Adaptive measurement filter: efficient strategy for optimal estimation of quantum Markov chains

Alfred Godley and Mădălin Guţă

School of Mathematical Sciences, University of Nottingham, University Park, NG8 2SD Nottingham, United Kingdom

Continuous-time measurements are instrumental for a multitude of tasks in quantum engineering and quantum control, including the estimation of dynamical parameters of open quantum systems monitored through the environment. However, such measurements do not extract the maximum amount of information available in the output state, so finding alternative optimal measurement strategies is a major open problem.

In this paper we solve this problem in the setting of discrete-time input-output quantum Markov chains. We present an efficient algorithm for optimal estimation of one-dimensional dynamical parameters which consists of an iterative procedure for updating a 'measurement filter' operator and determining successive measurement bases for the output units. A key ingredient of the scheme is the use of a coherent quantum absorber as a way to physically 'post-process' the output after the interaction with the system. The absorber is designed such that the joint system plus absorber stationary state is pure at parameter value provided by a preliminary estimator. The scheme offers an exciting prospect for optimal continuous-time adaptive measurements, but more work is needed to find realistic practical implementations.

1 Introduction

The quantum input-output (I-O) formalism is an effective framework for describing the evolution, monitoring and control of Markovian quantum open systems [1–3]. 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 [4–8].

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 I-O formalism is ideally suited for this statistical inference task, and more generally for implementing online system identification methods [9]. Unlike direct measurement techniques which require repeated system re-preparations and fast control operations [10–16], 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 [17]. Subsequent works have addressed a variety of related problems including the dependence on measurement choice [18], adaptive estimation [19, 20] filtering with uncertain parameters [21] particle filters for estimation [22, 23] achieving Heisenberg scaling [24–26], sensing with error correction [27], Bayesian estimation [28–32], quantum smoothing [33–36], waveform estimation [37, 38] estimation of linear systems [39–42], classical and quantum Fisher informations of the output channel [28, 40, 43–48].

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

[49–51], 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 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 [52].

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 1. We assume that the interaction unitary U_θ acting on $\mathbb{C}^d \otimes \mathbb{C}^k$ depends on a parameter $\theta \in \mathbb{R}$,
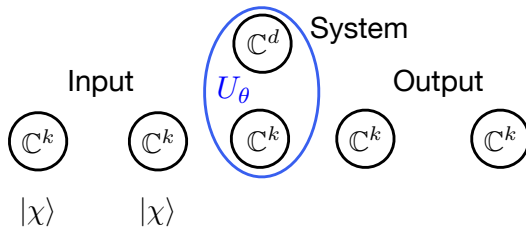

Figure 1: Quantum input-output discrete-time dynamics with θ dependent unitary interaction U_θ

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 [53] to the joint pure system-output state; indeed this has been shown to attain the quantum Cramér-Rao bound. 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 6.1 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 output measurement is optimal in the leading contribution to the 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 [54]. 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 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 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 $D = d^2$
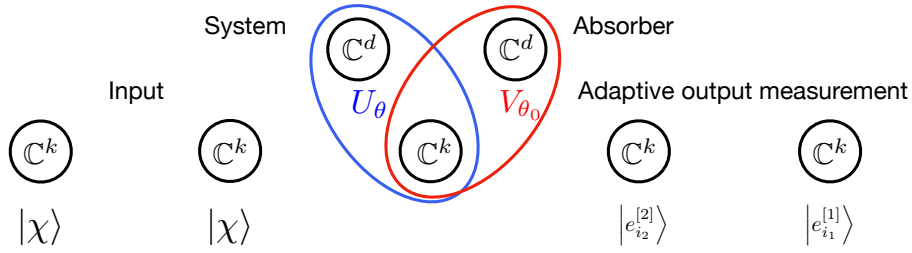
Figure 2: Adaptive measurement filter: the output units undergo post-processing with a coherent quantum absorber, followed by applying an adaptive measurement computed with the algorithm

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 [55], 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 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 1.

In section 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 7.1), 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 [56, 57]. Although we do not treat this in detail here, we comment on this extension at the end of the paper.

The paper is organised as follows. In section 2 we briefly review the adaptive algorithm for optimal separable measurements developed in [53]. Section 3 introduces the Markov dynamics setting and reviews a key result on the asymptotic QFI of the output. Section 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 5, which details the adaptive measurement procedure including the key 'measurement filter' algorithm. In section 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 7 presents simulation results using an elaboration of an amplitude decay qubit model.

2 Optimal separable measurements

In this section we review a result by Zhou, Zou and Jiang [53] 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 quantum Cramér-Rao bound (QCRB) [49–51], the variance of any unbiased estimator $\hat{\theta}$ is lower bounded as

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

where F_θ is the quantum Fisher information (QFI) defined as $F(\theta) = \mathrm{Tr}(\rho_\theta \mathcal{L}_\theta^2)$, and $\mathcal{L}_\theta$ is the symmetric logarithmic derivative (SLD) satisfying $\frac{d}{d\theta}\rho_\theta = \mathcal{L}_\theta \circ \rho_\theta$. In general, for any given parameter value θ_0, the QCRB is saturated[1] by measuring the 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 [51]. 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 [53] 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

$$F_\theta = 4\|\dot{\psi}_\theta\|^2. \tag{1}$$

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

$$M = |\psi_\theta\rangle \langle\dot{\psi}_\theta| - |\dot{\psi}_\theta\rangle \langle\psi_\theta|. \tag{2}$$

The authors of [53] note that if a projective rank-one measurement $\{E_i = |e_i\rangle\langle e_i|\}$ satisfies the conditions

$$\langle e_i|M|e_i\rangle = 0, \quad \text{and} \quad p_\theta(i) = |\langle e_i|\psi_\theta\rangle|^2 = 1/k, \quad k = \dim(\mathcal{H}) \tag{3}$$

then it fulfils the general criteria of [51] 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

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

---

[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 [54, 58].

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 [53] to show that a separable measurement satisfying the above conditions can be constructed by using the algorithm outlined below. For any set of indices $A \subset \{1, \ldots, n\}$ 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 $\{1, \ldots, m\}$. 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 [53]. The concrete construction in two dimensions is described in section 7.

After this, the following procedure is applied sequentially to determine the measurement basis for system $j+1$ with $j = 1, \ldots, n-1$: given the outcomes $i_{\underline{j}} := \{i_1, \ldots, i_j\}$ 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

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

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

$$M_{j+1}(i_{\underline{j}}) = \left\langle e_{i_{\underline{j}}}^{[j]} \right| M_{\underline{j+1}} \left| e_{i_{\underline{j}}}^{[j]} \right\rangle, \quad \rho_{j+1}(i_{\underline{j}}) = \left\langle e_{i_{\underline{j}}}^{[j]} \right| \rho_{\underline{j+1}} \left| e_{i_{\underline{j}}}^{[j]} \right\rangle, \quad \text{and} \quad \left| e_{i_{\underline{j}}}^{[j]} \right\rangle = \left| e_{i_1}^{[1]} \right\rangle \otimes \cdots \otimes \left| e_{i_j}^{[j]} \right\rangle$$

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

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

and one can verify that such a measurement satisfies the general condition (3).

