Quantum Incompatibility Witnesses

Claudio Carmeli*
DIME, Università di Genova, Via Magliotto 2, I-17100 Savona, Italy

Teiko Heinosaari[†]
QTF Centre of Excellence, Turku Centre for Quantum Physics,
Department of Physics and Astronomy, University of Turku, FI-20014 Turku, Finland

Alessandro Toigo[‡]
Dipartimento di Matematica, Politecnico di Milano,
Piazza Leonardo da Vinci 32, I-20133 Milano, Italy and
I.N.F.N., Sezione di Milano, Via Celoria 16, I-20133 Milano, Italy

We demonstrate that quantum incompatibility can always be detected by means of a state discrimination task with partial intermediate information. This is done by showing that only incompatible measurements allow for an efficient use of premeasurement information in order to improve the probability of guessing the correct state. Thus, the gap between the guessing probabilities with pre- and postmeasurement information is a witness of the incompatibility of a given collection of measurements. We prove that all linear incompatibility witnesses can be implemented as some state discrimination protocol according to this scheme. As an application, we characterize the joint measurability region of two noisy mutually unbiased bases.

Introduction. — Quantum incompatibility is one of the key features that separate the quantum from the classical world [1]. It gives rise to several among the most intriguing quantum phenomena, including measurement uncertainty relations [2], contextuality [3] and nonlocality [4]. So far, however, the direct experimental verification of quantum incompatibility has been a demanding task, as the known detection methods, based on Bell experiments [5–7] and steering protocols [8–11], rely on entanglement.

In this paper, we show that quantum incompatibility can be detected by means of a state discrimination task with partial intermediate information. More precisely, we consider a scenario where Alice sends Bob a quantum system that she has prepared into a state chosen from one of n disjoint state ensembles, but she reveals to him the chosen ensemble only at a later time. Bob can then decide to perform his measurement either before or after Alice's announcement and, importantly, the achievable success probabilities can be compared. We show that Bob can benefit from prior compared to posterior measurement information and improve his probability of guessing the correct state only if his measurements are incompatible.

Looking at it from another perspective, the difference between Bob's guessing probabilities with pre- and postmeasurement information is a witness of the incompatibility of the collection of measurements he uses in the discrimination task. Since the complement set of incompatible collections of measurements is the closed and convex set of all the compatible collections of measurements, this observation sets the previous detection scheme for incompatibility within the broader framework of witnesses.

In general, a witness is any experimentally assessable linear function whose value is greater than or equal to zero whenever the measured object does not have the investigated property, but gives a negative value at least for some object with that property. The paradigmatic example of witnesses is that of entanglement witnesses, which have become one of the main methods to detect entanglement [12, 13]. Other examples include the detection of non-Gaussianity of states [14], dimensionality of correlations [15], or for the unital channels the detection of not being a random unitary channel [16]. The fact that witnesses can be applied to detect incompatibility has been recently noted in [17, 18].

We prove that any incompatibility witness essentially arises as a state discrimination task with intermediate information of the type described above. By standard separation results for convex sets, this implies that all incompatible sets of measurements can be detected by performing some state discrimination where premeasurement information is strictly better than postmeasurement information. This yields a novel operational interpretation of quantum incompatibility, and provides a method to detect it in a physically feasible experiment. In particular, this proves that entanglement is not needed to reveal incompatibility.

General framework of witnesses. — We briefly recall the general setting of witnesses as this clarifies our main results on incompatibility witnesses and makes the reasoning behind them easy to follow.

Let $\mathcal{V}$ be a real linear space and $C \subset \mathcal{V}$ a compact convex subset that mathematically describes the objects we are interested in. This set is further divided into two disjoint subsets C_0 and $\overline{C}_0$, with C_0 being closed and convex. We can think of C_0 and $\overline{C}_0$ as properties – either an element $x \in C$ is in C_0 or in $\overline{C}_0$. A witness of the property $\overline{C}_0$, or $\overline{C}_0$-*witness*, is a map $\xi : C \to \mathbb{R}$ such that

---

* claudio.carmeli@gmail.com
[†] teiko.heinosaari@utu.fi
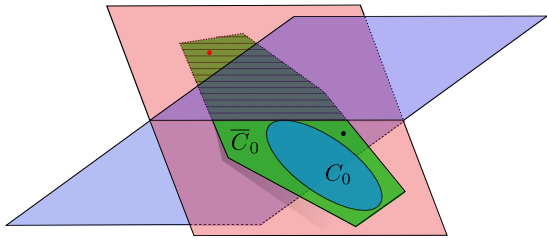[‡] alessandro.toigo@polimi.it

FIG. 1. Witnesses are associated with hyperplanes, and they are detection equivalent if they yield the same separation of the set C. Here, two tight equivalent witnesses detect the red point, but not the black one.

(W1) $\xi(x) \geq 0$ for all $x \in C_0$ and $\xi(x) < 0$ at least for some $x \in \overline{C}_0$;

(W2) $\xi(tx + (1-t)y) = t\xi(x) + (1-t)\xi(y)$ for all $x, y \in C$ and $t \in [0, 1]$.

By condition (W2), each witness generates a hyperplane separating $\mathcal{V}$ into two half-spaces. Condition (W1) then asserts that one of the two halves entirely contains C_0, but still does not contain all of C (see Fig. 1).

We say that an element $x \in \overline{C}_0$ is *detected* by ξ if $\xi(x) < 0$, and we denote by $\mathcal{D}(\xi)$ the subset of all elements of $\overline{C}_0$ that are detected by ξ. Another $\overline{C}_0$-witness ξ' is called *finer* than ξ if $\mathcal{D}(\xi') \supseteq \mathcal{D}(\xi)$, and in this case we write $\xi \preceq \xi'$. If $\mathcal{D}(\xi') = \mathcal{D}(\xi)$, we say that ξ and ξ' are *detection equivalent* and denote this by $\xi \approx \xi'$ (see Fig. 1). As we typically aim to detect as many elements as possible, we favor witnesses that cannot be made any finer. A necessary condition for ξ being optimal in that sense is that ξ is *tight*, meaning that $\xi(x) = 0$ for some $x \in C_0$.

Any $\overline{C}_0$-witness ξ can be written in the form

$$\xi(x) = \delta - v^*(x) \qquad \forall x \in C, \tag{1}$$

where $v^* : \mathcal{V} \to \mathbb{R}$ is a linear map and $\delta \in \mathbb{R}$ is a constant. An essential point for our later developments is that the representation (1) of a witness ξ is not unique but there is some freedom in the choice of v^* and δ. In addition, if we are only interested in the set of detected elements $\mathcal{D}(\xi)$, we have a further degree of freedom, coming from the possibility to switch from ξ to an equivalent $\overline{C}_0$-witness $\xi' = \alpha\xi$ for some constant $\alpha > 0$.

Detecting quantum incompatibility.— A measurement with a finite outcome set X is mathematically described as a positive operator valued measure (POVM), i.e., a map A from X to the set $\mathcal{L}_s(\mathcal{H})$ of self-adjoint linear operators on a Hilbert space $\mathcal{H}$ such that the operators $\mathsf{A}(x)$ are positive (meaning that $\langle \psi | \mathsf{A}(x)\psi \rangle \geq 0$ for all $\psi \in \mathcal{H}$) and they satisfy the normalization condition $\sum_x \mathsf{A}(x) = \mathbb{1}$.

For clarity, we limit our discussion to pairs of measurements. The treatment of finite collections of measurements is similar. Two measurements A and B, having outcome sets X and Y, respectively, are *compatible* if

there exists a measurement M, called their *joint measurement*, with outcome set $X \times Y$, such that $\sum_y \mathsf{M}(x, y) = \mathsf{A}(x)$ and $\sum_x \mathsf{M}(x, y) = \mathsf{B}(y)$. Otherwise, A and B are *incompatible*.

By $\mathcal{O}_{X,Y}$ we denote the compact set of all pairs of measurements (A, B) with outcome sets X, Y, respectively. This set is divided into compatible pairs $\mathcal{O}^{\mathrm{com}}_{X,Y}$ and incompatible pairs $\mathcal{O}^{\mathrm{inc}}_{X,Y} \equiv \overline{\mathcal{O}^{\mathrm{com}}_{X,Y}}$. We define convex combinations in $\mathcal{O}_{X,Y}$ componentwise, and it follows that the subset $\mathcal{O}^{\mathrm{com}}_{X,Y}$ of compatible pairs is closed and convex. Hence we can consider $\mathcal{O}^{\mathrm{inc}}_{X,Y}$-witnesses; we call them *incompatibility witnesses* (IWs).

Discrimination scenario as an incompatibility witness.— In the standard state discrimination scenario [19–21], Alice picks a label z from a given set Z with probability $p(z)$. She encodes the label into a quantum state ϱ_z and delivers the state to Bob. Bob knows the set $\{\varrho_z\}_{z \in Z}$ of states used in the encoding. He is trying to recover the label by making a measurement on the quantum system that he has received. It is convenient to merge the a priori probability distribution p and the state encoding into a single map $\mathcal{E}$, given as $\mathcal{E}(z) = p(z)\varrho_z$. We call this map a *state ensemble*; its defining properties are that $\mathcal{E}(z)$ is positive for all z, and $\sum_z \mathrm{tr}[\mathcal{E}(z)] = 1$. The guessing probability depends on the measurement M that Bob uses, and it is given as

$$P_{\mathrm{guess}}(\mathcal{E}; \mathsf{M}) = \sum_z \mathrm{tr}\left[\mathcal{E}(z)\mathsf{M}(z)\right].$$

Further, we denote

$$P_{\mathrm{guess}}(\mathcal{E}) = \max_{\mathsf{M}} P_{\mathrm{guess}}(\mathcal{E}; \mathsf{M}), \tag{2}$$

where the optimization is done over all measurements with outcome set Z.

We are then considering two modifications of the standard state discrimination scenario, where partial classical information concerning the correct label is given either before or after the measurement is performed [22–25]. The form of the partial information is given as a partitioning $Z = X \cup Y$ of Z into two disjoint subsets. By conditioning the state ensemble $\mathcal{E}$ to the occurrence of a label in X or Y, we obtain new state ensembles $\mathcal{E}_X$ and $\mathcal{E}_Y$, which we call *subensembles* of $\mathcal{E}$; they are given as

$$\mathcal{E}_X(x) = \frac{1}{p(X)}\mathcal{E}(x), \qquad \mathcal{E}_Y(y) = \frac{1}{p(Y)}\mathcal{E}(y),$$

and their label sets are X and Y, respectively. Here we have denoted $p(X) = \sum_{z \in X} p(z)$ and $p(Y) = \sum_{z \in Y} p(z)$. We write $\hat{\mathcal{E}} = (\mathcal{E}, \{X, Y\})$ for the partitioned state ensemble, i.e., the state ensemble $\mathcal{E}$ with the partitioning of Z into disjoint subsets X and Y.

If Alice announces the correct subensemble before Bob chooses his measurement, we call the task *discrimination with premeasurement information*. In this case, Bob can choose a measurement A with the outcome set X to discriminate $\mathcal{E}_X$ and a measurement B with the outcome

set Y to discriminate $\mathcal{E}_Y$. At each round of the experiment he measures either A or B, depending on Alice's announcement. Bob's total guessing probability is

$$P_{\text{guess}}^{\text{prior}}(\hat{\mathcal{E}}; \mathsf{A}, \mathsf{B}) = p(X) P_{\text{guess}}(\mathcal{E}_X; \mathsf{A}) + p(Y) P_{\text{guess}}(\mathcal{E}_Y; \mathsf{B}) \tag{3}$$

and its maximal value is

$$\begin{aligned}
P_{\text{guess}}^{\text{prior}}(\hat{\mathcal{E}}) &= \max_{(\mathsf{A},\mathsf{B}) \in \mathcal{O}_{X,Y}} P_{\text{guess}}^{\text{prior}}(\hat{\mathcal{E}}; \mathsf{A}, \mathsf{B}) \\
&= p(X) P_{\text{guess}}(\mathcal{E}_X) + p(Y) P_{\text{guess}}(\mathcal{E}_Y).
\end{aligned} \tag{4}$$

In the other variant of the discrimination scenario, Alice announces the correct subensemble only after Bob has performed his measurement. Bob has to use a fixed measurement at each round but he can postprocess the obtained measurement outcome according to the additional information. We call this task *discrimination with postmeasurement information.* It has been shown in [24] that now the maximal guessing probability, denoted as $P_{\text{guess}}^{\text{post}}(\hat{\mathcal{E}})$, is given by

$$P_{\text{guess}}^{\text{post}}(\hat{\mathcal{E}}) = \max_{(\mathsf{A},\mathsf{B}) \in \mathcal{O}_{X,Y}^{\text{com}}} P_{\text{guess}}^{\text{prior}}(\hat{\mathcal{E}}; \mathsf{A}, \mathsf{B}). \tag{5}$$

A comparison of (4) and (5) reveals that the maximal guessing probabilities $P_{\text{guess}}^{\text{prior}}(\hat{\mathcal{E}})$ and $P_{\text{guess}}^{\text{post}}(\hat{\mathcal{E}})$ result in optimizing the same mathematical quantity, with the important difference that in the latter the optimization is restricted to compatible pairs of measurements. From this, we already conclude that if $P_{\text{guess}}^{\text{prior}}(\hat{\mathcal{E}}; \mathsf{A}, \mathsf{B}) > P_{\text{guess}}^{\text{post}}(\hat{\mathcal{E}})$ for some partitioned state ensemble $\hat{\mathcal{E}}$, then A and B are incompatible. This conclusion is essentially [24, Thm. 1], stated in slightly different words. In the following, we develop this observation into a necessary and sufficient condition for incompatibility by using the framework of witnesses.

We first notice that, for a partitioned state ensemble $\hat{\mathcal{E}} = (\mathcal{E}, \{X, Y\})$ with $P_{\text{guess}}^{\text{prior}}(\hat{\mathcal{E}}) > P_{\text{guess}}^{\text{post}}(\hat{\mathcal{E}})$, the function

$$\xi_{\hat{\mathcal{E}}}(\mathsf{A}, \mathsf{B}) = P_{\text{guess}}^{\text{post}}(\hat{\mathcal{E}}) - P_{\text{guess}}^{\text{prior}}(\hat{\mathcal{E}}; \mathsf{A}, \mathsf{B}) \tag{6}$$

is a tight IW for pairs of measurements in $\mathcal{O}_{X,Y}$; we call it the *incompatibility witness associated with* $\hat{\mathcal{E}}$. In some cases, the exact evaluation of $P_{\text{guess}}^{\text{post}}(\hat{\mathcal{E}})$ may be a difficult task, but still by finding a number δ such that $P_{\text{guess}}^{\text{post}}(\hat{\mathcal{E}}) \le \delta < P_{\text{guess}}^{\text{prior}}(\hat{\mathcal{E}})$ one obtains an IW by setting

$$\xi_{\hat{\mathcal{E}}}^{\delta}(\mathsf{A}, \mathsf{B}) = \delta - P_{\text{guess}}^{\text{prior}}(\hat{\mathcal{E}}; \mathsf{A}, \mathsf{B}). \tag{7}$$

Clearly, we then have $\xi_{\hat{\mathcal{E}}}^{\delta} \preccurlyeq \xi_{\hat{\mathcal{E}}}$.

An important feature of the witnesses arising from partitioned state ensembles is that their physical implementation is straightforward. Namely, the quantities $P_{\text{guess}}(\mathcal{E}_X; \mathsf{A})$ and $P_{\text{guess}}(\mathcal{E}_Y; \mathsf{B})$ are obtained by performing standard state discrimination experiments, and $P_{\text{guess}}^{\text{prior}}(\hat{\mathcal{E}}; \mathsf{A}, \mathsf{B})$ is then given via (3). The constant term $P_{\text{guess}}^{\text{post}}(\hat{\mathcal{E}})$ must be calculated analytically or numerically,

or at least upper bounded tightly enough. It has been shown in [24] that the calculation of $P_{\text{guess}}^{\text{post}}(\hat{\mathcal{E}})$ reduces to the evaluation of the standard guessing probability $P_{\text{guess}}(\mathcal{E}')$ of an auxiliary state ensemble $\mathcal{E}'$, and the techniques for calculating the standard guessing probability (see, e.g., [26]) are thereby applicable.

Characterization of incompatibility witnesses.— The following two theorems are the main results of this paper.

