Aberystwyth University

Quantum estimation via sequential measurements

Burgarth, Daniel; Giovannetti, Vittorio; Kato, Airi N.; Yuasa, Kazuya

Quantum estimation via sequential measurements

You may also be interested in:

Building versatile bipartite probes for quantum metrology
Alessandro Farace, Antonella De Pasquale, Gerardo Adesso et al.

Quantum metrology from a quantum information science perspective
Géza Tóth and Iagoba Apellaniz

Open-system dynamics of entanglement:a key issues review
Leandro Aolita, Fernando de Melo and Luiz Davidovich

Optimal quantum states for frequency estimation
F Fröwis, M Skotiniotis, B Kraus et al.

Maximum likelihood versus likelihood-free quantum system identification in the atom maser
Catalin Catana, Theodore Kypraios and Mdlin Gu

Trapping phenomenon of the parameter estimation in asymptotic quantum states
K Berrada

Quantum parameter estimation with imperfect reference frames
Dominik Šafránek, Mehdi Ahmadi and Ivette Fuentes

New Journal of Physics
The open access journal at the forefront of physics

PAPER

CrossMark

Quantum estimation via sequential measurements

Daniel Burgarth[1], Vittorio Giovannetti[2], Airi N Kato[3] and Kazuya Yuasa[4,5]

[1] Department of Mathematics, Aberystwyth University, SY23 3BZ Aberystwyth, UK
[2] NEST, Scuola Normale Superiore and Istituto Nanoscienze-CNR, I-56126 Pisa, Italy
[3] Department of Physics, University of Tokyo, Tokyo 113-0033, Japan
[4] Department of Physics, Waseda University, Tokyo 169-8555, Japan
[5] Author to whom any correspondence should be addressed.
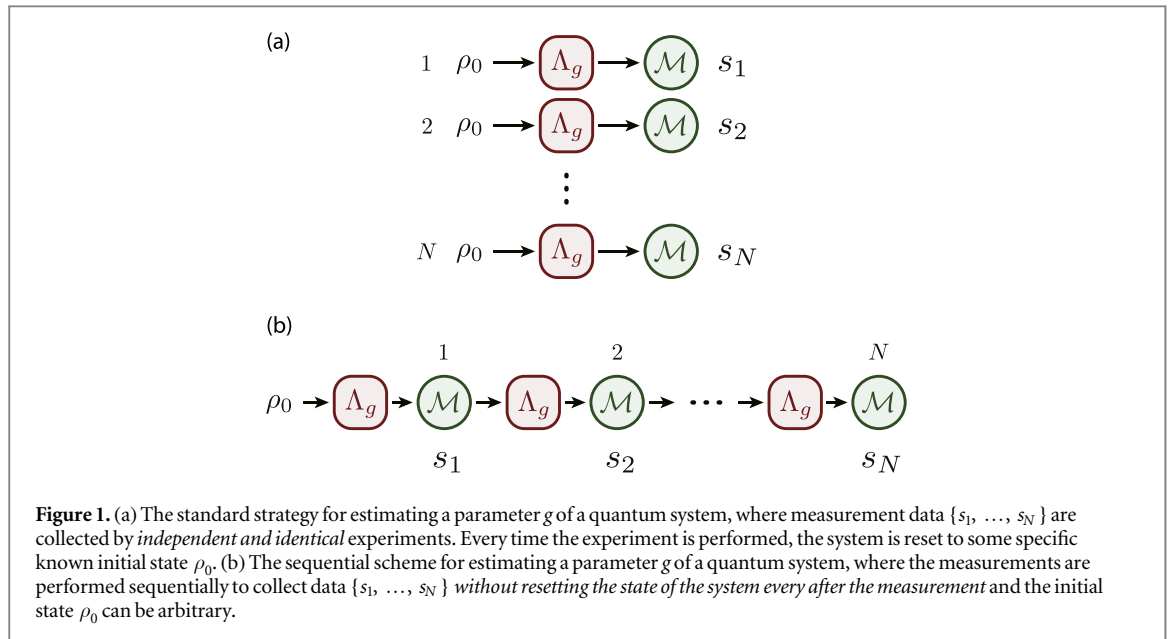
E-mail: yuasa@waseda.jp

Abstract

The problem of estimating a parameter of a quantum system through a series of measurements performed sequentially on a quantum probe is analyzed in the general setting where the underlying statistics is explicitly non-i.i.d. We present a generalization of the central limit theorem in the present context, which under fairly general assumptions shows that as the number N of measurement data increases the probability distribution of functionals of the data (e.g., the average of the data) through which the target parameter is estimated becomes asymptotically normal and independent of the initial state of the probe. At variance with the previous studies (Guţă M 2011 *Phys. Rev.* A **83** 062324; van Horssen M and Guţă M 2015 *J. Math. Phys.* **56** 022109) we take a diagrammatic approach, which allows one to compute not only the leading orders in N of the moments of the average of the data but also those of the correlations among subsequent measurement outcomes. In particular our analysis points out that the latter, which are not available in usual i.i.d. data, can be exploited in order to improve the accuracy of the parameter estimation. An explicit application of our scheme is discussed by studying how the temperature of a thermal reservoir can be estimated via sequential measurements on a quantum probe in contact with the reservoir.

1. Introduction

Seeking the most efficient way to recover the value of a parameter g encoded in the state ρ_g of a quantum system is the fundamental problem of a branch of quantum information technologies [1], which goes under the name of quantum metrology [2, 3]. It goes without mentioning that this topic has applications in a variety of different research areas, ranging e.g. from the interferometric estimation of the phase shifts induced by gravitational waves [4], high-precision quantum magnetometry [5, 6], to remote probing of targets.

In the standard approach one typically focuses on the case where several (say N) identical copies of ρ_g are available to experimentalists, who can hence rely on the statistical inference extracted from independent and identically distributed (i.i.d.) measurement outcomes to estimate the value of g. This scenario is particularly well formulated by those configurations where the unknown parameter g is associated with some black-box transformation Λ_g (say a phase shift induced in one arm of an interferometric setup) which acts on the input state ρ_0 of a probing system (say the light beam injected into the interferometer) yielding $\rho_g = \Lambda_g(\rho_0)$ as the output density matrix to be measured, with such test repeated N times to collect data $\{s_1, \ldots, s_N\}$. See figure 1(a). In this context, the ultimate limits on the attainable precision in the estimation of g, optimized with respect to the general detection strategy, can be computed, resulting in the so-called quantum Cramér–Rao bound, which exhibits the functional dependence upon ρ_g via the quantum Fisher information. See e.g. [2, 3, 7–12].

In many situations of physical interest, however, the possibility of reinitializing the setup to the same state is not necessarily guaranteed. In the present study we are going to consider a different scheme, in which a single probing system undergoes multiple applications of Λ_g while being monitored during the process without being

Figure 1. (a) The standard strategy for estimating a parameter g of a quantum system, where measurement data $\{s_1, \ldots, s_N\}$ are collected by *independent and identical* experiments. Every time the experiment is performed, the system is reset to some specific known initial state ρ_0. (b) The sequential scheme for estimating a parameter g of a quantum system, where the measurements are performed sequentially to collect data $\{s_1, \ldots, s_N\}$ *without resetting the state of the system every after the measurement* and the initial state ρ_0 can be arbitrary.

reinitialized to the same input state. See figure 1(b). The data $\{s_1, \ldots, s_N\}$ collected by such sequential measurements will be non-i.i.d. in general. Still, we are able to estimate the target parameter g from the data under certain conditions. We will see that the property of the channel describing the process is important. The idea is to let the probing system forget about the past by the mixing of the channel [13–16] (the channel being intrinsically mixing or designed to be mixing), which clusters the data and allows the central limit theorem to hold for appropriately chosen functionals of the data. This sequential scheme is suited to account for estimation procedures where one aims to recover g via a sequence of weak measurements that slightly perturb the probe. In particular it can be adapted to study physical setups where the probing system is a proper subset of a many-body quantum system which is directly affected by the black-box generator Λ_g (an explicit example of this scenario will be analyzed in the final part of this paper).

Various schemes for quantum parameter estimation based on repetitive or continuous measurements have been studied: see e.g. [17–25]. Among them, analogous setups were analyzed in [19, 24], where the problem was formalized in terms of quantum Markov chains. Specifically in [19] it has been shown that, under rather general assumptions, the statistics of the associated estimation problem converges asymptotically to a normal one, generalizing the similar results which were known to apply to purely classical settings [26–28]. In the present paper we first provide an independent derivation of the previous result [19] via a diagrammatic approach to compute the leading-order contributions to the moments of the associated estimating functional of the data $\{s_1, \ldots, s_N\}$, i.e., the moments of the average $S = \frac{1}{N} \sum_{i=1}^{N} s_i$. This approach allows us to prove the central limit theorem including other estimating functionals capturing the correlations among different measurement outcomes, e.g., $C_\ell = \frac{1}{N - \ell} \sum_{i=1}^{N-\ell} s_i s_{i+\ell}$. The asymptotic normality of the empirical measure associated to chains of subsequent measurement outcomes is proved in [24], but in contrast to this previous work we provide explicit formulas which allow us to evaluate the elements of the covariance matrix of the normal distribution of the variables S and C_ℓ. Moreover, we point out that the inclusion of the correlations C_ℓ for estimation, which do not contain any useful information in the usual i.i.d. data, can help improve the accuracy of the estimation. This result, while not conclusive, is a preliminary (yet nontrivial) step towards the determination of the ultimate accuracy limit attainable in the non-i.i.d. settings.

This paper is organized as follows. In section 2 we introduce the notation and recall some basic mathematical facts which will be used in the paper. The non-i.i.d. estimation model is then presented in section 3. In section 4 we focus on the simplest estimating functional S of the measurement data, and prove its asymptotic normality under the assumption of mixing of the process. The central limit theorem is generalized to include the correlations C_ℓ and their role in the estimation problem is addressed in section 5. An explicit example is then presented in section 6, where we discuss the estimation of the temperature of a thermal reservoir via local measurements on a quantum probe in contact with the reservoir. Conclusions and perspectives are summarized in section 7, while some technical elements are presented in the appendices.

2. Notation and mathematical background

In this section we introduce the notation and recall some basics facts on the theory of quantum channels.

2.1. Quantum ergodic/mixing channels

Quantum channels are completely positive and trace-preserving (CPTP) maps, transforming density operators to density operators [29–31]. Every CPTP map $\mathcal{E}$ admits at least one *fixed point* [13, 14], namely, a stationary state ρ_*,

$$\mathcal{E}(\rho_*) = \rho_*, \tag{1}$$

which is hermitian, positive-semidefinite, and of unit trace. In other words, the fixed point ρ_* is an eigenstate of the map $\mathcal{E}$ belonging to its unit eigenvalue 1.

   If the fixed point ρ_* is unique, the quantum channel $\mathcal{E}$ is called *ergodic* [14–16]. It implies

$$\frac{1}{N} \sum_{n=0}^{N-1} \mathcal{E}^n(\rho_0) \to \rho_* \quad (N \to \infty), \quad \forall \text{ states } \rho_0, \tag{2}$$

where $\mathcal{E}^n = \mathcal{E} \circ \cdots \circ \mathcal{E}$ denotes n recursive applications of the channel $\mathcal{E}$, and the convergence is in the superoperator norm with corrections whose leading order scales as $1/N$. Moreover, if the fixed point ρ_* is unique and the unit eigenvalue 1 is the only *peripheral eigenvalue* (eigenvalue of unit magnitude), the quantum channel $\mathcal{E}$ is converging as

$$\mathcal{E}^N(\rho_0) \to \rho_* \quad (N \to \infty), \quad \forall \text{ states } \rho_0, \tag{3}$$

and is called *mixing*, with the convergence being as in (2) [13–16]. Mixing implies ergodicity, but the converse is not necessarily true.

   As commented above, the fixed point ρ_* of a quantum channel $\mathcal{E}$ is an eigenstate of $\mathcal{E}$ belonging to its unit eigenvalue 1. In a matrix representation of $\mathcal{E}$, it is a 'right eigenvector.' The corresponding 'left eigenvector' belonging to the same eigenvalue can be different from the right eigenvector in general. For a quantum channel $\mathcal{E}$, the trace Tr is a left eigenvector belonging to the unit eigenvalue 1, since the quantum channel $\mathcal{E}$ is trace-preserving

$$\text{Tr} \{ \mathcal{E}(\rho) \} = \text{Tr} \, \rho. \tag{4}$$

Let us hence write the fixed point ρ_* and the trace Tr in the vectorized notation as

$$\rho_* \leftrightarrow |\rho_*), \qquad \text{Tr} \leftrightarrow (1|, \tag{5}$$

respectively, and a couple of eigenvalue equations for the unit eigenvalue 1 read

$$\mathcal{E} |\rho_*) = |\rho_*), \qquad (1| \, \mathcal{E} = (1|. \tag{6}$$

   More explicitly, given any complete set of orthonormal basis states $\{ |n\rangle \}_n$ of the system, an operator $A = \sum_{n,n'} A_{nn'} |n\rangle \langle n'|$ is vectorized by $|A) = \sum_{n,n'} A_{nn'} |n\rangle \otimes |n'\rangle$ [32]. The trace $(1| = \sum_n \langle n| \otimes \langle n|$ is (the hermitian conjugate of) the vectorized version of the identity operator: that is why it is denoted by $(1|$. In addition, the inner product $(A|B) = \text{Tr} \{ A^\dagger B \}$ is the Hilbert–Schmidt inner product. In this representation, the quantum channel $\mathcal{E} \leftrightarrow \sum_{m,n,m',n'} \mathcal{E}_{mn,m'n'} |m\rangle \langle m'| \otimes |n\rangle \langle n'|$ is a matrix with the matrix elements $\mathcal{E}_{mn,m'n'} = \langle m| \, \mathcal{E}(|m'\rangle \langle n'|) |n\rangle$ in the original representation, and the application of a map $\mathcal{E}$ is expressed by the multiplication of the corresponding matrix. By abuse of notation we use the same symbol $\mathcal{E}$ for its matrix representation.

   In this matrix representation, the eigenvalue equation for $\mathcal{E}$ reads

$$\mathcal{E} |u_n) = \lambda_n |u_n), \qquad (v_n| \, \mathcal{E} = \lambda_n (v_n|. \tag{7}$$

In particular, $|u_0) = |\rho_*)$ and $(v_0| = (1|$ with $\lambda_0 = 1$. The eigenvectors belonging to different eigenvalues are orthogonal to each other and normalized as

$$(v_n|u_n) = 1, \qquad (v_m|u_n) = 0 \quad \text{for} \quad \lambda_m \neq \lambda_n. \tag{8}$$

Note that the matrix $\mathcal{E}$ might not be diagonalizable but is cast in the Jordan canonical form in general [13, 14].

   In this paper, ergodic or mixing channels will play a central role. The unit eigenvalue 1 of such a channel $\mathcal{E}$ is not degenerated by definition, and the ergodic/mixing channel $\mathcal{E}$ can always be decomposed as

$$\mathcal{E} = \mathcal{P}_* + \mathcal{E}', \tag{9}$$

where

$$\mathcal{P}_* = \big|\rho_*\big)(1| = \rho_* \, \mathrm{Tr}\,\{\cdot\} \tag{10}$$

is the eigenprojection belonging to the nondegenerate unit eigenvalue 1 of $\mathcal{E}$, and the remaining part $\mathcal{E}'$ (which is not CPTP) is built on the eigenvectors $\{|u_n)\}_{n\neq 0}$ and $\{(v_n|\}_{n\neq 0}$ belonging to the eigenvalues λ_n different from 1. By construction $\mathcal{E}'$ is orthogonal to $\mathcal{P}_*$, i.e., $\mathcal{P}_*\mathcal{E}' = \mathcal{E}'\mathcal{P}_* = 0$. Moreover, since $\mathcal{E}'$ does not admit a unit eigenvalue 1, the inverse $(1 - \mathcal{E}')^{-1}$ exists, and we have

$$\frac{1}{N}\sum_{n=0}^{N-1}\mathcal{E}^n = \mathcal{P}_* + \frac{1}{N}\frac{1-\mathcal{E}'^N}{1-\mathcal{E}'}\mathcal{Q}_*, \qquad \mathcal{Q}_* = 1 - \mathcal{P}_*, \tag{11}$$

which allows us to prove the convergence to the fixed point ρ_* in (2). If the channel $\mathcal{E}$ is not only ergodic but also mixing, the spectral radius of $\mathcal{E}'$ is strictly smaller than 1, and we get

$$\mathcal{E}^N = \mathcal{P}_* + \mathcal{E}'^N \to \mathcal{P}_* \quad (N \to \infty), \tag{12}$$

which proves (3). If the channel $\mathcal{E}$ is ergodic but not mixing, $\mathcal{E}'$ admits a peripheral eigenvalue, and $\mathcal{E}'^N$ does not decay: we lose the convergence (3), but the averaged channel converges as (2).

Note again that $\mathcal{E}'$ might not be diagonalizable if some of the eigenvalues λ_n of $\mathcal{E}$ are degenerated, but it is not a problem for the convergence: see [13–16].

