This is the published version of the publication, made available in accordance with the publisher's policy.

- -

Bounding and Simulating Contextual Correlations in Quantum Theory

- -

Tavakoli, Armin; Cruzeiro, Emmanuel Zambrini; Uola, Roope Kristian; Abbott, Alastair

Bounding and Simulating Contextual Correlations in Quantum Theory

Armin Tavakoli,[1,2,*] Emmanuel Zambrini Cruzeiro,[3] Roope Uola,[2] and Alastair A. Abbott[2,4]

[1] *Institute for Quantum Optics and Quantum Information – IQOQI Vienna, Austrian Academy of Sciences, Boltzmanngasse 3, Vienna 1090, Austria*

[2] *Département de Physique Appliquée, Université de Genève, Genève CH-1211, Switzerland*

[3] *Laboratoire d'Information Quantique, CP 225, Université libre de Bruxelles (ULB), Av. F. D. Roosevelt 50, Bruxelles 1050, Belgium*

[4] *Univ. Grenoble Alpes, Inria, Grenoble 38000, France*

We introduce a hierarchy of semidefinite relaxations of the set of quantum correlations in generalized contextuality scenarios. This constitutes a simple and versatile tool for bounding the magnitude of quantum contextuality. To illustrate its utility, we use it to determine the maximal quantum violation of several noncontextuality inequalities whose maximum violations were previously unknown. We then go further and use it to prove that certain preparation-contextual correlations cannot be explained with pure states, thereby showing that mixed states are an indispensable resource for contextuality. In the second part of the paper, we turn our attention to the simulation of preparation-contextual correlations in general operational theories. We introduce the information cost of simulating preparation contextuality, which quantifies the additional, otherwise forbidden, information required to simulate contextual correlations in either classical or quantum models. In both cases, we show that the simulation cost can be efficiently bounded using a variant of our hierarchy of semidefinite relaxations, and we calculate it exactly in the simplest contextuality scenario of parity-oblivious multiplexing.

I. INTRODUCTION

The contextuality of quantum theory is a fundamental sign of its nonclassicality that has been investigated for several decades. While contextuality was originally established as a property specific to the formalism of quantum theory [1,2], it has, in more recent times, been further generalized as a property of nonclassical probability distributions that can arise in operational theories [3]. This operational notion of contextuality is applicable to a broad range of physical scenarios and has been shown to be linked to a variety of foundational and applied topics in quantum theory (see, e.g., Refs. [4–12]).

The principle of noncontextuality holds that operationally equivalent physical procedures must correspond to identical descriptions in any underlying ontological model [3]. This assumption imposes constraints on the correlations that can be obtained in prepare-and-measure

scenarios involving operationally equivalent preparations and measurements. In such scenarios, which we term "contextuality scenarios," the correlations obtainable by noncontextual models can be characterized in terms of linear programming [13]. In contrast, quantum models that nonetheless respect the operational equivalences may produce "contextual correlations" unobtainable by any such noncontextual model [3]. This leads to a conceptually natural question: how can we determine if, for a given contextuality scenario, a given set of contextual correlations is compatible with quantum theory? This question is crucial for understanding the extent of nonclassicality manifested in quantum theory, and hence also for the development of quantum-information protocols powered by quantum contextuality. While an explicit quantum model is sufficient to prove compatibility with quantum theory, proving the converse—that no such model exists—is more challenging.

Here, we provide an answer to the question by introducing a hierarchy of semidefinite relaxations of the set of quantum correlations arising in contextuality scenarios involving arbitrary operational equivalences between preparations and measurements. This constitutes a sequence of increasingly precise necessary conditions that contextual correlations must satisfy in order to admit

*armin.tavakoli@oeaw.ac.at

of a quantum model. Thus, if a given contextual probability distribution fails one of the tests, it is incompatible with any quantum model satisfying the specified operational equivalences. We exemplify their practical utility by determining the maximal quantum violations of several different noncontextuality inequalities (for noisy state discrimination [14], for three-dimensional parity-oblivious multiplexing [15], for the communication task experimentally investigated in Ref. [16], and for the polytope inequalities obtained in Ref. [13]). Then, we apply our method to solve a foundational problem in quantum contextuality: we present a correlation inequality satisfied by all quantum models based on pure states and show that it can be violated by quantum strategies exploiting mixed states. Thus, we prove that mixed states are an indispensable resource for strong forms of quantum contextuality.

Equipped with the ability to bound the magnitude of quantum contextuality, we ask what additional resources are required to simulate preparation contextual correlations with classical or quantum models. We identify this resource as the preparation of states deviating from the required operational equivalences, and quantify this deviation in terms of the information extractable about the operational equivalences via measurement. This allows us to interpret preparation contextuality scenarios, and experiments aiming to simulate their results, as particular types of informationally restricted correlation experiments [17,18]. For both classical and quantum models, we show that the simulation cost can be lower bounded using variants of our hierarchy of semidefinite relaxations. We apply these concepts to the simplest preparation contextuality scenario [19], where we explicitly derive both the classical and quantum simulation costs of contextuality.

II. CONTEXTUALITY

Consider a prepare-and-measure experiment in which Alice receives an input $x \in [n_X] := \{1, \ldots, n_X\}$, prepares a system using the preparation P_x, and sends it to Bob. Bob receives an input $y \in [n_Y]$, performs a measurement M_y, and obtains an outcome $b \in [n_B]$; this event is called the *measurement effect* and is denoted $[b|M_y]$. When the experiment is repeated many times it gives rise to the conditional probability distribution $p(b|x, y) := p(b|P_x, M_y)$.

An ontological model provides a realist explanation of the observed correlations $p(b|x, y)$ [3]. In an ontological model, the preparation is associated to an ontic variable λ subject to some distribution (i.e., an epistemic state) $p(\lambda|x)$ and the measurement is represented by a probabilistic response function depending on the ontic state, $p(b|y, \lambda)$ [20]. The observed correlations are then written

$$p(b|x, y) = \sum_{\lambda} p(\lambda|x) p(b|y, \lambda). \tag{1}$$

Notice that every probability distribution admits of an ontological model.

A. Operational equivalences

The notion of noncontextuality becomes relevant when certain operational procedures (either preparations or measurements) are *operationally equivalent* [3]. Two preparations P and P' are said to be operationally equivalent, denoted $P \simeq P'$, if no measurement [21] can distinguish them, i.e.,

$$\forall [b|M]: \quad p(b|P, M) = p(b|P', M). \tag{2}$$

Similarly, two measurement effects $[b|M]$ and $[b'|M']$ are operationally equivalent, denoted $[b|M] \simeq [b'|M']$, if no preparation can distinguish them, i.e.,

$$\forall P: \quad p(b|P, M) = p(b'|P, M'). \tag{3}$$

In prepare-and-measure experiments, we are particularly interested in operationally equivalent procedures obtained by combining preparations P_x or measurement effects $[b|M_y]$. Specifically, one may have (hypothetical) preparations $P_\alpha = \sum_{x=1}^{n_X} \alpha_x P_x$ and $P_\beta = \sum_{x=1}^{n_X} \beta_x P_x$, where $\{\alpha_x\}_x$ and $\{\beta_x\}_x$ are convex weights (i.e., nonnegative and summing to one) and, likewise, measurement effects $[b_\alpha|M_\alpha] = \sum_{b=1}^{n_B} \sum_{y=1}^{n_Y} \alpha_{b|y}[b|M_y]$ and $[b_\beta|M_\beta] = \sum_{b=1}^{n_B} \sum_{y=1}^{n_Y} \beta_{b|y}[b|M_y]$, where $\{\alpha_{b|y}\}_{b,y}$ and $\{\beta_{b|y}\}_{b,y}$ are sets of convex weights, with $P_\alpha \simeq P_\beta$ and $[b_\alpha|M_\alpha] \simeq [b_\beta|M_\beta]$.

Such operationally equivalent procedures can naturally be grouped into equivalence classes, and it will be convenient for us to specify equivalent procedures in a slightly different, yet equivalent, way as follows.

Definition 1. (a) A *preparation operational equivalence* is a set $\mathcal{E}_\mathcal{P} = \{(S_k, \{\xi_k(x)\}_{x \in S_k})\}_{k=1}^K$, where $\{S_k\}_k$ is a partition of $[n_X]$ into K disjoint sets and, for each k, $\{\xi_k(x)\}_{x \in S_k}$ are convex weights [i.e., with $\xi_k(x) \geq 0$ and $\sum_{x \in S_k} \xi_k(x) = 1$]. We say that the preparations $\{P_x\}_{x=1}^{n_X}$ satisfy $\mathcal{E}_\mathcal{P}$ if for all $k, k' \in [K]$

$$\sum_{x \in S_k} \xi_k(x) P_x \simeq \sum_{x \in S_{k'}} \xi_{k'}(x) P_x. \tag{4}$$

(b) A *measurement operational equivalence* is a set $\mathcal{E}_\mathcal{M} = \{(T_\ell, \{\zeta_\ell(b, y)\}_{(b,y) \in T_\ell})\}_{\ell=1}^L$, where $\{T_\ell\}_\ell$ is a partition of $[n_B] \times [n_Y]$ into L disjoint sets and, for each ℓ, $\{\zeta_\ell(b, y)\}_{(b,y) \in T_\ell}$ are convex weights. We say the measurements $\{M_y\}_{y=1}^{n_Y}$ with effects $\{[b|M_y]\}_{b=1}^{n_B}$ satisfy $\mathcal{E}_\mathcal{M}$ if for all $\ell, \ell' \in [L]$

$$\sum_{(b,y) \in T_\ell} \zeta_\ell(b, y)[b|M_y] \simeq \sum_{(b,y) \in T_{\ell'}} \zeta_{\ell'}(b, y)[b|M_y]. \tag{5}$$

Note that any operational equivalence of the form $\sum_x \alpha_x P_x \simeq \sum_x \beta_x P_x$ or $\sum_{b,y} \alpha_{b|y}[b|M_y] \simeq \sum_{b,y} \beta_{b|y}$ $[b|M_y]$ can be specified in this way [22]. The formulation of Definition 1 allows us to consider natural partitions into $K \geq 2$ or $L \geq 2$ sets, which will prove useful later. For example, if one had three operationally equivalent preparations of the form $(1/2)(P_1 + P_2) \simeq (1/2)(P_3 + P_4) \simeq (1/2)(P_5 + P_6)$, we can express this as a single operational equivalence rather than several pairwise equivalences.

B. Contextuality scenarios and noncontextuality

With these basic notions, we can now more precisely define the kind of scenario in which we study noncontextuality and its precise definition in such settings. In particular, we consider prepare-and-measure scenarios of the form described above in which Alice's preparations and Bob's measurements must obey fixed sets of operational equivalences.

Definition 2. A *contextuality scenario* is a tuple $(n_X, n_Y, n_B, \{\mathcal{E}_\mathcal{P}^{(r)}\}_{r=1}^R, \{\mathcal{E}_\mathcal{M}^{(q)}\}_{q=1}^Q)$, where $\mathcal{E}_\mathcal{P}^{(r)}$ and $\mathcal{E}_\mathcal{M}^{(q)}$ are preparation and measurement operational equivalences, respectively.

Note that the normalization of the probability distribution $p(b|x, y)$ implies that $\sum_b [b|M_y] = \sum_b [b|M_{y'}]$ for all y, y', and hence every ontological model must satisfy the corresponding operational equivalence. We generally omit this trivial operational equivalence from the specification of a contextuality scenario.

The notion of (operational) noncontextuality formalizes the idea that operationally identical procedures must have identical representations in the underlying ontological model [3].

Definition 3. An ontological model is said to be the following.

(a) *Preparation noncontextual* if it assigns the same epistemic state to operationally equivalent preparation procedures; i.e., if the preparations P_x satisfy an operational equivalence $\mathcal{E}_\mathcal{P} = \{(S_k, \{\xi_k(x)\}_{x \in S_k})\}_{k=1}^K$ then, for all k, $k' \in [K]$

$$\forall \lambda : \sum_{x \in S_k} \xi_k(x) p(\lambda|x) = \sum_{x \in S_{k'}} \xi_{k'}(x) p(\lambda|x). \quad (6)$$

(b) *Measurement noncontextual* if it endows operationally equivalent measurement procedures with the same response function; i.e., if the measurement effects $[b|M_y]$ satisfy an operational equivalence $\mathcal{E}_\mathcal{M} = \{(T_\ell, \{\zeta_\ell(b, y)\}_{(b,y) \in T_\ell})\}_{\ell=1}^L$ then, for all $\ell, \ell' \in [L]$

$$\forall \lambda : \sum_{(b,y) \in T_\ell} \zeta_\ell(b, y) p(b|y, \lambda) = \sum_{(b,y) \in T_{\ell'}} \zeta_{\ell'}(b, y) p(b|y, \lambda). \quad (7)$$

Finally, if an ontological model is both preparation and measurement noncontextual, we simply say that it is *noncontextual*.

The assumption of noncontextuality imposes nontrivial constraints on the probability distributions that can arise in an ontological model [3].

Definition 4. Given a contextuality scenario, the correlations $p(b|x, y)$ are said to be (preparation or measurement) noncontextual if there exists a (preparation or measurement) noncontextual ontological model satisfying the operational equivalences of the scenario and reproducing the desired correlations. If no so much model exists, we say that the correlations are *(preparation or measurement) contextual*.

It is known that the set of noncontextual correlations (and, likewise, the sets of preparation or measurement noncontextual correlations) forms, for a given contextuality scenario, a convex polytope delimited by *noncontextuality inequalities* [13].

