Communications in Mathematical Physics manuscript No.
(will be inserted by the editor)

Measurement uncertainty relations for discrete observables: Relative entropy formulation

Alberto Barchielli[1,2,3], **Matteo Gregoratti**[1,2], **Alessandro Toigo**[1,3]

[1] Politecnico di Milano, Dipartimento di Matematica, Piazza Leonardo da Vinci 32, I-20133 Milano, Italy
[2] Istituto Nazionale di Alta Matematica (INDAM-GNAMPA)
[3] Istituto Nazionale di Fisica Nucleare (INFN), Sezione di Milano

The date of receipt and acceptance will be inserted by the editor

Abstract We introduce a new information-theoretic formulation of quantum measurement uncertainty relations, based on the notion of relative entropy between measurement probabilities. In the case of a finite-dimensional system and for any approximate joint measurement of two target discrete observables, we define the entropic divergence as the maximal total loss of information occurring in the approximation at hand. For fixed target observables, we study the joint measurements minimizing the entropic divergence, and we prove the general properties of its minimum value. Such a minimum is our uncertainty lower bound: the total information lost by replacing the target observables with their optimal approximations, evaluated at the worst possible state. The bound turns out to be also an entropic incompatibility degree, that is, a good information-theoretic measure of incompatibility: indeed, it vanishes if and only if the target observables are compatible, it is state-independent, and it enjoys all the invariance properties which are desirable for such a measure. In this context, we point out the difference between general approximate joint measurements and sequential approximate joint measurements; to do this, we introduce a separate index for the tradeoff between the error of the first measurement and the disturbance of the second one. By exploiting the symmetry properties of the target observables, exact values, lower bounds and optimal approximations are evaluated in two different concrete examples: (1) a couple of spin-1/2 components (not necessarily orthogonal); (2) two Fourier conjugate mutually unbiased bases in prime power dimension. Finally, the entropic incompatibility degree straightforwardly generalizes to the case of many observables, still maintaining all its relevant properties; we explicitly compute it for three orthogonal spin-1/2 components.

1. Introduction

In the foundations of Quantum Mechanics, a remarkable achievement of the last years has been the clarification of the differences between *preparation uncertainty relations* (PURs) and *measurement uncertainty relations* (MURs) [1–13], both of them arising from Heisenberg's heuristic considerations about the precision with which the position and the momentum of a quantum particle can be determined [14].

One speaks of PURs when some lower bound is given on the "spreads" of the distributions of two observables A and B measured in the same state ρ. The most known formulation of PURs, due to Robertson [15], involves the product of the two standard deviations; more recent formulations are given in terms of distances among probability distributions [10] or entropies [13, 16–22].

On the other hand, one refers to MURs when some lower bound is given on the "errors" of any approximate joint measurement M of two target observables A and B. When M is realized as a sequence of two measurements, one for each target observable, MURs are regarded also as relations between the "error" allowed in an approximate measurement of the first observable and the "disturbance" affecting the successive measurement of the second one.

Although the recent developments of the theory of approximate quantum measurements [11, 23–26] and nondisturbing quantum measurements [27, 28] have generated a considerable renewed interest in MURs, no agreement has yet been reached about the proper quantifications of the "error" or "disturbance" terms. Here, the main problem is how to compare the target observables A and B with their approximate or perturbed versions provided by the marginals $\mathsf{M}_{[1]}$ and $\mathsf{M}_{[2]}$ of M; indeed, A, $\mathsf{M}_{[1]}$, $\mathsf{M}_{[2]}$ and B may typically be incompatible. The proposals then range from operator formulations of the error [1–4, 29, 30] to distances for probability distributions [6–12] and conditional entropies [31–33].

In this paper, we propose and develop a new approach to MURs based on the notion of *relative entropy*. Here we deal with the case of discrete observables for a finite dimensional quantum system. The extension to position and momentum is given in [34].

In the spirit of Busch, Lahti, Werner [6–10], we quantify the "error" in the approximation $\mathsf{A} \simeq \mathsf{M}_{[1]}$ by comparing the respective outcome distributions A^ρ and $\mathsf{M}_{[1]}^\rho$ in every possible state ρ; however, differently from [6–10], the comparison is done from the point of view of information theory. Then, the natural choice is to consider $S\big(\mathsf{A}^\rho \| \mathsf{M}_{[1]}^\rho\big)$, the relative entropy of A^ρ with respect to $\mathsf{M}_{[1]}^\rho$, as a quantification of the information loss when A^ρ is approximated with $\mathsf{M}_{[1]}^\rho$. Similarly, in order to quantify either the "error" or – if A and B are measured in sequence – the "disturbance" related to the approximation $\mathsf{B} \simeq \mathsf{M}_{[2]}$, we employ the relative entropy $S\big(\mathsf{B}^\rho \| \mathsf{M}_{[2]}^\rho\big)$. Relative entropy appears to be the fundamental quantity from which the other entropic notions can be derived, cf. [35–37]. It should be noticed that relative entropy, of classical or quantum type, has already been used in quantum measurement theory to give proper measures of information gains and losses in various scenarios [37–41].

The relative entropy formulation of MURs, given in Section 2.3, is: for every approximate joint measurement M of A and B, there exists a state ρ such that

$$S\big(\mathsf{A}^\rho \| \mathsf{M}_{[1]}^\rho\big) + S\big(\mathsf{B}^\rho \| \mathsf{M}_{[2]}^\rho\big) \geq c(\mathsf{A}, \mathsf{B}), \tag{1}$$

where the uncertainty lower bound

$$c(\mathsf{A}, \mathsf{B}) = \inf_{\mathsf{M}} \sup_{\rho} \Big\{ S\big(\mathsf{A}^\rho \| \mathsf{M}_{[1]}^\rho\big) + S\big(\mathsf{B}^\rho \| \mathsf{M}_{[2]}^\rho\big) \Big\} \tag{2}$$

depends on the allowed joint measurements M. In the above definition, the same state ρ appears in both error terms $S\big(\mathsf{A}^\rho \| \mathsf{M}_{[1]}^\rho\big)$ and $S\big(\mathsf{B}^\rho \| \mathsf{M}_{[2]}^\rho\big)$; thus, by making their sum, all possible error compensations are taken into account in the maximization. The quantity $\sup_\rho \left\{ S\big(\mathsf{A}^\rho \| \mathsf{M}^\rho_{[1]}\big) + S\big(\mathsf{B}^\rho \| \mathsf{M}^\rho_{[2]}\big) \right\}$ gives a state-independent quantification of the total inefficiency of the approximate joint measurement M at hand, and we call it entropic divergence of M from (A, B).

By considering any possible approximate joint measurement in the definition of $c(\mathsf{A}, \mathsf{B})$, we get an uncertainty lower bound $c_{\mathrm{inc}}(\mathsf{A}, \mathsf{B})$ that turns out to be a proper measure of the incompatibility of A and B. On the other hand, by considering only sequential measurements, we derive an uncertainty lower bound $c_{\mathrm{ed}}(\mathsf{A}, \mathsf{B})$ that provides a suitable quantification of the error/disturbance tradeoff for the two (sequentially ordered) target observables. Indeed, such lower bounds share a lot of desirable properties: they are zero if and only if the target observables are compatible (respectively, sequentially compatible); they are invariant under unitary transformations and relabelling of the output values of the measurements; and finally, they are bounded from above by a value that is independent of both the dimension of the Hilbert space and the number of the possible outcomes. As a main result, we show also that, for a generic couple of observables A and B, considering only their sequential measurements is a real restriction, because in general $c_{\mathrm{ed}}(\mathsf{A}, \mathsf{B})$ may be larger than $c_{\mathrm{inc}}(\mathsf{A}, \mathsf{B})$; actually, the two indexes are guaranteed to coincide only if one makes some extra assumptions on A and B (e.g. if the second observable B is supposed to be sharp).

Thus, every time A and B are incompatible, the total loss of information $S\big(\mathsf{A}^\rho \| \mathsf{M}^\rho_{[1]}\big) + S\big(\mathsf{B}^\rho \| \mathsf{M}^\rho_{[2]}\big)$ in the approximations $\mathsf{A} \simeq \mathsf{M}_{[1]}$ and $\mathsf{B} \simeq \mathsf{M}_{[2]}$ depends on both the joint measurement M and the state ρ; however, since $c_{\mathrm{inc}}(\mathsf{A}, \mathsf{B}) > 0$, inequality (1) states that there is a minimum potential loss that no joint measurement M can avoid. Similar remarks hold for sequential measurements and the corresponding error/disturbance coefficient. Note that, even if A and B are incompatible, the left hand side of (1) can vanish if the state ρ and the approximate joint measurement M are suitably chosen (see Section 2.3). Of course, this is not a contradiction, as the formulation (1), (2) of MURs is about the size of the total information loss in the worst – but not all – input states. In this sense, the bound (2) is a state-independent quantification of the minimal inefficiency of the approximations $\mathsf{A} \simeq \mathsf{M}_{[1]}$ and $\mathsf{B} \simeq \mathsf{M}_{[2]}$.

Our MURs directly compare with those of [6–10], from which however they differ in one essential aspect: the latter quantify the inaccuracy of the approximate joint measurement M by maximizing the errors of the approximations $\mathsf{A}^{\rho_1} \simeq \mathsf{M}^{\rho_1}_{[1]}$ and $\mathsf{B}^{\rho_2} \simeq \mathsf{M}^{\rho_2}_{[2]}$ over independently chosen states ρ_1 and ρ_2; instead, in (2) we maximize the total approximation error $S\big(\mathsf{A}^\rho \| \mathsf{M}^\rho_{[1]}\big) + S\big(\mathsf{B}^\rho \| \mathsf{M}^\rho_{[2]}\big)$ over a single state ρ. On the conceptual level, this amounts to say that our MURs are a statement about the inaccuracy of the approximation $(\mathsf{A}, \mathsf{B}) \simeq (\mathsf{M}_{[1]}, \mathsf{M}_{[2]})$ that occurs in one preparation of the system; those of [6–10] rather refer to the inefficiencies of two separate uses of the approximate joint measurement M, namely, for approximating $\mathsf{A} \simeq \mathsf{M}_{[1]}$ in a first preparation, and $\mathsf{B} \simeq \mathsf{M}_{[2]}$ in a second one. Similar considerations hold for the conditional entropy approach of [31–33], where the "noise" and "disturbance" terms are defined through different preparations in a sort of calibration procedure. In this respect, our MURs are reminiscent of the traditional entropic PURs, which relate the spreads of the distributions A^ρ and B^ρ evaluated at the same state ρ (see Section 2.5).

Whenever A and B are incompatible, we will look for the exact value of $c_{\mathrm{inc}}(\mathsf{A}, \mathsf{B})$, or at least some lower bound for it, as well as we will try to determine the optimal approximate joint measurements M which saturate the minimum. In particular, we will prove that in some relevant applications there is actually a unique such M, thus showing that in these cases the entropic optimality criterium unambiguously fixes the best approximate joint measurement.

The generalization of our MURs to the case of more than two target observables is rather straightforward by the very structure of the relative entropy formulation. It is worth noticing that there are triples of observables whose optimal approximate joint measurements are not unique, even if all their possible pairings do have the corresponding binary uniqueness property (see e.g. the two and three orthogonal spin-1/2 components in Sections 3.2 and 4.2).

Now, we summarize the structure of the paper. In Section 2, we state our entropic MURs for two target observables, and we introduce and study the main mathematical objects which are involved in their formulation. In Section 3, we undertake the explicit computation of the incompatibility indexes $c_{\rm inc}(\mathsf{A},\mathsf{B})$ and $c_{\rm ed}(\mathsf{A},\mathsf{B})$ and their respective optimal approximate joint measurements M for several examples of incompatible target observables. Some general results are proved, which show how the symmetry properties of the quantum system can help in the task. Then, two cases are studied: two spin-1/2 components, which we do not assume to be necessarily orthogonal, and two Fourier conjugate observables associated with a pair of mutually unbiased bases (MUBs) in prime power dimension. In Section 4, we generalize the relative entropy formulation of MURs to the case of many target observables. As an example, the case of three orthogonal spin-1/2 components is completely solved. Finally, Section 5 contains a conclusive discussion and presents some open problems. Three further appendices are provided at the end of the paper: in Appendix A, a couple of examples show that the coefficients $c_{\rm inc}(\mathsf{A},\mathsf{B})$, $c_{\rm ed}(\mathsf{A},\mathsf{B})$ and $c_{\rm ed}(\mathsf{B},\mathsf{A})$ may be different in general; Appendices B and C collect all the technical details and proofs for the cases studied in Sections 3.2, 3.3, and 4.2.

1.1. Observables and instruments. We start by fixing our quantum system and recalling the notions and basic facts on observables and measurements that we will use in the article [23–25, 27, 42–45].

The Hilbert space $\mathcal{H}$ *and the spaces* $\mathcal{L}(\mathcal{H})$, $\mathcal{T}(\mathcal{H})$, $\mathcal{S}(\mathcal{H})$. We consider a quantum system described by a finite-dimensional complex Hilbert space $\mathcal{H}$, with $\dim\mathcal{H}=d$; then, the spaces $\mathcal{L}(\mathcal{H})$ of all linear bounded operators on $\mathcal{H}$ and the trace-class $\mathcal{T}(\mathcal{H})$ coincide. Let $\mathcal{S}(\mathcal{H})$ denote the convex set of all states on $\mathcal{H}$ (positive, unit trace operators), which is a compact subset of $\mathcal{T}(\mathcal{H})$. The extreme points of $\mathcal{S}(\mathcal{H})$ are the pure states (rank-one projections) $\rho=|\,\psi\,\rangle\langle\,\psi\,|$, with $\psi\in\mathcal{H}$ and $\|\psi\|=1$.

The space of observables $\mathcal{M}(\mathcal{X})$ *and the space of probabilities* $\mathcal{P}(\mathcal{X})$. In the general formulation of quantum mechanics, an *observable* is identified with a *positive operator valued measure* (POVM). We will consider only observables with outcomes in a finite set $\mathcal{X}$. Then, a POVM on $\mathcal{X}$ is identified with its discrete density $\mathsf{A}:\mathcal{X}\to\mathcal{L}(\mathcal{H})$, whose values $\mathsf{A}(x)$ are positive operators on $\mathcal{H}$ such that $\sum_{x\in\mathcal{X}}\mathsf{A}(x)=\mathbb{1}$; here, the sum involves a finite number $|\mathcal{X}|$ of terms ($|\mathcal{X}|$ denotes the cardinality of $\mathcal{X}$). Similarly, a probability on $\mathcal{X}$ is identified with its discrete probability density (or mass function) $p:\mathcal{X}\to\mathbb{R}$, where $p(x)\geq 0$ and $\sum_{x\in\mathcal{X}}p(x)=1$.

For $\rho\in\mathcal{S}(\mathcal{H})$, the function $\mathsf{A}^\rho(x)=\operatorname{Tr}\{\rho\mathsf{A}(x)\}$ is the discrete probability density on $\mathcal{X}$ which gives the outcome distribution in a measurement of the observable A performed on the quantum system prepared in the state ρ.

We denote by $\mathcal{M}(\mathcal{X})$ the set of the observables which are associated with the system at hand and have outcomes in $\mathcal{X}$; $\mathcal{M}(\mathcal{X})$ is a convex, compact subset of $\mathcal{L}(\mathcal{H})^{\mathcal{X}}$, the finite dimensional linear space of all functions from $\mathcal{X}$ to $\mathcal{L}(\mathcal{H})$. Both mappings $\rho \mapsto \mathsf{A}^\rho$ and $\mathsf{A} \mapsto \mathsf{A}^\rho$ are continuous and affine (i.e. preserving convex combinations) from the respective domains into the convex set $\mathcal{P}(\mathcal{X})$ of the probabilities on $\mathcal{X}$. As a subset of $\mathbb{R}^{\mathcal{X}}$, the set $\mathcal{P}(\mathcal{X})$ is convex and compact. The extreme points of $\mathcal{P}(\mathcal{X})$ are the (Kronecker) delta distributions δ_x, with $x \in \mathcal{X}$.

Trivial and sharp observables. An observable A is *trivial* if $\mathsf{A} = p\mathbb{1}$ for some probability p, where $\mathbb{1}$ is the identity of $\mathcal{H}$. In particular, we will make use of the uniform distribution $u_{\mathcal{X}}$ on $\mathcal{X}$, $u_{\mathcal{X}}(x) = 1/|\mathcal{X}|$, and the *trivial uniform observable* $\mathsf{U}_{\mathcal{X}} = u_{\mathcal{X}}\mathbb{1}$.

An observable A is *sharp* if $\mathsf{A}(x)$ is a projection $\forall x \in \mathcal{X}$. Note that we allow $\mathsf{A}(x) = 0$ for some x, which is required when dealing with sets of observables sharing the same outcome space. Of course, for every sharp observable we have $|\{x : \mathsf{A}(x) \neq 0\}| \leq d$.

Bi-observables and compatible observables. When the outcome set has the product form $\mathcal{X} \times \mathcal{Y}$, we speak of bi-observables. In this case, given the POVM $\mathsf{M} \in \mathcal{M}(\mathcal{X} \times \mathcal{Y})$, we can introduce also the *marginal observables* $\mathsf{M}_{[1]} \in \mathcal{M}(\mathcal{X})$ and $\mathsf{M}_{[2]} \in \mathcal{M}(\mathcal{Y})$ by

$$\mathsf{M}_{[1]}(x) = \sum_{y \in \mathcal{Y}} \mathsf{M}(x, y), \qquad \mathsf{M}_{[2]}(y) = \sum_{x \in \mathcal{X}} \mathsf{M}(x, y).$$

In the same way, for $p \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$, we get the *marginal probabilities* $p_{[1]} \in \mathcal{P}(\mathcal{X})$ and $p_{[2]} \in \mathcal{P}(\mathcal{Y})$. Clearly, $(\mathsf{M}_{[i]})^\rho = (\mathsf{M}^\rho)_{[i]}$; hence there is no ambiguity in writing $\mathsf{M}^\rho_{[i]}$ for both probabilities.

Two observables $\mathsf{A} \in \mathcal{M}(\mathcal{X})$ and $\mathsf{B} \in \mathcal{M}(\mathcal{Y})$ are *jointly measurable* or *compatible* if there exists a bi-observable $\mathsf{M} \in \mathcal{M}(\mathcal{X} \times \mathcal{Y})$ such that $\mathsf{M}_{[1]} = \mathsf{A}$ and $\mathsf{M}_{[2]} = \mathsf{B}$; then, we call M a *joint measurement* of A and B.

Two classical probabilities $p \in \mathcal{P}(\mathcal{X})$ and $q \in \mathcal{P}(\mathcal{Y})$ are always compatible, as they can be seen as the marginals of at least one joint probability in $\mathcal{P}(\mathcal{X} \times \mathcal{Y})$. Indeed, one can take the *product probability* $p \otimes q$ given by $(p \otimes q)(x, y) = p(x)q(y)$. Clearly, nothing similar can be defined for two non-commuting quantum observables, for which instead compatibility usually is a highly nontrivial requirement.

The space of instruments $\mathcal{J}(\mathcal{X})$. Given a pre-measurement state ρ, a POVM allows to compute the probability distribution of the measurement outcome. In order to describe also the state change produced by the measurement, we need the more general mathematical notion of *instrument*, i.e. a measure $\mathcal{J}$ on the outcome set $\mathcal{X}$ taking values in the set of the completely positive maps on $\mathcal{L}(\mathcal{H})$. In our case of finitely many outcomes, an instrument is described by its discrete density $x \mapsto \mathcal{J}_x$, $x \in \mathcal{X}$, whose general structure is $\mathcal{J}_x[\rho] = \sum_\alpha J_x^\alpha \rho J_x^{\alpha\,*}$, $\forall \rho \in \mathcal{S}(\mathcal{H})$; here, the Kraus operators $J_x^\alpha \in \mathcal{L}(\mathcal{H})$ are such that $\sum_{x \in \mathcal{X}} \sum_\alpha J_x^{\alpha\,*} J_x^\alpha = \mathbb{1}$ and, since $\mathcal{H}$ is finite-dimensional, the index α can be restricted to finitely many values. The *adjoint instrument* is given by $\mathcal{J}_x^*[F] = \sum_\alpha J_x^{\alpha\,*} F J_x^\alpha, \forall F \in \mathcal{L}(\mathcal{H})$. The sum $\mathcal{J}_{\mathcal{X}} = \sum_{x \in \mathcal{X}} \mathcal{J}_x$ is a *quantum channel*, i.e. a completely positive trace preserving map on $\mathcal{S}(\mathcal{H})$. We denote by $\mathcal{J}(\mathcal{X})$ the convex and compact set of all $\mathcal{X}$-valued instruments for our quantum system.

By setting $\mathsf{A}(x) = \mathcal{J}_x^*[\mathbb{1}] = \sum_\alpha J_x^{\alpha\,*} J_x^\alpha$, a POVM $\mathsf{A} \in \mathcal{M}(\mathcal{X})$ is defined, which is the observable measured by the instrument $\mathcal{J}$; we say that *the instrument* $\mathcal{J}$ *implements the observable* A. The state of the system after the measurement, conditioned on the outcome x, is $\mathcal{J}_x[\rho]/\mathsf{A}^\rho(x)$. We recall that, given an observable A, one can always find an instrument $\mathcal{J}$ implementing A, but $\mathcal{J}$ is not uniquely determined by A, i.e. different instruments $\mathcal{J}$, with different actions on the quantum system, may be used to measure the same observable A.

Sequential measurements and sequentially compatible observables. Employing the notion of instrument, we can describe a measurement of an observables $\mathsf{A} \in \mathcal{M}(\mathcal{X})$ followed by a measurement of an observable $\mathsf{B} \in \mathcal{M}(\mathcal{Y})$: a *sequential measurement* of A followed by B is a bi-observable $\mathsf{M}(x,y) = \mathcal{J}_x^*[\mathsf{B}(y)]$, where $\mathcal{J}$ is any instrument implementing A. Its marginals are $\mathsf{M}_{[1]}(x) = \mathcal{J}_x^*[\mathbb{1}] = \mathsf{A}(x)$ and $\mathsf{M}_{[2]}(y) = \mathcal{J}_{\mathcal{X}}^*[\mathsf{B}(y)]$. We write $\mathsf{M} = \mathcal{J}^*(\mathsf{B})$, which is a measurement in which one first applies the instrument $\mathcal{J}$ to measure A, and then he measures the observable B on the resulting output state; in this way, he obtains a joint measurement of A and $\mathcal{J}_{\mathcal{X}}^*[\mathsf{B}(\cdot)]$, a perturbed version of B.

An observable $\mathsf{A} \in \mathcal{M}(\mathcal{X})$ *can be measured without disturbing* $\mathsf{B} \in \mathcal{M}(\mathcal{Y})$ [27], or shortly A and B are *sequentially compatible observables*, if there exists a sequential measurement $\mathsf{M} = \mathcal{J}^*(\mathsf{B})$ such that

$$\mathsf{M}_{[1]} \equiv \mathcal{J}_\cdot^*[\mathbb{1}] = \mathsf{A}, \qquad \mathsf{M}_{[2]} \equiv \mathcal{J}_{\mathcal{X}}^*[\mathsf{B}(\cdot)] = \mathsf{B}.$$

So, a measurement of B at time 1 (i.e. after the measurement of A) has the same outcome distribution as a measurement of B at time 0 (i.e. before the measurement of A).

If A and B are sequentially compatible observables, they clearly are also jointly measurable. However, the opposite is not true; two counterexamples are shown in [27] and are reported in Appendix A. This happens because we demand to measure just B at time 1, i.e. we do not content ourselves with getting at time 1 the same outcome distribution of a measurement of B performed at time 0. Indeed, this second requirement is weaker: it can be satisfied by any couple of jointly measurable observables A and B, by measuring a suitable third observable C after A (with A implemented by an instrument $\mathcal{J}$ which possibly increases the dimension of the Hilbert space). The definition of sequentially compatible observables is not symmetric, and indeed there exist couples of observables such that A can be measured without disturbing B, but for which the opposite is not true. This asymmetry is also reflected in the remarkable fact that, if the second observable is sharp, then the compatibility of A and B turns out to be equivalent to their sequential compatibility.

Target observables. In this paper, we fix two target observables with finitely many values, $\mathsf{A} \in \mathcal{M}(\mathcal{X})$ and $\mathsf{B} \in \mathcal{M}(\mathcal{Y})$, and we study how to characterize their uncertainty relations. For any $\rho \in \mathcal{S}(\mathcal{H})$, the associated probability distributions A^ρ and B^ρ can be estimated by measuring either A or B in many identical preparations of the quantum system in the state ρ. No joint or sequential measurement of A and B is required at this stage. In Section 2 we develop a general theory to quantify the error made by approximating A and B with compatible observables and we introduce the notion of optimal approximate joint measurement for A and B.

1.2. Relative and Shannon entropies. In this paper, we will be concerned with entropic quantities of classical type [35, 36]; we express them in "bits", which means to use logarithms with base 2: $\log \equiv \log_2$.

The fundamental quantity is the *relative entropy*; although it can be defined for general probability measures, here we only recall the discrete case. Given two probabilities

$p,\ q \in \mathcal{P}(\mathcal{X})$, the relative entropy of p with respect to q is

$$S\big(p\|q\big) = \begin{cases} \displaystyle\sum_{x\in\operatorname{supp} p} p(x)\log\frac{p(x)}{q(x)} & \text{if } \operatorname{supp} p \subseteq \operatorname{supp} q, \\ +\infty & \text{otherwise;} \end{cases} \tag{3}$$

it defines an extended real valued function on the product set $\mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{X})$. Also the terms *Kullback-Leibler divergence* and *information for discrimination* are used for $S\big(p\|q\big)$.

The relative entropy $S\big(p\|q\big)$ is a measure of the inefficiency of assuming that the probability is q when the true probability is p [36, Sect. 2.3]; in other words, it is the amount of information lost when q is used to approximate p [35, p. 51]. It appears in data compression theory [36, Theor. 5.4.3], model selection problems [35], and it is related to the error probability in the context of hypothesis tests that discriminate the two distributions p and q [36, Theor. 11.8.3]. We stress that $S\big(p\|q\big)$ compares p and q, but it is not a distance since it is not symmetric. As such, the use of S is particularly convenient when the two probabilities have different roles; for instance, if p is the true distribution of a given random variable, while q is the distribution actually used as an approximation of p. This will be our case, where the role of p is played by the distribution A^ρ (or B^ρ) of the target observable A (or B) and q will be the distribution of some allowed approximation; in particular, no joint distribution of p and q is involved.

In comparing our results with entropic PURs, we need also the *Shannon entropy* of a probability $p \in \mathcal{P}(\mathcal{X})$. It is defined by

$$H(p) = -\sum_{x\in\mathcal{X}} p(x)\log p(x), \tag{4}$$

and it provides a measure of the uncertainty of a random variable with distribution p [36, Sect. 2.1].

We collect in the following proposition the main properties of the relative and Shannon entropies [35–37, 43, 46]. For the definition and main properties of lower semicontinuous (LSC) functions, we refer to [47, Sect. 1.5].

Proposition 1. *The following properties hold.*

(i) $0 \leq H(p) \leq \log|\mathcal{X}|$ *and* $S\big(p\|q\big) \geq 0$*, for all* $p,\ q \in \mathcal{P}(\mathcal{X})$.
(ii) $H(p) = 0$ *if and only if* $p = \delta_x$ *for some* x*, where* δ_x *is the delta distribution at* x. $S\big(p\|q\big) = 0$ *if and only if* $p = q$.
(iii) $H(u_\mathcal{X}) = \log|\mathcal{X}|$*, and* $H(p) = \log|\mathcal{X}| - S\big(p\|u_\mathcal{X}\big)$ *for all* $p \in \mathcal{P}(\mathcal{X})$*, where* $u_\mathcal{X}$ *is the uniform probability on* $\mathcal{X}$.
(iv) H *and* S *are invariant for relabelling of the outcomes; that is, if* $f : \mathcal{X}' \to \mathcal{X}$ *is a bijective map, then* $H(p\circ f) = H(p)$ *and* $S\big(p\circ f\|q\circ f\big) = S\big(p\|q\big)$.
(v) H *is a concave function on* $\mathcal{P}(\mathcal{X})$*, and* S *is jointly convex on* $\mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{X})$*, namely*

$$S\big(\lambda p_1 + (1-\lambda)p_2 \|\lambda q_1 + (1-\lambda)q_2\big) \leq \lambda S\big(p_1\|q_1\big) + (1-\lambda)S\big(p_2\|q_2\big), \ \forall\lambda\in[0,1].$$

(vi) *The function* $p \mapsto H(p)$ *is continuous on* $\mathcal{P}(\mathcal{X})$*. The function* $(p,q) \mapsto S\big(p\|q\big)$ *is LSC on* $\mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{X})$.
(vii) *If* $p_1, p_2 \in \mathcal{P}(\mathcal{X})$ *and* $q_1, q_2 \in \mathcal{P}(\mathcal{Y})$*, then* $S\big(p_1\otimes q_1\|p_2\otimes q_2\big) = S\big(p_1\|p_2\big) + S\big(q_1\|q_2\big)$.

In order to derive some further specific properties of the relative entropy that will be needed in the following, it is useful to introduce the extended real function $s : [0,1] \times [0,1] \to [-1/2, +\infty]$, with

$$s(u,v) = \begin{cases} u \log \dfrac{u}{v} & \text{if } 0 < u \leq 1 \text{ and } 0 < v \leq 1, \\ 0 & \text{if } u = 0 \text{ and } 0 \leq v \leq 1, \\ +\infty & \text{if } u > 0 \text{ and } v = 0. \end{cases} \tag{5}$$

In terms of s, the relative entropy can be rewritten as $S(p\|q) = \sum_{x\in\mathcal{X}} s(p(x), q(x))$. Note that, unlike the relative entropy, the function s can take also negative values, and its minimum is $s(1/2, 1) = -1/2$. As a function of (u,v), s is continuous at all the points of the square $[0,1] \times [0,1]$ except at the origin $(0,0)$, where it is easily proved to be LSC.

Proposition 2. *For all $\lambda \in (0,1]$ and $q \in \mathcal{P}(\mathcal{X})$, the map $g_\lambda(p) = S\big(p\|\lambda p + (1-\lambda)q\big)$ is finite and continuous in $p \in \mathcal{P}(\mathcal{X})$. It attains the maximum value*

$$\max_{p\in\mathcal{P}(\mathcal{X})} S\big(p\|\lambda p + (1-\lambda)q\big) = \log \frac{1}{\lambda + (1-\lambda)\min_{x\in\mathcal{X}} q(x)}, \tag{6}$$

which is a strictly decreasing function of $\lambda \in (0,1]$.

Proof. Let $\lambda \in (0,1]$. For all $u, v \in [0,1]$, the condition $u > 0$ implies $\lambda u + (1-\lambda)v > 0$, hence

$$s(u, \lambda u + (1-\lambda)v) = \begin{cases} u \log \dfrac{u}{\lambda u + (1-\lambda)v} & \text{if } 0 < u \leq 1, \\ 0 & \text{if } u = 0. \end{cases}$$

Clearly, this is a continuous function of $u \in (0,1]$. To see that it is continuous also at 0, we take the limit

$$\lim_{u\to 0^+} u \log \frac{u}{\lambda u + (1-\lambda)v} = \lim_{u\to 0^+} u \log u - \lim_{u\to 0^+} u \log[\lambda u + (1-\lambda)v]$$
$$= -\lim_{u\to 0^+} u \log[\lambda u + (1-\lambda)v] = \begin{cases} 0 & \text{if } v \neq 0, \\ -\frac{1}{\lambda} \lim_{u\to 0^+} \lambda u \log(\lambda u) = 0 & \text{if } v = 0. \end{cases}$$

Since $g_\lambda(p) = \sum_x s\left(p(x), \lambda p(x) + (1-\lambda)q(x)\right)$, the continuity of g_λ then follows. Since g_λ is also convex on $\mathcal{P}(\mathcal{X})$ by Proposition 1, item (v), and the set $\mathcal{P}(\mathcal{X})$ is compact, the function g_λ takes its maximum at some extreme point δ_x of $\mathcal{P}(\mathcal{X})$. It follows that

$$\sup_{p\in\mathcal{P}(\mathcal{X})} S\big(p\|\lambda p + (1-\lambda)q\big) = \max_{x\in\mathcal{X}} S\big(\delta_x\|\lambda\delta_x + (1-\lambda)q\big)$$
$$= \log \frac{1}{\lambda + (1-\lambda)\min_{x\in\mathcal{X}} q(x)}.$$

Setting $q_{\min} = \min_{x\in\mathcal{X}} q(x)$, the derivative in λ of the last expression is

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\left(\log \frac{1}{\lambda + (1-\lambda)q_{\min}}\right) = \frac{q_{\min} - 1}{(1-q_{\min})\lambda + q_{\min}},$$

which is negative for all $\lambda \in (0,1]$ since $q_{\min} \leq 1/|\mathcal{X}| < 1$. Thus, the right hand side of (6) is strictly decreasing in λ. □

Note that, if $\lambda = 0$, then $g_0(p) = S(p\|q)$ is an extended real LSC function on $\mathcal{P}(\mathcal{X})$ by Proposition 1, item (vi). However, it is not difficult to show along the lines of the previous proof that the maximum in (6) is still attained, and

$$\max_{p\in\mathcal{P}(\mathcal{X})} S(p\|q) = \begin{cases} \log \dfrac{1}{\min_x q(x)} & \text{if } \operatorname{supp} q = \mathcal{X}, \\ +\infty & \text{otherwise.} \end{cases}$$

2. Entropic measurement uncertainty relations

In general, the two target observables A and B, introduced at the end of Section 1.1, are incompatible, and only "approximate" joint measurements are possible for them. Moreover, any measurement of A may disturb a subsequent measurement of B, in a way that the resulting distribution of B can be very far from its unperturbed version; this disturbance may be present even when the two observables are compatible. Typically, such a disturbance of A on B can not be removed, nor just made arbitrarily small, unless we drop the requirement of exactly measuring A. However, in both cases, the measurement uncertainties on A and B can not always be made equally small. The quantum nature of A and B relates their measurement uncertainties, so that improving the approximation of A affects the quality of the corresponding approximation of B and vice versa. Incompatibility of A and B on the one hand, and the disturbance induced on B by a measurement of A on the other hand, are alternative manifestations of the quantum relation between the two observables, and as such deserve different approaches.