Theorem 1. *For any incompatibility witness ξ, there exists a partitioned state ensemble $\hat{\mathcal{E}}$ such that the associated incompatibility witness $\xi_{\hat{\mathcal{E}}}$ is finer than ξ. Further, if ξ is tight, there exists a partitioned state ensemble $\hat{\mathcal{E}}$ such that ξ is detection equivalent to $\xi_{\hat{\mathcal{E}}}$.*

In the case of IWs, the natural choice for the ambient vector space $\mathcal{V}$ containing $\mathcal{O}_{X,Y}$ is the Cartesian product $\mathcal{F}(X) \times \mathcal{F}(Y)$, where $\mathcal{F}(X)$ is the vector space of all operator valued functions $F: X \to \mathcal{L}_s(\mathcal{H})$. All linear maps on $\mathcal{F}(X) \times \mathcal{F}(Y)$ are expressible in terms of scalar products with elements $(F, G) \in \mathcal{F}(X) \times \mathcal{F}(Y)$, so that the basic representation (1) of witnesses takes the form

$$\xi(\mathsf{A}, \mathsf{B}) = \delta - \sum_x \text{tr}\left[F(x) \mathsf{A}(x)\right] - \sum_y \text{tr}\left[G(y) \mathsf{B}(y)\right] \tag{8}$$

for all $(\mathsf{A}, \mathsf{B}) \in \mathcal{O}_{X,Y}$. The proof of Thm. 1 is based on the freedom in the choice of (F, G) and δ.

Proof of Thm. 1. Starting from an IW ξ of the general form (8), we similarly define a map ξ' by choosing $F'(x) = \alpha[F(x) - \mu\mathbb{1}]$, $G'(y) = \alpha[G(y) - \mu\mathbb{1}]$ and $\delta' = \alpha(\delta - 2\mu d)$, where d is the dimension of the Hilbert space and $\alpha, \mu \in \mathbb{R}$ are constants that we determine next. A direct calculation shows that $\xi' = \alpha\xi$ on $\mathcal{O}_{X,Y}$. First, we fix the value of μ by setting

$$-\mu = \sum_{x \in X} \|F(x)\| + \sum_{y \in Y} \|G(y)\|,$$

where $\|\cdot\|$ denotes the uniform operator norm on $\mathcal{L}_s(\mathcal{H})$. With this choice, all the operators $\mathcal{E}(x) = |\alpha|\left[F(x) - \mu\mathbb{1}\right]$ and $\mathcal{E}(y) = |\alpha|\left[G(y) - \mu\mathbb{1}\right]$ are positive. Secondly, we fix the value of α by setting

$$\frac{1}{\alpha} = \sum_{x \in X} \text{tr}\left[F(x) - \mu\mathbb{1}\right] + \sum_{y \in Y} \text{tr}\left[G(y) - \mu\mathbb{1}\right].$$

The right-hand side of this expression is strictly positive, as otherwise $F(x) = G(y) = \mu\mathbb{1}$ for all x, y and so the original IW (8) would be constant on $\mathcal{O}_{X,Y}$, which is impossible. Thereby, $\alpha > 0$; hence, the map $\xi' = \alpha\xi$ is an IW and $\xi' \approx \xi$. Moreover, in this way we have obtained a partitioned state ensemble $\hat{\mathcal{E}} = (\mathcal{E}, \{X, Y\})$, for which the witness ξ' has the form (7): $\xi'(\mathsf{A}, \mathsf{B}) = \delta' - P_{\text{guess}}^{\text{prior}}(\hat{\mathcal{E}}; \mathsf{A}, \mathsf{B})$. Since ξ' is an IW and hence satisfies (W1), we must have $P_{\text{guess}}^{\text{post}}(\hat{\mathcal{E}}) \le \delta' < P_{\text{guess}}^{\text{prior}}(\hat{\mathcal{E}})$. Thereby, $\xi' \preccurlyeq \xi_{\hat{\mathcal{E}}}$. If in addition ξ is tight, then $\delta' = P_{\text{guess}}^{\text{post}}(\hat{\mathcal{E}})$, and thus $\xi' = \xi_{\hat{\mathcal{E}}}$. $\square$

An important consequence of Thm. 1 is the following novel operational interpretation for quantum incompatibility.

Theorem 2. *Two measurements* A *and* B *are incompatible if and only if there exists a partitioned state ensemble* $\hat{\mathcal{E}}$ *such that* $P_{\mathrm{guess}}^{\mathrm{prior}}(\hat{\mathcal{E}}; \mathsf{A}, \mathsf{B}) > P_{\mathrm{guess}}^{\mathrm{post}}(\hat{\mathcal{E}})$.

The probability $P_{\mathrm{guess}}^{\mathrm{prior}}(\hat{\mathcal{E}}; \mathsf{A}, \mathsf{B})$ is assessable by using Alice's classical information, and then performing quantum measurements only on Bob's side. Since no entangled state is shared in the state discrimination protocol, Thm. 2 provides a much more practical way to detect incompatibility than schemes based on Bell experiments or steering. In particular, as a fundamental fact, entanglement is not needed to detect incompatibility.

Proof of Thm. 2. The "if" statement has already been observed earlier, so here we prove the "only if" part. Let us assume that $(\mathsf{A}, \mathsf{B}) \notin \mathcal{O}_{X,Y}^{\mathrm{com}}$. Then, by the usual separation results for compact convex sets [27, Cor. 11.4.2], there exist $(F, G) \in \mathcal{F}(X) \times \mathcal{F}(Y)$ and $\delta \in \mathbb{R}$ such that, defining ξ as in (8), we have $\xi(\mathsf{A}', \mathsf{B}') \geq 0$ for all $(\mathsf{A}', \mathsf{B}') \in \mathcal{O}_{X,Y}^{\mathrm{com}}$ and $\xi(\mathsf{A}, \mathsf{B}) < 0$. By Thm. 1 there exists a partitioned state ensemble $\hat{\mathcal{E}}$ such that $\xi \preccurlyeq \xi_{\hat{\mathcal{E}}}$. It follows that $\xi_{\hat{\mathcal{E}}}(\mathsf{A}, \mathsf{B}) < 0$, i.e., $P_{\mathrm{guess}}^{\mathrm{prior}}(\hat{\mathcal{E}}; \mathsf{A}, \mathsf{B}) > P_{\mathrm{guess}}^{\mathrm{post}}(\hat{\mathcal{E}})$. $\qquad\square$

Bounding the compatibility region by means of two mutually unbiased bases.— As we have seen, constructing an IW involves the solution of two convex optimization problems: the evaluation of the maximal guessing probabilities defined in (4) and (5). In particular, if $\hat{\mathcal{E}}$ is a partitioned state ensemble for which the two probabilities differ, whenever the maximum in the right-hand side of (5) admits an analytical computation, one can insert the resulting value of $P_{\mathrm{guess}}^{\mathrm{post}}(\hat{\mathcal{E}})$ into (6) and thus write the tight IW associated with $\hat{\mathcal{E}}$ in an explicit form.

Interestingly, solving the optimization problem (5) yields even more. Indeed, evaluating a constrained maximum typically requires finding some feasible points where the maximum is attained; if the optimization problem is convex, these points are necessarily located on the relative boundary of the feasible domain. In our specific case, it means that, as a byproduct of solving (5), we get points lying on the relative boundary $\partial \mathcal{O}_{X,Y}^{\mathrm{com}}$ of the convex set $\mathcal{O}_{X,Y}$. Then, by taking convex combinations of these points, we can even have an insight into the set $\mathcal{O}_{X,Y}^{\mathrm{com}}$ itself. We thus see that the solution of (5) has a twofold purpose: on the one hand, through the IW constructed in (6), it provides a simple method to detect the incompatibility of many measurement pairs; on the other hand, by using the resulting optimal points, some information on the set of compatible pairs can be inferred.

An interesting special case in which the optimization problems (4) and (5) admit an analytical solution is when the partitioned state ensemble $\hat{\mathcal{E}}$ is made up of two mutually unbiased bases (MUB) of the system Hilbert space $\mathcal{H}$, or, more generally, smearings of two

MUB. Indeed, suppose $\{\varphi_h\}_{h \in \{1,\dots,d\}}$ and $\{\psi_k\}_{k \in \{1,\dots,d\}}$ is a fixed pair of MUB; then, we can use it to construct a partitioned state ensemble as follows. First, we choose $Z = \{1,\dots,d\} \times \{\varphi, \psi\}$ as the overall label set of the ensemble

$$\mathcal{E}_{\boldsymbol{\mu}}(j, \ell) = \frac{1}{2d} \left[\mu_\ell |\ell_j\rangle\langle\ell_j| + (1 - \mu_\ell)\frac{1}{d}\mathbb{1} \right], \qquad (9)$$

where $\boldsymbol{\mu} = (\mu_\varphi, \mu_\psi)$ and $\mu_\varphi, \mu_\psi \in [1/(1-d), 1]$ are real parameters. Next, we partition Z into the subsets $X = \{(1, \varphi), \dots (d, \varphi)\}$ and $Y = \{(1, \psi), \dots (d, \psi)\}$; here, the letters φ and ψ are just symbols, which are needed to distinguish labels in different subsets. Finally, we set $\hat{\mathcal{E}}_{\boldsymbol{\mu}} = (\mathcal{E}_{\boldsymbol{\mu}}, \{X, Y\})$.

The detailed solution to the optimization problems (4) and (5) for the partitioned state ensemble $\hat{\mathcal{E}}_{\boldsymbol{\mu}}$ is provided in Supplemental Material. It turns out that the pair of measurements

$$
\begin{aligned}
\mathsf{A}(h, \varphi) &= \gamma_\varphi |\varphi_h\rangle\langle\varphi_h| + (1 - \gamma_\varphi)\frac{1}{d}\mathbb{1} \\
\mathsf{B}(k, \psi) &= \gamma_\psi |\psi_k\rangle\langle\psi_k| + (1 - \gamma_\psi)\frac{1}{d}\mathbb{1}
\end{aligned}
\qquad (10)
$$

is a feasible maximum point for a suitable choice of real numbers γ_φ and γ_ψ, which depend on $\boldsymbol{\mu}$. The next two theorems then follow by our earlier observations.

Theorem 3. *Let* $\boldsymbol{\mu} = (\mu_\varphi, \mu_\psi) \in [1/(1-d), 1] \times [1/(1-d), 1]$ *with* $\boldsymbol{\mu} \neq (0, 0)$. *Then* $P_{\mathrm{guess}}^{\mathrm{post}}(\hat{\mathcal{E}}_{\boldsymbol{\mu}}) < P_{\mathrm{guess}}^{\mathrm{prior}}(\hat{\mathcal{E}}_{\boldsymbol{\mu}})$ *if and only if* $\mu_\varphi\mu_\psi \neq 0$ *and either* $d = 2$ *or* $\max\{\mu_\varphi, \mu_\psi\} > 0$. *In this case, the tight incompatibility witness associated with the partitioned state ensemble* $\hat{\mathcal{E}}_{\boldsymbol{\mu}}$ *is*

$$
\begin{aligned}
\xi_{\hat{\mathcal{E}}_{\boldsymbol{\mu}}}(\mathsf{A}, \mathsf{B}) &= \frac{1}{4}\left[\mu_\varphi + \mu_\psi + \sqrt{\mu_\varphi^2 + \mu_\psi^2 - 2\left(1 - \frac{2}{d}\right)\mu_\varphi\mu_\psi} \right] \\
&\quad - \frac{1}{2d}\sum_{j=1}^{d}\left[\mu_\varphi \langle \varphi_j | \mathsf{A}(j, \varphi)\varphi_j\rangle + \mu_\psi \langle \psi_j | \mathsf{B}(j, \psi)\psi_j\rangle \right].
\end{aligned}
\qquad (11)
$$

Finally, the ensembles $\hat{\mathcal{E}}_{\boldsymbol{\mu}}$ *and* $\hat{\mathcal{E}}_{\boldsymbol{\nu}}$ *determine detection equivalent incompatibility witnesses if and only if* $\boldsymbol{\nu} = \alpha\boldsymbol{\mu}$ *for some* $\alpha > 0$.

By the equivalence statement in the previous theorem, no generality is lost if we express the vector $\boldsymbol{\mu}$ in terms of a single real parameter θ. Consequently, also the vector $\boldsymbol{\gamma} = (\gamma_\varphi, \gamma_\psi)$ parametrizing the optimal measurements (10) becomes a function of θ. Thus, solving the optimization problem (5) for the present case actually yields a curve in the relative boundary $\partial\mathcal{O}_{X,Y}^{\mathrm{com}}$.

Theorem 4. *The pair of measurements* (A, B) *of* (10) *lies on the relative boundary* $\partial\mathcal{O}_{X,Y}^{\mathrm{com}}$ *if*

$$\boldsymbol{\gamma} = \left(\frac{d - 2 - d\cos(\theta + \theta_0)}{2(d-1)}, \frac{d - 2 - d\cos(\theta - \theta_0)}{2(d-1)} \right) \qquad (12)$$

for $\theta \in [-\theta_0, \theta_0]$ *and* $\theta_0 = \pi - \arctan\sqrt{d-1}$.

When $\theta = 0$, the common value of the two components of (12) is the *noise robustness* of the two MUB at hand; it was already derived by different methods in [28, 29]. On the other hand, under the assumption that the two MUB are Fourier conjugate, the portion of the curve (12) with $\gamma_\varphi > 0$ and $\gamma_\psi > 0$ was found in [30].
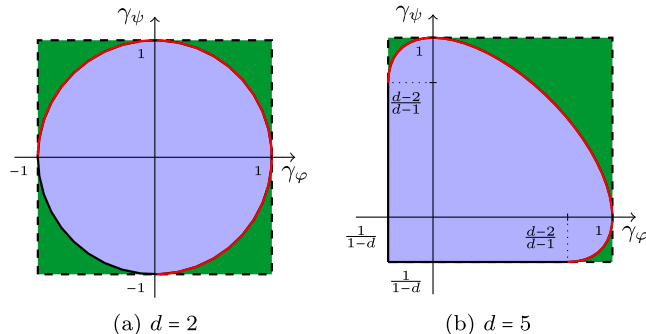

(a) $d = 2$ (b) $d = 5$

FIG. 2. The set of $\boldsymbol{\gamma} = (\gamma_\varphi, \gamma_\psi)$ for which Eq. (10) defines two measurements (green square), and the one for which these measurements are compatible (blue region) for different values of the dimension d. The red line is the curve (12). The case $d = 2$ is special, and was already treated in [31].

The operators in (10) are positive if and only if $\boldsymbol{\gamma} \in [1/(1-d), 1] \times [1/(1-d), 1]$. Thus, all pairs of measurements of the form (10) constitute a square-shaped section of the set $\mathcal{O}_{X,Y}$. Remarkably, the lower-left vertex $(1/(1-d), 1/(1-d))$ of this square corresponds to a compatible pair of measurements if and only if $d \geq 3$; on the contrary, when $d = 2$ the relative boundary is symmetric

around $(0,0)$ [31]. Combining these considerations and Thm. 4, we can give a partial inspection of the two sets $\mathcal{O}_{X,Y}$ and $\mathcal{O}_{X,Y}^{\text{com}}$, as shown in Fig. 2.

Discussion.— The framework of witnesses is an effective tool in the detection of properties described by sets with compact and convex complements. We have shown that for incompatibility of measurements, witnesses are not only a mathematical tool, but can be implemented in simple discrimination experiments. An important feature of this implementation is that it does not require entanglement.

Our characterization yields a novel operational interpretation of incompatibility: a collection of measurements is incompatible if and only if there is a state discrimination task where premeasurement information is strictly better than postmeasurement information.

Entanglement witnesses have been used not only to detect entanglement but also to quantify entanglement [32]. Further, one can drop the condition (W2) and consider nonlinear witnesses [33]. These and other modifications or generalizations will be an interesting matter of investigation in the case of incompatibility witnesses.

ACKNOWLEDGMENTS

Note added.— Recently we have been informed about two works that contain related results [34, 35].

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

[2] P. Busch, P. Lahti, and R.F. Werner. Measurement uncertainty relations. *J. Math. Phys.*, 55:042111, 2014.

[3] Y.-C. Liang, R.W. Spekkens, and H.M. Wiseman. Specker's parable of the overprotective seer: A road to contextuality, nonlocality and complementarity. *Phys. Rep.*, 506:1–39, 2011.

[4] A. Fine. Hidden variables, joint probability, and the Bell inequalities. *Phys. Rev. Lett.*, 48:291–295, 1982.

