Shared randomness and device-independent dimension witnessing

Julio I. de Vicente[*]

Departamento de Matemáticas, Universidad Carlos III de Madrid, Avda. de la Universidad 30, E-28911, Leganés (Madrid), Spain

(Received 15 November 2016; published 30 January 2017)

It has been shown that the conditional probability distributions obtained by performing measurements on an uncharacterized physical system can be used to infer its underlying dimension in a device-independent way both in the classical and the quantum setting. We analyze several aspects of the structure of the sets of probability distributions corresponding to a certain dimension, taking into account whether shared randomness is available as a resource. We first consider the so-called prepare-and-measure scenario. We show that quantumness and shared randomness are not comparable resources. That is, on the one hand there exist behaviors that require a quantum system of arbitrarily large dimension in order to be observed while they can be reproduced with a classical physical system of minimal dimension together with shared randomness. On the other hand, there exist behaviors that require exponentially larger dimensions classically than quantumly even if the former is supplemented with shared randomness. We also show that in the absence of shared randomness, the sets corresponding to a sufficiently small dimension are negligible (zero measure and nowhere dense) both classically and quantumly. This is in sharp contrast to the situation in which this resource is available, and it explains the exceptional robustness of dimension witnesses in the setting in which devices can be taken to be uncorrelated. We finally consider the Bell scenario in the absence of shared randomness, and we prove some nonconvexity and negligibility properties of these sets for sufficiently small dimensions. This shows again the enormous difference induced by the availability (or lack thereof) of this resource.

DOI: 10.1103/PhysRevA.95.012340

I. INTRODUCTION

Is it possible to estimate the degrees of freedom of an uncharacterized physical system? This question has received much attention in recent years in what is known as device-independent dimension witnessing (DIDW). It turns out that it is indeed possible to make tests about the underlying dimension of a physical system without making any assumption about it or on the internal functioning of the measurement devices used to interact with it. Dimension estimates can be constructed based only on the measurement data, i.e., on the observed probabilities of obtaining certain outcomes conditioned on the different possible choices of measurement. These results are not only interesting from a fundamental point of view, but they also play a role in quantum information processing. Besides allowing for experimental tests of the physical dimension [1], which might be considered as a resource, these investigations allow us to constrain the correlations that are achievable when the setting limits the underlying dimension of the physical systems used in a protocol. These scenarios are known as semi-device-independent quantum information processing: no assumption is made on the working of the devices nor on the physical system used except for its dimension. Ideas from DIDW have allowed us to prove the security of certain cryptographic schemes [2] and to provide randomness-expansion protocols in this framework [3]. Moreover, DIDW is intimately related to the field of quantum communication complexity, which studies the minimal amount of communication that parties must exchange to successfully carry out distributed computational tasks [4]. Indeed, communication can be quantified by the dimensionality of the physical systems used to encode the messages.

The first proposals for DIDW considered the Bell scenario of quantum nonlocality, since violating Bell inequalities by a certain amount might require quantum systems of at least a certain dimension [5]. Subsequently, the structure of quantum correlations under dimensionality constraints has been studied extensively [6]. Although other settings have been considered [7], a different general and simple formalism for DIDW was presented in [8] in the so-called prepare-and-measure scenario, which was largely explored afterward [9,10]. Both the Bell and the prepare-and-measure scenarios rely on different parties holding devices that interact with the physical system. It is usually assumed that the action of these devices might be correlated by the parties having access to a common random variable. This induces convexity into the sets of observable probability distributions corresponding to a given dimension, and separation theorems can be used to obtain linear functionals that enable DIDW. However, shared randomness can be viewed as a resource, and in certain settings it might be more natural to assume that all devices are independent (this is the case, for example, when the devices are trusted and are not jointly conspiring to mimic higher-dimensional behaviors). Conditions for DIDW with uncorrelated devices have been presented in [11] (the prepare-and-measure scenario) and more recently in [12] (the Bell scenario) and [13] (the prepare-and-measure scenario).

In this paper, we explore the differences for DIDW, taking into account whether shared randomness is available as a resource. To do this, we analyze in detail the structure of the sets of probability distributions corresponding to a certain dimension, taking into account both possibilities. We first consider the prepare-and-measure scenario, and we show that quantumness and shared randomness are not comparable resources. That is, on the one hand there exist behaviors

[*] jdvicent@math.uc3m.es

that require a quantum system of arbitrarily large dimension in order to be observed while they can be reproduced with a classical physical system of minimal dimension together with shared randomness. On the other hand, using results from communication complexity, it can be seen that there exist behaviors that require exponentially larger dimensions classically than quantumly even if the former is supplemented with shared randomness. We also show another clear difference depending on whether shared randomness is available. In the absence of it, the sets corresponding to a sufficiently small dimension are negligible both classically and quantumly: they are zero-measure and nowhere-dense subsets in the set of all possible behaviors. However, this is never the case in the other setting as these sets are never negligible independently of how small the dimension might be. This negligibility property also explains the exceptional robustness of dimension witnesses when devices are taken to be independent, as observed in [11]. In the second part of this article, we consider the Bell scenario. The availability (or lack thereof) of shared randomness is known to make a difference, and nonconvexity results for the sets of observable probability distributions of a fixed dimension are known [14,15]. Here we extend these results and prove systematically some nonconvexity properties for these sets for sufficiently small underlying dimension when the parties do not have access to shared randomness. Furthermore, contrary again to the case of correlated devices, we also show that in this case these sets have measure zero and are nowhere dense in the set of all quantum behaviors. To obtain all these results, we use some very simple dimension estimates based on the rank of a matrix.

II. PREPARE-AND-MEASURE SCENARIOS

The prepare-and-measure scenario [8] for witnessing dimensions in a device-independent way is the following. There are two parties, Alice (or A) and Bob (or B), which receive, respectively, inputs x and y from finite alphabets $\mathcal{X}$ and $\mathcal{Y}$. Their only chance to communicate is by A sending a classical or quantum physical system to B depending on her input. The dimension of this system, to be defined precisely below, quantifies the amount of communication used. Upon receival of the message, B interacts with the system by performing a measurement depending on his input and produces an output b that can take values in a finite alphabet. For simplicity, we will consider this output to be binary, i.e., $b \in \{0,1\}$. Then, we can record the conditional probabilities with which each output occurs for any given pair of inputs: $P(b|xy)$. This is the main object in a device-independent scenario, and we will refer to it as behavior and denote it by $\mathbf{P}$. Of course, as conditional probabilities, behaviors are characterized by $P(b|xy) \geqslant 0\ \forall\ b,x,y$ and $\sum_b P(b|xy) = 1\ \forall\ x,y$.

The question to be addressed in this setting is the following. Without using any knowledge on how A and B process their information, what is the minimal amount of classical or quantum communication sent from A to B that is compatible with the observation of a given behavior? The possible classical messages $m(x)$ are given by dits, i.e., $m \in \{1,\dots,d\}$. Thus, the amount of classical communication is measured by the dimension of the message d. A always has the chance to use a random strategy, i.e., she can send a message m given x with probability $s(m|x)$, and so does B, i.e., he can produce an output b given y and the reception of m with probability $t(b|ym)$. In the quantum case, A sends quantum states ρ_x. The dimension of her message is thus

$$d = \dim \sum_x \operatorname{supp} \rho_x, \tag{1}$$

where supp stands for the support of an operator. To produce his output, B can interact with the message through a quantum measurement conditioned on his input.

Thus, we define the set of behaviors obtained by sending classical messages of dimension at most d by $\mathcal{C}_d$ (this and the other sets to be defined below also depend on $|\mathcal{X}|$ and $|\mathcal{Y}|$, which we drop to ease the notation since these quantities should be clear in general from the context). In other words, $\mathbf{P} \in \mathcal{C}_d$ when

$$P(b|xy) = \sum_{m=1}^{d} s(m|x)t(b|my). \tag{2}$$

On the other hand, $\mathcal{Q}_d$ denotes the set of behaviors achievable by sending quantum states of dimension at most d. That is, $\mathbf{P} \in \mathcal{Q}_d$ if there exists measurements for B, $\{\Pi_b^y \geqslant 0\}$ with $\sum_b \Pi_b^y = \mathbf{1}\ \forall\ y$, such that

$$P(b|xy) = \operatorname{tr}\big(\rho_x \Pi_b^y\big), \tag{3}$$

where the $\{\rho_x\}$ are of dimension less than or equal to d [cf. Eq. (1)].

