HAL
open science

Covariance Bell inequalities

Victor Pozsgay, Flavien Hirsch, Cyril Branciard, Nicolas Brunner

To cite this version:

Victor Pozsgay, Flavien Hirsch, Cyril Branciard, Nicolas Brunner. Covariance Bell inequalities. Physical Review A, 2017, 96 (6), pp.062128. 10.1103/PhysRevA.96.062128 . hal-01926843

HAL Id: hal-01926843

https://hal.science/hal-01926843

Submitted on 11 May 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Covariance Bell Inequalities

Victor Pozsgay,[1] Flavien Hirsch,[1] Cyril Branciard,[2] and Nicolas Brunner[1]

[1] *Département de Physique Appliquée, Université de Genève, 1211 Genève, Switzerland*
[2] *Univ. Grenoble Alpes, CNRS, Grenoble INP, Institut Néel, 38000 Grenoble, France*

(Dated: 9th October 2017)

We introduce Bell inequalities based on covariance, one of the most common measures of correlation. Explicit examples are discussed, and violations in quantum theory are demonstrated. A crucial feature of these covariance Bell inequalities is their nonlinearity; this has nontrivial consequences for the derivation of their local bound, which is not reached by deterministic local correlations. For our simplest inequality, we derive analytically tight bounds for both local and quantum correlations. An interesting application of covariance Bell inequalities is that they can act as "shared randomness witnesses": specifically, the value of the Bell expression gives device-independent lower bounds on both the dimension and the entropy of the shared random variable in a local model.

arXiv:1710.02445v1 [quant-ph] 6 Oct 2017

Bell inequalities limit the strength of possible correlations for any model satisfying a natural definition of locality formulated by Bell [1]. Thus, the violation of a Bell inequality indicates that no local model (in the sense of Bell) can reproduce the observed data, which is therefore said to be (Bell) nonlocal. This can be the case in quantum theory where, by performing well-chosen local measurements on an entangled quantum system, one may obtain nonlocal quantum correlations.

Beyond their fundamental interest, Bell inequalities are widely used tools, in particular in quantum information science [2]. Indeed, a Bell inequality violation represents a simple and strong test for certifying the presence of entanglement. Importantly this test is device-independent, in the sense that, assuming quantum theory, a Bell inequality violation implies the presence of entanglement without any assumption on the measuring devices, nor on the Hilbert space dimension. Moreover, the violation of a Bell inequality can also be used to certify the presence of genuine quantum randomness [3, 4], or the security of a cryptographic key [5].

Consider two separate observers, performing local measurement on a shared physical system. The experiment results in some data, namely the joint conditional probabilities of observing a pair of outputs (measurement results) given a pair of inputs (choice of measurement settings). As they capture the strength of correlations in the data, it is intuitive that Bell inequalities are constructed based on some measure of correlation. For instance, the simplest and most famous Bell inequality of Clauser-Horne-Shimony-Holt (CHSH) [6] is a linear combination of simple correlation functions (see below). As these correlation functions are themselves linear combinations of the joint conditional probabilities, the CHSH inequality is a so-called linear Bell inequality. More generally, it is known that the set of distributions that are Bell local can be fully characterized via linear Bell inequalities [7], and great efforts have been dedicated to find such inequalities, see e.g. Refs. [8–13].

It is however natural to ask if other measures of correlations can be used for devising Bell inequalities. Beyond the purely conceptual interest, this could be relevant in a practical context in which the full data is not available, but only certain specific (not necessarily linear) functions of the joint probabilities are. Moreover, it would be interesting to devise new types of Bell inequalities for the case of continuous measurement outcomes, for which the standard approach for constructing linear inequalities does not work anymore.

Several works have explored these ideas. First, entropic Bell inequalities [14–16] were obtained by considering the mutual information between the measurement outcomes. Later, inequalities based on higher moments of the distribution were also derived [17–20]. Finally, nonlinear Bell inequalities have been developed for discussing the generalization of Bell nonlocality to networks [21–25].

Here we introduce a novel class of nonlinear Bell inequalities based on the covariance of measurement results, a natural measure of correlations that is widely used in many different areas of science. We propose a general way to construct such Bell inequalities for covariances, and present, in the simplest case, a method for computing their local bound—a nontrivial problem due to the nonlinearity of the Bell expression. We show on a few explicit examples that our inequalities can detect quantum nonlocality. Moreover, we show that covariance Bell inequalities can be used to characterize shared randomness in a Bell test, providing device-independent lower bounds on the dimension and the entropy of the shared random variable in any local model.

Standard Bell inequalities.— Consider an experiment in which two parties, Alice and Bob, can both locally test different properties of a shared physical system. Let us label by x the measurement choice of Alice and y that of Bob, with the corresponding outcomes defining the random variables A_x and B_y, which take the values a and b, respectively. We will only consider non-signalling scenarios, in which the statistics of A_x do not depend on the choice of y, and similarly the statistics of B_y do not depend on x.