C. Quantum models

Here, we are particularly interested in what correlations can be obtained in contextuality scenarios within quantum mechanics. In quantum theory, a preparation P corresponds to a density matrix ρ [i.e., satisfying $\rho \succeq 0$ and $\text{tr}(\rho) = 1$], and two preparations ρ and ρ' are operationally equivalent if and only if $\rho = \rho'$. Preparation operational equivalences thus correspond to different decompositions of the same density matrix. Likewise, a measurement corresponds to a positive operator-valued measure (POVM) $\{E_b\}$ (defined by $E_b \succeq 0$ and $\sum_b E_b = \mathbb{1}$), where the E_b are the measurement effects. Measurement effects E_b and $E'_{b'}$ are thus operationally equivalent if and only if $E_b = E'_{b'}$.

We can thus specify precisely what a quantum model for a contextuality scenario corresponds to.

Definition 5. A *quantum model* for a contextuality scenario $(n_X, n_Y, n_B, \{\mathcal{E}_\mathcal{P}^{(r)}\}_{r=1}^R, \{\mathcal{E}_\mathcal{M}^{(q)}\}_{q=1}^Q)$ is given by two sets of Hermitian positive semidefinite operators $\{\rho_x\}_{x=1}^{n_X}$ and $\{\{E_{b|y}\}_{b=1}^{n_B}\}_{y=1}^{n_Y}$, which satisfy

$$\forall x : \text{tr}(\rho_x) = 1, \quad (8)$$

$$\forall y : \sum_{b=1}^{n_B} E_{b|y} = \mathbb{1}, \quad (9)$$

as well as the operational equivalences

$$\forall r, k : \sum_{x \in S_k^{(r)}} \xi_k^{(r)}(x) \rho_x = \sigma_r, \quad (10)$$

$$\forall q, \ell : \sum_{(b,y) \in T_\ell^{(q)}} \zeta_\ell^{(q)}(b,y) E_{b|y} = \tau_q, \qquad (11)$$

for some operators σ_r and τ_q independent of k and ℓ.

If a quantum model consists only of pure states (i.e., if $\rho_x^2 = \rho_x$ for all x) or projective measurements (i.e., if $E_{b|y}^2 = E_{b|y}$ and $E_{b|y}E_{b'|y} = 0$ for all b, b', y), then we call the model *pure* or *projective*, respectively.

It turns out that quantum theory is conceptually different from standard realist models, in the sense that there exist quantum models for contextuality scenarios—that thus respect the specified operational equivalences—but nevertheless can give rise to contextual correlations [3]. Quantum theory is thus said to be contextual.

Interestingly, quantum models cannot provide any advantage over noncontextual ontological models in the absence of nontrivial preparation operational equivalences [23]. In this sense, quantum theory is measurement non-contextual. Conversely, however, quantum contextuality can be witnessed in contextuality scenarios involving only operational equivalences between the preparations [along with the trivial measurement operational equivalence arising from Eq. (9), which is necessarily satisfied by any quantum model for any contextuality scenario]. For this reason there has been particular interest in preparation non-contextual inequalities, although interesting contextuality scenarios involving both preparation and measurement operational equivalences have been proposed (see, e.g., Refs. [8,13,24,25]).

III. A HIERARCHY OF SDP RELAXATIONS

In recent years, hierarchies of semidefinite programming (SDP) relaxations of the set of quantum correlations have become an invaluable tool in the study of quantum correlations [26,27]. Such a hierarchy capable of bounding contextual correlations in contextuality scenarios, where operational equivalences must be taken into account, has thus far, however, proved elusive, and it is this problem we address here.

The fundamental question we are interested in is the following: given a contextuality scenario $(n_X, n_Y, n_B, \{\mathcal{E}_{\mathcal{P}}^{(r)}\}, \{\mathcal{E}_{\mathcal{M}}^{(q)}\})$ and a probability distribution $p(b|x,y)$, does there exist a quantum model for the scenario reproducing the observed correlations, i.e., satisfying $p(b|x,y) = \text{tr}(\rho_x E_{b|y})$?

Note that, in contrast to many scenarios in quantum information, such as Bell nonlocality, it is not *a priori* clear that, in the search for such a quantum model, one can restrict oneself to pure states and projective measurements despite the fact that no assumption on the Hilbert-space dimension is made. Indeed, while one can always purify a mixed state, or perform a Naimark dilation of the POVMs,

such extensions may no longer satisfy the operational equivalences of the contextuality scenario.

Although SDP hierarchies have previously been formulated for prepare-and-measure scenarios [18,28–30], the main challenge for contextuality scenarios is to represent the constraints arising from the operational equivalences. Here, we adopt an approach motivated by a recent hierarchy [18] bounding informationally restricted correlations [17] and the fact that operational equivalences can be interpreted as restrictions on the information obtainable about equivalent operational procedures (see also Sec. V A).

A. Necessary conditions for a quantum model

Similarly to other related SDP hierarchies, our approach to formulate increasingly strict necessary conditions for the existence of a quantum model is based on reformulating the problem in terms of the underlying *moment matrix* of a quantum model. To this end, let us define the set of operator variables

$$J = \{\mathbb{1}\} \cup \{\rho_x\}_x \cup \{E_{b|y}\}_{b,y} \cup \{\sigma_r, \tau_\ell\}_{r,\ell}, \qquad (12)$$

where σ_r, τ_ℓ (with $r \in [R]$, $\ell \in [L]$) are variables corresponding to the operators defined in Eqs. (10) and (11) and are used to enforce robustly the operational equivalences. Consider a list $\mathcal{S} = (\mathcal{S}_1, \dots, \mathcal{S}_{|\mathcal{S}|})$ of monomials (of degree at least one) of variables in J. We say that $\mathcal{S}$ represents the kth degree of the hierarchy if it contains all monomials over J of degree at most k [31]. The choice of $\mathcal{S}$ will lead to different semidefinite relaxations, but it should at least include all elements of J.

Given a monomial list $\mathcal{S}$, the existence of a quantum model implies the existence of a *moment matrix* Γ whose elements, labeled by the monomials in $u, v \in \mathcal{S}$, are

$$\Gamma_{u,v} = \text{tr}(u^\dagger v) \qquad (13)$$

and satisfy a number of properties that form our necessary conditions. Some of these constraints are common with those found in similar hierarchies [points (I)–(III) below], while others capture important aspects of quantum models for contextuality scenarios [points (IV)–(V)] and will be expressed through localizing matrices [32]. We outline these constraints below.

(I) Hermitian-positive semidefiniteness. By construction the moment matrix is Hermitian and it is easily seen to be positive semidefinite [26], i.e.,

$$\Gamma = \Gamma^\dagger \succeq 0. \qquad (14)$$

(II) Consistency with p. Since the quantum model must reproduce the correlations $p(b|x,y)$, Γ must satisfy

$$\forall x, y, b : \Gamma_{\rho_x, E_{b|y}} = p(b|x,y). \qquad (15)$$

(III) Validity of states and measurements. Since any quantum model must satisfy the constraints of Eqs. (8)

and (9), Γ must satisfy

$$\forall x : \Gamma_{\mathbb{1},\rho_x} = 1, \qquad (16)$$

as well as linear identities of the form

$$\sum_{u,v} c_{u,v}\Gamma_{u,v} = 0 \quad \text{if} \quad \sum_{u,v} c_{u,v}\text{tr}(u^\dagger v) = 0, \qquad (17)$$

where the sum is over all monomials u, v in $\mathcal{S}$. These constraints are, in particular, those satisfied by any quantum model that follow from the validity of the states and measurements making up the model and the cyclicity of the trace.

For example, Eq. (17) includes constraints of the form $\sum_b \Gamma_{E_{b|y},E_{b'|y'}} = \Gamma_{\mathbb{1},E_{b'|y'}}$, as well as constraints such as $\Gamma_{E_{b|y},\rho_x E_{b'|y'}} = \Gamma_{\rho_x,E_{b'|y'}E_{b|y}}$, which follows from the fact that $\text{tr}(E_{b|y}\rho_x E_{b'|y'}) = \text{tr}(\rho_x E_{b'|y'}E_{b|y})$. It thus includes the constraints implied by the trivial operational equivalence following from Eq. (9) that are satisfied by any quantum model, thereby justifying the fact that we generally do not explicitly include this operational equivalence relation when specifying contextuality scenarios.

Note that if we were to assume the quantum model is either pure or projective (so that, respectively, either $\rho_x^2 = \rho_x$, or $E_{b|y}^2 = E_{b|y}$ and $E_{b|y}E_{b'|y} = 0$), then this implies further constraints of the form of Eq. (17). In particular, one can always make this assumption if there are no nontrivial operational equivalences of the corresponding type, allowing the SDP hierarchy we formulate to be simplified, but can also be considered as an additional assumption of interest (see Sec. IV B).

(IV) Operational equivalences. A quantum model must satisfy the operational equivalences of Eqs. (10) and (11). While this implies that the traces of each side of those equations must be equal—which in turn imposes the corresponding linear identities on the moment matrix—this alone does not fully capture the constraints implied by the operational equivalences, and notably is not enough to provide a good hierarchy. To properly enforce these constraints, we draw inspiration from the hierarchy of informationally restricted quantum correlations [18] and make use of *localizing matrices*. These are additional matrices of moments whose elements (or a subset thereof) are linear combinations of elements of Γ, and which themselves must be positive semidefinite [32].

We thus define, for all $r \in [R]$, $k \in [K^{(r)}]$ and all $q \in [Q]$, $\ell \in [L^{(q)}]$, the localizing matrices $\tilde{\Lambda}^{(r,k)}$ and $\hat{\Lambda}^{(q,\ell)}$ with elements

$$\tilde{\Lambda}_{u,v}^{(r,k)} = \text{tr}\left[u^\dagger\left(\sigma_r - \sum_{x\in S_k^{(r)}}\xi_k^{(r)}(x)\rho_x\right)v\right] \qquad (18)$$

$$\hat{\Lambda}_{u,v}^{(q,\ell)} = \text{tr}\left[u^\dagger\left(\tau_q - \sum_{(b,y)\in T_\ell^{(q)}}\zeta_\ell^{(q)}(b,y)E_{b|y}\right)v\right], \qquad (19)$$

which are labeled now by monomials from a monomial list $\mathcal{L}$, in general different from $\mathcal{S}$ (and which, in principal, could differ for each localizing matrix). Ideally, $\mathcal{L}$ should be chosen so that the elements of the localizing matrices are linear combinations of elements of the moment matrix Γ.

For a quantum model exactly satisfying the operational equivalences $\mathcal{E}_{\mathcal{P}}^{(r)}$ and $\mathcal{E}_{\mathcal{M}}^{(q)}$, with σ_r and τ_q defined as in Eqs. (10) and (11) one has $\tilde{\Lambda}^{(r,k)} = \hat{\Lambda}^{(q,\ell)} = 0$, $\text{tr}(\sigma_r) = 1$ and $\text{tr}(\tau_q) = \sum_{(b,y)\in T_\ell^{(q)}}\zeta_\ell^{(q)}(b,y)\text{tr}(E_{b|y})$. Such complicated matrix equality constraints (which one could in principle enforce without defining the localizing matrices), however, tend to lead to poor results in practice due to the numerical instability of SDP solvers. Instead, we impose the more robust constraints that $\tilde{\Lambda}^{(r,k)}, \hat{\Lambda}^{(q,\ell)} \succeq 0$ (along with the equality constraints on the traces of σ_r, τ_q, which serve to "normalize" the localizing matrices), which follow from the existence, for any quantum model, of Hermitian operators σ_r, τ_q satisfying $\sigma_r \succeq \sum_{x\in S_k^{(r)}}\xi_k^{(r)}(x)\rho_x$ and $\tau_q \succeq \sum_{(b,y)\in T_\ell^{(q)}}\zeta_\ell^{(q)}(b,y)E_{b|y}$. We thus have, for all r, k, q, ℓ

$$\tilde{\Lambda}^{(r,k)} \succeq 0, \quad \hat{\Lambda}^{(q,\ell)} \succeq 0, \qquad (20)$$

$$\Gamma_{\mathbb{1},\sigma_r} = 1, \quad \Gamma_{\mathbb{1},\tau_q} = \sum_{(b,y)\in T_\ell^{(q)}}\zeta_\ell^{(q)}(b,y)\Gamma_{\mathbb{1},E_{b|y}}. \qquad (21)$$

Moreover, whenever the monomials $u, \sigma_r u,$ and $\rho_x u$ are in $\mathcal{S}$ we have

$$\tilde{\Lambda}_{u,v}^{(r,k)} = \Gamma_{u,\sigma_r v} - \sum_{x\in S_k^{(r)}}\xi_k^{(r)}\Gamma_{u,\rho_x v}, \qquad (22)$$

and, when $u, \tau_q u,$ and $E_{b|y}u$ are similarly in $\mathcal{S}$,

$$\hat{\Lambda}_{u,v}^{(q,\ell)} = \Gamma_{u,\tau_q v} - \sum_{(b,y)\in T_\ell^{(q)}}\zeta_\ell^{(q)}\Gamma_{u,E_{b|y}v}, \qquad (23)$$

thereby relating the localizing matrices to the moment matrix Γ.

We note that the operators σ_r and τ_q, and the localizing matrices expressing the deviation of their moments from those of the operational equivalences, hence play the role of slack variables to robustly enforce the operational equivalencies. As we see in Sec. V, the formulation we adopt here will also allow a natural generalization allowing us to study the simulation cost of preparation contextuality, where the trace of σ_r has a natural interpretation, further

motivating our choice to present the constraints in the form given here.

(V) Positivity of states and measurements. In most SDP hierarchies used in quantum information, one can assume without loss of generality that the states and measurements in question are projective (see, e.g., Refs. [26, 27]); since all projective operators are positive semidefinite, it is not necessary in such cases to consider explicitly the constraints the positive semidefiniteness of the operators in a quantum model imposes on a moment matrix. As already mentioned, however, for contextuality scenarios this is not *a priori* the case, and to capture the constraints implied by the positive semidefiniteness of states and measurements (i.e., $\rho_x, E_{b|y} \succeq 0$) we again exploit localizing matrices.