Our aim is now to quantify both these types of measurement uncertainty relations between A and B by means of suitable informational quantities. In the case of incompatible observables, we will find an *entropic incompatibility degree*, encoding the minimum total error affecting any approximate joint measurement of A and B. Similarly, when the observable B is measured after an approximate version of A, the resulting uncertainties on both observables will produce an *error/disturbance tradeoff* for A and B. In both cases, we will look for an optimal bi-observable M whose marginals $\mathsf{M}_{[1]}$ and $\mathsf{M}_{[2]}$ are the best approximations of the two target observables A and B. However, the different points of view will be reflected in the fact that we will optimize over M in two different sets, according to the case at hand.

2.1. Error function and entropic divergence for observables. We now regard any bi-observable $\mathsf{M} \in \mathcal{M}(\mathcal{X} \times \mathcal{Y})$ as an *approximate joint measurement* of A and B and we want an informational quantification of how far its marginals $\mathsf{M}_{[1]}$ and $\mathsf{M}_{[2]}$ are from correctly approximating the two target observables A and B. Following [6–8], these two approximations will be judged by comparing (within our entropic approach) the distribution $\mathsf{M}_{[1]}^{\rho}$ with A^{ρ}, and the distribution $\mathsf{M}_{[2]}^{\rho}$ with B^{ρ}, for all states ρ. Note that we can not compare the output of $\mathsf{M}_{[1]}$ with that of A, and the output of $\mathsf{M}_{[2]}$ with that of B, in one and the same experiment. Indeed, although our bi-observable M is a joint measurement of $\mathsf{M}_{[1]}$ and $\mathsf{M}_{[2]}$, there is no way to turn it into a joint measurement of the four observables A, $\mathsf{M}_{[1]}$, $\mathsf{M}_{[2]}$ and B, when A and B are not compatible. Nevertheless, even if A and B are incompatible, each of them can be measured in independent repetitions of a preparation (state) ρ of the system. Similarly, any bi-observable M can be measured in other independent repetitions of the same preparation. So, all the three probability distributions A^{ρ}, B^{ρ}, M^{ρ} can be estimated from independent experiments, and then they can be compared without any hypothesis of compatibility among A, B and M.

The first step is to quantify the inefficiency of the distribution approximations $\mathsf{A}^\rho \simeq \mathsf{M}^\rho_{[1]}$ and $\mathsf{B}^\rho \simeq \mathsf{M}^\rho_{[2]}$, given the bi-observable M. According to the discussion in Section 1.2, the natural way to quantify the loss of information in each approximation is to use the relative entropy. Remarkably, the relative entropy properties allow us to give a single quantification for the whole couple approximation $(\mathsf{A}^\rho, \mathsf{B}^\rho) \simeq (\mathsf{M}^\rho_{[1]}, \mathsf{M}^\rho_{[2]})$: since $S\big(\mathsf{A}^\rho \| \mathsf{M}^\rho_{[1]}\big)$ and $S\big(\mathsf{B}^\rho \| \mathsf{M}^\rho_{[2]}\big)$ are homogeneous and dimensionless, they can be added to give the total amount of information loss.

Definition 1. *For any bi-observable* $\mathsf{M} \in \mathcal{M}(\mathcal{X} \times \mathcal{Y})$, *the* error function *of the approximation* $(\mathsf{A}, \mathsf{B}) \simeq (\mathsf{M}_{[1]}, \mathsf{M}_{[2]})$ *is the state-dependent quantity*

$$S[\mathsf{A}, \mathsf{B} \| \mathsf{M}](\rho) = S\big(\mathsf{A}^\rho \| \mathsf{M}^\rho_{[1]}\big) + S\big(\mathsf{B}^\rho \| \mathsf{M}^\rho_{[2]}\big). \tag{7}$$

Note that the approximating distributions $\mathsf{M}^\rho_{[i]}$ appear in the second entry of the relative entropy, consistently with the discussion following its definition (3).

By Proposition 1, item (vii), we can rewrite (7) in the form

$$S[\mathsf{A}, \mathsf{B} \| \mathsf{M}](\rho) = S\big(\mathsf{A}^\rho \otimes \mathsf{B}^\rho \| \mathsf{M}^\rho_{[1]} \otimes \mathsf{M}^\rho_{[2]}\big). \tag{8}$$

It is important to note that the error function itself is a relative entropy; this can be mathematically useful in some situations (see e.g. the proof of Theorem 4). Note that, whether A and B are compatible or not, $\mathsf{A}^\rho \otimes \mathsf{B}^\rho$ is the distribution of their measurements in two independent preparations of the same state ρ.

The second step is to quantify the inefficiency of the observable approximations $\mathsf{A} \simeq \mathsf{M}_{[1]}$ and $\mathsf{B} \simeq \mathsf{M}_{[2]}$ by means of the marginals of a given bi-observable M, without reference to any particular state. In order to construct a state-independent quantity, we take the worst case in (7) with respect to the system state ρ.

Definition 2. *The* entropic divergence of $\mathsf{M} \in \mathcal{M}(\mathcal{X} \times \mathcal{Y})$ from (A, B) is *the quantity*

$$D\big(\mathsf{A}, \mathsf{B} \| \mathsf{M}\big) = \sup_{\rho \in \mathcal{S}(\mathcal{H})} S[\mathsf{A}, \mathsf{B} \| \mathsf{M}](\rho) \equiv \sup_{\rho \in \mathcal{S}(\mathcal{H})} \Big\{ S\big(\mathsf{A}^\rho \| \mathsf{M}^\rho_{[1]}\big) + S\big(\mathsf{B}^\rho \| \mathsf{M}^\rho_{[2]}\big) \Big\}. \tag{9}$$

The entropic divergence $D\big(\mathsf{A}, \mathsf{B} \| \mathsf{M}\big)$ quantifies the worst total loss of information due to the couple approximation $(\mathsf{A}, \mathsf{B}) \simeq (\mathsf{M}_{[1]}, \mathsf{M}_{[2]})$. Note that there is a unique supremum over ρ, so that $D\big(\mathsf{A}, \mathsf{B} \| \mathsf{M}\big)$ takes into account any possible balancing and compensation between the information losses in the first and in the second approximation. The entropic divergence depends only on $\mathsf{M}_{[1]}$ and $\mathsf{M}_{[2]}$, and so it is the same for different bi-observables with equal marginals. If A and B are compatible and M is any of their joint measurements, then $D\big(\mathsf{A}, \mathsf{B} \| \mathsf{M}\big) = 0$ by Proposition 1, item (ii).

Theorem 1. *Let* $\mathsf{A} \in \mathcal{M}(\mathcal{X})$, $\mathsf{B} \in \mathcal{M}(\mathcal{Y})$ *be the target observables. The error function and the entropic divergence defined above have the following properties.*

(i) The function $S[\mathsf{A}, \mathsf{B} \| \mathsf{M}] : \mathcal{S}(\mathcal{H}) \to [0, +\infty]$ *is convex and LSC,* $\forall \mathsf{M} \in \mathcal{M}(\mathcal{X} \times \mathcal{Y})$.
(ii) The function $D\big(\mathsf{A}, \mathsf{B} \| \cdot \big) : \mathcal{M}(\mathcal{X} \times \mathcal{Y}) \to [0, +\infty]$ *is convex and LSC.*
(iii) For any $\mathsf{M} \in \mathcal{M}(\mathcal{X} \times \mathcal{Y})$, *the following three statements are equivalent:*
  (a) $D\big(\mathsf{A}, \mathsf{B} \| \mathsf{M}\big) < +\infty$,
  (b) $\ker \mathsf{M}_{[1]}(x) \subseteq \ker \mathsf{A}(x),\ \forall x$, *and* $\ker \mathsf{M}_{[2]}(y) \subseteq \ker \mathsf{B}(y),\ \forall y$,

(c) $S[\mathsf{A},\mathsf{B}\|\mathsf{M}]$ *is bounded and continuous.*

(iv) $D\big(\mathsf{A},\mathsf{B}\|\mathsf{M}\big) = \max_{\rho\in\mathcal{S}(\mathcal{H}),\ \rho \text{ pure}} S[\mathsf{A},\mathsf{B}\|\mathsf{M}](\rho)$*, where the maximum can be any value in the extended interval* $[0,+\infty]$.

(v) The error $S[\mathsf{A},\mathsf{B}\|\mathsf{M}](\rho)$ *is invariant under an overall unitary conjugation of* A, B, M *and* ρ*, and a relabelling of the outcome spaces* $\mathcal{X}$ *and* $\mathcal{Y}$.

(vi) The entropic divergence $D\big(\mathsf{A},\mathsf{B}\|\mathsf{M}\big)$ *is invariant under an overall unitary conjugation of* A, B *and* M*, and a relabelling of the outcome spaces* $\mathcal{X}$ *and* $\mathcal{Y}$.

Proof. (i) The function $S[\mathsf{A},\mathsf{B}\|\mathsf{M}]$ is the sum of two terms which are convex, because the mapping $\rho\mapsto\mathsf{X}^\rho$ is affine for any observable X and by Proposition 1, item (v); hence $S[\mathsf{A},\mathsf{B}\|\mathsf{M}]$ is convex. Moreover, each term is LSC, since $\rho\mapsto\mathsf{X}^\rho$ is continuous and because of Proposition 1, item (vi); so the sum $S[\mathsf{A},\mathsf{B}\|\mathsf{M}]$ is LSC by [47, Prop. 1.5.12].

(ii) Each mapping $\mathsf{M}\mapsto\mathsf{M}^\rho_{[i]}$ is affine and continuous, and the functions $S\big(\mathsf{A}^\rho\|\cdot\big)$, $S\big(\mathsf{B}^\rho\|\cdot\big)$ are convex and LSC by Proposition 1, items (v) and (vi). It follows that $\mathsf{M}\mapsto S\big(\mathsf{A}^\rho\|\mathsf{M}^\rho_{[1]}\big)$ and $\mathsf{M}\mapsto S\big(\mathsf{B}^\rho\|\mathsf{M}^\rho_{[2]}\big)$ are also convex and LSC functions on $\mathcal{M}(\mathcal{X}\times\mathcal{Y})$; hence, such are their sum and the supremum $D\big(\mathsf{A},\mathsf{B}\|\cdot\big)$ [47, Prop. 1.5.12].

(iii) Let us show (a)⇒(b)⇒(c)⇒(a).

(a)⇒(b). If $\ker\mathsf{M}_{[1]}(x)\nsubseteq\ker\mathsf{A}(x)$ for some x, then we could take a pure state $\rho = |\,\psi\,\rangle\langle\,\psi\,|$ with ψ belonging to $\ker\mathsf{M}_{[1]}(x)$ but not to $\ker\mathsf{A}(x)$, so that $\mathsf{M}^\rho_{[1]}(x) = 0$ while $\mathsf{A}^\rho(x) > 0$; thus, we would get $S\big(\mathsf{A}^\rho\|\mathsf{M}^\rho_{[1]}\big) = +\infty$ and the contradiction $D\big(\mathsf{A},\mathsf{B}\|\mathsf{M}\big) = +\infty$.

(b)⇒(c). The function $S[\mathsf{A},\mathsf{B}\|\mathsf{M}]$ is a finite sum of terms of the kind $s\big(\mathsf{A}^\rho(x),\mathsf{M}^\rho_{[1]}(x)\big)$ or $s\big(\mathsf{B}^\rho(y),\mathsf{M}^\rho_{[2]}(y)\big)$, where s is the function defined in (5). Under the hypothesis (b), each of these terms is a bounded and continuous function of ρ by Lemma 1 below. We thus conclude that $S[\mathsf{A},\mathsf{B}\|\mathsf{M}]$ is bounded and continuous.

(c)⇒(a). Trivial, as $D\big(\mathsf{A},\mathsf{B}\|\mathsf{M}\big) = \sup_{\rho\in\mathcal{S}(\mathcal{H})} S[\mathsf{A},\mathsf{B}\|\mathsf{M}](\rho)$.

(iv) If $D\big(\mathsf{A},\mathsf{B}\|\mathsf{M}\big) < +\infty$, then $S[\mathsf{A},\mathsf{B}\|\mathsf{M}]$ is a bounded and continuous function on the compact set $\mathcal{S}(\mathcal{H})$ by item (iii) above, and thus it attains a maximum; moreover, $S[\mathsf{A},\mathsf{B}\|\mathsf{M}]$ is convex, hence it has at least a maximum point among the extreme points of $\mathcal{S}(\mathcal{H})$, which are the pure states. If instead $D\big(\mathsf{A},\mathsf{B}\|\mathsf{M}\big) = +\infty$, then $\ker\mathsf{M}_{[1]}(x)\nsubseteq\ker\mathsf{A}(x)$ for some x, or $\ker\mathsf{M}_{[2]}(y)\nsubseteq\ker\mathsf{B}(y)$ for some y again by item (iii). In this case, every pure state $\rho = |\,\psi\,\rangle\langle\,\psi\,|$ with $\psi\in\ker\mathsf{M}_{[1]}(x)\setminus\ker\mathsf{A}(x)$, or $\psi\in\ker\mathsf{M}_{[2]}(y)\setminus\ker\mathsf{B}(y)$, is such that $S[\mathsf{A},\mathsf{B}\|\mathsf{M}](\rho) = +\infty$, and thus it is a maximum point of $S[\mathsf{A},\mathsf{B}\|\mathsf{M}]$.

(v) For any unitary operator U on $\mathcal{H}$, we have $(U^*\mathsf{A}U)^{U^*\rho U} = \mathsf{A}^\rho$, $(U^*\mathsf{B}U)^{U^*\rho U} = \mathsf{B}^\rho$, and, since $(U^*\mathsf{M}U)_{[i]} = U^*\mathsf{M}_{[i]}U$, also $(U^*\mathsf{M}U)^{U^*\rho U}_{[i]} = \mathsf{M}^\rho_{[i]}$. Therefore, by the definition (7) of the error function, we get the equality

$$S[U^*\mathsf{A}U,\, U^*\mathsf{B}U\|U^*\mathsf{M}U](U^*\rho U) = S[\mathsf{A},\mathsf{B}\|\mathsf{M}](\rho).$$

The invariance under relabelling of the outcomes is an immediate consequence of the analogous property of the relative entropy (Proposition 1, item (iv)).

(vi) The two invariances immediately follow by the previous item. We check only the first one:

$$\begin{aligned} D\big(U^*\mathsf{A}U,\, U^*\mathsf{B}U\|U^*\mathsf{M}U\big) &= \sup_{\rho\in\mathcal{S}(\mathcal{H})} S[U^*\mathsf{A}U,\, U^*\mathsf{B}U\|U^*\mathsf{M}U](\rho) \\ &= \sup_{\rho\in\mathcal{S}(\mathcal{H})} S[U^*\mathsf{A}U,\, U^*\mathsf{B}U\|U^*\mathsf{M}U](U^*\rho U) = \sup_{\rho\in\mathcal{S}(\mathcal{H})} S[\mathsf{A},\mathsf{B}\|\mathsf{M}](\rho) \\ &= D\big(\mathsf{A},\mathsf{B}\|\mathsf{M}\big), \end{aligned}$$

where in the second equality we have used the fact that $U\mathcal{S}(\mathcal{H})U^* = \mathcal{S}(\mathcal{H})$. □

An essential step in the last proof is the following lemma.

Lemma 1. *Suppose* $A, B \in \mathcal{L}(\mathcal{H})$ *are such that* $0 \leq A \leq \mathbb{1}$ *and* $0 \leq B \leq \mathbb{1}$*, and assume that* $\ker B \subseteq \ker A$*. Let* $A^\rho = \operatorname{Tr}\{A\rho\}$*,* $B^\rho = \operatorname{Tr}\{B\rho\}$*, and let* s *be the function defined in* (5)*. Then, the function* $s_{A,B} : \mathcal{S}(\mathcal{H}) \to [0,+\infty]$*, with* $s_{A,B}(\rho) = s(A^\rho, B^\rho)$*, is bounded and continuous.*

Proof. We will show that $s_{A,B}$ is a continuous function on $\mathcal{S}(\mathcal{H})$; since $\mathcal{S}(\mathcal{H})$ is compact, this will also imply that $s_{A,B}$ is bounded. The case $B = 0$ is trivial, hence we will suppose $B \neq 0$. By the hypotheses, the condition $B^\rho = 0$ implies that $A^\rho = 0$. The definition (5) of s then gives

$$s_{A,B}(\rho) = \begin{cases} A^\rho \log \dfrac{A^\rho}{B^\rho} & \text{if } A^\rho > 0 \text{ and } B^\rho > 0 \\ 0 & \text{if } A^\rho = 0 \text{ and } B^\rho > 0 \\ 0 & \text{if } A^\rho = 0 \text{ and } B^\rho = 0 \end{cases} = \begin{cases} B^\rho h\left(\dfrac{A^\rho}{B^\rho}\right) & \text{if } B^\rho > 0 \\ 0 & \text{if } B^\rho = 0 \end{cases}$$

where we have introduced the continuous function $h : [0,+\infty) \to [-1/2,+\infty)$, with $h(t) = t\log t$ if $t > 0$, and $h(0) = 0$. The function $s_{A,B}$ is clearly continuous on the open subset $\mathcal{U} = \{\rho \in \mathcal{S}(\mathcal{H}) : B^\rho > 0\}$ of the state space $\mathcal{S}(\mathcal{H})$. It remains to show that it is also continuous at all the points of the set $\mathcal{U}^c = \{\rho \in \mathcal{S}(\mathcal{H}) : B^\rho = 0\}$. To this aim, observe that

$$A \leq c_{\max}(A)P_A \leq c_{\max}(A)P_B \qquad \text{and} \qquad B \geq c_{\min}(B)P_B,$$

where $c_{\max}(A)$ is the maximum eigenvalue of A, $c_{\min}(B)$ is the minimum positive eigenvalue of B, and we denote by P_A and P_B the orthogonal projections onto $\ker A^\perp$ and $\ker B^\perp$, respectively. Since $P_B^\rho \neq 0$ for all ρ such that $B^\rho > 0$, it follows that

$$0 \leq \frac{A^\rho}{B^\rho} \leq \frac{c_{\max}(A)}{c_{\min}(B)}\,, \qquad \forall \rho \in \mathcal{U}.$$

Hence, by continuity of h and boundedness of the interval $[0, c_{\max}(A)/c_{\min}(B)]$, there is a constant $M > 0$ such that

$$|s_{A,B}(\rho)| = \left| B^\rho h\left(\frac{A^\rho}{B^\rho}\right)\right| \leq MB^\rho, \qquad \forall \rho \in \mathcal{U}.$$

On the other hand, for $\rho \in \mathcal{U}^c$ we have $s_{A,B}(\rho) = 0$. If $(\rho_k)_k$ is a sequence in $\mathcal{S}(\mathcal{H})$ converging to $\rho_0 \in \mathcal{U}^c$, then $|s_{A,B}(\rho_k) - s_{A,B}(\rho_0)| \leq MB^{\rho_k} \underset{k\to\infty}{\longrightarrow} 0$, which shows that $s_{A,B}$ is continuous at ρ_0. □

2.2. Incompatibility degree, error/disturbance coefficient, and optimal approximate joint measurements. After introducing the error function $S[\mathsf{A},\mathsf{B}\|\mathsf{M}](\rho)$, which describes the total information lost by measuring the bi-observable M in place of A and B in the state ρ, and after defining its maximum value $D\big(\mathsf{A},\mathsf{B}\|\mathsf{M}\big)$ over all states, the third step is to quantify the intrinsic measurement uncertainties between A and B, dropping any reference to a particular state or approximating joint measurement. When we are interested in incompatibility, this is done by taking the minimum of the divergence $D\big(\mathsf{A},\mathsf{B}\|\mathsf{M}\big)$ over all possible bi-observables $\mathsf{M}\in\mathcal{M}(\mathcal{X}\times\mathcal{Y})$. The resulting quantity is the minimum inefficiency which can not be avoided when the (possibly incompatible) observables A and B are approximated by the compatible marginals $\mathsf{M}_{[1]}$ and $\mathsf{M}_{[2]}$ of any bi-observable M. This minimum can be understood as an "incompatibility degree" of the two observables A and B.

Definition 3. *The* entropic incompatibility degree $c_{\rm inc}(\mathsf{A},\mathsf{B})$ *of the observables* A *and* B *is*

$$c_{\rm inc}(\mathsf{A},\mathsf{B}) = \inf_{\mathsf{M}\in\mathcal{M}(\mathcal{X}\times\mathcal{Y})} D\big(\mathsf{A},\mathsf{B}\|\mathsf{M}\big) \equiv \inf_{\mathsf{M}\in\mathcal{M}(\mathcal{X}\times\mathcal{Y})} \sup_{\rho\in\mathcal{S}(\mathcal{H})} S[\mathsf{A},\mathsf{B}\|\mathsf{M}](\rho). \qquad (10)$$

The definition is consistent, as obviously $c_{\rm inc}(\mathsf{A},\mathsf{B})\geq 0$, and $c_{\rm inc}(\mathsf{A},\mathsf{B})=0$ when A and B are compatible. As the notion of incompatibility is symmetric by exchanging the observables A and B, we would expect that also the incompatibility degree satisfies the property $c_{\rm inc}(\mathsf{A},\mathsf{B}) = c_{\rm inc}(\mathsf{B},\mathsf{A})$. Indeed, this is actually true, as $D\big(\mathsf{A},\mathsf{B}\|\mathsf{M}\big) = D\big(\mathsf{B},\mathsf{A}\|\mathsf{M}'\big)$ for all $\mathsf{M}\in\mathcal{M}(\mathcal{X}\times\mathcal{Y})$, where $\mathsf{M}'\in\mathcal{M}(\mathcal{Y}\times\mathcal{X})$ is defined by $\mathsf{M}'(y,x) = \mathsf{M}(x,y)$. Note that the symmetry of $c_{\rm inc}$ comes from the fact that, in defining the error function $S[\mathsf{A},\mathsf{B}\|\mathsf{M}]$, we have chosen equal weights for the contributions of the two approximation errors of A and B.

On the other hand, when we deal with the error/disturbance uncertainty relation, our analysis is restricted to the bi-observables describing sequential measurements of an approximate version A' of A, followed by an exact measurement of B. In other words, we focus on

$$\begin{aligned}\mathcal{M}(\mathcal{X};\mathsf{B}) &= \{\mathcal{J}^*(\mathsf{B}) : \mathcal{J}\in\mathcal{J}(\mathcal{X})\}\\ &= \{\mathsf{M}\in\mathcal{M}(\mathcal{X}\times\mathcal{Y}) : \mathsf{M}(x,y) = \mathcal{J}^*_x[\mathsf{B}(y)]\ \forall x,y, \text{ for some } \mathcal{J}\in\mathcal{J}(\mathcal{X})\}, \qquad (11)\end{aligned}$$

the subset of $\mathcal{M}(\mathcal{X}\times\mathcal{Y})$ consisting of the sequential measurements where the first outcome set $\mathcal{X}$ and the second observable B are fixed. If $\mathsf{M} = \mathcal{J}^*(\mathsf{B})\in\mathcal{M}(\mathcal{X};\mathsf{B})$, then $\mathsf{A}' = \mathsf{M}_{[1]} = \mathcal{J}^*_\cdot[\mathbb{1}]$ is the observable approximating A, and $\mathsf{B}' = \mathcal{J}^*_{\mathcal{X}}[\mathsf{B}(\cdot)]$ is the version of B perturbed by the measurement of A'. In general, it may equally well be $\mathsf{A}'\neq\mathsf{A}$ and $\mathsf{B}'\neq\mathsf{B}$, unless the observable A can be measured without disturbing B [27].

In order to quantify the measurement uncertainties due to the error/disturbance tradeoff, we then consider the minimum of the entropic divergence $D\big(\mathsf{A},\mathsf{B}\|\mathsf{M}\big)$ for $\mathsf{M}\in\mathcal{M}(\mathcal{X};\mathsf{B})$. If we read $S\big(\mathsf{A}^\rho\|\mathsf{M}^\rho_{[1]}\big)$ as the error made by $\mathcal{J}$ in measuring A in the state ρ, and $S\big(\mathsf{B}^\rho\|\mathsf{M}^\rho_{[2]}\big)$ as the amount of disturbance introduced by $\mathcal{J}$ on the subsequent measurement of B, then the divergence $D\big(\mathsf{A},\mathsf{B}\|\mathsf{M}\big)$ expresses the sum error + disturbance maximized over all states for the sequential measurement M. Minimizing $D\big(\mathsf{A},\mathsf{B}\|\mathsf{M}\big)$ over all sequential measurements, we then obtain the following entropic quantification of the error/disturbance tradeoff between A and B.

Definition 4. *The* entropic error/disturbance coefficient $c_{\mathrm{ed}}(\mathsf{A},\mathsf{B})$ *of* A *followed by* B *is*

$$c_{\mathrm{ed}}(\mathsf{A},\mathsf{B}) = \inf_{\mathsf{M}\in\mathcal{M}(\mathcal{X};\mathsf{B})} D\big(\mathsf{A},\mathsf{B}\|\mathsf{M}\big) \equiv \inf_{\mathsf{M}\in\mathcal{M}(\mathcal{X};\mathsf{B})} \sup_{\rho\in\mathcal{S}(\mathcal{H})} S[\mathsf{A},\mathsf{B}\|\mathsf{M}](\rho). \tag{12}$$

Similarly to the incompatibility degree, the error/disturbance coefficient is always nonnegative, and $c_{\mathrm{ed}}(\mathsf{A},\mathsf{B}) = 0$ when A can be measured without disturbing B, i.e. A and B are sequentially compatible. Contrary to c_{inc}, we stress that in general the two indexes $c_{\mathrm{ed}}(\mathsf{A},\mathsf{B})$ and $c_{\mathrm{ed}}(\mathsf{B},\mathsf{A})$ can be different, as shown in Remark 1 below.

When the approximate measurement of the first observable A is described by the instrument $\mathcal{J}$, the measurement of the second fixed observable B could be preceded by any kind of correction taking into account the observed outcome x [7]. This can be formalized by inserting a quantum channel $\mathcal{C}_x$ in between the measurements of A and B. As the composition $\mathcal{J}'_x = \mathcal{C}_x \circ \mathcal{J}_x$ gives again an instrument $\mathcal{J}' \in \mathcal{J}(\mathcal{X})$, we then see that any possible correction is considered when we take the infimum in $\mathcal{M}(\mathcal{X};\mathsf{B})$. The latter fact shows that Definition 4 is consistent, since only by taking into account all possible corrections we can properly speak of pure unavoidable disturbance and of error/disturbance tradeoff.

Comparing the two indexes c_{inc} and c_{ed}, the inequality $c_{\mathrm{inc}}(\mathsf{A},\mathsf{B}) \leq c_{\mathrm{ed}}(\mathsf{A},\mathsf{B})$ trivially follows from the inclusion $\mathcal{M}(\mathcal{X};\mathsf{B}) \subseteq \mathcal{M}(\mathcal{X}\times\mathcal{Y})$. This means that, even if one is interested in c_{ed}, the most symmetric index c_{inc} is at least a lower bound for it. We stress that the inclusion $\mathcal{M}(\mathcal{X};\mathsf{B}) \subseteq \mathcal{M}(\mathcal{X}\times\mathcal{Y})$ may be strict in general. For example, there may exist observables which are compatible with B, but can not be measured before B without disturbing it. Then, taken such an observable A, a joint measurement of A and B clearly belongs to $\mathcal{M}(\mathcal{X}\times\mathcal{Y})$ but can not be in $\mathcal{M}(\mathcal{X};\mathsf{B})$. When $\mathcal{M}(\mathcal{X};\mathsf{B}) \subsetneq \mathcal{M}(\mathcal{X}\times\mathcal{Y})$, the incompatibility and error/disturbance approaches definitely are not equivalent. Nevertheless, there is one remarkable situation in which they are the same.

Proposition 3. *If* $\mathsf{B} \in \mathcal{M}(\mathcal{Y})$ *is a sharp observable, then* $\mathcal{M}(\mathcal{X};\mathsf{B}) = \mathcal{M}(\mathcal{X}\times\mathcal{Y})$.

Proof. The proof directly follows from the argument at the end of [27, Sect. II.D]. Indeed, for any $\mathsf{M} \in \mathcal{M}(\mathcal{X}\times\mathcal{Y})$, we can define the instrument $\mathcal{J} \in \mathcal{J}(\mathcal{X})$ with

$$\mathcal{J}_x[\rho] = \sum_{y\in\mathcal{Y}:\mathsf{B}(y)\neq 0} \mathrm{Tr}\,\{\rho\mathsf{M}(x,y)\}\,\frac{\mathsf{B}(y)}{\mathrm{Tr}\,\{\mathsf{B}(y)\}}.$$

For such an instrument, the equality $\mathsf{M}(x,y) = \mathcal{J}_x^*[\mathsf{B}(y)]$ is immediate. □

As an immediate consequence of this result, we have $c_{\mathrm{ed}}(\mathsf{A},\mathsf{B}) = c_{\mathrm{inc}}(\mathsf{A},\mathsf{B})$ whenever the second measured observable B is sharp.

By Theorem 2 below, the two infima in the definitions of $c_{\mathrm{inc}}(\mathsf{A},\mathsf{B})$ and $c_{\mathrm{ed}}(\mathsf{A},\mathsf{B})$ are actually two minima. It is convenient to give a name to the corresponding sets of minimizing bi-observables:

$$\mathcal{M}_{\mathrm{inc}}(\mathsf{A},\mathsf{B}) = \operatorname*{arg\,min}_{\mathsf{M}\in\mathcal{M}(\mathcal{X}\times\mathcal{Y})} D\big(\mathsf{A},\mathsf{B}\|\mathsf{M}\big), \qquad \mathcal{M}_{\mathrm{ed}}(\mathsf{A},\mathsf{B}) = \operatorname*{arg\,min}_{\mathsf{M}\in\mathcal{M}(\mathcal{X};\mathsf{B})} D\big(\mathsf{A},\mathsf{B}\|\mathsf{M}\big).$$

We can say that $\mathcal{M}_{\mathrm{inc}}(\mathsf{A},\mathsf{B})$ is the set of the *optimal approximate joint measurements of* A *and* B. Similarly, $\mathcal{M}_{\mathrm{ed}}(\mathsf{A},\mathsf{B})$ contains the sequential measurements optimally approximating A and B.

The next theorem summarizes the main properties of c_{inc} and c_{ed} contained in the above discussion, and states some further relevant facts about the two indexes.