[5] M.M. Wolf, D. Perez-Garcia, and C. Fernandez. Measurements incompatible in quantum theory cannot be measured jointly in any other no-signaling theory. *Phys. Rev. Lett.*, 103:230402, 2009.

[6] M. Banik, M.R. Gazi, S. Ghosh, and G. Kar. Degree of complementarity determines the nonlocality in quantum mechanics. *Phys. Rev. A*, 87:052125, 2013.

[7] G. Kar, S. Ghosh, S.K. Choudhary, and M. Banik. Role of measurement incompatibility and uncertainty in determining nonlocality. *Mathematics*, 4:52, 2016.

[8] R. Uola, T. Moroder, and O. Gühne. Joint measurability of generalized measurements implies classicality. *Phys. Rev. Lett.*, 113:160403, 2014.

[9] M.T. Quintino, T. Vertesi, and N. Brunner. Joint mea-

surability, Einstein-Podolsky-Rosen steering, and Bell nonlocality. *Phys. Rev. Lett.*, 113:160402, 2014.

[10] R. Uola, C. Budroni, O. Gühne, and J.-P. Pellonpää. One-to-one mapping between steering and joint measurability problems. *Phys. Rev. Lett.*, 115:230402, 2015.

[11] S.-L. Chen, C. Budroni, Y.-C. Liang, and Y.-N. Chen. Natural framework for device-independent quantification of quantum steerability, measurement incompatibility, and self-testing. *Phys. Rev. Lett.*, 116:240401, 2016.

[12] O. Gühne and G. Tóth. Entanglement detection. *Phys. Rep.*, 474:1–75, 2009.

[13] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki. Quantum entanglement. *Rev. Mod. Phys.*, 81:865–942, 2009.

[14] C. Hughes, M.G. Genoni, T. Tufarelli, M.G.A. Paris, and M.S. Kim. Quantum non-Gaussianity witnesses in phase space. *Phys. Rev. A*, 90:013810, 2014.

[15] N. Brunner, S. Pironio, A. Acin, N. Gisin, A. Méthot, and V. Scarani. Testing the dimension of Hilbert spaces. *Phys. Rev. Lett.*, 100:210503, 2008.

[16] C.B. Mendl and M.M. Wolf. Unital quantum channels— convex structure and revivals of Birkhoff's theorem. *Comm. Math. Phys.*, 289:1057–1086, 2009.

[17] A. Jenčová. Incompatible measurements in a class of general probabilistic theories. *Phys. Rev. A*, 98:012133,

2018.

[18] A. Bluhm and I. Nechita. Compatibility of quantum measurements and inclusion constants for the matrix jewel. arXiv:1809.04514 [quant-ph].

[19] A.S. Holevo. Statistical decision theory for quantum systems. *J. Multivariate Anal.*, 3:337–394, 1973.

[20] H.P. Yuen, R.S. Kennedy, and M. Lax. Optimum testing of multiple hypotheses in quantum detection theory. *IEEE Trans. Inform. Theory*, IT-21:125–134, 1975.

[21] C.W. Helstrom. *Quantum Detection and Estimation Theory*. Academic Press, New York, 1976.

[22] M.A. Ballester, S. Wehner, and A. Winter. State discrimination with post-measurement information. *IEEE Trans. Inf. Theory*, 54:4183–4198, 2008.

[23] D. Gopal and S. Wehner. Using postmeasurement information in state discrimination. *Phys. Rev. A*, 82:022326, 2010.

[24] C. Carmeli, T. Heinosaari, and A. Toigo. State discrimination with postmeasurement information and incompatibility of quantum measurements. *Phys. Rev. A*, 98:012126, 2018.

[25] S. Akibue, G. Kato, and N. Marumo. Perfect discrimination of nonorthogonal quantum states with posterior classical partial information. *Phys. Rev. A*, 99:020102, 2019

[26] J. Bae. Structure of minimum-error quantum state discrimination. *New J. Phys.*, 15:073037, 2013.

[27] R.T. Rockafellar. *Convex analysis*. Princeton University Press, Princeton, N.J., 1997.

[28] R. Uola, K. Luoma, T. Moroder, and T. Heinosaari. Adaptive strategy for joint measurements. *Phys. Rev. A*, 94:022109, 2016.

[29] S. Designolle, P. Skrzypczyk, F. Fröwis, and N. Brunner. Quantifying measurement incompatibility of mutually unbiased bases. *Phys. Rev. Lett.*, 122:050402, 2019.

[30] C. Carmeli, T. Heinosaari, and A. Toigo. Informationally complete joint measurements on finite quantum systems. *Phys. Rev. A*, 85:012109, 2012.

[31] P. Busch. Unsharp reality and joint measurements for spin observables. *Phys. Rev. D*, 33:2253–2261, 1986.

[32] J. Eisert, F.G.S. Brandão, and K.M.R. Audenaert. Quantitative entanglement witnesses. *New J. Phys.*, 9:46, 2007.

[33] O. Gühne and N. Lütkenhaus. Nonlinear entanglement witnesses. *Phys. Rev. Lett.*, 96:170502, 2006.

[34] R. Uola, T. Kraft, J. Shang, X-D. Yu, and O. Gühne. Quantifying quantum resources with conic programming. arXiv:1812.09216 [quant-ph].

[35] P. Skrzypczyk, I. Supić, and D. Cavalcanti. All sets of incompatible measurements give an advantage in quantum state discrimination. arXiv:1901.00816 [quant-ph].

Supplementary Material: Quantum Incompatibility Witnesses

This supplement is planned as follows. In Sec. I, we provide the general framework for witnesses, recalling the basic related notions from convex analysis and deriving some useful results; in particular, by making a natural assumption upon the convex set to be detected, we fully characterize the equivalence classes of witnesses under the detection equivalence relation. In Sec. II, these results are applied to incompatibility witnesses; to do it, we describe the convex structure of the set of all pairs of measurements and its subset of all the pairs that are compatible. Finally, in Sec. III, we solve the optimization problem for the guessing probability with postmeasurement information for a state ensemble constructed by means of two mutually unbiased bases; we thus provide the detailed proofs of Thms. 3 and 4 of the main paper.

Notation: In this supplement, the numberings of equations, theorems etc. are preceeded by the letter 'S' (e.g.: Eq. (S1), Thm. S1 etc.). When we refer to results in the main paper, we simply drop the 'S' (Eq. (1), Thm. 1 etc.).

I. WITNESSES FOR GENERAL CONVEX SETS

A. Preliminaries from convex analysis

In the following, we will need some standard terminology and notations from convex analysis. We refer to [1] for further details.

Suppose $\mathcal{V}$ is a finite dimensional, real and normed linear space. If $\mathcal{V}_0 \subseteq \mathcal{V}$ is a linear subspace, we denote by $\mathcal{V}_0^*$ the dual linear space of $\mathcal{V}_0$, and $\langle v^*, v \rangle$ the canonical pairing between an element $v \in \mathcal{V}_0$ and a dual vector $v^* \in \mathcal{V}_0^*$.

We recall that

- an *affine* [respectively, *convex*] *combination* of elements $v_1, \ldots, v_n \in \mathcal{V}$ is any linear combination $\lambda_1 v_1 + \ldots + \lambda_n v_n$ such that $\lambda_1 + \ldots + \lambda_n = 1$ [resp., such that $\lambda_1 + \ldots + \lambda_n = 1$ and $\lambda_i \in [0, 1]$ for all $i = 1, \ldots, n$];

- an *affine* [resp., *convex*] *set* is any subset $\mathcal{S} \subseteq \mathcal{V}$ such that all affine [resp., convex] combinations of elements of $\mathcal{S}$ are still contained in $\mathcal{S}$;

- an *affine* [resp., *c-affine*] *map* on an affine [resp., convex] set $\mathcal{S}$ is a function $\xi : \mathcal{S} \to \mathbb{R}$ such that

$$\xi(\lambda_1 v_1 + \ldots + \lambda_n v_n) = \lambda_1 \xi(v_1) + \ldots + \lambda_n \xi(v_n)$$

for all $v_1, \ldots, v_n \in \mathcal{S}$ and any affine [resp., convex] combination $\lambda_1 v_1 + \ldots + \lambda_n v_n$.

If $\mathcal{M}$ is an affine set, there is a unique vector subspace $\mathcal{V}(\mathcal{M}) \subseteq \mathcal{V}$ such that $\mathcal{M} = v_0 + \mathcal{V}(\mathcal{M})$ for some (hence for any) $v_0 \in \mathcal{M}$. Moreover, if $\xi : \mathcal{M} \to \mathbb{R}$ is an affine map and $v_0 \in \mathcal{M}$ is fixed, there exist unique $v_\xi^* \in \mathcal{V}(\mathcal{M})^*$ and $\delta_\xi \in \mathbb{R}$ such that $\xi(v_0 + v) = \delta_\xi - \langle v_\xi^*, v \rangle$ for all $v \in \mathcal{V}(\mathcal{M})$. In particular, given an affine map $\Xi : \mathcal{V} \to \mathbb{R}$, there exist unique $v^* \in \mathcal{V}^*$ and $\delta \in \mathbb{R}$ such that

$$\Xi(v) = \delta - \langle v^*, v \rangle =: \Xi[v^*, \delta](v) \qquad \forall v \in \mathcal{V}. \tag{S1}$$

By picking any $v^* \in \mathcal{V}^*$ whose restriction to $\mathcal{V}(\mathcal{M})$ coincides with v_ξ^* and choosing $\delta = \delta_\xi - \langle v^*, v_0 \rangle$, we see that the previous affine map $\xi : \mathcal{M} \to \mathbb{R}$ extends to the affine map $\Xi[v^*, \delta] : \mathcal{V} \to \mathbb{R}$. Clearly, this extension is not unique unless $\mathcal{V}(\mathcal{M}) = \mathcal{V}$.

Now, suppose $C \subseteq \mathcal{V}$ is a convex set. Then, the *affine hull* of C is the smallest affine set $\mathcal{M}(C)$ containing C; equivalently, it is the set of all affine combinations of elements of C. We abbreviate $\mathcal{V}(C) = \mathcal{V}(\mathcal{M}(C))$; further, we introduce the following *annihilator subspace* of $\mathcal{V}(C)$ in $\mathcal{V}^*$:

$$\mathcal{V}(C)^\perp = \{v^* \in \mathcal{V}^* \mid \langle v^*, v \rangle = 0 \ \ \forall v \in \mathcal{V}(C)\}.$$

Any c-affine map $\xi : C \to \mathbb{R}$ uniquely extends to an affine map $\tilde{\xi} : \mathcal{M}(C) \to \mathbb{R}$; as we have already seen, such a map $\tilde{\xi}$ can be further extended to an affine map $\Xi : \mathcal{V} \to \mathbb{R}$, the latter extension being in general not unique. Actually, the first assertion is a particular case of the following more general result [2].

Proposition S1. *Suppose C_1 and C_2 are convex sets, and let $\phi : C_1 \to C_2$ be a map such that $\phi(\lambda_1 v_1 + \ldots + \lambda_n v_n) = \lambda_1 \phi(v_1) + \ldots + \lambda_n \phi(v_n)$ for all convex combinations of elements $v_1, \ldots, v_n \in C_1$. Then, there exists a unique map $\tilde{\phi} : \mathcal{M}(C_1) \to \mathcal{M}(C_2)$ such that the restriction $\tilde{\phi}|_{C_1}$ coincides with ϕ and $\tilde{\phi}(\lambda_1 v_1 + \ldots + \lambda_n v_n) = \lambda_1 \tilde{\phi}(v_1) + \ldots + \lambda_n \tilde{\phi}(v_n)$ for all affine combinations of elements $v_1, \ldots, v_n \in \mathcal{M}(C_1)$.*

Proof. Any element of the affine hull $\mathcal{M}(C_i)$ is an affine combination of elements of C_i. Then, if $x = \lambda_1 v_1 + \ldots + \lambda_n v_n \in \mathcal{M}(C_1)$ is any affine combination of $v_1, \ldots, v_n \in C_1$, we define

$$\tilde{\phi}(x) = \lambda_1 \phi(v_1) + \ldots + \lambda_n \phi(v_n).$$

First of all, we claim that this definition is independent of the chosen representation of x. Indeed, suppose $x = \mu_1 w_1 + \ldots + \mu_m w_m$ for some other $w_1, \ldots, v_m \in C_1$ and $\mu_1, \ldots, \mu_m \in \mathbb{R}$ with $\mu_1 + \ldots + \mu_m = 1$. We set

$$\sigma = \sum_{i|\lambda_i>0} \lambda_i - \sum_{j|\mu_j<0} \mu_j = \sum_{j|\mu_j>0} \mu_j - \sum_{i|\lambda_i<0} \lambda_i,$$

and observe that $\sigma > 0$ since at least one among $\lambda_1, \ldots, \lambda_n$ is necessarily strictly positive. Then, as

$$\sum_{i|\lambda_i>0} \frac{\lambda_i}{\sigma} v_i + \sum_{j|\mu_j<0} \frac{-\mu_j}{\sigma} w_j = \sum_{j|\mu_j>0} \frac{\mu_j}{\sigma} w_j + \sum_{i|\lambda_i<0} \frac{-\lambda_i}{\sigma} v_i,$$

and the two sides of the latter equation are convex combinations of elements of C_1, it follows that

$$\sum_{i=1}^{n} \lambda_i \phi(v_i) - \sum_{j=1}^{m} \mu_j \phi(w_j) = \sigma \left[\phi \left(\sum_{i|\lambda_i>0} \frac{\lambda_i}{\sigma} v_i + \sum_{j|\mu_j<0} \frac{-\mu_j}{\sigma} w_j \right) - \phi \left(\sum_{j|\mu_j>0} \frac{\mu_j}{\sigma} w_j + \sum_{i|\lambda_i<0} \frac{-\lambda_i}{\sigma} v_i \right) \right] = 0,$$

thus showing that $\tilde{\phi}$ is well defined.

Secondly, we prove that $\tilde{\phi}$ is an affine map, that is

$$\tilde{\phi} \left(\sum_{i=1}^{n} \lambda_i u_i \right) = \sum_{i=1}^{n} \lambda_i \tilde{\phi}(u_i) \quad \text{for all} \quad u_1, \ldots, u_n \in \mathcal{M}(C_1) \quad \text{and} \quad \lambda_1, \ldots, \lambda_n \in \mathbb{R} \quad \text{with} \quad \sum_{i=1}^{n} \lambda_i = 1. \qquad (\circ)$$

To do it, write each u_i as an affine combination $u_i = \mu_{i,1} v_{i,1} + \ldots + \mu_{i,m_i} v_{i,m_i}$ of elements $v_{i,1}, \ldots, v_{i,m_i} \in C_1$. Then,

$$\sum_{i=1}^{n} \lambda_i u_i = \sum_{i=1}^{n} \sum_{j=1}^{m_i} \lambda_i \mu_{i,j} v_{i,j},$$

where the right hand side is an affine combination of the elements $v_{i,j} \in C_1$. Hence,

$$\tilde{\phi} \left(\sum_{i=1}^{n} \lambda_i u_i \right) = \sum_{i=1}^{n} \sum_{j=1}^{m_i} \lambda_i \mu_{i,j} \phi(v_{i,j}) \qquad \text{and} \qquad \tilde{\phi}(u_i) = \sum_{j=1}^{m_i} \mu_{i,j} \phi(v_{i,j})$$

by definition of $\tilde{\phi}$. This proves $(\circ)$.

Finally, the equality $\tilde{\phi}\big|_{C_1} = \phi$ also follows by the very definition of $\tilde{\phi}$. $\qquad \square$

The *relative interior* $\mathrm{ri}(C)$ of the convex set C is the set of all its interior points with respect to the relative topology of $\mathcal{M}(C)$. The *relative boundary* of C is the set-theoretic difference $\partial C = C \smallsetminus \mathrm{ri}(C)$. An element $z \in C$ is an *extreme point* of C if the equality $z = \lambda x + (1 - \lambda) y$ with $x, y \in C$ and $\lambda \in (0, 1)$ implies $x = y = z$.

If C is a compact convex set, its *support function* is the map

$$\delta_C : \mathcal{V}^* \to \mathbb{R}, \qquad \delta_C(v^*) = \max\{\langle v^*, x \rangle \mid x \in C\}.$$

Clearly, if $C_0 \subseteq C$ is a compact convex subset, then $\delta_{C_0}(v^*) \le \delta_C(v^*)$ for all $v^* \in V^*$. The support function satisfies $\delta_C(\alpha v^*) = \alpha \delta_C(v^*)$ for all $\alpha > 0$; moreover, if $v^* - u^* \in \mathcal{V}(C)^\perp$, then $\delta_C(v^*) - \delta_C(u^*) = \langle v^* - u^*, v_0 \rangle$ for some (hence for all) $v_0 \in \mathcal{M}(C)$.

