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

Deutsche Physikalische Gesellschaft DPG IOP Institute of Physics

PAPER • OPEN ACCESS

Incompatibility in quantum parameter estimation

To cite this article: Federico Belliardo and Vittorio Giovannetti 2021 *New J. Phys.* **23** 063055

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

Deutsche Physikalische Gesellschaft DPG
IOP Institute of Physics

Published in partnership with: Deutsche Physikalische Gesellschaft and the Institute of Physics

CrossMark

OPEN ACCESS

RECEIVED
2 March 2021

REVISED
17 May 2021

ACCEPTED FOR PUBLICATION
25 May 2021

PUBLISHED
18 June 2021

PAPER

Incompatibility in quantum parameter estimation

Federico Belliardo[1,*] **and Vittorio Giovannetti**[2]

[1] NEST, Scuola Normale Superiore, I-56126 Pisa, Italy
[2] NEST, Scuola Normale Superiore and Istituto Nanoscienze-CNR, I-56126 Pisa, Italy
[*] Author to whom any correspondence should be addressed.

E-mail: federico.belliardo@sns.it and vittorio.giovannetti@sns.it

Keywords: quantum parameter estimation, quantum metrology, quantum measurements, incompatibility

Abstract

In this paper we introduce a measure of genuine quantum incompatibility in the estimation task of multiple parameters, that has a geometric character and is backed by a clear operational interpretation. This measure is then applied to some simple systems in order to track the effect of a local depolarizing noise on the incompatibility of the estimation task. A semidefinite program is described and used to numerically compute the figure of merit when the analytical tools are not sufficient, among these we include an upper bound computable from the symmetric logarithmic derivatives only. Finally we discuss how to obtain compatible models for a general unitary encoding on a finite-dimensional probe.

1. Introduction

Quantum metrology [1–5] is a special branch of quantum information theory that focuses on the possibility of using quantum effects for improving the accuracy of conventional estimation procedures. Thanks to the huge variety of potential applications (which among other include the probing of delicate biological systems [6], optical interferometry [7, 8], gravitational wave detection [9, 10], magnetometry [11–15] and atomic clocks [16–18]), this research field is likely to play a fundamental role in the looming quantum technology revolution. As evident from the seminal works of Holevo [19] and Helstrom [20], this research field can be thought as a quantum counterpart of experimental design [21, 22]. Specifically the main goal of quantum metrology is to efficiently plan different types of experiments by minimizing the invested effort to overcome noisy fluctuations that originate by fabrication errors, external fields, microscopic degrees of freedom that are only statistically taken into account, and intrinsic limitations related to the formal structure of the quantum theory itself (e.g. the Heisenberg uncertainty principle). In recent years many significant results have accumulated in the domain of multi-parameter quantum metrology [23, 24], i.e. processes where an agent tries to recover two or more attributes of a physical system (modeled by real numbers) via properly chosen measurements. The first studies that lit up the experimental interest in this subject have been done on the joint estimation of phase and phase diffusion [25–29], on quantum imaging [30–37], and on magnetometry [38, 39]. What makes the problem intriguing is that in a purely quantum setting, due to constraints ultimately related to the incompatibility of non-commuting observables [40], it could be that an efficient experiment for the determination of one specific parameter leads to poor results in the precision of the others (while this may also be true in classical mechanics, since here the phenomenon is related to the technological limits of the experimenter, there is no reason to believe it to be fundamental). Aim of the present work is to quantify the genuine quantum incompatibility associated with the estimation task of multiple parameters. The analysis is then applied to some simple systems of qubits and qutrits in order to track the effect of a local depolarizing noise on the incompatibility of the estimation task. A semidefinite program (SDP) is described and used to numerically compute the figure of merit when the analytical tools are not sufficient. Finally we notice that the strategies that allow us to codify information without incompatibility in the two-qubits scenario can be generalized to the case of a general unitary encoding on a finite-dimensional probe. Before proceeding with the presentation, we add

here a terminology clarification: with 'quantum parameter' estimation we denote the task of extracting a parameter encoded on a certain given fixed state of a quantum system, while if we use 'quantum metrology' it means that we have the possibility of choosing the probe that will undergo the encoding process. In this perspective the problem of parameter estimation is hence a sub-problem of quantum metrology. In this paper we will take the probe to be fixed and therefore we will be dealing with parameter estimation.

An outlook of the manuscript follows. In section 2 we introduce the setting of quantum metrology, and isolate the form of incompatibility that we will characterize later on. In section 3.1 the incompatibility figure of merit for quantum estimation is defined and its well-definedness is proved in section 3.2. The geometric interpretation of the figure of merit is presented in sections 3.3 and in 3.4 we express it in terms of the Holevo–Cramér–Rao bound [19, 41], proved to be achievable thanks to the quantum central limit theorem and the quantum local asymptotic normality [42–48]. This allows us to compute the incompatibility via the SDP reported in appendix C, which is derived from the one presented in [49]. In section 3.5 an analytic upper bound for the incompatibility is presented, and in section 3.6 a version of the figure of merit for separable measurements is given. Section 4 is dedicated to some examples with systems of qubits and qutrits subject to local depolarizing noise, here we put at work the linear program and some peculiar behavior of the incompatibility is observed. In section 5 we describe three strategies to build a compatible statistical model for a quantum metrological task involving D-dimensional probe states. The mathematical environments definition, theorem, and corollary will be used to highlight the most important concepts that we introduce.

2. Multiparameter quantum estimation

2.1. Setting and definitions

A prototypical example of multi-parameter quantum metrology is provided by magnetometry [11–15, 50] where a spin particle is used as a probe for evaluating the three components of a magnetic field $\boldsymbol{B} := \left(B_x, B_y, B_z\right)$. In the most basic scenario the evolution of the particle is given by the unitary transformation $U_{\boldsymbol{B}} := \exp\left[\mathrm{i}\left(B_x S_x + B_y S_y + B_z S_z\right) t\right]$ where S_i for $i = x, y, z$ are the components of the spin. By measuring the evolved state of the probe we can hence try infer the values of B_x, B_y, B_z, following the post-processing of the measurements output. What makes this procedure truly quantum in nature is that, fixing the number of experimental repetitions, due to the non-commuting nature of the generators S_i, any attempt to improve the estimation accuracy of one of the Cartesian components of $\boldsymbol{B}$ will have a negative impact on the accuracies of the other two [51]. An exact formalization of this problem can be obtained by considering a more general model where one is asked to determine d parameters $\boldsymbol{\theta} := (\theta_1, \theta_2, \ldots, \theta_d) \in \Theta$ (an open subset of $\mathbb{R}^d$) that have been encoded in the input state ρ of a probing quantum system via a mapping of the form

$$\rho \to \rho_{\boldsymbol{\theta}} := \mathcal{E}_{\boldsymbol{\theta}}(\rho), \tag{1}$$

where now $\mathcal{E}_{\boldsymbol{\theta}}$ is a completely positive, trace-preserving transformation [52] which parametrically depends on $\boldsymbol{\theta}$ and which, at variance with the simplified scenario detailed at the beginning of the section, might include a noise disturbing the process. Given N copies of $\rho_{\boldsymbol{\theta}}$ we can now try to recover the needed information by performing on them some (possibly joint) positive operator valued measure (POVM) $M_N := \{E^{(N)}_{\hat{\boldsymbol{\theta}}}\}_{\hat{\boldsymbol{\theta}}}$ whose elements are labeled by a classical outcome variable $\hat{\boldsymbol{\theta}}$ that, without loss of generality [53], can be assumed to belong to the same set Θ of $\boldsymbol{\theta}$. Accordingly M_N can hence be thought as an operation which, starting from $\rho_{\boldsymbol{\theta}}^{\otimes N}$, induces a measure on Θ, defined by the conditional probability distribution

$$P_{M_N}(\hat{\boldsymbol{\theta}}|\boldsymbol{\theta}) := \mathrm{Tr}[E^{(N)}_{\hat{\boldsymbol{\theta}}} \rho_{\boldsymbol{\theta}}^{\otimes N}], \tag{2}$$

with the stochastic outcome $\hat{\boldsymbol{\theta}} := (\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_d) \in \Theta$ playing the role of the estimator of $\boldsymbol{\theta}$. The two most important properties of the estimator $\hat{\boldsymbol{\theta}}$ are the bias vector $\boldsymbol{b}^{(N)}(\boldsymbol{\theta}) := (b_1(\boldsymbol{\theta}), b_2(\boldsymbol{\theta}), \ldots b_d(\boldsymbol{\theta}))$, of components

$$b_i(\boldsymbol{\theta}) := \mathbb{E}[\hat{\theta}_i] - \theta_i, \tag{3}$$

and the mean square error (MSE) $d \times d$ matrix $\Sigma^{(N)}(\boldsymbol{\theta})$, of elements

$$\Sigma^{(N)}_{ij}(\boldsymbol{\theta}) := \mathbb{E}[(\hat{\theta}_i - \theta_i)(\hat{\theta}_j - \theta_j)], \tag{4}$$

with $\mathbb{E}$ representing the statistical average computed with the probability measure in (2). Ideally we would like to deal with estimators that are unbiased, meaning that $\boldsymbol{b}^{(N)}(\boldsymbol{\theta}) = 0$ for all $\boldsymbol{\theta} \in \Theta$, but this may not always be possible. Accordingly in what follows we shall focus on sensing, i.e. we shall measure small variations of the parameters $\boldsymbol{\theta}$ around a known value and assume that we are allowed to employ locally

unbiased POVMs at such special point, that is measurements which bias vector $\boldsymbol{b}^{(N)}(\boldsymbol{\theta})$ satisfy the following conditions

$$\begin{cases} \boldsymbol{b}^{(N)}(\boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = 0, \\ \dfrac{\partial}{\partial\theta_j}\boldsymbol{b}^{(N)}(\boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = 0 \quad \forall\, j = 1,\ldots,d. \end{cases} \tag{5}$$

For these measurements the quantum Cramér–Rao (QCR) bound [3, 54] gives a limit on the precision of the sensing task, formulated as a lower bound on the associated MSE matrix, i.e.

$$\Sigma^{(N)}(\boldsymbol{\theta}) \geqslant \frac{F^{-1}(\boldsymbol{\theta})}{N}. \tag{6}$$

In this expression $F(\boldsymbol{\theta})$ is the so called quantum Fisher information (QFI) matrix which no longer depends on the selected POVM M_N and whose elements can be computed as

$$F_{ij}(\boldsymbol{\theta}) := \frac{1}{2}\,\mathrm{Tr}\left[\rho_{\boldsymbol{\theta}}\left(L_i(\boldsymbol{\theta})L_j(\boldsymbol{\theta}) + L_j(\boldsymbol{\theta})L_i(\boldsymbol{\theta})\right)\right], \tag{7}$$

with $L_i(\boldsymbol{\theta})$ the symmetric logarithmic derivative (SLD) [54] associated to the ith component of the parameter vector $\boldsymbol{\theta}$, i.e. the operator (possibly dependent on $\boldsymbol{\theta}$) fulfilling the identity

$$\frac{\partial\rho_{\boldsymbol{\theta}}}{\partial\theta_i} = \frac{1}{2}\left(\rho_{\boldsymbol{\theta}}L_i(\boldsymbol{\theta}) + L_i(\boldsymbol{\theta})\rho_{\boldsymbol{\theta}}\right). \tag{8}$$

For a pure state $\rho_{\boldsymbol{\theta}} = |\psi_{\boldsymbol{\theta}}\rangle\langle\psi_{\boldsymbol{\theta}}|$ the above equation admits as solution

$$L_i(\boldsymbol{\theta}) = 2\frac{\partial\rho_{\boldsymbol{\theta}}}{\partial\theta_i}, \tag{9}$$

while in general a solution is [54]

$$L_i(\boldsymbol{\theta}) = 2\int_0^{+\infty} \mathrm{e}^{-s\rho_{\boldsymbol{\theta}}}\frac{\partial\rho_{\boldsymbol{\theta}}}{\partial\theta_i}\,\mathrm{e}^{-s\rho_{\boldsymbol{\theta}}}\,\mathrm{d}s. \tag{10}$$

Throughout the paper we will assume the QFI to be limited (i.e. $\| F(\boldsymbol{\theta}) \| < \infty$) and non-singular (i.e. $F(\boldsymbol{\theta}) > 0$). In particular the last requirement imposes that the maximum value of d (the number of parameters) we can allow in our study is upper bounded by $D^2 - 1$ with D being the dimension of the Hilbert space associated with the probing system (indeed values of d greater than such limit will necessarily force a linear dependence between the SLD operators $L_i(\boldsymbol{\theta})$, leading to a singular QFI matrix).

2.2. Achievability of the multi-parameter QCR bound

In general the multiparameter QCR bound (6) cannot be saturated, meaning that there is no locally unbiased POVM M_N with a $\Sigma^{(N)}(\boldsymbol{\theta})$ matrix equal to $F^{-1}(\boldsymbol{\theta})/N$ or, equivalently, which is capable of saturating the inequality

$$\mathrm{Tr}[G\cdot\Sigma^{(N)}(\boldsymbol{\theta})] \geqslant \frac{1}{N}\,\mathrm{Tr}[G\cdot F^{-1}(\boldsymbol{\theta})] := \frac{C_{\mathrm{S}}\left(G,\boldsymbol{\theta}\right)}{N}, \tag{11}$$

for all choices of a positive weight matrix $G \geqslant 0$. This is the form of metrological incompatibility that will be extensively studied in this paper. In order to better appreciate the meaning of this, suppose that we are interested in the estimation of an analytic function $f \in \mathcal{C}^{\omega}(\Theta)$ of the unknown parameters vector $\boldsymbol{\theta}$. The function f will be evaluated on the estimator $\hat{\boldsymbol{\theta}}$ extracted from the observations. By expanding to first order the expectation value of $f(\hat{\boldsymbol{\theta}}) - f(\boldsymbol{\theta})$ we get the expression for the error

$$\varepsilon := \mathbb{E}[(f(\hat{\boldsymbol{\theta}}) - f(\boldsymbol{\theta}))^2], \tag{12}$$

$$\simeq \sum_{i,j}\mathbb{E}[\partial_i f(\boldsymbol{\theta})(\hat{\theta}_i - \theta_i)\partial_j f(\boldsymbol{\theta})(\hat{\theta}_j - \theta_j)]$$

$$= \sum_{i,j}\partial_i f(\boldsymbol{\theta})\partial_j f(\boldsymbol{\theta})\mathbb{E}[(\hat{\theta}_i - \theta_i)(\hat{\theta}_j - \theta_j)], \tag{13}$$

which can be equivalently written as:

$$\varepsilon = \langle\partial f(\boldsymbol{\theta})|\Sigma^{(N)}(\boldsymbol{\theta})|\partial f(\boldsymbol{\theta})\rangle = \mathrm{Tr}[G(\boldsymbol{\theta})\cdot\Sigma^{(N)}(\boldsymbol{\theta})], \tag{14}$$

where we introduced the rank-1 weight matrix $G_{ij}(\boldsymbol{\theta}) = \partial_i f(\boldsymbol{\theta})\partial_j f(\boldsymbol{\theta}) = |\partial f(\boldsymbol{\theta})\rangle\langle\partial f(\boldsymbol{\theta})|$, with $|\partial f(\boldsymbol{\theta})\rangle \in \mathbb{R}^3$. Written in this form we can now use (11) to cast a bound on the accuracy of the estimation of $f(\boldsymbol{\theta})$. As a matter of fact a rank-1G can always be thought as the weight matrix of some function $f(\boldsymbol{\theta})$. We will see that according to our definitions a rank-1G manifests no incompatibility, indeed we will see that the error associated to a single tangent vector $|\partial f(\boldsymbol{\theta})\rangle$ on the statistical manifold can saturate the ultimate QFI (this can be understood e.g. from the upper bound (28) discussed in section 3.5 below, which, for G rank-1, collapses to $C_S(G, \boldsymbol{\theta})$). On the contrary the gap manifests itself when the weight matrix G is at least rank-2. This situations arises as we try to estimate at the same time multiple functions of the parameters $\boldsymbol{\theta}$, named $f_1(\boldsymbol{\theta}), f_2(\boldsymbol{\theta}), \ldots, f_K(\boldsymbol{\theta})$, which could also just be the components $\theta_1, \theta_2, \ldots, \theta_d$ of the vector $\boldsymbol{\theta}$. To each of the functions we associate a weight $g_i \geqslant 0$, then the total error is the weighted sum of the errors for the estimation of each $f_i(\boldsymbol{\theta})$, i.e.

$$\varepsilon := \sum_{i=1}^{K} g_i \ \mathrm{Tr}[|\partial f_i(\boldsymbol{\theta})\rangle\langle\partial f_i(\boldsymbol{\theta})|\Sigma^{(N)}] = \mathrm{Tr}[G(\boldsymbol{\theta}) \cdot \Sigma^{(N)}], \tag{15}$$

with $G(\boldsymbol{\theta}) := \sum_{i=1}^{K} g_i |\partial f_i(\boldsymbol{\theta})\rangle\langle\partial f_i(\boldsymbol{\theta})| \geqslant 0$.

3. Incompatibility measure

In this section we introduce a figure of merit to gauge the incompatibility of multi-parameter estimation procedures, which is based on the assumption that the agent is allowed to perform on the probes arbitrary locally unbiased POVMs. After showing its well-definedness we clarify its interpretation in the framework of information geometry. We then provide a linear program to compute this incompatibility measure and an analytical upper bound. The figure of merit is then generalized to separable measurements.