The statistics of the whole experiment, and thereby the correlations between Alice and Bob's measurement

outcomes, are characterized by the joint conditional probability distribution $P(a,b|x,y) = P(A_x = a, B_y = b)$. Such a distribution is called local (in the sense of Bell) if it admits a decomposition of the form

$$P(a,b|x,y) = \int \rho(\lambda)\, P_A(a|x,\lambda)\, P_B(b|y,\lambda)\, d\lambda, \quad (1)$$

where λ represents the possible values a shared classical variable Λ (shared randomness), distributed with the density function $\rho(\lambda)$. The local response functions, defined by the distributions $P_A(a|x,\lambda)$ and $P_B(b|y,\lambda)$ represent the local behavior of Alice and Bob's subsystems.

The set of local distributions—the "local set"—is constrained by linear Bell inequalities of the form

$$\mathrm{B} = \sum_{x,y,a,b} \alpha_{ab|xy}\, P(a,b|x,y) \leq \beta, \quad (2)$$

with some real coefficients $\alpha_{ab|xy}$, and where the "local bound" β is the maximal value of the quantity B over all local distributions. For any given (finite) number of possible measurement settings and outcomes, the local set forms a polytope, and is thus fully characterized by a finite set of such linear Bell inequalities [7].

As a concrete example, consider the case of two measurements for each party, labelled by $x, y \in \{0,1\}$, with binary outcomes $a, b \in \{+1,-1\}$. Here we have the well-known CHSH inequality [6]

$$\mathrm{CHSH} \equiv \langle A_0 B_0\rangle + \langle A_0 B_1\rangle + \langle A_1 B_0\rangle - \langle A_1 B_1\rangle \leq 2, \quad (3)$$

where the correlation functions $\langle A_x B_y\rangle$ are simply defined as the expectation values of the products of outcomes, $\langle A_x B_y\rangle = \sum_{a,b} a\, b\, P(a,b|x,y)$. Like many other linear Bell inequalities, the CHSH inequality can detect quantum nonlocal correlations, obtained by performing well chosen local measurements on a shared entangled state.

A Bell inequality for covariances.— Instead of using the correlation functions $\langle A_x B_y\rangle$, other quantities can also be considered to characterize local distributions; indeed, Bell inequalities were for instance constructed for entropic quantities [14–16]. In this work we focus on another natural and widely used measure of correlation, namely the covariance. As mentioned previously, this may be of practical interest for situations where the full data is not available, and only covariances can be estimated. An advantage of using covariances is also that, like entropies, they are naturally defined for any number of possible measurement outcomes, and even for continuous outcomes. Unlike entropies however, they depend on the specific values given to the measurement outcomes; this allows one, in particular, to distinguish correlations versus anti-correlations.

The covariance of the two output variables A_x and B_y is defined as

$$\mathrm{cov}(A_x, B_y) = \langle A_x B_y\rangle - \langle A_x\rangle\langle B_y\rangle, \quad (4)$$

with $\langle A_x B_y\rangle$ defined as above, and similarly with $\langle A_x\rangle = \sum_{a,b} a\, P(a,b|x,y)$ and $\langle B_y\rangle = \sum_{a,b} b\, P(a,b|x,y)$ (for continuous values, the sums can simply be replaced by integrals, and the probabilities by probability density functions; note also that because of the non-signalling assumption, $\langle A_x\rangle$ does not depend on y, and $\langle B_y\rangle$ does not depend on x). We emphasize that because of the product term in Eq. (4), the covariance is a nonlinear function of the joint probabilities $P(a,b|x,y)$.

We start by again considering the case of two measurements per party ($x, y \in \{0,1\}$), but with now outcomes A_x and B_y that can take any values a, b in the interval $[-1,+1]$; this implies in particular that $-1 \leq \mathrm{cov}(A_x, B_y) \leq 1$ (note that if a, b are taken in any other bounded interval, they can, together with the corresponding covariance, simply be rescaled so that they lie in $[-1,+1]$). Our goal is to bound the set of possible values of the four covariances $\mathrm{cov}(A_x, B_y)$ for local distributions—the "local set" for covariances. To get some intuition, and inspired by the form of CHSH in Eq. (3), one may look for instance at the projection of this local set onto the 2-dimensional $(\mathrm{covCHSH}, \mathrm{covCHSH}')$ plane, with

$$\begin{aligned}\mathrm{covCHSH} \equiv\ & \mathrm{cov}(A_0, B_0) + \mathrm{cov}(A_0, B_1) \\ & + \mathrm{cov}(A_1, B_0) - \mathrm{cov}(A_1, B_1), \end{aligned} \quad (5)$$

and with $\mathrm{covCHSH}'$ defined similarly, except that $\mathrm{cov}(A_0, B_1)$ and $\mathrm{cov}(A_1, B_0)$ come with minus signs. By optimizing the values of $\mathrm{covCHSH}^{(\prime)}$ numerically (but with reliable enough results), we obtained the local set represented in Fig. 1.