As already mentioned in the Introduction, this does not exhaust all possibilities. Depending on the physical setting, A and B may be granted with another resource to build their strategy: shared randomness. This means that A and B may preestablish the strategy each will follow depending on the value of a random variable to which they both have access. This boils down to the fact that they can prepare any convex combination of their previously allowed behaviors. Thus, we define the sets of behaviors obtained by sending classical or quantum messages of dimension at most d together with shared randomness by

$$\mathcal{C}'_d = \operatorname{Conv}(\mathcal{C}_d), \quad \mathcal{Q}'_d = \operatorname{Conv}(\mathcal{Q}_d), \tag{4}$$

where Conv$(\cdot)$ stands for the convex hull. The sets $\mathcal{C}_d$ and $\mathcal{Q}_d$ need not be convex, so in general we have strict inclusions $\mathcal{C}_d \subset \mathcal{C}'_d$ and $\mathcal{Q}_d \subset \mathcal{Q}'_d$ [9]. This means that shared randomness is indeed a resource that can allow us to perform some tasks using less communication. On the other hand, we clearly have as well the inclusions $\mathcal{C}_d \subseteq \mathcal{Q}_d$ and $\mathcal{C}'_d \subseteq \mathcal{Q}'_d$, which can also be seen in general to be strict. That is, quantum strategies are also a resource over classical strategies in order to reduce the amount of communication.

Notice that if $d \geqslant |\mathcal{X}|$, the scenario is trivial. A can unambiguously encode the value of her input into her message to B, who can then use his private randomness to output any possible behavior. Hence, $\mathcal{C}_{|\mathcal{X}|} = \mathcal{Q}_{|\mathcal{X}|} = \mathcal{C}'_{|\mathcal{X}|} = \mathcal{Q}'_{|\mathcal{X}|}$, which constitute the set of all behaviors in a given setting. Thus, given any valid behavior there always exist values of the dimension in which it is realizable in any of the aforementioned sets, since in the worst case we have $d = |\mathcal{X}|$. Therefore, it is always well defined to ask what the minimal value of the dimension is to obtain some behavior in any of the four sets of

$$\begin{tikzcd}
\mathcal{C}'_d & \subset & \mathcal{Q}'_d \\
\cup & \boxed{?} \arrow[ul] \arrow[dr] & \cup \\
\mathcal{C}_d & \subset & \mathcal{Q}_d
\end{tikzcd}$$

FIG. 1. For a fixed value of d, all inclusions among sets are clear except for those corresponding to classical communication together with shared randomness and quantum communication without shared randomness.

possible strategies. Taking into account the inclusions pointed out above, it comes as a natural question what the relation between $\mathcal{C}'_d$ and $\mathcal{Q}_d$ is (see Fig. 1). Moreover, given the status of both quantumness and shared randomness as a resource, it is interesting to know if one can exchange one for the other or if one is strictly more powerful than the other. Actually, this is a standard question in the context of communication complexity [16]. Here, one usually seeks differences in d that are larger than a logarithmic cost over $|\mathcal{X}|$ as this is considered negligible with respect to the size of the input: the so-called exponential separations. We will show that there exist scenarios in which $\mathcal{C}'_d \nsubseteq \mathcal{Q}_d$ and $\mathcal{Q}_d \nsubseteq \mathcal{C}'_d$, with both separations being exponential (or even arbitrary). The second inclusion is a straightforward observation from known results in communication complexity. To establish the first one, we will first observe in the following subsection some very simple dimension estimates based on the rank of a matrix associated with $\mathbf{P}$. Using again these estimates, we will finish this section by showing the negligibility of low-dimensional sets in the absence of shared randomness. This explains the exceptional robustness to noise of dimension witnessing in this scenario, and it provides a clear contrast to the case in which this resource is available.

A. Dimension estimates

The fact that $\mathcal{C}'_d$ and $\mathcal{Q}'_d$ are convex sets allows us to separate each set from its complement by linear functionals on the behaviors. This gives rise to the so-called linear dimension witnesses [8,9]. The case of $\mathcal{C}_d$ and $\mathcal{Q}_d$ was recently addressed in [11], which obtained some nonlinear dimension witness for these nonconvex sets. Specifically, they consider the scenario in which $|\mathcal{X}| = 2|\mathcal{Y}| = 2k$ and show that the $k \times k$ matrix W_k with entries $W_k(i,j) = P(0|2j-1,i) - P(0|2j,i)$ $(i,j = 1,\dots,k)$ is such that $\det W_k = 0$ for all behaviors in $\mathcal{C}_d$ $(\mathcal{Q}_d)$ with $d \leqslant k$ $(d \leqslant \sqrt{k})$. Thus, the determinant of W_k being nonzero allows one to establish nontrivial lower bounds on the required dimensionality both in the classical and quantum cases.

In the following, we obtain more refined estimates. To do so, and to deal with behaviors, we will arrange the array of numbers given by $\mathbf{P}$ into a matrix $P \in \mathbb{R}^{|\mathcal{X}|\times 2|\mathcal{Y}|}$ according to the rule

$$P = \sum_{bxy} P(b|xy)|x\rangle\langle yb|, \tag{5}$$

where in the standard notation of quantum mechanics $|yb\rangle = |y\rangle \otimes |b\rangle$ and $\{|y\rangle\}$ denotes the computational basis of $\mathbb{R}^{|\mathcal{Y}|}$, and similarly for the other alphabet elements. In other words, P takes the form

$$P = \begin{pmatrix} P(0|11) & P(1|11) & P(0|12) & P(1|12) & \cdots & P(0|1|\mathcal{Y}|) & P(1|1|\mathcal{Y}|) \\ P(0|21) & P(1|21) & P(0|22) & P(1|22) & \cdots & P(0|2|\mathcal{Y}|) & P(1|2|\mathcal{Y}|) \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ P(0||\mathcal{X}|1) & P(1||\mathcal{X}|1) & P(0||\mathcal{X}|2) & P(1||\mathcal{X}|2) & \cdots & P(0||\mathcal{X}||\mathcal{Y}|) & P(1||\mathcal{X}||\mathcal{Y}|) \end{pmatrix}. \tag{6}$$

Now, in the case $\mathbf{P} \in \mathcal{C}_d$, we can make the following immediate observation. Since the behavior is given by Eq. (2), we have that

$$P = \sum_{m=1}^{d} \left(\sum_x s(m|x)|x\rangle \right) \left(\sum_{by} t(b|my)\langle yb| \right) = \sum_{m=1}^{d} u_m v_m^T \tag{7}$$

for some real (actually non-negative) vectors $\{u_m\}$ and $\{v_m\}$ of size $|\mathcal{X}|$ and $2|\mathcal{Y}|$, respectively. Thus, we clearly see that if $\mathbf{P} \in \mathcal{C}_d$, then $\operatorname{rank} P \leqslant d$. On the other hand, if the behavior is quantum, Eq. (3) tells us that its entries are given by the Hilbert-Schmidt inner product of pairs of Hermitian matrices of size $d \times d$. This set of matrices forms a subspace that is isomorphic to $\mathbb{R}^{d^2}$. Thus, there exists vectors $\{w_x\}$ and $\{t_{by}\}$ in this space such that $P(b|xy) = w_x^T t_{by}$. Therefore, we now have that if $\mathbf{P} \in \mathcal{Q}_d$, then $\operatorname{rank} P \leqslant d^2$.

Observation 1. If $\mathbf{P} \in \mathcal{C}_d$, then $d \geqslant \operatorname{rank} P$ while if $\mathbf{P} \in \mathcal{Q}_d$, then $d \geqslant \sqrt{\operatorname{rank} P}$.

These simple observations generalize the previous aforementioned result of [11] in several ways. First, we do not put any constraint on the size of the alphabets $\mathcal{X}$ and $\mathcal{Y}$. Second, the estimates are finer since we do not rely on a matrix being of full rank or not, but the bound is sensitive to the different possible values of the rank. It should be mentioned that the result of Bowles *et al.* based on the W matrix is also obtained by showing that the entries of this matrix are given by the inner product of a set of vectors. Thus, rank estimates are also possible in this case. In more detail, one obtains that $d \geqslant \operatorname{rank} W_k + 1$ if $\mathbf{P} \in \mathcal{C}_d$ and $d \geqslant \sqrt{\operatorname{rank} W_k + 1}$ if $\mathbf{P} \in \mathcal{Q}_d$. Hence, it comes as a natural question whether it is better to use P or W to get the strongest estimate. Notice that, in the $|\mathcal{X}| = 2|\mathcal{Y}| = 2k$ setting, the maximal possible rank of P is $k+1$ [this is because several columns are surely linearly dependent due to the condition $\sum_b P(b|xy) = 1\ \forall\, x,y$] while for W it is obviously k. Thus, in the case of maximal rank both approaches yield equal estimates. In the Appendix, we show that $\operatorname{rank} W \leqslant \operatorname{rank} P$. Thus, this suggests that it is generally better to use P. In fact, it can only be worse in cases for which rank $W =$ rank P. However, this can only lead to a difference of 1 in the estimate (and not always in the quantum case since it holds for many natural numbers n that $\lceil\sqrt{n+1}\rceil = \lceil\sqrt{n}\rceil$).

