I. INTRODUCTION

Quantum metrology is one of the pillars of quantum science and technology [1–4]. This field deals with fundamental precision limits of parameter estimation imposed by quantum physics. Notably, it seeks to use non-classical effects to enhance the estimation precision of unknown parameters in quantum systems, which has led to the development of improved sensing protocols in various experimental platforms [5–9]. The quantum Fisher information (QFI) characterizes the amount of information a quantum state carries about an unknown parameter [10, 11]. It sets an asymptotically saturable precision bound for parameter estimation in quantum experiments, called the quantum Cramér–Rao bound (QCRB) [12, 13]. Therefore, the QFI is one of the most useful and celebrated tools in quantum metrology, where a considerable amount of the research in the field deals with different ways to calculate and bound it [14–19].

Although the QFI applies extensively in quantum sensing and is a well-behaved mathematical quantity [20], it is defined assuming arbitrary quantum measurements can be applied on quantum states to extract information about the unknown parameter. In reality, however, measurement on quantum states in actual experimental platforms, e.g., nitrogen-vacancy centers [21–24], superconducting qubits [25], trapped ions [26], etc., are often noisy, time-expensive, and cannot attain the QFI, rendering the sensitivity of practical quantum devices far away from the theoretical limits given by the QCRB. In particular, measurement noise is a significant source of noise in quantum sensing experiments. Other sources of noise include system evolution noise and state preparation noise, and a plethora of methods to reduce them were studied in previous works [27–36].

To tackle the effect of measurement noise on quantum metrology, interaction-based readouts were proposed [37–41] and demonstrated experimentally [42, 43], where specially designed interparticle interactions that enhance phase estimation precision in spin ensembles are applied before the noisy measurement step and after the probing step. The idea of employing unitary controls in the preprocessing manner was later formulated as the imperfect (or noisy) QFI problem [41, 44], where the (classical) Fisher information (FI) of the noisy measurement statistics is optimized over all possible unitary controls, setting a precision bound for metrology under noisy measurements. Note that, classical postprocessing methods, e.g., measurement error mitigation [45–47] work in complement to the quantum preprocessing method for parameter estimation under noisy measurements.

Besides a few specific cases, e.g., qubit sensors with lossy photon detection [44], the imperfect QFI remains an obscure concept with no easy way to compute and its practical applications are limited. In this work, we propose a more general measurement optimization scheme where arbitrary quantum controls (i.e., general quantum channels that can be implemented utilizing unitary gates and ancillas), are applied before the noisy measurement. The goal is to identify, for general quantum states and measurements, the FI optimized over all quantum channels before a noisy measurement, which we call the quantum preprocessing-optimized FI (QPIF), and the corresponding optimal preprocessing controls.
We systematically study the preprocessing optimization problem in this work. In Sec. II, we first define the QPFI and review related concepts. We then introduce the error observable formulation in Sec. III, based on which we show that the QPFI problem can be cast as a numerically tractable biconvex optimization problem [48]. Furthermore, this allows us to find analytical conditions for optimality, and to identify optimal controls to be applied before commuting-operator measurements (i.e., measurement of which each measurement operator commutes with each other) that saturate the QPFI for pure states (see Sec. IV) and classically mixed states (i.e., states of which the unknown parameter is encoded in the eigenvalues) (see Sec. V). We also show unitary controls are optimal for pure states under general measurements and coarse graining controls are optimal for classically mixed states under commuting-operator measurements, with the latter implying the non-optimality of unitary controls in general. For general mixed states, we further prove useful bounds on the QPFI in Sec. VI. In terms of the asymptotic behavior of identical local measurements acting on multi-probe systems, in Sec. VII, we identify a sufficient condition for the convergence of the QPFI to the QFI using an optimal encoding protocol based on the Holevo–Schumacher–Westmoreland (HSW) theorem [49, 50]. We show that the sufficient condition is satisfied by a generic class of quantum states including low-rank states, permutation-invariant states and Gibbs states (with an unknown temperature); while previously only the pure state case was proven [44]. Our results provide a theoretically accessible precision bound for quantum metrology under noisy measurements and a roadmap towards preprocessing optimization in sensing experiments.

II. DEFINITIONS

Given a quantum state \( \rho_0 \) as a function of an unknown parameter \( \theta \), the procedure to estimate \( \theta \) goes as follows (see Fig. 1a): (1) Perform a quantum measurement \( \{ M_i \} \) on \( \rho_0 \) which gives a measurement outcome \( i \) with probability \( p_{i,\theta} = \text{Tr}(\rho_0 M_i) \); (2) Infer the value of \( \theta \) using an estimator \( \hat{\theta} \) which is a function of \( i \); (3) Repeat the above two steps multiple times and use the average of \( \hat{\theta}_i \) as the final estimate of \( \theta \). Here the quantum measurement \( \{ M_i \} \) is mathematically formulated as a positive operator-valued measure [51] that satisfies \( M_i \geq 0 \) and \( \sum_i M_i = 1 \) (we use \( A \geq 0 \) to indicate a matrix \( A \) is positive semidefinite). We also assume in this work that \( \rho_0 \) lies in a finite-dimensional Hilbert state and the measurement outcomes are contained in a finite set.

In estimation theory, the Cramér–Rao bound (CRB) [52, 53] provides a lower bound on the estimation error for any locally unbiased estimator \( \hat{\theta} \) at a local point \( \theta \), satisfying

\[
\mathbb{E}[\hat{\theta}] = \theta, \quad \text{and} \quad \frac{\partial}{\partial \theta} \mathbb{E}[\hat{\theta}] = 1,
\]

where we use \( \mathbb{E}[\cdot] \) to denote the expectation value. The above condition indicates that locally unbiased estimators \( \hat{\theta} \) provide an unbiased estimation of \( \theta \) at the point \( \theta \), and is also precise up to the first order in its neighborhood. The CRB states that the estimation error \( \Delta \hat{\theta} \) (i.e., the standard deviation of the estimator \( \hat{\theta} \)) has the following lower bound:

\[
\Delta \hat{\theta} := (\mathbb{E}[(\hat{\theta} - \theta)^2])^{\frac{1}{2}} \geq \frac{1}{\sqrt{N_{\text{expr}} F(\rho_0, \{ M_i \})}},
\]

where \( N_{\text{expr}} \) is the number of experiments and \( F(\rho_0, \{ M_i \}) \) is the FI of the probability distribution \( \{ p_{i,\theta} = \text{Tr}(\rho_0 M_i) \} \) [52, 53], defined by

\[
F(\rho_0, \{ M_i \}) := \sum_{i: \text{Tr}(\rho_0 M_i) \neq 0} \frac{(\text{Tr}(\rho_0 M_i))^2}{\text{Tr}(\rho_0 M_i)}. \tag{3}
\]

The CRB is saturable asymptotically (i.e., when \( N_{\text{expr}} \to \infty \)) using the maximum likelihood estimator [52, 53] and therefore the FI, which is inversely proportional to the variance of the estimator, serves as a good measure of the degree of sensitivity of \( \{ p_{i,\theta} \} \) with respect to \( \theta \).

The QFI of \( \rho_0 \) is the FI maximized over all possible quantum measurements on \( \rho_0 \) (see Appx. A for further details) and we will refer to the optimal measurements as QFI-attainable measurements. Formally, the QFI is defined by [10, 12, 13]

\[
J(\rho_0) = \max_{\{ M_i \}} F(\rho_0, \{ M_i \}),
\]

giving rise to the QCRB

\[
\Delta \hat{\theta} \geq \frac{1}{\sqrt{N_{\text{expr}} J(\rho_0)}}, \tag{5}
\]

which characterizes the ultimate lower bound on the estimation error. We will also overload the notation and use

\[
J(\{ p_{i,\theta} \}) = \sum_{i: p_{i,\theta} \neq 0} \frac{(\partial p_{i,\theta})^2}{p_{i,\theta}}, \tag{6}
\]

to denote the FI of a classical probability distribution \( \{ p_{i,\theta} \} \) in this work. Note that we will implicitly assume that the summation is taken over terms with non-zero denominators in discussions later.

In practice, the optimal measurements achieving the QFI are not always implementable, restricting the range of applications of the QCRB. For example, the projective measurement onto the basis of the symmetric logarithmic operators, which is usually a correlated measurement among multiple probes, is known to be optimal [10], while quantum measurements in experiments are usually noisy
FIG. 1. (a) Standard parameter estimation procedure of a quantum state $\rho_\theta$ using a quantum measurement $\{M_i\}$. The estimation of $\theta$ is through an unbiased estimator $\hat{\theta}$ as a function of measurement outcomes $i$. The CRB states $\Delta\hat{\theta} \geq 1/\sqrt{N} \mathcal{F}(\rho_\theta, \{M_i\})$. (b) Preprocessing protocol where the measurement device is fixed, and the quantum control acting before the measurement is optimized over all quantum channels. The CRB states $\Delta\hat{\theta} \geq 1/\sqrt{N} \mathcal{F}(\rho_\theta, \{M_i\})$. (c) Preprocessing protocol where the measurement device is fixed, and the quantum control acting before the measurement is optimized over all unitary channels. The CRB states $\Delta\hat{\theta} \geq 1/\sqrt{N} \mathcal{F}(\rho_\theta, \{M_i\})$. Different types of FIs discussed in this work satisfy $F(\rho_\theta, \{M_i\}) \leq F^P(\rho_\theta, \{M_i\}) \leq F^P(\rho_\theta, \{M_i\}) \leq J(\rho_\theta)$ (and each inequality can be strict).

and not exactly projective. Here we consider a metrological protocol where arbitrary quantum controls can be implemented before a quantum measurement (see Fig. 1b) and we will call this “preprocessing” (“pre-measurement-processing” in full) quantum controls on the quantum states. This model effectively describes quantum experiments where the measurement error is dominant, while the gate implementation error and the state preparation error is small, a noise model that arises naturally in modern quantum devices such as nitrogen-vacancy centers [21–24] and superconducting qubits [25].

To quantify the sensitivity of estimating $\theta$ on $\rho_\theta$ with the measurement $\{M_i\}$ fixed, we define the FI optimized over all preprocessing quantum channels or the quantum preprocessing-optimized Fisher information (QPFI) to be

$$F^P(\rho_\theta, \{M_i\}) = \sup_\mathcal{E} F(\mathcal{E}(\rho_\theta), \{M_i\}),$$

where $\mathcal{E}$ is an arbitrary quantum channel (or a CPTP map [54]). See Appx. B for mathematical properties of the QPFI. In particular, when the quantum measurement is fixed, the CRB induced by the QPFI, i.e.,

$$\Delta\hat{\theta} \geq \frac{1}{\sqrt{N} \mathcal{F}(\rho_\theta, \{M_i\})}$$

provides a practical and tighter Cramér–Rao-type bound, compared to the QCRB, for parameter estimation under noisy measurements. We assume in the following discussions that all measurements are non-trivial (i.e., $\exists M_i \neq \mathbb{1}$ for all $\{M_i\}$) and $\partial_\theta \rho_\theta \neq 0$ so that the QPFI is always positive.

Unless stated otherwise, we will denote the systems $\rho_\theta$ and $\{M_i\}$ act on by $\mathcal{H}_S$ and $\mathcal{H}_{S'}$, respectively, and we will refer to $\mathcal{H}_S$ as the input system and $\mathcal{H}_{S'}$ as the output system. We do not assume $\mathcal{H}_S \cong \mathcal{H}_{S'}$ here. This assumption is of particular interest when the quantum state $\rho_\theta$ cannot be directly measured (e.g., the readout of superconducting qubits is performed on a resonator coupled to it [25]); or when the quantum state is restricted to a subsystem of the entire system while quantum measurement can be performed globally.

Note that for generic noisy measurements, the supremum in Eq. (7) is usually attainable, i.e., there exists an optimal $\mathcal{E}$ such that $F(\mathcal{E}(\rho_\theta), \{M_i\})$ is maximized (see Appx. C). However, there exists singular cases where $F(\mathcal{E}(\rho_\theta), \{M_i\})$ has no maximum, due to the singularity of the FI at the point $\text{Tr}(\mathcal{E}(\rho_\theta)M_i) = 0$ (see Sec. IV C for an example). In such cases, there still exist near-optimal quantum controls that attain $\sup_\mathcal{E} F(\mathcal{E}(\rho_\theta), \{M_i\}) - \eta$ for any small $\eta > 0$. In fact, we prove in Appx. C that

**Theorem 1.** Let $M_i^{(\epsilon)} = (1 - \epsilon)M_i + \epsilon \text{Tr}(M_i)\mathbb{1}$ where $d = \dim(\mathcal{H}_{S'})$ and $0 < \epsilon < 1$. Then

$$F^P(\rho_\theta, \{M_i\}) = \lim_{\epsilon \to 0^+} F^P(\rho_\theta, \{M_i^{(\epsilon)}\}),$$

and the QPFI $F^P(\rho_\theta, \{M_i^{(\epsilon)}\})$ is attainable for any $\epsilon \in (0, 1]$.

In the following, we will focus mostly on the case where the QPFI is attainable. We will discuss the behaviors of the QPFI, exploring numerical optimization algorithms and analytical solutions to the optimal controls for some practically relevant quantum states and measurements.

We will also examine the FI optimized over all unitary preprocessing channels or the quantum unitary-preprocessing-optimized Fisher information (QUPFI) [41, 44]

$$F^U(\rho_\theta, \{M_i\}) = \sup_{U} F(U \rho_\theta U^\dagger, \{M_i\}),$$

where $U$ is an arbitrary unitary gate. (Note that the QUPFI was called the imperfect QFI in [44].) Unlike the QPFI, we assume $\mathcal{H}_{S'} \cong \mathcal{H}_S$ (and do not distinguish between $S'$ and $S$) when we talk about the QUPFI so that it is well defined. Note that Theorem 1 also holds for the QUPFI.

The optimal preprocessing controls that attain the QPFI and the QUPFI usually depend on $\theta$, whose value...
should be roughly known before the experiment. Otherwise, one might use the two-step method by first using $\sqrt{N_{\text{expr}}}$ states to obtain a rough estimation $\hat{\theta} \approx \theta$ of $\theta$ and then performing the optimal controls based on $\hat{\theta}$ on the remaining $N_{\text{expr}} - \sqrt{N_{\text{expr}}}$ states [55–57]. The two-step procedure introduces a negligible amount of error asymptotically.

Before we proceed, we prove a relation between and the QPFI and the QUPFI that will be useful later.

**Proposition 2.** Let $\mathcal{H}_S$ and $\mathcal{H}_{S'}$ be the input and output systems of $\mathcal{E}(\cdot)$. Suppose $\mathcal{H}_{A_1}$ and $\mathcal{H}_{A_2}$ are ancillary systems such that $\mathcal{H}_{A_1} \otimes \mathcal{H}_S \cong \mathcal{H}_{A_2} \otimes \mathcal{H}_{S'}$ and $\dim(\mathcal{H}_{A_1}) \geq \dim(\mathcal{H}_{A_2})^2$. Then

$$F^P((\rho_0)_S, \{(M_i)_{S'}\}) = F^U((\rho_0)_S \otimes |0\rangle_{A_1}\langle 0|, \{(M_i)_{S'} \otimes |1\rangle_{A_2}\}),$$

Eq. (11) then follows.

**III. ERROR OBSERVABLE FORMULATION**

In this section, we will formalize the optimization of FI over quantum preprocessing controls as a biconvex optimization problem using the concepts of error observables. Using this new formulation, the preprocessing optimization problem becomes numerically tractable using standard algorithms for biconvex optimization [48]; and also analytically tractable for some practically relevant quantum states (see later in Sec. IV and Sec. V).

Here we consider the preprocessing optimization problem in Eq. (7). On the surface, it may appear from the definition of FI (Eq. (3)) that the target function $F(\mathcal{E}(\rho), \{M_i\})$ is mathematically formidable. To simplify the target function, we introduce the error observable $X$ and the squared error observable $X_2$, defined by

$$X = \sum_i x_i M_i, \quad \text{and} \quad X_2 = \sum_i x_i^2 M_i,$$

where $x_i$ is interpreted as the difference between the estimator value $\hat{\theta}(i)$ and the true value $\theta$, i.e., $x_i = \hat{\theta}(i) - \theta$. We assume there are $r$ measurement outcomes and use $x$ to denote the vector $(x_1, \ldots, x_r)$. The locally unbiasedness conditions (Eq. (1)) for a single-shot measurement then become

$$\text{Tr}(\rho_0 X) = 0, \quad \text{and} \quad \text{Tr}(\partial\rho_0 X) = 0.$$  \hspace{1cm} (14)

It can be verified mathematically (which is essentially a proof of the CRB) that the minimum of the variance of the estimator under the locally unbiased conditions is the inverse of the FI, that is,

$$F(\rho_0, \{M_i\})^{-1} = \min_{\rho} \langle x | F(\rho_0 X^2), \text{ s.t. Eq. (14)}. \hspace{1cm} (15)$$

The problem above is a convex optimization over variables $x$, which can be solved using, e.g., the method of Lagrange multipliers. The optimal solution to $x$ is

$$x_i = \frac{1}{F(\rho_0, \{M_i\})} \frac{\text{Tr}(\partial\rho_0 M_i)}{\text{Tr}(\partial\rho_0 M_i)}, \hspace{1cm} (16)$$

where $\text{Tr}(\rho_0 M_i) \neq 0$ and $x_i = 0$ when $\text{Tr}(\rho_0 M_i) = 0$.

Note that the error observable formulation was previously used to derive the QCRB [58], where the QFI satisfies

$$J(\rho_0)^{-1} = \min_{\text{Hermitian } X} \text{Tr}(\rho_0 X^2), \text{ s.t. Eq. (14)}, \hspace{1cm} (17)$$

where $X$ can be an arbitrary Hermitian matrix. It has many useful applications [59–61], e.g., in [44], an algorithm was proposed to optimize the QFI over quantum channels using the error observable formulation of the QFI.

Combining Eq. (15) and Eq. (7), we have that

$$F^P(\rho_0, \{M_i\})^{-1} = \inf_{(x, \mathcal{E})} \text{Tr}(\mathcal{E}(\rho_0) X^2), \text{ s.t. Eq. (18)}.$$  \hspace{1cm} (18)

Let $\mathcal{H}_S$ and $\mathcal{H}_{S'}$ be the input and output systems of $\mathcal{E}(\cdot)$. It is convenient to represent a CPTP map $\mathcal{E}(\cdot)$ using a linear operator acting on $\mathcal{H}_{S'} \otimes \mathcal{H}_S$. Let $\mathcal{E}(\cdot) = \sum_{j} |\mathcal{K}_j \rangle \langle \mathcal{K}_j| \sigma^T$ be the Kraus representation of $\mathcal{E}(\cdot)$, the linear operator $\Omega = \sum |\mathcal{K}_j \rangle \langle \mathcal{K}_j|$ is usually called the Choi matrix of $\mathcal{E}$ [54], where $|\mathcal{K}_j \rangle := \sum_{jk} |\mathcal{K}_j \rangle \langle j| \sigma^T |k\rangle$. $\Omega$ corresponds to a CPTP map if and only if $\Omega \geq 0$ and $\text{Tr}_{S'}(\Omega) = \mathbb{1}_S$. $\mathcal{E}(\cdot)$ acting on any density operator $\sigma$ can be expressed using $\Omega$ through $\mathcal{E}(\sigma) = \text{Tr}_S(\mathbb{1} \otimes \sigma^T \Omega)$. (we use $(\cdot)^T$ to denote matrix transpose). Using the Choi matrix representation in Eq. (18), we have

**Theorem 3.** The optimal value of the following biconvex optimization problem gives the inverse of the QPFI

$$F^P(\rho_0, \{M_i\})^{-1} = \inf_{(x, \Omega)} \text{Tr}(X_2 \otimes \rho_0^T \Omega), \hspace{1cm} (19)$$

$s.t. \Omega \geq 0$, $\text{Tr}_{S'}(\Omega) = \mathbb{1}_S$, $\text{Tr}(X \otimes \rho_0^T \Omega) = 0$, $\text{Tr}(X \otimes \partial \rho_0^T \Omega) = 1$. 


Eq. (19) is a biconvex optimization problem of variables $\mathbf{x}$ and $\Omega$. Fixing $\Omega$, Eq. (19) is a quadratic program with respect to $\mathbf{x}$, and fixing $\mathbf{x}$, Eq. (19) is a semidefinite program with respect to $\Omega$: each of which is efficiently solvable when the system dimensions are moderate and the domain of variables is compact.