3.1. Definition of the figure of merit

Given the encoding (1) and a generic weight matrix G, from (11) it follows that a bona-fide evaluation of the precision attainable with a locally unbiased POVM M_N can be obtained by considering the ratio

$$r_N(G, M_N, \boldsymbol{\theta}) := \frac{N \ \mathrm{Tr}\left[G \cdot \Sigma^{(N)}(\boldsymbol{\theta})\right]}{\mathrm{Tr}\left[G \cdot F^{-1}(\boldsymbol{\theta})\right]} \geqslant 1, \tag{16}$$

where $\Sigma^{(N)}(\boldsymbol{\theta})$ is the MSE matrix (4) associated with M_N. As indicated by the notation the quantity (16) exhibits an explicitly functional dependence on G and M_N which we remove by considering the term

Definition 3.1. (Incompatibility figure of merit for *N* probes)

$$r_N(\boldsymbol{\theta}) := \inf_{M_N \in \mathcal{M}_N^{(\mathrm{LU})}} \sup_{G \geqslant 0} r_N(G, M_N, \boldsymbol{\theta}), \tag{17}$$

where now $\mathcal{M}_N^{(\mathrm{LU})}$ indicates the set of locally unbiased POVM on N copies of the probes. For any given elements M_N of $\mathcal{M}_N^{(\mathrm{LU})}$ the $\sup_{G \geqslant 0}$ selects the weight matrix that has the reachable precision $\mathrm{Tr}\left[G \cdot \Sigma^{(N)}(\boldsymbol{\theta})\right]$ as far away from the information content $\mathrm{Tr}\left[G \cdot F^{-1}(\boldsymbol{\theta})\right]/N$ as possible. Then we minimize on $M_N \in \mathcal{M}_N^{(\mathrm{LU})}$ to compute the best worst case scenario, as in a typical min-max definition [55]. The figure of merit $r_N(\boldsymbol{\theta})$ quantifies the competition between optimal measurements for different parameters, and has a clear operational meaning. Because of the QCR bound in (16) we have $r_N(\boldsymbol{\theta}) \geqslant 1$ and the N define a fully compatible model only when $r_N(\boldsymbol{\theta}) = 1$. This is true if and only if $\exists\, M_N \in \mathcal{M}_N^{(\mathrm{LU})}$ (possibly dependent on $\boldsymbol{\theta}$) for which in (16) equality holds $\forall G$. On the contrary $r_N(\boldsymbol{\theta}) > 1$ indicates the presence of incompatibility and happens if and only if $\forall\, M_N \in \mathcal{M}_N^{(\mathrm{LU})}\ \exists\, G \geqslant 0$ such that in (16) the strict inequality holds. In the asymptotic scenario of infinitely many probes available we introduce

Definition 3.2. (Incompatibility figure of merit)

$$\underline{r}(\boldsymbol{\theta}) := \liminf_{N \to \infty} r_N(\boldsymbol{\theta}), \tag{18}$$

which always exists and from $r_N(\boldsymbol{\theta})$ inherits the property $\underline{r}(\boldsymbol{\theta}) \geqslant 1$. In particular in this case we have $\underline{r}(\boldsymbol{\theta}) = 1$ if and only if there exists a sequence of $M_N \in \mathcal{M}_N^{(\mathrm{LU})}$, which, for all $G \geqslant 0$, allows us to saturate the inequality (16) asymptotically in N. It is worth noticing that the incompatibility figure of merit could be defined for locally asymptotic covariant measurements (LAC) [42, 48] as well, and it would be exactly equal to $\underline{r}(\boldsymbol{\theta})$, see appendix A for the details.

3.2. Well-definedness of the figure of merit

We now briefly show that $r_N(\boldsymbol{\theta})$ in (17) is invariant under reparametrization and therefore a well-defined quantity. This translates to $\underline{r}(\boldsymbol{\theta})$, which is therefore a well defined property of the statistical manifold (see section 3.3). Consider a reparametrization $\boldsymbol{\theta} = \boldsymbol{\theta}(\boldsymbol{\eta})$ having an invertible Jacobian $J_{ij} := \frac{\partial \theta_i(\boldsymbol{\eta})}{\partial \eta_j}$. Then the MSE matrix for the parameters $\boldsymbol{\theta}$, defined in (4), can be written $\Sigma^{(N)}(\boldsymbol{\theta}) = J\Sigma^{(N)}(\boldsymbol{\eta})J^t$, where $\Sigma^{(N)}_{ij}(\boldsymbol{\eta}) := \mathbb{E}[(\hat{\eta}_i - \eta_i)(\hat{\eta}_j - \eta_j)]$ is the MSE matrix for the parameters $\boldsymbol{\eta}$. Similarly we write the inverse of the QFI matrix as $F^{-1}(\boldsymbol{\theta}) = JF^{-1}(\boldsymbol{\eta})J^t$, with $F(\boldsymbol{\eta})$ computed from the SLDs $L_i(\boldsymbol{\eta})$, which differ from the definition in (8) in the derivatives, which are taken with respect to η_i. Its now easy to show that $r_N(\boldsymbol{\theta}) = r_N(\boldsymbol{\eta})$. The action of the Jacobian matrix on the MSE matrix and on the QFI can be moved on G, that becomes J^tGJ both at numerator and at denominator of the ratio in (16), while multiplying respectively $\Sigma^{(N)}(\boldsymbol{\eta})$ and $F^{-1}(\boldsymbol{\eta})$. Then we observe that the set of positive matrices is invariant under congruence for an invertible matrix, i.e. $J^t\{G \geqslant 0\}J = \{G \geqslant 0\}$, and therefore we get

$$r_N(\boldsymbol{\theta}) = \inf_{M_N \in \mathcal{M}_N^{(\mathrm{LU})}} \sup_{G \geqslant 0} \frac{N \ \mathrm{Tr}\left[G \cdot \Sigma^{(N)}(\boldsymbol{\eta})\right]}{\mathrm{Tr}\left[G \cdot F^{-1}(\boldsymbol{\eta})\right]} := r_N(\boldsymbol{\eta}). \tag{19}$$

It worth stressing that, by construction the quantity $\underline{r}(\boldsymbol{\theta})$ only depends on the input probe state ρ, the encoding $\mathcal{E}_{\boldsymbol{\theta}}$, and the specific point of interest $\boldsymbol{\theta}$. It is an intrinsic property of the statistical manifold defined by the trajectories (1). The need for a reparametrization invariant measure of incompatibility was already pointed out in [56], in which the figure of merit $\underline{r}(\boldsymbol{\theta})$ was independently discovered.

3.3. Geometric interpretation

The parameters $\boldsymbol{\theta} \in \Theta$ can be interpreted as coordinates defining via the map (1) a submanifold of the space of states $\mathfrak{S}(\mathcal{H})$, called the statistical manifold. The QFI matrix, being a positive semidefinite matrix can be thought as a Riemannian metric on this manifold. This metric is generally non trivial as it explicitly depends on the coordinates $\boldsymbol{\theta}$ and may have intrinsic curvature. The QFI is said to be a distinguishability metric [57, 58]: given two very near states $\rho_{\boldsymbol{\theta}}$ and $\rho_{\boldsymbol{\theta}+\mathrm{d}\boldsymbol{\theta}}$, their infinitesimal distance in the QFI metric is

$$\mathrm{d}s^2 := \frac{1}{4}F_{ij}(\boldsymbol{\theta})\mathrm{d}\theta_i\,\mathrm{d}\theta_j = 2\left(1 - \sqrt{\mathcal{F}\left(\rho_{\boldsymbol{\theta}}, \rho_{\boldsymbol{\theta}+\mathrm{d}\boldsymbol{\theta}}\right)}\right), \tag{20}$$

which is negatively correlated with the fidelity $\mathcal{F}(\rho_{\boldsymbol{\theta}}, \rho_{\boldsymbol{\theta}+\mathrm{d}\boldsymbol{\theta}})$ between $\rho_{\boldsymbol{\theta}}$ and $\rho_{\boldsymbol{\theta}+d\boldsymbol{\theta}}$, defined as $\mathcal{F}(\rho, \sigma) := \left[\mathrm{Tr}\left(\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}}\right)\right]^2$ [59]. In order to gain information about $\boldsymbol{\theta}$ it is thus better to choose the probe state ρ such that in the statistical manifold the codified state $\rho_{\boldsymbol{\theta}}$ is highly distinguishable from its neighbors $\rho_{\boldsymbol{\theta}+\mathrm{d}\boldsymbol{\theta}}$, and has therefore the highest statistical distance from them as possible. This picture clarifies why the inverse of the distinguishability metric, i.e. $F^{-1}(\boldsymbol{\theta})$, gives the precision to which a single point $\boldsymbol{\theta}$ can be identified in Θ, given the quantum state $\rho_{\boldsymbol{\theta}}$. For $\rho_{\boldsymbol{\theta}}^{\otimes N}$ the relevant metric is $F^{-1}(\boldsymbol{\theta})/N$. When a measurement is performed and an estimator $\hat{\boldsymbol{\theta}}$ is chosen there is a new Riemannian metric insisting on the statistical manifold: the positive semidefinite $\Sigma^{(N)}(\boldsymbol{\theta})$ matrix. The key question is if one can find a POVM $M_N \in \mathcal{M}_N^{(\mathrm{LU})}$ with an MSE metric that fully adapts to the underling quantum metric $F^{-1}(\boldsymbol{\theta})/N$ of the manifold, i.e. if the inequality (6) can be saturated (at a certain point $\boldsymbol{\theta}$). In general this is not possible. Let us introduce a representation of G as a sum of projectors $|v_i\rangle\langle v_i|$, each weighted with $g_i \geqslant 0$, i.e. $G := \sum_i g_i|v_i\rangle\langle v_i|$, where $|v_i\rangle$ are vectors in the tangent plane of the statistical manifold at point $\boldsymbol{\theta}$, then

$$\frac{1}{N}\ \mathrm{Tr}[G \cdot F^{-1}(\boldsymbol{\theta})] = \frac{1}{N}\sum_i g_i\langle v_i|F^{-1}(\boldsymbol{\theta})|v_i\rangle. \tag{21}$$

According to the above expression, the information content is a weighted combination of the distinguishability of the manifold in different directions defined by $|v_i\rangle$, see figure 1 for a 2D representation. This has to be compared with the experimental weighted distinguishability, i.e.

$$\mathrm{Tr}[G \cdot \Sigma^{(N)}(\boldsymbol{\theta})] = \sum_i g_i\langle v_i|\Sigma^{(N)}(\boldsymbol{\theta})|v_i\rangle, \tag{22}$$

given by a particular measurement. The whole point of the non commutative nature of the manifold is the impossibility to saturate the distinguishability in more than one direction at the same time. By taking

$$\sup_{G \geqslant 0} \frac{N \ \mathrm{Tr}\left[G \cdot \Sigma^{(N)}(\boldsymbol{\theta})\right]}{\mathrm{Tr}\left[G \cdot F^{-1}(\boldsymbol{\theta})\right]} = \sup_{g_i \geqslant 0, |v_i\rangle} \frac{N\sum_i g_i\langle v_i|\Sigma^{(N)}(\boldsymbol{\theta})|v_i\rangle}{\sum_i g_i\langle v_i|F^{-1}(\boldsymbol{\theta})|v_i\rangle}, \tag{23}$$

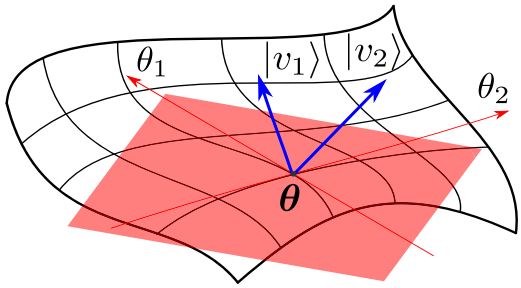

Figure 1. Representation of a 2D statistical manifold with its tangent space at a point $\boldsymbol{\theta}$ and two directions $|v_1\rangle$ and $|v_2\rangle$ on this plane.

we measure the worst case fitting of the $\Sigma^{(N)}(\boldsymbol{\theta})$ matrix on the metric $F^{-1}(\boldsymbol{\theta})/N$ at a point $\boldsymbol{\theta}$, spanning all possible sets of tangent vectors and weights. Then we minimize on the classical metric (and hence on the POVM) to find the most adapt one. By taking the asymptotic limit of infinitely many probes (through the lim inf) we have completed the analysis of the definition (18) from the geometrical point of view. We sum up everything and say that $\underline{r}(\boldsymbol{\theta})$ measures, in the asymptotic scenario, the failure of finding a metric on the statistical manifold, stemming from a measurement, which fully adapts to the underlying quantum metric (in all directions) at a specific point $\boldsymbol{\theta}$.

3.4. Computation of the figure of merit

We would like to apply the existing results in local estimation theory to compute the incompatibility figure of merit. This requires the exchange of the sup and the inf in (17). In appendix B we do that and show

$$\underline{r}(\boldsymbol{\theta}) = \sup_{G\geqslant 0} \frac{C(G,\boldsymbol{\theta})}{C_{\mathrm{S}}(G,\boldsymbol{\theta})}, \tag{24}$$

with $C_{\mathrm{S}}(G,\boldsymbol{\theta})$ defined in (11), and with the numerator given by the quantity

$$C(G,\boldsymbol{\theta}) := \lim_{n\to\infty} \min_{N\leqslant n} \inf_{M_N\in\mathcal{M}_N^{(\mathrm{LU})}} N\ \mathrm{Tr}[G\cdot\Sigma^{(N)}(\boldsymbol{\theta})]. \tag{25}$$

In [42] it has been proved that $C(G,\boldsymbol{\theta}) = C_{\mathrm{H}}(G,\boldsymbol{\theta})$, where $C_{\mathrm{H}}(G,\boldsymbol{\theta})$ is the Holevo–Cramér–Rao bound functional $C_{\mathrm{H}}(G,\boldsymbol{\theta})$ [41, 60]. Exploiting this facts we can use (24) to deduce the following equality

$$\underline{r}(\boldsymbol{\theta}) = \sup_{G\geqslant 0} \frac{C_{\mathrm{H}}(G,\boldsymbol{\theta})}{C_{\mathrm{S}}(G,\boldsymbol{\theta})}. \tag{26}$$

In [61] the upper bound $C_{\mathrm{H}}(G,\boldsymbol{\theta}) \leqslant 2C_{\mathrm{S}}(G,\boldsymbol{\theta})$ is given, which implies $\underline{r}(\boldsymbol{\theta}) \leqslant 2$. Because of this it makes sense to introduce

Definition 3.3. (Incompatibility measure)

$$\mathcal{I}(\boldsymbol{\theta}) := \underline{r}(\boldsymbol{\theta}) - 1, \tag{27}$$

as a proper quantifier of incompatibility: by construction it belongs to the interval $[0,1]$ with $\mathcal{I}(\boldsymbol{\theta}) = 0$ indicating full compatibility, while $\mathcal{I}(\boldsymbol{\theta}) = 1$ maximal incompatibility. The Holevo–Cramér–Rao bound can be computed via the semidefinite linear program in [49], which can be adapted to compute $\mathcal{I}(\boldsymbol{\theta})$, as reported in appendix C.

3.5. Upper bound on $\underline{r}(\boldsymbol{\theta})$

In this section we propose an upper bound on $\underline{r}(\boldsymbol{\theta})$ that relies only on the computation of the SLDs defined in (8). It is essentially based on $C_{\mathrm{Z}}(G,\boldsymbol{\theta})$ [62], a well know upper bound on $C_{\mathrm{H}}(G,\boldsymbol{\theta})$, which reads

$$C_{\mathrm{H}}(G,\boldsymbol{\theta}) \leqslant C_{\mathrm{Z}}(G,\boldsymbol{\theta}) := \mathrm{Tr}\left[GF^{-1}(\boldsymbol{\theta})\right] + \mathrm{Tr}\ \mathrm{Abs}\left[GF^{-1}(\boldsymbol{\theta})A(\boldsymbol{\theta})F^{-1}(\boldsymbol{\theta})\right], \tag{28}$$

where $A(\boldsymbol{\theta})$ contains the expectation values of the commutators of the SLDs:

$$A_{ij}(\boldsymbol{\theta}) := \frac{1}{2\mathrm{i}}\ \mathrm{Tr}\left[\rho_{\boldsymbol{\theta}}\left[L_i(\boldsymbol{\theta}), L_j(\boldsymbol{\theta})\right]\right]. \tag{29}$$

In writing (28) we used $\mathrm{Tr}\ \mathrm{Abs}\,[G\cdot R] := \mathrm{Tr}\,|\sqrt{G}R\sqrt{G}|$, with $|X| := \sqrt{XX^\dagger}$. Combining (28) and (26) we get

$$\underline{r}(\boldsymbol{\theta}) \leqslant 1 + \sup_{G\geqslant 0} \frac{\mathrm{Tr}\ \mathrm{Abs}\left[G\cdot F^{-1}(\boldsymbol{\theta})A(\boldsymbol{\theta})F^{-1}(\boldsymbol{\theta})\right]}{\mathrm{Tr}\left[G\cdot F^{-1}(\boldsymbol{\theta})\right]} := \underline{r}^\star(\boldsymbol{\theta}). \tag{30}$$

The above inequality shows that a sufficient condition to have compatibility is $A(\boldsymbol{\theta}) = 0$. In appendix D we compute explicitly $\sup_{G\geqslant 0}$ in (30) and obtain

$$\underline{r}^\star(\boldsymbol{\theta}) = 1 + \|F^{-\frac{1}{2}}(\boldsymbol{\theta})A(\boldsymbol{\theta})F^{-\frac{1}{2}}(\boldsymbol{\theta})\|, \tag{31}$$

where $\|\cdot\|$ is the operator norm. This translates to an upper bound on $\mathcal{I}(\boldsymbol{\theta})$, i.e.

Theorem 3.1. (Upper bound on the incompatibility measure)

$$\mathcal{I}(\boldsymbol{\theta}) \leqslant \mathcal{I}^\star(\boldsymbol{\theta}) := \underline{r}^\star(\boldsymbol{\theta}) - 1 = \|F^{-\frac{1}{2}}(\boldsymbol{\theta})A(\boldsymbol{\theta})F^{-\frac{1}{2}}(\boldsymbol{\theta})\|. \tag{32}$$

This strengthen the interpretation of $A(\boldsymbol{\theta})$ as a measure of incompatibility [51]. The upper bound $\mathcal{I}^\star(\boldsymbol{\theta})$ was first defined in [63] and called $\mathcal{R}$. It has already been used as a measure of incompatibility and 'quantumness' and applied to qubits [56] and many-body systems [63, 64]. By defining $\mathcal{I}(\boldsymbol{\theta})$ we offer a more informative definition of incompatibility. It is noteworthy that for a D-invariant model [65] this bound is saturated and $\mathcal{I}(\boldsymbol{\theta}) = \mathcal{I}^\star(\boldsymbol{\theta})$.