It is worth mentioning that stronger bounds on d can be placed by using generalizations of the rank [17]. Recent literature has established an intimate relation between the non-negative rank and the classical dimension and the positive-semidefinite rank and the quantum dimension in similar scenarios [17,18]. It is immediately apparent that the non-negative rank of P and the positive-semidefinite rank of P are lower bounds for d in the classical and quantum case, respectively, in the scenario considered here. Reference [13] also offers related strategies to bound the dimension. However, we stick here to the weaker rank estimates because they seem much easier to use. In fact, we do not know efficient algorithms to compute these other notions of the rank [17,19].

B. Behaviors more expensive quantumly than classically together with shared randomness

In this subsection, we will show that $\mathcal{C}'_d \nsubseteq \mathcal{Q}_d$ with an arbitrary separation: there exist behaviors requiring a constant amount of classical communication together with shared randomness (actually just one bit) while the necessary quantum communication increases at least as $\sqrt{|\mathcal{Y}|}$. In more detail, we will consider general scenarios such that $|\mathcal{Y}| = k$ and $|\mathcal{X}| = m \geqslant k+1$ (this condition is only to make it possible that the matrices of behaviors can have the largest possible rank, $k+1$) and we will construct behaviors $\mathbf{P}_k$ for any natural k that they all belong to $\mathcal{C}'_2$ but cannot belong to $\mathcal{Q}_{\lfloor\sqrt{k}\rfloor}$. The idea is to mix a sufficient number of behaviors in $\mathcal{C}_2$ such that the corresponding matrix P_k has its rank as large as possible so that Observation 1 leads us to conclude that $d \geqslant \sqrt{k+1}$ in order for $\mathbf{P}_k \in \mathcal{Q}_d$ to hold.

An example of such a construction goes as follows. Here and throughout this paper, we will denote by $e_i^{(n)}$ the vector of $\mathbb{R}^n$ that has zeros everywhere except a 1 in the ith entry. Consider the behavior $\mathbf{D}_1$ in the aforementioned setting whose $m \times 2k$ matrix is given by (to ease the notation we drop the dependence on k)

$$D_1 = \begin{pmatrix} \left(e_2^{(2)}\right)^T & \left(e_1^{(2)}\right)^T & \cdots & \left(e_1^{(2)}\right)^T \\ \left(e_1^{(2)}\right)^T & \left(e_1^{(2)}\right)^T & \cdots & \left(e_1^{(2)}\right)^T \\ \vdots & \vdots & \ddots & \vdots \\ \left(e_1^{(2)}\right)^T & \left(e_1^{(2)}\right)^T & \cdots & \left(e_1^{(2)}\right)^T \\ \left(e_1^{(2)}\right)^T & \left(e_1^{(2)}\right)^T & \cdots & \left(e_1^{(2)}\right)^T \\ \vdots & \vdots & \vdots & \vdots \\ \left(e_1^{(2)}\right)^T & \left(e_1^{(2)}\right)^T & \cdots & \left(e_1^{(2)}\right)^T \end{pmatrix}$$

$$= \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \left[\left(e_1^{(2)}\right)^T \quad \cdots \quad \left(e_1^{(2)}\right)^T \right]$$

$$+ e_1^{(m)} \left[\left(e_2^{(2)}\right)^T \quad \left(e_1^{(2)}\right)^T \quad \cdots \quad \left(e_1^{(2)}\right)^T \right]. \tag{8}$$

The second way to write D_1 shows clearly that $\mathbf{D}_1 \in \mathcal{C}_2$: B outputs all the time $b = 0$ except maybe when $y = 1$ depending on the bit sent by A, her action relying on whether she gets the input $x = 1$ or any other. We can similarly define the behaviors $\mathbf{D}_i$ ($i = 1, \ldots, k$) whose matrices are all made by 1×2 blocks given by $(e_1^{(2)})^T$ except at the position (i,i) where it is given by $(e_2^{(2)})^T$. By the same arguments as above, we have that $\mathbf{D}_i \in \mathcal{C}_2\ \forall\ i$. It is easy to see that any nontrivial mixture of all these behaviors has maximal rank. Taking, for instance,

$$\mathbf{P}_k = \sum_{i=1}^{k} \frac{1}{k} \mathbf{D}_i, \tag{9}$$

we find that

$$P_k = \begin{pmatrix} c_k^T & \left(e_1^{(2)}\right)^T & \cdots & \left(e_1^{(2)}\right)^T \\ \left(e_1^{(2)}\right)^T & c_k^T & \cdots & \left(e_1^{(2)}\right)^T \\ \vdots & \vdots & \ddots & \vdots \\ \left(e_1^{(2)}\right)^T & \left(e_1^{(2)}\right)^T & \cdots & c_k^T \\ \left(e_1^{(2)}\right)^T & \left(e_1^{(2)}\right)^T & \cdots & \left(e_1^{(2)}\right)^T \\ \vdots & \vdots & \vdots & \vdots \\ \left(e_1^{(2)}\right)^T & \left(e_1^{(2)}\right)^T & \cdots & \left(e_1^{(2)}\right)^T \end{pmatrix}, \tag{10}$$

where

$$c_k^T = (1 - 1/k \quad 1/k). \tag{11}$$

Now, one can see that P_k has the largest possible rank, i.e., rank $P_k = k+1$. This is because, on the one hand, all even columns are clearly linearly independent, i.e., $\text{col}_{2j}(P_k) = e_j^{(m)}/k$ for $j = 1,2,\ldots,k$. On the other hand, if we add to this set any other odd column, the set remains linearly independent because this column has nonzero entries where all the others have a zero entry [from the $(k+1)$th entry to the mth]. Thus, using Observation 1, we finally obtain that if $\mathbf{P}_k \in \mathcal{Q}_d$ it must hold that $d \geqslant \sqrt{k+1}$ while, by construction, $\mathbf{P}_k \in \mathcal{C}'_2\ \forall\ k$. In passing, since obviously $\mathbf{P}_k \in \mathcal{Q}'_2\ \forall\ k$, this also shows the nonconvexity of $\mathcal{Q}_d$ (see also [9] and [13]).

C. Behaviors cheaper quantumly than classically together with shared randomness

The results of [11] discussed before show that there is a quadratic gap between $\mathcal{C}$ and $\mathcal{Q}$. Interestingly, this gap can be seen to be exponential and extended to $\mathcal{C}'$. Testing classical and quantum dimensions in the prepare-and-measure scenario, is intimately connected to the field of communication complexity when restricted to the scenario of one-way communication complexity. In fact, this setting is the same, with the only difference being that it is task-oriented. In this case, under the same restrictions, A and B now have the goal of evaluating with high probability of success a binary function f of their inputs that is known to both of them. That is, their strategies should aim at preparing behaviors $P(b|xy)$ for which the result $b = f(x,y)$ is much more likely than $b \neq f(x,y)$. The field of communication complexity studies what is the least amount of communication (from A to B in the one-way case) necessary to evaluate different functions. The possible benefits of using quantum communication over classical communication have been studied extensively in recent years, and there are several scenarios for which it is known that certain functions can be evaluated with a given probability of success requiring exponentially less communication in the quantum case than in the case of classical messages [4]. Interestingly, the one-way scenario is not an exception, and Refs. [20,21] provide instances of this situation for the case of partial functions. In more detail, [21] considers a function, f_P, for which A receives an n-bit string x (i.e., $|\mathcal{X}| = 2^n$) and B an $n \times n$ permutation matrix M (i.e., $|\mathcal{Y}| = n!$). The goal is to output 1 if $Mx = x$ and 0 if Mx and x are sufficiently different (in a precise way that is irrelevant here). This is an example of a partial function or a function with a promise; A and B are guaranteed to receive a strict subset of the inputs x and y (those for which any of the above conditions hold). In [21], it is shown that a quantum strategy solves this function using $O(\log n)$ qubits of communication [i.e., $d = O(n)$] while there cannot exist any classical strategy solving f_P using less than of the order of $n^{7/16}$ bits. This immediately implies that $\mathcal{Q}_d \nsubseteq \mathcal{C}'_d$. To see this, consider any behavior corresponding to the aforementioned quantum strategies that solve f_P. It must then be that $\mathbf{P}_n \in \mathcal{Q}_d$ for some $d = O(n)$. However, it cannot be that $\mathbf{P}_n \in \mathcal{C}'_d$ as this would be in contradiction to the result of [21]. Indeed, any behavior in $\mathcal{C}'_d$ cannot have the same entries as $\mathbf{P}_n$ over the subset of promised inputs x and y since it could then be used to solve f_P. Moreover, it must be that $\mathbf{P}_n \in \mathcal{C}'_{d'}$ with d' scaling at least as $2^{n^{7/16}}$. Interestingly, there is roughly no difference between $\mathcal{C}_d$ and $\mathcal{C}'_d$ for the evaluation of functions. Newman's theorem [22] shows in the one-way communication scenario that classical strategies with shared randomness that solve some function can be turned into successful strategies without shared randomness with just a logarithmic overhead. Thus, if there exists some exponential gap for the solution of a function with quantum and classical resources, it must persist if we allow classical resources supplemented with shared randomness.