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 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

$$|\Psi(n)\rangle = U(n)|\phi \otimes \chi^{\otimes n}\rangle = U^{(n)} \cdot \ldots \cdot U^{(2)} \cdot U^{(1)} |\phi \otimes \chi^{\otimes n}\rangle \in \mathcal{H}_s \otimes \mathcal{H}_u^{\otimes n} \tag{4}$$

where $U^{(i)}$ is the unitary acting on the system and the i-th noise unit. From equation (4) we find that the reduced state of the system at time n is given by

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

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

$$|\Psi(n)\rangle = \sum_{i_1, \ldots, i_n = 1}^{k} K_{i_n} \ldots K_{i_1} |\phi\rangle \otimes |i_1\rangle \otimes \cdots \otimes |i_n\rangle \qquad (5)$$

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

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

The problem of estimating θ in this formulation has been investigate in both the discrete time [43, 47] and the continuous time [44, 46, 48] settings. For our purposes, we summarise here the relevant results of [47]. 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. $T(\rho_{\text{ss}}) = \rho_{\text{ss}}$), 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 $T^n(\rho_{\text{in}}) \to \rho_{\text{ss}}$ 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 3.1. *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 $\rho_\theta^{\text{out}}(n)$ scales linearly with n and its rate is equal to*

$$\lim_{n \to \infty} \frac{1}{n} F_\theta(n) = f_\theta = 4 \sum_{i=1}^{k} \left[\text{Tr}\left[\rho_{ss} \dot{K}_i^* \dot{K}_i\right] + 2\text{Tr}\left[\text{Im}(K_i \rho_{ss} \dot{K}_i^*) \cdot \mathcal{R}(\text{Im}\sum_j \dot{K}_j^* K_j)\right] \right] \qquad (6)$$

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

Following standard quantum Cramér-Rao theory [49–51], 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 [47] 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 CR bound 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 3.1 which shows that the same result holds for rank-deficient stationary states of ergodic chains, and in particular for pure states [59].

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 [44]. 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 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 [55] for continuous-time dynamics. In section 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 3 characterised by a unitary U on $\mathcal{H}_s \otimes \mathcal{H}_u$. We now modify the setup as illustrated in Figure 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 [55].

Lemma 4.1. *Any given primitive Markov 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 $\rho_{\mathrm{ss}} = \sum_i \lambda_i |f_i\rangle\langle f_i|$ 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 $\mathrm{Tr}_a(|\tilde{\psi}\rangle\langle\tilde{\psi}|) = \rho_{\mathrm{ss}}$ 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. $\square$

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 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 [54, 58], 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 $\delta\theta = \theta - \theta_0 = O(n^{-1/2+\epsilon})$ [44], 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 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 [53]. 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 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 and Lemma 4.1 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 $V_{\theta_0}^{(n)} \ldots V_{\theta_0}^{(1)}$. 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). 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 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

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

and

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

In Appendix 9 we show that the adaptive measurement of [53] reduces to the following iterative algorithm which is conceptually similar to quantum state filtering [4, 5], 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:

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

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

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

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

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

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

$$\left\langle e_i^{[j]}\right| B_j \left|e_i^{[j]}\right\rangle = 0, \qquad \text{and} \qquad \left|\langle e_i^{[j]}|\chi\rangle\right|^2 = \frac{1}{k}, \qquad \text{for all } i = 1, \ldots k. \tag{7}$$

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

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

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

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

The system and absorber is 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

$$p_\theta(i_1, \ldots i_n, i_0) = \left|\left\langle e_{i_0}^{[s+a]} \otimes e_{i_1}^{[1]} \otimes \cdots \otimes e_{i_n}^{[n]} \mid \Psi_\theta(n)\right\rangle\right|^2. \tag{8}$$

For later use, we denote by $p_\theta(i_1, \ldots i_n)$ the marginal distribution of the output measurement record only.

6   Fisher informations considerations

In this section we investigate the relationship between the classical Fisher information (CFI) 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.

6.1   Achievability of the QFI with adaptive output measurements

The adaptive measurement scheme described in section 5 insures the CFI of the full measurement (output and s+a) is equal to the QFI of the full pure state model

$$I_{\theta_0}^{(s+a+o)}(n) = F_{\theta_0}^{(s+a+o)}(n) = F_{\theta_0}^{(s+o)}(n) \tag{9}$$

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 6.1 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 7 we will illustrate the result on a specific model.

Proposition 6.1. *Consider the setup described in section 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 (6).

6.2 Computing the classical Fisher information of the output

The *classical Fisher information* of the output measurement process at θ_0 is

$$I^{(\mathrm{o})}(n) = \mathbb{E}_{\theta_0}\left(\frac{d\log p_\theta}{d\theta}\right)^2 = \sum_{i_1,\ldots,i_n} p_{\theta_0}(i_1,\ldots i_n)^{-1}\left(\left.\frac{dp_\theta(i_1,\ldots,i_n)}{d\theta}\right|_{\theta_0}\right)^2$$

where the sum runs over indices such that $p_\theta(i_1,\ldots,i_n) > 0$. In general

$$I^{(\mathrm{o})}(n) \le F^{(\mathrm{o})}(n) \le F^{(\mathrm{s+o})}(n)$$

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 6.1. 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 6.2 below, we will use the fact that at $\theta = \theta_0$ the vector $|\psi\rangle$ is the stationary state, and therefore

$$K_i^{[j]}|\psi\rangle = c_i^{[j]}|\psi\rangle \tag{10}$$

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 3). The proof of Lemma 6.2 can be found in Appendix 11.

Lemma 6.2. *Consider the setup described in section 5. The output CFI at $\theta = \theta_0$ is given by*

$$I^{(\mathrm{o})}(n) = \mathbb{E}_{\theta_0}(f^2) = \sum_{i_1,\ldots,i_n} p_{\theta_0}(i_1,\ldots,i_n)f^2(i_1,\ldots,i_n) \tag{11}$$

where f is the function

$$f(i_1,\ldots,i_n) = 2\mathrm{Re}\sum_{j=1}^n \frac{\langle\psi|K_{i_n}^{[n]}\ldots K_{i_{j+1}}^{[j+1]}\dot{K}_{i_j}^{[j]}|\psi\rangle}{c_{i_j}^{[j]}\ldots c_{i_n}^{[n]}}$$

and the constants $c_{i_j}^{[j]}$ are defined by equation 10. In particular, $I^{(\mathrm{o})}(n)$ can be estimated by computing the empirical average of f^2 over sampled trajectories.