Theorem 2. *Let* $\mathsf{A} \in \mathcal{M}(\mathcal{X})$*,* $\mathsf{B} \in \mathcal{M}(\mathcal{Y})$ *be the target observables. For the entropic coefficients defined above the following properties hold.*

(i) The coefficients $c_{\rm inc}(\mathsf{A},\mathsf{B})$ *and* $c_{\rm ed}(\mathsf{A},\mathsf{B})$ *are invariant under an overall unitary conjugation of the observables* A *and* B*, and they do not depend on the labelling of the outcomes in* $\mathcal{X}$ *and* $\mathcal{Y}$*.*

(ii) The incompatibility degree has the exchange symmetry $c_{\rm inc}(\mathsf{A},\mathsf{B}) = c_{\rm inc}(\mathsf{B},\mathsf{A})$*.*

(iii) We have $0 \leq c_{\rm inc}(\mathsf{A},\mathsf{B}) \leq c_{\rm ed}(\mathsf{A},\mathsf{B}) \leq \log|\mathcal{X}| - \inf_{\rho\in\mathcal{S}(\mathcal{H})} H(\mathsf{A}^\rho)$ *and* $c_{\rm inc}(\mathsf{A},\mathsf{B}) \leq \log|\mathcal{Y}| - \inf_{\rho\in\mathcal{S}(\mathcal{H})} H(\mathsf{B}^\rho)$.

(iv) The sets $\mathcal{M}_{\rm inc}(\mathsf{A},\mathsf{B})$ *and* $\mathcal{M}_{\rm ed}(\mathsf{A},\mathsf{B})$ *are nonempty convex compact subsets of* $\mathcal{M}(\mathcal{X}\times\mathcal{Y})$.

(v) $c_{\rm inc}(\mathsf{A},\mathsf{B}) = 0$ *if and only if the observables* A *and* B *are compatible, and in this case* $\mathcal{M}_{\rm inc}(\mathsf{A},\mathsf{B})$ *is the set of all their joint measurements.*

(vi) $c_{\rm ed}(\mathsf{A},\mathsf{B}) = 0$ *if and only if the observables* A *and* B *are sequentially compatible, and in this case* $\mathcal{M}_{\rm ed}(\mathsf{A},\mathsf{B})$ *is the set of all the sequential measurements of* A *followed by* B.

(vii) If B *is sharp, then* $\mathcal{M}_{\rm inc}(\mathsf{A},\mathsf{B}) = \mathcal{M}_{\rm ed}(\mathsf{A},\mathsf{B})$ *and* $c_{\rm inc}(\mathsf{A},\mathsf{B}) = c_{\rm ed}(\mathsf{A},\mathsf{B})$.

Proof. (i) The invariance under unitary conjugation follows from the corresponding property of the entropic divergence (Theorem 1, item (vi)). We will prove it only for $c_{\rm ed}$, the case of $c_{\rm inc}$ being even simpler. We have

$$c_{\rm ed}(U^*\mathsf{A}U, U^*\mathsf{B}U) = \inf_{\mathsf{M}\in\mathcal{M}(\mathcal{X};U^*\mathsf{B}U)} D\big(U^*\mathsf{A}U, U^*\mathsf{B}U\|\mathsf{M}\big) = \inf_{\mathsf{M}'\in U\mathcal{M}(\mathcal{X};U^*\mathsf{B}U)U^*} D\big(\mathsf{A},\mathsf{B}\|\mathsf{M}'\big),$$

and, in order to show that $c_{\rm ed}(U^*\mathsf{A}U, U^*\mathsf{B}U) = c_{\rm ed}(\mathsf{A},\mathsf{B})$, it only remains to prove the set equality $U\mathcal{M}(\mathcal{X};U^*\mathsf{B}U)U^* = \mathcal{M}(\mathcal{X};\mathsf{B})$. If $\mathsf{M} = \mathcal{J}^*(U^*\mathsf{B}U) \in \mathcal{M}(\mathcal{X};U^*\mathsf{B}U)$, then, defining the instrument $\mathcal{J}'_x[\rho] = U\mathcal{J}_x[U^*\rho U]U^*$, $\forall\rho, x$, we have $U\mathsf{M}U^* = \mathcal{J}'^*(\mathsf{B}) \in \mathcal{M}(\mathcal{X};\mathsf{B})$, as claimed. In a similar way, the invariance under relabelling of the outcomes is a consequence of the analogous property of the entropic divergence.

(ii) This property has already been noticed.

(iii) The positivity the inequality between the two indexes have already been noticed. Then, let $\mathcal{U} \in \mathcal{J}(\mathcal{X})$ be the trivial uniform instrument $\mathcal{U}_x[\rho] = u_\mathcal{X}(x)\rho$. Taking the sequential measurement $\mathcal{U}^*(\mathsf{B}) \in \mathcal{M}(\mathcal{X};\mathsf{B})$, we get $\mathcal{U}^*(\mathsf{B})^\rho = u_\mathcal{X}\otimes\mathsf{B}^\rho$ and

$$S\big(\mathsf{A}^\rho\|\mathcal{U}^*(\mathsf{B})^\rho_{[1]}\big) + S\big(\mathsf{B}^\rho\|\mathcal{U}^*(\mathsf{B})^\rho_{[2]}\big) = S\big(\mathsf{A}^\rho\|u_\mathcal{X}\big) = \log|\mathcal{X}| - H\big(\mathsf{A}^\rho\big),$$

where the last equality follows from Proposition 1, item (iii). By taking the supremum over all the states, we get $D\big(\mathsf{A},\mathsf{B}\|\mathcal{U}^*(\mathsf{B})\big) = \log|\mathcal{X}| - \inf_{\rho\in\mathcal{S}(\mathcal{H})} H(\mathsf{A}^\rho)$, hence $c_{\rm ed}(\mathsf{A},\mathsf{B}) \leq \log|\mathcal{X}| - \inf_{\rho\in\mathcal{S}(\mathcal{H})} H(\mathsf{A}^\rho)$ by definition. The last inequality then follows by item (ii).

(iv) By item (ii) of Theorem 1 and item (iii) just above, $D\big(\mathsf{A},\mathsf{B}\|\cdot\big)$ is a convex LSC proper (i.e. not identically $+\infty$) function on the compact set $\mathcal{M}(\mathcal{X}\times\mathcal{Y})$. This implies that $\mathcal{M}_{\rm inc}(\mathsf{A},\mathsf{B}) \neq \emptyset$ [47, Exerc. E.1.6]. Closedness and convexity of $\mathcal{M}_{\rm inc}(\mathsf{A},\mathsf{B})$ are then easy and standard consequences of $D\big(\mathsf{A},\mathsf{B}\|\cdot\big)$ being convex and LSC. On the other hand, the set $\mathcal{M}(\mathcal{X};\mathsf{B})$ is a convex and compact subset of $\mathcal{M}(\mathcal{X}\times\mathcal{Y})$; indeed, this follows from convexity and compactness of $\mathcal{J}(\mathcal{X})$ and continuity of the mapping $\mathcal{J}\mapsto\mathcal{J}^*(\mathsf{B})$ in the definition (11). The proof that the subset $\mathcal{M}_{\rm ed}(\mathsf{A},\mathsf{B}) \subseteq \mathcal{M}(\mathcal{X};\mathsf{B})$ is nonempty, convex and compact then follows along the same lines of $\mathcal{M}_{\rm inc}(\mathsf{A},\mathsf{B})$.

(v) Assume $c_{\rm inc}(\mathsf{A},\mathsf{B})=0$. Then $\mathcal{M}_{\rm inc}(\mathsf{A},\mathsf{B})$ exactly consists of all the joint measurements of A and B, which therefore turn out to be compatible, as $\mathcal{M}_{\rm inc}(\mathsf{A},\mathsf{B})\neq\emptyset$ by (iv). Indeed, if $\mathsf{M}\in\mathcal{M}_{\rm inc}(\mathsf{A},\mathsf{B})$, then $0=c_{\rm inc}(\mathsf{A},\mathsf{B})=D\big(\mathsf{A},\mathsf{B}\|\mathsf{M}\big)$, which gives $S\big(\mathsf{A}^\rho\|\mathsf{M}^\rho_{[1]}\big)=S\big(\mathsf{B}^\rho\|\mathsf{M}^\rho_{[2]}\big)=0$ for all ρ. By Proposition 1, item (ii), this yields $\mathsf{A}^\rho=\mathsf{M}^\rho_{[1]}$, $\mathsf{B}=\mathsf{M}^\rho_{[2]}$, $\forall\rho$, and so $\mathsf{A}=\mathsf{M}_{[1]}$, $\mathsf{B}=\mathsf{M}_{[2]}$, which means that M is a joint measurement of A and B. The converse implication was already noticed in the text.

(vi) Similarly to the previous item, if $c_{\rm ed}(\mathsf{A},\mathsf{B})=0$, then $\mathcal{M}_{\rm ed}(\mathsf{A},\mathsf{B})$ consists exactly of all the sequential measurements of A followed by B. Indeed, by the same argument of (v), if $\mathsf{M}\in\mathcal{M}_{\rm ed}(\mathsf{A},\mathsf{B})$, then M is a joint measurement of A and B; since $\mathcal{M}_{\rm ed}(\mathsf{A},\mathsf{B})\subseteq\mathcal{M}(\mathcal{X};\mathsf{B})$, such a M is also a sequential measurement. As $\mathcal{M}_{\rm ed}(\mathsf{A},\mathsf{B})\neq\emptyset$ by (iv), this proves that A and B are sequentially compatible. The other implication is trivial and was already remarked.

(vii) As observed above, if B is sharp, then by Proposition 3 we have $\mathcal{M}(\mathcal{X};\mathsf{B})=\mathcal{M}(\mathcal{X}\times\mathcal{Y})$, which implies the claim. □

Item (iii) implies that the two indexes $c_{\rm inc}$ and $c_{\rm ed}$ are always finite, although the relative entropy $S\big(p\|q\big)$ is infinite whenever $\operatorname{supp}q\not\supseteq\operatorname{supp}p$. Actually, such a feature of S has a role: because of Theorem 1, item (iii), a bi-observable M is immediately discarded as a very bad approximation of A and B whenever $\ker\mathsf{M}_{[1]}(x)\not\subseteq\ker\mathsf{A}(x)$ for some x, or $\ker\mathsf{M}_{[2]}(y)\not\subseteq\ker\mathsf{B}(y)$ for some y.

We see in items (v) and (vi) that $c_{\rm inc}$ and $c_{\rm ed}$ have the desirable feature of being zero exactly when the two observables A and B satisfy the corresponding compatibility or nondisturbance property. We also stress that, by their very definitions, $c_{\rm inc}(\mathsf{A},\mathsf{B})$ and $c_{\rm ed}(\mathsf{A},\mathsf{B})$ are independent of both the preparations ρ and the approximating bi-observables M, as well as they satisfy the natural invariance properties of item (i). In view of these facts, we are allowed once more to say that the two bounds $c_{\rm inc}(\mathsf{A},\mathsf{B})$ and $c_{\rm ed}(\mathsf{A},\mathsf{B})$ are proper quantifications of the intrinsic incompatibility and error/disturbance affecting the two observables A and B.

We stress that the definitions of $c_{\rm inc}(\mathsf{A},\mathsf{B})$ and $c_{\rm ed}(\mathsf{A},\mathsf{B})$ are rather implicit. Indeed, even if we proved that they are strictly positive when A and B are incompatible (or sequentially incompatible), their evaluation requires the two optimizations "sup" on the states and "inf" on the measurements. Nevertheless, in some cases explicit computations are possible (even including the evaluation of the optimal approximate joint measurements) or explicit lower bounds can be exhibited, see Sections 3.2 and 3.3.

Remark 1. Item (vii) of Theorem 2 says that the two indexes coincide in the important case in which B is sharp. However, this is not true in general, as shown e.g. by the two examples in Appendix A (taken from [27]). In the first example, $\dim\mathcal{H}=3$, $|\mathcal{X}|=2$, $|\mathcal{Y}|=5$, and we have $c_{\rm ed}(\mathsf{A},\mathsf{B})>c_{\rm ed}(\mathsf{B},\mathsf{A})=c_{\rm inc}(\mathsf{A},\mathsf{B})=0$. The second example is more symmetric and simpler ($|\mathcal{X}|=|\mathcal{Y}|=2$), and it yields $c_{\rm ed}(\mathsf{A},\mathsf{B})>c_{\rm inc}(\mathsf{A},\mathsf{B})=0$ and also $c_{\rm ed}(\mathsf{B},\mathsf{A})>0$.

2.3. Entropic MURs. By definition, the two coefficients (10) and (12) are lower bounds for the entropic divergence (9) of every bi-observable M from (A,B):

$$\begin{aligned} &D\big(\mathsf{A},\mathsf{B}\|\mathsf{M}\big)\geq c_{\rm inc}(\mathsf{A},\mathsf{B}), \quad \forall\mathsf{M}\in\mathcal{M}(\mathcal{X}\times\mathcal{Y});\\ &D\big(\mathsf{A},\mathsf{B}\|\mathsf{M}\big)\geq c_{\rm ed}(\mathsf{A},\mathsf{B}), \quad \forall\mathsf{M}\in\mathcal{M}(\mathcal{X};\mathsf{B}). \end{aligned} \tag{13}$$

By items (v) and (vi) of Theorem 2, the two inequalities are non trivial and, by item (iv), both bounds are tight. As $D\big(\mathsf{A},\mathsf{B}\|\mathsf{M}\big)$ is a state-independent quantification of the inefficiency of the observable approximations $\mathsf{A}\simeq\mathsf{M}_{[1]}$ and $\mathsf{B}\simeq\mathsf{M}_{[2]}$, the inequalities (13) are two state-independent formulations of entropic MURs.

Since the definition of $D\big(\mathsf{A},\mathsf{B}\|\mathsf{M}\big)$ involves a unique supremum over ρ, by Theorem 1, item (iv), we can also reformulate the entropic MURs (13) as statements about the total loss of information that occurs in one preparation of the system:

$$\begin{aligned}\forall\mathsf{M}\in\mathcal{M}(\mathcal{X}\times\mathcal{Y}),\ \exists\rho\in\mathcal{S}(\mathcal{H}):\ S\big(\mathsf{A}^\rho\|\mathsf{M}^\rho_{[1]}\big)+S\big(\mathsf{B}^\rho\|\mathsf{M}^\rho_{[2]}\big)\geq c_{\mathrm{inc}}(\mathsf{A},\mathsf{B});\\ \forall\mathsf{M}\in\mathcal{M}(\mathcal{X};\mathsf{B}),\ \exists\rho\in\mathcal{S}(\mathcal{H}):\ S\big(\mathsf{A}^\rho\|\mathsf{M}^\rho_{[1]}\big)+S\big(\mathsf{B}^\rho\|\mathsf{M}^\rho_{[2]}\big)\geq c_{\mathrm{ed}}(\mathsf{A},\mathsf{B}).\end{aligned}\tag{14}$$

So, in an approximate joint measurement of A and B, the total loss of information can not be arbitrarily reduced: it depends on the state ρ, but potentially it can be as large as $c_{\mathrm{inc}}(\mathsf{A},\mathsf{B})$. Similarly, in a sequential measurement of A and B, there is a tradeoff between the information lost in the first measurement (because of the approximation error) and the information lost in the second measurement (because of the disturbance): they both depend on the state ρ, but potentially their sum can be as large as $c_{\mathrm{ed}}(\mathsf{A},\mathsf{B})$.

The indexes $c_{\mathrm{inc}}(\mathsf{A},\mathsf{B})$ and $c_{\mathrm{ed}}(\mathsf{A},\mathsf{B})$ are state-independent by their very definitions; however, the corresponding MURs (14) only refer to the worst possible state ρ for the measurement M at hand. Such a state-dependency is a general feature of MURs [7, Sect. C]: no MUR can provide a non trivial bound for the error of the approximation $(\mathsf{A}^\rho,\mathsf{B}^\rho)\simeq(\mathsf{M}^\rho_{[1]},\mathsf{M}^\rho_{[2]})$, holding for all states ρ in any approximate joint measurement M. Indeed, for a fixed $\rho\in\mathcal{S}(\mathcal{H})$, the trivial bi-observable $\mathsf{M}(x,y)=\mathsf{A}^\rho(x)\mathsf{B}^\rho(y)\mathbb{1}$ gives $(\mathsf{A}^\rho,\mathsf{B}^\rho)=(\mathsf{M}^\rho_{[1]},\mathsf{M}^\rho_{[2]})$; hence, it perfectly approximates the target observables in the state ρ whatever criterion one chooses for defining the error.

Here, in some detail, let us compare our MURs with Busch, Lahti and Werner's approach based on Wasserstein (or transport) distances (in the following, BLW approach; see [6–8]). As for BLW, our starting point is just giving a quantification of the error in the distribution approximation $\mathsf{A}^\rho\simeq\mathsf{M}^\rho_{[1]}$ (or $\mathsf{B}^\rho\simeq\mathsf{M}^\rho_{[2]}$). Anyway, employing the relative entropy in place of a Wasserstein distance reflects a different point of view, with some immediate consequences. BLW use a Wasserstein distance $d(\mathsf{A}^\rho,\mathsf{M}^\rho_{[1]})$ because they want that the error reflects the metric structure of the underlying outputs $\mathcal{X}$; since the units of measurement of $\mathcal{X}$ and $\mathcal{Y}$ may not be homogeneous, this essentially leads to quantifying the error of the whole couple approximation $(\mathsf{A}^\rho,\mathsf{B}^\rho)\simeq(\mathsf{M}^\rho_{[1]},\mathsf{M}^\rho_{[2]})$ with the dimensional pair $\big(d(\mathsf{A}^\rho,\mathsf{M}^\rho_{[1]}),d(\mathsf{B}^\rho,\mathsf{M}^\rho_{[2]})\big)$. On the contrary, the relative entropy is homogeneous and scale invariant; thus, it allows us to quantify the error of the couple approximation $(\mathsf{A}^\rho,\mathsf{B}^\rho)\simeq(\mathsf{M}^\rho_{[1]},\mathsf{M}^\rho_{[2]})$ with the single, dimensionless and scalar total error $S\big(\mathsf{A}^\rho\|\mathsf{M}^\rho_{[1]}\big)+S\big(\mathsf{B}^\rho\|\mathsf{M}^\rho_{[2]}\big)$.

A second difference arises in the quantification of the inefficiency of the observable approximations $\mathsf{A}\simeq\mathsf{M}_{[1]}$ and $\mathsf{B}\simeq\mathsf{M}_{[2]}$. The BLW approach naturally leads to using the two *deviations* $d(\mathsf{A},\mathsf{M}_{[1]})=\sup_\rho d(\mathsf{A}^\rho,\mathsf{M}^\rho_{[1]})$ and $d(\mathsf{B},\mathsf{M}_{[2]})=\sup_\rho d(\mathsf{B}^\rho,\mathsf{M}^\rho_{[2]})$, that is, the dimensional couple $\big(d(\mathsf{A},\mathsf{M}_{[1]}),d(\mathsf{B},\mathsf{M}_{[2]})\big)$. Instead, the entropic approach gives the entropic divergence $D\big(\mathsf{A},\mathsf{B}\|\mathsf{M}\big)$ as a natural, dimensionless and scalar measure of the approximation inefficiency.

Note that, for fixed M, the divergence $D\big(\mathsf{A},\mathsf{B}\|\mathsf{M}\big)$ tells us how badly M^ρ can approximate the probabilities A^ρ and B^ρ when the three observables are measured in one state ρ, but the same is not true for $\big(d(\mathsf{A},\mathsf{M}_{[1]}),d(\mathsf{B},\mathsf{M}_{[2]})\big)$. Indeed, BLW evaluate the worst possible errors separately, so that the two suprema for the Wasserstein distances $d(\mathsf{A}^{\rho_1}, \mathsf{M}_{[1]}^{\rho_1})$ and $d(\mathsf{B}^{\rho_2}, \mathsf{M}_{[2]}^{\rho_2})$ are attained at possibly different states ρ_1 and ρ_2.

Now, when MURs are derived, the difference of the two approaches is reflected in the distinct aims of the respective statements.

For BLW, proving a MUR means showing that the two deviations $d(\mathsf{A}, \mathsf{M}_{[1]})$ and $d(\mathsf{B}, \mathsf{M}_{[2]})$ can not both be too small; that is, all the couples $\big(d(\mathsf{A}, \mathsf{M}_{[1]}), d(\mathsf{B}, \mathsf{M}_{[2]})\big)$ must lie above some curve in the real plane, away from the origin. One can even look for the exact characterisation of all the admissible points

$$\Big\{\big(d(\mathsf{A}, \mathsf{M}_{[1]}), d(\mathsf{B}, \mathsf{M}_{[2]})\big) : \mathsf{M} \in \mathcal{M}(\mathcal{X} \times \mathcal{Y})\Big\};$$

this is the *uncertainty region* (or diagram) of A and B. Then, any constraint on the shape of the uncertainty region yields a relation between the worst errors occurring in two separate uses of an approximate joint measurement M: namely, for approximating $\mathsf{A} \simeq \mathsf{M}_{[1]}$ in a first preparation, and $\mathsf{B} \simeq \mathsf{M}_{[2]}$ in a second one.

On the other hand, in our entropic approach, proving a MUR amounts to giving a strictly positive lower bound for $D\big(\mathsf{A}, \mathsf{B} \| \mathsf{M}\big)$; the sharpest statements are achieved when $c_{\mathrm{inc}}(\mathsf{A}, \mathsf{B})$ or $c_{\mathrm{ed}}(\mathsf{A}, \mathsf{B})$ are explicitly evaluated. This is the state-independent formulation (13); it can be further rephrased as the statement (14) about the inefficiency of an arbitrary approximation $(\mathsf{A}, \mathsf{B}) \simeq (\mathsf{M}_{[1]}, \mathsf{M}_{[2]})$ that occurs in one preparation of the system, the same for both observables.

2.4. Noisy observables and uncertainty upper bounds. Before trying to exactly compute $c_{\mathrm{inc}}(\mathsf{A}, \mathsf{B})$ and $c_{\mathrm{ed}}(\mathsf{A}, \mathsf{B})$ in some concrete examples, let us improve their general upper bound given in Theorem 2, item (iii). For this task, we introduce an important class of bi-observables M that are known to give good approximations of A and B. Even if these M were not optimal, we expect that they should have a small divergence from (A, B) and thus they should give a good upper bound for its minimum.

Two incompatible observables A and B can always be turned into a compatible pair by adding enough classical noise to their measurements. Indeed, for any choice of trivial observables $\mathsf{T}_\mathsf{A} = p_\mathsf{A} \mathbb{1}$, $p_\mathsf{A} \in \mathcal{P}(\mathcal{X})$, and $\mathsf{T}_\mathsf{B} = p_\mathsf{B} \mathbb{1}$, $p_\mathsf{B} \in \mathcal{P}(\mathcal{Y})$, the observables $\lambda \mathsf{A} + (1-\lambda)\mathsf{T}_\mathsf{A}$ and $\gamma \mathsf{B} + (1-\gamma)\mathsf{T}_\mathsf{B}$, which are *noisy versions* of A and B with *noise intensities* $1-\lambda$ and $1-\gamma$, are compatible for all $\lambda, \gamma \in [0,1]$ such that $\lambda + \gamma \leq 1$ (sufficient condition) [48, Prop. 1]. A bi-observable with the given marginals is

$$\mathsf{M}(x,y) = \lambda \mathsf{A}(x) p_\mathsf{B}(y) + \gamma p_\mathsf{A}(x) \mathsf{B}(y) + \Big(1 - \lambda - \gamma\Big) p_\mathsf{A}(x) p_\mathsf{B}(y) \mathbb{1}.$$

Anyway, depending on A, B, p_A and p_B, it may be possible to go outside the region $\lambda + \gamma \leq 1$, and so reduce the noise intensities. In the following, for every $0 \leq \lambda \leq 1$, we will consider the couple of equally noisy observables

$$\begin{aligned} \mathsf{A}_\lambda(x) &= \lambda \mathsf{A}(x) + (1-\lambda) \mathsf{A}^{\rho_0}(x) \mathbb{1}, \\ \mathsf{B}_\lambda(y) &= \lambda \mathsf{B}(y) + (1-\lambda) \mathsf{B}^{\rho_0}(y) \mathbb{1}, \end{aligned} \tag{15}$$

where $\rho_0 = (1/d)\mathbb{1}$ is the maximally chaotic state. Note that, if A is a rank-one sharp observable, then $\mathsf{A}^{\rho_0} = u_\mathcal{X}$; a similar consideration holds for B. If $\lambda \leq 1/2$, the two observables are compatible, but, depending on the specific A and B, they could be compatible also for larger λ. In any case, by (6) and (9) we get the bound

$$c_{\rm inc}(\mathsf{A},\mathsf{B}) \leq D\big(\mathsf{A},\mathsf{B}\|\mathsf{M}\big) \leq \log \frac{1}{\lambda+(1-\lambda)\min_{x\in\mathcal{X}} \mathsf{A}^{\rho_0}(x)} + \log \frac{1}{\lambda+(1-\lambda)\min_{y\in\mathcal{Y}} \mathsf{B}^{\rho_0}(y)} \quad (16)$$

for all $\lambda \in [0,1]$ such that A_λ and B_λ are compatible, and any joint measurement M of A_λ and B_λ. Since the two terms in the right hand side of (16) are decreasing functions of λ, in order to obtain the best bound we are led to find the maximal value $\lambda_{\rm max}$ of λ for which the noisy observables A_λ and B_λ are compatible. This problem was addressed in [49], where a complete solution was given for a couple of Fourier conjugate sharp observables. Moreover, it was shown that in the general case a nontrivial lower bound for $\lambda_{\rm max}$ can always be achieved by means of *optimal approximate cloning* [50].

Following the same idea, we are going to find a nontrivial upper bound for $c_{\rm inc}(\mathsf{A},\mathsf{B})$ by means of the optimal approximate 2-cloning channel

$$\Phi : \mathcal{S}(\mathcal{H}) \to \mathcal{S}(\mathcal{H}\otimes\mathcal{H}), \qquad \Phi(\rho) = \frac{2}{d+1}\, S_2(\rho\otimes\mathbb{1})S_2,$$

where $S_2 : \mathcal{H}\otimes\mathcal{H} \to \mathcal{H}\otimes\mathcal{H}$ is the orthogonal projection of $\mathcal{H}\otimes\mathcal{H}$ onto its symmetric subspace Sym$(\mathcal{H}\otimes\mathcal{H})$, defined by $S_2(\phi_1\otimes\phi_2) = (\phi_1\otimes\phi_2+\phi_2\otimes\phi_1)/2$. Performing a measurement of the tensor product observable $\mathsf{A}\otimes\mathsf{B}$ in the state $\Phi(\rho)$ then amounts to measure the bi-observable $\mathsf{M}_{\rm cl} = \Phi^*(\mathsf{A}\otimes\mathsf{B})$ in ρ; its marginals are (see [51])

$$\mathsf{M}_{{\rm cl}\,[1]} = \mathsf{A}_{\lambda_{\rm cl}} \qquad \text{and} \qquad \mathsf{M}_{{\rm cl}\,[2]} = \mathsf{B}_{\lambda_{\rm cl}} \qquad \text{where} \qquad \lambda_{\rm cl} = \frac{d+2}{2(d+1)}.$$

Of course $\lambda_{\rm cl} \leq \lambda_{\rm max}$, but the important point is that $\lambda_{\rm cl} > 1/2$. Inserting the above $\lambda_{\rm cl}$ in the bound (16) and using $d\mathsf{A}^{\rho_0}(x) = {\rm Tr}\,\{\mathsf{A}(x)\}$, we obtain

$$c_{\rm inc}(\mathsf{A},\mathsf{B}) \leq D\big(\mathsf{A},\mathsf{B}\|\mathsf{M}_{\rm cl}\big) \leq \log\frac{2(d+1)}{d+2+\min_x {\rm Tr}\,\{\mathsf{A}(x)\}} + \log\frac{2(d+1)}{d+2+\min_y {\rm Tr}\,\{\mathsf{B}(y)\}},$$

holding for all observables A and B.

It is worth noticing that the bi-observable $\mathsf{M}_{\rm cl}$ describes a sequential measurement having B as second measured observable. Indeed, define the instrument $\mathcal{J} \in \mathcal{J}(\mathcal{X})$, with

$$\mathcal{J}_x[\rho] = {\rm Tr}_1\,\{(\mathsf{A}(x)\otimes\mathbb{1})\Phi(\rho)\}\,,$$

where ${\rm Tr}_1$ denotes the partial trace with respect to the first factor. It is easy to check that $\mathsf{M}_{\rm cl} = \mathcal{J}^*(\mathsf{B})$, so that $\mathsf{M}_{\rm cl} \in \mathcal{M}_{\rm ed}(\mathcal{X};\mathsf{B})$. Therefore, the upper bound we have found for $D\big(\mathsf{A},\mathsf{B}\|\mathsf{M}_{\rm cl}\big)$ actually provides a bound also for the entropic error/disturbance coefficient $c_{\rm ed}(\mathsf{A},\mathsf{B})$.

Summarizing the above discussion, we thus arrive at the main conclusion of this section.

Theorem 3. *For any couple of observables* A *and* B*, we have*

$$c_{\mathrm{inc}}(\mathsf{A},\mathsf{B}) \leq c_{\mathrm{ed}}(\mathsf{A},\mathsf{B}) \leq \log \frac{2(d+1)}{d+2+\min_{x\in\mathcal{X}} \operatorname{Tr}\{\mathsf{A}(x)\}} + \log \frac{2(d+1)}{d+2+\min_{y\in\mathcal{Y}} \operatorname{Tr}\{\mathsf{B}(y)\}} \leq 2\log\frac{2(d+1)}{d+k} \leq 2, \quad (17)$$

where in the second to last expression, $k=2$ *in general, or even* $k=3$ *if* $|\mathcal{X}|=|\mathcal{Y}|=d$ *and both* A *and* B *are sharp with* $\operatorname{rank}\mathsf{A}(x)=\operatorname{rank}\mathsf{B}(y)=1$ *for all* x,y*.*

The striking result is that the two uncertainty indexes lie between 0 and 2, independently of the target observables A and B, the numbers $|\mathcal{X}|$ and $|\mathcal{Y}|$ of the possible outcomes, and the Hilbert space dimension d. Note that the bound $2\log[2(d+1)]/(d+k)$ tends to 2 from below as $d\to\infty$.

For sharp observables, the bound (17) is much better than the bound given in Theorem 2, item (iii). However, the case of two trivial uniform observables $\mathsf{A}=\mathsf{U}_{\mathcal{X}}$ and $\mathsf{B}=\mathsf{U}_{\mathcal{Y}}$ is an example where the bound of Theorem 2 is better than the bound (17).

As a final consideration, we will later show that there are observables A and B such that their compatible noisy versions (15) do not optimally approximate A and B. Equivalently, for these observables all the elements $\mathsf{M}\in\mathcal{M}_{\mathrm{inc}}(\mathsf{A},\mathsf{B})$ (or $\mathsf{M}\in\mathcal{M}_{\mathrm{ed}}(\mathsf{A},\mathsf{B})$) have marginals $\mathsf{M}_{[1]}\neq\mathsf{A}_\lambda$ and $\mathsf{M}_{[2]}\neq\mathsf{B}_\lambda$ for all $\lambda\in[0,1]$. Indeed, an example is provided by the two nonorthogonal sharp spin-1/2 observables in Section 3.2. The motivation of this feature comes from the fact that we are not making any extra assumption about our approximate joint measurements, as we optimize over the whole sets $\mathcal{M}(\mathcal{X}\times\mathcal{Y})$ or $\mathcal{M}(\mathcal{X};\mathsf{B})$, according to the case at hand. This is the main difference with the approach e.g. of [45, 49], where a degree of compatibility is defined by considering the minimal noise which one needs to add to A and B in order to make them compatible. It should also be remarked that the non-optimality of the noisy versions is true also in other contexts [26].

2.5. Connections with preparation uncertainty. The entropic incompatibility degree and error/disturbance coefficient are the non trivial and tight lower bounds of the entropic MURs stated in Section 2.3. As we recalled in the Introduction, MURs are different from PURs, which have been formulated in the information-theoretic framework by using different types of entropies (Shannon, Rényi,...) [16–22]. Here we consider only the Shannon entropy (4), and, to facilitate the connections with our indexes, we introduce the *entropic preparation uncertainty coefficient*

$$c_{\mathrm{prep}}(\mathsf{A},\mathsf{B}) = \inf_{\rho\in\mathcal{S}(\mathcal{H})}\left[H(\mathsf{A}^\rho)+H(\mathsf{B}^\rho)\right]. \quad (18)$$

According to the previous sections, the target observables A and B are general POVMs. With this definition, the lower bound proved in [18, Cor. 2.6] can be written as

$$c_{\mathrm{prep}}(\mathsf{A},\mathsf{B}) \geq -\log \max_{x\in\mathcal{X},\, y\in\mathcal{Y}} \left\|\mathsf{A}(x)^{1/2}\mathsf{B}(y)^{1/2}\right\|^2. \quad (19)$$

When the observables are sharp, this lower bound reduces to the one conjectured in [16] and proved in [17].