B. Detecting a convex property

We are interested in a set of objects C, in which we consider the subset $C_0 \subset C$ of all the objects sharing some given property. We assume $\varnothing \ne C_0 \ne C$, and we denote by $\overline{C}_0 = C \smallsetminus C_0$ the subset of all the objects which *do not* possess the property at hand; we want to find simple sufficient conditions guaranteeing that an object $x \in C$ actually belongs to $\overline{C}_0$.

In the following, we always suppose that both C and C_0 are convex and compact subsets of a finite dimensional, real and normed linear space $\mathcal{V}$. Then, the simplest conditions involve some specific c-affine map $\xi : C \to \mathbb{R}$, related to both C and C_0, and the value that ξ takes at $x \in C$.

Definition S1. *A $\overline{C}_0$-witness on the convex set C is a c-affine map $\xi : C \to \mathbb{R}$ such that*

(i) $\xi(y) \geq 0$ for all $y \in C_0$;

(ii) $\xi(x) < 0$ for some $x \in \overline{C}_0$.

A $\overline{C}_0$-witness is tight *if it satisfies the further condition*

(iii) $\xi(z) = 0$ for some $z \in C_0$.

If ξ is a $\overline{C}_0$-witness, the inequality $\xi(x) < 0$ entails that $x \in C$ does not possess the property C_0. The set $\mathcal{D}(\xi) = \{x \in C \mid \xi(x) < 0\}$ is thus the subset of all the objects of $\overline{C}_0$ which are *detected* by ξ. Note that there always exists a tight $\overline{C}_0$-witness ξ' detecting at least as many objects of $\overline{C}_0$ as ξ. Namely, it is enough to set $\xi'(x) = \xi(x) - \min\{\xi(y) \mid y \in C_0\}$. In general, whenever two $\overline{C}_0$-witnesses ξ and ξ' are such that $\mathcal{D}(\xi) \subseteq \mathcal{D}(\xi')$, we say that ξ' is *finer* than ξ, and in this case we write $\xi \preccurlyeq \xi'$. Moreover, if $\mathcal{D}(\xi) = \mathcal{D}(\xi')$, we say that ξ and ξ' are *detection equivalent* and write $\xi \approx \xi'$.

C. Structure of $\overline{C}_0$-witnesses

Proposition S2. *A c-affine map $\xi : C \to \mathbb{R}$ is a $\overline{C}_0$-witness if and only if there exist $v^* \in \mathcal{V}^*$ and $\delta \in \mathbb{R}$ such that*

(i) $\delta_{C_0}(v^*) \leq \delta < \delta_C(v^*)$;

(ii) $\xi(x) = \delta - \langle v^*, x \rangle$ for all $x \in C$.

In (i), *the equality is attained if and only if ξ is tight.*

Proof. Any c-affine map ξ on C extends to an affine map $\Xi = \Xi[v^*, \delta]$ on $\mathcal{V}$, and then $\xi(x) = \delta - \langle v^*, x \rangle \quad \forall x \in C$ by (S1). In this case, ξ is a $\overline{C}_0$-witness if and only if, for all $y \in C_0$ and some $x \in \overline{C}_0$,

$$\Xi[v^*, \delta](y) \geq 0 > \Xi[v^*, \delta](x) \quad \Leftrightarrow \quad \langle v^*, y \rangle \leq \delta < \langle v^*, x \rangle.$$

This is equivalent to $\delta_{C_0}(v^*) \leq \delta < \delta_C(v^*)$, where the equality is attained if and only if $\langle v^*, z \rangle = \delta$ for some $z \in C_0$, that is, $\xi(z) = 0$. $\qquad\square$

If $\mathcal{V}$ is an Euclidean space and v^* is given by the scalar product $\langle v^*, v \rangle = \hat{e} \cdot v \quad \forall v \in \mathcal{V}$ for some unit vector $\hat{e} \in \mathcal{V}$ not belonging to $\mathcal{V}(C_0)^\perp$, the nonnegative gap $\delta_C(v^*) - \delta_{C_0}(v^*)$ is the distance between the affine hyperplane of $\mathcal{V}$ which is orthogonal to $\hat{e}$, touches C at its relative boundary and has C on its side opposite to $\hat{e}$, and the analogous hyperplane touching the relative boundary of C_0.

D. Tight witnesses

The following proposition establishes the natural connection between tight $\overline{C}_0$-witnesses and the extremality property for points of the set C_0. Although it is an easy consequence of more general and standard results (see e.g. [1, Thm. 32.1 and Cor. 32.3.1]), for the reader's convenience we provide a simple proof adapted to the present case.

Proposition S3. *Suppose ξ is a tight $\overline{C}_0$-witness on C. Then, the following facts hold.*

(a) $\xi(z_0) = 0$ for some extreme point z_0 of C_0.

(b) If $\mathcal{V}(C_0) = \mathcal{V}(C)$ and $z \in C_0$, the equality $\xi(z) = 0$ implies that $z \in \partial C_0$.

The proof relies on the following lemma, that is sometimes useful by itself.

Lemma S1. *For $v^* \in \mathcal{V}^*$, the following facts hold.*

(a) $\delta_{C_0}(v^*) = \langle v^*, z_0 \rangle$ for some extreme point z_0 of C_0.

(b) Suppose $v^* \in \mathcal{V}^* \smallsetminus \mathcal{V}(C_0)^\perp$. Then, if $z \in C_0$, the equality $\langle v^*, z \rangle = \delta_{C_0}(v^*)$ implies that $z \in \partial C_0$.

Proof. (a) The set $Z = \{z \in C_0 \mid \langle v^*, z \rangle = \delta_{C_0}(v^*)\}$ is a nonempty closed and convex subset of C_0. By Krein-Milman theorem [1, Cor. 18.5.1], Z has some extreme point z_0. We claim that z_0 is extreme also for C_0. Indeed, suppose that $z_0 = \lambda y_1 + (1 - \lambda)y_2$ with $y_1, y_2 \in C_0$ and $\lambda \in (0, 1)$. The conditions $\delta_{C_0}(v^*) = \langle v^*, z_0 \rangle = \lambda \langle v^*, y_1 \rangle + (1 - \lambda)\langle v^*, y_2 \rangle$ and $\langle v^*, y_i \rangle \le \delta_{C_0}(v^*)$ for $i = 1, 2$ then imply that $y_1, y_2 \in Z$, hence $y_1 = y_2 = z_0$.

(b) Suppose by contradiction that $z \in \mathrm{ri}(C_0)$. Then, for any $v \in \mathcal{V}(C_0)$ with $\langle v^*, v \rangle \ne 0$, there exists $\varepsilon \in \mathbb{R}$ such that $z + \varepsilon v \in C_0$ and $\varepsilon\langle v^*, v \rangle > 0$. It follows that $\langle v^*, z + \varepsilon v \rangle > \delta_{C_0}(v^*)$, which is impossible. $\square$

Proof of Prop. S3. By Prop. S2, for some $v^* \in \mathcal{V}^*$ we have $\xi(x) = \delta_{C_0}(v^*) - \langle v^*, x \rangle$ for all $x \in C$, and $\delta_{C_0}(v^*) < \delta_C(v^*)$; in particular, v^* is not constant on C, hence $v^* \in \mathcal{V}^* \setminus \mathcal{V}(C)^\perp$. Since $\xi(z) = 0$ is then equivalent to $\langle v^*, z \rangle = \delta_{C_0}(v^*)$, the two claims follow by the analogous statements of Lemma S1. $\square$

E. Detection equivalent witnesses

If the subset C_0 is sufficiently large in C, we have the following characterization of detection equivalence.

Proposition S4. *Suppose $C_0 \cap \mathrm{ri}(C) \ne \varnothing$. Then, for two $\overline{C}_0$-witnesses ξ_1 and ξ_2 on C, the following facts are equivalent.*

(a) $\xi_1 \approx \xi_2$.

(b) $\xi_1 = \alpha\xi_2$ for some $\alpha > 0$.

(c) If $v_i^* \in \mathcal{V}^*$ and $\delta_i \in \mathbb{R}$ are such that $\xi_i(x) = \delta_i - \langle v_i^*, x \rangle$ for all $x \in C$ and $i = 1, 2$, then there exists $\alpha > 0$ such that

(i) $v_1^* - \alpha v_2^* \in \mathcal{V}(C)^\perp$;

(ii) $\delta_1 - \alpha\delta_2 = \langle v_1^* - \alpha v_2^*, v_0 \rangle$ for some (hence for all) $v_0 \in \mathcal{M}(C)$.

Note that, if the two $\overline{C}_0$-witnesses in the above proposition are tight, then in statement (c) we have $\delta_i = \delta_{C_0}(v_i^*)$ by Prop. S2, hence item (c.ii) is automatically implied by (c.i) and the properties of the support function δ_{C_0} recalled at the end of Sec. I A.

Prop. S4 exhibits the degree of freedom one has in choosing detection equivalent $\overline{C}_0$-witnesses: namely, if we have a $\overline{C}_0$-witness $\xi(x) = \delta - \langle v^*, x \rangle$ $\forall x \in C$, we can turn it into an equivalent one by: (1) replacing the dual vector $v^* \in \mathcal{V}^*$ with $v'^* = \alpha(v^* + u^*)$, where $u^* \in \mathcal{V}(C)^\perp$ and $\alpha > 0$; (2) redefining the constant δ according to Prop. S4.(c.ii). The smaller is the linear space $\mathcal{V}(C)$ inside $\mathcal{V}$, the larger is the freedom in the choice of the dual vector v^*. This freedom is the crux of the proof of Thm. 1 in the main paper, and therefore it is at the heart of the interpretation of incompatibility witnesses as postmeasurement discrimination problems.

Proof of Prop. S4. (a) $\Rightarrow$ (b): Assuming statement (a), we preliminarily show that the two nonempty sets $Z_i = \{z \in C \mid \xi_i(z) = 0\}$ ($i = 1, 2$) actually coincide. Indeed, suppose by contradiction that $Z_1 \ne Z_2$. We can assume with no restriction that there is some $z \in C$ such that $\xi_1(z) = 0$ and $\xi_2(z) \ne 0$, hence $\xi_2(z) > 0$ since $\mathcal{D}(\xi_1) = \mathcal{D}(\xi_2) =: \mathcal{D}$. Picking any $x \in \mathcal{D}$, we have $\xi_1(\lambda x + (1 - \lambda)z) = \lambda\xi_1(x) < 0$, or, equivalently, $\xi_2(\lambda x + (1 - \lambda)z) < 0$ for all $\lambda \in (0, 1)$, that contradicts continuity of the mapping $\lambda \mapsto \xi_2(\lambda x + (1 - \lambda)z)$ at $\lambda = 0$.

We next claim that there is $z_0 \in \mathrm{ri}(C)$ such that $\xi_i(z_0) = 0$ for all $i = 1, 2$. To this aim, let $y \in C_0 \cap \mathrm{ri}(C)$. Then, for all $i = 1, 2$, either $\xi_i(y) = 0$ and we are done, or $\xi_i(y) > 0$. In the latter case, again by a continuity argument, for any $x \in \mathcal{D}$ there is some $\lambda \in (0, 1)$ such that $\xi_1(\lambda x + (1 - \lambda)y) = \lambda\xi_1(x) + (1 - \lambda)\xi_1(y) = 0$. Setting $z_0 = \lambda x + (1 - \lambda)y$, we thus see that $z_0 \in Z_1 = Z_2$, and $z_0 \in \mathrm{ri}(C)$ by [1, Thm. 6.1].

For $i = 1, 2$, let $\tilde{\xi}_i$ be the extension of ξ_i to an affine map on $\mathcal{M}(C)$. Then, the mapping $\mathcal{V}(C) \ni v \mapsto \tilde{\xi}_1(z_0 + v) \in \mathbb{R}$ is linear and nonzero, hence there exists a linear basis $\{v_1, \ldots, v_m\}$ of $\mathcal{V}(C)$ such that $\tilde{\xi}_1(z_0 + v_k) = 0$ if $k = 1, \ldots, m - 1$, and $\tilde{\xi}_1(z_0 + v_m) < 0$. By possibly replacing all the v_k's with μv_k for some $\mu \in (0, 1)$, we can assume that $z_k := z_0 + v_k \in \mathrm{ri}(C)$ for all $k = 1, \ldots, m$, and so $\xi_1(z_k) = 0$ if $k \le m - 1$, and $\xi_1(z_m) < 0$. Hence, also $\xi_2(z_k) = 0$ if $k \le m - 1$, and $\xi_2(z_m) < 0$. It follows that

$$\left[\tilde{\xi}_1 - \frac{\xi_1(z_m)}{\xi_2(z_m)}\tilde{\xi}_2\right](z_k) = 0 \qquad \forall k = 0, \ldots, m,$$

which implies $\tilde{\xi}_1 - (\xi_1(z_m)/\xi_2(z_m))\tilde{\xi}_2 = 0$ because any element of $\mathcal{M}(C)$ is an affine combination of $z_0, \ldots, z_m$. This yields statement (b).

(b) $\Rightarrow$ (a): The implication is clear.

(b) $\Leftrightarrow$ (c): Suppose $v_i^* \in \mathcal{V}^*$ and $\delta_i \in \mathbb{R}$ are as in statement (c), and extend the c-affine map ξ_i to an affine map $\tilde{\xi}_i$ on $\mathcal{M}(C)$ by means of the relation $\tilde{\xi}_i(x) = \delta_i - \langle v_i^*, x \rangle \quad \forall x \in \mathcal{M}(C)$. Then, since affine extensions are unique, statement (b) is equivalent to $\tilde{\xi}_1 = \alpha \tilde{\xi}_2$. Picking any $v_0 \in \mathcal{M}(C)$, this is in turn equivalent to

$$0 = (\tilde{\xi}_i - \alpha \tilde{\xi}_i)(v_0 + v) = [(\delta_1 - \alpha \delta_2) - \langle v_1^* - \alpha v_2^*, v_0 \rangle] - \langle v_1^* - \alpha v_2^*, v \rangle \qquad \forall v \in \mathcal{V}(C),$$

that is the same as statement (c). $\qquad\qquad\square$

II. INCOMPATIBILITY WITNESSES

We fix a finite dimensional complex Hilbert space $\mathcal{H}$, with $\dim \mathcal{H} = d$. We denote by $\mathcal{L}_s(\mathcal{H})$ the real linear space of all selfadjoint operators on $\mathcal{H}$, endowed with the uniform operator norm $\|\cdot\|$. We write $\mathbb{1}$ for the identity operator. If Z is a set, we let $|Z|$ be its cardinality. A *measurement* with outcomes in a finite set Z is any map $\mathsf{M} : Z \to \mathcal{L}_s(\mathcal{H})$ such that $\mathsf{M}(z) \geq 0$ for all $z \in Z$ and $\sum_{z \in Z} \mathsf{M}(z) = \mathbb{1}$. The *uniform measurement* with outcomes in Z is given by $\mathsf{U}_Z(z) = \mathbb{1}/|Z|$ for all $z \in Z$.

All measurements with outcomes in Z constitute a closed and bounded convex subset $\mathcal{O}(Z)$ in the real linear space of all operator valued functions $H : Z \to \mathcal{L}_s(\mathcal{H})$. We denote by $\mathcal{F}(Z)$ the latter linear space of functions, and we regard it as a normed space with the sup-norm $\|H\|_\infty = \max\{\|H(z)\| \mid z \in Z\}$; the dimension of $\mathcal{F}(Z)$ is $d^2|Z|$. For any $A \in \mathcal{L}_s(\mathcal{H})$, we define the affine set $\mathcal{F}_A(Z) = \{H \in \mathcal{F}(Z) \mid \sum_{z \in Z} H(z) = A\}$; the inclusion $\mathcal{O}(Z) \subset \mathcal{F}_{\mathbb{1}}(Z)$ is clear.