We now consider the case where a (projective) measurement $\{P_i^{(\mathrm{s+a})}\}$ is performed on s+a, after obtaining the output measurement record $(i_1,\ldots,i_n)$. We denote the additional outcome by i_0. The CFI of the full process is

$$I^{(\mathrm{s+a+o})}(n) = \mathbb{E}_{\theta_0}\left(\left.\frac{d\log p_\theta}{d\theta}\right|_{\theta_0}\right)^2$$

where

$$p_\theta(i_1,\ldots,i_n,i_0) = \left\|P_{i_0}^{(\mathrm{s+a})}K_{i_n}^{[n]}\ldots K_{i_1}^{[1]}\psi\right\|^2$$

is the likelihood function of a trajectory augmented by the system measurement outcome i_0. The relevant upper bound in this case is

$$I^{(\mathrm{s+a+o})}(n) \le F^{(\mathrm{s+a+o})}(n) = F^{(\mathrm{s+o})}(n). \tag{12}$$

A similar computation to that of Lemma 6.2 gives the system-output classical Fisher information

$$I^{(\mathrm{s+a+o})}(n) = \mathbb{E}_{\theta_0}(\tilde{f}^2)$$

where $\tilde{f}$ is the function

$$\tilde{f}(i_1,\ldots,i_n,i_0) = 2\mathrm{Re}\sum_{j=1}^n \frac{\langle\psi|P_{i_0}^{(\mathrm{s})}K_{i_n}^{[n]}\ldots K_{i_{j+1}}^{[j+1]}\dot{K}_{i_j}^{[j]}|\psi\rangle}{c_{i_j}^{[j]}\ldots c_{i_n}^{[n]}}$$

For fixed (non-adaptive) measurements, the bound (12) 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 9. This is confirmed by our simulation study which also investigates the performance of the fixed measurement scenario.

7 Numerical simulations

We now test the key properties of the adaptive measurement scheme developed in section 5, in two separate numerical investigations.

The first investigation described in subsections 7.1 and 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 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.

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

$$U_\theta : |00\rangle \longrightarrow \cos(\theta)\sqrt{1-\theta^2}\,|00\rangle + i\sin(\theta)\sqrt{1-\theta^2}\,|10\rangle + \theta\,|11\rangle ,$$
$$U_\theta : |10\rangle \longrightarrow i\sin(\theta)\sqrt{1-\lambda}\,|00\rangle + \cos(\theta)\sqrt{1-\lambda}\,|10\rangle + \sqrt{\lambda}e^{i\phi}\,|01\rangle . \qquad (13)$$

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 5. In both cases, given the measurement record $\{i_1, \ldots, i_n\} \in \{0,1\}^n$, the conditional state of the system is

$$|\psi_n(i_1, \ldots i_n)\rangle = \frac{K_{i_n}^{[n]} \ldots K_{i_1}^{[1]}|\psi\rangle}{\|K_{i_n}^{[n]} \ldots K_{i_1}^{[1]}\psi\|} \qquad (14)$$

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

$$p_\theta(i_1, \ldots i_n) = \|K_{i_n}^{[n]} \ldots K_{i_1}^{[1]}\psi\|^2. \qquad (15)$$

Recall that the optimal measurement basis needs to satisfy the conditions (7) in section 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 5 as $B_j = i\vec{r}^{\,[j]} \cdot \vec{\sigma}$ where $\vec{r}^{\,[j]} = (r_x^{[j]}, r_y^{[j]}, 0)$ is its Bloch vector, and similarly, we let $\pm\vec{s}^{\,[j]}$ be the Bloch vectors of the basis vectors $\{|e_1^{[j]}\rangle, |e_2^{[j]}\rangle\}$

satisfying the conditions (7). Then one finds that the conditions (7) are satisfied if $\vec{s}^{\,[j]}$ is taken to be $\vec{s}^{\,[j]} = (r_y^{[j]}, -r_x^{[j]}, 0)$.

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 6.2 with the *quantum* Fisher information of the system-output state. In the latter scenario, Proposition 6.1 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 3.1, the system-output QFI scales linearly with time, i.e.

$$F_\theta^{(\mathrm{s+o})}(n) = n f_\theta + o(n)$$

where the QFI *rate* is given by the equation (6) and the $o(n)$ term depends on the specific system parameters. Applying this to our model we obtain

$$f_{\theta_0} = \frac{8}{1 - \sqrt{1 - \lambda}}. \tag{16}$$

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 (16) and of the following lemma can be found in Appendix 12.

Lemma 7.1. *The system-output QFI at $\theta = \theta_0$ is given by the formula*

$$F_\theta^{(\mathrm{s+o})}(n) \quad = \quad = \frac{8n}{1-a} + 4\left[\frac{2(a^2 - a^n)}{(1-a)^2} - 2\frac{b^2(a - a^n)}{(1-a)^3} + \frac{2(a^2 - a^{2n})}{b^2}\right] \tag{17}$$

where $a = \sqrt{1 - \lambda}$ and $b = \sqrt{\lambda}$. In particular, the leading term in n is given by (16) while the remaining terms are bounded.

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 7.1 while the CFIs are estimated by sampling using the expression in Lemma 6.2.

The results are illustrated in Figure 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. 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 we plot 2φ on circles of radius increasing linearly with time. We note that the initial steps of the trajectory
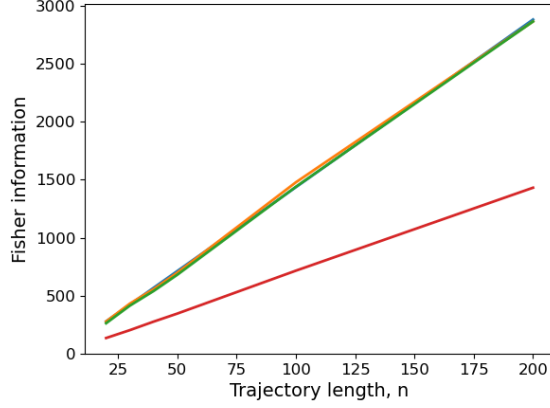
Figure 3: Fisher informations as function of output lenght n: quantum Fisher information (blue), classical Fisher information for the adaptive measurement with/without system measurement (orange/green), classical Fisher information for a regular (non-adaptive) measurement.
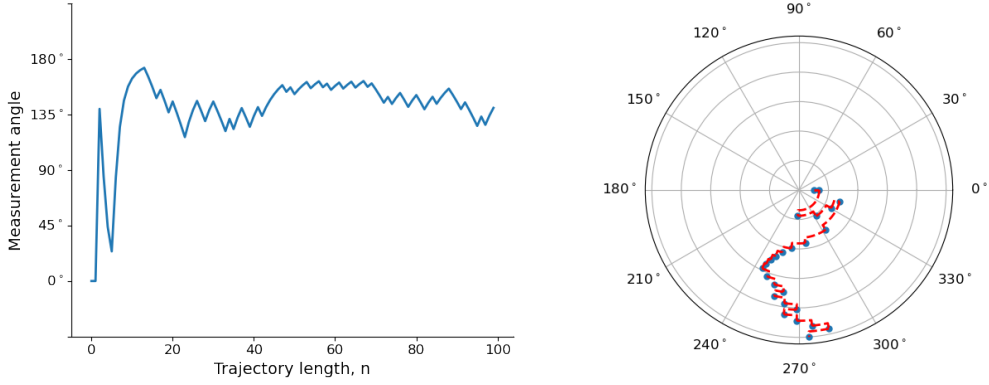

Figure 4: Adaptive measurement trajectory with basis angle φ on left and φ plotted on the circle on the right.