2.2. Measurement and back-action

We recall that in quantum mechanics the most general detection scheme can be formalized in terms of positive operator-valued measure (POVM). See e.g. [29]. Expressed in the superoperator language this accounts to assigning a collection $\mathfrak{M} = \{\mathcal{M}_s\}_s$ of trace-decreasing channels $\mathcal{M}_s$ describing the statistics of the measurement and the back-action on the probed system. In particular, given ρ the density matrix of the system before the measurement, the probability of getting outcome s by the measurement $\mathfrak{M}$ is given by

$$p(s|\rho) = \mathrm{Tr}\left\{\mathcal{M}_s(\rho)\right\} = \Big(1\Big|\mathcal{M}_s\Big|\rho\Big), \tag{13}$$

with $\mathcal{M}_s(\rho)$ being the conditional (not normalized) state immediately after the event. By construction the map

$$\mathcal{M} = \sum_s \mathcal{M}_s, \tag{14}$$

obtained by summing over all possible values of s, is CPTP and describes the evolution of the system when no record of the measurement outcome is kept. We also notice that given $\mathcal{D}$ and $\mathcal{E}$ two CPTP maps, the set of channels $\mathcal{M}'_s = \mathcal{E}\mathcal{M}_s\mathcal{D}$ defines a new POVM measurement $\mathfrak{M}' = \{\mathcal{M}'_s\}_s$, where immediately before and after the measurement $\mathfrak{M}$ one transforms the state of the probed system through the actions of $\mathcal{D}$ and $\mathcal{E}$, respectively. Finally we observe that given $\mathfrak{M} = \{\mathcal{M}_r\}_r$ and $\mathfrak{N} = \{\mathcal{N}_s\}_s$ two POVMs, the operator $(\mathcal{N}_s \circ \mathcal{M}_r)(\rho)$ represents the conditional (not normalized) state obtained when the measurements $\mathfrak{M}$ and $\mathfrak{N}$ are performed on a system in the state ρ yielding measurement outcomes r and s, respectively, the associated probability given by $p(r, s|\rho) = (1|\mathcal{N}_s\mathcal{M}_r|\rho)$.

3. Sequential scheme

The problem we study is the following: we wish to recover an unknown parameter g of a quantum system, which is encoded in the state of a quantum probe via the action of a quantum channel Λ_g,

$$\rho_0 \mapsto \Lambda_g(\rho_0). \tag{15}$$

Here ρ_0 is the input state of the probe, which (possibly) is initialized by us, while $\Lambda_g(\rho_0)$ is the associated output state, on which we are allowed to perform measurement in order to learn about g. In a standard i.i.d. approach [10], one is supposed to perform the same experiment several times collecting i.i.d. outcomes $\{s_1, \ldots, s_N\}$, from which the value of g is to be extrapolated via some suitable data processing. See figure 1(a). More precisely, in every experimental run of such an i.i.d. scheme the probe should be initialized in the *same* input state ρ_0 and the *same* POVM measurement $\mathfrak{M} = \{\mathcal{M}_s\}_s$ should be performed after Λ_g has operated on the probe. On the contrary, in the protocol we are going to discuss here, while we keep performing the same measurement $\mathfrak{M}$ on the probe, the probe is not reset to ρ_0 after each measurement step. Instead, we just repeat the application of Λ_g followed by a measurement many times to get a sequence of outcomes $\{s_1, \ldots, s_N\}$, whose statistics is not necessarily i.i.d. anymore. See figure 1(b). In this scenario, following the framework detailed in section 2, the state of the probe undergoes a conditional evolution described by the (not necessarily normalized) density matrix

$$\rho_0 \mapsto \left(\mathcal{E}_{s_N} \circ \cdots \circ \mathcal{E}_{s_1}\right)(\rho_0), \qquad \mathcal{E}_s = \mathcal{M}_s \circ \Lambda_g, \tag{16}$$

whose trace

$$p\left(s_1, \ldots, s_N \middle| \rho_0\right) = \left(1 \middle| \mathcal{E}_{s_N} \cdots \mathcal{E}_{s_1} \middle| \rho_0\right) \tag{17}$$

defines the probability of the associated measurement event.

It is worth observing that this mathematical setting includes the i.i.d. scenario as a special case, where Λ_g is identified with $\Lambda_g \circ \mathcal{P}_0$, with $\mathcal{P}_0 = |\rho_0)(1| = \rho_0 \operatorname{Tr} \{\bullet\}$ being the map resetting the state of the probe into ρ_0. Indeed, with this choice the probability (17) coincides with the one for the case where the measurements $\mathfrak{M} = \{\mathcal{M}_s\}_s$ are performed independently on N copies of $\Lambda_g(\rho_0)$, i.e.

$$p\left(s_1, \ldots, s_N \middle| \rho_0\right) = \left(1 \middle| \mathcal{E}_{s_N} \middle| \rho_0\right) \cdots \left(1 \middle| \mathcal{E}_{s_2} \middle| \rho_0\right)\left(1 \middle| \mathcal{E}_{s_1} \middle| \rho_0\right)$$
$$= p\left(s_N \middle| \rho_0\right) \cdots p\left(s_2 \middle| \rho_0\right) p\left(s_1 \middle| \rho_0\right). \tag{18}$$

From (14) it follows that the map $\mathcal{E}$ obtained by summing $\mathcal{E}_s$ in (16) over all s,

$$\mathcal{E} = \sum_s \mathcal{E}_s = \mathcal{M} \circ \Lambda_g, \tag{19}$$

is CPTP, ensuring the proper normalization of the probability (17). As an additional constraint we will require it to be mixing (in some cases, e.g., in section 4.1, however, we will weaken this requirement by imposing $\mathcal{E}$ to be just ergodic). This is not a strong assumption, as mixing channels actually form an open and dense set. Under this condition we will be able to prove that the parameter g can be estimated from the single sequence of data $\{s_1, \ldots, s_N\}$ collected by the sequential measurements, irrespective of the initial state ρ_0 [19]. The rough idea is that, thanks to the mixing (3), repeated applications of the channel force the quantum system to forget its initial state and at the same time decorrelate the data separated beyond the correlation length, which clusters the data and allows us to define *self-averaging* quantities as estimating functionals, whose fluctuations diminish as N increases, i.e., the central limit theorem holds.

Inferring g from $\{s_1, \cdots, s_N\}$
A standard estimating functional of the measured data $\{s_1, \ldots, s_N\}$, through which one tries to infer the value of g, is the average

$$S = \frac{1}{N} \sum_{i=1}^{N} s_i. \tag{20}$$

In [19] it was noted that under the assumption that the average channel $\mathcal{E}$ in (19) is mixing the central limit theorem holds for S, and for large N the probability distribution $P(S)$ of S asymptotically becomes a Gaussian peaked at a value $\langle S \rangle_*$ with a shrinking variance σ^2/N, which are both independent of the input state ρ_0 of the probe, i.e.

$$P(S) \simeq \frac{1}{\sqrt{2\pi\sigma^2/N}} e^{-\frac{\left(S - \langle S \rangle_*\right)^2}{2\sigma^2/N}}. \tag{21}$$

The explicit expressions for $\langle S \rangle_*$ and σ will be provided in (27) and (35), respectively, in section 4 below. This ensures that the quantity S evaluated from the single sequence of measurement outcomes is expected, with a high probability, to be very close to its expectation value $\langle S \rangle_*$ with a vanishingly small variance σ^2/N for large N. Therefore, by comparing the observed value of S with the formula for the expectation value $\langle S \rangle_*$ as a function of g, one can infer the parameter g. It is worth stressing once more that in the sequential scheme the measurement data are not independent of each other. Therefore, it is not trivial whether the central limit theorem holds, which is usually based on i.i.d. data set. The mixing, however, is strong enough to kill the correlations between two data if they are sufficiently far away from each other, and clusters the data, allowing the central limit theorem to hold.

Thanks to (21) the uncertainty in the estimation of g through the quantity S can be evaluated via the Cramér-Rao bound as [3, 7–12, 33]

$$\delta g \simeq \frac{1}{\sqrt{\mathcal{F}(g)}}, \tag{22}$$

where $\mathcal{F}(g)$ is the Fisher information of the problem given by

$$\mathcal{F}(g) = \int dS\, P(S) \left(\frac{\partial}{\partial g} \ln P(S)\right)^2 \simeq \frac{N}{\sigma^2} \left(\frac{\partial \langle S \rangle_*}{\partial g}\right)^2. \tag{23}$$

Accordingly, as long as $\langle S \rangle_*$ exhibits a nontrivial functional dependence upon g, the Fisher information $\mathcal{F}(g)$ increases linearly in N, yielding an estimation error (22) which diminishes as $\delta g \simeq 1/\sqrt{N}$ (in (23) we have omitted the contribution from $\partial \sigma / \partial g$ to the Fisher information $\mathcal{F}(g)$ since it does not grow with N).

It may happen however that the quantity $\langle S \rangle_*$ does not depend upon g. In such a case $\mathcal{F}(g)$ nullifies, signaling that it is impossible to recover g through S (a problem which cannot be fixed by properly choosing the input state ρ_0 of the probe, the asymptotic distribution (21) being independent of ρ_0). Nonetheless, even in this particular case, the sequence of data $\{s_1, \ldots, s_N\}$, which is not i.i.d. in general, can still contain some functional dependence upon g, which can be exploited for the estimation of g. In particular, the aim of the present work is to show that the correlations among the measurement data, which are absent in the usual i.i.d. data, can be used for this purpose. It turns out that, under the same mixing assumption on the channel $\mathcal{E}$ that leads to the central limit theorem for S in (21), the correlations are also self-averaging and become asymptotically normal for large N, enabling one to estimate g through them. See (67) and (68) in section 5 below. Even in the case where $\langle S \rangle_*$ depends upon g, looking also at the correlations help enhance the precision of the estimation of g, which will be demonstrated in figure 9 with the example studied in section 6.

We first present an alternative derivation of the results of [19], i.e., the asymptotic normality of S, on the basis of a diagrammatic approach in section 4. While our approach is more involved than the elegant perturbative approach taken in [19], it allows us to generalize the scheme to include the correlations among the measurement data in a straightforward manner to enhance the precision of the estimation. We shall indeed prove the asymptotic normality of variables including the correlation functionals in section 5.

4. Statistical behavior of S

This section is devoted to provide an alternative derivation of the results of [19], which ultimately leads to the asymptotic normality of S in (21). We start in section 4.1 by proving that under the hypothesis that the average channel $\mathcal{E}$ in (19) is ergodic the quantity S is self-averaging, converging to a fixed value $\langle S \rangle_*$ independent of the input state ρ_0. Then in section 4.2 we introduce the mixing property and show that under this stronger condition the distribution $P(S)$, which rules the statistics of S, becomes asymptotically normal.

4.1. Law of large numbers by ergodicity

Consider the expectation value of the quantity S with respect to the probability (17) governing the statistical distribution of the measurement outcomes, i.e.

$$
\begin{aligned}
\langle S \rangle_N &= \sum_{s_1} \cdots \sum_{s_N} S \, p\left(s_1, \ldots, s_N \middle| \rho_0\right) \\
&= \frac{1}{N} \sum_{i=1}^{N} \sum_{s_1} \cdots \sum_{s_N} s_i \left(1 \middle| \mathcal{E}_{s_N} \cdots \mathcal{E}_{s_1} \middle| \rho_0\right) \\
&= \frac{1}{N} \sum_{i=1}^{N} \left(1 \middle| \mathcal{E}^{(1)} \mathcal{E}^{i-1} \middle| \rho_0\right),
\end{aligned}
\tag{24}
$$

where $\mathcal{E}^{(1)}$ is defined by

$$
\mathcal{E}^{(1)} = \sum_s s \, \mathcal{E}_s.
\tag{25}
$$

Assume here that the channel $\mathcal{E}$ is ergodic with unique fixed point ρ_*: using (11), the right-hand side of (24) can be written as

$$
\langle S \rangle_N = \langle S \rangle_* + \frac{1}{N}\left(1 \middle| \mathcal{E}^{(1)} \frac{1 - \mathcal{E}'^N}{1 - \mathcal{E}'} \mathcal{Q}_* \middle| \rho_0\right).
\tag{26}
$$

The first contribution

$$
\langle S \rangle_* = \left(1 \middle| \mathcal{E}^{(1)} \middle| \rho_*\right) = \sum_s s \, \mathrm{Tr}\left\{\mathcal{E}_s(\rho_*)\right\} = \langle s \rangle_*
\tag{27}
$$

is the value of $\langle S \rangle_N$ when the input state ρ_0 of the probe coincides with the fixed point ρ_* of $\mathcal{E}$. As stressed by the last identity, it also coincides with the expectation value associated with the i.i.d. measurement on ρ_* with the POVM $\mathfrak{M}$. The second contribution in (26) instead is a correction which scales at most as $1/N$ for any other choice of ρ_0. Accordingly in the large N limit we get

$$
\langle S \rangle_N \to \langle S \rangle_* \quad (N \to \infty),
\tag{28}
$$

irrespectively of ρ_0.

In a similar way we can compute the variance of S, obtaining

$$(\Delta S)_N^2 = \langle S^2 \rangle_N - \langle S \rangle_N^2$$

$$= \frac{1}{N}(\Delta s)_*^2 + \frac{2}{N}\left(1 \left| \tilde{\mathcal{E}}^{(1)} \frac{\mathcal{Q}_*}{1 - \mathcal{E}'} \tilde{\mathcal{E}}^{(1)} \right| \rho_* \right)$$

$$- \frac{2}{N^2} \sum_{j=1}^{N-1} \left(1 \left| \tilde{\mathcal{E}}^{(1)} \frac{\mathcal{E}'^{N-j}}{1 - \mathcal{E}'} \mathcal{Q}_* \tilde{\mathcal{E}}^{(1)} \mathcal{E}'^{j-1} \mathcal{Q}_* \right| \rho_0 \right) + O\left(1/N^2\right), \tag{29}$$

where $\langle S^2 \rangle_N$ is defined similarly to $\langle S \rangle_N$ in (24), and

$$\tilde{\mathcal{E}}^{(m)} = \sum_s (\delta s)^m \mathcal{E}_s, \qquad \delta s = s - \langle s \rangle_*, \tag{30}$$

while

$$(\Delta s)_*^2 = \langle s^2 \rangle_* - \langle s \rangle_*^2 = \left(1 \left| \tilde{\mathcal{E}}^{(2)} \right| \rho_* \right) \tag{31}$$

is the variance of s in the stationary state ρ_* as in (27). Equation (29) shows that the variance shrinks as $(\Delta S)_N^2 \sim 1/N$, and the fluctuation of S around $\langle S \rangle_N$ becomes smaller and smaller as we proceed with the measurements. As a result, the probability of finding a single-shot value S close to its expectation value $\langle S \rangle_N$ becomes very high. Indeed, Chebyshev's inequality bounds the probability of S deviating from $\langle S \rangle_N$ as

$$\text{Prob}\left(\left| S - \langle S \rangle_N \right| > K (\Delta S)_N \right) < \frac{1}{K^2} \tag{32}$$

for any positive K. In this way, S is self-averaging: each single S is very close to its expectation value with very high probability. In addition, as shown in (28), $\langle S \rangle_N$ becomes independent of the initial state ρ_0.

4.2. Beyond the law of large numbers: central limit theorem by mixing

We have so far assumed the ergodicity of the channel $\mathcal{E}$: this is necessary and sufficient for the convergence $\langle S \rangle_N \to \langle S \rangle_*$ in (28), and for the shrinking variance $(\Delta S)_N^2 \sim 1/N$ in (29). If we further assume that $\mathcal{E}$ is mixing, we can say more. For instance, the third contribution to the variance $(\Delta S)_N^2$ in (29) decays as $\mathcal{E}'^N/N$ (note that the sum over j accumulates to $O(N)$), i.e., faster than $1/N$ (it is not guaranteed under the ergodicity, since $\mathcal{E}'^N$ does not decay), and the variance $(\Delta S)_N^2$ asymptotically becomes independent of the initial state ρ_0. This is because the mixing makes the system forget the initial state ρ_0 as (3) without averaging along the time trace.

Most importantly, if $\mathcal{E}$ is mixing, one can prove that the probability distribution of S asymptotically becomes normal, converging to the Gaussian distribution (21). The asymptotic normality of S under the mixing condition was proved in [19]. Here we derive the same result by introducing a diagrammatic approach. Specifically, in the following subsection we shall compute the moments of the variable $S - \langle S \rangle_*$, showing that for large N they admit the scaling

$$\left\langle \left(S - \langle S \rangle_*\right)^n \right\rangle_N \sim O\left(1/N^{\lceil \frac{n}{2} \rceil}\right), \tag{33}$$

where $\lceil x \rceil$ denotes the smallest integer not less than x. In particular for even n we shall see that the leading-order term is given by

$$\left\langle \left(S - \langle S \rangle_*\right)^n \right\rangle_N = \frac{n!}{(2N)^{n/2}(n/2)!} \sigma^n + O\left(1/N^{n/2+1}\right) \tag{34}$$

with

$$\sigma^2 = \left(1 \left| \tilde{\mathcal{E}}^{(2)} \right| \rho_* \right) + 2\left(1 \left| \tilde{\mathcal{E}}^{(1)} \frac{\mathcal{Q}_*}{1 - \mathcal{E}'} \tilde{\mathcal{E}}^{(1)} \right| \rho_* \right), \tag{35}$$

where $\tilde{\mathcal{E}}^{(1)}$ and $\tilde{\mathcal{E}}^{(2)}$ are defined as in (30). These results allow us to conclude that the characteristic function for the scaled variable $x = \sqrt{N}(S - \langle S \rangle_*)$ becomes asymptotically normal in the limit $N \to \infty$. Indeed by direct substitution we have

$$
\begin{aligned}
\chi(k) &= \left\langle e^{ik\sqrt{N}\left(S-\langle S\rangle_*\right)}\right\rangle_N \\
&= \sum_{n=0}^{\infty}\frac{1}{n!}\left(ik\sqrt{N}\right)^n\left\langle\left(S-\langle S\rangle_*\right)^n\right\rangle_N \\
&\to \sum_{r=0}^{\infty}\frac{1}{2^r r!}(ik\sigma)^{2r} \quad (N\to\infty) \\
&= e^{-k^2\sigma^2/2}.
\end{aligned}
\tag{36}
$$

Accordingly the central limit theorem holds and in the limit $N\to\infty$ the probability distribution $P(x)$ of x converges to a Gaussian peaked at $x=0$ with variance σ^2, i.e., $P(x)\to e^{-x^2/2\sigma^2}/\sqrt{2\pi\sigma^2}$, which in the original variable implies (21).