One immediately sees that contrary to the standard local polytope in the probability space, the local set in the covariance space is nonconvex, and that (unsurprisingly) it cannot be fully described by a finite number of Bell inequalities based on linear combinations of the covariances. Its full characterization thus looks much more complicated than that of the standard local polytope. Nevertheless, one can still derive some Bell inequalities for covariances that provide an outer approximation of the local set. As illustrated on Fig. 1, an example of such covariance Bell inequalities is the following:

$$\mathrm{covCHSH} \leq \frac{16}{7}. \quad (6)$$

In order to prove that this inequality indeed holds for any local distribution, it is in fact sufficient to restrict to distributions with binary outputs ± 1. Those can be written as convex mixtures of finitely many deterministic (local) distributions, and one can then optimize the expression covCHSH over the corresponding weights. To take into account the constraint that these weights must be nonnegative and sum up to 1, one can introduce Karush-Kuhn-Tucker (KKT) multipliers [26], which provide necessary conditions for a solution to be optimal. By considering decompositions onto different sets of deterministic distributions with nonzero weights, these KKT

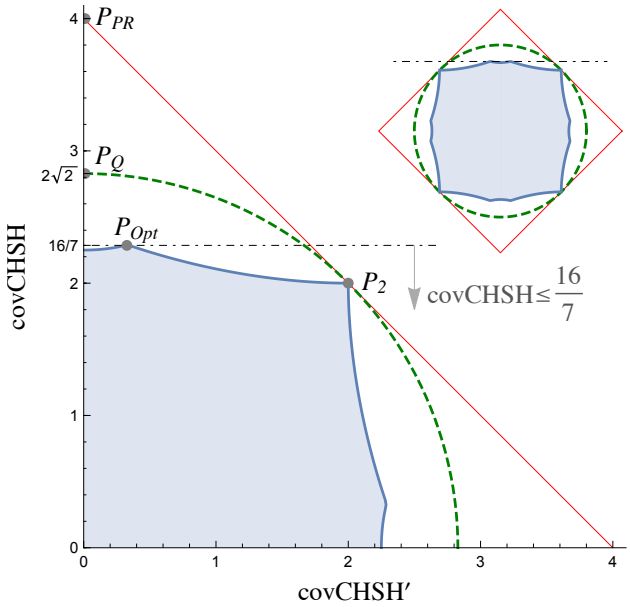

Figure 1. Projection of the local set for covariances (shaded region) onto the upper-right quadrant of the (covCHSH, covCHSH′) plane; the other three quadrants are obtained by symmetry, see inset. All local distributions give values in the local set and satisfy the covariance Bell inequality covCHSH $\leq \frac{16}{7}$. This inequality is saturated by the distributions that project onto P_{Opt}; those require mixing at least 3 deterministic strategies, while mixing 2 allows one to obtain for instance the point P_2 (see main text). Quantum distributions can reach the dashed green circle and can thus violate the covariance Bell inequality, with P_Q corresponding to a distribution that also maximally violates the CHSH inequality. The no-signaling set is limited by the solid red line (as it satisfies covCHSH + covCHSH′ $= 2[\mathrm{cov}(A_0, B_0) - \mathrm{cov}(A_1, B_1)] \leq 4$), where P_{PR} corresponds to a PR nonlocal box.

conditions simplify to a number of linear systems, which can easily be solved. The full details of the proof are given (together with an alternative approach) in Appendix A.

It is instructive to look more closely in the proof at decompositions with a given number d of deterministic distributions with nonzero weights. First note that for $d = 1$, i.e., for deterministic distributions, all covariances are 0, and therefore one also obtains covCHSH $= 0$. To obtain a nonzero value, one thus needs to consider mixtures of at least 2 distributions. For $d = 2$, one finds that the maximal possible value is covCHSH $= 2$, obtained for instance by the distribution $P_2 = \frac{1}{2}\big[(++/+-) + (--/-+)\big]$ (which also gives covCHSH′ $= 2$, see Fig. 1)—where the notation (A_0A_1/B_0B_1) denotes a strategy in which Alice and Bob deterministically output A_x and B_y for inputs x and y, respectively. In order to reach the local upper bound in Eq. (6), one needs to go to $d = 3$; the distribution $P_{Opt} = \frac{3}{7}(++/++) + \frac{2}{7}(-+/--) + \frac{2}{7}(--/-+)$, for instance, gives covCHSH $= \frac{16}{7}$ (and covCHSH′ $= \frac{16}{49}$, see Fig. 1). One thus sees that reaching a given value of covCHSH requires mixing a certain number of deterministic strategies; this is an interesting feature, which will allow one to use the covariance inequality as a shared randomness witness (see below).