If X and Y are finite sets, two measurements $\mathsf{A} \in \mathcal{O}(X)$ and $\mathsf{B} \in \mathcal{O}(Y)$ are *compatible* if there exists a third measurement $\mathsf{M} \in \mathcal{O}(X \times Y)$ such that A and B are the *margins* of M, that is,

$$\mathsf{M}_X(x) := \sum_{y \in Y} \mathsf{M}(x, y) = \mathsf{A}(x) \qquad \text{and} \qquad \mathsf{M}_Y(y) := \sum_{x \in X} \mathsf{M}(x, y) = \mathsf{B}(y) \qquad\qquad \text{(S2)}$$

for all x and y. In this case, we say that M is a *joint measurement* of A and B. We denote by $\mathcal{O}_{X,Y} = \mathcal{O}(X) \times \mathcal{O}(Y)$ the set of all pairs of measurements on X and Y, and by $\mathcal{O}_{X,Y}^{\mathrm{com}}$ the subset of all pairs made up of compatible measurements. If $(\mathsf{A}, \mathsf{B}) \in \overline{\mathcal{O}_{X,Y}^{\mathrm{com}}} = \mathcal{O}_{X,Y} \smallsetminus \mathcal{O}_{X,Y}^{\mathrm{com}}$, the two measurements A and B are *incompatible*.

As it is well known, unless $\mathcal{H} = \mathbb{C}$ or $\min\{|X|, |Y|\} = 1$, the inclusion $\mathcal{O}_{X,Y}^{\mathrm{com}} \subset \mathcal{O}_{X,Y}$ is strict. So, in the following we will always assume $d \geq 2$ and $\min\{|X|, |Y|\} \geq 2$ to avoid degeneracies.

The sets $\mathcal{O}_{X,Y}$ and $\mathcal{O}_{X,Y}^{\mathrm{com}}$ are convex and compact in the direct product linear space $\mathcal{V} = \mathcal{F}(X) \times \mathcal{F}(Y)$; here, as the norm of $\mathcal{V}$ we choose the ℓ_∞-norm $\|(F, G)\|_\infty = \max\{\|F\|_\infty, \|G\|_\infty\}$. Indeed, only the compactness of $\mathcal{O}_{X,Y}^{\mathrm{com}}$ needs to be checked; it follows by the compactness of $\mathcal{O}(X \times Y)$ and the continuity of the mapping $\mathcal{O}(X \times Y) \ni \mathsf{M} \mapsto (\mathsf{M}_X, \mathsf{M}_Y) \in \mathcal{O}(X) \times \mathcal{O}(Y)$.

The next proposition gives some further insight into the convex structure of the sets $\mathcal{O}_{X,Y}$ and $\mathcal{O}_{X,Y}^{\mathrm{com}}$.

Proposition S5. *The following properties hold.*

(a) $\mathcal{M}(\mathcal{O}_{X,Y}^{\mathrm{com}}) = \mathcal{M}(\mathcal{O}_{X,Y}) = \mathcal{F}_{\mathbb{1}}(X) \times \mathcal{F}_{\mathbb{1}}(Y)$.

(b) $\mathcal{V}(\mathcal{O}_{X,Y}^{\mathrm{com}}) = \mathcal{V}(\mathcal{O}_{X,Y}) = \mathcal{F}_0(X) \times \mathcal{F}_0(Y)$.

(c) $\mathrm{ri}(\mathcal{O}_{X,Y}^{\mathrm{com}}) \subset \mathrm{ri}(\mathcal{O}_{X,Y})$ *and* $(\mathsf{U}_X, \mathsf{U}_Y) \in \mathrm{ri}(\mathcal{O}_{X,Y}^{\mathrm{com}})$.

Proof. Clearly,

$$\mathcal{M}(\mathcal{O}_{X,Y}^{\mathrm{com}}) \subseteq \mathcal{M}(\mathcal{O}_{X,Y}) \subseteq \mathcal{F}_{\mathbb{1}}(X) \times \mathcal{F}_{\mathbb{1}}(Y) = (\mathsf{U}_X, \mathsf{U}_Y) + \mathcal{F}_0(X) \times \mathcal{F}_0(Y). \qquad (*)$$

In order to prove that the previous inclusions actually are equalities (item (a)), it is enough to show that

$$(\mathsf{U}_X, \mathsf{U}_Y) + \varepsilon B \subseteq \mathcal{O}_{X,Y}^{\mathrm{com}} \quad \text{for} \quad \varepsilon = \frac{1}{2} \min\left\{\frac{1}{|X|}, \frac{1}{|Y|}\right\} \quad \text{and} \quad B = \{(F, G) \in \mathcal{F}_0(X) \times \mathcal{F}_0(Y) \mid \|(F, G)\|_\infty < 1\}; \quad (**)$$

here, B is the open unit ball in $\mathcal{F}_0(X) \times \mathcal{F}_0(Y)$. Indeed, if $(**)$ holds, then

$$(\mathsf{U}_X, \mathsf{U}_Y) + \mathcal{F}_0(X) \times \mathcal{F}_0(Y) = \mathcal{M}((\mathsf{U}_X, \mathsf{U}_Y) + \varepsilon B) \subseteq \mathcal{M}(\mathcal{O}_{X,Y}^{\mathrm{com}}).$$

Now, $(**)$ immediately follows, since for $(F, G) \in B$, the formula

$$\mathsf{M}(x, y) = \mathsf{U}_{X \times Y}(x, y) + \varepsilon \left[\frac{1}{|Y|} F(x) + \frac{1}{|X|} G(y) \right]$$

defines an element $\mathsf{M} \in \mathcal{O}(X \times Y)$ such that $(\mathsf{M}_X, \mathsf{M}_Y) = (\mathsf{U}_X, \mathsf{U}_Y) + \varepsilon(F, G)$. Having proven that in $(*)$ all inclusions actually are equalities, item (b) is obvious, while item (c) follows from $(**)$ and [1, Cor. 6.5.2]. $\qquad\square$

The dual space $\mathcal{V}^*$ of $\mathcal{V} = \mathcal{F}(X) \times \mathcal{F}(Y)$ can be identified with $\mathcal{V}$ itself by means of the pairing

$$\langle (F_1, G_1), (F_2, G_2) \rangle = \sum_{x \in X} \text{tr} \left[F_1(x) F_2(x) \right] + \sum_{y \in Y} \text{tr} \left[G_1(y) G_2(y) \right]. \tag{S3}$$

With this identification, Prop. S5.(b) and a simple dimension counting lead to the equalities

$$\mathcal{V}(\mathcal{O}_{X,Y}^{\text{com}})^\perp = \mathcal{V}(\mathcal{O}_{X,Y})^\perp = \{ (F, G) \in \mathcal{F}(X) \times \mathcal{F}(Y) \mid F(x_1) = F(x_2)$$
$$\text{and} \quad G(y_1) = G(y_2) \quad \forall x_1, x_2 \in X, \, y_1, y_2 \in Y \}. \tag{S4}$$

We recall from the main paper that any $\overline{\mathcal{O}_{X,Y}^{\text{com}}}$-witness on the convex set $\mathcal{O}_{X,Y}$ is an *incompatibility witness* (IW). Thus, by Prop. S2 and (S3), any IW is of the form

$$\xi(\mathsf{A}, \mathsf{B}) = \delta - \langle (F, G), (\mathsf{A}, \mathsf{B}) \rangle \qquad \forall (\mathsf{A}, \mathsf{B}) \in \mathcal{O}_{X,Y}$$

for some $(F, G) \in \mathcal{F}(X) \times \mathcal{F}(Y)$ and $\delta \in [\delta_{\mathcal{O}_{X,Y}^{\text{com}}}(F, G), \delta_{\mathcal{O}_{X,Y}}(F, G))$. In particular, ξ is tight if in the above formula $\delta = \delta_{\mathcal{O}_{X,Y}^{\text{com}}}(F, G)$.

Combining Props. S3 and S5.(b) immediately proves the following connection between tight IWs and the relative boundary of the set $\mathcal{O}_{X,Y}^{\text{com}}$.

Proposition S6. *Suppose ξ is a tight IW on the set $\mathcal{O}_{X,Y}$ and $(\mathsf{A}, \mathsf{B}) \in \mathcal{O}_{X,Y}^{\text{com}}$. Then, the equality $\xi(\mathsf{A}, \mathsf{B}) = 0$ implies that $(\mathsf{A}, \mathsf{B}) \in \partial \mathcal{O}_{X,Y}^{\text{com}}$. Moreover, there always exists some extreme point $(\mathsf{A}_0, \mathsf{B}_0)$ of $\mathcal{O}_{X,Y}^{\text{com}}$ such that $\xi(\mathsf{A}_0, \mathsf{B}_0) = 0$.*

As a consequence of Prop. S5.(c), also the hypothesis of Prop. S4 is satisfied by the sets $\mathcal{O}_{X,Y}$ and $\mathcal{O}_{X,Y}^{\text{com}}$. We then obtain the following characterization of detection equivalence for two IWs.

Proposition S7. *If ξ_1, ξ_2 are two IWs on the set $\mathcal{O}_{X,Y}$ and $\xi_i(\mathsf{A}, \mathsf{B}) = \delta_i - \langle (F_i, G_i), (\mathsf{A}, \mathsf{B}) \rangle$ for all $(\mathsf{A}, \mathsf{B}) \in \mathcal{O}_{X,Y}$ and $i = 1, 2$, then $\xi_1 \approx \xi_2$ if and only if*

$$F_2(x) = \alpha F_1(x) + A, \qquad G_2(y) = \alpha G_1(y) + B, \qquad \delta_2 = \alpha \delta_1 + \text{tr} \left[A + B \right]$$

for some $\alpha > 0$ and $A, B \in \mathcal{L}_s(\mathcal{H})$.

Proof. By Props. S4 and S5.(c), the equivalence $\xi_1 \approx \xi_2$ holds if and only if, for some $\alpha > 0$, both the following conditions are satisfied:

- $(F_1, G_1) - \alpha(F_2, G_2) \in \mathcal{V}(\mathcal{O}_{X,Y}^{\text{com}})^\perp \iff$ by (S4), there exist $A, B \in \mathcal{L}_s(\mathcal{H})$ such that $F_1(x) - \alpha F_2(x) = A$ for all $x \in X$ and $G_1(y) - \alpha G_2(y) = B$ for all $y \in Y$;

- $\delta_1 - \alpha \delta_2 = \langle (F_1, G_1) - \alpha(F_2, G_2), (\mathsf{U}_X, \mathsf{U}_Y) \rangle \iff \delta_1 - \alpha \delta_2 = \text{tr} \left[A \right] + \text{tr} \left[B \right]$ with A and B given by the previous item.

This concludes the proof. $\qquad\qquad\square$

We have already seen in tha main paper that, up to detection equivalence, any tight IW can be associated with a state discrimination problem with postmeasurement information. This problem consists in discriminating the classical labels of some *partitioned state ensemble*, that is, a couple $\hat{\mathcal{E}} = (\mathcal{E}, \{X, Y\})$, in which

(i) X and Y are disjoint finite sets, and

(ii) $\mathcal{E}$ is a *state ensemble* with label set $X \cup Y$, i.e., an element of $\mathcal{F}(X \cup Y)$ such that $\mathcal{E}(z) \geq 0$ for all $z \in X \cup Y$ and $\sum_{z \in X \cup Y} \text{tr} \left[\mathcal{E}(z) \right] = 1$.

Notice that the pairing (S3) for the restrictions $F_1 = \mathcal{E}|_X$ and $G_1 = \mathcal{E}|_Y$ rewrites

$$\langle (\mathcal{E}|_X, \mathcal{E}|_Y), (\mathsf{A}, \mathsf{B}) \rangle = P_{\text{guess}}^{\text{prior}}(\hat{\mathcal{E}}; \mathsf{A}, \mathsf{B}) \qquad \forall (\mathsf{A}, \mathsf{B}) \in \mathcal{O}_{X,Y}, \tag{S5}$$

where $P_{\text{guess}}^{\text{prior}}(\hat{\mathcal{E}}; \mathsf{A}, \mathsf{B})$ is the guessing probability with premeasurement information defined in (4) of the main text. It follows that

$$\delta_{\mathcal{O}_{X,Y}}(\mathcal{E}|_X, \mathcal{E}|_Y) = P_{\text{guess}}^{\text{prior}}(\hat{\mathcal{E}}), \qquad\qquad \delta_{\mathcal{O}_{X,Y}^{\text{com}}}(\mathcal{E}|_X, \mathcal{E}|_Y) = P_{\text{guess}}^{\text{post}}(\hat{\mathcal{E}}), \tag{S6}$$

where $P_{\text{guess}}^{\text{prior}}(\hat{\mathcal{E}})$ and $P_{\text{guess}}^{\text{post}}(\hat{\mathcal{E}})$ are the optimal guessing probabilities with premeasurement and postmeasurement information given by (5) and (6) of the paper. In particular, whenever the strict inequality $P_{\text{guess}}^{\text{post}}(\hat{\mathcal{E}}) < P_{\text{guess}}^{\text{prior}}(\hat{\mathcal{E}})$ holds, we can define the tight IW associated with the partitioned state ensemble $\hat{\mathcal{E}}$ as we did in (7):

$$\xi_{\hat{\mathcal{E}}}(\mathsf{A},\mathsf{B}) = P_{\text{guess}}^{\text{post}}(\hat{\mathcal{E}}) - P_{\text{guess}}^{\text{prior}}(\hat{\mathcal{E}};\mathsf{A},\mathsf{B}) \qquad \forall (\mathsf{A},\mathsf{B}) \in \mathcal{O}_{X,Y}. \tag{S7}$$

We remark that the evaluation of $P_{\text{guess}}^{\text{prior}}(\hat{\mathcal{E}})$ consists in solving two separate standard state discrimination problems: one for the subensemble $\mathcal{E}_X = p(X)^{-1}\,\mathcal{E}|_X$ and another one for $\mathcal{E}_Y = p(Y)^{-1}\,\mathcal{E}|_Y$. (Note: Although we use similar notations for subensembles and margin measurements, their correct interpretation is always clear from the context.) On the other hand, the optimization problem in the definition of $P_{\text{guess}}^{\text{post}}(\hat{\mathcal{E}})$ can be turned into a single standard state discrimination problem by means of [3, Thm. 2]. Therefore, in order to evaluate both probabilities in (S6), one can resort to techniques from standard quantum state discrimination, as those described e.g. in [3, Sec. IV B], [4–7].

If the partitioned state ensemble $\hat{\mathcal{E}}$ is not trivial, any compatible pair of measurements solving the optimization problem in the definition of $P_{\text{guess}}^{\text{post}}(\hat{\mathcal{E}})$ necessarily lies on the relative boundary of the set $\mathcal{O}_{X,Y}^{\text{com}}$. More precisely, we have the following fact. It should be compared with the detailed characterization of the relative boundary and the extreme points of the compact convex set $\mathcal{O}_{X,Y}^{\text{com}}$ provided in [8].

Proposition S8. *Suppose the partitioned state ensemble $\hat{\mathcal{E}} = (\mathcal{E},\{X,Y\})$ is such that either one of the restrictions $\mathcal{E}|_X$ or $\mathcal{E}|_Y$ is not a constant function. Then, for two compatible measurements A and B, the equality $P_{\text{guess}}^{\text{post}}(\hat{\mathcal{E}}) = P_{\text{guess}}^{\text{prior}}(\hat{\mathcal{E}};\mathsf{A},\mathsf{B})$ entails that $(\mathsf{A},\mathsf{B}) \in \partial\mathcal{O}_{X,Y}^{\text{com}}$. Moreover, there always exists some extreme point $(\mathsf{A}_0,\mathsf{B}_0)$ of $\mathcal{O}_{X,Y}^{\text{com}}$ such that $P_{\text{guess}}^{\text{post}}(\hat{\mathcal{E}}) = P_{\text{guess}}^{\text{prior}}(\hat{\mathcal{E}};\mathsf{A}_0,\mathsf{B}_0)$.*

Proof. By combining Lemma S1 and (S5), (S6), the second claim is always true, while the first one holds whenever $(\mathcal{E}|_X\,,\,\mathcal{E}|_Y)$ is not an element of $\mathcal{V}(\mathcal{O}_{X,Y}^{\text{com}})^{\perp}$. By (S4), this is equivalent to either $\mathcal{E}|_X$ or $\mathcal{E}|_Y$ being not constant. $\square$

Note that the last proposition does not require $P_{\text{guess}}^{\text{post}}(\hat{\mathcal{E}}) < P_{\text{guess}}^{\text{prior}}(\hat{\mathcal{E}})$. Indeed, if two compatible measurements A and B attain both the equalities $P_{\text{guess}}^{\text{post}}(\hat{\mathcal{E}}) = P_{\text{guess}}^{\text{prior}}(\hat{\mathcal{E}}) = P_{\text{guess}}^{\text{prior}}(\hat{\mathcal{E}};\mathsf{A},\mathsf{B})$, then actually $(\mathsf{A},\mathsf{B}) \in \partial\mathcal{O}_{X,Y}^{\text{com}} \cap \partial\mathcal{O}_{X,Y}$ by an easy extension of the argument in the last proof.

