of an environment channel coupled to the system. This allows the experimenter to track the conditional state of the system by means of stochastic filtering equations [16, 22, 35, 45, 173] and control it using feedback. As these techniques require full knowledge of the system's dynamical parameters, it is important to devise tools for estimating such parameters from the stochastic trajectory of the measurement record. Since the early works [62, 117], many aspects of continuous-time estimation have been investigated, including

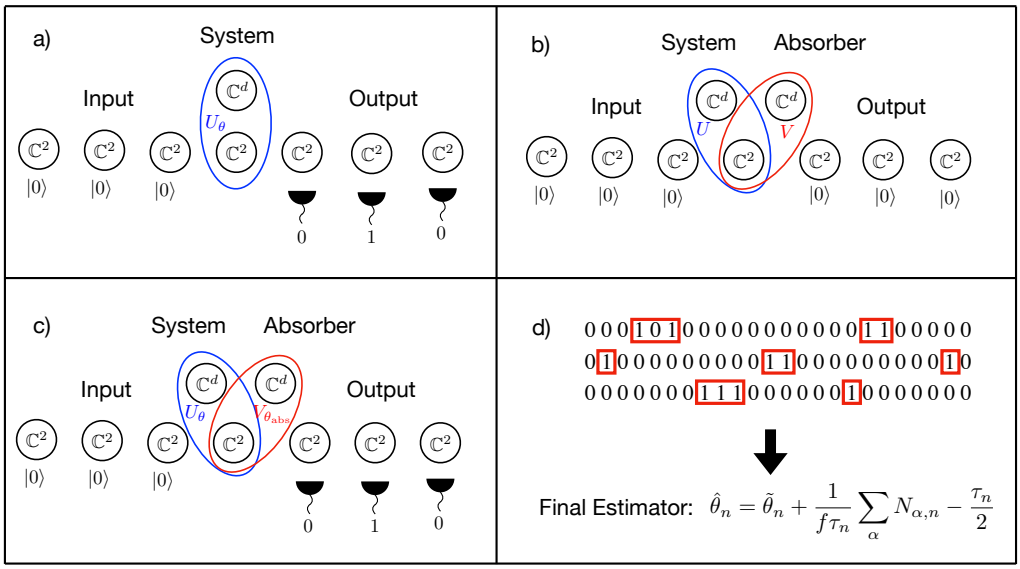

Figure 6.1: Basic elements of the pattern counting estimator. Panel a) A quantum Markov chain as a system interacting sequentially with the environment via a parameter dependent unitary U_θ. The first stage estimator $\tilde{\theta}_n$ is obtained by performing a standard sequential measurement on the output and equating empirical and expected counts. Panel b) Post-processing the output using a coherent absorber. When system and absorber parameters match, the output is identical to the input (vacuum) Panel c) After the first estimation stage the absorber is fixed at a value $\theta_{\text{abs}} = \tilde{\theta}_n - \delta_n$ where $\tilde{\theta}_n$ is the preliminary estimator and δ_n is the parameter shift required by the displaced-null measurement theory [76]. The output generated by the system and absorber dynamics with unitary $V_{\theta_{\text{abs}}} U_\theta$ is measured sequentially in the standard basis. Panel d) Given a measurement trajectory, excitation patterns are identified as binary sequences starting and ending with a 1 separated by long sequences of 0s. The final estimator is a correction to the preliminary estimator which depends only on the total number of patterns $\sum_\alpha N_{\alpha,n}$, the QFI f at $\tilde{\theta}_n$ and the displacement parameter τ_n.

adaptive estimation [20, 140] filtering methods [38, 142, 154], Heisenberg scaling [6, 7, 98, 118], sensing with error correction [138], Bayesian estimation [64, 105, 126, 143, 180], quantum smoothing [82, 161, 165, 166], estimation of linear systems [68, 85, 112, 113], central limit and large deviations theory for trajectories [31, 36, 67, 96], concentration bounds for time averaged observables [19, 74], estimation with feedback control [56].

However, a major problem in this area has been that standard measurement protocols such as counting and homodyne do not achieve the ultimate precision limit prescribed by the QCRB in terms of the *quantum Fisher information* (QFI) of the output state [37, 65, 68, 83, 84, 86].

Two recent papers [78, 177] have addressed this problem by introducing the idea of *quantum post-processing* of the output state using a *quantum coherent absorber* (CA) [157]. For a given QMC reference dynamics, the absorber takes the system's output as its own input and is characterised by the property that it "reverts" the action of the system, so that its output is a trivial product state (vacuum in continuous-time dynamics), cf. Figure 6.1 b). In the statistical estimation

framework, the absorber is set to a particular reference parameter of the QMC dynamics, so that a small deviation of the true parameter from this value will lead to non-trivial output statistics which can be used to estimate the deviation as illustrated in Figure 6.1 c). In [78], two of the present authors proved that the QCRB can be achieved by performing sequential, adaptive measurements on the output units, after the interaction with the coherent absorber. In addition, the adaptive measurement can be implemented efficiently in a Markovian fashion. The work [177] deals with the same problem but in the continuous-time setting, and proposes to perform a standard counting measurement (instead of an adaptive one) in conjunction with post-processing using the coherent absorber. In both papers the final estimator was computed from the measurement trajectory using the maximum likelihood method, which becomes inefficient for long trajectories.

The strategy proposed here is similar to that of [177], albeit in discrete rather than continuous time, but strengthens it in several important aspects. Firstly, we employ the technique of *displaced-null measurements* [76] to provide a precise recipe for choosing the coherent absorber parameter. As we explain below, this is an important technical detail, as the intuitive choice of parameters *fails* to achieve the QCRB in the limit of long trajectories. Secondly, our strategy employs a two step adaptive procedure which allows us to compute the final estimator as a simple linear transformation of the total number of "pattern counts" which can be easily extracted from the measurement trajectory, cf. Figure 6.1 d). This circumvents the computational issues associated with the maximum likelihood estimator. Thirdly, we provide strong theoretical evidence that the final estimator achieves the QCRB in the limit of large times. This is based on a novel representation of the output in terms of *translationally invariant modes* which are shown to satisfy the *quantum local asymptotic normality* property [32, 87, 89, 102].

We now give a brief summary of the two steps estimation strategy proposed here and the related mathematical results. We consider a QMC whose dynamics depends on a one-dimensional parameter θ which we aim to estimate by measuring the output state produced after n time steps. In the first stage we run the QMC dynamics with the unknown parameter θ for $1 \ll \tilde{n} \ll n$ time steps and measure the noise units in a fixed basis, cf. Figure 6.1 a). From the total counts statistics we construct a rough estimator $\tilde{\theta}_n$ by matching the empirical frequency to its expected value. In general this estimator is not optimal but its mean square error has the standard $1/\tilde{n}$ scaling ([74]). In the second stage we run the system *and* absorber QMC for the remaining $n - \tilde{n}$ time steps and measure the output in the standard basis. If the absorber parameter matched the true system parameter θ, this measurement would produce a string of 0s (corresponding to no counts in continuous-time) cf. Figure 6.1 b). Therefore, it would seem natural to choose the absorber parameter to be $\tilde{\theta}_n$, our best guess at the unknown parameter θ. However, this choice is unsuitable since for small deviations $\Delta_n = \theta - \tilde{\theta}_n$, the counting statistics depends quadratically on Δ_n, which prevents the estimation of θ at standard $1/n$ rate. This non-identifiability issue is explained in detail in [76], which also provides the solution to this problem. We deliberately set the absorber parameter at $\theta_{\text{abs}} = \tilde{\theta}_n - \delta_n$, which is away from the best guess by a small "displacement" $\delta_n \downarrow 0$ chosen to be larger than the uncertainty $|\Delta_n|$. This allows us to unambiguously identify θ from counts statistics. Stage two of the estimation procedure is illustrated in Figure 6.1 c).

We turn now to the question of estimating θ from the counts trajectory $\omega = (\omega_1, \omega_2, \ldots, \omega_n)$ of the second stage measurement (setting $n - \tilde{n}$ to n for simplicity). Since $\theta - \theta_{\text{abs}}$ is vanishingly small (even with the extra displacement), ω will typically contain a small number of 1s while most of the outcomes will be 0, cf. Figure 6.1 d). This allows us split the trajectory into long sequences of 0s and in between them, binary "excitation patterns" starting and ending with a 1. For each pattern α (e.g.- 1, 11, 101 etc.) we count the number of occurrences $N_{\alpha,n}$. In Theorem 20 we show that in the limit of large n the counts $N_{\alpha,n}$ become independent Poisson variables whose intensities are $\lambda_\alpha u^2$ where $u = \sqrt{n}(\theta - \tilde{\theta}_{\text{abs}})$ is the "local parameter" and λ_α is a model dependent coefficient. Moreover the total Fisher information of the Poisson variables is equal to the output QFI, which shows that the pattern counts statistics capture the full information of the output state. Using this asymptotic behaviour, we construct a simple estimator $\hat{\theta}_n$ (cf. equations (6.23) and (6.22)) which is linear in the total pattern count, and we argue why it should achieve the QCRB in the limit of large n. To summarise, the two step procedure provides a computationally and statistically efficient estimation method which involves only standard basis measurements and a minimal amount of "quantum post-processing" implemented by the coherent absorber.

For a more in-depth understanding of why the excitation pattern counts have asymptotically Poisson distributions, we refer to sections 6.4 and 6.5 where we develop a theory of *translationally invariant modes* (TIMs) of the output. These modes turn out to capture all statistical information about the unknown parameter, and can be measured simultaneously and optimally by performing the sequential standard output measurement. For each excitation pattern $\alpha = (\alpha_1, \ldots, \alpha_k) \in \{0,1\}^k$ we define the creation operator $A^*_\alpha(n)$ on the output chain of length n. This consists of a running average

$$A^*_\alpha(n) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n-k+1} \sigma^\alpha_i$$

where σ^α_i is the tensor product of the the type $\sigma^\alpha = \sigma^{\alpha_1} \otimes \cdots \otimes \sigma^{\alpha_k}$ where $\sigma^0 = \mathbf{1}$ and $\sigma^1 = \sigma^+ = |1\rangle\langle 0|$, with first tensor acting on position i of the output chain. In Proposition 8 and Corollary 2 we show that asymptotically with n, by applying the creation operators to the reference (vacuum) state $|0\rangle^{\otimes n}$, we obtain Fock-type states with different excitation pattern numbers. The creation and annihilation operators satisfy the Bosonic commutation relations with each excitation pattern being an independent mode. For large n, a Fock state is a superposition of basis states consisting of long sequences of 0s interspersed with the corresponding patterns appearing in any possible order. One of our key results, Theorem 19 shows that when the gap between system and absorber parameters scales as $\theta - \theta_{\text{abs}} = u/\sqrt{n}$, the quadratures of the excitation pattern modes satisfy the Central Limit Theorem and the corresponding joint state is a product of coherent states whose amplitudes are linear in the local parameter u. The total QFI of this multimode coherent state is equal to the output QFI, showing that the TIMs contain all statistical information about the dynamics. We also prove separately, that the number operators of the TIMs have asymptotic Poisson distributions, as expected for a coherent state. Together with the result of Theorem 20, this completes a circle of ideas, which played a crucial role in formulating our

estimation strategy. In a nutshell, when looking at the output from the perspective of the TIMs, one deals with a simple Gaussian estimation problem. Using the displaced-null method, we can achieve the QCRB by measuring the number operators of the TIMs and such a measurement can be implemented by simple sequential counting measurements followed by the extraction of pattern counts from the measurement trajectory ω.

The chapter is organised as follows. In section 6.2 we give a brief review of quantum estimation theory and the displaced-null measurement technique developed in [76]. In section 6.3 we introduce the notion of QMC and the estimation problem, together with the idea of quantum post-processing using a coherent absorber. In section 6.4 we define the translationally invariant modes of the output and establish their Fock space properties. In section 6.5 we show that the restriction of the output state to the TIMs is a coherent state whose amplitude is linear in the local parameter and whose QFI is equal to the output QFI (cf. Theorem 19 and Corollary 3). In section 6.6 we establish that the excitation pattern counts obtained from the sequential output measurement have asymptotically Poisson distribution (cf. Theorem 20). In section 6.7 we formulate our measurement and estimation strategy and define the "pattern counts" estimator. Finally in section 6.8 we present results of a simulation study confirming the earlier theoretical results.

6.2 QUANTUM ESTIMATION AND THE DISPLACED NULL MEASUREMENT TECHNIQUE

In this section we give a brief overview of the quantum parameter estimation theory [4, 12, 49, 92, 130, 153, 160] used in this chapter, with an emphasis on asymptotic theory and the displaced-null measurement technique [76] developed in the previous chapter. In particular, we explain why this method is asymptotically optimal, by employing the Gaussian approximation technique called local asymptotic normality [32, 59, 60, 70, 87–89, 102, 176]. Later on, this picture will guide our intuition when dealing with the Markov estimation problem. For our purposes it suffices to discuss the case of one-dimensional parameters, and we refer to [76] for the multi-dimensional setting.

Let $\rho_\theta \in S(\mathbb{C}^d)$ be a family of quantum states depending smoothly on a one-dimensional parameter θ. Consider a measurement described by a positive operator valued measure $\{M_1, \ldots, M_k\}$ and let X be the measurement outcome with probability distribution $p_\theta(X = i) = \mathrm{Tr}(\rho_\theta M_i)$. As we have seen, the quantum Cramér-Rao bound (QCRB) [15, 25, 26, 93, 95, 179] states that the variance of any unbiased estimator $\hat{\theta} = \hat{\theta}(X)$ is lower bounded as

$$\mathrm{Var}(\hat{\theta}) = \mathbb{E}_\theta(\hat{\theta} - \theta)^2 \geq F_\theta^{-1}$$

where F_θ is the quantum Fisher information (QFI) and $\mathcal{L}_\theta$ is the SLD. In general, the QCRB is not achievable when only a single copy of ρ_θ is available. However, the bound *is* attainable in the asymptotic limit of large number of samples by the following two step adaptive procedure [71]. Given n copies of ρ_θ one can use a small proportion of the samples (e.g.- $\tilde{n} = n^{1-\epsilon}$ for a small $\epsilon > 0$) to compute a preliminary (non-optimal) estimator $\tilde{\theta}_n$ of θ; reasonable estimators

will concentrate around θ such that $|\tilde{\theta}_n - \theta| = O(n^{-1/2+\epsilon})$ with high probability, which will be assumed throughout. In the second step, one measures the SLD operator $\mathcal{L}_{\tilde{\theta}_n}$ on each of the remaining copies. If $X_1, \ldots, X_{n'}$ are the outcomes of these measurements (with $n' = n - \tilde{n}$) then the estimator

$$\hat{\theta}_n := \tilde{\theta}_n + \frac{1}{F_{\tilde{\theta}_n} n'} \left(\sum_{i=1}^{n'} X_i \right) \tag{6.1}$$

is *asymptotically optimal* in the sense that

$$n \mathbb{E}_\theta (\hat{\theta}_n - \theta)^2 \to F_\theta^{-1} \tag{6.2}$$

in the limit of large n and in addition $\hat{\theta}_n$ is *asymptotically normal*, i.e., $\sqrt{n}(\hat{\theta}_n - \theta)$ converges in distribution to the normal $N(0, F_\theta^{-1})$.

However, for certain models including that considered in this chapter, measuring the SLD may not be feasible experimentally. Instead, we will use a different method called *displaced-null measurement* [76], which aims to estimate the parameter of a *pure state model* $\rho_\theta = |\psi_\theta\rangle\langle\psi_\theta|$ by measuring each copy in a basis that contains the vector $|\psi_{\tilde{\theta}}\rangle$ with $\tilde{\theta}$ close to the true parameter θ. The Fisher information of such a measurement is known to converge to the QFI as $\tilde{\theta}$ approaches θ [115, 136, 137]. This suggests that the QCRB can be achieved asymptotically by using a two-step strategy similar to the SLD case: one first obtains a preliminary estimator $\tilde{\theta}_n$ and then measures each copy in a basis containing the vector $|\psi_{\tilde{\theta}_n}\rangle$. However, it turns out that this "null measurement" strategy fails due to the fact that for small deviations from $\tilde{\theta}_n$, the outcome probabilities depend on $(\theta - \tilde{\theta}_n)^2$ and one cannot distinguish between left and right deviations from $\tilde{\theta}_n$, cf. [76] for the precise mathematical statement. This non-identifiability issue can be sidestepped by deliberately changing the reference parameter from $\tilde{\theta}_n$ to $\theta_0 := \tilde{\theta}_n - \delta_n$ where $\delta_n = n^{-1/2+3\epsilon}$, so that $\theta = \theta_0 + (u + \tau_n)/\sqrt{n}$ with $\tau_n = n^{3\epsilon}$. The choice of τ_n is not unique but we refer to [76] for the general requirements. Since $|\theta - \tilde{\theta}_n| = O(n^{-1/2+\epsilon}) \ll \delta_n$, it means that θ lies on the right side of θ_0 and can be unambiguously identified from the outcomes of a measurement in a basis $\{|e_0\rangle, \ldots, |e_{d-1}\rangle\}$ such that $|e_0\rangle \equiv |\psi_{\theta_0}\rangle$. In addition, as θ_0 approaches θ in the limit of large n, the displaced-null measurement exhibits the optimality properties of the "null measurements" without sharing their non-identifiability issues.

Let $X_1, \ldots, X_{n'} \in \{0, 1, \ldots, d-1\}$ be the independent outcomes of basis $\{|e_0\rangle, \ldots, |e_{d-1}\rangle\}$ measurements performed on the remaining $n' = n - \tilde{n}$ systems, and let $N_{j,n}$ denote the counts of the outcome $j = 0, \ldots d-1$. The displaced-null estimator based on the two-stage measurement strategy is defined as follows

$$\hat{\theta}_n := \tilde{\theta}_n + \hat{u}_n / \sqrt{n}$$

with local parameter estimator

$$\hat{u}_n = \frac{2}{\tau_n f} \sum_{j=1}^{d-1} N_{j,n} - \frac{\tau_n}{2}$$

where $f = 4\|\dot{\psi}_{\tilde{\theta}_n}\|^2$ is the QFI at $\tilde{\theta}_n$. The estimator $\hat{\theta}_n$ is asymptotically optimal in the sense of equation (6.1) and asymptotically normal.

6.3 QUANTUM MARKOV CHAINS AND POST-PROCESSING USING COHERENT ABSORBERS

We start this section by reviewing the problem of estimating dynamical parameters of quantum Markov chains (QMCs). We then introduce the notion of quantum coherent absorber, which will play a key role in designing an optimal sequential measurement strategy.

A quantum Markov chain consists of a system interacting successively with a chain of independent "noise units" (the input) modelling the environment. In this chapter the system's space is taken to be $\mathbb{C}^d$ while the "noise units" are two dimensional systems prepared in the state $|0\rangle$ where $\{|0\rangle, |1\rangle\}$ is the standard basis in $\mathbb{C}^2$. We expect that the theory developed here works for general finite dimensional inputs, but we restrict here to this minimal setup which can be used to represent a discretised version of a continuous-time Markovian model with a single Bosonic field [10].

At each time step the system interacts with the input unit via a unitary U on $\mathbb{C}^d \otimes \mathbb{C}^2$. If the system is initially prepared in a state $|\varphi\rangle$, the joint state of system and noise units (output) after n times steps is

$$\begin{aligned} |\Psi_n\rangle &= U_n|\varphi \otimes 0^{\otimes n}\rangle \\ &= U^{(n)} \cdot \cdots \cdot U^{(2)} \cdot U^{(1)}|\varphi \otimes 0^{\otimes n}\rangle \in \mathbb{C}^d \otimes \left(\mathbb{C}^2\right)^{\otimes n} \end{aligned} \tag{6.3}$$

where $U^{(i)}$ is the unitary acting on the system and the i-th noise unit. By expanding the state (6.3) with respect to the standard product basis in the output we have

$$|\Psi_n\rangle = \sum_{i_1,\ldots,i_n \in \{0,1\}} K_{i_n} \ldots K_{i_1}|\varphi\rangle \otimes |i_1\rangle \otimes \cdots \otimes |i_n\rangle \tag{6.4}$$

where $K_i = \langle i|U|0\rangle$ are Kraus operators acting on $\mathbb{C}^d$.

From equation (6.3) it follows that the reduced system state of the system at time n is given by

$$\rho_n^{\text{sys}} := \text{Tr}_{\text{out}}(|\Psi_n\rangle\langle\Psi_n|) = T_*^n(\rho_{\text{in}}^{\text{sys}}), \quad \rho_{\text{in}}^{\text{sys}} = |\varphi\rangle\langle\varphi|,$$

where the partial trace is taken over the output noise units, and $T_* : S(\mathbb{C}^d) \to S(\mathbb{C}^d)$ is the Markov transition operator (Schrödinger picture)

$$T_* : \rho \mapsto \sum_{i \in \{0,1\}} K_i \rho K_i^*$$

whose dual (Heisenberg picture) will be denoted by T.

On the other hand, the reduced state of the output is

$$\begin{aligned} \rho_n^{\text{out}} &:= \text{Tr}_{\text{sys}}(|\Psi_n\rangle\langle\Psi_n|) \\ &= \sum_{\mathbf{i},\mathbf{j} \in \{0,1\}^n} \langle\varphi|K_{\mathbf{j}}^* K_{\mathbf{i}}|\varphi\rangle \cdot |\mathbf{i}\rangle\langle\mathbf{j}| \end{aligned} \tag{6.5}$$

where $K_{\mathbf{i}} := K_{i_n} \ldots K_{i_1}$ for $\mathbf{i} = (i_1, \ldots i_n)$.

Hypothesis 1. *Throughout the chapter we will assume that the dynamics is* primitive *in the sense that T_* has a unique stationary state $\rho_{ss} > 0$ so that $T_*(\rho_{ss}) = \rho_{ss}$ and it is* aperiodic, *i.e., the only eigenvalue of T_* with unit absolute value is* 1.

Estimation of dynamical parameters We investigate the following quantum estimation problem: assuming that the dynamics depends smoothly on an unknown parameter $\theta \in \mathbb{R}$, we would like to estimate θ by performing measurements on the output state ρ_n^{out} generated after a number n of interaction steps. In particular, we are interested in designing measurement strategies which achieve the highest possible precision, at least in the limit of large times.

Let $\theta \mapsto U_\theta$ be a smooth map describing how the dynamics depends on an unknown parameter θ, which is assumed to belong to an open bounded interval Θ of $\mathbb{R}$. We use similar notations $|\Psi_{\theta,n}\rangle, K_{\theta,i}, T_\theta$ to denote the dependence on θ of the system-output state, Kraus operators, transition operator, etc. Two sequences of quantum statistical models indexed by time are of interest here: the system-output state $\mathcal{SO}_n := \{|\Psi_{\theta,n}\rangle : \theta \in \Theta\}$ defined in equation (6.3) and the output state $\mathcal{O}_n := \{\rho_{\theta,n}^{\text{out}} : \theta \in \Theta\}$ defined in equation (6.5). While the former is more informative than the latter and easier to analyse, we are particularly interested in estimation strategies which involve only measurements on the output, hence the importance of the model $\mathcal{O}_n$. The following Theorem [83] shows that for primitive dynamics the QFI of both models scale linearly with n with the *same* rate, so having access to the system does not change the asymptotic theory. To simplify the expression of the QFI rate (6.7) we assume the following "gauge condition"

$$\sum_j \text{Tr}(\rho_\theta^{\text{ss}} \dot{K}_{\theta,j}^* K_{\theta,j}) = 0. \tag{6.6}$$

The condition (6.6) means that $\sum_j \dot{K}_{\theta,j}^* K_{\theta,j}$ belongs to the subspace $\{X : \text{Tr}(\rho_\theta^{\text{ss}} X) = 0\} \subseteq \mathcal{B}(\mathbb{C}^d)$ on which the resolvent $\mathcal{R}_\theta := (\text{Id} - T_\theta)^{-1}$ is well defined as the Moore-Penrose inverse. The condition can be satisfied by choosing the complex phase of the Kraus operators appropriately, or equivalently the phase of the standard basis in the noise unit space $\mathbb{C}^2$.

Theorem 18. *Consider a primitive discrete time Markov chain whose unitary U_θ depends smoothly on $\theta \in \Theta \subset \mathbb{R}$, and assume that condition (6.6) holds true. The QFI $F_n^{s+o}(\theta)$ of the system and output state $|\Psi_{\theta,n}\rangle$ and the QFI $F_n^{out}(\theta)$ of the output state $\rho_{\theta,n}^{out}$ scale linearly with n with the same rate:*

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{n} F_\theta^{s+o}(n) = \lim_{n\to\infty} \frac{1}{n} F_\theta^{out}(n) &= f_\theta \\ &= 4 \sum_{i=1}^{k} Tr\left[\rho_\theta^{ss} \dot{K}_{\theta,i}^* \dot{K}_{\theta,i}\right] \\ &\quad + 8 \sum_{i=1}^{k} Tr\left[Im(K_{\theta,i} \rho_\theta^{ss} \dot{K}_{\theta,i}^*) \cdot \mathcal{R}_\theta(Im \sum_j \dot{K}_{\theta,j}^* K_{\theta,j}) \right] \end{aligned} \tag{6.7}$$

where $\mathcal{R}_\theta$ is the Moore-Penrose inverse of $Id - T_\theta$.

In the following, we will always assume that $f_\theta > 0$ for every $\theta \in \Theta$. In general, the classical Fisher information associated to simple repeated measurements (measuring the same observable on each output unit) does not achieve the QFI rate f_θ. However, the class of available measurements can be enlarged by unitarily "post-processing" the output before performing a standard measurement, so that effectively one measures the original output in a rotated basis. While this shifts

the difficulty from measurement to "quantum computation", it turns out that the post-processing can be implemented with minimal computational cost by employing the concept of a coherent absorber introduced in [157]. Indeed [78] demonstrated that the QFI rate *is* achievable by combining post-processing by a coherent absorber with a simple adaptive sequential measurement scheme. Furthermore, [177], argued that one can also achieve the QFI by performing simple counting measurements in the output, without the need for adaptive measurements. Our goal is to revisit this scheme and to provide a new, computationally and statistically effective estimation strategy.

6.3.1 *Quantum postprocessing with a coherent absorber*

The working of the coherent absorber (CA) is illustrated in Figure 6.1 b). Consider a QMC with a fixed and known unitary U. After interacting with the system, each output noise unit interacts with a separate d dimensional system (the coherent absorber), via a unitary V. The system and absorber can now be regarded as a single doubled-up system which interacts with the input via the unitary $W := VU$ on $\mathbb{C}^d \otimes \mathbb{C}^d \otimes \mathbb{C}^2$, where U acts on first and third tensors and V on second and third. The defining feature is that the system plus absorber have a pure stationary state. One can arrange this by requiring

$$VU : |\chi^{\rm ss}\rangle \otimes |0\rangle \mapsto |\chi^{\rm ss}\rangle \otimes |0\rangle$$

where $|\chi^{\rm ss}\rangle \in \mathbb{C}^d \otimes \mathbb{C}^d$ is a purification of the system stationary state, i.e., $\rho^{\rm ss} = {\rm Tr}_{\rm abs}(|\chi^{\rm ss}\rangle\langle\chi^{\rm ss}|)$. This implies that in the stationary regime the output is decoupled from system and absorber, and is in the "vacuum" state $|0\rangle^{\otimes n}$. We briefly recall a few expressions related to the construction of V that will be useful later on, and refer to Lemma 4.1 in [78] for more details; for clarity, in the following we will use the labels S, A, N to indicate system, absorber and noise unit. Let us consider a spectral representation of $\rho^{\rm ss}$ and the corresponding purification:

$$\rho^{\rm ss} = \sum_{i=1}^{d} \lambda_i |i_S\rangle \langle i_S| , \quad |\chi^{\rm ss}\rangle = \sum_{i=1}^{d} \sqrt{\lambda_i}\, |i_S\rangle \otimes |i_A\rangle .$$

For simplicity we assume that the eigenvalues λ_i are strictly positive and are ordered in decreasing order; one can check that the following vectors are orthonormal:

$$|v_i\rangle = \sum_{k=0}^{1} \sum_{j=1}^{d} \sqrt{\frac{\lambda_j}{\lambda_i}}\, \langle i_S| K_k |j_S\rangle\, |j_A\rangle \otimes |k_N\rangle , \quad i = 1, \ldots, d.$$

For any choice of $v_{d+1}, \ldots, v_{2d}$ such that $\{v_1, \ldots, v_{2d}\}$ is an orthonormal basis for $\mathbb{C}^d \otimes \mathbb{C}^2$ (the Hilbert space corresponding to the absorber and the ancilla), a suitable choice for V is given by

$$V = \mathbf{1}_S \otimes \left(\sum_{i=1}^{d} |i_A \otimes 0_N\rangle \otimes \langle v_i| + \sum_{i=d+1}^{2d} |i_A \otimes 1_N\rangle \otimes \langle v_i| \right) .$$

Note that V is not uniquely defined: there is freedom in the spectral resolution of $\rho^{\rm ss}$ if there are degenerate eigenvalues and in picking $v_{d+1}, \ldots, v_{2d}$. The Kraus

operators corresponding to the reduced dynamics $W = VU$ of the system and the absorber together are given by the following expression:

$$\begin{aligned}\tilde{K}_k : &= \langle k|W|0\rangle = \sum_{l=0}^{1} \langle k_N| V |l_N\rangle \langle l_N| U |0_N\rangle \\ &= \sum_{l=0}^{1} K_l \otimes V_{kl}\end{aligned}$$

where V_{kl} and K_l are the Kraus operators of V and U considered without amplification. The following Lemma prescribes the structure of the blocks V_{kl}. We first define the "recovery" channel [14, 133, 134] with Kraus operators $K'_i = \sqrt{\rho} K_i^* \sqrt{\rho^{-1}}$ and note that they satisfy the normalisation condition and the recovery channel has ρ as invariant state.

Lemma 7. *The absorber operators V_{kl} are of the following form. The blocks V_{0l} are determined as $V_{0l} = K_l'^T$ where the transpose is taken with respect to the eigenbasis of ρ. Assuming that $\mathbf{1} - |V_{00}|^2$ and $\mathbf{1} - |V_{01}|^2$ are strictly positive, then the V_{1l} blocks are determined up to an overall arbitrary unitary u*

$$V_{10} = u|V_{10}|, V_{11} = uw|V_{11}|$$

where $|V_{10}| = \sqrt{\mathbf{1} - |V_{00}|^2}$, $|V_{11}| = \sqrt{\mathbf{1} - |V_{01}|^2}$ are fixed, as well as the unitary $w = -|V_{10}|^{-1} V_{00}^* V_{01} |V|_{11}^{-1}$.

Proof. From the definition of $|v_i\rangle$ we have

$$V_{0k} = \sum_{i,j=1}^{d} \sqrt{\frac{\lambda_j}{\lambda_i}} \langle j_S| K_k^* |i_S\rangle |i_A\rangle \langle j_A| = K_k'^T.$$

which proves the first statement. From the fact that V is unitary we obtain

$$\begin{aligned} V_{00}^* V_{00} + V_{10}^* V_{10} &= \mathbf{1} \\ V_{01}^* V_{01} + V_{11}^* V_{11} &= \mathbf{1} \end{aligned}$$

from which we get

$$|V_{10}| = \sqrt{\mathbf{1} - |V_{00}|^2}, \quad |V_{11}| = \sqrt{\mathbf{1} - |V_{01}|^2}$$

which means that the absolute values of V_{10}, V_{11} are fixed. Let $V_{10} = U_0|V_{10}|$ and $V_{11} = U_1|V_{11}|$ be their polar decompositions. Then from

$$V_{00}^* V_{01} + V_{10}^* V_{11} = 0$$

we get $V_{10}^* V_{11} = -V_{00}^* V_{01}$ and

$$U_0^* U_1 = |V_{10}|^{-1} V_{10}^* V_{11} |V_{11}|^{-1} = -|V_{10}|^{-1} V_{00}^* V_{01} |V_{11}|^{-1}$$

which is a fixed unitary w. This proves the claim. □

Later on we will require that the system and absorber transition operator $\tilde{T}(\cdot) = \sum_k \tilde{K}_k^* \cdot \tilde{K}_k$ is primitive, in addition to the system's transition operator T satisfying the same property. At the moment we are not able to establish what is the connection between the two properties. However we performed extensive numerical simulations with randomly chosen QMC dynamics and corresponding absorbers which indicate that "generically" with respect to the original dynamics, for every primitive T there exists a corresponding absorber such that $\tilde{T}$ is primitive. In fact, a stronger statement seems to hold, which is that the spectral gap of $\tilde{T}$ is smaller or equal to that of T and one can always choose the absorber such that the two are equal. For more details on the absorber theory we refer to the recent paper [162].