Let us thus introduce the localizing matrices (for all x, y, b) $\tilde{\Upsilon}^x$ and $\hat{\Upsilon}^{(b,y)}$ with elements

$$\tilde{\Upsilon}^x_{u,v} = \text{tr}(u^\dagger \rho_x v), \tag{24}$$

$$\hat{\Upsilon}^{(b,y)}_{u,v} = \text{tr}(u^\dagger E_{b|y} v), \tag{25}$$

which are labeled by monomials from a monomial list $\mathcal{O}$, in general different from $\mathcal{S}$ (and which, as for $\mathcal{L}$, in principle could differ for each x, y, b). Ideally, $\mathcal{O}$ should be chosen so that the elements of the localizing matrices are also elements of the moment matrix Γ.

It is easily seen that the positive semidefiniteness of ρ_x and $E_{b|y}$ implies

$$\forall x : \quad \tilde{\Upsilon}^x \succeq 0, \tag{26}$$

$$\forall y, b : \quad \hat{\Upsilon}^{(b,y)} \succeq 0, \tag{27}$$

which in turn (for well chosen $\mathcal{O}$) constrains Γ. Moreover, for all u, v in $\mathcal{O}$, whenever the monomials $u, \rho_x v$ or, respectively, $u, E_{b|y}v$ are in $\mathcal{S}$ we have

$$\tilde{\Upsilon}^x_{u,v} = \Gamma_{u,\rho_x v}, \quad \hat{\Upsilon}^{(b,y)}_{u,v} = \Gamma_{u,E_{b|y}v}, \tag{28}$$

thereby relating the localizing matrices to the main moment matrix.

For given choices of the moment lists $\mathcal{S}$, $\mathcal{L}$, and $\mathcal{O}$, the constraints presented above thus provide necessary conditions for a given correlation to have a quantum realization in the contextuality scenario. Note moreover that, by standard arguments [26], one can actually assume the moment matrix (and localizing matrices) are real since the above constraints involve only real coefficients. These conditions are all semidefinite constraints, which leads us to the following proposition summarizing our hierarchy of SDP relaxations.

Proposition 6. *Let $\mathcal{S}, \mathcal{L}, \mathcal{O}$ be fixed lists of monomials from J. A necessary condition for the existence of a*

quantum model in a given contextuality scenario reproducing the correlations $\{p(b|x,y)\}_{b,x,y}$ is the feasibility of the following SDP:

$$\text{find} \quad \Gamma, \{\tilde{\Lambda}^{(r,k)}\}_{r,k}, \{\hat{\Lambda}^{(q,\ell)}\}_{q,\ell}, \{\tilde{\Upsilon}^x\}_x, \{\hat{\Upsilon}^{(b,y)}\}_{b,y},$$

$$\text{such that} \quad \Gamma \succeq 0, \quad \tilde{\Lambda}^{(r,k)} \succeq 0, \quad \hat{\Lambda}^{(q,\ell)} \succeq 0,$$

$$\tilde{\Upsilon}^x \succeq 0, \quad \hat{\Upsilon}^{(b,y)} \succeq 0, \tag{29a}$$

$$\Gamma_{\rho_x, E_{b|y}} = p(b|x,y), \tag{29b}$$

$$\Gamma_{\mathbb{1},\rho_x} = 1, \tag{29c}$$

$$\sum_{u,v} c_{u,v}\Gamma_{u,v} = 0 \quad \text{if} \quad \sum_{u,v} c_{u,v}\text{tr}(u^\dagger v) = 0, \tag{29d}$$

$$\Gamma_{\mathbb{1},\sigma_r} = 1, \quad \Gamma_{\mathbb{1},\tau_q} = \sum_{(b,y)\in T_\ell^{(q)}} \zeta_\ell^{(q)}(b,y)\Gamma_{\mathbb{1},E_{b|y}}, \tag{29e}$$

$$\tilde{\Lambda}^{(r,k)}_{u,v} = \Gamma_{u,\sigma_r v} - \sum_{x\in S_k^{(r)}} \xi_k^{(r)}\Gamma_{u,\rho_x v}, \tag{29f}$$

$$\hat{\Lambda}^{(q,\ell)}_{u,v} = \Gamma_{u,\tau_q v} - \sum_{(b,y)\in T_\ell^{(q)}} \zeta_\ell^{(q)}\Gamma_{u,E_{b|y}v}, \tag{29g}$$

$$\tilde{\Upsilon}^x_{u,v} = \Gamma_{u,\rho_x v}, \quad \hat{\Upsilon}^{(b,y)}_{u,v} = \Gamma_{u,E_{b|y}v}, \tag{29h}$$

where the above operators are all symmetric real matrices.

By taking increasingly long monomials lists $\mathcal{S}$, $\mathcal{L}$, and $\mathcal{O}$, one thus obtains increasingly strong necessary conditions for a quantum realization, and which can be efficiently checked by standard numerical solvers for SDPs.

While the above hierarchy applies to arbitrary contextuality scenarios, in many scenarios or situations of interest, it can be somewhat simplified. In particular, if one wishes to determine whether a given correlation is compatible with a pure and/or projective quantum model, the extra constraints imposed on the states and measurement effects (cf. Definition 5) correspond to further linear constraints in Eq. (29d), meaning that the corresponding localizing matrices $\tilde{\Upsilon}^x$ and/or $\hat{\Upsilon}^{(b,y)}$ [and subsequent constraints in Eq. (29h)] are not required. Similarly, if there are either no preparation or no measurement operational equivalences present in the problem (i.e., if $R = 0$ or $Q = 0$) then the corresponding localizing matrices $\tilde{\Lambda}^{(r,k)}$ or $\hat{\Lambda}^{(q,\ell)}$ [and subsequent constraints in Eqs. (29e) and (29f)] are also not required. The later case is particularly relevant in many (preparation) contextuality scenarios of interest, including the examples we consider in the following section. To illustrate this, in Appendix A we show how the SDP simplifies for the case of preparation noncontextuality, where only nontrivial preparation operational equivalences are considered.

Although the above hierarchy solves a feasibility problem, asking whether a distribution $p(b|x,y)$ is compatible with a quantum model for the contextuality scenario, in practice one is often interested with maximizing a linear functional of the probability distribution over all possible quantum models—i.e., a *noncontextuality inequality*—perhaps subject to some further constraints on the distribution. It is easily seen that, following standard techniques, the hierarchy of necessary conditions we present also allows one to bound such optimization problems by instead maximizing the corresponding functional over all feasible solutions to the SDP of Proposition 6.

As we see in the following section, the hierarchy of Proposition 6 allows us to readily obtain tight bounds on quantum contextual correlations in many scenarios of interest. However, in some cases involving both nontrivial preparation and measurement operational equivalences and no assumptions of pure states or projective measurements, it performs relatively poorly in practice. This appears to stem from the fact that, in such cases, the probabilities $p(b|x,y)$ do not appear on the diagonal of the moment matrix or any of the localizing matrices. In Appendix B we show how these difficulties can be overcome by presenting a modified version of our hierarchy, obtained by taking the operators $\{\sqrt{\rho_x}\}_x$ and/or $\{\sqrt{E_{b|y}}\}_{b,y}$ in the operator set J [cf. Eq. (12)] instead of $\{\rho_x\}_x$ and $\{E_{b|y}\}_{b,y}$, an approach that we believe may be of independent technical interest.

IV. APPLICATIONS OF THE SDP HIERARCHY

We implement a version of this hierarchy (and the variant described in Appendix B) in MATLAB, exploiting the SDP interface YALMIP [33], and our code is freely available [34]. Our implementation can handle arbitrary contextuality scenarios, restrictions to pure or projective quantum models or to classical (commuting) models, and solve either the feasibility SDP of Proposition 6 or maximize a linear functional of the correlations $p(b|x,y)$ subject to linear constraints on the probabilities. In solving large SDP problems that would otherwise be numerically intractable, it can make use of RepLAB [35,36] (a recently developed tool for manipulating finite groups with an emphasis on SDP applications) to exploit symmetries in noncontextuality inequalities, a capability we exploit in obtaining some of the results presented below.

A. Quantum violations of established preparation noncontextuality inequalities

To illustrate the usefulness of the hierarchy described in Proposition 6, we first exploit it to derive tight bounds on the maximal quantum violation of three preparation noncontextuality inequalities introduced in the previous literature. In Appendix C we detail the analysis of two examples based on the inequalities derived in Ref. [15]

and the inequalities experimentally explored in Ref. [16]. Here, we focus on the noncontextuality inequalities for state discrimination presented in Ref. [14].

To reveal a contextual advantage in state discrimination, Ref. [14] considers a scenario with $x \in [4]$, $y \in [3]$, and $b \in [2]$ and attempts to discriminate the preparations P_1 and P_2, while P_3 and P_4 are symmetric extensions that ensure the operational equivalence $(1/2)P_1 + (1/2)P_3 \simeq (1/2)P_2 + (1/2)P_4$. The first two measurements ($y = 1, 2$) correspond to distinguishing preparations P_1 and P_3, and P_2 and P_4, respectively (in the noiseless case, these should be perfectly discriminable); while the third ($y = 3$) corresponds to the state discrimination task, i.e., discriminating P_1 and P_2. There are three parameters of interest: the probability of a correct discrimination, s; the probability of confusing the two states, c; and the noise parameter, ϵ. Under the symmetry ansatz considered, the observed statistics are thus required to satisfy

$$
\begin{aligned}
s &= p(1|1,3) = p(2|2,3) = p(2|3,3) = p(1|4,3), \\
c &= p(1|2,1) = p(1|1,2) = p(2|4,1) = p(2|3,2), \\
1 - \epsilon &= p(1|2,2) = p(1|1,1) = p(2|4,2) = p(2|3,1).
\end{aligned}
\tag{30}
$$

The authors show that, for $\epsilon \leq c \leq 1 - \epsilon$, the following noncontextuality inequality holds:

$$
s \leq 1 - \frac{c - \epsilon}{2}.
\tag{31}
$$

What is the maximal quantum advantage in the task? Ref. [14] presented a specific family of quantum models that achieve

$$
s = \frac{1}{2}\left(1 + \sqrt{1 - \epsilon + 2\sqrt{\epsilon(1-\epsilon)c(1-c)} + c(2\epsilon - 1)}\right),
\tag{32}
$$

which violates the bound, Eq. (31), and conjectured it to be optimal for qubit systems. The semidefinite programming hierarchy presented in the previous section allows us to place upper bounds on s for given values of (c, ϵ) by maximizing s under the above constraints. Using a moment matrix of size 42 and localizing matrices of size 7 [37], we systematically perform this maximization with a standard numerical SDP solver [38] for different values of (c, ϵ) by dividing the space of valid parameters (i.e., satisfying $\epsilon \leq c \leq 1 - \epsilon$) into a grid with spacing of 0.01. We consistently obtain in every case an upper bound agreeing with the value in Eq. (32) to within 10^{-5}, which is consistent with the precision of the SDP solver. We thus find that Eq. (32) indeed gives the maximal quantum contextual advantage in state discrimination.

For the interested reader, in Appendix D we use this example to show more explicitly what form the constraints

of the SDP hierarchy take and how they relate the moment matrix and localizing matrices.

B. Mixed states as resources for quantum contextuality

In many forms of nonclassicality, such as Bell nonlocality, steering, and quantum dimension witnessing, the strongest quantum correlations are necessarily obtained with pure states. In the former two, this stems from the fact that any mixed state can be purified in a larger Hilbert space. In the latter, it follows from the possibility to realize a mixed state as a convex combination of pure states of the same dimension. Interestingly, however, it is *a priori* unclear whether mixed states should play a more fundamental role in quantum contextuality: both purifications of mixed states and postselections on pure-state components of mixed states may break the operational equivalences between preparation in contextuality scenarios. Here we show that this intuition turns out to be correct: preparation contextuality indeed is exceptional as mixed states are needed to obtain some contextual quantum correlations.

To prove this, we consider the noncontextuality scenario of Hameedi-Tavakoli-Marques-Bourennane (HTMB) [16]. In this scenario, Alice receives two trits, $x := x_1 x_2 \in \{0, 1, 2\}^2$ and Bob receives a bit $y \in [2]$ and produces a ternary outcome $b \in \{0, 1, 2\}$. There are two operational equivalences involved, corresponding to Alice sending zero information about the value of the sums $x_1 + x_2$ and $x_1 + 2x_2$ (modulo 3), respectively. Each of these corresponds to a partition of Alice's nine preparations into three sets. Under these constraints, Alice and Bob evaluate a random access code [39]. The HTMB inequality bounds the success probability of the task in a noncontextual model [16]:

$$\mathcal{A}_{\text{HTMB}} := \frac{1}{18} \sum_{x,y} p(b = x_y | x, y) \leq \frac{2}{3}. \quad (33)$$

We revisit this scenario and employ our semidefinite relaxations to determine a bound on the largest value of $\mathcal{A}_{\text{HTMB}}$ attainable in a quantum model in which all nine preparations are pure. As described following Proposition 6, this scenario can easily be considered with our hierarchy by simply including the linear constraints following from $\rho_x^2 = \rho_x$ (for all x) in Eq. (29d) and noting that the localizing matrices $\tilde{\Upsilon}^x$ are no longer required. Using a moment matrix of size 2172 and localizing matrices of size 187, we find that $\mathcal{A}_{\text{HTMB}} \lesssim 0.667$ up to solver precision. To make such a large SDP problem numerically tractable, we use RepLAB [35,36] to make the moment matrix invariant under the symmetries of the random access code, thereby significantly reducing the number of variables in the SDP problem. This gives us strong evidence (i.e., up to numerical precision) that pure states cannot violate the HTMB inequality (33), and we conjecture this to indeed

be the case exactly [40]. Importantly, however, mixed states are known to enable a violation of the inequality: six-dimensional quantum systems can achieve $\mathcal{A}_{\text{HTMB}} \approx 0.698$ [16,41]. This shows that sufficiently strong contextual quantum correlations can require the use of mixed states.