Note that the domain of $\mathbf{x}$ is unbounded in Eq. (19). In practice, one may impose a bounded domain on $\mathbf{x}$ so that the minimum of Eq. (19) always exists. For cases where the QPFI is attainable, the optimal value of the bounded version will be equal to the one of Eq. (19) when the size of the bounded domain is sufficiently large. For singular cases where the QPFI is not attainable, the optimal value of the bounded version will approach the one of Eq. (19) with an arbitrarily small error as the size of the domain increases. We describe an algorithm called the global optimization algorithm [62] in Appx. D that can solve the bounded version of Eq. (19).

Finally, we note that Theorem 3 does not directly generalize to the case of QUPFI because the Choi matrices of unitary operators do not form a convex set. (Although the target function in Eq. (19) is a concave function over $\Omega$; each of which is efficiently solvable when the system dimensions are moderate and the domain of variables is compact.) On the other hand, besides the set of quantum channels, our approach is also useful in optimizing the FI over other sets of quantum channels when the constraints on their Choi matrices can be represented using semidefinite constraints, e.g., the set of quantum channels that act only on a subsystem of the entire system.

### IV. PURE STATES

In this section, we consider the special case where $\rho_0 = |\psi_0\rangle \langle \psi_0|$ is pure, which is the most common type of quantum states in sensing experiments. We first consider the optimization of the FI over the error vector $\mathbf{x}$ and the unitary control $U$ and obtain two necessary conditions for the optimality of $(\mathbf{x}, U)$. We use them to prove the equality between the QPFI and the QUPFI for pure states, that is, unitary controls are always optimal for pure states (when $H_S \cong H_{S'}$). We also obtain an analytical expression of the QPFI for binary measurements (i.e., measurements with only two outcomes) and a semi-analytical expression for general commuting-operator measurements (i.e., measurements $\{M_i\}$ that satisfy $[M_i, M_j] = 0$ for all $i, j$). In particular, we prove that the optimal control is always to rotate the pure state and its derivative into a two-dimensional subspace spanned by two of the eigenstates of the commuting-operator measurement.

A. Necessary conditions for optimal controls

Proposition 2 shows that the optimization for the QPFI can be reduced to an optimization for the QUPFI using the ancillary system. Thus, here we first focus on the following optimization problem over the unitary control

$$F_U(\rho_0, \{M_i\})^{-1} = \inf_{x(U)} \text{Tr}(U \rho_0 U^\dagger X_2), \quad (20)$$

s.t. \(\text{Tr}(U \rho_0 U^\dagger X) = 0, \quad (21)\)

\(\text{Tr}(U \partial_\theta \rho_0 U^\dagger X) = 1. \quad (22)\)

and obtain necessary conditions for the optimality of $(\mathbf{x}, U)$ that will be useful later.

**Lemma 4.** If $(\mathbf{x}, U)$ is an optimal point for Eq. (20), it must satisfy

$$\frac{\text{Tr}(U \partial_\theta \rho_0 U^\dagger X)}{\text{Tr}(U \rho_0 U^\dagger X_2)} = 2[X, U \partial_\theta \rho_0 U^\dagger]. \quad (23)$$

In particular, suppose $\rho_0 = |\psi_0\rangle \langle \psi_0|$ is pure. Let

$$|\phi\rangle = U|\psi_0\rangle, \quad (24)$$

$$|\phi^\perp\rangle = \frac{1}{\sqrt{n}} U(1 - |\psi_0\rangle \langle \psi_0|) \partial_\theta |\psi_0\rangle, \quad (25)$$

where the normalization factor $n = \langle \partial_\theta |\psi_0\rangle (1 - |\psi_0\rangle \langle \psi_0|) |\partial_\theta |\psi_0\rangle$. Then Eq. (23) is equivalent to the following two conditions:

1. $X |\phi\rangle = 1/(2\sqrt{n}) |\phi^\perp\rangle$.

2. $(\langle \phi | X_2 |\phi\rangle X^2 - \langle \phi | X |\phi\rangle X_2) |\phi\rangle = 0$.

**Proof.** Assume $(\mathbf{x}, U)$ satisfies the constraints Eq. (21) and Eq. (22). Then for any unitary operator $V$ such that $\text{Tr}(U \partial_\theta \rho_0 U^\dagger V^\dagger X V) \neq 0$,

$$\left( \frac{\mathbf{x} - \text{Tr}(U \rho_0 U^\dagger V^\dagger X V) \mathbf{1}}{\text{Tr}(U \partial_\theta \rho_0 U^\dagger V^\dagger X V)}, VU \right) \quad (26)$$

also satisfies the constraints Eq. (21) and Eq. (22), where $\mathbf{1}$ is a $r$-dimensional vector of which each element is 1. We call the transformation above a “$V$-transformation” on $(\mathbf{x}, U)$. After a $V$-transformation, the target function becomes

$$\frac{\text{Tr}(U \rho_0 U^\dagger X_2 V) - \text{Tr}(U \rho_0 U^\dagger V^\dagger X V)^2}{\text{Tr}(U \rho_0 U^\dagger V^\dagger X V)^2}, \quad (27)$$

which shall be no smaller than $\text{Tr}(U \rho_0 U^\dagger X_2)$ when $(\mathbf{x}, U)$ is optimal. Let $V = e^{idG}$ where $dG$ is an arbitrary infinitesimally small Hermitian matrix. The first order derivative of Eq. (27) with respect to $dG$ must be zero, which then implies Eq. (23).

For pure states, Eq. (23) can be further simplified. Using the definitions of $|\phi\rangle$ and $|\phi^\perp\rangle$, we have $U \rho_0 U^\dagger$ =
trivially satisfied. Furthermore, choose (\(x\)) where \(H_\langle\it is always equal to 1\{ \\)

Eq (23) becomes

\[
|u\rangle \langle φ| - |φ\rangle \langle u| + |φ\rangle \langle φ^+| - |φ^+\rangle \langle φ| = 0, \tag{28}
\]

where \(|u\rangle = X|φ^+\rangle - \frac{\text{Re}[\langle φ|X^+|φ^+\rangle]}{\langle φ|X^+|φ\rangle}X_2|φ\rangle\) and \(|v\rangle = X|φ\rangle\). Eq (28) is equivalent to \(|u\rangle, |v\rangle \in \text{span}\{|φ\rangle, |φ^+\rangle\}\}, \langle φ|u\rangle, \langle φ^+|v\rangle \in \mathbb{R}\) and \(\langle φ|v\rangle = (u|φ^+\rangle\). Combining these conditions with the locally unbiasedness constraints \(\langle φ|X|φ\rangle = 0\) and \(2\sqrt{n}\text{Re}[\langle φ|X^+|φ^+\rangle] = 1\), the two conditions in Lemma 4 are then proven. Note that we explicitly put \(\langle φ|X^2|φ\rangle\) in Condition (2) for later use, although it is always equal to \(1/(4n)\) from Condition (1).

As a sanity check, consider the special case where \(\{M_i = |i\rangle \langle i|\}_{i=1}^{\dim(H_\mathcal{S})}\) is a projection onto an orthonormal basis of \(H_\mathcal{S}\). Then we have \(X_2 = X^2\), so Condition (2) is trivially satisfied. Furthermore, choose \((x, U)\) such that the error observable \(X = \frac{1}{2\sqrt{n}}(\langle φ^+\rangle \langle φ| + \langle φ| \langle φ^+\rangle\), Condition (1) is satisfied. Moreover, the variance of the estimation is

\[
\langle φ|X_2|φ\rangle = \langle φ|X^2|φ\rangle = \frac{1}{4n} = J(ρ_0)^{-1}, \tag{29}
\]

implying that the QFI is achievable using the above projective measurement and error observable. Here we use the fact that \(J(ρ_0) = 4n\) for pure states [10, 64]. For general quantum measurements, the inequality in

\[
\langle φ|X_2|φ\rangle \geq \langle φ|X^2|φ\rangle = J(ρ_0)^{-1} \tag{30}
\]

might strictly holds, in which case even for optimal controls the QUPFI is not equal to the QFI.

**B. Unitary controls are optimal**

Using the definitions of \(|φ\rangle\) and \(|φ^+\rangle\) in Eq (24) and Eq (25), we observe that Eq (20) can be rewritten as

\[
F^U(ψ_θ, M_i) = \inf_{(x|, |φ)} \langle φ|X_2|φ\rangle, \tag{31}
\]

\[\text{s.t. } \langle φ|φ^+\rangle = 0, \langle φ|X|φ\rangle = 0, \text{ Re}[\langle φ|X^+|φ^+\rangle] = 1/(2\sqrt{n}), \]

where \(n = J(ψ_θ)/4\) and \(ψ_θ = |ψ_θ\rangle \langle ψ_θ|\) is pure. Here \(|φ\rangle\) and \(|φ^+\rangle\) are two arbitrary normal vectors that are orthogonal. It is clear that \(F^U(ρ_θ, M_i)\) will be a product of \(n\) and a state-independent constant. We have

\[
F^U(ψ_θ, M_i) = γ(\{M_i\})J(ψ_θ), \tag{32}
\]

where

\[
γ(\{M_i\}) = \inf_{(x|, |φ)} \langle φ|X_2|φ\rangle, \tag{33}
\]

\[\text{s.t. } \langle φ|φ^+\rangle = 0, \langle φ|X|φ\rangle = 0, \text{ Re}[\langle φ|X^+|φ^+\rangle] = 1. \]

Or more explicitly,

\[
γ(\{M_i\}) = \sup_{|φ\rangle, |φ^+\rangle} \sum_i \text{Re}[\langle φ|M_i|φ^+\rangle]^2. \tag{34}
\]

\(γ(\{M_i\})\) is the normalized QUPFI for any pure states with unit QFIs and it is a function of \(\{M_i\}\) that lies in \([0, 1]\), which is the ratio between the QUPFI and the QFI for any pure states. It is independent of the exact \(ψ_θ\) and can fully characterize the power of quantum measurements in terms of estimation on pure states. Note that Eq (32) and Eq (34) were also proven using a different method in [44].

Using Condition (1) in Lemma 4, we now prove that unitary controls are always optimal, that is, the QUPFI is equal to the QUPFI when \(H_\mathcal{S} \cong H_\mathcal{S}_r\). We have the following theorem

**Theorem 5.** Consider a pure state \(ψ_θ\) and a quantum measurement \(\{M_i\}\) acting on the same system. Unitary preprocessing controls are always optimal among quantum preprocessing controls for optimizing the FI, i.e.,

\[
F^U(ψ_θ, \{M_i\}) = F^U(ψ_θ, \{M_i\}). \tag{35}
\]

Or equivalently,

\[
γ(\{M_i\}) = γ(\{M_i \otimes I_A\}), \tag{36}
\]

where \(A\) is an ancillary system of an arbitrary size.

**Proof.** We first consider the situation where the QUPFI is attainable, that is, that always exists an \((x, U)\) such that the infimum in Eq (20) is attainable. Using Condition (1) in Lemma 4, we can rewrite Eq (31) as

\[
F^U(ψ_θ, \{M_i\})^{-1} = \min_{(x|, |φ)} \langle φ|X_2|φ\rangle, \tag{37}
\]

\[\text{s.t. } \langle φ|φ^+\rangle = 0, \langle φ|X|φ\rangle = 1/(4n), \]

where \(n = J(ψ_θ)/4\). Let \(\dim(H_\mathcal{A}) \geq \dim(H_\mathcal{S})\), \(J(ψ_θ) = J(ψ_θ \otimes |0_A\rangle \langle 0_A|\) and Proposition 2 imply

\[
F^P(ψ_θ, \{M_i\})^{-1} = F^P(ψ_θ \otimes |0_A\rangle \langle 0_A|, \{M_i \otimes I_A\})^{-1} = \min_{(x|, |φ)} \langle φ|X \otimes I_A|φ\rangle \tag{38}
\]

\[\text{s.t. } \langle φ|X \otimes I_A|φ\rangle = 0, \langle φ|X^2 \otimes I_A|φ\rangle = 1/(4n).\]

It is equivalent to the optimization problem

\[
\min_{(x, σ)} \text{Tr}(σX_2), \tag{39}
\]

\[\text{s.t. } \text{Tr}(σX) = 0, \text{ Tr}(σX^2) = 1/(4n), \]

where \(σ\) is an arbitrary density operator and corresponds to \(Tr_A(\langle φ| \langle φ)\). Since the target function in Eq (39) is a concave function over \(σ\) and the contraints are convex over \(σ\), there always exists an optimal solution to \(σ\) that
is pure. Thus, the optimal values of Eq. (37) and Eq. (39) must be the same, proving Eq. (35).

When the QPFI of Eq. (20) is not attainable, we take \( M^{(e)}_i = (1 - \epsilon)M_i + \epsilon \text{Tr}(M_i) \frac{1}{2} \) and using Theorem 1, we have

\[
F[\psi_0, \{ M_i \}] = \lim_{\epsilon \to 0^+} F[\psi_0, \{ M_i^{(e)} \}]
\]

\[
= \lim_{\epsilon \to 0^+} F[U(\psi_0, \{ M_i^{(e)} \})] = F[U(\psi_0, \{ M_i \})],
\]

where in the second step we use the equality between the QPFI and the QUPFI in the case where the QUPFI is attainable.

So far, we have proven that Eq. (36) is true when \( \text{dim}(H_A) \geq \text{dim}(H_S) \), due to Proposition 2 and the equality between the QPFI and the QUPFI. It also holds for any \( H_{A'} \) such that \( \text{dim}(H_{A'}) \leq \text{dim}(H_S) \) because we have \( \gamma(M_i \otimes I_{A'}) \geq \gamma(M_i \otimes I_{A'}) \geq \gamma(M_i) \) by definition.

\[ \square \]

C. Analytical solution for binary measurements

Here we provide an analytical solution to the QPFI and the corresponding optimal preprocessing control using Proposition 2 for binary measurements where \( r = 2 \).

1. Measurement on a qubit

We first consider the simplest case where the measurement is on a single qubit. Let \( X = x_1 M_1 + x_2 M_2 \) where \( M_1 = M \) and \( M_2 = 1 - M \). Without loss of generality, we assume

\[
M = m_1 |1\rangle \langle 1| + m_2 |2\rangle \langle 2|,
\]

for some \( m_1, m_2 \in [0, 1] \), where \( \{|1\rangle, |2\rangle\} \) is an orthonormal basis. Moreover, we assume \( m_1 > m_2 \) and \( 1 - m_1 \geq m_2 \). (When \( m_1 = m_2 \), we must have \( \gamma(M_i) = 0 \) because the measurement outcome does not depend on \( \theta \).) Here \( m_2 \) and \( 1 - m_1 \) can be interpreted as the error probabilities that state \( |2\rangle \) is mistaken for \( |1\rangle \), and state \( |1\rangle \) is mistaken for \( |2\rangle \), respectively.

Consider first the case where \( 1 > m_1 > m_2 > 0 \), that is, the error probabilities are both non-zero. We show in Appx. E 1 that all solutions that satisfy the two necessary conditions in Lemma 4 give the same optimal FI. One optimal solution to the preprocessed state is

\[
|\phi \rangle = \sqrt{p} |1\rangle + \sqrt{1-p} |2\rangle,
\]

\[
|\phi^+ \rangle = \sqrt{1-p} |1\rangle - \sqrt{p} |2\rangle,
\]

where

\[
p = \frac{\sqrt{m_2(1-m_2)}}{\sqrt{m_1(1-m_1) + \sqrt{m_2(1-m_2)}}}.
\]

Here the optimal unitary control \( U \) can be chosen as any unitary such that Eq. (24) and Eq. (25) are true. Note that the symmetry transformations \( |\phi^+ \rangle \mapsto -|\phi^+ \rangle \), \( |1\rangle \mapsto e^{i\omega}|1\rangle \) and \( |2\rangle \mapsto e^{i\omega'}|2\rangle \) for any \( \omega, \omega' \in \mathbb{R} \) will generate alternative optimal solutions, and they all provide the same optimal normalized FI:

\[
\gamma(M_i) = 1 - (\sqrt{m_1 m_2} + \sqrt{(1 - m_1)(1 - m_2)})^2.
\]

Note that this result was obtained also in [44] using a different method based on the Bloch sphere representation. Here \( \sqrt{1 - \gamma(M_i)} \) is exactly equal to the fidelity between two binary probability distributions \( (m_1, 1 - m_1) \) and \( (m_2, 1 - m_2) \).

Take the symmetric binary measurement as an example, where \( m_1 = 1 - m, m_2 = m \) and \( m < 1/2 \), and \( m \) represents the probability of a bit-flip error in the measurement. Then we have \( p = 1/2 \) (as expected from the bit-flip symmetry), and \( \gamma(M_i) = 1 - 4m(1 - m) \), which is equal to 1 in the noiseless case, and drops to 0 when \( m \to 1/2 \).

In the case of perfect projective measurements where \( 1 = m_1 > m_2 = 0 \), we show in Appx. E 1 that the QPFI is equal to the QFI and is attainable for any \( 0 < p < 1 \). The case where \( 1 > m_1 > m_2 = 0 \) is singular, in the sense that the QPFI is no longer attainable but only approachable. It corresponds to the situation where one type of error (|2\rangle mistaken for |1\rangle) is zero, while the other is non-zero. In this case, we have \( \gamma(M_i) = m_1 \) using Eq. (45) and Theorem 1.

2. Measurement on a qubit

Next, we consider the general case where the measurement is on a qubit and we assume \( \text{dim}(H_S) = d \geq 2 \). Without loss of generality, we assume

\[
M = \sum_{j=1}^d m_j |j\rangle \langle j|,
\]

where \( \{|j\rangle\}_{j=1}^d \) is an orthonormal basis of \( H_S \). We also assume \( m_i \geq m_j \) for all \( i \leq j \) without loss of generality. Here we assume \( 1 > m_1 > m_d > 0 \), which guarantees the attainability of the QPFI (see Lemma S1 in Appx. C) and the non-triviality of quantum measurements. (The singular cases where \( m_1 = 1 \) or \( m_d = 0 \) can be solved using Theorem 1.) We show in Appx. E 2 that the optimal solution to \( |\phi\rangle \) is supported on basis states corresponding to at most two different values of \( m_i \) and the problem is simplified to selecting the optimal basis states and applying the qubit-case results. We show that

\[
|\phi \rangle = \sqrt{p} |1\rangle + \sqrt{1-p} |d\rangle,
\]

\[
|\phi^+ \rangle = \sqrt{1-p} |1\rangle - \sqrt{p} |d\rangle,
\]

is an optimal solution, where

\[
p = \frac{\sqrt{m_d(1-m_d)}}{\sqrt{m_1(1-m_1) + \sqrt{m_d(1-m_d)}}}.
\]
The normalized QPFI is given by
\[ \gamma(\{M_i\}) = 1 - \left( \sum \frac{m_k^{(i)} m_k^{(j)}}{m_k^{(i)} + (1 - p_{kl})m_k^{(j)}} \right)^2. \]  
(50)

Viewing \(\{m_{1i}, 1 - m_{1i}\}\) as \(d\) binary probability distributions, the optimal strategy is always to select the two probability distributions that has the minimum fidelity (i.e., the largest distance) between each other.

D. Semi-analytical solution for commuting-operator measurements

Here we consider commuting-operator measurements, where all measurement operators commutes, which characterize the most common type of measurements in quantum sensing experiments, e.g., projective measurements affected by detection errors.

Assume \(\dim(\mathcal{H}_S) = d \geq 2\). Without loss of generality, we assume
\[ M_i = \sum_{j=1}^d m_j^{(i)} |j\rangle \langle j|, \]  
(51)

where \(\{|j\rangle\}^d_{j=1}\) is an orthonormal basis of \(\mathcal{H}_S\) and \(\sum_{j=1}^d m_j^{(i)} = 1\) for all \(j\). Again, we assume \(m_j^{(i)} > 0\) for all \(i, j\) to exclude the singular cases where the QPFI is not attainable.

In order to find the optimal control, we first prove the following theorem which states that the optimal \(|\phi\rangle\) can be restricted to a two-dimensional subspace spanned by two basis states, i.e., the optimal unitary controls rotate the pure state and its derivative to a subspace spanned by two of the eigenstates of the commuting-operator measurement.