Diagrammatic approach to evaluate the moments of S

The expression for the first moment of $S-\langle S\rangle_*$ follows from (26) and is equal to

$$
\left\langle S-\langle S\rangle_*\right\rangle_N = \frac{1}{N}\left(1\left|\tilde{\mathcal{E}}^{(1)}\frac{1-\mathcal{E}'^N}{1-\mathcal{E}'}\mathcal{Q}_*\right|\rho_0\right),
\tag{37}
$$

which scales as $1/N$ as anticipated. Analogously the second moment is readily obtained from (29) by noticing that

$$
(\Delta S)_N^2 = \left\langle\left(S-\langle S\rangle_*\right)^2\right\rangle_N - \left(\langle S\rangle_N - \langle S\rangle_*\right)^2.
\tag{38}
$$

For future reference we find it useful to rederive it:

$$
\begin{aligned}
\left\langle\left(S-\langle S\rangle_*\right)^2\right\rangle_N &= \sum_{s_1}\cdots\sum_{s_N}\left(S-\langle S\rangle_*\right)^2 p\left(s_1,\ldots,s_N\middle|\rho_0\right) \\
&= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{s_1}\cdots\sum_{s_N}\delta s_i\,\delta s_j\left(1\left|\mathcal{E}_{s_N}\cdots\mathcal{E}_{s_1}\right|\rho_0\right) \\
&= \frac{1}{N^2}\sum_{i=1}^{N}\left(1\left|\tilde{\mathcal{E}}^{(2)}\mathcal{E}^{i-1}\right|\rho_0\right) + \frac{2}{N^2}\sum_{j=2}^{N}\sum_{i=1}^{j-1}\left(1\left|\tilde{\mathcal{E}}^{(1)}\mathcal{E}^{j-i-1}\tilde{\mathcal{E}}^{(1)}\mathcal{E}^{i-1}\right|\rho_0\right).
\end{aligned}
\tag{39}
$$

To simplify this we insert the decomposition of the ergodic channel $\mathcal{E}$ given in (9), namely, we insert $\mathcal{P}_* = |\rho_*)(1|$ or $\mathcal{E}'$ in place of $\mathcal{E}$. Notice however that

$$
\langle\delta s\rangle_* = \left(1\left|\tilde{\mathcal{E}}^{(1)}\right|\rho_*\right) = 0.
\tag{40}
$$

Due to this condition, the places in which we can insert $\mathcal{P}_*$ are limited. The nonvanishing contributions to the second moment hence read
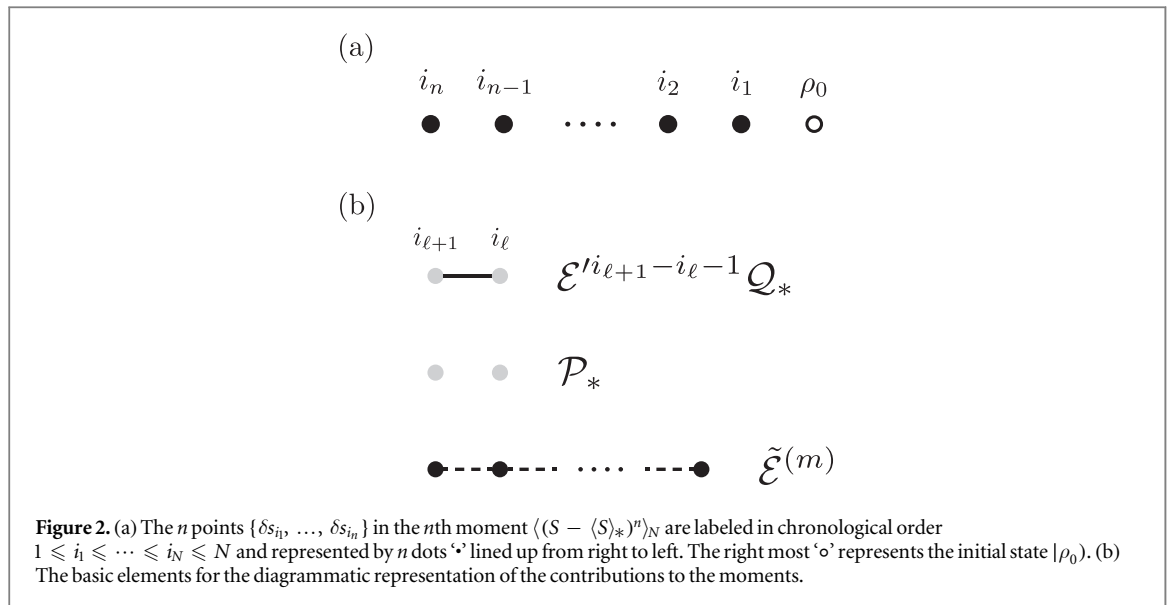
$$
\begin{aligned}
\left\langle\left(S-\langle S\rangle_*\right)^2\right\rangle_N &= \frac{1}{N^2}\sum_{i=1}^{N}\left(1\left|\tilde{\mathcal{E}}^{(2)}\right|\rho_*\right)\left(1\middle|\rho_0\right) + \frac{1}{N^2}\sum_{i=1}^{N}\left(1\left|\tilde{\mathcal{E}}^{(2)}\mathcal{E}'^{i-1}Q_*\right|\rho_0\right) \\
&\quad + \frac{2}{N^2}\sum_{j=2}^{N}\sum_{i=1}^{j-1}\left(1\left|\tilde{\mathcal{E}}^{(1)}\mathcal{E}'^{j-i-1}Q_*\tilde{\mathcal{E}}^{(1)}\right|\rho_*\right)\left(1\middle|\rho_0\right) \\
&\quad + \frac{2}{N^2}\sum_{j=2}^{N}\sum_{i=1}^{j-1}\left(1\left|\tilde{\mathcal{E}}^{(1)}\mathcal{E}'^{j-i-1}Q_*\tilde{\mathcal{E}}^{(1)}\mathcal{E}'^{i-1}Q_*\right|\rho_0\right),
\end{aligned}
\tag{41}
$$

and the direct computation of the summations yields (A.1) and (29).

It is now clear why the second moment $(\Delta S)_N^2$ in (29) as well as the second moment $\langle(S-\langle S\rangle_*)^2\rangle_N$ scales as $1/N$. There are two cases, as we saw in (39): (i) two points δs_i and δs_j coincide ($i=j$) and we have a single summation $\frac{1}{N^2}\sum_i$; (ii) two points δs_i and δs_j do not coincide ($i\neq j$) and we have double summations $\frac{1}{N^2}\sum\sum_{i\neq j}$. In any case, once $\mathcal{E}^k$ with some power k ($=j-i-1$ or $i-1$ in the above formula for the second moment) is substituted by $\mathcal{P}_*$, a summation accumulates as $\frac{1}{N}\sum\to O(1)$, while the contribution from $\mathcal{E}'$ does not: recall the geometric series in (11), where the contribution from $\mathcal{E}'$ remains $O(1/N)$. Thus, the substitution rules for estimating the scaling are:

$$
\frac{1}{N}\sum\mathcal{P}_*\to O(1) \quad\text{and}\quad \frac{1}{N}\sum\mathcal{E}'^k\to O(1/N).
\tag{42}
$$

Due to $(1|\tilde{\mathcal{E}}^{(1)}|\rho_*)=0$ and the coincidence of δs_i and δs_j, we can insert at most one $\mathcal{P}_*$ in place of $\mathcal{E}$ for the second moment $\langle(S-\langle S\rangle_*)^2\rangle_N$: see (41). Therefore, the second moment $\langle(S-\langle S\rangle_*)^2\rangle_N$ is at most $O(1/N)$,

Figure 2. (a) The n points $\{\delta s_{i_1}, \ldots, \delta s_{i_n}\}$ in the nth moment $\langle (S - \langle S \rangle_*)^n \rangle_N$ are labeled in chronological order $1 \leqslant i_1 \leqslant \cdots \leqslant i_N \leqslant N$ and represented by n dots '•' lined up from right to left. The right most '∘' represents the initial state $|\rho_0\rangle$. (b) The basic elements for the diagrammatic representation of the contributions to the moments.

and so is the variance $(\Delta S)_N^2$. Note that the second substitution rule in (42) is not valid if the channel $\mathcal{E}$ is ergodic but not mixing. Indeed, in such a case, the last term in (41) yields $O(1/N)$, as mentioned in the beginning of this subsection. The rule is safe if $\mathcal{E}$ is mixing.

We can generalize the above way of estimating the scaling to higher central moments, but a bit more sophisticated rules are required to check the asymptotic normality: we need to care about not only the scalings but also their coefficients. Anyway, the basic strategy to collect the leading-order contributions is to try to insert $\mathcal{P}_*$ as many times as possible in place of $\mathcal{E}$ avoiding $(1| \tilde{\mathcal{E}}^{(1)} |\rho_*) = 0$. Another important observation is that the insertion of $\mathcal{P}_* = |\rho_*)(1|$ 'breaks' the process into pieces. Recognizing these points, we introduce a diagrammatic way of representing the contributions to the moments.

The nth moment is given by

$$\left\langle \left(S - \langle S \rangle_* \right)^n \right\rangle_N = \frac{1}{N^n} \sum_{i_1=1}^{N} \cdots \sum_{i_n=1}^{N} \sum_{s_1} \cdots \sum_{s_N} \delta s_{i_1} \cdots \delta s_{i_n} \left(1 \, \Big| \, \mathcal{E}_{s_N} \cdots \mathcal{E}_{s_1} \Big| \rho_0 \right). \tag{43}$$

Within the summations over $\{i_1, \ldots, i_n\}$ we relabel the n points $\{\delta s_{i_1}, \ldots, \delta s_{i_n}\}$ in chronological order $1 \leqslant i_1 \leqslant \cdots \leqslant i_N \leqslant N$ and represent them by n dots '•' lined up in chronological order from right to left. See figure 2(a). The right most '∘' represents the initial state $|\rho_0\rangle$, and a trace $(1|$ is supposed to be at the left end. The points can coincide ($i_\ell = i_{\ell+1}$, as in the case $i = j$ for the second moment: see (39)), while between nondegenerate points ($i_\ell < i_{\ell+1}$) there are $\mathcal{E}^{i_{\ell+1}-i_\ell-1}$ (with a convention $i_0 = 0$), which are to be substituted by $\mathcal{P}_*$ or $\mathcal{E}'^{i_{\ell+1}-i_\ell-1}\mathcal{Q}_*$, as we did for the second moment (we need $\mathcal{Q}_*$ to remove $\mathcal{P}_*$ from $\mathcal{E}'^{i_{\ell+1}-i_\ell-1}$ when $i_{\ell+1} - i_\ell - 1 = 0$: see (41)). Now,

(i) When $\mathcal{E}^{i_{\ell+1}-i_\ell-1}$ between two points is substituted by $\mathcal{E}'^{i_{\ell+1}-i_\ell-1}\mathcal{Q}_*$, we connect the two points by a solid line.

(ii) When $\mathcal{E}^{i_{\ell+1}-i_\ell-1}$ between two points is substituted by $\mathcal{P}_*$, we leave the two points disconnected.

(iii) In the case where two or more points coincide, we connect the points by dashed lines. (Note that '∘' cannot be connected by a dashed line.)

See figure 2(b). There are two constraints due to $(1|\tilde{\mathcal{E}}^{(1)}|\rho_*) = 0$:

(a) The left most two points are surely connected either by a solid line or by a dashed line, since we cannot insert $\mathcal{P}_*$ between them due to $(1| \tilde{\mathcal{E}}^{(1)} |\rho_*) = 0$ with the left most trace $(1|$.

(b) Each point (except for '∘') must be connected with at least one adjacent point either by a solid line or by a dashed line, since we cannot insert $\mathcal{P}_*$ on both sides of a point due to $(1|\tilde{\mathcal{E}}^{(1)}|\rho_*) = 0$.

Then, it is easy to draw the diagrams relevant to the leading-order contributions, with the largest possible number of $\mathcal{P}_*$ inserted. The relevant diagrams for $n = 2, 3, 4$ are shown in figure 3 (see how the two diagrams for $n = 2$ correspond to the two leading-order terms in (41)). For each diagram contributing to the nth moment:
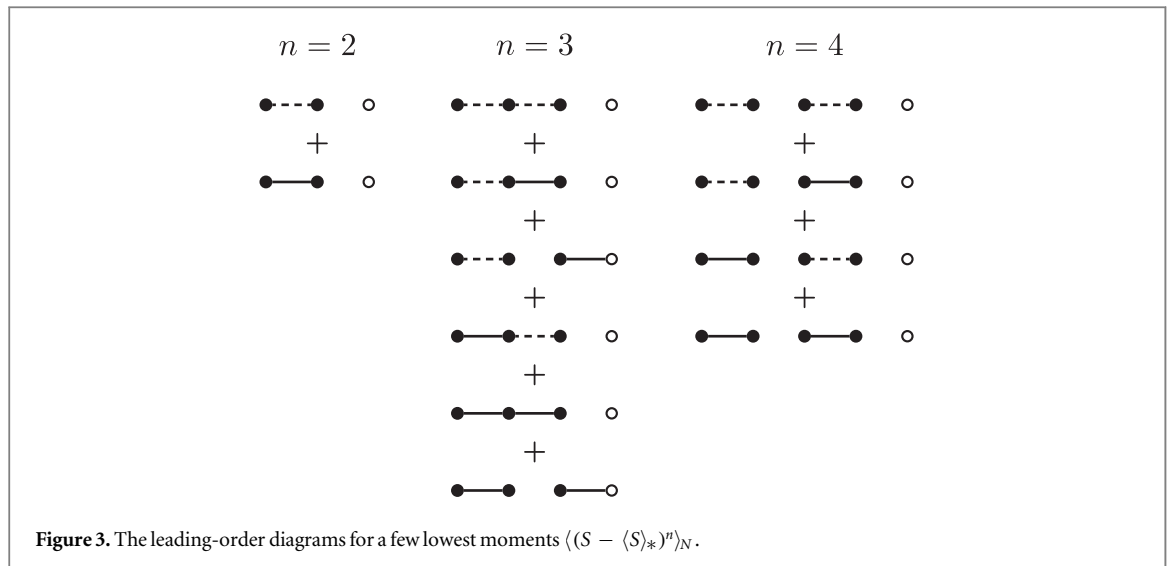
Figure 3. The leading-order diagrams for a few lowest moments $\langle (S - \langle S \rangle_*)^n \rangle_N$.

(1) Assign $\mathcal{E}'^{i_{\ell+1}-i_\ell-1}Q_*$ to each solid line.

(2) Insert $\mathcal{P}_* = |\rho_*)(1|$ for each space between disconnected points.

(3) Turn each (group of) dot(s) '•' (connected by dashed lines) into $\tilde{\mathcal{E}}^{(m)}$ (where m is the number of connected dots) while '∘' into $|\rho_0)$.

(4) Close each diagram with a trace $(1|$ at the left end.

(5) Put $\dfrac{1}{N^n}\sum\cdots\sum$ to sum the contributions over all possible distances between nondegenerate points respecting the chronological ordering of the points, with an appropriate coefficient counting how many times such a diagram (the specific ordering of the points) appears in the original full range summation $\dfrac{1}{N^n}\sum\cdots\sum$ exploring all possible orderings of the points. The right coefficient reads $n!/m_1!\,m_2!\cdots$, where m_i are the numbers of coincident points connected by dashed lines in the relevant diagram and the factors $m_i!$ are to disregard the orderings among the coincident points.

It is easily recognized from figure 3 that the maximum number of $\mathcal{P}_*$ we can insert for the nth moment is given by $\left\lfloor \dfrac{n}{2} \right\rfloor$ (where now $\lfloor x \rfloor$ denotes the largest integer not greater than x). Therefore, the substitution rules in (42) tell us that the nth moment scales as anticipated in (33) (the power of N is obtained by $n - \left\lfloor \dfrac{n}{2} \right\rfloor = \left\lceil \dfrac{n}{2} \right\rceil$). As discussed in the beginning of this subsection, this is the right scaling for the central limit theorem, and only the even moments ($n = 2, 4, 6, \ldots$) are relevant. An important observation is that the leading-order contributions to the even moments are independent of the initial state ρ_0, since '∘' representing the initial state $|\rho_0)$ is always disconnected from the first '•'. See figure 3.