show large variations which then 'stabilise' around a certain range of values. This has to do with the fact that the initial angle can be chosen arbitrarily in this model as $B_1 = 0$, cf. section 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

$$\hat{\theta}_n := \arg\max_{\tau} p_\tau(i_1, \ldots i_n).$$

where the likelihood $p_\tau(i_1, \ldots i_n)$ is computed as in equation (15). In numerics, we maximise the log-likelihood function which can be computed as the sum

$$\log p_\tau(i_i, \ldots i_n) = \sum_{j=1}^{n} \log \|K_{i_j}^{[j]} \psi(i_1, \ldots i_{j-1})\|^2$$

where $\psi(i_1, \ldots i_{j-1})$ is the system's state conditional on the output trajectory (filter), cf. equation (14). 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 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
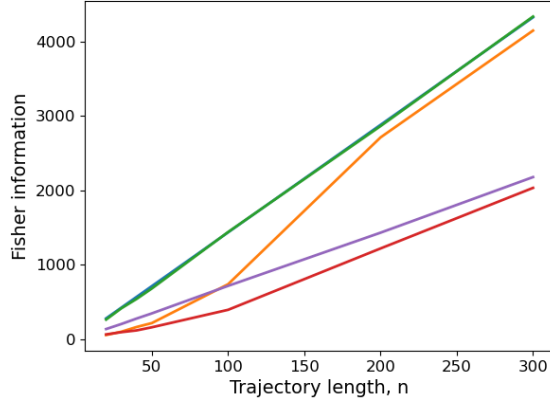
Figure 5: Comparison of different Fisher informations of the output state as function of trajectory length n, obtained by averaging over $N = 10000$ trajectories, at $\lambda = 0.8, \varphi = \pi/4$. The quantum Fisher information (QFI) (blue) and the classical Fisher information (CFI) of adapted measurement (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.
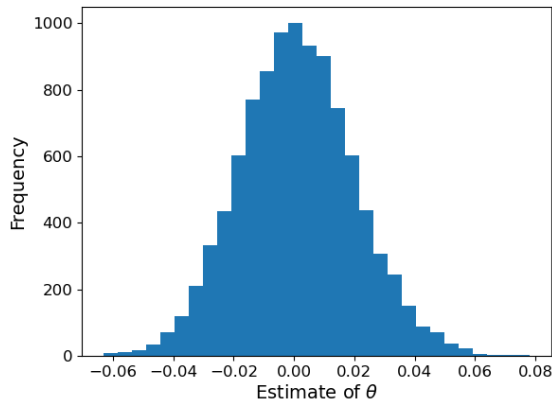

Figure 6: Histogram of the MLE distribution in the adaptive measurement scenario, with $n = 200, \theta = 0$ and 10000 samples, at $\lambda = 0.8, \varphi = \pi/4$.

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 [60, 61]. 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 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* [60, 61] (i.e. has asymptotically normal distribution with variance equal to the inverse of the Fisher information).

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 paper, 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 (13), 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
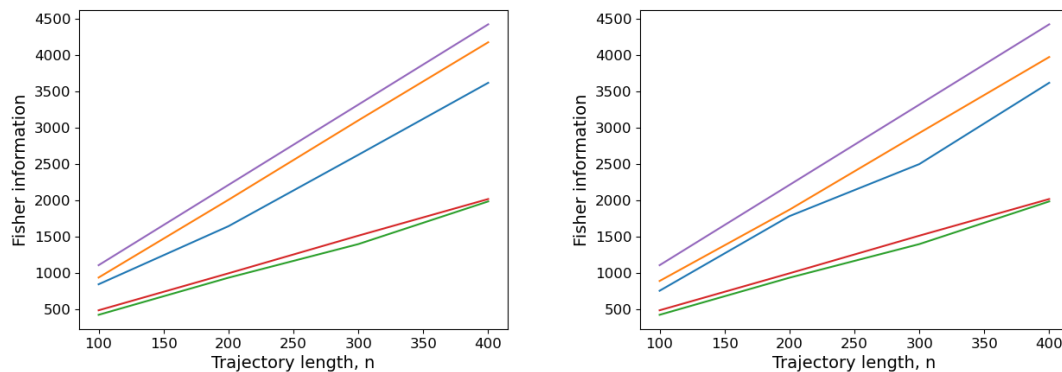
Figure 7: Comparison of different Fisher informations of the output state as function of trajectory length n, in the fully adaptive numerical investigation, for two proportions of samples used in the preliminary stage: $q = 0.15$ (left panel) and $q = 0.25$ (right panel). The results are obtained by 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).

this value the stationary state is mixed and we will apply the full protocol described in section 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). 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 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 $(i_1, \ldots i_n)$, the observed Fisher information is defined as the second derivative of the log-likelihood function evaluated at the maximum likelihood estimator $\hat{\theta}_n$:

$$I_{\text{obs}}(i_1, \ldots i_n) = - \left. \frac{d^2 \log p_\theta(i_1, \ldots i_n)}{d\theta^2} \right|_{\theta = \hat{\theta}_n}$$

While a full theoretical justification of I_{obs} goes beyond the scope of this paper, 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 [62].

Figure 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 3.1, which 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 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.

8 Conclusions and Outlook

In this paper 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 [53] which achieves the quantum Fisher information for pure state models of multipartite systems with one dimensional unknown parameters. However, unlike the scheme of [53] 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 [55] 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 achieve 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' study involving the use of the coherent absorber.

Our discrete-time procedure raises interesting questions about the 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 [56, 57]. 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 behavior 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 processes. 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 direction. 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.

References

[1] 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, Berlin and New York, 2004).

[2] H. M. Wiseman and G. J. Milburn, *Quantum Measurement and Control* (Cambridge University Press, 2010).

[3] J. Combes, J. Kerckhoff, and M. Sarovar, The SLH framework for modeling quantum input-output networks, Adv. Phys. X **2**, 784 (2017).

[4] V. P. Belavkin, Nondemolition principle of quantum measurement theory, Found. Phys. **24**, 685 (1994).

[5] L. Bouten, R. van Handel, and M. R. James, An introduction to quantum filtering, SIAM J. Control Optim. **46**, 2199 (2007).

[6] H. J. Carmichael, *An Open Systems Approach to Quantum Optics* (Springer-Verlag, Berlin, 1993).

[7] J. Dalibard, Y. Castin, and K. Mølmer, Wave-function approach to dissipative processes in quantum optics, Phys. Rev. Lett. **68**, 580 (1992).

[8] H. M. Wiseman and G. J. Milburn, Quantum theory of field-quadrature measurements, Phys. Rev. A **47**, 642 (1993).

[9] L. Ljung, *System Identification: Theory for the User*, 2nd ed. (Prentice-Hall, 1999).

[10] S. F. Huelga, C. Macchiavello, T. Pellizzari, A. K. Ekert, M. B. Plenio, and J. I. Cirac, Improvement of frequency standards with quantum entanglement, Phys. Rev. Lett. **79**, 3865 (1997).

[11] B. Escher, R. de Matos Filho, and L. Davidovich, General framework for estimating the ultimate precision limit in noisy quantum-enhanced metrology, Nature Phys. **7**, 406–411 (2011).

[12] R. Demkowicz-Dobrzański, J. Kołodyński, and M. Guţă, The elusive heisenberg limit in quantum-enhanced metrology, Nature Communications **3**, 1063 (2012).

[13] F. Benatti, S. Alipour, and A. T. Rezakhani, Dissipative quantum metrology in manybody systems of identical particles, New J. Phys. **16**, 015023 (2014).