**Theorem 6.** For commuting-operator measurements (Eq. (51)), there always exists an optimal solution to \(|\phi\rangle, |\phi^\perp\rangle\) such that \(|\phi\rangle = \sqrt{p} |k\rangle + \sqrt{1-p} |l\rangle\) and \(|\phi^\perp\rangle = \sqrt{1-p} |k\rangle - \sqrt{p} |l\rangle\) for two basis states \(|k\rangle\) and \(|l\rangle\) and \(0 < p < 1\).

The proof is provided in Appx. F 1. Then we see that the normalized QPFI for commuting-operator measurements will be
\[ \gamma(\{M_i\}) = \max_{1 \leq k < l \leq d} \gamma_{kl}, \]  
(52)

using Theorem 6, where
\[ \gamma_{kl} = \gamma(\{M_i\}|_{\text{span}\{|k\rangle, |l\rangle\}}), \]  
(53)

and \(\{M_i\}|_{\text{span}\{|k\rangle, |l\rangle\}}\) is the quantum measurement restricted in the subspace spanned by \(|k\rangle\) and \(|l\rangle\).

We show in Appx. F 2 that
\[ \gamma_{kl} = \sum_i p_{kl}(1 - p_{kl}) \left( \frac{m_k^{(i)} - m_l^{(i)}}{m_k^{(i)} + (1 - p_{kl})m_l^{(i)}} \right)^2, \]  
(54)

where \(p_{kl} \in [0, 1]\) is the unique solution to
\[ \sum_{i=1}^r \frac{m_k^{(i)} m_j^{(i)}}{m_k^{(i)} + (1 - p_{kl})m_j^{(i)}} = \sum_{i=1}^r \frac{m_k^{(i)} m_j^{(i)} - m_l^{(i)} m_j^{(i)}}{m_k^{(i)} + (1 - p_{kl})m_l^{(i)}} \]  
(55)

and it corresponds to the optimal preprocessed state
\[ |\phi_{kl}\rangle = \sqrt{p_{kl}} |k\rangle + \sqrt{1 - p_{kl}} |l\rangle, \]  
(56)
\[ |\phi^\perp_{kl}\rangle = \sqrt{1 - p_{kl}} |k\rangle - \sqrt{p_{kl}} |l\rangle. \]  
(57)

(The symmetry transformations \(|\phi^\perp\rangle \mapsto -|\phi^\perp\rangle, |k\rangle \mapsto e^{i\omega} |k\rangle\) and \(|l\rangle \mapsto e^{i\omega'} |l\rangle\) for any \(\omega, \omega' \in \mathbb{R}\) will generate alternative optimal solutions.) For the special case where \(r = 2\), the problem reduces to the binary measurement problem discussed in Sec. IV C and \(p_{kl}\) can be solved analytically. In general, however, the analytical solution to \(p_{kl}\) might not exist since it is a root of a high degree polynomial (Eq. (55)). The optimal preprocessed state \(|\langle \phi^\perp |, |\phi^\perp\rangle\rangle\) that achieves Eq. (52) is chosen as \(|\langle \phi_{kl} |, |\phi^\perp_{kl} \rangle\rangle\) for \((k, l)\) that maximizes \(\gamma_{kl}\).

Note that although Theorem 6 does not tell us how to choose the two optimal basis states, it directly helps us identify the optimal control when such a choice is obvious. For example, consider a \(n\)-qubit system \((\text{span}\{|1\rangle, |2\rangle\})^\otimes n\) measured by \(\{M, 1 - M\}\) (independently on each subsystem) and \(M = (1 - m)|1\rangle \langle 1| + m|2\rangle \langle 2|\). Then using Theorem 6, due to the bit-flip symmetry and the fact that tracing out some parts of the quantum state will not increase its QPFI, it is clear that rotating \(|\langle \phi |, |\phi^\perp\rangle\rangle\) into \(\text{span}\{|1\rangle^\otimes n, |2\rangle^\otimes n\}\), or any other basis states, e.g., \(|\{121 \cdots 1\}, |212 \cdots 2\rangle\) that are distinct on each qubit, must be an optimal choice. In general, it remains open if there is a simple criteria to help us select the optimal \(k\) and \(l\) besides a direct calculation of Eq. (54) for different \(k\) and \(l\).

V. CLASSICALLY MIXED STATES

In this section, we consider another type of quantum states, which we called classically mixed states, with commuting-operator measurements. A classically mixed state is a state which commutes with its derivative, e.g., Gibbs states whose temperature is to be estimated [65]. In this section, we use the following form of classically mixed states:
\[ \zeta_\theta = \sum_{i=1}^D \lambda_i |i\rangle \langle i|, \]  
(58)

where \(D = \dim(\mathcal{H}_S), \lambda_i\) are functions of \(\theta\) (we omit the subscript \(\theta\) for notation simplicity), \(|\{i\}\rangle\) is an orthonormal basis of \(\mathcal{H}_S\) that is independent of \(\theta\) and we use \(\zeta_\theta\) to represent classically mixed states. Note that the QFI of Eq. (58) \(J(\zeta_\theta) = \sum_{i=1}^D \partial_\theta \lambda_i^2 / \lambda_i\) is equal to the FI \(J(\{\lambda_i\})\) of the classical distribution \(\{\lambda_i\}_{i=1}^D\), which is the reason we call Eq. (58) a classically mixed state. Also,
note that we assume in this section, without loss of generality, that the commuting-operator measurement \( \{M_i\} \) and the classically mixed state \( \rho \) share the same eigenstates \( \{|i\rangle\}_{i=1}^{max\{d,D\}} \), as it is always possible to apply a unitary rotation in the preprocessing control so that they are aligned.

We first show that optimizing the FI over quantum channels is equivalent to finding optimal stochastic matrices (which describes the transitions of a classical Markov chain) for the classical noisy measurement problem. Then we prove that the optimal control always corresponds to a stochastic matrix that has only elements 0 or 1, which we call a coarse graining stochastic matrix.

It implies that the QPFI is always attainable, and that the OPQFI can in some cases be strictly larger than the QUPFI. Finally, we closely examine the case of a binary measurement on a single qubit.

### A. Optimization over stochastic matrices

**Lemma 7.** Consider classically mixed states Eq. (58) and commuting-operator measurements Eq. (51). Then

\[
F^P(\rho, \{M_i\}) = \sup_{P \in S_{d,D}} J(\{m^{(i)}T P \lambda\}), \tag{59}
\]

and when \( d = D \),

\[
F^U(\rho, \{M_i\}) \leq \sup_{P \in S_{d}^{st}} J(\{m^{(i)}T P \lambda\}), \tag{60}
\]

where \( S_{d,D} \) represents the set of \( d \times D \) stochastic matrices of which every column vector sums up to one and \( S_{d}^{st} \) represents the set of \( d \) doubly stochastic matrices of which every column vector or row vector sums up to one, \( m^{(i)} \) is a column vector whose entries are \( m_{ij}^{(i)} \), \( \lambda \) is a column vector whose entries are \( \lambda_i \).

**Proof.** Let \( \mathcal{E}(\cdot) = \sum_j K_j \cdot \mathcal{E}(\cdot) \) be any arbitrary quantum channel, then we have

\[
\text{Tr}(M_i \mathcal{E}(\rho)) = m^{(i)}T P \lambda, \tag{61}
\]

where \( P_{\ell k} = |(K_{j})_{\ell k}|^2 \), which implies

\[
J(\{\text{Tr}(M_i \mathcal{E}(\rho))\}) = J(\{m^{(i)}T P \lambda\}). \tag{62}
\]

Since \( \sum_j K_j \cdot K_j = \mathbb{1} \), we must have \( \sum_{\ell} P_{\ell k} = 1 \). Thus, \( P \) is a stochastic matrix. For any quantum channel, there exists a stochastic matrix such that Eq. (61) holds true, proving the left-hand side is no larger than the right-hand side in Eq. (59). Moreover, when \( \mathcal{E}(\cdot) = U \cdot \mathcal{E}(\cdot) \) is a unitary channel, \( P_{\ell k} = |(U_j)_{\ell k}|^2 \) must be doubly stochastic, implying Eq. (60).

On the other hand, for any stochastic matrix \( P \), we define \( K_{(\ell,k)} = \sqrt{P_{\ell k}} |\ell\rangle \langle k| \) for \( 1 \leq \ell \leq d \) and \( 1 \leq k \leq D \). Then we have \( \sum_{\ell k} K_{(\ell,k)}^\dagger K_{(\ell,k)} = \sum_{\ell k} P_{\ell k} |k\rangle \langle k| = \mathbb{1} \).

And \( \mathcal{E}(\cdot) = \sum_{(\ell,k)} K_{(\ell,k)}^\dagger \rho K_{(\ell,k)} \) is then a quantum channel. For any stochastic matrix, there exists a quantum channel such that Eq. (61) holds true, proving the left-hand side is no smaller than the right-hand side in Eq. (59).

We show in Lemma 7 that the problem of optimizing preprocessing quantum controls on classically mixed states with commuting-operator measurements is equivalent to a classical version of preprocessing optimization where

\[
F^P(\lambda, \{m^{(i)}\}) := \sup_{P \in S_{d,D}} J(\{m^{(i)}T P \lambda\}) \tag{63}
\]

represents the classical FI with respect to a classical distribution \( \lambda \) and a noisy measurement \( m^{(i)} \) satisfying \( \sum_i m_{ij}^{(i)} = 1 \) (1 is a vector with all elements equal to 1), optimized over any stochastic mapping described by stochastic matrices. In particular, for perfect measurements where \( (m^{(i)})_{ij} = \delta_{ij} \), \( F^P(\lambda, \{m^{(i)}\}) = J(\lambda) = \frac{1}{\lambda_i} \) is the classical FI. Note that Theorem 9 presented later implies that the supremum of the FI over stochastic matrices is always attainable using some \( P \in S_{d,D} \) and it means we are allowed to replace \( \sup_{P \in S_{d,D}} \) by \( \max_{P \in S_{d,D}} \) in the definition (Eq. (63)).

### B. Coarse graining controls are optimal

We first consider the classical case and prove Eq. (63) can always be attained using some \( d \times D \) stochastic matrix \( P \) where every element of \( P \) is either 0 or 1. We call this type of stochastic matrix a coarse graining stochastic matrix in the sense that \( P \) maps the sum of one or multiple entries of \( \lambda \) to one entry in \( P \lambda \), which is a coarse graining of the measurement outcomes.

**Lemma 8.** Given a classical probability distribution \( \lambda \in \mathbb{R}^d \) and a measurement \( \{m^{(i)}\} \subseteq \mathbb{R}^d \) (satisfying \( \sum_i m_{ij}^{(i)} = 1 \)). When \( F^P(\lambda, \{m^{(i)}\}) \) is attainable, there exists a \( d \times D \) coarse graining stochastic matrix \( P \) such that

\[
F^P(\lambda, \{m^{(i)}\}) = J(\{m^{(i)}T P \lambda\}). \tag{64}
\]

**Proof.** Suppose \( F^P(\lambda, \{m^{(i)}\}) \) is attainable and \( P^* \) is an optimal solution. We will show that there exists an optimal solution \( P \) whose every column vector contains one (and only) element equal to 1. If \( P^* \) does not satisfy this condition, without loss of generality, assume \( P^*_{11} = t_1^* \) and \( P^*_{21} = a_1^* - t_1^* \) where \( 0 < t_1^* < a_1^* \leq 1 \). Let \( P(t) \) be a stochastic matrix function of \( t \) in \([0, a_1^*] \) where \( P(t)_{11} = t_1 \), \( P(t)_{21} = a_1^* - t_1 \) and \( P(t)_{\ell k} = (P^*)_{\ell k} \) for
(ℓ, k) ≠ (1, 1), (2, 1). We have the FI equal to
\[ f(t_1) = \sum_i \frac{(\partial_i (m^{(i)} T P(t_1) \lambda_\theta))^2}{m^{(i)} T P(t_1) \lambda_\theta} = \sum_i \frac{((m^{(i)}_1 - m^{(i)}_2) \partial_i \lambda_1 t_1 + b^{(i)})^2}{(m^{(i)}_1 - m^{(i)}_2) \lambda_1 t_1 + a^{(i)}}, \]
where \( a^{(i)} \) and \( b^{(i)} \) are constants, independent of \( t_1 \). The second order derivative of \( f(t_1) \) is
\[ \frac{\partial^2 f(t_1)}{\partial t_1^2} = \sum_i \frac{2 (m^{(i)}_1 - m^{(i)}_2)^2 (a^{(i)} \partial_i \lambda_1 - b^{(i)} \lambda_1)^2}{((m^{(i)}_1 - m^{(i)}_2) \lambda_1 t_1 + a^{(i)})^3}, \]
(65)
which is always non-negative. Therefore, \( f(t_1) \) is a convex function and always attains its maximum at the boundary \( t_1 = 0 \) or \( t_1 = a^*_1 \). Repeat this argument many times, one can show that there exists an optimal solution \( P \) such that there is only one positive entry in every column.

Note that it is not necessarily true that the optimal coarse graining stochastic matrix that maximizes \( J(\{m^{(i)} T P \lambda_\theta\}) \) is a full-rank matrix. Consider the following example. Let \( d = D = 3 \), \( r = 2 \), \( \lambda_\theta = (\cos^2 \theta, \frac{1}{2} \sin^2 \theta, \frac{1}{2} \sin^2 \theta) \), \( m^{(1)} = (1, \frac{1}{4}, 0) \), and \( m^{(2)} = (0, \frac{1}{4}, 1) \). Then it is clear that the following stochastic matrix is optimal,
\[ P^* = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix}, \]
(66)
because \( J(\{m^{(i)} T P^* \lambda_\theta\}) = J(\lambda_\theta) = 4 \). However, it can be verified by enumeration that \( J(\{m^{(i)} T P \lambda_\theta\}) \leq 3 \), whenever \( P \) is a permutation matrix, showing the non-optimality of the full-rank stochastic matrices.

Using Lemma 7, we can show a similar result to Lemma 8 in the quantum case, that is, coarse graining channels are optimal quantum controls.

**Theorem 9.** Consider classically mixed states Eq. (58) and commuting-operator measurements Eq. (51). The QPFI is always attainable using the following type of quantum channels, which we call coarse graining channels,
\[ \mathcal{E}(\cdot) = \sum_{\ell k} P_{\ell k} |\ell\rangle \langle k| \cdot |k\rangle \langle \ell|, \]
(67)
where \( P_{\ell k} \) is a \( d \times D \) stochastic matrix of rank \( \min\{d, D\} \) satisfying \( \sum_{\ell} P_{\ell k} = 1 \) and \( P_{\ell k} = 0 \) or 1.

**Proof.** By definition, there exists a sequence of channels \( \{\mathcal{E}_1, \ldots, \mathcal{E}_n, \ldots\} \) such that \( \lim_{n \to \infty} F(\mathcal{E}_n, \{M_i\}) = F^\theta(\zeta_\theta, \{M_i\}) \). According to Eq. (62) in the proof of Lemma 7 and the arguments in Lemma 8, for every \( \mathcal{E}_n \) there exists a channel \( \tilde{\mathcal{E}}_n \) of the form Eq. (67) such that \( F(\tilde{\mathcal{E}}_n, \{M_i\}) \leq F(\mathcal{E}_n, \{M_i\}) \). Therefore, \( \lim_{n \to \infty} F(\tilde{\mathcal{E}}_n, \{M_i\}) = F^\theta(\zeta_\theta, \{M_i\}) \). Since there are finite number of channels of the form Eq. (67), there must exist a \( \mathcal{E}^* = \tilde{\mathcal{E}}_n \) for some \( n \) such that
\[ F(\mathcal{E}^*, \{M_i\}) = F^\theta(\zeta_\theta, \{M_i\}), \]
(68)
proving the attainability of the QPFI.

**Theorem 9** also implies that there is a gap between the QUPFI and the QPFI for general quantum states, unlike for pure states where the QUPFI is equal to the QPFI.

**Theorem 10.** There exists a classically mixed state \( \zeta_\theta \) and a commuting-operator measurement \( \{M_i\} \) such that
\[ F^U(\zeta_\theta, \{M_i\}) < F^\theta(\zeta_\theta, \{M_i\}). \]
(69)

**Proof.** Consider the example discussed below Lemma 8 and here we fix \( \theta = \pi/4 \). **Theorem 9** implies that for \( \zeta_\theta = \cos^2 \theta |1\rangle \langle 1| + \frac{1}{2} \sin^2 \theta |2\rangle \langle 2| + \frac{1}{2} \sin^2 \theta |3\rangle \langle 3| \),
\[ J(\zeta_\theta) = F^\theta(\zeta_\theta, \{M_i\}) = 4, \]
(70)
where \( M_1 = |1\rangle \langle 1| + \frac{1}{2} |2\rangle \langle 2| \) and \( M_2 = \frac{1}{2} |2\rangle \langle 2| + |3\rangle \langle 3| \). In general, given any stochastic matrix \( P \), the probabilities for measurement outcomes 1 and 2 must have the form
\[ p_1 = m^{(1)} T P \lambda_\theta = a \cos^2 \theta + b \sin^2 \theta, \]
(71)
\[ p_2 = m^{(2)} T P \lambda_\theta = (1 - a) \cos^2 \theta + (1 - b) \sin^2 \theta, \]
(72)
for some \( 0 \leq a, b \leq 1 \). Moreover, \( J(\{p_1, p_2\}) = (4 - b)^2/(2 - (a + b))/(a + b) \). And \( J(\{p_1, p_2\}) = J(\zeta_\theta) \) if and only if \( (a, b) = (1, 0) \) or \( (0, 1) \). Noting that the situation where \( (a, b) = (1, 0) \) or \( (0, 1) \) is not possible if \( P \) is doubly stochastic. Eq. (69) is then proven applying Lemma 7.

**C. Binary measurement on a single qubit**

With **Theorem 9**, in principle, one can find the QPFI for classically mixed states and commuting-operator measurements by exhausting all channels of the form Eq. (67) which is contained in a finite set. However, since the number of coarse graining stochastic matrices is large, the exhaustion procedure will be too costly. Here we closely examine a special case where a classically mixed state is measured by a binary measurement on a single qubit. The time to exhausting all coarse graining matrices is exponentially large with respect to the state dimension \( D \). We will show that the time to find a solution can be reduced to a linear complexity by narrowing down the possible forms of the optimal controls.

To be specific, consider the binary measurement \( M_1 = M = m_1 |1\rangle \langle 1| + m_2 |2\rangle \langle 2| \), \( M_2 = \mathbb{1} - M \) (assuming \( m_2 \leq \min\{m_1, 1 - m_1\} \)), and \( \zeta_\theta = \sum_{i=1}^D \lambda_i |i\rangle \langle i| \). Then using
Lemma 7, we have $F^P(\zeta_0, \{M_i\}) = \max_t f(t)$ where
\[
f(t) = \frac{(m_1 - m_2)^2(t^T \partial_0 \lambda_0)^2}{(m_2 + (m_1 - m_2)t^T \lambda_0)(1 - m_2 - (m_1 - m_2)t^T \lambda_0)},
\]
where $t$ is a column vector in $[0, 1]^D$, corresponding to the first row of the stochastic matrix $P$ in the proof of Lemma 7.

Without loss of generality, we can arrange the order of the elements in $\lambda_0$ such that
\[
(\partial_0 \lambda_i)/\lambda_i \geq (\partial_0 \lambda_j)/\lambda_j, \quad \forall i < j \text{ and } \lambda_{i,j} > 0.
\]
Note that we do not care about elements equal to 0 in $\lambda_0$ as they provide no contribution to the QPFI. Then we assert that
\[
F^P(\zeta_0, \{M_i\}) = \max_{i \in [1, D-1]} \max_{j \in [1,D-1]} \{f(1_{\leq i}), f(1_{\geq i})\},
\]
where $1_{\leq i}$ represents the vector whose the first $i$ elements are equal to 1 and the rest are zero. $1_{\geq i+1} = 1 - 1_{\leq i}$. Now we prove Eq. (75). Without loss of generality, assume $\lambda_i > 0$ for all $i$ (we can simply take $t_i = 0$ for any $\lambda_i = 0$). Choose an optimal $t^* \in [0, 1]^D$ that maximizes $f(t)$. Consider the following three cases: (i) $t^T \partial_0 \lambda_0 = 0$. Then the QPFI is zero and Eq. (75) is trivial. (ii) $t^T \partial_0 \lambda_0 > 0$. If there exists $i < j$ such that $t_i^* = 1$ and $t_j^* > 0$. Then define $t_i^{**} = t_i^* + \epsilon/\lambda_i$ and $t_j^{**} = t_j^* - \epsilon/\lambda_j$ where $\epsilon = \min\{(1 - t_i^*) \lambda_i, t_j^* \lambda_j\}$ and $t_{k \neq i,j}^{**} = t_{k \neq i,j}^*$. Then we have either $i^{**} = 1$ or $j^{**} = 0$. Moreover, since $f(t^{**}) \geq f(t^*)$, $t^{**}$ is also optimal. Repeating this procedure, we can always find an optimal $t$ of the form $(1 \cdots 1 t 0 \cdots 0)$ for some $t \in [0, 1]$. (iii) $t^T \partial_0 \lambda_0 < 0$. If there exists $i < j$ such that $t_i^* > 0$ and $t_j^* < 1$. Then define $t_i^{**} = t_i^* - \epsilon/\lambda_i$ and $t_j^{**} = t_j^* + \epsilon/\lambda_j$ where $\epsilon = \min\{(1 - t_j^*) \lambda_j, t_i^* \lambda_i\}$ and $t_{k \neq i,j}^{**} = t_{k \neq i,j}^*$. Repeating this procedure, we can always find an optimal $t$ of the form $(0 \cdots 0 t 1 \cdots 1)$ for some $t \in [0, 1]$. Combining the discussion above with Theorem 9, Eq. (75) is proven.