In light of the communication complexity, the reader might wonder whether the results of the previous section showing behaviors that required overwhelmingly more quantum communication to be prepared than classical communication together with shared randomness could be used to devise functions whose solution has a similar gap, i.e., functions that are at least exponentially cheaper to solve classically if shared randomness is allowed than quantumly. However, Newman's theorem forbids this possibility. With regard to the evaluation of functions, the differences with and without shared randomness can be at most logarithmic in the one-way scenario even if one just uses classical messages.

D. Structure of $\mathcal{C}_d$ and $\mathcal{Q}_d$ and robustness of dimensionality detection

As discussed in the Introduction, the prepare-and-measure scenario was introduced to certify the dimension of uncharacterized physical systems in a device-independent way, i.e., based solely on the observed statistics and without any assumption on the internal working of the devices used. In this context, any condition expressed in terms of the observed behavior that guarantees that A and B exchange physical systems of at least a certain dimension is usually referred to as a dimension witness. The rank estimates introduced in Sec. II A are therefore an example of such an object. In practice, the measurement device cannot be perfectly isolated from external noise introducing errors in the experimentally reconstructed behavior, which can make it less dimensional. The robustness of a dimension witness characterizes its noise tolerance in these scenarios and plays a crucial role in dimensionality certification. Although not strictly necessary, in this setting a natural assumption is that the preparing and measuring devices are uncorrelated (i.e., the preparer and the measurer are not maliciously conspiring to fool the certifier) and, hence, in this case one takes that shared randomness is not available. Thus, the problem here boils down to identifying what is the smallest d such that $\mathbf{P} \in \mathcal{C}_d$ or $\mathbf{P} \in \mathcal{Q}_d$. This was the motivation of [11] to introduce the dimension witness based on the determinant of the matrix W_k that we have reviewed in Sec. II A. It was observed there that these witnesses are extraordinarily robust tolerating arbitrary amounts of noise. In this section, we investigate the structure of the sets $\mathcal{C}_d$ and $\mathcal{Q}_d$ from this point of view, and we find reasons for this exceptional robustness. Low-dimensional sets are negligibly small in the set of all possible behaviors: they are nowhere dense and have measure zero. Hence, very contrived forms of noise are required to drastically reduce the dimension. This is in sharp contrast to the case in which shared randomness is available since, as we also discuss here, the sets $\mathcal{C}'_d$ and $\mathcal{Q}'_d$ are not negligible $\forall\, d \geqslant 2$. We finish this section by observing that rank estimates are also extremely robust under any physically reasonable form of noise.

Notice that evaluating the rank of a matrix is an ill-conditioned problem. Due to the estimates presented in Sec. II A, this indicates that small perturbations of a behavior could increase considerably the required dimension to prepare it. Furthermore, for general matrices it is well known that lower-rank matrices are of measure zero and nowhere dense among matrices of higher rank. This suggests that lower-dimensional sets of behaviors might be negligible. We formalize this in the following:

Theorem 1. In every scenario with $|\mathcal{Y}| = k$ and $|\mathcal{X}| = m \geqslant k + 1$, the sets $\mathcal{C}_d$ with $d < k + 1$ and $\mathcal{Q}_d$ with $d < \sqrt{k+1}$ have measure zero and are nowhere dense in the set of all possible behaviors $\mathcal{C}_m = \mathcal{Q}_m$.

Proof. We will show that the set of rank-deficient behaviors (i.e., rank $P < k + 1$), which we will denote by S, is of measure zero and nowhere dense in $\mathcal{C}_m = \mathcal{Q}_m$. Since in the classical case we have seen that rank $P \leqslant d$, when $d < k + 1$ we have that $\mathcal{C}_d \subset S$ and the result follows. The same applies to the quantum case. The proof of the claim for rank-deficient behaviors follows basically in a straightforward manner the analogous case for general matrices.

Let us first show that S has measure zero. Notice that the set of behaviors is a subset of $\mathbb{R}^{km}$ determined by specifying an arbitrary collection of values $0 \leqslant P(0|xy) \leqslant 1\ \forall\, x,y$ and it has nonzero Lebesgue measure. When $P \in S$, this additionally imposes that the determinants of certain square submatrices vanish, which is a polynomial in the matrix entries, i.e., in the $\{P(0|xy)\}$. However, the zero set of a polynomial must have measure zero (unless it is the zero polynomial). Hence, S has Lebesgue measure equal to zero.

Let us now see that S is nowhere dense. For this we have to see that the closure of S, $\overline{S}$, has an empty interior. Since we are dealing with finite-dimensional matrices, for these topological considerations we can take any matrix norm $||\cdot||$. First of all, it is useful to notice that S is closed. This is because S is characterized by the determinant of all $(k+1)\times(k+1)$ submatrices of P being zero. Hence, the set is the preimage of a closed set under a continuous map (the determinant is a polynomial of the matrix entries) and it is therefore closed. Now, since $S=\overline{S}$, we just need to check that S has an empty interior. Clearly, general rank-deficient matrices can always be approximated by full-rank matrices. It remains to see the same, being careful that the full-rank approximation can be chosen to be a behavior as well. For this, take any $P\in S$ and define for any $\epsilon\in[0,1]$,

$$P_\epsilon=(1-\epsilon)P+\epsilon Q, \tag{12}$$

where

$$Q=\sum_{j=1}^{k}e_j^{(m)}v_j^T+\left(\sum_{j=k+1}^{m}e_j^{(m)}\right)v_{k+1}^T \tag{13}$$

with the $2k$-dimensional vectors $\{v_j\}$ defined by

$$v_1=\begin{pmatrix}e_2^{(2)}\\ e_1^{(2)}\\ \vdots\\ e_1^{(2)}\end{pmatrix},\ v_2=\begin{pmatrix}e_1^{(2)}\\ e_2^{(2)}\\ e_1^{(2)}\\ \vdots\\ e_1^{(2)}\end{pmatrix},\ldots,\ v_k=\begin{pmatrix}e_1^{(2)}\\ \vdots\\ e_1^{(2)}\\ e_2^{(2)}\end{pmatrix},\ v_{k+1}=\begin{pmatrix}e_1^{(2)}\\ \vdots\\ e_1^{(2)}\\ e_1^{(2)}\end{pmatrix}. \tag{14}$$

Notice that Q is a valid behavior and therefore so is $P_\epsilon\ \forall\,\epsilon$. It is important to notice that the $\{v_j\}$ are linearly independent (LI) vectors. That the first k of them are LI is clear because each of them has a nonzero entry where all the others are zero. To see that adding v_{k+1} to the set keeps it LI, we notice the following. This vector has its second entry equal to zero, which is the case for all of the others except v_1. Thus if v_{k+1} could be obtained as a linear combination of the other vectors, the weight of v_1 has to be zero. Iterating this argument for all even entries of v_{k+1}, we obtain the claim. Now, because P and Q are behaviors and the $\{v_j\}$ are non-negative vectors, we have that Pv_j and Qv_j are non-negative vectors too $\forall\ j$. Moreover, by construction $Qv_j\neq 0\ \forall\ j$ and, therefore, $P_\epsilon v_j\neq 0\ \forall\ j$ and all $\epsilon>0$. Since the $\{v_j\}$ are LI, this implies that $\dim\ker P_\epsilon\leqslant k-1$ and, hence, given that the dimension of the kernel and the rank must add up to the number of columns $2k$, rank $P_\epsilon=k+1\ \forall\ \epsilon>0$, i.e., $P_\epsilon\notin S\ \forall\ \epsilon>0$. Thus, we finally see that $\forall\ \delta>0$ and $\forall\ P\in S, \exists P'\notin S$ such that $||P-P'||<\delta$ (for this it suffices to take $P'=P_\epsilon$ with ϵ sufficiently small). Hence, S does not contain any nonempty open set, i.e., it has an empty interior, as we wanted to prove [23]. ■

Thus, the sets $\mathcal{C}_d$ with $d<k+1$ and $\mathcal{Q}_d$ with $d<\sqrt{k+1}$ are negligibly small and $\mathcal{C}_m\backslash\mathcal{C}_d$ and $\mathcal{Q}_m\backslash\mathcal{Q}_d$ have full measure and a dense interior. It might be that the conditions $d<k+1$ and $d<\sqrt{k+1}$ are an artifact of the proof due to the rank estimates and that the above claim can be extended to larger values of $d<m$. Moreover, it could be that $\mathcal{C}_{d-1}$ ($\mathcal{Q}_{d-1}$) has zero measure and is nowhere dense in $\mathcal{C}_d$ ($\mathcal{Q}_d$) for all d such that $2\leqslant d\leqslant m$.