3.6. Incompatibility for separable measurements

We now go back to the first definition of a figure of merit presented in (18), but consider the minimization in (17) to be performed only on the locally unbiased separable measurements subset $\mathcal{M}_N^{(\mathrm{LU-S})}$ of $\mathcal{M}_N^{(\mathrm{LU})}$ which operate locally on $\rho_{\boldsymbol{\theta}}^{\otimes N}$. This brings to the definitions

$$r_N^{\mathrm{s}}(\boldsymbol{\theta}) := \inf_{M_N\in\mathcal{M}_N^{(\mathrm{LU-S})}} \sup_{G\geqslant 0} r_N\left(G, M_N, \boldsymbol{\theta}\right), \tag{33}$$

and

Definition 3.4. (Incompatibility figure of merit for separable measurements)

$$\underline{r}^{\mathrm{s}}(\boldsymbol{\theta}) := \liminf_{N\to\infty} r_N^{\mathrm{s}}(\boldsymbol{\theta}). \tag{34}$$

Now we apply the result of [66], which gives us a lower bound on the precision of the estimation with N probes when we use a measurement $M_N \in \mathcal{M}_N^{(\mathrm{LU-S})}$. The bound reads

$$N\ \mathrm{Tr}[G\cdot\Sigma^{(N)}(\boldsymbol{\theta})] \geqslant \frac{1}{D-1}\left(\mathrm{Tr}\sqrt{F^{-\frac{1}{2}}(\boldsymbol{\theta})GF^{-\frac{1}{2}}(\boldsymbol{\theta})}\right)^2, \tag{35}$$

where $\Sigma^{(N)}(\boldsymbol{\theta})$ is the MSE matrix of M_N and D is the size of the Hilbert space of the single probe $\rho_{\boldsymbol{\theta}}$. This translates to a lower bound on $r_N^{\mathrm{s}}\left(G, M_N, \boldsymbol{\theta}\right) \forall\ N$, i.e.

$$r_N^{\mathrm{s}}\left(G, M_N, \boldsymbol{\theta}\right) \geqslant \frac{\left(\mathrm{Tr}\sqrt{F^{-\frac{1}{2}}(\boldsymbol{\theta})GF^{-\frac{1}{2}}(\boldsymbol{\theta})}\right)^2}{(D-1)\,\mathrm{Tr}\left[F^{-\frac{1}{2}}(\boldsymbol{\theta})GF^{-\frac{1}{2}}(\boldsymbol{\theta})\right]}, \tag{36}$$

which propagates to the definition of $\underline{r}^{\mathrm{s}}(\boldsymbol{\theta})$, giving

Theorem 3.2. (Lower bound for separable measurements)

$$\underline{r}^{\mathrm{s}}(\boldsymbol{\theta}) \geqslant \frac{d}{D-1}, \tag{37}$$

where we have compute explicitly $\sup_{G\geqslant 0}$ using the AM-QM inequality and its saturation. Observe that the inequality (37) bares no reference to the details of the encoding process (1) and that it is non trivial only if the number d of parameters we have to estimate is larger than or equal to $D-1$. The manipulations of appendix B are valid also for the class of measurements $\mathcal{M}_N^{(\mathrm{LU-S})}$, because only the local unbiasedness is required in their proof. Therefore we can write

$$\underline{r}^{\mathrm{s}}(\boldsymbol{\theta}) = \sup_{G\geqslant 0} \frac{C^{\mathrm{s}}\left(G, \boldsymbol{\theta}\right)}{C_{\mathrm{S}}\left(G, \boldsymbol{\theta}\right)}, \tag{38}$$

where now

$$C^{\mathrm{s}}(G,\boldsymbol{\theta}) := \lim_{n\to\infty} \min_{N\leqslant n} \inf_{M_N\in\mathcal{M}_N^{(\mathrm{LU-S})}} N\ \mathrm{Tr}[G\cdot\Sigma^{(N)}(\boldsymbol{\theta})]. \tag{39}$$

At least for the case of a qubit probe ($D=2$) the above expression allows us to exactly compute $\underline{r}^{\mathrm{s}}(\boldsymbol{\theta})$. Indeed as shown in [60, 66] for this model one has

$$C^{\mathrm{s}}(G,\boldsymbol{\theta}) = \left(\mathrm{Tr}\sqrt{F^{-\frac{1}{2}}(\boldsymbol{\theta})GF^{-\frac{1}{2}}(\boldsymbol{\theta})}\right)^2, \tag{40}$$

leading to

Corollary 1.

$$D=2 \quad\Rightarrow\quad \underline{r}^{\mathrm{s}}(\boldsymbol{\theta}) = d, \tag{41}$$

which shows that in the case of a single qubit, multi-parameter estimation always exhibit incompatibility for separable locally unbiased measurements (remember that our analysis is explicitly restricted to the cases where $d\leqslant D^2-1=3$).

3.7. Hierarchy of incompatibility measures

Whether a certain estimation process is compatible or not depends on the set of measurements $\mathcal{M}_N$ that we are allowed to perform. Consider a hierarchy of POVM sets

$$\mathcal{M}_N^{(1)}\subseteq\mathcal{M}_N^{(2)}\subseteq\ldots\subseteq\mathcal{M}_N^{(k)}\quad\forall\ N, \tag{42}$$

we define the figure of merit $r_N^{(i)}(\boldsymbol{\theta})$ as in (17), but taking $\mathcal{M}_N^{(i)}$ as the domain of the infimum. By construction the $r_N^{(i)}(\boldsymbol{\theta})$ satisfy the following hierarchy of inequalities

$$r_N^{(1)}(\boldsymbol{\theta})\geqslant r_N^{(2)}(\boldsymbol{\theta})\geqslant\ldots\geqslant r_N^{(k)}(\boldsymbol{\theta})\quad\forall\ N, \tag{43}$$

which carries over to

$$\underline{r}^{(1)}(\boldsymbol{\theta})\geqslant\underline{r}^{(2)}(\boldsymbol{\theta})\geqslant\ldots\geqslant\underline{r}^{(k)}(\boldsymbol{\theta}), \tag{44}$$

when taking the proper $N\to\infty$ limits (18). For example the space of separable locally unbiased measurement is a subset of the set of all locally unbiased measurements, i.e.

$$\mathcal{M}_N^{(\mathrm{LU-S})}\subseteq\mathcal{M}_N^{(\mathrm{LU})}, \tag{45}$$

which means $\underline{r}^{\mathrm{s}}(\boldsymbol{\theta})\geqslant\underline{r}(\boldsymbol{\theta})$.

4. Incompatibility of a noisy estimation task

In this section, by using the previously defined figures of merit $\underline{r}(\boldsymbol{\theta})$ in (26) and $\underline{r}^{\mathrm{s}}(\boldsymbol{\theta})$ in (34), we study the incompatibility of the estimation process in a few simple cases concerning the sensing of two phases θ_1 and θ_2 encoded by the unitary transformation

$$U_{\boldsymbol{\theta}} := \exp\left[\mathrm{i}\left(\theta_1\sigma_y+\theta_2\sigma_z\right)\right], \tag{46}$$

acting on individual qubits. The probes will be states of one and three qubits subject to local depolarizing noise, which is given by the map

$$\Lambda_\lambda(\rho) := \lambda\rho+(1-\lambda)\frac{\mathbb{1}}{2}, \tag{47}$$

with $\lambda\in[-1/3,1]$ being a characteristic parameter of the model [67]. The transformation Λ_λ induces a shrinking of the qubit Bloch vector by a factor given by the modulus $|\lambda|$ which can be used to gauge the intensity of the noise. In particular for $\lambda=1$ the map (47) corresponds to the noiseless evolution, and for $\lambda=0$ to the complete depolarization process, while negative values of λ indicate the presence of an inversion of the Bloch sphere with respect to the origin [68]. We are interested in investigating if the noise can force the system to a more classical behavior and therefore ensure compatibility in the estimation scenario, as it does for measurements [40]. We then turn to D-dimensional system, and with the opportune generalizations of (46) and (47) we explore the upper bound $\mathcal{I}^\star(\boldsymbol{\theta})$ in (27) for a generic system and the incompatibility $\mathcal{I}(\boldsymbol{\theta})$ in (32) for a qutrit. Notice that the chosen noise is covariant and therefore in all our examples it could be applied before or after the encoding without changing the final output $\rho_{\boldsymbol{\theta}}$. Table 1 contains a recap of the improvements and observed phenomena in the following examples.

Table 1. Recap of the examples of section 4 with the improvements we propose and/or the observed relevant phenomena.

System	Known results	Improvements/observed phenomena
1 qubit	Computation of $\mathcal{I}(\boldsymbol{\theta})$ ($\mathcal{R}$) for qubit tomography and two phase estimation with pure states [56]	Computation of $\mathcal{I}(\boldsymbol{\theta},\lambda)$ for two phase estimation with depolarizing noise. Incompatibility with separable measurements
3 qubit	Computation of $\parallel A(\boldsymbol{\theta}) \parallel_F$ and $1 - C_{\mathrm{H}}(\mathbb{1})/C_{\mathrm{S}}(\mathbb{1})$ [49]	Efficient computation of $\mathcal{I}(\boldsymbol{\theta},\lambda)$. Observation of gap between $\mathcal{I}^\star(\boldsymbol{\theta},\lambda)$ and $\mathcal{I}(\boldsymbol{\theta},\lambda)$ and its behavior
1 D-dim	—	Asymmetry around $\lambda = 0$ of the upper bound $\mathcal{I}^\star(\boldsymbol{\theta},\lambda)$
1 qutrit	—	Asymmetry around $\lambda = 0$ of the incompatibility measure $\mathcal{I}(\boldsymbol{\theta},\lambda)$

4.1. Incompatibility for a one-qubit probe

First of all we analyze the case of a single qubit probe. The fact that the figure of merit is parameterization invariant allows for an elegant exact solution of the qubit model for whatever probe state and encoded phases under depolarization noise. In this example the measure $\mathcal{I}(\boldsymbol{\theta})$ and its upper bound $\mathcal{I}^\star(\boldsymbol{\theta})$ will coincide. After the encoding by $U_{\boldsymbol{\theta}}$ in (46), the probe undergoes the action of the noise map Λ_λ in (47), so that its final state $\rho_{\boldsymbol{\theta}}$ is described by the mapping (1) with $\mathcal{E}_{\boldsymbol{\theta}}$ given by

$$\mathcal{E}_{\boldsymbol{\theta}}(\rho) := \Lambda_\lambda(U_{\boldsymbol{\theta}}\rho U_{\boldsymbol{\theta}}^\dagger) = U_{\boldsymbol{\theta}}\Lambda_\lambda(\rho)U_{\boldsymbol{\theta}}^\dagger. \tag{48}$$

The purity of the encoded state $\rho_{\boldsymbol{\theta}} = \mathcal{E}_{\boldsymbol{\theta}}(\rho)$ is independent on $\boldsymbol{\theta}$, this makes the statistical model D-invariant [62, 65], and allows us to conclude that the Holevo–Cramér–Rao bound $C_{\mathrm{H}}(G,\boldsymbol{\theta})$ coincides with $C_{\mathrm{Z}}(G,\boldsymbol{\theta})$ defined in (28), therefore the inequality (30) is saturated ($\mathcal{I}(\boldsymbol{\theta}) = \mathcal{I}^\star(\boldsymbol{\theta})$), and the incompatibility can be computed from the SLDs only. We consider an arbitrary qubit probe state $\rho := \frac{1}{2}(\mathbb{1} + \boldsymbol{a}\cdot\boldsymbol{\sigma})$. Its Bloch vector is $\boldsymbol{a} := (a_x, a_y, a_z)$, with $\mathrm{Tr}\,\rho^2 = \frac{1}{2}(1 + \|\boldsymbol{a}\|^2)$. After the encoding the Bloch vector of $\rho_{\boldsymbol{\theta}} = \mathcal{E}_{\boldsymbol{\theta}}(\rho)$ is $\boldsymbol{a}_{\boldsymbol{\theta}} := \lambda\left(a_x(\boldsymbol{\theta}), a_y(\boldsymbol{\theta}), a_z(\boldsymbol{\theta})\right)$. We can perform an implicitly defined change of variables $(\theta_1,\theta_2) \to (\alpha,\beta)$, that brings us to $\boldsymbol{a}_{(\alpha,\beta)} = \lambda\sqrt{2\,\mathrm{Tr}\,\rho^2 - 1}(\cos\alpha\cos\beta, \cos\alpha\sin\beta, \sin\alpha)$. For this model [69] we compute the matrices $F(\alpha,\beta)$ and the $A(\alpha,\beta)$, which are

$$F(\alpha,\beta) = \sqrt{2\,\mathrm{Tr}\,\rho^2 - 1}\begin{pmatrix}\lambda^2 & 0\\ 0 & \lambda^2\cos^2\alpha\end{pmatrix},$$

$$A(\alpha,\beta) = \sqrt{2\,\mathrm{Tr}\,\rho^2 - 1}\begin{pmatrix}0 & -\lambda^3\cos\alpha\\ \lambda^3\cos\alpha & 0\end{pmatrix},$$

that substituted in (32) give

Theorem 4.1. (Incompatibility measure for a depolarized qubit two phase model)

$$\mathcal{I}(\boldsymbol{\theta},\lambda) = \mathcal{I}(\alpha,\beta,\lambda) = \sqrt{2\,\mathrm{Tr}\,\rho^2 - 1}|\lambda| \quad \forall\,\boldsymbol{\theta},\ \forall\,\rho. \tag{49}$$

Equation (49) reveals that the noise level intensity controls directly the compatibility. Indeed for fixed input the value of $\mathcal{I}(\boldsymbol{\theta},\lambda)$ reaches its maximum in the noiseless scenario ($\lambda \to 1$) providing full incompatibility $\mathcal{I}(\boldsymbol{\theta},\lambda) \to 1$ for pure input states. On the contrary as the noise sends $\rho_{\boldsymbol{\theta}}$ to the completely mixed state ($\lambda \to 0$) the codified information is dissipated and the compatibility increases, indeed $\mathcal{I}(\boldsymbol{\theta},\lambda) \to 0$. Fundamentally the same result was discover in [56] for qubit tomography. We finally remind the reader that, as anticipated at the end of section 3.6, for a single qubit we get $\underline{r}^{\mathrm{s}}(\boldsymbol{\theta}) = d = 2$ independently on the noise. Again this result is valid $\forall\boldsymbol{\theta}$ and for every input probe ρ.

4.2. Incompatibility for three entangled qubits

Consider now the scenario in which we have at disposal multiple copies of three entangled qubits and we codify them through $U_{\boldsymbol{\theta}} \otimes U_{\boldsymbol{\theta}} \otimes U_{\boldsymbol{\theta}}$, with $U_{\boldsymbol{\theta}}$ given in (46). This more complicate scenario gives us the opportunity to compute $\mathcal{I}(\boldsymbol{\theta},\lambda)$ with the SDP and show the presence of a gap between $\mathcal{I}(\boldsymbol{\theta},\lambda)$ and $\mathcal{I}^\star(\boldsymbol{\theta},\lambda)$. In this example we will not be able to compute $\underline{r}(\boldsymbol{\theta})$ for every probe state, therefore we will concentrate on

$$|\psi\rangle := \frac{|\psi_z\rangle + |\psi_y\rangle}{\sqrt{2}}, \tag{50}$$

with

$$|\psi_z\rangle := \frac{1}{\sqrt{2}}\left(|000\rangle + |111\rangle\right),$$

$$|\psi_y\rangle := \frac{1}{\sqrt{2}}\left(|\phi^+\phi^+\phi^+\rangle + |\phi^-\phi^-\phi^-\rangle\right),$$

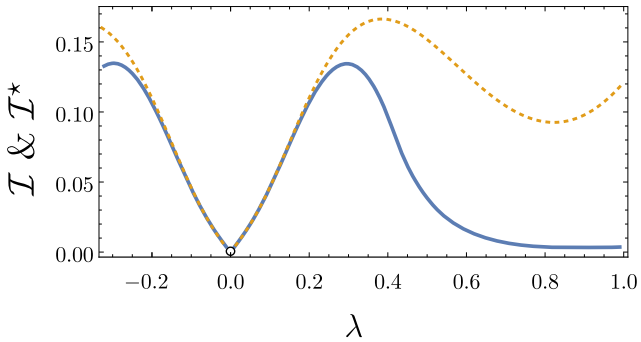

Figure 2. The orange dashed curve is the upper bound $\mathcal{I}^\star(\boldsymbol{\theta}=0)$, defined in (32), computed for the encoded three qubits state (50) as a function of the local noise intensity λ. The blue solid curve is the figure of merit $\mathcal{I}(\boldsymbol{\theta}=0)$ defined in (27) referred to the same scenario and computed numerically as explained in appendix C. The curves are symmetric around $\lambda=0$. The empty point in $\lambda=0$ indicates that at this point the information quantities are not defined.

where $|\phi^+\rangle$ and $|\phi^-\rangle$ are the eigenvectors of σ_y corresponding to the positive and negative eigenvalue respectively. In [38] it is proved that the analogous state for the estimation of three phases with N entangled qubits reaches Heisenberg scaling in the QFI in all the three parameters. At difference with the previous example, here we are able to compute the figure of merit for the probe $|\psi\rangle$ only at the point $\boldsymbol{\theta}=0$ through numerical evaluations via the SDP reported in appendix C, these indicate a non-null $\mathcal{I}(\boldsymbol{\theta}=0)$. We add a local depolarization noise Λ_λ on each qubit and compute $\mathcal{I}(\boldsymbol{\theta}=0,\lambda)$ and its upper bound $\mathcal{I}^\star(\boldsymbol{\theta}=0,\lambda)$ as functions of λ to see if the noise increases compatibility, the results are reported in figure 2. $\mathcal{I}(\boldsymbol{\theta}=0,\lambda)$ and $\mathcal{I}^\star(\boldsymbol{\theta}=0,\lambda)$ have been computed for 100 values of λ uniformly distributed in $\left(-1/3,1\right)$. The addition of noise does not necessarily diminish the incompatibility, on the contrary $\mathcal{I}(\boldsymbol{\theta}=0,\lambda)$ and $\mathcal{I}^\star(\boldsymbol{\theta}=0,\lambda)$ both display a non-monotonic behavior with respect to $|\lambda|$. This behavior of the incompatibility has already been observed in [49]. We notice that as the noise destroys the information codified in $\rho_{\boldsymbol{\theta}}$ both the compatibility and its upper bound $\mathcal{I}^\star(\boldsymbol{\theta},\lambda)$ go to 0, but this does not seem to be a universal behavior [56]. We confirm a separation between $\mathcal{I}^\star(\boldsymbol{\theta},\lambda)$ and $\mathcal{I}(\boldsymbol{\theta},\lambda)$, that has been evidenced in [56], and we conjecture that $\mathcal{I}(\boldsymbol{\theta},\lambda)-\mathcal{I}^\star(\boldsymbol{\theta},\lambda)$ shrinks to zero as the amount of encoded information diminish, as it happens in this example for $\lambda\to 0$. In this model also the relative gap $(\mathcal{I}(\boldsymbol{\theta},\lambda)-\mathcal{I}^\star(\boldsymbol{\theta},\lambda))/\mathcal{I}(\boldsymbol{\theta},\lambda)$ shrinks to zero as $\lambda\to 0$. For a generic noise this phenomenon depends on the behavior of $A(\boldsymbol{\theta},\lambda)$ as the disturbance is increased. The figure of merit $\mathcal{I}(\boldsymbol{\theta},\lambda)$ appears to be not correlated with the information quantities $F_{11}^{-1}(\boldsymbol{\theta},\lambda)$ and $F_{22}^{-1}(\boldsymbol{\theta},\lambda)$ or with the purity of the encoded state, as these measures are all monotonic in the noise λ. Also because of this we think of $\mathcal{I}(\boldsymbol{\theta})$ as a genuine non trivial new property of the estimation process. Notice that for $\lambda=0$, the state is unable to codify information ($F(\boldsymbol{\theta})=0$).