Returning to the parameter estimation setting where $U = U_\theta$ depends on the unknown parameter θ, we note that one cannot implement a CA which precisely matches the system dynamics. Instead, one can implement the absorber for an approximate value θ_0 of θ and try to estimate the offset $\theta - \theta_0$ by measuring the output. This setting is closely related to that of displaced null measurements discussed in section 6.2. Indeed, the joint state of system, absorber and output is pure for all θ and at $\theta = \theta_0$ it is of the product form $|\chi_{\theta_0}^{\mathrm{ss}}\rangle \otimes |0\rangle^{\otimes n}$, assuming that system and absorber are initially in the stationary state. Therefore, repeated standard basis measurements on the output units constitute a null measurement (in conjunction with a final appropriate measurement on system and absorber).

The exact procedure for determining θ_0 will be described in section 6.7 and follows the important displacement prescription outlined in section 6.2. For the moment it suffices to say that θ_0 will be informed by the outcome of a preliminary estimation stage involving simple (non-optimal) measurements on the output (without post-processing), and it will converge to θ in the limit of large n. While for $\theta_0 = \theta$ the output state is the "vacuum", for $\theta \neq \theta_0$ the output could be seen as carrying a certain amount of "excitations" which increases with the parameter mismatch $|\theta - \theta_0|$.

In section 6.4 we show how these "excitations" can be given a precise meaning by fashioning the output Hilbert space into a Fock space carrying modes labelled by certain "excitation patterns". In section 6.5 we show that from this perspective, the output state converges to a joint coherent state of the excitation pattern modes whose displacement depends linearly on $\theta - \theta_0$. This will allow us to devise a simple "pattern counting" algorithm for estimating θ in section 6.7.

6.4 TRANSLATIONALLY INVARIANT MODES IN THE OUTPUT

In this section we introduce the concept of translationally invariant modes (TIMs) of a spin chain. We show that in the limit of large chain size, certain translationally invariant states acquire the characteristic Fock space structure, and that the corresponding creation and annihilation operators satisfy the bosonic commutation relations. This construction will then be used in analysing the stationary Markov output state in section 6.5.

Let $\left(\mathbb{C}^2\right)^{\otimes n}$ be a spin chain of length n and let $|\Omega_n\rangle := |0\rangle^{\otimes n}$ be the reference "vacuum" state. For every pair of integers (k, l) with $1 \leq l \leq k$ we define an *excitation pattern* of length k and number of excitations l to be an ordered sequence

$\alpha := (\alpha_1, \dots \alpha_k) \in \{0,1\}^k$ such that $\alpha_1 = \alpha_k = 1$ and $\sum_{i=1}^k \alpha_i = l$. For instance, for $k = 1, 2$ the only patterns are 1 and respectively 11 while for $k = 3$ the possible patterns are 111 and 101.

For each pattern α of length k we define 'creation and annihilation' operators

$$A_\alpha^*(n) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n-k+1} \sigma_i^\alpha, \quad A_\alpha(n) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n-k+1} \sigma_i^{\alpha *}$$

where $\sigma_i^\alpha = \prod_{j=0}^{k-1} \sigma_{i+j}^{\alpha_{j+1}}$, with $\sigma^0 := \mathbf{1}$, $\sigma^1 := \sigma^+ := |1\rangle\langle 0|$ and the index i denotes the position in the chain.

In particular for $\alpha = 1$ we have

$$A_1^*(n) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sigma_i^+, \quad A_1(n) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sigma_i^-.$$

We further define the "canonical coordinates" and "number operator" of the "mode" α as

$$Q_\alpha(n) = \frac{A_\alpha(n) + A_\alpha^*(n)}{\sqrt{2}}, \tag{6.8}$$

$$P_\alpha(n) = \frac{A_\alpha(n) - A_\alpha^*(n)}{\sqrt{2}i}, \tag{6.9}$$

$$N_\alpha(n) = A_\alpha^*(n) A_\alpha(n). \tag{6.10}$$

We now introduce "Fock states" obtained by applying creation operators to the vacuum.

Let $\mathcal{P}$ denote the ordered set of all patterns, where the order is the usual one when regarding patterns as integer numbers in binary representation. Let $\mathbf{n} : \mathcal{P} \to \mathbb{N}$ be pattern counts $\mathbf{n} = (n_\alpha)_{\alpha \in \mathcal{P}}$ such that all but a finite number of counts are zero, and let $\mathbf{n}! := \prod_{\alpha \in \mathcal{P}} n_\alpha!$ and $|\mathbf{n}| := \sum_\alpha n_\alpha$ the total number of patterns.

We define the approximate Fock state associated to the set of counts $\mathbf{n}$ as

$$|\mathbf{n}; n\rangle := \frac{1}{\sqrt{\mathbf{n}!}} \prod_{\alpha \in \mathcal{P}} A_\alpha^*(n)^{n_\alpha} |\Omega_n\rangle, \quad n \geq 1, \tag{6.11}$$

where the product is ordered according to the order on $\mathcal{P}$. For example, the state $|n_{101} = 1; n = 5\rangle$ is

$$|n_{101} = 1; n = 5\rangle = \frac{1}{\sqrt{5}} \left[|10100\rangle + |01010\rangle + |00101\rangle \right].$$

For any fixed n these vectors are not normalised or orthogonal to each other, and indeed they are not linearly independent since the Hilbert space is finite dimensional; however, Proposition 8 shows that the expected Fock structure emerges in the limit of large n. Let us first illustrate this with a simple example. The state containing 2 patterns $\alpha = 1$ is given by

$$\begin{aligned} |n_1 = 2; n\rangle &:= \frac{1}{\sqrt{2}} \left(A_1^*(n)\right)^2 |\Omega_n\rangle \\ &= \frac{1}{\sqrt{2}n} \sum_{i \neq j = 1}^{n} |0 \dots 010 \dots 010 \dots 0\rangle \end{aligned}$$

where $i \neq j$ indicate the positions of the excitations, while the state containing a single 11 pattern is

$$\begin{aligned} |n_{11} = 1; n\rangle &:= A^*_{11}(n)|\Omega_n\rangle \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n-1} |0 \dots 0110 \dots 0\rangle. \end{aligned}$$

Now it is easy to check that as $n \to \infty$

$$\begin{aligned} \langle n_1 = 2; n | n_1 = 2; n\rangle &= \frac{1}{2n^2} \frac{4n(n-1)}{2} \to 1, \\ \langle n_{11} = 1; n | n_{11} = 1; n\rangle &= \frac{n-1}{n} \to 1, \\ \langle n_1 = 2; n | n_{11} = 1; n\rangle &= \frac{2(n-1)}{2\sqrt{n}n} = O\left(\frac{1}{\sqrt{n}}\right). \end{aligned}$$

This is generalised in the following Proposition which establishes the familiar structure of the bosonic Fock space in the limit of large n.

Proposition 8. *Let $|\boldsymbol{n}; n\rangle$ be the "Fock states" defined in equation (6.11). In the limit of large n the "Fock states" become normalised and are orthogonal to each other*

$$\lim_{n\to\infty} \langle \boldsymbol{n}; n | \boldsymbol{m}; n\rangle = \delta_{\boldsymbol{n},\boldsymbol{m}}.$$

Moreover, the order of the creation operators in (6.11) becomes irrelevant in the limit of large n.

The proof of Proposition 8 can be found in section 6.10. From Proposition 8 we obtain the following Corollary which shows that the action of the creation operators on the "Fock states" converges to that of a bosonic creation operator in the limit of large n.

Corollary 2. *Let β be a pattern and let $\boldsymbol{\delta}^{(\beta)}$ be the counts set with $\boldsymbol{\delta}^{(\beta)}_\alpha = \delta_{\alpha,\beta}$. Let $|\boldsymbol{n}; n\rangle$ be a "Fock state" as defined in equation (6.11). In the limit of large n the action of creation and annihilation operators $A^*_\beta(n)$ and $A_\beta(n)$ satisfy*

$$A^*_\beta(n)|\boldsymbol{n}; n\rangle = \sqrt{n_\beta + 1}|\boldsymbol{n} + \boldsymbol{\delta}^\beta; n\rangle + o(1) \tag{6.12}$$

$$A_\beta(n)|\boldsymbol{n}; n\rangle = \sqrt{n_\beta}|\boldsymbol{n} - \boldsymbol{\delta}^\beta; n\rangle + o(1) \tag{6.13}$$

The proof of Corollary 2 can be found in section 6.10.

6.5 LIMIT DISTRIBUTION OF QUADRATURES AND NUMBER OPERATORS

In this section we analyse the structure of the output state obtained by post-processing the output of a QMC with a coherent absorber, as described in section 6.3.1. Motivated by the fact that the statistical uncertainty scales as $1/\sqrt{n}$ we choose the absorber parameter θ_0 to be fixed and known, and write the system parameter as $\theta = \theta_0 + u/\sqrt{n}$, where u is an unknown local parameter.

Since the output state becomes stationary for long times, it is natural to focus on the state of the translationally invariant modes introduced in section 6.4. In

Theorem 19 we show that asymptotically with n, the restricted state of these modes is a joint coherent state whose amplitude depends linearly on u, i.e., a Gaussian shift model. Moreover, Corollary 3 shows that the QFI of this model is equal to the QFI rate f_{θ_0} of the output state characterised in Theorem 18. This means that the TIMs capture the entire QFI of the output. Together with Theorem 20 of section 6.6, these results will be the theoretical underpinning the estimator proposed in section 6.7.

We consider the system and absorber together as an open system with space $\mathbb{C}^D = \mathbb{C}^d \otimes \mathbb{C}^d$ with $D = d^2$, interacting with the noise units. The corresponding unitary is $W_\theta = V_{\theta_0} U_\theta$ where the absorber parameter is a fixed value θ_0 (which later will be determined based on a preliminary estimation procedure) and θ is the unknown parameter to be estimated. We distinguish between the system Kraus operators $K_{\theta,i} = \langle i|U_\theta|0\rangle \in \mathcal{B}(\mathbb{C}^d)$ and the system and absorber Kraus operators $\tilde{K}_{\theta,i} = \langle i|W_\theta|0\rangle \in \mathcal{B}(\mathbb{C}^D)$, and similarly between the system transition operator T_θ and the system and absorber transition operator $\tilde{T}_\theta(X) = \sum_{i=0}^{1} \tilde{K}^*_{\theta,i} X \tilde{K}_{\theta,i}$.

We will be interested in the probabilistic and statistical properties of the output state $\tilde{\rho}^{\text{out}}_\theta$ of the system and absorber dynamics, for parameters θ in the neighbourhood of a given θ_0. For clarity we state the precise mathematical properties we assume throughout.

Hypothesis 2. *The following properties of the system-absorber QMC are assumed to be true:*

1. *$\tilde{T}_\theta$ has a unique faithful invariant state $\tilde{\rho}^{ss}_\theta$ and is aperiodic for θ in a neighborhood of θ_0;*
2. *The Kraus operators $\tilde{K}_{\theta,i}$ and the stationary state $\tilde{\rho}^{ss}_\theta$ are analytic functions of θ around θ_0;*
3. *At θ_0 the stationary state is pure $\tilde{\rho}^{ss}_{\theta_0} = |\chi^{ss}_{\theta_0}\rangle\langle\chi^{ss}_{\theta_0}|$ and $\tilde{K}_{\theta_0,i}|\chi^{ss}_{\theta_0}\rangle = (1-i)|\chi^{ss}_{\theta_0}\rangle$ for $i = 0, 1$.*

To formulate the result we use the local parametrisation $\theta = \theta_0 + u/\sqrt{n}$ where u is to be seen as a local parameter to be estimated from the output of length n. To simplify the notation, we denote the derivatives at θ_0 as

$$\dot{K}_i := \left.\frac{dK_{\theta,i}}{d\theta}\right|_{\theta_0}, \quad \dot{\tilde{K}}_i := \left.\frac{d\tilde{K}_{\theta,i}}{d\theta}\right|_{\theta_0},$$

and drop the subscript θ_0 in $K_{\theta_0,i} =: K_i, \tilde{K}_{\theta_0,i} =: \tilde{K}_i$, etc.; we also use the local parameter instead of θ, e.g., $K_{\theta_0+u/\sqrt{n},i} =: K_{u,i}$.

Note that properties 1. and 2. in Hyothesis 2 imply that $\dot{\rho}^{\text{ss}} = \tilde{\mathcal{R}}_* \dot{\tilde{T}}_*(\tilde{\rho}^{\text{ss}})$ where $\tilde{\mathcal{R}}$ is the Moore-Penrose inverse of $\text{Id} - \tilde{T}$. In addition, $\tilde{K}_0 |\chi^{\text{ss}}\rangle = \tilde{K}_0^* |\chi^{\text{ss}}\rangle = |\chi^{\text{ss}}\rangle$. Indeed by construction $\tilde{K}_0|\chi^{\text{ss}}\rangle = |\chi^{\text{ss}}\rangle$ and $\tilde{K}_1|\chi^{\text{ss}}\rangle = 0$, and by applying $\tilde{K}_0^*\tilde{K}_0 + \tilde{K}_1^*\tilde{K}_1 = \mathbf{1}$ to $|\chi^{\text{ss}}\rangle$ we get $\tilde{K}_0^*|\chi^{\text{ss}}\rangle = |\chi^{\text{ss}}\rangle$.

Since for large n the dynamics reaches stationarity, we consider the output state corresponding to the system starting in the stationary state $\tilde{\rho}^{\text{ss}}_u$

$$\tilde{\rho}^{\text{out}}_{u,n} = \sum_{\mathbf{i},\mathbf{j}\in\{0,1\}^n} \text{Tr}\left[\tilde{\rho}^{\text{ss}}_u \tilde{K}^*_{u,\mathbf{j}} \tilde{K}_{u,\mathbf{i}}\right] |\mathbf{i}\rangle\langle\mathbf{j}| \tag{6.14}$$

To formulate our results below we need to introduce several superoperators acting on the system and absorber space. For $x \in \mathcal{B}(\mathbb{C}^D)$, we define

$$\mathcal{A}_i(x) = \begin{cases} \tilde{T}(x) & i = 0 \\ \tilde{K}_1^* x \tilde{K}_0 & i = 1 \\ \tilde{K}_0^* x \tilde{K}_1 & i = -1, \end{cases} \tag{6.15}$$

and

$$\dot{\mathcal{A}}_1(x) = \dot{\tilde{K}}_1^* x \tilde{K}_0 + \tilde{K}_1^* x \dot{\tilde{K}}_0$$

Furthermore, for every pattern α of length l, we denote

$$\mathcal{A}^\alpha = \mathcal{A}_{\alpha_1} \cdots \mathcal{A}_{\alpha_l} \text{ and } \tilde{\mathcal{A}}^\alpha = \dot{\mathcal{A}}_1 \mathcal{A}_{\alpha_2} \cdots \mathcal{A}_{\alpha_l}.$$

With a slight abuse of notation we will denote the expectation with respect to a density matrix ρ as $\rho(X) = \mathrm{Tr}(\rho X)$.

Theorem 19. *Let θ_0 be a fixed parameter and assume that the dynamics satisfies the assumptions in Hypothesis (2). Let $\theta = \theta_0 + u/\sqrt{n}$ be the system parameter with fixed local parameter u. Let α be a fixed pattern and let $z = \beta + i\gamma \in \mathbb{C}$ with $|z| = 1$. Then the following convergence in distribution hold in the limit of large n with respect to the output state $\tilde{\rho}^{out}_{u,n}$ as defined in equation (6.14).*

i) ***The quadratures (6.8) and (6.9) of the TIM mode α satisfy the joint Central Limit Theorem***

$$\beta Q_\alpha(n) + \gamma P_\alpha(n) \xrightarrow{\mathcal{L}} N(u\mu_{\alpha,z}, 1/2),$$

where $N(\mu, V)$ denotes the normal distribution with mean μ and variance V and $\mu_{\alpha,z} = \sqrt{2}\Re(\bar{z}\mu_\alpha)$ with

$$\mu_\alpha = \tilde{\rho}^{ss}((\dot{\tilde{T}}\tilde{\mathcal{R}}\mathcal{A}^\alpha + \tilde{\mathcal{A}}^\alpha)(\mathbf{1})). \tag{6.16}$$

ii) ***The number operator (6.10) of the TIM mode α satisfies the Poisson limit:***

$$N_\alpha(n) \xrightarrow{\mathcal{L}} Poisson(u^2\lambda_\alpha),$$

where $\lambda_\alpha := |\mu_\alpha|^2$.

The proof of Theorem 19 can be found in section 6.11.

We now provide a simpler expression for the parameters μ_α in terms of Kraus operators and their first derivatives. As a by-product we show that the sum of all (limiting) QFIs of the Gaussian modes $(Q_\alpha(n), P_\alpha(n))$ is the QFI rate of the output state. This means that the TIMs capture all the QFI of the output state.

Lemma 8. *Let α be any excitation pattern and let μ_α be the constant defined in (6.16). Then μ_α can also be expressed as*

$$\begin{aligned} \mu_\alpha &= \langle \tilde{K}_{\alpha_{|\alpha|}} \cdots \tilde{K}_{\alpha_1} (\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 \chi^{ss} | \chi^{ss} \rangle \\ &+ \langle \tilde{K}_{\alpha_{|\alpha|}} \cdots \tilde{K}_{\alpha_2} \dot{\tilde{K}}_1 \chi^{ss} | \chi^{ss} \rangle. \end{aligned} \tag{6.17}$$

Moreover with $\lambda_\alpha = |\mu_\alpha|^2$, one has

$$\sum_\alpha \lambda_\alpha = \| (\tilde{K}_1 (\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 + \dot{\tilde{K}}_1) \chi^{ss} \|^2. \tag{6.18}$$

The proof of Lemma 8 can be found in section 6.11.

Corollary 3. *Asymptotically with n, the total QFI of the TIMs is equal to the QFI rate of the output state (6.7), that is*

$$4\sum_{\alpha}\lambda_{\alpha} = 4\|(\tilde{K}_1(\mathbf{1}-\tilde{K}_0)^{-1}\dot{\tilde{K}}_0 + \dot{\tilde{K}}_1)\chi^{ss}\|^2 = f_{\theta_0}.$$

The proof of Corollary 3 can be found in section 6.12.

The upshot of this section is that (asymptotically in n) the statistical information of the output state is concentrated in the TIMs and the state's restriction to the TIM Bosonic algebra is a coherent state. Formally, in order to optimally estimate the parameter, one would only need to measure the appropriate quadrature of the Gaussian shift model, as explained in section 6.2. However, it is not obvious how to perform such a measurement, and the theoretical insight does not seem to help on the practical side. Surprisingly, it turns out that the standard sequential counting measurement is an effective joint measurement of all the TIMs' number operators! This will be the main result of the next section, which in conjunction with the displaced null strategy provides the ingredients of a counting-based estimation strategy.

6.6 LIMIT THEOREM FOR COUNTING TRAJECTORIES

In this section we continue to investigate the probabilistic properties of the output state and consider the distribution of the stochastic process obtained measuring the output units sequentially in the canonical basis $\{|0\rangle, |1\rangle\}$. We consider the system and absorber dynamics with fixed absorber parameter θ_0 and system parameter $\theta = \theta_0 + u/\sqrt{n}$ for a fixed local parameter u. The output state is given by equation (6.14). The probability of observing a sequence $\omega = (\omega_1, \dots, \omega_n) \in \{0,1\}^n$ as the outcome of the first n measurements is given by

$$\nu_{u,n}(\omega) := \rho_u^{\mathrm{ss}}\left(\mathcal{B}_{u,\omega_1}\cdots\mathcal{B}_{u,\omega_n}(\mathbf{1})\right) \tag{6.19}$$

where

$$\mathcal{B}_{u,j}(x) = \begin{cases} \tilde{K}_{u,1}^* x \tilde{K}_{u,1} & j=1 \\ \tilde{K}_{u,0}^* x \tilde{K}_{u,0} & j=0 \end{cases}.$$

In order to state the main result of this section, we need to introduce a collection of events: first of all we define

$$B_{\mathbf{0}}(n) = \{(0,\dots,0)\} \subset \{0,1\}^n.$$

Let us fix any real number $0 < \gamma < 1$. Let $\mathbf{m} = \{m_\alpha\}_{\alpha\in\mathcal{P}}$ be a set of pattern excitation counts where all occupation numbers are zero except $(m_{\alpha^{(1)}}, \dots, m_{\alpha^{(k)}})$, and define $B_{\mathbf{m}}(n)$ as the empty set for every n strictly smaller than $\sum_{i=1}^k m_{\alpha^{(i)}}|\alpha^{(i)}| + (k-1)n^\gamma$ ($|\alpha|$ is the length of the pattern α), otherwise $B_{\mathbf{m}}(n)$ is the set of all binary sequences obtained concatenating $m_{\alpha^{(1)}}$ copies of $\alpha^{(1)}$s, up to $m_{\alpha^{(k)}}$ copies of $\alpha^{(k)}$s in any order, and inserting 0s before, between and after them in order to reach a total length of n and making sure that between two consecutive patterns there are at least n^γ 0s.

The following Theorem shows that the distribution of the pattern counts $\mathbf{m}$ converges to a product of Poisson distributions with the same intensities as those of the number operators of the TIM in Theorem 19. This means that performing a standard output measurement and extracting the pattern counts provides an effective joint measurement of the number operators of the TIMs. This finding is essential in constructing an optimal estimator in section 6.7.

Theorem 20. *For every positive constant $C > 0$ and finite collection of excitation patterns counts $\boldsymbol{m} = (m_{\alpha^{(1)}}, \ldots, m_{\alpha^{(k)}})$, the following limit holds true:*

$$\lim_{n\to+\infty} \sup_{|u|<C} \left| \nu_{u,n}\left(B_{\boldsymbol{m}}(n)\right) - e^{-\lambda_{tot}u^2} \prod_{i=1}^{k} \frac{\left(\lambda_{\alpha^{(i)}}u^2\right)^{m_{\alpha^{(i)}}}}{m_{\alpha^{(i)}}!} \right| = 0. \tag{6.20}$$

where

$$\begin{aligned} \lambda_{tot} &:= -\frac{1}{2}\langle \chi^{ss}| \left(2\dot{\mathcal{B}}_{0*}\mathcal{R}_{0*}\dot{\mathcal{B}}_{0*} + \ddot{\mathcal{B}}_{0*}\right)\left(|\chi^{ss}\rangle\langle\chi^{ss}|\right)|\chi^{ss}\rangle \\ &= -\Re(\langle \chi^{ss}, (2\dot{\tilde{K}}_0(\mathbf{1}-\tilde{K}_0)^{-1}\dot{\tilde{K}}_0 + \ddot{\tilde{K}}_0)\chi^{ss}\rangle). \end{aligned} \tag{6.21}$$

Moreover,

$$\sum_{\alpha} \lambda_\alpha = \lambda_{tot}.$$

The Proof of Theorem 20 can be found in section 6.13.

The previous result has some relevant consequences. Let us define the "pattern extraction" function which associates to each trajectory $\omega \in \{0,1\}^n$ a set of pattern counts $\{N_{\alpha,n}(\omega) : \alpha \in \mathcal{P}\} \in \mathbb{N}^{\mathcal{P}}$, which is uniquely determined by the condition that ω is a maximal union of contiguous patterns separated by sequences of 0s of length at least n^γ with a fixed $0 < \gamma < 1$; moreover, let us consider the stochastic process given by the infinite collection of independent random variables $\{N_\alpha : \alpha \in \mathcal{P}\}$ where N_α is a Poisson random variable with parameter $\lambda_\alpha u^2$.

Corollary 4. *For every $u \in \mathbb{R}$ the law of the stochastic process $\{N_{\alpha,n} : \alpha \in \mathcal{P}\}$ under the measure $\nu_{u,n}$ converges to the one of $\{N_\alpha : \alpha \in \mathcal{P}\}$. Moreover, for every $\alpha \in \mathcal{P}$, $p \geq 1$ one has*

$$\lim_{n\to+\infty} \mathbb{E}[N_{\alpha,n}^p] = \mathbb{E}[N_\alpha^p].$$

The Proof of Corollary 4 can be found in section 6.14.

However, in the following we will be interested in local parameters with growing size, i.e., $|u| \leq n^{\epsilon'}$ for some $0 < \epsilon' < 1/2$. In this case we can show the following result.

Proposition 9. *For $0 < \epsilon' < 1/6$ and for every finite collection of excitation patterns counts $\boldsymbol{m}$ the following holds true:*

$$\lim_{n\to+\infty} \sup_{|u|\leq n^{\epsilon'}} \left| \frac{\nu_{u,n}\left(B_{\mathbf{m}}(n)\right)}{e^{-\lambda_{tot}u^2}\prod_{i=1}^{k} \frac{\lambda_{\alpha^{(i)}}^{m_i} u^{2m_i}}{m_i!}} - 1 \right| = 0.$$

The Proof of Proposition 9 can be found in section 6.13 as well. Upgrading this result to a weak convergence one similar to Theorem 20 remains the subject of future research.

6.7 PATTERN COUNTING ESTIMATOR

In this section we describe our adaptive estimation scheme which exploits the asymptotic results presented in sections 6.5 and 6.6. The scheme involves four key ingredients:

i) perform a simple output measurement (no absorber) to compute a preliminary estimator;

ii) set the absorber parameter by using the displaced-null measurement technique developed in [76] and run the system-absorber dynamics for the remainder of the time;

iii) perform a sequential counting measurement in the output and extract counts for the TIMs from the outcomes trajectory;

iv) construct a simple estimator expressed in terms of total counts of patterns for different TIM modes.

The first step of the adaptive protocol is to use $\tilde{n} := n^{1-\epsilon} \ll n$ output units to produce a rough preliminary estimator $\tilde{\theta}_n$. This can be done by performing a repeated standard basis measurement on the output (without using an absorber). Typically, the estimator $\tilde{\theta}_n$ will have variance scaling with the standard rate $\tilde{n}^{-1} = n^{-1+2\epsilon}$, and $\tilde{n}^{1/2}(\tilde{\theta}_n - \theta)$ will satisfy the central limit theorem and a concentration bound ensuring that $|\tilde{\theta}_n - \theta| = O(n^{-1/2+\epsilon})$ with high probability. For instance, one can define $\tilde{\theta}_n$ by equating the empirical counting rate with the theoretical rate $n_\tau := \mathrm{Tr}(\rho_\tau^{\mathrm{ss}} K_{\tau,1}^* K_{\tau,1})$, see [37, 74].

In the second step we set the absorber at a parameter value θ_0 and run the system-absorber quantum Markov chain for the reminder of the time $n' = n - \tilde{n}$. The naive choice for θ_0 is our best guess $\tilde{\theta}_n$ about θ, based on the first stage measurement. However, with this choice, the counting measurement suffers from the non-identifiability issue described in section 6.2. This can be resolved by further displacing the absorber parameter by an amount $\delta_n := \tau_n/\sqrt{n}$ where $\tau_n = n^{3\epsilon}$ so that $\theta_{\mathrm{abs}} := \tilde{\theta}_n - \delta_n$. As usual we write $\theta = \tilde{\theta}_n + u_n/\sqrt{n}$ where u_n is a local parameter satisfying $|u_n| \leq n^\epsilon$, so that $\theta = \theta_{\mathrm{abs}} + (u_n + \tau_n)/\sqrt{n}$.

In the third step we perform *standard basis measurements* in the output of the modified dynamics which includes the absorber. For simplicity, in the discussion below we ignore the fact that we have $n' = n - \tilde{n}$ rather than n output units, which does not affect the error scaling of the estimator. Let $\omega = (\omega_1, \ldots, \omega_n)$ be the measurement outcome with distribution (6.19), where the local parameter u_n is replaced by $u_n + \tau_n$ to take into account the displacement.

We now describe the construction of the estimator from the outcomes of this last stage measurement. The estimator will be built using the "pattern extraction" function $\{N_{\alpha,n}(\omega) : \alpha \in \mathcal{P}\} \in \mathbb{N}^{\mathcal{P}}$ that we defined in the previous section; we recall that it associates to each trajectory $\omega \in \{0,1\}^n$ a set of pattern counts, which is uniquely determined by the condition that ω is a maximal union of contiguous patterns separated by sequences of 0s of length at least n^γ with a fixed $0 < \gamma < 1$, cf. Figure 6.1 d) for illustration. This means that the algorithm will not detect any pattern which contains a sequence of zeros of length larger than n^γ, since this would be seen as being made up of several identified patterns. We now introduce

the final estimator using an intuitive argument based on extrapolating the results of Theorem 20 from fixed to slowly growing local parameters $u_n + \tau_n$. This means that $N_{\alpha,n}(\omega)$ is approximately distributed as Poisson$((u_n+\tau_n)^2|\mu_\alpha|^2)$, for large n; since $\tau_n = n^{3\epsilon}$ is larger than $u_n = O(n^\epsilon)$, the intensity of the Poisson distribution diverges with n, and the distribution can be approximated further by the normal $N((u_n+\tau_n)^2|\mu_\alpha|^2, (u_n+\tau_n)^2|\mu_\alpha|^2)$ with the same mean and variance. Using

$$\begin{aligned}\frac{1}{\tau_n}(u_n+\tau_n)^2 &= 2u_n + \tau_n + o(1)\\ \frac{1}{\tau_n^2}(u_n+\tau_n)^2 &= 1 + o(1)\end{aligned}$$

we obtain that

$$Y_{\alpha,n} := \frac{1}{|\mu_\alpha|}\left(\frac{N_{\alpha,n}}{\tau_n} - \tau_n|\mu_\alpha|^2\right)$$

has approximate distribution $N(2u_n|\mu_\alpha|, 1)$. A simple computation shows that the optimal estimator of u_n based on the (approximately) normal variables $Y_{\alpha,n}$ is the linear combination

$$\hat{u}_n := Y_n := \frac{2}{f_{\theta_{\text{abs}}}}\sum_\alpha |\mu_\alpha| Y_{\alpha,n} = \frac{2}{f_{\theta_{\text{abs}}}\tau_n}\sum_\alpha N_{\alpha,n} - \frac{\tau_n}{2} \tag{6.22}$$

where $f_{\theta_{\text{abs}}} = 4\sum_\alpha |\mu_\alpha|^2$ is the quantum Fisher information rate of the output by Corollary 3. Note that Y_n depends only on the *total number of patterns* of the trajectory ω, not to be confused with the total number of 1s.

Since $Y_{\alpha,n}$ is approximately normal with distribution $N(2u_n|\mu_\alpha|, 1)$, we obtain that $\hat{u}_n$ has approximate distribution $N(u_n, f^{-1}_{\theta_{\text{abs}}})$. The final estimator of θ is

$$\hat{\theta}_n := \tilde{\theta}_n + \frac{\hat{u}_n}{\sqrt{n}}. \tag{6.23}$$

and it attains the QCRB in the sense that $\sqrt{n}(\hat{\theta}_n - \theta)$ converges in distribution to $N(0, f_\theta^{-1})$.

At the moment we only have a rigorous proof of this statement assuming a stronger version of Proposition 9, which we were not able to obtain; however, we point out that Theorem 2 in [76] establishes a similar optimality result in the case of multi-parameter estimation with independent, identical copies.

Before proceeding, we need to introduce some more notations. Let us consider the following collection of random variables:

$$\overline{N}_{n,\theta,\tilde{\theta}} \sim \text{Poisson}(\lambda_{\text{tot}}(\tilde{\theta})(\sqrt{n}(\theta-\tilde{\theta}) + \tau_n)^2),$$

where $n \in \mathbb{N}$, and $\tilde{\theta}, \theta \in \Theta$. We recall that $\tau_n = n^{3\epsilon}$ is the displacement size. Let us define

$$\overline{Y}_{n,\theta,\tilde{\theta}} := \frac{1}{2\tau_n\lambda_{\text{tot}}(\tilde{\theta})}\overline{N}_{n,\theta,\tilde{\theta}} - \frac{\tau_n}{2}.$$

Theorem 21. *Let fix $\theta \in \Theta$ at which f_θ is continuous and let $\tilde{\theta}_n$ be a preliminary estimator which uses $\tilde{n} := n^{1-\epsilon}$ samples with ϵ small enough, such that it satisfies the concentration bound*

$$\mathbb{P}_\theta(|\tilde{\theta}_n - \theta| > n^{-1/2+\epsilon}) \leq Ce^{-n^\epsilon r} \tag{6.24}$$

for some constants $C, r > 0$. Let

$$\hat{\theta}_n := \tilde{\theta}_n + \frac{Y_n}{\sqrt{n}},$$

be the final estimator as defined in (6.22) and (6.23).

