Discontinuities of the quantum Fisher information and the Bures metric

Dominik Šafránek[*]

SCIPP, University of California, Santa Cruz, Natural Sciences II No. 337, 1156 High Street, Santa Cruz, California 95064, USA

We show that two quantities in quantum metrology that were thought to be the same, the quantum Fisher information matrix and the Bures metric, are not the same. They differ at points at which the rank of the density matrix changes. The quantum Fisher information matrix is discontinuous at these points. However, these discontinuities are removable in some sense. We show that the expression given by the Bures metric represents the continuous version of the quantum Fisher information matrix. We also derive an explicit formula for the Bures metric for both singular and nonsingular density matrices.

I. INTRODUCTION

The quantum Fisher information and the Bures metric are cornerstones of modern quantum metrology and quantum information geometry. They give the ultimate precision bound on the estimation of a parameter encoded in a quantum state known as the Cramér-Rao bound [1,2]. This bound gives the theoretical framework for maximizing the sensitivity of new-era quantum detectors, such as the recently improved [3] gravitational wave detector LIGO that confirmed the last missing piece in the Einstein's theory of relativity [4]. The quantum Fisher information and the Bures metric have been also used in the description of criticality and quantum phase transitions under the name of "fidelity susceptibility", where they help to describe a sudden change of a quantum state when an external parameter such as temperature is varied [5–8]. Last but not least, these measures also give the speed limits on the evolution of quantum states [9,10]. They have been used, for example, to estimate the speed limits of quantum computation [11] or speed limits in charging of batteries [12].

Introduced by Holevo [13], Helstrom [14,15], by Bures [16], and later popularized by Braunstein and Caves [1], the quantum Fisher information and the Bures metric describe limits in distinguishability of infinitesimally close quantum states $\hat{\rho}_\epsilon$ and $\hat{\rho}_{\epsilon+d\epsilon}$ that differ only by a small variation in parameters that parametrize them. To explain, assume we perform a measurement on these two states to distinguish them. We obtain two different statistics of measurement outcomes, and how well we can distinguish between these two statistics is given by a measure known as the Fisher information. Since statistics of measurement outcomes depend on both the quantum states and the chosen measurement, and because some measurements can lead to statistics that are easier to distinguish, to obtain the ultimate precision with what we can distinguish between the two close states we have to optimize over all such measurements. This then gives rise to the quantum Fisher information, which is a function only of the density matrix $\hat{\rho}_\epsilon$ [2]. Since the density matrix depends on the parameter to be estimated, distinguishing between two close density matrices $\hat{\rho}_\epsilon$ and $\hat{\rho}_{\epsilon+d\epsilon}$ is equivalent to distinguishing between two close parameters ϵ and $\epsilon + d\epsilon$ that

parametrize them. As a result, the quantum Fisher information measures how well the parameter ϵ itself can be estimated.

Despite the fact that both the quantum Fisher information and the Bures metric have been widely used before, they still contain a large number of strange and unexplored properties. For example, although these quantities are widely believed to be equal, finding the true connection between them is rather elusive. It has been shown that these quantities are the same when two infinitesimally close states that are being compared are pure [17], or when they are both described by a full-rank density matrix [1,18]. It has been suggested that this is also true when the density matrices are of arbitrary rank [19]; however, we will show that this is not true in general.

The quantum Fisher information and the Bures metric also exhibit mathematical features that make them difficult to handle and that are uncommon in physics. Expressions for the quantum Fisher information or the Bures metric that are valid for a certain quantum state are often undefined for a state of a lower rank. Formulas for the full-rank or one-rank density matrices are usually easy to obtain but hard to connect to each other, even when using appropriate limits, and deriving expressions for density matrices of an arbitrary rank is much harder. To connect known expressions or to derive new ones, unintuitive regularization procedures have to be employed [20,21], or uncommon operators such as the Moore-Penrose pseudoinverse have to be introduced [20]. Due to these difficulties, it is common in literature that discussions of glaringly pathological behavior such as $\frac{0}{0}$ of the derived expressions are often omitted. It is implicitly assumed that such expressions are either invalid when they are undefined, or that such expressions are still valid and the pathological terms are either set to zero [22] or to some other value, depending on a particular limit involved [21]. Performing those limits shows that expressions for the quantum Fisher information exhibit strange jumps (discontinuities) when a mixed state approaches a pure state, suggesting that the physics of pure states should differ from the physics of mixed states, a surprising statement on its own.

In this paper we study and describe this strange behavior and expose places where we can expect discontinuities in the figures of merit, even for density matrices that are analytical functions of the estimated parameters. We find that these discontinuities happen when a small change in the parameter of the density matrix changes the rank of the density matrix. This is also when two figures of merit of the local estimation theory—the quantum Fisher information matrix and the Bures

———————
[*]dsafrane@ucsc.edu

metric—do not coincide. Such scenarios are common when estimating noise in a quantum system [23–25] or when the parameter that we want to estimate is encoded into some larger quantum state, while an experimentalist has access only to a smaller subsystem [26–28]. This is inevitable in a quantum field theory in curved space-time because there are infinitely many modes that need to be traced over [29–34]. In all of these scenarios the change in the parameter changes the purity of a state. Therefore the rank of the density matrix can also change, which ultimately leads to a discontinuity.

This paper is structured as follows. We first give the necessary background and we review literature published on the topic. Then we present our main results: the relation between the quantum Fisher information matrix and the Bures metric (Theorem 1, Corollary 1, and Theorem 2); continuity of the Bures metric (Theorem 2); discontinuities of the quantum Fisher information matrix and the Bures metric (Corollary 2 and Theorem 3); and an expression for the quantum Fisher information matrix of any state as a limit of the quantum Fisher information matrix of a mixed state (Theorem 4). We accompany our text by three examples and four figures for better understanding. Finally, we discuss possible physical interpretations of points of discontinuity and conclude.

II. BACKGROUND

We use the following notation: We denote the vector of parameters as $\boldsymbol{\epsilon} = (\epsilon_1,...,\epsilon_n)$, and we denote the density matrix dependent on this vector as $\hat{\rho}_{\boldsymbol{\epsilon}}$. If a symbol with an index appears under the sum, the sum goes over all values of the index such that the property is satisfied. For example, $\sum_{p_k>0}$ means that the sum goes over all k such that $p_k > 0$. If there is no condition present, the sum goes over all indices written under the sum. We also usually drop writing the explicit dependence on the vector of parameters $\boldsymbol{\epsilon}$ unless we want to stress out this dependence. For example, instead of $p_i(\boldsymbol{\epsilon})$ we often write only p_i, but for $p_i(\boldsymbol{\epsilon} + d\boldsymbol{\epsilon})$ we write the full form. $d\boldsymbol{\epsilon} = (d\epsilon_1,...,d\epsilon_n)$ denotes a small variation in vector $\boldsymbol{\epsilon}$. We denote partial derivatives as $\partial_i \equiv \partial_{\epsilon_i}$ and $\partial_{ij} \equiv \partial_{\epsilon_i}\partial_{\epsilon_j}$. Derivatives with respect to elements of $d\boldsymbol{\epsilon}$ will be denoted as $\partial_{d\epsilon_i}$ for the first derivatives and $\partial_{d\epsilon_i d\epsilon_j}$ for the second derivatives. Elements of a matrix will be denoted by upper indices, e.g., H^{ij}, while different matrices or operators will be denoted by lower indices, e.g., $\hat{L}_i$. We also write the spectral decomposition of the density matrix as

$$\hat{\rho}_{\boldsymbol{\epsilon}} = \sum_k p_k |k\rangle\langle k|. \qquad (1)$$

We define symmetric logarithmic derivatives $\hat{L}_i$ [2] as operator solutions to equations

$$\tfrac{1}{2}(\hat{L}_i \hat{\rho}_{\boldsymbol{\epsilon}} + \hat{\rho}_{\boldsymbol{\epsilon}} \hat{L}_i) = \partial_i \hat{\rho}_{\boldsymbol{\epsilon}}. \qquad (2)$$

The quantum Fisher information matrix is then a symmetric positive or a positive semidefinite matrix defined as [2]

$$H^{ij}(\boldsymbol{\epsilon}) := \tfrac{1}{2}\text{tr}[(\hat{L}_i \hat{L}_j + \hat{L}_j \hat{L}_i)\hat{\rho}_{\boldsymbol{\epsilon}}]. \qquad (3)$$

Using the spectral decomposition (1) of the density matrix it is relatively easy[1] to check that Eq. (2) has solutions $\hat{L}_i = 2\sum_{p_k+p_l>0} \frac{\langle k|\partial_i\hat{\rho}_{\boldsymbol{\epsilon}}|l\rangle}{p_k+p_l}|k\rangle\langle l|$. Inserting these expressions into Eq. (3) gives the quantum Fisher information matrix,

$$H^{ij}(\boldsymbol{\epsilon}) = 2 \sum_{p_k+p_l>0} \frac{\text{Re}(\langle k|\partial_i\hat{\rho}_{\boldsymbol{\epsilon}}|l\rangle\langle l|\partial_j\hat{\rho}_{\boldsymbol{\epsilon}}|k\rangle)}{p_k+p_l}, \qquad (4)$$

where Re denotes the real part.

The quantum Fisher information matrix is the figure of merit in the multiparameter quantum Cramér-Rao bound, which gives a lower bound on the covariance matrix of the vector of locally unbiased estimators $\hat{\boldsymbol{\epsilon}}$ [2,35],

$$\text{Cov}[\hat{\boldsymbol{\epsilon}}] \geqslant H^{-1}(\boldsymbol{\epsilon}). \qquad (5)$$

$\text{Cov}[\hat{\boldsymbol{\epsilon}}] = \langle\hat{\epsilon}_i\hat{\epsilon}_j\rangle - \langle\hat{\epsilon}_i\rangle\langle\hat{\epsilon}_j\rangle$ is the covariance matrix and $H^{-1}(\boldsymbol{\epsilon})$ the inverse of the matrix defined in Eq. (3). The above equation should be understood as an operator inequality. It states that $\text{Cov}[\hat{\boldsymbol{\epsilon}}] - H^{-1}$ is a positive semidefinite or a positive definite matrix.

In this paper we study the connection between the quantum Fisher information matrix and the Bures metric (also known the "statistical distance" in older literature [1,17] and as the "fidelity susceptibility" in the condensed matter theory literature [7]). To define the Bures metric [16] we first introduce the Bures distance. The Bures distance is a measure of distinguishability between two quantum states $\hat{\rho}_{1,2}$ and it is defined through the Uhlmann fidelity [36]

$$\mathcal{F}(\hat{\rho}_1,\hat{\rho}_2) := (\text{tr}\sqrt{\sqrt{\hat{\rho}_1}\,\hat{\rho}_2\,\sqrt{\hat{\rho}_1}})^2 \qquad (6)$$

as

$$d_B^2(\hat{\rho}_1,\hat{\rho}_2) = 2(1 - \sqrt{\mathcal{F}(\hat{\rho}_1,\hat{\rho}_2)}). \qquad (7)$$

The Bures distance gives rise to the Bures metric g^{ij} through the definition for the line element,

$$\sum_{i,j} g^{ij}(\boldsymbol{\epsilon})d\epsilon_i d\epsilon_j := d_B^2(\hat{\rho}_{\boldsymbol{\epsilon}},\hat{\rho}_{\boldsymbol{\epsilon}+d\boldsymbol{\epsilon}}). \qquad (8)$$

This definition shows that the Bures metric measures the amount of distinguishability between two close density matrices $\hat{\rho}_{\boldsymbol{\epsilon}}$ and $\hat{\rho}_{\boldsymbol{\epsilon}+d\boldsymbol{\epsilon}}$ in the coordinate system $\boldsymbol{\epsilon}$. Precisely speaking, the above equation defines a metric tensor (or simply metric) g^{ij} induced by the Bures distance. The coordinate system $\boldsymbol{\epsilon}$ is not required to describe the entire manifold of density matrices but can rather define a submanifold. The metric (8) is then an induced metric on this submanifold. We define quantity H_c (which we will later call the continuous quantum Fisher information matrix) as four times the Bures metric,

$$H_c := 4g. \qquad (9)$$

The connection between the quantum Fisher information matrix and the Bures metric was extensively studied in

---

[1]Assuming $\hat{\rho}_{\boldsymbol{\epsilon}} \in C^{(1)}$, inserting $\hat{L}_i$ to Eq. (2) gives the left-hand side $LHS = \sum_{p_k+p_l>0}\langle k|\partial_i\hat{\rho}_{\boldsymbol{\epsilon}}|l\rangle|k\rangle\langle l|$, which together with $\sum_{p_k=p_l=0}\langle k|\partial_i\hat{\rho}_{\boldsymbol{\epsilon}}|l\rangle|k\rangle\langle l| = 0$ gives the right-hand side. The second identity comes from the fact that for $\boldsymbol{\epsilon}$ such that $p_k(\boldsymbol{\epsilon}) = 0$, also $\partial_i p_k(\boldsymbol{\epsilon}) = 0$ because p_k reaches the local minimum at point $\boldsymbol{\epsilon}$.

literature, particularly in several papers deriving explicit formulas for the statistical distance, the Bures metric, or the infinitesimal Bures distance [1,2,6,13,15–19,36–41]. We also point out papers related to the continuity of the quantum Fisher information [42,43].