VI. GENERAL QUANTUM STATES

In Sec. IV and Sec. V, we have obtained fruitful results on preoptimization for pure states and classically mixed states. Here, we consider the QPFI for general mixed states and derive useful upper and lower bounds on them.

A. Upper bound

Theorem 11. Given any density operator $\rho_0$ and quantum measurement $\{M_i\}$, we have
\[
F^P(\rho_0, \{M_i\}) \geq \gamma(\{M_i\}) J(\rho_0).
\]

Proof. Suppose $\mathcal{H}_A_1$ and $\mathcal{H}_A_2$ are ancillary systems such that $\mathcal{H}_{A_1} \otimes \mathcal{H}_S \cong \mathcal{H}_{A_2} \otimes \mathcal{H}_S$ and dim($\mathcal{H}_{A_1}$) $\geq$ dim($\mathcal{H}_S$)$^2$, where $\mathcal{H}_S$ and $\mathcal{H}_{A_2}$ are the systems $\rho_0$ and $\{M_i\}$ act on. We also define an additional environmental system $\mathcal{H}_E$ satisfying dim($\mathcal{H}_E$) = dim($\mathcal{H}_S$). Let $\psi_0 = |\psi_0\rangle \langle \psi_0|_{ES}$ denote the purifications of $\rho_0$ in $\mathcal{H}_E \otimes \mathcal{H}_S$. Using the purification-based definition of QFI [15, 66], we have
\[
J(\rho_0) = \min_{\psi_0, \rho_0 = \text{Tr}_E(\psi_0)} J(\psi_0).
\]
Choose $\psi^*_0$ to be the optimal purification $\psi^*_0$ that minimizes $J(\psi_0)$ such that $J(\rho_0) = J(\psi^*_0)$. Then
\[
F^P(\rho_0, \{M_i\}) = F^U(\rho_0 \otimes |0\rangle \langle 0|_{A_1} , \{M_i \otimes I_{A_2}\}) \leq F(\psi^*_0 \otimes |0\rangle \langle 0|_{A_1} , \{I_E \otimes M_i \otimes I_{A_2}\}) \leq \gamma(\{I_E \otimes M_i \otimes I_{A_2}\}) J(\psi^*_0 \otimes |0\rangle \langle 0|_{A_1}) = \gamma(\{M_i\}) J(\rho_0),
\]
where we use Proposition 2, Eq. (32) and Theorem 5.

Theorem 11 provides an upper bound on the QPFI for general quantum states. In particular, it shows the ratio between the QPFI and the QFI is always upper bounded by a state-independent constant $\gamma(\{M_i\})$ which is attainable when the state is pure and gives rise to the following CRB for general quantum states under noisy measurement $\{M_i\}$:
\[
\Delta \hat{\theta} \geq \frac{1}{\sqrt{N_{\text{exp}} \gamma(\{M_i\}) J(\rho_0)}}.
\]

B. Lower bound

Lemma 12. Consider a density operator $\rho_0$ and quantum measurement $\{M_i\}$. Assume $\{T_i\}$ is a QFI-attainable measurement, i.e., $F(\rho_0, \{T_i\}) = J(\rho_0)$. Let the quantum-classical channel $\mathcal{T}(\cdot) = \sum_i \text{Tr}(\cdot)T_i |i\rangle \langle i|_{C}$ where $\{|i\rangle_C\}$ is an orthonormal basis of an auxiliary system $\mathcal{H}_C$. Then
\[
F^P(\rho_0, \{M_i\}) \geq F^P(\mathcal{T}(\rho_0), \{M_i\}).
\]

The proof Lemma 12 is straightforward—it immediately follows from the definition of the QFPI. However, it is simple but powerful, as we will see later Sec. VII.

Note that the equality in Lemma 12 holds when $\rho_0$ is a classically mixed state, because the projective measurement onto the basis of $\mathcal{H}_S$ is QFI-attainable and $\mathcal{T}(\rho_0) = \rho_0$ for classically mixed states. For general mixed states, since $\mathcal{T}(\rho_0)$ is a classically mixed state, the results in Sec. V can be applied here to compute $F^P(\mathcal{T}(\rho_0), \{M_i\})$ and derive lower bounds for general mixed states. For example, one can divide the measurement operators into two subsets, restrict the measurement in a two-dimensional subspace, and then use our
previous result of the binary measurement on a qubit for classically mixed states to derive an efficiently computable lower bound on the QPFI.

Note that unlike the upper bound (Theorem 11), there are no constant lower bounds independent of $\rho_n$ on the ratio between $F^\mathcal{Q}(\rho_n, \{M_i\})$ and $J(\rho_n)$. For example, consider the single qubit case where $\rho_n = |1\rangle \langle 1| + |2\rangle \langle 2|$, $M_1 = (1 - \eta) |1\rangle \langle 1| + \eta |2\rangle \langle 2|$, and $M_2 = I - M_1$ ($0 < \eta < 1/2$). We have, from Sec. V, that

$$F^\mathcal{Q}(\rho_n, \{M_i\}) = \frac{4(1 - 2\eta)^2 \sin^2(2\theta)^2}{1 - (1 - 2\eta) \cos(2\theta)^2},$$

which tends to zero as $\theta \to 0$. On the other hand, $J(\rho_n) = 4$ is a constant, showing that $F^\mathcal{Q}(\rho_n, \{M_i\})/J(\rho_n)$ has no state-independent constant lower bounds.

### VII. ASYMPOTIC LIMIT

In this section, we consider the power of quantum pre-processing in the asymptotic limit (see Fig. 2). We consider a multi-partite system $\mathcal{H}_S = \mathcal{H}_E^\otimes n$ and $\mathcal{H}_S' = \mathcal{H}_E^\otimes n$, a set of quantum states $\rho_n^{(i)}$ in $\mathcal{H}_E^\otimes n$, and quantum measurements $\{M_i\}^\otimes n$ that can be written as tensor products of identical measurements on each subsystem $\mathcal{H}_E$. Arbitrary pre-processing quantum channels $\mathcal{E}(\cdot)$ can be applied before the measurement. We will show that for a generic class of quantum states, the QPFI can reach the QFI asymptotically for large $n$.

#### A. Attaining the QFI with noisy measurements

**Theorem 13.** Given a set of quantum states $\{\rho^{(n)}_\theta\}_n$, where $\rho^{(n)}_\theta$ is a function of $\theta$ and acts on $\mathcal{H}_E^\otimes n$ for each $n$, we have

$$\lim_{n \to \infty} \frac{F^\mathcal{Q}(\rho^{(n)}_\theta, \{M_i\}^\otimes n)}{J(\rho^{(n)}_\theta)} = 1,$$

if for each $\rho^{(n)}_\theta$ the following are true:

- There exists a quantum measurement $\{T_i^{(n)}\}$ whose number of measurement outcomes is $r_n$ such that

$$\lim_{n \to \infty} \frac{F(\rho^{(n)}_\theta, \{T_i^{(n)}\})}{J(\rho^{(n)}_\theta)} = 1, \quad \text{and} \quad \lim_{n \to \infty} \frac{\log r_n}{n} < C(\mathcal{M}),$$

where $\log$ is the binary logarithm and $C(\mathcal{M})$ is the classical capacity of the quantum-classical channel $\mathcal{M}(\cdot) = \sum_i \text{Tr}(\cdot | i\rangle \langle i|_C) \{|i\rangle\langle i|_C\}$ is an orthonormal basis of an auxiliary system $\mathcal{H}_C$.

- The regularity conditions are satisfied:

  1. When $\partial_\theta \lambda_i \neq 0$, $\lambda_i = 1/e^{\lambda_i}$, where $\lambda_i := \text{Tr}(\rho^{(n)}_\theta T_i^{(n)})$ and $\{T_i^{(n)}\}$ is defined above.

**Proof of Theorem 13.** We first choose an $\alpha$ such that $\lim_{n \to \infty} \log r_n/n < \alpha < C(\mathcal{M})$. According to the definition of the classical capacity of quantum channels [54], for any $\epsilon > 0$, there exists an $n_0$ such that for any $n > n_0$, there exist an encoding channel $\mathcal{E}_E$ and a decoding channel $\mathcal{D}_D$ such that

$$\|\mathcal{D}_D \circ \mathcal{M}_{\mathcal{E}}^{\otimes n} \circ \mathcal{E}_E - \mathcal{D}_2^{\otimes [\alpha n]}\|_\diamond \leq \epsilon$$

**FIG. 2.** A quantum state $\rho^{(n)}_\theta$ in an $n$-partite system is estimated using $n$ identical noisy measurements acting on each subsystem, described by $\{M_i\}^\otimes n$. The QPFI can approach the QFI in the asymptotic limit $n \to \infty$ if the sufficient condition in Theorem 13 is satisfied. The optimal control is the composition of a quantum-classical channel $\mathcal{T}_n(\cdot) = \sum_{i=1}^n \text{Tr}(\cdot | T_i^{(n)}\rangle \langle T_i^{(n)}|) \{|i\rangle \langle i| \}$ where the measurement $\{T_i^{(n)}\}$ is asymptotically QFI-attainable, and an encoding channel $\mathcal{E}_E$ chosen as the optimal encoding channel for $\mathcal{M}_{\mathcal{E}}^{\otimes n}$ from the HSW theorem. Note that the decoding channel $\mathcal{D}_D$ from the HSW theorem only needs to be performed classically in the data processing step.

(2) $\Omega(1) \leq J(\rho^{(n)}_\theta) \leq e^{\lambda(n)}$. Theorem 13 provides a sufficient condition to attain the QFI using noisy measurements in the asymptotic limit $n \to \infty$. We will first provide a proof of Theorem 13, and return to the physical understandings of the sufficient condition later. Readers who are not interested in the technical details can skip the technical proof and advance to the discussion part.

In the proof, we will show that there exists a quantum-classical channel $\mathcal{T}_n$ defined using $\{T_i^{(n)}\}$, and an encoding channel $\mathcal{E}_E$, such that $F(\mathcal{E}_E \circ \mathcal{T}_n(\rho^{(n)}_\theta), \{M_i\}^\otimes n)$ approaches $J(\rho^{(n)}_\theta)$ asymptotically (see Fig. 2). Intuitively speaking, the first step $\mathcal{T}_n(\cdot)$ is to simulate the (asymptotically) QFI-attainable measurement $\{T_i^{(n)}\}$ on $\rho^{(n)}_\theta$ to transform it into a classically mixed state $\mathcal{T}_n(\rho^{(n)}_\theta)$ such that $J(\mathcal{T}_n(\rho^{(n)}_\theta)) = J(\rho^{(n)}_\theta)$. The second step is to choose a suitable encoding channel $\mathcal{E}_E$ such that the classical information in $\mathcal{T}_n(\rho^{(n)}_\theta)$ is fully preserved under $\mathcal{M}$, i.e., $J(\mathcal{M}\circ \mathcal{E}_E \circ \mathcal{T}_n(\rho^{(n)}_\theta)) \approx J(\mathcal{T}_n(\rho^{(n)}_\theta))$, leading to the asymptotic attainability of the QFI. Here $\mathcal{E}_E$, along with a corresponding decoding channel $\mathcal{D}_D$, is chosen such that $\mathcal{E}_E \circ \mathcal{M}_{\mathcal{E}}^{\otimes n} \circ \mathcal{D}_D$ is asymptotically equal to a completely dephasing channel with a transmission rate $\approx C(\mathcal{M})$, which is guaranteed to exist using the HSW theorem [49, 50].

**Proof of Theorem 13.** We first choose an $\alpha$ such that $\lim_{n \to \infty} \log r_n/n < \alpha < C(\mathcal{M})$. According to the definition of the classical capacity of quantum channels [54], for any $\epsilon > 0$, there exists an $n_0$ such that for any $n > n_0$, there exist an encoding channel $\mathcal{E}_E$ and a decoding channel $\mathcal{D}_D$ such that

$$\|\mathcal{D}_D \circ \mathcal{M}_{\mathcal{E}}^{\otimes n} \circ \mathcal{E}_E - \mathcal{D}_2^{\otimes [\alpha n]}\|_\diamond \leq \epsilon$$
where $D_2$ is a completely dephasing qubit channel acting on qubit Hilbert space $\mathcal{H}_n$, i.e., $D_2(\cdot) = |0\rangle \langle 0| + |1\rangle \langle 1|$, and $\|\cdot\|$ is the diamond norm of a quantum channel [54] defined by $\|\Phi\|_\diamond = \max\{\|\Phi \otimes I(X)\|_1, \|X\|_1 \leq 1\}$ ($\Phi$ and $I$ act on systems of the same dimension, and $\|\cdot\|$ denotes the trace norm). Moreover, $\epsilon = e^{-\Theta(n)}$ (see a proof in Appx. G).

For any operator $\sigma$, we have

$$\|\Xi_D \circ \mathcal{M} \otimes n \circ \Xi_E(\sigma) - D_2(\otimes n)\|_1 \leq \epsilon \|\sigma\|_1. \quad (88)$$

We also assume $n_0$ is large enough such that for any $n > n_0$, $r_n \leq 2^{[an]}$.

Let $T_n(\cdot) := \sum_{i=1}^{r_n} \text{Tr}(\rho^{(n)}_i T_i^{(n)} | \epsilon_i \rangle \langle \epsilon_i |)$ where we choose $\{|\epsilon_i\rangle\}_{i=1}^{r_n}$ to be a subset of the computational basis in $\mathcal{H}_n^{\otimes [an]}$. Without loss of generality, we assume $\lambda_i = \text{Tr}(\rho^{(n)}_i T_i^{(n)}) > 0$ for all $i$ (we can always exclude the terms that are equal to zero), then $T_n(\rho^{(n)}_0) := \sum_{i=1}^{r_n} \lambda_i |\epsilon_i\rangle \langle \epsilon_i |$ and

$$F(\rho^{(n)}_0, \{T_i^{(n)}\}) \geq F(\Xi_D \circ T_n(\rho^{(n)}_0), \{M_i\} \otimes n) = J(\mathcal{M} \otimes n \circ T_n(\rho^{(n)}_0)) \geq J(\Xi_D \circ \mathcal{M} \otimes n \circ \Xi_E \circ T_n(\rho^{(n)}_0)), \quad (89)$$

where we use the monotonicity of the QFI in the first and third inequalities and $J(\{p_{i,\theta}\}) = J(\sum_{i} p_{i,\theta} |i\rangle \langle i|)$ for any classical probability distribution $\{p_{i,\theta}\}$ in the second equality. Then we have

$$J(\Xi_D \circ \mathcal{M} \otimes n \circ \Xi_E \circ T_n(\rho^{(n)}_0)) \leq \frac{F(\rho^{(n)}_0, \{T_i^{(n)}\})}{F(\rho^{(n)}_0, \{T_i^{(n)}\})} \leq 1. \quad (90)$$

Next we aim to show $\frac{J(\Xi_D \circ \mathcal{M} \otimes n \circ \Xi_E \circ T_n(\rho^{(n)}_0))}{F(\rho^{(n)}_0, \{T_i^{(n)}\})}$ is lower bounded by a constant that approaches 1 for large $n$. First, assume $n > n_0$, we have

$$J(D_2^{[an]} \circ T_n(\rho^{(n)}_0)) = J(T_n(\rho^{(n)}_0)) = F(\rho^{(n)}_0, \{T_i^{(n)}\}) = \sum_{i=1}^{r_n} \frac{(\partial \rho_i \lambda_i)^2}{\lambda_i^2}, \quad (91)$$

where we use $D_2^{[an]}(\{\epsilon_i\}) = |\epsilon_i\rangle \langle \epsilon_i |$ in the first equality. On the other hand, consider

$$J(D' \circ \Xi_D \circ \mathcal{M} \otimes n \circ \Xi_E \circ T_n(\rho^{(n)}_0)) = \sum_{i=1}^{r_n} \frac{(\partial \rho_i \lambda_i)^2}{\eta_i} \leq J(\Xi_D \circ \mathcal{M} \otimes n \circ \Xi_E \circ T_n(\rho^{(n)}_0)) \quad (92)$$

where $D' \circ \Xi_D \circ \mathcal{M} \otimes n \circ \Xi_E \circ T_n(\rho^{(n)}_0) = \sum_{i=1}^{r_n} (\partial \rho_i \lambda_i)^2 \eta_i$. We will also assume $n$ is large enough such that $\delta_i < \lambda_i$, which is possible due to Eq. (88) and the regularity condition (1).

Then we have

$$\sum_{i=1}^{r_n} \frac{(\partial \rho_0 \lambda_i)^2}{\eta_i} = \sum_{i=1}^{r_n} \frac{(\partial \rho_0 \lambda_i + \partial \rho_0 \delta_i)^2}{\lambda_i^2 + \delta_i^2} = \sum_{i=1}^{r_n} (\partial \rho_0 \lambda_i + \partial \rho_0 \delta_i)^2 \left( \frac{1}{\lambda_i^2} - \frac{\delta_i}{\lambda_i^2(1 + \lambda_i^2)} \right) \geq \sum_{i=1}^{r_n} ((\partial \rho_0 \lambda_i)^2 + 2 \partial \rho_0 \lambda_i \partial \rho_0 \delta_i) \frac{1}{\lambda_i} \left( 1 - \frac{|\delta_i|}{\lambda_i} \right) \geq F(\rho^{(n)}_0, \{T_i^{(n)}\}) - \sum_{i=1}^{r_n} \frac{(\partial \rho_0 \lambda_i)^2}{\lambda_i} |\delta_i| - \sum_{i=1}^{r_n} \frac{2 |\partial \rho_0 \lambda_i \partial \rho_0 \delta_i|}{\lambda_i} \quad (93)$$

In the second equality above, we use the Taylor expansion $\frac{1}{1 + \xi} = 1 - \frac{\delta_i}{\lambda_i} \xi$ for some $\xi_i \in [0, \delta_i / \lambda_i]$. In the last inequality above, we use Eq. (88) to derive that $\sum_i |\delta_i| \leq \epsilon \|\rho^{(n)}_0\|_1 = \epsilon$ and

$$\sum_i |\partial \rho_0 \delta_i| \leq \epsilon |\partial \rho_0 T_n(\rho^{(n)}_0)| \leq 1 \leq \epsilon |J(T_n(\rho^{(n)}_0))|^{1/2} \leq \epsilon J(\rho^{(n)}_0, \{T_i^{(n)}\})^{1/2} \quad (94)$$

Here we use the inequality $|\partial \rho_0 \sigma| \leq J(\sigma_0)^{1/2}$ for any classically mixed state $\sigma_0 = \sum_i \mu_i |i\rangle \langle i|$, which is true because $(\sum_i |\partial \rho_0 \mu_i|)^2 \leq \sum_i (\partial \rho_0 \mu_i)^2 / \mu_i$ from the Cauchy–Schwarz inequality. Note that it also holds that $|\partial \rho_0 \sigma| \leq J(\sigma_0)^{1/2}$ for general mixed states $\sigma_0$ [67, 68].