[14] A. Smirne, J. Kołodyński, S. F. Huelga, and R. Demkowicz-Dobrzański, Ultimate precision limits for noisy frequency estimation, Phys. Rev. Lett. **116**, 120801 (2016).

[15] R. Demkowicz-Dobrzański, J. Czajkowski, and P. Sekatski, Adaptive quantum metrology under general markovian noise, Phys. Rev. X **7**, 041009 (2017).

[16] S. Zhou, M. Zhang, J. Preskill, and L. Jiang, Achieving the heisenberg limit in quantum metrology using quantum error correction, Nature Communications **78**, 78 (2018).

[17] H. Mabuchi, Dynamical identification of open quantum systems, Quantum Semiclass. Opt. **8**, 1103 (1996).

[18] J. Gambetta and H. M. Wiseman, State and dynamical parameter estimation for open quantum systems, Phys. Rev. A **64**, 042105 (2001).

[19] D. W. Berry and H. M. Wiseman, Adaptive quantum measurements of a continuously varying phase, Phys. Rev. A **65**, 043803 (2002).

[20] 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, Phys. Rev. A **70**, 043812 (2004).

[21] J. F. Ralph, K. Jacobs, and C. D. Hill, Frequency tracking and parameter estimation for robust quantum state estimation, Phys. Rev. A **84**, 052119 (2011).

[22] B. Chase and J. M. Geremia, Single-shot parameter estimation via continuous quantum measurement, Phys. Rev. A **79**, 022314 (2009).

[23] P. Six, P. Campagne-Ibarcq, L. Bretheau, B. Huard, and P. Rouchon, Parameter estimation from measurements along quantum trajectories, in *2015 54th IEEE Conference on Decision and Control (CDC)* (2015) pp. 7742–7748.

[24] K. Macieszczak, M. Guta, I. Lesanovsky, and J. P. Garrahan, Dynamical phase transitions as a resource for quantum enhanced metrology, Physical Review A **93**, 022103 (2016).

[25] F. Albarelli, M. A. C. Rossi, M. G. A. Paris, and M. G. Genoni, Ultimate limits for quantum magnetometry via time-continuous measurements, New J. Phys. **19**, 123011 (2017).

[26] F. Albarelli, M. A. C. Rossi, D. Tamascelli, and M. G. Genoni, Restoring heisenberg scaling in noisy quantum metrology by monitoring the environment, Quantum **2**, 110 (2018).

[27] M. B. Plenio and S. F. Huelga, Sensing in the presence of an observed environment, Phys. Rev. A **93**, 032123 (2016).

[28] S. Gammelmark and K. Mølmer, Bayesian parameter inference from continuously monitored quantum systems, Phys. Rev. A **87**, 032115 (2013).

[29] A. Negretti and K. Mølmer, Estimation of classical parameters via continuous probing of complementary quantum observables, New J. Phys. **15**, 125002 (2013).

[30] A. H. Kiilerich and K. Mølmer, Bayesian parameter estimation by continuous homodyne detection, Phys. Rev. A **94**, 032103 (2016).

[31] J. F. Ralph, S. Maskell, and K. Jacobs, Multiparameter estimation along quantum trajectories with sequential monte carlo methods, Phys. Rev. A **96**, 052306 (2017).

[32] C. Zhang, K. Zhou, W. Feng, and X.-Q. Li, Estimation of parameters in circuit qed by continuous quantum measurement, Phys. Rev. A **99**, 022114 (2019).

[33] M. Tsang, Time-symmetric quantum theory of smoothing, Phys. Rev. Lett. **102**, 250403 (2009).

[34] M. Tsang, Optimal waveform estimation for classical and quantum systems via time-symmetric smoothing, Phys. Rev. A **80**, 033840 (2009).

[35] M. Tsang, Optimal waveform estimation for classical and quantum systems via time-symmetric smoothing. ii. applications to atomic magnetometry and hardy's paradox, Phys. Rev. A **81**, 013824 (2010).

[36] I. Guevara and H. Wiseman, Quantum state smoothing, Phys. Rev. Lett. **115**, 180407 (2015).

[37] M. Tsang, H. M. Wiseman, and C. Caves, Fundamental quantum limit to waveform estimation, Phys. Rev. Lett. **106**, 090401 (2011).

[38] M. Tsang and R. Nair, Fundamental quantum limits to waveform detection, Phys. Rev. A **86**, 042115 (2012).

[39] M. Guţă and N. Yamamoto, System identification for passive linear quantum systems, IEEE Trans. Automat. Contr. **61**, 921 (2016).

[40] M. G. Genoni, Cramér-rao bound for time-continuous measurements in linear gaussian quantum systems, Phys. Rev. A **95**, 012116 (2017).

[41] M. Levitt and M. Guţă, Identification of single-input–single-output quantum linear systems, Phys. Rev. A **95**, 033825 (2017).

[42] M. Levitt, M. Guţă, and H. I. Nurdin, Power spectrum identification for quantum linear systems, Automatica **90**, 255 (2018).

[43] M. Guţă, Fisher information and asymptotic normality in system identification for quantum markov chains, Physical Review A **83**, 062324 (2011).

[44] C. Catana, L. Bouten, and M. Guţă, Fisher informations and local asymptotic normality for continuous-time quantum markov processes, Journal of Physics A: Mathematical and Theoretical **48**, 365301 (2015).

[45] A. H. Kiilerich and K. Mølmer, Estimation of atomic interaction parameters by photon counting, Phys. Rev. A **89**, 052110 (2014).

[46] S. Gammelmark and K. Mølmer, Fisher information and the quantum cramér-rao sensitivity limit of continuous measurements, Phys. Rev. Lett. **112**, 170401 (2014).

[47] M. Guţă and J. Kiukas, Equivalence classes and local asymptotic normality in system identification for quantum markov chains, Communications in Mathematical Physics **335**, 1397–1428 (2014).

[48] M. Guţă and J. Kiukas, Information geometry and local asymptotic normality for multiparameter estimation of quantum markov dynamics, J. Math. Phys. **58**, 052201 (2017).

[49] C. W. Helstrom, *Quantum detection and estimation theory* (Academic press, 1976).

[50] A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (North Holland, Amsterdam, 1982).

[51] S. L. Braunstein and C. M. Caves, Statistical distance and the geometry of quantum states, Phys. Rev. Lett. **72**, 3439 (1994).

[52] J. Gough and M. R. James, The series product and its application to quantum feedforward and feedback networks, IEEE Trans. Automat. Control **54**, 2530 (2009).

[53] S. Zhou, C.-L. Zou, and L. Jiang, Saturating the quantum cramér–rao bound using locc, Quantum Sci. Technol. **5**, 025005 (2020).

[54] R. D. Gill and S. Massar, State estimation for large ensembles, Phys. Rev. A **61**, 042312 (2000).

[55] K. Stannigel, P. Rabl, and P. Zoller, Driven-dissipative preparation of entangled states in cascaded quantum-optical networks, New J. Phys. **12**, 063014 (2012).

[56] T. J. Osborne, J. Eisert, and F. Verstraete, Holographic quantum states, Phys. Rev. Lett. **105**, 260401 (2010).