There are three papers directly related to our study. In the first paper [1] Braunstein and Caves generalized the notion of the statistical distance from pure states to mixed states by maximizing the Fisher information over all possible quantum measurements. In today's terms, the resulting statistical distance is an equivalent of the quantum Fisher information defined in Eq. (4). Moreover, it was noted in [1] that the derived expression of the statistical distance is proportional to the infinitesimal Bures distance that was explicitly computed by Hübner [18]. What was not mentioned, however, is that the results of paper [18] are applicable only to nonsingular density matrices.[2] It is discussed there that in the case when $\hat{\rho}$ becomes singular the metric can be regularized by switching to a new set of coordinates and that the metric tensor g^{ij} remains finite. On the other hand, paper [18] does not provide an explicit expression for the infinitesimal Bures distance for the case of singular density matrices. Sommers and Życzkowski went a bit further by considering also singular density matrices [19]. However, the entire discussion of this topic is reduced to one sentence,[3] which leads to incorrect conclusions.

Performing the proof in Ref. [19] in detail using the same argumentation reveals that the resulting expression differs from the one that was published. The infinitesimal Bures distance is actually $d_B^2(\hat{\rho},\hat{\rho}+d\hat{\rho}) = \frac{1}{2}\sum_{p_k>0,\,p_l>0}\frac{|\langle k|d\hat{\rho}|l\rangle|^2}{p_k+p_l}$ and not $d_B^2(\hat{\rho},\hat{\rho}+d\hat{\rho}) = \frac{1}{2}\sum_{p_k+p_l>0}\frac{|\langle k|d\hat{\rho}|l\rangle|^2}{p_k+p_l}$ as stated in the paper. Intuitively, the reason why extra terms given by $p_k > 0$, $p_l = 0$ and $p_k = 0, p_l > 0$ do not appear in the sum can be understood in the following way: the argument $d_B^2(\hat{\rho},\hat{\rho}+d\hat{\rho})$ takes into account only the first-order correction $\hat{\rho}+d\hat{\rho}$, while the right-hand side depends on the second order $d\hat{\rho}^2$. The extra terms that are missing come from the second-order correction to the argument. However, to obtain these extra terms it is necessary to consider the expression given by $d_B^2(\hat{\rho}_\epsilon,\hat{\rho}_{\epsilon+d\epsilon})$ from Eq. (8) instead of $d_B^2(\hat{\rho},\hat{\rho}+d\hat{\rho})$, as has been done in Ref. [19] (see Appendix A for more detail).

We will show in detail that even when considering the right figure of merit $d_B^2(\hat{\rho}_\epsilon,\hat{\rho}_{\epsilon+d\epsilon})$ for the Bures metric, there is still a discrepancy between the Bures metric and the quantum Fisher information matrix in certain cases. To motivate this paper we illustrate this discrepancy in the following example.

Example 1. Consider a state where the parameter we estimate characterizes the purity of a quantum state,

$$\hat{\rho}_\epsilon = \sin^2\epsilon\,|0\rangle\langle0| + \cos^2\epsilon\,|1\rangle\langle1|. \tag{10}$$

The fidelity between two close states can be easily calculated as $\sqrt{\mathcal{F}(\hat{\rho}_\epsilon,\hat{\rho}_{\epsilon+d\epsilon})} = |\sin(\epsilon)\sin(\epsilon+d\epsilon)| + |\cos(\epsilon)\cos(\epsilon+d\epsilon)|$, which inserting into Eq. (8) and using

---
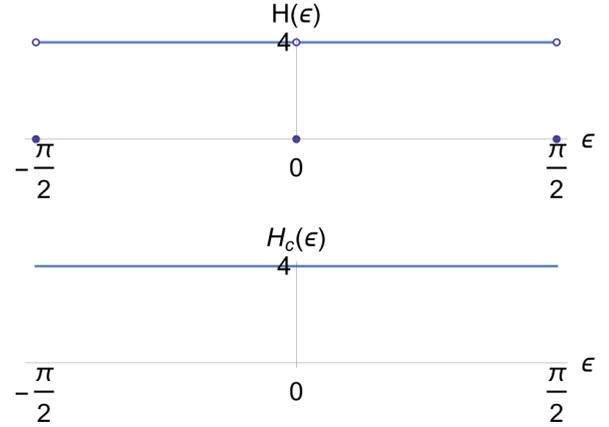
FIG. 1. The quantum Fisher information H and the (four times) Bures metric H_c give different results for the same density matrix (10).

definition (9) gives a constant function

$$H_c(\epsilon) = 4. \tag{11}$$

Using Eq. (4), for $\epsilon \neq k\frac{\pi}{2}$, $k \in \mathbb{Z}$ we find $H(\epsilon) = 4$. For $\epsilon = k\frac{\pi}{2}$ one term in the sum has its denominator equal to zero ($p_k + p_l = 0$) and therefore it is not considered in the sum, while other terms are zero. Together we have

$$H(\epsilon) = \begin{cases} 4 & \epsilon \neq k\frac{\pi}{2} \\ 0 & \epsilon = k\frac{\pi}{2}. \end{cases} \tag{12}$$

Graphs of functions H and H_c are shown in Fig. 1 .

This example shows that although the expression given by the quantum Fisher information matrix H and the four times Bures metric H_c give the same results everywhere where the density matrix is full-rank (nonsingular), the expressions differ at points ϵ at which an eigenvalue vanishes. As we will show in the following section, this is a completely general behavior. For parameterized quantum states $\hat{\rho}_\epsilon$ in which a slight change in the parameter ϵ results in an eigenvalue of the density matrix to vanish (or equivalently, results in an eigenvalue to "pop out"), the two figures of merit do not coincide. This discrepancy will be then connected to the discontinuous behavior of the quantum Fisher information matrix.

III. RESULTS

We assume $\hat{\rho}_\epsilon \in C^{(2)}$ in all following theorems, i.e., we assume that the second derivative of the density matrix exists and that it is a continuous function. Although the first theorem could be easily modified to require only the existence of the second derivative (with its discontinuity possibly resulting in $\partial_{ij}p_k \neq \partial_{ji}p_k$), the continuity of the second derivative is crucial for other theorems that speak about continuity of the quantum Fisher information matrix and the Bures metric.

Theorem 1. The Bures metric is connected to the quantum Fisher information matrix through the relation

$$H_c^{ij}(\boldsymbol{\epsilon}) = H^{ij}(\boldsymbol{\epsilon}) + 2\sum_{p_k(\boldsymbol{\epsilon})=0}\partial_{ij}p_k(\boldsymbol{\epsilon}). \tag{13}$$

$p_k(\boldsymbol{\epsilon}) = 0$ denotes that the sum goes over all values k such that their respective eigenvalue p_k vanishes at point $\boldsymbol{\epsilon}$. Defining the
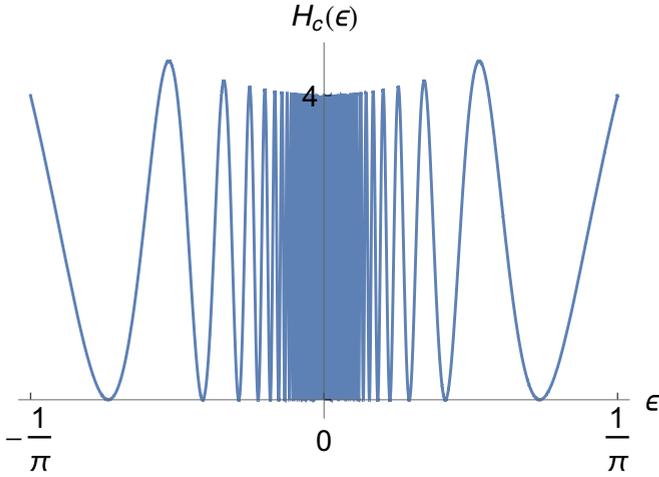
FIG. 2. The continuous quantum Fisher information of density matrix $\hat{\rho}_\epsilon = \{^{\epsilon^4 \sin^2 \frac{1}{\epsilon} |0\rangle\langle 0| + (1 - \epsilon^4 \sin^2 \frac{1}{\epsilon})|1\rangle\langle 1| \quad \epsilon \neq 0}_{|1\rangle\langle 1| \quad \epsilon = 0}$. The second derivative $\partial_{\epsilon\epsilon}\hat{\rho}_\epsilon$ exists everywhere, but it is discontinuous at point $\epsilon = 0$. Theorem 2 does not apply anymore and H_c does not have to be continuous. An explicit computation shows $H_c(\epsilon) = \{^{\frac{4(2\epsilon \sin \frac{1}{\epsilon} - \cos \frac{1}{\epsilon})^2}{1 - \epsilon^4 \sin^2 \frac{1}{\epsilon}} \quad \epsilon \neq 0}_{0 \quad \epsilon = 0}$.

Hessian matrices as $\mathcal{H}_k^{ij} := \partial_{ij} p_k$ we can also write Eq. (13) in an elegant matrix form, $H_c = H + 2\sum_{p_k=0} \mathcal{H}_k$.

Proof. See Appendix B.

Theorem 1 gives an explicit formula for the Bures metric for both singular and nonsingular density matrices, and it generalizes the result of [18] by including singular matrices. It also shows that the (four times) Bures metric and the quantum Fisher information matrix do not coincide only at certain points ϵ at which an eigenvalue vanishes. When the density matrix $\hat{\rho}_\epsilon$ is full rank (nonsingular), or when the change of the parameter does not result in the change of purity, for example, when the operation encoding ϵ is a unitary operation, the (four times) Bures metric H_c and the quantum Fisher information matrix H are identical. It is worth noting that the Hessian matrix $\mathcal{H}_k(\epsilon)$ is a positive or a positive semidefinite matrix because p_k reaches the local minimum at point ϵ for which $p_k(\epsilon) = 0$. We can sum these findings in the following matrix inequality.

Corollary 1.

$$H_c \geqslant H, \tag{14}$$

and $H_c = H$ if and only if for all k and ϵ such that $p_k(\epsilon) = 0$, $\mathcal{H}_k(\epsilon) = 0$.

Proof. The inequality comes from the fact that $\mathcal{H}_k(\epsilon)$ is a positive semidefinite or a positive definite matrix. The equality condition comes directly from Eq. (13). ∎

Next we show that the quantity H_c is in a certain sense a continuous version of the quantum Fisher information matrix. The discontinuous points of the quantum Fisher information matrix are redefined as the limits of the quantum Fisher information matrix of nearby points.