One may also look at possible violations of the covariance Bell inequality (6). One finds that it can indeed detect quantum nonlocality: that is, one can obtain covCHSH $> \frac{16}{7}$ by performing local measurements on an entangled state. For example, the quantum correlations that reach the maximal quantum value (the "Tsirelson bound" [27]) of CHSH$=2\sqrt{2}$ have vanishing marginals $\langle A_x\rangle = \langle B_y\rangle = 0$, and therefore also give covCHSH$=2\sqrt{2} > \frac{16}{7}$ (point P_Q on Fig. 1). In fact, this is the largest possible quantum violation, as one can prove that all quantum correlations must satisfy the inequality

$$\mathrm{covCHSH} \leq 2\sqrt{2}\,. \tag{7}$$

This is shown in Appendix B, where we also give a tighter characterization of the quantum set in terms of covariances. Note that in contrast to the local bound, marginals do not play any role in reaching the Tsirelson bound for covCHSH. Nevertheless, the nonlinearity of the Bell expression covCHSH also has interesting consequences in the quantum case. For instance, one can find two pure entangled states such that none of the states can individually violate the inequality (6), whereas a mixture of the two states can violate it; see Appendix C for details.

Note, finally, that given the range of output variables, we have $|\mathrm{cov}(A_x, B_y)| \leq 1$, and therefore the algebraic maximum possible value of covCHSH is 4 (just like for CHSH). This value can be reached by the non-signalling distribution known as the Popescu-Rohrlich (PR) nonlocal box [28] (point P_{PR} on Fig. 1), which again has vanishing marginals and also reaches the algebraic maximum of the CHSH expression.

Constructing other covariance Bell inequalities.— One can follow similar ideas to the ones developed above and derive other Bell inequalities based on covariances.

In the case of ternary inputs $x, y \in \{0, 1, 2\}$ and binary outputs $a, b \in \{+1, -1\}$ for instance, the local set in the space of expectation values $\langle A_xB_y\rangle$ is again fully characterized by CHSH-like inequalities; however, when considering the full probability space, i.e., including the marginals $\langle A_x\rangle$, $\langle B_y\rangle$, one gets a new inequivalent family of Bell inequalities of the form [9, 10, 29]

$$\begin{aligned} I_{3322} \equiv\ & \langle A_0B_0\rangle + \langle A_0B_1\rangle + \langle A_0B_2\rangle + \langle A_1B_0\rangle \\ & + \langle A_1B_1\rangle - \langle A_1B_2\rangle + \langle A_2B_0\rangle - \langle A_2B_1\rangle \\ & + \langle A_0\rangle + \langle A_1\rangle - \langle B_0\rangle - \langle B_1\rangle \ \leq\ 4. \end{aligned} \tag{8}$$

Inspired by the previous example of CHSH, one may look at the local bound when the expectation values $\langle A_xB_y\rangle$ are replaced by the covariances $\mathrm{cov}(A_x, B_y)$. Marginal terms, viewed e.g. as $\langle A_x\rangle = \langle A_x\mathbb{1}_B\rangle$ (where $\mathbb{1}_B$ is the identity measurement operator for Bob, that always outputs $b = 1$) then simply drop out, as $\mathrm{cov}(A_x, \mathbb{1}_B) = 0$.

We thus get the covariance Bell inequality (that allows for any outputs $a, b \in [-1, +1]$)

$$\begin{aligned}\text{cov3322} \equiv\ & \text{cov}(A_0, B_0) + \text{cov}(A_0, B_1) + \text{cov}(A_0, B_2) \\ & + \text{cov}(A_1, B_0) + \text{cov}(A_1, B_1) - \text{cov}(A_1, B_2) \\ & + \text{cov}(A_2, B_0) - \text{cov}(A_2, B_1) \ \leq\ \frac{9}{2}\,. \end{aligned} \tag{9}$$

Here the local bound was obtained through numerical optimization (up to machine precision); nevertheless, the reproducibility and reliability of the numerical results make us quite confident that the bound is correct. In Appendix D we present local distributions that reach it, and discuss quantum and super-quantum (non-signalling) violations.

More generally, one may consider constructing covariance Bell inequalities following the above recipe, starting from an arbitrary linear Bell inequality with binary outcomes. Note that instead of simply dropping the marginal terms, one could as well keep them, or replace them with other functions, e.g. with variances.

Generalising even further our approach for covariance Bell inequalities, one can also investigate Bell inequalities based on Pearson correlators, which can be thought of as normalised covariances. The Pearson correlator for two variables A_x and B_y with variances $\sigma(A_x)^2 = \langle A_x^2\rangle - \langle A_x\rangle^2$ and $\sigma(B_y)^2 = \langle B_y^2\rangle - \langle B_y\rangle^2$ is defined as $\text{r}(A_x, B_y) = \frac{\text{cov}(A_x,B_y)}{\sigma(A_x)\sigma(B_y)}$ (if either $\sigma(A_x)$ or $\sigma(B_y)$ is zero, we define $\text{r}(A_x, B_y) = 0$). Following a similar construction as above, we get, for the case of binary inputs and now also restricting to binary outputs, the Bell inequality