4.3. Estimation on D-dimensional probes

In this section we study the incompatibility for a generic unitary encoding of d parameters on a D-dimensional probe in $\mathfrak{S}(\mathcal{H})$, i.e.

$$U_{\boldsymbol{\theta}} := \exp\left(\mathrm{i}\sum_{j=1}^{d}\theta_j H_j\right), \tag{51}$$

where H_j are null-trace Hermitian operators acting on $\mathcal{H}$. For the estimation around $\boldsymbol{\theta}=0$, these operators are the infinitesimal generators of the encoding. However for a generic point $\boldsymbol{\theta}\neq 0$ this is not necessarily true. As explained in appendix E, for a given probe state, the sensing procedure around a point $\boldsymbol{\theta}\neq 0$ can however be described in terms of an effective set of new generators $H_j^{\mathrm{eff}}(\boldsymbol{\theta})$. Accordingly, since the results of the preset section are valid for estimations around $\boldsymbol{\theta}=0$ for all possible choices of H_j, we can conclude that they hold true also $\forall\boldsymbol{\theta}$ encoded by (51). Finally as for the noise model we replace (47) with

$$\Lambda_\lambda(\rho) := \lambda\rho + (1-\lambda)\,\frac{\mathbb{1}}{D}, \tag{52}$$

which for $\lambda\in[-1/(D^2-1),1]$ is a proper generalization of the depolarization channel for a D-dimensional system [67, 68].

4.4. Incompatibility for a D-dimensional probe

Let us consider a single-probe scenario where the state of the system is described by the density matrix

$$\rho_{\boldsymbol{\theta}} := \Lambda_\lambda(U_{\boldsymbol{\theta}}|\psi\rangle\langle\psi|U_{\boldsymbol{\theta}}^\dagger) = U_{\boldsymbol{\theta}}\Lambda_\lambda(|\psi\rangle\langle\psi|)U_{\boldsymbol{\theta}}^\dagger = \lambda|\psi_{\boldsymbol{\theta}}\rangle\langle\psi_{\boldsymbol{\theta}}| + (1-\lambda)\frac{\mathbb{1}}{D}, \tag{53}$$

with $|\psi\rangle$ being the pure input state of the system, and with $|\psi_{\boldsymbol{\theta}}\rangle := U_{\boldsymbol{\theta}}|\psi\rangle$. If we now call $L_i(\boldsymbol{\theta})$ the SLD associated to the parameter θ_i in the absence of noise, i.e. the SLD of $|\psi_{\boldsymbol{\theta}}\rangle$, given in (9), then it can be seen that for $\lambda \neq 1$

$$L_i(\boldsymbol{\theta}, \lambda) = \frac{\lambda D}{2 + \lambda(D-2)} L_i(\boldsymbol{\theta}), \tag{54}$$

is the SLD in the noisy scenario. We obtain this expression by substituting $\rho_{\boldsymbol{\theta}}$ defined in (53) in (10). From this result the QFI matrix $F(\boldsymbol{\theta}, \lambda)$ and the commutator matrix $A(\boldsymbol{\theta}, \lambda)$ are both found to be proportional to their noiseless counterparts $F(\boldsymbol{\theta})$ and $A(\boldsymbol{\theta})$ computed from $L_i(\boldsymbol{\theta})$, i.e.

$$F(\boldsymbol{\theta}, \lambda) = \frac{\lambda^2 D}{2 + \lambda(D-2)} F(\boldsymbol{\theta}), \tag{55}$$

$$A(\boldsymbol{\theta}, \lambda) = \frac{\lambda^3 D^2}{[2 + \lambda(D-1)]^2} A(\boldsymbol{\theta}). \tag{56}$$

Replaced into (32) the above expressions lead to

$$\mathcal{I}^\star(\boldsymbol{\theta}, \lambda) = \frac{|\lambda| D}{2 + \lambda(D-2)} \mathcal{I}^\star(\boldsymbol{\theta}), \tag{57}$$

with $\mathcal{I}^\star(\boldsymbol{\theta})$ being the upper bound on the noiseless incompatibility figure of merit defined in (32). Notice that this expression is not symmetric around $\lambda = 0$, i.e. $\mathcal{I}^\star(\boldsymbol{\theta}, \lambda) \neq \mathcal{I}^\star(\boldsymbol{\theta}, -\lambda)$ for $\lambda \geqslant 0$. We define

Definition 4.1. (Asymmetry factor for $\mathcal{I}^\star(\boldsymbol{\theta}, \lambda)$)

$$\kappa^\star(\lambda) := \frac{|\mathcal{I}^\star(\boldsymbol{\theta}, \lambda) - \mathcal{I}^\star(\boldsymbol{\theta}, -\lambda)|}{\mathcal{I}^\star(\boldsymbol{\theta}, \lambda)} = \frac{2|\lambda|(D-2)}{2 - \lambda(D-2)}. \tag{58}$$

The presence of an asymmetry in the properties of the D-dimensional depolarizing channel around $\lambda = 0$ was already pointed out in the context of communication in [68]. For a qubit model $D = 2 \Rightarrow \kappa^\star(\lambda) = 0$. We show through a numerical example that this asymmetry exists not only for the upper bound $\mathcal{I}^\star(\boldsymbol{\theta}, \lambda)$ but also for the actual figure of merit $\mathcal{I}(\boldsymbol{\theta}, \lambda)$. Consider the encoding of two near-zero phases ($d = 2$) on a qutrit ($D = 3$) via the unitary operator (51) where the generators are chosen to be

$$H_1 = \begin{pmatrix} 0 & -\mathrm{i} & 0 \\ +\mathrm{i} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad H_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{59}$$

and the probe state is

$$|\psi\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}.$$

As in section 4.2 the figure of merit has been computed with the SDP presented in appendix C, for 500 equally spaced values of λ in the allowed region. In figure 3 the plot of $\mathcal{I}(\boldsymbol{\theta} = 0, \lambda)$ is reported for $\lambda \in (-1/8, 1)$, with a zoom on $\lambda \in (-1/8, 1/8)$. The dashed curve for $\lambda < 0$ reported in the insert is the reflection of the curve for $\lambda > 0$. It has been plotted in order to highlight the presence of the asymmetry. For this model, at the point $\boldsymbol{\theta} = 0$, the upper bound $\mathcal{I}^\star(\boldsymbol{\theta} = 0, \lambda)$ and the figure of merit $\mathcal{I}(\boldsymbol{\theta} = 0, \lambda)$ coincide. While in this qutrit example there is no gap between $\mathcal{I}(\boldsymbol{\theta}, \lambda)$ and $\mathcal{I}^\star(\boldsymbol{\theta}, \lambda)$, in general for a D-dimensional model this could be the case. From (57) we see that

$$\lim_{\lambda \to 0} \mathcal{I}^\star(\boldsymbol{\theta}, \lambda) = 0 \Rightarrow \lim_{\lambda \to 0} \mathcal{I}(\boldsymbol{\theta}, \lambda) = 0, \tag{60}$$

which means $\mathcal{I}^\star(\boldsymbol{\theta}, \lambda) - \mathcal{I}(\boldsymbol{\theta}, \lambda) \to 0$ for $\lambda \to 0$.

5. Design of compatible models for quantum metrology

The following section is somewhat disconnected from the previous discussions on the incompatibility measure $\mathcal{I}(\boldsymbol{\theta}, \lambda)$. Here we want to analyze some strategies that have been proposed in the past and some

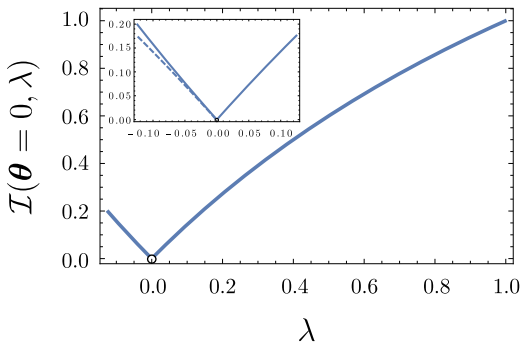

Figure 3. This curve is the incompatibility figure of merit $\mathcal{I}$ defined in (27), for the asymptotic covariant measurements, numerically computed for the qutrit example of section 4.4 for $\lambda \in (-1/8, 1)$. The empty circle at $\lambda = 0$ indicates that at this point the information quantities are not defined. The dashed curve in the small insert is the mirrored figure of merit for $\lambda > 0$.

generalizations that allow to produce a fully compatible statistical model in quantum metrology. We will only need to asses the condition $A(\boldsymbol{\theta}) = 0$ to claim compatibility, according to the bound (32), and therefore we will not need the linear program for $\mathcal{I}(\boldsymbol{\theta})$. We first review what is already known for 2 qubits, each codified via (46), and then generalize it for the D-dimensional encoding (51) when a couple of D-dimensional systems are available.

5.1. Known results for a two-qubits probe

In this section we analyze the compatibility of three different two qubit encoding scenarios in the absence of noise and for some special instances of the input states. We do not claim paternity of these results, they are only reviewed here in order to be generalized in the next section. Consider first the ancilla-aided model, in which only one of the two qubits is subject to the unitary encoding (46). This means that the total evolution of the two qubits is $\mathbb{1} \otimes U_{\boldsymbol{\theta}}$. As input state for the probes we take a Bell state, which is known in the literature to be optimal for the estimation of $SU(2)$ operations [70], for example

$$|\psi\rangle := \frac{|00\rangle + |11\rangle}{\sqrt{2}}, \tag{61}$$

where $|0\rangle$ and $|1\rangle$ are the eigenvectors of σ_z. From a direct computation we see that for this state $A(\boldsymbol{\theta}) = 0 \forall \boldsymbol{\theta}$, which from (32) gives

$$\mathcal{I}(\boldsymbol{\theta}) = 0, \quad \forall\, \boldsymbol{\theta}, \tag{62}$$

leading to compatibility. This result was first reported in [70]. Interestingly enough compatibility can also be obtained when operating on the maximally entangled state (61) with the encoding $U_{\boldsymbol{\theta}} \otimes U_{\boldsymbol{\theta}}$. Indeed by explicit computation we get again $A(\boldsymbol{\theta}) = 0 \forall \boldsymbol{\theta}$ that leads once more to (62). Such result can be found in [51]. We will see in section 5.2 that these effects are just a special instance of a more general trend since a maximally entangled state of a finite-dimensional probe always gives full compatibility, both for one and two uses of the encoding unitary channel.

We now give a last example, which we here name 'anti-parallel spin strategy' for future reference. Take the input state going through the encoding $U_{\boldsymbol{\theta}} \otimes U_{\boldsymbol{\theta}}$ to be $|+\hat{\boldsymbol{n}}\rangle \otimes |-\hat{\boldsymbol{n}}\rangle$, where $|+\hat{\boldsymbol{n}}\rangle$ and $|-\hat{\boldsymbol{n}}\rangle$ have opposite Bloch vectors $+\hat{\boldsymbol{n}}$ and $-\hat{\boldsymbol{n}}$. This state has the same QFI of the state of two parallel spins $|+\hat{\boldsymbol{n}}\rangle \otimes |+\hat{\boldsymbol{n}}\rangle$, but has $A(\boldsymbol{\theta}) = 0 \forall \boldsymbol{\theta}$ (in contrast to $|+\hat{\boldsymbol{n}}\rangle \otimes |+\hat{\boldsymbol{n}}\rangle$), which means that it is fully compatible and a superior probe for the sensing task. This result can be obtained from direct computation or thanks to the observation of section 5.3, where we generalize this ideas to finite-dimensional probes. The superiority of the anti-parallel spin state was already observed by Gisin and Popescu in [71] and in the context of parameter estimation in [72]. In all these three examples, being the encoded state pure, a measure entangled across two qubits only is sufficient to get compatibility [73].

5.2. Compatibility of the maximally entangled states

In this subsection we will show that the results of section 5.1 are only a particular case of a general observation, by proving that the use of an ancilla, maximally entangled with the probe, can completely remove the incompatibility, leading to the identity

$$\mathcal{I}(\boldsymbol{\theta}) = 0, \quad \forall\, \boldsymbol{\theta}. \tag{63}$$

Consider hence as input the following pure state

$$|\psi\rangle := \frac{1}{\sqrt{D}}\sum_{i=1}^{d}|i\rangle\otimes|i\rangle \in \mathcal{H}\otimes\mathcal{H}, \tag{64}$$

on which the evolution $\mathbb{1}\otimes U_{\boldsymbol{\theta}}$ acts to produce the output state

$$|\psi_{\boldsymbol{\theta}}\rangle := \mathbb{1}\otimes U_{\boldsymbol{\theta}}|\psi\rangle, \tag{65}$$

(i.e. the Choi–Jamiołkowski state of the channel $U_{\boldsymbol{\theta}}$ [74, 75]). Following section 4.3 and appendix E, the associated SLDs (9) of $\rho_{\boldsymbol{\theta}} := |\psi_{\boldsymbol{\theta}}\rangle\langle\psi_{\boldsymbol{\theta}}|$, can be expressed as

$$L_k(\boldsymbol{\theta}) = \frac{2\mathrm{i}}{D}\sum_{ij}|i\rangle\langle j|\otimes\left(H_k^{\mathrm{eff}}(\boldsymbol{\theta})|i\rangle\langle j| - |i\rangle\langle j|H_k^{\mathrm{eff}}(\boldsymbol{\theta})\right), \tag{66}$$

which lead to the following expressions for the $F(\boldsymbol{\theta})$ and $A(\boldsymbol{\theta})$ matrices:

$$F_{lm}(\boldsymbol{\theta}) = 2\ \mathrm{Tr}\left(\{H_l^{\mathrm{eff}}(\boldsymbol{\theta}), H_m^{\mathrm{eff}}(\boldsymbol{\theta})\}\right)/D, \tag{67}$$

$$A_{lm}(\boldsymbol{\theta}) = -2\mathrm{i}\ \mathrm{Tr}\left(\left[H_l^{\mathrm{eff}}(\boldsymbol{\theta}), H_m^{\mathrm{eff}}(\boldsymbol{\theta})\right]\right)/D = 0, \tag{68}$$

where in the last identity we used the fact that $\left[H_l^{\mathrm{eff}}(\boldsymbol{\theta}), H_m^{\mathrm{eff}}(\boldsymbol{\theta})\right]$ is a traceless operator. Accordingly the upper bound (32) imposes (63), hence the thesis: the addition of a sufficiently large ancilla permits to remove entirely the quantum incompatibility for the LAC measurements.

A similar result holds true also when we let evolve the maximally entangled state (64) through $U_{\boldsymbol{\theta}}\otimes U_{\boldsymbol{\theta}}$. In this case (65) gets replaced by

$$|\psi_{\boldsymbol{\theta}}\rangle := U_{\boldsymbol{\theta}}\otimes U_{\boldsymbol{\theta}}|\psi\rangle, \tag{69}$$

which leads to

$$F_{lm}(\boldsymbol{\theta}) = 8\ \mathrm{Tr}\left(\{H_l^{\mathrm{eff}}(\boldsymbol{\theta}), H_m^{\mathrm{eff}}(\boldsymbol{\theta})\}\right)/D, \tag{70}$$

$$A_{lm}(\boldsymbol{\theta}) = -4\mathrm{i}\ \mathrm{Tr}\left(\left[H_l^{\mathrm{eff}}(\boldsymbol{\theta}), H_m^{\mathrm{eff}}(\boldsymbol{\theta})\right]\right)/D = 0, \tag{71}$$

which gives again the full compatibility condition (63).

5.3. Generalized anti-parallel spin strategy

Now we generalize the 'anti-parallel spin' strategy of section 5.1. Suppose that we only have two parameters to estimate ($d = 2$) and we take for probe the separable input state $|\psi_1\rangle\otimes|\psi_2\rangle$ that evolves through the mapping induced by $U_{\boldsymbol{\theta}}\otimes U_{\boldsymbol{\theta}}$. The sufficient condition for compatibility $A(\boldsymbol{\theta}) = 0$ can be expanded as

$$\langle\psi_1|[H_1^{\mathrm{eff}}(\boldsymbol{\theta}), H_2^{\mathrm{eff}}(\boldsymbol{\theta})]|\psi_1\rangle + \langle\psi_2|[H_1^{\mathrm{eff}}(\boldsymbol{\theta}), H_2^{\mathrm{eff}}(\boldsymbol{\theta})]|\psi_2\rangle = 0. \tag{72}$$