The result of Theorem 1 is in sharp contrast to the case when shared randomness is available. The sets $\mathcal{C}'_d$ and $\mathcal{Q}'_d$ are not negligible in the set of all possible behaviors $\forall\ d\geqslant 2$, as we show below. Thus, this negligibility property provides a crucial difference for DIDW in the presence (or lack thereof) of shared randomness.

Proposition 1. In every scenario with $|\mathcal{Y}|=k$ and $|\mathcal{X}|=m\geqslant k+1$, the sets $\mathcal{C}'_d$ and $\mathcal{Q}'_d\ \forall\ d\geqslant 2$ have nonzero measure and are not nowhere dense in the set of all possible behaviors $\mathcal{C}_m=\mathcal{Q}_m$.

Proof. First of all, notice that $\mathcal{C}'_2\subset\mathcal{C}'_d$ $(d>2)$ and $\mathcal{C}'_2\subset\mathcal{Q}'_d$ $(d\geqslant 2)$. Hence, it suffices to prove the claim for $\mathcal{C}'_2$. Notice, moreover, that both $\mathcal{C}'_2$ and $\mathcal{C}_m$ are (convex) polytopes [8]. We have discussed in the proof of Theorem 1 that $\dim\mathcal{C}_m=km$. Therefore, one only needs to see that $\dim\mathcal{C}'_2=km$ too. For this, we have to find $km+1$ points in $\mathcal{C}'_2$ that are affinely independent. We give such a construction in the following. Notice that, as in Eq. (8), behaviors whose matrix has all rows except one equal are in $\mathcal{C}_2\subset\mathcal{C}'_2$. In analogy with Eq. (8), we denote then by $\{\mathbf{D}_{ij}\}$ $(1\leqslant i\leqslant m,\ 1\leqslant j\leqslant k)$ the behaviors whose $m\times 2k$ matrices have 1×2 blocks equal to $(e_1^{(2)})^T$, except for the block at position (i,j), which is equal to $(e_2^{(2)})^T$. We will also consider the behavior $\mathbf{D}_0$, whose matrix has all blocks equal to $(e_1^{(2)})^T$. Arguing in a similar manner as with the set of vectors of Eq. (14), it is easy to see that the $km+1$ points $\{\mathbf{D}_0,\mathbf{D}_{ij}\}$ (which all happen to be vertices of the polytope $\mathcal{C}'_2$) are LI and, hence, affinely independent. ■

Theorem 1 should not be interpreted as a physical impossibility of preparing low-dimensional behaviors. If the setting limits the amount of communication A can send to B, we are bound to observe such a low-dimensional behavior. What we learn from it is that if the underlying dimension is sufficiently large, and B's measurements are subject to noise, it must have a very particular form in order to drastically reduce the dimension of the observed behavior. Actually, for any behavior P one can see that under any physically reasonable form of noise P_n, the observed behavior,

$$P_\eta=\eta P+(1-\eta)P_n\quad(\eta\in[0,1]), \tag{15}$$

maintains the rank $\forall\ \eta>0$. That is, rank $P_\eta=$ rank $P\ \forall\ \eta>0$ and the rank estimates are completely robust against noise. Thus, if P can be certified to have dimension $d\geqslant k+1$ in the classical case (or $d\geqslant\sqrt{k+1}$ in the quantum case) by means of its rank, this will not change for its noisy version (unless in the extremal case of full noise $\eta=0$).

We finish this section by proving the above claim that rank $P_\eta=$ rank $P\ \ \forall\ \eta>0$. First we need to discuss what P_n can be. The most reasonable and general form for the noise is that it is independent of A's input, i.e., $P_n(b|xy)=P_n(b|y)\ \forall\ b,x,y$. This is because the errors only occur in the measurement process carried out by B, and A has no control over it to affect the encoding of her message. Thus, $P_n=\mathbf{1}v^T$, where $\mathbf{1}$ is a column vector in $\mathbb{R}^{|\mathcal{X}|}$ with all entries equal to 1 and $v=\sum_{b,y}P_n(b|y)|yb\rangle$. This implies that the noisy behavior P_η is given by a rank-1 perturbation to P. This means that the rank of P and P_η can at most differ by 1. However, we will see now that they are actually equal (as long as $\eta\neq 0$). Since the image of P_n is spanned by $\mathbf{1}$, rank $P_\eta=$ rank $P-1$ can only hold if there exists some vector u such that $Pu\propto\mathbf{1}$ in such a way that $P_\eta u = 0$. Clearly, this vector must be of the form $u = \alpha\mathbf{1} + w$, where $\alpha \in \mathbb{R}$ and $w \in \ker P$ because $P\mathbf{1} = |\mathcal{Y}|\mathbf{1}$ for any behavior. However, since we also have that $P_\eta\mathbf{1} = |\mathcal{Y}|\mathbf{1}$, if it is possible to have a vector in the kernel of P_η that is not in the kernel of P, $P_\eta u = [\alpha|\mathcal{Y}| + (1-\eta)(v^T w)]\mathbf{1} = 0$, it must be such that $v^T w \neq 0$. In this case, we then have that $P_\eta w \neq 0$, that is, we can also find a vector such that it is in the kernel of P but not in that of P_η. This shows that $\operatorname{rank} P_\eta \neq \operatorname{rank} P - 1$ when $\eta \neq 0$. A similar argument shows that $\operatorname{rank} P \neq \operatorname{rank} P_\eta - 1$ when $\eta \neq 0$. Hence, we obtain the desired result.

III. BELL SCENARIOS

The first scheme [5] that was proposed to test in a device-independent way the dimension of a quantum system used the Bell scenario of quantum nonlocality [24]. In this setting, we also have two parties A and B, but they cannot communicate in this case. However, both A and B perform measurements dependent, respectively, on some inputs x and y on a bipartite quantum state ρ they share. Each measurement leads to outputs a and b for A and B, respectively. This scenario can also be catalogued according to the (finite) size of the input and output alphabets $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{A}$, and $\mathcal{B}$. In the following, we will consider that $|\mathcal{X}| = |\mathcal{Y}| = m$ and $|\mathcal{A}| = |\mathcal{B}| = n$, and we will refer to (m,n) scenarios. Similarly to the prepare-and-measure setting, the object to which we have access to here is the set of conditional probabilities of obtaining the outputs (a,b) given the choice of inputs (x,y), $P(ab|xy)$. We will use again the term behavior to refer to this collection of numbers. Obviously, it must hold that $P(ab|xy) \geqslant 0\ \forall\ a,b,x,y$ and $\sum_{a,b} P(ab|xy) = 1\ \forall\, x,y$. All behaviors attainable classically (together with shared randomness) satisfy

$$P(ab|xy) = \sum_\lambda p_\lambda P_\lambda^A(a|x) P_\lambda^B(b|y)\ \forall\, a,b,x,y \tag{16}$$

for some convex weights $\{p_\lambda\}$ and sets of conditional probabilities $\{P_\lambda^A\}$ and $\{P_\lambda^B\}$. Alternatively, the set of all such behaviors, the so-called local set $\mathcal{L}$, can be characterized to be the convex hull of all local deterministic behaviors (LDBs). The LDBs correspond to all possible deterministic uncorrelated behaviors, i.e., to those of the form $D(ab|xy) = \delta_{a,f(x)}\delta_{b,g(y)}$, where f is any function mapping elements of $\mathcal{X}$ to $\mathcal{A}$ and similarly for g. That is, for every party a unique output occurs with probability 1 for every choice of input. Given a scenario, there is a finite number (actually n^{2m}) of possible LDBs, and hence $\mathcal{L}$ is a polytope. We will denote by $\mathcal{Q}$ here the set of all behaviors that can be obtained by performing measurements on bipartite quantum states ρ_{AB}, i.e.,

$$P(ab|xy) = \operatorname{tr}\!\left(\rho_{AB} E_a^x \otimes F_b^y\right) \tag{17}$$

for some positive-semidefinite operators $\{E_a^x, F_b^y\}$ such that $\sum_a E_a^x$ and $\sum_b F_b^y$ equal the identity in each party's Hilbert space $\forall\ x,y$. The celebrated conclusion of Bell's theorem is that $\mathcal{L} \subsetneq \mathcal{Q}$.