C. Quantum violation of contextuality inequalities involving nontrivial measurement operational equivalences

The examples discussed above focused on preparation contextuality scenarios, in which there are no nontrivial measurement operational equivalences. Nonetheless, quantum contextuality can also be observed in scenarios involving measurement operational equivalences (in addition to preparation operational equivalences), and we demonstrate the ability of our hierarchy to provide tight bounds in such scenarios by applying to the noncontextuality inequalities derived in Ref. [13].

In Ref. [13], the authors consider a scenario with $x \in [6]$, $y \in [3]$, and $b \in [2]$ where the preparations satisfy the operational equivalence $(1/2)(P_1 + P_2) \simeq (1/2)(P_3 + P_4) \simeq (1/2)(P_5 + P_6)$ and the measurements satisfy the operational equivalence $(1/3)\sum_y [1|M_y] \simeq (1/3)\sum_y [2|M_y]$. The authors completely characterized the polytope of noncontextual correlations in this contextuality scenario, finding the following six inequivalent (under symmetries), nontrivial noncontextuality "facet" inequalities [where we use the notation $p_{xy} := p(1|x, y)$]:

$$I_1 = p_{11} + p_{32} + p_{53} \leq 2.5, \quad (34a)$$

$$I_2 = p_{11} + p_{22} + p_{53} \leq 2.5, \quad (34b)$$

$$I_3 = p_{11} - 2p_{22} + 2p_{32} - p_{41} - 2p_{51} + 2p_{53} \leq 3, \quad (34c)$$

$$I_4 = 2p_{11} - p_{22} + 2p_{32} \leq 3, \quad (34d)$$

$$I_5 = p_{11} + p_{22} + p_{32} - p_{51} + 2p_{53} \leq 4, \quad (34e)$$

$$I_6 = p_{11} + 2p_{22} - p_{51} + 2p_{53} \leq 4. \quad (34f)$$

While it was shown in Ref. [24] that a quantum model can violate the first of these inequalities and obtain the logical maximum of $I_1 = 3$, the degree to which the other inequalities can be violated has not, to our knowledge, previously been studied. We note that this question is also addressed in the parallel work of Ref. [42].

In this scenario, where we have both nontrivial preparation and measurement operational equivalences, we fail to obtain nontrivial bounds on these inequalities using the basic hierarchy described by Proposition 6. Instead, we employ the variant of the hierarchy described in Appendix B, which uses the principal square roots of the states ρ_x and/or measurements $E_{b|y}$ in the operator list, but otherwise follows the same approach. This hierarchy, which is a strict extension of the one described by Proposition 6, allows us to place strong bounds on all the above

inequalities. Indeed, using moment matrices of size 1191 and localizing matrices of size 85 (and monomials involving square roots of measurement operators, but not of states; see Appendix B), we obtain the following quantum bounds [43]:

$$I_1 \lesssim 3.000, \quad I_2 \lesssim 2.866, \quad I_3 \lesssim 3.500,$$
$$I_4 \lesssim 3.366, \quad I_5 \lesssim 4.689, \quad I_6 \lesssim 4.646. \tag{35}$$

Using a see-saw optimization approach, for all six inequalities we are able to obtain quantum strategies saturating the bounds from the hierarchy, showing that they are in fact tight up to the precision of the SDP solver.

Interestingly, we are moreover able to show that the maximum quantum violation of the third inequality (34c) cannot be obtained with projective measurements. Indeed, by using the hierarchy of Proposition 6 and imposing the constraints following from the projectivity of POVM elements (and using the same monomial lists as for the above results) we are able to show that $I_3 \lesssim 3.464$ for projective quantum models. Using a see-saw optimization, we are able to obtain projective quantum models saturating this bound to numerical precision, thereby confirming its tightness and showing that nonprojective measurements, just like mixed states, are resources for quantum contextuality.

V. SIMULATING PREPARATION CONTEXTUALITY

Quantum correlations are famously capable of going beyond those achievable in classical theories in numerous scenarios, as highlighted by the violation of Bell inequalities and, indeed, noncontextuality inequalities. One can likewise consider correlations that are even stronger than those observed in nature, which we call "postquantum" correlations. Interest in postquantum theories stems from them nonetheless respecting physical principles such as no signalling, and understanding what physical principles distinguish quantum and postquantum correlations can lead to new insights into quantum theory itself [44–46].

An interesting strategy to study the correlations obtained by different physical theories is to ask what kind of resource, and how much of it, one should supplement a theory with to achieve stronger correlations. This question has been extensively studied in the context of simulating Bell correlations with classical theory and additional resources. Two such resources that can be used in that case are classical communication [47,48] and measurement dependence [49]. Similarly, various resources have also been investigated in Kochen-Specker contextuality experiments with the goal of simulating quantum correlations within a classical theory [50–52]. To our knowledge, however, nothing is known about what resources would be necessary to simulate operationally contextual correlations, and

in particular the especially relevant resource of preparation contextuality.

In this section, we begin by casting preparation contextuality scenarios as information-theoretic games, and show how these allow us to formalize a notion of simulation cost, for both classical and quantum models. The resource used is the preparation of states, which deviate from the required operational equivalences. This is a natural figure of merit as the defining feature of a model for noncontextual correlations within a given theory is that the underlying ontological model obeys the specified operational equivalences; it is thus this condition that must be violated in some way if stronger correlations are to be simulated. We leverage our hierarchy of semidefinite relaxations to quantify both the simulation of quantum contextual correlations using classical theory, and the simulation of postquantum correlations using quantum theory.

A. Zero-information games

To show how the cost of simulating preparation contextuality can be quantified in information-theoretic terms, we begin by giving an alternative interpretation for preparation contextuality scenarios (i.e., contextuality scenarios involving only nontrivial operational equivalences between sets of preparations). In particular, we describe how preparation contextuality scenarios can be interpreted as games in which Alice is required to hide some knowledge about her input x (see, e.g., Ref. [53]).

Consider thus a contextuality experiment involving R preparation operational equivalences. For a given such equivalence $r \in [R]$ involving a partition into K_r sets $S_k^{(r)}$, let Alice randomly choose a set $S_k^{(r)}$ [with uniform prior $p(S_k^{(r)}) = 1/K^{(r)}$] and a state from that set with prior $p(x|S_k^{(r)}) = \xi_k^{(r)}(x)$. How well could a receiver hope to identify which of the sets $\{S_1^{(r)}, \ldots, S_{K^{(r)}}^{(r)}\}$ the state they receive is sampled from? The optimal discrimination probability in an operational theory is

$$G^{(r)} := \max_{\tilde{p}(\cdot|x)} \frac{1}{K^{(r)}} \sum_{k=1}^{K^{(r)}} \sum_{x \in S_k^{(r)}} \xi_k^{(r)}(x) \tilde{p}(k|x), \tag{36}$$

where $\tilde{p}$ is the response distribution for the discrimination. Using that $\sum_{k=1}^{K_r} \tilde{p}(k|x) = 1$, it straightforwardly follows (see Appendix E) that the discrimination probability is $G^{(r)} = (1/K^{(r)})$ (i.e., random) if and only if the rth operational equivalence is satisfied. The discrimination probability constitutes an operational interpretation of the min-entropic accessible information about the set membership of x [54], and is convenient to work with. More precisely, the accessible information is given by

$$\mathcal{I}_r = \log_2(K^{(r)}) + \log_2(G^{(r)}). \tag{37}$$

Thus, we can associate the operational equivalences to an information tuple $\bar{\mathcal{I}} = (\mathcal{I}_1, \ldots, \mathcal{I}_R)$. A contextuality experiment is a zero-information game since $G^{(r)} = (1/K^{(r)})$ for all r is equivalent to vanishing information: $\bar{\mathcal{I}} = \bar{0}$.

B. Information cost of simulating preparation contextuality

Since a vanishing information tuple $\bar{\mathcal{I}}$ is necessary for a faithful realization of a contextuality scenario in a given physical model, it follows that contextual correlations that cannot be explained in said model require an overhead information, i.e., an information tuple $\bar{\mathcal{I}} \neq \bar{0}$. In both classical (noncontextual) models and quantum theory, this means that the preparations are allowed to deviate from the operational equivalences specified by the contextuality scenario to an extent quantified by the overhead information. By doing so, one necessarily goes beyond a standard model for the scenario, as defined in Definition 3 for classical models and Definition 5 for quantum theory.

For the simplest case of a single operational equivalence (i.e., $R = 1$), we define the information cost, $\mathcal{Q}$, of simulating $p(b|x, y)$ in quantum theory as the smallest amount of overhead information required for quantum theory to reproduce the correlations:

$$\mathcal{Q}[p] := \min \mathcal{I},$$
$$\text{such that} \quad \rho_x \succeq 0, \quad \text{tr}(\rho_x) = 1, \quad E_{b|y} \succeq 0,$$
$$\sum_b E_{b|y} = \mathbb{1}, \text{ and } p(b|x, y) = \text{tr}\left(\rho_x E_{b|y}\right). \quad (38)$$

However, when several operational equivalences are involved, the information is represented by a tuple $\bar{\mathcal{I}}$ and it is unclear how the information cost of simulation should be defined (note, in particular, that the operational equivalences may not be independent, so information about one may also provide information about another). We thus focus here on the simpler case described above, and leave the more general case of $R > 1$ for future research.

It is not straightforward to evaluate $\mathcal{Q}$. However, by modifying our semidefinite relaxations of contextual quantum correlations we can efficiently obtain lower bounds on $\mathcal{Q}$ in general scenarios. Indeed, note that from Eq. (36), interpreted in a quantum model, it follows that if σ_r satisfies $\sigma_r \succeq \sum_{x \in S_k^{(r)}} \xi_k^{(r)}(x)\rho_x$ for every $k \in [K^{(r)}]$ then one has $G^{(r)} \leq (1/K_r)\text{tr}(\sigma_r)$. Thus, rather than imposing the constraint arising from $\text{tr}(\sigma_r) = 1$ in our hierarchy of semidefinite relaxations, we can instead minimize $[(1/K_r)$ times] the term corresponding to $\text{tr}(\sigma_r)$ in the moment matrix, which thus provides an upper bound on $G^{(r)}$. Note that this provides an alternative interpretation to the constraint that $\Gamma_{\mathbb{1},\sigma_r} = 1$ in Eq. (29e): it enforces the fact that Bob should have no information about which set $S_k^{(r)}$ Alice's state is chosen from. This interpretation makes an interesting link to the recently developed approach

to bounding informationally constrained correlations [18], and which indeed is the initial motivation for the approach we take in this paper.

Considering still the case of $R = 1$, we thereby bound the information cost of a quantum simulation by evaluating the semidefinite relaxation as follows.

Proposition 7. *For any fixed lists $\mathcal{S}$, $\mathcal{L}$, $\mathcal{O}$ of monomials from J, the quantum simulation cost $\mathcal{Q}[p]$ is lower bounded as*

$$\log_2(K^{(1)}) + \log_2(G^*) \leq \mathcal{Q}[p], \quad (39)$$

where G^* is obtained as

$$G^* = \min \quad \frac{\Gamma_{\mathbb{1},\sigma_1}}{K^{(1)}},$$
$$\text{such that} \quad \Gamma \succeq 0, \quad \tilde{\Lambda}^{(1,k)} \succeq 0, \quad \tilde{\Upsilon}^x \succeq 0,$$
$$\Gamma_{\mathbb{1},\rho_x} = 1,$$
$$\sum_{u,v} c_{u,v}\Gamma_{u,v} = 0 \quad \text{if} \quad \sum_{u,v} c_{u,v}\text{tr}(u^\dagger v) = 0,$$
$$\tilde{\Lambda}_{u,v}^{(1,k)} = \Gamma_{u,\sigma_1 v} - \sum_{x \in S_k^{(1)}} \xi_k^{(1)}\Gamma_{u,\rho_x v},$$
$$\tilde{\Upsilon}_{u,v}^x = \Gamma_{u,\rho_x v}, \quad (40)$$

where the above operators are all taken to be Hermitian.

The correctness of Proposition 7 follows immediately from Eq. (37) and the fact that G^* is an upper bound on $G^{(1)}$.

Furthermore, one can similarly consider the information cost of simulation in classical models. In analogy with the quantum simulation cost, we define the classical simulation cost, $\mathcal{C}$, as the smallest overhead information required for a classical noncontextual model to reproduce given correlations:

$$\mathcal{C}[p] := \min \mathcal{I} \quad \text{such that} \quad p(b|x, y)$$
$$= \sum_\lambda p(\lambda|x)p(b|y, \lambda),$$
$$\forall x : \sum_\lambda p(\lambda|x) = 1, \quad \forall(\lambda, y) : \sum_b p(b|y, \lambda) = 1. \quad (41)$$

Naturally, in contrast to quantum simulation, every contextual distribution $p(b|x, y)$ is associated to a nonzero classical simulation cost. In analogy with the quantum case, we can place lower bounds on the classical simulation cost using the SDP hierarchy we discuss and assuming that all variables commute, thereby introducing many further constraints on the SDP and providing necessary conditions for a classical model to exist for a given value of G.

However, it turns out that a precise characterization of the classical simulation cost, in terms of a linear program, is also possible by exploiting the fact that the set of classical, informationally restricted, correlations forms a convex polytope [17,18,55].

Finally, we make the interesting observation that the discrimination probability G can be given a resource-theoretic interpretation in terms of a robustness measure. As we discuss in Appendix F, this can be used to give an alternative interpretation of the simulation cost $\mathcal{I}$.