$$\begin{aligned}\text{rCHSH} \equiv\ & \text{r}(A_0, B_0) + \text{r}(A_0, B_1) \\ & + \text{r}(A_1, B_0) - \text{r}(A_1, B_1) \ \leq\ \frac{5}{2}, \end{aligned} \tag{10}$$

where the local bound was again obtained through numerical optimization. It must be emphasized that unlike for covariances, the local set for Pearson correlators is not the same when considering binary outputs or more possible outputs; for instance, with ternary outputs in $\{+1, 0, -1\}$, one can locally reach the value $\text{rCHSH} = 2\sqrt{2}$, which turns out to also be the Tsirelson bound for rCHSH (whether we restrict to binary outputs or not), as proven already in Ref. [30] (we also give a proof of this in Appendix B). In Appendix D we present local distributions that reach these two local bounds of $\frac{5}{2}$ and $2\sqrt{2}$, and discuss quantum and super-quantum violations of them. Of course, similar Bell inequalities with Pearson correlators could also be constructed, following a similar recipe as suggested above, starting e.g. from I_{3322} or any other linear Bell inequality with binary outcomes.

Shared randomness witnesses.— Let us come back now to covCHSH. As emphasized before, the value of covCHSH that can be obtained locally depends on the number of deterministic strategies involved in the local strategy being used. This readily allows one to obtain a lower bound on the classical dimension d of the shared variable Λ (i.e., the number of different values it can take), or equivalently its max-entropy $H_{\max}(\Lambda) \equiv \log_2 d$: from the discussion above, it follows that as soon as covCHSH > 0, one requires $d \geq 2$, i.e. $H_{\max}(\Lambda) \geq 1$; if covCHSH > 2, then $d \geq 3$, i.e. $H_{\max}(\Lambda) \geq \log_2 3$.

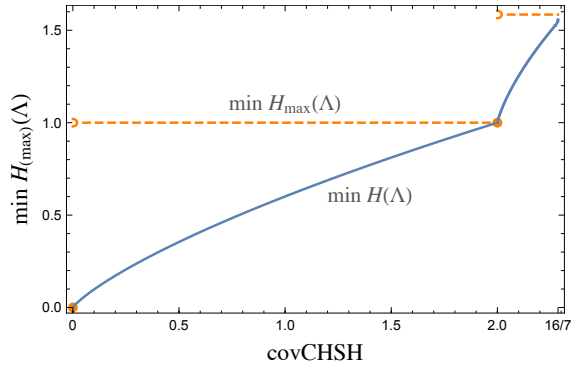

Figure 2. Minimum Shannon entropy $H(\Lambda)$ (blue curve) and max-entropy $H_{\max}(\Lambda) = \log_2 d$ (orange dashed curve) of the shared variable Λ that is required to obtain any given value of the Bell expression covCHSH with a local model. For instance, obtaining covCHSH $= 2$ requires at least one bit of (Shannon or max-) entropy, while reaching the maximum local value covCHSH $= \frac{16}{7} \simeq 2.29$ requires $H(\Lambda) \gtrsim 1.56$ and $H_{\max}(\Lambda) \geq \log_2 3 \simeq 1.58$, i.e. $d \geq 3$.

One may also quantify the amount of shared randomness in terms of the Shannon entropy of Λ, $H(\Lambda) \equiv -\sum_\lambda q_\lambda \log_2 q_\lambda$ (defined here, for simplicity, for the case of a discrete variable Λ that takes the value λ with probability $P(\Lambda = \lambda) = q_\lambda$). For a given value of covCHSH between 0 and $\frac{16}{7}$, one can also obtain a lower bound on $H(\Lambda)$ by minimizing it for all local strategies. As it turns out, it suffices to optimise $H(\Lambda)$ over decompositions onto deterministic strategies with binary outputs ± 1—intuitively, local randomness does not help increase covCHSH. The details of our optimization are given in Appendix E, and the results we obtained are plotted on Fig. 2. Unsurprisingly, we find that $\min H(\Lambda)$ increases with the value of covCHSH. As soon as covCHSH > 0, one requires $H(\Lambda) > 0$; for covCHSH $= 2$, one must have $H(\Lambda) \geq 1$ (recall that covCHSH $= 2$ can be reached by mixing two deterministic strategies with equal weights); finally, reaching the local bound of covCHSH $= \frac{16}{7}$ requires at least as much shared randomness as in the distribution P_{Opt} given previously, i.e. $H(\Lambda) \geq -\frac{3}{7}\log_2\frac{3}{7} - 2\times\frac{2}{7}\log_2\frac{2}{7} \simeq 1.56$.

Thus covCHSH acts as a device-independent witness for characterizing the dimension and the entropy of the shared randomness of a local strategy in a Bell experiment. This complements recent works on device-independent tests of the dimension of quantum systems [31], and of the dimension [32] and entropy [33] of classical communications.