Note that the infimum in (18) actually is a minimum, because the two entropies are continuous in ρ. Moreover, the equality $c_{\rm prep}(\mathsf{A},\mathsf{B})=0$ is attained if and only if there exist two outcomes x and y such that both positive operators $\mathsf{A}(x)$ and $\mathsf{B}(y)$ have at least one common eigenvector with eigenvalue 1.

For sharp observables, we immediately deduce that the absence of measurement uncertainty implies the absence of preparation uncertainty. Indeed, $c_{\rm inc}(\mathsf{A},\mathsf{B})=0$ is the same as A and B being compatible, which in turn is equivalent to the existence of a whole basis of common eigenvectors $\{\psi_i : i=1,\ldots,d\}$ for which both distributions $\langle\psi_i|\mathsf{A}(x)\psi_i\rangle$ and $\langle\psi_i|\mathsf{B}(y)\psi_i\rangle$ reduce to Kronecker deltas [52, Cor. 5.3]. Therefore, we have the implication $c_{\rm inc}(\mathsf{A},\mathsf{B})=0 \implies c_{\rm prep}(\mathsf{A},\mathsf{B})=0$. However, the same relation fails for general POVMs: for any couple of trivial observables A and B such that $\mathsf{A}\neq\delta_x\mathbb{1}$ or $\mathsf{B}\neq\delta_y\mathbb{1}$, we have $c_{\rm inc}(\mathsf{A},\mathsf{B})=0$ and $c_{\rm prep}(\mathsf{A},\mathsf{B})>0$. On the converse direction, the example of two non commuting sharp observables with a common eigenspace shows that in general $c_{\rm prep}(\mathsf{A},\mathsf{B})=0 \;\not\Longrightarrow\; c_{\rm inc}(\mathsf{A},\mathsf{B})=0$. The failure of this implication exhibits a striking difference between preparation and measurement uncertainties: actually, the entropic incompatibility degree vanishes if and only if the two observables are compatible (Theorem 2, item (v)), while in the preparation case nothing similar happens.

Nevertheless, there exists a link between the entropic incompatibility degree $c_{\rm inc}$ and the preparation uncertainty coefficient $c_{\rm prep}$. Indeed, let us consider the trivial uniform observable $\mathsf{U}\in\mathcal{M}(\mathcal{X}\times\mathcal{Y})$, with $\mathsf{U}=(u_\mathcal{X}\otimes u_\mathcal{Y})\mathbb{1}$ and $\mathsf{U}_{[1]}=u_\mathcal{X}\mathbb{1}$, $\mathsf{U}_{[2]}=u_\mathcal{Y}\mathbb{1}$. By Proposition 1, item (iii), we have

$$S\big(\mathsf{A}^\rho\|\mathsf{U}^\rho_{[1]}\big)+S\big(\mathsf{B}^\rho\|\mathsf{U}^\rho_{[2]}\big)=\log|\mathcal{X}|+\log|\mathcal{Y}|-H(\mathsf{A}^\rho)-H(\mathsf{B}^\rho).$$

By taking the supremum over all states, Definitions 2 and 3 give

$$c_{\rm inc}(\mathsf{A},\mathsf{B})\leq D\big(\mathsf{A},\mathsf{B}\|\mathsf{U}\big)=\log|\mathcal{X}|+\log|\mathcal{Y}|-c_{\rm prep}(\mathsf{A},\mathsf{B}).$$

The final result is the following tradeoff bound:

$$c_{\rm inc}(\mathsf{A},\mathsf{B})+c_{\rm prep}(\mathsf{A},\mathsf{B})\leq\log|\mathcal{X}|+\log|\mathcal{Y}|\,. \tag{20}$$

Note that this bound is saturated at least in the trivial case $\mathsf{A}=u_\mathcal{X}\mathbb{1}$, $\mathsf{B}=u_\mathcal{Y}\mathbb{1}$, for which we have $c_{\rm prep}(\mathsf{A},\mathsf{B})=\log|\mathcal{X}|+\log|\mathcal{Y}|$ and $c_{\rm inc}(\mathsf{A},\mathsf{B})=0$. We also remark that (20) is not the trivial sum of the two upper bounds $c_{\rm inc}(\mathsf{A},\mathsf{B})\leq 2$ (Theorem 3) and $c_{\rm prep}(\mathsf{A},\mathsf{B})\leq\log|\mathcal{X}|+\log|\mathcal{Y}|$ (following from the definition (18) of $c_{\rm prep}$ and the bound for the Shannon entropy of Proposition 1, item (i)).

3. Symmetries and uncertainty lower bounds

In quantum mechanics, many fundamental observables are directly related to symmetry properties of the quantum system at hand. That is, in many concrete situations there is some symmetry group G acting on both the measurement outcome space and the set of system states, in such a way that the two group actions naturally intertwine. The observables that preserve the symmetry structure are usually called *G-covariant*.

In the present setting, covariance will help us to find the incompatibility degree $c_{\rm inc}(\mathsf{A},\mathsf{B})$ and characterize the optimal set $\mathcal{M}_{\rm inc}(\mathsf{A},\mathsf{B})$ for a couple of sharp observables A and B sharing suitable symmetry properties. In Section 3.1 below we provide a general result in this sense, which we then apply to the cases of two spin-1/2 components (Section 3.2) and two observables that are conjugated by the Fourier transform of a finite field (Section 3.3).

3.1. Symmetries and optimal approximate joint measurments. We now suppose that the joint outcome space $\mathcal{X} \times \mathcal{Y}$ carries the action of a finite group G, acting on the left, so that each $g \in G$ is associated with a bijective map on the finite set $\mathcal{X} \times \mathcal{Y}$. Moreover, we also assume that there is a projective unitary representation U of G on $\mathcal{H}$. The following natural left actions are then defined for all $g \in G$:

- on $\mathcal{S}(\mathcal{H})$: $g\rho = U(g)\rho U(g)^*$;
- on $\mathcal{P}(\mathcal{X} \times \mathcal{Y})$: $gp(x, y) = p(g^{-1}(x, y))$ for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$;
- on $\mathcal{M}(\mathcal{X} \times \mathcal{Y})$: $g\mathsf{M}(x, y) = U(g)\mathsf{M}(g^{-1}(x, y))U(g)^*$ for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$.

While the two actions on $\mathcal{S}(\mathcal{H})$ and $\mathcal{P}(\mathcal{X} \times \mathcal{Y})$ have a clear physical interpretation, the action on $\mathcal{M}(\mathcal{X} \times \mathcal{Y})$ is understood by means of the fundamental relation

$$g(\mathsf{M}^\rho) = (g\mathsf{M})^{g\rho}, \tag{21}$$

which asserts that $g\mathsf{M}$ is defined in such a way that measuring it on the transformed state $g\rho$ just gives the translated probability $g(\mathsf{M}^\rho)$. Note that the parenthesis order actually matters in (21).

A fixed point M for the action of G on $\mathcal{M}(\mathcal{X} \times \mathcal{Y})$ is a *G-covariant observable*, i.e. $U(g)\mathsf{M}(x, y)U(g)^* = \mathsf{M}(g(x, y))$ for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$ and $g \in G$. On the other hand, if $\mathsf{M} \in \mathcal{M}(\mathcal{X} \times \mathcal{Y})$ is any observable, then

$$\mathsf{M}_G = \frac{1}{|G|} \sum_{g \in G} g\mathsf{M} \tag{22}$$

is a G-covariant element in $\mathcal{M}(\mathcal{X} \times \mathcal{Y})$, which we call the *covariant version* of M.

Now we state some sufficient conditions on the observables A, B and the action of the group G ensuring that the entropic divergence $D(\mathsf{A}, \mathsf{B} \| \cdot)$ is G-invariant, and then we derive their consequences on the optimal approximate joint measurements of A and B.

Note that the relative entropy is always invariant for a group action, that is,

$$S(gp \| gq) = S(p \| q), \qquad \forall p, q \in \mathcal{P}(\mathcal{X} \times \mathcal{Y}), \quad g \in G, \tag{23}$$

by Proposition 1, (iv). Note also that, for $p \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$, the expression $gp_{[i]} = (gp)_{[i]}$ is unambiguous, as the action of g is defined on $\mathcal{P}(\mathcal{X} \times \mathcal{Y})$ and not on $\mathcal{P}(\mathcal{X})$ or $\mathcal{P}(\mathcal{Y})$.

Theorem 4. *Let $\mathsf{A} \in \mathcal{M}(\mathcal{X})$, $\mathsf{B} \in \mathcal{M}(\mathcal{Y})$ be the target observables. Let G be a finite group, acting on $\mathcal{X} \times \mathcal{Y}$ and with a projective unitary representation U on $\mathcal{H}$. Suppose the group G is generated by a subset $S_G \subseteq G$, such that each $g \in S_G$ satisfies either one condition between:*

(i) there exist maps $f_{g,\mathcal{X}} : \mathcal{X} \to \mathcal{X}$ and $f_{g,\mathcal{Y}} : \mathcal{Y} \to \mathcal{Y}$ such that, for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$,
(a) $g(x, y) = (f_{g,\mathcal{X}}(x), f_{g,\mathcal{Y}}(y))$
(b) $U(g)\mathsf{A}(x)U(g)^* = \mathsf{A}(f_{g,\mathcal{X}}(x))$ and $U(g)\mathsf{B}(y)U(g)^* = \mathsf{B}(f_{g,\mathcal{Y}}(y))$;

(ii) there exist maps $f_{g,\mathcal{X}} : \mathcal{X} \to \mathcal{Y}$ and $f_{g,\mathcal{Y}} : \mathcal{Y} \to \mathcal{X}$ such that, for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$,
(a) $g(x, y) = (f_{g,\mathcal{Y}}(y), f_{g,\mathcal{X}}(x))$
(b) $U(g)\mathsf{A}(x)U(g)^* = \mathsf{B}(f_{g,\mathcal{X}}(x))$ and $U(g)\mathsf{B}(y)U(g)^* = \mathsf{A}(f_{g,\mathcal{Y}}(y))$.

Then, $D(\mathsf{A}, \mathsf{B} \| g\mathsf{M}) = D(\mathsf{A}, \mathsf{B} \| \mathsf{M})$ for all $\mathsf{M} \in \mathcal{M}(\mathcal{X} \times \mathcal{Y})$ and $g \in G$.

Proof. If two elements $g_1, g_2 \in G$ satisfy the above hypotheses, so does their product $g_1 g_2$. Since S_G generates G, we can then assume that $S_G = G$. In this case, condition (i.a) or (ii.a) easily implies the relation

$$gp_{[1]} \otimes gp_{[2]} = g(p_{[1]} \otimes p_{[2]}), \qquad \forall p \in \mathcal{P}(\mathcal{X} \times \mathcal{Y}),\ g \in G. \tag{24}$$

On the other hand, by condition (i.b) or (ii.b), we get

$$\mathsf{A}^{g\rho} \otimes \mathsf{B}^{g\rho} = g(\mathsf{A}^{\rho} \otimes \mathsf{B}^{\rho}), \qquad \forall \rho \in \mathcal{S}(\mathcal{H}),\ g \in G. \tag{25}$$

For any $\mathsf{M} \in \mathcal{M}(\mathcal{X} \times \mathcal{Y})$, we then have

$$\begin{aligned}
S[\mathsf{A}, \mathsf{B} \| g^{-1}\mathsf{M}](\rho) &= S\big(\mathsf{A}^{\rho} \otimes \mathsf{B}^{\rho} \| (g^{-1}\mathsf{M})^{\rho}_{[1]} \otimes (g^{-1}\mathsf{M})^{\rho}_{[2]}\big) && \text{by (8)} \\
&= S\big(\mathsf{A}^{\rho} \otimes \mathsf{B}^{\rho} \| g^{-1}(\mathsf{M}^{g\rho})_{[1]} \otimes g^{-1}(\mathsf{M}^{g\rho})_{[2]}\big) && \text{by (21)} \\
&= S\big(\mathsf{A}^{\rho} \otimes \mathsf{B}^{\rho} \| g^{-1}(\mathsf{M}^{g\rho}_{[1]} \otimes \mathsf{M}^{g\rho}_{[2]})\big) && \text{by (24)} \\
&= S\big(g(\mathsf{A}^{\rho} \otimes \mathsf{B}^{\rho}) \| \mathsf{M}^{g\rho}_{[1]} \otimes \mathsf{M}^{g\rho}_{[2]}\big) && \text{by (23)} \\
&= S\big(\mathsf{A}^{g\rho} \otimes \mathsf{B}^{g\rho} \| \mathsf{M}^{g\rho}_{[1]} \otimes \mathsf{M}^{g\rho}_{[2]}\big) && \text{by (25)} \\
&= S[\mathsf{A}, \mathsf{B} \| \mathsf{M}](g\rho).
\end{aligned}$$

Taking the supremum over ρ and observing that $\mathcal{S}(\mathcal{H}) = g\mathcal{S}(\mathcal{H})$, it follows that $D\big(\mathsf{A}, \mathsf{B} \| g^{-1}\mathsf{M}\big) = D\big(\mathsf{A}, \mathsf{B} \| \mathsf{M}\big)$. □

Remark 2.
1. The occurrence of either hypothesis (i) or (ii) may depend on the generator $g \in S_G$; however, in both cases g does not mix the $\mathcal{X}$ and $\mathcal{Y}$ outcomes together.
2. Conditions (i.a), (ii.a) are hypotheses about the action of G on the outcome space $\mathcal{X} \times \mathcal{Y}$. Note that each one implies that the maps $f_{g,\mathcal{X}}$ and $f_{g,\mathcal{Y}}$ are bijective. In particular, one can have some generator g satisfying (ii.a) only if $|\mathcal{X}| = |\mathcal{Y}|$.
3. Conditions (i.b), (ii.b) involve also the observables A and B. Even if A and B are not compatible, they are required to behave as if they were the marginals of a covariant bi-observable.
4. The symmetries allowed in hypothesis (ii) of Theorem 4 essentially are of permutational nature. They directly follow from the exchange symmetry of the error function (7), in which the approximation errors $S\big(\mathsf{A}^{\rho} \| \mathsf{M}^{\rho}_{[1]}\big)$ and $S\big(\mathsf{B}^{\rho} \| \mathsf{M}^{\rho}_{[2]}\big)$ are equally weighted.

Corollary 1. *Under the hypotheses of Theorem 4,*

- *the set* $\mathcal{M}_{\mathrm{inc}}(\mathsf{A}, \mathsf{B})$ *is G-invariant;*
- *for any* $\mathsf{M} \in \mathcal{M}_{\mathrm{inc}}(\mathsf{A}, \mathsf{B})$*, we have* $\mathsf{M}_G \in \mathcal{M}_{\mathrm{inc}}(\mathsf{A}, \mathsf{B})$*;*
- *there exists a G-covariant observable in* $\mathcal{M}_{\mathrm{inc}}(\mathsf{A}, \mathsf{B})$.

Proof. Since $D\big(\mathsf{A}, \mathsf{B} \| \cdot\big)$ is G-invariant by Theorem 4, then the set $\mathcal{M}_{\mathrm{inc}}(\mathsf{A}, \mathsf{B})$ is G-invariant. This fact and the convexity of $\mathcal{M}_{\mathrm{inc}}(\mathsf{A}, \mathsf{B})$ implies that $\mathsf{M}_G \in \mathcal{M}_{\mathrm{inc}}(\mathsf{A}, \mathsf{B})$ for all $\mathsf{M} \in \mathcal{M}_{\mathrm{inc}}(\mathsf{A}, \mathsf{B})$. Since the latter set is nonempty by Theorem 2, item (iv), it then always contains a G-covariant observable. □

Remark 3. Since the covariance requirement reduces the many degrees of freedom in the choice of a bi-observable $\mathsf{M} \in \mathcal{M}(\mathcal{X} \times \mathcal{Y})$, we expect that the larger is the symmetry group G, the fewer amount of free parameters will be needed to describe a G-covariant element M. This will be a big help in the computation of $c_{\mathrm{inc}}(\mathsf{A}, \mathsf{B})$, as Corollary 1 allows to minimize $D\big(\mathsf{A},\mathsf{B}\|\cdot\big)$ just on the set of G-covariant bi-observables. More precisely, under the hypotheses of Theorem 4,

$$c_{\rm inc}(\mathsf{A},\mathsf{B}) = \min_{\substack{\mathsf{M}\in\mathcal{M}(\mathcal{X}\times\mathcal{Y})\\ \mathsf{M}\ G\text{-covariant}}} \max_{\substack{\rho\in\mathcal{S}(\mathcal{H})\\ \rho \text{ pure}}} \left\{ S\big(\mathsf{A}^\rho\|\mathsf{M}^\rho_{[1]}\big) + S\big(\mathsf{B}^\rho\|\mathsf{M}^\rho_{[2]}\big)\right\},$$

where the minimum has to be computed only with respect to the parameters describing a G-covariant bi-observable M. In particular, it is only the dependence of the marginals $\mathsf{M}_{[1]}$ and $\mathsf{M}_{[2]}$ on such parameters that comes into play. Of course, solving this double optimization problem yields the value of $c_{\rm inc}(\mathsf{A},\mathsf{B})$ and all the covariant optimal joint measurement of A and B, but not the whole optimal set $\mathcal{M}_{\rm inc}(\mathsf{A},\mathsf{B})$.

In the cases of two othogonal spin-1/2 components (Section 3.2.1) and two Fourier conjugate observables (Section 3.3), covariance will reduce the number of parameters to just a single one.

If B is not sharp, the two sets $\mathcal{M}_{\rm inc}(\mathsf{A},\mathsf{B})$ and $\mathcal{M}_{\rm ed}(\mathsf{A},\mathsf{B})$ may be different, and we need a specific corollary for $\mathcal{M}_{\rm ed}(\mathsf{A},\mathsf{B})$. Indeed, stronger hypotheses are required to ensure that the sequential measurement set $\mathcal{M}(\mathcal{X};\mathsf{B})$ is G-invariant.

Corollary 2. *Under the hypotheses of Theorem 4, and supposing in addition that all the generators $g \in S_G$ enjoy only condition (i) of that theorem,*

- *the set* $\mathcal{M}(\mathcal{X};\mathsf{B})$ *is G-invariant;*
- *the set* $\mathcal{M}_{\rm ed}(\mathsf{A},\mathsf{B})$ *is G-invariant;*
- *for any* $\mathsf{M}\in\mathcal{M}_{\rm ed}(\mathsf{A},\mathsf{B})$*, we have* $\mathsf{M}_G\in\mathcal{M}_{\rm ed}(\mathsf{A},\mathsf{B})$*;*
- *there exists a G-covariant observable in* $\mathcal{M}_{\rm ed}(\mathsf{A},\mathsf{B})$.

Proof. We know that $D\big(\mathsf{A},\mathsf{B}\|\cdot\big)$ is G-invariant by Theorem 4, and so we only have to prove that $\mathcal{M}(\mathcal{X};\mathsf{B})$ is G-invariant; then the subsequent claims follow as in Corollary 1. Since we can assume $S_G = G$, any element $g\in G$ maps a sequential measurement $\mathsf{M} = \mathcal{J}^*(\mathsf{B})$ to another sequential measurement $\mathcal{J}'^*(\mathsf{B})$, due to condition (i) of Theorem 4:

$$\begin{aligned} g\mathsf{M}(x,y) &= U(g)\mathsf{M}\Big(g^{-1}(x,y)\Big)U(g)^* = U(g)\mathsf{M}\Big(f^{-1}_{g,\mathcal{X}}(x), f^{-1}_{g,\mathcal{Y}}(y)\Big)U(g)^* \\ &= U(g)\mathcal{J}^*_{f^{-1}_{g,\mathcal{X}}(x)}\Big[\mathsf{B}(f^{-1}_{g,\mathcal{Y}}(y))\Big]U(g)^* \\ &= U(g)\mathcal{J}^*_{f^{-1}_{g,\mathcal{X}}(x)}\Big[U(g)^*\mathsf{B}(y)U(g)\Big]U(g)^* =: \mathcal{J}'^*(\mathsf{B})(x,y). \end{aligned}$$

□

Remark 4. Corollary 2 does not admit elements g satisfying condition (ii) of Theorem 4 because this hypothesis alone can not guarantee the G-invariance of the set $\mathcal{M}(\mathcal{X};\mathsf{B})$. Of course, it works for a sharp B, but it could fail, for example, for a trivial B. Indeed, take $\mathcal{X} = \mathcal{Y}$ and $\mathsf{A} = \mathsf{B} = \mathsf{U}_\mathcal{X}$; then $\mathcal{M}(\mathcal{X};\mathsf{B}) = \{\mathsf{M}\in\mathcal{M}(\mathcal{X}\times\mathcal{Y}) : \mathsf{M}(x,y) = \mathsf{M}_1(x)u_\mathcal{X}(y),\ \forall x,y, \text{ for some } \mathsf{M}_1\in\mathcal{M}(\mathcal{X})\}$, and $\mathsf{M}_{[2]}(y) = \mathsf{B}(y)$ has rank d for every $\mathsf{M}\in\mathcal{M}(\mathcal{X};\mathsf{B})$ and $y\in\mathcal{Y}$. Nevertheless, if g satisfies (ii.a), then (ii.b) is obvious, but g could send a sequential measurement M outside $\mathcal{M}(\mathcal{X};\mathsf{B})$. Indeed, $(g\mathsf{M})_{[2]}(y) = U(g)\mathsf{M}_1\Big(f^{-1}_{g,\mathcal{X}}(y)\Big)U(g)^*$ has rank equal to the rank of $\mathsf{M}_1\Big(f^{-1}_{g,\mathcal{X}}(y)\Big)$, which can be chosen smaller than d.

3.2. Two spin-1/2 components. As a first application of Theorem 4 and its corollaries, we take as target observables two spin-1/2 components along the directions defined by two unit vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ in $\mathbb{R}^3$. They are represented by the sharp observables

$$\mathsf{A}(x) = \frac{1}{2}\left(\mathbb{1} + x\,\boldsymbol{a}\cdot\boldsymbol{\sigma}\right), \quad \mathsf{B}(y) = \frac{1}{2}\left(\mathbb{1} + y\,\boldsymbol{b}\cdot\boldsymbol{\sigma}\right), \tag{26}$$

where $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$ is the vector of the three Pauli matrices on $\mathcal{H} = \mathbb{C}^2$, and $\mathcal{X} = \mathcal{Y} = \{-1, +1\}$. Let $\alpha \in [0, \pi]$ be the angle formed by $\boldsymbol{a}$ and $\boldsymbol{b}$; by item (i) of Theorem 2, the coefficient $c_{\mathrm{inc}}(\mathsf{A}, \mathsf{B})$ does not depend on the choice of the values of the outcomes, and this allows us to take $\alpha \in [0, \pi/2]$. Indeed, when $\alpha > \pi/2$, it is enough to change $y \to -y$ and $\boldsymbol{b} \to -\boldsymbol{b}$ to recover the previous case. Without loss of generality, we take the two spin directions in the $\boldsymbol{ij}$-plane and choose the $\boldsymbol{i}$- and $\boldsymbol{j}$-axes in such a way that the bisector of the angle formed by $\boldsymbol{a}$ and $\boldsymbol{b}$ coincides with the bisector $\boldsymbol{n}$ of the first quadrant; $\boldsymbol{m}$ is the bisector of the second quadrant. This choice is illustrated in Figure 1, where $\alpha \in [0, \pi/2]$, $a_1^2 + a_2^2 = 1$, and

$$\begin{aligned} &\boldsymbol{a} = a_1\boldsymbol{i} + a_2\boldsymbol{j}, \qquad \boldsymbol{b} = a_2\boldsymbol{i} + a_1\boldsymbol{j}, \qquad \boldsymbol{n} = \frac{\boldsymbol{i}+\boldsymbol{j}}{\sqrt{2}}, \qquad \boldsymbol{m} = \frac{\boldsymbol{j}-\boldsymbol{i}}{\sqrt{2}}, \\ &a_1 = \sqrt{\frac{1+\sin\alpha}{2}} \in \left[\frac{1}{\sqrt{2}}, 1\right], \qquad a_2 = \frac{\cos\alpha}{\sqrt{2(1+\sin\alpha)}} \in \left[0, \frac{1}{\sqrt{2}}\right]. \end{aligned} \tag{27}$$

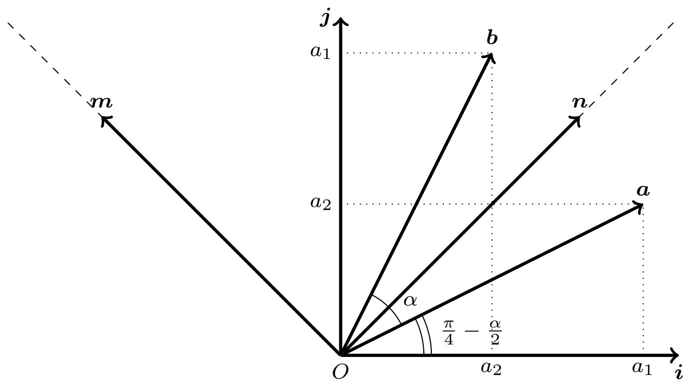

Figure 1. The unit vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ characterizing the target spin-1/2 observables (26).

In the next part, we will see that the compatible observables optimally approximating the two target spins (26) are noisy versions of another two spin-1/2 components; however, in general their directions may be different from the original $\boldsymbol{a}$ and $\boldsymbol{b}$. For this reason, we need to introduce the family of observables $\mathsf{A}_{\boldsymbol{c}}, \mathsf{B}_{\boldsymbol{c}} \in \mathcal{M}(\{+1, -1\})$, with

$$\mathsf{A}_{\boldsymbol{c}}(x) = \frac{1}{2}\left[\mathbb{1} + x\,(c_1\sigma_1 + c_2\sigma_2)\right], \quad \mathsf{B}_{\boldsymbol{c}}(y) = \frac{1}{2}\left[\mathbb{1} + y\,(c_2\sigma_1 + c_1\sigma_2)\right], \tag{28}$$

where $\boldsymbol{c} = c_1\boldsymbol{i} + c_2\boldsymbol{j}$, $c_i \in \mathbb{R}$. Note that the components of $\boldsymbol{c}$ appear in $\mathsf{A}_{\boldsymbol{c}}$ and $\mathsf{B}_{\boldsymbol{c}}$ in the reverse order; moreover, $\mathsf{A} = \mathsf{A}_{\boldsymbol{a}}$ and $\mathsf{B} = \mathsf{B}_{\boldsymbol{a}}$. Formula (28) defines two observables if and only if $|\boldsymbol{c}| \leq 1$, that is, $\boldsymbol{c}$ belongs to the disk

$$C = \left\{c_1\boldsymbol{i} + c_2\boldsymbol{j} : c_1^2 + c_2^2 \leq 1\right\}. \tag{29}$$

Note that, for $|\boldsymbol{c}| = 1$, the observable $\mathsf{A}_{\boldsymbol{c}}$ is sharp, and it is the spin-1/2 component along the direction $\boldsymbol{c}$; on the other hand, for $|\boldsymbol{c}| \in (0,1)$, $\mathsf{A}_{\boldsymbol{c}}$ is a noisy version of $\mathsf{A}_{\boldsymbol{c}/|\boldsymbol{c}|}$ with noise intensity $1-\lambda = 1-|\boldsymbol{c}|$ (cf. (15)). Analogue considerations hold for $\mathsf{B}_{\boldsymbol{c}}$.

3.2.1. Entropic incompatibility degree and optimal measurements. When the angle between the spin directions $\boldsymbol{a}$ and $\boldsymbol{b}$ is $\alpha = \pi/2$, the target observables become the two orthogonal spin-1/2 components along the $\boldsymbol{i}$- and $\boldsymbol{j}$-axes:

$$\mathsf{A}(x) = \mathsf{X}(x) = \frac{1}{2}\left(\mathbb{1} + x\sigma_1\right), \qquad \mathsf{B}(y) = \mathsf{Y}(y) = \frac{1}{2}\left(\mathbb{1} + y\sigma_2\right). \tag{30}$$

In Appendix B.2, we use Theorem 4 and the many rotational symmetries of these observables to drastically simplify the problem of finding both the value of $c_{\mathrm{inc}}(\mathsf{X},\mathsf{Y})$ and the explicit expression of a bi-observable in $\mathcal{M}_{\mathrm{inc}}(\mathsf{X},\mathsf{Y})$. Remarkably, it also turns out that $\mathcal{M}_{\mathrm{inc}}(\mathsf{X},\mathsf{Y})$ is a singleton set. Indeed, the following theorem is proved.

Theorem 5. *Let* X *and* Y *be the two orthogonal spin-1/2 components* (30). *Then, there is a unique optimal approximate joint measurement of* X *and* Y, *that is the bi-observable*

$$\mathsf{M}_0(x,y) = \frac{1}{4}\left(\mathbb{1} + \frac{x}{\sqrt{2}}\,\sigma_1 + \frac{y}{\sqrt{2}}\,\sigma_2\right), \tag{31}$$

i.e. $\mathcal{M}_{\mathrm{inc}}(\mathsf{X},\mathsf{Y}) = \{\mathsf{M}_0\}$. *If* ρ_e *is the projection on any eigenvector of* σ_1 *or* σ_2, *then*

$$c_{\mathrm{inc}}(\mathsf{X},\mathsf{Y}) = S[\mathsf{X},\mathsf{Y}\|\mathsf{M}_0](\rho_e) = \log\left[2\left(2-\sqrt{2}\right)\right] \simeq 0.228447\,. \tag{32}$$

Note that $\mathsf{M}_0(x,y)$ is a rank-one operator for all $(x,y) \in \mathcal{X}\times\mathcal{Y}$, and its marginals

$$\mathsf{M}_{0\,[1]}(x) = \frac{1}{2}\left(\mathbb{1} + \frac{x}{\sqrt{2}}\,\sigma_1\right), \qquad \mathsf{M}_{0\,[2]}(y) = \frac{1}{2}\left(\mathbb{1} + \frac{y}{\sqrt{2}}\,\sigma_2\right)$$

turn out to be the noisy versions $\mathsf{X}_{1/\sqrt{2}}$, $\mathsf{Y}_{1/\sqrt{2}}$ of the target observables X, Y (cf. (15)).

When the two spin directions $\boldsymbol{a}$ and $\boldsymbol{b}$ are not orthogonal, the system loses the 180° rotational symmetries around the $\boldsymbol{i}$- and $\boldsymbol{j}$-axes. According to Remark 3, this makes the evaluation of $c_{\mathrm{inc}}(\mathsf{A},\mathsf{B})$ a more difficult task. The best we can do is to express $c_{\mathrm{inc}}(\mathsf{A},\mathsf{B})$ as the solution of a maximization/minimization (minimax) problem for an explicit function of two parameters. The analysis of the symmetries of two nonorthogonal spin-1/2 components, and the consequent proof of the next theorem are given in Appendix B.1.

Theorem 6. *Let* A *and* B *be the spin-1/2 components* (26) *with the angle* $\alpha \in [0,\pi/2]$. *For all* $\phi \in [0,2\pi)$, $\gamma \in [-1,1]$ *and* $x,y \in \{-1,+1\}$, *define*

$$\rho(\phi) = \frac{1}{2}\left(\mathbb{1} + \cos\phi\,\sigma_1 + \sin\phi\,\sigma_2\right), \qquad \boldsymbol{c}(\gamma) = \frac{\boldsymbol{i}+\gamma\boldsymbol{j}}{\sqrt{2}}, \tag{33}$$

$$\mathsf{M}_\gamma(x,y) = \frac{1}{4}\left[(1+\gamma xy)\,\mathbb{1} + \frac{1}{\sqrt{2}}\left(x\sigma_1 + y\sigma_2\right) + \frac{\gamma}{\sqrt{2}}\left(y\sigma_1 + x\sigma_2\right)\right]. \tag{34}$$

Then, $\mathsf{M}_\gamma \in \mathcal{M}(\mathcal{X}\times\mathcal{Y})$, *and we have*

$$c_{\mathrm{inc}}(\mathsf{A},\mathsf{B}) = \min_{\gamma\in[-1,1]}\ \max_{\phi\in[0,2\pi)} S[\mathsf{A},\mathsf{B}\|\mathsf{M}_\gamma](\rho(\phi)), \tag{35}$$

$$S[\mathsf{A},\mathsf{B}\|\mathsf{M}_\gamma](\rho) = S\big(\mathsf{A}^\rho\|\mathsf{A}^\rho_{\boldsymbol{c}(\gamma)}\big) + S\big(\mathsf{B}^\rho\|\mathsf{B}^\rho_{\boldsymbol{c}(\gamma)}\big). \tag{36}$$

Moreover, γ *solves the minimization problem* (35) *if and only if* $\mathsf{M}_\gamma \in \mathcal{M}_{\mathrm{inc}}(\mathsf{A},\mathsf{B})$.