The operator $[H_1^{\mathrm{eff}}(\boldsymbol{\theta}), H_2^{\mathrm{eff}}(\boldsymbol{\theta})]$ is skew-Hermitian and therefore is diagonalizable and has purely imaginary eigenvalues $\pm\mathrm{i}a_j$, where $a_j > 0$, for $j = 1, 2, \ldots, \lfloor D/2\rfloor$, each associated with an eigenvector $|\pm\mathrm{i}a_j\rangle$. If the dimension D is odd, then we have an extra unique zero eigenvalue. Let us denote with V the unitary operator that performs such diagonalization, i.e. $V^\dagger\left[H_1^{\mathrm{eff}}(\boldsymbol{\theta}), H_2^{\mathrm{eff}}(\boldsymbol{\theta})\right]V = \mathrm{diag}\,(\pm\mathrm{i}a_1, \pm\mathrm{i}a_2, \ldots)$. Let us define $\mathcal{S}\subseteq\{1, 2, \ldots, \lfloor D/2\rfloor\}$, then a couple of states that guarantees compatibility is

$$|\psi_1\rangle := \frac{1}{|\mathcal{S}|}\sum_{j\in\mathcal{S}}\mathrm{e}^{\mathrm{i}\varphi_j^1}V|\mathrm{i}(-1)^{s_j}a_j\rangle, \tag{73}$$

$$|\psi_2\rangle := \frac{1}{|\mathcal{S}|}\sum_{j\in\mathcal{S}}\mathrm{e}^{\mathrm{i}\varphi_j^2}V|-\mathrm{i}(-1)^{s_j}a_j\rangle, \tag{74}$$

where φ_j^1 and φ_j^2 are arbitrary phases, $s_j\in\{0, 1\}$, and $|\mathcal{S}|$ is the cardinality of $\mathcal{S}$. Notice that we are also free to add in the definition of whichever $|\psi_1\rangle$ or $|\psi_2\rangle$ the state $V|0\rangle$, with $|0\rangle$ being the eigenvector with null eigenvalue (in case there is one). With the above choice of probes the compatibility condition (72) is realized. Notice also that the QFI matrix of $|\psi_1\rangle\otimes|\psi_2\rangle$ is the sum of the QFIs of $|\psi_1\rangle$ and $|\psi_2\rangle$. This two states taken individually manifest incompatibility, but when measured jointly they gain full compatibility. This construction is the analogue of the 'anti-parallel spin strategy' of section 5.1. We observe that a state $|\psi\rangle$, being an equal superposition of $|\psi_1\rangle$ and $|\psi_2\rangle$ is also fully compatible. The condition (72) justifies also

the compatibility of $|+\widehat{\boldsymbol{n}}\rangle \otimes |-\widehat{\boldsymbol{n}}\rangle \ \forall\ \boldsymbol{\theta}$ in section 5.1. These two states are an orthogonal basis for the qubit Hilbert space, therefore

$$\langle +\widehat{\boldsymbol{n}}|[H_1^{\rm eff}(\boldsymbol{\theta}), H_2^{\rm eff}(\boldsymbol{\theta})]|+\widehat{\boldsymbol{n}}\rangle + \langle -\widehat{\boldsymbol{n}}|[H_1^{\rm eff}(\boldsymbol{\theta}), H_2^{\rm eff}(\boldsymbol{\theta})]|-\widehat{\boldsymbol{n}}\rangle = {\rm Tr}\left([H_1^{\rm eff}(\boldsymbol{\theta}), H_2^{\rm eff}(\boldsymbol{\theta})]\right) = 0,$$

hence condition (72) is satisfied and (63) holds. Incidentally we observe also that for a d-parameter estimation, the state $|\psi_1\rangle \otimes |\psi_2\rangle \otimes \ldots \otimes |\psi_D\rangle$ is compatible when $\{|\psi_i\rangle\}_{i=1}^D$ is base for the Hilbert space $\mathcal{H}$, because then we would have

$$A_{ij}(\boldsymbol{\theta}) \propto {\rm Tr}\left(\left[H_i^{\rm eff}(\boldsymbol{\theta}), H_j^{\rm eff}(\boldsymbol{\theta})\right]\right) = 0. \tag{75}$$

6. Conclusions

One of the defining properties of quantum mechanics is the intrinsic incompatibility between the possible experiments that could be carried out on a quantum system. This causes the appearance of information limits on the precision to which different characteristics of a certain quantum system can be known. Formally, this comes always from the non-commutativity of quantum operations. The main role of this paper is to give a theoretical foundation to the measure of incompatibility in the quantum estimation task. For this purpose we define in section 3 a figure of merit having a well defined operational and geometrical meaning. The figure of merit $\underline{r}(\boldsymbol{\theta})$ in (18) is built to be easily liked to the asymptotic results of estimation theory [42]. This allows us to easily compute it, as explained in appendix C, and to give the upper bound in section 3.5. The definition of incompatibility depends on the operations that we are able to perform, which is our level of control over the system. If we are only able to perform separable measurements on the probes then the relevant incompatibility measure is $\underline{r}^{\rm s}(\boldsymbol{\theta})$ defined in (34). In section 4 the estimation is studied in the scenario where a depolarizing noise acts, this is a form of disturbance which is often used to model the decoherence dynamics in metrology [76, 77]. We observed some interesting phenomena like the asymmetry of the incompatibility for inversion in the space of states in section 4.3. In section 5 we discuss some strategies able to produce compatible models for quantum metrology with generic D-dimension systems, which are for example the use of maximally entangled states. As a further development it would be interesting to determine which state gives the maximum and minimum incompatibility for a certain encoding at a fixed point $\boldsymbol{\theta}$. This optimization is hard because the figure of merit is a complicated non linear function of the state. In a sense the probe that maximizes incompatibility is the one which captures at most the quantum properties of the encoding process. Finally we would like to look for a link between the incompatibility and the Heisenberg scaling. In this context the only relevant measures are the one that assume no constraints on the separability of the input, because a single giant entangled probe would be used.

Acknowledgments

We thank Angelo Carollo, Francesco Albarelli, Marco Genoni, and Yuxiang Yang for discussions. We acknowledges support by MIUR via PRIN 2017 (Progetto di Ricerca di Interesse Nazionale): Project QUSHIP (2017SRNBRK).

Data availability statement

No new data were created or analysed in this study.

Appendix A. Figure of merit for LAC measurements

In this appendix we will define a version of the incompatibility figure of merit for local asymptotic covariant (LAC) measurements [42, 48]. Consider the sequences of POVMs $(M_k)_{k\in\mathbb{N}_0} \in \mathcal{M}_{\mathcal{C}}^{(\rm LAC)}$ that satisfy local asymptotic covariance (LAC) at the point $\boldsymbol{\theta} \in \Theta$, as defined in [42, 48]. The kth measurement M_k of a sequence in $\mathcal{M}_{\mathcal{C}}^{(\rm LAC)}$ acts on k probes and has $\Sigma^{(k)}(\boldsymbol{\theta})$ as associated MSE matrix. Because of the definition of LAC, $(M_k)_{k\in\mathbb{N}_0} \in \mathcal{M}_{\mathcal{C}}^{(\rm LAC)}$ admits a limiting MSE matrix, i.e. $\lim_{k\to\infty} k\Sigma^{(k)}(\boldsymbol{\theta}) := \Sigma(\boldsymbol{\theta})$ (we also ask $\| \Sigma(\boldsymbol{\theta}) \| < \infty$). The new figure of merit is hence defined as

$$r(\boldsymbol{\theta}) := \inf_{\mathcal{M}_{\mathcal{C}}^{(\rm LAC)}} \sup_{G\geqslant 0} \frac{{\rm Tr}\left[G\cdot\Sigma(\boldsymbol{\theta})\right]}{{\rm Tr}\left[G\cdot F^{-1}(\boldsymbol{\theta})\right]}. \tag{A.1}$$

In [48] the authors have proven the validity of the Holevo bound for the LAC measurements and its achievability in the same class (for a full rank $\rho_{\boldsymbol{\theta}}$ with non degenerate spectrum), that is

$$\inf_{\mathcal{M}_{\mathcal{C}}^{(\mathrm{LAC})}} \mathrm{Tr}\left[G\cdot\Sigma(\boldsymbol{\theta})\right] = C_{\mathrm{H}}\left(G,\boldsymbol{\theta}\right). \tag{A.2}$$

We now prove that th inf and sup of (A.1) can be commuted. It is easy to prove that the set $\mathcal{M}_{\mathcal{C}}^{(\mathrm{LAC})}$ is convex. The set of POVMs acting on $\mathfrak{S}\left(\mathcal{H}^{\otimes N}\right)$ and having $\boldsymbol{\theta}$ as outcome set can be thought as a convex subset of a certain dual vector space $\mathcal{V}'$ [78]. The set containing the infinite sequences of $\mathcal{V}'$ is a vector space, and the set of sequences thereof that are sequences of POVMs is convex. Furthermore the defining properties of LAC [48] is stable under convex combination of two measurement sequences. In the Minimax theorem of [79] the both spaces are required to be be locally convex. Banach spaces like $\mathcal{G}$ and $\mathcal{V}'\supset\mathcal{M}_N$ are locally convex, and the countable infinite product space of multiple $\mathcal{V}'$ (which is the space of sequences) is also a locally convex space. Furthermore given $\Sigma^1(\boldsymbol{\theta})$ and $\Sigma^2(\boldsymbol{\theta})$ the limiting MSE matrices of two sequences $\left(M_k^1\right)_{k\in\mathbb{N}_0}$ and $\left(M_k^2\right)_{k\in\mathbb{N}_0}$, the asymptotic MSE of the convex combination $\left(\alpha M_k^1+(1-\alpha)\,M_k^2\right)_{k\in\mathbb{N}_0}$ is $\Sigma(\boldsymbol{\theta})=\alpha\Sigma^1(\boldsymbol{\theta})+(1-\alpha)\,\Sigma^2(\boldsymbol{\theta})$. Consequently, just like in appendix B, the Minimax theorem of Kneser [79, 80] can be applied to swap the order of inf over $\mathcal{M}_{\mathcal{C}}^{(\mathrm{LAC})}$ and $\sup_{G\geqslant 0}$. It is understood that the argument of (A.1) can be cast into the same form of (B.6), from which the linearity in $(M_k)_{k\in\mathbb{N}_0}$ and G, and the continuity in G easily follow. We arrive therefore at

$$r(\boldsymbol{\theta}) = \sup_{G\geqslant 0}\inf_{\mathcal{M}_{\mathcal{C}}^{(\mathrm{LAC})}} \frac{\mathrm{Tr}\left[G\cdot\Sigma(\boldsymbol{\theta})\right]}{\mathrm{Tr}\left[G\cdot F^{-1}(\boldsymbol{\theta})\right]} = \sup_{G\geqslant 0}\frac{C_{\mathrm{H}}(G,\boldsymbol{\theta})}{C_{\mathrm{S}}(G,\boldsymbol{\theta})} = \underline{r}(\boldsymbol{\theta}). \tag{A.3}$$

Appendix B. Exchanging sup and inf in the figure of merit definition

We will arrive at (24) through a series of lemmas.

Lemma 1. *The function $r_N\left(G,M_N,\boldsymbol{\theta}\right)$ is continuous in $G\in\mathcal{G}$ at fixed $M_N\in\mathcal{M}_N^{(LU)}$.*

The denominator of $r_N\left(G,M_N,\boldsymbol{\theta}\right)$ can be bounden as $\mathrm{Tr}[G\cdot F(\boldsymbol{\theta})^{-1}]\geqslant\lambda_{\min}\left(F(\boldsymbol{\theta})^{-1}\right)=\lambda_{\max}(F(\boldsymbol{\theta}))^{-1}=1/\|F(\boldsymbol{\theta})\|$, this means

$$r_N\left(G,M_N,\boldsymbol{\theta}\right)\leqslant N\|F(\boldsymbol{\theta})\|\ \mathrm{Tr}[G\cdot\Sigma^{(N)}(\boldsymbol{\theta})], \tag{B.1}$$

it follows

$$|r_N\left(G_1,M_N,\boldsymbol{\theta}\right)-r_N\left(G_2,M_N,\boldsymbol{\theta}\right)|\leqslant N\|F(\boldsymbol{\theta})\|\|\Sigma^{(N)}(\boldsymbol{\theta})\|\|G_1-G_2\|.$$

Its important to assume that the set $M_N\in\mathcal{M}_N^{(\mathrm{LU})}$ contains only measurements with bounded MSE matrices, i.e. $\|\ \Sigma^{(N)}(\boldsymbol{\theta})\ \|\leqslant C$. Therefore we have

$$|r_N\left(G_1,M_N,\boldsymbol{\theta}\right)-r_N\left(G_2,M_N,\boldsymbol{\theta}\right)|\leqslant NC\|F(\boldsymbol{\theta})\|\|G_1-G_2\|, \tag{B.2}$$

which means the function $r_N\left(G,M_N,\boldsymbol{\theta}\right)$ is Lipschitz continuous with constant $NC\ \|\ F(\boldsymbol{\theta})\ \|$ and therefore continuous. It will be useful in the latter to notice that $r_N(F(\boldsymbol{\theta})^{\frac{1}{2}}GF(\boldsymbol{\theta})^{\frac{1}{2}},M_N,\boldsymbol{\theta})=N\ \mathrm{Tr}[G\cdot F(\boldsymbol{\theta})^{\frac{1}{2}}\Sigma^{(N)}(\boldsymbol{\theta})F(\boldsymbol{\theta})^{\frac{1}{2}}]$ is also continuous in $G\in\mathcal{G}$, because is the composition of the continuous functions $r_N(G,M_N,\boldsymbol{\theta})$ and $G\to F(\boldsymbol{\theta})^{\frac{1}{2}}GF(\boldsymbol{\theta})^{\frac{1}{2}}$.

In this paper only the upper semicontinuity of $r_N(G,M_N,\boldsymbol{\theta})$ is actually used.

Lemma 2. *The figure of merit $r_N(\boldsymbol{\theta})$ defined in (17) can be equivalently expressed as*

$$r_N(\boldsymbol{\theta}) = \sup_{G\geqslant 0}\inf_{M_N\in\mathcal{M}_N^{(LU)}} r_N\left(G,M_N,\boldsymbol{\theta}\right). \tag{B.3}$$

This lemma is based on the application of the Minimax theorem of Kneser [79, 80]. First of all we need to cast $\mathrm{r_N}(\Theta)$ in a suitable form. We start from the observation that the set of positive weight matrices G is invariant under congruence for the positive matrix $F(\boldsymbol{\theta})^{\frac{1}{2}}$, i.e.

$$\{G\geqslant 0\}=F(\boldsymbol{\theta})^{\frac{1}{2}}\{G\geqslant 0\}F(\boldsymbol{\theta})^{\frac{1}{2}}=\{F(\boldsymbol{\theta})^{\frac{1}{2}}GF(\boldsymbol{\theta})^{\frac{1}{2}}\geqslant 0\}. \tag{B.4}$$

This means that the sup can be taken on the matrices $F(\boldsymbol{\theta})^{\frac{1}{2}}GF(\boldsymbol{\theta})^{\frac{1}{2}}\geqslant 0$ without changing $r_N(\boldsymbol{\theta})$, so we have

$$r_N(\boldsymbol{\theta}) = \inf_{M_N\in\mathcal{M}_N^{(\mathrm{LU})}}\ \sup_{F(\boldsymbol{\theta})^{\frac{1}{2}}GF(\boldsymbol{\theta})^{\frac{1}{2}}\geqslant 0}\frac{N\ \mathrm{Tr}\left[F(\boldsymbol{\theta})^{\frac{1}{2}}GF(\boldsymbol{\theta})^{\frac{1}{2}}\cdot\Sigma^{(N)}(\boldsymbol{\theta})\right]}{\mathrm{Tr}\left[F(\boldsymbol{\theta})^{\frac{1}{2}}GF(\boldsymbol{\theta})^{\frac{1}{2}}\cdot F(\boldsymbol{\theta})^{-1}\right]}, \tag{B.5}$$

$$= \inf_{M_N \in \mathcal{M}_N^{(LU)}} \sup_{G \geqslant 0} \frac{N \ \mathrm{Tr}\left[G \cdot F(\boldsymbol{\theta})^{\frac{1}{2}} \Sigma^{(N)}(\boldsymbol{\theta}) F(\boldsymbol{\theta})^{\frac{1}{2}}\right]}{\mathrm{Tr}[G]}$$

$$= \inf_{M_N \in \mathcal{M}_N^{(LU)}} \sup_{G \in \mathcal{G}} N \ \mathrm{Tr}\left[G \cdot F(\boldsymbol{\theta})^{\frac{1}{2}} \Sigma^{(N)}(\boldsymbol{\theta}) F(\boldsymbol{\theta})^{\frac{1}{2}}\right]. \tag{B.6}$$

In the second line, in the domain of the supremum, we have again used that every $F(\boldsymbol{\theta})^{\frac{1}{2}} G F(\boldsymbol{\theta})^{\frac{1}{2}} \geqslant 0$ corresponds to a $G \geqslant 0$. In the last equation the sup is restricted to the set $\mathcal{G} = \{G \geqslant 0, \mathrm{Tr}\ G = 1\}$, which is compact and convex. The set of locally unbiased measurements $M_N \in \mathcal{M}_N^{(\mathrm{LU})}$ can be thought as a (non-empty) subset of a dual vector space [78, 81], which is a convex set because the locally unbiased measurements are stable under convex combination. The function $N\ \mathrm{Tr}[G \cdot F(\boldsymbol{\theta})^{\frac{1}{2}} \Sigma^{(N)}(\boldsymbol{\theta}) F(\boldsymbol{\theta})^{\frac{1}{2}}]$ is linear in both its arguments. The linearity in G is self evident, so we only show linearity in the measurement. Suppose that we are given two POVMs denoted by M^1 and M^2, characterized respectively by the effects $E^1_{\hat{\boldsymbol{\theta}}}$ and $E^2_{\hat{\boldsymbol{\theta}}}$ associated to the outcome $\hat{\boldsymbol{\theta}}$. We have dropped for simplicity the subscript N in M^1 and M^2 and we will also drop the superscript (N) in the MSE matrix $\Sigma(\boldsymbol{\theta})$. Consider the POVM being the linear combination $M := \lambda M^1 + (1-\lambda) M^2$. By definition its effects are

$$E_{\hat{\boldsymbol{\theta}}} := \lambda E^1_{\hat{\boldsymbol{\theta}}} + (1-\lambda) E^2_{\hat{\boldsymbol{\theta}}}. \tag{B.7}$$

The $\Sigma(\boldsymbol{\theta})$ matrix associated to M is computed as expectation value on the probability distribution

$$\begin{aligned} p(\hat{\boldsymbol{\theta}}) &:= \mathrm{Tr}(\rho_{\boldsymbol{\theta}} \mathrm{E}_{\hat{\boldsymbol{\theta}}}) \\ &= \alpha\ \mathrm{Tr}(\rho_{\boldsymbol{\theta}} \mathrm{E}^1_{\hat{\boldsymbol{\theta}}}) + (1-\alpha)\, \mathrm{Tr}(\rho_{\boldsymbol{\theta}} \mathrm{E}^2_{\hat{\boldsymbol{\theta}}}) \\ &= \alpha p^1(\hat{\boldsymbol{\theta}}) + (1-\alpha) p^2(\hat{\boldsymbol{\theta}}). \end{aligned} \tag{B.8}$$

The linearity of $p(\hat{\boldsymbol{\theta}})$ translates to the linearity of $\Sigma(\boldsymbol{\theta})$, i.e.

$$\Sigma_{ij}(\boldsymbol{\theta}) = \alpha \Sigma^1_{ij}(\boldsymbol{\theta}) + (1-\alpha) \Sigma^2_{ij}(\boldsymbol{\theta}). \tag{B.9}$$

This means that the whole argument of the inf sup in (B.6) is linear in the POVM, and it is additionally upper semicontinuous in G at fixed M_N (In lemma 1 we proved continuity, which implies upper semicontinuity). We have now proved all the required hypothesis for the Minimax theorem of Kneser [79, 80], which allows us to write

$$r_N(\boldsymbol{\theta}) = \sup_{G \geqslant 0} \inf_{M_N \in \mathcal{M}_N^{(\mathrm{LU})}} \frac{N\ \mathrm{Tr}\left[G \cdot \Sigma^{(N)}(\boldsymbol{\theta})\right]}{\mathrm{Tr}\left[G \cdot F(\boldsymbol{\theta})^{-1}\right]}. \tag{B.10}$$

It is worth stressing that without such argument the quantity

$$r_N^{(\downarrow)}(\boldsymbol{\theta}) := \sup_{G \geqslant 0} \inf_{M_N \in \mathcal{M}_N^{(\mathrm{LU})}} \frac{N\ \mathrm{Tr}\left[G \cdot \Sigma^{(N)}(\boldsymbol{\theta})\right]}{\mathrm{Tr}\left[G \cdot F(\boldsymbol{\theta})^{-1}\right]}, \tag{B.11}$$

is by construction always smaller than or equal to

$$r_N^{(\uparrow)}(\boldsymbol{\theta}) := \inf_{M_N \in \mathcal{M}_N^{(\mathrm{LU})}} \sup_{G \geqslant 0} \frac{N\ \mathrm{Tr}\left[G \cdot \Sigma^{(N)}(\boldsymbol{\theta})\right]}{\mathrm{Tr}\left[G \cdot F(\boldsymbol{\theta})^{-1}\right]}, \tag{B.12}$$

i.e. $r_N^{(\downarrow)}(\boldsymbol{\theta}) \leqslant r_N^{(\uparrow)}(\boldsymbol{\theta})$.