If for every $a \in \mathbb{R}$ one has

$$\lim_{n\to+\infty} \sup_{\tilde{\theta}:|\tilde{\theta}-\theta|<n^{-1/2+\epsilon}} |\mathbb{E}_\theta[e^{iaY_n}|\tilde{\theta}_n = \tilde{\theta}] - \mathbb{E}[e^{ia\overline{Y}_{n,\theta,\tilde{\theta}}}]| = 0, \tag{6.25}$$

then $\hat{\theta}$ is asymptotically optimal and asymptotically normal, i.e., the following convergence in law holds for large n

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}_\theta} \mathcal{N}\left(0, \frac{1}{f_\theta}\right).$$

The proof of the following theorem and the relationship between the extra hypothesis in Eq. (6.25) and Proposition 9 can be found in section 6.15.

In the following section we illustrate our method with results of numerical simulations on a qubit model.

6.8 NUMERICAL EXPERIMENTS

In this section we illustrate the estimation protocol through numerical simulations, using a qubit model inspired by the previous work [76]. The simulations are implemented in Python using the QuTiP package [99, 100].

The quantum Markov chain model consists of a two-dimensional system coupled to two-dimensional noise units by a unitary U_θ with unknown parameter $\theta \in \mathbb{R}$, where the noise units are all prepared in the same initial state $|0\rangle$. Since the system interacts with a fresh noise unit at each step, it suffices to specify the action of U_θ on the states $|00\rangle$ and $|10\rangle$, and we define

$$\begin{aligned} U_\theta : |00\rangle &\to \cos(\theta)\sqrt{1-\theta^2}\,|00\rangle \\ &\quad + i\sin(\theta)\sqrt{1-\theta^2}\,|10\rangle + \theta\,|11\rangle\,, \\ U_\theta : |10\rangle &\to i\sin(\theta)\sqrt{1-\lambda}\,|00\rangle \\ &\quad + \cos(\theta)\sqrt{1-\lambda}\,|10\rangle + \sqrt{\lambda}e^{i\phi}\,|01\rangle\,, \end{aligned}$$

where λ, ϕ are known parameters. In simulations we used $\phi = \pi/4, \lambda = 0.8$ and $\theta = 0.2$ for the true values of the parameter.

In a first simulation study we verify the predictions made in Corollary 4. We run the dynamics for a total of $n = 6 \times 10^5$ time steps, with absorber parameter $\theta_{\text{abs}} = \theta_0$ (cf. section 6.3.1) and system parameter $\theta = \theta_0 + u/\sqrt{n}$ such that $\theta = 0.2$ and the local parameter is $u = 2$.

For each run we perform repeated measurements in the standard basis of the output units, to produce a measurement record $\omega = (\omega_1, \ldots, \omega_n)$. For each such trajectory, the pattern counts $N_{\alpha,n}(\omega)$ are obtained by identifying patterns (sequences starting and ending with a 1) which are separated by at least n^γ 0s and no pattern contains more than n^γ 0s, where $\gamma > 0$ is a small parameter. This can be done by combing through the sequence and identifying occurrences of a

given such pattern padded by n^γ 0s to the left and right (taking care of the special case of the first and last patterns). Note that for any given n this procedure will not count patterns with more than n^γ successive 0s. Since the mean counts for each pattern α is $|\mu_\alpha|^2 u^2$ and $|\mu_\alpha|^2$ decays exponentially with $|\alpha|$, we find that patterns of such length are unlikely to occur for large n. The results of $N = 2000$ independent repetitions of the experiment are illustrated in Figure 6.2 which shows a good match between the counts histograms corresponding to several patterns (in blue) and the theoretical Poisson distributions Poisson$(|\mu_\alpha|^2 u^2)$ (in orange) , as predicted by Corollary 4.

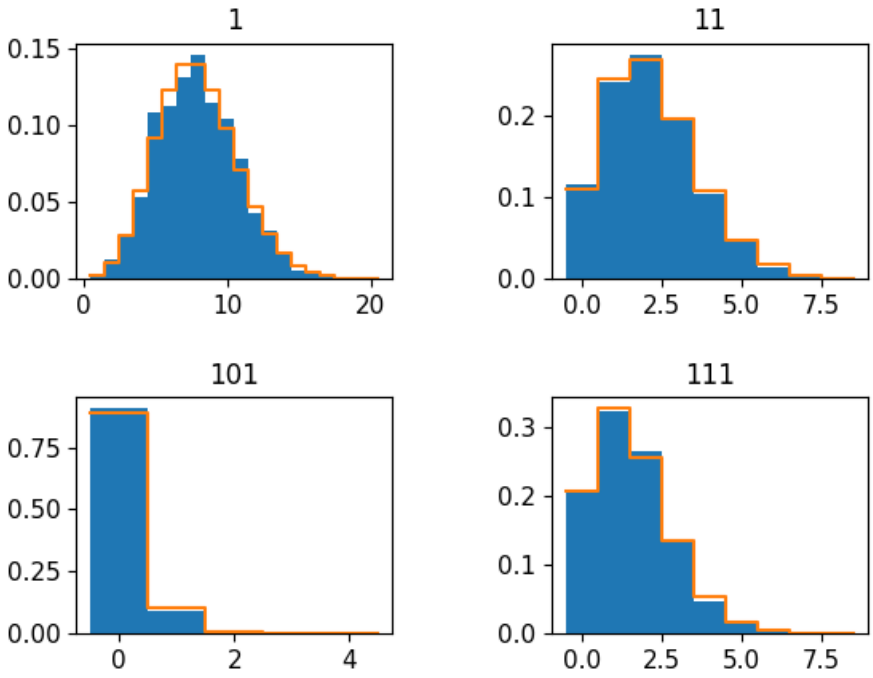

Figure 6.2: (Blue) Counts histograms for patterns $\alpha = 1, 11, 101, 111$ from N=2000 trajectories. (Orange line) The Poisson distribution with intensity given by $|\mu_\alpha|^2 u^2$ matches well the empirical counts distribution, as prescribed by Corollary 4.

In a second simulation study we use system parameter $\theta = 0.2$ and set the absorber at $\theta_{\text{abs}} = \theta - \delta_n$, with displacement $\delta_n = n^{-1/2}\tau_n$ for $n = 6 \times 10^5$ and $\tau_n = 7$. We perform the same measurement as above and extract the pattern counts $N_{\alpha,n}(\omega)$ for each trajectory ω. We then use the pattern counts estimator (6.23) to estimate θ, taking $\tilde{\theta}_n = \theta$. This amounts to assuming that the first stage of the general estimation procedure outlined in section 6.7 gives a perfect estimator, which is then used in setting the absorber parameter in the second step. While this procedure cannot be used in a practical situation, the study has theoretical value in that it allows us to study the performance of the pattern counts estimator in *its own right*, rather than in conjunction with the first step estimator. Figure 6.3 shows that the final estimator $\hat{\theta}_n$ has Gaussian distribution with variance closely matching $1/(nf_\theta)$, thus achieving the QCRB in this idealised setup. This can also be seen by comparing the "effective" Fisher information $F_{\text{eff}} := (n\overline{(\hat{\theta}_n - \theta)^2})^{-1}$

where the mean square error is estimated from the data, with the QFI rate f_θ; the former is equal to $F_{\text{eff}} = 13.8$ while the latter is $f_\theta = 13.5$.

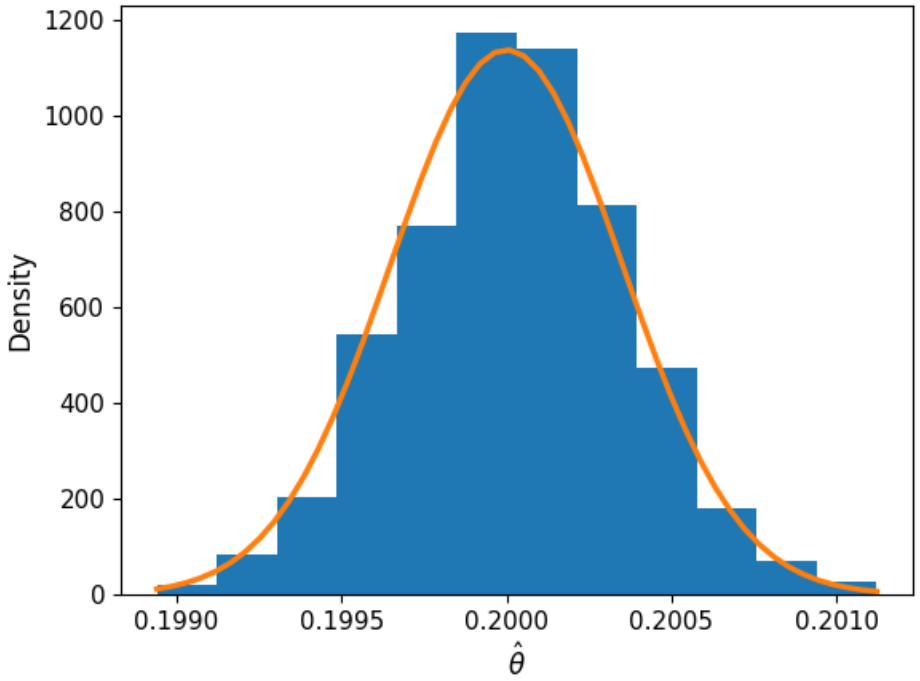

Figure 6.3: (Blue) Histogram of the final estimator $\hat{\theta}_n$ from N=1000 trajectories with no first stage estimation ($\theta = \tilde{\theta}_n$). The effective Fisher information (inverse of rescaled estimated variance) $F_{\text{eff}} \approx 13.8$ matches closely the QFI $f_\theta = 13.5$. (Orange line) For comparison we plot the density of the normal distribution with mean $\bar{\theta} = 0.2$ and variance $\sigma^2 = (nf_\theta)^{-1}$.

In the third simulations study we implement the full estimation procedure described in section 6.7 including the first stage estimator. In the first step we fix $\theta = 0.2$, and run the Markov chain (without the absorber) for $\tilde{n} = 4 \times 10^5$. We then perform sequential measurements in the standard basis on the output noise units to obtain a measurement trajectory from which we compute $\tilde{\theta}_n$ by equating the empirical mean (the average number of 1 counts) with the stationary expected value $c_\tau := \text{Tr}(\rho_\tau^{\text{ss}} K_{\tau,1}^* K_{\tau,1})$, where $K_{\tau,i}$ are the system's Kraus operators. The effective Fisher information of this estimator is $F_{\text{eff}} = 4.06$, significantly lower than the QFI $f_\theta = 13.5$.

We then set the absorber to $\theta_{\text{abs}} = \tilde{\theta}_n - n^{-1/2}\tau_n$ where $\tau_n = 25.5$ and $n = 6.6 \times 10^6$ and we run the system and absorber chain for $n' = n - \tilde{n}$ steps. The system and absorber is initialised in the pure stationary state corresponding to $\tilde{\theta}_n$, but since the system and absorber dynamics is assumed to be primitive, any other initial state will result in equivalent asymptotic results. We then perform the counting measurement as in the previous simulation studies to obtain a trajectory $\omega \in \{0,1\}^{n'}$ and extract the pattern counts $N_{\alpha,n'}(\omega)$.

From these average counts we compute the estimator (6.23), where the Fisher information f is computed at $\tilde{\theta}_n$ and multipled by n'/n to account for the smaller number of samples used in the last step. The results are illustrated in Figure

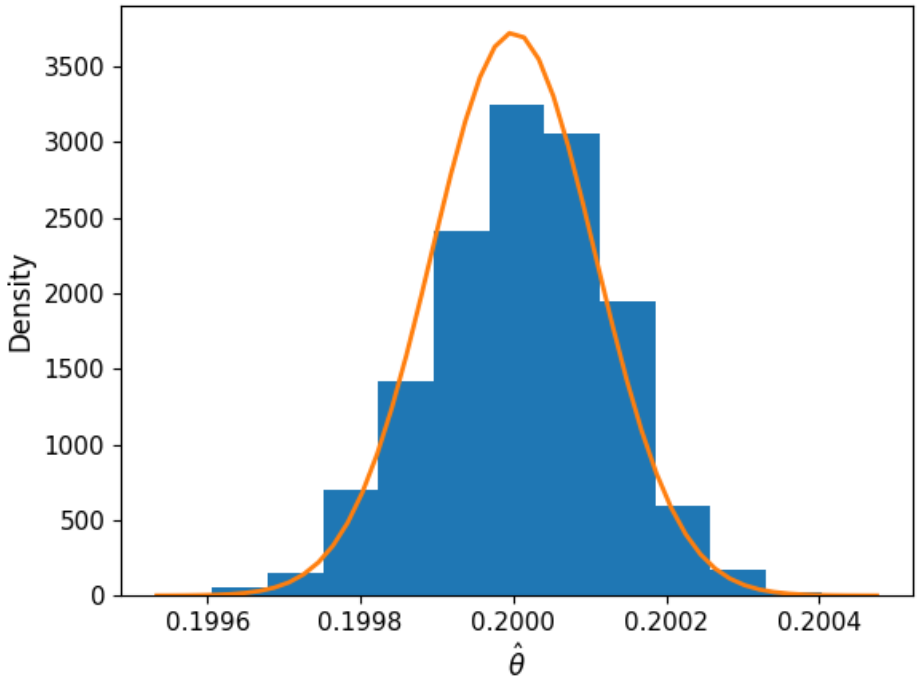

Figure 6.4: (Blue) Histogram of final estimator $\hat{\theta}_n$ from N=1070 trajectories; the effective Fisher information is $F_{\text{eff}} = 10.8$ compared to the QFI $f_\theta = 13.5$. (Orange line) The density of the normal distribution with mean $\bar{\theta} = 0.2$ and variance $\sigma^2 = (nf_\theta)^{-1}$.

.

6.4 which compares the histogram of $\hat{\theta}_n$ (in blue) with the density of a normal distribution with mean θ and variance $1/(nf_\theta)$. We find that there is a good fit with the normal distribution but less accurate that that of the second simulation study, cf. Figure 6.3. This is expected, since the final estimator is based on a two stage estimation process and does not use any prior information about θ. A more accurate measure of the protocol's performance is given by the effective Fisher information which works out as $F_{\text{eff}} = 10.8$ compared to the QFI rate $f_\theta = 13.5$, while the effective Fisher information of the first stage was only 4.06. These simulation results are in agreement with the theoretical arguments put forward in section 6.7 which indicate that the two stage estimator attains the QCRB asymptotically.

6.9 CONCLUSIONS AND OUTLOOK

In this chapter we developed a computationally efficient strategy to estimate dynamical parameters of a quantum Markov chain and provided strong theoretical evidence that the estimator achieves the quantum Cramér-Rao bound in the large time limit. In addition, we established asymptotic results pertaining to the mathematical structure of the output state which are of more general interest.

The estimation strategy consisted of two estimation stages. In the first stage a rough estimator is computed from outcomes of simple output measurements by

using a fast but non-optimal procedure, e.g. equating the empirical counts with the expected value at the estimated parameter. In the second stage, we used a coherent quantum absorber [157] to post-process the output [78, 177]. When tuned to the true value of the system parameter, the absorber "reverts" the evolution such that, in the stationary regime, the post-processed output is identical to the input "vacuum" product state. On the other hand, small mismatches between system and absorber parameters lead to slight rotations away from this product state, which can be detected by simple sequential measurements on the output "noise units". To achieve the perfect "null" setup it would seem natural to use the first stage estimator as absorber parameter, but as shown in [76], this leads to non-identifiability issues and sub-optimal final estimators. Instead, we applied the displaced-null technique [76] which prescribes an extra parameter shift, calibrated to remove the non-identifiability issue while preserving the optimality of the sequential measurement.

The key theoretical contributions of this work are related to the understanding of the output state and the stochastic measurement process. We introduced the concept of translationally invariant modes (TIMs) of the output and showed how they generate a Bosonic algebra in the asymptotic limit. Each mode is labelled by a binary sequence called a "pattern" and its creation operator is an average of shifted blocks consisting of tensor products operators. We then showed that when the mismatch between system and absorber parameter scales at the estimation rate $n^{-1/2}$, the restriction of the output state to the TIMs becomes a multi-mode coherent state whose displacement depends linearly on the mismatch (quantum Gaussian shift model). Moreover, we showed that the TIMs carry all the quantum Fisher information of the output state and are therefore the relevant quantum statistics of the problem. While homodyne is the standard optimal measurement for such models, in the presence of the additional parameter displacement the modes amplitudes become large and counting measurements become effectively equivalent to homodyne. Surprisingly, we discovered that the sequential counting measurement acts as a joint measurement of all TIMs number operators. Due to the proximity to the vacuum state, typical trajectories consist of a relatively small number of patterns separated by long sequences of 0s. We showed that for large times the patterns counts distribution converges to the Poisson distribution of the TIMs coherent state. These insights allowed us to devise a simple "pattern counting" estimator for estimating the unknown parameter.

Our discrete-time results open the way for fast and optimal continuous-time estimation strategies based on coherent absorber post-processing [177] and simple pattern counting estimation, as opposed to expensive maximum likelihood methods. Interesting and important topics for future investigations concerns the robustness of the pattern counting method with respect to various types of noises, improving the estimation accuracy for short times, extensions to multi-parameter models and the relationship between the general theory of quantum absorbers [162] and quantum estimation.

6.10 PROOFS OF PROPOSITION 8 AND COROLLARY 2

Proof of Proposition 8. Let $\boldsymbol{\alpha} = (\alpha^{(1)}, \ldots, \alpha^{(p)}), \boldsymbol{\beta} = (\beta^{(1)}, \ldots, \beta^{(q)})$ be the patterns with non-zero counts for $\mathbf{n}$ and respectively $\mathbf{m}$; let $(n_1, \ldots, n_p)$ and $(m_1, \ldots, m_q)$ be their counts, and let M be the maximum length of all patterns in $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$.

Since pattern creation operators $A^*_\alpha(n)$ involve sums of σ_i^α for different positions i, the Fock state $|\mathbf{n}; n\rangle$ is a superposition of vectors obtained by applying the following type of *ordered* products to the vacuum

$$\prod_{k=1}^{n_p} \sigma_{i_{p,k}}^{\alpha^{(p)}} \cdots \prod_{k=1}^{n_1} \sigma_{i_{1,k}}^{\alpha^{(1)}}$$

where $i_{j,k}$ is the index marking the location of the left end of the k-th pattern $\alpha^{(j)}$, with $j \in \{1, \ldots, p\}$ and $k \in \{1, \ldots, n_j\}$. In general, some of these operators may 'overlap' (act on the same spin) and the computation of the superposition becomes cumbersome. However, since the total length of the patterns is finite, for large n, the main contribution in the superposition comes from arrangements in which there are no overlapping patterns. Even more, we can restrict to arrangements where the patterns are separated by at least M zeros, in which case each pattern in the sequence of zeros and ones can be identified unambiguously.

Therefore

$$\begin{aligned} &|\mathbf{n}; n\rangle \\ &= \frac{1}{\sqrt{n^{|\mathbf{n}|}}} \frac{1}{\sqrt{\mathbf{n}!}} \sum_{\mathbf{i} \in \mathcal{I}(\mathbf{n}; n)} \prod_{j=1}^{p} \prod_{k=1}^{n_j} \sigma_{i_{j,k}}^{\alpha^{(j)}} |\Omega_n\rangle + o(1) \\ &= \frac{1}{\sqrt{n^{|\mathbf{n}|}}} \frac{1}{\sqrt{\mathbf{n}!}} \sum_{\mathbf{i} \in \mathcal{I}(\mathbf{n}; n)} |\omega(\mathbf{i}, \mathbf{n})\rangle + o(1) \end{aligned} \tag{6.26}$$

where $\mathbf{n}! := n_1! \ldots n_p!$ and $\mathcal{I}(\mathbf{n}; n)$ is the subset of locations $\mathbf{i} = \{i_{j,k}\}$ leading to non-overlapping patterns with counts set $\mathbf{n}$, such that all patterns are at distance of at least M from each other, and $\omega(\mathbf{i}, \mathbf{n}) \in \{0, 1\}^n$ is the basis vector (trajectory) obtained by placing the all the patterns corresponding to the counts set $\mathbf{n}$ at locations prescribed by $\mathbf{i}$. More precisely $\mathbf{i} \in \mathcal{I}(\mathbf{n}; n)$ if for any two pairs (j, k) and $(\tilde{j}, \tilde{k})$ with $i_{j,k} \leq i_{\tilde{j},\tilde{k}}$ one has $i_{j,k} + |\alpha^{(j)}| + M - 1 < i_{\tilde{j},\tilde{k}}$. The basis vector $|\omega(\mathbf{i}, \mathbf{n})\rangle$ consist of zeros except patterns $\alpha^{(j)}$ written at positions $i_{j,k}$ for $j \in \{1, \ldots, p\}$ and $k \in \{1, \ldots, n_j\}$.

To show that the remainder term is $o(1)$ note that any term in the superposition $|\mathbf{n}, n\rangle$ corresponding to a specific product of σs is a vector of the standard basis, and any such vector $|\omega\rangle$ has at most a fixed number $M \cdot |\mathbf{n}|$ of 1s. Note also that the action of applying a pattern σ_i^α to the vacuum cannot be reversed by applying subsequent σs, since these contain only creation operators. This means that for a given $|\omega\rangle$, the locations $\mathbf{i} = \{i_{j,k}\}$ of the σs producing this vector are limited to an area of size $2M$ around each 1 in the sequence, or in other words, the coefficient of $|\omega\rangle$ is bounded by $(2M)^{M|\mathbf{n}|}$. On the other hand, the number of basis vectors $|\omega\rangle$ obtained by applying patterns such that at least two overlap or are at distance smaller than M from each other is $o(n^{|\mathbf{n}|})$ since the number of possible locations for two such σs is $O(n)$. Therefore, the remainder term in (6.26) is $o(1)$.

Using equation (6.26) we obtain

$$\langle \mathbf{m}; n|\mathbf{n}; n\rangle = \frac{1}{\sqrt{n^{|\mathbf{n}|}\cdot n^{|\mathbf{m}|}\cdot \mathbf{n}!\cdot \mathbf{m}!}} \sum_{\mathbf{i},\tilde{\mathbf{i}}} \langle \omega(\mathbf{i},\mathbf{n})|\omega(\tilde{\mathbf{i}},\mathbf{m})\rangle + o(1)$$

where the sum runs over $\mathbf{i} \in \mathcal{I}(\mathbf{n}; n)$ and $\tilde{\mathbf{i}} \in \mathcal{I}(\mathbf{m}; n)$. Since each $\mathbf{i}$ and $\tilde{\mathbf{i}}$ uniquely determines the patterns it contains, the basis vectors $|\omega(\mathbf{i},\mathbf{n})\rangle$ and $|\omega(\tilde{\mathbf{i}},\mathbf{m})\rangle$ have non-zero overlap (coincide) only if their sets of patterns coincide, i.e., $\mathbf{n} = \mathbf{m}$. Therefore, if $\mathbf{n} \neq \mathbf{m}$ hold then

$$\lim_{n\to\infty} \langle \mathbf{m}; n|\mathbf{n}; n\rangle = 0.$$

On the other hand, if $\mathbf{m} = \mathbf{n}$ then

$$\lim_{n\to\infty} \langle \mathbf{n}; n|\mathbf{n}; n\rangle = \lim_{n\to\infty} \frac{1}{n^{|\mathbf{n}|}\mathbf{n}!}(\mathbf{n}!)^2 \frac{|\mathcal{I}(\mathbf{n}, n)|}{\mathbf{n}!} = 1$$

where we took into account that each basis vector $|\omega(\mathbf{i},\mathbf{n})\rangle$ appears $\mathbf{n}!$ times in the sum over $\mathbf{i} \in \mathcal{I}(\mathbf{n}; n)$, and that

$$\lim_{n\to\infty} \frac{|\mathcal{I}(\mathbf{n}, n)|}{n^{|\mathbf{n}|}} = 1.$$

□

Proof of Corollary 2. As shown in the proof of Proposition 8 the "Fock state" $|\mathbf{n}; n\rangle$ can be approximated by a superposition of basis states in which the patterns in $\mathbf{n}$ are non-overlapping and are situated at least at a certain distance from each other. Moreover the remainder term contains $o(n^{|\mathbf{n}|})$ basis vectors with bounded coefficients. After applying $A^*_\beta(n)$, the multiplicity of each basis vector $|\omega\rangle$ is finite and the number of possible basis vectors is $o(n^{|\mathbf{n}|+1})$. Taking account of the factor $n^{-(|\mathbf{n}|+1)/2}$ we find that the action of $A^*_\beta(n)$ on the $o(1)$ remainder in (8) is still $o(1)$. On the other hand, the action on the main term in (8) is to add a pattern β separated by the other patterns by $\max(M, |\beta|)$, with a negligible term coming from locations in which β overlaps with or is too close to one of the existing patterns. The factor $\sqrt{n_\beta + 1}$ comes from the definition of the "Fock state" $|\mathbf{n} + \delta^{(\beta)}; n\rangle$.

The action of $A_\beta(n)$ on one of the vectors $|\omega(\mathbf{i},\mathbf{n})\rangle$ is to produce a superposition of basis vectors in which the pattern β has been removed from the set $\mathbf{n}$ of patters, at all possible locations. This may include removing part of a existing pattern $\alpha^{(j)}$ which coincides with β. These two case will produce orthogonal vectors and can be evaluated separately. In the first case, the pattern β is removed from one of the locations where such pattern existed.

In the Fourier decomposition of $\sqrt{n_\beta}|\mathbf{n} - \delta^\beta; n\rangle$, a basis vector $|\omega\rangle$ given by non-overlapping patterns has coefficient

$$c(\omega) = \sqrt{n_\beta}\frac{\mathbf{n}!}{n_\beta}\frac{\sqrt{n\cdot n_\beta}}{\sqrt{n^{|\mathbf{n}|}\mathbf{n}!}} = \mathbf{n}!\frac{\sqrt{n}}{\sqrt{n^{|\mathbf{n}|}\mathbf{n}!}}$$

Such a vector can be obtained in approximately n ways by removing a pattern β from a basis vector which appears in the decomposition of $|\mathbf{n}; n\rangle$. Therefore its coefficient in $A_\beta|\mathbf{n}, n\rangle$ is approximately

$$\frac{\mathbf{n}!}{\sqrt{n}}\cdot n \cdot \frac{1}{\sqrt{n^{|\mathbf{n}|}\mathbf{n}!}}$$

where we took into account a factor $1/\sqrt{n}$ from the definition of A_β. We therefore obtain that the coefficients of the non-overlapping terms in $A_\beta|\mathbf{n};n\rangle$ and $\sqrt{n_\beta}|\mathbf{n}-\delta^\beta;n\rangle$ agree asymptotically. The fact that the $o(1)$ terms remain small after applying A_β can be shown similarly to the above.

□

6.11 PROOFS OF THEOREM 19 AND LEMMA 8

Proof of Theorem 19. In order to prove the theorem, we will make use of the method of moments; this can be done since the moment problem corresponding to the moments of Poisson and Gaussian random variables admits a unique solution (this can be seen for instance using Cramér condition, see [114, 147]).

Number operators. We will show the convergence of the moments of $N_\alpha(n)$ in the state $\rho^{\text{out}}_{u,n}$ to those of a Poisson random variable of intensity $\lambda(u) := u^2\lambda_\alpha$. We recall the r-th moment of a Poisson random variable of intensity $\lambda(u)$ is equal to $\sum_{m=1}^r S(r,m)\lambda(u)^m$ where $S(r,m)$ are the Stirling numbers of the second type: a combinatorial interpretation of $S(r,m)$ is the number of partitions in m non-empty subsets of a set of cardinality r.

Let us focus on the expression of the r-th moment of $N_\alpha(n)$. For simplicity we denote $\langle X\rangle_u := \tilde{\rho}^{\text{out}}_{u,n}(X)$. From (6.10) we have

$$\langle N_\alpha(n)^r\rangle_u = \frac{1}{n^r}\sum_{\substack{i_1,\dots,i_r=1\\ j_1,\dots,j_r=1}}^{n-|\alpha|+1}\langle \sigma^\alpha_{i_1}\sigma^{\alpha*}_{j_1}\cdots\sigma^\alpha_{i_r}\sigma^{\alpha*}_{j_r}\rangle_u. \tag{6.27}$$

Splitting of the sum based on non-overlapping groups of σs. Let us consider a term $\sigma^\alpha_{i_1}\sigma^{\alpha*}_{j_1}\cdots\sigma^\alpha_{i_r}\sigma^{\alpha*}_{j_r}$ and represent each σ^α_i or $\sigma^{\alpha*}_i$ as a block of length $|\alpha|$ covering positions $\{i,i+1,\dots i+|\alpha|-1\}$ of the string $\{1,2,\dots n\}$. Depending on the overlapping pattern of the blocks, the indices can be split (uniquely) in a number $s = s(i_1,j_1,\dots,i_r,j_r)$ of groups ($1\le s\le 2r$) such that blocks in different groups do not overlap, and each group cannot be split into further non-overlapping sub-groups. Among these groups we identify $g = g(i_1,j_1,\dots,i_r,j_r)$ special groups characterised by the fact that they are made up of one or more pairs of blocks of the type $(\sigma^{\alpha*}_z\sigma^\alpha_z)$ for some z. We call $\mathcal{P}_0$ the set of such groups. Note that not all groups associated to a product of σs can be in $\mathcal{P}_0$ because the order of $\sigma^{\alpha*}_z$ and σ^α_z in the latter is opposite to that in which such terms appear in $N_\alpha(n)$.

For example, consider the pattern α ="11" for $n=10$ and the term

$$\sigma^\alpha_8\sigma^{\alpha*}_1\sigma^\alpha_6\sigma^{\alpha*}_8\sigma^\alpha_1\sigma^{\alpha^*}_2\sigma^\alpha_2\sigma^{\alpha*}_5$$

This has 3 non-overlapping groups of operators (that commute with each others): the first one

$$\sigma^{\alpha*}_1\sigma^\alpha_1\sigma^{\alpha*}_2\sigma^\alpha_2$$

belongs to $\mathcal{P}_0$, while the second and third ones

$$\sigma^\alpha_6\sigma^{\alpha*}_5 \qquad \text{and} \qquad \sigma^\alpha_8\sigma^{\alpha*}_8$$

are not in $\mathcal{P}_0$. Therefore $s=3$ and $g=1$.

Let us now look at the expected value of a given product of σs. We have

$$\langle \sigma^{\alpha}_{i_1} \sigma^{\alpha *}_{j_1} \cdots \sigma^{\alpha}_{i_r} \sigma^{\alpha *}_{j_r} \rangle_u = \rho^{\mathrm{ss}}_u \left(\mathcal{C}_{u,1} \tilde{T}^{x_1}_u \cdots \tilde{T}^{x_{s-1}}_u \mathcal{C}_{u,s}(\mathbf{1}) \right) \tag{6.28}$$

where x_i's are the distances between the non-overlapping groups and $\mathcal{C}_{u,i}$ is the map corresponding to the i-th group, which is computed according to the following rule. In every group we multiply the operators on a given position (we may have several σ^+, σ^- or σ^0 on one position) and the result will be an element of the set $\mathcal{O} := \{\sigma^+, \sigma^-, \sigma^0, |0\rangle\langle 0|, |1\rangle\langle 1|, \mathbf{0}\}$. For example the first group above gives

$$\sigma^{\alpha *}_1 \sigma^{\alpha}_1 \sigma^{\alpha *}_2 \sigma^{\alpha}_2 = (|0\rangle\langle 0|)_1 (|0\rangle\langle 0|)_2 (|0\rangle\langle 0|)_3$$

while

$$\sigma^{\alpha}_6 \sigma^{\alpha *}_5 = \sigma^-_5 (|1\rangle\langle 1|)_6 \sigma^+_7$$

and

$$\sigma^{\alpha}_8 \sigma^{\alpha *}_8 = (|1\rangle\langle 1|)_8 (|1\rangle\langle 1|)_9.$$

Suppose the result of this computation is $O^1_i \cdots O^k_{i+k-1}$, with $O^k \in \mathcal{O}$; then in the expectation, this translates into a superoperator $\mathcal{C}_u = \mathcal{C}_u[O^1] \circ \cdots \circ \mathcal{C}_u[O^k]$ obtained by composing in the same order basic maps $\mathcal{C}_u[O]$ defined as follows

$$\begin{aligned} &\mathcal{C}_u[\sigma^+] := \mathcal{A}_{u,1}, \quad \mathcal{C}_u[\sigma^0] := \mathcal{A}_{u,0}, \quad \mathcal{C}_u[\sigma^-] := \mathcal{A}_{u,-1}, \\ &\mathcal{C}_u[\mathbf{0}] := 0, \mathcal{C}_u[|j\rangle\langle j|] := \mathcal{B}_{u,j}, \; j \in \{0,1\}. \end{aligned}$$

where $\mathcal{A}_{u,i}$ are defined as in equation (6.15) by replacing $\tilde{K}_i$ with $\tilde{K}_{u,i}$ and

$$\mathcal{B}_{u,j}(x) = \begin{cases} \tilde{K}^*_{u,1} x \tilde{K}_{u,1} & j = 1 \\ \tilde{K}^*_{u,0} x \tilde{K}_{u,0} & j = 0 \end{cases}.$$

In what follows we will drop the label u when $u = 0$.