In Section 3.2.2, we provide a numerical evaluation of the entropic incompatibility degree (35) for some angles $\alpha \in [0, \pi/2]$. Moreover, using the family of approximate joint measurements in (34), we analytically find a lower bound for $c_{\rm inc}(\mathsf{A}, \mathsf{B})$. Note that, for $\alpha \in (0, \pi/2)$, it is not clear if there is a unique γ solving (35), and if the set $\mathcal{M}_{\rm inc}(\mathsf{A}, \mathsf{B})$ is only made up of the corresponding bi-observables M_γ (see also Remark 7 in Appendix B.1).

The noisy spin-1/2 components $\mathsf{A}_{\boldsymbol{c}(\gamma)}$ and $\mathsf{B}_{\boldsymbol{c}(\gamma)}$ appearing in (36) are the two marginals of the bi-observable M_γ in (34). When M_γ is optimal, we stress that for $\alpha \neq \pi/2$ they may not be noisy versions of the target observables A and B. Indeed, in Table 1 below and the subsequent discussion, we numerically show this for the case $\alpha = \pi/4$.

It is worth noticing that every bi-observable (31) or (34) can be rewritten as a *mixture* (convex combination) of two sharp joint measurements of compatible spin components, along the bisector $\boldsymbol{n}$ in the case of the first bi-observable, and along the bisector $\boldsymbol{m}$ for the other one. More precisely, we introduce the sharp bi-observables

$$\begin{aligned} \mathsf{M}_+(x,y) &= \left[\frac{1}{2}(\mathbb{1} + x\boldsymbol{n}\cdot\boldsymbol{\sigma})\right]\left[\frac{1}{2}(\mathbb{1} + y\boldsymbol{n}\cdot\boldsymbol{\sigma})\right] \equiv \mathsf{A}_{\boldsymbol{n}}(x)\mathsf{B}_{\boldsymbol{n}}(y), \\ \mathsf{M}_-(x,y) &= \left[\frac{1}{2}(\mathbb{1} - x\boldsymbol{m}\cdot\boldsymbol{\sigma})\right]\left[\frac{1}{2}(\mathbb{1} + y\boldsymbol{m}\cdot\boldsymbol{\sigma})\right] \equiv \mathsf{A}_{-\boldsymbol{m}}(x)\mathsf{B}_{-\boldsymbol{m}}(y). \end{aligned} \tag{37}$$

Then, we have

$$\mathsf{M}_0 = \frac{1}{2}(\mathsf{M}_+ + \mathsf{M}_-), \qquad \mathsf{M}_\gamma = \frac{1+\gamma}{2}\,\mathsf{M}_+ + \frac{1-\gamma}{2}\,\mathsf{M}_-. \tag{38}$$

In terms of M_0, the bi-observable M_γ can be expressed also as the mixture

$$\mathsf{M}_\gamma = \begin{cases} \gamma\mathsf{M}_+ + (1-\gamma)\mathsf{M}_0 & \text{if } \gamma \geq 0, \\ |\gamma|\,\mathsf{M}_- + (1-|\gamma|)\mathsf{M}_0 & \text{if } \gamma \leq 0. \end{cases}$$

3.2.2. Numerical and analytical results for nonorthogonal components. In the case of two arbitrarily oriented spin components, the minimax problem (35), giving $c_{\rm inc}$ and γ for the optimal bi-observable M_γ, is hard to be solved analytically. Nevertheless, the double optimization over the angle ϕ and the parameter γ can be tackled numerically, and the resulting $c_{\rm inc}(\mathsf{A}, \mathsf{B})$ for 100 equally distant values α in the interval $[0, \pi/2]$ are plotted in Figure 2.

A good analytical lower bound for $c_{\rm inc}(\mathsf{A}, \mathsf{B})$ can be found by fixing a trial state $\rho(\phi)$, considering the bi-observables M_γ of (34), and then minimizing the error function $S[\mathsf{A}, \mathsf{B}\|\mathsf{M}_\gamma](\rho(\phi))$ with respect to $\gamma \in [-1, 1]$. Indeed, equation (35) yields the inequality $c_{\rm inc}(\mathsf{A}, \mathsf{B}) \geq \min_{\gamma\in[-1,1]} S[\mathsf{A}, \mathsf{B}\|\mathsf{M}_\gamma](\rho(\phi))$ for all $\phi \in [0, 2\pi)$. A convenient choice for ϕ, suggested by the results in the case of two orthogonal components, is to take $\phi \in \{\pi/4 \pm \alpha/2,\ 5\pi/4 \pm \alpha/2\}$, so that the corresponding state $\rho(\phi)$ is any eigenprojection of $\boldsymbol{a}\cdot\boldsymbol{\sigma}$ or $\boldsymbol{b}\cdot\boldsymbol{\sigma}$; say we take the eigenprojection $\rho_e = \rho(\pi/4 - \alpha/2)$ of $\boldsymbol{a}\cdot\boldsymbol{\sigma}$ with positive eigenvalue. Then, we get

$$c_{\rm inc}(\mathsf{A}, \mathsf{B}) \geq \min_{\gamma\in[-1,1]} S[\mathsf{A}, \mathsf{B}\|\mathsf{M}_\gamma](\rho_e) =: LB(\alpha). \tag{39}$$

In Appendix B.3, the explicit expression of $S[\mathsf{A}, \mathsf{B}\|\mathsf{M}_\gamma](\rho_e)$ is given in (80), its minimum over γ is computed and, for $\alpha \neq \pi/2$, it is found at the point

$$\gamma = \frac{\sqrt{2}\ell - a_2}{a_1}, \tag{40}$$

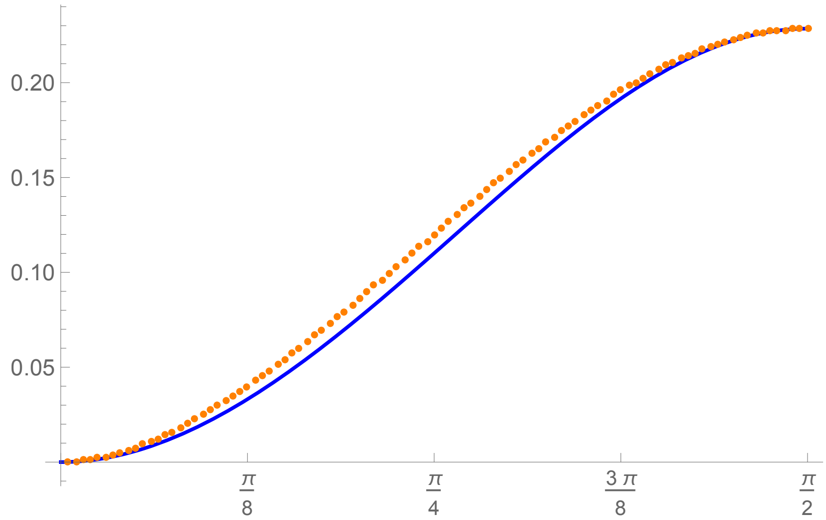

Figure 2. Dots: numerical evaluations of $c_{\mathrm{inc}}(\mathsf{A},\mathsf{B})$ as a function of α. Continuous line: the analytical lower bound $LB(\alpha)$ in (43).

where

$$\ell = \frac{1}{2\sqrt{2}\,a_2}\left(\sqrt{u^2+8(1+u)a_2^2}-u\right), \tag{41}$$

$$u = \left(a_1+\frac{1}{\sqrt{2}}\right)\frac{a_1^2-a_2^2}{\sqrt{2}} = \left(1+\sqrt{1+\sin\alpha}\right)\frac{\sin\alpha}{2}\,. \tag{42}$$

In particular, the value (40) for γ, together with the fact that the bi-observable M_γ has marginals $\mathsf{M}_{\gamma\,[1]} = \mathsf{A}_{\boldsymbol{c}(\gamma)}$ and $\mathsf{M}_{\gamma\,[2]} = \mathsf{B}_{\boldsymbol{c}(\gamma)}$, show that the marginals of the bi-observable giving the lower bound (39) are not noisy versions of the target observables A and B; indeed, $\boldsymbol{c}(\gamma) \not\propto \boldsymbol{a}$ in this case. Finally, the lower bound turns out to be

$$LB(\alpha) = -\log w + \frac{1}{2}\,(1+\cos\alpha)\log\frac{1+\cos\alpha}{1+\ell} + \frac{1}{2}\,(1-\cos\alpha)\log\frac{1-\cos\alpha}{1-\ell}\,, \tag{43}$$

with

$$w = \frac{1}{2} + \frac{\sqrt{u^2+8(1+u)a_2^2}}{4\sqrt{2}\,a_1} + \frac{\sin\alpha}{8}\left(\frac{3}{\sqrt{2}\,a_1}-1\right). \tag{44}$$

The plot of $LB(\alpha)$ is the continuous line in Figure 2.

For $\alpha = 0$, the target observables are compatible and $c_{\mathrm{inc}}(\mathsf{A},\mathsf{B}) = 0$. For $\alpha \to 0$ the previous formulae give $u = 0$, $\ell = 1$, $\boldsymbol{c}(\gamma) = \boldsymbol{a} \equiv \boldsymbol{n}$, and one can check that also the lower bound (43) vanishes, as it must be.

For two orthogonal components, i.e. $\alpha = \pi/2$, the expression (43) gives the exact value (32) of the entropic incompatibility degree, and it is not only a lower bound. This value can be computed by going to the limit $\alpha \to \pi/2$ in (43), or directly by Remark 8 in Appendix B.3.

For $\alpha \in (0, \pi/2)$, Figure 2 shows that the analytical lower bound (43) is not so far from the numerical value.

We now compare our optimal approximate joint measurements with other proposals coming from different approaches. Of course, every approximate joint measurement M that is optimal with respect to some other criterium will have a divergence from the target observables (A, B) larger or equal than $c_{\mathrm{inc}}(\mathsf{A}, \mathsf{B})$. We stress that the other two proposals we will consider yield optimal bi-observables of the form M_γ, in which however the parameter γ is different from ours.

We have seen that, when $\alpha = \pi/2$, the incompatibility degree of A and B, as well as their unique optimal approximate joint measurement M_0, can be evaluated analytically. In this special case, it turns out that M_0 is optimal also with respect to the other criteria we are going to consider in this section. However, as already said, this is not true for general α. In order to show it, we fix the angle $\alpha = \pi/4$, and compare the results of the different criteria in Table 1. We also add a LB column summarizing the parameters for the analytical lower bound (39). The rows provide: (1) the parameter γ characterizing the measurement M_γ; (2) the angle characterizing the pure state $\rho(\phi)$ at which $S[\mathsf{A}, \mathsf{B} \| \mathsf{M}_\gamma]$ is computed, that is the trial angle $\pi/4 - \alpha/2$ in the first column, and the angle maximizing $S[\mathsf{A}, \mathsf{B} \| \mathsf{M}_\gamma](\rho(\phi))$ in the other ones; (3) the value of $S[\mathsf{A}, \mathsf{B} \| \mathsf{M}_\gamma](\rho(\phi))$ for the parameters chosen in (1) and (2), which gives $LB(\pi/4)$ in the first column and the entropic divergence $D\big(\mathsf{A}, \mathsf{B} \| \mathsf{M}_\gamma\big)$ in the other ones. The

Table 1. Incompatibility degree and its bounds for $\alpha = \pi/4$.

criterium	LB	c_{inc}	BLW	NV
measurement: $\gamma \simeq$	0.795559	0.743999	0.541195	0.414213
state: $\phi \simeq$	0.392699	0.282743	0.391128	0.416889
value: $S[\mathsf{A}, \mathsf{B} \| \mathsf{M}_\gamma](\rho(\phi)) \simeq$	0.110081	0.120035	0.160886	0.212079

description of the columns is as follows.

LB: The choice of the parameters is the one described in the computation of the analytical lower bound for c_{inc}. The parameter γ comes from (40), the angle $\phi = \pi/8$ corresponds to the trial state $\rho_e = \rho(\pi/8)$ (i.e. the eigenprojection of $\boldsymbol{a}\cdot\boldsymbol{\sigma}$ for $\alpha = \pi/4$), and the corresponding value of $S[\mathsf{A}, \mathsf{B} \| \mathsf{M}_\gamma](\rho_e)$ is the lower bound $LB(\pi/4)$.

c_{inc}: The parameters are chosen following the relative entropy approach to MURs. They are the numerical solution of the minimax problem (35). Thus, the value of $S[\mathsf{A}, \mathsf{B} \| \mathsf{M}_\gamma](\rho(\phi))$ is the one found numerically for $c_{\mathrm{inc}}(\mathsf{A}, \mathsf{B})$, i.e. the dot at $\alpha = \pi/4$ in Figure 2; γ is the corresponding minimum point giving the optimal approximate joint measurement M_γ of A and B; the angle ϕ corresponds to the state at which the error function $S[\mathsf{A}, \mathsf{B} \| \mathsf{M}_\gamma]$ attains its maximum.

BLW: As discussed in Section 2.3, in [6–8] a different approach is proposed. In particular, its application to the case of two spin-1/2 components is given in [9] (see also [26], where the same final results are obtained in a slightly different context). There, the authors find a strictly positive lower bound for the sum $d(\mathsf{A}, \mathsf{M}_{[1]})^2 + d(\mathsf{B}, \mathsf{M}_{[2]})^2$, which holds for all approximate joint measurements M. Moreover, they find a couple of compatible observables $(\mathsf{A}_{\boldsymbol{c}}, \mathsf{B}_{\boldsymbol{c}})$ saturating the lower bound, and thus optimally approximating the target observables (A, B); this couple is given by a vector $\boldsymbol{c}$ yielding compatible $\mathsf{A}_{\boldsymbol{c}}$ and $\mathsf{B}_{\boldsymbol{c}}$, and lying as close as possible to the target direction $\boldsymbol{a}$. Referring to Figure 3 in Appendix B.1, this amounts to requiring that $\boldsymbol{c}$ is the orthogonal projection of $\boldsymbol{a}$ on the right edge of the square Q in the $\boldsymbol{ij}$-plane; such a square is the region of the plane where the approximating observables $\mathsf{A}_{\boldsymbol{c}}$ and $\mathsf{B}_{\boldsymbol{c}}$ are compatible (see Proposition 4, item (ii), in Appendix B.1). Using the parametrization $\boldsymbol{c}(\gamma)$ given in (33) for the right edge of Q, we see that this approach fixes $\gamma = \sqrt{2}a_2$. The entropic divergence of the corresponding bi-observable M_γ from (A, B) and the angle of the state $\rho(\phi)$ at which it is attained are the content of the BLW column.

NV: At the end of Section 2.4, we briefly discussed the proposal of [45, 49] to use approximating joint measurements whose marginals are noisy versions (NV) of the two target observables. In this approach, one approximates the target observables by means of a compatible couple $(\mathsf{A}_{\boldsymbol{c}}, \mathsf{B}_{\boldsymbol{c}})$ with $\boldsymbol{c} \parallel \boldsymbol{a}$. Still making reference to Figure 3 in the appendix, the best choice is then picking $\boldsymbol{c}$ as large as possible; in this way, $\boldsymbol{c}$ lies where the right edge of the compatibility square Q intersects the line joining $\boldsymbol{a}$ and the origin. With our parametrization $\boldsymbol{c}(\gamma)$ of the edge, this implies $\gamma = a_2/a_1$. The results for this choice (together with the corresponding maximizing state) are reported in the last column.

3.3. Two conjugate observables in prime power dimension. We now consider two complementary observables in prime power dimension, realized by a couple of MUBs that are conjugated by the Fourier transform of a finite field. In general, the construction of a maximal set of MUBs in a prime power dimensional Hilbert space by using finite fields is well known since Wootters and Fields' seminal paper [53]; see also [54, Sect. 2] for a review, and [55–57] for a group theoretic perspective on the topic.

Let $\mathbb{F}$ be a finite field with characteristic p. We refer to [58, Sect. V.5] for the basic notions on finite fields. Here we only recall that p is a prime number, and $\mathbb{F}$ has cardinality $|\mathbb{F}| = p^n$ for some positive integer n. We need also the field trace $\operatorname{tr} : \mathbb{F} \to \mathbb{Z}_p$ defined by $\operatorname{tr} x = \sum_{k=0}^{n-1} x^{p^k}$ (see [58, Sect. VI.5] for its definition and properties).

We consider the Hilbert space $\mathcal{H} = \ell^2(\mathbb{F})$, with dimension $d = p^n$, and we let our target observables be the two sharp rank-one observables Q and P with outcome spaces $\mathcal{X} = \mathcal{Y} = \mathbb{F}$, given by

$$\mathsf{Q}(x) = |\,\delta_x\,\rangle\langle\,\delta_x\,|, \qquad \mathsf{P}(y) = |\,\omega_y\,\rangle\langle\,\omega_y\,|, \qquad \forall x, y \in \mathbb{F}. \tag{45}$$

In this formula, δ_x is the delta function at x, and

$$\omega_y(z) = \frac{1}{\sqrt{d}}\,\mathrm{e}^{\frac{2\pi\mathrm{i}}{p}\operatorname{tr} yz} \equiv (F^*\delta_y)(z) \quad \text{with} \quad F\phi(z) = \frac{1}{\sqrt{d}}\sum_{t\in\mathbb{F}} \mathrm{e}^{-\frac{2\pi\mathrm{i}}{p}\operatorname{tr} zt}\phi(t). \tag{46}$$

Since $|\,\langle\,\delta_x\,|\,\omega_y\,\rangle\,| = 1/\sqrt{d}$ for all x and y, the two orthonormal bases $\{\delta_x\}_{x\in\mathbb{F}}$ and $\{\omega_y\}_{y\in\mathbb{F}}$ satisfy the MUB condition. In particular, as a consequence of the bound in [17], their preparation uncertainty coefficient (18) is

$$c_{\mathrm{prep}}(\mathsf{Q}, \mathsf{P}) = \log d. \tag{47}$$

In (46), the operator $F : \mathcal{H} \to \mathcal{H}$ is the unitary *discrete Fourier transform* over the field $\mathbb{F}$. The observables Q and P are then an example of *Fourier conjugate MUBs*, as $\mathsf{P}(y) = F^*\mathsf{Q}(y)F$ for all $y \in \mathbb{F}$.

The definitions (45) and (46) should be compared with the analogous ones for MUBs that are conjugated by means of the Fourier transform over the cyclic ring $\mathbb{Z}_d$, see e.g. [59]. In the latter case, the Hilbert space is $\mathcal{H} = \ell^2(\mathbb{Z}_d)$, and the operator F in (46) is replaced by

$$\mathcal{F}\phi(z) = \frac{1}{\sqrt{d}}\sum_{t\in\mathbb{Z}} \mathrm{e}^{-\frac{2\pi\mathrm{i}}{d}zt}\phi(t) \tag{48}$$

(cf. [59, Eq. (4)]; note that no field trace appears in this formula). The two definitions are clearly the same if $\mathbb{F}$ coincides with the cyclic field $\mathbb{Z}_p$ (i.e. $n = 1$ and so $d = p$), but they are essentially different in general. Indeed, as observed in [54, Sect. 5.3], they are inequivalent already for $d = 2^2$.

The following theorem is the main result of this section.

Theorem 7. *For the two sharp observables* Q *and* P *defined in* (45)*, we have*

$$\log \frac{2\sqrt{d}}{\sqrt{d}+1} \leq c_{\mathrm{inc}}(\mathsf{Q},\mathsf{P}) = \max_{\substack{\rho\in\mathcal{S}(\mathcal{H})\\ \rho \text{ pure}}} \left[S\big(\mathsf{Q}^\rho \| \mathsf{Q}^\rho_{\lambda_0}\big) + S\big(\mathsf{P}^\rho \| \mathsf{P}^\rho_{\lambda_0}\big)\right] \leq 2\log\frac{2(d+1)}{d+3}\,, \quad (49)$$

where $\mathsf{Q}_{\lambda_0} = \lambda_0\mathsf{Q} + (1-\lambda_0)\mathsf{U}_{\mathbb{F}}$ *and* $\mathsf{P}_{\lambda_0} = \lambda_0\mathsf{P} + (1-\lambda_0)\mathsf{U}_{\mathbb{F}}$ *are the uniformly noisy versions of the observables* Q *and* P *with noise intensity*

$$1-\lambda_0 = \frac{\sqrt{d}}{2(\sqrt{d}+1)}\,. \quad (50)$$

An optimal approximate joint measurement $\mathsf{M} \in \mathcal{M}_{\mathrm{inc}}(\mathsf{A},\mathsf{B})$ *is given by*

$$\mathsf{M}_0(x,y) = \frac{1}{2(d+\sqrt{d})}\,|\,\psi_{x,y}\,\rangle\langle\,\psi_{x,y}\,| \quad \textit{with} \quad \psi_{x,y} = \delta_x + \mathrm{e}^{-\frac{2\pi\mathrm{i}}{p}\,\mathrm{tr}\,xy} F\delta_{-y}. \quad (51)$$

If $p \neq 2$*, then* M_0 *is the unique optimal approximate joint measurement of* Q *and* P*, i.e.* $\mathcal{M}_{\mathrm{inc}}(\mathsf{Q},\mathsf{P}) = \{\mathsf{M}_0\}$.

As in the case of the two spin-1/2 components, the proof of this theorem relies on a detailed study of the symmetries of the pair of observables (Q,P), and a subsequent application of Theorem 4. The symmetries and the proof of the theorem are given in Appendix C. Here we briefly comment on its statements and provide a simple example.

Remark 5. 1. Since Q and P are sharp, the inequality (49) also gives a bound for the index $c_{\mathrm{ed}}(\mathsf{Q},\mathsf{P}) = c_{\mathrm{inc}}(\mathsf{Q},\mathsf{P})$.
2. The two bounds in (49) are not asymptotically optimal for $d \to \infty$, as the lower bound tends to 1 while the upper bound goes to 2.
3. The value in (50) is the minimal noise intensity making the two uniformly noisy observables Q_{λ_0} and P_{λ_0} compatible [59, Prop. 5 and Ex. 1].
4. In the terminology of [24, 25], the bi-observable in (51) is the *covariant phase-space observable* generated by the state $[\sqrt{d}/(2\sqrt{d}+2)]\,|\,\psi_{0,0}\,\rangle\langle\,\psi_{0,0}\,| = d\mathsf{M}_0(0,0)$ (see (86) and the discussion below it for further details on covariant phase-space observables).
5. Our choice of using the field $\mathbb{F}$ instead of the ring $\mathbb{Z}_d$ in defining the Fourier operator in (46), and the consequent restriction to only prime power dimensional systems, comes from the fact that the resulting MUBs (45) share dilational symmetries that are not present in the $\mathcal{F}$-conjugate ones. These extra symmetries drastically reduce the number of parameters to be optimized for finding an element of $\mathcal{M}_{\mathrm{inc}}(\mathsf{A},\mathsf{B})$ (see Remark 10.2 for further details).

6. The uniqueness property of the optimal approximate joint measurement M_0 in odd prime power dimensions should be compared with the measurement uncertainty region for two qudit observables found in [10, Sect. 5.3]. In particular, we remark that there is a whole family of covariant phase-space observables saturating the uncertainty bound of [10, Eq. (38)]. Our optimal bi-observable M_0 just corresponds to one of them, that is, the one generated by the state $\rho = d\mathsf{M}_0(0,0)$.
7. When $d = 2^n$ with $n \geq 2$, it is not clear whether or not the set $\mathcal{M}_{\mathrm{inc}}(\mathsf{Q},\mathsf{P})$ is made up of a unique bi-observable. However, in the simplest case $d = 2$ we have already shown that $\mathcal{M}_{\mathrm{inc}}(\mathsf{Q},\mathsf{P}) = \{\mathsf{M}_0\}$ (see Theorem 5).

Example 1 (Two orthogonal spin-1/2 components). Let us consider as target observables the two sharp spin-1/2 components $\mathsf{X},\mathsf{Y} \in \mathcal{M}(\{+1,-1\})$ associated with the first two Pauli matrices, defined in (30). This is the easiest example of two Fourier conjugate MUBs. To see this, take the cyclic field $\mathbb{F} = \mathbb{Z}_2 \equiv \{0,1\}$, corresponding to the choice $d = p = 2$, $n = 1$, $\operatorname{tr} x = x$, and identify the observables $\mathsf{Q}(x) = \mathsf{X}\left((-1)^x\right)$ and $\mathsf{P}(y) = \mathsf{Y}\left((-1)^y\right)$ $(x, y = 0,1)$ by setting $\sigma_1 = |\delta_0\rangle\langle\delta_0| - |\delta_1\rangle\langle\delta_1|$, and $\sigma_2 = |\delta_0\rangle\langle\delta_1| + |\delta_1\rangle\langle\delta_1|$. With this identification, the discrete Fourier transform becomes $F = (\sigma_1 + \sigma_2)/\sqrt{2} \equiv \mathrm{i}\exp\{-\mathrm{i}\pi \boldsymbol{n}\cdot\boldsymbol{\sigma}/2\}$. We have already found in (31) the optimal joint observable of X and Y, together with the value of the entropic incompatibility degree. These are precisely the bi-observable and the lower bound found in Theorem 7.

4. Entropic measurement uncertainty relations for n observables

Uncertainty relations have been studied also in the case of more than two observables, see e.g. [13,19,22] for the case of entropic PURs. Both our entropic coefficients (10) and (12) (and the related MURs) can be generalized to the case of $n > 2$ target observables. However, in the case of $c_{\mathrm{ed}}(\mathsf{A}_1,\ldots,\mathsf{A}_n)$ an order of observation has to be fixed, and one needs to point out the subset of the observables for which imprecise measurements are allowed (the analogues of the observable A in the binary case of $c_{\mathrm{ed}}(\mathsf{A},\mathsf{B})$) from those observables that are kept fixed and get disturbed by the other measurements (similar to B in $c_{\mathrm{ed}}(\mathsf{A},\mathsf{B})$). Thus, different definitions of c_{ed} are possible in the n-observable case. This leads us to generalize only the entropic incompatibility degree $c_{\mathrm{inc}}(\mathsf{A}_1,\ldots,\mathsf{A}_n)$, whose definition is straightforward and gives a lower bound for c_{ed}, independently of its possible definitions.

4.1. Entropic incompatibility degree and MURs. Let $\mathsf{A}_1,\ldots,\mathsf{A}_n$ be n fixed observables with outcome sets $\mathcal{X}_1,\ldots,\mathcal{X}_n$, respectively. As usual, we assume that all the sets $\mathcal{X}_i$ are finite. The observables with outcomes in the product set $\mathcal{X}_{1\cdots n} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_n$ are called multi-observables, and we use the notation $\mathcal{M}(\mathcal{X}_{1\cdots n})$ for the set of all such observables. If $\mathsf{M} \in \mathcal{M}(\mathcal{X}_{1\cdots n})$, its *i-th marginal observable* is the element $\mathsf{M}_{[i]} \in \mathcal{M}(\mathcal{X}_i)$, with

$$\mathsf{M}_{[i]}(x) = \sum_{x_j \in \mathcal{X}_j : j \neq i} \mathsf{M}(x_1,\ldots,x_{i-1},x,x_{i+1},\ldots,x_n).$$

The notion of compatibility straightforwardly extends to the case of n observables.

As in the $n = 2$ case, we regard any $\mathsf{M} \in \mathcal{M}(\mathcal{X}_{1\cdots n})$ as an approximate joint measurement of $\mathsf{A}_1,\ldots,\mathsf{A}_n$. For all $\rho \in \mathcal{S}(\mathcal{H})$, the total amount of information loss in the distribution approximations $\mathsf{A}_i^\rho \simeq \mathsf{M}_{[i]}^\rho$, $i = 1, \ldots, n$, is the sum of the respective relative entropies. Then, we have the following generalization of Definitions 1, 2 and 3.

Definition 5. *For any multi-observable* $\mathsf{M} \in \mathcal{M}(\mathcal{X}_{1\cdots n})$*, the* error function *of the approximation* $(\mathsf{A}_1, \ldots, \mathsf{A}_n) \simeq (\mathsf{M}_{[1]}, \ldots, \mathsf{M}_{[n]})$ *is the state-dependent quantity*

$$S[\mathsf{A}_1, \ldots, \mathsf{A}_n \| \mathsf{M}](\rho) = \sum_{i=1}^{n} S\big(\mathsf{A}_i^\rho \| \mathsf{M}_{[i]}^\rho\big). \tag{52}$$

The entropic divergence *of* $\mathsf{M} \in \mathcal{M}(\mathcal{X}_{1\cdots n})$ *from* $(\mathsf{A}_1, \ldots, \mathsf{A}_n)$ *is*

$$D\big(\mathsf{A}_1, \ldots, \mathsf{A}_n \| \mathsf{M}\big) = \sup_{\rho \in \mathcal{S}(\mathcal{H})} S[\mathsf{A}_1, \ldots, \mathsf{A}_n \| \mathsf{M}](\rho). \tag{53}$$

The entropic incompatibility degree *of the observables* $\mathsf{A}_1, \ldots, \mathsf{A}_n$ *is*

$$c_{\mathrm{inc}}\big(\mathsf{A}_1, \ldots, \mathsf{A}_n\big) = \inf_{\mathsf{M} \in \mathcal{M}(\mathcal{X}_{1\cdots n})} D\big(\mathsf{A}_1, \ldots, \mathsf{A}_n \| \mathsf{M}\big). \tag{54}$$

We still denote by

$$\mathcal{M}_{\mathrm{inc}}(\mathsf{A}_1, \ldots, \mathsf{A}_n) = \operatorname*{arg\,min}_{\mathsf{M} \in \mathcal{M}(\mathcal{X}_{1\cdots n})} D\big(\mathsf{A}_1, \ldots, \mathsf{A}_n \| \mathsf{M}\big)$$

the set of the *optimal approximate joint measurements* of $\mathsf{A}_1, \ldots, \mathsf{A}_n$. As in the case with $n = 2$, the optimality of a multi-observable M depends only on its marginals $\mathsf{M}_{[i]}$, since the entropic divergence itself depends only on such marginals.

We have the following extension of Theorems 1, 2 and 3.

Theorem 8. *Let* $\mathsf{A}_i \in \mathcal{M}(\mathcal{X}_i)$*,* $i = 1, \ldots, n$*, be the target observables. The error function, entropic divergence and incompatibility degree satisfy the following properties.*

(i) The function $S[\mathsf{A}_1, \ldots, \mathsf{A}_n \| \mathsf{M}] : \mathcal{S}(\mathcal{H}) \to [0, +\infty]$ *is convex and LSC,* $\forall \mathsf{M} \in \mathcal{M}(\mathcal{X}_{1\cdots n})$*.*
(ii) The function $D\big(\mathsf{A}_1, \ldots, \mathsf{A}_n \| \cdot\,\big) : \mathcal{M}(\mathcal{X}_{1\cdots n}) \to [0, +\infty]$ *is convex and LSC.*
(iii) For any $\mathsf{M} \in \mathcal{M}(\mathcal{X}_{1\cdots n})$*, the following three statements are equivalent:*
  (a) $D\big(\mathsf{A}_1, \ldots, \mathsf{A}_n \| \mathsf{M}\big) < +\infty$*,*
  (b) $\ker \mathsf{M}_{[i]}(x) \subseteq \ker \mathsf{A}_i(x) \quad \forall x,\ \forall i$*,*
  (c) $S[\mathsf{A}_1, \ldots, \mathsf{A}_n \| \mathsf{M}]$ *is bounded and continuous.*
(iv) $D\big(\mathsf{A}_1, \ldots, \mathsf{A}_n \| \mathsf{M}\big) = \max_{\substack{\rho \in \mathcal{S}(\mathcal{H}) \\ \rho \text{ pure}}} S[\mathsf{A}_1, \ldots, \mathsf{A}_n \| \mathsf{M}](\rho)$*, where the maximum can be any value in the extended interval* $[0, +\infty]$*.*
(v) The quantities $S[\mathsf{A}, \mathsf{B} \| \mathsf{M}](\rho)$*,* $D\big(\mathsf{A}_1, \ldots, \mathsf{A}_n \| \mathsf{M}\big)$ *and* $c_{\mathrm{inc}}(\mathsf{A}_1, \ldots, \mathsf{A}_n)$ *are invariant under an overall unitary conjugation of the state* ρ *and the observables* $\mathsf{A}_1, \ldots, \mathsf{A}_n$ *and* M*, and they do not depend on the labelling of the outcomes in* $\mathcal{X}_1, \ldots, \mathcal{X}_n$*.*
(vi) $c_{\mathrm{inc}}(\mathsf{A}_{\sigma(1)}, \ldots, \mathsf{A}_{\sigma(n)}) = c_{\mathrm{inc}}(\mathsf{A}_1, \ldots, \mathsf{A}_n)$ *for any permutation* σ *of the index set* $\{1, \ldots, n\}$*.*
(vii) The entropic incompatibility coefficient $c_{\mathrm{inc}}(\mathsf{A}_1, \ldots, \mathsf{A}_n)$ *is always finite, and it satisfies the bounds*

$$c_{\mathrm{inc}}(\mathsf{A}_1, \ldots, \mathsf{A}_n) \leq \sum_{i=1}^{n} \log |\mathcal{X}_i| - \inf_{\rho \in \mathcal{S}(\mathcal{H})} \sum_{i=1}^{n} H(\mathsf{A}_i^\rho), \tag{55}$$

$$c_{\mathrm{inc}}(\mathsf{A}_1,\ldots,\mathsf{A}_n) \leq \sum_{i=1}^n \log \frac{n(d+1)}{d+n+(n-1)\min_{x\in\mathcal{X}_i} \operatorname{Tr}\{\mathsf{A}_i(x)\}} \\ \leq n\log\frac{n(d+1)}{d+n} \leq n\log n. \qquad (56)$$

(viii) The set $\mathcal{M}_{\mathrm{inc}}(\mathsf{A}_1,\ldots,\mathsf{A}_n)$ *is a nonempty convex compact subset of* $\mathcal{M}(\mathcal{X}_{1\cdots n})$.