[57] I. Lesanovsky, M. van Horssen, M. Guţă, and J. P. Garrahan, Characterization of dynamical phase transitions in quantum jump trajectories beyond the properties of the stationary state, Phys. Rev. Lett. **110**, 150401 (2013).

[58] R. Demkowicz-Dobrzański, W. Górecki, and M. Guţă, Multi-parameter estimation beyond quantum fisher information, Journal of Physics A: Mathematical and Theoretical **53**, 363001 (2020).

[59] J. Kiukas and M. Guţă, in preparation (2022).

[60] E. L. Lehmann and G. Casella, *Theory of point estimation* (Springer, 1998).

[61] P. J. Bickel, Y. Ritov, and T. Rydén, Asymptotic normality of the maximum likelihood estimator for general hidden markov models, Ann. Statist. **26**, 1614–1635 (1998).

[62] B. Efron and D. V. Hinkley, Assessing the accuracy of the maximum likelihood estimator: Observed versus expected fisher information, Biometrika **65**, 457–487 (1978).

9 Appendix: Derivation of the measurement filter

We start by applying the algorithm [53] 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

$$M^{(n)} := |\Psi(n)\rangle \langle \dot{\Psi}(n)| - |\dot{\Psi}(n)\rangle \langle \Psi(n)| , \qquad (18)$$

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

$$\left\langle e_i^{[1]} \right| M_1^{(n)} \left| e_i^{[1]} \right\rangle = 0 \qquad \text{and} \qquad \left| \left\langle e_i^{[1]} \middle| \chi \right\rangle \right|^2 = \frac{1}{k}, \qquad \text{for all } i.$$

where $M_1^{(n)} = \text{Tr}_{0,2,\ldots,n} M^{(n)}$. 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

$$\left\langle e_i^{[2]} \right| M_2^{(n)}(i_1) \left| e_i^{[2]} \right\rangle = 0 \qquad \text{and} \qquad \left| \left\langle e_i^{[2]} \middle| \chi \right\rangle \right|^2 = \frac{1}{k}, \qquad \text{for all } i$$

where

$$M_2^{(n)}(i_1) = \text{Tr}_{0,3,\ldots,n} \left\langle e_{i_1}^{[1]} \right| M^{(n)} \left| e_{i_1}^{[1]} \right\rangle .$$

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

$$M^{(n)}(i_{\underline{j}}) := \left\langle e_{i_1}^{[1]} \otimes e_{i_2}^{[2]} \otimes \cdots \otimes e_{i_j}^{[j]} \right| M^{(n)} \left| e_{i_1}^{[1]} \otimes e_{i_2}^{[2]} \otimes \cdots \otimes e_{i_j}^{[j]} \right\rangle , \qquad M_{j+1}^{(n)}(i_{\underline{j}}) := \text{Tr}_{0,j+2,\ldots,n} M^{(n)}(i_{\underline{j}}).$$

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

$$\left\langle e_i^{[j+1]}(i_{\underline{j}}) \right| M_{j+1}^{(n)}(i_{\underline{j}}) \left| e_i^{[j+1]}(i_{\underline{j}}) \right\rangle = 0 \qquad \text{and} \qquad \left| \left\langle e_i^{[j+1]}(i_{\underline{j}}) \middle| \chi \right\rangle \right|^2 = \frac{1}{k}, \qquad \text{for all } i.$$

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 9.1. *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

$$P(n) := |\psi \otimes \chi^{\otimes n}\rangle \langle \psi \otimes \chi^{\otimes n}| = |\Psi_{\theta_0}(n)\rangle \langle \Psi_{\theta_0}(n)|.$$

Then

$$|\dot{\Psi}(n)\rangle = \sum_{i=1}^n W^{(n)} \ldots \dot{W}^{(i)} \ldots W^{(1)} |\psi \otimes \chi^{\otimes n}\rangle = \sum_{i=1}^n W^{(n)} \ldots \dot{W}^{(i)} |\psi \otimes \chi^{\otimes n}\rangle$$

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.

$$
\begin{aligned}
M_1^{(n)} = \text{Tr}_{0,2,\ldots,n}\left[M^{(n)}\right] &= \sum_i \text{Tr}_{0,2,\ldots,n}\left[W^{(n)} \cdots W^{(i)} \cdots W^{(1)} P(n) W^{(1)*} \cdots \dot{W}^{(i)*} \cdots W^{(n)*} - c.c\right] \\
&= \sum_i \text{Tr}_{0,2,\ldots,n}\left[W^{(n)} \cdots W^{(i)} P(n) \dot{W}^{(i)*} \cdots W^{(n)*} - c.c\right] \\
&= \sum_i \text{Tr}_{0,2,\ldots,n}\left[W^{(i)} P(n) \dot{W}^{(i)*} - c.c\right] \\
&= \text{Tr}_0\left[W^{(1)} P(1) \dot{W}^{(1)*} - c.c\right] + \sum_{i=2}^n \text{Tr}_{0,2,\ldots,n}\left[W^{(i)} P(n) \dot{W}^{(i)*} - c.c\right] \\
&= \text{Tr}_0\left[M^{(1)}\right] = M_1^{(1)}, \quad\quad\quad\quad (19)
\end{aligned}
$$

where $c.c$, denotes the adjoint. In the third equality we used

$$\text{Tr}_{0,2}[W_{0,2} A_{0,1,2} W_{0,2}^*] = \text{Tr}_{0,2}[A_{0,1,2}]$$

where $W_{0,2}$ is a unitary acting on subsystems $0, 2$ of a tripartite system, and $A_{0,1}$ acts on subsystems $0, 1$. 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.

$$
\begin{aligned}
M_2^{(n)}(i_1) = \text{Tr}_{0,3,\ldots,n}&\left[M^{(n)}(i_1)\right] \\
= &\sum_i \text{Tr}_{0,3,\ldots,n}\left[\left\langle e_{i_1}^{[1]} \right| W^{(n)} \cdots W^{(i)} \cdots W^{(1)} P(n) W^{(1)*} \cdots \dot{W}^{(i)*} \cdots W^{(n)*} \left| e_{i_1}^{[1]} \right\rangle - c.c\right] \\
= &\text{Tr}_{0,3,\ldots,n}\left[\left\langle e_{i_1}^{[1]} \right| W^{(n)} \cdots W^{(1)} P(n) \dot{W}^{(1)*} \cdots W^{(n)*} \left| e_{i_1}^{[1]} \right\rangle - c.c\right] \\
&+ \text{Tr}_{0,3,\ldots,n}\left[\left\langle e_{i_1}^{[1]} \right| W^{(n)} \cdots W^{(2)} W^{(1)} P(n) W^{(1)*} \dot{W}^{(2)*} \cdots W^{(n)*} \left| e_{i_1}^{[1]} \right\rangle - c.c\right] \\
&+ \sum_{i=3}^n \text{Tr}_{0,3,\ldots,n}\left[\left\langle e_{i_1}^{[1]} \right| W^{(n)} \cdots W^{(i)} \cdots W^{(1)} P(n) W^{(1)*} \cdots \dot{W}^{(i)*} \cdots W^{(n)*} \left| e_{i_1}^{[1]} \right\rangle - c.c\right] \\
= &\text{Tr}_0\left[W^{(2)} \left\langle e_{i_1}^{[1]} \right| W^{(1)} P(2) \dot{W}^{(1)*} \left| e_{i_1}^{[1]} \right\rangle W^{(2)*} - c.c\right] \\
&+ \frac{1}{k} \text{Tr}_0\left[W^{(1)} P(1) \dot{W}^{(1)*} - c.c\right] \\
= &\text{Tr}_0\left[W^{(2)} \left(\left\langle e_{i_1}^{[1]} \right| M^{(1)} \left| e_{i_1}^{[1]} \right\rangle \otimes P_\chi\right) W^{(2)*}\right] + \frac{1}{k} \text{Tr}_0\left[M^{(1)}\right] = M_2^{(2)}(i_1).
\end{aligned}
$$