Discussion.— In this Letter we introduced Bell inequalities based on covariances, a natural measure of correlations, rather than on linear combinations of probabilities. We presented the simplest example of such a covariance Bell inequality with binary inputs, that echoes the well-known CHSH inequality, and investigated some of its properties. Proving analytically the local bound is not straightforward here, due to the nonlinearity of the Bell expression, but we could nevertheless provide a method to do so. Its quantum bound was also proven analytically.

Beyond this simplest example, we proposed a general recipe to construct, from any standard Bell inequality with binary outcomes, new Bell inequalities for covariances as well as for Pearson correlators. While our analytical method for proving the local bound generalizes in principle to any covariance Bell inequality obtained with our construction, in practice it becomes intractable for larger numbers of inputs. Nevertheless, for the examples we presented the local bounds could reliably be obtained numerically. It would be interesting to find a more efficient way to prove local bounds, as well as quantum bounds, for covariance-based Bell inequalities—or even to find other ways to characterize the local and quantum sets of admissible covariances (note that a tighter characterization of the quantum set than Eq. (7) can already be given, see Appendix B). One possible idea would be to look for some hierarchy of criteria—similar in spirit to that of Navascués-Pironio-Acín (NPA) for quantum correlations—that covariances must satisfy; indeed, one finds that covariances and Pearson correlators appear naturally in the NPA hierarchy [34].

An interesting feature of our covariance Bell inequalities is that they serve as a shared randomness witness. Indeed, in contrast with standard linear Bell inequalities where the local bound can always be attained by a single deterministic strategy, reaching the local bound—or in fact, any nontrivial value—for a Bell expression defined in terms of covariances requires, for a local strategy, to make use of shared randomness. We showed explicitly, on our simplest example, how the value of the Bell expression allows one to place device-independent lower bounds on the amount of shared randomness in a local model (both in terms of its dimension and its entropy). Such witnesses may help addressing certain problems in quantum nonlocality related to shared randomness [35–37], in particular finding what is the minimal amount of shared randomness necessary to simulate the correlations of entangled states admitting a local model [38]. Our example with binary inputs allows one to certify the use of relatively little shared randomness (e.g. a dimension of at least 3); it would be interesting to find (families of) covariance Bell inequalities that can certify larger amounts of shared randomness. While covariance Bell inequalities were found to provide a natural way to provide device-independent witnesses for shared randomness, such witnesses could also be studied in other frameworks, with other types of Bell inequalities that require shared randomness to be saturated—e.g. with entropic inequalities [14–16].

From a more practical perspective, our covariance Bell inequality could be useful in experimental situations where access to measurement data is limited, e.g. if only covariances can be measured, and where standard Bell inequalities cannot be used. This could be relevant to certain mesoscopic electronics experimental setups. With such applications in mind, it would also be interesting to also investigate similar covariance inequalities to the ones constructed here for other types of quantum nonlocality, e.g. for entanglement [39] or quantum steering [40] (we note that inequalities were constructed in these contexts for covariance matrices, although following very different approaches [41–45]).

Finally, Bell inequalities based on covariances could be useful for the study of nonlocal correlations in networks. When a network features several sources that are assumed to be independent, the set of local distributions (or rather, "N-local distributions" for a network with N independent sources) is typically nonconvex, and nonlinear Bell inequalities are necessary to give an effective description of the (N-)local set [21–25]. In fact, some recent work [46] has already made use of covariances in the context of quantum networks, however not constructing explicit Bell inequalities. The ideas developed here could also find some nice applications in providing ways to obtain new types of Bell inequalities, and possibly shared randomness witnesses, for networks.

Acknowledgements.— We thank Alastair A. Abbott, Joe Bowles, Michael J. W. Hall and Paul Skrzypczyk for discussions. We acknowledge financial support from the Swiss National Science Foundation (Starting grant DIAQ and QSIT) and from the French National Research Agency ('Retour Post-Doctorants' program ANR-13-PDOC-0026).

[1] J. S. Bell, *Physics* **1**, 195–200 (1964).
[2] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, Rev. Mod. Phys. **86**, 419 (2014).
[3] R. Colbeck, PhD Thesis, Univ. of Cambridge (2007).
[4] S. Pironio, A. Acín, S. Massar, A. Boyer de la Giroday, D. N. Matsukevich, P. Maunz, S. Olmschenk, D. Hayes, L. Luo, T. A. Manning, and C. Monroe, Nature **464**, 1021 (2010).
[5] A. Acín, N. Brunner, N. Gisin, S. Massar, S. Pironio, and V. Scarani, Phys. Rev. Lett. **98**, 230501 (2007).
[6] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, Phys. Rev. Lett. **23**, 880 (1969).
[7] I. Pitowsky, Quantum Probability-Quantum Logic, Lecture Notes in Physics, Vol. 321 (Springer-Verlag, New York, 1989).
[8] I. Pitowsky and K. Svozil, Phys. Rev. A **64**, 014102 (2001).