(ix) $c_{\mathrm{inc}}(\mathsf{A}_1,\ldots,\mathsf{A}_n) = 0$ *if and only if the observables* $\mathsf{A}_1,\ldots,\mathsf{A}_n$ *are compatible, and in this case* $\mathcal{M}_{\mathrm{inc}}(\mathsf{A}_1,\ldots,\mathsf{A}_n)$ *is the set of all the joint measurements of* $\mathsf{A}_1,\ldots,\mathsf{A}_n$.

(x) If $\mathsf{A}_{n+1} \in \mathcal{M}(\mathcal{X}_{n+1})$ *is another observable, then we have* $c_{\mathrm{inc}}(\mathsf{A}_1,\ldots,\mathsf{A}_{n+1}) \geq c_{\mathrm{inc}}(\mathsf{A}_1,\ldots,\mathsf{A}_n)$.

Proof. The proofs of items (i)–(vi), (viii) and (ix) are straightforward extensions of the analogous ones for two observables.

In item (vii), the upper bound (55) follows by evaluating the entropic divergence of the uniform observable $\mathsf{U} = (u_{\mathcal{X}_1}\otimes\cdots\otimes u_{\mathcal{X}_n})\,\mathbb{1}$ from $(\mathsf{A}_1,\ldots,\mathsf{A}_n)$:

$$c_{\mathrm{inc}}(\mathsf{A}_1,\ldots,\mathsf{A}_n) \leq D\big(\mathsf{A}_1,\ldots,\mathsf{A}_n\|\mathsf{U}\big) = \sup_{\rho\in\mathcal{S}(\mathcal{H})}\sum_{i=1}^n S\big(\mathsf{A}_i^\rho\|u_{\mathcal{X}_i}\big)$$

$$= \sup_{\rho\in\mathcal{S}(\mathcal{H})}\sum_{i=1}^n \left[\log|\mathcal{X}_i| - H(\mathsf{A}_i^\rho)\right] \qquad \text{by Proposition 1, item (iii);}$$

this yields (55).

The upper bound (56) follows by using an approximate cloning argument, just as in the case of only two observables. Indeed, the optimal approximate n-cloning channel is the map

$$\Phi : \mathcal{S}(\mathcal{H}) \to \mathcal{S}(\mathcal{H}^{\otimes n}), \qquad \Phi(\rho) = \frac{d!n!}{(d+n-1)!} S_n(\rho\otimes\mathbb{1}^{\otimes(n-1)})S_n,$$

where S_n is the orthogonal projection of $\mathcal{H}^{\otimes n}$ onto its symmetric subspace $\operatorname{Sym}(\mathcal{H}^{\otimes n})$ [50]. Evaluating the marginals of the multi-observable $\mathsf{M}_{\mathrm{cl}} = \Phi^*(\mathsf{A}_1\otimes\cdots\otimes\mathsf{A}_n)$, we obtain the noisy versions

$$\mathsf{M}_{\mathrm{cl}\,[i]} = \mathsf{A}_{i\,\lambda_{\mathrm{cl}}}, \qquad \text{where} \qquad \lambda_{\mathrm{cl}} = \frac{d+n}{n(d+1)}$$

(see [51]). Since $c_{\mathrm{inc}}(\mathsf{A}_1,\ldots,\mathsf{A}_n) \leq D\big(\mathsf{A}_1,\ldots,\mathsf{A}_n\|\mathsf{M}_{\mathrm{cl}}\big)$, a computation similar to the one for obtaining the bound (17) in Section 2.4 then yields the bounds (56).

Finally, in order to prove item (x), take any $\mathsf{M}' \in \mathcal{M}(\mathcal{X}_1\times\cdots\times\mathcal{X}_{n+1})$, and let

$$\mathsf{M}(x_1,\ldots,x_n) = \sum_{x\in\mathcal{X}_{n+1}} \mathsf{M}'(x_1,\ldots,x_n,x).$$

We have $\mathsf{M}'_{[i]} = \mathsf{M}_{[i]}$ for all $i = 1,\ldots,n$, hence

$$c_{\mathrm{inc}}(\mathsf{A}_1,\ldots,\mathsf{A}_n) \leq D\big(\mathsf{A}_1,\ldots,\mathsf{A}_n\|\mathsf{M}\big) = \sup_\rho\sum_{i=1}^n S\big(\mathsf{A}_i^\rho\|\mathsf{M}_{[i]}^\rho\big) \\ \leq \sup_\rho\sum_{i=1}^{n+1} S\big(\mathsf{A}_i^\rho\|\mathsf{M}_{[i]}'^\rho\big) = D\big(\mathsf{A}_1,\ldots,\mathsf{A}_{n+1}\|\mathsf{M}'\big).$$

Item (x) then follows by taking the infimum over M'. □

The monotonicity property (x), which is specific of the many observable case, is another desirable feature for an incompatibility coefficient: the amount of incompatibility cannot decrease when an extra observable is added.

Remark 6 (MURs). Theorem 8 gives the following extension of the entropic MURs (13) and (14):

$$D\big(\mathsf{A}_1,\ldots,\mathsf{A}_n\|\mathsf{M}\big)\geq c_{\mathrm{inc}}(\mathsf{A}_1,\ldots,\mathsf{A}_n),\qquad \forall\mathsf{M}\in\mathcal{M}(\mathcal{X}_{1\cdots n}),\tag{57}$$

$$\forall\mathsf{M}\in\mathcal{M}(\mathcal{X}_{1\cdots n}),\ \exists\rho\in\mathcal{S}(\mathcal{H}):\ \sum_{i=1}^{n}S\big(\mathsf{A}_i^\rho\|\mathsf{M}_{[i]}^\rho\big)\geq c_{\mathrm{inc}}(\mathsf{A}_1,\ldots,\mathsf{A}_n).\tag{58}$$

Finally, suppose the product space $\mathcal{X}_{1\cdots n}$ carries the action of a finite symmetry group G, which also acts on the quantum system Hilbert space $\mathcal{H}$ by means of a projective unitary representation U. These actions then extend to the set of states $\mathcal{S}(\mathcal{H})$, the set of probabilities $\mathcal{P}(\mathcal{X}_{1\cdots n})$ and the set of multi-observables $\mathcal{M}(\mathcal{X}_{1\cdots n})$ exactly as in Section 3.1. Similarly, for any $\mathsf{M}\in\mathcal{M}(\mathcal{X}_{1\cdots n})$, we can define its covariant version M_G. Then, the content of Theorem 4 and Corollary 1 can be translated to the case of n observables as follows.

Theorem 9. *Let $\mathsf{A}_i\in\mathcal{M}(\mathcal{X}_i)$, $i=1,\ldots,n$, be the target observables. Suppose the finite group G acts on both the output space $\mathcal{X}_{1\cdots n}$ and the index set $\{1,\ldots,n\}$, and it also acts with a projective unitary representation U on $\mathcal{H}$. Moreover, assume that G is generated by a subset $S_G\subseteq G$ such that, for every $g\in S_G$ and $i\in\{1,\ldots,n\}$, there exists a bijective map $f_{g,i}:\mathcal{X}_i\to\mathcal{X}_{gi}$ for which*

(a) $g(x_1,\ldots,x_n)_{gi}=f_{g,i}(x_i)$ for all $(x_1,\ldots,x_n)\in\mathcal{X}_{1\cdots n}$,
(b) $U_g\mathsf{A}_i(x_i)U_g^*=\mathsf{A}_{gi}(f_{g,i}(x_i))$ for all $x_i\in\mathcal{X}_i$.

Then,

- $D\big(\mathsf{A}_1,\ldots,\mathsf{A}_n\|g\mathsf{M}\big)=D\big(\mathsf{A}_1,\ldots,\mathsf{A}_n\|\mathsf{M}\big)$ for all $\mathsf{M}\in\mathcal{M}(\mathcal{X}_{1\cdots n})$ and $g\in G$;
- the set $\mathcal{M}_{\mathrm{inc}}(\mathsf{A}_1,\ldots,\mathsf{A}_n)$ is G-invariant;
- for any $\mathsf{M}\in\mathcal{M}_{\mathrm{inc}}(\mathsf{A}_1,\ldots,\mathsf{A}_n)$, we have $\mathsf{M}_G\in\mathcal{M}_{\mathrm{inc}}(\mathsf{A}_1,\ldots,\mathsf{A}_n)$;
- there exists a G-covariant observable in $\mathcal{M}_{\mathrm{inc}}(\mathsf{A}_1,\ldots,\mathsf{A}_n)$.

Proof. As in the proof of Theorem 4, it is not restrictive to assume that $S_G=G$. For all $p\in\mathcal{P}(\mathcal{X}_{1\cdots n})$, condition (a) implies

$$\begin{aligned}
gp_{[i]}(x_i)&=\sum_{\substack{x_j\in\mathcal{X}_j\\ \text{s.t. } j\neq i}}gp(x_1,\ldots,x_n)=\sum_{\substack{x_{gj}\in\mathcal{X}_{gj}\\ \text{s.t. } gj\neq i}}p(f_{g^{-1},g1}(x_{g1}),\ldots,f_{g^{-1},gn}(x_{gn}))\\
&=\sum_{\substack{y_j\in\mathcal{X}_j\\ \text{s.t. } gj\neq i}}p(y_1,\ldots,y_n)\qquad\qquad \text{where } y_j=f_{g^{-1},gj}(x_{gj})\\
&=\sum_{\substack{y_j\in\mathcal{X}_j\\ \text{s.t. } j\neq g^{-1}i}}p(y_1,\ldots,y_n)=p_{[g^{-1}i]}(y_{g^{-1}i})\\
&=p_{[g^{-1}i]}(f_{g^{-1},i}(x_i)),
\end{aligned}$$

and hence

$$(gp_{[1]} \otimes \cdots \otimes gp_{[n]})(x_1, \ldots, x_n) = \prod_{i=1}^n gp_{[i]}(x_i) = \prod_{i=1}^n p_{[g^{-1}i]}(f_{g^{-1},i}(x_i))$$
$$= \prod_{i=1}^n p_{[i]}(f_{g^{-1},gi}(x_{gi})) = g(p_{[1]} \otimes \cdots \otimes p_{[n]})(x_1, \ldots, x_n).$$

Therefore,

$$gp_{[1]} \otimes \cdots \otimes gp_{[n]} = g(p_{[1]} \otimes \cdots \otimes p_{[n]}). \tag{59}$$

On the other hand, by condition (b) we have $\mathsf{A}_i^{g\rho}(x_i) = \mathsf{A}^\rho_{g^{-1}i}(f_{g^{-1},i}(x_i))$, and then

$$(\mathsf{A}_1^{g\rho} \otimes \cdots \otimes \mathsf{A}_n^{g\rho})(x_1, \ldots, x_n) = \prod_{i=1}^n \mathsf{A}^\rho_{g^{-1}i}(f_{g^{-1},i}(x_i)) = \prod_{i=1}^n \mathsf{A}_i^\rho(f_{g^{-1},gi}(x_{gi}))$$
$$= g(\mathsf{A}_1^\rho \otimes \cdots \otimes \mathsf{A}_n^\rho)(x_1, \ldots, x_n),$$

that is,

$$\mathsf{A}_1^{g\rho} \otimes \cdots \otimes \mathsf{A}_n^{g\rho} = g(\mathsf{A}_1^\rho \otimes \cdots \otimes \mathsf{A}_n^\rho). \tag{60}$$

Having established (59) and (60), the proof of the equality $D(\mathsf{A}_1, \ldots, \mathsf{A}_n \| g\mathsf{M}) = D(\mathsf{A}_1, \ldots, \mathsf{A}_n \| \mathsf{M})$ follows along the same lines of the proof of Theorem 4. The remaining statements are then proved as in Corollary 1. □

In the next section we will use Theorem 9 to solve the case of $n = 3$ orthogonal target spin-1/2 components. This is the basic example of a maximal set of $d+1$ MUBs in a d-dimensional Hilbert space. It is an open problem whether similar arguments lead to find the incompatibility index $c_{\mathrm{inc}}(\mathsf{Q}_1, \ldots, \mathsf{Q}_{d+1})$ of a maximal set of $d+1$ MUBs $\mathsf{Q}_1, \ldots, \mathsf{Q}_{d+1}$ whenever such a set of MUBs is known to exist, that is, for all prime powers d.

4.2. Three orthogonal spin-1/2 components. Let the target observables A_1, A_2, A_3 be three mutually orthogonal spin-1/2 components, that is, the sharp observables X, Y, $\mathsf{Z} \in \mathcal{M}(\{+1, -1\})$ associated with the three Pauli matrices; the observables X, Y are given in (30), and

$$\mathsf{Z}(z) = \frac{1}{2}\left(\mathbb{1} + z\sigma_3\right), \qquad \forall z \in \mathcal{Z} = \{+1, -1\}. \tag{61}$$

Then, we have the following three-spin version of Theorem 5.

Theorem 10. *Let* X, Y *and* Z *be the three orthogonal spin-1/2 components* (30), (61). *Then, for the following two tri-observables* $\mathsf{M}_0, \mathsf{M}_1 \in \mathcal{M}(\mathcal{X} \times \mathcal{Y} \times \mathcal{Z})$

$$\mathsf{M}_0(x, y, z) = \frac{1}{8}\left[\mathbb{1} + \frac{1}{\sqrt{3}}(x\sigma_1 + y\sigma_2 + z\sigma_3)\right] \tag{62}$$

$$\begin{gathered} \mathsf{M}_1(1,1,-1) = 2\mathsf{M}_0(1,1,-1), \quad \mathsf{M}_1(1,-1,1) = 2\mathsf{M}_0(1,-1,1), \\ \mathsf{M}_1(-1,1,1) = 2\mathsf{M}_0(-1,1,1), \quad \mathsf{M}_1(-1,-1,-1) = 2\mathsf{M}_0(-1,-1,-1), \\ \mathsf{M}_1(x,y,z) = 0 \quad \textit{otherwise,} \end{gathered} \tag{63}$$

we have $\mathsf{M}_0, \mathsf{M}_1 \in \mathcal{M}_{\rm inc}(\mathsf{X},\mathsf{Y},\mathsf{Z})$. *If* ρ_e *is the projection on any eigenvector of* σ_1, σ_2 *or* σ_3, *then, for* $i = 0, 1$,

$$c_{\rm inc}(\mathsf{X},\mathsf{Y},\mathsf{Z}) = S[\mathsf{X},\mathsf{Y},\mathsf{Z}\|\mathsf{M}_i](\rho_e) = \log\left(3-\sqrt{3}\right) \simeq 0.342497\,. \tag{64}$$

The description of the symmetry group of X, Y and Z, and the consequent application of Theorem 9 yielding the proof of Theorem 10, are provided in Appendix B.4.

Note that, differently from the case with only $n = 2$ spins, for $n = 3$ orthogonal spin-1/2 components there is not a unique optimal approximate joint measurement. It should be also remarked that, although $\mathsf{M}_0 \neq \mathsf{M}_1$, both optimal approximate joint measurements given in (62), (63) have the same marginals, so that $S[\mathsf{X},\mathsf{Y},\mathsf{Z}\|\mathsf{M}_0] = S[\mathsf{X},\mathsf{Y},\mathsf{Z}\|\mathsf{M}_1]$. Indeed, they are the equally noisy observbles

$$\mathsf{M}_{0[1]} = \mathsf{M}_{1[1]} = \mathsf{X}_{\frac{1}{\sqrt{3}}}, \qquad \mathsf{M}_{0[2]} = \mathsf{M}_{1[2]} = \mathsf{Y}_{\frac{1}{\sqrt{3}}}, \qquad \mathsf{M}_{0[3]} = \mathsf{M}_{1[3]} = \mathsf{Z}_{\frac{1}{\sqrt{3}}}. \tag{65}$$

The construction of the tri-observable M_1 is taken from [60, Sect. VI]. It is an open question whether $\left(\mathsf{X}_{1/\sqrt{3}},\mathsf{Y}_{1/\sqrt{3}},\mathsf{Z}_{1/\sqrt{3}}\right)$ is the unique triple of compatible observables optimally approximating $(\mathsf{X},\mathsf{Y},\mathsf{Z})$; see also Remark 9 in Appendix B.4 for further comments.

5. Conclusions

We have formulated and proved entropic MURs for discrete observables in a finite-dimensional Hilbert space. In doing so, we have considered target observables A and B described by general POVMs, not only sharp observables. Our formulation employs the relative entropy to quantify the total amount of information that is lost when A and B are approximated with the marginals $\mathsf{M}_{[1]}$ and $\mathsf{M}_{[2]}$ of a bi-observable M. Such an information loss is the state-dependent error (8); maximizing it over all states ρ and then minimizing the result over the bi-observables M, we have derived our MURs (14): for every approximating bi-observable M, there is always a state ρ such that the total information loss of the approximation $(\mathsf{A}^\rho,\mathsf{B}^\rho) \simeq (\mathsf{M}^\rho_{[1]},\mathsf{M}^\rho_{[2]})$ is not less than a minimal threshold $c(\mathsf{A},\mathsf{B})$, independent of M and strictly positive when the target observables are incompatible.

Minimizing over different sets of bi-observables yields MURs with different meanings. If M varies over all the POVMs on the product set of the A- and B-outcomes, then the resulting index $c_{\rm inc}(\mathsf{A},\mathsf{B})$ is the minimal error potentially affecting all possible approximate joint measurements of A and B. On the other hand, if M is only allowed to vary over the subset of all the sequential measurements of an approximation of A followed by B, we obtain the minimal information loss $c_{\rm ed}(\mathsf{A},\mathsf{B})$ due to the error/disturbance tradeoff. We have proved that the two indexes remarkably coincide when the second observable B is sharp; we have also given explicit examples, involving general POVMs, where the two indexes actually differ.

The two coefficients $c_{\rm inc}$ and $c_{\rm ed}$ play a double role: on the one side, they are the lower bounds of our entropic MURs, as just described above; on the other side, they also properly quantify the degree of (total or sequential) incompatibility of the target observables. The latter interpretation is justified since $c_{\rm inc}$ and $c_{\rm ed}$ only depend on A and B, as well as by the remarkable properties of the two indexes (Theorem 2). In particular, the existence of an index allows to establish whether a couple of observables is more or less incompatible than another one. For instance, the incompatibility degree of two spin-1/2 components grows by increasing the angle between their directions, as naturally expected (see Figure 2).

Due to the double optimization in the definitions of $c_{\rm inc}(\mathsf{A},\mathsf{B})$ and $c_{\rm ed}(\mathsf{A},\mathsf{B})$, it is not easy to explicitly compute them and their corresponding optimal approximate joint measurements. Anyway, in Theorem 4 we have shown how one can use general symmetry arguments in order to simplify the problem. We have then applied this method to two spin-1/2 components (Theorems 5 and 6), and two Fourier conjugate MUBs in prime power dimension (Theorem 7).

A peculiar feature of our MURs is that in several cases there is actually a unique optimal approximate joint measurement. Indeed, in the two spin and MUB examples, we have uniqueness for all the cases in which we have managed to completely characterize the sets $\mathcal{M}_{\rm inc}$ and $\mathcal{M}_{\rm ed}$ of the optimally approximating joint measurements. We conjecture that this is still true also for the two nonorthogonal spin-1/2 components, and the Fourier conjugated MUBs in even prime power dimensions, for which up to now we have only partial results.

One nice aspect of our approach is that it naturally and easily generalizes to more than two observables. We have done this extension only for the entropic incompatibility degree $c_{\rm inc}$ (Theorems 8 and 9), as the multi-observable interpretation of $c_{\rm ed}$ is less transparent. As an application, we have computed the index $c_{\rm inc}$ for three orthogonal spin-1/2 components X, Y and Z; although in this case there is still a unique covariant approximate joint measurement, the main difference with the two spin case is that $\mathcal{M}_{\rm inc}(\mathsf{X},\mathsf{Y},\mathsf{Z})$ is not a singleton set now.

Many problems still remain open, as it is not clear how to analytically or at least numerically compute $c_{\rm inc}$ and $c_{\rm ed}$ and the corresponding optimal approximate joint measurements for an arbitrary couple of target observables. Explicit results would be desirable for physically relevant observables other than those considered in Sections 3.2, 3.3 and 4.2 (e.g. two or more spin-s components with $s > 1/2$, two or more MUBs in arbitrary dimensions and possibly not Fourier conjugate, etc.). A possible generalization is to include also systems in presence of “quantum memories”; indeed, this extension has recently been studied in the case of entropic PURs [22, 61, 62]. More importantly, the theory we have developed is restricted to discrete observables in a finite-dimensional Hilbert space. The bound $c_{\rm inc}(\mathsf{A}_1,\ldots,\mathsf{A}_n) \leq n\log n$ appearing in (17) and (56), which is independent of the number of the outcomes and the dimension of $\mathcal{H}$, suggests that it would be possible to generalize the theory to arbitrary observables in a separable Hilbert space. However, this is not a straightforward extension; indeed, the first results on position and momentum [34] already show that the error function (7) needs to be restricted to only particular classes of states, in order to avoid $c_{\rm inc} = +\infty$, merely due to classical effects.

A. Examples of compatible but not sequentially compatible observables

First example from [27]. Apart from an exchange of A and B and some explicit computations, this example is taken from [27, Sect. III.C, and the end of Sect. III.A]. With $\mathcal{H} = \mathbb{C}^3$, $\mathcal{X} = \{1,2\}$ and $\mathcal{Y} = \{1,\ldots,5\}$, the two target observables are defined by

$$\mathsf{A}(1) = \frac{1}{2}\begin{pmatrix} 2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad \mathsf{A}(2) = \frac{1}{2}\begin{pmatrix} 0 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix};$$

$$\mathsf{B}(1)=\frac{1}{4}\begin{pmatrix}2&0&-\sqrt{2}\\0&0&0\\-\sqrt{2}&0&1\end{pmatrix},\quad \mathsf{B}(2)=\frac{1}{10}\begin{pmatrix}0&0&0\\0&1&-2\\0&-2&4\end{pmatrix},\quad \mathsf{B}(3)=\frac{1}{2}\begin{pmatrix}0&0&0\\0&1&0\\0&0&0\end{pmatrix},$$

$$\mathsf{B}(4)=\frac{1}{10}\begin{pmatrix}0&0&0\\0&4&2\\0&2&1\end{pmatrix},\qquad \mathsf{B}(5)=\frac{1}{4}\begin{pmatrix}2&0&\sqrt{2}\\0&0&0\\\sqrt{2}&0&1\end{pmatrix}.$$

These two observables are compatible, and one can check that a joint observable is

$$\mathsf{M}(1,1)=\mathsf{B}(1),\quad \mathsf{M}(1,5)=\mathsf{B}(5),\quad \mathsf{M}(2,2)=\mathsf{B}(2),\quad \mathsf{M}(2,3)=\mathsf{B}(3),$$

$$\mathsf{M}(2,4)=\mathsf{B}(4),\qquad \mathsf{M}(1,2)=\mathsf{M}(1,3)=\mathsf{M}(1,4)=\mathsf{M}(2,1)=\mathsf{M}(2,5)=0.$$

This implies $c_{\mathrm{inc}}(\mathsf{A},\mathsf{B})=0$. Moreover, in [27, Sect. III.C] it is proved that: (1) there exists an instrument implementing B which does not disturb A; (2) any instrument implementing A disturbs B. By Theorem 2, item (vi), this implies $c_{\mathrm{ed}}(\mathsf{B},\mathsf{A})=0$ and $c_{\mathrm{ed}}(\mathsf{A},\mathsf{B})>0$.

Second example from [27]. This is the first example of [27, Sect. III.A], which we report in the particular case in which the noise parameters are fixed and equal; let us call them λ, with $\lambda\in\left(\frac{1}{2},\frac{2}{3}\right]$. The observables are two-valued ($\mathcal{X}=\mathcal{Y}=\{1,2\}$), and they are built up by using two noncommuting orthogonal projections P and Q: $[P,Q]\neq 0$. The joint observable M and its marginals are given by

$$\mathsf{M}(1,1)=(1-\lambda)\mathbb{1},\qquad \mathsf{M}(1,2)=(2\lambda-1)P,\qquad \mathsf{M}(2,1)=(2\lambda-1)Q,$$

$$\mathsf{M}(2,2)=\left(1-\frac{3}{2}\lambda\right)(P+Q)+\frac{\lambda}{2}(\mathbb{1}-P+\mathbb{1}-Q),$$

$$\mathsf{A}(1)=\mathsf{M}_{[1]}(1)=\lambda P+(1-\lambda)(\mathbb{1}-P),\quad \mathsf{A}(2)=\mathsf{M}_{[1]}(2)=\lambda(\mathbb{1}-P)+(1-\lambda)P,$$

$$\mathsf{B}(1)=\mathsf{M}_{[2]}(1)=\lambda Q+(1-\lambda)(\mathbb{1}-Q),\quad \mathsf{B}(2)=\mathsf{M}_{[2]}(2)=\lambda(\mathbb{1}-Q)+(1-\lambda)Q.$$

The observables A and B are compatible by construction, and so $c_{\mathrm{inc}}(\mathsf{A},\mathsf{B})=0$. In [27], it is proved that there does not exist any instrument implementing A which does not disturb B; it follows that $c_{\mathrm{ed}}(\mathsf{A},\mathsf{B})>0$ and, by exchanging P and Q, $c_{\mathrm{ed}}(\mathsf{B},\mathsf{A})>0$.

B. Symmetries and proofs for target spin-1/2 components

In this appendix, we describe the symmetry groups for two arbitrary and three orthogonal spin-1/2 components. Then, by using Theorem 4, we prove our main Theorems 6 (Appendix B.1), 5 (Appendix B.2) and 10 (Appendix B.4), and we provide the missing calculations in Section 3.2.2. Since the proof of Theorem 5 follows from Theorem 6 with the angle $\alpha=\pi/2$, here we prefer to reverse the order of the two proofs.

B.1. Incompatibility degree and optimal measurements for two spin-1/2 components. In this section, A and B are the spin-1/2 components defined in (26), with directions spanning an arbitrary angle $\alpha \in [0, \pi/2]$; the respective outcome spaces are $\mathcal{X} = \mathcal{Y} = \{-1, +1\}$. The symmetry group of A and B is the order 4 dihedral group $D_2 \subset SO(3)$ generated by the rotations $S_{D_2} = \{R_{\boldsymbol{n}}(\pi),\ R_{\boldsymbol{m}}(\pi)\}$, i.e. the 180° rotations around the bisectors $\boldsymbol{n}$ and $\boldsymbol{m}$ of the first two quadrants (see Figure 1). Here and in the following, our reference for the discrete subgroups of the rotation group is [63, pp. 77–79]. The natural action of the group D_2 on the outcome space $\mathcal{X} \times \mathcal{Y}$ is given by

$$R_{\boldsymbol{n}}(\pi)\,(x, y) = (y, x), \quad R_{\boldsymbol{m}}(\pi)\,(x, y) = (-y, -x), \qquad \forall (x, y) \in \mathcal{X} \times \mathcal{Y}. \tag{66a}$$

We then see that condition (ii.a) of Theorem 4 is satisfied for all $g \in S_{D_2}$. As the representation U of D_2 on $\mathbb{C}^2$, we take the restriction of the usual spin-1/2 projective representation of $SO(3)$; this gives

$$U\big(R_{\boldsymbol{n}}(\pi)\big) = \mathrm{e}^{-\mathrm{i}\pi\,\boldsymbol{n}\cdot\boldsymbol{\sigma}/2} \equiv -\mathrm{i}\,\boldsymbol{n}\cdot\boldsymbol{\sigma}, \quad U\big(R_{\boldsymbol{m}}(\pi)\big) = \mathrm{e}^{-\mathrm{i}\pi\,\boldsymbol{m}\cdot\boldsymbol{\sigma}/2} \equiv -\mathrm{i}\,\boldsymbol{m}\cdot\boldsymbol{\sigma}. \tag{66b}$$

It is easy to see that the observables A and B satisfy the relations

$$\begin{aligned} &U\big(R_{\boldsymbol{n}}(\pi)\big)\mathsf{A}(x)U\big(R_{\boldsymbol{n}}(\pi)\big)^* = \mathsf{B}(x), \quad U\big(R_{\boldsymbol{m}}(\pi)\big)\mathsf{A}(x)U\big(R_{\boldsymbol{m}}(\pi)\big)^* = \mathsf{B}(-x), \\ &U\big(R_{\boldsymbol{n}}(\pi)\big)\mathsf{B}(y)U\big(R_{\boldsymbol{n}}(\pi)\big)^* = \mathsf{A}(y), \quad U\big(R_{\boldsymbol{m}}(\pi)\big)\mathsf{B}(y)U\big(R_{\boldsymbol{m}}(\pi)\big)^* = \mathsf{A}(-y). \end{aligned} \tag{67}$$

This implies that also condition (ii.b) of Theorem 4 is fulfilled for all $g \in S_{D_2}$. Then, because of Remark 3, in order to find $c_{\mathrm{inc}}(\mathsf{A}, \mathsf{B})$, we are led to study the most general form of a D_2-covariant bi-observable and its marginals.