We can now compute the expectation of our example product of σs as

$$\begin{aligned} &\langle \sigma^{\alpha}_8 \sigma^{\alpha *}_1 \sigma^{\alpha}_6 \sigma^{\alpha *}_8 \sigma^{\alpha}_1 \sigma^{\alpha^*}_2 \sigma^{\alpha}_2 \sigma^{\alpha *}_5 \rangle_u = \\ &\rho^{\mathrm{ss}}_u \left(\mathcal{B}^3_{u,0} \tilde{T}_u \mathcal{A}_{u,-1} \mathcal{B}_{u,1} \mathcal{A}_{u,1} \mathcal{B}^2_{u,1}(\mathbf{1}) \right). \end{aligned}$$

Let us define the space

$$\mathbf{1}^{\perp} := \{x \in \mathcal{B}(\mathbb{C}^D) : \tilde{\rho}^{\mathrm{ss}}(x) = 0\},$$

and note that the following properties hold true:

- $\mathcal{A}_0$ and $\mathcal{B}_0$ leave $\mathbf{1}^{\perp}$ invariant;
- the image of $\mathcal{A}_i$ and $\mathcal{B}_j$ for $i, j \neq 0$ is contained in $\mathbf{1}^{\perp}$;
- for every $m \geq 1$, the norm of $\mathcal{A}^m_0 = \tilde{T}^m$ restricted to $\mathbf{1}^{\perp}$ is less or equal than $c\lambda^m$ for some $c > 0$ and $1 > \lambda > 0$.

The reason why we singled out groups in $\mathcal{P}_0$ is given by the following key observation: such groups are *the only ones* for which the corresponding map $\mathcal{C}_{u,i}$ is only composed by $\mathcal{A}_{u,0}$s and $\mathcal{B}_{u,0}$s. Any other group will contain at least one

$\mathcal{A}_{u,\pm 1}$ or $\mathcal{B}_{u,1}$ factor. From this and the properties above it follows that if $\mathcal{C}$ is not in $\mathcal{P}_0$ then

$$\|\tilde{T}^m \circ \mathcal{C}\| \leq c\lambda^m. \tag{6.29}$$

We will often identify the maps with the corresponding patterns saying, for instance, that the map is in $\mathcal{P}_0$ when the corresponding group is.

We will prove the convergence of moments by expanding in Taylor series in u and showing the convergence at each order. The moment of order r of the Poisson distribution with intensity $u^2\lambda_\alpha$ is

$$m_r = \sum_{k=1}^{r} S(r,k) u^{2k} \lambda_\alpha^k \tag{6.30}$$

where $S(r,k)$ are the Stirling numbers of second kind; the mth derivative with respect to u at $u = 0$ is $S(r, m/2)m!\lambda_\alpha^{m/2}$ for m even, and zero otherwise.

Recall that each term in (6.27) gives rise to a certain set of groups of indices made up of overlapping blocks, and each group corresponds to a map $\mathcal{C}_{u,i}$ such that the expectation is expressed as in equation in (6.28). We will show that in limit of large n, the only terms which contribute to the m-th derivative are those coming from certain configurations with $s - g = m$ where s is the number of groups and g is the number of groups in $\mathcal{P}_0$.

Taylor approximation for a given set $(\mathcal{C}_{u,1}, \dots \mathcal{C}_{u,s})$ up to order $m = s - g$. Let us consider the sum of all the terms in Eq. (6.27) coming from all the correlations corresponding to a given sequence of maps $\boldsymbol{\mathcal{C}} := (\mathcal{C}_{u,1}, \dots, \mathcal{C}_{u,s})$. This contribution is given by the product between a combinatorial factor (independent of n) counting how many products of σs in (6.27) lead to the same set of maps $\boldsymbol{\mathcal{C}}$ and the following sum

$$\frac{1}{n^r} \sum_{x_0 + \dots + x_s = n - K} \tilde{\rho}_u^{\mathrm{ss}} \left(\mathcal{C}_{u,1} \tilde{T}_u^{x_1} \cdots \tilde{T}_u^{x_{s-1}} \mathcal{C}_{u,s}(\mathbf{1}) \right), \tag{6.31}$$

where K is the total length of all the s blocks in $\boldsymbol{\mathcal{C}}$, which is smaller than $2|\alpha|r$. Note that the factors $\tilde{T}_u^{x_0}$ and $\tilde{T}_u^{x_s}$ have been suppressed due to stationarity but the indices x_0, x_s are still present in the sum. We denote by g the number of maps in $(\mathcal{C}_{u,1}, \dots, \mathcal{C}_{u,s})$ that belong to $\mathcal{P}_0$.

Let us now consider the Taylor expansion of the correlations in (6.31). The 0-th order term is 0 because at least one $\mathcal{C}_i$ is not in $\mathcal{P}_0$ and hence $\tilde{\rho}^{\mathrm{ss}}$ annihilates the result. This is because on one hand, $C_{i*}(\tilde{\rho}^{\mathrm{ss}}) = \tilde{\rho}^{\mathrm{ss}}$ for all $C_i \in \mathcal{P}_0$, but on the other hand at least one $\mathcal{C}_i$ is not in $\mathcal{P}_0$ and hence it contains a term $\mathcal{A}_{\pm 1}$ or $\mathcal{B}_1$ for which $\mathcal{A}_{\pm 1 *}(\tilde{\rho}^{\mathrm{ss}}) = \mathcal{B}_{1*}(\tilde{\rho}^{\mathrm{ss}}) = 0$ since $\tilde{K}_1 |\chi^{\mathrm{ss}}\rangle = 0$.

Before addressing in detail the derivatives of (6.31) we make some remarks concerning the magnitude of the sum. Note that this contains $O(n^s)$ terms which are uniformly bounded and the largest possible value of s is $2r$ (all patterns are non-overlapping). On the other hand the sum is preceded by the factor n^{-r} and any derivative further multiplies it by $n^{-1/2}$ since all operators depend on u via $u/\sqrt{n}$. These arguments alone are not sufficient to deduce the convergence of (low order) derivatives, at least for configurations $\boldsymbol{\mathcal{C}}$ with large s.

The key additional ingredient is the fact that for $\mathcal{C}_i \notin \mathcal{P}_0$, the factors $\tilde{T}^{x_i}\mathcal{C}_i$ are exponentially decreasing, cf. equation (6.29). This will provide more conservative upper bounds for the derivatives of the sum (6.31), as detailed below. Before

considering that, note that the exponential bound can also be used to get an alternative proof of the convergence to 0 of

$$\frac{1}{n^r} \sum_{x_0+\cdots+x_s=n-K} \mathcal{C}_{u,1}\tilde{T}_u^{x_1}\cdots\tilde{T}_u^{x_{s-1}}\mathcal{C}_{u,s}(\mathbf{1}).$$

Indeed, by summing over x_is for groups not in $\mathcal{P}_0$, we get that the sum (6.31) is $O(n^g/n^r)$ rather than $O(n^s/n^r)$. Since in any configuration $\boldsymbol{\mathcal{C}}$, at least one group is not in $\mathcal{P}_0$, and each group in $\mathcal{P}_0$ has at least two of the original $2r$ blocks, we see that $g < r$ so that the whole sum is $O(n^{-1})$. The upshot for derivatives will be that in order to create contributions that do not decay, the derivatives have to be applied in an "efficient" way, namely to factors of the type $T^{x_i}\circ\mathcal{C}_i$ with $\mathcal{C}_i$ not in $\mathcal{P}_0$. This will break the exponential decay and allow for the balancing of the terms in the derivative with the pre-factor n^{-r}.

Consider now the case of the first order derivative of the sum (6.31). This will split into a sum of sub-sums, one for each of the terms $\mathcal{C}_{u,i}$ or $T_u^{x_i}$ that are differentiated. Each sub-sum will be shown to have a decaying contribution to the derivative.

i) If a term $\mathcal{C}_{u,i}$ in $\mathcal{P}_0$ is differentiated, then the sub-sum is similar to the original sum with the difference that $\mathcal{C}_{u,i}$ is replaced by another bounded term $\dot{\mathcal{C}}_i$). The overall contribution is then of the order $n^{-r}\times n^g\times n^{-1/2}$ which decays since $g < r$.

ii) If a term $\tilde{T}_u^{x_{i-1}}$ is differentiated, for which $\mathcal{C}_{u,i}$ is in $\mathcal{P}_0$ then the product $\tilde{T}_u^{x_{i-1}}\mathcal{C}_{u,i}$ becomes the sum

$$S_i = \sum_{l=1}^{x_{i-1}} \tilde{T}^{x_{i-1}-l}\dot{\tilde{T}}\tilde{T}^{l-1}\mathcal{C}_i(\cdot) \tag{6.32}$$

By ergodicity $\tilde{T}^{l-1}(\mathcal{C}_i(\cdot))$ converges exponentially fast to $\tilde{\rho}^{\text{ss}}(\mathcal{C}_i(\cdot))\mathbf{1}$ for large l so

$$S_i = \sum_{l'=1}^{x_{i-1}} \tilde{T}^{l'}(\dot{\tilde{T}}(\mathbf{1}))\cdot\tilde{\rho}^{\text{ss}}(\mathcal{C}_i(\cdot)) + O(1)$$

By the same argument $\tilde{T}^{l'}\dot{\tilde{T}}(\mathbf{1})$ converges to $\tilde{\rho}^{\text{ss}}(\dot{\tilde{T}}(\mathbf{1})) = 0$ for large values of l'. The latter follows from differentiating $\tilde{\rho}_u^{\text{ss}}(\tilde{T}_u(\cdot)) = \tilde{\rho}_u^{\text{ss}}(\cdot)$. Therefore S_i is $O(1)$ which does not change the magnitude of the overall sum before differentiation, so this contribution decays as well, as argued in point i).

iii) If a term $\tilde{T}_u^{x_{i-1}}$ is differentiated, for which $\mathcal{C}_{u,i}$ is not in $\mathcal{P}_0$, one obtains a sum S_i as in equation (6.32). Since $\mathcal{C}_i$ leaves $\mathbf{1}^\perp$ invariant we have $\|\tilde{T}^l\circ\mathcal{C}_i\| \le c\lambda^l$ so the terms with large l become negligible while for small l (large $x_{i-1}-l$) $\tilde{T}^{x_{i-1}-l}(\cdot)$ converges to $\tilde{\rho}^{\text{ss}}(\cdot)\mathbf{1}$. Therefore, using $\sum_{l'=0}^{\infty}\tilde{T}^{l'}\mathcal{C}_i = \tilde{\mathcal{R}}\mathcal{C}_i$ where $\mathcal{R}$ is the Moore-Penrose inverse of $\text{Id}-\tilde{T}$, we obtain

$$S_i = \rho^{\text{ss}}(\dot{\tilde{T}}\tilde{\mathcal{R}}\mathcal{C}_i(\cdot))\mathbf{1} + o(1).$$

Hence differentiating $\tilde{T}_u^{x_{i-1}}$ for which $\mathcal{C}_{u,i}$ is not in $\mathcal{P}_0$ removes the exponential decay associated to un-differentiated term $\tilde{T}_u^{x_{i-1}}\mathcal{C}_{u,i}$ so that the upper bound for this contribution to the derivative is $n^{g+1}n^{-1/2}n^{-r}$. Since $g < r$

we have $g+1-1/2-r<0$ so the contribution to the derivative decays (even though it decays slower than the previous sub-sums, which will be important when considering higher derivatives).

iv) If a term $\mathcal{C}_{u,i}$ not in $\mathcal{P}_0$ is differentiated then for large x_{i-1}

$$\tilde{T}^{x_{i-1}}\dot{\mathcal{C}}_i(\cdot)=\tilde{\rho}^{\mathrm{ss}}(\dot{\mathcal{C}}_i(\cdot))\mathbf{1}+o(1)$$

This means that such a derivative breaks the exponential decay of the product $\tilde{T}^{x_{i-1}}\mathcal{C}_i$ but the contribution to the sum is still bounded by $n^{g+1}n^{-1/2}n^{-r}$.

The upshot of the argument for the first order derivative is that differentiating a $\mathcal{C}_{u,i}$ in $\mathcal{P}_0$ or the factor $T^{x_{i-1}}$ in front of it, does not increase the magnitude of the sum and the contribution to the overall sum decays. On the other hand, differentiating a $\mathcal{C}_{u,i}$ which is not in $\mathcal{P}_0$ or the factor $T^{x_{i-1}}$ in front of it, breaks the overall exponential decay of the product and contributes with an additional factor $n^{1/2}$ compared to the un-differentiated term.

We now consider higher order derivatives. One can again consider the possible positions where derivatives are applied and evaluate their separate contributions to the derivative. Following the same argument as in point i) above, one can see that differentiating a term $\mathcal{C}_{u,i}$ in $\mathcal{P}_0$ once or multiple times does not bring any change compared to the term before differentiation. Similarly, as in point ii) one finds that differentiating $T_u^{x_{i-1}}$ in front of $\mathcal{C}_{u,i}$ which is in $\mathcal{P}_0$, once or multiple times does not bring any change (this is due to the fact that $\frac{d^k}{du^k}T_u^x(\mathbf{1})=0$ for every $k\geq 1$). Now we move our attention to terms $T_u^{x_{i-1}}\mathcal{C}_{u,i}$ for which $\mathcal{C}_{u,i}$ is not in $\mathcal{P}_0$. As in cases iii) and iv) above, differentiating either one of the two factors will break the exponential decay and bring an extra overall multiplicative factor $n^{1/2}$ compared to the un-differentiated terms. However if any of the two factors is differentiated further (more than one derivative in the product $T_u^{x_{i-1}}\mathcal{C}_{u,i}$) then the additional derivative does not change the bound except for the multiplicative factor $n^{-1/2}$ due to differentiation. This shows that in order to obtain contributions that do not vanish asymptotically, the derivatives have to be placed on the products $T_u^{x_{i-1}}\mathcal{C}_{u,i}$ for which $\mathcal{C}_{u,i}$ is not in $\mathcal{P}_0$, with at most one derivative for each product.

Recall that we are considering derivatives up to order $m=s-g$ which is equal to the number of $\mathcal{C}_{u,i}$ not in $\mathcal{P}_0$. According to the argument above, the most favourable positions for the derivatives is on different terms $T_u^{x_{i-1}}\mathcal{C}_{u,i}$. Therefore, for a derivative of order $k\leq m$ the entire derivative can be upper bounded as

$$n^{-r}\times n^{-k/2}\times n^{g}\times n^{k}=n^{-r+g+k/2}$$

Since $k\leq m=s-g$ the exponent is smaller than $t:=-r+g+(s-g)/2$. We will show that $t\leq 0$ with equality if and only if $m=s-g$ and the configuration $\boldsymbol{\mathcal{C}}$ is such that all groups in $\mathcal{P}_0$ consists of 2 perfectly overlapping σs and the groups not in $\mathcal{P}_0$ consists of single blocks. Indeed for given total number $2r$ of σ blocks, $s+g$ is maximum if all groups in $\mathcal{P}_0$ have just two σ blocks and all others have a single block, so $2g+s-g=s+g\leq 2r$, which implies $t\leq 0$. In conclusion, for any configuration $\boldsymbol{\mathcal{C}}$, the derivatives below order $m=s-g$ decay asymptotically, and that of order m does as well unless $\boldsymbol{\mathcal{C}}$ is of the special type described above. Derivatives of order above m will be treated below. Note that for the special configurations, m is even so all odd derivatives decay, as is expected from the

fact that the intensity of the limit Poisson distribution is proportional to u^2. We will now compute the limit of the mth derivative for the special configuration.

If $\mathcal{C}_i \in \mathcal{P}_0$ consists of two blocks of σs then the corresponding product is

$$\sigma^{\alpha *}\sigma^{\alpha} = O_{\alpha_1} \dots O_{\alpha_{|\alpha|}}$$

where $\alpha = (\alpha_1, \dots, \alpha_{|\alpha|})$ and $O_0 := \mathbf{1}$, $O_1 = |0\rangle\langle 0|$. Therefore

$$\mathcal{C}_i = C^{(\alpha_1)} \dots C^{(\alpha_{|\alpha|})}$$

with $C^{(0)} = \mathcal{A}_0 = T$ and $C^{(1)} = \mathcal{B}_0$. Therefore when $x_1, \dots, x_s$ are large the following holds

$$T^{x_{i-1}}\mathcal{C}_i(\cdot) = \rho^{\text{ss}}(\mathcal{C}_i(\cdot))\mathbf{1} + o(1) = \rho^{\text{ss}}(\cdot)\mathbf{1} + o(1)$$

since $\rho^{\text{ss}} \circ \mathcal{A}_0 = \rho^{\text{ss}}$ and $\rho^{\text{ss}} \circ \mathcal{B}_0 = \rho^{\text{ss}}$ This means that for large $x_1, \dots, x_s$ the derivative of order $m = 2(m/2)$ factorises as

$$|\rho^{\text{ss}}(\dot{T}\mathcal{R}\mathcal{A}^{\alpha}(\mathbf{1}) + \rho^{\text{ss}}(\tilde{\mathcal{A}}^{\alpha}(\mathbf{1}))|^{2(m/2)}$$

where $\tilde{\mathcal{A}}^{\alpha}$ is obtained from $\mathcal{A}^{\alpha}$ differentiating the first $\mathcal{A}_1$. This formula follows from the contributions obtained at points iii) and iv) above and the fact that the groups that are differentiated contain a single σ block.

We remark that the term for which we computed the limit appears in the m-th derivative with a factor $m!$ in front.

Finally, we need to compute the combinatorial factor counting the number of terms in the expectation which produce the desired set $\boldsymbol{\mathcal{C}}$ consisting of $g = r - m/2$ groups of two σs in $\mathcal{P}_0$ plus m groups of single σs (with $s = r + m/2$ total number of groups).

We will show that this is exactly $S(r, m/2)$ (Stirling number of second type), i.e., the number of partitions of a set of r elements into $m/2$ non-empty subsets. For this it is enough to show that the numbers of ways in which we can pair the σs is in a bijection with the partitions in $m/2$ classes of r elements. Consider the collection of σs

$$\sigma^{\alpha}_{i_1}\sigma^{\alpha *}_{j_1} \cdots \sigma^{\alpha}_{i_r}\sigma^{\alpha *}_{j_r},$$

which, for our purposes, we can identify with the set of "dipoles"

$$(\alpha, \alpha*)_1, \cdots, (\alpha, \alpha*)_r$$

In order to create a $\mathcal{C}$ in $\mathcal{P}_0$, we need to pair a $\sigma^{\alpha *}_{j_z}$ with a $\sigma^{\alpha}_{i_w}$ such that $z < w$. This induces an equivalence relation on the collection of dipoles where we identify two dipoles $(\alpha, \alpha*)_z$, $(\alpha, \alpha*)_w$ if $j_z = i_w$ and we impose transitivity of this relation. By construction, in each equivalence class $\{(\alpha, \alpha*)_{z_1}, (\alpha, \alpha*)_{z_k}\}$ we have $j_{z_l} = i_{z_{l+1}}$ so all positions of $\alpha*$s are equal to the position of α in the next dipole. In other words, an equivalence class uniquely determines a set of pairs of equal indices, and altogether the set of equivalence classes uniquely determine the splitting into $\mathcal{P}_0$ groups. Since i_{z_1} and j_{z_k} are the only indices that are not paired in the chosen equivalence class, the number of classes is $m/2$, i.e., half of the number of groups not in $\mathcal{P}_0$. The number of ways to group the r dipoles in $m/2$ equivalence classes is the Stirling number of second type $S(r, m/2)$ which provided the combinatorial factor for our mth order derivative.

With this we conclude that the mth order derivative of (6.27) converges to the corresponding derivative of (6.30).

Reminder of the Taylor approximation (order $s-g+1$). The last thing we need to check is that the remainder corresponding to the sum of all the terms in Eq. (6.27) with s non-overlapping blocks is negligible; the reminder is given by

$$\frac{u^{m'}}{m'!n^{r+\frac{m'}{2}}}\sum_{x_1+\cdots+x_s=n-K}\frac{d^{m'}}{du^{m'}}f_{x_1,\ldots,x_s}(u)\Big|_{u=\eta_{x_1,\ldots,x_{s-1}}}$$

for some $|\eta_{x_1,\ldots,x_{s-1}}|\leq|u|/\sqrt{n}$, where $m'=s-g+1$ and

$$f_{x_1,\ldots,x_s}(u)=\rho_u^{\mathrm{ss}}\left(\mathcal{C}_{1,u}T_u^{x_1}\cdots T_u^{x_{s-1}}\mathcal{C}_{s,u}(\mathbf{1})\right).$$

Notice that $r+m'/2>s$: indeed,

$$r+\frac{s-g+1}{2}>s\Leftrightarrow 2r+1>g+s.$$

Therefore, it is enough to show that

$$\frac{d^{l+1}}{du^{l+1}}f_{x_1,\ldots,x_s}(u)\Big|_{u=\eta_{x_1,\ldots,x_{s-1}}},\quad 2\leq l\leq 2r$$

are uniformly bounded: the only terms that requires some care are the derivatives of T_u^x, which are of the type

$$\sum_{0\leq l_1\leq\cdots\leq l_k\leq x}T_\eta^{x-l_k}T_\eta^{(m_k)}\cdots T_\eta^{(m_1)}T_\eta^{l_1-1},$$

where $T_\eta^{(m)}$ stays for the m-th derivative of T_u evaluated at η. Using

- the spectral decomposition of $T_\eta(\cdot)=\rho_\eta^{\mathrm{ss}}(\cdot)\mathbf{1}+\mathcal{R}_\eta$, with $\rho_\eta^{\mathrm{ss}}(\mathcal{R}_\eta(\cdot))=0$ and $\|\mathcal{R}_\eta\|\leq\lambda<1$ (for n big enough) and
- the fact that $T^{(m)}(\mathbf{1})=0$ $(m\geq 1)$,

we have that

$$\begin{aligned}&\sum_{0\leq l_1\leq\cdots\leq l_k\leq x}\|T_\eta^{x-l_k}T_\eta^{(m_k)}\cdots T_\eta^{(m_1)}T_\eta^{l_1-1}\|=\\&\sum_{0\leq l_1\leq\cdots\leq l_k\leq x}\|T_\eta^{x-l_k}T_\eta^{(m_k)}\mathcal{R}_\eta^{l_k-l_{k-1}-1}\cdots T_\eta^{(m_1)}\mathcal{R}_\eta^{l_1-1}\|\leq\\&C\sum_{0\leq l_1\leq\cdots\leq l_k\leq x}\lambda^{l_k},\end{aligned}$$

which is bounded.

Quadratures. For the sake of keeping notation simple, we will show the proof in the case of $z=1$, but the same reasoning applies to the general case inserting z and $\bar{z}$ where needed. We recall that $k=|\alpha|$.

First of all, notice that the mean of $Q_\alpha(n)$ converges to $u\mu_{\alpha,1}$: indeed, its first moment is given by

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n-k+1}\tilde{\rho}^{\mathrm{ss}}(u/\sqrt{n})\left(\sqrt{2}\Re(\mathcal{A}^\alpha)(\mathbf{1})\right)\to u\mu_{\alpha,1}.$$

In order to simplify the proof, we consider the standardised random variable

$$\sqrt{2}\left(Q_\alpha(n) - \frac{n-k+1}{n} u\mu_{\alpha,1}\right) = A^*_\alpha(n) + A_\alpha(n) - 2\Re(\mu_\alpha)u = \frac{1}{\sqrt{n}} \sum_{i=1}^{n-k+1} \tilde{\sigma}^\alpha_i(u) + \tilde{\sigma}^{\alpha *}_i(u),$$

where

$$\tilde{\sigma}^\alpha(u) = \sigma^\alpha - \frac{u}{\sqrt{n}} \rho((\dot{T}\mathcal{R}\mathcal{A}^\alpha + \tilde{\mathcal{A}}^\alpha)(\mathbf{1}))\mathbf{1}.$$

In order to prove the statement, we need to show that the sequence of standardized quadratures converges in law to a standard Gaussian random variable.

For $r \geq 2$, the r-th moment of the standardised random process at time n has the following expression:

$$\sum_{j_1,\dots,j_r \in \{\alpha,\alpha *\}} \frac{1}{n^{r/2}} \sum_{i_1,\dots,i_r=1}^{n-k+1} \langle \tilde{\sigma}^{j_1}_{i_1} \cdots \tilde{\sigma}^{j_r}_{i_r} \rangle_{u/\sqrt{n}}.$$

Consider the correlation corresponding to a choice of indices $i_1, \dots, i_r$ such that there are exactly s non-overlapping groups and let g be the number of groups with overlapping terms of the form $\tilde{\sigma}^{\alpha *}_z(u)\tilde{\sigma}^\alpha_z(u)$ (as before, we denote the set of such groups as $\mathcal{P}_0$). As before, these are the only overlapping groups that at $u = 0$ will produce an operator which is not in $\mathbf{1}^\perp$.

First suppose that $s - g > 0$. We can prevent the maps which do not belong to $\mathcal{P}_0$ to be annihilated by ρ and to cause an exponential decay differentiating and repeating the same computations as in the case of number operators. If we consider the m-th term in the Taylor expansion up to $m = s - g$, we can see that it grows at most as n to the power $g + m/2 - r/2$; if we want the exponent to be bigger or equal than 0, we need that $m \geq r - 2g$; since $m \leq s - g$, one realises that $g + m/2 - r/2$ can at most be equal to 0 and this is true when $m = s - g$ and $s + g = r$, which means that every group in $\mathcal{P}_0$ is of the form $\tilde{\sigma}^{\alpha *}_z(u)\tilde{\sigma}^\alpha_z(u)$ and all the other groups are singletons. However, if a group is composed by a single element, since we centered the random process, the first derivative will not be enough to cancel the exponential decay. Therefore, if $s - g > 0$, the corresponding terms will decrease to 0.

On the other hand, if r is even and $s = g = r/2$, the 0-th order term is equal to 1, hence the leading term comes from the case when r is even and $g = r/2$. In this case one can see that the limit quantity is equal to 1 times the way we can pair the $\tilde{\sigma}(u)$'s in groups of the type $\tilde{\sigma}^{\alpha *}_z(u)\tilde{\sigma}^\alpha_z(u)$ and this is given by $(r-1)!!$ (which is the number of partition into pairs of a set of r elements).

The reminder can be controlled as in the case of number operators. □

Proof of Lemma 8. **Alternative expression for μ_α.** Using that $\tilde{K}_0 |\chi^{\text{ss}}\rangle = |\chi^{\text{ss}}\rangle$ and $\tilde{K}_1 |\chi^{\text{ss}}\rangle = 0$, one can write

$$\begin{aligned}\dot{\tilde{T}}_*(\tilde{\rho}^{\text{ss}}) &= \sum_{i=0}^{1} \dot{\tilde{K}}_i |\chi^{\text{ss}}\rangle \langle\chi^{\text{ss}}| \tilde{K}^*_i + \tilde{K}_i |\chi^{\text{ss}}\rangle \langle\chi^{\text{ss}}| \dot{\tilde{K}}^*_i \\ &= |\dot{\tilde{K}}_0 \chi^{\text{ss}}\rangle \langle\chi^{\text{ss}}| + |\chi^{\text{ss}}\rangle \langle\dot{\tilde{K}}_0 \chi^{\text{ss}}|\end{aligned}.$$

Under the "gauge condition (6.6) and using the explicit expression of $\tilde{K}_{\theta,i}$'s in Eq. (6.35) we obtain $\langle \chi^{ss} | \dot{\tilde{K}}_0 \chi^{ss} \rangle = 0$.

From $\tilde{K}_0^* \tilde{K}_0 + \tilde{K}_1^* \tilde{K}_1 = \mathbf{1}$ we obtain $\tilde{K}_0^* |\chi^{ss}\rangle = |\chi^{ss}\rangle$. Therefore

$$\tilde{K}_0 = |\chi^{ss}\rangle\langle\chi^{ss}| + P_\perp^{ss} \tilde{K}_0 P_\perp^{ss}$$

where $P_\perp^{ss} = \mathbf{1} - |\chi^{ss}\rangle\langle\chi^{ss}|$, which implies

$$\sum_{k\geq 0} \tilde{K}_0^k |\dot{\tilde{K}}_0 \chi^{ss}\rangle = |(\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 \chi^{ss}\rangle$$

and

$$\begin{aligned} \tilde{\mathcal{R}}_* \dot{\tilde{T}}_*(\tilde{\rho}^{ss}) &= |(\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 \chi^{ss}\rangle \langle \chi^{ss}| \\ &+ |\chi^{ss}\rangle \langle (\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 \chi^{ss}| . \end{aligned} \tag{6.33}$$

When we evaluate it against $\mathcal{A}^\alpha(\mathbf{1})$, the first term in the previous equation gets killed, while the second one produces the term

$$\langle \tilde{K}_{\alpha_{|\alpha|}} \cdots \tilde{K}_{\alpha_1} (\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 \chi^{ss} | \chi^{ss} \rangle .$$

The rest of the proof is just a trivial check.

Expression for the total intensity. Since $\lambda_\alpha \geq 0$, one has that $\sum_\alpha \lambda_\alpha = C \in [0, +\infty]$ and the limit is always the same irrespectively of the choice of partial sums. Notice that

$$\begin{aligned} \lambda_{(1)} &= |\langle (\tilde{K}_1 (\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 + \dot{\tilde{K}}_1) \chi^{ss} | \chi^{ss} \rangle|^2 \\ &= \mathrm{Tr}(|\chi^{ss}\rangle \langle \chi^{ss}| Y), \end{aligned}$$

where

$$\begin{aligned} Y &:= |(\tilde{K}_1 (\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 + \dot{\tilde{K}}_1) \chi^{ss}\rangle \\ &\quad \langle (\tilde{K}_1 (\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 + \dot{\tilde{K}}_1) \chi^{ss}| . \end{aligned} \tag{6.34}$$

For any α such that $|\alpha| \geq 2$, λ_α is equal to:

$$\begin{aligned} &|\langle \tilde{K}_{\alpha_{|\alpha|}} \cdots \tilde{K}_{\alpha_2} (\tilde{K}_{\alpha_1} (\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 + \dot{\tilde{K}}_1) \chi^{ss} | \chi^{ss} \rangle|^2 = \\ &|\langle \tilde{K}_1 \tilde{K}_{\alpha_{|\alpha|-1}} \cdots \tilde{K}_{\alpha_2} (\tilde{K}_1 (\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 + \dot{\tilde{K}}_1) \chi^{ss} | \chi^{ss} \rangle|^2 = \\ &\mathrm{Tr}(X \tilde{K}_{\alpha_{|\alpha|-1}} \cdots \tilde{K}_{\alpha_2} Y \tilde{K}_{\alpha_2}^* \cdots \tilde{K}_{\alpha_{|\alpha|-1}}^*), \end{aligned}$$

where $X := |\tilde{K}_1^* \chi^{ss}\rangle \langle \tilde{K}_1^* \chi^{ss}|$ and Y is the same as in Eq. (6.34).