Remarkably, as a consequence of Prop. S8, whenever there is a unique pair of compatible measurements A and B attaining the equality $P_{\text{guess}}^{\text{post}}(\hat{\mathcal{E}}) = P_{\text{guess}}^{\text{prior}}(\hat{\mathcal{E}};\mathsf{A},\mathsf{B})$, then (A,B) is necessarily an extreme point of the compact convex set $\mathcal{O}_{X,Y}^{\text{com}}$; some examples can be found in [3, Secs. V and VI].

III. INCOMPATIBILITY WITNESSES WITH TWO MUTUALLY UNBIASED BASES

In this section, we suppose $\{\varphi_h \mid h = 1,\dots,d\}$ and $\{\psi_k \mid k = 1,\dots,d\}$ are two fixed mutually unbiased bases (MUB) of the d-dimensional Hilbert space $\mathcal{H}$. We will show how these bases can be used to construct a family of IWs for pairs of measurements with outcomes $X = \{(h,\varphi) \mid h = 1,\dots,d\}$ and $Y = \{(k,\psi) \mid k = 1,\dots,d\}$. Moreover, as a byproduct of this construction, we will also characterize the amount of uniform noise that is needed in order to make the two given MUB compatible. Note that, although when one regards X and Y as separate sets, the extra symbols φ and ψ are clearly redundant, nonetheless they are needed to define the disjoint union $Z = X \cup Y$, and then consider the partition $\{X,Y\}$ of Z.

A. Construction of the IWs

For all $\boldsymbol{\mu} = (\mu_{\varphi},\mu_{\psi}) \in [0,1]\times[0,1]$, we can define the state ensemble $\mathcal{E}_{\boldsymbol{\mu}}$ with label set Z, given as

$$\mathcal{E}_{\boldsymbol{\mu}}(j,\ell) = \frac{1}{2d}\left[\mu_{\ell}|\ell_j\rangle\langle\ell_j| + (1-\mu_{\ell})\frac{1}{d}\,\mathbb{1}\right] \qquad \forall j \in \{1,\dots,d\},\, \ell \in \{\varphi,\psi\}, \tag{S8}$$

and the corresponding partitioned state ensemble $\hat{\mathcal{E}}_{\boldsymbol{\mu}} = (\mathcal{E},\{X,Y\})$. For this state ensemble, all labels $z \in Z$ occur with the same probability $p(z) = \text{tr}\,[\mathcal{E}_{\boldsymbol{\mu}}(z)] = 1/(2d)$; moreover, $p(X) = p(Y) = 1/2$ are the probabilities that a label occur in the set X or Y, respectively. The subensemble $\mathcal{E}_{\boldsymbol{\mu},X}(x) = (1/p(X))\mathcal{E}_{\boldsymbol{\mu}}(x)\ \ \forall x \in X$ is given by

$$\mathcal{E}_{\boldsymbol{\mu},X}(h,\varphi) = \frac{1}{d}\left[\mu_{\varphi}|\varphi_h\rangle\langle\varphi_h| + (1-\mu_{\varphi})\frac{1}{d}\,\mathbb{1}\right], \tag{S9}$$

and it corresponds to a situation in which Alice randomly picks any of the vectors within the first basis with equal probability, and then she adds it uniform noise with intensity $1 - \mu_\varphi$. A similar interpretation holds for the subensemble $\mathcal{E}_{\boldsymbol{\mu},Y}$, with a possibly different choice of the noise parameter μ_ψ.

In order to obtain some more inequivalent IWs, we can let either μ_φ or μ_ψ take even slightly negative values, in which case $\mathcal{E}_{\boldsymbol{\mu},X}$ and $\mathcal{E}_{\boldsymbol{\mu},Y}$ can no longer be interpreted as noisy ensembles. The most general choice still making (S8) define a state ensemble is indeed $\boldsymbol{\mu} = [1/(1-d),1] \times [1/(1-d),1]$. In the following, we always use the latter values, and we further assume $\boldsymbol{\mu} \neq (0,0)$ to avoid the trivial case.

In order to use the state ensemble $\mathcal{E}_{\boldsymbol{\mu}}$ for constructing a tight IW as in (S7), first of all we need to evaluate the pre- and postmeasurement guessing probabilities $P_{\mathrm{guess}}^{\mathrm{prior}}(\hat{\mathcal{E}}_{\boldsymbol{\mu}})$ and $P_{\mathrm{guess}}^{\mathrm{post}}(\hat{\mathcal{E}}_{\boldsymbol{\mu}})$. To this aim, we recall the following two useful results.

Proposition S9 (Prop. 2 of [3])**.** *Suppose $\mathcal{E}$ is a state ensemble with label set X. For all $x \in X$, denote by $\lambda(x)$ the largest eigenvalue of $\mathcal{E}(x)$, and by $\Pi(x)$ the orthogonal projection onto the $\lambda(x)$-eigenspace of $\mathcal{E}(x)$. Define*

$$\lambda_\mathcal{E} = \max_{x \in X} \lambda(x), \qquad X_\mathcal{E} = \{x \in X : \lambda(x) = \lambda_\mathcal{E}\}. \tag{S10}$$

Then, if there exists $\nu \in \mathbb{R}$ such that

$$\sum_{x \in X_\mathcal{E}} \Pi(x) = \nu \mathbb{1}, \tag{S11}$$

we have the following consequences:

(a) $\nu = \frac{1}{d} \sum_{x \in X_\mathcal{E}} \mathrm{rank}\,\Pi(x)$;

(b) $P_{\mathrm{guess}}(\mathcal{E}) = d\lambda_\mathcal{E}$;

(c) a (not necessarily unique) measurement M_0 attaining the maximum guessing probability $P_{\mathrm{guess}}(\mathcal{E})$ is

$$\mathsf{M}_0(x) = \begin{cases} \nu^{-1}\Pi(x) & \text{if } x \in X_\mathcal{E} \\ 0 & \text{if } x \notin X_\mathcal{E} \end{cases}.$$

Theorem S1 (Thm. 2 of [3])**.** *For any partitioned state ensemble $\hat{\mathcal{E}} = (\mathcal{E}, \{X,Y\})$, we have*

$$P_{\mathrm{guess}}^{\mathrm{post}}(\hat{\mathcal{E}}) = (|X|\,p(Y) + |Y|\,p(X))\,P_{\mathrm{guess}}(\mathcal{F}),$$

where $\mathcal{F}$ is the state ensemble having the Cartesian product $X \times Y$ as its label set, and given by

$$\mathcal{F}(x,y) = \frac{\mathcal{E}(x) + \mathcal{E}(y)}{|X|\,p(Y) + |Y|\,p(X)} \qquad \forall (x,y) \in X \times Y. \tag{S12}$$

Moreover, for a measurement $\mathsf{M} : X \times Y \to \mathcal{L}_s(\mathcal{H})$, we have the equivalence

$$P_{\mathrm{guess}}(\mathcal{F}; \mathsf{M}) = P_{\mathrm{guess}}(\mathcal{F}) \quad \Leftrightarrow \quad P_{\mathrm{guess}}^{\mathrm{prior}}(\hat{\mathcal{E}}; \mathsf{M}_X, \mathsf{M}_Y) = P_{\mathrm{guess}}^{\mathrm{post}}(\hat{\mathcal{E}}),$$

where M_X and M_Y are the two margin measurements of M defined in (S2).

We can immediately use Prop. S9 to evaluate the optimal guessing probability for the subensemble $\mathcal{E}_{\boldsymbol{\mu},X}$. Indeed, the largest eigenvalue of (S9) is

$$\lambda(h,\varphi) = \begin{cases} \dfrac{1}{d^2}\left[1 + (d-1)\mu_\varphi\right] & \text{if } \mu_\varphi \geq 0 \\[2mm] \dfrac{1}{d^2}\left(1 - \mu_\varphi\right) & \text{if } \mu_\varphi < 0 \end{cases}$$

$$= \frac{1}{d^2}\left\{1 + \frac{1}{2}\left[(d-2)\mu_\varphi + d\,|\mu_\varphi|\right]\right\} \qquad \forall h = 1,\dots,d.$$

Thus, in (S10), the overall largest eigenvalue $\lambda_{\mathcal{E}_{\boldsymbol{\mu},X}}$ is given by the latter expression, and it is attained on the whole label set $X_{\mathcal{E}_{\boldsymbol{\mu},X}} = X$. Eq. (S11) is then easily verified, so that Prop. S9 can be applied. Hence, Prop. S9.(b) yields

$$P_{\mathrm{guess}}(\mathcal{E}_{\boldsymbol{\mu},X}) = \frac{1}{d}\left\{1 + \frac{1}{2}\left[(d-2)\mu_\varphi + d\,|\mu_\varphi|\right]\right\}.$$

A similar formula holds also for $P_{\text{guess}}(\mathcal{E}_{\boldsymbol{\mu},Y})$. We finally obtain

$$P_{\text{guess}}^{\text{prior}}(\hat{\mathcal{E}}_{\boldsymbol{\mu}}) = p(X)P_{\text{guess}}(\mathcal{E}_{\boldsymbol{\mu},X}) + p(Y)P_{\text{guess}}(\mathcal{E}_{\boldsymbol{\mu},Y})$$

$$= \frac{1}{4}\left[2 + |\mu_\varphi| + |\mu_\psi| + \left(1 - \frac{2}{d}\right)(\mu_\varphi + \mu_\psi - 2)\right]. \tag{S13}$$

Now we tackle the more difficult problem of evaluating $P_{\text{guess}}^{\text{post}}(\hat{\mathcal{E}}_{\boldsymbol{\mu}})$. In order to apply Thm. S1, we first write the auxiliary state ensemble (S12), which is

$$\mathcal{F}((h,\varphi),(k,\psi)) = \frac{1}{2d^2}\left[\mu_\varphi|\varphi_h\rangle\langle\varphi_h| + \mu_\psi|\psi_k\rangle\langle\psi_k| + \frac{1}{d}(2 - \mu_\varphi - \mu_\psi)\,\mathbb{1}\right] \tag{S14}$$

for the partitioned state ensemble $\hat{\mathcal{E}}_{\boldsymbol{\mu}}$. Then, we check if we can apply Prop. S9 in order to calculate $P_{\text{guess}}(\mathcal{F})$. To this aim, we need to find the spectral decomposition of (S14) for all h, k. The next lemma is useful to this purpose.

Lemma S2. *Let $\varphi, \psi \in \mathcal{H}$ be two unit vectors such that $|\langle\varphi|\psi\rangle|^2 = 1/d$. Denote $Q = |\varphi\rangle\langle\varphi|$ and $P = |\psi\rangle\langle\psi|$, and let*

$$S = qQ + pP \quad \text{with} \quad q, p \in \mathbb{R} \quad \text{and} \quad (q,p) \neq (0,0). \tag{S15}$$

Then, the eigenvalues of the selfadjoint operator S are

$$\lambda_+ = \frac{1}{2}\left[(q+p) + \sqrt{q^2 + p^2 - 2\Delta qp}\right] \qquad \text{with multiplicity} \geq 1, \tag{S16}$$

$$\lambda_- = \frac{1}{2}\left[(q+p) - \sqrt{q^2 + p^2 - 2\Delta qp}\right] \qquad \text{with multiplicity} \geq 1, \tag{S17}$$

$$\lambda_0 = 0 \qquad \text{with multiplicity} \geq d - 2, \tag{S18}$$

where

$$\Delta = 1 - \frac{2}{d} \geq 0.$$

They satisfy the following inequalities:

- if $q \geq 0$ and $p \geq 0$, then $\lambda_+ > \lambda_- \geq 0$, with equality if and only if $qp = 0$;

- if either $q > 0$ and $p < 0$, or $q < 0$ and $p > 0$, then $\lambda_+ > 0 > \lambda_-$;

- if $q \leq 0$ and $p \leq 0$, then $0 \geq \lambda_+ > \lambda_-$, with equality if and only if $qp = 0$.

Moreover, the three selfadjoint operators

$$\Pi_+ = \frac{1}{\lambda_+ - \lambda_-}\left\{qQ + pP - \frac{d\lambda_-}{d-1}[Q + P - (QP + PQ)]\right\}, \tag{S19}$$

$$\Pi_- = \frac{1}{\lambda_- - \lambda_+}\left\{qQ + pP - \frac{d\lambda_+}{d-1}[Q + P - (QP + PQ)]\right\}, \tag{S20}$$

$$\Pi_0 = \mathbb{1} - \frac{d}{d-1}[Q + P - (QP + PQ)] \tag{S21}$$

constitute an orthogonal resolution of the identity, with $S\Pi_k = \Pi_k S = \lambda_k S$ for all $k \in \{+,-,0\}$, and $\operatorname{rank}\Pi_+ = \operatorname{rank}\Pi_- = 1$, $\operatorname{rank}\Pi_0 = d - 2$.

Proof. After noticing that the vectors $\{\varphi, \psi\}$ are linearly independent, we define the following two subspaces of $\mathcal{H}$ with respective dimensions 2 and $d - 2$

$$\mathcal{H}_1 = \operatorname{span}\{\varphi, \psi\}, \qquad \mathcal{H}_0 = \mathcal{H}_1^\perp.$$

Then, $S\mathcal{H}_1 \subseteq \mathcal{H}_1$ and $S\mathcal{H}_0 = \{0\}$. In particular, the eigenvalue 0 has multiplicity greater than or equal to $d - 2$ in S. The matrix form of the restriction $S_1 := S|_{\mathcal{H}_1}$ with respect to the (nonorthogonal) basis $\{\varphi, \psi\}$ of $\mathcal{H}_1$ is

$$S_1 = \begin{pmatrix} q & q\langle\varphi|\psi\rangle \\ p\langle\psi|\varphi\rangle & p \end{pmatrix}.$$

The roots of the characteristic polynomial of S_1 are the eigenvalues λ_+ and λ_- in (S16), (S17). Since the quadratic form $q^2 + p^2 - 2\Delta qp$ is positive definite, the square root in (S16) and (S17) is nonzero, hence $\lambda_+ > \lambda_-$. The remaining inequalities involving 0 are straightforward calculations. The multiplicities of λ_+ and λ_- in S can not be less than the respective multiplicities in S_1, which are 1. This completes the proof of the statements about the eigenvalues. Now, we claim that

$$\Pi_1 = \frac{d}{d-1}[Q + P - (QP + PQ)] \tag{$\sim$}$$

is the orthogonal projection onto $\mathcal{H}_1$. Indeed,

- $\Pi_1^* = \Pi_1$ (immediate);

- $\Pi_1|_{\mathcal{H}_0} = 0$ (immediate);

- $\Pi_1 Q = Q$ (because $Q^2 = Q$ and $QPQ = (1/d)Q$) $\Rightarrow$ $\Pi_1\varphi = \varphi$;

- $\Pi_1 P = P$ (because $P^2 = P$ and $PQP = (1/d)P$) $\Rightarrow$ $\Pi_1\psi = \psi$.

By applying the spectral theorem to S_1, there exist two rank-1 orthogonal projections Π_+ and Π_- defined on $\mathcal{H}$ and satisfying the relations

$$\Pi_+ + \Pi_- = \Pi_1\,, \qquad\qquad S = \lambda_+\Pi_+ + \lambda_-\Pi_-\,.$$

Inserting the expressions (S15) for S and ($\sim$) for Π_1 into these relations, and solving with respect to Π_+, Π_-, we get (S19) and (S20). Since the operator (S21) is the orthogonal projection $\Pi_0 = \mathbb{1} - \Pi_1$ onto $\mathcal{H}_0$, the proof of the last claim of the lemma is concluded. $\qquad\square$

The expression in the square root of (S16), (S17) is the quadratic form associated with the 2×2 matrix

$$G = \begin{pmatrix} 1 & -\Delta \\ -\Delta & 1 \end{pmatrix}.$$

Since G is positive definite, this quadratic form actually is the squared norm

$$\|(q,p)\|_G^2 = (q,p)G(q,p)^t = q^2 + p^2 - 2\Delta qp$$

defined by the Euclidean scalar product $\langle (q_1,p_1), (q_2,p_2) \rangle_G = (q_1,p_1)G(q_2,p_2)^t$ in $\mathbb{R}^2$. Remarkably, in the special case $d = 2$, we have $\Delta = 0$, hence $\langle \cdot, \cdot \rangle_G$ is the canonical Euclidean scalar product of $\mathbb{R}^2$. Also note that in this case, independently of the values of q and p, the largest eigenvalue of S is λ_+ with multiplicity 1, its associated eigenprojection is Π_+, and the orthogonal projection Π_0 is zero.