For a fixed (m,n) setting, one can now define $\mathcal{Q}_d$ here as the set of all behaviors in $\mathcal{Q}$ that are obtainable by a quantum state such that $\min_{X=A,B}(\dim \operatorname{supp} \rho_X) \leqslant d$, i.e., all behaviors obtainable by measuring quantum states of minimum local dimension at most d. As discussed in the Introduction, the characterization of these sets has raised a considerable amount of attention from the point of view of both dimensionality certification and semi-device-independent quantum information protocols. It is interesting to notice that, when the dimension of the physical system is not restricted, shared randomness is not a resource. Its availability is irrelevant when it comes to which behaviors can be observed with quantum preparations because $\mathcal{Q} = \mathcal{Q}_\infty$ is a convex set [24]. However, this is not the case when the physical dimension of the underlying system plays a role. If shared randomness is freely available, this leads us to consider the sets $\operatorname{Conv}(\mathcal{Q}_d)$. Thus, it is interesting to explore the differences given by whether this resource is available or not and, in particular, whether $\mathcal{Q}_d \neq \operatorname{Conv}(\mathcal{Q}_d)$ in general. In fact, it is easy to see that $\mathcal{Q}_1$ is nonconvex in every scenario [15]. By definition, this set can only include uncorrelated behaviors. However, the convex hull of all LDBs gives rise to the full local polytope $\mathcal{L}$, and it is well known that this set includes correlated behaviors. Reference [14] was the first to observe that the sets $\mathcal{Q}_d$ need not be convex in general. In more detail, it is shown therein that in the scenarios $(m,2)$ with m even, every set $\mathcal{Q}_d$ such that $d < \sqrt{m+1}$ is nonconvex. In particular, this implies that $\mathcal{Q}_2$ is nonconvex already in the reasonably simple scenario (4,2). This was observed to be the case in [15] even in the simplest possible scenario (2,2) [25]. In the following, we prove the nonconvexity properties of the sets $\mathcal{Q}_d$ in the general scenario (m,n). In analogy with Sec. II, we will finish this section by proving that the sets $\mathcal{Q}_d$ are negligible in $\mathcal{Q}$ when the dimension is sufficiently small, a property that is not true for $\operatorname{Conv}(\mathcal{Q}_d)$.

A. Dimension estimates

To prove the nonconvexity and negligibility of $\mathcal{Q}_d$ in (m,n) scenarios, we first derive dimension estimates. In analogy with the previous section, and adapting the techniques of [14], we will obtain lower bounds on the quantum dimension in terms of the rank of some matrix associated with the behavior. More explicitly, for every (m,n) scenario we will arrange every behavior $\mathbf{P}$ to form the $mn \times mn$ real matrix,

$$P = \sum_{abxy} P(ab|xy)|xa\rangle\langle yb|, \tag{18}$$

where in the standard notation of quantum mechanics $|xa\rangle = |x\rangle \otimes |a\rangle$ and $\{|x\rangle\}$ denotes the computational basis of $\mathbb{R}^m$ and similarly for the other alphabet elements. Thus, P can be partitioned as a block matrix,

$$P = \begin{pmatrix} P_{11} & \cdots & P_{1m} \\ \vdots & \ddots & \vdots \\ P_{m1} & \cdots & P_{mm} \end{pmatrix} \in \mathbb{R}^{mn\times mn}$$

with blocks

$$P_{xy} = \begin{pmatrix} P(11|xy) & \cdots & P(1n|xy) \\ \vdots & \ddots & \vdots \\ P(n1|xy) & \cdots & P(nn|xy) \end{pmatrix} \in \mathbb{R}^{n\times n}.$$

It will be relevant in the next subsection to note that the matrix associated with LDBs $D(ab|xy) = \delta_{a,f(x)}\delta_{b,g(y)}$ is of rank 1, i.e.,

$$
\begin{aligned}
D &= \left(\sum_{ax} \delta_{a,f(x)} |xa\rangle\right)\left(\sum_{by} \delta_{b,g(y)} \langle yb|\right)\\
&= \left(\sum_{x} |xf(x)\rangle\right)\left(\sum_{y} \langle yg(y)|\right). \qquad (19)
\end{aligned}
$$

Suppose now that $\mathbf{P} \in \mathcal{Q}_d$ and that the optimal quantum state is such that $d = \dim \operatorname{supp} \rho_A$. This means that the operators $\{E_a^x\}$ act on $\mathbb{C}^d$. Since they are Hermitian, they must then belong to a real vector space of dimension d^2 and, thus, at most d^2 of them can be linearly independent. In other words, we can express all the $\{E_a^x\}$ as real linear combinations of a fixed set of d^2 Hermitian operators [e.g., the identity and the generators of SU(d)]. By linearity of the trace, this means that there are at most d^2 linearly independent rows in the matrix P and hence $\operatorname{rank} P \leqslant d^2$. If it was the case that $d = \dim \operatorname{supp} \rho_B$, then we can make the same reasoning with the operators $\{F_b^y\}$ and the columns of P, arriving again at the same conclusion that $d \geqslant \sqrt{\operatorname{rank} P}$.

Observation 2. If $\mathbf{P} \in \mathcal{Q}_d$, then $d \geqslant \sqrt{\operatorname{rank} P}$.

It is important to notice for the following that the largest rank a matrix of a behavior can attain is $mn - m + 1$. This is because quantum behaviors must obey the no-signaling constraints

$$
\begin{aligned}
&\sum_b P(ab|xy) = \sum_b P(ab|xy'), \quad \forall\, a,x,y,y',\\
&\sum_a P(ab|xy) = \sum_a P(ab|x'y), \quad \forall\, b,x,x',y. \qquad (20)
\end{aligned}
$$

The set of all behaviors fulfilling these conditions will be denoted by $\mathcal{NS}$, which is also a polytope.

It should be mentioned that [12] already provides the means to obtain lower bounds for the dimension and, actually, one can also use for these matters the positive-semidefinite rank of P. However, as in the previous section, although weaker, rank estimates turn out to be more easily applied.

B. Nonconvexity of $\mathcal{Q}_d$

To prove the nonconvexity of $\mathcal{Q}_d$, we will use the following strategy. We will construct a local behavior $\mathbf{L}$ such that L has the largest possible rank $mn - m + 1$. Through Observation 2 this implies that $\mathbf{L} \notin \mathcal{Q}_d$ if $d < \sqrt{\operatorname{rank} L}$. However, since all local behaviors can be written as a convex combination of LDBs, it must hold that $\mathbf{L} \in \operatorname{Conv}(\mathcal{Q}_1) \subset \operatorname{Conv}(\mathcal{Q}_d)\ \forall\, d$. Thus, for sufficiently small values of d, $\mathcal{Q}_d$ cannot be convex.

Lemma 1. In every (m,n) scenario there exists $\mathbf{L} \in \mathcal{L}$ such that $\operatorname{rank} L = mn - m + 1$.

Proof. Consider the set of $n - 1$ vectors of size mn given by (we use here the same notation for the $\{e_i^{(n)}\}$ as in Sec. II)

$$
v_i^{(1)} = \begin{pmatrix} e_i^{(n)} \\ e_1^{(n)} \\ \vdots \\ e_1^{(n)} \end{pmatrix}, \quad i = 2,3,\ldots,n. \qquad (21)
$$

We will also consider similar sets of the same cardinality,

$$
\{v_i^{(2)}\} = \left\{\begin{pmatrix} e_1^{(n)} \\ e_i^{(n)} \\ e_1^{(n)} \\ \vdots \\ e_1^{(n)} \end{pmatrix}\right\}, \ldots, \{v_i^{(m)}\} = \left\{\begin{pmatrix} e_1^{(n)} \\ \vdots \\ e_1^{(n)} \\ e_i^{(n)} \end{pmatrix}\right\}. \qquad (22)
$$

Notice now that the set $\{v_i^{(j)}\}$ $(i = 2,\ldots,n,\ j = 1,\ldots,m)$ contains $mn - m$ LI vectors. This can be seen by noticing that each vector has a nonzero entry where all the others are zero. Notice, moreover, that if we add the vector

$$
v^{(0)} = \begin{pmatrix} e_1^{(n)} \\ e_1^{(n)} \\ \vdots \\ e_1^{(n)} \end{pmatrix} \qquad (23)
$$

to this set, the vectors are still LI [cf. the reasoning after Eq. (14)]. Finally, notice that the matrices

$$
L_0 = v^{(0)}\left(v^{(0)}\right)^T, \quad \{L_{ij}\} = \left\{v_i^{(j)}\left(v_i^{(j)}\right)^T\right\} \qquad (24)
$$

clearly correspond to LDBs in the (m,n) scenario. Hence, the matrix

$$
L = \frac{1}{mn - m + 1}\left(L_0 + \sum_{ij} L_{ij}\right) \qquad (25)
$$

corresponds to a local behavior and has the desired property that $\operatorname{rank} L = mn - m + 1$. This is because by construction $Lv^{(0)} \neq 0$ and $Lv_i^{(j)} \neq 0\ \forall\, i,j$ and thus $\dim\ker L \leqslant m - 1$, which leads to the claim using that $\operatorname{rank} L + \dim\ker L = mn$. To see that indeed none of the above vectors is in the kernel of L, notice that $L_0 v^{(0)} = m v^{(0)}$ while $L_{ij} v^{(0)}$ is a non-negative vector $\forall\, i,j$ and similarly for the $\{v_i^{(j)}\}$. ■

This shows that $\mathbf{L} \notin \mathcal{Q}_d$ for $d < \sqrt{mn - m + 1}$ and as discussed above we obtain the following corollary.