Lemma 3.

$$r_{N_1}(\boldsymbol{\theta}) \geqslant r_{N_1 N_2}(\boldsymbol{\theta}). \tag{B.13}$$

Let N := N1N2 with N1, N2 integers. Given N copies of the probe we can organize them into N2 distinct subgroups, each of them containing N1 probes. We now perform exactly the same measurement M_{N_1} on each group and use the N_2 outcomes to estimate θ by taking their arithmetic mean. Calling this measurement $M^\star_N$ it follows that its MSE matrix $\Sigma^\star_N(\boldsymbol{\theta})$ corresponds to $\Sigma_{N_1}(\boldsymbol{\theta})/N_2$, being $\Sigma_{N_1}(\boldsymbol{\theta})$ the MSE matrix of M_{N_1}. This holds true because the estimators are unbiased at θ. Therefore we have

$$r_N(G, M^\star_N, \boldsymbol{\theta}) = \frac{N\ \mathrm{Tr}\left[G \cdot \Sigma^\star_N(\boldsymbol{\theta})\right]}{\mathrm{Tr}\left[G \cdot F(\boldsymbol{\theta})^{-1}\right]}, \tag{B.14}$$

$$= \frac{N \ \mathrm{Tr}\left[G \cdot \Sigma_{N_1}(\boldsymbol{\theta})\right]}{N_2 \ \mathrm{Tr}\left[G \cdot F(\boldsymbol{\theta})^{-1}\right]}, \tag{B.15}$$

$$= \frac{N_1 \ \mathrm{Tr}\left[G \cdot \Sigma_{N_1}(\boldsymbol{\theta})\right]}{\mathrm{Tr}\left[G \cdot F(\boldsymbol{\theta})^{-1}\right]}, \tag{B.16}$$

$$= r_{N_1}(G, M_{N_1}, \boldsymbol{\theta}). \tag{B.17}$$

We now need to introduce a new quantity:

$$r_N(G, \boldsymbol{\theta}) := \inf_{M_N \in \mathcal{M}_N^{(\mathrm{LU})}} r_N(G, M_N, \boldsymbol{\theta}), \tag{B.18}$$

where the supremum on $G \geqslant 0$ has been removed. We can always take $M_{N_1}^{\varepsilon}$ such that

$$r_{N_1}(G, M_{N_1}^{\varepsilon}, \boldsymbol{\theta}) \leqslant \inf_{M_{N_1} \in \mathcal{M}_{N_1}^{(\mathrm{LU})}} r_{N_1}\left(G, M_1, \boldsymbol{\theta}\right) + \varepsilon, \tag{B.19}$$

with $\varepsilon > 0$. Then we have

$$r_N\left(G, \boldsymbol{\theta}\right) \leqslant r_N(G, M_N^{\star}, \boldsymbol{\theta}) = r_{N_1}(G, M_{N_1}^{\varepsilon}, \boldsymbol{\theta}) \leqslant \inf_{M_{N_1} \in \mathcal{M}_{N_1}^{(\mathrm{LU})}} r_{N_1}\left(G, M_{N_1}, \boldsymbol{\theta}\right) + \varepsilon = r_{N_1}\left(G, \boldsymbol{\theta}\right) + \varepsilon,$$

where $r_N(G, \boldsymbol{\theta})$ has been defined in (B.18). Because of the arbitrariness of ε we have $r_N\left(G, \boldsymbol{\theta}\right) \leqslant r_{N_1}\left(G, \boldsymbol{\theta}\right)$. Taking $\sup_{G \geqslant 0}$ on both the members of this inequality gives finally the thesis.

Lemma 4.

$$\underline{r}(\boldsymbol{\theta}) = \inf_{N \geqslant 1} r_N(\boldsymbol{\theta}) = \lim_{n \to \infty} \min_{N \leqslant n} r_N(\boldsymbol{\theta}). \tag{B.20}$$

Recall the definition of $\underline{r}(\boldsymbol{\theta})$ in (18), it can be expressed as $\underline{r}(\boldsymbol{\theta}) = \lim_{n \to \infty} \underline{r}_n(\boldsymbol{\theta})$, where

$$\underline{r}_n(\boldsymbol{\theta}) := \inf_{N \geqslant n} r_N(\boldsymbol{\theta}), \tag{B.21}$$

which is by construction non-decreasing in n, i.e.

$$\underline{r}_m(\boldsymbol{\theta}) \geqslant \underline{r}_n(\boldsymbol{\theta}) \quad \forall\ m \geqslant n. \tag{B.22}$$

Our goal is to show that due to lemma 3, the inequality in the above expression is always saturated, or equivalently that

$$\underline{r}_m(\boldsymbol{\theta}) = \underline{r}_1(\boldsymbol{\theta}) = \inf_{N \geqslant 1} r_N(\boldsymbol{\theta}) \quad \forall m \ \geqslant 1, \tag{B.23}$$

which will lead automatically to $\underline{r}(\boldsymbol{\theta}) = \underline{r}_1(\boldsymbol{\theta}) = \inf_{N \geqslant 1} r_N(\boldsymbol{\theta})$. We can prove (B.23) by contradiction: assume that there exists m such that $\underline{r}_m(\boldsymbol{\theta}) > \underline{r}_1(\boldsymbol{\theta})$. This implies that there must exist $k < m$ integer such that

$$\underline{r}_1(\boldsymbol{\theta}) = r_k(\boldsymbol{\theta}) < \underline{r}_m(\boldsymbol{\theta}). \tag{B.24}$$

This however cannot be true because thanks to lemma 3 we must have

$$r_k(\boldsymbol{\theta}) \geqslant r_{mk}(\boldsymbol{\theta}) \geqslant \inf_{N \geqslant m} r_N(\boldsymbol{\theta}) = \underline{r}_m(\boldsymbol{\theta}). \tag{B.25}$$

For the second equality in (B.20) it is easy to notice that the sequence defined by

$$\underline{r}_n^{(<)}(\boldsymbol{\theta}) := \min_{N \leqslant n} r_N(\boldsymbol{\theta}), \tag{B.26}$$

is explicitly decreasing, i.e.

$$\underline{r}_{n+1}^{(<)}(\boldsymbol{\theta}) \leqslant \underline{r}_n^{(<)}(\boldsymbol{\theta}), \tag{B.27}$$

therefore its limit exists and it is easy to show it coincides with $\inf_{N \geqslant 1} r_N(\boldsymbol{\theta})$ as we see in the following. Take $\varepsilon > 0$, then $\exists N_\varepsilon$ such that

$$r_{N_\varepsilon}(\boldsymbol{\theta}) \leqslant \inf_{N \geqslant 1} r_N(\boldsymbol{\theta}) + \varepsilon, \tag{B.28}$$

therefore $\forall \varepsilon$, $\exists n_\varepsilon := N_\varepsilon$ such that

$$\underline{r}_{n_\varepsilon}^{(<)}(\boldsymbol{\theta}) \leqslant r_{N_\varepsilon}(\boldsymbol{\theta}) \leqslant \inf_{N \geqslant 1} r_N(\boldsymbol{\theta}) + \varepsilon, \tag{B.29}$$

furthermore

$$\underline{r}_n^{(<)}(\boldsymbol{\theta}) \geqslant \inf_{N\geqslant 1} r_N(\boldsymbol{\theta}), \tag{B.30}$$

because in $\underline{r}_n^{(<)}(\boldsymbol{\theta})$ the domain of minimization is smaller. Together (B.29) and (B.30) mean

$$\lim_{n\to\infty} \underline{r}_n^{(<)}(\boldsymbol{\theta}) = \inf_{N\geqslant 1} r_N(\boldsymbol{\theta}) = \underline{r}_1(\boldsymbol{\theta}). \tag{B.31}$$

Before proceeding further let us make yet another observation: the above construction can be applied even in the absence of the optimization over G. Specifically, for all $G \geqslant 0$ and $N \in \mathbb{N}_0$, we define the the quantities

$$\underline{r}_n(G,\boldsymbol{\theta}) := \inf_{N\geqslant n} r_N(G,\boldsymbol{\theta}), \tag{B.32}$$

$$\underline{r}_n^{(<)}(G,\boldsymbol{\theta}) := \min_{N\leqslant n} r_N(G,\boldsymbol{\theta}), \tag{B.33}$$

with $r_N(G,\boldsymbol{\theta})$ defined in (B.18). By construction, for all given G, $\underline{r}_n(G,\boldsymbol{\theta})$ is explicitly non-decreasing, while $\underline{r}_n^{(<)}(G,\boldsymbol{\theta})$ is explicitly non-increasing, i.e.

$$\underline{r}_{n+1}(G,\boldsymbol{\theta}) \geqslant \underline{r}_n(G,\boldsymbol{\theta}), \tag{B.34}$$

$$\underline{r}_{n+1}^{(<)}(G,\boldsymbol{\theta}) \leqslant \underline{r}_n^{(<)}(G,\boldsymbol{\theta}). \tag{B.35}$$

Moreover following the same arguments presented in lemma 4 it turns out that $\underline{r}_n(G,\boldsymbol{\theta})$ is indeed constant, i.e.

$$\underline{r}_n(G,\boldsymbol{\theta}) = \underline{r}_1(G,\boldsymbol{\theta}) \quad \forall\, n, \tag{B.36}$$

and

$$\underline{r}_1(G,\boldsymbol{\theta}) = \inf_{N\geqslant 1} r_N(G,\boldsymbol{\theta}), \tag{B.37}$$

$$= \lim_{n\to\infty} \min_{N\leqslant n} r_N(G,\boldsymbol{\theta}), \tag{B.38}$$

$$= \lim_{n\to\infty} \underline{r}_n^{(<)}(G,\boldsymbol{\theta}). \tag{B.39}$$

Lemma 5. *The function $r_N(G,\boldsymbol{\theta})$ is upper semicontinuous in $G \in \mathcal{G}\ \forall\ N$.*

The function $r_N(G, M_N, \boldsymbol{\theta})$, defined in (16), is continuous in $G \in \mathcal{G}$ for fixed M_N because of lemma 1, and in particular upper semicontinuous. The function $r_N(G,\boldsymbol{\theta})$ defined in (B.18) is upper semicontinuous in $G \in \mathcal{G}$ because the infimum of a family of upper semicontinuous functions (here labeled by the measurement M_N) is upper semicontinuous.

Consider next the supremum over G of $\underline{r}_1(G,\boldsymbol{\theta})$, this can be evaluated as

$$\sup_{G\geqslant 0} \underline{r}_1(G,\boldsymbol{\theta}) = \sup_{G\geqslant 0} \inf_{N\geqslant 1} r_N(G,\boldsymbol{\theta}), \tag{B.40}$$

$$= \sup_{G\geqslant 0} \inf_{N\geqslant 1} \inf_{M_N\in\mathcal{M}_N^{(\mathrm{LU})}} r_N(G, M_N, \boldsymbol{\theta})$$

$$= \inf_{N\geqslant 1} \inf_{M_N\in\mathcal{M}_N^{(\mathrm{LU})}} \sup_{G\geqslant 0} r_N(G, M_N, \boldsymbol{\theta})$$

$$= \inf_{N\geqslant 1} r_N(\boldsymbol{\theta}) = \underline{r}_1(\boldsymbol{\theta}) = \underline{r}(\boldsymbol{\theta}). \tag{B.41}$$

Going from (B.40) to (B.41) requires the application of two different versions of the Minimax theorem. First of all we need to commute $\inf_{N\geqslant 1}$ and $\sup_{G\geqslant 0}$ ($\sup_{\mathcal{G}}$). This can be accomplished with the Ky Fan Minimax theorem [80, 82]. In order to use this result it must be proved that $r_N(G,\boldsymbol{\theta}) = \inf_{M_N\in\mathcal{M}_N^{(\mathrm{LU})}} r_N(G, M_N, \boldsymbol{\theta})$ is Ky Fan concave-convex on $\mathcal{G} \times \mathbb{N}_0$. This condition is equivalent to having Ky Fan concavity in G for every fixed N and Ky Fan convexity in N for every fixed G. Let us fix an arbitrary $N \in \mathbb{N}_0$. Given the combination $G_\alpha := \alpha G_1 + (1-\alpha)\, G_2$ with $\alpha \in [0,1]$, $\forall\ \varepsilon\ \exists\, M_N^\varepsilon$ such that

$$r_N(G_\alpha, \boldsymbol{\theta}) = \inf_{M_N\in\mathcal{M}_N^{(\mathrm{LU})}} r_N(G_\alpha, M_N, \boldsymbol{\theta}), \tag{B.42}$$

$$\geqslant r_N\left(G_\alpha, M_N^\varepsilon, \boldsymbol{\theta}\right) - \varepsilon. \tag{B.43}$$

By expanding G_α in $r_N\left(G_\alpha, M_N^\varepsilon, \boldsymbol{\theta}\right)$ we have

$$r_N\left(G_\alpha, M_N^\varepsilon, \boldsymbol{\theta}\right) = \alpha\, r_N\left(G_1, M_N^\varepsilon, \boldsymbol{\theta}\right) + (1-\alpha)\, r_N\left(G_2, M_N^\varepsilon, \boldsymbol{\theta}\right), \tag{B.44}$$

which thanks to the definition of $\inf_{M_N\in\mathcal{M}_N^{(\mathrm{LU})}}$ becomes

$$r_N\left(G_\alpha, M_N^\varepsilon, \boldsymbol{\theta}\right) \geqslant \alpha \inf_{M_N\in\mathcal{M}_N^{(\mathrm{LU})}} r_N(G_1, M_N, \boldsymbol{\theta}) + (1-\alpha) \inf_{M_N\in\mathcal{M}_N^{(\mathrm{LU})}} r_N(G_2, M_N, \boldsymbol{\theta}),$$

finally, substituting $r_N(G, \boldsymbol{\theta})$, we get

$$r_N\left(G_\alpha, M_N^\varepsilon, \boldsymbol{\theta}\right) \geqslant \alpha r_N\left(G_1, \boldsymbol{\theta}\right) + (1-\alpha)\, r_N\left(G_2, \boldsymbol{\theta}\right). \tag{B.45}$$

Putting together (B.43) and (B.45) gives

$$r_N\left(G_\alpha, \boldsymbol{\theta}\right) \geqslant \alpha r_N\left(G_1, \boldsymbol{\theta}\right) + (1-\alpha)\, r_N\left(G_2, \boldsymbol{\theta}\right) - \varepsilon, \tag{B.46}$$

which for $\epsilon \to 0$ is the (Ky Fan) concavity condition for $r_N\left(G, \boldsymbol{\theta}\right)$. Let us now prove the Ky Fan convexity in N. Consider $N_1, N_2 \in \mathbb{N}_0$ and an arbitrary $G \in \mathcal{G}$, we have

$$r_{N_1N_2}\left(G, \boldsymbol{\theta}\right) \leqslant \alpha\, r_{N_1}\left(G, \boldsymbol{\theta}\right) + (1-\alpha) r_{N_2}\left(G, \boldsymbol{\theta}\right) \quad \forall\, \alpha \in [0,1]. \tag{B.47}$$

This is true because thanks to lemma 3 we have $r_{N_1N_2}(G, \boldsymbol{\theta}) \leqslant r_{N_1}(G, \boldsymbol{\theta})$ and $r_{N_1N_2}(G, \boldsymbol{\theta}) \leqslant r_{N_2}(G, \boldsymbol{\theta})$. Lemma 5 proves that $r_N\left(G, \boldsymbol{\theta}\right)$ is upper semicontinuous in G for every fixed N, this concludes the hypothesis check for the application of the Ky Fan Minimax theorem, according to which

$$\sup_{\mathcal{G}} \inf_{N\geqslant 1} r_N\left(G, \boldsymbol{\theta}\right) = \inf_{N\geqslant 1} \sup_{\mathcal{G}} r_N\left(G, \boldsymbol{\theta}\right). \tag{B.48}$$

In order to get (B.41) from (B.40) we still need

$$\sup_{\mathcal{G}} \inf_{M_N\in\mathcal{M}_N^{(\mathrm{LU})}} r_N\left(G, M_N, \boldsymbol{\theta}\right) = \inf_{M_N\in\mathcal{M}_N^{(\mathrm{LU})}} \sup_{\mathcal{G}} r_N\left(G, M_N, \boldsymbol{\theta}\right). \tag{B.49}$$

This is the content of lemma 2. By putting together (B.39) and (B.41) we get the expression

$$\underline{r}(\boldsymbol{\theta}) = \sup_{G\geqslant 0} \lim_{n\to\infty} \underline{r}_n^{(<)}(G, \boldsymbol{\theta}). \tag{B.50}$$

Expanding this expression we arrive at (24), with $C(G, \boldsymbol{\theta})$ defined in (25).