Theorem 2. We denote a unit vector with number 1 at the l'th position as $\boldsymbol{e}_l = (0, \ldots, 0, 1, 0 \ldots, 0)$. Then

$$H_c^{ij}(\boldsymbol{\epsilon}) = \lim_{d\epsilon \to 0} H^{ij}(\boldsymbol{\epsilon} + d\epsilon\, \boldsymbol{e}_i) = \lim_{d\epsilon \to 0} H^{ij}(\boldsymbol{\epsilon} + d\epsilon\, \boldsymbol{e}_j). \tag{15}$$

Moreover, H_c^{ij} is a continuous function in parameter ϵ_i, ϵ_j, respectively, for any fixed parameters ϵ_k such that $k \neq i, k \neq j$, respectively.[4]

Proof. See Appendix C.

The next corollary will show that the quantum Fisher information matrix is not in general a continuous function even for density matrices that are analytical functions of its parameters.

Corollary 2. If $\sum_{p_k(\epsilon)=0} \partial_{ij} p_k(\epsilon) \neq 0$, then the element of the quantum Fisher information matrix H^{ij} is not continuous at point ϵ.

Proof. Combining Theorem 1 and Theorem 2, we find

$$\lim_{d\epsilon \to 0} H^{ij}(\boldsymbol{\epsilon} + d\epsilon\, \boldsymbol{e}_i) - H^{ij}(\boldsymbol{\epsilon}) = 2 \sum_{p_k(\boldsymbol{\epsilon})=0} \partial_{ij} p_k(\boldsymbol{\epsilon}) \neq 0, \tag{16}$$

which by definition means that H^{ij} is not continuous in ϵ_i at point $\boldsymbol{\epsilon}$ and thus neither is it a continuous function at point $\boldsymbol{\epsilon}$. ∎

Theorem 2 says when a single parameter ϵ is estimated, H_c is a continuous function in this parameter. For that reason we call H_c *the continuous quantum Fisher information matrix*. It is important to point out that the assumption required in all of our theorems, $\hat{\rho}_\epsilon \in C^{(2)}$, is crucial for Theorem 2 to hold (see Fig. 2). Also, similarly to the quantum Fisher information matrix, the continuous quantum Fisher information matrix is not in general continuous in the topology of multiple parameters $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)$. This is precisely stated in the following theorem.

Theorem 3. If there exists a unit vector $\boldsymbol{u} = (u_1, \ldots, u_n)$ such that

$$\Delta_{\boldsymbol{u}}^{ij}(\boldsymbol{\epsilon}) := 2 \sum_{\substack{p_k = 0, \\ \sum_{s,t} \partial_{st} p_k u_s u_t > 0}} \left(\frac{\left(\sum_m \partial_{im} p_k u_m\right)\left(\sum_n \partial_{jn} p_k u_n\right)}{\sum_{s,t} \partial_{st} p_k u_s u_t} - \partial_{ij} p_k \right) \neq 0, \tag{17}$$

---

[4]In the proof of this continuity property we also assume that the number of eigenvalues is finite, which leads to $\sum_{p_k=0} O(d\epsilon) = O(d\epsilon)$. This assumption might be problematic, for example, when estimating parameters encoded in Gaussian quantum states because such states live in an infinite-dimensional Hilbert space. However, we believe that this assumption might not be necessary, and it should

be possible to derive $\sum_{p_k=0} O(d\epsilon) = O(d\epsilon)$ by showing that the sum converges sufficiently fast. This would nevertheless come from very careful considerations, and it could require expanding relevant quantities up to the third order in $d\epsilon$.

then the element of the continuous quantum Fisher information matrix H_c^{ij} is not continuous at point ϵ. We can also write Eq. (17) in an elegant matrix form, $\Delta_{\boldsymbol{u}}^{ij}(\boldsymbol{\epsilon}) := 2\sum_{p_k=0,\boldsymbol{u}^T\mathcal{H}_k\boldsymbol{u}>0}(\frac{(\mathcal{H}_k\boldsymbol{u})_i(\mathcal{H}_k\boldsymbol{u})_j}{\boldsymbol{u}^T\mathcal{H}_k\boldsymbol{u}} - \mathcal{H}_k^{ij}) \neq 0$.

We denote a unit vector with number 1 at the l'th position as $\boldsymbol{e}_l = (0,\ldots,0,1,0\ldots,0)$. H_c^{ij} is continuous in ϵ_l at point $\boldsymbol{\epsilon}$ if and only if $\Delta_{\boldsymbol{e}_l}^{ij}(\boldsymbol{\epsilon}) = 0$.

$\Delta_{\boldsymbol{u}}^{ij}$ measures the jump of function H_c^{ij} at point $\boldsymbol{\epsilon}$ when coming from direction $\boldsymbol{u}$,

$$\Delta_{\boldsymbol{u}}^{ij}(\boldsymbol{\epsilon}) = \lim_{d\epsilon\to 0} H_c^{ij}(\boldsymbol{\epsilon} + d\epsilon\,\boldsymbol{u}) - H_c^{ij}(\boldsymbol{\epsilon}). \qquad (18)$$

Proof. See Appendix D.

We illustrate the discontinuous behavior of the quantum Fisher information matrix on the following example.

Example 2. Consider a quantum state depending on two parameters,

$$\hat{\rho}_{\boldsymbol{\epsilon}} = \tfrac{1}{2}(\sin^2\epsilon_1 + \sin^2\epsilon_2)|0\rangle\langle 0| + \tfrac{1}{2}\cos^2\epsilon_1|1\rangle\langle 1|$$
$$+ \tfrac{1}{2}\cos^2\epsilon_2|2\rangle\langle 2|. \qquad (19)$$

We are going to study the first element of the quantum Fisher information matrix H^{11} which measures the mean squared error in the estimating parameter ϵ_1. While the expression for the quantum Fisher information matrix (4) assigns value $H^{11}(0,0) = 0$ to the problematic point $\boldsymbol{\epsilon} = (0,0)$, the continuous quantum Fisher information matrix assigns value $H_c^{11}(0,0) = 2$. According to Theorem 2, this definition of the problematic point makes the function H_c^{11} a continuous function in ϵ_1 but not necessarily in ϵ_2.

Using Theorem 3, we are going to prove that H_c is not continuous in ϵ_2 at point $\boldsymbol{\epsilon} = (0,0)$. To do that we will study $\Delta_{\boldsymbol{u}}^{11}(0,0)$ from Eq. (17). The only relevant eigenvalue is the first one because $p_1(0,0) = 0$, while for others $p_2(0,0) = p_3(0,0) = \frac{1}{2}$. The respective Hessian matrix is

$$\mathcal{H}_1(0,0) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad (20)$$

that gives

$$\Delta_{\boldsymbol{u}}^{11}(0,0) = 2\left(\frac{u_1^2}{u_1^2 + u_2^2} - 1\right). \qquad (21)$$

Setting $\boldsymbol{u} = \boldsymbol{e}_2$ we have $\Delta_{\boldsymbol{e}_2}^{11}(0,0) = -2$, which according to Theorem 3 means that H_c^{11} is not continuous in ϵ_2 at point $(0,0)$ and thus neither is it a continuous function of $\boldsymbol{\epsilon}$. Graphs of H^{11} and H_c^{11} are shown in Fig. 3.

Theorem 3 also states that it is not generally possible to use the multiparameter Taylor's expansion of the quantum Fisher information matrix at problematic points $\boldsymbol{\epsilon}$ at which $p_k(\boldsymbol{\epsilon}) = 0$ and $\mathcal{H}_k(\boldsymbol{\epsilon}) > 0$. These are often exactly the points around which we would like to do this expansion, for example, when considering a state with a slight impurity that is almost pure. For example, consider a task of estimating both phase and temperature of a quantum state, $\boldsymbol{\epsilon} = (\theta,T)$. Theorem 3 says it is not possible in general to approximate the quantum Fisher information matrix by expanding this function in both the phase parameter θ and the small temperature parameter T at the same time. The directional Taylor's expansion, for example, in parameter $\frac{\theta}{T}$, has to be employed instead. An
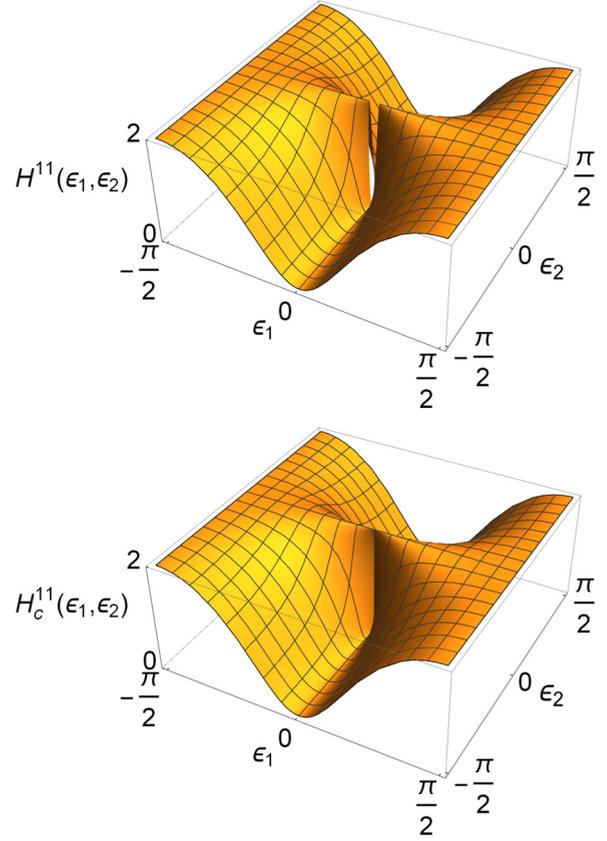
FIG. 3. Graphs of the first element of the quantum Fisher information matrix H^{11} and the first element of the continuous quantum Fisher information matrix H_c^{11}, for the estimation of parameters of the density matrix (19). These graphs are identical everywhere apart from point $(\epsilon_1,\epsilon_2) = (0,0)$. Clearly, neither function is a continuous function in both parameters at the same time; however, H_c^{11} is guaranteed to be a continuous function in ϵ_1 for any ϵ_2.

example that utilizes a variation of this method in a quantum field theory in curved space-time can be found in Ref. [33]. We will return to this in a moment.

We mentioned in the Introduction that it is often necessary to use different expressions when calculating the quantum Fisher information matrix for density matrices of different ranks. However, we will design a method which will require only one expression to obtain every other expression by performing a certain limit. We are going to show that the quantum Fisher information matrix of any state can be calculated as a limit of the quantum Fisher information matrix of a full-rank state. We call this process *the regularization procedure*, in analogy with the result for Gaussian states [21].

Theorem 4. We define a density matrix

$$\hat{\rho}_{\boldsymbol{\epsilon},\nu} := (1-\nu)\hat{\rho}_{\boldsymbol{\epsilon}} + \nu\hat{\rho}_0, \qquad (22)$$

where $0 < \nu < 1$ is a real parameter and $\hat{\rho}_0$ is any $\boldsymbol{\epsilon}$-independent full-rank density matrix that is diagonal in the eigenbasis of the density matrix $\hat{\rho}_{\boldsymbol{\epsilon}}$. Then the resulting matrix $\hat{\rho}_{\boldsymbol{\epsilon},\nu}$ is a full-rank matrix and

$$H(\boldsymbol{\epsilon}) = \lim_{\nu\to 0} H(\hat{\rho}_{\boldsymbol{\epsilon},\nu}), \qquad (23)$$

$$H_c(\epsilon) = \lim_{\nu \to 0} H(\hat{\rho}_{\epsilon,\nu}) + 2 \sum_{p_k(\epsilon)=0} \mathcal{H}_k(\epsilon). \tag{24}$$

In finite-dimensional Hilbert spaces $\hat{\rho}_0$ can be defined as a multiple of identity, $\hat{\rho}_0 = \frac{1}{\dim \mathcal{H}} \hat{I}$.