Assuming $d \geq 3$, only for $q > 0$ or $p > 0$ the largest eigenvalue of S is still given by λ_+; otherwise, for $q \leq 0$ and $p \leq 0$, the largest eigenvalue of S is 0. In the former case, the eigenvalue λ_+ has still multiplicity 1 and associated eigenprojection Π_+; in the latter case, the eigenvalue 0 has either multiplicity $d-2$ and associated eigenprojection Π_0 (subcase $qp \neq 0$), or multiplicity $d-1$ and associated eigenprojection $\Pi_0 + \Pi_+$ (subcase $qp = 0$).

As a consequence of the last two paragraphs, the spectral decomposition of the operators (S14) is different according to the cases:

(C1) $d = 2$ or $\max\{\mu_\varphi, \mu_\psi\} > 0$;

(C2) $d \geq 3$ and $\max\{\mu_\varphi, \mu_\psi\} < 0$;

(C3) $d \geq 3$ and $\max\{\mu_\varphi, \mu_\psi\} = 0$.

More precisely, all the operators $\{\mathcal{F}(x,y) \mid (x,y) \in X \times Y\}$ always have the same largest eigenvalue, which is

$$\lambda_{\mathcal{F}} = \begin{cases} \dfrac{1}{4d^2}\left[\left(1 - \dfrac{2}{d}\right)(\mu_\varphi + \mu_\psi) + \dfrac{4}{d} + \|\boldsymbol{\mu}\|_G\right] & \text{in case (C1)} \\[4mm] \dfrac{1}{2d^3}\left[2 - (\mu_\varphi + \mu_\psi)\right] & \text{in cases (C2)-(C3)} \end{cases}$$

by Lemma S2; in particular, the state ensemble $\mathcal{F}$ attains its largest eigenvalue on the whole label set $X \times Y$, that is, $(X \times Y)_{\mathcal{F}} = X \times Y$ with the notation of Prop. S9. Moreover, combining (S19) and (S21) according to the case at hand, the orthogonal projection onto the $\lambda_{\mathcal{F}}$-eigenspace of (S14) is

$$\Pi((h,\varphi),(k,\psi)) = a\mathbb{1} + b(\mu_\varphi\mathsf{Q}(h) + \mu_\psi\mathsf{P}(k)) - \frac{dc}{d-1}\left[\mathsf{Q}(h) + \mathsf{P}(k) - (\mathsf{Q}(h)\mathsf{P}(k) + \mathsf{P}(k)\mathsf{Q}(h))\right], \tag{S22a}$$

where

$$Q(h) = |\varphi_h\rangle\langle\varphi_h|, \qquad P(k) = |\psi_k\rangle\langle\psi_k| \qquad (S22b)$$

and

$$
\begin{array}{llll}
a = 0\,, & b = \dfrac{1}{\|\boldsymbol{\mu}\|_G}\,, & c = \dfrac{1}{2}\left(\dfrac{\mu_\varphi + \mu_\psi}{\|\boldsymbol{\mu}\|_G} - 1\right) & \text{in case (C1)} \\[2ex]
a = 1\,, & b = 0\,, & c = 1 & \text{in case (C2)} \\[1ex]
a = 1\,, & b = \dfrac{1}{\|\boldsymbol{\mu}\|_G}\,, & c = 0 & \text{in case (C3)}\,.
\end{array}
\qquad (S22c)
$$

Since $\sum_{h=1}^d Q(h) = \sum_{k=1}^d P(k) = \mathbb{1}$ and the coefficients a, b, c do not depend on h, k, we have

$$\sum_{(x,y)\in(X\times Y)_{\mathcal{F}}} \Pi(x,y) = \sum_{h,k=1}^{d} \Pi((h,\varphi),(k,\psi)) = \nu\mathbb{1}$$

where

$$\nu = d^2 a + db(\mu_\varphi + \mu_\psi) - 2dc = \begin{cases} d & \text{in case (C1)} \\ d(d-2) & \text{in case (C2)} \\ d(d-1) & \text{in case (C3)} \end{cases}. \qquad (S23)$$

In all cases, Eq. (S11) holds, hence we can apply Prop. S9 to determine $P_{\text{guess}}(\mathcal{F}) = d\lambda_{\mathcal{F}}$ and the optimal measurement $M_0(x,y) = \nu^{-1}\Pi(x,y) \;\; \forall (x,y) \in X\times Y$. Note that all projections $\Pi(x,y)$ have the same rank, hence rank $M_0(x,y) = \nu/d$ for all x, y by Prop. S9.(a).

Now that we have found $P_{\text{guess}}^{\text{post}}(\mathcal{F})$ and characterized a measurement attaining it, by making use of Thm. S1 we can translate these results into the state discrimination problem with postmeasurement information for the partitioned state ensemble $\hat{\mathcal{E}}_{\boldsymbol{\mu}}$. The maximal guessing probability in the postmeasurement scenario is $P_{\text{guess}}^{\text{post}}(\hat{\mathcal{E}}_{\boldsymbol{\mu}}) = dP_{\text{guess}}(\mathcal{F}) = d^2\lambda_{\mathcal{F}}$; it is attained on the compatible pair given by the margins of the measurement $M_0 = \nu^{-1}\Pi$, which are easily calculated from (S22) and the normalization of Q and P. We collect the conclusions in the next proposition.

Proposition S10. *For all $\boldsymbol{\mu} = (\mu_\varphi, \mu_\psi) \in [1/(1-d), 1] \times [1/(1-d), 1]$ with $\boldsymbol{\mu} \neq (0,0)$, we have*

$$
P_{\text{guess}}^{\text{post}}(\hat{\mathcal{E}}_{\boldsymbol{\mu}}) = \begin{cases} \dfrac{1}{4}\left[\left(1 - \dfrac{2}{d}\right)(\mu_\varphi + \mu_\psi) + \dfrac{4}{d} + \|\boldsymbol{\mu}\|_G\right] & \text{in case (C1)} \\[2ex] \dfrac{1}{2d}\left[2 - (\mu_\varphi + \mu_\psi)\right] & \text{in cases (C2)-(C3)} \end{cases}. \qquad (S24)
$$

Moreover, $P_{\text{guess}}^{\text{post}}(\hat{\mathcal{E}}_{\boldsymbol{\mu}}) = P_{\text{guess}}^{\text{prior}}(\hat{\mathcal{E}}_{\boldsymbol{\mu}}; M_{0,X}, M_{0,Y})$, where the measurement $M_0 = \nu^{-1}\Pi$ is expressed in terms of the constant ν defined in (S23) and the rank-(ν/d) orthogonal projections $\Pi(x,y)$ given by (S22) (which both depend on $\boldsymbol{\mu}$). Explicitly,

$$M_{0,X}(h,\varphi) = \gamma_\varphi Q(h) + (1 - \gamma_\varphi)\frac{1}{d}\mathbb{1}\,, \qquad M_{0,Y}(k,\psi) = \gamma_\psi P(k) + (1 - \gamma_\psi)\frac{1}{d}\mathbb{1}\,, \qquad (S25)$$

where

$$\gamma_\ell = \frac{d}{\nu}\left[b\mu_\ell - \frac{(d-2)c}{d-1}\right] = \begin{cases} \dfrac{d\mu_\ell - (d-2)(\mu_{\bar\ell} - \|\boldsymbol{\mu}\|_G)}{2(d-1)\|\boldsymbol{\mu}\|_G} & \text{in case (C1)} \\[2ex] \dfrac{1}{1-d}(1 - \delta_{\mu_\ell,0}) & \text{in cases (C2)-(C3)} \end{cases}. \qquad (S26)$$

In the last equation, $(\ell, \bar\ell)$ denotes either (φ, ψ) or (ψ, φ); moreover, $\delta_{\mu_\varphi,0}$ and $\delta_{\mu_\psi,0}$ are the usual Kronecker deltas.

Remarkably, in case (C2) there exists an optimal pair of compatible measurements which is unaffected by the value of $\boldsymbol{\mu}$; it is the pair (A, B) given by $A(h,\varphi) = [1/(1-d)]Q(h) + [1/(d-1)]\mathbb{1}$ and $B(k,\psi) = [1/(1-d)]P(h) + [1/(d-1)]\mathbb{1}$ for all h, k. Note that (A, B) is optimal also in case (C3), where the margins of M_0 however yield a different solution.

We can now determine the values of $\boldsymbol{\mu}$ for which $P_{\text{guess}}^{\text{post}}(\hat{\mathcal{E}}_{\boldsymbol{\mu}}) < P_{\text{guess}}^{\text{prior}}(\hat{\mathcal{E}}_{\boldsymbol{\mu}})$, and for these values explicitly evaluate the tight IW (S7) associated with $\hat{\mathcal{E}}_{\boldsymbol{\mu}}$. This yields the first main result of the present section.

Theorem S2 (Thm. 3 of the main paper). *Let $\boldsymbol{\mu} = (\mu_\varphi, \mu_\psi) \in [1/(1-d), 1] \times [1/(1-d), 1]$ with $\boldsymbol{\mu} \neq (0,0)$. Then, we have the strict inequality $P_{\mathrm{guess}}^{\mathrm{post}}(\hat{\mathcal{E}}_{\boldsymbol{\mu}}) < P_{\mathrm{guess}}^{\mathrm{prior}}(\hat{\mathcal{E}}_{\boldsymbol{\mu}})$ if and only if $\mu_\varphi \mu_\psi \neq 0$ and either $d = 2$ or $\max\{\mu_\varphi, \mu_\psi\} > 0$. In this case, the tight IW associated with the partitioned state ensemble $\hat{\mathcal{E}}_{\boldsymbol{\mu}}$ by means of (S7) is*

$$\xi_{\hat{\mathcal{E}}_{\boldsymbol{\mu}}}(\mathsf{A}, \mathsf{B}) = \frac{1}{4}(\mu_\varphi + \mu_\psi + \|\boldsymbol{\mu}\|_G) - \frac{1}{2d}\left\{\mu_\varphi \sum_{h=1}^{d} \mathrm{tr}\left[\mathsf{A}(h, \varphi)\mathsf{Q}(h)\right] + \mu_\psi \sum_{k=1}^{d} \mathrm{tr}\left[\mathsf{B}(k, \psi)\mathsf{P}(k)\right]\right\}. \tag{S27}$$

Finally, the ensembles $\hat{\mathcal{E}}_{\boldsymbol{\mu}}$ and $\hat{\mathcal{E}}_{\boldsymbol{\nu}}$ yield detection equivalent IWs if and only if $\boldsymbol{\nu} = \alpha\boldsymbol{\mu}$ for some $\alpha > 0$.

Proof. By (S13) and (S24),

$$P_{\mathrm{guess}}^{\mathrm{prior}}(\hat{\mathcal{E}}_{\boldsymbol{\mu}}) - P_{\mathrm{guess}}^{\mathrm{post}}(\hat{\mathcal{E}}_{\boldsymbol{\mu}}) = \begin{cases} \frac{1}{4}(|\mu_\varphi| + |\mu_\psi| - \|\boldsymbol{\mu}\|_G) & \text{in case (C1)} \\ 0 & \text{in cases (C2)-(C3)} \end{cases}.$$

The triangular inequality for the norm $\|\cdot\|_G$ implies that the expression for case (C1) strictly positive unless $\mu_\varphi = 0$ or $\mu_\psi = 0$. This proves the first claim.
Eq. (S27) then follows by combining (S24) with

$$P_{\mathrm{guess}}^{\mathrm{prior}}(\hat{\mathcal{E}}_{\boldsymbol{\mu}}; \mathsf{A}, \mathsf{B}) = p(X)P_{\mathrm{guess}}(\mathcal{E}_{\boldsymbol{\mu}, X}; \mathsf{A}) + p(Y)P_{\mathrm{guess}}(\mathcal{E}_{\boldsymbol{\mu}, Y}; \mathsf{B})$$
$$= \frac{1}{2d}\left\{\mu_\varphi \sum_h \mathrm{tr}\left[\mathsf{A}(h, \varphi)\mathsf{Q}(h)\right] + \mu_\psi \sum_k \mathrm{tr}\left[\mathsf{B}(k, \psi)\mathsf{P}(k)\right] + 2 - (\mu_\varphi + \mu_\psi)\right\}.$$

Finally, we prove the equivalence statement. If $\boldsymbol{\nu} = \alpha\boldsymbol{\mu}$ with $\alpha > 0$, then $\xi_{\hat{\mathcal{E}}_{\boldsymbol{\nu}}} = \alpha\xi_{\hat{\mathcal{E}}_{\boldsymbol{\mu}}}$ by (S27), hence the two witnesses are detection equivalent. Conversely, if $\xi_{\hat{\mathcal{E}}_{\boldsymbol{\mu}}} \approx \xi_{\hat{\mathcal{E}}_{\boldsymbol{\nu}}}$, then by Prop. S7 there exist $\alpha > 0$ and $A, B \in \mathcal{L}_s(\mathcal{H})$ such that

$$\mathcal{E}_{\boldsymbol{\nu}}(h, \varphi) = \alpha\mathcal{E}_{\boldsymbol{\mu}}(h, \varphi) + A \quad \forall h \quad \text{and} \quad \mathcal{E}_{\boldsymbol{\nu}}(k, \psi) = \alpha\mathcal{E}_{\boldsymbol{\mu}}(k, \psi) + B \quad \forall k.$$

In particular, $\mu_\ell = 0 \Leftrightarrow \nu_\ell = 0$, as any of the two equalities imply that both the maps $\mathcal{E}_{\boldsymbol{\mu}}(\cdot, \ell)$ and $\mathcal{E}_{\boldsymbol{\nu}}(\cdot, \ell)$ are constant. Since

$$\mathcal{E}_{\boldsymbol{\nu}}(j, \ell) = \frac{\nu_\ell}{\mu_\ell}\mathcal{E}_{\boldsymbol{\mu}}(j, \ell) + \frac{1}{2d^2}\left(1 - \frac{\nu_\ell}{\mu_\ell}\right)\mathbb{1} \qquad \forall j, \ell \qquad \left(\text{with } \frac{0}{0} := 0\right),$$

it must be $\alpha = \nu_\varphi/\mu_\varphi = \nu_\psi/\mu_\psi$ and $A = B = [(1-\alpha)/(2d^2)]\mathbb{1}$. In particular, $\boldsymbol{\nu} = \alpha\boldsymbol{\mu}$. $\qquad\square$

B. Incompatibility of noisy MUB

As a byproduct, the evaluation of $P_{\mathrm{guess}}^{\mathrm{post}}(\hat{\mathcal{E}}_{\boldsymbol{\mu}})$ and the solution to the related optimization problem provided in the last subsection allow to characterize the compatibility of the following pair of measurements $(\mathsf{A}_{\boldsymbol{\gamma}}, \mathsf{B}_{\boldsymbol{\gamma}}) \in \mathcal{O}_{X,Y}$

$$\mathsf{A}_{\boldsymbol{\gamma}}(h, \varphi) = \gamma_\varphi \mathsf{Q}(h) + (1 - \gamma_\varphi)\frac{1}{d}\mathbb{1}, \qquad \mathsf{B}_{\boldsymbol{\gamma}}(k, \psi) = \gamma_\psi \mathsf{P}(k) + (1 - \gamma_\psi)\frac{1}{d}\mathbb{1}, \tag{S28}$$

where $\boldsymbol{\gamma} = (\gamma_\varphi, \gamma_\psi) \in [1/(1-d), 1] \times [1/(1-d), 1]$. This set constitues all the values of the parameters $\boldsymbol{\gamma}$ such that the operator valued maps $\mathsf{A}_{\boldsymbol{\gamma}}, \mathsf{B}_{\boldsymbol{\gamma}}$ constitute two POVMs. Note that, for $\gamma_\varphi, \gamma_\psi \geq 0$, the measurements (S28) can be interpreted as uniformly noisy versions of the sharp measurements Q and P [9].