C. Simulation cost in the simplest scenario

We illustrate the above discussion of the classical and quantum simulation costs of contextuality by applying it to arguably the simplest contextuality experiment, namely parity-oblivious multiplexing (POM) [19]. In POM, Alice has four preparations ($x \in [4]$ written in terms of two bits $x := x_1 x_2 \in [2]^2$) and Bob has two binary-outcome measurements ($y \in [2]$ and $b \in [2]$). The sole operational equivalence is $(1/2)P_{11} + (1/2)P_{22} \simeq (1/2)P_{12} + (1/2)P_{21}$, which corresponds to Alice's preparations carrying no information about the parity of her input x. The task is for Bob to guess the value of her yth input bit. The average success probability in a noncontextual model obeys

$$\mathcal{A}_{\text{POM}} := \frac{1}{8} \sum_{x,y} p\left(b = x_y | x, y\right) \leq \frac{3}{4}. \quad (42)$$

In contrast, quantum models obey the tight bound $\mathcal{A}_{\text{POM}} \leq (1/2)[1 + (1/\sqrt{2})]$ [19]. However, a postquantum probability theory can achieve the algebraically maximal success probability of $\mathcal{A}_{\text{POM}} = 1$ [56].

We consider the information cost of simulating a given value of $\mathcal{A}_{\text{POM}}$ (i.e., the minimal information cost over all distributions compatible with that value, which can easily be evaluated by modifying the linear and semidefinite programs defined above) in both classical and quantum models. The results are illustrated in Fig. 1. The classical simulation cost is analytically given by

$$\mathcal{C}_{\text{POM}} = \log_2\left(4\mathcal{A}_{\text{POM}} - 2\right). \quad (43)$$

In Appendix G we present an explicit simulation strategy that saturates this result, while the results of the linear program and the classical version of the hierarchy coincide with this value up to numerical precision.

For quantum models, we employ the described semidefinite relaxations using a moment matrix of size 547 and localizing matrices of size 89. The results are illustrated in Fig. 1. Importantly, we find that this lower bound on the quantum simulation cost is tight since we can saturate it with an explicit quantum strategy (detailed in
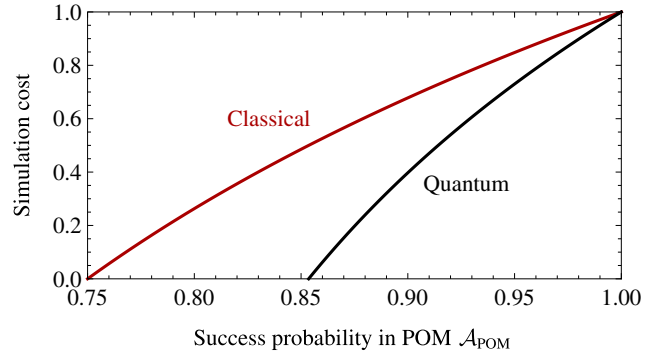

FIG. 1. The information cost of simulating contextuality in parity-oblivious multiplexing using classical and quantum models.

Appendix G). The quantum simulation cost is analytically given by

$$\mathcal{Q} = \log_2\left(2\mathcal{A}_{\text{POM}} - \sqrt{2}\right) + \log_2\left(2 + \sqrt{2}\right). \quad (44)$$

VI. CONCLUSIONS

In this paper we introduce a semidefinite relaxation hierarchy for bounding the set of contextual quantum correlations and demonstrate its usefulness by applying it to solve several open problems in quantum contextuality. This approach opens the door to the investigation of the limits of quantum contextuality in general prepare-and-measure experiments, as well as potential applications thereof. Moreover, it provides the building blocks with which to explore several interesting, related questions, such as whether our approach can be extended to contextuality scenarios involving more than two parties, and whether it can be adapted to bound quantum correlations in Kochen-Specker-type contextuality experiments.

By leveraging the interpretation of contextuality experiments as zero-information games, we introduce a measure of the cost of simulating preparation contextual correlations in restricted physical models, and show how this simulation cost can be bounded in both classical and quantum models. This raises three fundamental questions: (1) How can the definition of the simulation cost be extended to scenarios with multiple preparation operational equivalences which, *a priori*, may not be independent? (2) How does the simulation cost of contextuality scale in prepare-and-measure scenarios with increasingly many settings? and (3) For a given number of inputs and outputs, what is the largest simulation cost possible in order for classical correlations to reproduce quantum correlations? Additionally, it would be interesting to investigate how the simulation cost of operational contextuality relates to other notions of simulation, e.g., in Bell nonlocality, Kochen-Specker contextuality, and communication complexity. In particular, can our semidefinite relaxation techniques be

adapted to also bound simulation costs in such correlation experiments?

Our work thus provides both a versatile tool for bounding quantum contextuality and a general framework for analyzing the simulation of contextual correlations.

Note added.—Recently, we became aware of the related work of Ref. [42]. This work also addresses the problem of bounding the set of contextual quantum correlations. It uses a hierarchy of semidefinite programming relaxations that is considerably different to the one introduced here. For contextuality scenarios featuring measurement operational equivalences, as well as general mixed states and nonprojective measurements, the hierarchy of Ref. [42] appears to provide faster convergence [they recover, for example, more readily the bounds of Eq. (35)]. In contrast, the hierarchy we introduce here appears particularly well suited to preparation contextuality scenarios, admits a generalization to quantifying the simulation cost of contextuality, and makes an interesting conceptual connection to informationally restricted quantum correlations [17,18].

APPENDIX A: SDP HIERARCHY FOR PREPARATION NONCONTEXTUAL CORRELATIONS

There has been particular interest studying noncontextual correlations in contextuality scenarios involving only nontrivial operational equivalences for preparations; i.e., preparation noncontextual correlations. Since many of the examples we apply our SDP hierarchy to are of this form, we show here explicitly how Proposition 6 simplifies in such scenarios.

The case of preparation noncontextuality is particularly simplified by noting that, in this particular case, we can assume the measurements to be projective. Indeed, we can always invoke Naimark's dilation theorem to obtain projective measurements on a larger Hilbert space that give the same statistics on the states in a given quantum model. Crucially, since there are no (nontrivial) measurement operational equivalences, these dilated projective measurements also provide a valid quantum model for the contextuality scenario in question.

Proposition 8. *Let* $\mathcal{S}, \mathcal{L}, \mathcal{O}$ *be fixed lists of monomials from* J. *A necessary condition for the existence of a quantum model in a given preparation contextuality scenario reproducing the correlations* $\{p(b|x,y)\}_{b,x,y}$ *is the feasibility of the following SDP:*

$$\text{find} \quad \Gamma, \{\tilde{\Lambda}^{(r,k)}\}_{r,k}, \{\tilde{\Upsilon}^x\}_x,$$

$$\text{such that} \quad \Gamma \succeq 0, \quad \tilde{\Lambda}^{(r,k)} \succeq 0, \quad \tilde{\Upsilon}^x \succeq 0, \tag{A1a}$$

$$\Gamma_{\rho_x, E_{b|y}} = p(b|x,y), \tag{A1b}$$

$$\Gamma_{\mathbb{1},\rho_x} = 1, \quad \Gamma_{\mathbb{1},\sigma_r} = 1, \tag{A1c}$$

$$\sum_{u,v} c_{u,v} \Gamma_{u,v} = 0 \quad \text{if} \quad \sum_{u,v} c_{u,v} \text{tr}(u^\dagger v) = 0, \tag{A1d}$$

$$\tilde{\Lambda}^{(r,k)}_{u,v} = \Gamma_{u,\sigma_r v} - \sum_{x \in S_k^{(r)}} \xi_k^{(r)} \Gamma_{u,\rho_x v}, \tag{A1e}$$

$$\tilde{\Upsilon}^x_{u,v} = \Gamma_{u,\rho_x v}, \tag{A1f}$$

where the above operators are all taken to be real symmetric matrices.

APPENDIX B: VARIANT OF SDP HIERARCHY USING PRINCIPAL-SQUARE-ROOT OPERATORS

In the hierarchy described in Proposition 6, if the measurements are taken to be projective or the states pure (so they are likewise described by projectors), then all of the probabilities $p(b|x,y)$ appear on the diagonal either of one of the localizing matrices $\tilde{\Upsilon}^x$ or $\hat{\Upsilon}^{(b,y)}$ or, if both these sets of operators are projective, the moment matrix Γ. For example, in the case of projective measurements (as can always be assumed for preparation noncontextuality), one has $\tilde{\Upsilon}^x_{E_{b|y}, E_{b|y}} = \text{tr}(E_{b|y} \rho_x E_{b|y}) = \text{tr}(\rho_x E_{b|y}) = p(b|x,y)$. The positive semidefiniteness of these matrices thereby imposes strong constraints on the probability distribution, even at low levels of the hierarchy (notably, that they are non-negative, although the constraints are strictly stronger that this).

In the most general case, however, when no assumption of projective measurements or pure states can be made, the probabilities only appear on off-diagonal entries. In practice, we find that a consequence of this is the need to go to much higher levels of the hierarchy to obtain nontrivial constraints. Indeed, for the inequalities discussed in Sec. IV C we are unable to obtain useful constraints with the hierarchy of Proposition 6. Here, we show how this hierarchy can be modified and generalized to overcome this shortcoming.

Our approach exploits the simple fact that, since the states ρ_x and POVM elements $E_{b|y}$ are positive semidefinite, they have positive semidefinite principal square roots $\sqrt{\rho_x}$ and $\sqrt{E_{b|y}}$ such that $\sqrt{\rho_x}\sqrt{\rho_x} = \rho_x$

and $\sqrt{E_{b|y}}\sqrt{E_{b|y}} = E_{b|y}$, respectively. Instead of taking the operator set J defined in Eq. (12), we reformulate our hierarchy using the finer-grained operator set

$$J' = \{\mathbb{1}\} \cup \{\sqrt{\rho_x}\}_x \cup \{\sqrt{E_{b|y}}\}_{b,y} \cup \{\sigma_r, \tau_\ell\}_{r,\ell}. \quad (B1)$$

The moment matrix Γ and localizing matrices $\tilde{\Lambda}^{(r,k)}, \hat{\Lambda}^{(q,\ell)}$ can be constructed in the same way as for the original hierarchy, while the localizing matrices $\tilde{\Upsilon}^x$ and $\hat{\Upsilon}^{(b,y)}$ are now used to enforce the positive semidefiniteness of the principal roots, and thus have elements

$$\tilde{\Upsilon}^x_{u,v} = \text{tr}(u^\dagger \sqrt{\rho_x} v), \quad (B2)$$

$$\hat{\Upsilon}^{(b,y)}_{u,v} = \text{tr}(u^\dagger \sqrt{E_{b|y}} v). \quad (B3)$$

While this modification may appear to change little, an immediate consequence is that the probabilities $p(b|x,y)$ now appear on the diagonal of Γ; indeed one has $\Gamma_{\sqrt{E_{b|y}}\sqrt{\rho_x}, \sqrt{E_{b|y}}\sqrt{\rho_x}} = \text{tr}(\sqrt{\rho_x}\sqrt{E_{b|y}}\sqrt{E_{b|y}}\sqrt{\rho_x}) = \text{tr}(E_{b|y}\rho_x) = p(b|x,y)$.

Apart from this change in operator set, the conceptual approach of the hierarchy remains unchanged. The constraints (II)–(IV) described in Sec. III A are thus enforced in the same way, but now on the squares of the operators $\sqrt{\rho_x}$ and $\sqrt{E_{b|y}}$ around which the hierarchy is constructed. For example, the constraint that, for all x, $\text{tr}(\rho_x) = 1$ in any quantum model is now imposed by requiring that Γ satisfy

$$\forall x : \Gamma_{\sqrt{\rho_x}, \sqrt{\rho_x}} = 1. \quad (B4)$$

Following analogous reasoning to that of Sec. III A, we thus arrive at the following proposition describing the modified hierarchy.

Proposition 9. *Let $\mathcal{S}, \mathcal{L}, \mathcal{O}$ be fixed lists of monomials from J'. A necessary condition for the existence of a quantum model in a given contextuality scenario reproducing the correlations $\{p(b|x,y)\}_{b,x,y}$ is the feasibility of the following SDP:*

$$\begin{aligned}
\text{find} \quad & \Gamma, \{\tilde{\Lambda}^{(r,k)}\}_{r,k}, \{\hat{\Lambda}^{(q,\ell)}\}_{q,\ell}, \{\tilde{\Upsilon}^x\}_x, \{\hat{\Upsilon}^{(b,y)}\}_{b,y}, \\
\text{such that} \quad & \Gamma \succeq 0, \quad \tilde{\Lambda}^{(r,k)} \succeq 0, \quad \hat{\Lambda}^{(q,\ell)} \succeq 0,
\end{aligned}$$

$$\tilde{\Upsilon}^x \succeq 0, \quad \hat{\Upsilon}^{(b,y)} \succeq 0, \quad (B5a)$$

$$\Gamma_{\sqrt{E_{b|y}}\sqrt{\rho_x}, \sqrt{E_{b|y}}\sqrt{\rho_x}} = p(b|x,y), \quad (B5b)$$

$$\Gamma_{\sqrt{\rho_x}, \sqrt{\rho_x}} = 1, \quad (B5c)$$

$$\sum_{u,v} c_{u,v} \Gamma_{u,v} = 0 \quad \text{if} \quad \sum_{u,v} c_{u,v} \text{tr}(u^\dagger v) = 0, \quad (B5d)$$

$$\Gamma_{\mathbb{1},\sigma_r} = 1, \quad (B5e)$$

$$\Gamma_{\mathbb{1},\tau_q} = \sum_{(b,y)\in T_\ell^{(q)}} \zeta_\ell^{(q)}(b,y) \Gamma_{\sqrt{E_{b|y}}, \sqrt{E_{b|y}}}, \quad (B5f)$$

$$\tilde{\Lambda}^{(r,k)}_{u,v} = \Gamma_{u,\sigma_r v} - \sum_{x \in S_k^{(r)}} \xi_k^{(r)} \Gamma_{\sqrt{\rho_x}u, \sqrt{\rho_x}v}, \quad (B5g)$$

$$\hat{\Lambda}^{(q,\ell)}_{u,v} = \Gamma_{u,\tau_q v} - \sum_{(b,y)\in T_\ell^{(q)}} \zeta_\ell^{(q)} \Gamma_{\sqrt{E_{b|y}}u, \sqrt{E_{b|y}}v}, \quad (B5h)$$

$$\tilde{\Upsilon}^x_{u,v} = \Gamma_{u,\sqrt{\rho_x}v}, \quad \hat{\Upsilon}^{(b,y)}_{u,v} = \Gamma_{u,\sqrt{E_{b|y}}v}, \quad (B5i)$$

where the above operators are all symmetric real matrices.