Proof. $\hat{\rho}_{\epsilon,\nu}$ has eigenvalues equal to $(1-\nu)p_k + \nu p_{0k} > 0$, where $p_{0k} > 0$ are eigenvalues of $\hat{\rho}_0$. Therefore the density matrix is full rank, and the sum in Eq. (4) for evaluating $H(\hat{\rho}_{\epsilon,\nu})$ goes over all terms. We evaluate this sum while inserting $\partial_i \hat{\rho}_{\epsilon,\nu} = (1-\nu)\partial_i \hat{\rho}_\epsilon$ and perform the limit $\nu \to 0$. Limits of terms for which $p_k + p_l = 0$ are zero, and limits of terms for which $p_k + p_l > 0$ are identical to terms in the sum for $H(\hat{\rho}_\epsilon) \equiv H(\epsilon)$. Therefore, performing such a limit gives exactly the quantum Fisher information matrix $H(\epsilon)$. The rest of the statement follows directly from Theorem 1. ∎

Theorem 4 shows that the quantum Fisher information matrix of a pure state can be calculated as a limit of the quantum Fisher information matrix of mixed states, while obtaining all of its possible discontinuities. These discontinuities reflect the results of Corollary 2 and results of Theorem 3, with ν acting as an additional parameter with the only difference that now ν is not the parameter we estimate. It is, for example, possible to use Eq. (18) to study directional limits in the extended unit vector $\tilde{u} = (u_1, \ldots, u_n, u_\nu)$, where u_ν denotes the amount of direction in the mixedness parameter ν. Such expressions can serve as a good substitute for a possibly nonexistent multiparameter Taylor's expansion: defining $\tilde{\epsilon} = \tilde{\epsilon}_0 + d\epsilon\, \tilde{u}$, where $\tilde{\epsilon}_0 = (\epsilon_1^{(0)}, \ldots, \epsilon_n^{(0)}, \nu^{(0)})$ is the point around which we expand, according to Eq. (18) the quantum Fisher information matrix can be approximated to the zeroth order by

$$H_c^{ij}(\tilde{\epsilon}) \approx H_c^{ij}(\tilde{\epsilon}_0) + \Delta_{\tilde{u}}^{ij}(\epsilon_0), \tag{25}$$

where $i, j = 1, \ldots, n$. Of course, other parameters can be used instead of ν, such as the previously mentioned temperature parameter T using the extended unit vector $\tilde{u} = (u_1, \ldots, u_n, u_T)$. The above equation is then what we could call the zeroth order of the directional Taylor's expansion. In case the function H_c is discontinuous at point ϵ_0, the value of $\Delta_{\tilde{u}}^{ij}(\epsilon_0)$ is nonzero and this value measures the amount of jump in function H_c in the direction u from the point ϵ_0.

Both the regularization procedure and the directional Taylor's expansion can be demonstrated on the following example. We use the state from the first example, Eq. (10), and study the quantum Fisher information of the respective regularized state $\hat{\rho}_{\epsilon,\nu}$.

Example 3. Consider a density matrix dependent on parameter ϵ and the mixedness parameter ν,

$$\hat{\rho}_{\epsilon,\nu} = (1-\nu)(\sin^2\epsilon\, |0\rangle\langle 0| + \cos^2\epsilon\, |1\rangle\langle 1|) + \frac{\nu}{2}\hat{I}. \tag{26}$$

The quantum Fisher information for estimating ϵ from this state is

$$H(\hat{\rho}_{\epsilon,\nu}) \equiv H^{11}(\epsilon,\nu) = \frac{4(1-\nu)^2 \sin^2(2\epsilon)}{1 - (1-\nu)^2 \cos^2(2\epsilon)}. \tag{27}$$

(The notation was taken from Theorem 4, where ϵ is the only estimated parameter while ν is auxiliary at this point.) This function is depicted in Fig. 4. It is undefined at points $(\epsilon,\nu) = (k\frac{\pi}{2}, 0)$, $k \in \mathbb{Z}$; however, as Theorem 4 shows, for $\nu \to 0$ this
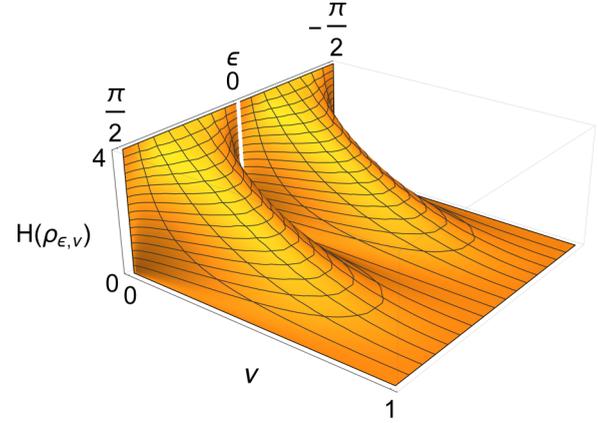
FIG. 4. The quantum Fisher information of the regularized state, Eq. (27). Clearly, as ν approaches zero, the function approaches the (discontinuous) quantum Fisher information (12), as shown in Fig. 1.

function must converge to the quantum Fisher information given by Eq. (12).

Now we illustrate how the approximation (25) performs compared to the exact value (27). We will derive the quantum Fisher information at point $(\epsilon,\nu) = (0.1, 0.04)$. Inserting this point into Eq. (27) yields the exact value of the quantum Fisher information for the estimation of ϵ,

$$H(\epsilon) \equiv H^{11}(0.1, 0.04) = 1.27. \tag{28}$$

The extended unit vector for the approximation is defined by $\tilde{\epsilon} \equiv (\epsilon,\nu) = (0.1, 0.04) = (0,0) + d\epsilon\, \tilde{u}$, which gives $\tilde{u} = \frac{(0.1, 0.04) - (0,0)}{||(0.1, 0.04) - (0,0)||}$. All eigenvalues of $\hat{\rho}_{\epsilon,\nu}$ are nonzero at $(\epsilon,\nu) = (0.1, 0.04)$, which according to Corollary 1 means that the quantum Fisher information and the continuous quantum Fisher information are identical at this point. This in combination with Eq. (25) yields

$$H(\epsilon) = H_c(\epsilon) = H_c^{11}(\tilde{\epsilon}) \approx H_c^{11}(0,0) + \Delta_{\tilde{u}}^{11}(0,0). \tag{29}$$

Clearly $H_c^{11}(0,0) = 4$. The only relevant eigenvalue needed for calculating $\Delta_{\tilde{u}}^{11}(0,0)$ is $p_1(\epsilon,\nu) = (1-\nu)\sin^2\epsilon + \frac{\nu}{2}$, for which $p_1(0,0) = 0$. The respective Hessian matrix is

$$\mathcal{H}_1(0,0) = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}, \tag{30}$$

from which we calculate $\Delta_{\tilde{u}}^{11}(0,0) = 0$. This results in $H(\epsilon) \approx 4$, which is obviously not a good approximation.

To fix this problem we employ a simple trick: we substitute $\nu_1 := \sqrt{\nu}$, which reparametrizes Eq. (27) and effectively changes its graph, and then we perform the identical procedure at point $(\epsilon,\nu_1) = (\epsilon,\sqrt{\nu}) = (0.1, 0.2)$. The relevant Hessian matrix is now

$$\mathcal{H}_1(0,0) = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}, \tag{31}$$

resulting in $\Delta_{\tilde{u}}^{11}(0,0) = -\frac{8}{3}$. Finally, we have

$$H(\epsilon) \approx H_c^{11}(0,0) + \Delta_{\tilde{u}}^{11}(0,0) = 4 - \frac{8}{3} = 1.33, \tag{32}$$

which approximates the exact value very well.

IV. DISCUSSION AND CONCLUSION

We presented a theory that describes discontinuities of the quantum Fisher information matrix and we linked these discontinuities to the discrepancy between two figures of merit in quantum metrology, the quantum Fisher information matrix and the (four times) Bures metric. Although we have shown that the Bures metric represents in some sense a continuous version of the quantum Fisher information matrix, both the quantum Fisher information matrix and the Bures metric can be discontinuous in the topology of multiple parameters. These discontinuities and discrepancies appear at sets of measure zero and therefore can be often ignored. They also appear only when varying estimation parameters changes the rank of the density matrix describing a quantum state. Specifically, these problems never appear in the estimation of unitary channels using quantum probe states.

However, in certain scenarios these problems show up and can be a source of great confusion, as it it is not usual to see discontinuous functions in physics. These scenarios involve common tasks such as estimation of decoherence parameters, space-time parameters, temperature, or simultaneous estimation of multiple parameters. Moreover, it is often assumed that common tools such as the Taylor's expansion can be always employed, especially when assumptions are difficult to check, for instance, when using a perturbative quantum field theory. But the Taylor's expansion of the quantum Fisher information cannot exist at points of discontinuity, and such points can appear even when the density matrix is an analytical function of its parameters.

It is not very clear how to interpret these discontinuities from a physical point of view. Expressions for the quantum Fisher information show that when the rank of the density matrix changes, there is a sudden drop in the precision with what we can estimate the parameter of interest. This can be connected to the fact that in such scenarios ϵ is not identifiable, i.e., ϵ and $-\epsilon$ produce exactly the same statistics of measurement results and therefore cannot be distinguished (remember example 1 in which $\hat{\rho}_\epsilon = \hat{\rho}_{-\epsilon}$). However, it is possible to design scenarios where ϵ can be identified by different means, for example, by simultaneously observing the change in phase, but the drop in the quantum Fisher information still does not disappear. It is clear, however, that the sudden drop is always connected to the information that can be extracted from the change of purity, since it always depends on the derivatives of eigenvalues of the density matrix. This drop might also be a demonstration of a quantum phase transition which occurs at absolute zero. When temperature of a quantum probe goes to zero, the thermal state describing this probe suddenly becomes a pure state, resulting in a change of the rank of the density matrix, and consequently resulting in a discontinuity of the quantum Fisher information for the estimation of parameter of interest. A similar type of behavior has been reported in several papers [6,8,44,45]. The physical meaning of these discontinuities, also in scenarios of bosonic systems, is discussed in more detail in Ref. [46].

The quantum Cramér-Rao bound holds for the possibly discontinuous quantum Fisher information matrix. The last open question is whether such a bound can be derived for the Bures metric and under what circumstances is this possible. We leave answering these questions for future work.

APPENDIX A: INFINITESIMAL BURES DISTANCE FOR SINGULAR DENSITY MATRICES

In this Appendix we are going to show that the logic used in paper [19] to derive the infinitesimal Bures distance $d_B^2(\hat{\rho}, \hat{\rho} + d\hat{\rho})$ leads to a different result than the one stated in that paper. At the same time we argue that other papers based on this result are not usually affected in large part. This is because an expression similar to the one published in [19] can be obtained by considering $d_B^2(\hat{\rho}_\epsilon, \hat{\rho}_{\epsilon+d\epsilon})$ from Eq. (8) instead of $d_B^2(\hat{\rho}, \hat{\rho} + d\hat{\rho})$. In other words, the right (expected) result comes from a different starting point. Moreover, when $\hat{\rho}$ is nonsingular these two approaches lead to the same results.

To derive the correct expression for $d_B^2(\hat{\rho}, \hat{\rho} + d\hat{\rho})$, we are going to redo the proof from [19] in detail using the same argumentation. By the same argumentation we mean that we will inherently assume that $d\hat{\rho}$ is the first-order correction to $\hat{\rho}$. This assumption also means that the starting point $d_B^2(\hat{\rho}, \hat{\rho} + d\hat{\rho})$ cannot be used for the derivation of the Bures metric g^{ij}, because the proof we are going to show will be applicable only for $d\hat{\rho}$ linearly proportional to the estimation parameters, $d\hat{\rho} \sim d\epsilon$, and *will not* be applicable in case of $d\hat{\rho} \approx d\epsilon + d\epsilon^2$. We will also use a notation similar to [19] so the reader can easily see the differences.