Therefore

$$\sum_{2\leq|\alpha|\leq N} \lambda_\alpha = \mathrm{Tr}(\sum_{k=0}^{N-1} \tilde{T}^k(X) Y)$$

and

$$\lim_{N\to+\infty} \sum_{2\leq|\alpha|\leq N} \lambda_\alpha = \sum_{2\leq|\alpha|} \lambda_\alpha = \mathrm{Tr}(\tilde{\mathcal{R}}(X) Y),$$

therefore

$$\sum_\alpha \lambda_\alpha = \mathrm{Tr}((\mathrm{Id} + \tilde{\mathcal{R}} \mathcal{B}_1)(|\chi^{ss}\rangle \langle \chi^{ss}|) Y).$$

Let us massage a little the expression we obtained:

$$\begin{aligned}
&\mathrm{Tr}((\mathrm{Id}+\tilde{\mathcal{R}}\mathcal{B}_1)(|\chi^{ss}\rangle\langle\chi^{ss}|)Y) = \\
&\langle\chi^{ss}|Y\chi^{ss}\rangle + \langle\chi^{ss}|\mathcal{B}_{1*}\tilde{\mathcal{R}}_*(Y-\mathrm{Tr}(Y)\,|\chi^{ss}\rangle\langle\chi^{ss}|)\chi^{ss}\rangle = \\
&\langle\chi^{ss}|Y\chi^{ss}\rangle + \langle\chi^{ss}|\tilde{T}_*\tilde{\mathcal{R}}_*(Y-\mathrm{Tr}(Y)\,|\chi^{ss}\rangle\langle\chi^{ss}|)\chi^{ss}\rangle - \\
&\langle\chi^{ss}|\mathcal{R}_*(Y-\mathrm{Tr}(Y)\,|\chi^{ss}\rangle\langle\chi^{ss}|)\chi^{ss}\rangle = \\
&\langle\chi^{ss}|Y\chi^{ss}\rangle - \langle\chi^{ss}|(Y-\mathrm{Tr}(Y)\,|\chi^{ss}\rangle\langle\chi^{ss}|)\chi^{ss}\rangle = \mathrm{Tr}(Y).
\end{aligned}$$

We used the fact that $\mathcal{B}_0(|\chi^{ss}\rangle\langle\chi^{ss}|) = |\chi^{ss}\rangle\langle\chi^{ss}|$ and that $\tilde{T}_*\tilde{\mathcal{R}}_* = \tilde{\mathcal{R}}_* - \mathrm{Id}$. Finally,

$$\mathrm{Tr}(Y) = \|(\tilde{K}_1(\mathbf{1}-\tilde{K}_0)^{-1}\dot{\tilde{K}}_0 + \dot{\tilde{K}}_1)\chi^{ss}\|^2$$

□

6.12 PROOF OF COROLLARY 3

Proof of Corollary 3. We will use the expression of $\sum_\alpha \lambda_\alpha$ given by

$$-\Re(\langle\chi^{ss}, 2\dot{\tilde{K}}_0(\mathbf{1}-\tilde{K}_0)^{-1}\dot{\tilde{K}}_0 + \ddot{\tilde{K}}_0\chi^{ss}\rangle)$$

as states in equation (6.21) in Theorem 20. The proof of this identity can be found in section 6.13; the expression we use here has the advantage that it immediately shows that $\sum_\alpha \lambda_\alpha$ does not change for different choices of the postprocessing. Before proceeding, we recall the expression of all the terms appearing in the previous equation using quantities of the dynamics of the system alone.

- $|\chi^{ss}\rangle = \sum_{i=1}^d \sqrt{\lambda_i}\,|\varphi_i\rangle_S \otimes |\varphi_i\rangle_A$, where $\rho^{ss} = \sum_{i=1}^d \lambda_i\,|\varphi_i\rangle\langle\varphi_i|$ is the spectral resolution of the stationary state of the system dynamics;
- The Kraus operator $\tilde{K}_0$ is uniquely determined as follows (left tensor is the system)

$$\tilde{K}_0 = \sum_{k=0}^{1}\sum_{i,j=1}^{d}\sqrt{\frac{\lambda_j}{\lambda_i}}\langle\varphi_j, K_k^*\varphi_i\rangle K_k \otimes |\varphi_i\rangle\langle\varphi_j|\,. \tag{6.35}$$

- The derivative of $\dot{\tilde{K}}_0$ is (where we keep in mind than only the system unitary depends on θ)

$$\dot{\tilde{K}}_0 = \sum_{k=0}^{1}\sum_{i,j=1}^{d}\sqrt{\frac{\lambda_j}{\lambda_i}}\langle\varphi_j, K_k^*\varphi_i\rangle \dot{K}_k \otimes |\varphi_i\rangle\langle\varphi_j|\,,$$

- The second derivative is

$$\ddot{\tilde{K}}_0 = \sum_{k=0}^{1}\sum_{i,j=1}^{d}\sqrt{\frac{\lambda_j}{\lambda_i}}\langle\varphi_j, K_k^*\varphi_i\rangle \ddot{K}_k \otimes |\varphi_i\rangle\langle\varphi_j|\,.$$

Note that

$$\begin{aligned}
\langle\chi^{ss}, \ddot{\tilde{K}}_0\chi^{ss}\rangle &= \sum_{k=0}^{1}\sum_{i,j=1}^{d}\lambda_j\langle\varphi_i, \ddot{K}_k\varphi_j\rangle_S\langle\varphi_j, K_k^*\varphi_i\rangle_S \\
&= \sum_{k=0}^{1}\mathrm{Tr}(\ddot{K}_k\rho^{ss}K_k^*).
\end{aligned}$$

Hence, using that $\sum_{k=0}^{1} \ddot{K}_k^* K_k + K_k^* \ddot{K}_k + 2\dot{K}_k^* \dot{K}_k = 0$, one has

$$-\Re(\langle \chi^{ss}, \ddot{\tilde{K}}_0 \chi^{ss} \rangle) = \sum_{k=0} \mathrm{Tr}(\rho^{ss} \dot{K}_k^* \dot{K}_k).$$

Let us consider the rest of the total intensity: using that $(\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 |\chi^{ss}\rangle = \sum_{l=0}^{+\infty} \tilde{K}_0^l \dot{\tilde{K}}_0 |\chi^{ss}\rangle$, one gets

$$\begin{aligned}
&\langle \chi^{ss}, \dot{\tilde{K}}_0 (\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 \chi^{ss} \rangle \\
&= \sum_{i,j=1}^{d} \sum_{a,b=0}^{1} \sum_{l=0}^{+\infty} \sum_{k_1,\dots k_l=0} \lambda_j \langle \varphi_i, \dot{K}_a K_{k_l} \cdots K_{k_1} \dot{K}_b \varphi_j \rangle \cdot \\
&\cdot \langle \varphi_j, K_b^* K_{k_1}^* \cdots K_{k_l}^* K_a^* \varphi_i \rangle \\
&= \sum_{a,b=0}^{1} \sum_{l=0}^{+\infty} \sum_{k_1,\dots k_l=0}^{1} \mathrm{Tr}(\dot{K}_a K_{k_l} \cdots K_{k_1} \dot{K}_b \rho^{ss} K_b^* K_{k_1}^* \cdots K_{k_l}^* K_a^*) \\
&= \mathrm{Tr}(\sum_{a=0}^{1} K_a^* \dot{K}_a \mathcal{R} \left(\sum_{b=0}^{1} \dot{K}_b \rho^{ss} K_b^* \right)).
\end{aligned}$$

Therefore,

$$\begin{aligned}
&-2\Re(\langle \chi^{ss}, \dot{\tilde{K}}_0 (\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 \chi^{ss} \rangle) \\
&= 2\mathrm{Tr}(\Im \left(\sum_{a=0}^{1} K_a^* \dot{K}_a \right) \mathcal{R} \left(\Im \left(\sum_{b=0}^{1} \dot{K}_b \rho^{ss} K_b^* \right) \right)).
\end{aligned}$$

□

6.13 PROOF OF THEOREM 20 AND PROPOSITION 9

Proof of Theorem 20 and Proposition 9. First of all, let us show that λ_{tot} given by Eq. (6.21) is equal to $\sum_\alpha \lambda_\alpha$, whose expression is given by equation (6.18). We have

$$\begin{aligned}
&\|(\tilde{K}_1 (\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 + \dot{\tilde{K}}_1) \chi^{ss}\|^2 = \\
&\langle (\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 \chi^{ss}, \tilde{K}_1^* \tilde{K}_1 (\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 \chi^{ss} \rangle \\
&+ \langle (\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 \chi^{ss}, \dot{\tilde{K}}_1^* \dot{\tilde{K}}_1 \chi^{ss} \rangle \\
&+ \langle \chi^{ss}, \dot{\tilde{K}}_1^* \tilde{K}_1 (\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 \chi^{ss} \rangle \\
&+ \langle \chi^{ss}, \dot{\tilde{K}}_1^* \dot{\tilde{K}}_1 \chi^{ss} \rangle.
\end{aligned}$$

Note that

$$\begin{aligned}
&\langle \chi^{ss}, \dot{\tilde{K}}_1^* \dot{\tilde{K}}_1 \chi^{ss} \rangle = \frac{1}{2} \langle \chi^{ss}, \ddot{\mathcal{B}}_1(\mathbf{1}) \chi^{ss} \rangle = -\frac{1}{2} \langle \chi^{ss}, \ddot{\mathcal{B}}_0(\mathbf{1}) \chi^{ss} \rangle \\
&= -\Re(\langle \chi^{ss}, \ddot{\tilde{K}}_0 \chi^{ss} \rangle) - \|\dot{\tilde{K}}_0 \chi^{ss}\|^2.
\end{aligned}$$

Moreover, since $\tilde{K}_0^* \tilde{K}_0 + \tilde{K}_1^* \tilde{K}_1 = \mathbf{1}$,

$$\begin{aligned}
&\langle (\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 \chi^{ss}, \tilde{K}_1^* \tilde{K}_1 (\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 \chi^{ss} \rangle \\
&= \|(\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 \chi^{ss}\|^2 - \|\tilde{K}_0 (\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 \chi^{ss}\|^2.
\end{aligned}$$

As for the other terms, one has

$$\begin{aligned}
&\langle(\mathbf{1}-\tilde{K}_0)^{-1}\dot{\tilde{K}}_0\chi^{\mathrm{ss}},\tilde{K}_1^*\dot{\tilde{K}}_1\chi^{\mathrm{ss}}\rangle\\
&=\langle(\mathbf{1}-\tilde{K}_0)^{-1}\dot{\tilde{K}}_0\chi^{\mathrm{ss}},\dot{\mathcal{B}}_1(\mathbf{1})\chi^{\mathrm{ss}}\rangle\\
&=-\langle(\mathbf{1}-\tilde{K}_0)^{-1}\dot{\tilde{K}}_0\chi^{\mathrm{ss}},\dot{\mathcal{B}}_0(\mathbf{1})\chi^{\mathrm{ss}}\rangle\\
&=-\langle\dot{\tilde{K}}_0(\mathbf{1}-\tilde{K}_0)^{-1}\dot{\tilde{K}}_0\chi^{\mathrm{ss}},\chi^{\mathrm{ss}}\rangle\\
&-\langle\tilde{K}_0(\mathbf{1}-\tilde{K}_0)^{-1}\dot{\tilde{K}}_0\chi^{\mathrm{ss}},\dot{\tilde{K}}_0\chi^{\mathrm{ss}}\rangle
\end{aligned}$$

and analogously

$$\begin{aligned}
&\langle\chi^{\mathrm{ss}},\dot{\tilde{K}}_1^*\tilde{K}_1(\mathbf{1}-\tilde{K}_0)^{-1}\dot{\tilde{K}}_0\chi^{\mathrm{ss}}\rangle\\
&=-\langle\chi^{\mathrm{ss}},\dot{\tilde{K}}_0(\mathbf{1}-\tilde{K}_0)^{-1}\dot{\tilde{K}}_0\chi^{\mathrm{ss}}\rangle\\
&-\langle\dot{\tilde{K}}_0\chi^{\mathrm{ss}},\tilde{K}_0(\mathbf{1}-\tilde{K}_0)^{-1}\dot{\tilde{K}}_0\chi^{\mathrm{ss}}\rangle.
\end{aligned}$$

Putting everything together, one gets

$$\begin{aligned}
&\sum_\alpha\lambda_\alpha=\|(\tilde{K}_1(\mathbf{1}-\tilde{K}_0)^{-1}\dot{\tilde{K}}_0+\dot{\tilde{K}}_1)\chi^{\mathrm{ss}}\|^2\\
&=\|(\mathbf{1}-\tilde{K}_0)^{-1}\dot{\tilde{K}}_0\chi^{\mathrm{ss}}\|^2-\|\tilde{K}_0(\mathbf{1}-\tilde{K}_0)^{-1}\dot{\tilde{K}}_0\chi^{\mathrm{ss}}\|^2\\
&-2\Re(\langle\dot{\tilde{K}}_0(\mathbf{1}-\tilde{K}_0)^{-1}\dot{\tilde{K}}_0\chi^{\mathrm{ss}},\chi^{\mathrm{ss}}\rangle)\\
&-2\Re(\langle\tilde{K}_0(\mathbf{1}-\tilde{K}_0)^{-1}\dot{\tilde{K}}_0\chi^{\mathrm{ss}},\dot{\tilde{K}}_0\chi^{\mathrm{ss}}\rangle)\\
&-\Re(\langle\chi^{\mathrm{ss}},\ddot{\tilde{K}}_0\chi^{\mathrm{ss}}\rangle)-\|\dot{\tilde{K}}_0\chi^{\mathrm{ss}}\|^2\\
&=\lambda_{\mathrm{tot}}+\|(\mathbf{1}-\tilde{K}_0)^{-1}\dot{\tilde{K}}_0\chi^{\mathrm{ss}}\|^2\\
&-\|(\tilde{K}_0(\mathbf{1}-\tilde{K}_0)^{-1}+\mathbf{1})\dot{\tilde{K}}_0\chi^{\mathrm{ss}}\|^2=\lambda_{\mathrm{tot}}.
\end{aligned}$$

Let us now prove the first part of the theorem. What we are actually going to show is that for $n\to+\infty$ and $|u|\leq n^{\epsilon'}$, one has

$$e^{\lambda_{tot}u^2}\nu_{u,n}\left(B_{\mathbf{m}}(n)\right)\asymp\prod_{i=1}^{k}\frac{(\lambda_{\alpha^{(i)}}u^2)^{m_{\alpha^{(i)}}}}{m_{\alpha^{(i)}}!}.$$

Let us consider a fixed *ordered* sequence of excitation patterns $\alpha^{(1)},\dots,\alpha^{(k)}$ (here we do not require them to be distinct). For an observation time n big enough, the probability of observing such a sequence of patterns separated one from the other by more than n^γ consecutive 0s is given by

$$\begin{aligned}
\sum_{\substack{x_1+\cdots+x_{k+1}=n-K\\ x_2,\dots,x_k\geq n^\gamma}}\quad&\tilde{\rho}_u^{\mathrm{ss}}\Big(\mathcal{B}_{u,0}^{x_1}\mathcal{B}_{u,\alpha^{(1)}}\mathcal{B}_{u,0}^{x_2}\cdots\\
&\cdots\mathcal{B}_{u,0}^{x_k}\mathcal{B}_{u,\alpha^{(k)}}\mathcal{B}_{u,0}^{x_{k+1}}(\mathbf{1})\Big)\Big).
\end{aligned}\tag{6.36}$$

where $K=\sum_{i=1}^k|\alpha_i|$ and $\mathcal{B}_{u,\alpha}(x):=\mathcal{B}_{u,\alpha_1}\cdots\mathcal{B}_{u,\alpha_{|\alpha|}}(x)$.

The rest of the proof follows a similar line as the proof of Theorem 19: we study the Taylor expansion of the series and identify the leading terms in the limit $n\to+\infty$. We will often use the spectral decomposition of $\mathcal{B}_0$, i.e.,

$$\mathcal{B}_0(x)=\tilde{\rho}^{\mathrm{ss}}(x)\,|\chi^{\mathrm{ss}}\rangle\langle\chi^{\mathrm{ss}}|+\mathcal{E}_0(x)$$

such that

- for any $k \geq 1$, for some constants $C > 0$, $0 < \lambda < 1$, one has $\|\mathcal{E}_0^k\| \leq C\lambda^k$,
- $\tilde{\rho}^{ss}(\mathcal{E}_0(\cdot)) = 0$ and $\mathcal{E}_0(|\chi^{ss}\rangle\langle\chi^{ss}|) = 0$.

Notice that, in general, $\tilde{\rho}^{ss}(u)$ and the eigenvector of $\mathcal{B}_{u,0*}$ corresponding to the spectral radius only coincide for $u = 0$.

We will also use the following identities

- $\tilde{\rho}^{ss}(\dot{\tilde{T}}\tilde{\mathcal{R}}\mathcal{B}_0^x\mathcal{B}_\alpha(\cdot)) = 0$ follows from (6.33) and the fact that $\tilde{K}_1|\chi^{ss}\rangle = 0$.
- $\tilde{\rho}^{ss}(\dot{\mathcal{B}}_\alpha(\cdot)) = 0$ follows from $\tilde{K}_1|\chi^{ss}\rangle = 0$
- $\tilde{\rho}^{ss}(\dot{\mathcal{B}}_0(\cdot)) = 0$ follows from the "gauge condition" (6.6) and using the explicit expression (6.35).

Significant terms in the Taylor approximation (up to the $2k$-th term). We will show that up to derivatives of order $2k$ the only contribution in the Taylor expansion of (6.36) which does not decay with n is that coming from the part of the order $2k$ derivative in which one takes the second order derivative to each of the blocks $\mathcal{B}_{u,0}^{x_i}\mathcal{B}_{u,\alpha^{(i)}}$, for $i = 1, \ldots k$. This follows from the observations below:

1) the first derivative of (6.36) at $u = 0$ is zero since $\tilde{\rho}^{ss}(\mathcal{B}_\alpha(\cdot)) = 0$ and the first derivative of $\tilde{\rho}_u^{ss}(\mathcal{B}_{u,0}^x\mathcal{B}_{u,\alpha}(\cdot))$ at $u = 0$ is equal to

$$\tilde{\rho}^{ss}(\dot{\tilde{T}}\tilde{\mathcal{R}}\mathcal{B}_0^x\mathcal{B}_\alpha(\cdot)) + \sum_{l=0}^{x-1}\tilde{\rho}^{ss}(\dot{\mathcal{B}}_0\mathcal{B}_0^l\mathcal{B}_\alpha(\cdot)) + \tilde{\rho}^{ss}(\dot{\mathcal{B}}_\alpha(\cdot))$$

and the three terms have been shown to be equal to zero above.

2) Each block of the type $\mathcal{B}_{u,0}^x\mathcal{B}_{u,\alpha}$ needs to be differentiated twice. Indeed if it is not differentiated at all, then it will bring an exponential decaying contribution since

$$\sum_{n^\gamma \leq x \leq n}\|\mathcal{B}_0^x\mathcal{B}_\alpha\| \leq C\sum_{n^\gamma \leq x \leq n}\lambda^x = \frac{C(\lambda^{n^\gamma} - \lambda^{n+1})}{1-\lambda} \to 0;$$

This follows from the fact that since α contains at least one 1, we have $\tilde{\rho}^{ss}(\mathcal{B}_\alpha(\cdot)) = 0$ which means that $\mathcal{B}_\alpha(\cdot)$ belongs to the subspace on which $\mathcal{B}_0$ acts as the strict contraction $\mathcal{R}_0$.

Alternatively, if we take a first order derivative we get a vanishing contribution since

$$\sum_{n^\gamma \leq x \leq n}\|\mathcal{B}_0^x\dot{\mathcal{B}}_\alpha\| \leq C\sum_{n^\gamma \leq x \leq n}\lambda^x$$

which decays as the term in the previous equation and

$$\sum_{n^\gamma \leq k < x \leq n}\|\mathcal{B}_0^{x-k}\dot{\mathcal{B}}_0\mathcal{B}_0^{k-1}\mathcal{B}_\alpha\|$$
$$\leq C\sum_{n^\gamma \leq k < x \leq n}\lambda^x \asymp Cn^\gamma\lambda^{n^\gamma} \to_{n\to+\infty} 0.$$

On the other hand, by differentiating a block of the form $\mathcal{B}_{u,0}^x\mathcal{B}_{u,\alpha}$ twice, one stops the exponential decay and obtains a linear growth in n; however, we need to remember that every time we use a derivative, everything gets multiplied by $u/\sqrt{n}$, which is of the order $n^{-1/2+\epsilon'}$. Therefore, if we consider the terms in the Taylor expansion up to order $2k$, the only one that does not decay to 0 is the one

with $2k$ derivatives where we use two of them on the block $\tilde{\rho}^{\mathrm{ss}}(\mathcal{B}^{x_1}_{u,0}\mathcal{B}_{u,\alpha^{(1)}}(\cdot))$ at the beginning and other two on each following block of the form $\mathcal{B}^{x_i}_{u,0}\mathcal{B}_{u,\alpha^{(i)}}(\cdot)$. Any other term where we have less derivatives involved or where we spend them in a different way is either 0 or decays at least as (for $k \geq 2$)

$$(n^{\frac{1}{2}+\gamma}\lambda^{n^\gamma})n^{2k\epsilon'}$$

uniformly in u for $|u| \leq n^{\epsilon'}$; indeed, for any derivative used in a different way (they cannot be more than $2(k-1)$ because two of them need to be used for the first block), one gains a growth of $n^{1/2}$, but suffers a decay of at least $n^\gamma\lambda^{n^\gamma}$.

We now focus on the leading (order $2k$) term of the Taylor expansion. The second derivative of a block of the type $\mathcal{B}^x_{u,0}\mathcal{B}_{u,\alpha}$ looks like

$$\begin{aligned}
&\mathcal{B}_0^x\ddot{\mathcal{B}}_\alpha + \sum_{k=1}^{x}\mathcal{B}_0^{k-1}\ddot{\mathcal{B}}_0\mathcal{B}^{x-k}\mathcal{B}_\alpha\\
&+2\sum_{k=1}^{x}\mathcal{B}_0^{k-1}\dot{\mathcal{B}}_0\mathcal{B}_0^{x-k}\dot{\mathcal{B}}_\alpha\\
&+2\sum_{1\leq k<s\leq x}\mathcal{B}_0^{k-1}\dot{\mathcal{B}}_0\mathcal{B}_0^{s-k-1}\dot{\mathcal{B}}_0\mathcal{B}_0^{x-s}\mathcal{B}_\alpha
\end{aligned}$$

For large n this becomes

$$\begin{aligned}
&\tilde{\rho}^{\mathrm{ss}}\left[(\ddot{\mathcal{B}}_\alpha + 2\dot{\mathcal{B}}_0\mathcal{R}_0\dot{\mathcal{B}}_\alpha)(\cdot)\right]|\chi^{\mathrm{ss}}\rangle\langle\chi^{\mathrm{ss}}|\\
&+\tilde{\rho}^{\mathrm{ss}}\left[(\ddot{\mathcal{B}}_0 + 2\dot{\mathcal{B}}_0\mathcal{R}_0\dot{\mathcal{B}}_0)\mathcal{R}_0\mathcal{B}_\alpha)(\cdot)\right]|\chi^{\mathrm{ss}}\rangle\langle\chi^{\mathrm{ss}}|
\end{aligned}$$

where $\mathcal{R}_0$ is the Moore-Penrose inverse of $\mathrm{Id}-\mathcal{B}_0$, and we have used the spectral decomposition of $\mathcal{B}_0$. Moreover, the rightmost term $B_0^{x_{k+1}}(\mathbf{1})$ (which is not differentiated) converges to $|\chi^{\mathrm{ss}}\rangle\langle\chi^{\mathrm{ss}}|$. This means that the full $2k$ derivative of (6.36) becomes

$$\frac{u^{2k}}{k!}\prod_{i=1}^{k}\frac{1}{2}\tilde{\rho}^{\mathrm{ss}}\left[\mathcal{B}^{(2)}_{\alpha^{(i)}}(|\chi^{\mathrm{ss}}\rangle\langle\chi^{\mathrm{ss}}|)\right]$$

where

$$\mathcal{B}^{(2)}_\alpha := \ddot{\mathcal{B}}_\alpha + 2\dot{\mathcal{B}}_0\mathcal{R}_0\dot{\mathcal{B}}_\alpha + (\ddot{\mathcal{B}}_0 + 2\dot{\mathcal{B}}_0\mathcal{R}_0\dot{\mathcal{B}}_0)\mathcal{R}_0\mathcal{B}_\alpha.$$

The number of terms with two derivatives in each block is equal to $(2k)!/2^k$ (it is the same as the number of terms of the form $2^k = \left(\frac{d^2}{dx}{}^2(x^2)\right)^k$ in the $2k$-th derivative of $(x^2)^k$): $(2k)!$ simplifies with the one coming from the Taylor expansion, while the factor 2^{-k} can be distributed to each factor. The expression of the factors and the upper bound on the error can be obtained differentiating and using the spectral decomposition of $\mathcal{B}_0$; the $1/k!$ in front comes from the sum: indeed one can see that

$$\sum_{\substack{x_1+\cdots+x_{k+1}=n-K\\ x_2,\ldots,x_k\geq n^\gamma}} \asymp \sum_{x_1+\cdots+x_{k+1}=n} \asymp \frac{n^k}{k!},$$

which is the number of ways one can choose k numbers out of n.

The first term of the form

$$\frac{1}{2}\tilde{\rho}^{ss}(\ddot{\mathcal{B}}_\alpha(|\chi^{ss}\rangle\langle\chi^{ss}|)) = |\langle\chi^{ss}|\tilde{K}_{\alpha_{|\alpha|}}\cdots\tilde{K}_{\alpha_2}\dot{\tilde{K}}_1\chi^{ss}\rangle|^2.$$

This follows from the fact that $K_1|\chi^{ss}\rangle = 0$, so the derivatives need to be applied the first K_1 terms of $\mathcal{B}_\alpha$. For the second term, one gets

$$\begin{aligned}&\tilde{\rho}^{ss}(\dot{\mathcal{B}}_0\mathcal{R}_0\dot{\mathcal{B}}_\alpha(|\chi^{ss}\rangle\langle\chi^{ss}|)) = \\ &2\Re\Big(\langle\tilde{K}_{\alpha_{|\alpha|}}\cdots\tilde{K}_{\alpha_2}\dot{\tilde{K}}_1\chi^{ss}|\chi^{ss}\rangle\times \\ &\langle\chi^{ss}|\tilde{K}_{\alpha_{|\alpha|}}\cdots\tilde{K}_{\alpha_1}(\mathbf{1}-\tilde{K}_0)^{-1}\dot{\tilde{K}}_0\chi^{ss}\rangle\Big).\end{aligned}$$

Finally, for the last term, below we will show that

$$\begin{aligned}&\frac{1}{2}\tilde{\rho}^{ss}((\ddot{\mathcal{B}}_0+2\dot{\mathcal{B}}_0\mathcal{R}_0\dot{\mathcal{B}}_0)\mathcal{R}_0\mathcal{B}_\alpha(|\chi^{ss}\rangle\langle\chi^{ss}|)) = \\ &|\langle\chi^{ss}|\tilde{K}_{\alpha_{|\alpha|}}\cdots\tilde{K}_{\alpha_1}(\mathbf{1}-\tilde{K}_0)^{-1}\dot{\tilde{K}}_0\chi^{ss}\rangle|^2.\end{aligned} \tag{6.37}$$

Note that the sum of the three terms is equal to $|\mu_\alpha|^2$ where μ_α is given in equation (6.17) in Lemma 8.

In conclusion, we obtained that for large n, the probability of any sequence of n outcomes showing the ordered sequence of excitation patterns given by $\alpha^{(1)},\ldots,\alpha^{(k)}$ is asymptotically equivalent to

$$\frac{1}{k!}\prod_{i=1}^{k}(\lambda_{\alpha^{(i)}}u^2) \tag{6.38}$$

plus a reminder coming from neglecting the terms of order bigger than $2k$ in the Taylor expansion. If we are able to show that the reminder is negligible compared to the term in Eq. (6.38), then we can prove the statement in Eq. (6.20). Indeed, suppose that the sequence we are analysing belongs to $B_{\mathbf{m}}$; then we can partition $B_{\mathbf{m}}$ into

$$\frac{k!}{\prod_{i=1}^{k} m_{\alpha^{(i)}}!}$$

disjoint subsets containing the excitation patters in $(\alpha^{(1)},m_{\alpha^{(1)}}),\ldots,(\alpha^{(k)},m_{\alpha^{(k)}})$ in a fixed order and whose probability asymptotically behaves as (as we just showed)

$$\frac{1}{k!}\prod_{i=1}^{k}(u^2\lambda_{\alpha^{(i)}})^{m_{\alpha^{(i)}}}.$$

We now prove (6.37). Notice that, since

$$\begin{aligned}&|\langle\chi^{ss}|\tilde{K}_{\alpha_{|\alpha|}}\cdots\tilde{K}_{\alpha_1}(\mathbf{1}-\tilde{K}_0)^{-1}\dot{\tilde{K}}_0\chi^{ss}\rangle|^2 = \\ &\langle(\mathbf{1}-\tilde{K}_0)^{-1}\dot{\tilde{K}}_0\chi^{ss}|\mathcal{B}_\alpha(|\chi^{ss}\rangle\langle\chi^{ss}|)|(\mathbf{1}-\tilde{K}_0)^{-1}\dot{\tilde{K}}_0\chi^{ss}\rangle,\end{aligned}$$

we need to prove that

$$\begin{aligned}&\mathcal{R}_{0*}(\ddot{\mathcal{B}}_{0*}+2\dot{\mathcal{B}}_{0*}\mathcal{R}_{0*}\dot{\mathcal{B}}_{0*})(\tilde{\rho}^{ss})) = \\ &2\,|(\mathbf{1}-\tilde{K}_0)^{-1}\dot{\tilde{K}}_0\chi^{ss}\rangle\langle(\mathbf{1}-\tilde{K}_0)^{-1}\dot{\tilde{K}}_0\chi^{ss}| + \text{rem},\end{aligned}$$

where by $\mathcal{R}_{0*}$ we mean the Moore-Penrose inverse and "rem" is a term which is gives 0 when evaluated against $\mathcal{B}_\alpha(|\chi^{ss}\rangle\langle\chi^{ss}|)$. Equivalently,

$$(\ddot{\mathcal{B}}_{0*} + 2\dot{\mathcal{B}}_{0*}\mathcal{R}_{0*}\dot{\mathcal{B}}_{0*})(\tilde{\rho}^{ss}) =$$
$$2(\mathrm{Id} - \mathcal{B}_{0*})(|(\mathbf{1}-\tilde{K}_0)^{-1}\dot{\tilde{K}}_0\chi^{ss}\rangle\langle(\mathbf{1}-\tilde{K}_0)^{-1}\dot{\tilde{K}}_0\chi^{ss}|) + \mathrm{rem}',$$

where $\mathcal{R}_{0*}(\mathrm{rem}') = \mathrm{rem}$.