Let us note firstly that Proposition 9 is strictly stronger than Proposition 6. Indeed, the latter can be seen as a special case of the former in which the monomial lists $\mathcal{S}, \mathcal{L}, \mathcal{O}$ are chosen so that the square root operators only ever appear in "matching" pairs.

While one may worry that one must go to higher levels of the hierarchy to obtain similarly strong constraints when employing this modified hierarchy, in practice we find that the situation is more subtle. Even in the case where either the measurements are assumed to be projective, or the states pure, we generally find that equally tight bounds could be obtained using either hierarchy. On the other hand, in the fully general case, we find that the modified hierarchy of Proposition 9 provides a clear advantage.

Finally, we note that one could likewise consider the intermediate possibility of taking the principal roots of only the states or the POVM elements in the operator set. In this case, the probabilities instead appear on the diagonal of the localizing matrices $\tilde{\Upsilon}^x$ or $\hat{\Upsilon}^{(b,y)}$. We find that, in practice, this option generally provided the best results for moment and localizing matrices of a given size. Indeed, the results for the example of Sec. IV C are obtained using the operator set

$$J'' = \{\mathbb{1}\} \cup \{\rho_x\}_x \cup \{\sqrt{E_{b|y}}\}_{b,y} \cup \{\sigma_r, \tau_\ell\}_{r,\ell}. \quad (B6)$$

The implementation of our hierarchy, which is freely available [34], allows one to choose between all these different variants of the hierarchy.

We finish by noting that, to our knowledge, this approach of building a SDP hierarchy from principal square root operators is novel, at least within quantum information, and may be of independent interest in other applications.

APPENDIX C: MAXIMAL QUANTUM VIOLATIONS OF NONCONTEXTUALITY INEQUALITIES

Here we present two further case studies illustrating the practical usefulness of the hierarchy of semidefinite relaxations of the set of quantum correlations in contextuality experiments that we describe in the main text.

1. The inequality of Ref. [16]

Ref. [16] experimentally implemented a test of contextuality based on the communication games introduced in Ref. [57]. In the scenario considered there, Alice receives one of six preparations $x := x_1 x_2$, where $x_1 \in \{0, 1\}$ is a bit and $x_2 \in \{0, 1, 2\}$ a trit. Bob receives a binary input $y \in \{0, 1\}$ and produces a ternary outcome $b \in \{0, 1, 2\}$. The authors then present the following noncontextuality inequality:

$$\mathcal{A} := \frac{1}{12} \sum_{x,y,m} (-1)^m p(b = T_m | x, y) \leq \frac{1}{2}, \qquad \text{(C1)}$$

where $m = 0, 1$ and $T_m = x_2 - (-1)^{x_1 + y + m} m - x_1 y \bmod 3$. This inequality is valid under the operational equivalence

$$\frac{1}{3} \sum_{x_2} P_{0x_2} \simeq \frac{1}{3} \sum_{x_2} P_{1x_2}, \qquad \text{(C2)}$$

i.e., when no information is relayed about the bit x_1. Notably, this noncontextuality inequality is isomorphic to the Collins-Gisin-Linden-Massar-Popescu Bell inequality [58]. It is shown in Ref. [16] that a quantum strategy (based on qutrits) can achieve the violation $\mathcal{A}^Q = (3 + \sqrt{33}/12) \approx 0.7287$, but the optimality of this violation was not proved.

Using our semidefinite relaxations we evaluate an upper bound on the largest possible value of $\mathcal{A}$ attainable in quantum theory. Specifically, using a moment matrix of size 386 and localizing matrices of size 49 (with $\mathcal{L} = \mathcal{O}$) and evaluating the corresponding semidefinite program, we obtain the value $\mathcal{A}^Q$ (up to the precision of approximately 10^{-7}). Hence, up to solver precision, this shows that the quantum protocol considered in Ref. [16] is indeed optimal.

2. The inequality of Ref. [15]

Ref. [15] introduced noncontextuality inequalities based on the task of random access coding. The authors consider a scenario in which Alice has an input $x \in [d^2]$ represented as two d-valued entries $x = x_1 x_2 \in \{0, \ldots, d-1\}^2$ while Bob receives a binary input $y \in [2]$ and produces an output $b \in \{0, \ldots, d-1\}$. Alice is required to communicate no information about the modular sum $x_1 + x_2 \bmod d$, i.e., her preparations must respect the operational equivalences

$$\forall (s, s') : \quad \frac{1}{d} \sum_{x_1 + x_2 = s} P_x \simeq \frac{1}{d} \sum_{x_1 + x_2 = s'} P_x, \qquad \text{(C3)}$$

where the addition is modulo d. Ref. [15] shows that the success probability in the random access code in a

noncontextual model obeys

$$\mathcal{A}_d := \frac{1}{2d^2} \sum_{x,y} p(b = x_y | x, y) \leq \frac{d+1}{2d}. \qquad \text{(C4)}$$

Notably, these noncontextuality inequalities are isomorphic to known Bell inequalities for random access codes [59].

Let us focus on the case of $d = 3$ (note that the case of $d = 2$ was solved in Ref. [19]). It was shown in Ref. [15] that there exists a quantum strategy (based on qutrits), which achieves the quantum violation $\mathcal{A}^Q = (7/9)$. However, the authors were unable to prove that a better quantum implementation cannot be found. Using a semidefinite relaxation corresponding to a moment matrix of size 563 and localizing matrices of size 52 (with $\mathcal{L} = \mathcal{O}$), we evaluate an upper bound on $\mathcal{A}$ valid for general quantum models. Up to solver precision, we recover the result $\mathcal{A}^Q$ (it agrees up to order 10^{-8}), thus showing that the explicit quantum strategy of Ref. [15] is optimal.

APPENDIX D: PEDAGOGICAL ILLUSTRATION OF SDP HIERARCHY CONSTRAINTS

To give some further understanding into the SDP hierarchy we present in Proposition 6 and in particular the form of the moment and localizing matrices and the constraints imposed upon them, we show here somewhat more explicitly the form that they take in the example we treat in Sec. IV A based on state discrimination.

In this example, the only operational equivalence is the preparation operational equivalence $(1/2)P_1 + (1/2)P_3 \simeq (1/2)P_2 + (1/2)P_4$, which in the form of Definition 1 is

$$\mathcal{E}_\mathcal{P} = \left[\left(\{1, 3\}, \{\tfrac{1}{2}, \tfrac{1}{2}\} \right), \left(\{2, 4\}, \{\tfrac{1}{2}, \tfrac{1}{2}\} \right) \right]. \qquad \text{(D1)}$$

As a result, we assume the measurements are projective and use the simplified version of the SDP hierarchy given in Proposition 8.

In order to derive the tight bounds discussed in the main text, we take the moment lists to be $\mathcal{S} = (\mathbb{1}, \boldsymbol{\rho}, \mathbf{E}, \sigma, \rho_1 \mathbf{E}, \rho_2 \mathbf{E}, \rho_3 \mathbf{E}, \rho_4 \mathbf{E}, \sigma \mathbf{E})$ and $\mathcal{L} = \mathcal{O} = (\mathbb{1}, \mathbf{E})$, where we use the shorthand $\boldsymbol{\rho} = (\rho_1, \rho_2, \rho_3, \rho_4)$ and $\mathbf{E} = (E_{1|1}, E_{2|1}, E_{1|2}, E_{2|2}, E_{1|3}, E_{2|3})$, so that $\rho_x \mathbf{E} = (\rho_x E_{1|1}, \rho_x E_{2|1}, \ldots)$, etc., and we denote $\sigma_1 = \sigma$ since we have only one operational equivalence ($R = 1$). We thus have $|\mathcal{S}| = 42$ and $|\mathcal{L}| = |\mathcal{O}| = 7$. The moment matrix thus has the following block structure:

$$
\Gamma = \left(\begin{array}{c|c|c|c|c|c|c|c|c}
\gamma_{\mathbb{1}} & (1,1,1,1) & \Gamma_{\mathbb{1},\mathbf{E}} & 1 & \mathbf{p}(b|1,y) & \mathbf{p}(b|2,y) & \mathbf{p}(b|3,y) & \mathbf{p}(b|4,y) & \Gamma_{\mathbb{1},\sigma\mathbf{E}} \\
\hline
& \Gamma_{\boldsymbol{\rho},\boldsymbol{\rho}} & \Gamma_{\boldsymbol{\rho},\mathbf{E}} & \Gamma_{\boldsymbol{\rho},\sigma} & \Gamma_{\boldsymbol{\rho},\rho_1\mathbf{E}} & \Gamma_{\boldsymbol{\rho},\rho_2\mathbf{E}} & \Gamma_{\boldsymbol{\rho},\rho_3\mathbf{E}} & \Gamma_{\boldsymbol{\rho},\rho_4\mathbf{E}} & \Gamma_{\boldsymbol{\rho},\sigma\mathbf{E}} \\
\hline
& & \Gamma_{\mathbf{E},\mathbf{E}} & \Gamma_{\mathbf{E},\sigma} & \Gamma_{\mathbf{E},\rho_1\mathbf{E}} & \Gamma_{\mathbf{E},\rho_2\mathbf{E}} & \Gamma_{\mathbf{E},\rho_3\mathbf{E}} & \Gamma_{\mathbf{E},\rho_4\mathbf{E}} & \Gamma_{\mathbf{E},\sigma\mathbf{E}} \\
\hline
& & & \gamma_{\sigma,\sigma} & \Gamma_{\sigma,\rho_1\mathbf{E}} & \Gamma_{\sigma,\rho_2\mathbf{E}} & \Gamma_{\sigma,\rho_3\mathbf{E}} & \Gamma_{\sigma,\rho_4\mathbf{E}} & \Gamma_{\sigma,\sigma\mathbf{E}} \\
\hline
& & & & \Gamma_{\rho_1\mathbf{E},\rho_1\mathbf{E}} & \Gamma_{\rho_1\mathbf{E},\rho_2\mathbf{E}} & \Gamma_{\rho_1\mathbf{E},\rho_3\mathbf{E}} & \Gamma_{\rho_1\mathbf{E},\rho_4\mathbf{E}} & \Gamma_{\rho_1\mathbf{E},\sigma\mathbf{E}} \\
\hline
& & & & & \Gamma_{\rho_2\mathbf{E},\rho_2\mathbf{E}} & \Gamma_{\rho_2\mathbf{E},\rho_3\mathbf{E}} & \Gamma_{\rho_2\mathbf{E},\rho_4\mathbf{E}} & \Gamma_{\rho_2\mathbf{E},\sigma\mathbf{E}} \\
\hline
& & & & & & \Gamma_{\rho_3\mathbf{E},\rho_3\mathbf{E}} & \Gamma_{\rho_3\mathbf{E},\rho_4\mathbf{E}} & \Gamma_{\rho_3\mathbf{E},\sigma\mathbf{E}} \\
\hline
& & & & & & & \Gamma_{\rho_4\mathbf{E},\rho_4\mathbf{E}} & \Gamma_{\rho_4\mathbf{E},\sigma\mathbf{E}} \\
\hline
& & & & & & & & \Gamma_{\sigma\mathbf{E},\sigma\mathbf{E}}
\end{array}\right),
\tag{D2}
$$

where the blocks correspond to the block specification of $\mathcal{S}$ given above, and we give only the upper triangle since the matrix is symmetric. The vectors $\mathbf{p}(b|x,y)$ are to be understood as $\mathbf{p}(b|x,y) = [p(1|x,1), p(2|x,1), p(1|x,2), p(2|x,2), p(1|x,3), p(2|x,3)]$. Let us note immediately that, by Eq. (A1b), one has

$$
\Gamma_{\rho,E} = \begin{pmatrix} \mathbf{p}(1|1,y) \\ \mathbf{p}(2|1,y) \\ \mathbf{p}(3|1,y) \\ \mathbf{p}(4|1,y) \end{pmatrix}.
\tag{D3}
$$

The localizing matrices $\tilde{\Upsilon}^x$ and $\tilde{\Lambda}^{(k)}$ can readily be identified as

$$
\tilde{\Upsilon}^x = \left(\begin{array}{c|c} \gamma_{\mathbb{1}} & \mathbf{p}(b|x,y) \\ \hline & \Gamma_{\mathbf{E},\rho_x\mathbf{E}} \end{array}\right)
\tag{D4}
$$

and

$$
\tilde{\Lambda}^{(k)} = \left(\begin{array}{c|c} \gamma_{\mathbb{1}} & \Gamma_{\sigma\mathbf{E}} \\ \hline & \Gamma_{\mathbf{E},\sigma\mathbf{E}} \end{array}\right) - \frac{1}{2}\sum_{x \in S_k} \tilde{\Upsilon}^x,
\tag{D5}
$$

where $S_1 = \{1,3\}$ and $S_2 = \{2,4\}$.