We are going to compute the Bures distance $d_B^2(\hat{\rho}_A, \hat{\rho}_B)$ between two infinitesimally close density matrices of size N. Let us set $\hat{\rho}_A = \hat{\rho}$, $\hat{\rho}_B = \hat{\rho} + d\hat{\rho}$:

$$\sqrt{\hat{\rho}_A^{1/2} \hat{\rho}_B \hat{\rho}_A^{1/2}} = \hat{\rho} + \hat{X} + \hat{Y}, \tag{A1}$$

where the operator $\hat{X}$ is of order 1 in $d\hat{\rho}$, while $\hat{Y}$ is of order 2. Squaring this equation we obtain the first and the second order

$$\hat{\rho}^{1/2} d\hat{\rho} \hat{\rho}^{1/2} = \hat{X}\hat{\rho} + \hat{\rho}\hat{X}, \quad -\hat{X}^2 = \hat{Y}\hat{\rho} + \hat{\rho}\hat{Y}, \tag{A2}$$

or in the basis $\{|k\rangle\}$ in which $\hat{\rho}$ is diagonal with eigenvalues p_k, $\hat{\rho} = \sum_k p_k |k\rangle\langle k|$, for $p_k + p_l > 0$ we obtain

$$\langle k|\hat{X}|l\rangle = \langle k|d\hat{\rho}|l\rangle \frac{p_k^{1/2} p_l^{1/2}}{p_k + p_l}, \quad \langle k|\hat{Y}|l\rangle = -\langle k|\hat{X}^2|l\rangle \frac{1}{p_k + p_l}. \tag{A3}$$

We point out that the above equation also gives $\langle k|\hat{X}|l\rangle = 0$ and $\langle k|\hat{Y}|l\rangle = 0$ for either $p_k > 0, p_l = 0$ or $p_k = 0, p_l > 0$. (*Now we start to differ from the proof in [19].*) Because the left-hand side of Eq. (A1) belongs to the subspace $\mathcal{L}(\mathcal{H}_{>0})$, where $\mathcal{H}_{>0}$ is the Hilbert space spanned by the eigenvectors

associated with nonzero eigenvalue p_k, also the right-hand side belongs to the same subspace. Hence for $p_k p_l = 0$ we have

$$\langle k|\hat{X}|l\rangle = 0, \quad \langle k|\hat{Y}|l\rangle = 0. \quad (A4)$$

Combining Eqs. (A3) and (A4) we obtain

$$\hat{X} = \sum_{p_k>0,p_l>0} \langle k|d\hat{\rho}|l\rangle \frac{p_k^{1/2} p_l^{1/2}}{p_k + p_l} |k\rangle\langle l|. \quad (A5)$$

Applying the trace on this operator we find

$$\text{tr}\hat{X} = \tfrac{1}{2}\text{tr}\big[\hat{P}_{\mathcal{H}_{>0}}d\hat{\rho}\big] = \tfrac{1}{2}\text{tr}[(\hat{I} - \hat{P}_0)d\hat{\rho}] = -\tfrac{1}{2}\text{tr}[\hat{P}_0 d\hat{\rho}], \quad (A6)$$

where $\hat{P}_{\mathcal{H}_{>0}} = \hat{I} - \hat{P}_0$ is the projector onto the previously mentioned subspace $\mathcal{H}_{>0}$ and $\hat{P}_0$ is the projector onto the kernel of the density matrix $\hat{\rho}$. The last inequality is due to the fact that $\text{tr}\hat{\rho} = 1$, and hence $\text{tr}[d\hat{\rho}] = 0$. Using the spectral decomposition of the density matrix $\hat{\rho} = \sum_k p_k|k\rangle\langle k|$ we have

$$\text{tr}[\hat{P}_0 d\hat{\rho}] = \text{tr}\left[\hat{P}_0\left(\sum_k dp_k|k\rangle\langle k| + p_k|dk\rangle\langle k|\right.\right.$$
$$\left.\left. + p_k|k\rangle\langle dk|\right)\right] = 0. \quad (A7)$$

The above expression vanishes because in every term

(1) either $p_k = 0$ and then also $dp_k = 0$ because p_k achieves a local minimum,

(2) or $p_k > 0$ and then by definition $\hat{P}_0|k\rangle = 0$.

Combining Eqs. (A6) and (A7) we have

$$\text{tr}\hat{X} = 0. \quad (A8)$$

Using Eqs. (A3), (A4), and (A5) we obtain

$$\text{tr}\hat{Y} = -\sum_{p_k>0} \frac{\langle k|\hat{X}^2|k\rangle}{2p_k} = -\sum_{p_k>0,p_l>0} \frac{1}{4}\frac{|\langle k|d\hat{\rho}|l\rangle|^2}{p_k + p_l}. \quad (A9)$$

Using definition (7) we find $d_B^2(\hat{\rho}, \hat{\rho} + d\hat{\rho}) = -2(\text{tr}\hat{X} + \text{tr}\hat{Y})$, which gives the result for the infinitesimal Bures distance,

$$(ds)_B^2 = d_B^2(\hat{\rho}, \hat{\rho} + d\hat{\rho}) = \frac{1}{2}\sum_{p_k>0,p_l>0} \frac{|\langle k|d\hat{\rho}|l\rangle|^2}{p_k + p_l}. \quad (A10)$$

Note that in contrast to the results of Sommers and Zyczkowski [19], who arrived at $(ds)_B^2 = d_B^2(\hat{\rho}, \hat{\rho} + d\hat{\rho}) = \frac{1}{2}\sum_{p_k+p_l>0} \frac{|\langle k|d\hat{\rho}|l\rangle|^2}{p_k+p_l}$, by applying the same argumentation in detail we arrived at the result where mixed terms for which either $p_k > 0, p_l = 0$ or $p_k = 0, p_l > 0$ are missing. Such a result does not seem to give the result for the metric tensor g^{ij} that we would expect to obtain: we would expect something similar or identical to the quantum Fisher information matrix (4). In fact, plainly inserting $d\hat{\rho} = \sum_i \partial_i \hat{\rho} d\epsilon_i$ into Eq. (A10) yields

$$g^{ij}(\epsilon) = \frac{1}{2}\sum_{p_k>0,p_l>0} \frac{\text{Re}(\langle k|\partial_i\hat{\rho}_\epsilon|l\rangle\langle l|\partial_j\hat{\rho}_\epsilon|k\rangle)}{p_k + p_l}, \quad (A11)$$

which is not compatible with the quantum Fisher information matrix (4) as long as $\hat{\rho}$ is singular. However, we are getting this unexpected result because we misused the inherent assumption that $d\hat{\rho} \approx d\epsilon$ is the first-order correction, while the defining relation for the Bures metric (8) is defined by the second-order terms in $d\epsilon$. A simple fix by expanding $d\hat{\rho}$ into the second order, $d\hat{\rho} = \sum_i \partial_i\hat{\rho} d\epsilon_i + \frac{1}{2}\sum_{ij} \partial_{ij}\hat{\rho} d\epsilon_i d\epsilon_j$, also does not work and leads to the same result (A11). This result for g^{ij} is, however, incorrect. This is simply because the proof in which we derived an expression for $d_B^2(\hat{\rho}, \hat{\rho} + d\hat{\rho})$ is not applicable for deriving the expression for the Bures metric g^{ij}. This is because if we wanted to expand $\hat{\rho}_B$ up to the second order in $d\epsilon$ there would be extra terms in the second expression in Eq. (A2). These extra terms can be found in an equivalent of this equation, in Eq. (B10). Considering these extra terms then leads to the correct expression for the Bures metric as given by Theorem 1,

$$g^{ij}(\epsilon) = \frac{1}{4}H^{ij}(\epsilon) + \frac{1}{2}\sum_{p_k(\epsilon)=0} \partial_{ij}p_k(\epsilon)$$

$$= \frac{1}{2}\sum_{p_k+p_l>0} \frac{\text{Re}(\langle k|\partial_i\hat{\rho}_\epsilon|l\rangle\langle l|\partial_j\hat{\rho}_\epsilon|k\rangle)}{p_k + p_l}$$

$$+ \frac{1}{2}\sum_{p_k(\epsilon)=0} \partial_{ij}p_k(\epsilon). \quad (A12)$$

APPENDIX B: PROOF OF THEOREM 1

Proof. We generalize the proof from Ref. [18] to include singular density matrices. By combining the defining relation for the Bures metric (8), the definition of the Bures distance (7), and the definition of the Uhlmann fidelity (6) we obtain an expression for the Bures metric,

$$\sum_{i,j} g^{ij}(\epsilon)d\epsilon_i d\epsilon_j = 2(1 - \text{tr}[\sqrt{\sqrt{\hat{\rho}_\epsilon}\hat{\rho}_{\epsilon+d\epsilon}\sqrt{\hat{\rho}_\epsilon}}]). \quad (B1)$$

We define the operator $\hat{O}(d\epsilon) := \sqrt{\hat{\rho}_\epsilon}\hat{\rho}_{\epsilon+d\epsilon}\sqrt{\hat{\rho}_\epsilon}$. Because operator $\hat{O}(d\epsilon)$ is given by applying $\sqrt{\hat{\rho}_\epsilon} = \sum_k \sqrt{p_k}|k\rangle\langle k|$ on both sides of $\hat{\rho}_{\epsilon+d\epsilon}$, it can be written as

$$\hat{O}(\epsilon) = \sum_{p_k>0,\,p_l>0} o_{kl}(d\epsilon)|k\rangle\langle l|. \quad (B2)$$

(We omitted writing the explicit dependence on ϵ.) As a result, this operator clearly belongs to the subspace of linear operators acting on the Hilbert space spanned by eigenvectors associated with nonzero eigenvalues p_k, i.e., $\hat{O} \in \mathcal{L}(\mathcal{H}_{>0})$, where $\mathcal{H}_{>0} :=$ span$\{|k\rangle\}_{p_k>0}$. Now we define its square root $\hat{A}(d\epsilon)$,

$$\hat{A}(d\epsilon)\hat{A}(d\epsilon) = \hat{O}(d\epsilon). \quad (B3)$$

Because $\hat{O} \in \mathcal{L}(\mathcal{H}_{>0})$, also $\hat{A} \in \mathcal{L}(\mathcal{H}_{>0})$ together with all of its derivatives. To show that, we assume that $\hat{O}$ has a spectral decomposition[5] $\hat{O}(d\epsilon) = \sum_m o_m^{\text{diag}}(d\epsilon)P_m(d\epsilon)$, where $P_m(d\epsilon) = \sum_{p_k>0,\,p_l>0} c_{kl}^{(m)}(d\epsilon)|k\rangle\langle l|$. Such an expression is

---

[5]Spectral decomposition exists, because $\hat{O}$ is a Hermitian operator.

valid because operator $\hat{O}$ lies in the previously mentioned subspace $\mathcal{L}(\mathcal{H}_{>0})$. The square root is then given by

$$\hat{A}(d\epsilon) = \sum_m \sqrt{o_m^{\text{diag}}(d\epsilon)} P_m(d\epsilon)$$
$$= \sum_{p_k>0,\, p_l>0} \left(\sum_m \sqrt{o_m^{\text{diag}}(d\epsilon)} c_{kl}^{(m)}(d\epsilon) \right) |k\rangle\langle l|. \quad \text{(B4)}$$

Clearly, any derivatives of $\hat{A}(d\epsilon)$ with respect to $d\epsilon_i$ will change only the factors, so the resulting operator will still remain in the same subspace $\mathcal{L}(\mathcal{H}_{>0})$.