By explicit computations, one can see that

$$(\mathrm{Id} - \mathcal{B}_{0*})(|(\mathbf{1}-\tilde{K}_0)^{-1}\dot{\tilde{K}}_0\chi^{ss}\rangle\langle(\mathbf{1}-\tilde{K}_0)^{-1}\dot{\tilde{K}}_0\chi^{ss}|) =$$
$$|(\mathbf{1}-\tilde{K}_0)^{-1}\dot{\tilde{K}}_0\chi^{ss}\rangle\langle(\mathbf{1}-\tilde{K}_0)^{-1}\dot{\tilde{K}}_0\chi^{ss}| -$$
$$|\tilde{K}_0(\mathbf{1}-\tilde{K}_0)^{-1}\dot{\tilde{K}}_0\chi^{ss}\rangle\langle\tilde{K}_0(\mathbf{1}-\tilde{K}_0)^{-1}\dot{\tilde{K}}_0\chi^{ss}|.$$

Using that $\tilde{K}_0(\mathbf{1}-\tilde{K}_0)^{-1} = (\mathbf{1}-\tilde{K}_0)^{-1} - \mathbf{1}$, one gets

$$(\mathrm{Id} - \mathcal{B}_{0*})(|(\mathbf{1}-\tilde{K}_0)^{-1}\dot{\tilde{K}}_0\chi^{ss}\rangle\langle(\mathbf{1}-\tilde{K}_0)^{-1}\dot{\tilde{K}}_0\chi^{ss}|) =$$
$$|(\mathbf{1}-\tilde{K}_0)^{-1}\dot{\tilde{K}}_0\chi^{ss}\rangle\langle\dot{\tilde{K}}_0\chi^{ss}| + |\dot{\tilde{K}}_0\chi^{ss}\rangle\langle(\mathbf{1}-\tilde{K}_0)^{-1}\dot{\tilde{K}}_0\chi^{ss}| -$$
$$|\dot{\tilde{K}}_0\chi^{ss}\rangle\langle\dot{\tilde{K}}_0\chi^{ss}|.$$

On the other hand,

$$\begin{aligned}
&(\ddot{\mathcal{B}}_{0*} + 2\dot{\mathcal{B}}_{0*}\mathcal{R}_{0*}\dot{\mathcal{B}}_{0*})(\rho)\\
&= |\ddot{\tilde{K}}_0\chi^{ss}\rangle\langle\chi^{ss}| + 2\,|\dot{\tilde{K}}_0\chi^{ss}\rangle\langle\dot{\tilde{K}}_0\chi^{ss}| + |\chi^{ss}\rangle\langle\ddot{\tilde{K}}_0\chi^{ss}|\\
&+2\,|\dot{\tilde{K}}_0(\mathbf{1}-\tilde{K}_0)^{-1}\dot{\tilde{K}}_0\chi^{ss}\rangle\langle\chi^{ss}|\\
&+2\,|\chi^{ss}\rangle\langle\dot{\tilde{K}}_0(\mathbf{1}-\tilde{K}_0)^{-1}\dot{\tilde{K}}_0\chi^{ss}|\\
&+2\,|\tilde{K}_0(\mathbf{1}-\tilde{K}_0)^{-1}\dot{\tilde{K}}_0\chi^{ss}\rangle\langle\dot{\tilde{K}}_0\chi^{ss}|\\
&+2\,|\dot{\tilde{K}}_0\chi^{ss}\rangle\langle\dot{\tilde{K}}_0(\mathbf{1}-\tilde{K}_0)^{-1}\dot{\tilde{K}}_0\chi^{ss}|\\
&= |(\ddot{\tilde{K}}_0 + 2\dot{\tilde{K}}_0(\mathbf{1}-\tilde{K}_0)^{-1}\dot{\tilde{K}}_0)\chi^{ss}\rangle\langle\chi^{ss}|\\
&+ |\chi^{ss}\rangle\langle(\ddot{\tilde{K}}_0 + 2\dot{\tilde{K}}_0(\mathbf{1}-\tilde{K}_0)^{-1}\dot{\tilde{K}}_0)\chi^{ss}|\\
&+2\,|(\mathbf{1}-\tilde{K}_0)^{-1}\dot{\tilde{K}}_0\chi^{ss}\rangle\langle\dot{\tilde{K}}_0.\chi^{ss}|\\
&+2\,|\dot{\tilde{K}}_0\chi^{ss}\rangle\langle(\mathbf{1}-\tilde{K}_0)^{-1}\dot{\tilde{K}}_0\chi^{ss}| - 2\,|\dot{\tilde{K}}_0\chi^{ss}\rangle\langle\dot{\tilde{K}}_0\chi^{ss}|.
\end{aligned}$$

Note that the last two lines are exactly equal to $2(\mathrm{Id} - \mathcal{B}_{0*})(|(\mathbf{1}-\tilde{K}_0)^{-1}\dot{\tilde{K}}_0\chi^{ss}\rangle\langle(\mathbf{1}-\tilde{K}_0)^{-1}\dot{\tilde{K}}_0\chi^{ss}|)$. Let us look at the remaining part:

$$\begin{aligned}
\mathrm{rem}' = &|(\ddot{\tilde{K}}_0 + 2\dot{\tilde{K}}_0(\mathbf{1}-\tilde{K}_0)^{-1}\dot{\tilde{K}}_0)\chi^{ss}\rangle\langle\chi^{ss}| +\\
&|\chi^{ss}\rangle\langle(\ddot{\tilde{K}}_0 + 2\dot{\tilde{K}}_0(\mathbf{1}-\tilde{K}_0)^{-1}\dot{\tilde{K}}_0)\chi^{ss}|.
\end{aligned}$$

one can easily see that

$$\mathcal{R}_{0*}(\mathrm{rem}') =$$
$$|(\mathbf{1}-\tilde{K}_0)^{-1}(\ddot{\tilde{K}}_0 + 2\dot{\tilde{K}}_0(\mathbf{1}-\tilde{K}_0)^{-1}\dot{\tilde{K}}_0)\chi^{ss}\rangle\langle\chi^{ss}|$$
$$+ |\chi^{ss}\rangle\langle(\mathbf{1}-\tilde{K}_0)^{-1}(\ddot{\tilde{K}}_0 + 2\dot{\tilde{K}}_0(\mathbf{1}-\tilde{K}_0)^{-1}\dot{\tilde{K}}_0)\chi^{ss}|.$$

In the previous Eq. we used $(\mathbf{1}-K_0)^{-1}$ for the Moore-Penrose inverse: in general, $|(\ddot{\tilde{K}}_0 + 2\dot{\tilde{K}}_0(\mathbf{1}-\tilde{K}_0)^{-1}\dot{\tilde{K}}_0)\chi^{ss}\rangle$ is not orthogonal to $|\chi\rangle^{ss}$. It is now clear that $\mathrm{Tr}(())\mathcal{R}_{0*}(\mathrm{rem}')\mathcal{B}_\alpha(|\chi^{ss}\rangle\langle\chi^{ss}|) = 0$ which proves (6.37).

Remainder. Now, we need to take care of the reminder: it is enough to show that the following expression is $o(n^{2k\epsilon})$:

$$\frac{u^{2k+1}}{n^{k+1/2}} \sum_{\substack{x_1+\cdots+x_{k+1}=n-K \\ x_2,\ldots,x_k\geq n^\gamma}} \frac{d^{2k+1}}{u^{2k+1}} \tilde{\rho}^{ss}\left(\frac{u}{\sqrt{n}}\right)\left(\tilde{\mathcal{B}}_{0,u}^{x_1}\tilde{\mathcal{B}}_{\alpha_1,u}\tilde{\mathcal{B}}_{0,u}^{x_2}\cdots \right. \\ \left. \cdots \tilde{\mathcal{B}}_{0,u}^{x_k}\tilde{\mathcal{B}}_{\alpha_k,u}\tilde{\mathcal{B}}_{0,u}^{x_{k+1}}(\mathbf{1})\right)\Big|_{u=\eta} \tag{6.39}$$

for any $|\eta| \leq u/\sqrt{n}$, where

$$\tilde{\mathcal{B}}_{\alpha,u} = e^{\lambda_{tot}|\alpha|u^2/n}\mathcal{B}_{\alpha,u}$$

for any string α.

Let us first point out some properties of the maps $\tilde{\mathcal{B}}_{\alpha,u}$:

- for u small enough,

  $$\tilde{\mathcal{B}}_{0,u}(\cdot) = a(u)l(u)(\cdot)r(u) + \tilde{\mathcal{E}}_{0,u}(\cdot)$$

  where
  - $a(u) = 1 + O(u^3/n^{3/2})$,
  - $l(u)(r(u)) \equiv 1$,
  - $l(u)(\tilde{\mathcal{E}}_{0,u}(\cdot)) = 0$,
  - $\tilde{\mathcal{E}}_{0,u}(r(u)) = 0$ and
  - $\|\tilde{\mathcal{E}}_{0,u}^k\| \leq C\lambda^k$ for some $C \geq 0$ and $0 < \lambda < 1$.

  The order of the reminder in the expression $a(u)$ is due to the fact that $\dot{\lambda}$ is equal to the first derivative at 0 of the spectral radius of $\mathcal{B}_{0,u}$ which is equal to 0 because it attains a maximum there; the fact that $\ddot{\lambda} = 0$ as well is due to the multiplicative factor in front of $\mathcal{B}_{0,u}$ in the definition of $\tilde{\mathcal{B}}_{0,u}$.

- $\tilde{\rho}^{ss}(u/\sqrt{n})(\tilde{\mathcal{E}}_{0,u}(\cdot)) = O(u/\sqrt{n})$.

- For any excitation pattern α one has

  $$l(u)(\tilde{\mathcal{B}}_{\alpha,u}(\cdot)) = O(u^2/n),$$
  $$l(u)(\dot{\tilde{\mathcal{B}}}_{\alpha,u}(\cdot)) = O(u/\sqrt{n}).$$

  Indeed, the first derivative of $l(u)(\tilde{\mathcal{B}}_{\alpha,u}(\cdot))$ at 0 is given by

  $$\tilde{\rho}^{ss}(\dot{\mathcal{B}}_\alpha(\cdot)) + \dot{l}(\mathcal{B}_\alpha(\cdot)).$$

  That the first addend is 0 has been shown earlier, while for the second one, it is clear using $\dot{l} = \tilde{\rho}^{ss}(\dot{\mathcal{B}}_\alpha(\mathrm{Id} - \mathcal{B}_0)^{-1}(\cdot))$.

- Moreover,

  $$l(u)(\dot{\tilde{\mathcal{B}}}_{0,u}(r(u))) = O(u^2/n) \text{ and}$$
  $$l(u)((\ddot{\tilde{\mathcal{B}}}_{0,u} + 2\dot{\tilde{\mathcal{B}}}_{0,u}\tilde{\mathcal{E}}_{0,u}\dot{\tilde{\mathcal{B}}}_{0,u}(r(u))) = O(u/\sqrt{n}). \tag{6.40}$$

  This can be seen differentiating

  $$l(u)\tilde{\mathcal{B}}_{0,u}(x(u)) = a(u)$$

  and evaluating at 0.

- Finally,
$$l(u)(\dot{\tilde{\mathcal{B}}}_{0,u}\tilde{\mathcal{E}}_{0,u}\tilde{\mathcal{B}}_{\alpha,u}(\cdot)) = O(u/\sqrt{n}),$$
since $\tilde{\rho}^{ss}(\dot{\mathcal{B}}_0)(\cdot) = 0$.

Let us now study the growth of the derivatives of $\tilde{\mathcal{B}}^x_{0,u}\tilde{\mathcal{B}}_{\alpha,u}(\cdot)$ for $0 \leq x \leq n \to +\infty$:

1. 0^{th} order:
$$\begin{aligned}\tilde{\mathcal{B}}^x_{0,u}\tilde{\mathcal{B}}_{\alpha,u}(\cdot) &= a(u)l(u)(\tilde{\mathcal{B}}_{\alpha,u}(\cdot))x(u) + O(\lambda^x)\\ &= O\left(\frac{u^2}{n} + \lambda^x\right);\end{aligned}$$

2. 1^{st} order:
$$\begin{aligned}&\sum_{1\leq l\leq x}\tilde{\mathcal{B}}^{x-l}_{0,u}\dot{\tilde{\mathcal{B}}}_{0,u}\tilde{\mathcal{B}}^{l-1}_{0,u}\tilde{\mathcal{B}}_{\alpha,u}(\cdot) + \tilde{\mathcal{B}}^x_{0,u}\dot{\tilde{\mathcal{B}}}_{\alpha,u}(\cdot) =\\ &xa(u)^2l(u)(\dot{\tilde{\mathcal{B}}}_{0,u}(x(u)))l(u)(\tilde{\mathcal{B}}_{\alpha,u}(\cdot))+\\ &a(u)\tilde{\mathcal{R}}_{0,u}\dot{\tilde{\mathcal{B}}}_{0,u}(x(u)))l(u)(\tilde{\mathcal{B}}_{\alpha,u}(\cdot))+\\ &a(u)l(u)(\dot{\tilde{\mathcal{B}}}_{0,u}\tilde{\mathcal{R}}_{0,u}\tilde{\mathcal{B}}_{\alpha,u}(\cdot))x(u) + O\left(\frac{u}{\sqrt{n}} + x\lambda^x\right)\\ &= O\left(\frac{u}{\sqrt{n}} + x\lambda^x\right),\end{aligned}$$
where we used that $\epsilon < 1/6$;

We remark that in the case where the block $\tilde{\mathcal{B}}^x_{0,u}\tilde{\mathcal{B}}_{\alpha,u}(\cdot)$ is the first one, due to the action of $\tilde{\rho}^{ss}(u/\sqrt{n})$, the 0^{th}-order term becomes $O(u^2/n + u\lambda^x/\sqrt{n})$ and the 1^{st}-order one becomes $O(u/\sqrt{n})$, while if the block is not the first one, λ^x decays exponentially fast in n since $x \geq n^\gamma$.

3. 2^{nd} order:
$$\begin{aligned}&\sum_{1\leq l\leq x}\tilde{\mathcal{B}}^{x-l}_{0,u}\ddot{\tilde{\mathcal{B}}}_{0,u}\tilde{\mathcal{B}}^{l-1}_{0,u}\tilde{\mathcal{B}}_{\alpha,u}(\cdot)+\\ &2\sum_{1\leq l<k\leq x}\tilde{\mathcal{B}}^{x-k}_{0,u}\dot{\tilde{\mathcal{B}}}_{0,u}\tilde{\mathcal{B}}^{k-l-1}_{0,u}\dot{\tilde{\mathcal{B}}}_{0,u}\tilde{\mathcal{B}}^{l-1}_{0,u}\tilde{\mathcal{B}}_{\alpha,u}(\cdot)+\\ &2\sum_{1\leq l\leq x}\tilde{\mathcal{B}}^{x-l}_{0,u}\dot{\tilde{\mathcal{B}}}_{0,u}\tilde{\mathcal{B}}^{l-1}_{0,u}\dot{\tilde{\mathcal{B}}}_{\alpha,u}(\cdot)+\\ &\tilde{\mathcal{B}}^x_{0,u}\ddot{\tilde{\mathcal{B}}}_{\alpha,u}(\cdot) =\\ &2x^2a(u)^3(l(u)(\dot{\tilde{\mathcal{B}}}_{0,u}(x(u)))^2l(u)(\tilde{\mathcal{B}}_{\alpha,u}(\cdot))+\\ &xa(u)^2l(u)((\ddot{\tilde{\mathcal{B}}}_{0,u} + 2\dot{\tilde{\mathcal{B}}}_{0,u}\tilde{\mathcal{R}}_{0,u}\dot{\tilde{\mathcal{B}}}_{0,u})(x(u)))l(u)(\tilde{\mathcal{B}}_{\alpha,u}(\cdot))+\\ &2xa(u)^2l(u)(\dot{\tilde{\mathcal{B}}}_{0,u}(x(u))l(u)(\dot{\tilde{\mathcal{B}}}_{\alpha,u}(\cdot)) + O(1) = O(1);\end{aligned}$$

4. 3rd order:

$$\sum_{1\le l\le x} \tilde{\mathcal{B}}_{0,u}^{x-l}\dddot{\tilde{\mathcal{B}}}_{0,u}\tilde{\mathcal{B}}_{0,u}^{l-1}\tilde{\mathcal{B}}_{\alpha,u}(\cdot)+$$
$$3\sum_{1\le l<k\le x} \tilde{\mathcal{B}}_{0,u}^{x-k}\ddot{\tilde{\mathcal{B}}}_{0,u}\tilde{\mathcal{B}}_{0,u}^{k-l-1}\dot{\tilde{\mathcal{B}}}_{0,u}\tilde{\mathcal{B}}_{0,u}^{l-1}\tilde{\mathcal{B}}_{\alpha,u}(\cdot)+$$
$$3\sum_{1\le l<k\le x} \tilde{\mathcal{B}}_{0,u}^{x-k}\dot{\tilde{\mathcal{B}}}_{0,u}\tilde{\mathcal{B}}_{0,u}^{k-l-1}\ddot{\tilde{\mathcal{B}}}_{0,u}\tilde{\mathcal{B}}_{0,u}^{l-1}\tilde{\mathcal{B}}_{\alpha,u}(\cdot)+$$
$$6\sum_{1\le l<k<m\le x} \tilde{\mathcal{B}}_{0,u}^{x-m}\dot{\tilde{\mathcal{B}}}_{0,u}\tilde{\mathcal{B}}_{0,u}^{m-k-1}\dot{\tilde{\mathcal{B}}}_{0,u}\tilde{\mathcal{B}}_{0,u}^{k-l-1}\dot{\tilde{\mathcal{B}}}_{0,u}\tilde{\mathcal{B}}_{0,u}^{l-1}\tilde{\mathcal{B}}_{\alpha,u}(\cdot)+$$
$$3\sum_{1\le l\le x} \tilde{\mathcal{B}}_{0,u}^{x-l}\ddot{\tilde{\mathcal{B}}}_{0,u}\tilde{\mathcal{B}}_{0,u}^{l-1}\dot{\tilde{\mathcal{B}}}_{\alpha,u}(\cdot)+$$
$$6\sum_{1\le l<k\le x} \tilde{\mathcal{B}}_{0,u}^{x-k}\dot{\tilde{\mathcal{B}}}_{0,u}\tilde{\mathcal{B}}_{0,u}^{k-l-1}\dot{\tilde{\mathcal{B}}}_{0,u}\tilde{\mathcal{B}}_{0,u}^{l-1}\dot{\tilde{\mathcal{B}}}_{\alpha,u}(\cdot)+$$
$$3\sum_{1\le l\le x} \tilde{\mathcal{B}}_{0,u}^{x-l}\dot{\tilde{\mathcal{B}}}_{0,u}\tilde{\mathcal{B}}_{0,u}^{l-1}\ddot{\tilde{\mathcal{B}}}_{\alpha,u}(\cdot)+$$
$$\tilde{\mathcal{B}}_{0,u}^{x}\dddot{\tilde{\mathcal{B}}}_{\alpha,u}(\cdot) = O(n^{2\epsilon});$$

5. m^{th} order for $m \ge 4$: first, notice that

$$\frac{d^m}{du^m}\left(\tilde{\mathcal{B}}_{0,u}^{x}\tilde{\mathcal{B}}_{\alpha,u}(\cdot)\right) = \sum_{l=0}^{m}\binom{m}{l}\frac{d^l}{du^l}(\tilde{\mathcal{B}}_{0,u}^{x})\frac{d^{(m-l)}}{du^{(m-l)}}(\tilde{\mathcal{B}}_{\alpha,u}(\cdot)).$$

The term $d^l(\tilde{\mathcal{B}}_{0,u}^{x})/du^l$ is a sum over all possible ways of distributing the derivatives among the factors $\tilde{\mathcal{B}}_{0,u}$.

Then one can use the spectral decomposition of the $\tilde{\mathcal{B}}_{0,u}$ which have not been differentiated and glue together differentiated terms using $\mathcal{R}_{0,u}$ or turning them into products using the projection $l(u)(\cdot)x(u)$ as we did in the previous items. From Eq. (6.40), one can see that the blocks with a single derivative of a single term $\tilde{\mathcal{B}}_{0,u}$ bring a growth of the order $n^{2\epsilon'}$, the terms with two derivatives $(\ddot{\tilde{\mathcal{B}}}_{0,u} + 2\dot{\tilde{\mathcal{B}}}_{0,u}\tilde{\mathcal{R}}_{0,u}\dot{\tilde{\mathcal{B}}}_{0,u})$ bring a growth of the order $n^{1/2+\epsilon'}$ while all other blocks (with at least 3 elements), cause a growth of the order at most n. Notice that the *highest growth is attained by making as many groups of three derivatives as possible* and this will be used in obtaining the following estimate.

The term $d^{(m-l)}(\tilde{\mathcal{B}}_{\alpha,u})/du^{(m-l)}$ does not cause any reduction in the growth if $m-l \ge 2$, while causes a decay equal to $n^{-1/2+\epsilon'}$ if $m = l+1$ and $n^{-1+2\epsilon'}$ if $m = l$.

Therefore,

$$\left\|\frac{d^m}{du^m}\left(\tilde{\mathcal{B}}_{0,u}^{x}\tilde{\mathcal{B}}_{\alpha,u}(\cdot)\right)\right\| \lesssim n^{\max\{f(m-2),f(m-1)-\frac{1}{2}+\epsilon',f(m)-1+2\epsilon'\}}, \tag{6.41}$$

where

$$f(m) = a(m) + \left(\frac{1}{2}+\epsilon'\right)b(m) + 2\epsilon'(m-(3a(m)+2b(m)))$$

$$a(m) := \left\lfloor \frac{m}{3} \right\rfloor, \quad b(m) := \left\lfloor \frac{1}{2}\left(m - 3\left\lfloor \frac{m}{3} \right\rfloor\right) \right\rfloor.$$

$a(m)$ is the maximum number of groups with three derivatives and $b(m)$ is the maximum number of groups with two derivatives that we can make with the derivatives left. Notice that if $\epsilon' < 1/6$, then $f(m) \leq m/3$: indeed,

$$\begin{aligned}
&a(m) + \left(\frac{1}{2} + \epsilon'\right) b(m) + 2\epsilon'(m - (3a(m) + 2b(m))) = \\
&2\epsilon' m + (1 - 6\epsilon')a(m) + \left(\frac{1}{2} - 3\epsilon'\right) b(m) \leq \\
&2\epsilon' m + (1 - 6\epsilon')a(m) + \frac{1}{2}\left(\frac{1}{2} - 3\epsilon'\right)(m - 3a(m)) \leq \\
&\frac{1}{2}\left(\left(\frac{1}{2} + \epsilon'\right) m + \left(\frac{1}{2} - 3\epsilon'\right) a(m)\right) \leq \\
&\frac{1}{2}\left(\left(\frac{1}{2} + \epsilon'\right) m + \left(\frac{1}{6} - \epsilon'\right) m\right) \leq \frac{m}{3}.
\end{aligned}$$

Therefore, the growth of the term in Eq. (6.41) is upper bounded by

$$n^{\max\{\frac{m-2}{3}, \frac{m-1}{3} - \frac{1}{2} + \epsilon', \frac{m}{3} - 1 + 2\epsilon'\}}. \tag{6.42}$$

This implies that the first two derivatives spent on a block of the form $\tilde{\mathcal{B}}^x_{0,u}\tilde{\mathcal{B}}_{\alpha,u}$ cause a growth equal to $n^{1/2-\epsilon'}$, while the other $m-2$ only brings a growth at most equal to the term in Eq. (6.42). Notice that it is more convenient to spend them in a way that every block of the form $\tilde{\mathcal{B}}^x_{0,u}\tilde{\mathcal{B}}_{\alpha,u}$ has two derivatives, since one has that

$$\left(\frac{1}{2} - \epsilon'\right)(m-2) \geq \frac{m-2}{3} \Leftrightarrow m \geq 2,$$

moreover

$$\left(\frac{1}{2} - \epsilon'\right)(m-2) \geq \frac{m-1}{3} - \frac{1}{2} + \epsilon' \Leftrightarrow m \geq 1$$

and

$$\left(\frac{1}{2} - \epsilon'\right)(m-2) \geq \frac{m}{3} - 1 + 2\epsilon' \Leftrightarrow m \geq 0.$$

Therefore the growth of

$$\begin{aligned}
\sum_{\substack{x_1+\cdots+x_{k+1}=n-K \\ x_2,\ldots,x_k \geq n^\gamma}} \frac{d^{2k+1}}{u^{2k+1}} \tilde{\rho}^{\mathrm{ss}}\left(\frac{u}{\sqrt{n}}\right) \Big(\tilde{\mathcal{B}}^{x_1}_{0,u}\tilde{\mathcal{B}}_{\alpha_1,u}\tilde{\mathcal{B}}^{x_2}_{0,u}\cdots \\
\cdots \tilde{\mathcal{B}}^{x_k}_{0,u}\tilde{\mathcal{B}}_{\alpha_k,u}\tilde{\mathcal{B}}^{x_{k+1}}_{0,u}(\mathbf{1})\Big)\Big|_{u=\eta},
\end{aligned}$$

is of the order of $O(n^{2\epsilon'})$ uniformly in u and it is attained by the term where we spend at least two derivatives in each block of the form $\tilde{\mathcal{B}}^x_{0,u}\tilde{\mathcal{B}}_{\alpha,u}$ and the last derivative in any of such blocks. One can see that it is not convenient to use any derivative in the final term $\tilde{\mathcal{B}}^x_{0,u}(\mathbf{1})$, since it is better to contrast the decay induced by the other blocks. To conclude, the term in Eq. (6.39) grows at most as $n^{(2k+3)\epsilon' - 1/2} = o(n^{2k\epsilon'})$. We proved Proposition 9 and notice that Theorem 20 follows considering $\epsilon' = 0$. □

6.14 PROOF OF COROLLARY 4

Proof. Let us define the measurable space given by the set $\Omega = \mathbb{N}^{\mathbb{N}}$ together with the σ-field $\mathcal{F}$ generated by cylindrical sets; we can consider on $(\Omega, \mathcal{F})$ the law $\overline{\nu}_u$ of $\{N_\alpha : \alpha \in \mathcal{P}\}$ and the law of $\{N_\alpha(n) : \alpha \in \mathcal{P}\}$, which, with a slight abuse of notation, we still denote by $\nu_{u,n}$ as well. We know that for every finite set A of patterns, one has

$$\lim_{n\to+\infty} \sup_{|u|<C} |\nu_{u,n}(A) - \overline{\nu}_u(A)| = 0. \tag{6.43}$$

Notice that, for every $\epsilon > 0$, there exists a set A_ϵ of finitely many patter such that

$$\inf_{|u|<C} \overline{\nu}_u(A_\epsilon) > 1 - \epsilon.$$

Therefore, using Eq. (6.43), one has that there exists N_ϵ such that $\forall n \geq N_\epsilon$,

$$\inf_{|u|<C} \nu_{u,n}(A_\epsilon) > 1 - \epsilon, \text{ and } \sup_{|u|<C} |\nu_{u,n}(A_\epsilon) - \overline{\nu}_u(A_\epsilon)| < \epsilon.$$

Therefore, given a bounded function $f : \mathbb{N}^{\mathcal{P}} \to \mathbb{R}$, for every $\epsilon > 0$, $\forall n \geq N_\epsilon$, one has

$$\begin{aligned}
&\sup_{|u|<C} |\mathbb{E}_{\nu_{u,n}}[f] - \mathbb{E}_{\overline{\nu}_u}[f]| \leq \\
&\|f\|_\infty \left(\sup_{|u|<C} |\nu_{u,n}(A_\epsilon) - \overline{\nu}_u(A_\epsilon)| + \right. \\
&\left. \sup_{|u|<C} \nu_{u,n}(A_\epsilon^C) + \overline{\nu}_u(A_\epsilon^C) \right) \leq \\
&3\|f\|_\infty \epsilon.
\end{aligned}$$

For the arbitrariness of ϵ, we proved the first statement about the weak convergence.

Consider the random process $N_{\text{tot}}(n)$ that counts all the occurrences of 1's in the output up to time n and notice that for every $\alpha \in \mathcal{P}$, $N_\alpha(n) \leq N_{\text{tot}}(n)$. Therefore for every $p \geq 1$ and $m \in \mathbb{N}$, one has

$$\mathbb{E}_{\nu_{u,n}}[N_\alpha(n)^p \mathbf{1}_{\{N_\alpha(n)>m\}}] \leq \mathbb{E}_{\nu_{u,n}}[N_{\text{tot}}(n)^p \mathbf{1}_{\{N_{\text{tot}}(n)>m\}}].$$

If we show that the moments of every order of $N_{\text{tot}}(n)$ converge to some finite limit and that it converges in law to some limit random variable X_u, we obtain the second statement as well. Indeed, let us call $C(p)$ the limit of the p-moment of $N_{\text{tot}}(n)$; notice that for every $p \geq 1$, $m \in \mathbb{N}$

$$\begin{aligned}
&\limsup_{n\to+\infty} \mathbb{E}_{\nu_{u,n}}[N_\alpha(n)^p \mathbf{1}_{\{N_\alpha(n)>m\}}] \leq \\
&\limsup_{n\to+\infty} \mathbb{E}_{\nu_{u,n}}[N_{\text{tot}}(n)^p \mathbf{1}_{\{N_{\text{tot}}(n)>m\}}] \leq \\
&\lim_{n\to+\infty} \mathbb{E}_{\nu_{u,n}}[N_{\text{tot}}(n)^{pq}]^{1/q} \nu_{u,n}(N_{\text{tot}}(n) > m)^{1/q'} = \\
&C(pq)^{1/q} \mathbb{P}(X_u > m)^{1/q'}.
\end{aligned}$$

Notice that in the last inequality we made use of Hölder inequality for some pair of conjugate indices (q, q'). Therefore, if we fix $p \geq 1$, for every $\epsilon > 0$, one can choose m_ϵ such that

$$\limsup_{n\to+\infty} \mathbb{E}_{\nu_{u,n}}[N_\alpha(n)^p \mathbf{1}_{\{N_\alpha(n)>m_\epsilon\}}] \leq \epsilon/2,$$
$$\mathbb{E}_{\bar{\nu}_u}[N_\alpha^p \mathbf{1}_{\{N_\alpha>m_\epsilon\}}] \leq \epsilon/2$$

and one gets

$$\limsup_{n\to+\infty} |\mathbb{E}_{\nu_{u,n}}[N_\alpha(n)^p] - \mathbb{E}_{\bar{\nu}_u}[N_\alpha^p]| \leq$$
$$\lim_{n\to+\infty} |\mathbb{E}_{\nu_{u,n}}[N_\alpha(n)^p \mathbf{1}_{\{N_\alpha(n)\leq m_\epsilon\}}] - \mathbb{E}_{\bar{\nu}_u}[N_\alpha^p \mathbf{1}_{\{N_\alpha\leq m_\epsilon\}}]| +$$
$$\limsup_{n\to+\infty} \mathbb{E}_{\nu_{u,n}}[N_\alpha(n)^p \mathbf{1}_{\{N_\alpha(n)>m_\epsilon\}}] +$$
$$\mathbb{E}_{\bar{\nu}_u}[N_\alpha^p \mathbf{1}_{\{N_\alpha>m_\epsilon\}}] \leq \epsilon.$$

Since this holds for every $\epsilon > 0$, we proved the statement.

We need to show that, under $\nu_{u,n}$, $N_{\text{tot}}(n)$ converges in law to a random variable X_u with finite moments of every order and that we have convergence of the moments as well. One can see that the Laplace transform of $N_{\text{tot}}(n)$ can be expressed as

$$\mathbb{E}_{\nu_{u,n}}[e^{zN_{\text{tot}}(n)}] = \tilde{\rho}^{\text{ss}}(u/\sqrt{n})(\tilde{T}^n_{u,z,n}(\mathbf{1})) \quad z \in \mathbb{C},$$

where

$$\tilde{T}_{u,z,n}(\cdot) = \tilde{K}_0^*(u/\sqrt{n}) \cdot \tilde{K}_0(u/\sqrt{n}) + e^z \tilde{K}_1^*(u/\sqrt{n}) \cdot \tilde{K}_1^*(u/\sqrt{n}).$$

Notice that $\tilde{T}_{0,z} := \tilde{T}_{0,z,n}$ is independent from n and is an analytic perturbation of $\tilde{T}$. If we pick z small enough in modulus, perturbation theory ensures that $\tilde{T}_{0,z}$ has 1 as eigenvalue with maximum modulus with $|\chi^{\text{ss}}\rangle \langle\chi^{\text{ss}}|$ as left eigenvector. Let x_z be the corresponding right eigenvector such that $\text{Tr}(\rho^{\text{ss}} x_z) = \langle\chi^{\text{ss}}| x_z |\chi^{\text{ss}}\rangle = 1$, and let

$$x_z = \begin{pmatrix} 1 & a \\ b & c \end{pmatrix}$$

be its block matrix form with respect to the decomposition of $\mathbb{C}^D$ into $\mathbb{C}|\chi^{\text{ss}}\rangle$ and its orthogonal complement. Then one can prove that $a = b = 0$ by using the fact that the Kraus operators are of the form

$$K_0 = \begin{pmatrix} 1 & 0 \\ 0 & \beta \end{pmatrix}, \quad K_1 = \begin{pmatrix} 0 & \gamma \\ 0 & \delta \end{pmatrix}$$

for some blocks β, γ, δ such that $|\beta|^2 + |\gamma|^2 + |\delta|^2 = \mathbf{1}$. Let us first fix z small enough, then for n big enough, $\tilde{T}_{u,z,n}$ has a unique eigenvalue $\lambda_z(u,n)$ of maximum modulus with corresponding left and right eigenvectors $l_z(u,n)$, $x_z(u,n)$ and one has that

$$\mathbb{E}_{\nu_{u,n}}[e^{zN_{\text{tot}}(n)}] =$$
$$\lambda_z(u,n)^n l_z(u,n)(\mathbf{1})\tilde{\rho}^{\text{ss}}(u/\sqrt{n})(x_z(u)) + o(n)$$

where $o(n)$ is uniform in u (and z in compact sets). Notice that both $l_z(u,n)(\mathbf{1})$ and $\tilde{\rho}^{\mathrm{ss}}(u/\sqrt{n})(x_z(u))$ converge to 1 for $n \to +\infty$; regarding the behaviour of $\lambda_z(u,n)$ consider the Taylor expansion up to second order in u around 0:

$$\lambda_z(u,n) = 1 + \frac{u}{\sqrt{n}}\lambda_z^{(1)} + \frac{u^2}{n}\lambda_z^{(2)} + o(n^{-1}).$$

We can choose $l_z(u,n)$ and $x_z(u,n)$ such that $\mathrm{Tr}(()l_z(u,n)x_z(u,n)) \equiv 1$, therefore differentiating $\mathrm{Tr}(()l_z(u,n)\tilde{T}_{z,u,n}(x_z(u,n))) = \lambda_z(u,n)$ at 0 one gets

$$\lambda_z^{(1)} = \mathrm{Tr}(()l_z\partial_u\tilde{T}_{z|u=0}x_z)$$

which can be easily seen to be 0. Summing up, we proved that

$$\lim_{n\to+\infty} \mathbb{E}_{\nu_{u,n}}[e^{zN_{\mathrm{tot}}(n)}] = e^{\lambda_z^{(2)}}$$

uniformly in u and in z in compact small neighborhoods of 0. Since for every u, $f_n(z) := \mathbb{E}_{\nu_{u,n}}[e^{zN_{\mathrm{tot}}(n)}]$ are analytic functions around 0, then we can deduce that $e^{\lambda_z^{(2)}}$ is analytic as well and that we have uniform convergence (on compact small neighborhoods of 0) of all the derivatives, which consists exactly in the convergence of moments of all orders and we are done. □

6.15 ACHIEVABILITY OF THE QCRB UNDER ADDITIONAL ASSUMPTIONS

In this section we present a proof of Theorem 21 and we comment on the gap between the hypothesis that we need to assume and what we proved in Proposition 9.