Let us look more carefully at the fourth moment. The leading-order contributions represented by the diagrams in figure 3 read

$$
\begin{aligned}
\left\langle \left(S - \langle S \rangle_*\right)^4 \right\rangle_N &= \frac{4!}{(2!)^2} \frac{1}{N^4} \sum_{i_3=2}^{N} \sum_{i_1=1}^{i_3-1} \left(1 \left| \tilde{\mathcal{E}}^{(2)} \right| \rho_* \right) \left(1 \left| \tilde{\mathcal{E}}^{(2)} \right| \rho_* \right) \\
&+ \frac{4!}{2!} \frac{1}{N^4} \sum_{i_3=3}^{N} \sum_{i_2=2}^{i_3-1} \sum_{i_1=1}^{i_2-1} \left(1 \left| \tilde{\mathcal{E}}^{(2)} \right| \rho_* \right) \left(1 \left| \tilde{\mathcal{E}}^{(1)} \mathcal{E}'^{i_2-i_1-1}Q_* \tilde{\mathcal{E}}^{(1)} \right| \rho_* \right) \\
&+ \frac{4!}{2!} \frac{1}{N^4} \sum_{i_4=3}^{N} \sum_{i_3=2}^{i_4-1} \sum_{i_1=1}^{i_3-1} \left(1 \left| \tilde{\mathcal{E}}^{(1)} \mathcal{E}'^{i_4-i_3-1}Q_* \tilde{\mathcal{E}}^{(1)} \right| \rho_* \right) \left(1 \left| \tilde{\mathcal{E}}^{(2)} \right| \rho_* \right) \\
&+ 4! \frac{1}{N^4} \sum_{i_4=4}^{N} \sum_{i_3=3}^{i_4-1} \sum_{i_2=2}^{i_3-1} \sum_{i_1=1}^{i_2-1} \left(1 \left| \tilde{\mathcal{E}}^{(1)} \mathcal{E}'^{i_4-i_3-1}Q_* \tilde{\mathcal{E}}^{(1)} \right| \rho_* \right) \left(1 \left| \tilde{\mathcal{E}}^{(1)} \mathcal{E}'^{i_2-i_1-1}Q_* \tilde{\mathcal{E}}^{(1)} \right| \rho_* \right) + O\left(1/N^3\right)
\end{aligned}
$$

$$
\begin{aligned}
= \;& \frac{4!}{(2!)^2} \frac{1}{N^4} \frac{1}{2} N(N-1) \left(1 \left| \tilde{\mathcal{E}}^{(2)} \right| \rho_* \right) \left(1 \left| \tilde{\mathcal{E}}^{(2)} \right| \rho_* \right) \\
& + \frac{4!}{2!} \frac{1}{N^4} \frac{1}{2} N(N-1) \left(1 \left| \tilde{\mathcal{E}}^{(2)} \right| \rho_* \right) \left(1 \left| \tilde{\mathcal{E}}^{(1)} \frac{Q_*}{1-\mathcal{E}'} \tilde{\mathcal{E}}^{(1)} \right| \rho_* \right) \\
& + \frac{4!}{2!} \frac{1}{N^4} \frac{1}{2} N(N-1) \left(1 \left| \tilde{\mathcal{E}}^{(1)} \frac{Q_*}{1-\mathcal{E}'} \tilde{\mathcal{E}}^{(1)} \right| \rho_* \right) \left(1 \left| \tilde{\mathcal{E}}^{(2)} \right| \rho_* \right) \\
& + 4! \frac{1}{N^4} \frac{1}{2} N(N-1) \left(1 \left| \tilde{\mathcal{E}}^{(1)} \frac{Q_*}{1-\mathcal{E}'} \tilde{\mathcal{E}}^{(1)} \right| \rho_* \right) \left(1 \left| \tilde{\mathcal{E}}^{(1)} \frac{Q_*}{1-\mathcal{E}'} \tilde{\mathcal{E}}^{(1)} \right| \rho_* \right) + O\left(1/N^3\right).
\end{aligned}
\tag{44}
$$

This suggests that by the summations each of the leading-order contributions to an even moment $\langle (S - \langle S \rangle_*)^n \rangle_N$ acquires a common factor (binomial coefficient) $N!/(n/2)!(N-n/2)!$ while each $\mathcal{E}'^{i_{\ell+1}-i_\ell-1}$ is transformed into $(1-\mathcal{E}')^{-1}$. Indeed, each leading-order diagram for an even moment consists of pairs of points (i_{2r-1}, i_{2r}) $(r = 1, \ldots, n/2)$ connected by dashed or solid lines (see figure 3), and its evaluation proceeds two points by two points with the help of the following formulas for $r = 1, \ldots, n/2$ (with a convention $i_{n+1} = N + 1$): for pairs of coincident points connected by dashed lines

$$
\sum_{i_{2r-1}=r}^{i_{2r+1}-1} \frac{\left(i_{2r-1}-1\right)!}{(r-1)!\left(i_{2r-1}-r\right)!} = \frac{\left(i_{2r+1}-1\right)!}{r!\left(i_{2r+1}-r-1\right)!}
\tag{45}
$$

(we have a single sum for each coincident pair: see (44)), while for pairs of nondegenerate points connected by solid lines

$$
\begin{aligned}
& \sum_{i_{2r}=r+1}^{i_{2r+1}-1} \sum_{i_{2r-1}=r}^{i_{2r}-1} \mathcal{E}'^{i_{2r}-i_{2r-1}-1} \frac{\left(i_{2r-1}-1\right)!}{(r-1)!\left(i_{2r-1}-r\right)!} \\
& = \sum_{i_{2r-1}=r}^{i_{2r+1}-1} \frac{1-\mathcal{E}'^{i_{2r+1}-i_{2r-1}-1}}{1-\mathcal{E}'} \frac{\left(i_{2r-1}-1\right)!}{(r-1)!\left(i_{2r-1}-r\right)!} \\
& = \frac{\left(i_{2r+1}-1\right)!}{r!\left(i_{2r+1}-r-1\right)!} \frac{1}{1-\mathcal{E}'} + O\left(i_{2r+1}^{r-1}\right).
\end{aligned}
\tag{46}
$$

(The actual ranges of the summations in the leading-order contributions to the even moments are slightly different from these in (45) and (46), but the corrections are finite and become negligible in the asymptotic regime $N \gg n$.) In this way, each leading-order diagram for an even moment acquires the binomial coefficient $N!/(n/2)!(N-n/2)!$ with $\mathcal{E}'^{i_{\ell+1}-i_\ell-1}$ being transformed into $(1-\mathcal{E}')^{-1}$. This leads us to the following recipe for obtaining the expressions for the leading-order contributions to any even moment directly from the relevant diagrams:

(1′) Assign $(1|\tilde{\mathcal{E}}^{(2)}|\rho_*)/2$ to each pair of points connected by a dashed line.

(2′) Assign $(1|\tilde{\mathcal{E}}^{(1)}(1-\mathcal{E}')^{-1}Q_*\tilde{\mathcal{E}}^{(1)}|\rho_*)$ to each pair of points connected by a solid line.

(3′) Give a common factor $n!N!/N^n(n/2)!(N-n/2)! \sim n!/N^{n/2}(n/2)!$ to each diagram.

Then, the leading-order contributions to the even moments factorize as shown in figure 4 and yield (34).

5. Use of correlations

An important difference from the standard strategy for parameter estimation, where independent identical experiments are performed to collect data, is that in the present sequential scheme the correlations among the measurement data are available for estimation. Combining the information attainable from the correlations with that from the average S, the precision of the estimation can be enhanced. The primary motivation of the present paper is to explore this possibility.

For instance, one can compute

$$
C_\ell = \frac{1}{N-\ell} \sum_{i=1}^{N-\ell} s_i s_{i+\ell}
\tag{47}
$$

from a single sequence of N measurement outcomes $\{s_1, \ldots, s_N\}$, which captures the correlation between two data separated by a distance ℓ. In the presence of the correlations among the data, C_ℓ may depend on the target
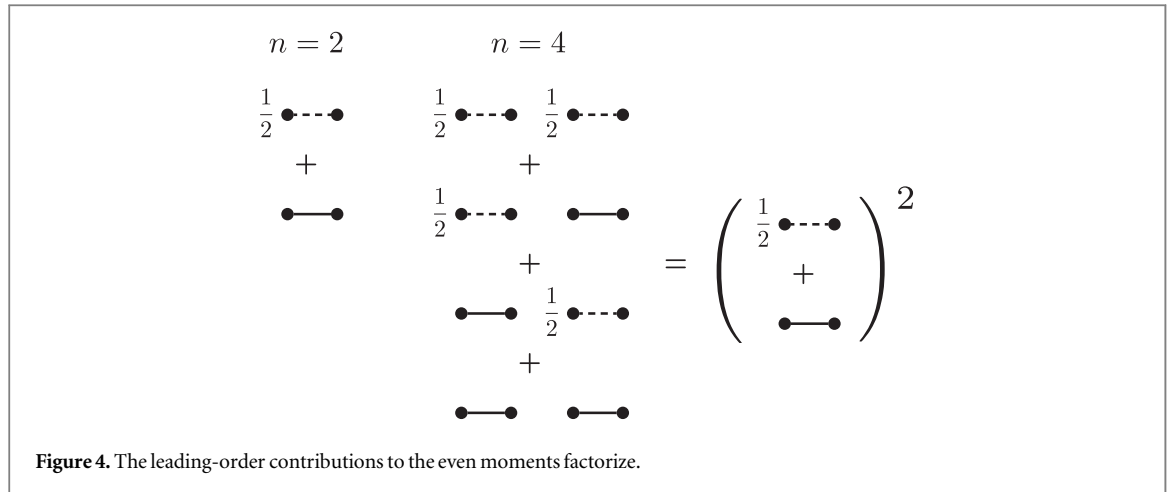
Figure 4. The leading-order contributions to the even moments factorize.

parameter *g* in a way that cannot be deduced solely from *S*. This might provide additional knowledge on how the parameter *g* is encoded in the process and can enhance the precision of the estimation of *g*.

In principle, ℓ ranges $\ell = 1, \ldots, N - 1$, but recall that the correlation between two data are expected to decay exponentially as ℓ increases under a mixing channel: C_ℓ with ℓ greater than the correlation length would not contain useful information. In addition, *N* should be much greater than ℓ so that the number $N - \ell$ of data used to evaluate C_ℓ is large enough. Therefore, we will require $N \gg L \geqslant \ell$, with *L* being the maximum ℓ we take to estimate the parameter *g*.

The correlations C_ℓ are also self-averaging quantities. Moreover, we are able to prove that the central limit theorem holds for the set of quantities $\boldsymbol{X} = (S, C_1, \ldots, C_L)$. First, the expectation value of C_ℓ is evaluated as

$$
\begin{aligned}
\left\langle C_\ell \right\rangle_N &= \sum_{s_1} \cdots \sum_{s_N} C_\ell \, p\left(s_1, \ldots, s_N \big| \rho_0\right) \\
&= \frac{1}{N - \ell} \sum_{i=1}^{N-\ell} \sum_{s_1} \cdots \sum_{s_N} s_i s_{i+\ell} \left(1 \, \big| \, \mathcal{E}_{s_N} \cdots \mathcal{E}_{s_1} \big| \rho_0\right) \\
&= \frac{1}{N - \ell} \sum_{i=1}^{N-\ell} \left(1 \, \big| \, \mathcal{E}^{(1)} \mathcal{E}^{\ell-1} \mathcal{E}^{(1)} \mathcal{E}^{i-1} \big| \rho_0\right) \\
&= \left\langle C_\ell \right\rangle_* + \frac{1}{N - \ell} \left(1 \, \big| \, \mathcal{E}^{(1)} \mathcal{E}^{\ell-1} \mathcal{E}^{(1)} \frac{1 - \mathcal{E}'^{N-\ell}}{1 - \mathcal{E}'} \mathcal{Q}_* \big| \rho_0\right) \\
&\to \left\langle C_\ell \right\rangle_* \quad (N \to \infty),
\end{aligned}
\tag{48}
$$

which approaches

$$
\left\langle C_\ell \right\rangle_* = \left(1 \, \big| \, \mathcal{E}^{(1)} \mathcal{E}^{\ell-1} \mathcal{E}^{(1)} \big| \rho_*\right) = \left\langle s_i s_{i+\ell} \right\rangle_*
\tag{49}
$$

under the ergodicity of the channel $\mathcal{E}$. Then, let us look at the *n*th moment

$$
\mu_n(\boldsymbol{k}) = \left\langle \left(k_0 \left(S - \langle S \rangle_* \right) + \sum_{\ell=1}^{L} k_\ell \left(C_\ell - \left\langle C_\ell \right\rangle_* \right) \right)^n \right\rangle_N.
\tag{50}
$$

It is an *n*th-order polynomial of $\boldsymbol{k} = (k_0, \ldots, k_L)$, and is a collection of all the *n*th moments among $\boldsymbol{X} = (S, C_1, \ldots, C_L)$ as its coefficients. Since we are interested in the asymptotic limit $N \to \infty$, we collect the leading-order contributions to $\mu_n(\boldsymbol{k})$ for large *N*. The idea to do that is basically the same as that for *S*: we try to insert $\mathcal{P}_*$ as many times as possible in place of $\mathcal{E}$ between points δs_i from $S - \langle S \rangle_*$ and pairs of points

$$
\delta\left(s_i s_{i+\ell}\right)_\ell = s_i s_{i+\ell} - \left\langle C_\ell \right\rangle_*
\tag{51}
$$

Figure 5. Each pair of points $\delta(s_i s_{i+\ell})_\ell$ is also represented by a dot '•', and if the pair overlaps with another pair or a point δs_i, we connect the couple of dots '•' by a dashed line. Several terms are involved in such a single connected diagram as shown here. These diagrams represent $\tilde{\mathcal{E}}_\ell^{(2)}$ and $\tilde{\mathcal{E}}_{\ell\ell'}^{(2)}$ in (58) and (59).
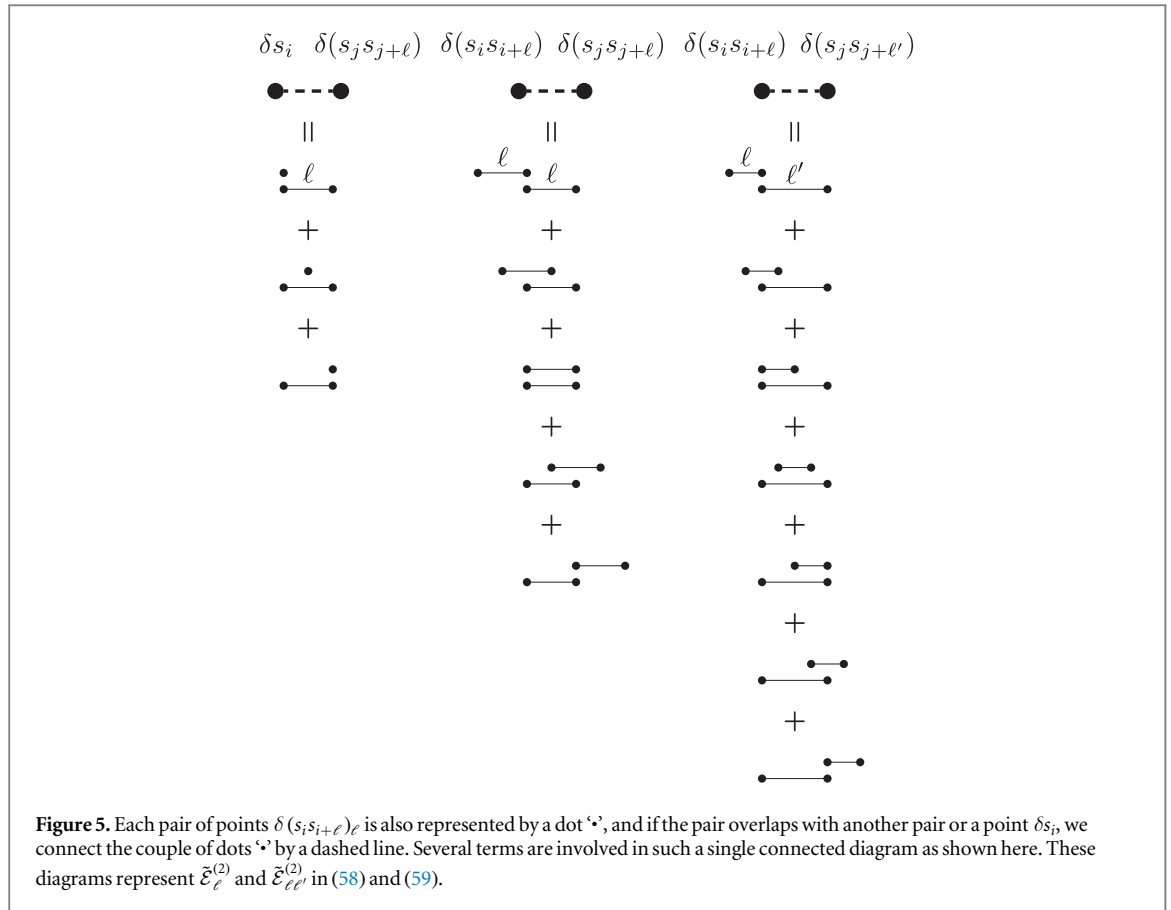
from $C_\ell - \langle C_\ell \rangle_*$. Since we have $\langle \delta s_i \rangle_* = (1|\tilde{\mathcal{E}}^{(1)}|\rho_*) = 0$ (equation (40)) and