Finally, from the monotonicity of the QFI (i.e., $\sum_i (\partial \rho_0 \lambda_i)^2 \leq J(\rho^{(n)}_0)$) and the regularity conditions (1) and (2), we have

$$\max_i \left( \frac{(\partial \rho_0 \lambda_i)^2}{\lambda_i^2} + 2 \frac{\partial \rho_0 \lambda_i}{\lambda_i} |\partial \rho_0| \right) \leq J(\rho^{(n)}_0, \{T_i^{(n)}\}) \quad (96)$$

Taking the limit $n \to \infty$ in Eq. (93), from $\epsilon = e^{-\Theta(n)}$ and Eq. (95), we have

$$\lim_{n \to \infty} \frac{1}{F(\rho^{(n)}_0, \{T_i^{(n)}\})} \sum_{i=1}^{r_n} \frac{(\partial \rho_0 \eta_i)^2}{\eta_i} \geq 1. \quad (96)$$

Combining Eq. (90), Eq. (92), and Eq. (96), we have

$$\lim_{n \to \infty} \frac{J(\Xi_D \circ \mathcal{M} \otimes n \circ \Xi_E \circ T_n(\rho^{(n)}_0))}{F(\rho^{(n)}_0, \{T_i^{(n)}\})} = 1. \quad (97)$$

Since $\lim_{n \to \infty} \frac{F(\rho^{(n)}_0, \{T_i^{(n)}\})}{J(\rho^{(n)}_0, \{T_i^{(n)}\})} = 1$ and

$$F(\rho^{(n)}_0, \{M_i\} \otimes n) \geq J(\Xi_D \circ \mathcal{M} \otimes n \circ \Xi_E \circ T_n(\rho^{(n)}_0)), \quad (98)$$
we must have
\[
\lim_{n \to \infty} \frac{F^n \left( \rho_\theta^{(n)}, \{ M_i \}^{\otimes n} \right)}{J(\rho_\theta^{(n)})} = 1, \tag{99}
\]
proving the theorem.

\[\Box\]

B. Discussion

Now we explain the physical meaning of the sufficient condition in Theorem 13 and discuss the relevant situations where it is satisfied. We will see that the sufficient condition is satisfied for a generic class of quantum states \(\rho_\theta^{(n)}\) and noisy measurements \(\{ M_i \}\).

Let us first explain the meaning of the condition Eq. (86). It states that there exists an (asymptotically) QFI-attainable measurement for \(\rho_\theta^{(n)}\) that has a small number of measurement outcomes. Specifically, the number of measurement outcomes \(r_n\) should be smaller than \(2^{C(M)n}\) (asymptotically) where \(C(M)\) is the classical capacity of the quantum measurement \(\{ M_i \}\) under consideration, i.e., Theorem 13 applies when
\[
\log r_n \lesssim C(M)n. \tag{100}
\]

The requirement (Eq. (100)) is satisfied by many practically relevant quantum states and measurements. In fact, for any non-trivial quantum measurement with a positive classical capacity, Eq. (100) is satisfied whenever \(r_n\) is subexponential. For simplicity, in the following we will call a quantum measurement that has a subexponential (i.e., \(e^{\Theta(n)}\)) number of measurement outcomes a coarse measurement.

Now we provide several typical examples where the (asymptotically) QFI-attainable coarse measurement exists and Theorem 13 applies. See Appx. A for additional details.

1. Low-rank states. For pure states, it was known that there exist 2-outcome QFI-attainable measurements [10]. (Note that [44] contains another proof of Theorem 13 when \(\rho_\theta^{(n)}\) is pure.) More generally, any \(\rho_\theta^{(n)}\) that is supported on a subspace with a subexponential dimension also has a QFI-attainable coarse measurement.

2. Symmetric states. The second example with a QFI-attainable coarse measurement is symmetric (permutation-invariant) state (e.g., tensor product of \(n\) identical mixed states). According to the Schur–Weyl duality [69, 70], \(\mathcal{H}_n = (\mathbb{C}^D)^{\otimes n}\) can be decomposed as \(\bigoplus_\nu (\mathcal{H}_n(\mathbb{U}(D)) \otimes \mathcal{H}_n(S_n))\), where \(\mathcal{H}_n(\mathbb{U}(D))\) and \(\mathcal{H}_n(S_n)\) are irreducible representation spaces of the unitary group \(\mathbb{U}(D)\) and the permutation group \(S_n\) with index \(\nu\). Any symmetric state \(\rho_\theta^{(n)}\) can be written as
\[
\rho_\theta^{(n)} = \bigoplus_\nu \left( p_\nu \rho_\nu^{(n)} \otimes \frac{1_\nu}{\dim(\mathcal{H}_n(S_n))} \right), \tag{101}
\]
where \(\rho_\nu^{(n)}\) are mixed states acting on \(\mathcal{H}_n(\mathbb{U}(D))\) and \(p_\nu\) satisfies \(\sum_\nu p_\nu = 1\) (that can be functions of \(\theta\)). Then a QFI-attainable coarse measurement \(\{ \bigoplus_\nu (T_\nu)_{ij} \otimes 1_\nu \}\) for \(\rho_\theta^{(n)}\) can be constructed from a QFI-attainable measurement \(\{ \bigoplus_\nu (T^\nu)_{ij} \}\) of \(\bigoplus_\nu p_\nu \rho_\nu^{(n)}\). Let us estimate the number of measurement outcomes: \(\nu\) corresponds to Young diagrams (i.e., partitions of \(n\) into \(D\) parts), implying the number of different indices \(\nu\) is \(O(n^{D-1})\). For any \(\nu\), \(\dim(\mathcal{H}_n(\mathbb{U}(D)))\) is equal to the number of semistandard Young tableaux, which is at most \(O(n^{D(D-1)/2})\) according to the Weyl dimension formula [71]. The number of measurement outcomes is thus upper bounded by \(\sum_\nu \dim(\mathcal{H}_n(\mathbb{U}(D))) = O(n^{D(D-1)/2})\).

3. Gibbs states. For classically mixed states \(\rho_\theta^{(n)}\), the projection onto the eigenstates of \(\rho_\theta^{(n)}\) is QFI-attainable but has exponentially many measurement outcomes. However, we argue that in many cases, coarse projections onto direct sums of eigenspaces are sufficient to attain the QFI up to the leading order, so that Theorem 13 applies. For instance, consider the Gibbs state
\[
\rho_\theta^{(n)} = \frac{1}{\sum_\nu e^{-\theta E_\nu}} \sum_\nu e^{-\theta E_\nu} |\nu\rangle \langle \nu|, \tag{102}
\]
where \(|\nu\rangle\) are energy eigenstates with eigenvalues \(\{ E_\nu \}\) and \(\theta\) is the inverse temperature to be estimated. The QFI is equal to the variance of energy, i.e.,
\[
J(\rho_\theta^{(n)}) = \sum_\nu p_\nu E_\nu^2 - \left( \sum_\nu p_\nu E_\nu \right)^2, \tag{103}
\]
where \(p_\nu = \frac{e^{-\theta E_\nu}}{\sum_\nu e^{-\theta E_\nu}}\). Assume the energy eigenvalues lie in \([0, \bar{E}]\), where \(\bar{E} = \Theta(n)\) (which is a standard assumption in condensed matter systems) and divide them into intervals \(\{ I_k = [E_k, E_{k+1}] \}_{k=1}^n\) such that \(E_0 = 0\), \(E_n = \bar{E}\) and \(E_{k+1} - E_k = \Delta E = \bar{E}/n\). Consider the coarse projection \(\{ \Pi_k \}_{k=1}^n\) onto the direct sums of eigenspaces corresponding to all eigenvalues in \(I_k\). The FI is
\[
F(\rho_\theta^{(n)}, \{ \Pi_k \}) = \sum_{k=1}^n p_k E_k^2 - \left( \sum_\nu p_\nu E_\nu \right)^2, \tag{104}
\]
where \(p_k = \sum_{\nu: E_\nu \in I_k} p_\nu\) and \(E_k = \frac{1}{p_k} \sum_{\nu: E_\nu \in I_k} p_\nu E_\nu\). Then we have \(J(\rho_\theta^{(n)}) - F(\rho_\theta^{(n)}, \{ \Pi_k \}) \leq \sum_k p_k (E_{k+1} - E_k) \leq 2\bar{E} \Delta E = \Theta(1)\). Combining with the regularity condition (2), it implies that \(F(\rho_\theta^{(n)}, \{ \Pi_k \})\) is equal to \(J(\rho_\theta^{(n)})\) up to the leading order.

Finally, let us explain the intuitions behind the regularity conditions:

1. Regularity condition (1) states that when the probability of obtaining measurement outcome \(i\) depends on \(\theta\) (i.e., \(\rho_\theta \lambda_i \neq 0\)), it must be no smaller than an inverse of a subexponential function of \(n\), that is, the probability to detect \(i\) cannot be exponentially small. This is also a
practically reasonable assumption as we would want to exclude the singular cases where an exponentially small signal provides a non-trivial contribution to the QFI.

(2) Regularity condition (2) requires that the QFI of $\rho^n(\theta)$ does not decrease with $n$ asymptotically, which should be satisfied in any practically relevant cases. It also requires the QFI to be subexponential, which is a natural assumption in quantum sensing experiments (note that the Heisenberg limit implies $J(\rho^n(\theta)) = O(n^2)$).

VIII. CONCLUSIONS AND OUTLOOK

We conducted a systematic study of the preprocessing optimization problem for noisy quantum measurements in quantum metrology. The QPFI (i.e., the FI of noisy measurement statistics optimized over all preprocessing quantum channels), that we defined and investigated in depth, sets an ultimate precision bound for noisy detection of quantum states. Our results provide, in many cases, both numerically and analytically, approaches to identifying the optimal preprocessing controls that will be of great importance in alleviating the effect of measurement noise in quantum sensing experiments.

We also considered, specifically, the asymptotic limit of the QPFI in multi-probe systems with individual measurement on each probe. We proved the convergence of the QPFI to the QFI when there exists an (asymptotically) QFI-attainable measurement with a sufficiently small number of measurement outcomes, by establishing a connection to the classical channel capacity theorem. It would be interesting to explore, in future works, if the number of outcomes for QFI-attainable measurements can be easily bounded given a quantum state.

Although we’ve discussed only two types of quantum preprocessing controls, CPTP maps and unitary maps, our biconvex formulation might be generalized to cover other more restricted types of quantum controls. We also narrowed the analytical forms of optimal controls for pure states and classically mixed states down to rotations onto the span of two basis states and coarse graining channels, respectively, but it remains open whether a simple method exists to help us determine the exact operations.

Finally, there are a few important directions to extend our results to, e.g., incorporating the state preparation optimization into the FI optimization problem and considering the preprocessing optimization in multi-parameter estimation where the incompatibility of optimal preprocessings for different parameters might become an issue.

ACKNOWLEDGMENTS

We thank Senrui Chen, Jan Kolodyński, Yaodong Li, Zi-Wen Liu, John Preskill, Tianci Zhou for helpful discussions. S.Z. acknowledges funding provided by the Institute for Quantum Information and Matter, an NSF Physics Frontiers Center (NSF Grant PHY-1733907). T.G. acknowledges funding provided by the Institute for Quantum Information and Matter, and the Quantum Science and Technology Scholarship of the Israel Council for Higher Education.
[15] BM Escher, RL de Matos Filho, and L Davidovich, “General framework for estimating the ultimate precision limit in noisy quantum-enhanced metrology,” Nature Physics 7, 406 (2011).

[16] Rafal Demkowicz-Dobrzański, Jan Kolodyński, and Mădălina Guţă, “The elusive heisenberg limit in quantum-enhanced metrology,” Nature communications 3, 1063 (2012).

[17] Haidong Yuan and Chi-Hang Fred Fung, “Fidelity and fisher information on quantum channels,” New Journal of Physics 19, 113039 (2017).

[18] Akira Sone, M Cerezo, Jacob L Beckey, and Patrick J Coles, “Generalized measure of quantum fisher information,” Physical Review A 104, 062602 (2021).

[19] Anian Altherr and Yuxiang Yang, “Quantum metrology for non-markovian processes,” Physical Review Letters 127, 060501 (2021).

[20] Dénes Petz and Csaba Sudár, “Geometries of quantum states,” Journal of Mathematical Physics 37, 2662–2673 (1996).

[21] MV Gurudev Dutt, L Childress, L Jiang, E Togan, J Maze, F Jelezo, AS Zibrov, PR Hamer, and MD Lukin, “Quantum register based on individual electronic and nuclear spin qubits in diamond,” Science 316, 1312–1316 (2007).

[22] R Hanson, VV Dobrovitski, AE Feiguin, O Gywat, and DD Awschalom, “Coherent dynamics of a single spin interacting with an adjustable spin bath,” Science 320, 352–355 (2008).

[23] JM Taylor, Paola Cappellaro, L Childress, Liang Jiang, Dmitry Budker, PR Hamer, Amir Yacoby, R Walsworth, and MD Lukin, “High-sensitivity diamond magnetometer with nanoscale resolution,” Nature Physics 4, 810–816 (2008).

[24] Liang Jiang, JS Hodges, J Maze, Peter Maurer, JM Taylor, DG Cory, PR Hamer, Ronald Lee Walsworth, Amir Yacoby, Alexander S Zibrov, et al., “Repetitive readout of a single electronic spin via quantum logic with nuclear spin ancillae,” Science 326, 267–272 (2009).

[25] Philip Krantz, Morten Kjaergaard, Fei Yan, Terry P Orlando, Simon Gustavsson, and William D Oliver, “A quantum engineer’s guide to superconducting qubits,” Applied Physics Reviews 6, 021318 (2019).

[26] AH Myerson, DJ Szwer, SC Webster, DTC Allcock, MJ Curtis, G Imreh, JA Sherman, DN Stacey, AM Steane, and DM Lucas, “High-fidelity readout of trapped-ion qubits,” Physical Review Letters 100, 200502 (2008).

[27] Lorenza Viola, Emanuel Knill, and Seth Lloyd, “Dynamical decoupling of open quantum systems,” Physical Review Letters 82, 2417 (1999).

[28] Eric M Kessler, Igor Lovechinsky, Alexander O Sushkov, and Mikhail D Lukin, “Quantum error correction for metrology,” Physical Review Letters 112, 150802 (2014).

[29] W Dürr, M Skotiniotis, Florian Froewis, and B Kraus, “Improved quantum metrology using quantum error correction,” Physical Review Letters 112, 080801 (2014).

[30] Gilad Arrad, Yuval Vinkler, Dorit Aharonov, and Alex Retzker, “Increasing sensing resolution with error correction,” Physical Review Letters 112, 150801 (2014).

[31] Jing Liu and Haidong Yuan, “Quantum parameter estimation with optimal control,” Physical Review A 96, 012117 (2017).

[32] Kaoru Yamamoto, Suguru Endo, Hideaki Hakoshima, Yuichiro Matsuzaki, and Yuuki Tokunaga, “Error-mitigated quantum metrology,” arXiv:2112.01850 (2021).

[33] Francesco Albarelli, Matteo AC Rossi, Dario Tamascelli, and Marco G Genoni, “Restoring heisenberg scaling in noisy quantum metrology by monitoring the environment,” Quantum 2, 110 (2018).

[34] Frank Verstraete, Michael M Wolf, and J Ignacio Cirac, “Quantum computation and quantum-state engineering driven by dissipation,” Nature physics 5, 633–636 (2009).

[35] Liang Jiang, Ana Maria Rey, Oriol Romero-Isat, Juan José Garcia-Ripoll, Anna Sanpera, and Mikhail D Lukin, “Preparation of decoherence-free cluster states with optical superlattices,” Physical Review A 79, 022309 (2009).

[36] Mattias T Johnsson, Nahomita Roy Mukty, Daniel Burgarth, Thomas Volz, and Gavin K Brennen, “Geometric pathway to scalable quantum sensing,” Physical Review Letters 125, 190403 (2020).

[37] Emily Davis, Gregory Bentsen, and Monika Schleier-Smith, “Approaching the heisenberg limit without single-particle detection,” Physical Review Letters 116, 053601 (2016).

[38] Florian Fröwis, Pavel Sekatski, and Wolfgang Dür, “Detecting large quantum fisher information with finite measurement precision,” Physical Review Letters 116, 090801 (2016).

[39] Tommaso Macrì, Augusto Smerzi, and Luca Pezzè, “Loschmidt echo for quantum metrology,” Physical Review A 94, 010102 (2016).

[40] Samuel P Nolan, Stuart S Szigeti, and Simon A Haine, “Optimal and robust quantum metrology using interaction-based readouts,” Physical Review Letters 119, 193601 (2017).

[41] Simon A Haine, “Using interaction-based readouts to approach the ultimate limit of detection-noise robustness for quantum-enhanced metrology in collective spin systems,” Physical Review A 98, 030303 (2018).

[42] Onur Hosten, Radha Krishnakumar, Nils J Engelsen, and Mark A Kasevich, “Quantum phase magnification,” Science 352, 1552–1555 (2016).

[43] Daniel Linnemann, Helmut Strobel, Wolfgang Mues sel, Jonas Schulz, Robert J Lewis-Swan, Karen V Khe runtsyan, and Markus K Oberthaler, “Quantum-enhanced sensing based on time reversal of nonlinear dynamics,” Physical Review Letters 117, 013001 (2016).

[44] Yink Loong Len, Tuvia Gefen, Alex Retzker, and Jan Kolodyński, “Quantum metrology with imperfect measurements,” arXiv:2109.01160 (2021).

[45] Filip B Maciejewski, Zoltán Zimborás, and Michal Oszmaniec, “Mitigation of readout noise in near-term quantum devices by classical post-processing based on detector tomography,” Quantum 4, 257 (2020).

[46] Michael R Geller, “Rigorous measurement error correction,” Quantum Science and Technology 5, 03LT01 (2020).

[47] Sergey Bravyi, Sarah Sheldon, Abhinav Kandala, David C Mckay, and Jay M Gambetta, “Mitigating measurement errors in multiquit experiments,” Physical Review A 103, 042605 (2021).

[48] Jochen Gorski, Frank Pfeuffer, and Kathrin Klannroth, “Biconvex sets and optimization with biconvex functions: a survey and extensions,” Mathematical methods of op-
erations research 66, 373–407 (2007).
[49] Alexander S Holevo, “The capacity of the quantum channel with general signal states,” IEEE Transactions on Information Theory 44, 269–273 (1998).
[50] Benjamin Schumacher and Michael D Westmoreland, “Sending classical information via noisy quantum channels,” Physical Review A 56, 131 (1997).
[51] Michael A Nielsen and Isaac L Chuang, Quantum Computation and Quantum Information (Cambridge University Press, 2010).
[52] Erich L Lehmann and George Casella, Theory of point estimation (Springer, 2006).

H. K. Mark, “The capacity of the quantum channel with general signal states,” IEEE Transactions on Information Theory 44, 269–273 (1998).

B. Schumacher and M. D. Westmoreland, “Sending classical information via noisy quantum channels,” Physical Review A 56, 131 (1997).

M. A. Nielsen and I. L. Chuang, Quantum Computation and Quantum Information (Cambridge University Press, 2010).

E. L. Lehmann and G. Casella, Theory of point estimation (Springer, 2006).

Hisashi Kobayashi, Brian L Mark, and William Turin, Probability, random processes, and statistical analysis: applications to communications, signal processing, queueing theory and mathematical finance (Cambridge University Press, 2011).

John Watrous, The theory of quantum information (Cambridge University Press, 2018).

OE Barndorff-Nielsen and RD Gill, “Fisher information in quantum statistics,” Journal of Physics A: Mathematical and General 33, 4481 (2000).

Masahito Hayashi, “Comparison between the cramer-rao and the mini-max approaches in quantum channel estimation,” Communications in Mathematical Physics 304, 689–709 (2011).

Yuxiang Yang, Giulio Chiribella, and Masahito Hayashi, “Attaining the ultimate precision limit in quantum state estimation,” Communications in Mathematical Physics 368, 223–293 (2019).

Carl Helstrom, “The minimum variance of estimates in quantum signal detection,” IEEE Transactions on Information Theory 14, 234–242 (1968).

Luca Pezzé and Augusto Smerzi, “Entanglement, nonlinear dynamics, and the heisenberg limit,” Physical review letters 102, 100401 (2009).

BM Escher, “Quantum noise-to-sensitivity ratio,” arXiv preprint arXiv:1212.2533 (2012).

Katarzyna Macieszczak, Martin Fraas, and Rafał Demkowicz-Dobrzański, “Bayesian quantum frequency estimation in presence of collective dephasing,” New Journal of Physics 16, 113002 (2014).