From Eq. (B1) we obtain

$$\sum_{i,j} g^{ij}(\epsilon) d\epsilon_i d\epsilon_j = 2(1 - \text{tr}[A(d\epsilon)]), \quad \text{(B5)}$$

which gives an expression for elements of the Bures metric,

$$g^{ij}(\epsilon) = -\text{tr}\big[\partial_{d\epsilon_i d\epsilon_j} A(0)\big], \quad \text{(B6)}$$

if the second derivatives exist. For that reason we assume that $\hat{\rho} \in C^{(2)}$, i.e., the second derivatives of $\hat{\rho}$ exist and are continuous.[6] To obtain these second partial derivatives we rewrite Eq. (B3) while expanding $\hat{\rho}_{\epsilon+d\epsilon}$ around point ϵ,

$$\hat{A}(d\epsilon)\hat{A}(d\epsilon)$$
$$= \sqrt{\hat{\rho}}\left(\hat{\rho} + \sum_k \partial_k \hat{\rho}\, d\epsilon_k + \frac{1}{2}\sum_{k,l} \partial_{kl}\hat{\rho}\, d\epsilon_k d\epsilon_l \right)\sqrt{\hat{\rho}}. \quad \text{(B7)}$$

By differentiating this equation with respect to $d\epsilon_i$ and setting $d\epsilon = 0$ we obtain

$$\partial_{d\epsilon_i}\hat{A}(0)\,\hat{\rho} + \hat{\rho}\,\partial_{d\epsilon_i}\hat{A}(0) = \sqrt{\hat{\rho}}\partial_i\hat{\rho}\sqrt{\hat{\rho}}, \quad \text{(B8)}$$

where we used $\hat{A}(0) = \hat{\rho}$. By applying $\langle k| \; |l\rangle$ for $p_k > 0$ and $p_l > 0$ we obtain the matrix elements of $\partial_{d\epsilon_i}\hat{A}(0)$,

$$\langle k|\partial_{d\epsilon_i}\hat{A}(0)|l\rangle = \frac{\sqrt{p_k p_l}\langle k|\partial_i\hat{\rho}|l\rangle}{p_k + p_l}. \quad \text{(B9)}$$

Elements $\langle k|\partial_{d\epsilon_i}\hat{A}(0)|l\rangle$ such that $p_k = 0$ or $p_l = 0$ are identically zero, because as we proved earlier all derivatives of $\hat{A}$ lie in the subspace $\mathcal{L}(\mathcal{H}_{>0})$. Differentiating Eq. (B7) for the second time and setting $d\epsilon = 0$ yields[7]

$$\partial_{d\epsilon_i d\epsilon_j}\hat{A}(0)\,\hat{\rho} + \big\{\partial_{d\epsilon_i}\hat{A}(0),\partial_{d\epsilon_j}\hat{A}(0)\big\} + \hat{\rho}\,\partial_{d\epsilon_i d\epsilon_j}\hat{A}(0)$$
$$= \sqrt{\hat{\rho}}\partial_{ij}\hat{\rho}\sqrt{\hat{\rho}}, \quad \text{(B10)}$$

---

[6]Actually, this assumption can be slightly weakened. We can assume that the second derivatives exist but may not be necessarily continuous. But the continuity of the second derivatives implies $\partial_{ij}\hat{\rho} = \partial_{ji}\hat{\rho}$, which will be useful later in Theorem 2 when discussing the continuity of the Bures metric.

[7]This is where our proof starts to effectively differ from finding the expression for $d_B^2(\hat{\rho}, \hat{\rho} + d\hat{\rho})$. In both [18,19] the right-hand side of Eq. (B10) is set to zero as a result of a not explicitly stated (and easy to miss) assumption that $d\hat{\rho}$ can be linear only in $d\epsilon$, $d\hat{\rho} \approx d\epsilon$.

where $\{,\}$ denotes an anticommutator. Now, restricting ourselves to the subspace $\mathcal{L}(\mathcal{H}_{>0})$, the density matrix has the inverse matrix $\hat{\rho}^{-1}$ in this subspace. We multiply the above equation by this matrix and perform the trace on this subspace,

$$\text{tr}_{\mathcal{L}(\mathcal{H}_{>0})}\big[\rho^{-1}\big\{\partial_{d\epsilon_i}\hat{A}(0),\partial_{d\epsilon_j}\hat{A}(0)\big\}\big] + 2\text{tr}_{\mathcal{L}(\mathcal{H}_{>0})}\big[\partial_{d\epsilon_i d\epsilon_j}\hat{A}(0)\big]$$
$$= \text{tr}_{\mathcal{L}(\mathcal{H}_{>0})}[\partial_{ij}\hat{\rho}]. \quad \text{(B11)}$$

Because all derivatives of $\hat{A}$ lie in the subspace $\mathcal{L}(\mathcal{H}_{>0})$, traces of such operators are identical on both the subspace and the full space, $\text{tr}_{\mathcal{L}(\mathcal{H}_{>0})}[\partial_{d\epsilon_i d\epsilon_j}\hat{A}(0)] = \text{tr}[\partial_{d\epsilon_i d\epsilon_j}\hat{A}(0)]$. However, that is not necessarily true for the last element for which $\text{tr}_{\mathcal{L}(\mathcal{H}_{>0})}[\partial_{ij}\hat{\rho}] = \text{tr}[\hat{P}_{\mathcal{H}_{>0}}\partial_{ij}\hat{\rho}]$, where $\hat{P}_{\mathcal{H}_{>0}}$ denotes the projector on the Hilbert space $\mathcal{H}_{>0}$. Because $\text{tr}[\partial_{ij}\hat{\rho}] = 0$, this term can be equivalently written as $\text{tr}[\hat{P}_{\mathcal{H}_{>0}}\partial_{ij}\hat{\rho}] = -\text{tr}[\hat{P}_0\partial_{ij}\hat{\rho}]$, where the projector $\hat{P}_0 := \hat{I} - \hat{P}_{\mathcal{H}_{>0}}$ projects onto the subspace spanned by the eigenvectors of the density matrix $\hat{\rho}$ associated with the zero eigenvalue. Therefore combining Eqs. (B6) and (B11) yields

$$g^{ij} = \frac{1}{2}\big(\text{tr}_{\mathcal{L}(\mathcal{H}_{>0})}\big[\rho^{-1}\big\{\partial_{d\epsilon_i}\hat{A}(0),\partial_{d\epsilon_j}\hat{A}(0)\big\}\big] + \text{tr}[\hat{P}_0\partial_{ij}\hat{\rho}]\big). \quad \text{(B12)}$$

The first term of the right-hand side can be readily computed from Eq. (B9) while the antisymmetric part vanishes under the sum,

$$\text{tr}_{\mathcal{L}(\mathcal{H}_{>0})}\big[\rho^{-1}\big\{\partial_{d\epsilon_i}\hat{A}(0),\partial_{d\epsilon_j}\hat{A}(0)\big\}\big]$$
$$= \sum_{p_k>0,\, p_l>0} \frac{\text{Re}(\langle k|\partial_i\hat{\rho}|l\rangle\langle l|\partial_j\hat{\rho}|k\rangle)}{p_k + p_l}. \quad \text{(B13)}$$

Now we compute the second term. The second derivative of $\hat{\rho}$ is given by

$$\partial_{ij}\hat{\rho} = \sum_k \partial_{ij}p_k|k\rangle\langle k| + p_k(|\partial_i k\rangle\langle\partial_j k| + |\partial_j k\rangle\langle\partial_i k|)$$
$$+ p_k(|\partial_{ij}k\rangle\langle k| + |k\rangle\langle\partial_{ij}k|) + \partial_j p_k(|\partial_i k\rangle\langle k| + |k\rangle\langle\partial_i k|)$$
$$+ \partial_i p_k(|\partial_j k\rangle\langle k| + |k\rangle\langle\partial_j k|). \quad \text{(B14)}$$

We stress out that the summation goes over all values of k, even over those for which $p_k = 0$. When using the above equation to calculate $\text{tr}[\hat{P}_0\partial_{ij}\hat{\rho}]$ we find that many terms vanish because

(1) for k such that $p_k > 0$, $\hat{P}_0|k\rangle = 0$,

(2) for k such that $p_k = 0$, also $\partial_i p_k = \partial_j p_k = 0$, because p_k reaches the local minimum at point ϵ such that $p_k(\epsilon) = 0$.

Only parts of the first two terms of Eq. (B14) remain,

$$\text{tr}[\hat{P}_0\partial_{ij}\hat{\rho}] = \sum_{p_k=0} \partial_{ij}p_k + 2\sum_{p_k>0,\, p_l=0} p_k\,\text{Re}(\langle l|\partial_i k\rangle\langle\partial_j k|l\rangle)$$
$$= \sum_{p_k=0} \partial_{ij}p_k + 2\sum_{p_k>0,\, p_l=0} \frac{\text{Re}(\langle k|\partial_i\hat{\rho}|l\rangle\langle l|\partial_j\hat{\rho}|k\rangle)}{p_k + p_l}. \quad \text{(B15)}$$

Combining Eqs. (B12), (B13), and (B15), and the expression for the quantum Fisher information matrix (4)

we derive

$$g^{ij} = \frac{1}{2} \sum_{p_k + p_l > 0} \frac{\mathrm{Re}(\langle k|\partial_i \hat{\rho}|l\rangle \langle l|\partial_j \hat{\rho}|k\rangle)}{p_k + p_l} + \frac{1}{2} \sum_{p_k = 0} \partial_{ij} p_k$$

$$= \frac{1}{4} \left(H^{ij} + 2 \sum_{p_k = 0} \partial_{ij} p_k \right), \tag{B16}$$

which considering the definition of H_c, Eq. (9), proves the theorem.

APPENDIX C: PROOF OF THEOREM 2

Proof. First we will study the neighborhood of the quantum Fisher information matrix, i.e., we will study the function $H^{ij}(\epsilon + d\epsilon)$, where we define $d\epsilon = (d\epsilon_1, \dots, d\epsilon_n)$. Then we show that the (four times) Bures metric H_c is given by the limits stated in the theorem. Finally, we prove the continuity property stated in the theorem.

Equation (4) gives

$$H^{ij}(\epsilon + d\epsilon) = 2 \sum_{p_{k\,\epsilon+d\epsilon} + p_{l\,\epsilon+d\epsilon} > 0} \frac{\mathrm{Re}(\langle k_{\epsilon+d\epsilon}|\partial_i \hat{\rho}_{\epsilon+d\epsilon}|l_{\epsilon+d\epsilon}\rangle \langle l_{\epsilon+d\epsilon}|\partial_j \hat{\rho}_{\epsilon+d\epsilon}|k_{\epsilon+d\epsilon}\rangle)}{p_{k\,\epsilon+d\epsilon} + p_{l\,\epsilon+d\epsilon}}$$

$$= 2 \sum_{p_{k\,\epsilon} + p_{l\,\epsilon} > 0} \frac{\mathrm{Re}(\langle k_\epsilon|\partial_i \hat{\rho}_\epsilon|l_\epsilon\rangle \langle l_\epsilon|\partial_j \hat{\rho}_\epsilon|k_\epsilon\rangle + O(d\epsilon))}{p_{k\,\epsilon} + p_{l\,\epsilon} + O(d\epsilon)}$$

$$+ 2 \sum_{\substack{p_{k\,\epsilon} + p_{l\,\epsilon} = 0, \\ p_{k\,\epsilon+d\epsilon} + p_{l\,\epsilon+d\epsilon} > 0}} \frac{\mathrm{Re}(\langle k_{\epsilon+d\epsilon}|\partial_i \hat{\rho}_{\epsilon+d\epsilon}|l_{\epsilon+d\epsilon}\rangle \langle l_{\epsilon+d\epsilon}|\partial_j \hat{\rho}_{\epsilon+d\epsilon}|k_{\epsilon+d\epsilon}\rangle)}{p_{k\,\epsilon+d\epsilon} + p_{l\,\epsilon+d\epsilon}}. \tag{C1}$$

$O(d\epsilon)$ (the big O notation) denotes the remainder after expanding $p_{k\,\epsilon+d\epsilon} + p_{l\,\epsilon+d\epsilon}$. In the above equation we have chosen $d\epsilon$ small enough such that $|O(d\epsilon)| < p_{k\,\epsilon} + p_{l\,\epsilon}$ for $p_{k\,\epsilon} + p_{l\,\epsilon} > 0$.