Proof of Theorem 21. By hypothesis, we know that θ belongs to the (random) confidence interval

$$I_n = (\tilde{\theta}_n - n^{-1/2+\epsilon},\ \tilde{\theta}_n + n^{-1/2+\epsilon})$$

with high probability. In order to prove the statement, it suffices to show that

$$|(\mathbb{E}_\theta[e^{ia\sqrt{n}(\hat{\theta}_n-\theta)}|\tilde{\theta}_n] - e^{-\frac{a^2}{2f_\theta}})|\chi_{\{\theta\in I_n\}}(\tilde{\theta}_n)$$

can be upper bounded uniformly in $\tilde{\theta}_n$ by a sequence converging to 0. Indeed, notice that for every $a \in \mathbb{R}$

$$\begin{aligned}\mathbb{E}_\theta[e^{ia\sqrt{n}(\hat{\theta}_n-\theta)}] = e^{-\frac{a^2}{2f_\theta}}\mathbb{P}_\theta(\theta\in I_n)&\\ + \int_{\theta\in I_n} p_\theta(d\tilde{\theta}_n)(\mathbb{E}_\theta[e^{ia\sqrt{n}(\hat{\theta}_n-\theta)}|\tilde{\theta}_n] - e^{-\frac{a^2}{2f_\theta}})&\\ + \int_{\theta\notin I_n} p_\theta(d\tilde{\theta}_n)\mathbb{E}_\theta[e^{ia\sqrt{n}(\hat{\theta}_n-\theta)}|\tilde{\theta}_n].&\end{aligned}$$

Since $\mathbb{P}_\theta(\theta \notin I_n)$ goes to zero, the first term goes to $e^{-\frac{a^2}{2f_\theta}}$ and the third one vanishes. If we show that the second term vanishes then we obtain the convergence in distribution of $\sqrt{n}(\hat{\theta}_n - \theta)$ as in the statement.

First of all notice that for every $a \in \mathbb{R}$, for every $\tilde{\theta} \in \Theta$

$$\mathbb{E}[e^{ia(\overline{Y}_{n,\theta,\tilde{\theta}} - \sqrt{n}(\theta-\tilde{\theta}))}] = e^{\lambda_{\mathrm{tot}}(\tilde{\theta})(\sqrt{n}(\theta-\tilde{\theta})+\tau_n)^2(e^{ia/(2\lambda_{\mathrm{tot}}(\tilde{\theta})\tau_n)}-1)-ia\left(\frac{\tau_n}{2}+\sqrt{n}(\tilde{\theta}-\theta)\right)}.$$

Therefore for every $a \in \mathbb{R}$ one has

$$\lim_{n\to+\infty} \sup_{\tilde{\theta}:|\tilde{\theta}-\theta|<n^{-1/2+\epsilon}} \left| \mathbb{E}[e^{ia(\overline{Y}_{n,\theta,\tilde{\theta}} - \sqrt{n}(\theta-\tilde{\theta}))}] - e^{-\frac{a^2}{f_\theta}} \right| = 0$$

We will denote by $u = \sqrt{n}(\theta - \tilde{\theta}_n)$. The definition of $\hat{\theta}_n$ implies that

$$\sqrt{n}(\hat{\theta}_n - \theta) = Y_n - u.$$

Therefore we obtain that

$$\begin{aligned} &\sup_{\tilde{\theta}:|\tilde{\theta}-\theta|<n^{-1/2+\epsilon}} \left| \mathbb{E}_\theta[e^{ia(Y_n-u)}|\tilde{\theta}_n = \tilde{\theta}] - e^{-\frac{a^2}{2f_\theta}} \right| \leq \\ &\sup_{\tilde{\theta}:|\tilde{\theta}-\theta|<n^{-1/2+\epsilon}} |\mathbb{E}_\theta[e^{iaY_n}|\tilde{\theta}_n = \tilde{\theta}] - \mathbb{E}[e^{ia\overline{Y}_{n,\theta,\tilde{\theta}}}]| + \\ &\sup_{\tilde{\theta}:|\tilde{\theta}-\theta|<n^{-1/2+\epsilon}} \left| \mathbb{E}[e^{ia(\overline{Y}_{n,\theta,\tilde{\theta}} - \sqrt{n}(\theta-\tilde{\theta}))}] - e^{-\frac{a^2}{f_\theta}} \right|. \end{aligned}$$

Hence,

$$\lim_{n\to+\infty} \sup_{\tilde{\theta}:|\tilde{\theta}-\theta|<n^{-1/2+\epsilon}} \left| \mathbb{E}_\theta[e^{ia(Y_n-u)}|\tilde{\theta}_n = \tilde{\theta}] - e^{-\frac{a^2}{2f_\theta}} \right| = 0$$

and we are done. □

Let us briefly comment on the relationship between the additional hypothesis we introduced (Eq. (6.25)) and the result in Proposition 9; let us consider the following family of stochastic processes: for every $n \in \mathbb{N}$, $\tilde{\theta}$, $\theta \in \Theta$, consider the collection of independent random variables

$$\overline{N}_{n,\theta,\tilde{\theta},\alpha} \sim \mathrm{Poisson}(\lambda_\alpha(\tilde{\theta}(\sqrt{n}(\theta-\tilde{\theta})+\tau_n)^2), \quad \alpha \in \mathcal{P}$$

and their law $\overline{\nu}_{n,\theta,\tilde{\theta}}$ on $\mathbb{N}^{\mathcal{P}}$. (together with the σ-field of cylindrical sets). Notice that $\sum_{\alpha\in\mathcal{P}} \overline{N}_{n,\theta,\tilde{\theta},\alpha}$ converges in mean square and has the same law as $\overline{N}_{n,\theta,\tilde{\theta}}$.

Inspecting the proof, one can notice that the convergence in the statement of Proposition 9 holds uniformly in a small neighborhood of the reference parameter θ_0, therefore we can restate it in the following way: for ϵ small enough and for every finite collections of excitation patterns counts $\mathbf{m}$ one has

$$\lim_{n\to+\infty} \sup_{\tilde{\theta}:|\tilde{\theta}-\theta|<n^{-1/2+\epsilon}} \left| \frac{\nu_{\sqrt{n}(\theta-\tilde{\theta}),n}(B_{\mathbf{m}}(n))}{\overline{\nu}_{\theta,\tilde{\theta},n}(\mathbf{m})} - 1 \right| = 0.$$

If we were able to show that the previous result still holds integrating with respect to $\overline{\nu}_{\theta,\tilde{\theta},n}$, i.e.,

$$\begin{aligned} &\lim_{n\to+\infty} \sup_{\tilde{\theta}:|\tilde{\theta}-\theta|<n^{-1/2+\epsilon}} \sum_{\mathbf{m}} \overline{\nu}_{\theta,\tilde{\theta},n}(\mathbf{m}) \left| \frac{\nu_{\sqrt{n}(\theta-\tilde{\theta}),n}(B_{\mathbf{m}}(n))}{\overline{\nu}_{\theta,\tilde{\theta},n}(\mathbf{m})} - 1 \right| = \\ &\lim_{n\to+\infty} \sup_{\tilde{\theta}:|\tilde{\theta}-\theta|<n^{-1/2+\epsilon}} \sum_{\mathbf{m}} \left| \nu_{\sqrt{n}(\theta-\tilde{\theta}),n}(B_{\mathbf{m}}(n)) - \overline{\nu}_{\theta,\tilde{\theta},n}(\mathbf{m}) \right| = 0, \end{aligned} \tag{6.44}$$

this would imply the condition in Eq. (6.25) (it can be seen using the fact that $\sum_{\alpha\in\mathcal{P}} \overline{N}_{n,\theta,\tilde{\theta},\alpha}$ and $\overline{N}_{n,\theta,\tilde{\theta}}$ have the same law). Unfortunately, we are not able to prove this.

The last remark we make is that Eq. (6.44) cannot be true unless

$$\lim_{n\to+\infty} \sup_{\tilde{\theta}:|\tilde{\theta}-\theta|<n^{-1/2+\epsilon}} \overline{\nu}_{\theta,\tilde{\theta},n}(G(n)^C) = 0,$$

where $G(n)$ is the set of all $\mathbf{m}$'s such that $\nu_{\sqrt{n}(\theta-\tilde{\theta}),n}(B_{\mathbf{m}}(n)) > 0$. We can prove that this is indeed the case.

Lemma 9. *If ϵ is small enough, then*

$$\lim_{n\to+\infty} \sup_{\tilde{\theta}:|\tilde{\theta}-\theta|<n^{-1/2+\epsilon}} \overline{\nu}_{\theta,\tilde{\theta},n}(G(n)^C) = 0.$$

Proof. Notice that $G(n)$ is the set of all those patterns counts $\mathbf{m} = (m_{\alpha^{(1)}}, \ldots, m_{\alpha^{(k)}})$ such that the following conditions are satisfied

1. $\alpha^{(i)}$ does not contain more than n^γ consecutive 0s for every $i = 1, \ldots, k$;
2. $\sum_{i=1}^k m_{\alpha^{(i)}}|\alpha^{(i)}| + (k-1)n^\gamma \leq n$.

We recall that $|\alpha|$ is the length of the pattern α.

Let us consider a positive number $\eta < 1-\gamma$ and notice that one has $\tilde{G}(n) \subseteq G(n)$, where $\tilde{G}(n)$ is the set of all those patterns counts $\mathbf{m} = (m_{\alpha^{(1)}}, \ldots, m_{\alpha^{(k)}})$ such that

1. $|\alpha^{(i)}| \leq \eta \log_2(n)$ for every $i = 1, \ldots, k$ and
2. $\sum_{i=1}^k m_{\alpha^{(i)}}|\alpha^{(i)}| + (k-1)n^\gamma \leq n$.

We denote by $A(n)$ the set of patterns satisfying 1. and $B(n)$ the set of patterns satisfying 2., hence $\tilde{G}(n) = A(n) \cap B(n)$. In order to prove the statement, it suffices to show that

$$\lim_{n\to+\infty} \sup_{\tilde{\theta}:|\tilde{\theta}-\theta|<n^{-1/2+\epsilon}} \overline{\nu}_{\theta,\tilde{\theta},n}(\tilde{G}(n)^C) = 0.$$

Notice that $\tilde{G}(n)^C = A(n)^C \sqcup (B(n)^C \cap A(n))$; let us first show that

$$\lim_{n\to+\infty} \inf_{\tilde{\theta}:|\tilde{\theta}-\theta|<n^{-1/2+\epsilon}} \overline{\nu}_{\theta,\tilde{\theta},n}(A(n)) \to 1.$$

Notice that,

$$\lim_{n\to+\infty} \inf_{\tilde{\theta}:|\tilde{\theta}-\theta|<n^{-1/2+\epsilon}} \overline{\nu}_{\theta,\tilde{\theta},n}(A(n)) =$$

$$\lim_{n\to+\infty} \inf_{\tilde{\theta}:|\tilde{\theta}-\theta|<n^{-1/2+\epsilon}} e^{-\left(\sum_{|\alpha|>\eta\log_2(n)} \lambda_\alpha\right)(\sqrt{n}(\theta-\tilde{\theta})-\tau_n)^2} = 1,$$

since, for every $\tilde{\theta}$ such that $|\theta-\tilde{\theta}| < n^{-1/2+\epsilon}$, one has

$$\left(\sum_{|\alpha|>\eta\log_2(n)} \lambda_\alpha\right)(\sqrt{n}(\theta-\tilde{\theta})-\tau_n)^2 \lesssim n^{\eta\log_2(\lambda)+3\epsilon} \to 0$$

if $\epsilon < -\eta \log_2(\lambda)/3$. Let us now study the probability of $B(n)^C \cap A(n)$: first notice that

$$B(n)^C \cap A(n) \subseteq$$
$$\{\mathbf{m} : |\alpha^{(i)}| \leq \eta \log_2(n), \sum_{i=1}^{k} m_{\alpha^{(i)}} |\alpha^{(i)}| + 2^{\eta \log_2(n)} n^{\gamma} > n\},$$

because $2^{\eta \log_2(n)}$ upper bounds the cardinality of all the patterns of length smaller or equal than $\eta \log_2(n)$. Therefore

$$\overline{\nu}_{\theta,\tilde{\theta},n}(B(n)^C \cap A(n)) \leq \overline{\nu}_{\theta,\tilde{\theta},n}(C(n))$$
$$\leq \overline{\nu}_{\theta,\tilde{\theta},n}\left(\left\{\mathbf{m} : \sum_{|\alpha| \leq \eta \log_2(n)} m_{\alpha} > \frac{n - n^{\eta+\gamma}}{\eta \log(n)}\right\}\right).$$

where $C(n)$ is the set of pattern counts such that $|\alpha^{(i)}| \leq \eta \log_2(n)$ and $\sum_{i=1}^{k} m_{\alpha^{(i)}} \eta \log_2(n) + n^{\eta+\gamma} n > n$. The last term amounts to the probability that a Poisson random variable of parameter $\sum_{|\alpha| \leq \eta \log_2(n)} \lambda_{\alpha} (\sqrt{n}(\theta - \tilde{\theta}) - \tau_n)^2 \lesssim n^{3\epsilon}$ is bigger than something that grows as $n/(\eta \log(n))$. Such a probability goes to 0 uniformly in $\tilde{\theta}$ if $3\epsilon < 1$ and we are done.

□

Part III

APPENDIX

A

NUMERICAL SIMULATIONS PYTHON CODE

In this appendix we introduce the Python code used for all the numerical simulations found in this thesis. The code has been made available on GitHub at

https://github.com/AGodley/ThesisCode.

The primary aim of this appendix is to provide a guide to the most recent code, explaining the role of each of the main files so that the reader can run a simulation themselves with minimal difficulty.

To this end, we start by describing the structure of the GitHub repository. On this GitHub page there are two main folders: one containing the Python code itself and one containing the data from our numerical simulations. Both folders are split again into distinct projects. Theses project correspond with the numerical simulations for the adaptive measurement scheme [78], the exploratory simulations for an independent identically distributed (IID) displaced null measurement scheme [77] and the numerical simulations for the quantum Markov chain (QMC) displaced null measurement scheme [75]. The code for this last project is the most recent and up to date, so we will mostly discuss this project.

The model that we are simulating is a simple qubit model outlined in section 6.8 with parameters θ,λ,ϕ. The parameter of interest is θ, while both $\lambda = 0.8$ and $\phi = \pi/4$ are fixed values. In particular, the environment, or input units, of a quantum Markov chain (QMC) are modelled as a series of qubits initialised in the $|0\rangle\langle 0|$ state, the system itself is modelled as a qubit and the absorber is modelled as a qubit. In the /code/displacedNullMeasurements folder there is a total of 9 Python files:

```
absorber.py
analysis.py
dnm_main.py
dnm_main_fixed.py
dnm_main_repeated.py
initial.py
kraus.py
patterns.py
qfi.py
```

The model itself is specified within kraus.py. This file contains functions that calculate the model's Kraus operators for specified parameter values, both with and without the absorber:

```
k(tht, lmbd, phi, meas)
k_dot(tht, lmbd, phi, meas)
```

```
k_abs(tht, tht_r, lmbd, phi, meas)
k_abs_dot(tht, tht_r, lmbd, phi, meas)
```

In each function the model's pseudo-unitary U is manually coded; this is not a true unitary as we only have to define its action on states $|i\rangle \otimes |0\rangle$ for an ONB $\{|i\rangle \in \mathbb{C}^2\}$ since each noise unit of the quantum Markov chain (QMC) is prepared in the same initial state $|0\rangle \langle 0|$. In `k( )` the Kraus operators are then calculated as

$$K_i = \mathrm{Tr}(\mathbf{1} \otimes |0\rangle \langle e_i| U)$$

and these Kraus operators are returned by the function to be used elsewhere. For example:

```
# Kraus operators without the absorber
def k(tht, lmbd, phi, meas):
    # Inputs the unitary U; this is not a true unitary as only it's action ...
    # At each step a new auxiliary system |0> interacts with our soi.
    U = Qobj([[np.cos(tht) * sqrt(1 - tht ** 2), 0, 1j * np.sin(tht) \
    * sqrt(1 - lmbd), 0],
              [0, 0, sqrt(lmbd) * np.exp(1j * phi), 0],
              [1j * np.sin(tht)
              * sqrt(1 - tht ** 2), 0, np.cos(tht) * sqrt(1 - lmbd), 0],
              [tht, 0, 0, 0]],
             dims=[[2, 2], [2, 2]])

    # Checks whether the function was sent 1 or 2 measurements
    if len(meas) == 2:
        # WHen sent two measurements, calculates both corresponding ...
        K_0 = (tensor(qeye(2), fock(2, 0) * meas[0].dag()) * U).ptrace([0])
        K_1 = (tensor(qeye(2), fock(2, 0) * meas[1].dag()) * U).ptrace([0])

        # Kraus operators are returned in a list
        K = [K_0, K_1]

        # For checking they're proper Kraus operators
        # print('Kraus check:')
        # print(K[0].dag()*K[0] + K[1].dag()*K[1])
    else:
        # Kraus corresponding to a single measurement
        K = (tensor(qeye(2), fock(2, 0) * meas[0].dag()) * U).ptrace([0])
    return K, U
```

Crucially, this function is passed an orthonormal basis (ONB) as the measurement choice in a list. This is always

```
meas = [fock(2, 0), fock(2, 1)],
```

but it could be varied. A legacy feature allows a single measurement to be passed to the function instead, and the function will calculate the corresponding Kraus operator.

The function `k_abs( )` calculates the Kraus operators for the combined system+absorber in a similar manner using total unitary VU, where the unitary V corresponding to the absorber is obtained from a function within the file absorber.py. For this model, we calculated an analytical formula for the system's

stationary state for any θ. This file simply uses these analytical formula for an input value of θ to calculate the absorber's unitary V based off the desired property in Lemma 1. These functions are pretty well documented, so we suggest looking at the files instead. The result is a new set of Kraus operators on a now 4-dimensional space (system+absorber).

The file kraus.py also contains another function

```
true_ss(tht, tht_r, lmbd, phi, meas)
```

which calculates the actual stationary state of the channel associated with the system+absorber by vectorising its Kraus operators and finding the eigenvector with eigenvalue 1. This is mostly used in some analysis of the generated trajectories. The other two functions calculate the derivatives of the Kraus operators (with/without absorber) in a similar manner using a hard-coded derivative of the pseudo-unitary.

The most important file is dnm_main.py. This is the main file, from which we configure and run simulations of the estimation scheme described in Chapter 6, specifically section 6.8. This configuration is mostly done by changing the values found in the dictionary:

```
# Global parameter values
setup = {
    'N': 3000,  # No. of samples to generate
    'n': 6 * 10**5,  # No. of ancillary systems
    'theta': 0.2,  # True value of theta for trajectory generation
    'lambda': 0.8,  # Transition parameter, 0<=lmbd<=1
    'phi': pi / 4,  # Phase parameter
    'initial': False,  # Controls initial estimation; ...
    'eps': 0.065,  # Prop. of traj. to use in initial est.
    'gamma': 2.1   # Controls how far the absorber is offset, gamma>1
}
```

As explained by the comments, these variables are the number of trajectories (individual simulations) to generate, the length of each trajectory, the 3 parameters of the qubit model that we are simulating (see equation (4.20)), a variable that configures whether we want to include an initial estimation step, a variable that specifies the proportion of the input systems that we use in the initial estimation ($n^{1-\epsilon}$), and a variable that controls the displacement in the null measurement.

When we run the main file, the following code repeatedly calls a function

```
trajectory(id, setup)
```

to generate multiple trajectories:

```
# Pool object to run the multiprocessing
pool = Pool()
results = pool.starmap(trajectory, zip(ids, repeat(setup)))
pool.close()
pool.join()
```

This handles multiprocessing, i.e.- generating multiple trajectories at the same time.

Inside `trajectory( )`, an initial if statement determines whether `setup` included initial estimation. If so, the total length n is partitioned into parts for initial and final estimation, `n_init` $= n^{1-\epsilon}$ and `n_final`:

```
# Splits the samples into samples used in the initial and final est
n_init = int(np.floor(n ** (1 - eps)))
n_final = n - n_init
theta_rough, *_ = initial_est(theta, lmbd, phi, n_init)
```

An initial estimate is then calculated through the function `initial_est( )`, found within the initial.py file. This function calculates an initial estimate through a rudimentary method; this involves first generating a trajectory through the following algorithm:

1. First, initialise a qubit in the state $\rho = |0\rangle\langle 0|$ state and calculate the Kraus operators K_0, K_1 at $\theta = 0.2$.
2. Pick a Kraus operator to apply randomly with probabilities $p_i = \mathrm{Tr}(K_i \rho K_i)$, record the choice i in a list `x`.
3. Apply the Kraus operator to the state
$$\rho \to \frac{K_i \rho K_i}{p_i}.$$
4. Repeat the last two steps for the remaining $n_i nit - 1$ qubits.

We then calculate the expected number of 1s on a grid of values of θ between 0.18 and 0.22. This is done by calculating the stationary state $\rho_*(\theta')$ for each value θ' from this grid using the analytical function found within absorber.py. We then calculate the corresponding Kraus operators at θ' and calculate the expected number of 1s as

$$\texttt{n\_init}.\mathrm{Tr}(\rho K_1^* K_1).$$

Fitting this to the observed number of 1s provides our initial estimate `theta_rough`. Otherwise, we set `theta_rough = 0.2`, the true value of θ used in all our simulations. This covers the initial estimation.

We can then calculate the Kraus operators that include the absorber using the rough estimate plus the offset required by the null measurement scheme:

```
offset = n ** (-0.5 + gamma * eps)
absorber = theta_rough + offset

# Stage 2: measurement in the fixed basis specified by the function
M = measurement_choice()

# Finds kraus operators
K, *_ = k_abs(theta, absorber, lmbd, phi, M)
```

This corresponds to total unitary

$$V_{\text{absorber}} U_\theta.$$

The final stage trajectory is then generated in a very similar manner with these new Kraus operators, starting with the purified stationary state at $\theta =$ `theta_rough`, which is also returned by the function `uV( )` from absorber.py that calculates the unitary V for the absorber. This is our best guess at the true stationary state.

```
# Initializes the state of the s+a; estimate of true ss
a, rho_0 = uV(theta_rough, lmbd, phi)
rho = rho_0

# List to record which outcomes occurred
x = [None] * n_final

# Trajectory generation
for j in np.arange(n_final):
    # Defines probability that outcome 0 occurs
    p0 = np.real_if_close( (K[0] * rho * K[0].dag()).tr() )

    # Choice of {0, 1} with prob {p(0), p(1)}
    x_j = np.random.choice([0, 1], p=[p0, 1 - p0])

    # Records the outcome in x_s2
    x[j] = x_j

    # Updates the state by applying the measurement projection and ...
    rho = K[x_j] * rho * K[x_j].dag()
    rho = rho / rho.tr()
```

At this stage the result is a list `x` of mostly 0s that contains excitation patterns as described in chapter 6. E.g.-

```
x = [0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, ...]
```

To extract these patterns, there is a choice of two functions: `stochastic_patterns( )` or `pattern_check(x, k)`. The first is a new experimental method of counting the patterns that relies upon a probabilistic algorithm. The second is the method described in section 6.7, which looks for an excitation pattern α padded with 0s either side, where `k` is the max pattern length that we search for. These pattern counts are then saved in a csv file such as counting_Markov_excitations_analysis.csv.

Additionally, the expected rates of each pattern are calculated using the forumlae in section 6.5:

```
# Calculates the expected counts for the pattern
expected_counts, mpn_pat = expected(rho_t0, K,
abs(theta - absorber), n_final)

# Alternative stationary state used in Taylor expansion formula for ...
rho_ss = true_ss(theta_rough, theta_rough, lmbd, phi, M)

# Best guess at kraus operators for calculating expected values ...
K, *_ = k_abs(theta_rough, theta_rough, lmbd, phi, M)
K_dot, *_ = k_abs_dot(theta_rough, theta_rough, lmbd, phi, M)  # ...

# Calculates the mus for the pattern using alternative formula
alt_mus, FI_pat, alt_expected = alternative(rho_ss, K, K_dot,
            abs(theta - absorber), n_final)
```

Crucially, this is also done using `theta_rough`; we use these rates in the pattern counts estimator, so we cannot use the true value in their calculation. These rates are also saved in csv files.

The Python file analysis.py then processes these saved csv files patterns.py to produce the figures. This involves the pattern counting estimator, see section 6.7, contained in the function `u_method( )`. This function is quite long, containing several different methods that specify a cutoff on the patterns we include in our estimator. Therefore, we recommend looking at the file itself.

The remaining files are dnm_main_fixed.py, dnm_main_repeated.py and qfi. The first simply fixes the initial estimation so a that fixed `theta_rough` is used to generate multiple trajectories, which we used in some figures. The second increases the length of the trajectory used in the final estimation by a multiplicative factor. This was useful when we wanted to have realistic initial estimation, but get much closer to an asymptotic regime where the proportion of n used in the initial estimation should be very small. And finally, the last file simple calculate the model's quantum Fisher information.

BIBLIOGRAPHY

[1] J. Abadie et al. "A gravitational wave observatory operating beyond the quantum shot-noise limit". In: *Nature Physics* 7.12 (Dec. 2011), pp. 962–965 (cit. on pp. 1, 2, 50, 51).

[2] B. P. Abbott et al. "Observation of Gravitational Waves from a Binary Black Hole Merger". In: *Physical Review Letters* 116.6 (Feb. 2016), p. 061102 (cit. on pp. 1, 2).

[3] Felix Abramovich and Ya'acov Ritov. *Statistical Theory*. Dec. 2022, pp. 1–237 (cit. on pp. 34–37, 42).

[4] F. Albarelli et al. "A perspective on multiparameter quantum metrology: From theoretical tools to applications in quantum imaging". In: *Physics Letters A* 384.12 (Apr. 2020), p. 126311 (cit. on pp. 2, 50, 134).

[5] Francesco Albarelli, Jamie F Friel, and Animesh Datta. "Evaluating the Holevo Cramér-Rao Bound for Multiparameter Quantum Metrology". In: *Phys. Rev. Lett.* 123.20 (Nov. 2019), p. 200503 (cit. on p. 2).

[6] Francesco Albarelli et al. "Ultimate limits for quantum magnetometry via time-continuous measurements". In: *New Journal of Physics* 19.12 (Dec. 2017), p. 123011 (cit. on pp. 55, 131).

[7] Francesco Albarelli et al. "Restoring Heisenberg scaling in noisy quantum metrology by monitoring the environment". In: *Quantum* 2 (Dec. 2018), p. 110 (cit. on pp. 55, 131).

[8] S Aravinda, Shilpak Banerjee, and Ranjan Modak. "Ergodic and mixing quantum channels: From two-qubit to many-body quantum systems". In: (Oct. 2023) (cit. on p. 25).

[9] Alain Aspect, Philippe Grangier, and Gérard Roger. "Experimental Realization of Einstein-Podolsky-Rosen-Bohm Gedankenexperiment : A New Violation of Bell's Inequalities". In: *Physical Review Letters* 49.2 (July 1982), pp. 91–94 (cit. on p. 15).

[10] Stéphane Attal and Yan Pautrat. "From Repeated to Continuous Quantum Interactions". In: *Annales Henri Poincaré* 7.1 (Jan. 2006), pp. 59–104 (cit. on pp. 4, 51, 130, 136).

[11] M. Bailes et al. "Gravitational-wave physics and astronomy in the 2020s and 2030s". In: *Nature Reviews Physics* 3.5 (Apr. 2021), pp. 344–366 (cit. on p. 2).

[12] K. Banaszek, M. Cramer, and D. Gross. "Focus on quantum tomography". In: *New Journal of Physics* 15.12 (Dec. 2013), p. 125020 (cit. on p. 134).

[13] Ole E. Barndorff-Nielsen, Richard D. Gill, and Peter E. Jupp. "On Quantum Statistical Inference, II". In: *Journal of the Royal Statistical Society. Series B: Statistical Methodology* 65.4 (July 2003), pp. 775–804 (cit. on p. 1).

[14] H Barnum and E Knill. "Sufficiency of channels over von Neumann algebras". In: *Journal of Mathematical Physics* 43 (2002), 2097–2106 (cit. on p. 139).

[15] V P Belavkin. "Generalized uncertainty relations and efficient measurements in quantum systems". In: *Theoretical and Mathematical Physics* 26.3 (Mar. 1976), pp. 213–222 (cit. on pp. 83, 84, 134).

[16] V P Belavkin. "Nondemolition principle of quantum measurement theory". In: *Foundations of Physics* 24.5 (May 1994), pp. 685–714 (cit. on pp. 55, 64, 130).

[17] J. S. Bell. "On the Einstein Podolsky Rosen paradox". In: *Physics Physique Fizika* 1.3 (Nov. 1964), pp. 195–200 (cit. on p. 15).

[18] F Benatti, S Alipour, and A T Rezakhani. "Dissipative quantum metrology in manybody systems of identical particles". In: *New Journal of Physics* 16.1 (Jan. 2014), p. 015023 (cit. on pp. 55, 130).

[19] Tristan Benoist, Lisa Hänggli, and Cambyse Rouzé. "Deviation bounds and concentration inequalities for quantum noises". In: *Quantum* 6 (Aug. 2022), p. 772 (cit. on p. 131).

[20] D W Berry and H M Wiseman. "Adaptive quantum measurements of a continuously varying phase". In: *Phys. Rev. A* 65.4 (2002), p. 43803 (cit. on pp. 55, 131).

[21] P J Bickel, Y Ritov, and T Rydén. "Asymptotic normality of the maximum likelihood estimator for general hidden Markov models". In: *Ann. Statist.* 26 (1998), 1614–1635 (cit. on pp. 74, 75).

[22] Luc Bouten, Ramon van Handel, and Matthew James. "An introduction to quantum filtering". In: *SIAM J. Control Optim.* 46.6 (Jan. 2006), pp. 2199–2241 (cit. on pp. 55, 64, 130).

[23] J. B. Brask, R. Chaves, and J. Kołodyński. "Improved Quantum Magnetometry beyond the Standard Quantum Limit". In: *Physical Review X* 5.3 (July 2015), p. 031010 (cit. on pp. 1, 2, 50).

[24] Jonatan Bohr Brask. "Gaussian states and operations – a quick reference". In: (Feb. 2021) (cit. on p. 17).

[25] Samuel L. Braunstein and Carlton M. Caves. "Statistical distance and the geometry of quantum states". In: *Physical Review Letters* 72.22 (May 1994), pp. 3439–3443 (cit. on pp. 1, 55, 58, 59, 62, 83, 84, 89, 134).

[26] Samuel L. Braunstein, Carlton M. Caves, and G.J. Milburn. "Generalized Uncertainty Relations: Theory, Examples, and Lorentz Invariance". In: *Annals of Physics* 247.1 (Apr. 1996), pp. 135–173 (cit. on pp. 2, 89, 134).

[27] Heinz-Peter Breuer and Francesco Petruccione. *The Theory of Open Quantum Systems*. Vol. 9780199213. Jan. 2007, pp. 1–656 (cit. on pp. 31, 32).

[28] Dmitry Budker and Derek F Jackson Kimball. *Optical Magnetometry*. Mar. 2013, pp. 1–2 (cit. on pp. 2, 50).

[29] Dmitry Budker and Michael Romalis. "Optical magnetometry". In: *Nature Physics* 3.4 (Apr. 2007), pp. 227–234 (cit. on p. 50).

[30] D Burgarth et al. "Ergodic and mixing quantum channels in finite dimensions". In: *New Journal of Physics* 15.7 (July 2013), p. 073045 (cit. on pp. 4, 21, 25, 28).

[31] Daniel Burgarth et al. "Quantum estimation via sequential measurements". In: *New Journal of Physics* 17.11 (Nov. 2015), p. 113055 (cit. on p. 131).

[32] Cristina Butucea, Mădălin Guţă, and Michael Nussbaum. "Local asymptotic equivalence of pure states ensembles and quantum Gaussian white noise". In: *The Annals of Statistics* 46.6B (Dec. 2018), pp. 3676–3706 (cit. on pp. 106, 132, 134).

[33] Raffaella Carbone and Yan Pautrat. "Irreducible Decompositions and Stationary States of Quantum Channels". In: *Reports on Mathematical Physics* 77.3 (June 2016), pp. 293–313 (cit. on p. 4).

[34] Raffaella Carbone and Yan Pautrat. "Open Quantum Random Walks: Reducibility, Period, Ergodic Properties". In: *Annales Henri Poincaré* 17.1 (Jan. 2016), pp. 99–135 (cit. on p. 4).