Proposition 4. *Let the dihedral group D_2 act on $\mathcal{X} \times \mathcal{Y}$ and $\mathcal{H}$ as in* (66)*. Then, the following facts hold.*

(i) The most general D_2-covariant bi-observable on $\mathcal{X} \times \mathcal{Y}$ is

$$\mathsf{M}(x, y) = \frac{1}{4}\left[(1 + \gamma xy)\,\mathbb{1} + (c_1 x + c_2 y)\,\sigma_1 + (c_2 x + c_1 y)\,\sigma_2\right], \tag{68}$$

with $\gamma \in \mathbb{R}$ and $\boldsymbol{c} = c_1\boldsymbol{i} + c_2\boldsymbol{j} \in \mathbb{R}^2$ such that

$$\sqrt{2}\,|c_1 + c_2| - 1 \le \gamma \le 1 - \sqrt{2}\,|c_1 - c_2|\,. \tag{69}$$

The marginals of M are $\mathsf{M}_{[1]} = \mathsf{A}_{\boldsymbol{c}}$ and $\mathsf{M}_{[2]} = \mathsf{B}_{\boldsymbol{c}}$, with $\mathsf{A}_{\boldsymbol{c}}$, $\mathsf{B}_{\boldsymbol{c}}$ defined in (28)*.*

(ii) Equation (28) *defines the marginals of a D_2-covariant bi-observable on $\mathcal{X} \times \mathcal{Y}$ if and only if the vector $\mathbf{c}$ belongs to the square*

$$Q = \{c_1\boldsymbol{i} + c_2\boldsymbol{j} : |c_1| \le 1/\sqrt{2},\ |c_2| \le 1/\sqrt{2}\}. \tag{70}$$

Proof. (i) The set $\mathcal{M}(\mathcal{X}\times\mathcal{Y})$ is a subset of the linear space $\mathcal{L}(\mathbb{C}^2)^{\mathcal{X}\times\mathcal{Y}} = \mathbb{C}^{\mathcal{X}\times\mathcal{Y}} \otimes \mathcal{L}(\mathbb{C}^2)$, where the set of the 16 products between one of the functions 1, x, y, xy and one of the operators $\mathbb{1}, \sigma_1, \sigma_2, \sigma_3$ provides a basis of linearly independent elements. Then, the most general bi-observable on $\mathcal{X} \times \mathcal{Y}$ is a linear combination of such products; it is easy to see that the covariance under the rotation $R_{\boldsymbol{n}}(\pi)R_{\boldsymbol{m}}(\pi)$ implies the vanishing of the coefficients of the products $x\mathbb{1}$, $y\mathbb{1}$, $xy\sigma_1$, $xy\sigma_2$, $1\sigma_1$, $1\sigma_2$, $x\sigma_3$, $y\sigma_3$. By taking into account also the normalization and selfadjointness conditions, we are left with

$$\mathsf{M}(x, y) = \frac{1}{4}\left[(1 + \gamma xy)\,\mathbb{1} + (c_1 x + c_2 y)\,\sigma_1 + (c_1' x + c_2' y)\,\sigma_2 + (c_3 + c_4 xy)\,\sigma_3\right],$$

with real coefficients γ, c_i and c_i'. By imposing the covariance under $R_{\boldsymbol{n}}(\pi)$, we get $c_1' = c_2$, $c_2' = c_1$, $c_3 = c_4 = 0$, and (68) is obtained. Finally, since $R_{\boldsymbol{m}}(\pi) = R_{\boldsymbol{n}}(\pi)R_{\boldsymbol{m}}(\pi)R_{\boldsymbol{n}}(\pi)$, the bi-observable (68) is covariant with respect to the whole group D_2. To impose the positivity of the operators $\mathsf{M}(x, y)$, it is enough to study the diagonal elements and the determinant of the 2×2–matrix representing (68). The positivity of the diagonal elements $\forall(x, y)$ gives $\gamma \in [-1, 1]$. By the positivity of the determinant,

$$(1 + \gamma xy)^2 \geq (c_1 x + c_2 y)^2 + (c_2 x + c_1 y)^2 , \qquad \forall(x, y) \in \mathcal{X} \times \mathcal{Y}.$$

The latter two conditions are equivalent to (69). Evaluating the marginals of (68) immediately yields the observables (28).

(ii) We begin by noticing that $\boldsymbol{c} \in Q$ is equivalent to

$$\sqrt{2}\,|c_1 + c_2| - 1 \leq 1 - \sqrt{2}\,|c_1 - c_2| . \tag{71}$$

For the marginals $\mathsf{A}_{\boldsymbol{c}}$ and $\mathsf{B}_{\boldsymbol{c}}$ of a D_2-covariant bi-observable, inequalities (69) trivially imply (71), and so $\boldsymbol{c} \in Q$ holds; alternatively, the same result follows from [26, Prop. 3]. Conversely, if $\mathsf{A}_{\boldsymbol{c}}$ and $\mathsf{B}_{\boldsymbol{c}}$ are as in (28) with $\boldsymbol{c} \in Q$, then by (71) we can always find γ as in (69). The D_2-covariant bi-observable corresponding to γ, c_1, c_2 then has marginals $\mathsf{A}_{\boldsymbol{c}}$ and $\mathsf{B}_{\boldsymbol{c}}$. □

Now we tackle the problem of evaluating the lower bound $c_{\rm inc}(\mathsf{A}, \mathsf{B})$ and finding the optimal covariant approximate joint measurements of the target spin-1/2 components (26). By Remark 3 and Proposition 4,

$$\begin{aligned} c_{\rm inc}(\mathsf{A}, \mathsf{B}) &= \min_{\substack{\mathsf{M}\in\mathcal{M}(\mathcal{X}\times\mathcal{Y}) \\ \mathsf{M}\ D_2\text{-covariant}}} \max_{\substack{\rho\in\mathcal{S}(\mathcal{H}) \\ \rho \text{ pure}}} \{S(\mathsf{A}^\rho \| \mathsf{M}^\rho_{[1]}) + S(\mathsf{B}^\rho \| \mathsf{M}^\rho_{[2]})\} \\ &= \min_{\boldsymbol{c}\in Q} \max_{\substack{\rho\in\mathcal{S}(\mathcal{H}) \\ \rho \text{ pure}}} \{S(\mathsf{A}^\rho \| \mathsf{A}^\rho_{\boldsymbol{c}}) + S(\mathsf{B}^\rho \| \mathsf{B}^\rho_{\boldsymbol{c}})\}, \end{aligned} \tag{72}$$

where Q is the square (70). Thus, the value of $c_{\rm inc}(\mathsf{A}, \mathsf{B})$ can be found by minimizing the function

$$D(\boldsymbol{c}) = \max_{\substack{\rho\in\mathcal{S}(\mathcal{H}) \\ \rho \text{ pure}}} \left\{S(\mathsf{A}^\rho \| \mathsf{A}^\rho_{\boldsymbol{c}}) + S(\mathsf{B}^\rho \| \mathsf{B}^\rho_{\boldsymbol{c}})\right\} \tag{73}$$

for $\boldsymbol{c}$ ranging inside Q.

Note that the domain of the function D can be extended to the whole disk C introduced in (29). In the domain C, $D(\boldsymbol{c}) = 0$ if and only if $\mathsf{A}_{\boldsymbol{c}} = \mathsf{A}$ and $\mathsf{B}_{\boldsymbol{c}} = \mathsf{B}$, which is equivalent to $\boldsymbol{c} = \boldsymbol{a}$. The regions C and Q in the $\boldsymbol{ij}$-plane are depicted in Figure 3.

We are now ready to prove our main result for the case of A and B being two arbitrary spin-1/2 components. Indeed, the key point is that, by convexity arguments, the minimization of the function D over the square Q fixes $c_1 = 1/\sqrt{2}$. This considerably simplifies the search of an optimal D_2-covariant bi-observable, as it reduces the involved parameters from the number of three (see (68)) to a single one (see (34)).

Proof (of Theorem 6). By (72), we have

$$c_{\rm inc}(\mathsf{A}, \mathsf{B}) = \min_{\boldsymbol{c}\in Q} D(\boldsymbol{c}). \tag{74}$$

Let us start with the case $\alpha \neq 0$. For $\boldsymbol{c} \in Q$, the observables $\mathsf{A}_{\boldsymbol{c}}$ and $\mathsf{B}_{\boldsymbol{c}}$ are compatible, and $D(\boldsymbol{c}) = D(\mathsf{A}, \mathsf{B} \| \mathsf{M})$ for any of their joint measurements M. By Theorem 1,

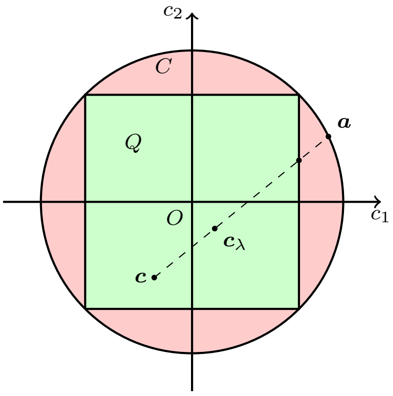

Figure 3. The existence disk (29) for the observables $\mathsf{A}_{\boldsymbol{c}}$ and $\mathsf{B}_{\boldsymbol{c}}$, and their compatibility square (70). The disk is the domain of the function D defined in (73), and the square is the subset over which D is minimized in (74).

item (iii), $D(\boldsymbol{c})$ is finite if and only if $\ker \mathsf{A}_{\boldsymbol{c}}(x) \subseteq \ker \mathsf{A}(x)$ and $\ker \mathsf{B}_{\boldsymbol{c}}(y) \subseteq \ker \mathsf{B}(y)$ for all x, y. In turn, this is equivalent to $\boldsymbol{c}$ not being any of the vertices V of the square Q, since $\mathsf{A}_{\boldsymbol{c}} = |\boldsymbol{c}|\, \mathsf{A}_{\boldsymbol{c}/|\boldsymbol{c}|} + (1 - |\boldsymbol{c}|)\mathsf{U}_{\mathcal{X}}$ and $\mathsf{B}_{\boldsymbol{c}} = |\boldsymbol{c}|\, \mathsf{B}_{\boldsymbol{c}/|\boldsymbol{c}|} + (1 - |\boldsymbol{c}|)\mathsf{U}_{\mathcal{Y}}$. Therefore, in the minimum (74) we can assume that $\boldsymbol{c} \notin V$, and so $D(\boldsymbol{c}) < +\infty$.

The mappings $\boldsymbol{c} \mapsto \mathsf{A}^\rho_{\boldsymbol{c}}$ and $\boldsymbol{c} \mapsto \mathsf{B}^\rho_{\boldsymbol{c}}$ are affine on the disk C for all $\rho \in \mathcal{S}(\mathcal{H})$, which, together with the convexity of the relative entropy, implies that the mappings $\boldsymbol{c} \mapsto S\big(\mathsf{A}^\rho \| \mathsf{A}^\rho_{\boldsymbol{c}}\big)$ and $\boldsymbol{c} \mapsto S\big(\mathsf{B}^\rho \| \mathsf{B}^\rho_{\boldsymbol{c}}\big)$ are convex; hence, such are their sum and the supremum D in (73). Moreover, we have already noticed that $D(\boldsymbol{c}) = 0$ if and only if $\boldsymbol{c} = \boldsymbol{a}$.

Making reference to Figure 3, let us take $\boldsymbol{c} \in Q \setminus V$ and introduce the line segment joining $\boldsymbol{c}$ and $\boldsymbol{a}$: $\boldsymbol{c}_\lambda = (1 - \lambda)\boldsymbol{c} + \lambda \boldsymbol{a}$, $\lambda \in [0, 1]$. By defining $D(\lambda) = D(\boldsymbol{c}_\lambda)$, a simple convexity argument (see Lemma 2 below) shows that the function $\lambda \mapsto D(\lambda)$ is finite and strictly decreasing on $[0, 1]$. Then, the minimum of $D(\boldsymbol{c}_\lambda)$ with respect to $\boldsymbol{c}_\lambda \in Q$ is attained where the line segment crosses the right side of the square, i.e. for $(\boldsymbol{c}_\lambda)_1 = 1/\sqrt{2}$. This is true for every point $\boldsymbol{c}$ in the set $Q \setminus V$. Therefore, the points $\boldsymbol{c}$ minimizing (74) need to be on the right edge $\{1/\sqrt{2}\boldsymbol{i} + c_2\boldsymbol{j} : |c_2| \leq 1/\sqrt{2}\}$ $= \{\boldsymbol{c}(\gamma) : \gamma \in [-1, 1]\}$ of the square Q; in the second equality, we have used the parametrization in (33). In conclusion,

$$c_{\rm inc}(\mathsf{A}, \mathsf{B}) = \min_{\gamma \in [-1,1]} D\left(\boldsymbol{c}(\gamma)\right). \tag{75}$$

Note that (75) is true also in the case $\alpha = 0$ (compatible A and B), for which we have $D(\boldsymbol{c}(1)) = 0$.

Now, for $\gamma \in [-1, 1]$, define M_γ as in (34). Then, M_γ has the form (68) with $\boldsymbol{c} = \boldsymbol{c}(\gamma)$. In particular, since γ, $c_1 = 1/\sqrt{2}$ and $c_2 = \gamma/\sqrt{2}$ satisfy (69), item (i) of Proposition 4 implies that M_γ is a POVM, and $\mathsf{M}_{\gamma\,[1]} = \mathsf{A}_{\boldsymbol{c}(\gamma)}$ and $\mathsf{M}_{\gamma\,[2]} = \mathsf{B}_{\boldsymbol{c}(\gamma)}$. Equation (36) then follows from the definition (7) of the error function. Moreover, by (36) and (73), we have $D\big(\mathsf{A}, \mathsf{B} \| \mathsf{M}_\gamma\big) = D(\boldsymbol{c}(\gamma))$, hence $\mathsf{M}_\gamma \in \mathcal{M}_{\rm inc}(\mathsf{A}, \mathsf{B})$ if and only if γ attains the minimum in (75).

In order complete the proof, it only remains to show that the minimization problem (75) is equivalent to (35). Indeed, for $\rho = (\mathbb{1} + \boldsymbol{v} \cdot \boldsymbol{\sigma})/2$ and $\rho' = (\mathbb{1} + \boldsymbol{v}' \cdot \boldsymbol{\sigma})/2$, with $\boldsymbol{v} = (v_1\,, v_2\,, v_3)$ and $\boldsymbol{v}' = (v_1\,, v_2\,, 0)$, we have $S[\mathsf{A}, \mathsf{B} \| \mathsf{M}_\gamma](\rho) = S[\mathsf{A}, \mathsf{B} \| \mathsf{M}_\gamma](\rho')$ by

(36). Therefore, by defining $\rho(\phi)$ as in (33),

$$\begin{aligned} D(\boldsymbol{c}(\gamma)) &= \max_{\substack{\rho\in\mathcal{S}(\mathcal{H})\\ \rho \text{ pure}}} S[\mathsf{A},\mathsf{B}\|\mathsf{M}_\gamma](\rho) && \text{by (36), (73)} \\ &= \max_{\phi\in[0,2\pi)} S[\mathsf{A},\mathsf{B}\|\mathsf{M}_\gamma](\rho(\phi)). \end{aligned}$$

By inserting this expression into (75), we get the desired equivalence. □

Remark 7. The last proof shows that the bi-observables M_γ given by (34) with γ yielding the minimum in (35) actually exhaust all D_2-covariant elements in $\mathcal{M}_{\rm inc}(\mathsf{A},\mathsf{B})$. Indeed, for the most general D_2-covariant bi-observable M parameterized with γ and $\boldsymbol{c}$ as in (68), we have $D\big(\mathsf{A},\mathsf{B}\|\mathsf{M}\big) = D(\boldsymbol{c}) > c_{\rm inc}(\mathsf{A},\mathsf{B})$ if $c_1 \neq 1/\sqrt{2}$, or, equivalently, $\mathsf{M} \neq \mathsf{M}_\gamma$. However, it is not clear whether any optimal bi-observables needs to be D_2-covariant, and, if this is the case, the minimum (35) is attained at a unique γ.

In the proof Theorem 6, we have made use of the following lemma, which will turn out useful also later.

Lemma 2. *Let $a \leq 0$, and suppose $D : [a,1] \to [0,+\infty]$ is a convex function such that $D(1) = 0$ and $D(0) < +\infty$. Then, D is nonincreasing on the interval $[a,0]$, and it is finite and strictly decreasing on $[0,1]$.*

Proof. For $a \leq x < y < 1$, the convexity of D implies

$$D(y) \leq \frac{1-y}{1-x}D(x) + \frac{y-x}{1-x}D(1) = \frac{1-y}{1-x}D(x). \tag{76}$$

In particular, $D(y) \leq D(x)$, and, choosing $x = 0$, $D(y) < +\infty$ for all $y \in (0,1)$. Then, another application of (76), now with $0 \leq x < y < 1$, yields $D(y) < D(x)$. Since the latter inequality implies $D(x) > 0$, for all $x \in [0,1)$, its extension to $y = 1$ is clear. □

B.2. The case of two orthogonal components. When the target observables are the orthogonal spin-1/2 components X and Y in (30), the symmetries of our system increase from D_2 to the enlarged dihedral group D_4. Here we recall that $D_4 \subset SO(3)$ is the order 8 group of the 90° rotations around the $\boldsymbol{k}$-axis, together with the 180° rotations around $\boldsymbol{i}$, $\boldsymbol{j}$, $\boldsymbol{n}$ and $\boldsymbol{m}$; clearly, $D_2 \subset D_4$. Now, the two rotations $S_{D_4} = \{R_{\boldsymbol{i}}(\pi), R_{\boldsymbol{n}}(\pi)\}$ generate D_4; for instance, we have $R_{\boldsymbol{j}}(\pi) = R_{\boldsymbol{n}}(\pi)R_{\boldsymbol{i}}(\pi)R_{\boldsymbol{n}}(\pi)$, $R_{\boldsymbol{m}}(\pi) = R_{\boldsymbol{i}}(\pi)R_{\boldsymbol{n}}(\pi)R_{\boldsymbol{i}}(\pi)$, $R_{\boldsymbol{k}}(\pi/2) = R_{\boldsymbol{m}}(\pi)R_{\boldsymbol{j}}(\pi)$.

The action of the group element $R_{\boldsymbol{n}}(\pi)$ on $\mathcal{X}\times\mathcal{Y}$, $\mathcal{H}$, $\mathsf{A} = \mathsf{X}$ and $\mathsf{B} = \mathsf{Y}$ is still given by (66) and (67); we have already seen that these actions satisfy condition (ii) of Theorem 4. Further, by introducing the natural actions

$$R_{\boldsymbol{i}}(\pi)\,(x,y) = (x,-y), \qquad U\big(R_{\boldsymbol{i}}(\pi)\big) = \mathrm{e}^{-\mathrm{i}\pi\,\boldsymbol{i}\cdot\boldsymbol{\sigma}/2} \equiv -\mathrm{i}\,\boldsymbol{i}\cdot\boldsymbol{\sigma}, \tag{77}$$

we have

$$U\big(R_{\boldsymbol{i}}(\pi)\big)\mathsf{X}(x)U\big(R_{\boldsymbol{i}}(\pi)\big)^* = \mathsf{X}(x), \qquad U\big(R_{\boldsymbol{i}}(\pi)\big)\mathsf{Y}(y)U\big(R_{\boldsymbol{i}}(\pi)\big)^* = \mathsf{Y}(-y).$$

In particular, we see that $R_{\boldsymbol{i}}(\pi)$ fulfills condition (i) of the same theorem. Therefore, all $g \in S_{D_4}$ satisfy the hypotheses of Theorem 4.

Again, in view of Remark 3, now we look for the general expression of a D_4-covariant bi-observable.

Proposition 5. *Let the dihedral group* D_4 *act on* $\mathcal{X} \times \mathcal{Y}$ *and* $\mathcal{H}$ *by* (66) *and* (77)*. Then, the most general* D_4*-covariant bi-observable on* $\mathcal{X} \times \mathcal{Y}$ *is given by* (68) *with* $\gamma = 0$*,* $c_2 = 0$ *and* $|c_1| \leq 1/\sqrt{2}$*, that is,*

$$\mathsf{M}(x,y) = \frac{1}{4}\left[\mathbb{1} + c_1\left(x\sigma_1 + y\sigma_2\right)\right], \qquad |c_1| \leq 1/\sqrt{2}. \tag{78}$$

Proof. By applying the extra transformation (77) to the D_2-covariant bi-observable (68) we get

$$\begin{aligned} R_{\boldsymbol{i}}(\pi)\mathsf{M}(x,y) &= U\big(R_{\boldsymbol{i}}(\pi)\big)\mathsf{M}(x,-y)U\big(R_{\boldsymbol{i}}(\pi)\big)^* \\ &= \frac{1}{4}\left[(1-\gamma xy)\,\mathbb{1} + (c_1x - c_2y)\,\sigma_1 - (c_2x - c_1y)\,\sigma_2\right]. \end{aligned}$$

In order to have covariance also under this transformation, it must be $\gamma = 0$ and $c_2 = 0$; then, condition (69) reduces to the inequality in (78). □

We are now ready to prove our main theorem for two orthogonal spin components.

Proof (of Theorem 5). By Theorem 4, there is at least one D_4-covariant bi-observable $\mathsf{M} \in \mathcal{M}_{\rm inc}(\mathsf{X},\mathsf{Y})$, which is necessarily of the form (78) by Proposition 5. Comparing it with (34), we see that they coincide if and only if $c_1 = 1/\sqrt{2}$ and $\gamma = 0$, and in this case both of them equal M_0 in (31). Thus, by Theorem 6, $\gamma = 0$ solves the minimization problem (35), and M_0 is the unique D_4-covariant element in $\mathcal{M}_{\rm inc}(\mathsf{X},\mathsf{Y})$. In particular, by (35) and (36) we have

$$\begin{aligned} c_{\rm inc}(\mathsf{X},\mathsf{Y}) &= \max_{\phi\in[0,2\pi)} S[\mathsf{A},\mathsf{B}\|\mathsf{M}_0](\rho(\phi)) \\ S[\mathsf{A},\mathsf{B}\|\mathsf{M}_0](\rho(\phi)) &= \tilde{s}(\cos\phi) + \tilde{s}(\sin\phi), \end{aligned} \tag{79}$$

where we have introduced the function

$$\tilde{s}(v) = \frac{1}{2}\sum_{k=\pm 1}(1+kv)\log\frac{1+kv}{1+kv/\sqrt{2}}, \qquad |v| \leq 1.$$

In (79), the best way to maximize $\tilde{s}(\cos\phi) + \tilde{s}(\sin\phi)$ is by means of a suitable integral representation. Namely, by direct inspection, we have

$$\tilde{s}(v) = \frac{1}{2\ln 2}\int_{\frac{1}{\sqrt{2}}}^{1}\frac{2v^2(1-\lambda)}{1-\lambda^2v^2}\,\mathrm{d}\lambda.$$

Then, by differentiation and simple computations, we get

$$\begin{aligned} f(\phi) &= \frac{\mathrm{d}}{\mathrm{d}\phi}\left(\tilde{s}(\cos\phi) + \tilde{s}(\sin\phi)\right) \\ &= -\frac{\sin(4\phi)}{2\ln 2}\int_{1/\sqrt{2}}^{1}\frac{\lambda^2(1-\lambda)(2-\lambda^2)}{\left(1-\lambda^2(\sin\phi)^2\right)^2\left(1-\lambda^2(\cos\phi)^2\right)^2}\,\mathrm{d}\lambda. \end{aligned}$$

The integrand is nonnegative for all $\lambda \in [1/\sqrt{2}, 1]$ and $\phi \in [0, 2\pi)$. We then see that $f(\phi) < 0$ for $0 < \phi < \pi/4$, $f(\pi/4) = 0$, $f(\phi) > 0$ for $\pi/4 < \phi < \pi/2$. So, for $\phi \in [0, \pi/2]$, the point $\phi = \pi/4$ gives a minimum of $\tilde{s}(\cos\phi) + \tilde{s}(\sin\phi)$, while we have two equal maxima at $\phi = 0$ and $\phi = \pi/2$; as $\tilde{s}$ is a continuous even function on

$[-1, 1]$, the maximum (79) is attained at $\phi = 0, \pi/2, \pi, 3\pi/2$. Such angles correspond to $\rho(\phi)$ being the eigenprojections of σ_1 or σ_2; this gives the first equality in (32). Then, in the last two ones, the numerical values follow by direct computation.

Finally, we still have to prove the uniqueness of M_0 in the set $\mathcal{M}_{\rm inc}(\mathsf{X},\mathsf{Y})$. Let M be any bi-observable in $\mathcal{M}_{\rm inc}(\mathsf{X},\mathsf{Y})$. By Corollary 1, its covariant version M_{D_4} is still in $\mathcal{M}_{\rm inc}(\mathsf{X},\mathsf{Y})$, and hence $\mathsf{M}_{D_4} = \mathsf{M}_0$ since M_0 is the unique D_4-covariant element of $\mathcal{M}_{\rm inc}(\mathsf{X},\mathsf{Y})$. Definition (22) implies $g\mathsf{M}(x,y) \leq |D_4|\mathsf{M}_{D_4}(x,y)$ for all g and x, y, hence in particular $\mathsf{M}(x,y) \leq |D_4|\mathsf{M}_{D_4}(x,y) = 8\mathsf{M}_0(x,y)$ for all x, y. Since $\mathsf{M}_0(x,y)$ has rank 1, it must then be $\mathsf{M}(x,y) = f(x,y)\mathsf{M}_0(x,y)$, $\forall x, y$, for some nonnegative coefficients $f(x,y)$. Writing f in the linear basis 1, x, y, xy of $\mathbb{C}^{\mathcal{X}\times\mathcal{Y}}$, the normalization constraint $\sum_{x,y}\mathsf{M}(x,y) = \sum_{x,y} f(x,y)\mathsf{M}_0(x,y) = \mathbb{1}$ gives $f(x,y) = 1 + \epsilon xy$ for some real parameter ϵ. For all x, y, we have the positivity constraint $\mathsf{M}(x,y) = f(x,y)\mathsf{M}_0(x,y) \geq 0$, which implies $f(x,y) \geq 0$; this gives $-1 \leq \epsilon \leq 1$.

Summing up, if $\mathsf{M} \in \mathcal{M}_{\rm inc}(\mathsf{X},\mathsf{Y})$, then $\mathsf{M}(x,y) = (1 + \epsilon xy)\mathsf{M}_0(x,y)$ for some $\epsilon \in [-1, 1]$. Let us show that the only possible parameter is $\epsilon = 0$. Indeed, the marginals of M are

$$\mathsf{M}_{[1]} = \mathsf{A}_{\boldsymbol{c}(\epsilon)}, \qquad \mathsf{M}_{[2]} = \mathsf{B}_{\boldsymbol{c}(\epsilon)}, \qquad \text{with} \qquad \boldsymbol{c}(\epsilon) = \frac{\boldsymbol{i} + \epsilon\boldsymbol{j}}{\sqrt{2}}.$$

Their distributions in the state $\rho_e = (\mathbb{1} + \sigma_1)/2$ are

$$\mathsf{M}_{[1]}^{\rho_e} = \frac{1}{\sqrt{2}}\,\delta_1 + \left(1 - \frac{1}{\sqrt{2}}\right)u_{\mathcal{X}}, \qquad \mathsf{M}_{[2]}^{\rho_e} = \frac{\epsilon}{\sqrt{2}}\,\delta_1 + \left(1 - \frac{\epsilon}{\sqrt{2}}\right)u_{\mathcal{Y}}.$$

On the other hand, we have $\mathsf{X}^{\rho_e} = \delta_1$ and $\mathsf{Y}^{\rho_e} = u_{\mathcal{Y}}$, so that

$$\begin{aligned} c_{\rm inc}(\mathsf{X},\mathsf{Y}) = D\big(\mathsf{X},\mathsf{Y}\|\mathsf{M}\big) \geq S[\mathsf{X},\mathsf{Y}\|\mathsf{M}](\rho_e) &= S\big(\mathsf{X}^{\rho_e}\|\mathsf{M}_{[1]}^{\rho_e}\big) + S\big(\mathsf{Y}^{\rho_e}\|\mathsf{M}_{[2]}^{\rho_e}\big) \\ &= \log\frac{2\sqrt{2}}{1+\sqrt{2}} + S\big(\mathsf{Y}^{\rho_e}\|\mathsf{M}_{[2]}^{\rho_e}\big) = c_{\rm inc}(\mathsf{X},\mathsf{Y}) + S\big(\mathsf{Y}^{\rho_e}\|\mathsf{M}_{[2]}^{\rho_e}\big), \end{aligned}$$

which implies $S\big(\mathsf{Y}^{\rho_e}\|\mathsf{M}_{[2]}^{\rho_e}\big) = 0$. Hence, $\mathsf{Y}^{\rho_e} = \mathsf{M}_{[2]}^{\rho_e}$, and $\epsilon = 0$ then follows. □

B.3. A lower bound for the incompatibility degree. In order to compute the lower bound (39), we have to minimize the following quantity over γ:

$$\begin{aligned} S[\mathsf{A},\mathsf{B}\|\mathsf{M}_\gamma](\rho_e) = \log\frac{2}{1+(a_1+a_2\gamma)/\sqrt{2}} &+ \frac{1+2a_1a_2}{2}\,\log\frac{1+2a_1a_2}{1+(a_1\gamma+a_2)/\sqrt{2}} \\ &+ \frac{1-2a_1a_2}{2}\,\log\frac{1-2a_1a_2}{1-(a_1\gamma+a_2)/\sqrt{2}}\,. \end{aligned} \tag{80}$$

By setting $\ell = (a_1\gamma + a_2)/\sqrt{2}$ and $f(\ell) = (\ln 2)S[\mathsf{A},\mathsf{B}\|\mathsf{M}_\gamma](\rho_e)$, we get

$$\begin{aligned} f(\ell) = \ln &\frac{2\sqrt{2}\,a_1}{\sqrt{2}\,a_1 + \sqrt{2}\,a_2\ell + a_1^2 - a_2^2} \\ &+ \frac{1}{2}\,(1+2a_1a_2)\ln\frac{1+2a_1a_2}{1+\ell} + \frac{1}{2}\,(1-2a_1a_2)\ln\frac{1-2a_1a_2}{1-\ell}\,, \end{aligned} \tag{81}$$

whose derivative is

$$f'(\ell) = -\frac{\sqrt{2}\,a_2}{\sqrt{2}\,a_1 + \sqrt{2}\,a_2\ell + a_1^2 - a_2^2} + \frac{\ell - 2a_1a_2}{1-\ell^2}\,.$$

Remark 8. For $\alpha = \pi/2$, i.e. $a_1 = 1$ and $a_2 = 0$, we immediately get that the expression (81) has a unique minimum at $\ell = 0$, which gives $\gamma = 0$ and the value (32) for the incompatibility degree.

For $\alpha \neq \pi/2$, the zeros of $f'(\ell)$ satisfy the algebraic equation $\ell^2 + u\ell/(\sqrt{2}\,a_2) - 1 - u = 0$, where u is defined in (42). By solving the algebraic equation and studying the sign of the derivative, we find that the minimum of (81) is at the point (41) and that the corresponding value of γ is (40). By using this result and $2a_1a_2 = \cos\alpha$, we get the lower bound (43).

B.4. Incompatibility degree and optimal measurements for three orthogonal spin-1/2 components. Here the target observables are X, Y and Z defined in (30) and (61). Their symmetry group is the order 24 octahedron group $O \subset SO(3)$, generated by the 90° rotations around the three coordinate axes: $S_O = \{R_{\boldsymbol{i}}(\pi/2),\ R_{\boldsymbol{j}}(\pi/2),\ R_{\boldsymbol{k}}(\pi/2)\}$. Note that for the dihedral groups introduced before we have $D_2 \subset D_4 \subset O$. Let us denote the three generators of O by $g_1 = R_{\boldsymbol{i}}(\pi/2)$, $g_2 = R_{\boldsymbol{j}}(\pi/2)$, $g_3 = R_{\boldsymbol{k}}(\pi/2)$. By using again the spin-1/2 projective representation of $SO(3)$, which we now restrict to O, we have the relations

$$\begin{aligned}
U_{g_1}\mathsf{X}(x)U_{g_1}^* &= \mathsf{X}(x), & U_{g_1}\mathsf{Y}(y)U_{g_1}^* &= \mathsf{Z}(y), & U_{g_1}\mathsf{Z}(z)U_{g_1}^* &= \mathsf{Y}(-z),\\
U_{g_2}\mathsf{X}(x)U_{g_2}^* &= \mathsf{Z}(-x), & U_{g_2}\mathsf{Y}(y)U_{g_2}^* &= \mathsf{Y}(y), & U_{g_2}\mathsf{Z}(z)U_{g_2}^* &= \mathsf{X}(z),\\
U_{g_3}\mathsf{X}(x)U_{g_3}^* &= \mathsf{Y}(x), & U_{g_3}\mathsf{Y}(y)U_{g_3}^* &= \mathsf{X}(-y), & U_{g_3}\mathsf{Z}(z)U_{g_3}^* &= \mathsf{Z}(z).
\end{aligned}$$

Moreover, the natural action of O on the outcome space $\mathcal{X} \times \mathcal{Y} \times \mathcal{Z} = \{+1, -1\}^3$ is

$$g_1\,(x,y,z) = (x,-z,y), \qquad g_2\,(x,y,z) = (z,y,-x), \qquad g_3\,(x,y,z) = (-y,x,z),$$

and the action on the index set is

$$g_i i = i, \quad g_1 2 = 3, \quad g_1 3 = 2, \quad g_2 1 = 3, \quad g_2 3 = 1, \quad g_3 1 = 2, \quad g_3 2 = 1.$$

Then, the hypotheses of Theorem 9 are satisfied by setting

$$f_{g_1,1}(x) = x, \quad f_{g_1,2}(y) = y, \quad f_{g_1,3}(z) = -z, \quad f_{g_2,1}(x) = -x,$$

$$f_{g_2,2}(y) = y, \quad f_{g_2,3}(z) = z, \quad f_{g_3,1}(x) = x, \quad f_{g_3,2}(y) = -y, \quad f_{g_3,3}(z) = z.$$

Therefore, we can apply Theorem 9 in order to prove the main result of Section 4.2.