Using $\hat{\rho}_\epsilon \in C^{(2)}$ (which is assumed in all theorems in this paper) we can write the following expansions:

$$p_{k\,\epsilon+d\epsilon} = p_k + \sum_m \partial_m p_k d\epsilon_m + \frac{1}{2} \sum_{m,n} \partial_{mn} p_k d\epsilon_m d\epsilon_n + O(d\epsilon^3),$$

$$\hat{\rho}_{\epsilon+d\epsilon} = \hat{\rho} + \sum_m \partial_m \hat{\rho} d\epsilon_m + \frac{1}{2} \sum_{m,n} \partial_{mn} \hat{\rho} d\epsilon_m d\epsilon_n + O(d\epsilon^3), \tag{C2}$$

$$|k_{\epsilon+d\epsilon}\rangle = |k\rangle + \sum_m |\partial_m k\rangle d\epsilon_m + \frac{1}{2} \sum_{m,n} |\partial_{mn} k\rangle d\epsilon_m d\epsilon_n + O(d\epsilon^3).$$

$O(d\epsilon^3)$ denotes the remainder that consists of sums of multiples of three or more elements of vector $d\epsilon$. Using these expansions and Eq. (B14), for k, l such that $p_{k\,\epsilon} + p_{l\,\epsilon} = 0$ we have $\partial_i p_{k\,\epsilon} + \partial_i p_{l\,\epsilon} = 0$ and

$$p_{k\,\epsilon+d\epsilon} + p_{l\,\epsilon+d\epsilon} = \frac{1}{2} \sum_{m,n} (\partial_{mn} p_k + \partial_{mn} p_l) d\epsilon_m d\epsilon_n + O(d\epsilon^3),$$

$$\langle k_{\epsilon+d\epsilon}|\partial_i \hat{\rho}_{\epsilon+d\epsilon}|l_{\epsilon+d\epsilon}\rangle = \sum_m (\langle \partial_m k|\partial_i \hat{\rho}|l\rangle + \langle k|\partial_i \hat{\rho}|\partial_m l\rangle + \langle k|\partial_{im} \hat{\rho}|l\rangle) d\epsilon_m + O(d\epsilon^2)$$

$$= \delta_{kl} \sum_m \partial_{im} p_k d\epsilon_m + O(d\epsilon^2), \tag{C3}$$

where we used $\langle k|\partial_i j\rangle = -\langle \partial_i k|j\rangle$, which comes from the orthonormality condition. Inserting Eq. (C2) for $p_k + p_l > 0$ and Eq. (C3) for $p_k + p_l = 0$ into Eq. (C1) yields

$$H^{ij}(\epsilon + d\epsilon) = 2 \sum_{p_k + p_l > 0} \frac{\mathrm{Re}(\langle k|\partial_i \hat{\rho}|l\rangle \langle l|\partial_j \hat{\rho}|k\rangle) + O(d\epsilon)}{p_k + p_l + O(d\epsilon)} + 2 \sum_{\substack{p_k = 0, \\ \sum_{s,t} \partial_{st} p_k d\epsilon_s d\epsilon_t > 0}} \frac{\left(\sum_m \partial_{im} p_k d\epsilon_m\right)\left(\sum_n \partial_{jn} p_k d\epsilon_n\right) + O(d\epsilon^3)}{\sum_{s,t} \partial_{st} p_k d\epsilon_s d\epsilon_t + O(d\epsilon^3)}. \tag{C4}$$

By setting $d\epsilon = d\epsilon e_i$ and performing the limit we find

$$\lim_{d\epsilon \to 0} H^{ij}(\epsilon + d\epsilon e_i) = H^{ij}(\epsilon) + 2 \sum_{\substack{p_k(\epsilon) = 0, \\ \partial_{ii} p_k(\epsilon) > 0}} \partial_{ij} p_k(\epsilon) = H^{ij}(\epsilon) + 2 \sum_{p_k(\epsilon) = 0} \partial_{ij} p_k(\epsilon), \tag{C5}$$

which is equal to $H_c^{ij}(\epsilon)$ according to Theorem 1. The second equality in Eq. (C5) is due to the Sylvester's criterion for positive semidefinite matrices [47], which gives $\partial_{ii} p_k(\epsilon)\partial_{jj} p_k(\epsilon) - \partial_{ij} p_k(\epsilon)^2 \geqslant 0$, i.e., for $p_k(\epsilon) = 0$ and $\partial_{ii} p_k(\epsilon) = 0$ also $\partial_{ij} p_k(\epsilon) = 0$. The same equality holds for $d\epsilon = d\epsilon e_j$, which proves the first part of the theorem.

Now we are going to prove the continuity property stated in the theorem. We set $d\epsilon = d\epsilon u$, where $u = (u_1, \ldots, u_n)$ is a unit vector. Using

$$\frac{x + O(d\epsilon)}{y + O(d\epsilon)} = \frac{x}{y\left(1 + \frac{O(d\epsilon)}{y}\right)} + \frac{O(d\epsilon)}{y + O(d\epsilon)} = \frac{x}{y}\left(1 - \frac{O(d\epsilon)}{y}\right) + \frac{O(d\epsilon)}{y + O(d\epsilon)} = \frac{x}{y} + O(d\epsilon), \tag{C6}$$

which holds for any $y \neq 0$, and Eq. (C4) while assuming the number of eigenvalues p_k is finite we derive

$$H^{ij}(\epsilon + d\epsilon u) = H^{ij}(\epsilon) + 2 \sum_{\substack{p_k = 0, \\ \sum_{s,t} \partial_{st} p_k u_s u_t > 0}} \frac{\left(\sum_m \partial_{im} p_k u_m\right)\left(\sum_n \partial_{jn} p_k u_n\right)}{\sum_{s,t} \partial_{st} p_k u_s u_t} + O(d\epsilon). \tag{C7}$$

By definition, the function H_c^{ij} is continuous in ϵ_i at point ϵ when

$$(\forall \gamma > 0)(\exists \delta > 0)(\forall d\epsilon, |d\epsilon| < \delta)\left(\left|H_c^{ij}(\epsilon + d\epsilon e_i) - H_c^{ij}(\epsilon)\right| < \gamma\right). \tag{C8}$$

Setting $u := e_i$, using Theorem 1, and Eq. (C7) we derive

$$\left|H_c^{ij}(\epsilon + d\epsilon e_i) - H_c^{ij}(\epsilon)\right| = \left|H^{ij}(\epsilon + d\epsilon e_i) - H^{ij}(\epsilon) + 2 \sum_{p_k(\epsilon + d\epsilon e_i)=0} \partial_{ij} p_k(\epsilon + d\epsilon e_i) - 2 \sum_{p_k(\epsilon)=0} \partial_{ij} p_k(\epsilon)\right|$$

$$= \left|2 \sum_{\substack{p_k(\epsilon) = 0, \\ \partial_{ii} p_k(\epsilon) > 0}} \frac{\partial_{ii} p_k(\epsilon) \partial_{ji} p_k(\epsilon) d\epsilon^2}{\partial_{ii} p_k(\epsilon) d\epsilon^2} + O(d\epsilon) - 2 \sum_{p_k(\epsilon)=0} \partial_{ij} p_k(\epsilon)\right.$$

$$\left. + 2 \sum_{p_k(\epsilon + d\epsilon e_i)=0} (\partial_{ij} p_k(\epsilon + d\epsilon e_i) - \partial_{ij} p_k(\epsilon)) + 2 \sum_{p_k(\epsilon + d\epsilon e_i)=0} \partial_{ij} p_k(\epsilon)\right|$$

$$\leqslant 2 \sum_{p_k(\epsilon + d\epsilon e_i)=0} |\partial_{ij} p_k(\epsilon + d\epsilon e_i) - \partial_{ij} p_k(\epsilon)| + 2 \sum_{p_k(\epsilon + d\epsilon e_i)=0} |\partial_{ij} p_k(\epsilon)| + |O(d\epsilon)| < \frac{\gamma}{3} + \frac{\gamma}{3} + \frac{\gamma}{3}$$

$$= \gamma. \tag{C9}$$

The first inequality is the triangle inequality. $\partial_{ij} p_k(\epsilon) = \partial_{ji} p_k(\epsilon)$ follows from $\hat{\rho}_\epsilon \in C^{(2)}$ ($p_k \in C^{(2)}$, i.e., the second derivative is continuous). The same property also implies $2 \sum_{p_k(\epsilon + d\epsilon e_i)=0} |\partial_{ij} p_k(\epsilon + d\epsilon e_i) - \partial_{ij} p_k(\epsilon)| < \frac{\gamma}{3}$ for small enough $d\epsilon$ (i.e., for all $d\epsilon$ such that $|d\epsilon| < \delta$ where $\delta > 0$ is some sufficiently small radius). $2 \sum_{p_k(\epsilon + d\epsilon e_i)=0} |\partial_{ij} p_k(\epsilon)| < \frac{\gamma}{3}$ holds for small enough $d\epsilon$ because

(1) either $p_k(\epsilon) = 0$ and then from $0 = p_k(\epsilon + d\epsilon e_i) = p_k(\epsilon) + \frac{1}{2}\partial_{ii} p_k(\epsilon)d\epsilon^2 = \frac{1}{2}\partial_{ii} p_k(\epsilon)d\epsilon^2$ follows $\partial_{ii} p_k(\epsilon) = 0$. Using the Sylvester's criterion again we have $\partial_{ij} p_k(\epsilon) = 0$, i.e., the corresponding term in the sum is zero.

(2) or $p_k(\epsilon) > 0$, and then the continuity of p_k implies that for small enough $d\epsilon$ also $p_k(\epsilon + d\epsilon e_i) > 0$, i.e., the corresponding term does not appear in the sum.

At last, $|O(d\epsilon)| < \frac{\gamma}{3}$ comes from the definition of $O(d\epsilon)$, which can be made arbitrarily small, i.e., we can choose δ such that for all $d\epsilon$, $|d\epsilon| < \delta$, $|O(d\epsilon)| < \frac{\gamma}{3}$, which proves the theorem.

APPENDIX D: PROOF OF THEOREM 3

Proof. We are going to prove Eq. (18) first. Other statements of the theorem will follow easily. To do that we generalize the second part of the proof of Theorem 2. Combining Theorem 1, Eq. (C7), and definition (17) yields

$$\left|H_c^{ij}(\epsilon + d\epsilon u) - H_c^{ij}(\epsilon) - \Delta_u^{ij}(\epsilon)\right| = \left|2 \sum_{p_k(\epsilon + d\epsilon u)=0} \partial_{ij} p_k(\epsilon + d\epsilon u) - 2 \sum_{\substack{p_k(\epsilon) = 0, \\ \sum_{s,t} \partial_{st} p_k(\epsilon) u_s u_t = 0}} \partial_{ij} p_k(\epsilon) + O(d\epsilon)\right|$$

$$\leqslant 2 \sum_{p_k(\epsilon + d\epsilon u)=0} |\partial_{ij} p_k(\epsilon + d\epsilon u) - \partial_{ij} p_k(\epsilon)|$$

$$+ 2 \left|\sum_{p_k(\epsilon + d\epsilon u)=0} \partial_{ij} p_k(\epsilon) - \sum_{\substack{p_k(\epsilon) = 0, \\ \sum_{s,t} \partial_{st} p_k(\epsilon) u_s u_t = 0}} \partial_{ij} p_k(\epsilon)\right| + |O(d\epsilon)| < \frac{\gamma}{3} + \frac{\gamma}{3} + \frac{\gamma}{3} = \gamma,$$

$$\tag{D1}$$

which proves Eq. (18). $2 \sum_{p_k(\boldsymbol{\epsilon}+d\boldsymbol{\epsilon}\boldsymbol{u})=0} |\partial_{ij} p_k(\boldsymbol{\epsilon}+d\boldsymbol{\epsilon}\boldsymbol{u}) - \partial_{ij} p_k(\boldsymbol{\epsilon})| < \frac{\gamma}{3}$ in Eq. (D1) comes from the continuity of second derivatives. $2|\sum_{p_k(\boldsymbol{\epsilon}+d\boldsymbol{\epsilon}\boldsymbol{u})=0} \partial_{ij} p_k(\boldsymbol{\epsilon}) - \sum_{\substack{p_k(\boldsymbol{\epsilon})=0, \\ \sum_{s,t} \partial_{st} p_k(\boldsymbol{\epsilon}) u_s u_t = 0}} \partial_{ij} p_k(\boldsymbol{\epsilon})| < \frac{\gamma}{3}$ because

(1) either $p_k(\boldsymbol{\epsilon}) = 0$ and then from $0 = p_k(\boldsymbol{\epsilon}+d\boldsymbol{\epsilon}\boldsymbol{u}) = p_k(\boldsymbol{\epsilon}) + \frac{1}{2}\sum_{s,t} \partial_{st} p_k(\boldsymbol{\epsilon}) u_s u_t d\epsilon^2$ follows $\sum_{s,t} \partial_{st} p_k(\boldsymbol{\epsilon}) u_s u_t = 0$. In this case the element $\partial_{ij} p_k(\boldsymbol{\epsilon})$ in the first sum $\sum_{p_k(\boldsymbol{\epsilon}+d\boldsymbol{\epsilon}\boldsymbol{u})=0} \partial_{ij} p_k(\boldsymbol{\epsilon})$ is compensated by the element $\partial_{ij} p_k(\boldsymbol{\epsilon})$ in the second sum.