Appendix C. Formulation of the semidefinite program

We start from (26) and write

$$\underline{r}(\boldsymbol{\theta}) = \sup_{G\geqslant 0} \frac{C_{\mathrm{H}}\left(G, \boldsymbol{\theta}\right)}{\mathrm{Tr}\left[G\cdot F(\boldsymbol{\theta})^{-1}\right]} = \sup_{G\in\mathcal{G}} C_{\mathrm{H}}\left(F(\boldsymbol{\theta})^{\frac{1}{2}} G F(\boldsymbol{\theta})^{\frac{1}{2}}, \boldsymbol{\theta}\right). \tag{C.1}$$

The SDP for $C_{\mathrm{H}}\left(G, \boldsymbol{\theta}\right)$ reported in [49] is

$$C_{\mathrm{H}}\left(G, \boldsymbol{\theta}\right) = \underset{V\in\mathbb{S}^n, X\in\mathbb{R}^{\tilde{d}\times n}}{\text{minimize}} \ \mathrm{Tr}[G\cdot V] \text{ subject to } \begin{pmatrix} V & X^T R_{\boldsymbol{\theta}}^\dagger \\ R_{\boldsymbol{\theta}} X & \mathbb{1}_{\tilde{r}} \end{pmatrix} \geqslant 0, X^T \frac{\partial \boldsymbol{s}_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}} = \mathbb{1}_n. \tag{C.2}$$

See the work [49] for the definitions of all the objects appearing in this program, it is not necessary to understand them in order to follow our derivation. equation (C.1) becomes

$$\underline{r}(\boldsymbol{\theta}) = \sup_{G\in\mathcal{G}} \underset{V\in\mathbb{S}^n, X\in\mathbb{R}^{\tilde{d}\times n}}{\text{minimize}} \ \mathrm{Tr}\left[F(\boldsymbol{\theta})^{\frac{1}{2}} G F(\boldsymbol{\theta})^{\frac{1}{2}} \cdot V\right] \text{ subject to } \begin{pmatrix} V & X^T R_{\boldsymbol{\theta}}^\dagger \\ R_{\boldsymbol{\theta}} X & \mathbb{1}_{\tilde{r}} \end{pmatrix} \geqslant 0, X^T \frac{\partial \boldsymbol{s}_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}} = \mathbb{1}_n. \tag{C.3}$$

The objective function $\mathrm{Tr}[F(\boldsymbol{\theta})^{\frac{1}{2}} G F(\boldsymbol{\theta})^{\frac{1}{2}} \cdot V]$ is linear and continuous in both G and V. The domain of the sup and the min are both convex, with $\mathcal{G}$ being compact. We can therefore apply again the Mimimax theorem of [79] as done in appendix B. Having $\sup_{G\in\mathcal{G}}$ as the innermost operation we can solve it and write

$$\sup_{G\in\mathcal{G}} \mathrm{Tr}\left[F(\boldsymbol{\theta})^{\frac{1}{2}} G F(\boldsymbol{\theta})^{\frac{1}{2}} \cdot V\right] = \|F(\boldsymbol{\theta})^{\frac{1}{2}} V F(\boldsymbol{\theta})^{\frac{1}{2}}\|. \tag{C.4}$$

Now the objective of the minimization is the spectral norm of $F^{\frac{1}{2}}(\boldsymbol{\theta})VF^{\frac{1}{2}}(\boldsymbol{\theta})$. We can introduce a dummy variable t and write the program as

$$\underline{r}(\boldsymbol{\theta}) = \underset{V\in\mathbb{S}^n, X\in\mathbb{R}^{\tilde{d}\times n}}{\text{minimize}}\ t \text{ subject to } \|F(\boldsymbol{\theta})^{\frac{1}{2}}VF(\boldsymbol{\theta})^{\frac{1}{2}}\| \leqslant t, \begin{pmatrix} V & X^T R_{\boldsymbol{\theta}}^{\dagger} \\ R_{\boldsymbol{\theta}} X & \mathbb{1}_{\tilde{r}} \end{pmatrix} \geqslant 0, X^T \frac{\partial \boldsymbol{s}_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}} = \mathbb{1}_n. \tag{C.5}$$

The condition on $\|F(\boldsymbol{\theta})^{\frac{1}{2}}VF(\boldsymbol{\theta})^{\frac{1}{2}}\|$ can be written as

$$\lambda_{\max}(F(\boldsymbol{\theta})^{\frac{1}{2}}VF(\boldsymbol{\theta})VF(\boldsymbol{\theta})^{\frac{1}{2}}) \leqslant t^2 \Rightarrow F(\boldsymbol{\theta})^{\frac{1}{2}}VF(\boldsymbol{\theta})VF(\boldsymbol{\theta})^{\frac{1}{2}} \leqslant t^2\mathbb{1} \Rightarrow$$
$$t\mathbb{1} - F(\boldsymbol{\theta})^{\frac{1}{2}}VF(\boldsymbol{\theta})^{\frac{1}{2}}(t\mathbb{1})^{-1}F(\boldsymbol{\theta})^{\frac{1}{2}}VF(\boldsymbol{\theta})^{\frac{1}{2}} \geqslant 0.$$

Because of the Schur complement condition for the positive semidefinite matrices [83] the optimization becomes

$$\underline{r}(\boldsymbol{\theta}) = \underset{\substack{V\in\mathbb{S}^n,\\ X\in\mathbb{R}^{\tilde{d}\times n}}}{\text{minimize}}\ t \text{ subject to } \begin{pmatrix} t\mathbb{1} & F(\boldsymbol{\theta})^{\frac{1}{2}}VF(\boldsymbol{\theta})^{\frac{1}{2}} \\ F(\boldsymbol{\theta})^{\frac{1}{2}}VF(\boldsymbol{\theta})^{\frac{1}{2}} & t\mathbb{1} \end{pmatrix} \geqslant 0, \begin{pmatrix} V & X^T R_{\boldsymbol{\theta}}^{\dagger} \\ R_{\boldsymbol{\theta}} X & \mathbb{1}_{\tilde{r}} \end{pmatrix} \geqslant 0, X^T \frac{\partial \boldsymbol{s}_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}} = \mathbb{1}_n. \tag{C.6}$$

From which we compute $\mathcal{I}(\boldsymbol{\theta})$ according to (27). This SDP is solved by means of the modeling system CVX developed on Matlab [84].

Appendix D. Explicit computation of $\underline{r}^\star(\boldsymbol{\theta})$

In this section we prove that $\sup_{G\geqslant 0}$ in the definition (30) of $\underline{r}^\star(\boldsymbol{\theta})$ can be computed exactly and we obtain the explicit expression for $\underline{r}^\star(\boldsymbol{\theta})$ in (31). First of all we define $A'(\boldsymbol{\theta}) := F(\boldsymbol{\theta})^{-\frac{1}{2}}A(\boldsymbol{\theta})F(\boldsymbol{\theta})^{-\frac{1}{2}}$. This means we can write (30) as

$$\begin{aligned} \underline{r}^\star(\boldsymbol{\theta}) - 1 &= \sup_{G\in\mathcal{G}} \text{Tr Abs}\left[G\cdot A'(\boldsymbol{\theta})\right] \\ &= \sup_{G\in\mathcal{G}} \text{Tr}\ |\sqrt{G}A'(\boldsymbol{\theta})\sqrt{G}| \\ &= \sup_{G\in\mathcal{G}} \text{Tr}\left[\sqrt{\sqrt{G}\left(-A'(\boldsymbol{\theta})GA'(\boldsymbol{\theta})\right)\sqrt{G}}\right], \end{aligned} \tag{D.1}$$

with the sup taken on $\mathcal{G}$. Because $A(\boldsymbol{\theta})^\dagger = -A(\boldsymbol{\theta})$ it holds that $-A(\boldsymbol{\theta})'GA(\boldsymbol{\theta})' \geqslant 0$. The trace in (D.1) can be associated to the definition of the (squared) fidelity between the states identified by the matrices G and $-A(\boldsymbol{\theta})'GA(\boldsymbol{\theta})'$. Notice that this last state must be normalized. Therefore we write

$$\underline{r}^\star(\boldsymbol{\theta}) - 1 = \sup_{G\in\mathcal{G}} \sqrt{\mathcal{F}}\left(G, -A(\boldsymbol{\theta})'GA(\boldsymbol{\theta})'\right), \tag{D.2}$$

$$= \sqrt{\text{Tr}\left[-A(\boldsymbol{\theta})'GA(\boldsymbol{\theta})'\right]} \cdot \sup_{G\in\mathcal{G}} \sqrt{\mathcal{F}}\left(G, \frac{-A(\boldsymbol{\theta})'GA(\boldsymbol{\theta})'}{\text{Tr}\left[-A(\boldsymbol{\theta})'GA(\boldsymbol{\theta})'\right]}\right). \tag{D.3}$$

We will prove that there is a choice of G that gives both the maximum of $\text{Tr}\left[-A(\boldsymbol{\theta})'GA(\boldsymbol{\theta})'\right]$ and of the squared fidelity. Let us write $A(\boldsymbol{\theta})'$ in the form $A(\boldsymbol{\theta})' = QMQ^T$ where M is a block diagonal matrix having 2×2 blocks

$$M_i := \begin{pmatrix} 0 & \lambda_i \\ -\lambda_i & 0 \end{pmatrix},$$

with $0 \leqslant \lambda_{i+1} \leqslant \lambda_i \in \mathbb{R}$. If $A'(\boldsymbol{\theta})$ is of odd size the matrix M has the last row and column full of 0. We have $\text{Tr}\left[-A(\boldsymbol{\theta})'GA(\boldsymbol{\theta})'\right] = \text{Tr}[-M\widetilde{G}M]$, with $\widetilde{G} = Q^TGQ$, which explicitly reads

$$\text{Tr}[-M\widetilde{G}M] = \lambda_1^2\left(\widetilde{G}_{11} + \widetilde{G}_{22}\right) + \lambda_2^2\left(\widetilde{G}_{33} + \widetilde{G}_{44}\right) + \lambda_3^2\left(\widetilde{G}_{55} + \widetilde{G}_{66}\right) + \cdots.$$

The maximum of the above expression is λ_1^2, realized for a $\widetilde{G}$ having $\widetilde{G}_{11} + \widetilde{G}_{22} = 1$ and all the other matrix elements null. Notice that $\|F^{-\frac{1}{2}}A(\boldsymbol{\theta})F^{-\frac{1}{2}}\| = \|A(\boldsymbol{\theta})'\| = \lambda_1$. For the square fidelity to reach its maximum ($\sqrt{\mathcal{F}} = 1$) it must be

$$G = -\frac{A(\boldsymbol{\theta})'GA(\boldsymbol{\theta})'}{\text{Tr}\left[-A(\boldsymbol{\theta})'GA(\boldsymbol{\theta})'\right]}, \tag{D.4}$$

this is realized for $\widetilde{G}_{11} = \widetilde{G}_{22} = \frac{1}{2}$. Therefore we have build implicitly a matrix G that saturates the sup and gives (31).

Appendix E. Effective generators for $\theta \neq 0$

Consider the single qubit encoding given in (46), in order to compute the relevant metrological quantities, for example the QFI and the Holevo–Cramér–Rao bound, it is necessary to take the derivatives of this evolution, evaluated at $\boldsymbol{\theta}$, i.e. $\frac{\partial U_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}}$. If we are sensing small deviations of the phases around $\boldsymbol{\theta} = 0$, then these expressions are fairly easily computable, they are indeed $\frac{\partial U_{\boldsymbol{\theta}}}{\partial \theta_1}\Big|_{\boldsymbol{\theta}} = \mathrm{i}\sigma_y$, and $\frac{\partial U_{\boldsymbol{\theta}}}{\partial \theta_2}\Big|_{\boldsymbol{\theta}} = \mathrm{i}\sigma_z$. But if the base point of the sensing process is $\boldsymbol{\theta} \neq 0$, then these derivatives became cumbersome, and can hinder the derivation of simple analytical results. To overcome this issue we show in this section that the metrological properties of the estimation at a point $\boldsymbol{\theta} \neq 0$ are equivalent to that of a sensing process around zero, where the encoding has the effective generators $H_1^{\mathrm{eff}}(\boldsymbol{\theta})$ and $H_2^{\mathrm{eff}}(\boldsymbol{\theta})$, which are null-trace Hermitian operators depending on the non null point $\boldsymbol{\theta}$ and are in general not simply σ_y and σ_z. We write explicitly the small variations $\delta\boldsymbol{\theta}$ from the base point $\boldsymbol{\theta}$ in the encoding (46), i.e.

$$U_{\boldsymbol{\theta}+\delta\boldsymbol{\theta}} := \exp\left(\mathrm{i}[(\theta_1 + \delta\theta_1)\sigma_y + (\theta_2 + \delta\theta_2)\sigma_z]\right), \tag{E.1}$$

$$:= \exp\left(\mathrm{i}H + \mathrm{i}\delta H\right). \tag{E.2}$$

The variables $\delta\boldsymbol{\theta}$ are now the unknown parameters, while $\boldsymbol{\theta}$ is known and fixed. The Hamiltonians $H := \theta_1\sigma_y + \theta_2\sigma_z$ and $\delta H := \delta\theta_1\sigma_y + \delta\theta_2\sigma_z$ have been defined. We expand the expression for $U_{\boldsymbol{\theta}+\delta\boldsymbol{\theta}}$ in terms of H and δH with the Baker–Campbell–Hausdorff formula, and keep only the first order terms in the infinitesimal variation $\delta\boldsymbol{\theta}$, obtaining

$$U_{\boldsymbol{\theta}+\delta\boldsymbol{\theta}} \simeq U_{\boldsymbol{\theta}} \exp\left(\mathrm{i}\delta H - \frac{1}{2}[\mathrm{i}H, \mathrm{i}\delta H] + \frac{1}{6}[\mathrm{i}H, [\mathrm{i}H, \mathrm{i}\delta H]] + \dots\right). \tag{E.3}$$

Now, the idea is to perform the rotation $U_{-\boldsymbol{\theta}}$ on the probe after the encoding with $U_{\boldsymbol{\theta}+\delta\boldsymbol{\theta}}$, in such way we compensate for the know component of the rotation $U_{\boldsymbol{\theta}+\delta\boldsymbol{\theta}}$ and leave only a term depending on the new unknown variables $\delta\boldsymbol{\theta}$.

$$U_{-\boldsymbol{\theta}}U_{\boldsymbol{\theta}+\delta\boldsymbol{\theta}} \simeq \exp\left(\mathrm{i}\delta H - \frac{1}{2}[\mathrm{i}H, \mathrm{i}\delta H] + \frac{1}{6}[\mathrm{i}H, [\mathrm{i}H, \mathrm{i}\delta H]] + \dots\right), \tag{E.4}$$

$$= \exp\left(\mathrm{i}\left[\delta\theta_1 H_1^{\mathrm{eff}}(\boldsymbol{\theta}) + \delta\theta_2 H_2^{\mathrm{eff}}(\boldsymbol{\theta})\right]\right). \tag{E.5}$$

In the last expression we have collected the terms multiplied by θ_1 and θ_2 respectively, which have been named $H_1^{\mathrm{eff}}(\boldsymbol{\theta})$ and $H_2^{\mathrm{eff}}(\boldsymbol{\theta})$. Notice that the commutator of two skew-Hermitian operators like $\mathrm{i}H$ and $\mathrm{i}\delta H$ is again skew-Hermitian, this applies to all the elements of the exponentiated sum in (E.4), and means that the right-hand side of (E.4) is a unitary operator even if we have neglected higher order terms in $\delta\boldsymbol{\theta}$. The exponentiated sum is either equal to $\mathrm{i}\delta\theta_1 H_1^{\mathrm{eff}}(\boldsymbol{\theta})$ or to $\mathrm{i}\delta\theta_2 H_2^{\mathrm{eff}}(\boldsymbol{\theta})$ when we set either $\delta\theta_2 = 0$ or $\delta\theta_1 = 0$, therefore the effective generators are also Hermitian operators. Consider a probe ρ codified by $U_{\boldsymbol{\theta}+\delta\boldsymbol{\theta}}$, i.e. $\rho_{\boldsymbol{\theta}+\delta\boldsymbol{\theta}} \simeq U_{\boldsymbol{\theta}+\delta\boldsymbol{\theta}}\rho U^\dagger_{\boldsymbol{\theta}+\delta\boldsymbol{\theta}}$. All the informational quantities remain the same if a known unitary is applied to the state, indeed its effects can be always absorbed at the measurement stage (if the selected measurements set allows to do so). By choosing $U_{-\boldsymbol{\theta}}$ to be this unitary we get $U_{-\boldsymbol{\theta}}U_{\boldsymbol{\theta}+\delta\boldsymbol{\theta}}\rho U^\dagger_{\boldsymbol{\theta}+\delta\boldsymbol{\theta}}U^\dagger_{-\boldsymbol{\theta}} = U_{\delta\boldsymbol{\theta}}\rho U^\dagger_{\delta\boldsymbol{\theta}}$, with $U_{\delta\boldsymbol{\theta}} := U_{-\boldsymbol{\theta}}U_{\boldsymbol{\theta}+\delta\boldsymbol{\theta}}$. We observe that the traces of $H_1^{\mathrm{eff}}(\boldsymbol{\theta})$ and $H_1^{\mathrm{eff}}(\boldsymbol{\theta})$ can be neglected without consequences, indeed they contribute only to a global phase. Also if the gate $U_{\boldsymbol{\theta}}$ is used multiple times on an entangled state, so that the encoding is $U_{\boldsymbol{\theta}} \otimes U_{\boldsymbol{\theta}} \otimes \cdots \otimes U_{\boldsymbol{\theta}}$, the traces of the generators only give an irrelevant global phase. We now further manipulate the encoding and look for a parameterization in which the generators are orthogonal. Two qubits operators H_1 and H_2 are said to be orthogonal if $\{H_1, H_2\} = 0$. As null-trace Hermitian operators on a qubit $H_1^{\mathrm{eff}}(\boldsymbol{\theta})$ and $H_2^{\mathrm{eff}}(\boldsymbol{\theta})$ can be written

$$H_1^{\mathrm{eff}}(\boldsymbol{\theta}) = \alpha_1(\boldsymbol{\theta})\sigma_x + \beta_1(\boldsymbol{\theta})\sigma_y + \gamma_1(\boldsymbol{\theta})\sigma_z, \tag{E.6}$$

$$H_2^{\mathrm{eff}}(\boldsymbol{\theta}) = \alpha_2(\boldsymbol{\theta})\sigma_x + \beta_2(\boldsymbol{\theta})\sigma_y + \gamma_2(\boldsymbol{\theta})\sigma_z, \tag{E.7}$$

with $\boldsymbol{\alpha}(\boldsymbol{\theta}) := (\alpha_1(\boldsymbol{\theta}), \alpha_2(\boldsymbol{\theta}), \alpha_3(\boldsymbol{\theta})) \in \mathbb{R}^3$ and $\boldsymbol{\beta}(\boldsymbol{\theta}) := (\beta_1(\boldsymbol{\theta}), \beta_2(\boldsymbol{\theta}), \beta_3(\boldsymbol{\theta})) \in \mathbb{R}^3$. The orthogonality condition is then $\{H_1^{\mathrm{eff}}(\boldsymbol{\theta}), H_2^{\mathrm{eff}}(\boldsymbol{\theta})\} = 2\boldsymbol{\alpha}(\boldsymbol{\theta}) \cdot \boldsymbol{\beta}(\boldsymbol{\theta})\,\mathbb{1}$. We can decompose $H_2^{\mathrm{eff}}(\boldsymbol{\theta})$ in a term proportional to $H_1^{\mathrm{eff}}(\boldsymbol{\theta})$ and one orthogonal as following