$$\left\langle \delta\left(s_i s_{i+\ell}\right)_\ell \right\rangle_* = \left(1 \left| \tilde{\mathcal{E}}_\ell^{(1)} \right| \rho_* \right) = 0, \tag{52}$$

where

$$\tilde{\mathcal{E}}_\ell^{(1)} = \sum_s \sum_{s'} \delta\left(ss'\right)_\ell \mathcal{E}_s \mathcal{E}^{\ell-1} \mathcal{E}_{s'}, \tag{53}$$

there should be at least two pieces (points δs_i and/or pairs of points $\delta(s_i s_{i+\ell})_\ell$) between two $\mathcal{P}_*$. We just have to generalize the diagrammatic rules in figure 2: we represent each pair of points $\delta(s_i s_{i+\ell})_\ell$ by a dot '•', too, and if the pair overlaps with another pair or a point δs_i we connect the couple of dots '•' by a dashed line (see figure 5), while if it does not we leave it disconnected from or connect it with its adjacent dot by a solid line depending on whether $\mathcal{P}_*$ is inserted between them or not. The ranges of the summations $\sum \cdots \sum$ exploring all possible distances between dots '•' should be carefully arranged depending on whether the dots represent points δs_i or pairs of points $\delta(s_i s_{i+\ell})_\ell$, and some of the prefactors in $1/N^n$ are replaced by $1/(N-\ell)$, but such details become irrelevant in the asymptotic regime $N \gg n, L$. Then, the analysis goes in the same way as before, the leading-order diagrams are again given by figure 3, and the nth moment $\mu_n(\boldsymbol{k})$ for an even n asymptotically factorizes pairwise as figure 4, where the pair of dots '•' connected by a dashed line or a solid line represents the collection of all pairwise combinations among δs_i and $\delta(s_i s_{i+\ell})_\ell$ ($\ell = 1, \ldots, L$), with some care on the coefficients to distinguish different orderings of the pieces, i.e., give a coefficient $1/2$ to the pair connected by the solid line in figure 4 and collect contributions with different orderings of the pieces (see the second and third terms in $\Sigma_{00}, \Sigma_{0\ell}$, and $\Sigma_{\ell\ell'}$ in (55)–(57) below). We get

$$\mu_n(\boldsymbol{k}) = \begin{cases} \dfrac{n!}{(2N)^{n/2}(n/2)!} \left(\sum_{\ell=0}^{L} \sum_{\ell'=0}^{L} k_\ell \Sigma_{\ell\ell'} k_{\ell'} \right)^{n/2} + O\left(1/N^{\frac{n}{2}+1}\right) & (n \text{ even}), \\ O\left(1/N^{\lceil \frac{n}{2} \rceil}\right) & (n \text{ odd}), \end{cases} \tag{54}$$

where

$$\Sigma_{00} = \left(1 \left| \tilde{\mathcal{E}}^{(2)} \right| \rho_* \right) + 2\left(1 \left| \tilde{\mathcal{E}}^{(1)} \frac{\mathcal{Q}_*}{1 - \mathcal{E}'} \tilde{\mathcal{E}}^{(1)} \right| \rho_* \right) = \sigma^2, \tag{55}$$

$$\Sigma_{0\ell} = \Sigma_{\ell 0} = \left(1 \left| \tilde{\mathcal{E}}_{\ell}^{(2)} \right| \rho_* \right) + \left(1 \left| \tilde{\mathcal{E}}^{(1)} \frac{\mathcal{Q}_*}{1-\mathcal{E}'} \tilde{\mathcal{E}}_{\ell}^{(1)} \right| \rho_* \right) + \left(1 \left| \tilde{\mathcal{E}}_{\ell}^{(1)} \frac{\mathcal{Q}_*}{1-\mathcal{E}'} \tilde{\mathcal{E}}^{(1)} \right| \rho_* \right), \tag{56}$$

$$\Sigma_{\ell\ell'} = \left(1 \left| \tilde{\mathcal{E}}_{\ell\ell'}^{(2)} \right| \rho_* \right) + \left(1 \left| \tilde{\mathcal{E}}_{\ell}^{(1)} \frac{\mathcal{Q}_*}{1-\mathcal{E}'} \tilde{\mathcal{E}}_{\ell'}^{(1)} \right| \rho_* \right) + \left(1 \left| \tilde{\mathcal{E}}_{\ell'}^{(1)} \frac{\mathcal{Q}_*}{1-\mathcal{E}'} \tilde{\mathcal{E}}_{\ell}^{(1)} \right| \rho_* \right), \tag{57}$$

with $\tilde{\mathcal{E}}^{(m)}$ and $\tilde{\mathcal{E}}_{\ell}^{(1)}$ defined in (30) and (53), respectively,

$$\tilde{\mathcal{E}}_{\ell}^{(2)} = \sum_{s_1} \sum_{s_2} \left(\delta s_2 + \delta s_1\right) \delta\left(s_2 s_1\right)_{\ell} \mathcal{E}_{s_2} \mathcal{E}^{\ell-1} \mathcal{E}_{s_1} + \sum_{k=1}^{\ell-1} \sum_{s_1} \sum_{s_2} \sum_{s_3} \delta s_2 \delta\left(s_3 s_1\right)_{\ell} \mathcal{E}_{s_3} \mathcal{E}^{k-1} \mathcal{E}_{s_2} \mathcal{E}^{\ell-k-1} \mathcal{E}_{s_1} \tag{58}$$

and

$$\tilde{\mathcal{E}}_{\ell\ell'}^{(2)} = \tilde{\mathcal{E}}_{\ell\ell'}^{\circ\star\bullet} + \sum_{k=1}^{\min(\ell,\ell')-1} \tilde{\mathcal{E}}_{\ell\ell',k}^{\circ\bullet\circ\bullet} + \delta_{\ell\ell'} \tilde{\mathcal{E}}_{\ell}^{\star\star} + \left(1 - \delta_{\ell\ell'}\right)\left(\tilde{\mathcal{E}}_{\ell\ell'}^{\star\circ\bullet} + \tilde{\mathcal{E}}_{\ell\ell'}^{\bullet\circ\star} + \tilde{\mathcal{E}}_{\ell\ell'}^{\bullet\circ\circ\bullet}\right) \tag{59}$$

composed of

$$\tilde{\mathcal{E}}_{\ell}^{\star\star} = \sum_{s_1} \sum_{s_2} \left[\delta\left(s_2 s_1\right)_{\ell}\right]^2 \mathcal{E}_{s_2} \mathcal{E}^{\ell-1} \mathcal{E}_{s_1}, \tag{60}$$

$$\tilde{\mathcal{E}}_{\ell\ell'}^{\circ\star\bullet} = \sum_{s_1} \sum_{s_2} \sum_{s_3} \delta\left(s_3 s_2\right)_{\ell} \delta\left(s_2 s_1\right)_{\ell'} \mathcal{E}_{s_3} \mathcal{E}^{\ell-1} \mathcal{E}_{s_2} \mathcal{E}^{\ell'-1} \mathcal{E}_{s_1} + \left(\ell \leftrightarrow \ell'\right), \tag{61}$$

$$\tilde{\mathcal{E}}_{\ell\ell',k}^{\circ\bullet\circ\bullet} = \sum_{s_1} \sum_{s_2} \sum_{s_3} \sum_{s_4} \delta\left(s_4 s_2\right)_{\ell} \delta\left(s_3 s_1\right)_{\ell'} \mathcal{E}_{s_4} \mathcal{E}^{\ell-k-1} \mathcal{E}_{s_3} \mathcal{E}^{k-1} \mathcal{E}_{s_2} \mathcal{E}^{\ell'-k-1} \mathcal{E}_{s_1} + \left(\ell \leftrightarrow \ell'\right), \tag{62}$$

$$\tilde{\mathcal{E}}_{\ell\ell'}^{\star\circ\bullet} = \sum_{s_1} \sum_{s_2} \sum_{s_3} \delta\left(s_3 s_2\right)_{\min(\ell,\ell')} \delta\left(s_3 s_1\right)_{\max(\ell,\ell')} \mathcal{E}_{s_3} \mathcal{E}^{\min(\ell,\ell')-1} \mathcal{E}_{s_2} \mathcal{E}^{|\ell-\ell'|-1} \mathcal{E}_{s_1}, \tag{63}$$

$$\tilde{\mathcal{E}}_{\ell\ell'}^{\bullet\circ\star} = \sum_{s_1} \sum_{s_2} \sum_{s_3} \delta\left(s_3 s_1\right)_{\max(\ell,\ell')} \delta\left(s_2 s_1\right)_{\min(\ell,\ell')} \mathcal{E}_{s_3} \mathcal{E}^{|\ell-\ell'|-1} \mathcal{E}_{s_2} \mathcal{E}^{\min(\ell,\ell')-1} \mathcal{E}_{s_1}, \tag{64}$$

$$\tilde{\mathcal{E}}_{\ell\ell'}^{\bullet\circ\circ\bullet} = \sum_{k=1}^{|\ell-\ell'|-1} \sum_{s_1} \sum_{s_2} \sum_{s_3} \sum_{s_4} \delta\left(s_3 s_2\right)_{\min(\ell,\ell')} \delta\left(s_4 s_1\right)_{\max(\ell,\ell')} \mathcal{E}_{s_4} \mathcal{E}^{k-1} \mathcal{E}_{s_3} \mathcal{E}^{\min(\ell,\ell')-1} \mathcal{E}_{s_2} \mathcal{E}^{|\ell-\ell'|-k-1} \mathcal{E}_{s_1}, \tag{65}$$

corresponding to the diagrams in figure 5. We provide the complete expressions for the covariances among S and C_ℓ valid for any (even small) N in appendix A, whose asymptotic forms coincide with the covariances (55)–(57) divided by N.

This result shows that the set of scaled variables $\sqrt{N}\left(S - \langle S \rangle_*\right)$ and $\sqrt{N}\left(C_\ell - \langle C_\ell \rangle_*\right)$ asymptotically become normal in the limit $N \to \infty$. The characteristic function reads

$$\begin{aligned}
\chi(\boldsymbol{k}) &= \left\langle e^{i\sqrt{N} \sum_{\ell=0}^{L} k_\ell \left(X_\ell - \langle X_\ell \rangle_*\right)} \right\rangle_N \\
&= \sum_{n=0}^{\infty} \frac{1}{n!} \left(i\sqrt{N}\right)^n \mu_n(\boldsymbol{k}) \\
&\to \sum_{r=0}^{\infty} \frac{1}{r!} \left(-\frac{1}{2} \sum_{\ell=0}^{L} \sum_{\ell'=0}^{L} k_\ell \Sigma_{\ell\ell'} k_{\ell'}\right)^r \quad (N \to \infty) \\
&= e^{-\boldsymbol{k}^T \Sigma \boldsymbol{k}/2},
\end{aligned} \tag{66}$$

where Σ is the $(L+1) \times (L+1)$ matrix with its matrix elements given by the covariances in (55)–(57). The central limit theorem holds, and the probability distribution $P(\boldsymbol{X})$ of $\boldsymbol{X} = (S, C_1, \ldots, C_L)$ becomes asymptotically Gaussian
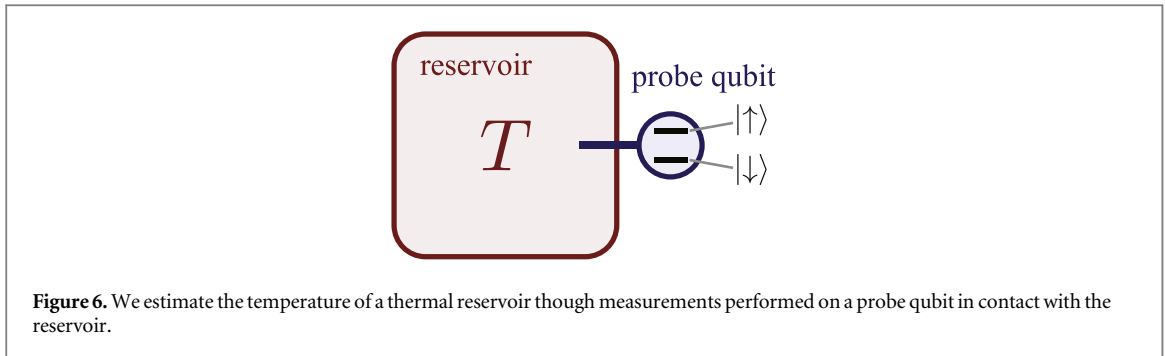
$$P(\boldsymbol{X}) \simeq \frac{1}{\sqrt{(2\pi/N)^{L+1} \det \Sigma}} e^{-\frac{1}{2}N(\boldsymbol{X}-\langle\boldsymbol{X}\rangle_*)^T \Sigma^{-1}(\boldsymbol{X}-\langle\boldsymbol{X}\rangle_*)}, \tag{67}$$

peaked at $\boldsymbol{X} = \langle\boldsymbol{X}\rangle_*$ with a shrinking covariance Σ/N. This ensures that the single-shot values $\boldsymbol{X}$ computed from a single sequence of measurement data well represent their expectation values $\langle\boldsymbol{X}\rangle_*$, through which we can estimate a parameter g. The uncertainty δg in the estimation of g is given by (22) with the Fisher information

$$\mathcal{F}_L(g) = \int d^{L+1}\boldsymbol{X} \, P(\boldsymbol{X}) \left(\frac{\partial}{\partial g} \ln P(\boldsymbol{X})\right)^2 \simeq N \frac{\partial \langle\boldsymbol{X}\rangle_*^T}{\partial g} \Sigma^{-1} \frac{\partial \langle\boldsymbol{X}\rangle_*}{\partial g}, \tag{68}$$

which increases linearly in N, and the uncertainty δg diminishes as $\delta g \simeq 1/\sqrt{N}$ (in (68) we have omitted the contribution from $\partial\Sigma/\partial g$ to the Fisher information $\mathcal{F}_L(g)$ since it does not grow with N). Moreover, this

Figure 6. We estimate the temperature of a thermal reservoir though measurements performed on a probe qubit in contact with the reservoir.

Fisher information $\mathcal{F}_L(g)$ for the estimation of g through a set of quantities $\boldsymbol{X} = (S, C_1, \ldots, C_L)$ can be greater than the Fisher information $\mathcal{F}(g) = \mathcal{F}_0(g)$ given in (23) for the estimation of the same g but solely through S. The precision of the estimation can be enhanced by looking at the correlation data C_ℓ in addition to the average S.

Here we have considered the two-point correlations C_ℓ as well as the average S. If we incorporate higher-order correlations with more points, the precision of the estimation can be further improved. On the other hand, correlations with too many points would not be helpful, since the number of data used to evaluate such correlations is reduced, and some of the points involved in the correlations are separated beyond the correlation length of the mixing channel supplying no more information than lower-order correlations. It is currently not clear to what extent we can improve the precision of the estimation by looking at higher-order correlations.

6. Example: estimation of the temperature of a reservoir

In this section we analyze an explicit example, where the correlations among the data collected by the sequential measurements would be useful for improving the estimation of a parameter. The setting we consider is related to quantum thermometry, which aims to use low-dimensional quantum systems (say qubits) as temperature probes to minimize the undesired disturbance on the sample (see e.g. [34, 35] and references therein). Specifically we focus on the paradigmatic example with a qubit probe in contact with a thermal reservoir at a finite temperature T. Our goal is to estimate the temperature T of the reservoir by monitoring the relaxation dynamics induced on the qubit, which effectively plays the role of a local 'thermometer' (figure 6). In our approach we describe the probe-reservoir coupling in terms of the resulting Markovian master equation [30, 36–39] operating on the probe, i.e.

$$\frac{\mathrm{d}}{\mathrm{d}t}\rho(t) = -\frac{\mathrm{i}}{2}\Omega\big[\sigma_z, \rho(t)\big] - \frac{1}{2}\gamma_+\big[\sigma_+\sigma_-\rho(t) + \rho(t)\sigma_+\sigma_- - 2\sigma_-\rho(t)\sigma_+\big]$$
$$- \frac{1}{2}\gamma_-\big[\sigma_-\sigma_+\rho(t) + \rho(t)\sigma_-\sigma_+ - 2\sigma_+\rho(t)\sigma_-\big], \tag{69}$$

where $\rho(t)$ represents the state of the qubit, $\hbar\Omega$ is the energy gap between the excited $|\uparrow\rangle$ and ground $|\downarrow\rangle$ states of the qubit, and

$$\sigma_z = |\uparrow\rangle\langle\uparrow| - |\downarrow\rangle\langle\downarrow|, \qquad \sigma_+ = |\uparrow\rangle\langle\downarrow|, \qquad \sigma_- = |\downarrow\rangle\langle\uparrow|. \tag{70}$$

The two relaxation constants γ_+ (for decay) and γ_- (for excitation) are related to the temperature of the reservoir T, respecting the detailed balance condition. For a bosonic thermal reservoir, they are given by [36–39]

$$\gamma_+ = \big(1 + n_{\mathrm{th}}\big)\gamma, \qquad \gamma_- = n_{\mathrm{th}}\gamma, \qquad n_{\mathrm{th}} = \frac{1}{e^{\hbar\Omega/k_B T} - 1}, \tag{71}$$

with k_B being the Boltzmann constant. We assume that the parameters Ω and γ (i.e., the characteristics of the thermometer) are known. Estimating the temperature T is then equivalent to estimating

$$\gamma_\beta = \gamma_+ + \gamma_- = \gamma \coth \frac{\hbar\Omega}{2k_B T}, \tag{72}$$

while $\gamma = \gamma_+ - \gamma_-$ is a known constant independent of the temperature T. The higher is the temperature, the larger is the decay rate γ_β.