Since $P_{\mathrm{guess}}^{\mathrm{prior}}(\hat{\mathcal{E}}_{\alpha\boldsymbol{\mu}}; \mathsf{A}, \mathsf{B}) = \alpha P_{\mathrm{guess}}^{\mathrm{prior}}(\hat{\mathcal{E}}_{\boldsymbol{\mu}}; \mathsf{A}, \mathsf{B}) + (1 - \alpha)/d$ for $\alpha > 0$, the optimization problem is the same for the partitioned state ensembles $\hat{\mathcal{E}}_{\boldsymbol{\mu}}$ and $\hat{\mathcal{E}}_{\alpha\boldsymbol{\mu}}$; hence, there is no restriction in parametrizing the family of state ensembles (S8) by means of the direction of the vector $\boldsymbol{\mu}$. We then choose the parametrization $\boldsymbol{\mu} = \boldsymbol{\mu}(\theta)$, given by

$$\mu_\varphi(\theta) = \alpha\left(\sqrt{d}\cos\theta + \sqrt{\frac{d}{d-1}}\sin\theta\right), \qquad \mu_\psi(\theta) = \alpha\left(\sqrt{d}\cos\theta - \sqrt{\frac{d}{d-1}}\sin\theta\right)$$

in terms of the single real parameter $\theta \in [-\pi, \pi]$. Here, $\alpha > 0$ is fixed in such a way that $|\mu_\varphi(\theta)| \leq 1/(d-1)$ and $|\mu_\psi(\theta)| \leq 1/(d-1)$ for all θ's; for example, $\alpha = 1/(d\sqrt{d-1})$. The curve $\{\boldsymbol{\mu}(\theta) \mid \theta \in [-\pi, \pi]\}$ is an ellipse centered at $(0,0)$, and thus it spans all directions in the $(\gamma_\varphi, \gamma_\psi)$-plane. With this parametrization, we always have $\|\boldsymbol{\mu}(\theta)\|_G = 2\alpha$.

In order to apply the classification into cases (C1), (C2) and (C3) of the the previous section, note that the inequality $\mu_\psi(\theta) < 0$ holds if and only if $\theta \in [-\pi, \pi] \smallsetminus [-\theta_0, \pi - \theta_0]$, where

$$\theta_0 = \pi - \arctan\sqrt{d-1} \in (\pi/2, 3\pi/4] ; \tag{S29}$$

in addition, the set $\{-\theta_0, \pi - \theta_0\}$ constitutes all the solutions to the equation $\mu_\psi(\theta) = 0$. Combining these facts with the symmetry $\mu_\varphi(\theta) = \mu_\psi(-\theta)$, we have $\max\{\mu_\varphi(\theta), \mu_\psi(\theta)\} \geq 0$ if and only if $\theta \in [-\theta_0, \theta_0]$. Moreover, $\{-\theta_0, \theta_0 - \pi, \pi - \theta_0, \theta_0\}$ are all the values of θ for which either $\mu_\varphi(\theta) = 0$ or $\mu_\psi(\theta) = 0$. Among the latter values, we have $\max\{\mu_\varphi(\theta), \mu_\psi(\theta)\} = 0$ if and only if $\theta \in \{-\theta_0, \theta_0\}$. Thus, the cases described in the previous section occur as follows:

(C1) $\Leftrightarrow$ $d = 2$ or $\theta \in (-\theta_0, \theta_0)$;

(C2) $\Leftrightarrow$ $d \geq 3$ and $\theta \in [-\pi, \pi] \smallsetminus [-\theta_0, \theta_0]$;

(C3) $\Leftrightarrow$ $d \geq 3$ and $\theta \in \{-\theta_0, \theta_0\}$.
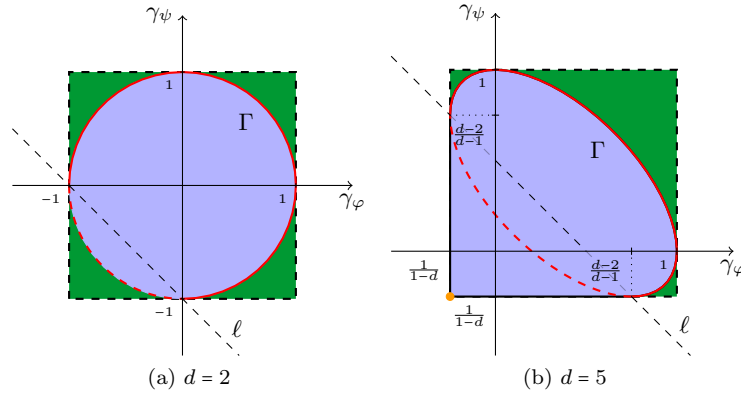

(a) $d = 2$ (b) $d = 5$

FIG. 1. The set of $\boldsymbol{\gamma} = (\gamma_\varphi, \gamma_\psi)$ for which (S28) defines two measurements (green square), and the one for which these measurements are compatible (blue region) for different values of the dimension d. The red line is the ellipse (S31), whose solid part is the curve (S30) for $\theta \in [-\theta_0, \theta_0]$. The dashed line ℓ is described by the equation $\gamma_\varphi + \gamma_\psi = (d-3)/(d-1)$. Note the symmetry around the origin in dimension $d = 2$, and the compatible pair corresponding to $\boldsymbol{\gamma}_0 = (1/(1-d), 1/(1-d))$ when $d \geq 3$ (orange dot).

If $\theta \in (-\theta_0, \theta_0)$, inserting $\mu_\varphi(\theta)$ and $\mu_\psi(\theta)$ into (S26), we obtain that the two measurements (S28) coincide with the margin measurements (S25) for $\boldsymbol{\gamma} = \boldsymbol{\gamma}(\theta) = (\gamma_\varphi(\theta), \gamma_\psi(\theta))$, where

$$\gamma_\varphi(\theta) = \frac{d - 2 - d\cos(\theta + \theta_0)}{2(d-1)}, \qquad \gamma_\psi(\theta) = \frac{d - 2 - d\cos(\theta - \theta_0)}{2(d-1)} ; \tag{S30}$$

moreover, the equality $P_{\text{guess}}^{\text{post}}(\hat{\mathcal{E}}_{\boldsymbol{\mu}(\theta)}) = P_{\text{guess}}^{\text{prior}}(\hat{\mathcal{E}}_{\boldsymbol{\mu}(\theta)}; \mathsf{A}_{\boldsymbol{\gamma}(\theta)}, \mathsf{B}_{\boldsymbol{\gamma}(\theta)})$ holds with this choice of $\boldsymbol{\gamma}$. By Prop. S8, this implies that $(\mathsf{A}_{\boldsymbol{\gamma}(\theta)}, \mathsf{B}_{\boldsymbol{\gamma}(\theta)}) \in \partial\mathcal{O}_{X,Y}^{\text{com}}$ for all $\theta \in (-\theta_0, \theta_0)$. Since the set $\partial\mathcal{O}_{X,Y}^{\text{com}}$ is closed, an easy continuity argument shows that the last inclusion is true also for $\theta \in \{-\theta_0, \theta_0\}$ (although $\mathsf{A}_{\boldsymbol{\gamma}(\theta)}$ and $\mathsf{B}_{\boldsymbol{\gamma}(\theta)}$ do no longer coincide with the margins (S25) for these values of θ). Note that for $\lambda > 1$ the two measurements $\mathsf{A}_{\lambda\boldsymbol{\gamma}(\theta)}$ and $\mathsf{B}_{\lambda\boldsymbol{\gamma}(\theta)}$ are necessarily incompatible, as otherwise we would get the contradiction $(\mathsf{A}_{\boldsymbol{\gamma}(\theta)}, \mathsf{B}_{\boldsymbol{\gamma}(\theta)}) = (1 - 1/\lambda)(\mathsf{U}_X, \mathsf{U}_Y) + (1/\lambda)(\mathsf{A}_{\lambda\boldsymbol{\gamma}(\theta)}, \mathsf{B}_{\lambda\boldsymbol{\gamma}(\theta)}) \in \text{ri}(\mathcal{O}_{X,Y}^{\text{com}})$ by Prop. S5.(c) and [1, Thm. 6.1].

In the $(\gamma_\varphi, \gamma_\psi)$-plane, the curve $\Gamma = \{\boldsymbol{\gamma}(\theta) \mid \theta \in [-\theta_0, \theta_0]\}$ given by (S30) is the part of the ellipse

$$d(\gamma_\varphi^2 + \gamma_\psi^2) + 2(d-2)\gamma_\varphi\gamma_\psi - 2(d-2)(\gamma_\varphi + \gamma_\psi) = 4 - d \tag{S31}$$

lying above the line $\gamma_\varphi + \gamma_\psi = (d-3)/(d-1)$, as depicted in Fig. 1. As we have just seen, all pairs of measurements corresponding to points beyond Γ are incompatible.

If $\theta \in [-\pi, \pi] \smallsetminus [-\theta_0, \theta_0]$, two different situations occur according to the dimension d. If $d = 2$, inserting $\mu_\varphi(\theta)$ and $\mu_\psi(\theta)$ into (S26) still gives the same results as before also for the new values of θ. The curve (S30) for $\theta \in [-\pi, \pi]$ is then the unit circle in the $(\gamma_\varphi, \gamma_\psi)$-plane. Since for $\boldsymbol{\gamma} = (0, 0)$ the two observables $\mathsf{A}_{\boldsymbol{\gamma}} = \mathsf{U}_X$ and $\mathsf{B}_{\boldsymbol{\gamma}} = \mathsf{U}_Y$ are trivially compatible, by convexity we thus conclude that the measurements $\mathsf{A}_{\boldsymbol{\gamma}}$ and $\mathsf{B}_{\boldsymbol{\gamma}}$ are compatible if and only if $\gamma_\varphi^2 + \gamma_\psi^2 \leq 1$.

On the other hand, if $d \geq 3$, then inserting $\mu_\varphi(\theta)$ and $\mu_\psi(\theta)$ into (S26) yields $\gamma_\varphi = \gamma_\psi = 1/(1-d)$ irrespectively of the value of θ. Setting $\boldsymbol{\gamma}_0 = (1/(1-d), 1/(1-d))$, it follows that $(\mathsf{A}_{\boldsymbol{\gamma}_0}, \mathsf{B}_{\boldsymbol{\gamma}_0}) \in \partial\mathcal{O}_{X,Y}^{\mathrm{com}}$, and, again by convexity, all $\boldsymbol{\gamma}$'s lying between $\boldsymbol{\gamma}_0$ and the curve Γ correspond to compatible pairs of measurements.

The previous discussion is summarized in the second main result of this section.

Theorem S3 (Thm. 4 of the main paper). *(a) Suppose $d = 2$. For $\boldsymbol{\gamma} \in [-1, 1] \times [-1, 1]$, the two measurements $\mathsf{A}_{\boldsymbol{\gamma}}$ and $\mathsf{B}_{\boldsymbol{\gamma}}$ of (S28) are compatible if and only if $\gamma_\varphi^2 + \gamma_\psi^2 \leq 1$. Moreover, $(\mathsf{A}_{\boldsymbol{\gamma}}, \mathsf{B}_{\boldsymbol{\gamma}}) \in \partial\mathcal{O}_{X,Y}^{\mathrm{com}}$ if and only if $\gamma_\varphi^2 + \gamma_\psi^2 = 1$.*

(b) Suppose $d \geq 3$. For $\boldsymbol{\gamma} \in [1/(1-d), 1] \times [1/(1-d), 1]$, the two measurements $\mathsf{A}_{\boldsymbol{\gamma}}$ and $\mathsf{B}_{\boldsymbol{\gamma}}$ of (S28) are compatible if and only if

$$\gamma_\varphi + \gamma_\psi \leq \frac{d-3}{d-1} \qquad or \qquad d(\gamma_\varphi^2 + \gamma_\psi^2) + 2(d-2)\gamma_\varphi\gamma_\psi - 2(d-2)(\gamma_\varphi + \gamma_\psi) \leq 4 - d.$$

Moreover, $(\mathsf{A}_{\boldsymbol{\gamma}}, \mathsf{B}_{\boldsymbol{\gamma}}) \in \partial\mathcal{O}_{X,Y}^{\mathrm{com}}$ if and only if

$$\boldsymbol{\gamma} = \left(\frac{t}{d-1}, \frac{1}{1-d}\right) \quad or \quad \boldsymbol{\gamma} = \left(\frac{1}{1-d}, \frac{t}{d-1}\right) \quad for \; some \; t \in [-1, d-2]$$

$$or \quad \boldsymbol{\gamma} = \left(\frac{d-2-d\cos(\theta+\theta_0)}{2(d-1)}, \frac{d-2-d\cos(\theta-\theta_0)}{2(d-1)}\right) \quad for \; some \; \theta \in [-\theta_0, \theta_0],$$

where θ_0 is given in (S29).

Statement (a) of the previous theorem is well known [10]. On the other hand, statement (b) was proved in the particular case in which the two measurements $\mathsf{A}_{\boldsymbol{\gamma}}$ and $\mathsf{B}_{\boldsymbol{\gamma}}$ are conjugate by the Fourier transform of the cyclic group $\mathbb{Z}_d$, and restricting only to $\boldsymbol{\gamma} \in [0, 1] \times [0, 1]$ [9]. Compared with the group theoretical approach of [9], the present derivation of Thm. S3 has the advantage of not requiring any symmetry condition on the two MUB $\{\varphi_1, \ldots, \varphi_d\}$ and $\{\psi_1, \ldots, \psi_d\}$; thus, for dimensions $d \geq 4$ it actually applies to many inequivalent pairs of MUB, and not only to the Fourier conjugate pairs considered in [9] (see [11] for a list of inequivalent pairs in dimensions $4 \leq d \leq 16$ and the proof of the equivalence of all pairs in dimensions $d = 2, 3$). In the general (not symmetric) case, a proof of statement (b) for $\boldsymbol{\gamma}$ constrained on the diagonal $\gamma_\varphi = \gamma_\psi$ is contained in [12, 13].

We finally remark that the essential differences that lead to separate results for the cases $d = 2$ and $d \geq 3$ are: (1) the additional symmetry $(\mathsf{A}_{\boldsymbol{\gamma}}, \mathsf{B}_{\boldsymbol{\gamma}}) \in \mathcal{O}_{X,Y}^{\mathrm{com}} \Leftrightarrow (\mathsf{A}_{-\boldsymbol{\gamma}}, \mathsf{B}_{-\boldsymbol{\gamma}}) \in \mathcal{O}_{X,Y}^{\mathrm{com}}$, which is specific of the $d = 2$ case; (2) the fact that, for $\boldsymbol{\gamma}_0 = (1/(1-d), 1/(1-d))$, the two measurements $(\mathsf{A}_{\boldsymbol{\gamma}_0}, \mathsf{B}_{\boldsymbol{\gamma}_0})$ are compatible if and only if $d \geq 3$.

---

[1] R.T. Rockafellar. *Convex analysis*. Princeton University Press, Princeton, N.J., 1997.

[2] L. Vesely, unpublished notes.

[3] C. Carmeli, T. Heinosaari, and A. Toigo. State discrimination with postmeasurement information and incompatibility of quantum measurements. *Phys. Rev. A*, 98:012126, 2018.

[4] A.S. Holevo. Statistical decision theory for quantum systems. *J. Multivariate Anal.*, 3:337–394, 1973.

[5] H.P. Yuen, R.S. Kennedy, and M. Lax. Optimum testing of multiple hypotheses in quantum detection theory. *IEEE Trans. Inform. Theory*, IT-21:125–134, 1975.

[6] C.W. Helstrom. *Quantum Detection and Estimation Theory*. Academic Press, New York, 1976.

[7] J. Bae. Structure of minimum-error quantum state discrimination. *New J. Phys.*, 15:073037, 2013.

[8] L. Guerini and M.T. Cunha. Uniqueness of the joint measurement and the structure of the set of compatible quantum measurements. *J. Math. Phys.*, 59:042106, 2018.

[9] C. Carmeli, T. Heinosaari, and A. Toigo. Informationally complete joint measurements on finite quantum systems. *Phys. Rev. A*, 85:012109, 2012.

[10] P. Busch. Unsharp reality and joint measurements for spin observables. *Phys. Rev. D*, 33:2253–2261, 1986.

[11] W. Tadej and K. Życzkowski. A concise guide to complex Hadamard matrices. *Open Syst. Inf. Dyn.*, 13:133–177, 2006.

[12] R. Uola, K. Luoma, T. Moroder, and T. Heinosaari. Adaptive strategy for joint measurements. *Phys. Rev. A*, 94:022109, 2016.

[13] S. Designolle, P. Skrzypczyk, F. Fröwis, and N. Brunner. Quantifying measurement incompatibility of mutually unbiased bases. *Phys. Rev. Lett.*, 122:050402, 2019.