Christodoulos A Floudas, Deterministic global optimization: theory, methods and applications, Vol. 37 (Springer Science & Business Media, 2013).

Shmuel Friedland and Raphael Loewy, “On the extreme points of quantum channels,” Linear Algebra and its Applications 498, 553–573 (2016).

William K Wootters, “Statistical distance and hilbert space,” Physical Review D 23, 357 (1981).

Stephen J Blundell and Katherine M Blundell, Concepts in thermal physics (Oxford University Press on Demand, 2010).

Akio Fujiiwara, “Strong consistency and asymptotic efficiency for adaptive quantum estimation problems,” Journal of Physics A: Mathematical and General 39, 12489–12504 (2006).

Michal Oszmaniec, Remigiusz Augusiak, Christian Gogolin, Jan Kolodyński, Antonio Acín, and Maciej Lewenstein, “Random bosonic states for robust quantum metrology,” Physical Review X 6, 041044 (2016).

Marcin Jarzyna and Jan Kolodyński, “Geometric approach to quantum statistical inference,” IEEE Journal on Selected Areas in Information Theory 1, 367–386 (2020).

William Fulton and Joe Harris, Representation theory: a first course, Vol. 129 (Springer Science & Business Media, 2013).

Masahito Hayashi, Group representation for quantum theory (Springer, 2017).

Roe Goodman and Nolan R Wallach, Representations and invariants of the classical groups (Cambridge University Press, 2000).

Stephen Boyd, Stephen P Boyd, and Lieven Vandenberghe, Convex optimization (Cambridge university press, 2004).

Mark M Wilde, Quantum information theory (Cambridge University Press, 2013).

Wassily Hoeffding, “Probability inequalities for sums of bounded random variables,” in The collected works of Wassily Hoeffding (Springer, 1994) pp. 409–426.
Appendix A: QFI-attainable measurements

In this appendix, we provide several simple facts about the QFI-attainable measurements for quantum states, that will be useful in the main text.

(1) **Necessary and sufficient condition [10].** Given a quantum state $\rho_\theta$, $\{M_i\}$ attains the QFI if and only if

$$\forall i=1, \ldots, r, \exists \epsilon_i \in \mathbb{R}, \text{ s.t. } M_i^{1/2} \rho_\theta^{1/2} = c_i M_i^{1/2} L_\theta \rho_\theta^{1/2},$$

where $L_\theta$ is the symmetric logarithmic derivative (SLD) operator which is a Hermitian operator defined by

$$\partial L_\theta = \frac{1}{2} \left( L_\theta \rho_\theta + \rho_\theta L_\theta \right)$$

and the equality holds if and only if $\partial \rho_\theta \neq 0$. The non-negativity follows from the faithfulness of the classical FI. In order to see when $\partial_\theta \rho_\theta \neq 0$, $F^E(\rho_\theta, \{M_i\}) > 0$, we use Lemma 12 in Sec. VI. Assume $\partial_\theta \rho_\theta \neq 0$. $F^E(\rho_\theta, \{M_i\}) > F^E(T(\rho_\theta), \{M_i\})$, where $T(\rho_\theta)$ is a classically mixed state that satisfies $J(T(\rho_\theta)) = J(\rho_\theta) > 0$ (from the faithfulness of the QFI). Consider the following simplification of measurements: dividing $\{M_i\}$ into two subsets and restricting them in a two-dimensional subspace such that the measurement becomes a non-trivial binary measurement on a qubit $\{\hat{M}, 1 - \hat{M}\}$. Then we see that $F^E(T(\rho_\theta), \{M_i\}) \geq F^E(T(\rho_\theta), \{\hat{M}, 1 - \hat{M}\}) > 0$ from the exact expression of the QFI in Eq. (75).

(2) **Low-rank states.** If $\rho_\theta$ and $\partial_\theta \rho_\theta$ are supported on a $D'$-dimensional subspace of $\mathcal{H}_S$ ($D' \leq D$) (which should be true when $\rho_\theta$ is supported on a $[D'/2]$-dimensional subspace), there exists an SLD $L_\theta$ supported on the $D'$-dimensional subspace. Moreover, let $L_\theta = \sum_{i=1}^{D'} \ell_i |i\rangle \langle i|$, where $\ell_i > 0$ and $\{|i\rangle\}$ is an orthonormal basis for the $D'$-dimensional subspace, it can be verified that $\{M_i\}_{i=1}^{D'}$ (that has $D'$ measurement outcomes) is QFI-attainable when $M_i = |i\rangle \langle i|$ for $i = 1, \ldots, D' - 1$ and $M_D = \mathbb{1} - \sum_{i=1}^{D'} M_i$. That means, a low-rank state must have a QFI-attainable measurement with a small number of measurement outcomes. In particular, for pure states, binary measurements are sufficient to attain the QFI.

(3) **Classically mixed states.** For classically mixed states $\rho_\theta = \sum_{i=1}^{D} \lambda_i |i\rangle \langle i|$, the SLD is also diagonal in the basis $\{|i\rangle\}$. It implies that the rank-one projection onto the basis is QFI-attainable for classically mixed states.

(4) **Block-diagonal states.** More generally, consider block-diagonal states $\rho_\theta = \bigoplus pF\rho_\theta$, where $\rho_\nu, \theta$ are supported on different orthogonal subspaces. The SLD can also be block-diagonal, implying that there exists a QFI-attainable measurement of the form $\bigoplus pF(\rho_\nu, \theta)$, whose number of measurement outcomes is at most the rank of $\rho_\theta$, where $\{|T_\nu\rangle\}$ is a quantum measurement for each fixed $\nu$.

Appendix B: Mathematical properties of the QFI

Here we briefly discuss the mathematical properties of the QFI. We will always assume quantum measurements are non-trivial in this appendix (i.e., $3M_i \not\propto \mathbb{1}$ for all $\{M_i\}$).

(1) **Faithfulness.** The QFI is faithful, that means,

$$F^E(\rho_\theta, \{M_i\}) \geq 0,$$

and the equality holds if and only if $\partial_\theta \rho_\theta \neq 0$. The non-negativity follows from the faithfulness of the classical FI. In order to see when $\partial_\theta \rho_\theta \neq 0$, $F^E(\rho_\theta, \{M_i\}) > 0$, we use Lemma 12 in Sec. VI. Assume $\partial_\theta \rho_\theta \neq 0$. $F^E(\rho_\theta, \{M_i\}) > F^E(T(\rho_\theta), \{M_i\})$, where $T(\rho_\theta)$ is a classically mixed state that satisfies $J(T(\rho_\theta)) = J(\rho_\theta) > 0$ (from the faithfulness of the QFI). Consider the following simplification of measurements: dividing $\{M_i\}$ into two subsets and restricting them in a two-dimensional subspace such that the measurement becomes a non-trivial binary measurement on a qubit $\{\hat{M}, 1 - \hat{M}\}$. Then we see that $F^E(T(\rho_\theta), \{M_i\}) \geq F^E(T(\rho_\theta), \{\hat{M}, 1 - \hat{M}\}) > 0$ from the exact expression of the QFI in Eq. (75).

(2) **Monotonicity.** The QFI is monotonic, i.e., $F^E(\rho_\theta, \{M_i\}) \geq F^E(\mathcal{E}(\rho_\theta), \{M_i\})$ for any CPTP map $\mathcal{E}(\cdot)$ by definition. Or equivalently, $F^E(\rho_\theta, \{M_i\}) \geq F^E(\rho_\theta, \{\mathcal{E}'(M_i)\})$ where $\mathcal{E}'(\cdot)$ is the dual map of $\mathcal{E}(\cdot)$.

(3) **Convexity.** The QFI is convex in quantum states when $\{M_i\}$ is fixed. That is, for $p \in (0, 1)$ independent of $\theta$,

$$F^E(p\rho_\theta + (1-p)\sigma_\theta, \{M_i\}) = \sup_{\mathcal{E}} F(\mathcal{E}(p\rho_\theta + (1-p)\sigma_\theta, \{M_i\}$$

$$\leq \sup_{\mathcal{E}} pF(\mathcal{E}(\rho_\theta, \{M_i\}) + (1-p)F(\mathcal{E}(\sigma_\theta, \{M_i\})$$

$$\leq pF^E(\rho_\theta, \{M_i\}) + (1-p)F^E(\sigma_\theta, \{M_i\}),$$

where we use the convexity of the classical FI in the second step. Similarly, we also have

$$F^E(p\rho_\theta, \{pM_i + (1-p)M_i'\}) \leq pF^E(\rho_\theta, \{M_i\}) + (1-p)F^E(\rho_\theta, \{M_i'\}).$$
(4) Additivity. When quantum states under consideration are pure, the QPFI is additive because of Eq. (32) and the additivity of the QFI, i.e., $F^p(\psi_\theta \otimes \psi'_\theta, \{M_i\}) = F^p(\psi_\theta, \{M_i\}) + F^p(\psi'_\theta, \{M_i\})$ when $\psi_\theta$ and $\psi'_\theta$ are pure.

For general mixed states, there is no general inequality relation between $F^p(\rho_\theta \otimes \sigma_\theta, \{M_i\})$ and $F^p(\rho_\theta, \{M_i\}) + F^p(\sigma_\theta, \{M_i\})$. Consider the simple example where one (or two identical) qubit state $\rho_\theta = \cos^2 \theta |1\rangle \langle 1| + \sin^2 \theta |2\rangle \langle 2|$ ($\theta \in (0, \pi/2)$), is measured using the binary measurement on a qubit: $M_1 = (1 - \eta) |1\rangle \langle 1| + \eta |2\rangle \langle 2|$ and $M_2 = I - M_1 (0 < \eta < 1/2)$. We have, from Eq. (75), that

$$F^p(\rho_\theta, \{M_i\}) = \frac{4(1 - 2\eta)^2 \sin(2\theta)^2}{1 - (1 - 2\eta) \cos(2\theta)^2},$$

(B6)

and

$$F^p(\rho_\theta \otimes \rho_\theta, \{M_i\}) = \max \left\{ \frac{16(1 - 2\eta)^2 \cos(\theta)^6 \sin(\theta)^2}{(1 - \eta) \eta + (1 - 2\eta)^2 (\cos \theta)^4 (1 - (\cos \theta)^4)}, \frac{16(1 - 2\eta)^2 \sin(\theta)^2 \cos(\theta)^6}{(1 - \eta) \eta + (1 - 2\eta)^2 (\sin \theta)^4 (1 - (\sin \theta)^4)} \right\}.$$ (B7)

Consider the limit $\eta \to 0$, we have $F^p(\rho_\theta \otimes \rho_\theta, \{M_i\}) < 2F^p(\rho_\theta, \{M_i\})$ (which is expected because $\{M_i\}$ is QFI-attainable for $\rho_\theta$ and the QFI is additive). On the other hand, one can immediately find cases where $F^p(\rho_\theta \otimes \rho_\theta, \{M_i\}) > 2F^p(\rho_\theta, \{M_i\})$, e.g., when $\eta = 0.1$ and $\theta = \pi/8$. For any fixed $\eta > 0$, there is a threshold of $\theta$ above which the sign of $F^p(\rho_\theta \otimes \rho_\theta, \{M_i\}) - 2F^p(\rho_\theta, \{M_i\})$ changes from positive to negative.

Finally, we can consider multiple states under multiple measurements. We have, by definition, $F^p(\rho_\theta \otimes \sigma_\theta, \{M_i\} \otimes \{M'_i\}) \geq F^p(\rho_\theta, \{M_i\}) + F^p(\sigma_\theta, \{M'_i\})$ and the inequality can be strict (see the convergence to the QFI in the asymptotic limit in Sec. VII).

### Appendix C: Attainability of the QPFI

Here we prove several results that are related to the attainability of the QPFI and the QUPFI.

We first show the existence of the optimal controls (or unitaries) for generic noisy measurement that has non-zero noise in all subspaces.

**Lemma S1.** For arbitrary quantum states $\rho_\theta$ and noisy measurements $\{M_i\}$ such that $\min_i \lambda_{\min}(M_i) > 0$, the suprema in Eq. (7) and Eq. (10) are attainable. Here $\lambda_{\min}(\cdot)$ represents the minimum eigenvalue of an operator.

**Proof.** By definition, there exists a sequence of quantum channels $(\mathcal{E}_1, \cdots, \mathcal{E}_n, \cdots)$ such that

$$F(\mathcal{E}_n(\rho_\theta), \{M_i\}) \geq F^p(\rho_\theta, \{M_i\}) - \eta_n,$$ (C1)

where $\lim_{n \to \infty} \eta_n = 0$. Since the set of quantum channels is bounded and closed, there exists a limiting point of $(\mathcal{E}_1, \cdots, \mathcal{E}_n, \cdots)$ that we denote by $\mathcal{E}$. Without loss of generality, we assume the sequence converges and $\mathcal{E} = \lim_{n \to \infty} \mathcal{E}_n$. And

$$\lim_{n \to \infty} F(\mathcal{E}_n(\rho_\theta), \{M_i\}) = \lim_{n \to \infty} \sum_i \frac{(\text{Tr}(\mathcal{E}_n(\rho_\theta)M_i))^2}{\text{Tr}(\mathcal{E}_n(\rho_\theta)M_i)}$$ (C2)

$$= \sum_i \frac{\lim_{n \to \infty} \text{Tr}(\mathcal{E}_n(\rho_\theta)M_i))^2}{\lim_{n \to \infty} \text{Tr}(\mathcal{E}_n(\rho_\theta)M_i)} = \sum_i \frac{(\text{Tr}(\mathcal{E}(\rho_\theta)M_i))^2}{\text{Tr}(\mathcal{E}(\rho_\theta)M_i)} = F(\mathcal{E}(\rho_\theta), \{M_i\}),$$ (C3)

where we use $\text{Tr}(\mathcal{E}_n(\rho_\theta)M_i) > \min_i \lambda_{\min}(M_i) > 0$ for all $n$. Then we must have $F(\mathcal{E}(\rho_\theta), \{M_i\}) \geq F^p(\rho_\theta, \{M_i\})$ using Eq. (C1). Since $F(\mathcal{E}(\rho_\theta), \{M_i\}) \leq F^p(\rho_\theta, \{M_i\})$ by definition, we have

$$F(\mathcal{E}(\rho_\theta), \{M_i\}) = F^p(\rho_\theta, \{M_i\}),$$ (C5)

proving the existence of the optimal channels. The existence of the optimal unitaries can also be proven analogously. \qed
For any quantum state \( \rho_0 \), quantum measurement \( \{ M_i \} \) and \( \eta > 0 \), there always exists \( \{ M_i^{(\epsilon)} \} \) and a constant \( c > 0 \) such that the corresponding QPFI and the QUPFI are attainable, and when \( \epsilon < c \),

\[
F^P(\rho_0, \{ M_i^{(\epsilon)} \}) \geq F^P(\rho_0, \{ M_i \}) - \eta, \tag{C6}
\]

\[
F^U(\rho_0, \{ M_i^{(\epsilon)} \}) \geq F^U(\rho_0, \{ M_i \}) - \eta. \tag{C7}
\]

**Proof.** Assume \( F^P(\rho_0, \{ M_i \}) - \eta \geq F^U(\rho_0, \{ M_i \}) - \eta > 0 \). By definition, we can pick \( E \) and \( U \) such that

\[
F(E(\rho_0), \{ M_i \}) \geq F^P(\rho_0, \{ M_i \}) - \eta/2, \tag{C8}
\]

\[
F(U \rho_0 U^\dagger, \{ M_i \}) \geq F^U(\rho_0, \{ M_i \}) - \eta/2. \tag{C9}
\]

Let

\[
c' = \min \left\{ \min_{i : \text{Tr}(E(\rho_0) M_i) \neq 0} \frac{\text{Tr}(E(\rho_0) M_i)}{\text{Tr}(M_i)}, \min_{i : \text{Tr}(U(\rho_0 U^\dagger) M_i) \neq 0} \frac{\text{Tr}(U(\rho_0 U^\dagger) M_i)}{\text{Tr}(M_i)} \right\}, \tag{C10}
\]

define (assuming \( d = \text{dim}(\mathcal{H}_{S'}) \))

\[
M_i^{(\epsilon)} = (1 - \epsilon) M_i + \epsilon \text{Tr}(M_i) \frac{I}{d}, \tag{C11}
\]

and assume \( \epsilon \) is small enough such that

\[
1 - \epsilon - \frac{\epsilon}{dc}(1 - \epsilon) > \frac{F^U(\rho_0, \{ M_i \}) - \eta}{F^U(\rho_0, \{ M_i \}) - \eta/2}. \tag{C12}
\]

Using Lemma S1, it is clear that the QPFI and the QUPFI for \( \{ M_i^{(\epsilon)} \} \) are attainable. Furthermore, we have

\[
F^P(\rho_0, \{ M_i^{(\epsilon)} \}) \geq \sum_{i : \text{Tr}(E(\rho_0) M_i^{(\epsilon)}) \neq 0} \frac{(\text{Tr}(E(\rho_0) M_i^{(\epsilon)}))^2}{\text{Tr}(E(\rho_0) M_i^{(\epsilon)})} \geq \sum_{i : \text{Tr}(E(\rho_0) M_i^{(\epsilon)}) \neq 0} \frac{(1 - \epsilon)^2 (\text{Tr}(E(\rho_0) M_i))^2}{(1 - \epsilon) \text{Tr}(E(\rho_0) M_i) + \text{Tr}(M_i) \epsilon/d} \tag{C13}
\]

\[
\geq \sum_{i : \text{Tr}(E(\rho_0) M_i) \neq 0} \frac{(\text{Tr}(E(\rho_0) M_i))^2}{\text{Tr}(E(\rho_0) M_i)} \left( 1 - \epsilon - \frac{\epsilon}{dc}(1 - \epsilon) \right) \tag{C14}
\]

\[
\geq F(E(\rho_0), \{ M_i \}) \frac{F^U(\rho_0, \{ M_i \}) - \eta}{F^U(\rho_0, \{ M_i \}) - \eta/2} \geq F^P(\rho_0, \{ M_i \}) - \eta, \tag{C15}
\]

proving Eq. (C6). Eq. (C7) is also true, similarly. When \( F^P(\rho_0, \{ M_i \}) - \eta \leq 0 \) or \( F^U(\rho_0, \{ M_i \}) - \eta \leq 0 \), the results also follow trivially. \( \square \)

Finally, we are ready to provide a proof of Theorem 1, which shows a way to calculate the QPFI by considering the limit of the QPFI for a set of generic noisy measurements in its neighborhood. (Note that the theorem stated below also holds for the QUPFI.)

**Theorem 1.** Let \( M_i^{(\epsilon)} = (1 - \epsilon) M_i + \epsilon \text{Tr}(M_i) \frac{I}{d} \) where \( d = \text{dim}(\mathcal{H}_{S'}) \) and \( 0 < \epsilon < 1 \). Then

\[
F^P(\rho_0, \{ M_i \}) = \lim_{\epsilon \to 0^+} F^P(\rho_0, \{ M_i^{(\epsilon)} \}), \tag{9}
\]

and the QPFI \( F^P(\rho_0, \{ M_i^{(\epsilon)} \}) \) is attainable for any \( \epsilon \in (0, 1] \).

**Proof.** For any \( \eta > 0 \), following Lemma S2, we have

\[
F^P(\rho_0, \{ M_i \}) \leq F^P(\rho_0, \{ M_i^{(\epsilon)} \}) + \eta, \tag{C16}
\]
when $\epsilon$ is small enough, where $M_i^{(\epsilon)} = (1 - \epsilon)M_i + \epsilon \frac{1}{d}$. Take the limit $\epsilon \to 0^+$ on both sides, we have $F^p(\rho_0, \{M_i\}) \leq \liminf_{\epsilon \to 0^+} F^p(\rho_0, \{M_i^{(\epsilon)}\}) + \eta$ for any $\eta > 0$, implying
\[
F^p(\rho_0, \{M_i\}) \leq \liminf_{\epsilon \to 0^+} F^p(\rho_0, \{M_i^{(\epsilon)}\}).
\] (C17)

On the other hand, consider a quantum channel for $0 < \epsilon < 1$,
\[
E_\epsilon(\sigma) = (1 - \epsilon)\sigma + \epsilon \text{Tr}(\sigma) \frac{1}{d}.
\] (C18)
Then we have $\text{Tr}(E_\epsilon(\sigma)M_i) = \text{Tr}(\sigma E_\epsilon(M_i)) = \text{Tr}(\sigma M_i^{(\epsilon)})$ for any $\sigma$. By definition, we have
\[
F^p(\rho_0, \{M_i\}) = \sup_{\epsilon} F(\mathcal{E}(\rho_0), \{M_i\}) \geq \sup_{\epsilon} F(E_\epsilon(\mathcal{E}(\rho_0)), \{M_i\}) = F^p(\rho_0, \{E_\epsilon(M_i)\}) = F^p(\rho_0, \{M_i^{(\epsilon)}\}).
\] (C19)
Take the limit $\epsilon \to 0^+$ on both sides, we have
\[
F^p(\rho_0, \{M_i\}) \geq \limsup_{\epsilon \to 0^+} F^p(\rho_0, \{M_i^{(\epsilon)}\}).
\] (C20)
The theorem then follows from Eq. (C17) and Eq. (C20).