6.1. Standard strategy

The information about the temperature T, namely, the parameter γ_β, is imprinted in the state of the qubit through the dynamics under the influence of the thermal reservoir, i.e., by the action of the quantum channel Λ_t which is the solution to the master equation (69). Then, the standard strategy to estimate the parameter γ_β is

(i) to prepare the qubit in a specific initial state ρ_0,

(ii) to let the qubit evolve $\rho(\tau) = \Lambda_\tau(\rho_0)$ for a certain time τ in contact with the thermal reservoir, and

(iii) to measure a specific observable in the state $\rho(\tau)$.

We repeat this process N times to collect measurement results, from which we estimate the parameter γ_β.

For instance, we prepare the qubit in a specific initial state ρ_0, say in the excited state $|\uparrow\rangle$, and after a fixed waiting time τ we measure the qubit to check whether it is in the excited state $|\uparrow\rangle$ or in the ground state $|\downarrow\rangle$. We repeat this process N times, and we estimate γ_β from the survival probability of the initial state $|\uparrow\rangle$ after time τ. Our measurement however can be weak and unsharp: here we consider the measurement which provokes the following back-action on the qubit

$$\rho \mapsto \mathcal{M}_s(\rho) = M_s \rho M_s^\dagger \quad (s = \pm 1) \tag{73}$$

with

$$\begin{cases} M_{+1} = \cos\eta \, |\uparrow\rangle\langle\uparrow| + \sin\eta \, |\downarrow\rangle\langle\downarrow|, \\ M_{-1} = \sin\eta \, |\uparrow\rangle\langle\uparrow| + \cos\eta \, |\downarrow\rangle\langle\downarrow|, \end{cases} \tag{74}$$

depending on the outcome of the measurement s. This measurement process can be simulated with a CNOT gate [40, 41]. The parameter η controls the precision and the strength of the measurement: $\eta = 0$ provides the projective measurement, while with $\eta = \pi/4$ the measurement gives totally random results with no disturbance on the measured system. The probability of obtaining the measurement outcome s in the state $\rho(\tau)$ is then given by

$$p_\tau(s|\rho_0) = \mathrm{Tr}\left\{\mathcal{M}_s(\rho(\tau))\right\} = \mathrm{Tr}\left\{\Pi_s \rho(\tau)\right\}, \tag{75}$$

where

$$\Pi_s = M_s^\dagger M_s = \begin{cases} \cos^2\eta \, |\uparrow\rangle\langle\uparrow| + \sin^2\eta \, |\downarrow\rangle\langle\downarrow| \quad (s = +1), \\ \sin^2\eta \, |\uparrow\rangle\langle\uparrow| + \cos^2\eta \, |\downarrow\rangle\langle\downarrow| \quad (s = -1) \end{cases} \tag{76}$$

are the POVM elements of this measurement. The uncertainty in the estimation is then bounded by the Cramér–Rao inequality [3, 7–12, 33]

$$\delta\gamma_\beta \gtrsim \frac{1}{\sqrt{NF(\gamma_\beta)}} \tag{77}$$

with the Fisher information given by

$$F(\gamma_\beta) = \sum_{s=\pm 1} p_\tau(s|\rho_0) \left(\frac{\partial}{\partial\gamma_\beta} \ln p_\tau(s|\rho_0)\right)^2. \tag{78}$$

For the present model, the Bloch vector of the qubit evolves as

$$\begin{cases} \langle\sigma_x\rangle_t = e^{-\gamma_\beta t/2}\left(\langle\sigma_x\rangle_0 \cos\Omega t - \langle\sigma_y\rangle_0 \sin\Omega t\right), \\ \langle\sigma_y\rangle_t = e^{-\gamma_\beta t/2}\left(\langle\sigma_x\rangle_0 \sin\Omega t + \langle\sigma_y\rangle_0 \cos\Omega t\right), \\ \langle\sigma_z\rangle_t = \left(\langle\sigma_z\rangle_0 + \frac{\gamma}{\gamma_\beta}\right)e^{-\gamma_\beta t} - \frac{\gamma}{\gamma_\beta}, \end{cases} \tag{79}$$

where $\sigma_x = \sigma_+ + \sigma_-$ and $\sigma_y = -i(\sigma_+ - \sigma_-)$. The equilibrium state ρ_{eq} is characterized by

$$\langle\sigma_x\rangle_{\mathrm{eq}} = \langle\sigma_y\rangle_{\mathrm{eq}} = 0, \quad \langle\sigma_z\rangle_{\mathrm{eq}} = -\frac{\gamma}{\gamma_\beta}, \tag{80}$$

namely

$$\rho_{\mathrm{eq}} = \frac{1}{2}\left(1 - \frac{\gamma}{\gamma_\beta}\sigma_z\right) = \frac{e^{-\beta\hbar\Omega\sigma_z/2}}{\mathrm{Tr}\, e^{-\beta\hbar\Omega\sigma_z/2}}. \tag{81}$$

Figure 7. The dependence of the quantum Fisher information $F_Q(\gamma_\beta)$ given in (84) on the polar angle θ of a generic pure initial state $|\psi_0\rangle = e^{-i\varphi/2}\cos(\theta/2)|\uparrow\rangle + e^{i\varphi/2}\sin(\theta/2)|\downarrow\rangle$ and on the waiting time τ, for different values of γ_β/γ, i.e., for different temperatures. Note that $F_Q(\gamma_\beta)$ is symmetric around the polar axis and is independent of the azimuthal angle φ.
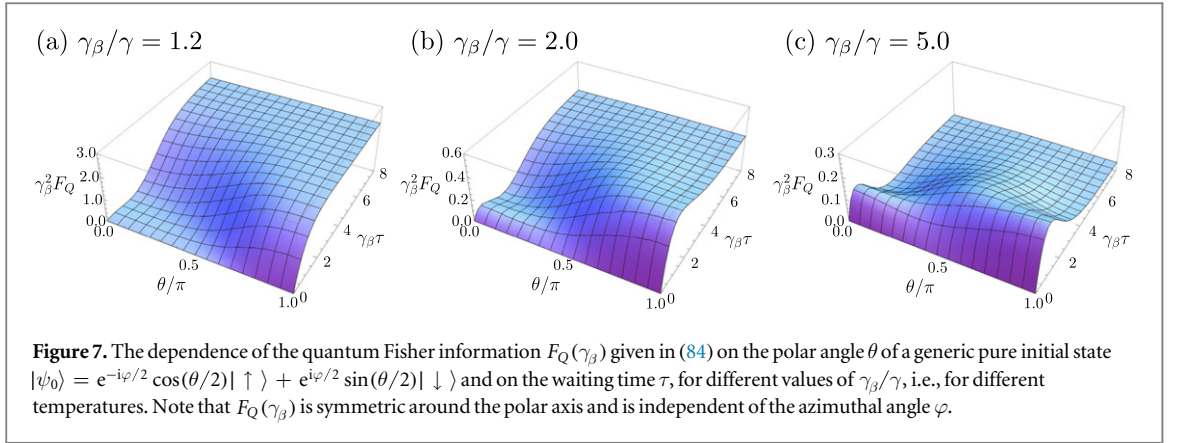
The probability distribution of the outcomes of the measurement (75) at time τ reads

$$p_\tau\left(\pm 1 \middle| \rho_0\right) = \frac{1}{2}\left(1 \pm \left\langle \sigma_z \right\rangle_\tau \cos 2\eta\right),\tag{82}$$

and the Fisher information $F(\gamma_\beta)$ in (78) is estimated to be

$$F(\gamma_\beta) = \frac{\cos^2 2\eta}{1 - \left\langle \sigma_z \right\rangle_\tau^2 \cos^2 2\eta}\left(\frac{\partial \left\langle \sigma_z \right\rangle_\tau}{\partial \gamma_\beta}\right)^2.\tag{83}$$

A larger Fisher information would be attainable by measuring a different observable. The maximum Fisher information one can reach with the optimal measurement is given by the quantum Fisher information [2, 3, 7–12],

$$F_Q(\gamma_\beta) = \text{Tr}\left\{\rho(\tau)L_{\gamma_\beta}^2\right\} = \frac{\partial \left\langle \boldsymbol{\sigma} \right\rangle_\tau^T}{\partial \gamma_\beta}V^{-1}\frac{\partial \left\langle \boldsymbol{\sigma} \right\rangle_\tau}{\partial \gamma_\beta},\tag{84}$$

with the symmetric logarithmic derivative

$$L_{\gamma_\beta} = \frac{\partial \left\langle \boldsymbol{\sigma} \right\rangle_\tau^T}{\partial \gamma_\beta}V^{-1}\left(\boldsymbol{\sigma} - \left\langle \boldsymbol{\sigma} \right\rangle_\tau\right),\tag{85}$$

where V is a 3×3 matrix whose matrix elements are given by

$$V_{ij} = \delta_{ij} - \left\langle \sigma_i \right\rangle_\tau \left\langle \sigma_j \right\rangle_\tau \qquad (i, j = x, y, z).\tag{86}$$

Notice here that both the Fisher information $F(\gamma_\beta)$ in (83) and the quantum Fisher information $F_Q(\gamma_\beta)$ in (84) depend on the choice of the initial state ρ_0. Because of the convexity of the quantum Fisher information, the maximum of the quantum Fisher information (the best estimation) is always achieved by choosing a pure input state $\rho_0 = |\psi_0\rangle\langle\psi_0|$ [3, 42]. Moreover, for the present problem, the ground state of the qubit $|\psi_0\rangle = |\downarrow\rangle$ is the optimal choice, in the sense that the maximum of $F_Q(\gamma_\beta)$ for a given temperature is achieved with $|\psi_0\rangle = |\downarrow\rangle$: see figure 7, and the temporal behavior of $F_Q(\gamma_\beta)$ for $|\psi_0\rangle = |\downarrow\rangle$ is plotted in figure 8(a). For this specific initial state $\rho_0 = |\downarrow\rangle\langle\downarrow|$, the Fisher information $F(\gamma_\beta)$ in (83) with $\eta = 0$ coincides with the quantum Fisher information $F_Q(\gamma_\beta)$ in (84), for any time τ and for any γ_β: the projective measurement to discriminate $|\uparrow\rangle$ and $|\downarrow\rangle$ is the optimal measurement. For nonvanishing $\eta > 0$ the Fisher information $F(\gamma_\beta)$ is reduced, and the weaker is the measurement, the smaller is the Fisher information $F(\gamma_\beta)$, as shown in figure 8(b).

6.2. Sequential scheme

Let us now turn our attention to the sequential scheme. First, it is important to check whether the channel $\mathcal{E}$ defined in (19) with $\mathcal{E}_s$ in (16) is mixing. For the present model, the spectrum of $\mathcal{E}$ is given by

$$\left\{1, \ e^{-\gamma_\beta\tau}, \ e^{-\left(\gamma_\beta/2 \pm i\Omega\right)\tau}\right\},\tag{87}$$

and therefore, $\mathcal{E}$ is mixing for any $\tau > 0$ with a unique fixed point (the eigenstate belonging to the eigenvalue 1)

$$\rho_* = \rho_{\text{eq}},\tag{88}$$
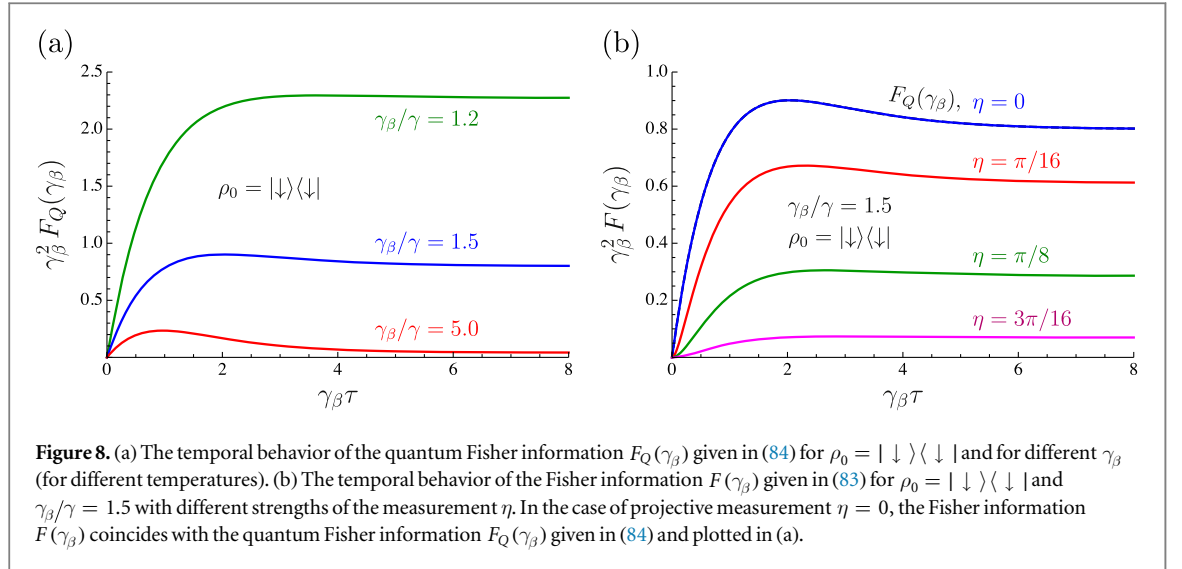
Figure 8. (a) The temporal behavior of the quantum Fisher information $F_Q(\gamma_\beta)$ given in (84) for $\rho_0 = |\downarrow\rangle\langle\downarrow|$ and for different γ_β (for different temperatures). (b) The temporal behavior of the Fisher information $F(\gamma_\beta)$ given in (83) for $\rho_0 = |\downarrow\rangle\langle\downarrow|$ and $\gamma_\beta/\gamma = 1.5$ with different strengths of the measurement η. In the case of projective measurement $\eta = 0$, the Fisher information $F(\gamma_\beta)$ coincides with the quantum Fisher information $F_Q(\gamma_\beta)$ given in (84) and plotted in (a).

which coincides with the equilibrium state ρ_{eq} in (81) of the free relaxation process. This mixing is apparently a direct consequence of the irreversibility of the relaxation process Λ_t of the probe qubit. Since $\mathcal{E}$ is mixing, the sequential scheme works for the present problem.

Let us take the average of the outcomes of a sequence of N measurements, S defined in (20), as a quantity through which we estimate γ_β. For the present model, its expectation value is computed to be

$$\langle S \rangle_N = -\left[\frac{\gamma}{\gamma_\beta} - \frac{1}{N} \frac{1 - \mathrm{e}^{-N\gamma_\beta\tau}}{\mathrm{e}^{\gamma_\beta\tau} - 1} \left(\left\langle \sigma_z \right\rangle_0 + \frac{\gamma}{\gamma_\beta} \right) \right] \cos 2\eta \quad (N \geqslant 1) \tag{89}$$
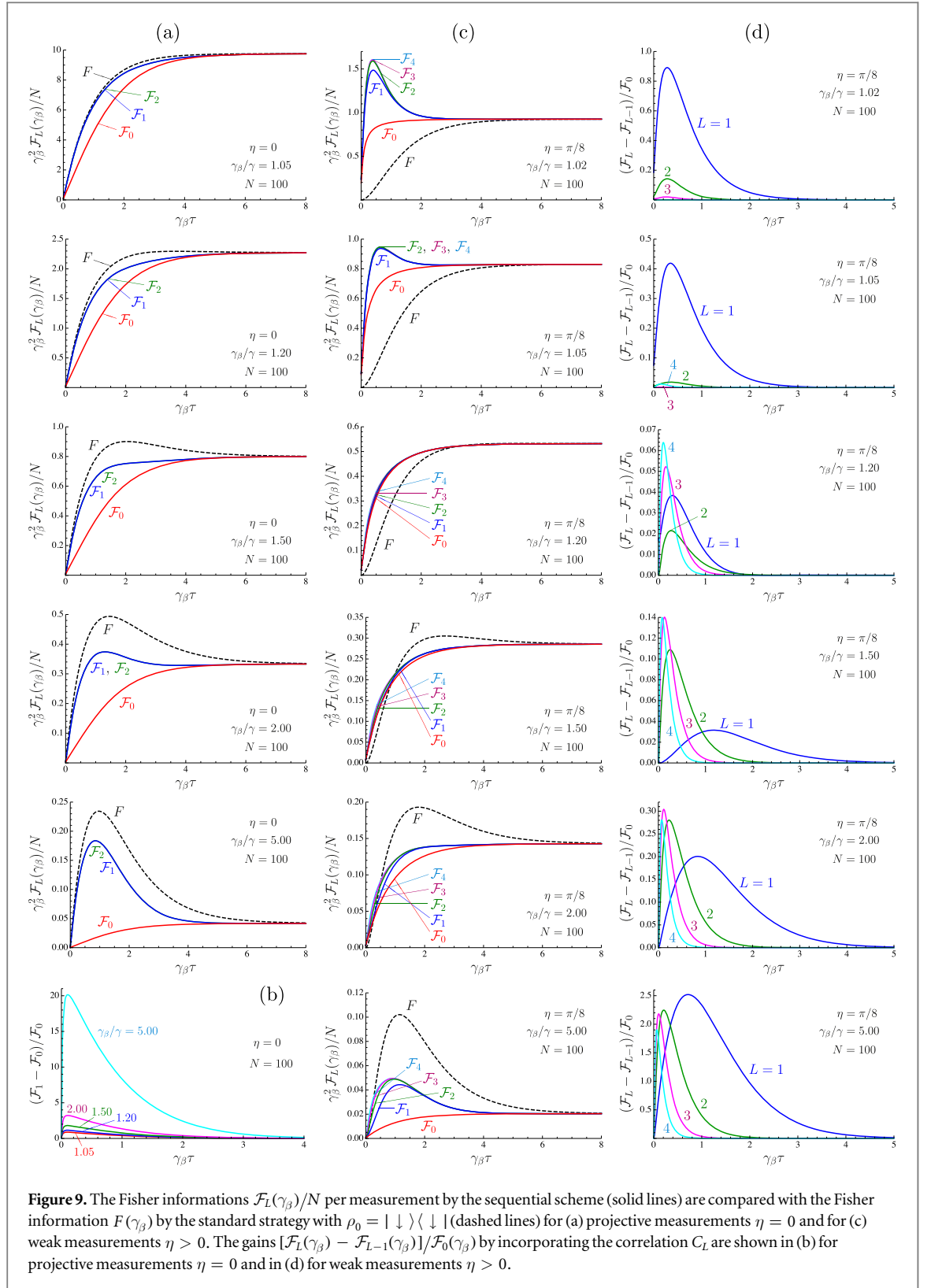
and the variance to be