The equalities follow in the same way as in (19), 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

$$
\begin{aligned}
M_j^{(n)}(i_{\underline{j-1}}) = \mathrm{Tr}_{0,j+1,\ldots,n}\left[M^{(n)}(i_{\underline{j-1}})\right] &= \frac{1}{k^{j-1}}\mathrm{Tr}_0\left[M^{(1)}\right] \\
&+ \mathrm{Tr}_0\left[W^{(j)}\left(\langle e_{i_{\underline{j-1}}}^{[j-1]}|M^{(j-1)}(i_{\underline{j-2}})|e_{i_{\underline{j-1}}}^{[j-1]}\rangle \otimes P_\chi\right)W^{(j)*}\right] \\
&= \mathrm{Tr}_0\left[M^{(j)}(i_{\underline{j-1}})\right] = M_j^{(j)}(i_{\underline{j-1}}).
\end{aligned}
\tag{20}
$$

$\square$

Next, we show that we can express $M^{(j)}(i_{\underline{j-1}})$ in terms of $M^{(1)}$, $M^{(j-1)}(i_{\underline{j-2}})$ and $|e_{j-1}^{[j-1]}\rangle$. Indeed by writing $W(j) = W^{(j)}W(j-1)$ we have

$$
\begin{aligned}
M^{(j)}(i_{\underline{j-1}}) &:= \left\langle e_{i_{j-1}}^{[j-1]}\middle| M^{(j)} \middle| e_{i_{j-1}}^{[j-1]}\right\rangle \\
&= \left\langle e_{i_{j-1}}^{[j-1]}\middle| W^{(j)}W(j-1)P(j)W^*(j-1)\dot{W}^{(j)*}\middle| e_{i_{j-1}}^{[j-1]}\right\rangle - c.c \\
&\quad + \left\langle e_{i_{j-1}}^{[j-1]}\middle| W^{(j)}W(j-1)P(j)\dot{W}(j-1)^*W^{(j)*}\middle| e_{i_{j-1}}^{[j-1]}\right\rangle - c.c \\
&= \frac{1}{k^{j-1}}M^{(1)} + W^{(j)}\left(\left\langle e_{i_{j-1}}^{[j-1]}\middle| M^{(j-1)}(i_{\underline{j-2}})\middle| e_{i_{j-1}}^{[j-1]}\right\rangle \otimes P_\chi\right)W^{(j)*}
\end{aligned}
\tag{21}
$$

This can then be used to determine the next measurement basis, producing the iterative procedure described in section 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 $A_j := M^{(j)}(i_{\underline{j-1}})$ for $j > 1$, and denote

$$
\Pi_j := \left\langle e_{i_j}^{[j]}\middle| A_j \middle| e_{i_j}^{[j]}\right\rangle = \left\langle e_{i_j}^{[j]}\middle| M^{(j)}(i_{\underline{j-1}})\middle| e_{i_j}^{[j]}\right\rangle
$$

Then equation (21) 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 $B_j := \mathrm{Tr}_0 A_j$.

10 Appendix: Proof of Proposition 6.1

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 [53]. 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

$$
I_\theta(X) + \mathbb{E}_X F(\psi_\theta(X)) \le F_\theta
$$

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

$$
F_\theta = I_\theta(X) + \mathbb{E}_X I_\theta(Y|X) \le I_\theta(X) + \mathbb{E}_X F(\psi_\theta(X)) \le F_\theta
$$

where the first equality is due to measurement optimality, while the second is the inequality between classical and quantum information. This implies that

$$F_\theta - I_\theta(X) = \mathbb{E}_X F_\theta(\psi_\theta(X)).\tag{22}$$

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 the output and the system+absorber are measured according to the optimal scheme presented in section 5. The joint state is given by

$$|\Psi_\theta(n)\rangle = W_\theta^{(n)} \cdot \dots \cdot W_\theta^{(1)} |\psi \otimes \chi^{\otimes n}\rangle$$

and the conditional states are

$$|\psi_\theta(i_1, \dots, i_n)\rangle = \frac{K_{\theta,i_n}^{[n]} \dots K_{\theta,i_1}^{[1]} |\psi\rangle}{\sqrt{p_\theta(i_1, \dots i_n)}}.$$

We will show that at $\theta = \theta_0$ the left side of (22) is bounded by a constant which does not depend on n. For simplicity, whenever possible we will use the compact notations such as

$$K_\mathbf{i} := K_{\theta,i_n}^{[n]} \dots K_{\theta,i_1}^{[1]}, \qquad \text{and} \qquad |e_\mathbf{i}\rangle = |e_{i_n}^{[n]} \otimes \dots \otimes e_{i_1}^{[1]}\rangle.$$

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 (22) is given by

$$F_s(\theta_0) = 4 \sum_\mathbf{i} p(\mathbf{i}) \|\psi_{\theta_0}^\perp(\mathbf{i})\|^2 \tag{23}$$

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 (23) 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

$$|\dot\Psi(n)\rangle = \sum_{j=1}^n W^{(n)} \dots W^{(j+1)} \dot W^{(j)} W^{(j-1)} \dots W^{(1)} |\psi \otimes \chi^{\otimes n}\rangle = \sum_{j=1}^n W^{(n)} \dots W^{(j+1)} \dot W^{(j)} |\psi \otimes \chi^{\otimes n}\rangle$$

and by triangle inequality

$$F_s(\theta_0) \le 4 \left(\sum_{j=1}^n \left\| (P_\psi^\perp \otimes I) W^{(n)} \dots W^{(j+1)} \dot W^{(j)} \psi \otimes \chi^{\otimes n} \right\| \right)^2$$

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

$$
\begin{aligned}
\left\|(P_\psi^\perp \otimes I)W^{(n)}\ldots W^{(j+1)}\dot{W}^{(j)}\psi \otimes \chi^{\otimes n}\right\|^2 &= \mathrm{Tr}_{0,1,\ldots n-j}\left((P_\psi^\perp \otimes I)W^{(n-j)}\ldots W^{(1)}\tau W^{(1)*}\ldots W^{(n-j)*}\right) \\
&= \mathrm{Tr}_0(P_\psi^\perp T^{n-j}(\tau))
\end{aligned}
$$

where in the last equality we have uses 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

$$
\mathrm{Tr}_0(P_\psi^\perp T^{n-j}(\tau)) \le a^{2(n-j)}
$$

for some $a < 1$. Therefore

$$
F_s(\theta_0) \le 4\left(\sum_{j=1}^n a^{n-j}\right)^2 \le 4\frac{1}{(1-a)^2}
$$

Note that if the spectral gap of T becomes small then the convergence to stationarity is slower and the upper bound increases.