[35] Howard Carmichael. *An Open Systems Approach to Quantum Optics*. Vol. 18. Lecture Notes in Physics Monographs. Springer Berlin Heidelberg, 1993 (cit. on pp. 55, 130).

[36] Federico Carollo, Robert L Jack, and Juan P Garrahan. "Unraveling the Large Deviation Statistics of Markovian Open Quantum Systems". In: *Phys. Rev. Lett.* 122.13 (Apr. 2019), p. 130605 (cit. on p. 131).

[37] Catalin Catana, Luc Bouten, and Mădălin Guţă. "Fisher informations and local asymptotic normality for continuous-time quantum Markov processes". In: *Journal of Physics A: Mathematical and Theoretical* 48.36 (Sept. 2015), p. 365301 (cit. on pp. 51, 55, 61–63, 118, 131, 147).

[38] B Chase and J M Geremia. "Single-shot parameter estimation via continuous quantum measurement". In: *Phys. Rev. A* 79 (2009), p. 22314 (cit. on pp. 55, 131).

[39] S. Choi and B. Sundaram. "Bose-Einstein condensate as a nonlinear Ramsey interferometer operating beyond the Heisenberg limit". In: *Physical Review A - Atomic, Molecular, and Optical Physics* 77.5 (May 2008), p. 053613 (cit. on p. 50).

[40] Nelson Christensen and Renate Meyer. "Parameter estimation with gravitational waves". In: *Reviews of Modern Physics* 94.2 (Apr. 2022), p. 025001 (cit. on pp. 50, 51).

[41] Mario A. Ciampini et al. "Quantum-enhanced multiparameter estimation in multiarm interferometers". In: *Scientific Reports 2016 6:1* 6.1 (July 2016), pp. 1–8 (cit. on p. 49).

[42] F Ciccarello et al. "Quantum collision models: Open system dynamics from repeated interactions". In: *Physics Reports* 954 (2022), pp. 1–70 (cit. on pp. 4, 51, 130).

[43] Joshua Combes, Joseph Kerckhoff, and Mohan Sarovar. "The SLH framework for modeling quantum input-output networks". In: *Advances in Physics: X* 2.3 (May 2017), pp. 784–888 (cit. on pp. 55, 130).

[44] Harald. Cramér. *Mathematical Methods of Statistics*. 1946 (cit. on p. 1).

[45] Jean Dalibard, Yvan Castin, and Klaus Mølmer. "Wave-function approach to dissipative processes in quantum optics". In: *Physical Review Letters* 68.5 (Feb. 1992), pp. 580–583 (cit. on pp. 55, 130).

[46] Antonella De Pasquale and Thomas M. Stace. "Quantum Thermometry". In: *Fundamental Theories of Physics* 195 (2018), pp. 503–527 (cit. on pp. 2, 50).

[47] C. L. Degen, F. Reinhard, and P. Cappellaro. "Quantum sensing". In: *Reviews of Modern Physics* 89.3 (July 2017), p. 035002 (cit. on pp. 2, 50).

[48] Rafał Demkowicz-Dobrzański, Jan Czajkowski, and Pavel Sekatski. "Adaptive Quantum Metrology under General Markovian Noise". In: *Physical Review X* 7.4 (Oct. 2017), p. 041009 (cit. on pp. 55, 130).

[49] Rafał Demkowicz-Dobrzański, Wojciech Górecki, and Mădălin Guţă. "Multi-parameter estimation beyond quantum Fisher information". In: *Journal of Physics A: Mathematical and Theoretical* 53.36 (Sept. 2020), p. 363001 (cit. on pp. 2, 37, 46–48, 59, 63, 88, 106, 110, 134).

[50] Rafał Demkowicz-Dobrzański, Jan Kołodyński, and Mădălin Guţă. "The elusive Heisenberg limit in quantum-enhanced metrology". In: *Nature Communications* 3.1 (Sept. 2012), p. 1063 (cit. on pp. 49, 50, 55).

[51] B Efron and D V Hinkley. "Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher Information". In: *Biometrika* 65 (1978), 457–487 (cit. on p. 76).

[52] A. Einstein, B. Podolsky, and N. Rosen. "Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?" In: *Physical Review* 47.10 (May 1935), pp. 777–780 (cit. on p. 15).

[53] B. M. Escher, R. L. de Matos Filho, and L. Davidovich. "General framework for estimating the ultimate precision limit in noisy quantum-enhanced metrology". In: *Nature Physics* 7.5 (May 2011), pp. 406–411 (cit. on p. 55).

[54] Franco Fagnola and Rely Pellicer. "Irreducible and periodic positive maps". In: *Communications on Stochastic Analysis* 3.3 (Dec. 2009) (cit. on pp. 4, 21, 25).

[55] Philippe Faist et al. "Time-Energy Uncertainty Relation for Noisy Quantum Metrology". In: *PRX Quantum* 4.4 (Dec. 2023), p. 040336 (cit. on p. 47).

[56] Alessio Fallani et al. "Learning Feedback Control Strategies for Quantum Metrology". In: *PRX Quantum* 3.2 (Apr. 2022), p. 020310 (cit. on pp. 118, 131).

[57] M Fannes, B Nachtergaele, and R F Werner. "Finitely correlated states on quantum spin chains". In: *Communications in Mathematical Physics* 144 (1992), pp. 443–490 (cit. on p. 130).

[58] M Fannes, B Nachtergaele, and R F Werner. "Finitely correlated pure states". In: *Journal of Functional Analysis* 120 (1994), pp. 511–534 (cit. on p. 130).

[59] Akio Fujiwara and Koichi Yamagata. "Noncommutative Lebesgue decomposition and contiguity with applications in quantum statistics". In: *Bernoulli* 26.3 (Aug. 2020), pp. 2105–2142 (cit. on p. 134).

[60] Akio Fujiwara and Koichi Yamagata. "Efficiency of estimators for locally asymptotically normal quantum statistical models". In: *The Annals of Statistics* 51.3 (June 2023), pp. 1159–1182 (cit. on p. 134).

[61] JOHN GOUGH. "Holevo-Ordering and the Continuous-Time Limit for Open Floquet Dynamics". In: *Letters in Mathematical Physics* 67.3 (Mar. 2004), pp. 207–221 (cit. on pp. 51, 130).

[62] Jay Gambetta and H M Wiseman. "State and dynamical parameter estimation for open quantum systems". In: *Physical Review A* 64.4 (Sept. 2001), p. 042105 (cit. on pp. 55, 118, 130).

[63] S. Gammelmark et al. "Hidden Markov model of atomic quantum jump dynamics in an optically probed cavity". In: *Physical Review A* 89.4 (Apr. 2014), p. 043839 (cit. on p. 84).

[64] Søren Gammelmark and Klaus Mølmer. "Bayesian parameter inference from continuously monitored quantum systems". In: *Physical Review A* 87.3 (Mar. 2013), p. 032115 (cit. on pp. 55, 131).

[65] Søren Gammelmark and Klaus Mølmer. "Fisher Information and the Quantum Cramér-Rao Sensitivity Limit of Continuous Measurements". In: *Physical Review Letters* 112.17 (Apr. 2014), p. 170401 (cit. on pp. 55, 61, 118, 131).

[66] C W Gardiner and P Zoller. *Quantum Noise: A Handbook of Markovian and Non-Markovian Quantum Stochastic Methods with Applications to Quantum Optics*. 3rd ed. Springer-Verlag, 2004 (cit. on pp. 55, 130).

[67] Juan P Garrahan and Igor Lesanovsky. "Thermodynamics of Quantum Jump Trajectories". In: *Physical Review Letters* 104.16 (Apr. 2010), p. 160601 (cit. on p. 131).

[68] M G Genoni. "Cramér-Rao bound for time-continuous measurements in linear Gaussian quantum systems". In: *Phys. Rev. A* 95.1 (2017), p. 12116 (cit. on pp. 55, 131).

[69] Manuel Gessner, Luca Pezzè, and Augusto Smerzi. "Sensitivity Bounds for Multiparameter Quantum Metrology". In: *Physical Review Letters* 121.13 (Sept. 2018), p. 130503 (cit. on p. 2).

[70] Richard D. Gill and Mădălin I. Guţă. "On asymptotic quantum statistical inference". In: *IMS Collections*. Vol. 9. IMS Collections. Institute of Mathematical Statistics, Jan. 2013, pp. 105–127 (cit. on pp. 88, 91, 101, 106, 134).

[71] Richard D. Gill and Serge Massar. "State estimation for large ensembles". In: *Physical Review A* 61.4 (Mar. 2000), p. 042312 (cit. on pp. 57, 59, 63, 83, 84, 91, 134).

[72] Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone. "Quantum Metrology". In: *Physical Review Letters* 96.1 (Jan. 2006), p. 010401 (cit. on pp. 2, 48–50).

[73] Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone. "Advances in quantum metrology". In: *Nature Photonics* 5.4 (Apr. 2011), pp. 222–229 (cit. on pp. 2, 48–50).

[74] Federico Girotti, Juan P Garrahan, and Măd˘ alin Gu\c t\u a. "Concentration Inequalities for Output Statistics of Quantum Markov Processes". In: *Annales Henri Poincaré* (2023) (cit. on pp. 131, 132, 147).

[75] Federico Girotti, Alfred Godley, and Mădălin Guţă. "Estimating quantum Markov chains using coherent absorber post-processing and pattern counting estimator". arxiv:2408.00626 (cit. on pp. ii, 6, 84, 119, 183).

[76] Federico Girotti, Alfred Godley, and Mădălin Guţă. "Optimal estimation of pure states with displaced-null measurements". In: *Journal of Physics A: Mathematical and Theoretical* 57.24 (June 2024), p. 245304 (cit. on pp. 131, 132, 134, 135, 147–149, 153).

[77] Federico Girotti, Alfred Godley, and Mădălin Guţă. "Optimal estimation of pure states with displaced-null measurements". In: *Journal of Physics A: Mathematical and Theoretical* 57 (24 June 2024), p. 245304 (cit. on pp. ii, 5, 183).

[78] Alfred Godley and Madalin Guta. "Adaptive measurement filter: efficient strategy for optimal estimation of quantum Markov chains". In: *Quantum* 7 (Apr. 2023), p. 973 (cit. on pp. ii, 4, 5, 119, 131, 132, 138, 153, 183).

[79] Wojciech Górecki et al. "Optimal probes and error-correction schemes in multi-parameter quantum metrology". In: *Quantum* 4 (July 2020), p. 288 (cit. on pp. 50, 130).

[80] Wojciech Górecki et al. "π-Corrected Heisenberg Limit". In: *Physical Review Letters* 124.3 (Jan. 2020), p. 030501 (cit. on pp. 83, 85).

[81] J Gough and M.R. James. "The Series Product and Its Application to Quantum Feedforward and Feedback Networks". In: *IEEE Transactions on Automatic Control* 54.11 (Nov. 2009), pp. 2530–2544 (cit. on pp. 56, 130).

[82] I Guevara and H Wiseman. "Quantum State Smoothing". In: *Phys. Rev. Lett.* 115.18 (2015), p. 180407 (cit. on pp. 55, 131).

[83] Madalin Guta and Jukka Kiukas. "Equivalence Classes and Local Asymptotic Normality in System Identification for Quantum Markov Chains". In: *Communications in Mathematical Physics* 335.3 (May 2015), pp. 1397–1428 (cit. on pp. 3, 51, 55, 61, 62, 84, 118, 131, 137).

[84] Madalin Guta and Jukka Kiukas. "Information geometry and local asymptotic normality for multi-parameter estimation of quantum Markov dynamics". In: *Journal of Mathematical Physics* 58.5 (May 2017), p. 52201 (cit. on pp. 3, 51, 55, 61, 84, 118, 131).

[85] Madalin Guta and Naoki Yamamoto. "System Identification for Passive Linear Quantum Systems". In: *IEEE Transactions on Automatic Control* 61.4 (Apr. 2016), pp. 921–936 (cit. on pp. 55, 131).

[86] Mădălin Guţă. "Fisher information and asymptotic normality in system identification for quantum Markov chains". In: *Physical Review A* 83.6 (June 2011), p. 062324 (cit. on pp. 51, 55, 61, 118, 131).

[87] Mădălin Guţă, Bas Janssens, and Jonas Kahn. "Optimal Estimation of Qubit States with Continuous Time Measurements". In: *Communications in Mathematical Physics* 277.1 (Nov. 2007), pp. 127–160 (cit. on pp. 102, 132, 134).

[88] Mădălin Guţă and Anna Jenčová. "Local Asymptotic Normality in Quantum Statistics". In: *Communications in Mathematical Physics* 276.2 (Oct. 2007), pp. 341–379 (cit. on pp. 42, 44, 51, 88, 100, 101, 134).

[89] Mădălin Guţă and Jonas Kahn. "Local asymptotic normality for qubit states". In: *Physical Review A* 73.5 (May 2006), p. 052108 (cit. on pp. 45, 51, 88, 100–103, 132, 134).

[90] Timothy C Ralph Hans-A. Bachor. *A Guide to Experiments in Quantum Optics*. Wiley-VCH Verlag GmbH & Co. KGaA, 2019 (cit. on p. 119).

[91] S Haroche and J.-M. Raimond. *Exploring the Quantum: atoms, cavities and photons*. Oxford University Press, 2006 (cit. on pp. 4, 51, 130).

[92] Masahito Hayashi. *Asymptotic Theory of Quantum Statistical Inference*. WORLD SCIENTIFIC, Feb. 2005, pp. 1–542 (cit. on pp. 2, 84, 134).

[93] Carl W. Helstrom. "Quantum detection and estimation theory". In: *Journal of Statistical Physics* 1.2 (1969), pp. 231–252 (cit. on pp. 1, 55, 58, 62, 83, 84, 89, 134).

[94] A. S. Holevo and V. Giovannetti. "Quantum channels and their entropic characteristics". In: *Reports on Progress in Physics* 75.4 (Apr. 2012), p. 046001 (cit. on p. 21).

[95] Alexander Holevo. *Probabilistic and Statistical Aspects of Quantum Theory*. Edizioni della Normale, 2011 (cit. on pp. 1, 48, 55, 58, 62, 83, 84, 87, 89, 100, 106, 134).

[96] Merlijn van Horssen and Mădălin Guţă. "Sanov and central limit theorems for output statistics of quantum Markov chains". In: *Journal of Mathematical Physics* 56.2 (Feb. 2015), p. 22109 (cit. on p. 131).

[97] S F Huelga et al. "Improvement of Frequency Standards with Quantum Entanglement". In: *Physical Review Letters* 79.20 (Nov. 1997), pp. 3865–3868 (cit. on pp. 55, 130).

[98] Theodoros Ilias et al. "Criticality-Enhanced Quantum Sensing via Continuous Measurement". In: *PRX Quantum* 3.1 (Mar. 2022), p. 010354 (cit. on pp. 119, 131).

[99] J. R. Johansson, P. D. Nation, and Franco Nori. "QuTiP: An open-source Python framework for the dynamics of open quantum systems". In: *Computer Physics Communications* 183.8 (2012), pp. 1760–1772 (cit. on p. 149).

[100] J. R. Johansson, P. D. Nation, and Franco Nori. "QuTiP 2: A Python framework for the dynamics of open quantum systems". In: *Computer Physics Communications* 184.4 (2013), pp. 1234–1240 (cit. on p. 149).

[101] Jonathan A. Jones et al. “Magnetic Field Sensing Beyond the Standard Quantum Limit Using 10-Spin NOON States”. In: *Science* 324.5931 (May 2009), pp. 1166–1168 (cit. on pp. 2, 50).

[102] Jonas Kahn and Mădălin Gut,ă. “Local Asymptotic Normality for Finite Dimensional Quantum Systems”. In: *Communications in Mathematical Physics* 289.2 (July 2009), pp. 597–652 (cit. on pp. 51, 88, 100, 101, 132, 134).

[103] Michael Keyl. “Fundamentals of quantum information theory”. In: *Physics Reports* 369.5 (Oct. 2002), pp. 431–548 (cit. on p. 21).

[104] Alexander Holm Kiilerich and Klaus Mølmer. “Estimation of atomic interaction parameters by photon counting”. In: *Physical Review A* 89.5 (May 2014), p. 052110 (cit. on p. 55).

[105] Alexander Holm Kiilerich and Klaus Mølmer. “Bayesian parameter estimation by continuous homodyne detection”. In: *Physical Review A* 94.3 (Sept. 2016), p. 032103 (cit. on pp. 55, 131).

[106] J Kiukas and Mădălin Guţă. “in preparation”. 2022 (cit. on p. 62).

[107] David Layden et al. “Ancilla-free quantum error correction codes for quantum metrology”. In: *Physical Review Letters* 122.4 (Nov. 2018) (cit. on p. 50).

[108] E L Lehmann and G Casella. *Theory of Point Estimation*. Springer Texts in Statistics. Springer-Verlag, 1998 (cit. on pp. 34, 37, 74, 75).

[109] U Leonhardt. *Measuring the Quantum State of Light*. Cambridge University Press, 1997 (cit. on p. 88).

[110] U. Leonhardt and H. Paul. “Measuring the quantum state of light”. In: *Progress in Quantum Electronics* 19.2 (Jan. 1995), pp. 89–130 (cit. on pp. 8, 17).

[111] I Lesanovsky et al. “Characterization of Dynamical Phase Transitions in Quantum Jump Trajectories Beyond the Properties of the Stationary State”. In: *Phys. Rev. Lett.* 110 (2013), p. 150401 (cit. on pp. 58, 79).

[112] Matthew Levitt and Mădălin Guţă. “Identification of single-input–single-output quantum linear systems”. In: *Physical Review A* 95.3 (Mar. 2017), p. 033825 (cit. on pp. 55, 131).

[113] Matthew Levitt, Mădălin Guţă, and Hendra I Nurdin. “Power spectrum identification for quantum linear systems”. In: *Automatica* 90 (Apr. 2018), pp. 255–262 (cit. on pp. 55, 131).

[114] Gwo Dong Lin. “Recent developments on the moment problem”. In: *Journal of Statistical Distributions and Applications* 4.1 (2017), p. 5 (cit. on p. 156).

[115] Jing Liu et al. “Quantum Fisher information matrix and multiparameter estimation”. In: *Journal of Physics A: Mathematical and Theoretical* 53.2 (Jan. 2020), p. 023001 (cit. on pp. 83–85, 88, 92, 105, 118, 135).

[116] L Ljung. *System Identification: Theory for the User*. 2nd ed. Prentice-Hall, 1999 (cit. on p. 55).

[117] H Mabuchi. "Dynamical identification of open quantum systems". In: *Quantum and Semiclassical Optics: Journal of the European Optical Society Part B* 8.6 (Dec. 1996), pp. 1103–1108 (cit. on pp. 55, 130).

[118] Katarzyna Macieszczak et al. "Dynamical phase transitions as a resource for quantum enhanced metrology". In: *Physical Review A* 93.2 (Feb. 2016), p. 022103 (cit. on pp. 55, 131).

[119] Daniel Manzano. "A short introduction to the Lindblad master equation". In: *AIP Advances* 10.2 (Feb. 2020) (cit. on pp. 31–33).

[120] K Matsumoto. "A new approach to the Cramér-Rao-type bound of the pure-state model". In: *Journal of Physics A: Mathematical and General* 35.13 (Apr. 2002), pp. 3111–3123 (cit. on pp. 114, 118, 125, 126).

[121] Yuichiro Matsuzaki. "Robust Quantum Sensing". In: (2021), pp. 289–314 (cit. on p. 2).

[122] Yuichiro Matsuzaki, Simon C. Benjamin, and Joseph Fitzsimons. "Magnetic field sensing beyond the standard quantum limit under the effect of decoherence". In: *Physical Review A* 84.1 (July 2011), p. 012103 (cit. on p. 1).

[123] Mohammad Mehboudi, Anna Sanpera, and Luis A. Correa. "Thermometry in the quantum regime: recent theoretical progress". In: *Journal of Physics A: Mathematical and Theoretical* 52.30 (July 2019), p. 303001 (cit. on pp. 2, 50).

[124] H Nagaoka. "A new approach to Cramer–Rao bounds for quantum state estimation". In: *Asymptotic Theory of Quantum Statistical Inference: Selected Papers*. World Scientific, 2005, 100–112 (cit. on pp. 83, 84).

[125] M. Napolitano et al. "Interaction-based quantum metrology showing scaling beyond the Heisenberg limit". In: *Nature* 471.7339 (Mar. 2011), pp. 486–489 (cit. on p. 50).

[126] A Negretti and K M\olmer. "Estimation of classical parameters via continuous probing of complementary quantum observables". In: *New J. Phys.* 15 (2013), p. 125002 (cit. on pp. 55, 131).

[127] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, June 2012 (cit. on pp. 9, 14, 21).

[128] T J Osborne, J Eisert, and F Verstraete. "Holographic Quantum States". In: *Phys. Rev. Lett.* 105 (2010), p. 260401 (cit. on pp. 58, 79).

[129] Shengshi Pang and Todd A. Brun. "Quantum metrology for a general Hamiltonian parameter". In: *Physical Review A* 90.2 (Aug. 2014), p. 022117 (cit. on pp. 49, 84).

[130] Matteo G.A. Paris. "Quantum Estimation for Quantum Technology". In: *International Journal of Quantum Information* 07.supp01 (Jan. 2009), pp. 125–137 (cit. on pp. 1, 46, 134).

[131] Asher Peres. "Measurement of time by quantum clocks". In: *American Journal of Physics* 48.7 (July 1980), pp. 552–557 (cit. on p. 2).

[132] D Perez-Garcia et al. "Matrix product state representations". In: *Quantum Info. Comput.* 7 (2007), 401–430 (cit. on p. 130).

[133] D Petz. "Sufficient subalgebras and the relative entropy of states of a von Neumann algebra". In: *Communications in Mathematical Physics* 105 (1986), 123–131 (cit. on p. 139).

[134] D Petz. "Sufficiency of channels over von Neumann algebras". In: *Quarterly Journal of Mathematics* 39 (1988), 97–108 (cit. on p. 139).

[135] Dénes Petz. *Quantum Information Theory and Quantum Statistics*. Theoretical and Mathematical Physics. 2008, pp. 1–209 (cit. on pp. 52, 102).

[136] Luca Pezze' and Augusto Smerzi. "Quantum theory of phase estimation". In: *Atom Interferometry* 188 (Nov. 2014), p. 691 (cit. on pp. 83–85, 88, 92, 105, 118, 135).

[137] Luca Pezzè et al. "Optimal Measurements for Simultaneous Quantum Estimation of Multiple Phases". In: *Physical Review Letters* 119.13 (Sept. 2017), p. 130504 (cit. on pp. 83–85, 88, 92, 105, 116, 118, 135).

[138] Martin B Plenio and Susana F Huelga. "Sensing in the presence of an observed environment". In: *Physical Review A* 93.3 (Mar. 2016), p. 032123 (cit. on pp. 55, 131).

[139] Martin B. Plenio and Shashank Virmani. "An introduction to entanglement measures". In: *Quantum Information and Computation* 7.1-2 (Apr. 2005), pp. 1–51 (cit. on pp. 16, 17).

[140] D T Pope, H M Wiseman, and N K Langford. "Adaptive phase estimation is more accurate than nonadaptive phase estimation for continuous beams of light". In: *Phys. Rev. A* 70 (2004), p. 43812 (cit. on pp. 55, 131).

[141] Sammy Ragy, Marcin Jarzyna, and Rafał Demkowicz-Dobrzański. "Compatibility in multiparameter quantum metrology". In: *Physical Review A* 94.5 (Nov. 2016), p. 052108 (cit. on p. 106).

[142] J F Ralph, K Jacobs, and C D Hill. "Frequency tracking and parameter estimation for robust quantum state estimation". In: *Phys. Rev. A* 84 (2011), p. 52119 (cit. on pp. 55, 131).

[143] J F Ralph, S Maskell, and K Jacobs. "Multiparameter estimation along quantum trajectories with sequential Monte Carlo methods". In: *Phys. Rev. A* 96 (2017), p. 52306 (cit. on pp. 55, 131).

[144] Ángel Rivas and Susana F. Huelga. "Open Quantum Systems. An Introduction". In: SpringerBriefs in Physics (Apr. 2011) (cit. on pp. 3, 31, 32).

[145] Jesús Rubio, Janet Anders, and Luis A. Correa. "Global Quantum Thermometry". In: *Physical Review Letters* 127.19 (Nov. 2021), p. 190402 (cit. on p. 2).

[146] J. J. Sakurai and Jim Napolitano. *Modern Quantum Mechanics*. Cambridge University Press, Sept. 2017 (cit. on pp. 9, 17).

[147] Konrad Schmüdgen. *Ten Lectures on the Moment Problem*. 2020 (cit. on p. 156).

[148] Roman Schnabel. "Squeezed states of light and their applications in laser interferometers". In: *Physics Reports* 684 (2017), pp. 1–51 (cit. on p. 119).

[149] C Schön et al. “Sequential Generation of Entangled Multiqubit States”. In: *Phys. Rev. Lett.* 95.11 (2005), p. 110503 (cit. on p. 130).

[150] Pavel Sekatski et al. “Quantum metrology with full and fast quantum control”. In: *Quantum* 1 (Sept. 2017), p. 27 (cit. on p. 130).

[151] Luigi Seveso and Matteo G. A. Paris. “Quantum enhanced metrology of Hamiltonian parameters beyond the Cram\‘er-Rao bound”. In: *International Journal of Quantum Information* 18.3 (Mar. 2020), p. 2030001 (cit. on pp. 2, 34–37, 46, 47, 49).

[152] Luigi Seveso, Matteo A. C. Rossi, and Matteo G. A. Paris. “Quantum metrology beyond the quantum Cramér-Rao theorem”. In: *Physical Review A* 95.1 (Jan. 2017), p. 012111 (cit. on pp. 34, 46).

[153] Jasminder S. Sidhu and Pieter Kok. “Geometric perspective on quantum parameter estimation”. In: *AVS Quantum Science* 2.1 (Feb. 2020), p. 14701 (cit. on p. 134).

[154] P Six et al. “Parameter estimation from measurements along quantum trajectories”. In: *2015 54th IEEE Conference on Decision and Control (CDC)*. 2015, pp. 7742–7748 (cit. on pp. 55, 131).

[155] Andrea Smirne et al. “Ultimate Precision Limits for Noisy Frequency Estimation”. In: *Physical Review Letters* 116.12 (Mar. 2016), p. 120801 (cit. on pp. 55, 130).

[156] Alexander R.H. Smith and Mehdi Ahmadi. “Quantizing time: Interacting clocks and systems”. In: *Quantum* 3 (July 2019), p. 160 (cit. on p. 2).

[157] K Stannigel, P Rabl, and P Zoller. “Driven-dissipative preparation of entangled states in cascaded quantum-optical networks”. In: *New Journal of Physics* 14.6 (June 2012), p. 063014 (cit. on pp. 4, 57, 62, 63, 78, 131, 138, 153).

[158] M Szczykulska, T Baumgratz, and A Datta. “Multi-parameter quantum metrology”. In: *Advances in Physics: X* 1 (2016), pp. 621–639 (cit. on p. 2).

[159] J. M. Taylor et al. “High-sensitivity diamond magnetometer with nanoscale resolution”. In: *Nature Physics 2008 4:10* 4.10 (Sept. 2008), pp. 810–816 (cit. on pp. 2, 50).

[160] Géza Tóth and Iagoba Apellaniz. “Quantum metrology from a quantum information science perspective”. In: *Journal of Physics A: Mathematical and Theoretical* 47.42 (Oct. 2014), p. 424006 (cit. on pp. 2, 134).

[161] M Tsang. “Time-Symmetric Quantum Theory of Smoothing”. In: *Phys. Rev. Lett.* 102 (2009), p. 250403 (cit. on pp. 55, 131).

[162] M Tsang. “Quantum reversal: a general theory of coherent quantum absorbers”. 2024 (cit. on pp. 140, 153).

[163] M Tsang and R Nair. “Fundamental quantum limits to waveform detection”. In: *Phys. Rev. A* 86.4 (2012), p. 42115 (cit. on p. 55).

[164] Makei Tsang. “Caveats of the Crámer-Rao bound” (cit. on pp. 83, 85).

[165] Mankei Tsang. “Optimal waveform estimation for classical and quantum systems via time-symmetric smoothing”. In: *Physical Review A* 80.3 (Sept. 2009), p. 033840 (cit. on pp. 55, 131).

[166] Mankei Tsang. "Optimal waveform estimation for classical and quantum systems via time-symmetric smoothing. II. Applications to atomic magnetometry and Hardy's paradox". In: *Physical Review A* 81.1 (Jan. 2010), p. 013824 (cit. on pp. 55, 131).

[167] Mankei Tsang, Howard M Wiseman, and Carlton M. Caves. "Fundamental Quantum Limit to Waveform Estimation". In: *Physical Review Letters* 106.9 (Mar. 2011), p. 090401 (cit. on p. 55).

[168] M. Tse et al. "Quantum-Enhanced Advanced LIGO Detectors in the Era of Gravitational-Wave Astronomy". In: *Physical Review Letters* 123.23 (Dec. 2019), p. 231107 (cit. on pp. 2, 50, 51).

[169] Thomas Unden et al. "Quantum metrology enhanced by repetitive quantum error correction". In: *Physical Review Letters* 116.23 (Feb. 2016) (cit. on p. 50).

[170] A. W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, Oct. 1998 (cit. on pp. 2, 34, 36–39, 41–43, 45, 84, 99).

[171] F Verstraete and J I Cirac. "Continuous Matrix Product States for Quantum Fields". In: *Physical Review Letters* 104.19 (May 2010), p. 190405 (cit. on pp. 51, 130).

[172] John Watrous. *The Theory of Quantum Information*. Apr. 2018, pp. 1–590 (cit. on p. 26).

[173] H M Wiseman and G J Milburn. "Quantum theory of field-quadrature measurements". In: *Physical Review A* 47.1 (Jan. 1993), pp. 642–662 (cit. on pp. 55, 130).

[174] Howard M Wiseman and Gerard J Milburn. *Quantum Measurement and Control*. Cambridge University Press, Nov. 2009 (cit. on pp. 55, 130).

[175] Michael M Wolf. *Quantum Channels & Operations Guided Tour*. Tech. rep. 2012 (cit. on pp. 4, 14, 21, 23).

[176] Koichi Yamagata, Akio Fujiwara, and Richard D. Gill. "Quantum local asymptotic normality based on a new quantum likelihood ratio". In: *The Annals of Statistics* 41.4 (Aug. 2013), p. 2197 (cit. on p. 134).

[177] Dayou Yang, Susana F. Huelga, and Martin B. Plenio. "Efficient Information Retrieval for Sensing via Continuous Measurement". In: *Physical Review X* 13.3 (July 2023), p. 031012 (cit. on pp. 4, 31, 84, 119, 131, 132, 138, 153).

[178] G A Young and R L Smith. *Essentials of Statistical Inference*. Cambridge University Press, 2005 (cit. on p. 99).

[179] H. Yuen and M. Lax. "Multiple-parameter quantum estimation and measurement of nonselfadjoint observables". In: *IEEE Transactions on Information Theory* 19.6 (Nov. 1973), pp. 740–750 (cit. on pp. 83, 84, 134).

[180] Cheng Zhang et al. "Estimation of parameters in circuit QED by continuous quantum measurement". In: *Phys. Rev. A* 99 (2019), p. 22114 (cit. on pp. 55, 131).

[181] Sisi Zhou and Liang Jiang. "Optimal approximate quantum error correction for quantum metrology". In: *Physical Review Research* 2.1 (Mar. 2020), p. 013235 (cit. on p. 50).

[182] Sisi Zhou and Liang Jiang. "Asymptotic Theory of Quantum Channel Estimation". In: *PRX Quantum* 2.1 (Mar. 2021), p. 010343 (cit. on pp. 2, 4, 48).

[183] Sisi Zhou, Chang-Ling Zou, and Liang Jiang. "Saturating the quantum Cramér–Rao bound using LOCC". In: *Quantum Science and Technology* 5.2 (Mar. 2020), p. 025005 (cit. on pp. 56, 58–60, 63–65, 78, 79).

[184] Sisi Zhou et al. "Achieving the Heisenberg limit in quantum metrology using quantum error correction". In: *Nature Communications* 9.1 (Jan. 2018), p. 78 (cit. on pp. 49, 50, 55, 84, 90, 130).

[185] Eric R. Ziegel, E. L. Lehmann, and George Casella. "Theory of Point Estimation". In: *Technometrics* 41.3 (Aug. 1999), p. 274 (cit. on pp. 84, 98).

[186] Analia Zwick and Gonzalo A. Álvarez. "Quantum sensing tools to characterize physical, chemical and biological processes with magnetic resonance". In: *Journal of Magnetic Resonance Open* 16-17 (Dec. 2023), p. 100113 (cit. on p. 2).