**Appendix D: Global optimization algorithm for biconvex optimization problems**

In Sec. III, we showed that the QPFI can be obtained from the following biconvex optimization problem (Eq. (19)):
\[
\begin{align*}
F^p(\rho_0, \{M_i\})^{-1} = \inf_{(x, \Omega)} \text{Tr}((X_2 \otimes \rho_0^T)\Omega), \\
\text{s.t.} \quad & \Omega \succeq 0, \\
& \text{Tr}_{S'}(\Omega) = 1, \text{Tr}
\end{align*}
\] (D1)
The constraints on $\Omega$ guarantee any feasible $\Omega$ is contained in a convex compact set $R_2$ (the absolute value of each entry of $\Omega$ should not be larger than $\dim(\mathcal{H}_S)$). We could also set a convex compact region $R_1$ on $x$, so that the following optimization problem generates the same optimal value as Eq. (19).
\[
\begin{align*}
\min_{(x, \Omega)} \text{Tr}((X_2 \otimes \rho_0^T)\Omega), \\
\text{s.t.} \quad & \Omega \succeq 0, \\
& \text{Tr}_{S'}(\Omega) = 1, \text{Tr}((X \otimes \rho_0^T)\Omega) = 0, \text{Tr}((X \otimes \partial_0 \rho_0^T)\Omega) = 1, \\
& x \in R_1, \Omega \in R_2.
\end{align*}
\] (D2)
As discussed in Sec. III, this is possible when the size of $R_1$ is sufficiently large, in normal cases when the infimum in Eq. (19) is attainable. Otherwise, the optimal value of Eq. (D2) can still approach that of Eq. (19) for sufficiently large size of $R_1$.

Here we describe the global optimization algorithm [62] for Eq. (D2) that is guaranteed to converge to the global optimum of Eq. (D2) in finite steps. One may seek [48] for a general survey on algorithms from biconvex optimization.

We first rewrite Eq. (D2) as
\[
\begin{align*}
\min_{(x, \Omega) \in R_1 \times R_2} f(x, \Omega), \\
\text{s.t.} \quad & \Omega \succeq 0, \forall i, \quad h_i(x, \Omega) = 0,
\end{align*}
\] (D3)
where $f(x, \Omega)$ is the biconvex target function and $h_i(x, \Omega)$ are bi-affine functions. The global optimization algorithm finds the global optimum of Eq. (D2) by solving a set of primal problems and relaxed dual problems which generate upper and lower bounds on the optimum respectively. The upper and lower bounds converge to the global optimum up to a small error in finite steps. The algorithm is described as follows.
Step 1: Initialization.
Define initial upper and lower bounds \((f^U, f^L)\) on the global optimum, where \(f^U\) and \(-f^L\) can be chosen as two very large numbers. Set the counter \(K = 1\). Set a convergence tolerance parameter \(\varepsilon\). Choose a starting point \(x^1\). Define three empty sets \(\mathcal{R}^{feas}\) (set of feasible problems), \(\mathcal{R}^{infeas}\) (set of infeasible problems), \(\mathcal{S}\) (set of candidates of lower bound).

Step 2: Primal problem.
(1) Consider the primal problem for \(x = x^K\) if it is feasible (that is, if there exists some \(\Omega \in R_2\) that satisfies the constraints):
\[
P(x^K) = \min_{\Omega \in R_2} f(x^K, \Omega),
\]
\[
\text{s.t. } \Omega \geq 0, \quad \forall i, h_i(x^K, \Omega) = 0.
\]
The strong duality theorem \([72]\) indicates that \(P(x^K)\) can be solved through
\[
P(x^K) = \max_{y, z \geq 0, \Omega \in R_2} \min \ L(x^K, \Omega, y, z),
\]
where the Lagrange function
\[
L(x, \Omega, y, z) := f(x, \Omega) + \sum_i y_i h_i(x, \Omega) - \text{Tr}(\Omega z),
\]
\(Z\) is a semidefinite positive matrix acting on \(\mathcal{H}_{S^c} \otimes \mathcal{H}_S\) and \(y\) is a vector of real numbers.
Solve Eq. (D5) and store the optimal values \((\Omega^K, y^K, Z^K)\). Set \(f^U = \min\{f^U, P(x^K)\}\) and \(\mathcal{R}^{feas} = \mathcal{R}^{feas} \cup \{K\}\).

(2) If Eq. (D4) is infeasible, solve the relaxed primal problem for \(x = x^K\) instead:
\[
\delta(x^K) = \min_{\Omega \in R_2, \alpha \geq 0} \alpha,
\]
\[
\text{s.t. } \Omega + \alpha 1_{S^c S} \geq 0, \quad \forall i, h_i(x^K, \Omega) = 0.
\]
The strong duality theorem implies
\[
\delta(x^K) = \max_{y, z \geq 0, \Omega \in R_2} \min \ L_1(x^K, \Omega, y, z),
\]
where the Lagrange function \(L_1(x, \Omega, y, z) := \sum_i y_i h_i(x, \Omega) - \text{Tr}(\Omega z)\).
Solve Eq. (D8) and store the optimal values \((\Omega^K, y^K, Z^K)\). Let \(\mathcal{R}^{infeas} = \mathcal{R}^{infeas} \cup \{K\}\).

Step 3: Determine the current region of \(x\).
Suppose \(\Omega\) is parameterized by a vector of real numbers \(\Omega_i\). Since \(\Omega\) is contained in a compact set, \(\Omega_i\) has upper and lower bounds that we denote by \(\Omega_i^U\) and \(\Omega_i^L\). Consider the partial derivatives of the Lagrange functions defined by \(g_i^l(x) := \frac{\partial}{\partial x^l} L(x, \Omega, y^k, Z^k)|_{\Omega^i}\) for \(k \in \mathcal{R}^{feas}\) and \(g_i^r(x) := \frac{\partial}{\partial x^r} L_1(x, \Omega, y^k, Z^k)|_{\Omega^i}\) for \(k \in \mathcal{R}^{infeas}\). Define the set of indices for connected variables \(I_k := \{i | g_i^r(x) = 0, \forall x\}\) (the last equality follows from the KKT conditions \([72]\)) and \(\Omega_i\) is called a connected variable of the Lagrange functions if \(i \in I_k\). We can also define the linearized Lagrange functions \(L(x, \Omega, y^k, Z^k)|_{\Omega^i}^{lin} := L(x, \Omega^i, y^k, Z^k) + \sum_{i \in I_k} g_i^r(x)(\Omega^i - \Omega_i^L)\) and \(L_1(x, \Omega, y^k, Z^k)|_{\Omega^i}^{lin} := L_1(x, \Omega^i, y^k, Z^k) + \sum_{i \in I_k} g_i^r(x)(\Omega^i - \Omega_i^L)\).
The linearized functions \(L(x, \Omega, y^k, Z^k)|_{\Omega^i}^{lin}\) and \(L_1(x, \Omega, y^k, Z^k)|_{\Omega^i}^{lin}\) are functions of the connected variables only and independent of \(\Omega_i\) if \(i \notin I_k\).
Let \(\mathcal{B}^k := \{\Omega_i^L, \Omega_i^U\}\) be the set of combinations of upper and lower bounds on \(\Omega_i\) for all \(i \in I_k\). We abuse the notation a bit and use \(\Omega \in \mathcal{B}^k\) to denote the case where the part of connected variables \(\Omega_{I_k}\) in \(\Omega\) is contained in \(\mathcal{B}^k\) and the other part is arbitrary. We will see that the other part is irrelevant in our calculations and can be ignored. In this sense, there are in total \(2^{|I_k|}\) number of \(\Omega \in \mathcal{B}^k\) which is finite. We also define \(R(k, \Omega)\) to be a region of \(x\) as a function of \(\Omega \in \mathcal{B}^k\) defined by
\[
R(k, \Omega) := \{x | \forall i \in I_k, g_i^r(x) \leq \Omega, 0\},
\]
where “$\leq \Omega_i$” represents “$\leq$” if $\Omega_i = \Omega_i^L$, and “$\geq$” if $\Omega_i = \Omega_i^R$.

Let $\mathcal{B}^{(k,K)} = \{ \Omega \in \mathcal{B}^k | x^K \in R(k, \Omega) \}$. The relaxed dual problem in the next step will be solved in the region of $x$ that is contained in $\bigcap_{k=1}^{K-1} \bigcap_{\Omega \in \mathcal{B}^{(k,K)}} R(k, \Omega)$.

Step 4: **Relaxed dual problem.**

Determine the set of indices for connected variables $I_K$. Note that for any $k$, $L(x, \Omega, y^k, Z^k)|_{\Omega^k}$ is a function of the connected variables only and is fixed if the connected variables $\Omega_k$ of $\Omega$ is fixed. Therefore we will also write $L(x, \Omega, y^k, Z^k)|_{\Omega^k} = L(x, \Omega, y^k, Z^k)|_{\Omega^k}$.

For each $\Omega_k \in \mathcal{B}^k = \mathcal{X}_{\Omega \in I_K} \{ \Omega_L^k, \Omega_U^k \}$ (there are $2^{|I_K|}$ different $\Omega_k$ in total), solve the following relaxed dual problem:

\[
\begin{align*}
\min_{x \in R_{1,\mu}, \mu} & \quad \mu \\
\text{s.t.} & \quad \mu \geq L(x, \Omega, y^k, Z^k)|_{\Omega^k} \quad \forall \Omega \in \mathcal{B}^{(k,K)}, 1 \leq k \leq K - 1, k \in \mathcal{R}^{feas}, \\
& \quad 0 \geq L_1(x, \Omega, y^k, Z^k)|_{\Omega^k} \quad \forall \Omega \in \mathcal{B}^{(k,K)}, 1 \leq k \leq K - 1, k \in \mathcal{R}^{infeas}, \\
& \quad \mu \geq L(x, \Omega, y^k, Z^k)|_{\Omega^k} \quad K \in \mathcal{R}^{feas}, \\
& \quad 0 \geq L_1(x, \Omega, y^k, Z^k)|_{\Omega^k} \quad K \in \mathcal{R}^{infeas}, \\
& \quad x \in R(k, \Omega), \\
& \quad x \in R(k, \Omega_k),
\end{align*}
\]  

(D10)

For each $\Omega_k$, store the solution $(\mu_k, x_k)$ of Eq. (D10) in $\mathcal{S}$.

Step 5: **Select a new lower bound and determine $x^{K+1}$**.

From the set $\mathcal{S}$, select the minimum $\mu^\text{min}$ and the corresponding $x^\text{min}$. Set $f^L = \mu^\text{min}$ and $x^{K+1} = x^\text{min}$. Delete $(\mu^\text{min}, x^\text{min})$ from the set of candidates of lower bound $\mathcal{S}$.

Step 6: **Check for convergence.**

Check if $f^L > f^U - \varepsilon$, if yes, STOP; otherwise, set $K = K + 1$ and return to Step 2.

The global optimization algorithm described above works in a branch-and-bound way where $x$ is partitioned into different regions and different candidates of lower bounds of the global optimum are explored in each iteration. The subproblems that are solved in each iteration are semidefinite programs (Eq. (D5) and Eq. (D8)) and quadratically constrained quadratic programs (Eq. (D10)) which can be solved efficiently (for a moderate system dimension) using algorithms for convex optimization [72]. The running time of the entire algorithm depends largely on the number of subproblems that are solved in each iteration which is exponential in the number of connected variables. Methods that can reduce this complexity were discussed in [62].

**Appendix E: Binary measurements on pure states**

1. **Measurement on a qubit**

Here consider a binary measurement on a single qubit where $X = x_1M_1 + x_2M_2$, $M_1 = M$ and $M_2 = I - M$. Without loss of generality, we assume

\[
M = \begin{pmatrix}
m_1 & 0 \\
0 & m_2
\end{pmatrix}, \quad m_1, m_2 \in [0, 1] \quad \text{and} \quad m_1 > m_2.
\]  

(E1)

Our goal is to find $(x, U)$ such that the two necessary conditions in Lemma 4 are satisfied.
Let $|\phi\rangle = \sqrt{p} |1\rangle + \sqrt{1-p} |2\rangle$ where $0 \leq p \leq 1$ (any additional phases in the amplitudes of $|\phi\rangle$ do not change the results). Conditions (1) translates into

$$
(\sqrt{p} \sqrt{1-p}) \begin{pmatrix}
  x_1 m_1 + x_2 (1-m_1) \\
  x_1 m_2 + x_2 (1-m_2)
\end{pmatrix} \begin{pmatrix}
  0 \\
  0
\end{pmatrix} = 0,
$$

(E2)

and Condition (2) is trivially true when $m \geq 1$.

When $m = 0$, we must have $\gamma(\{m_i\}) = \lim_{\epsilon \to 0^+} \gamma((\epsilon M_1)^\epsilon) = (\epsilon M_1)^\epsilon - 2\sqrt{\frac{\epsilon}{2}} \left(1 - \frac{\epsilon}{2}\right) m_1 (1 - \left(1 - \frac{\epsilon}{2}\right) m_1) = m_1$.

(E9)

2. Measurement on a qudit

Now consider a $d$-dimensional system with $d > 2$ and a binary measurement $M_1 = M$ and $M_2 = 1 - M$ where

$$
M = \begin{pmatrix}
  m_1 & m_2 & \cdots & m_d \\
  m_2 & m_1 & \cdots & m_d \\
  \vdots & \vdots & \ddots & \vdots \\
  m_d & m_2 & \cdots & m_1
\end{pmatrix},
$$

(E11)

and $1 > m_1 \geq \cdots \geq m_d > 0$. 

Solving the system of equations (Eq. (E2)–Eq. (E5)), we find the following results.

(1) When $1 > m_1 > m_2 > 0$, according to Lemma S1, the QPFI is attainable. Moreover, the only solution satisfying the necessary conditions (Eq. (E2)–Eq. (E5)) is

$$
x_2 = \frac{x_1}{m_1 m_2},
$$

(E6)

and

$$
p = \frac{\sqrt{m_2(1-m_2)}}{\sqrt{m_1(1-m_1)} + \sqrt{m_2(1-m_2)}},
$$

(E7)

which must be the optimal solution. It gives

$$
\gamma(\{M_i\}) = 1 - (\epsilon M_1)^\epsilon + \sqrt{(1 - \epsilon) m_1} + \sqrt{(1 - \epsilon) m_1} - 2\sqrt{\frac{\epsilon}{2}} \left(1 - \frac{\epsilon}{2}\right) m_1 (1 - \left(1 - \frac{\epsilon}{2}\right) m_1) = m_1.
$$

(E8)

(2) When $1 = m_1 > m_2 = 0$, we must have $X^2 = X_2$, implying $\gamma(\{M_i\}) = 1$. The QFI is achievable as long as a solution to $p x_1 + (1-p) x_2 = 0$ and $p x_1^2 + (1-p) x_2^2 = 1/(4n)$ exists, which means that any $0 < p < 1$ is optimal.

(3) When $1 > m_1 > m_2 = 0$, the necessary conditions (Eq. (E2)–Eq. (E5)) have no solutions. Thus, this is a singular case where the QPFI is not attainable. And we have from Theorem 1 that

$$
\gamma(\{M_i\}) = \lim_{\epsilon \to 0^+} \gamma((\epsilon M_1)^\epsilon)
$$

(E9)

and

$$
\gamma(\{M_i\}) = \lim_{\epsilon \to 0^+} \left(1 - \frac{\epsilon}{2}\right) m_1 + \frac{\epsilon}{2} - \left(1 - \frac{\epsilon}{2}\right) m_1 \epsilon - 2\sqrt{\frac{\epsilon}{2}} \left(1 - \frac{\epsilon}{2}\right) m_1 (1 - \left(1 - \frac{\epsilon}{2}\right) m_1) = m_1.
$$

(E10)
Let \( |\phi\rangle = \sum_{i=1}^{d} \phi_i |i\rangle \). We now show that the support of \( |\phi\rangle \): \( \text{supp}\{|\phi\rangle\} = \{i : \phi_i \neq 0\} \) must correspond to at most two different values of \( m_i \) when \( |\phi\rangle \) is optimal. We prove this by contradiction. Without loss of generality, assume \( |\phi_{1,2,3}\rangle > 0 \) and \( m_1 > m_2 > m_3 \). Condition (2) implies that

\[
\frac{\langle 1|X^2|1 \rangle}{\langle 1|X_2|1 \rangle} = \frac{\langle 2|X^2|2 \rangle}{\langle 2|X_2|2 \rangle} = \frac{\langle 3|X^2|3 \rangle}{\langle 3|X_2|3 \rangle},
\]

(E12)

\[
\Rightarrow \frac{(x_1m_1 + x_2(1-m_1))^2}{x_1^2m_1 + x_2^2(1-m_1)} = \frac{(x_1m_2 + x_2(1-m_2))^2}{x_1^2m_2 + x_2^2(1-m_2)} = \frac{(x_1m_3 + x_2(1-m_3))^2}{x_1^2m_3 + x_2^2(1-m_3)},
\]

(E13)

\[
\Rightarrow -\sqrt{\frac{m_1m_2}{(1-m_1)(1-m_2)}} = -\sqrt{\frac{m_1m_3}{(1-m_1)(1-m_3)}} = -\sqrt{\frac{m_2m_3}{(1-m_2)(1-m_3)}},
\]

(E14)

which contradicts \( m_1 > m_2 > m_3 \). Thus, we conclude that the support of \( |\phi\rangle \) must correspond to at most two different values of \( m_i \).

Therefore we have

\[
\gamma(\{M_i\}) = \max_{1 \leq k \leq d} \gamma_{kl} = \max_{kl} 1 - \left( \sqrt{m_km_l} + \sqrt{(1-m_k)(1-m_l)} \right)^2.
\]

(E15)

In fact, assume \( m_1 \geq m_2 \geq \cdots \geq m_d \), we must have

\[
\gamma(\{M_i\}) = 1 - \left( \sqrt{m_1m_d} + \sqrt{(1-m_1)(1-m_d)} \right)^2.
\]

(E16)

The reason is that when \( m_k \geq m_l \), increasing \( m_k \) or decreasing \( m_l \) while the other element is fixed will only increases \( \gamma_{kl} \). We see that by computing the derivative of \( \gamma_{kl} \) with respect to \( m_k \). We have

\[
\frac{\partial}{\partial m_k} \gamma_{kl} = 1 - 2m_l - \frac{\sqrt{(1-m_1)m_l}}{\sqrt{(1-m_k)m_k}} (1-2m_k) \geq 0,
\]

(E17)

when \( m_k \geq m_l \) because \( \frac{1-2m_l}{\sqrt{(1-m_1)m_l}} = \sqrt{\frac{1-m_l}{m_l}} - \sqrt{\frac{m_l}{1-m_l}} \geq \frac{1-2m_k}{\sqrt{(1-m_k)m_k}} \).

**Appendix F: Commuting-operator measurements on pure states**

We take one step further from binary measurements and consider the commuting-operator measurements where

\[
M_i = \begin{pmatrix}
   m_1^{(i)} \\
   m_2^{(i)} \\
   \vdots \\
   m_d^{(i)}
\end{pmatrix},
\]

(F1)

and \( \sum_i M_i = 1 \). We also assume \( m_j^{(i)} > 0 \) for all \( i,j \).

1. **Proof of Theorem 6**

We first prove Theorem 6:

**Theorem 6.** For commuting-operator measurements (Eq. (51)), there always exists an optimal solution to \( (|\phi\rangle, |\phi^1\rangle) \) such that \( |\phi\rangle = \sqrt{p} |k\rangle + \sqrt{\bar{p}} |l\rangle \) and \( |\phi^1\rangle = \sqrt{\bar{p}} |k\rangle - \sqrt{p} |l\rangle \) for two basis states \( |k\rangle \) and \( |l\rangle \) and \( 0 < p < 1 \).