[9] C. Śliwa, Phys. Lett. A **317**, 165 (2003).
[10] D. Collins and N. Gisin, J. Phys. A: Math. Gen. **37** 1775 (2004).
[11] N. Brunner and N. Gisin, Phys. Lett. A **372**, 3162 (2008).
[12] K.F. Pál and T. Vértesi, Phys. Rev. A **77**, 042105 (2008).
[13] D. Rosset, J.-D. Bancal, and N. Gisin, J. Phys. A: Math. Theor. **47**, 424022 (2014).
[14] S. L. Braunstein and C. M. Caves, Phys. Rev. Lett. **61**, 662 (1988).
[15] N. J. Cerf and C. Adami, Phys. Rev. A **55**, 3371 (1997).
[16] R. Chaves and T. Fritz, Phys. Rev. A **85**, 032113 (2012).
[17] E. G. Cavalcanti, C. J. Foster, M. D. Reid, and P. D. Drummond, Phys. Rev. Lett. **99**, 210405 (2007).
[18] A. Salles, D. Cavalcanti, A. Acín, D. Pérez-Garcia, and M. M. Wolf, Quant. Inf. Comp. **10**, 0703-0719 (2010).
[19] A. Bednorz, W. Bednorz, and W. Belzig, Phys. Rev. A **89**, 022125 (2014).
[20] A. Bednorz and W. Belzig, Phys. Rev. B **83**, 125304 (2011).
[21] C. Branciard, N. Gisin and S. Pironio, Phys. Rev. Lett. **104**, 170401 (2010).
[22] C. Branciard, D. Rosset, N. Gisin, and S. Pironio, Phys. Rev. A **85**, 032119 (2012).
[23] A. Tavakoli, P. Skrzypczyk, D. Cavalcanti, and A. Acín, Phys. Rev. A **90**, 062109 (2014).
[24] R. Chaves, Phys. Rev. Lett. **116**, 010402 (2016).
[25] D. Rosset, C. Branciard, T. J. Barnea, G. Pütz, N. Brunner, and N. Gisin, Phys. Rev. Lett. **116**, 010403 (2016).
[26] W. Karush, "Minima of Functions of Several Variables with Inequalities as Side Constraints". M.Sc. Dissertation. Dept. of Mathematics, Univ. of Chicago, Chicago, Illinois (1939); H. W. Kuhn and A. W. Tucker, "Nonlinear programming". Proceedings of 2nd Berkeley Symposium. Berkeley: University of California Press. pp. 481Ð492. MR 47303 (1951).
[27] B. S. Cirel'son, Lett. Math. Phys. **4**, 93 (1980).
[28] S. Popescu and D. Rohrlich, Found. Phys. **24**, 379 (1994).
[29] M. Froissart, Nuovo Cimento B **64**, 241 (1981).
[30] L. J. Landau, Found. Phys. **18**, 449 (1988).
[31] N. Brunner, S. Pironio, A. Acín, N. Gisin, A. A. Méthot, and V. Scarani, Phys. Rev. Lett. **100**, 210503 (2008).
[32] R. Gallego, N. Brunner, C. Hadley, and A. Acín, Phys. Rev. Lett. **105**, 230501 (2010).
[33] R. Chaves, J. B. Brask, N. Brunner, Phys. Rev. Lett. **115**, 110501 (2015).
[34] M. Navascués, S. Pironio, and A. Acín, Phys. Rev. Lett. **98**, 010401 (2007).
[35] K. F. Pal and T. Vértesi, Phys. Rev. A **80**, 042114 (2009).
[36] J. Sikora, A. Varvitsiotis, and Z. Wei, Phys. Rev. Lett. **117**, 060401 (2016).
[37] J. de Vicente, Phys. Rev. A **95**, 012340 (2017).
[38] J. Bowles, F. Hirsch, M. T. Quintino, and N. Brunner, Phys. Rev. Lett. **114**, 120401 (2015).
[39] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, Rev. Mod. Phys. **81**, 865 (2009).
[40] H. M. Wiseman, S. J. Jones, and A. C. Doherty, Phys. Rev. Lett. **98**, 140402 (2007).
[41] R. F. Werner and M. M. Wolf, Phys. Rev. Lett. **86**, 3658 (2001).
[42] G. Giedke, B. Kraus, M. Lewenstein, and J. I. Cirac, Phys. Rev. Lett. **87**, 167904 (2001).
[43] O. Gühne, P. Hyllus, O. Gittsovich, and J. Eisert, Phys. Rev. Lett. **99**, 130504 (2007).
[44] S. Jevtic, M. J. W. Hall, M. R. Anderson, M. Zwierz, and H. M. Wiseman, J. Opt. Soc. Am. B **32**, A40 (2015).
[45] L. Maccone, D. Bruss, and C. Macchiavello, Phys. Rev. Lett. **114**, 130401 (2015).
[46] A. Kela, K. von Prillwitz, J. Aberg, R. Chaves, and D. Gross, arXiv:1701.00652 (2017).
[47] C. Carathéodory, Math. Ann. **64**, 95 (1907).
[48] K.F. Pál and T. Vértesi, Phys. Rev. A **82**, 022116 (2010).