The remaining constraints of interest are those referred to in Eq. (A1d). To illustrate these, let us expand on the form of some of the blocks in Γ. From the completeness relation $\sum_b E_{b|y} = \mathbb{1}$ we can write $\Gamma_{\boldsymbol{\rho},\sigma\mathbf{E}}$ as

$$
\Gamma_{\boldsymbol{\rho},\sigma\mathbf{E}} = \begin{pmatrix}
\gamma_{\rho_1\sigma E_{1|1}} & \gamma_{\rho_1\sigma} - \gamma_{\rho_1\sigma E_{1|1}} & \gamma_{\rho_1\sigma E_{1|2}} & \gamma_{\rho_1\sigma} - \gamma_{\rho_1\sigma E_{1|2}} & \gamma_{\rho_1\sigma E_{1|3}} & \gamma_{\rho_1\sigma} - \gamma_{\rho_1\sigma E_{1|3}} \\
\gamma_{\rho_2\sigma E_{1|1}} & \gamma_{\rho_2\sigma} - \gamma_{\rho_2\sigma E_{1|1}} & \gamma_{\rho_2\sigma E_{1|2}} & \gamma_{\rho_2\sigma} - \gamma_{\rho_2\sigma E_{1|2}} & \gamma_{\rho_2\sigma E_{1|3}} & \gamma_{\rho_2\sigma} - \gamma_{\rho_2\sigma E_{1|3}} \\
\gamma_{\rho_3\sigma E_{1|1}} & \gamma_{\rho_3\sigma} - \gamma_{\rho_3\sigma E_{1|1}} & \gamma_{\rho_3\sigma E_{1|2}} & \gamma_{\rho_3\sigma} - \gamma_{\rho_3\sigma E_{1|2}} & \gamma_{\rho_3\sigma E_{1|3}} & \gamma_{\rho_3\sigma} - \gamma_{\rho_3\sigma E_{1|3}} \\
\gamma_{\rho_4\sigma E_{1|1}} & \gamma_{\rho_4\sigma} - \gamma_{\rho_4\sigma E_{1|1}} & \gamma_{\rho_4\sigma E_{1|2}} & \gamma_{\rho_4\sigma} - \gamma_{\rho_4\sigma E_{1|2}} & \gamma_{\rho_4\sigma E_{1|3}} & \gamma_{\rho_4\sigma} - \gamma_{\rho_4\sigma E_{1|3}}
\end{pmatrix},
\tag{D6}
$$

where $\gamma_{\rho_x\sigma}$ are the elements of $\Gamma_{\boldsymbol{\rho},\sigma}$. By the cyclicity of the trace and the projectivity of the measurement (i.e., $E_{b|y}E_{b'|y} = \delta_{b,b'}E_{b|y}$), the elements of $\Gamma_{\boldsymbol{\rho},\sigma\mathbf{E}}$ are then related to the elements of $\Gamma_{\rho_x\mathbf{E},\sigma\mathbf{E}}$ as [recalling that $\Gamma_{u,v} = \mathrm{tr}(u^\dagger v)$, so the elements of the monomial u are reversed]

$$
\Gamma_{\rho_x\mathbf{E},\sigma\mathbf{E}} = \begin{pmatrix}
\gamma_{\rho_x\sigma E_{1|1}} & 0 & \gamma_{\rho_x\sigma E_{1|2}E_{1|1}} & \gamma_{\rho_x\sigma E_{2|2}E_{1|1}} & \gamma_{\rho_x\sigma E_{1|3}E_{1|1}} & \gamma_{\rho_x\sigma E_{2|3}E_{1|1}} \\
0 & \gamma_{\rho_x\sigma E_{2|1}} & \gamma_{\rho_x\sigma E_{1|2}E_{2|1}} & \gamma_{\rho_x\sigma E_{2|2}E_{2|1}} & \gamma_{\rho_x\sigma E_{1|3}E_{2|1}} & \gamma_{\rho_x\sigma E_{2|3}E_{2|1}} \\
\gamma_{\rho_x\sigma E_{1|1}E_{1|2}} & \gamma_{\rho_x\sigma E_{2|1}E_{1|2}} & \gamma_{\rho_x\sigma E_{1|2}} & 0 & \gamma_{\rho_x\sigma E_{1|3}E_{1|2}} & \gamma_{\rho_x\sigma E_{2|3}E_{1|2}} \\
\gamma_{\rho_x\sigma E_{1|1}E_{2|2}} & \gamma_{\rho_x\sigma E_{2|1}E_{2|2}} & 0 & \gamma_{\rho_x\sigma E_{2|2}} & \gamma_{\rho_x\sigma E_{1|3}E_{2|2}} & \gamma_{\rho_x\sigma E_{2|3}E_{2|2}} \\
\gamma_{\rho_x\sigma E_{1|1}E_{1|3}} & \gamma_{\rho_x\sigma E_{2|1}E_{1|3}} & \gamma_{\rho_x\sigma E_{1|2}E_{1|3}} & \gamma_{\rho_x\sigma E_{2|2}E_{1|3}} & \gamma_{\rho_x\sigma E_{1|3}} & 0 \\
\gamma_{\rho_x\sigma E_{1|1}E_{2|3}} & \gamma_{\rho_x\sigma E_{2|1}E_{2|3}} & \gamma_{\rho_x\sigma E_{1|2}E_{2|3}} & \gamma_{\rho_x\sigma E_{2|2}E_{2|3}} & 0 & \gamma_{\rho_x\sigma E_{2|3}}
\end{pmatrix},
$$

where, for the sake of legibility, we do not yet apply the completeness relations. These, e.g., further impose that $\gamma_{\rho_x\sigma E_{2|2}E_{1|1}} = \gamma_{\rho_x\sigma E_{1|1}} - \gamma_{\rho_x\sigma E_{1|2}E_{1|1}}$, $\gamma_{\rho_x\sigma E_{2|2}E_{2|1}} = \gamma_{\rho_x\sigma} - \gamma_{\rho_x\sigma E_{1|2}} - \gamma_{\rho_x\sigma E_{1|1}} + \gamma_{\rho_x\sigma E_{1|2}E_{1|1}}$, etc.

The other blocks of Γ can be reduced and related in similar ways by applying similar simplifications.

In practice, our code (which is freely accessible [34]), works by applying reductions to every element of the moment matrix to reduce it to a canonical form, before identifying the unique elements. The completeness relations can then be applied to further reduce the number of variables in the optimization problem. We note, however, that when projective measurements are considered it is generally not actually necessary to apply the constraints arising from the completeness relation. Although one obtains a potentially weaker set of necessary conditions, in practice we rarely see any difference in the power of the hierarchy under this relaxation.

APPENDIX E: CONTEXTUALITY EXPERIMENTS AS ZERO-INFORMATION GAMES

Here we show that, for a given operational equivalence (i.e., a fixed $r \in [R]$), a uniform discrimination probability $G = 1/K$ (i.e., a vanishing information $\mathcal{I} = 0$) is equivalent to the corresponding operational equivalence

$$\sum_{x \in S_k(x)} \xi_k(x) P_x = \sum_{x \in S_{k'}(x)} \xi_{k'}(x) P_x \quad (\text{E1})$$

being satisfied. To this end, use that $\sum_{k=1}^{K} \tilde{p}(k|x) = 1$ to write the discrimination probability on the form

$$G = \frac{1}{K} + \max_{\tilde{p}(\cdot|x)} \frac{1}{K} \sum_{k=1}^{K-1} \left[\sum_{x \in S_k(x)} \xi_k(x) \tilde{p}(k|x) \right.$$
$$\left. - \sum_{x \in S_K(x)} \xi_K(x) \tilde{p}(k|x) \right]. \quad (\text{E2})$$

It then follows from the convex linearity of $\tilde{p}$ in x (cf. Ref. [20], noting that $\tilde{p}$ must by definition arise from an ontological model) that the operational equivalences (E1) imply that the bracket in the above expression vanishes, thus leading to $G = 1/K$. Conversely, the condition $G = 1/K$ is equivalent to

$$0 = \max_{\tilde{p}(\cdot|x)} \sum_{k=1}^{K-1} \left[\sum_{x \in S_k(x)} \xi_k(x) \tilde{p}(k|x) - \sum_{x \in S_K(x)} \xi_K(x) \tilde{p}(k|x) \right]. \quad (\text{E3})$$

If the bracket on the right-hand side does not vanish we can always find a $\tilde{p}(\cdot|x)$ such that the argument of the maximization becomes positive. Thus, the operational equivalences are implied.

APPENDIX F: SIMULATION COST FROM ROBUSTNESS OF OPERATIONAL INEQUIVALENCE

Let us first show how the discrimination probability G, as defined in Eq. (36), can be related to a robustness measure within a resource-theoretic framework (see Ref. [60] for an overview of robustness measures in such frameworks). To start with, let us write G, for a quantum model, as

$$G = \max_M \frac{1}{K} \sum_{k=1}^{K} \sum_{x \in S_k} \xi_k(x) \text{tr}\left[\rho_x E_k\right],$$

$$= \max_M \frac{1}{K} \sum_{k=1}^{K} \text{tr}\left[\hat{\sigma}_k E_k\right], \quad (\text{F1})$$

where $\hat{\sigma}_k := \sum_{x \in S_k} \xi_k(x) \rho_x$ and the maximization is taken over POVMs $\{E_k\}_k$. Note that this optimization can equally well be written as an optimization over sets of operators $\{E_k\}_k$ for which $E_k \succeq 0$ for all k and $\sum_k \text{tr}[\sigma E_k] \leq 1$ for all quantum states σ. In this way, the optimization variables $\{E_k\}_k$ can be interpreted as witnesses for the (nonmembership in the) set F consisting of tuples of states of the form $\{\hat{\sigma}_k\}_k = \{\sigma, \ldots, \sigma\}$, where σ is some quantum state. This "free set" F can be interpreted as the set of coarse-grained states $\sum_x \xi_k^{(r)}(x) \rho_x$ arising from operationally equivalent preparations. Indeed, any such preparations give, by definition, $\{\hat{\sigma}_k\}_k \in F$, while any $\{\sigma, \ldots, \sigma\} \in F$ can be obtained from such a set of preparations by taking $\rho_x = \sigma$ for all x.

Given a witness for a set, one can look to give an interpretation to its violation. Such an interpretation depends highly on the form of the involved optimization problem. In our case, our optimization problem is formulated in such a way that it corresponds to a commonly used resource quantifier in quantum-resource theories. Namely, the problem in Eq. (F1) is, up to scaling and shifting, the SDP dual of the generalized robustness measure R_F of a state tuple $\mathcal{T} = \{\hat{\sigma}_k\}_k$ with respect to the set F. The generalized robustness is defined as

$$R_F(\mathcal{T}) := \min \left\{ t \geq 0 \mid \frac{\mathcal{T} + t\tilde{\mathcal{T}}}{1 + t} \in F \right\}, \quad (\text{F2})$$

where the optimization is over all tuples of states $\tilde{\mathcal{T}}$ that are of the same size as $\mathcal{T}$. R_F is thus a measure of how much the operational equivalences between the states $\{\rho_x\}_x$ are violated (or, more precisely, how far their coarse grainings are from being equal), and we call it the (generalized) robustness of operational inequivalence. The dual of $R_F(\mathcal{T})$ is straightforward to obtain (see, e.g., Refs. [61,62]), leading to the relation

$$1 + R_F(\mathcal{T}) = K G. \quad (\text{F3})$$

Hence, for a given set of states $\{\rho_x\}_x$, the accessible information $\mathcal{I}$ can be related to the robustness of the corresponding tuple $\{\hat{\sigma}_k\}_k$ with respect to those tuples that contain no information about the index k, i.e.,

$$\mathcal{I} = \log_2[1 + R_F(\mathcal{T})]. \tag{F4}$$

In this way, the quantum simulation cost corresponds to the minimum robustness $R_F(\mathcal{T})$ taken over all tuples $\mathcal{T}$ arising from states $\{\rho_x\}_x$ compatible with the observed statistics $p(b|x, y)$.

We note that, in recent years, several links between robustness measures and advantages in discrimination tasks [61,62] and more general quantum games [63] have been uncovered. Free sets of the form of F have not previously been studied, and it remains an interesting open question to study how the robustness of operation inequivalence relates to advantages in such operational tasks, albeit one beyond the scope of the present paper.

APPENDIX G: SIMULATION STRATEGIES

Here we first present optimal strategies for both the classical and quantum simulation cost of contextual correlations in parity-oblivious multiplexing (POM).

1. Classical model

We first give an optimal strategy for the classical simulation cost of parity-oblivious multiplexing. A score of $\mathcal{A}_{\text{POM}} = 3/4$ can be obtained by the trivial classical strategy of sending only x_1, which reveals no information about the parity of $x_1 \oplus x_2$. A score $\mathcal{A}_{\text{POM}} > 3/4$ can be obtained by mixing this strategy (which has a simulation cost of 0) with another trivial strategy that has a simulation cost of 1 (e.g., by classically sending both x_1 and x_2). More precisely, a score of

$$\mathcal{A}_{\text{POM}} = 1 - q/4, \tag{G1}$$

for $q \in [0, 1]$, can be obtained by using these two strategies with probabilities q and $1 - q$, respectively. This trivially gives $G = 1 - (q/2)$ and thus $\mathcal{I} = \log_2(2 - q)$, providing an upper bound on $\mathcal{C}$.

Using the hierarchy of semidefinite relaxations with commuting operators as described in the main text, we tested 100 values of $\mathcal{A}_{\text{POM}}$ as given in Eq. (G1) (with $q \in [0, 1]$) and found that, in each case, the lower bound on $\mathcal{C}$ matched the above upper bound (up to the numerical precision of the solver). We thus find, as claimed in the main text, that

$$\mathcal{C} = \log_2(4\mathcal{A}_{\text{POM}} - 2). \tag{G2}$$

2. Quantum model

Here we describe an optimal quantum strategy for simulating postquantum correlations in parity-oblivious multiplexing. Recall that the optimal quantum strategy in parity-oblivious multiplexing gives $\mathcal{A}_{\text{POM}} = (1/2)[1 + (1/\sqrt{2})]$. This value can be obtained by taking the preparations

$$\rho_{x_1 x_2} = \frac{1}{2}\left(\mathbb{1} + \frac{(-1)^{x_1}\sigma_x + (-1)^{x_2}\sigma_y}{\sqrt{2}}\right) \tag{G3}$$

and performing projective measurements in the x and y bases if $y = 1$ or $y = 2$, respectively [19].