**Proof.** Assume \( (x,|\phi\rangle) \) satisfies Condition (2) in Lemma 4, where we write \( x_i = y(x + a_i) \), \( x \neq 0 \) and the support of \( |\phi\rangle \) contains \( |1\rangle \) and \( |2\rangle \). Then according to Condition (2), we must have

\[
\frac{(x + \langle a \rangle_1)^2}{x^2 + 2 \langle a \rangle_1 x + \langle a^2 \rangle_1} = \frac{(x + \langle a \rangle_2)^2}{x^2 + 2 \langle a \rangle_2 x + \langle a^2 \rangle_2} = \frac{\langle \phi | X^2 | \phi \rangle}{\langle \phi | X_2 | \phi \rangle},
\]

(F2)

where \( \langle a \rangle_j = \sum_i a_i m_j^{(i)} \) and \( \langle a^2 \rangle_j = \sum_i a_i^2 m_j^{(i)} \). We also define \( \langle \Delta a^2 \rangle_j = \langle a^2 \rangle_j - \langle a \rangle_j^2 \).

Assume \( \langle a \rangle_1 > \langle a \rangle_2 \), we have the following possible solutions of Eq. (F2).
(1) When \((\Delta a^2)_1 = (\Delta a^2)_2\),
\[
x = -\frac{1}{2}\langle a \rangle_1 + \langle a \rangle_2,
\]
and
\[
\frac{\langle \phi | X_2 | \phi \rangle}{\langle \phi | X^2 | \phi \rangle} = 1 + \left( \frac{\sqrt{\langle \Delta a^2 \rangle_1} + \sqrt{\langle \Delta a^2 \rangle_2}}{\langle a \rangle_1 - \langle a \rangle_2} \right)^2.
\] (F3)

(2) When \((\Delta a^2)_1 - (\Delta a^2)_2 \neq 0\), we have either
\[
x = -\langle a \rangle_1 \frac{\langle a \rangle_1 (\Delta a^2)_1 + \langle a \rangle_1 (\Delta a^2)_2 - (\langle a \rangle_1 - \langle a \rangle_2) \sqrt{(\Delta a^2)_1 (\Delta a^2)_2}}{(\Delta a^2)_1 - (\Delta a^2)_2}
\] (F4)
and
\[
\frac{\langle \phi | X_2 | \phi \rangle}{\langle \phi | X^2 | \phi \rangle} = 1 + \left( \frac{\sqrt{\langle \Delta a^2 \rangle_1} - \sqrt{\langle \Delta a^2 \rangle_2}}{\langle a \rangle_1 - \langle a \rangle_2} \right)^2,
\] (F5)
or
\[
x = -\langle a \rangle_2 \langle \Delta a^2 \rangle_1 + \langle a \rangle_1 \langle \Delta a^2 \rangle_2 + (\langle a \rangle_1 - \langle a \rangle_2) \sqrt{(\Delta a^2)_1 (\Delta a^2)_2}
\] (F6)
and
\[
\frac{\langle \phi | X_2 | \phi \rangle}{\langle \phi | X^2 | \phi \rangle} = 1 + \left( \frac{\sqrt{\langle \Delta a^2 \rangle_1} - \sqrt{\langle \Delta a^2 \rangle_2}}{\langle a \rangle_1 - \langle a \rangle_2} \right)^2.
\] (F7)

Next we show that there always is an optimal solution such that its support contains only two elements. Without loss of generality, assume \(|\phi^*\rangle = \sum_{i=1}^d \phi_i^* |i\rangle\) is an optimal solution (which is guaranteed to exist thanks to Lemma S1). The corresponding error vector \(x^*\) is written as \(x_i^* = y^*(x^* + a_i^*)\). We have from Condition (1) in Lemma 4 that \(\sum_{i=1}^d |\phi_i^*|^2 y^*(x^* + a_i^*) = 0\) and \(\sum_{i=1}^d |\phi_i^*|^2 (y^*(x^* + a_i^*))^2 = 1/(4n)\). Clearly, we must have some \(i \neq j\), such that \(\phi_i^* |i\rangle > 0\), \(\langle a_i^* |i\rangle \neq \langle a_j^* |j\rangle\) and \((x^* + a_i^*)(x^* + a_j^*) < 0\). Without loss of generality, we assume \(i = 1\), \(j = 2\) and \(\langle a_1^* |1\rangle > (\langle a_2^* |2\rangle\). Then \((x_i^* |\phi^*\rangle\) must satisfy either Eq. (F3) or Eq. (F4) and Eq. (F5) from the previous discussion. (Note that if Eq. (F6) and Eq. (F7) cannot be true because \((x^* + a_i^*)(x^* + a_j^*) < 0\).)

We then assert that \(|\phi^{**}\rangle = \sqrt{p}|1\rangle + \sqrt{1-p}|2\rangle \) (0 \(\leq p \leq 1\) is also an optimal solution, when \(p\) satisfies
\[
(\sqrt{p} \sqrt{1-p}) \begin{pmatrix} x^* + a_1^* & 0 \\ 0 & x^* + a_2^* \end{pmatrix} \begin{pmatrix} \sqrt{p} \\ \sqrt{1-p} \end{pmatrix} = 0.
\] (F8)

Using Eq. (F3)–Eq. (F5), it is easy to see that
\[
p = \frac{\sqrt{(\Delta a^2)_1^2} \sqrt{(\Delta a^2)_2^2}}{\sqrt{(\Delta a^2)_1^2} + \sqrt{(\Delta a^2)_2^2}}.
\] (F9)

We take \(x_i^{**} = y^{**}(x^* + a_i^*)\), where \(y^{**}\) is solved from
\[
(\sqrt{p} \sqrt{1-p}) \begin{pmatrix} (x^* + a_1^*)^2 & 0 \\ 0 & (x^* + a_2^*)^2 \end{pmatrix} \begin{pmatrix} \sqrt{p} \\ \sqrt{1-p} \end{pmatrix} = \frac{1}{4n(y^{**})^2}.
\] (F10)

Both equations are derived from Condition (1). Now we have a new solution \((x^{**}, |\phi^{**}\rangle)\) such that \(y^{**}\) and \(|\phi^{**}\rangle\) are solved by the equations above. Note that we still let \(a_i^{**} = a_i^*\) and \(x^{**} = x^*\). The new solution have the same FI as the original, i.e., \(\langle \phi | X_2 | \phi \rangle\) does not change, because \(\langle \phi | X_2 | \phi \rangle / \langle \phi | X^2 | \phi \rangle\) is independent of \(y\) (due to Eq. (F3) and Eq. (F5)) and \(\langle \phi | X^2 | \phi \rangle = 1/(4n)\) is invariant. The new solution is thus supported on a two-dimensional subspace spanned by \{|1\rangle, |2\rangle\}, proving Theorem 6. \(\square\)
2. Optimal solution for commuting-operator measurements

Now we proceed to compute general $\gamma(\{M_i\})$ for commuting-operator measurements. First, consider the optimization for measurements restricted in a two-dimensional subspace spanned by $|k\rangle, |l\rangle$ for some $k \neq l$, i.e.,

$$(M_i)_{kl} = m_k^{(i)} |k\rangle \langle k| + m_l^{(i)} |l\rangle \langle l|,$$  \hspace{1cm} (F11)

and $\sum_i (M_i)_{kl} = 1_{\text{span}\{|k\rangle, |l\rangle\}}$.

Let $(x^*, \phi^*)$ be an optimal solution when $|\phi\rangle, |\phi^+\rangle$ are restricted in $\text{span}\{|k\rangle, |l\rangle\}$ and $|\phi^+\rangle = \sqrt{p_{kl}} |k\rangle + \sqrt{1 - p_{kl}} |l\rangle$ (we also assume $\langle a^*_k \rangle > \langle a^*_l \rangle$). Using Eq. (16), we see that the optimal $a_i^*$

$$y^*(x^* + a_i^*) = x_i^* = \frac{y^* \langle \phi^+ | M_i X^* | \phi^+ \rangle}{\gamma_{kl} \langle \phi^+ | M_i | \phi^+ \rangle} = \frac{y^* \sqrt{p_{kl} m_k^{(i)}} (x^* + \langle a^*_k \rangle_k) + (1 - p_{kl}) m_i^{(i)} (x^* + \langle a^*_l \rangle_l)}{\gamma_{kl} m_k^{(i)} + (1 - p_{kl}) m_i^{(i)}}$$ \hspace{1cm} (F12)

where we use

$$\langle \phi^+ | X^* | \phi^+ \rangle = p_{kl} y^*(x^* + \langle a^*_k \rangle_k) + (1 - p_{kl}) y^*(x^* + \langle a^*_l \rangle_l) = 0,$$ \hspace{1cm} (F13)

in the last step. From Eq. (F12), we have,

$$(y^*)^2((x^*)^2 + 2 \langle a^*_k \rangle_k x^* + \langle (a^*_k)^2 \rangle_k) = \left(\frac{p_{kl} y^*(x^* + \langle a^*_k \rangle_k)}{\gamma_{kl}}\right)^2 \sum_k \frac{(m_k^{(i)} - m_i^{(i)})^2 m_k^{(i)}}{(p_{kl} m_k^{(i)} + (1 - p_{kl}) m_i^{(i)})^2},$$ \hspace{1cm} (F14)

$$(y^*)^2((x^*)^2 + 2 \langle a^*_l \rangle_l x^* + \langle (a^*_l)^2 \rangle_l) = \left(\frac{p_{kl} y^*(x^* + \langle a^*_l \rangle_l)}{\gamma_{kl}}\right)^2 \sum_l \frac{(m_l^{(i)} - m_i^{(i)})^2 m_l^{(i)}}{(p_{kl} m_l^{(i)} + (1 - p_{kl}) m_i^{(i)})^2}.$$ \hspace{1cm} (F15)

According to Condition (2),

$$\frac{(x^*)^2 + 2 \langle a^*_k \rangle_k x^* + \langle (a^*_k)^2 \rangle_k}{(x^* + \langle a^*_k \rangle_k)^2} = \frac{(x^*)^2 + 2 \langle a^*_l \rangle_l x^* + \langle (a^*_l)^2 \rangle_l}{(x^* + \langle a^*_l \rangle_l)^2}.$$ \hspace{1cm} (F16)

From Eq. (F13)–Eq. (F16), we have

$$p_{kl}^2 \left(\sum_k \frac{(m_k^{(i)} - m_i^{(i)})^2 m_k^{(i)}}{(p_{kl} m_k^{(i)} + (1 - p_{kl}) m_i^{(i)})^2}\right) = (1 - p_{kl})^2 \left(\sum_l \frac{(m_l^{(i)} - m_i^{(i)})^2 m_l^{(i)}}{(p_{kl} m_l^{(i)} + (1 - p_{kl}) m_i^{(i)})^2}\right).$$ \hspace{1cm} (F17)

It will give us a unique solution to $p_{kl}$ because the left-hand side is a monotonically increasing function in $[0, \sum (m_k^{(i)} - m_i^{(i)})^2]$ of $p_{kl} \in [0, 1]$ and the right-hand side is a monotonically decreasing function in $[0, \sum (m_l^{(i)} - m_i^{(i)})^2]$ of $p_{kl} \in [0, 1]$. However, a simple analytical solution to $p_{kl}$ from Eq. (F17) might not exist because it is a root of a high degree polynomial. Then we have

$$\gamma_{kl} = \sum_i \frac{\langle \phi | M_i | \phi \rangle}{\langle \phi | M_i | \phi^+ \rangle} = \sum_i \frac{p_{kl} (1 - p_{kl}) (m_k^{(i)} - m_i^{(i)})^2}{p_{kl} m_k^{(i)} + (1 - p_{kl}) m_i^{(i)}},$$ \hspace{1cm} (F18)

where $p_{kl}$ is the unique solution to Eq. (F17).

Finally,

$$\gamma(\{M_i\}) = \max_{k,l} \gamma_{kl}.$$ \hspace{1cm} (F19)

using Proposition 6. Note that although Eq. (F17) might only be solvable numerically in practice for a multiple-outcome measurement. Our solution for pure states and commuting-operator measurements still has a huge simplification compared to the original biconvex problem for general states and measurements.
Appendix G: Classical capacity of quantum channels

In this section, we prove the following lemma:

**Lemma S3.** Consider a quantum channel \( \Phi \), its classical channel capacity \( C(\Phi) \) and a constant \( \alpha \) satisfying \( 0 < \alpha < C(\Phi) \). Then for all but finitely many positive integers \( n \), there exists channels \( \Xi_E \) and \( \Xi_D \) such that \( \| \Xi_D \circ \Phi^\otimes n \circ \Xi_E - D_2^\otimes [\alpha n] \|_\infty \leq e^{-\beta n} \), for some \( \beta > 0 \).

**Lemma S3** essentially states that, fixing any \( \alpha \) that is smaller than the classical channel capacity of \( \Phi \), in the large \( n \) limit, \( \Phi^\otimes n \) with suitable encoding and decoding channels can be used to transmit classical binary information reliably at a rate \( \alpha \) with an exponentially small error with respect to \( n \). The proof of **Lemma S3** follows almost exactly from the proof of the HSW theorem [49, 50], with a slight refinement in the error analysis where the Hoeffding’s inequality is used to show the error is exponentially small. Therefore, we will only provide the error analysis part for the proof of **Lemma S3** that is different from the standard proof of the HSW theorem and skip the rest as they are lengthy and can easily be found in standard quantum information theory textbooks [54, 73].

Following the proof of Theorem 8.27 in [54], it is clear that in order to prove **Lemma S3**, it is sufficient to prove the following lemma (which is a refinement of Theorem 8.26 in [54]):

**Lemma S4.** Let \( \eta = (p(a), \sigma_a) \) be an ensemble of quantum states satisfying \( \sum_a p(a) = 1 \), where \( \sigma_a \) are density operators and \( a \in \Sigma \) (\( \Sigma \) is an alphabet whose order is equal to the square of the dimension of the system \( \sigma_a \) act on). Let

\[
\alpha < \chi(\eta) := H\left( \sum_{a \in \Sigma} p(a) \sigma_a \right) - \sum_{a \in \Sigma} p(a) H(\sigma_a),
\]

where \( H(\sigma_a) \) is the von Neumann entropy of \( \sigma_a \) and \( m = |\alpha n| \). For all but finite number of \( n \), there exists a function \( f : \{0,1\}^m \to \Sigma^n \) and a quantum measurement \( \{ M_{b \in \{0,1\}^m} \} \) such that

\[
\text{Tr}(M_{b \sigma_f(b)}) > 1 - e^{-\beta n},
\]

for every \( b = b_1 \cdots b_m \in \{0,1\}^m \), \( \sigma_f(b) = \sigma_f(b_1) \otimes \cdots \otimes \sigma_f(b_n) \) and some \( \beta > 0 \).

**Proof.** Choose a sufficiently small \( \varepsilon \) such that \( \alpha < \chi(\eta) - 3\varepsilon \). Following the proof of Theorem 8.26 in [54], there exists a function \( f : \{0,1\}^m \to \Sigma^n \) and a quantum measurement \( \{ M_{b \in \{0,1\}^m} \} \) such that

\[
\text{Tr}(M_{b \sigma_f(b)}) > 1 - \delta,
\]

for every \( b = b_1 \cdots b_m \in \{0,1\}^m \) where

\[
\delta = 4\left( 3 - 2\text{Tr}(P\sigma^\otimes n) - \sum_{a \in \Sigma^n} p(a_1) \cdots p(a_n) \text{Tr}(\Lambda_a \sigma_a) \right) + 2^{m+4-n(\chi(\eta)-2\varepsilon)}.
\]

Here \( a = a_1 \cdots a_n \in \Sigma^n \), \( \sigma = \sum_{a \in \Sigma} p(a) \sigma_a \), \( \sigma_a = \sigma_a \otimes \cdots \otimes \sigma_a \), \( \Pi \) is the projection onto the \( \varepsilon \)-typical subspace with respect to \( \sigma \), \( \Lambda_a \) is the projection onto the \( \varepsilon \)-typical subspace conditioned on \( a = a_1 \cdots a_n \). Specifically, let \( \sigma = \sum_{a \in \Sigma} p'(a) |u_a\rangle \langle u_a| \) where \( \{|u_a|, a \in \Sigma\} \) is an orthonormal basis and \( p(a) \sigma_a = \sum_{c \in \Gamma} p(a,c) |u_{ac}\rangle \langle u_{ac}| \) where \( \{|u_{ac}|, a \in \Gamma\} \) is an orthonormal basis for each \( a \in \Sigma \). Let \( p(a) = \sum_{c \in \Gamma} p(a,c) \) and \( H(p(a)) \) be the Shannon entropy of \( p(a) \). Then the definitions of \( \Pi \) and \( \Lambda_a \) are

\[
\Pi = \sum_{a \in T_e} |u_{a_1}\rangle \langle u_{a_1}| \otimes \cdots \otimes |u_{a_n}\rangle \langle u_{a_n}|,
\]

\[
\Lambda_a = \sum_{c \in K_{a,e}} |u_{a_1,c_1}\rangle \langle u_{a_1,c_1}| \otimes \cdots \otimes |u_{a_n,c_a}\rangle \langle u_{a_n,c_a}|,
\]

where \( T_e \) is the set of \( a \) satisfying \( 2^{-n(H(p'(a)) + \varepsilon)} < p'(a_1) \cdots p'(a_n) < 2^{-n(H(p(a)) - \varepsilon)} \) and \( K_{a,e} \) is the set of \( c \) satisfying \( 2^{-n(H(p(a,c)) - H(p(a))) + \varepsilon} \) for any \( a \) satisfying \( p(a_1) \cdots p(a_n) > 0 \). We have

\[
\text{Tr}(\Pi \sigma^\otimes n) = \sum_{a \in T_e} p'(a_1) \cdots p'(a_n),
\]

\[
\sum_{a \in \Sigma^n} \text{Tr}(\Lambda_a \sigma_a) = \sum_{a \in \Sigma^n} \sum_{c \in K_{a,e}} p(a_1,c_1) \cdots p(a_n,c_n).
\]
Define two random variables $X : \Sigma \to [0, x_{\text{upp}}]$ and $Y : \Sigma \times \Gamma \to [0, y_{\text{upp}}]$, where $x_{\text{upp}} = \max_{a: p'(a) \neq 0} - \log(p'(a))$ and $y_{\text{upp}} = \max_{a,b,c: p(a,c) \neq 0} - \log(p(a,c)) + \log(p(a))$, as

$$X(a) = -\log(p'(a)) \text{ if } p'(a) > 0, \text{ and } 0 \text{ otherwise,} \quad (G9)$$

$$Y(a,b) = -\log(p(a,c)) + \log(p(a)) \text{ if } p(a,c) > 0, \text{ and } 0 \text{ otherwise.} \quad (G10)$$

Let $X_1, \ldots, X_n$ be $n$ independent random variables each identically distributed to $X$. Using the Hoeffding inequality [74], we have

$$\Pr \left( \left| \frac{1}{n} \sum_{i=1}^{n} X_i \right| - H(p'(a)) \geq \varepsilon \right) \leq 2 \exp \left( -2n\varepsilon^2 / x_{\text{upp}}^2 \right). \quad (G11)$$

On the other hand, according to the definition of $T_\varepsilon$, we have

$$\Pr \left( \left| \frac{1}{n} \sum_{i=1}^{n} X_i \right| - H(p'(a)) \geq \varepsilon \right) = 1 - \sum_{a \in T_\varepsilon} p'(a_1) \cdots p'(a_n). \quad (G12)$$

It implies $\Tr(\Pi \sigma^n) \geq 1 - 2 \exp \left( -2n\varepsilon^2 / x_{\text{upp}}^2 \right)$ using Eq. (G7). Similarly, using the Hoeffding inequality for independent random variables distributed to $Y$ and Eq. (G8), we have $\sum_{a \in \Sigma^n} \Tr(\Lambda_a \sigma_a) \geq 1 - 2 \exp \left( -2n\varepsilon^2 / y_{\text{upp}}^2 \right)$. Plugging in these bounds in Eq. (G4), we have

$$\delta \leq 16e^{-2n\varepsilon^2 / x_{\text{upp}}^2} + 8e^{-2n\varepsilon^2 / y_{\text{upp}}^2} + 2^{4-n\varepsilon}, \quad (G13)$$

proving the lemma.