(2) or $p_k(\boldsymbol{\epsilon}) > 0$ and then the continuity of p_k implies that for small enough $d\epsilon$ also $p_k(\boldsymbol{\epsilon}+d\boldsymbol{\epsilon}\boldsymbol{u}) > 0$, i.e., for small enough $d\epsilon$ term $\partial_{ij} p_k(\boldsymbol{\epsilon})$ does not appear in the first sum. The corresponding term also does not appear in the second sum because only terms for which $p_k(\boldsymbol{\epsilon}) = 0$ are counted.

In total we have

$$
2 \left| \sum_{p_k(\boldsymbol{\epsilon}+d\boldsymbol{\epsilon}\boldsymbol{u})=0} \partial_{ij} p_k(\boldsymbol{\epsilon}) - \sum_{\substack{p_k(\boldsymbol{\epsilon})=0, \\ \sum_{s,t} \partial_{st} p_k(\boldsymbol{\epsilon}) u_s u_t = 0}} \partial_{ij} p_k(\boldsymbol{\epsilon}) \right|
$$

$$
\leqslant 2 \left| \sum_{\substack{p_k(\boldsymbol{\epsilon})=0, \\ p_k(\boldsymbol{\epsilon}+d\boldsymbol{\epsilon}\boldsymbol{u})=0}} \partial_{ij} p_k(\boldsymbol{\epsilon}) - \sum_{\substack{p_k(\boldsymbol{\epsilon})=0, \\ \sum_{s,t} \partial_{st} p_k(\boldsymbol{\epsilon}) u_s u_t = 0}} \partial_{ij} p_k(\boldsymbol{\epsilon}) \right| + 2 \left| \sum_{\substack{p_k(\boldsymbol{\epsilon})>0, \\ p_k(\boldsymbol{\epsilon}+d\boldsymbol{\epsilon}\boldsymbol{u})=0}} \partial_{ij} p_k(\boldsymbol{\epsilon}) \right|
$$

$$
\leqslant 0 + \frac{\gamma}{3}. \tag{D2}
$$

$|O(d\epsilon)| < \frac{\gamma}{3}$ comes from the definition of $O(d\epsilon)$, which can be made arbitrarily small.

By definition, H_c is continuous in ϵ_l if and only if $\lim_{d\epsilon \to 0} H_c^{ij}(\boldsymbol{\epsilon}+d\epsilon \, \boldsymbol{e}_l) - H_c^{ij}(\boldsymbol{\epsilon}) = 0$, which together with Eq. (18) proves the second part of the theorem.

The function H_c is continuous at point $\boldsymbol{\epsilon}$ if and only if

$$
(\forall \gamma > 0)(\exists \delta > 0)(\forall d\boldsymbol{\epsilon}, ||d\boldsymbol{\epsilon}|| < \delta)
$$
$$
\times \left(\left| H_c^{ij}(\boldsymbol{\epsilon}+d\boldsymbol{\epsilon}) - H_c^{ij}(\boldsymbol{\epsilon}) \right| < \gamma \right). \tag{D3}
$$

By negating this statement we obtain

$$
(\exists \gamma > 0)(\forall \delta > 0)(\exists d\boldsymbol{\epsilon}, ||d\boldsymbol{\epsilon}|| < \delta)
$$
$$
\times \left(\left| H_c^{ij}(\boldsymbol{\epsilon}+d\boldsymbol{\epsilon}) - H_c^{ij}(\boldsymbol{\epsilon}) \right| \geqslant \gamma \right). \tag{D4}
$$

Equation (18) yields

$$
\left| H_c^{ij}(\boldsymbol{\epsilon}+d\epsilon \, \boldsymbol{u}) - H_c^{ij}(\boldsymbol{\epsilon}) - \Delta_{\boldsymbol{u}}^{ij}(\boldsymbol{\epsilon}) \right| < \frac{|\Delta_{\boldsymbol{u}}^{ij}(\boldsymbol{\epsilon})|}{2}, \tag{D5}
$$

for all $d\epsilon$ such that $|d\epsilon| < \delta_1$, where $\delta_1 > 0$ is some sufficiently small radius. To show the validity of Eq. (D4) we choose $\gamma := \frac{|\Delta_{\boldsymbol{u}}^{ij}(\boldsymbol{\epsilon})|}{2}$ and $d\boldsymbol{\epsilon} := d\epsilon \boldsymbol{u}$ with $d\epsilon$ any such that $|d\epsilon| < \min\{\delta, \delta_1\}$. Then

$$
\left| H_c^{ij}(\boldsymbol{\epsilon}+d\boldsymbol{\epsilon}) - H_c^{ij}(\boldsymbol{\epsilon}) \right|
$$
$$
\geqslant \left| \left| \Delta_{\boldsymbol{u}}^{ij}(\boldsymbol{\epsilon}) \right| - \left| H_c^{ij}(\boldsymbol{\epsilon}+d\epsilon \, \boldsymbol{u}) - H_c^{ij}(\boldsymbol{\epsilon}) - \Delta_{\boldsymbol{u}}^{ij}(\boldsymbol{\epsilon}) \right| \right|
$$
$$
\geqslant \left| \Delta_{\boldsymbol{u}}^{ij}(\boldsymbol{\epsilon}) \right| - \frac{|\Delta_{\boldsymbol{u}}^{ij}(\boldsymbol{\epsilon})|}{2} = \frac{|\Delta_{\boldsymbol{u}}^{ij}(\boldsymbol{\epsilon})|}{2} = \gamma, \tag{D6}
$$

where the first inequality is a version $|a - b| \geqslant ||a| - |b||$ of the triangle inequality. This proves the first part of the theorem.

[1] S. L. Braunstein and C. M. Caves, Phys. Rev. Lett. **72**, 3439 (1994).

[2] M. G. A. Paris, Int. J. Quantum Inform. **07**, 125 (2009).

[3] J. Aasi, J. Abadie, B. Abbott, R. Abbott, T. Abbott, M. Abernathy, C. Adams, T. Adams, P. Addesso, R. Adhikari *et al.*, Nat. Photon. **7**, 613 (2013).

[4] B. P. e. a. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), Phys. Rev. Lett. **116**, 061102 (2016).

[5] G.-S. Paraoanu and H. Scutaru, Phys. Rev. A **58**, 869 (1998).

[6] P. Zanardi, L. Campos Venuti, and P. Giorda, Phys. Rev. A **76**, 062318 (2007).

[7] L. Campos Venuti and P. Zanardi, Phys. Rev. Lett. **99**, 095701 (2007).

[8] S.-J. Gu, Int. J. Mod. Phys. B **24**, 4371 (2010).

[9] N. Margolus and L. B. Levitin, Phys. D **120**, 188 (1998).

[10] D. P. Pires, M. Cianciaruso, L. C. Céleri, G. Adesso, and D. O. Soares-Pinto, Phys. Rev. X **6**, 021031 (2016).

[11] S. Lloyd, Nature (London) **406**, 1047 (2000).

[12] F. C. Binder, S. Vinjanampathy, K. Modi, and J. Goold, New J. Phys. **17**, 075015 (2015).

[13] A. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (Springer, Berlin, 2011).

[14] C. Helstrom, Phys. Lett. A **25**, 101 (1967).

[15] C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic Press, New York, 1976), Vol. 123.

[16] D. J. C. Bures, Trans. Amer. Math. Soc. **135**, 199 (1969).

[17] W. K. Wootters, Phys. Rev. D **23**, 357 (1981).

[18] M. Hübner, Phys. Lett. A **163**, 239 (1992).

[19] H.-J. Sommers and K. Zyczkowski, J. Phys. A: Math. Gen. **36**, 10083 (2003).

[20] A. Monras, arXiv:1008.2849.

[21] D. Šafránek, A. R. Lee, and I. Fuentes, New J. Phys. **17**, 073016 (2015).

[22] O. Pinel, P. Jian, N. Treps, C. Fabre, and D. Braun, Phys. Rev. A **88**, 040102 (2013).

[23] A. Monras and M. G. A. Paris, Phys. Rev. Lett. **98**, 160401 (2007).

[24] C. Invernizzi, M. G. A. Paris, and S. Pirandola, Phys. Rev. A **84**, 022334 (2011).

[25] G. Spedalieri, S. L. Braunstein, and S. Pirandola, arXiv:1602.05958.

[26] J. Gambetta and H. M. Wiseman, Phys. Rev. A **64**, 042105 (2001).

[27] M. Tsang, New J. Phys. **15**, 073005 (2013).

[28] S. Alipour, M. Mehboudi, and A. T. Rezakhani, Phys. Rev. Lett. **112**, 120405 (2014).

[29] M. Aspachs, G. Adesso, and I. Fuentes, Phys. Rev. Lett. **105**, 151301 (2010).

[30] A. Dragan, I. Fuentes, and J. Louko, Phys. Rev. D **83**, 085020 (2011).

[31] J. Wang, Z. Tian, J. Jing, and H. Fan, Sci. Rep. **4**, 7195 (2014).

[32] Z. Tian, J. Wang, H. Fan, and J. Jing, Sci. Rep. **5**, 7946 (2015).

[33] D. Šafránek, J. Kohlrus, D. E. Bruschi, A. R. Lee, and I. Fuentes, arXiv:1511.03905.

[34] S. P. Kish and T. C. Ralph, Phys. Rev. D **93**, 105013 (2016).

[35] M. Szczykulska, T. Baumgratz, and A. Datta, Adv. Phys. X **1**, 621 (2016).

[36] A. Uhlmann, Rep. Math. Phys. **9**, 273 (1976).

[37] M. Hübner, Phys. Lett. A **179**, 226 (1993).

[38] R. Jozsa, J. Mod. Opt. **41**, 2315 (1994).

[39] P. B. Slater, J. Phys. A: Math. Gen. **29**, L271 (1996).

[40] P. B. Slater, Phys. Lett. A **244**, 35 (1998).

[41] J. Dittmann, J. Phys. A: Math. Gen. **32**, 2663 (1999).

[42] W. Guo-You, Y. Ji-Bing, and Z. Hao-Sheng, Commun. Theor. Phys. **64**, 495 (2015).

[43] A. Rezakhani and S. Alipour, arXiv:1507.01736.

[44] L. Banchi, P. Giorda, and P. Zanardi, Phys. Rev. E **89**, 022102 (2014).

[45] W. Wu and J.-B. Xu, Quantum Inf. Process. **15**, 3695 (2016).

[46] D. Šafránek, arXiv:1610.03538.

[47] R. A. Horn and C. R. Johnson, *Matrix Analysis* (Cambridge University Press, Cambridge, UK, 2012).