Proof (of Theorem 10). By similar arguments as in the proofs of Propositions 4 and 5, one can prove that the most general O-covariant tri-observable in $\mathcal{M}(\mathcal{X} \times \mathcal{Y} \times \mathcal{Z})$ has the form

$$\mathsf{M}(x,y,z) = \frac{1}{8}\left[\mathbb{1} + c(x\sigma_1 + y\sigma_2 + z\sigma_3)\right] \qquad \text{with} \qquad |c| \leq \frac{1}{\sqrt{3}}\,. \tag{82}$$

Writing its marginals as

$$\mathsf{M}_{[1]} = \mathsf{X}_c, \qquad \mathsf{M}_{[2]} = \mathsf{Y}_c, \qquad \mathsf{M}_{[3]} = \mathsf{Z}_c,$$

we have

$$D\big(\mathsf{X},\mathsf{Y},\mathsf{Z}\|\mathsf{M}\big) = \max_{\substack{\rho\in\mathcal{S}(\mathcal{H})\\ \rho \text{ pure}}} \left[S\big(\mathsf{X}^\rho\|\mathsf{X}^\rho_c\big) + S\big(\mathsf{Y}^\rho\|\mathsf{Y}^\rho_c\big) + S\big(\mathsf{Z}^\rho\|\mathsf{Z}^\rho_c\big)\right]$$

for all $c \in [-1/\sqrt{3}, 1/\sqrt{3}]$. Denote by $D(c)$ the right hand side of the previous equation; then, the function D can be extended to all c's such that X_c, Y_c and Z_c define three POVMs on $\{-1,+1\}$. In particular, it is naturally defined also in the interval $(1/\sqrt{3}, 1]$, where X_c, Y_c and Z_c are the equally noisy versions of the sharp observables X, Y and Z (cf. (15)). We thus obtain a function $D : [-1/\sqrt{3}, 1] \to [0, +\infty]$. The mappings $c \mapsto \mathsf{X}^\rho_c$, $c \mapsto \mathsf{Y}^\rho_c$ and $c \mapsto \mathsf{Z}^\rho_c$ are affine on the interval $[-1/\sqrt{3}, 1]$, which, together with the convexity of the relative entropy, implies that such are the sum and the supremum in D. Moreover, $D(0) = D\big(\mathsf{X},\mathsf{Y},\mathsf{Z}\|\mathsf{U}_{\mathcal{X}\times\mathcal{Y}\times\mathcal{Z}}\big) < +\infty$ and $D(1) = 0$. Then, by Lemma 2, the divergence $D\big(\mathsf{X},\mathsf{Y},\mathsf{Z}\|\mathsf{M}\big)$, with M given by (82), attains its unique minimum when $c = 1/\sqrt{3}$; for such c, $\mathsf{M} = \mathsf{M}_0$ defined in (62). Since $\mathcal{M}_{\text{inc}}(\mathsf{X},\mathsf{Y},\mathsf{Z})$ contains at least one O-covariant tri-observable by Theorem 9, then M_0 is the unique O-covariant element in $\mathcal{M}_{\text{inc}}(\mathsf{X},\mathsf{Y},\mathsf{Z})$. The fact that also M_1 given by (63) is optimal follows since M_0 and M_1 have the same marginals (see (65)).

For the optimal approximate joint measurements M_0 and M_1, we have

$$\begin{aligned} c_{\text{inc}}(\mathsf{X},\mathsf{Y},\mathsf{Z}) = D\big(\mathsf{X},\mathsf{Y},\mathsf{Z}\|\mathsf{M}_i\big) &= \max_{\substack{\rho\in\mathcal{S}(\mathcal{H})\\ \rho \text{ pure}}} S[\mathsf{X},\mathsf{Y},\mathsf{Z}\|\mathsf{M}_i](\rho) \\ &= \max_{\substack{\rho\in\mathcal{S}(\mathcal{H})\\ \rho \text{ pure}}} \left[S\big(\mathsf{X}^\rho\|\mathsf{X}^\rho_{1/\sqrt{3}}\big) + S\big(\mathsf{Y}^\rho\|\mathsf{Y}^\rho_{1/\sqrt{3}}\big) + S\big(\mathsf{Z}^\rho\|\mathsf{Z}^\rho_{1/\sqrt{3}}\big)\right] \\ &= \max_{\substack{\phi\in[0,2\pi)\\ \theta\in[0,\pi)}} \left[\tilde{s}(\cos\phi\sin\theta) + \tilde{s}(\sin\phi\sin\theta) + \tilde{s}(\cos\theta)\right], \qquad (83) \end{aligned}$$

where we have used the parametrization $\rho = (\mathbb{1} + \cos\phi\sin\theta\,\sigma_1 + \sin\phi\sin\theta\,\sigma_2 + \cos\theta\,\sigma_3)/2$, inserted the marginals (65) of M_0, and introduced the function

$$\tilde{s}(v) = \frac{1}{2}\sum_{k=\pm1}(1+kv)\log\frac{1+kv}{1+kv/\sqrt{3}} = \frac{1}{2\ln 2}\int_{\frac{1}{\sqrt{3}}}^{1} \frac{2v^2(1-\lambda)}{1-\lambda^2v^2}\,\mathrm{d}\lambda, \qquad |v| \leq 1.$$

By using the integral representation of $\tilde{s}$,

$$\begin{aligned} &\frac{\partial}{\partial\phi}\big(\tilde{s}(\cos\phi\,\sin\theta) + \tilde{s}(\sin\phi\,\sin\theta) + \tilde{s}(\cos\theta)\big) \\ &\qquad = -\frac{\sin(4\phi)(\sin\theta)^4}{2\ln 2}\int_{1/\sqrt{3}}^{1} \frac{\lambda^2(1-\lambda)(2-\lambda^2)}{\left(1-\lambda^2v_1^2\right)^2\left(1-\lambda^2v_2^2\right)^2}\,\mathrm{d}\lambda; \end{aligned}$$

similar computations give the derivative with respect to θ. By the same arguments as in the case of two components, we obtain that in (83) the maximum is attained at all angles ϕ,θ corresponding to ρ being an eigenprojection of σ_1, σ_2 or σ_3. This fact and a final straightforward computation give (64). □

Remark 9. The last proof actually shows that M_0 given in (62) is the unique O-covariant optimal approximate joint measurement of X, Y and Z.

C. Symmetries and proofs for two Fourier conjugate MUBs

The natural symmetry group for the two Fourier conjugate observable Q and P of (45) is the group of the translations in the finite phase-space of the system, together with all its symplectic transformations; as usual, we identify the latter symplectic group with the group $SL(2,\mathbb{F})$ of the 2×2 matrices with entries in $\mathbb{F}$ and unit determinant. However, just a smaller subgroup of $SL(2,\mathbb{F})$ will be enough for us. Namely, for all $a\in\mathbb{F}_* = \mathbb{F}\setminus\{0\}$, we denote by $\mathsf{d}(a)$ and $\mathsf{f}(a)$ the $SL(2,\mathbb{F})$-matrices

$$\mathsf{d}(a) = \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix}, \qquad \mathsf{f}(a) = \begin{pmatrix} 0 & a \\ -a^{-1} & 0 \end{pmatrix}.$$

Then, the set $H = \{\mathsf{d}(a), \mathsf{f}(a) \mid a\in\mathbb{F}_*\}$ is an order $2(d-1)$ subgroup of the order $d(d^2-1)$ group $SL(2,\mathbb{F})$. It naturally acts by left multiplication on the additive abelian group $V=\mathbb{F}^2$ of the $\mathbb{F}$-valued 2-entries column vectors $\boldsymbol{u}=(u_1,u_2)^T$. We can then form the semidirect product group $G = H \ltimes V$, whose composition law is $(\mathsf{h},\boldsymbol{u})(\mathsf{k},\boldsymbol{v}) = (\mathsf{hk}, \mathsf{k}^{-1}\boldsymbol{u}+\boldsymbol{v})$.

The group G has a natural left action on the joint outcome space $\mathcal{X}\times\mathcal{Y}=\mathbb{F}^2$: by writing the points of $\mathcal{X}\times\mathcal{Y}=\mathbb{F}^2$ as columns, we have

$$(\mathsf{h},\boldsymbol{u})\begin{pmatrix} x \\ y \end{pmatrix} = \mathsf{h}\begin{pmatrix} x+u_1 \\ y+u_2 \end{pmatrix}. \tag{84}$$

In this context, the joint outcome space $\mathcal{X}\times\mathcal{Y}$ is called the *finite phase-space* of the system, and the subgroup $V\subset G$ is the group of its translations $(\mathsf{d}(1),\boldsymbol{u})$. The elements $(\mathsf{d}(a),\mathbf{0})\in H$ are diagonal symplectic transformations, while $(\mathsf{f}(1),\mathbf{0})$ just reverses the components x and y changing the sign of x (see e.g. [57] for more details on finite phase-spaces and their symmetries).

On the other hand, the group G has also a natural projective unitary representation on $\mathcal{H}$. In order to describe it, we first introduce the following unitary operators:

$$\begin{aligned} W(\boldsymbol{u})\phi(z) &= \mathrm{e}^{\frac{2\pi\mathrm{i}}{p}\,\mathrm{tr}\,u_2(z-u_1)}\phi(z-u_1), \qquad \forall\boldsymbol{u}\in\mathbb{F}^2, \\ D(a)\phi(z) &= \phi(a^{-1}z), \qquad \forall a\in\mathbb{F}_*=\mathbb{F}\setminus\{0\}. \end{aligned}$$

The operators $W(\boldsymbol{u})$ constitute the *Weyl operators* associated with the phase-space translations, and $D(a)$ are the *squeezing operators* by the nonzero scalars. Collected together with the Fourier transform F, they satisfy the composition rules

$$\begin{aligned} &W(\boldsymbol{u})W(\boldsymbol{v}) = \mathrm{e}^{\frac{2\pi\mathrm{i}}{p}\,\mathrm{tr}\,u_2v_1}W(\boldsymbol{u}+\boldsymbol{v}), && D(a)D(b)=D(ab), \\ &F^2 = D_{-1}, && FD(a)F^* = D(a^{-1}), \\ &D(a)W(\boldsymbol{u})D(a)^* = W(\mathsf{d}(a)\boldsymbol{u}), && FW(\boldsymbol{u})F^* = \mathrm{e}^{-\frac{2\pi\mathrm{i}}{p}\,\mathrm{tr}\,u_1u_2}W(\mathsf{f}(1)\boldsymbol{u}). \end{aligned}$$

Setting

$$U(\mathsf{d}(a),\boldsymbol{u}) = D(a)W(\boldsymbol{u}), \qquad U(\mathsf{f}(a),\boldsymbol{u}) = D(a)FW(\boldsymbol{u}),$$

we obtain a projective unitary representation of G on $\mathcal{H}$. It is easily checked that

$$\begin{aligned} U(\mathsf{d}(a),\boldsymbol{u})\mathsf{Q}(x)U(\mathsf{d}(a),\boldsymbol{u})^* &= \mathsf{Q}(a(x+u_1)), \\ U(\mathsf{d}(a),\boldsymbol{u})\mathsf{P}(y)U(\mathsf{d}(a),\boldsymbol{u})^* &= \mathsf{P}(a^{-1}(y+u_2)), \\ U(\mathsf{f}(a),\boldsymbol{u})\mathsf{Q}(x)U(\mathsf{f}(a),\boldsymbol{u})^* &= \mathsf{P}(-a^{-1}(x+u_1)), \\ U(\mathsf{f}(a),\boldsymbol{u})\mathsf{P}(y)U(\mathsf{f}(a),\boldsymbol{u})^* &= \mathsf{Q}(a(y+u_2)). \end{aligned} \tag{85}$$

The action (84) satisfies conditions (i.a) / (ii.a) of Theorem 4, with $S_G = G$. Moreover, by (85) the two sharp observables $\mathsf{A} = \mathsf{Q}$ and $\mathsf{B} = \mathsf{P}$ satisfy conditions (i.b) / (ii.b) of the same theorem. Therefore, by Corollary 1 we conclude that the set $\mathcal{M}_{\rm inc}(\mathsf{Q},\mathsf{P})$ contains a G-covariant element M_0.

Since in particular the bi-observable M_0 is covariant with respect to the group V of the phase-space translations, it must be of the form

$$\mathsf{M}_\tau(x,y) = \frac{1}{d}\, W((x,y)^T)\tau W((x,y)^T)^*, \qquad \forall x,y \in \mathbb{F}, \tag{86}$$

i.e. $\mathsf{M}_0 = \mathsf{M}_{\tau_0}$ for some state $\tau_0 \in \mathcal{S}(\mathcal{H})$ [42, Theor. 4.5.3]. According to [24, 25], we call an observable M_τ of the form (86) the *V-covariant phase-space observable generated by the state τ*. Since M_0 is also H-covariant and H is the stability subgroup of G at $(0,0)$, we see that $\tau_0 = d\,\mathsf{M}_0(0,0)$ can be any state commuting with the restriction $U|_H$ of the representation U to H.

By [64, Props. 1 and 2], the marginals of a V-covariant phase-space observable M_τ are

$$\mathsf{M}_{\tau\,[1]}(x) = \sum_{z\in\mathbb{F}} \mathsf{Q}^\tau(z-x)\mathsf{Q}(z), \qquad \mathsf{M}_{\tau\,[2]}(y) = \sum_{z\in\mathbb{F}} \mathsf{P}^\tau(z-y)\mathsf{P}(z). \tag{87}$$

Now, the fact that τ_0 commutes with $U|_H$ and the covariance relations (85) imply

$$\mathsf{Q}^{\tau_0}(x) = \mathrm{Tr}\,[\tau_0 U(\mathsf{f}(-1),\mathbf{0})\mathsf{Q}(x)U(\mathsf{f}(-1),\mathbf{0})^*] = \mathsf{P}^{\tau_0}(x), \qquad \forall x\in\mathbb{F},$$

$$\mathsf{Q}^{\tau_0}(x) = \mathrm{Tr}\,[\tau_0 U(\mathsf{d}(a),\mathbf{0})\mathsf{Q}(x)U(\mathsf{d}(a),\mathbf{0})^*] = \mathsf{Q}^{\tau_0}(ax), \qquad \forall x\in\mathbb{F},\ a\in\mathbb{F}_*.$$

By the second relation, the probability Q^{τ_0} is constant on the two subsets $\{0\}$ and $\mathbb{F}_*$ of $\mathbb{F}$, which are the orbits of the action of the multiplicative group $\mathbb{F}_*$ on $\mathbb{F}$. Therefore, we can write Q^{τ_0} as a linear combination of the two functions δ_0 and $u_\mathbb{F} - \delta_0/d$. The normalization of Q^{τ_0} requires

$$\mathsf{Q}^{\tau_0} = \lambda_0\delta_0 + (1-\lambda_0)u_\mathbb{F}$$

for some real λ_0. On the other hand, we must have $\lambda_0 \in [-1/(d-1)\,,\,1]$ by the positivity constraint. Equations (87) with $\tau = \tau_0$ then give

$$\mathsf{M}_{0\,[1]} = \lambda_0\mathsf{Q} + (1-\lambda_0)\mathsf{U}_\mathbb{F} =: \mathsf{Q}_{\lambda_0}\,, \qquad \mathsf{M}_{0\,[2]} = \lambda_0\mathsf{P} + (1-\lambda_0)\mathsf{U}_\mathbb{F} =: \mathsf{P}_{\lambda_0},$$

where $\mathsf{U}_\mathbb{F}$ is the trivial uniform observable on $\mathbb{F}$. If $\lambda_0 \geq 0$, then Q_{λ_0} and P_{λ_0} have the simple physical interpretation as uniformly noisy versions of Q and P with noise intensities $1-\lambda_0$, as it was explained in Section 2.4 (cf. (15)). However, we can not exclude that λ_0 takes its value in the negative interval $[-1/(d-1)\,,\,0)$, where this interpretation does not apply.

We finally come to the proof of our main result for two Fourier conjugate target observables.

Proof (of Theorem 7). For $\lambda \in [0,1]$, a straightforward extension of the argument in [59, Prop. 5] from the cyclic field $\mathbb{Z}_p$ to the finite field $\mathbb{F}$ yields that the minimal noise intensity making the two noisy observables Q_λ and P_λ compatible is

$$1-\lambda \geq 1-\lambda_* = \frac{\sqrt{d}}{2(\sqrt{d}+1)} \tag{88}$$

(see also Example 1 therein). Moreover, the same extension also proves that when in the previous bound the equality is attained, Q_{λ_*} and P_{λ_*} have a unique joint measurement in the whole set $\mathcal{M}(\mathcal{X}\times\mathcal{Y})$; it is the V-covariant phase-space observable M_{τ_*} generated by the pure state

$$\tau_* = \frac{\sqrt{d}}{2(1+\sqrt{d})}\,|\,\psi_{0,0}\,\rangle\,\langle\,\psi_{0,0}\,|\,,$$

with $\psi_{0,0}$ given in (51). As a consequence, for the two marginals Q_{λ_0} and P_{λ_0} of the optimal approximate joint measurement M_0, the inequalities $-1/(d-1)\leq\lambda_0\leq\lambda_*$ must hold. Note that the state τ_* commutes with $U|_H$, hence it is a valid candidate for generating the G-covariant phase-space observable M_0.

Now, by optimality of M_0 we have

$$c_{\rm inc}(\mathsf{Q},\mathsf{P}) = D\big(\mathsf{Q},\mathsf{P}\|\mathsf{M}_0\big) = \sup_\rho\left[S\big(\mathsf{Q}^\rho\|\mathsf{Q}^\rho_{\lambda_0}\big)+S\big(\mathsf{P}^\rho\|\mathsf{P}^\rho_{\lambda_0}\big)\right] =: D(\lambda_0).$$

The map $\lambda\mapsto D(\lambda)=\sup_\rho\left[S\big(\mathsf{Q}^\rho\|\mathsf{Q}^\rho_\lambda\big)+S\big(\mathsf{P}^\rho\|\mathsf{P}^\rho_\lambda\big)\right]$ is defined for all $\lambda\in\mathbb{R}$ such that Q_λ and P_λ are two POVMs. By affinity, these λ's form an interval I, which necessarily contains the subinterval $[0,1]$. On the interval I, the function D is nonnegative; moreover, the mappings $\lambda\mapsto\mathsf{Q}^\rho_\lambda$ and $\lambda\mapsto\mathsf{P}^\rho_\lambda$ are affine on I, which, together with the convexity of the relative entropy, implies that such are the sum and the supremum in D. Since $D(0)=D\big(\mathsf{Q},\mathsf{P}\|\mathsf{U}_{\mathcal{X}\times\mathcal{Y}}\big)<+\infty$ and $D(1)=0$, by Lemma 2 the function D is nonincreasing on I, and finite and strictly decreasing on $[0,1]$. This fact and inequality (88) for compatible Q_λ and P_λ then imply $\lambda_0=\lambda_*$. Moreover, the fact that M_{τ_*} is the unique joint observable of Q_{λ_*} and P_{λ_*} imposes $\tau_0=\tau_*$, that is $\mathsf{M}_0=\mathsf{M}_{\tau_*}$, which is (51). Therefore, M_{τ_*} is the unique G-covariant observable in $\mathcal{M}_{\rm inc}(\mathsf{Q},\mathsf{P})$, and

$$c_{\rm inc}(\mathsf{Q},\mathsf{P}) = D(\lambda_*) = \sup_\rho\left[S\big(\mathsf{Q}^\rho\|\mathsf{Q}^\rho_{\lambda_*}\big)+S\big(\mathsf{P}^\rho\|\mathsf{P}^\rho_{\lambda_*}\big)\right].$$

The first inequality in (49) then follows by evaluating the sum inside the sup at any eigenprojection $\rho=|\,\delta_x\,\rangle\,\langle\,\delta_x\,|$ of Q. On the other hand, the second inequality is the general bound for $c_{\rm inc}(\mathsf{Q},\mathsf{P})$ given in (17).

We finally prove the uniqueness of the optimal approximate joint measurement (51) in the case $p\neq 2$. If M is any observable in the optimal set $\mathcal{M}_{\rm inc}(\mathsf{Q},\mathsf{P})$, its covariant version M_G is still in $\mathcal{M}_{\rm inc}(\mathsf{Q},\mathsf{P})$ by Corollary 1, hence $\mathsf{M}_G=\mathsf{M}_{\tau_*}$ by the previous part. By (22), $\mathsf{M}(x,y)\leq|G|\mathsf{M}_G(x,y)=|G|\mathsf{M}_{\tau_*}(x,y)$ for all x,y. Since $\mathsf{M}_{\tau_*}(x,y)$ has rank 1, it must then be $\mathsf{M}(x,y)=f(x,y)\mathsf{M}_{\tau_*}(x,y)$ for some function $f:\mathbb{F}^2\to[0,|G|]$. The two normalization requirements $\sum_{x,y}\mathsf{M}_{\tau_*}(x,y)=\mathbb{1}$ and $\sum_{x,y}f(x,y)\mathsf{M}_{\tau_*}(x,y)=\sum_{x,y}\mathsf{M}(x,y)=\mathbb{1}$ impose constraints on the coefficients $f(x,y)$. If $d=p^n$ is odd, these constraints are enough to imply that $f(x,y)=1$ for all x,y. Indeed, this follows since in this case the observable M_{τ_*} is informationally complete. For $d=p$ odd, this is proved in [59, Prop. 9]. In the more general case $d=p^n$ odd, the same proof still holds, as it relies on the fact that the inverse Weyl transform of τ_*

$$\begin{aligned}\hat\tau_*(\boldsymbol{u}) &:= {\rm Tr}\,\{\tau_* W(\boldsymbol{u})\}\\ &= \frac{\sqrt{d}}{2(1+\sqrt{d})}\left[\delta_0(u_1)+\delta_0(u_2)+\frac{1}{\sqrt{d}}\left({\rm e}^{-\frac{2\pi i}{p}{\rm tr}\,u_1u_2}+1\right)\right]\end{aligned}$$

is nonzero for all $\boldsymbol{u}\in\mathbb{F}^2$ (see [65, Prop. 12]). The uniqueness statement is thus proved, and this concludes the proof of Theorem 7. □

Remark 10. 1. In the case $p = 2$, the above proof only shows that M_0 defined in (51) is the unique G-covariant observable in the set $\mathcal{M}_{\rm inc}(\mathsf{Q},\mathsf{P})$.
2. In the proof of Theorem 7, the dilational symmetries $\{\mathsf{d}(a) \mid a \in \mathbb{F}_*\}$ simplified the problem of characterizing the set $\mathcal{M}_{\rm inc}(\mathsf{Q},\mathsf{P})$, reducing it to the optimization of the single parameter λ.

References

1. Ozawa, M.: *Position measuring interactions and the Heisenberg uncertainty principle*, Phys. Lett. A **299** (2002) 1–7.
2. Ozawa, M.: *Physical content of Heisenberg's uncertainty relation: limitation and reformulation*, Phys. Lett. A **318** (2003) 21–29.
3. Ozawa, M.: *Universally valid reformulation of the Heisenberg uncertainty principle on noise and disturbance in measurement*, Phys. Rev. A **67** (2003) 042105.
4. Ozawa, M.: *Uncertainty relations for joint measurements of noncommuting observables*, Phys. Lett. A **320** (2004) 367–374.
5. Ozawa, M.: *Heisenberg's original derivation of the uncertainty principle and its universally valid reformulations*, Current Sci. **109** (2015) 2006–2016.
6. Werner, R.F.: *The uncertainty relation for joint measurement of position and momentum*, Quantum Inf. Comput. **4** (2004) 546–562.
7. Busch, P., Lahti, P., Werner, R.F.: *Measurement uncertainty relations*, J. Math. Phys. **55** (2014) 042111.
8. Busch, P., Lahti, P., Werner, R.F.: *Colloquium*: *Quantum root-mean-square error and measurement uncertainty relations*, Rev. Mod. Phys. **86** (2014) 1261–1281.
9. Busch, P., Lahti, P., Werner, R.F.: *Heisenberg uncertainty for qubit measurements*, Phys. Rev. A **89** (2014) 012129.
10. Werner, R.F.: *Uncertainty relations for general phase spaces*, Front. Phys. **11** (2016) 110305.
11. Busch, P., Heinonen, T., Lahti, P.: *Heisenberg's Uncertainty Principle*, Physics Reports **452** (2007) 155–176.
12. Dammeier, L., Schwonnek, R., Werner, R.F.: *Uncertainty relations for angular momentum*, New J. Phys. **17** (2015) 093046.
13. Abbott, A.A., Alzieu, P.-L. , Hall, M.J.W., Branciard, C.: *Tight state-independent uncertainty relations for qubits*, Mathematics, **4** (2016) 8.
14. Heisenberg, W.: *Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik*, Zeitschr. Phys. **43** (1927) 172–198.
15. Robertson, H.: *The uncertainty principle*, Phys. Rev. **34** (1929) 163–164.
16. Kraus, K.: *Complementary observables and uncertainty relations*, Phys. Rev. D **35** (1987) 3070–3075.
17. Maassen, H., Uffink, J.B.M.: *Generalized entropic uncertainty relations*, Phys. Rev. Lett. **60** (1988) 1103–1106.
18. Krishna, M., Parthasarathy, K.R.: *An Entropic Uncertainty Principle for Quantum Measurements*, Sankhya: Indian J. Stat. **64** (2002) 842–851.
19. Wehner, S., Winter, A.: *Entropic uncertainty relations — a survey*, New J. Phys. **12** (2010) 025009.
20. Kaniewski, J., Tomamichel, M., Wehner. S.: *Entropic uncertainty from effective anticommutators*, Phys. Rev. A **90** (2014) 012332.
21. Abdelkhalek, K., Schwonnek, R., Maassen, H., Furrer, F., Duhme, J., Raynal, P., Englert, B-G., Werner, R.F.: *Optimality of entropic uncertainty relations*, Int. J. Quantum Inf. **13** (2015) 1550045.
22. Coles, P.J., Berta, M., Tomamichel, M., Whener, S.: *Entropic uncertainty relations and their applications*, Rev. Mod. Phys. **89** (2017) 015002.
23. Holevo, A.S.: *Statistical Structure of Quantum Theory*, Lecture Notes in Physics **m 67** (Springer, Berlin, 2001).
24. Busch, P., Grabowski, M., Lahti, P.: *Operational Quantum Physics* (Springer, Berlin, 1997).
25. Busch, P., Lahti, P., Pellonpää, J.-P., Ylinen, K.: *Quantum Measurement* (Springer, Berlin, 2016).
26. Busch, P., Heinosaari, T.: *Approximate joint measurements of qubit observables*, Quantum Inf. Comp. **8** (2008) 797–818.
27. Heinosaari, T., Wolf, M.M.: *Nondisturbing quantum measurements*, J. Math. Phys. 51 (2010) 092201.
28. Heinosaari, T., Miyadera, T.: *Universality of sequential quantum measurements*, Phys. Rev. **91** (2015) 022110.
29. Appleby, D.M.: *Error principle*, Internat. J. Theoret. Phys. **37** (1998) 2557–2572.
30. Appleby, D.M.: *Quantum Errors and Disturbances: Response to Busch, Lahti and Werner*, Entropy **18** (2016) 174.

31. Buscemi, F., Hall, M.J.W., Ozawa, M., Wilde, M.M.: *Noise and disturbance in quantum measurements: an information-theoretic approach*, Phys. Rev. Lett. **112** (2014) 050401.
32. Abbot, A.A., Branciard, C.: *Noise and disturbance of Qubit measurements: An information-theoretic characterisation*, Phys. Rev. A **94** (2016) 062110.
33. Coles, P.J., Furrer, F.: *State-dependent approach to entropic measurement-disturbance relations*, Phys. Lett. A **379** (2015) 105–112.
34. Barchielli, A., Gregoratti, M., Toigo, A.: *Measurement uncertainty relations for position and momentum: Relative entropy formulation*, Entropy **19** (2017) 301.
35. Burnham, K.P., Anderson D.R.: *Model Selection and Multi-Model Inference*, 2nd edition (Springer, New York, 2002).
36. Cover, T.M., Thomas, J.A.: *Elements of Information Theory*, 2nd edition (Wiley, Hoboken, New Jersey, 2006).
37. Ohya, M., Petz, D.: *Quantum entropy and its use* (Springer, Berlin, 1993).
38. Barchielli, A., Lupieri, G.: *Instruments and channels in quantum information theory*, Optics and Spectroscopy **99** (2005) 425–432.
39. Barchielli, A., Lupieri, G.: *Quantum measurements and entropic bounds on information transmission*, Quantum Inf. Comput. **6** (2006) 16–45.
40. Barchielli, A., Lupieri, G.: *Instruments and mutual entropies in quantum information*, Banach Center Publ. **73** (2006) 65–80.
41. Maccone, L.: *Entropic information-disturbance tradeoff*, Europhys. Lett. **77** (2007) 40002.
42. Davies, E.B.: *Quantum Theory of Open Systems* (Academic, London, 1976).
43. Holevo, A.S.: *Quantum Systems, Channels, Information* (de Gruiter, Berlin, 2012).
44. Heinosaari, T., Ziman, M.: *The mathematical language of quantum theory: From uncertainty to entanglement* (Cambridge University Press, Cambridge, 2012).
45. Heinosaari, T., Miyadera, T., Ziman, M.: *An invitation to quantum incompatibility*, J. Phys. A: Math. Theor. **49** (2016) 123001.
46. Topsøe, F.: *Basic concepts, identities and inequalities — the toolkit of Information Theory*, Entropy **3** (2001) 162–190.
47. Pedersen, G.K.: *Analysis now* (Springer-Verlag, New York, 1989).
48. Busch, P., Heinosaari, T., Schultz, J., Stevens, N.: *Comparing the degrees of incompatibility inherent in probabilistic physical theories*, Europhys. Lett. **103** (2013) 10002.
49. Heinosaari, T., Schultz, J., Toigo, A., Ziman, M.: *Maximally incompatible quantum observables*, Phys. Lett. A **378** (2014) 1695–1699.
50. Keyl, M., Werner, R.F.: *Optimal cloning of pure states, testing single clones*, J. Math. Phys. **40** (1999) 3283–3299.
51. Werner, R.F.: *Optimal cloning of pure states*, Phys. Rev. A **58** (1998) 1827–1832.
52. P. Lahti, *Coexistence and Joint Measurability in Quantum Mechanics*, Int. J. Theor. Phys. **42** (2003) 893–906.
53. Wootters, W.K., Fields, D.B.: *Optimal state-determination by mutually unbiased measurements*, Ann. Physics **191** (1989) 363–381.
54. Durt, T., Englert, B.-G., Bengtsson, I., Zyczkowsky, K.: *On mutually unbiased bases*, Int. J. Quantum Inf. **8** (2010) 535–640.
55. Bandyopadhyay, S., Boykin, P.O., Roychowdhury, V., Vatan, F.: *A new proof for the existence of mutually unbiased bases*, Algorithmica **34** (2002) 512–528.
56. Appleby, D.M.: *Properties of the extended Clifford group with applications to SIC-POVMs and MUBs*, arXiv:0909.5233.
57. Carmeli, C., Schultz, J., Toigo, A.: *Covariant mutually unbiased bases*, Rev. Math. Phys. **28** (2016) 1650009.
58. Lang, S.: *Algebra*, 3rd edition, Graduate Texts in Mathematics, **211** (Springer, New York, 2002).
59. Carmeli, C., Heinosaari, T., Toigo, A.: *Informationally complete joint measurements on finite quantum systems*, Phys. Rev. A **85** (2012) 012109.
60. Heinosaari, T., Jivulescu, M.A., Reitzner, D., Ziman, M.: *Approximating incompatible von Neumann measurements simultaneously*, Phys. Rev. A **82** (2010) 032328.
61. Berta, M., Christandl, M., Colbeck, R., Renes, J.M., Renner, R.: *The uncertainty principle in the presence of quantum memory*, Nat. Phys. **6** (2010) 659.
62. Frank, R. L., Lieb, E.H.: *Extended Quantum Conditional Entropy and Quantum Uncertainty Inequalities*, Commun. Math. Phys. **323** (2013) 487–495.
63. Weyl, H.: *Symmetry* (Princeton University Press, Princeton, 1952).
64. Carmeli, C., Heinosaari, T., Toigo, A.: *Sequential measurements of conjugate observables*, J. Phys. A: Math. Theor. **44** (2011) 285304.
65. Carmeli, C., Heinosaari, T., Schultz, J., Toigo, A.: *Tasks and premises in quantum state determination*, J. Phys. A: Math. Theor. **47** (2014) 075302.