$\square$

11 Appendix: Proof of Lemma 6.2

The CFI of any output measurement can be computed explicitly by writing

$$
\begin{aligned}
\frac{d}{d\theta}p_\theta(i_1,\ldots i_n)\bigg|_{\theta_0} &= \frac{d}{d\theta}\left\|K_{i_n}^{[n]}\cdots K_{i_1}^{[1]}\psi\right\|^2\bigg|_{\theta_0} \\
&= 2\mathrm{Re}\sum_{j=1}^n \langle\psi|K_{i_1}^{[1]*}\cdots K_{i_n}^{[n]*}K_{i_n}^{[n]}\cdots \dot{K}_{i_j}^{[j]}\cdots K_{i_1}^{[1]}|\psi\rangle \\
&= \left|c_{i_1}^{[1]}\cdots c_{i_n}^{[n]}\right|^2 \cdot 2\mathrm{Re}\sum_{j=1}^n \frac{\langle\psi|K_{i_n}^{[n]}\ldots K_{i_{j+1}}^{[j+1]}\dot{K}_{i_j}^{[j]}|\psi\rangle}{c_{i_j}^{[j]}\ldots c_{i_n}^{[n]}}
\end{aligned}
$$

Therefore

$$
\begin{aligned}
I_{\theta_0}^{(\mathrm{out})}(n) &= \sum_{i_1,\ldots,i_n}\left|c_{i_1}^{[1]}\cdots c_{i_n}^{[n]}\right|^2\left(2\mathrm{Re}\sum_{j=1}^n \frac{\langle\psi|K_{i_n}^{[n]}\ldots K_{i_{j+1}}^{[j+1]}\dot{K}_{i_j}^{[j]}|\psi\rangle}{c_{i_j}^{[j]}\ldots c_{i_n}^{[n]}}\right)^2 \\
&= \sum_{i_1,\ldots,i_n}p_{\theta_0}(i_1,\ldots,i_n)f^2(i_1,\ldots,i_n) = \mathbb{E}_{\theta_0}(f^2)
\end{aligned}
\tag{24}
$$

where f is the function

$$
f(i_1,\ldots,i_n) = 2\mathrm{Re}\sum_{j=1}^n \frac{\langle\psi|K_{i_n}^{[n]}\ldots K_{i_{j+1}}^{[j+1]}\dot{K}_{i_j}^{[j]}|\psi\rangle}{c_{i_j}^{[j]}\ldots c_{i_n}^{[n]}}
$$

$\square$

12 Appendix: Computation of finite time system-output QFI

From (5) and (1) we have

$$
\begin{aligned}
F_\theta^{(\mathrm{s+o})}(n) &= 4\|\dot{\Psi}_\theta(n)\|^2 \\
&= 4\sum_{i_1,\ldots i_n}\left\|\sum_{j=1}^n K_{i_n}\ldots \dot{K}_{i_j}\ldots K_{i_1}\psi\right\|^2
\end{aligned}
\tag{25}
$$

where K_i are the (fixed) Kraus operators with respect to the standard basis, and $|psi\rangle = |0\rangle$. Our specific model ha the feature that both K_i and $\dot{K}_i$ map the basis vectors into each other:

$$K_0|0\rangle = |0\rangle \qquad \dot{K}_0|0\rangle = i|1\rangle \qquad K_0|1\rangle = \sqrt{1-\lambda}|1\rangle \qquad \dot{K}_0|1\rangle = i\sqrt{1-\lambda}|0\rangle$$

$$K_1|0\rangle = 0 \qquad \dot{K}_1|0\rangle = |1\rangle \qquad K_1|1\rangle = \sqrt{\lambda}e^{i\phi}|0\rangle \qquad \dot{K}_1|1\rangle = 0$$

This allows to compute the QFI explicitly by noting the terms in the sum (25) 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

$$F^{(0)} = 4\left\|\sum_{j=1}^{n} K_0 \ldots \dot{K}_0 \ldots K_0|0\rangle\right\|^2 = 4\left|\sum_{j=1}^{n} ia^{n-j}\right|^2 = 4\left(\frac{1-a^n}{1-a}\right)^2 \qquad (26)$$

where $a = \sqrt{1-\lambda}$.

Consider now a sequence $(0,\ldots,,0,1,0,\ldots 0)$ with a single one on position l. Since $K_0 \ldots \dot{K}_0 \ldots K_1 \ldots K_0|0\rangle = 0$ the only contributing terms will be those with derivative on the first $(l-1)$ K_0s or on K_1. This gives

$$
\begin{aligned}
F^{(1)} &= 4\sum_{l=2}^{n}\left\|i\sum_{r=1}^{l-1} e^{i\phi}a^{l-1-r}b|0\rangle + a^{n-l}|1\rangle\right\|^2 + 4\|a^{n-1}|1\rangle\|^2 \\
&= 4\sum_{l=2}^{n}\left(b^2\frac{1-a^{l-1}}{1-a} + a^{2(n-l)}\right) + 4a^{2(n-1)} \\
&= 4(n-1)\frac{b^2}{(1-a)^2} + 4\frac{a^2-a^{2n}}{(1-a)^2} - 8\frac{b^2(a-a^n)}{(1-a)^3} + 4\frac{1-a^{2(n-1)}}{b^2} + 4a^{2(n-1)} \qquad (27)
\end{aligned}
$$

where $b = \sqrt{\lambda}$.

Finally, consider the sequences of the type $(0,\ldots 0,1,0,\ldots 0,1,0\ldots 0)$ with 1s on positions $1 \leq i < k \leq n$. In this case the nonzero contributions come from terms where the derivative is on positions $j = i$. The Fisher contribution is

$$
\begin{aligned}
F^{(2)} &= 4\sum_{1\leq i<k\leq n} \|e^{i\phi}\sqrt{\lambda}(1-\lambda)^{(k-i-1)/2}|0\rangle\|^2 \\
&= 4\sum_{1\leq i<k\leq n} b^2 a^{2(k-i-1)} = 4(n-1) - 4\frac{a^2-a^{2n}}{b^2}. \qquad (28)
\end{aligned}
$$

Adding together the contributions (26), (27) and (28) we obtain the total QFI

$$
\begin{aligned}
F_\theta^{(\text{s+o})}(n) &= F^{(0)} + F^{(1)} + F^{(2)} \\
&= \frac{8n}{1-a} \\
&\quad + 4\left[\left(\frac{1-a^n}{1-a}\right)^2 - \frac{b^2}{(1-a)^2} + \frac{a^2-a^{2n}}{(1-a)^2} - 2\frac{b^2(a-a^n)}{(1-a)^3} + \frac{1-a^{2(n-1)}}{b^2} + a^{2(n-1)}\right] \\
&\quad - 4\left[1 + \frac{a^2-a^{2n}}{b^2}\right] \qquad (29)
\end{aligned}
$$

where the leading term is consistent with the QFI rate formula (16).