Corollary 1. Every set $\mathcal{Q}_d$ in an (m,n) scenario such that $d < \sqrt{mn - m + 1}$ is not convex.

Notice that in the case $n = 2$, we recover the result of [14] that allowed us to verify the nonconvexity of $\mathcal{Q}_2$ in the scenario (4,2). With our result, the simplest scenario for which we can see that $\mathcal{Q}_2$ is not convex is (2,3). However, as mentioned before, $\mathcal{Q}_2$ is known to be nonconvex in the simplest possible scenario (2,2). Since the maximal rank of the matrix of a behavior is $mn - m + 1$, Corollary 1 cannot be further improved using our techniques. Thus, more powerful constraints could be established in principle going beyond the estimates based on the rank.

Lemma 1 has also a similar interpretation to the result of Sec. II B in the prepare-and-measure scenario. If A and B are bound to local preparations but have access to shared randomness, they can obtain behaviors that, in order to be accessible quantumly without this resource, need an arbitrarily large dimension as the number of inputs and/or outputs grows. An analogous result to that of Sec. II C is also obviously true by Bell's theorem. There are behaviors observable quantumly without shared randomness (even with the smallest possible dimension $d = 2$) that cannot be attained by local strategies no matter how much access to this resource they have.

C. Negligibility of low-dimensional sets

As in the prepare-and-measure scenario, one can show as well that a significant difference between the availability (or lack thereof) of shared randomness is that in the latter case the sets of low-dimensional behaviors are negligible in the set of all quantum behaviors.

Theorem 2. Every set $\mathcal{Q}_d$ in an (m,n) scenario such that $d < \sqrt{mn - m + 1}$ is of measure zero and nowhere dense in the full set of quantum behaviors $\mathcal{Q}$.

Proof. The proof is given in two parts. We first show that with the given premise $\mathcal{Q}_d$ is of measure zero and nowhere dense in $\mathcal{NS}$. We then show that this implies the claim.

The first part follows closely the proof of Theorem 1 and we will only outline it. Similarly, we consider the set S of rank-deficient no-signaling behaviors (i.e., rank $P < mn - m + 1$) and prove the claim for this set, which is extended to $\mathcal{Q}_d$ with $d < \sqrt{mn - m + 1}$ because $\mathcal{Q}_d \subset S$. The no-signaling set $\mathcal{NS}$ is a polytope in $\mathbb{R}^t$ with [26]

$$t = m^2(n-1)^2 + 2m(n-1). \tag{26}$$

For behaviors in S, some polynomials of the t variables must additionally vanish, and S has measure zero in $\mathcal{NS}$. To see that S is nowhere dense in $\mathcal{NS}$, one should follow the same argumentation as before replacing P_ϵ in Eq. (12) by

$$P_\epsilon = (1-\epsilon)P + \epsilon L, \tag{27}$$

where L is given by Eq. (25).

We finish by showing that $\mathcal{Q}_d$ ($d < \sqrt{mn - m + 1}$) having zero measure and being nowhere dense in $\mathcal{NS}$ implies the same negligibility properties inside $\mathcal{Q}$. Since the local polytope $\mathcal{L}$ has the same dimension (t) as $\mathcal{NS}$ [26] and $\mathcal{L} \subset \mathcal{Q} \subset \mathcal{NS}$, we have that $\mathcal{Q}$ is not of measure zero in $\mathcal{NS}$ [27]. Hence $\mathcal{Q}_d$ has measure zero in $\mathcal{Q}$ as well. Regarding nowhere density, we use again that $\mathcal{Q}$ is full dimensional together with the fact that it is a convex set. Corollary 6.4.1 in [28] tells us then that for every P in the interior of $\mathcal{Q}$, $\mathring{\mathcal{Q}}$, we have that

$$\exists \epsilon > 0 \text{ such that } ||P - P'|| < \epsilon \Longrightarrow P' \in \mathcal{Q}. \tag{28}$$

Let us proceed by contradiction and assume that $\mathcal{Q}_d$ is not nowhere dense in $\mathcal{Q}$. Then, there would exist a $P \in \overline{\mathcal{Q}_d}$ and $\delta > 0$ such that $||P - P'|| < \delta$ and $P' \in \mathcal{Q}$ implies that $P' \in \overline{\mathcal{Q}_d}$. By definition, such P must belong to $\mathring{\overline{\mathcal{Q}_d}}$ and since $\mathcal{Q}_d \subset \mathcal{Q}$, it holds then that $\mathring{\overline{\mathcal{Q}_d}} \subset \mathring{\overline{\mathcal{Q}}} = \mathring{\mathcal{Q}}$, where the equality follows from the convexity of $\mathcal{Q}$ (cf. Theorem 6.3 in [28]). Thus, by condition (28) we can drop the assumption $P' \in \mathcal{Q}$ if we take $\min\{\epsilon,\delta\}$, i.e.,

$$||P - P'|| < \min\{\epsilon,\delta\} \Longrightarrow P' \in \overline{\mathcal{Q}_d}. \tag{29}$$

This means that $\mathcal{Q}_d$ is not nowhere dense in $\mathcal{NS}$ and we have reached a contradiction [29]. ■

Theorem 2 tells us that in such simple scenarios as (4,2) or (2,3) it is not only not enough to consider quantum systems with $d = 2$ to reproduce all quantum behaviors but that this is almost never the case.

This negligibility property is again in sharp contrast to the case in which the devices of the parties can be correlated. When shared randomness is granted to the parties, we have that $\mathcal{L} \subset \text{Conv}(\mathcal{Q}_d)\ \forall\ d$. Since, as we have already used in the proof of Theorem 2, $\dim \mathcal{L} = \dim \mathcal{Q} = \dim \mathcal{NS}$, the sets $\text{Conv}(\mathcal{Q}_d)$ have nonzero measure and are not nowhere dense in $\mathcal{Q}\ \forall d$.

Looking at Theorem 2, it comes as a natural question whether the bound $d < \sqrt{mn - m + 1}$ is optimal for the negligibility property to hold. Although we cannot answer this question completely, the above observation allows us to establish a lower bound on d for which negligibility does not hold anymore in every scenario [30].

Proposition 2. Every set $\mathcal{Q}_d$ in an (m,n) scenario such that $d \geqslant m^2(n-1)^2 + 2m(n-1) + 1$ has nonzero measure and is not nowhere dense in the full set of quantum behaviors $\mathcal{Q}$.

Proof. We first notice that if $\mathbf{P} \in \mathcal{Q}_d$ and $\mathbf{P}' \in \mathcal{Q}_{d'}$, then $\mathbf{P}_\lambda = \lambda \mathbf{P} + (1-\lambda)\mathbf{P}' \in \mathcal{Q}_{d+d'}\ \ \forall\ \lambda \in (0,1)$. This is a well-known argument that is used to show that $\mathcal{Q}$ is convex. Indeed, if

$$\begin{aligned} P(ab|xy) &= \text{tr}\big(\rho E_a^x \otimes F_b^y\big), \\ P'(ab|xy) &= \text{tr}\big[\rho'(E')_a^x \otimes (F')_b^y\big], \end{aligned} \tag{30}$$

then

$$P_\lambda(ab|xy) = \text{tr}\big\{[\lambda\rho \oplus (1-\lambda)\rho']\mathcal{E}_a^x \otimes \mathcal{F}_b^y\big\} \tag{31}$$

with $\mathcal{E}_a^x = E_a^x \oplus (E')_a^x\ \forall\ a,x$ and $\mathcal{F}_b^y = F_b^y \oplus (F')_b^y\ \forall\ b,y$.

On the other hand, Carathéodory's theorem [28] tells us that $\forall\ \mathbf{P} \in \mathcal{L} = \text{Conv}(\mathcal{Q}_1)$ we have the convex combination

$$P = \sum_{i=1}^{t+1} \lambda_i P_i, \quad P_i \in \mathcal{Q}_1\ \forall\ i, \tag{32}$$

where we are using that $\dim \mathcal{L} = t$ [cf. Eq. (26)]. This means that $\text{Conv}(\mathcal{Q}_1) \subset \mathcal{Q}_{t+1}$, and since $\text{Conv}(\mathcal{Q}_1)$ is not of measure zero nor nowhere dense, we obtain the claim. ■

As in the prepare-and-measure scenario, Theorem 2 also implies an exceptional robustness in the presence of noise for DIDW when shared randomness is not available. Notice that in this case it is natural to assume that the noisy behavior P_η [cf. Eq. (15)] is subjected to uncorrelated noise, i.e., $P_n(ab|xy) = P_A(a|x)P_B(b|y)\ \forall\ a,b,x,y$, since this affects independently the devices held by A and B. Therefore, the noise induces again a rank-1 perturbation and we have that rank $P - 1 \leqslant$ rank $P_\eta \leqslant$ rank $P + 1\ \forall\ \eta > 0$.