$$
\begin{aligned}
(\Delta S)_N^2 &= \left\langle S^2 \right\rangle_N - \langle S \rangle_N^2 \\
&= \frac{1}{N} \left[\sin^2 2\eta + \left(\frac{1 + \mathrm{e}^{-\gamma_\beta\tau}}{1 - \mathrm{e}^{-\gamma_\beta\tau}} - \frac{2}{N} \frac{1 - \mathrm{e}^{-N\gamma_\beta\tau}}{\left(1 - \mathrm{e}^{-\gamma_\beta\tau}\right)^2} \mathrm{e}^{-\gamma_\beta\tau} \right) \left(1 - \frac{\gamma^2}{\gamma_\beta^2} \right) \cos^2 2\eta \right] \\
&\quad - \frac{1}{N} \left(\frac{\mathrm{e}^{-N\gamma_\beta\tau}}{1 - \mathrm{e}^{-N\gamma_\beta\tau}} - \frac{1}{2N} \frac{1 + \mathrm{e}^{-\gamma_\beta\tau}}{1 - \mathrm{e}^{-\gamma_\beta\tau}} \right) \frac{1 - \mathrm{e}^{-N\gamma_\beta\tau}}{\mathrm{e}^{\gamma_\beta\tau} - 1} \left(\left\langle \sigma_z \right\rangle_0 + \frac{\gamma}{\gamma_\beta} \right) \frac{\gamma}{\gamma_\beta} \cos^2 2\eta \\
&\quad - \frac{1}{N^2} \left[\frac{1 - \mathrm{e}^{-N\gamma_\beta\tau}}{\mathrm{e}^{\gamma_\beta\tau} - 1} \left(\left\langle \sigma_z \right\rangle_0 + \frac{\gamma}{\gamma_\beta} \right) \cos 2\eta \right]^2 \qquad (N \geqslant 1). \tag{90}
\end{aligned}
$$

As N increases, both become independent of the initial state ρ_0, and the variance $(\Delta S)_N^2$ shrinks as $1/N$,

$$\langle S \rangle_N \to -\frac{\gamma}{\gamma_\beta} \cos 2\eta, \tag{91}$$

$$(\Delta S)_N^2 \to \frac{1}{N} \left[\sin^2 2\eta + \frac{1 + \mathrm{e}^{-\gamma_\beta\tau}}{1 - \mathrm{e}^{-\gamma_\beta\tau}} \left(1 - \frac{\gamma^2}{\gamma_\beta^2} \right) \cos^2 2\eta \right]. \tag{92}$$

In other words, S evaluated from a single sequence of measurements almost certainly exhibits a value very close to its expectation value $\langle S \rangle_N$, which is a function of γ_β. Therefore, by comparing S (obtained via a single experimental run) with its expectation value $\langle S \rangle_N$ (given by the formula (91)), the parameter γ_β is estimated with the uncertainty regulated by the variance $(\Delta S)_N^2$ in (92), i.e., with the precision given by the Fisher information $\mathcal{F}(\gamma_\beta) = \mathcal{F}_0(\gamma_\beta)$ in (23),

Figure 9. The Fisher informations $\mathcal{F}_L(\gamma_\beta)/N$ per measurement by the sequential scheme (solid lines) are compared with the Fisher information $F(\gamma_\beta)$ by the standard strategy with $\rho_0 = |\downarrow\rangle\langle\downarrow|$ (dashed lines) for (a) projective measurements $\eta = 0$ and for (c) weak measurements $\eta > 0$. The gains $[\mathcal{F}_L(\gamma_\beta) - \mathcal{F}_{L-1}(\gamma_\beta)]/\mathcal{F}_0(\gamma_\beta)$ by incorporating the correlation C_L are shown in (b) for projective measurements $\eta = 0$ and in (d) for weak measurements $\eta > 0$.

$$\mathcal{F}_0(\gamma_\beta) \to \frac{N}{\gamma_\beta^2} \frac{\dfrac{\gamma^2}{\gamma_\beta^2}\cos^2 2\eta}{\sin^2 2\eta + \dfrac{1 + e^{-\gamma_\beta\tau}}{1 - e^{-\gamma_\beta\tau}}\left(1 - \dfrac{\gamma^2}{\gamma_\beta^2}\right)\cos^2 2\eta}, \tag{93}$$

which is to be compared with the Fisher information $NF(\gamma_\beta)$ with (83) by the standard strategy (see figure 9).

As stressed above, the correlations among the acquired data are also available for the estimation in the sequential scheme. For instance, the two-point correlations C_ℓ defined in (47) can be used to estimate γ_β. Their expectation values (for a generic initial state ρ_0) are given by

$$\left\langle C_\ell \right\rangle_N = \left[\frac{\gamma^2}{\gamma_\beta^2} + e^{-\ell\gamma_\beta\tau}\left(1 - \frac{\gamma^2}{\gamma_\beta^2}\right) - \frac{1}{N-\ell}\frac{1 - e^{-(N-\ell)\gamma_\beta\tau}}{e^{\gamma_\beta\tau} - 1}\left(1 - e^{-\ell\gamma_\beta\tau}\right)\left(\left\langle \sigma_z \right\rangle_0 + \frac{\gamma}{\gamma_\beta}\right)\frac{\gamma}{\gamma_\beta} \right] \cos^2 2\eta$$

$$(\ell \geqslant 0; \ N \geqslant \ell + 1), \tag{94}$$

and their covariances (in the stationary state $\rho_0 = \rho_*$) by

$$\left\langle C_\ell C_{\ell'} \right\rangle_N - \left\langle C_\ell \right\rangle_N \left\langle C_{\ell'} \right\rangle_N$$

$$= \delta_{\ell\ell'}\frac{1}{N}\sin^4 2\eta + \frac{2}{N}\left(\left\langle C_{\ell-\ell'}\right\rangle_* + \left\langle C_{\ell+\ell'}\right\rangle_*\right)\sin^2 2\eta$$

$$- \frac{2}{N}\left[\ell' e^{-(\ell+\ell')\gamma_\beta\tau} - \frac{1}{2}\left(\ell - \ell' + \frac{1 + e^{-2\gamma_\beta\tau}}{1 - e^{-2\gamma_\beta\tau}}\right)\left(e^{-(\ell-\ell')\gamma_\beta\tau} - e^{-(\ell+\ell')\gamma_\beta\tau}\right)\right]\left(1 - \frac{\gamma^2}{\gamma_\beta^2}\right)^2 \cos^4 2\eta$$

$$- \frac{4}{N}\left[\ell'\left(e^{-\ell\gamma_\beta\tau} + e^{-\ell'\gamma_\beta\tau}\right) - \frac{1 + e^{-\gamma_\beta\tau}}{1 - e^{-\gamma_\beta\tau}}\left(1 - e^{-\ell'\gamma_\beta\tau}\right)\right]\left(1 - \frac{\gamma^2}{\gamma_\beta^2}\right)\frac{\gamma^2}{\gamma_\beta^2}\cos^4 2\eta + O\left(1/N^2\right)$$

$$(\ell \geqslant \ell' \geqslant 1; \ N \gg \ell), \tag{95}$$

$$\left\langle SC_\ell \right\rangle_N - \left\langle S \right\rangle_N \left\langle C_\ell \right\rangle_N$$

$$= \frac{2}{N}\left\langle S \right\rangle_* \left\{ \sin^2 2\eta - \left[\ell e^{-\ell\gamma_\beta\tau} - \left(\frac{1 + e^{-\gamma_\beta\tau}}{1 - e^{-\gamma_\beta\tau}} - \frac{1}{N-\ell}\frac{1 - e^{-(N-\ell)\gamma_\beta\tau}}{\left(1 - e^{-\gamma_\beta\tau}\right)^2}e^{-\gamma_\beta\tau}\right)\left(1 - e^{-\ell\gamma_\beta\tau}\right)\right]\right.$$

$$\left. \times \left(1 - \frac{\gamma^2}{\gamma_\beta^2}\right)\cos^2 2\eta \right\} \qquad (\ell \geqslant 1; \ N \geqslant \ell + 1) \quad (96)$$

(the complete expression for the covariance $\langle C_\ell C_{\ell'}\rangle_N - \langle C_\ell\rangle_N \langle C_{\ell'}\rangle_N$ valid for any $\ell, \ell' \geqslant 1$ and $N \geqslant \max(\ell, \ell') + 1$ (but for $\rho_0 = \rho_*$) is given in appendix B). All the covariances scale as $1/N$, and the Fisher information (68) increases linearly in N. This ensures that, by comparing the set of quantities $(S, C_1, \ldots, C_L)$ evaluated from a single sequence of measurement data with the set of their expectation values $(\langle S\rangle_N, \langle C_1\rangle_N, \ldots, \langle C_L\rangle_N)$, one can estimate γ_β with the precision given by the Fisher information $\mathcal{F}_L(\gamma_\beta)$ computed by the formula (68), which increases linearly in N. It is reasonable to expect that the estimation with the multiple quantities $(S, C_1, \ldots, C_L)$ is better in precision than the estimation solely through the average S, namely, the Fisher information $\mathcal{F}_L(\gamma_\beta)$ ($L > 0$) is larger than the Fisher information $\mathcal{F}_0(\gamma_\beta)$, and the more correlations are incorporated (the larger is the number L), the larger is the Fisher information $\mathcal{F}_L(\gamma_\beta)$.

Let us look at two different regimes.

6.2.1. Projective measurement $\eta = 0$

The Fisher informations $\mathcal{F}_L(\gamma_\beta)$ ($L = 0, 1, 2$) by the sequential scheme are plotted in the five panels in figure 9(a) and are compared with the Fisher information $F(\gamma_\beta)$ by the standard strategy with $\rho_0 = |\downarrow\rangle\langle\downarrow|$, for the case of projective measurements $\eta = 0$. In this case, the Fisher information $F(\gamma_\beta)$ coincides with the quantum Fisher information $F_Q(\gamma_\beta)$ in the standard strategy.

Compare first $F(\gamma_\beta)$ and $\mathcal{F}_0(\gamma_\beta)/N$ (per measurement). We observe that the standard strategy provides better estimation than the sequential scheme. Recall here that the input state $\rho_0 = |\downarrow\rangle\langle\downarrow|$ for $F(\gamma_\beta)$ is the optimal for the standard strategy. On the other hand, in the sequential scheme, the state of the qubit is projected into $|\uparrow\rangle$ or $|\downarrow\rangle$ depending on the outcome of the projective measurement. If it is projected into $|\uparrow\rangle$ by a measurement, it restarts to evolve from this non-optimal state for the next measurement. Not all the steps in the sequential measurements are optimal for the estimation. That is why the sequential scheme cannot beat the standard strategy, in the case of projective measurement.

One can improve the performance of the sequential scheme, by incorporating C_1 for the estimation. Indeed, as is clear from figure 9(a), the Fisher information $\mathcal{F}_1(\gamma_\beta)$ for the estimation through (S, C_1) is greater than the Fisher information $\mathcal{F}_0(\gamma_\beta)$ solely through S. Note that no additional resources or experiments are required to incorporate C_1: one simply needs to carry out additional data analysis to compute C_1 from the data used to

evaluate S. In figure 9(b), the gain in the Fisher information by incorporating C_1 is shown for different temperatures.

On the other hand, incorporating more correlation data, i.e., C_ℓ with $\ell > 1$, does not help improve the estimation. See figure 9(a) again. This is because every time one performs measurement the system is reset to a pure state by the projective measurement: there is no correlation between the measurement results separated over two steps. The system simply repeats the same dynamics, jumping between pure states $|\uparrow\rangle$ and $|\downarrow\rangle$, and the measurement after multiple steps gains no more information than that attainable by the measurement after a single step.

6.2.2. Weak measurement $\eta > 0$

Let us next look at the cases with weak measurements $\eta > 0$. As is clear from figure 9(c), the sequential scheme can be better than the standard strategy. In particular, at low temperatures, the Fisher informations $\mathcal{F}_L(\gamma_\beta)/N$ ($L \geqslant 0$) by the sequential scheme exceed the Fisher information $F(\gamma_\beta)$ by the standard strategy.

The reason is the following. In the standard strategy, the weak measurement is performed only once, and the system is reset to the specific initial state ρ_0 for the next measurement. The single weak measurement can acquire less information than a projective measurement, but if it is repeated many times, as in the sequential scheme, the information is accumulated, and better information is gained in our hands. At the same time, the system is gradually projected to one of the eigenstates of the measured observable by the repeated weak measurements [43]. In other words, the repetition of the weak measurements mimics a stronger measurement (closer to a projective measurement). That is why the sequential scheme can be better than the standard strategy, in the case of weak measurement.

It is also clear from figure 9(c) that the precision of the estimation is improved by incorporating the correlation data C_ℓ. The gain in the Fisher information $[\mathcal{F}_L(\gamma_\beta) - \mathcal{F}_{L-1}(\gamma_\beta)]/\mathcal{F}_0$ by adding a correlation C_L to $(S, C_1, \ldots, C_{L-1})$ is shown in figure 9(d). The enhancement is reminiscent when the time interval between measurements τ is short, i.e., $\gamma_\beta\tau \lesssim 1$. Moreover, the gain exhibits a peak at a smaller τ for a larger L. This is because the two points of each two-point correlation C_ℓ, separated by ℓ steps, should be within the correlation time $\tau_c \sim 2/\gamma_\beta$ (which is ruled by the second largest eigenvalues $\mathrm{e}^{-(\gamma_\beta/2\pm\mathrm{i}\Omega)\tau}$ of the mixing channel (87)), in order for the correlation C_ℓ to bear useful information.

It appears that the sequential scheme can beat the standard strategy only at low temperatures (small γ_β), but it should be noted that the standard strategy in figure 9 assumes the optimal initial state $\rho_0 = |\downarrow\rangle\langle\downarrow|$, while in the sequential scheme the system is around the stationary state ρ_* of the mixing channel $\mathcal{E}$, which is the thermal equilibrium state ρ_{eq} (see (88)). It would be more appropriate to compare the Fisher informations $\mathcal{F}_L(\gamma_\beta)/N$ by the sequential scheme with the Fisher information $F(\gamma_\beta)$ by the standard strategy in the large τ limit (which gives the Fisher information with the initial thermal state $\rho_0 = \rho_{\mathrm{eq}}$).

7. Conclusions

The estimation of a parameter encoded in a quantum probe, through a series of measurements performed sequentially on the probe, has been analyzed in a general non-i.i.d. setting. On the basis of a diagrammatic approach we have discussed the conditions under which the central limit theorem holds as the number of measurements increases, reproducing the previous results [19] and generalizing them to the case where the correlations among the measurement data are also taken into account in the estimation strategy. Our analysis explicitly shows that the latter strategy can yield a significant advantage over the standard procedure where only the average of the acquired data is considered.

At present however it is not clear whether this is the best strategy one can do: it is indeed possible that different data processing (including the evaluation of higher-order correlations commented at the end of section 5) can improve further the attainable accuracy. In the example studied in section 6, the sequential scheme surpassed the standard i.i.d. procedure when we are able to perform only weak measurements, but could not beat the standard procedure when we are allowed to perform strong measurements. A better strategy for the sequential scheme could beat the ultimate precision achievable by the standard strategy. The optimal strategy would require different measurements step by step, or moreover would require quantum-correlated measurements over different measurement probings. The use of entanglement is also an interesting possibility [44]. It is yet to be clarified what is the ultimate accuracy attainable in the sequential scheme for parameter estimation[6].

---

[6] In [19] a concise formula for the quantum Fisher information providing the maximum Fisher information attainable by the optimal measurement strategy applied on the target system plus probing systems in the sequential scheme is derived. It is however valid only when each probing process is unitary. Moreover, it is not applicable if one can measure the target system only through the probes.

Recently, quantum metrology in the presence of noise is under intense study [44–47]. The mixing property required for the sequential scheme is relevant to noisy channels, and connections with such issue would be interesting to be explored.