$$H_2^{\mathrm{eff}}(\boldsymbol{\theta}) := \frac{\boldsymbol{\alpha}(\boldsymbol{\theta}) \cdot \boldsymbol{\beta}(\boldsymbol{\theta})}{\|\boldsymbol{\alpha}(\boldsymbol{\theta})\|^2} H_1^{\mathrm{eff}}(\boldsymbol{\theta}) + H_2^{\perp\mathrm{eff}}(\boldsymbol{\theta}), \tag{E.8}$$

this is the definition of $H_2^{\perp\text{eff}}(\boldsymbol{\theta})$, which satisfies $\{H_1^{\text{eff}}(\boldsymbol{\theta}), H_2^{\perp\text{eff}}(\boldsymbol{\theta})\} = 0$. We define $x(\boldsymbol{\theta}) := \boldsymbol{\alpha}(\boldsymbol{\theta}) \cdot \boldsymbol{\beta}(\boldsymbol{\theta}) / \| \boldsymbol{\alpha}(\boldsymbol{\theta}) \|^2$ for ease of notation and substitute (E.8) in (E.5), thus getting

$$U_{\delta\boldsymbol{\theta}} = \exp\left(\mathrm{i}[(\delta\theta_1 + x(\boldsymbol{\theta})\delta\theta_2)H_1^{\perp\text{eff}}(\boldsymbol{\theta}) + \delta\theta_2 H_2^{\perp\text{eff}}(\boldsymbol{\theta})]\right), \tag{E.9}$$

where $H_1^{\text{eff}}(\boldsymbol{\theta})$ has been renamed $H_1^{\perp\text{eff}}(\boldsymbol{\theta})$. The final step is to normalize the generators, thus defining $\widetilde{H}_1^{\perp\text{eff}}(\boldsymbol{\theta}) := H_1^{\perp\text{eff}}(\boldsymbol{\theta}) / \operatorname{Tr}\left[H_1^{\perp\text{eff}}(\boldsymbol{\theta})^2\right]$, and $\widetilde{H}_1^{\perp\text{eff}}(\boldsymbol{\theta}) := H_2^{\perp\text{eff}}(\boldsymbol{\theta}) / \operatorname{Tr}\left[H_2^{\perp\text{eff}}(\boldsymbol{\theta})^2\right]$. Going from $H_i^{\text{eff}}(\boldsymbol{\theta})$ to $\widetilde{H}_i^{\perp\text{eff}}(\boldsymbol{\theta})$ corresponds to the following reparametrization

$$\begin{cases} \delta\theta_1' = \operatorname{Tr}\left[H_2^{\perp\text{eff}}(\boldsymbol{\theta})^2\right](\delta\theta_1 + x(\boldsymbol{\theta})\delta\theta_2), \\ \delta\theta_2' = \operatorname{Tr}\left[H_2^{\perp\text{eff}}(\boldsymbol{\theta})^2\right]\delta\theta_2. \end{cases} \tag{E.10}$$

A rotation of the reference system can align $\widetilde{H}_i^{\perp\text{eff}}(\boldsymbol{\theta})$ with σ_y and σ_z, remember thought that the probe state must also be transformed. Let us introduce the unitary $V_{\boldsymbol{\theta}}$ such that $V_{\boldsymbol{\theta}}\widetilde{H}_1^{\perp\text{eff}}(\boldsymbol{\theta})V_{\boldsymbol{\theta}}^\dagger = \sigma_y$ and $V_{\boldsymbol{\theta}}\widetilde{H}_2^{\perp\text{eff}}(\boldsymbol{\theta})V_{\boldsymbol{\theta}}^\dagger = \sigma_z$, then $V_{\boldsymbol{\theta}}U_{\delta\boldsymbol{\theta}}V_{\boldsymbol{\theta}}^\dagger = \mathrm{e}^{\mathrm{i}(\delta\theta_1'\sigma_y + \delta\theta_2'\sigma_z)}$, while the probe state becomes $V_{\boldsymbol{\theta}}\rho V_{\boldsymbol{\theta}}^\dagger$.

ORCID iDs

Federico Belliardo https://orcid.org/0000-0002-1466-396X

References

[1] Giovannetti V, Lloyd S and Maccone L 2004 Quantum-enhanced measurements: beating the standard quantum limit *Science* **306** 1330–6
[2] Giovannetti V, Lloyd S and Maccone L 2006 Quantum metrology *Phys. Rev. Lett.* **96** 010401
[3] Giovannetti V, Lloyd S and Maccone L 2011 Advances in quantum metrology *Nat. Photon.* **5** 222–9
[4] Braun D, Adesso G, Benatti F, Floreanini R, Marzolino U, Mitchell M W and Pirandola S 2018 Quantum-enhanced measurements without entanglement *Rev. Mod. Phys.* **90** 035006
[5] Polino E, Valeri M, Spagnolo N and Sciarrino F 2020 Photonic quantum metrology *AVS Quantum Sci.* **2** 024703
[6] Taylor M A and Bowen W P 2016 Quantum metrology and its application in biology *Phys. Rep.* **615** 1–59
[7] Caves C M 1981 Quantum-mechanical noise in an interferometer *Phys. Rev.* D **23** 1693
[8] Demkowicz-Dobrzański R, Jarzyna M and Kołodyński J 2015 Chapter four-quantum limits in optical interferometry *Prog. Opt.* **60** 345–435
[9] Acernese F *et al* 2019 Increasing the astrophysical reach of the advanced Virgo detector via the application of squeezed vacuum states of light *Phys. Rev. Lett.* **123** 231108
[10] Tse M *et al* 2019 Quantum-enhanced advanced LIGO detectors in the era of gravitational-wave astronomy *Phys. Rev. Lett.* **123** 231107
[11] Budker D and Romalis M 2007 Optical magnetometry *Nat. Phys.* **3** 227–34
[12] Koschorreck M, Napolitano M, Dubost B and Mitchell M W 2010 Sub-projection-noise sensitivity in broadband atomic magnetometry *Phys. Rev. Lett.* **104** 093602
[13] Wasilewski W, Jensen K, Krauter H, Renema J J, Balabas M V and Polzik E S 2010 Quantum noise limited and entanglement-assisted magnetometry *Phys. Rev. Lett.* **104** 133601
[14] Sewell R J, Koschorreck M, Napolitano M, Dubost B, Behbood N and Mitchell M W 2012 Magnetic sensitivity beyond the projection noise limit by spin squeezing *Phys. Rev. Lett.* **109** 253605
[15] Troiani F and Paris M G A 2018 Universal quantum magnetometry with spin states at equilibrium *Phys. Rev. Lett.* **120** 260503
[16] Ludlow A D, Boyd M M, Ye J, Peik E and Schmidt P O 2015 Optical atomic clocks *Rev. Mod. Phys.* **87** 637
[17] Louchet-Chauvet A, Appel J, Renema J J, Oblak D, Kjaergaard N and Polzik E S 2010 Entanglement-assisted atomic clock beyond the projection noise limit *New J. Phys.* **12** 065032
[18] Kessler E M, Kómár P, Bishof M, Jiang L, Sørensen A S, Ye J and Lukin M D 2014 Heisenberg-limited atom clocks based on entangled qubits *Phys. Rev. Lett.* **112** 190403
[19] Holevo A S 2011 *Probabilistic and Statistical Aspects of Quantum Theory* (Pisa: Edizioni della Normale) ch 6 section 6.7
[20] Helstrom C W 1969 Quantum detection and estimation theory *J. Stat. Phys.* **1** 231–52
[21] Fisher R A 1971 *The Design of Experiments* 9th edn (London: Macmillan)
[22] James F 2006 *Statistical Methods In Experimental Physics* 2nd edn (Singapore: World Scientific)
[23] Szczykulska M, Baumgratz T and Datta A 2016 Multi-parameter quantum metrology *Adv. Phys.:* X **1** 621–39
[24] Albarelli F, Barbieri M, Genoni M G and Gianani I 2020 A perspective on multiparameter quantum metrology: from theoretical tools to applications in quantum imaging *Phys. Lett.* A **384** 126311
[25] Vidrighin M D, Donati G, Genoni M G, Jin X-M, Kolthammer W S, Kim M S, Datta A, Barbieri M and Walmsley I A 2014 Joint estimation of phase and phase diffusion for quantum metrology *Nat. Commun.* **5** 3532
[26] Altorio M, Genoni M G, Vidrighin M D, Somma F and Barbieri M 2015 Weak measurements and the joint estimation of phase and phase diffusion *Phys. Rev.* A **92** 032114
[27] Roccia E, Gianani I, Mancino L, Sbroscia M, Somma F, Genoni M G and Barbieri M 2018 Entangling measurements for multiparameter estimation with two qubits *Quantum Sci. Technol.* **3** 01LT01
[28] Szczykulska M, Baumgratz T and Datta A 2017 Reaching for the quantum limits in the simultaneous estimation of phase and phase diffusion *Quantum Sci. Technol.* **2** 044004
[29] Crowley P J D, Datta A, Barbieri M and Walmsley I A 2014 Tradeoff in simultaneous quantum-limited phase and loss estimation in interferometry *Phys. Rev.* A **89** 023845

[30] Parniak M, Borówka S, Boroszko K, Wasilewski W, Banaszek K and Demkowicz-Dobrzański R 2018 Beating the Rayleigh limit using two-photon interference *Phys. Rev. Lett.* **121** 250503

[31] Pezzè L, Ciampini M A, Spagnolo N, Humphreys P C, Datta A, Walmsley I A, Barbieri M, Sciarrino F and Smerzi A 2017 Optimal measurements for simultaneous quantum estimation of multiple phases *Phys. Rev. Lett.* **119** 130504

[32] Humphreys P C, Barbieri M, Datta A and Walmsley I A 2013 Quantum enhanced multiple phase estimation *Phys. Rev. Lett.* **111** 070403

[33] Gagatsos C N, Branford D and Datta A 2016 Gaussian systems for quantum-enhanced multiple phase estimation *Phys. Rev.* A **94** 042342

[34] Knott P A, Proctor T J, Hayes A J, Ralph J F, Kok P and Dunningham J A 2016 Local versus global strategies in multiparameter estimation *Phys. Rev.* A **94** 062312

[35] Polino E, Riva M, Valeri M, Silvestri R, Corrielli G, Crespi A, Spagnolo N, Osellame R and Sciarrino F 2019 Experimental multiphase estimation on a chip *Optica* **6** 288–95

[36] Ciampini M A, Spagnolo N, Vitelli C, Pezzè L, Smerzi A and Sciarrino F 2016 Quantum-enhanced multiparameter estimation in multiarm interferometers *Sci. Rep.* **6** 28881

[37] Zhang L and Chan K W C 2017 Quantum multiparameter estimation with generalized balanced multimode NOON-like states *Phys. Rev.* A **95** 032321

[38] Baumgratz T and Datta A 2016 Quantum enhanced estimation of a multidimensional field *Phys. Rev. Lett.* **116** 030801

[39] Apellaniz I, Urizar-Lanz I, Zimborás Z, Hyllus P and Tóth G 2018 Precision bounds for gradient magnetometry with atomic ensembles *Phys. Rev.* A **97** 053603

[40] Heinosaari T, Miyadera T and Ziman M 2016 An invitation to quantum incompatibility *J. Phys. A: Math. Theor.* **49** 123001

[41] Holevo A S 1976 Noncommutative analogues of the Cramér-Rao inequality in the quantum measurement theory *Lect. Notes Math.* **550** 194–222

[42] Hayashi M 2009 Quantum estimation and the quantum central limit theorem *American Mathematical Society Translations Series 2* **277** 95–123

[43] Yamagata K, Fujiwara A and Gill R D 2013 Quantum local asymptotic normality based on a new quantum likelihood ratio *Ann. Stat.* **41** 2197–217

[44] Hayashi M and Matsumoto K 2008 Asymptotic performance of optimal state estimation in qubit system *J. Math. Phys.* **49** 102101

[45] Guţă M and Kahn J 2006 Local asymptotic normality for qubit states *Phys. Rev.* A **73** 052108

[46] Guţă M and Jenčová A 2007 Local asymptotic normality in quantum statistics *Commun. Math. Phys.* **276** 341–79

[47] Guţă M and Jenčová A 2009 Local asymptotic normality for finite dimensional quantum systems *Commun. Math. Phys.* **289** 597–652

[48] Yang Y, Chiribella G and Hayashi M 2019 Attaining the ultimate precision limit in quantum state estimation *Commun. Math. Phys.* **368** 223–93

[49] Albarelli F, Friel J F and Datta A 2019 Evaluating the Holevo Cramér–Rao bound for multiparameter quantum metrology *Phys. Rev. Lett.* **123** 200503

[50] Rondin L *et al* 2012 Nanoscale magnetic field mapping with a single spin scanning probe magnetometer *Appl. Phys. Lett.* **100** 153118

[51] Ragy S, Jarzyna M and Demkowicz-Dobrzański R 2016 Compatibility in multiparameter quantum metrology *Phys. Rev.* A **94** 052108

[52] Nielsen M A and Chuang I L Quantum Computation and Quantum Information (Cambridge: Cambridge University Press) ch 8 part 3 section 8.2.4.

[53] Suppose in fact that the POVM M_N produces outputs ω belonging to a numerical set $\Omega \neq \Theta$ which are then processed in order to build the estimator of θ through a function $\tilde{\boldsymbol{\theta}} : \Omega \to \Theta$. The associated statistics can then be equivalently described by means of a new POVM $\{E^{(N)}_{\hat{\boldsymbol{\theta}}}\}_{\hat{\boldsymbol{\theta}}}$ with output $\hat{\boldsymbol{\theta}} \in \Theta$ by setting $E^{(N)}_{\hat{\boldsymbol{\theta}}} = \int d\omega \delta(\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}(\omega)) Y_\omega$, where $\{Y\omega\}\omega \in \Omega$ are the elements of the original POVM and where $\delta(\dots)$ is the Dirac delta function

[54] Paris M G A 2009 Quantum estimation for quantum technology *Int. J. Quantum. Inf.* **07** 125–37

[55] Notice that by construction the infimum over M_N in (17) could in principle force the selected POVMs to bare a non trivial dependence upon the unknown value θ. This is a very well known fact in quantum metrology which is at the origin of a circular reasoning issue which in practice can be cured by adopting adaptive estimation approaches. In our case, however, the possible functional dependence of M_N on θ is not of particular concern as it simply reflects the local character of the figure of merit

[56] Razavian S, Paris M G A and Genoni M G 2020 On the quantumness of multiparameter estimation problems for qubit systems *Entropy* **22** 1197

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

[58] Braunstein S L, Caves C M and Milburn G J 1996 Generalized uncertainty relations: theory, examples, and lorentz invariance *Ann. Phys., NY* **247** 135–73

[59] Nielsen M A and Chuang I L 2010 *Quantum Computation and Quantum Information* (Cambridge: Cambridge University Press) ch 9 part 3 section 9.2.2

[60] Nagaoka H 2005 A new approach to Cramér–Rao bounds for quantum state estimation *Asymptotic Theory of Quantum Statistical Inference* ed M Hayashi (Singapore: World Scientific) ch 8

[61] Tsang M 2020 Quantum semiparametric estimation *Phys. Rev.* X **10** 031023

[62] Suzuki J 2016 Explicit formula for the Holevo bound for two-parameter qubit-state estimation problem *J. Math. Phys.* **57** 042201

[63] Carollo A, Spagnolo B, Dubkov A A and Valenti D 2019 On quantumness in multi-parameter quantum estimation *J. Stat. Mech.* 094010

[64] Carollo A, Valenti D and Spagnolo B 2020 Geometry of quantum phase transitions *Phys. Rep.* **838** 1–72

[65] Suzuki J 2019 Information geometrical characterization of quantum statistical models in quantum estimation theory *Entropy* **21** 703

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

[67] King C 2003 The capacity of the quantum depolarizing channel *IEEE Trans. Inf. Theory* **49** 221–9

[68] Rosati M and Giovannetti V 2018 Asymmetric information capacities of reciprocal pairs of quantum channels *Phys. Rev.* A **97** 052318

[69] The parameterization (α, β) is singular at the points $\alpha = \pi/2$ and $\alpha = 3\pi/3$, and so are the matrices $F(\alpha, \beta)$ and $A(\alpha, \beta)$. In case $a_{\boldsymbol{\theta}}$ exactly corresponds to one of these points we simply choose another set of spherical coordinates which have the singularity located at a different position.

[70] Fujiwara A 2001 Estimation of SU(2) operation and dense coding: an information geometric approach *Phys. Rev.* A **65** 012316
[71] Gisin N and Popescu S 1999 Spin flips and quantum information for antiparallel spins *Phys. Rev. Lett.* **83** 432
[72] Lina C, Nan L, Shunlong L and Hongting S 2014 Optimal extraction of information from two spins *Phys. Rev.* A **89** 042110
[73] Matsumoto K 2002 A new approach to the Cramér–Rao-type bound of the pure-state model *J. Phys. A: Math. Gen.* **35** 3111
[74] Jamiołkowski A 1972 Linear transformations which preserve trace and positive semidefiniteness of operators *Rep. Math. Phys.* **3** 275–8
[75] Jamiołkowski A 1975 Completely positive linear maps on complex matrices *Linear Algebr. Appl.* **10** 285–90
[76] Demkowicz-Dobrzański R, Kołodyński J and Guţă M 2012 The elusive Heisenberg limit in quantum-enhanced metrology *Nat. Commun.* **3** 1063
[77] Kołodyński J and Demkowicz-Dobrzański R 2013 Efficient tools for quantum metrology with uncorrelated noise *New J. Phys.* **15** 073043
[78] Chiribella G, D'Ariano G M and Schlingemann D M 2007 Realization of continuous-outcome measurements on finite dimensional quantum systems (arXiv:0703110v1 [quant-ph])
[79] Kneser H 1952 Sur un theoreme fondamental de la theorie des jeux [64–52b] *C. R. Acad. Sci., Paris* **234** 2418–20
[80] Frenk J B G, Kassay G and Kolumbán J 2004 On equivalent results in minimax theory *Eur. J. Oper. Res.* **157** 46–58
[81] Chiribella G, D'Ariano G M and Schlingemann D 2007 How continuous quantum measurements in finite dimensions are actually discrete *Phys. Rev. Lett.* **98** 190403
[82] Fan K 1953 Minimax theorems *Proc. Natl Acad. Sci.* **39** 42–7
[83] Zhang F 2005 The Schur complement and its applications *Numerical Methods and Algorithms* vol 4 (Berlin: Springer)
[84] Grant M and Boyd S 2020 CVX: MATLAB software for disciplined convex programming version 2.2 http://cvxr.com/cvx