IV. CONCLUSIONS

The task of DIDW has been receiving a lot of attention in recent years. It enables experiments to certify the underlying dimension of an uncharacterized physical system, and it provides a framework for semi-device-independent quantum information processing. The most common scenarios for DIDW involve different parties interacting with the physical system: the so-called prepare-and-measure and Bell scenarios. Depending on the context, it might or might not be the case that the parties are provided with an extra resource, shared randomness, that allows to correlate the different devices the parties hold. In this work, we have explored the differences that may arise for the task of DIDW in these two possible settings. We have seen that shared randomness is indeed a powerful resource: certain behaviors that can be obtained by sending just one classical bit (when the devices are correlated) need quantum systems of arbitrarily large dimension in the absence of shared randomness (the necessary quantum dimension grows with the number of possible inputs while the classical dimension remains 2). On the other hand, quantumness is also more powerful than classical systems even if the latter have access to shared randomness. There are behaviors that require an exponentially larger classical dimension, even though in the quantum setting the devices are not correlated. We have also shown that one of the main differences given by the availability (or lack thereof) of this resource is not only the lack of convexity of the corresponding sets of probability distributions, but the fact that for sufficiently small dimensions these sets are negligibly small (of measure zero and nowhere dense in the set of all possible distributions) if shared randomness is not granted. These results are obtained using very simple estimates for the dimension based on the rank of a matrix. In the future, it would be interesting to study whether the bounds on the dimension as a function of the number of possible inputs provided here for the sets to be nonconvex or negligible in both the prepare-and-measure and Bell scenarios can be improved by using more sophisticated tools. The results of [12,13] and the notions of non-negative and positive semidefinite rank might be helpful in this task.

ACKNOWLEDGMENTS

I thank A. Monràs, C. Palazuelos, and I. Villanueva for very useful discussions. This research was funded by the Spanish MINECO through Grants No. MTM2014-54692-P and No. MTM2014-54240-P and by the Comunidad de Madrid through Grant No. QUITEMAD+CM S2013/ICE-2801.

APPENDIX: RELATION BETWEEN THE RANKS OF W AND P

Taking the matrices defined in Sec. II A, here we prove that rank $W \leqslant$ rank P. To transform P into W, we have to subtract from every odd row i its subsequent row $i+1$. This is a so-called elementary row operation, and it can be achieved by multiplying P from the left with the matrix E_i, which is like the identity with the difference being that the entry $(i,i+1)$ should be equal to -1. Since the matrices $\{E_i\}$ are all full-rank, the matrix $\prod_{i \text{ odd}} E_i P$ has the same rank as P. To obtain W, it just remains to delete all even rows and columns, a process that certainly cannot increase the rank. Thus, we obtain the desired result.

[1] M. Hendrych, R. Gallego, M. Mičuda, N. Brunner, A. Acin, and J. P. Torres, Nat. Phys. **8**, 588 (2012); J. Ahrens, P. Badziag, A. Cabello, and M. Bourennane, *ibid.* **8**, 592 (2012).

[2] M. Pawlowski and N. Brunner, Phys. Rev. A **84**, 010302(R) (2011).

[3] H.-W. Li, Z.-Q. Yin, Y.-C. Wu, X.-B. Zou, S. Wang, W. Chen, G.-C. Guo, and Z.-F. Han, Phys. Rev. A **84**, 034301 (2011).

[4] See, e.g., the review H. Buhrman, R. Cleve, S. Massar, and R. de Wolf, Rev. Mod. Phys. **82**, 665 (2010).

[5] N. Brunner, S. Pironio, A. Acin, N. Gisin, A. A. Méthot, and V. Scarani, Phys. Rev. Lett. **100**, 210503 (2008).

[6] M. Navascués, G. de la Torre, and T. Vértesi, Phys. Rev. X **4**, 011011 (2014); M. Navascués and T. Vértesi, Phys. Rev. Lett. **115**, 020501 (2015); M. Navascués, A. Feix, M. Araujo, and T. Vértesi, Phys. Rev. A **92**, 042117 (2015).

[7] S. Wehner, M. Christandl, and A. C. Doherty, Phys. Rev. A **78**, 062112 (2008); M. M. Wolf and D. Pérez-Garcia, Phys. Rev. Lett. **102**, 190504 (2009).

[8] R. Gallego, N. Brunner, C. Hadley, and A. Acin, Phys. Rev. Lett. **105**, 230501 (2010).

[9] M. Dall'Arno, E. Passaro, R. Gallego, and A. Acin, Phys. Rev. A **86**, 042312 (2012).

[10] N. Brunner, M. Navascués, and T. Vértesi, Phys. Rev. Lett. **110**, 150501 (2013).

[11] J. Bowles, M. T. Quintino, and N. Brunner, Phys. Rev. Lett. **112**, 140407 (2014).

[12] J. Sikora, A. Varvitsiotis, and Z. Wei, Phys. Rev. Lett. **117**, 060401 (2016).

[13] J. Sikora, A. Varvitsiotis, and Z. Wei, Phys. Rev. A **94**, 042125 (2016).

[14] K. F. Pál and T. Vértesi, Phys. Rev. A **80**, 042114 (2009).

[15] J. M. Donohue and E. Wolfe, Phys. Rev. A **92**, 062120 (2015).

[16] H. Buhrman, R. Cleve, J. Watrous, and R. de Wolf, Phys. Rev. Lett. **87**, 167902 (2001); A. C.-C. Yao, *35th Annual ACM Symposium on Theory of Computing (STOC 03)* (ACM, New York, 2003), pp. 77–81; D. Gavinsky, J. Kempe, O. Regev, and R. de Wolf, *38th Annual ACM Symposium on Theory of Computing (STOC 06)* (ACM, New York, 2006), pp. 594–603; D. Gavinsky, T. Ito, and G. Wang, arXiv:1210.1535.

[17] H. Fawzi, J. Gouveia, P. A. Parrilo, R. Z. Robinson, and R. R. Thomas, Math. Program. **153**, 133 (2015).

[18] Y. Faenza, S. Fiorini, R. Grappe, and H. R. Tiwary, arXiv:1105.4127; S. Fiorini, S. Massar, S. Pokutta, H. R. Tiwary, and R. de Wolf, *44th Annual ACM Symposium on Theory of Computing (STOC 12)* (ACM, New York, 2012), pp. 95–106; J. ACM **62**, 17 (2015); S. Zhang, *Innovations in Theoretical Computer Science (ITCS 12)* (ACM, New York, 2012), pp. 39–59; R. Jain, Y. Shi, Z. Wei, and S. Zhang, IEEE Trans. Inf. Theor. **59**, 5171 (2013).

[19] S. A. Vavasis, SIAM J. Optim. **20**, 1364 (2009).

[20] D. Gavinsky, J. Kempe, I. Kerenidis, R. Raz, and R. de Wolf, SIAM J. Comput. **38**, 1695 (2008).

[21] A. Montanaro, Quantum Inf. Comput. **11**, 574 (2011).

[22] I. Newman, Inform. Process. Lett. **39**, 67 (1991).

[23] Actually, the fact that S is closed and has measure zero is already enough to prove that it is nowhere dense. However, similar constructions to that of Eq. (13) will be used throughout the paper.

[24] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, Rev. Mod. Phys. **86**, 419 (2014).

[25] The published version of [15] uses numerical evidence to provide a region in Conv($\mathcal{Q}_2$) which is not in $\mathcal{Q}_2$. A posterior version [arXiv:1506.01119v4] mentions that the tools of [12] can be used to see that some points in this region are indeed not in $\mathcal{Q}_2$.

[26] S. Pironio, J. Math. Phys. **46**, 062112 (2005).

[27] The fact that $\mathcal{Q}$ is a bounded convex set guarantees that it is measurable.

[28] R. T. Rockafellar, *Convex Analysis* (Princeton University Press, Princeton, NJ, 1970).

[29] An alternative way to prove that $\mathcal{Q}_d$ is nowhere dense in $\mathcal{Q}$ would be to show that $\mathcal{Q}_d$ is closed.

[30] Notice that a similar reasoning can be applied to the prepare-and-measure scenario. However, it yields a trivial bound since in this case $d = |\mathcal{X}|$ is enough to generate the set of all behaviors.