Acknowledgments

Appendix A. Complete expressions for the covariances

In section 5, we derived the asymptotic expression (54) for the even moments among S and C_ℓ. Here we provide the complete expressions for the covariances among S and C_ℓ valid for any (even small) N. Under the assumption that the quantum channel $\mathcal{E}$ is ergodic (not necessarily mixing), they read

$$
\begin{aligned}
\left\langle \left(S - \langle S \rangle_* \right)^2 \right\rangle_N &= \frac{1}{N}(\Delta s)_*^2 + \frac{2}{N}\left(1 \,\middle|\, \tilde{\mathcal{E}}^{(1)} \frac{\mathcal{Q}_*}{1-\mathcal{E}'} \tilde{\mathcal{E}}^{(1)} \,\middle|\, \rho_* \right) - \frac{2}{N^2}\left(1 \,\middle|\, \tilde{\mathcal{E}}^{(1)} \frac{1-\mathcal{E}'^N}{\left(1-\mathcal{E}'\right)^2} \mathcal{Q}_* \tilde{\mathcal{E}}^{(1)} \,\middle|\, \rho_* \right) \\
&\quad - \frac{2}{N^2}\sum_{j=1}^{N-1}\left(1 \,\middle|\, \tilde{\mathcal{E}}^{(1)} \frac{\mathcal{E}'^{N-j}}{1-\mathcal{E}'} \mathcal{Q}_* \tilde{\mathcal{E}}^{(1)} \mathcal{E}'^{j-1} \mathcal{Q}_* \,\middle|\, \rho_0 \right) + \frac{1}{N^2}\left(1 \,\middle|\, \tilde{\mathcal{E}}^{(2)} \frac{1-\mathcal{E}'^N}{1-\mathcal{E}'} \mathcal{Q}_* \,\middle|\, \rho_0 \right) \\
&\quad + \frac{2}{N^2}\left(1 \,\middle|\, \tilde{\mathcal{E}}^{(1)} \frac{\mathcal{Q}_*}{1-\mathcal{E}'} \tilde{\mathcal{E}}^{(1)} \frac{1-\mathcal{E}'^{N-1}}{1-\mathcal{E}'} \mathcal{Q}_* \,\middle|\, \rho_0 \right) \qquad (N \geqslant 1),
\end{aligned} \tag{A.1}
$$

$$
\begin{aligned}
&\left\langle \left(C_\ell - \langle C_\ell \rangle_* \right)\left(C_{\ell'} - \langle C_{\ell'} \rangle_* \right) \right\rangle_N \\
&= \frac{\theta\left(N - \ell - \ell'\right)}{\left(N-\ell\right)\left(N-\ell'\right)}\Bigg(\Bigg(1 \,\middle|\, \tilde{\mathcal{E}}_\ell^{(1)}\Bigg(\frac{N-\ell-\ell'}{1-\mathcal{E}'} - \frac{1-\mathcal{E}'^{N-\ell-\ell'}}{\left(1-\mathcal{E}'\right)^2} \Bigg)\mathcal{Q}_* \tilde{\mathcal{E}}_{\ell'}^{(1)} \,\middle|\, \rho_* \Bigg) + \left(\ell \leftrightarrow \ell' \right) \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad + \left(N - \ell - \ell' \right)\left(1 \,\middle|\, \tilde{\mathcal{E}}_{\ell\ell'}^{\circ\star\bullet} \,\middle|\, \rho_* \right)\Bigg) \\
&\quad + \frac{1}{\left(N-\ell\right)\left(N-\ell'\right)}\sum_{k=\max\left(1+\ell+\ell'-N,1\right)}^{\min\left(\ell,\ell'\right)-1}\left(N - \ell - \ell' + k \right)\left(1 \,\middle|\, \tilde{\mathcal{E}}_{\ell\ell',k}^{\circ\bullet\circ\bullet} \,\middle|\, \rho_* \right) \\
&\quad + \frac{1}{N-\min\left(\ell,\ell'\right)}\left(1 \,\middle|\, \delta_{\ell\ell'}\tilde{\mathcal{E}}_\ell^{\star\star} + \left(1-\delta_{\ell\ell'}\right)\left(\tilde{\mathcal{E}}_{\ell\ell'}^{\star\circ\bullet} + \tilde{\mathcal{E}}_{\ell\ell'}^{\bullet\circ\star} + \tilde{\mathcal{E}}_{\ell\ell'}^{\bullet\circ\circ\bullet} \right) \,\middle|\, \rho_* \right) \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad \left(\rho_0 = \rho_*; \ \ell, \ell' \geqslant 1; \ N \geqslant \max\left(\ell, \ell'\right) + 1 \right),
\end{aligned} \tag{A.2}
$$

and

$$
\begin{aligned}
&\left\langle \left(S - \langle S \rangle_* \right)\left(C_\ell - \langle C_\ell \rangle_* \right) \right\rangle_N \\
&= \frac{1}{N}\left(1 \,\middle|\, \tilde{\mathcal{E}}_\ell^{(2)} \,\middle|\, \rho_* \right) + \frac{1}{N\left(N-\ell\right)}\left(1 \,\middle|\, \tilde{\mathcal{E}}^{(1)}\Bigg(\frac{N-\ell}{1-\mathcal{E}'} - \frac{1-\mathcal{E}'^{N-\ell}}{\left(1-\mathcal{E}'\right)^2} \Bigg)\mathcal{Q}_* \tilde{\mathcal{E}}_\ell^{(1)} \,\middle|\, \rho_* \right) \\
&\quad + \frac{1}{N\left(N-\ell\right)}\left(1 \,\middle|\, \tilde{\mathcal{E}}_\ell^{(1)}\Bigg(\frac{N-\ell}{1-\mathcal{E}'} - \frac{1-\mathcal{E}'^{N-\ell}}{\left(1-\mathcal{E}'\right)^2} \Bigg)\mathcal{Q}_* \tilde{\mathcal{E}}^{(1)} \,\middle|\, \rho_* \right) \\
&\qquad\qquad\qquad\qquad\qquad\qquad \left(\rho_0 = \rho_*; \ \ell \geqslant 1; \ N \geqslant \ell + 1 \right),
\end{aligned} \tag{A.3}
$$

where $\tilde{\mathcal{E}}^{(m)}$ and $(\Delta s)_*^2$ are defined in (30) and (31), respectively, $\tilde{\mathcal{E}}_\ell^{(m)}$ are in (53) and (58), and the other components are given in (60)–(65).

Appendix B. Covariances among C_ℓ for the model

In (95) in section 6 we showed the asymptotic expression for the covariance between C_ℓ and $C_{\ell'}$ for large N for the model. Here we provide its complete expression valid for any (even small) N. In the stationary state $\rho_0 = \rho_*$, the covariances between C_ℓ and $C_{\ell'}$ ($\ell \geq \ell' \geq 1$) are given for $N \geq \ell + \ell'$ by

$$
\begin{aligned}
&\left\langle C_\ell C_{\ell'} \right\rangle_N - \left\langle C_\ell \right\rangle_N \left\langle C_{\ell'} \right\rangle_N \\
&= \delta_{\ell\ell'} \frac{1}{N-\ell'} \sin^4 2\eta + \frac{2}{N-\ell'} \left[\left\langle C_{\ell-\ell'} \right\rangle_* + \left(1 - \frac{\ell'}{N-\ell}\right)\left\langle C_{\ell+\ell'} \right\rangle_* \right] \sin^2 2\eta \\
&\quad - \frac{1}{N-\ell'} \Bigg\{ \ell'\left(2 - \frac{\ell'}{N-\ell}\right)e^{-(\ell+\ell')\gamma_\beta\tau} - \frac{1 + e^{-2\gamma_\beta\tau}}{1 - e^{-2\gamma_\beta\tau}}\left[e^{-(\ell-\ell')\gamma_\beta\tau} - \left(1 - \frac{\ell'}{N-\ell}\right)e^{-(\ell+\ell')\gamma_\beta\tau}\right] \\
&\quad - \left(\ell - \ell' - \frac{2}{N-\ell}\frac{e^{-2\gamma_\beta\tau}}{\left(1 - e^{-2\gamma_\beta\tau}\right)^2}\right)\left(e^{-(\ell-\ell')\gamma_\beta\tau} - e^{-(\ell+\ell')\gamma_\beta\tau}\right) \Bigg\}\left(1 - \frac{\gamma^2}{\gamma_\beta^2}\right)^2 \cos^4 2\eta \\
&\quad - \frac{2}{N-\ell'} \Bigg\{ \ell'\left(2 - \frac{\ell'}{N-\ell}\right)\left(e^{-\ell\gamma_\beta\tau} + e^{-\ell'\gamma_\beta\tau}\right) \\
&\quad - \frac{1 + e^{-\gamma_\beta\tau}}{1 - e^{-\gamma_\beta\tau}}\left[\left(2 - \frac{\ell'}{N-\ell}\right)\left(1 - e^{-\ell'\gamma_\beta\tau}\right) + \frac{\ell'}{N-\ell}e^{-\ell\gamma_\beta\tau}\right] \\
&\quad + \frac{1}{N-\ell}\frac{e^{-\gamma_\beta\tau}}{\left(1 - e^{-\gamma_\beta\tau}\right)^2}\left[\left(1 - e^{-(N-\ell-\ell')\gamma_\beta\tau}\right)\left(1 - e^{-\ell\gamma_\beta\tau}\right)\left(1 - e^{-\ell'\gamma_\beta\tau}\right)\right. \\
&\quad \left. + e^{-(\ell-\ell')\gamma_\beta\tau} - e^{-(\ell+\ell')\gamma_\beta\tau}\right] \Bigg\}\left(1 - \frac{\gamma^2}{\gamma_\beta^2}\right)\frac{\gamma^2}{\gamma_\beta^2} \cos^4 2\eta,
\end{aligned}
$$

$$(\text{B.1})$$

while for $\ell + \ell' > N \geq \ell + 1$ by

$$
\begin{aligned}
&\left\langle C_\ell C_{\ell'} \right\rangle_N - \left\langle C_\ell \right\rangle_N \left\langle C_{\ell'} \right\rangle_N \\
&= \delta_{\ell\ell'} \frac{1}{N-\ell'} \sin^4 2\eta + \frac{2}{N-\ell'} \left\langle C_{\ell-\ell'} \right\rangle_* \sin^2 2\eta \\
&\quad - \Bigg[e^{-(\ell+\ell')\gamma_\beta\tau} - \frac{1}{N-\ell'}\Bigg(\ell - \ell' + \frac{1 + e^{-2\gamma_\beta\tau}}{1 - e^{-2\gamma_\beta\tau}} \\
&\quad - \frac{2}{N-\ell}\frac{1 - e^{-2(N-\ell)\gamma_\beta\tau}}{\left(1 - e^{-2\gamma_\beta\tau}\right)^2}e^{-2\gamma_\beta\tau}\Bigg)e^{-(\ell-\ell')\gamma_\beta\tau} \Bigg]\left(1 - \frac{\gamma^2}{\gamma_\beta^2}\right)^2 \cos^4 2\eta \\
&\quad - 2\Bigg[\left(1 - \frac{\ell-\ell'}{N-\ell'}\right)\left(e^{-\ell\gamma_\beta\tau} + e^{-\ell'\gamma_\beta\tau}\right) - \frac{1}{N-\ell'}\frac{1 + e^{-\gamma_\beta\tau}}{1 - e^{-\gamma_\beta\tau}}\left(1 + e^{-\ell\gamma_\beta\tau} - e^{-\ell'\gamma_\beta\tau}\right) \\
&\quad + \frac{1}{(N-\ell)(N-\ell')}\frac{1 - e^{-(N-\ell)\gamma_\beta\tau}}{\left(1 - e^{-\gamma_\beta\tau}\right)^2}e^{-\gamma_\beta\tau}\left(1 + e^{-(\ell-\ell')\gamma_\beta\tau} - e^{(N-\ell-\ell')\gamma_\beta\tau} + e^{-\ell\gamma_\beta\tau}\right)\Bigg] \\
&\quad \times \left(1 - \frac{\gamma^2}{\gamma_\beta^2}\right)\frac{\gamma^2}{\gamma_\beta^2} \cos^4 2\eta.
\end{aligned}
$$

$$(\text{B.2})$$

References

[1] Bennett C H and Shor P W 1998 *IEEE Trans. Inf. Theory* **44** 2724
[2] Giovannetti V, Lloyd S and Maccone L 2006 *Phys. Rev. Lett.* **96** 010401

[3] Giovannetti V, Lloyd S and Maccone L 2011 *Nat. Photonics* **5** 222
[4] Singer L P *et al* 2014 *Astrophys. J.* **795** 105
[5] Auzinsh M, Budker D, Kimball D F, Rochester S M, Stalnaker J E, Sushkov A O and Yashchuk V V 2004 *Phys. Rev. Lett.* **93** 173002
[6] Wolfgramm F, Vitelli C, Beduini F A, Godbout N and Mitchell M W 2013 *Nat. Photonics* **7** 28
[7] Helstrom C W 1976 *Quantum Detection and Estimation Theory* (New York: Academic)
[8] Braunstein S L and Caves C M 1994 *Phys. Rev. Lett.* **72** 3439
[9] Braunstein S L, Caves C M and Milburn G J 1996 *Ann. Phys., NY* **247** 135
[10] Paris M G A and Řeháček J (ed) 2004 *Quantum State Estimation* (Berlin: Springer)
[11] Paris M G A 2009 *Int. J. Quantum Inf.* **7** 125
[12] Holevo A S 2011 *Probabilistic and Statistical Aspects of Quantum Theory* (Pisa: Edizioni della Normale)
[13] Terhal B M and DiVincenzo D P 2000 *Phys. Rev.* A **61** 022301
[14] Wolf M M *Quantum Channels & Operations: Guided Tour* (www-m5.ma.tum.de/foswiki/pub/M5/Allgemeines/MichaelWolf/QChannelLecture.pdf)
[15] Burgarth D and Giovannetti V 2007 *New J. Phys.* **9** 150
[16] Burgarth D, Chiribella G, Giovannetti V, Perinotti P and Yuasa K 2013 *New J. Phys.* **15** 073045
[17] Mabuchi H 1996 *Quantum Semiclass. Opt.* **8** 1103
[18] Gambetta J and Wiseman H M 2001 *Phys. Rev.* A **64** 042105
[19] Guţă M 2011 *Phys. Rev.* A **83** 062324
[20] Cătană C, van Horssen M and Guţă M 2012 *Phil. Trans. R. Soc.* A **370** 5308
[21] Gammelmark S and Mølmer K 2013 *Phys. Rev.* A **87** 032115
[22] Gammelmark S and Mølmer K 2014 *Phys. Rev. Lett.* **112** 170401
[23] Rybár T and Ziman M 2015 *Phys. Rev.* A **92** 042315
[24] van Horssen M and Guţă M 2015 *J. Math. Phys.* **56** 022109
[25] Kiilerich A H and Mølmer K 2015 *Phys. Rev.* A **92** 032124
[26] Höpfner R, Jacod J and Ladelli L 1990 *Probab. Theory Relat. Fields* **86** 105
[27] DasGupta A 2008 *Asymptotic Theory of Statistics and Probability* (New York: Springer) ch 10
[28] Billingsley P 2012 *Probability and Measure* anniversary edn (New York: Wiley) theorem 27.4
[29] Nielsen M A and Chuang I L 2000 *Quantum Computation and Quantum Information* (Cambridge: Cambridge University Press)
[30] Alicki R and Lendi K 2007 *Quantum Dynamical Semigroups and Applications* 2nd edn (Berlin: Springer)
[31] Holevo A S and Giovannetti V 2012 *Rep. Prog. Phys.* **75** 046001
[32] Bengtsson I and Życzkowski K 2006 *Geometry of Quantum States: An Introduction to Quantum Entanglement* (Cambridge: Cambridge University Press)
[33] Cramér H 1946 *Mathematical Methods of Statistics* (Princeton, NJ: Princeton University Press)
[34] Correa L A, Mehboudi M, Adesso G and Sanpera A 2015 *Phys. Rev. Lett.* **114** 220405
[35] De Pasquale A, Rossini D, Fazio R and Giovannetti V 2015 arXiv:1504.07787
[36] Scully M O and Zubairy M S 1997 *Quantum Optics* (Cambridge: Cambridge University Press)
[37] Breuer H P and Petruccione F 2002 *The Theory of Open Quantum Systems* (Oxford: Oxford University Press)
[38] Gardiner C W and Zoller P 2004 *Quantum Noise* 3rd edn (Berlin: Springer)
[39] Walls D F and Milburn G J 2008 *Quantum Optics* 2nd edn (Berlin: Springer)
[40] Lund A P and Wiseman H M 2010 *New J. Phys.* **12** 093011
[41] Rozema L A, Darabi A, Mahler D H, Hayat A, Soudagar Y and Steinberg A M 2012 *Phys. Rev. Lett.* **109** 100404
[42] Fujiwara A 2001 *Phys. Rev.* A **63** 042304
[43] Guerlin C, Bernu J, Deléglise S, Sayrin C, Gleyzes S, Kuhr S, Brune M, Raimond J M and Haroche S 2007 *Nature* **448** 889
[44] Demkowicz-Dobrzański R and Maccone L 2014 *Phys. Rev. Lett.* **113** 250801
[45] Escher B M, de Matos Filho R L and Davidovich L 2011 *Nat. Phys.* **7** 406
[46] Demkowicz-Dobrzański R, Kołodyński J and Guţă M 2012 *Nat. Commun.* **3** 1063
[47] Jarzyna M and Demkowicz-Dobrzański R 2015 *New J. Phys.* **17** 013010