A score $\mathcal{A}_{\text{POM}} > (1/2)[1 + (1/\sqrt{2})]$ can be obtained by mixing this optimal strategy (which has a simulation cost of 0 as it obeys the operational equivalences), and a trivial strategy with a simulation cost of 1 (e.g., by classically sending both x_1 and x_2). More precisely, a score of

$$\mathcal{A}_{\text{POM}} = 1 - \frac{q}{2}\left(1 - \frac{1}{\sqrt{2}}\right), \tag{G4}$$

for $q \in [0, 1]$, can be obtained by using these two strategies with probabilities q and $1 - q$, respectively. This trivially gives $G = 1 - (q/2)$ and thus $\mathcal{I} = \log_2(2 - q)$, providing an upper bound on $\mathcal{Q}$.

Using the hierarchy of semidefinite relaxations as described in the main text, we test 100 values of $\mathcal{A}_{\text{POM}}$ as given in Eq. (G4) (with $q \in [0, 1]$) and find that, in each case, the lower bound on $\mathcal{Q}$ matched the above upper bound (up to the numerical precision of the solver). We thus find, as claimed in the main text, that

$$\mathcal{Q} = \log_2\left(2\mathcal{A}_{\text{POM}} - \sqrt{2}\right) + \log_2\left(2 + \sqrt{2}\right). \tag{G5}$$

[1] J. S. Bell, On the problem of hidden variables in quantum mechanics, Rev. Mod. Phys. **38**, 447 (1966).

[2] S. Kochen and E. P. Specker, The problem of hidden variables in quantum mechanics, J. Math. Mech. **17**, 59 (1967).

[3] R. W. Spekkens, Contextuality for preparations, transformations, and unsharp measurements, Phys. Rev. A **71**, 052108 (2005).

[4] R. W. Spekkens, Negativity and Contextuality are Equivalent Notions of Nonclassicality, Phys. Rev. Lett. **101**, 020401 (2008).

[5] M. F. Pusey, Anomalous Weak Values are Proofs of Contextuality, Phys. Rev. Lett. **113**, 200401 (2014).

[6] M. S. Leifer and O. J. E. Maroney, Maximally Epistemic Interpretations of the Quantum State and Contextuality, Phys. Rev. Lett. **110**, 120401 (2013).

[7] M. Lostaglio, Quantum Fluctuation Theorems, Contextuality, and Work Quasiprobabilities, Phys. Rev. Lett. **120**, 040602 (2018).

[8] A. Tavakoli and R. Uola, Measurement incompatibility and steering are necessary and sufficient for operational contextuality, Phys. Rev. Res. **2**, 013011 (2020).

[9] M. Lostaglio and G. Senno, Contextual advantage for state-dependent cloning, Quantum **4**, 258 (2020).

[10] H. Anwer, N. Wilson, R. Silva, S. Muhammad, A. Tavakoli, and M. Bourennane, Noise-robust preparation contextuality shared between any number of observers via unsharp measurements, arXiv:1904.09766 [quant-ph] (2019).

[11] D. Saha and A. Chaturvedi, Preparation contextuality as an essential feature underlying quantum communication advantage, Phys. Rev. A **100**, 0221080 (2019).

[12] S. A. Yadavalli and R. Kunjwal, Contextuality in entanglement-assisted one-shot classical communication, arXiv:2006.00469 [quant-ph] (2020).

[13] D. Schmid, R. W. Spekkens, and E. Wolfe, All the non-contextuality inequalities for arbitrary prepare-and-measure experiments with respect to any fixed set of operational equivalences, Phys. Rev. A **97**, 062103 (2018).

[14] D. Schmid and R. W. Spekkens, Contextual Advantage for State Discrimination, Phys. Rev. X **8**, 011015 (2018).

[15] A. Ambainis, M. Banik, A. Chaturvedi, D. Kravchenko, and A. Rai, Parity oblivious d-level random access codes and class of noncontextuality inequalities, Quantum Inf. Process. **18**, 111 (2019).

[16] A. Hameedi, A. Tavakoli, B. Marques, and M. Bourennane, Communication Games Reveal Preparation Contextuality, Phys. Rev. Lett. **119**, 220402 (2017).

[17] A. Tavakoli, E. Zambrini Cruzeiro, J. Bohr Brask, N. Gisin, and N. Brunner, Informationally restricted quantum correlations, Quantum **4**, 332 (2020).

[18] A. Tavakoli, E. Zambrini Cruzeiro, E. Woodhead, and S. Pironio, Informationally restricted correlations: a general framework for classical and quantum systems, arXiv:2007.16145 [quant-ph] (2020).

[19] R. W. Spekkens, D. H. Buzacott, A. J. Keehn, B. Toner, and G. J. Pryde, Preparation Contextuality Powers Parity-Oblivious Multiplexing, Phys. Rev. Lett. **102**, 010401 (2009).

[20] These distributions must, respectively, be linear in preparations x (since the epistemic state of a mixture of preparations is the mixture of the epistemic states of the respective preparations) and measurement effects $[b|M_y]$ (since a measurement effect arising from a mixture of measurements or a postprocessing of outcomes must be represented by the corresponding mixture and postprocessing of response functions).

[21] Here, the quantifier is over all possible measurements (that, e.g., Bob could perform), not only the fixed set $\{M_y\}_y$ he uses in the prepare-and-measure experiment at hand.

[22] In particular, one can obtain such a bipartition (e.g., for preparations) by taking $S_1 = \{x : \alpha_x - \beta_x \geq 0\}$, $S_2 = \{x : \alpha_x - \beta_x < 0\}$, $\xi_1(x) = (\alpha_x - \beta_x)/[\sum_{x \in S_1}(\alpha_x - \beta_x)]$, and $\xi_2(x) = (\beta_x - \alpha_x)/[\sum_{x \in S_2}(\beta_x - \alpha_x)]$.

[23] R. W. Spekkens, The status of determinism in proofs of the impossibility of a noncontextual model of quantum theory, Found. Phys. **44**, 1125 (2014).

[24] M. D. Mazurek, M. F. Pusey, R. Kunjwal, K. J. Resch, and R. W. Spekkens, An experimental test of noncontextuality without unphysical idealizations, Nat. Commun. **7**, ncomms11780 (2016).

[25] R. Kunjwal and R. W. Spekkens, From the Kochen-Specker Theorem to Noncontextuality Inequalities Without Assuming Determinism, Phys. Rev. Lett. **115**, 110403 (2015).

[26] M. Navascués, S. Pironio, and A. Acín, A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations, New J. Phys. **10**, 073013 (2008).

[27] M. Navascués, S. Pironio, and A. Acín, Bounding the set of Quantum Correlations, Phys. Rev. Lett. **98**, 010401 (2007).

[28] M. Navascués and T. Vértesi, Bounding the set of Finite Dimensional Quantum Correlations, Phys. Rev. Lett. **115**, 020501 (2015).

[29] M. Navascués, A. Feix, M. Araújo, and T. Vértesi, Characterizing finite-dimensional quantum behavior, Phys. Rev. A **92**, 042117 (2015).

[30] Y. Wang, I. W. Primaatmaja, E. Lavie, A. Varvitsiotis, and C. C. W. Lim, Characterising the correlations of prepare-and-measure quantum networks, Npj Quantum Inf. **5**, 17 (2019).

[31] In practice, however, it is often preferable to use an intermediate hierarchy level, i.e., a monomial list that has largest degree k but does not contain all terms of degree at most k. We later exemplify this.

[32] M. Navascués, S. Pironio, and A. Acín, in Handbook on Semidefinite, Conic and Polynomial Optimization, edited by M. F. Anjos and J. B. Lasserre (Springer Science+Business Media, 2012), Chap. 21, p. 601.

[33] J. Lofberg, in 2004 IEEE International Conference on Robotics and Automation (IEEE Cat. No.04CH37508) (2004), p. 284.

[34] https://github.com/alastair-abbott/contextuality-hierarchy.

[35] D. Rosset, F. Montealegre-Mora, and J.-D. Bancal, in Quantum Theory and Symmetries, CRM Series in Mathematical Physics, edited by M. B. Paranjape, R. MacKenzie, Z. Thomova, P. Winternitz, and W. Witczak-Krempa (Springer, Cham, 2021).

[36] https://replab.github.io.

[37] The precise lists of moments used in this and all subsequent examples can be found along with our implementation of the SDP hierarchy, where the code generating these results is available [34]. In all these examples we take the monomial lists $\mathcal{L}$ and $\mathcal{O}$ for the localizing matrices to be the same (simply because this is sufficient to obtain the presented results), although one could indeed take these to be different if desired.

[38] MOSEK ApS, The MOSEK optimization toolbox for MATLAB manual. Version 9.0. (2019).

[39] A. Tavakoli, A. Hameedi, B. Marques, and M. Bourennane, Quantum Random Access Codes Using Single d-Level Systems, Phys. Rev. Lett. **114**, 170502 (2015).

[40] Note that the large size of the moment matrices meant that the solver precision we are able to obtain is somewhat reduced compared to the other examples discussed in this paper. Our numerical result agrees with the noncontextual bound of 2/3 to within 2×10^{-4}, which is within an acceptable range given the error metrics returned by the solver.

[41] We are similarly able to use our hierarchy to place an upper bound on the quantum violation of this inequality at $\mathcal{A}_{\text{HTMB}} \lesssim 0.704$ using a moment matrix of size 3295 and

localizing matrices of size 268 with the solver SCS [64]. We note that obtaining this bound requires using terms from the fourth level of the hierarchy. We leave it open as to what the tight quantum bound is.

[42] A. Chaturvedi, M. Farkas, and V. J. Wright, Characterising and bounding the set of quantum behaviours in contextuality scenarios, arXiv:2010.05853 [quant-ph] (2020).

[43] Due to the lack of symmetry in some of the inequalities and relatively large moment matrix size, we use the memory efficient solver SCS [64] to obtain some of the results of Eq. (35). This solver has the drawback of converging slower than more standard solvers [38], and we thus obtain a numerical precision of the order of $10^{-3} \sim 10^{-4}$.

[44] S. Popescu and D. Rohrlich, Quantum nonlocality as an axiom, Found. Phys. **24**, 379 (1994).

[45] J. Barrett, Information processing in generalized probabilistic theories, Phys. Rev. A **75**, 032304 (2007).

[46] M. Pawłowski, T. Paterek, D. Kaszlikowski, V. Scarani, A. Winter, and M. Żukowski, Information causality as a physical principle, Nature **461**, 1101 (2009).

[47] H. Buhrman, R. Cleve, S. Massar, and R. de Wolf, Nonlocality and communication complexity, Rev. Mod. Phys. **82**, 665 (2010).

[48] B. F. Toner and D. Bacon, Communication Cost of Simulating Bell Correlations, Phys. Rev. Lett. **91**, 187904 (2003).

[49] M. J. Hall, Relaxed Bell inequalities and Kochen-Specker theorems, Phys. Rev. A **84**, 022102 (2011).

[50] M. Kleinmann, O. Gühne, J. R. Portillo, J.-Å. Larsson, and A. Cabello, Memory cost of quantum contextuality, New J. Phys. **13**, 113011 (2011).

[51] S. Abramsky, R. S. Barbosa, and S. Mansfield, Contextual Fraction as a Measure of Contextuality, Phys. Rev. Lett. **119**, 050504 (2017).

[52] C. Duarte and B. Amaral, Resource theory of contextuality for arbitrary prepare-and-measure experiments, J. Math. Phys. **59**, 062202 (2018).

[53] I. Marvian, Inaccessible information in probabilistic models of quantum systems, non-contextuality inequalities and noise thresholds for contextuality, arXiv:2003.05984 [quant-ph] (2020).

[54] R. Koenig, R. Renner, and C. Schaffner, The operational meaning of min- and max-entropy, IEEE Trans, Inf. Theory **55**, 4337 (2009).

[55] This follows from the fact that it suffices to consider a finite alphabet size for the ontic variable λ [18].

[56] M. Banik, S. S. Bhattacharya, A. Mukherjee, A. Roy, A. Ambainis, and A. Rai, Limited preparation contextuality in quantum theory and its relation to the Cirel'son bound, Phys. Rev. A **92**, 030103(R) (2015).

[57] A. Tavakoli, M. Pawłowski, M. Żukowski, and M. Bourennane, Dimensional discontinuity in quantum communication complexity at dimension seven, Phys. Rev. A **95**, 020302(R) (2017).

[58] D. Collins, N. Gisin, N. Linden, S. Massar, and S. Popescu, Bell Inequalities for Arbitrarily High-Dimensional Systems, Phys. Rev. Lett. **88**, 040404 (2002).

[59] A. Tavakoli, B. Marques, M. Pawłowski, and M. Bourennane, Spatial versus sequential correlations for random access coding, Phys. Rev. A **93**, 032336 (2016).

[60] B. Regula, Convex geometry of quantum resource quantification, J. Phys. A: Math. Theor. **51**, 045303 (2017).

[61] R. Takagi, B. Regula, K. Bu, Z.-W. Liu, and G. Adesso, Operational Advantage of Quantum Resources in Subchannel Discrimination, Phys. Rev. Lett. **122**, 140402 (2019).

[62] R. Uola, T. Kraft, J. Shang, X.-D. Yu, and O. Gühne, Quantifying Quantum Resources with Conic Programming, Phys. Rev. Lett. **122**, 130404 (2019).

[63] R. Uola, T. Kraft, and A. A. Abbott, Quantification of quantum dynamics with input-output games, Phys. Rev. A **101**, 052306 (2020).

[64] B. O'Donoghue, E. Chu, N. Parikh, and S. Boyd, SCS: Splitting conic solver, version 2.1.2, https://github.com/cvxgrp/scs (2019).