Appendix A: Proofs of the covariance Bell inequality (6)

Obtaining the local bound for a standard (linear) Bell inequality is rather straightforward, as one can simply focus on deterministic strategies. However, such an approach does not work for covariance Bell inequalities because of their nonlinearity.

In this Appendix we give the details of the proof of the inequality (6), as sketched in the main text. We then present an alternative proof that also uses Karush-Kuhn-Tucker (KKT) multipliers and conditions [26], although in a different way. While the former can in principle be adapted to any other covariance Bell inequality (but may not actually be practical because of the very high number of cases to consider for increased numbers of parameters), the latter exploits specific properties of the CHSH (2-input / 2-output) scenario and uses fewer variables in the optimization; it is however less instructive with respect to the shared randomness required for obtaining a given value of covCHSH with a local model, and adapting it to another scenario will first require finding the right properties to exploit.

1. Optimizing over the weights of deterministic distributions

The proof here will be divided into three parts. The first one consists in showing that, without loss of generality, we can focus on a situation where only binary outputs are considered. The second part describes how standard methods for solving (quadratic) constrained problems can be applied. Finally, the third part shows how our quadratic problem can be simplified to a number of linear ones, and solved.

a. *Reduction to binary outputs*

Consider a local distribution $P(a,b|x,y)$ with outputs $a,b \in [-1,+1]$.[1] Let then Alice and Bob post-process their outputs locally, in the following way: they replace each output a (b) by $a' = +1$ ($b' = +1$) with probability $\frac{1+a}{2}$ ($\frac{1+b}{2}$), or by $a' = -1$ ($b' = -1$) with probability $\frac{1-a}{2}$ ($\frac{1-b}{2}$), thus defining a new local distribution $P'(a',b'|x,y) = \sum_{ab} \frac{1+a'a}{2}\frac{1+b'b}{2} P(a,b|x,y)$ with binary outputs $a',b' \in \{+1,-1\}$. One can easily check that the expectation values $\langle A_x^{(\prime)} B_y^{(\prime)}\rangle$, $\langle A_x^{(\prime)}\rangle$ and $\langle B_y^{(\prime)}\rangle$, and therefore also the covariances $\mathrm{cov}(A_x^{(\prime)}, B_y^{(\prime)})$, are the same for the two local distributions P and P'.[2]

Thus, any values of the covariances $\mathrm{cov}(A_x, B_y)$ (for the various inputs x,y) obtained by a local distribution with inputs in $[-1,+1]$ can also be obtained by a local distribution with inputs in $\{+1,-1\}$ (and vice-versa, obviously). The local set for covariances is the same in both cases, and to characterize it it thus suffices to restrict to local distributions with binary outputs ± 1.

b. *KKT conditions*

Our goal is now to find the largest possible value of the quantity covCHSH for any local strategy with binary inputs and outputs. Such a local distribution can be represented by a 16-dimensional vector $\vec{P}$, the components of which are the 16 joint probabilities $P(a,b|x,y)$[3]—and which in fact, because of the normalization and non-signaling constraints, live in a space of dimension only $\dim_{\vec{P}} = 8$.[4] Moreover, $\vec{P}$ admits a convex decomposition in terms of the vertices of the local polytope of the form $\vec{P} = \sum_k q_k \vec{P}_k^{\mathrm{det}}$ with $q_k \geq 0$, $\sum_k q_k = 1$, and where the vectors $\vec{P}_k^{\mathrm{det}}$ represent the 16 deterministic local strategies. (Note already that due to Carathéodory's theorem [47], for any local distribution $\vec{P}$ it will in fact be sufficient to consider at most $\dim_{\vec{P}} + 1 = 9$ deterministic strategies in the decomposition.) Evaluating covCHSH, we get

$$\mathrm{covCHSH}(\{q_k\}_k) = \sum_k q_k C_{kk} - \sum_{i,j} q_i q_j C_{ij}, \quad \text{(A1)}$$

where $C_{ij} = A_0^i B_0^j + A_0^i B_1^j + A_1^i B_0^j - A_1^i B_1^j = \pm 2$ and A_x^i refers to Alice's output for the measurement x given that strategy $\vec{P}_i^{\mathrm{det}}$ is used, and similarly for B_y^j—so that in particular, $C_{kk} = \mathrm{CHSH}(\vec{P}_k^{\mathrm{det}})$ is the CHSH value obtained for the deterministic strategy $\vec{P}_k^{\mathrm{det}}$.

We shall now maximize $\mathrm{covCHSH}(\{q_k\}_k)$ with respect to the variables q_k—which, because of the second term in Eq. (A1), is a nonlinear optimization problem. To account for the constraints that $q_k \geq 0$ and $\sum_k q_k = 1$, we introduce KKT multipliers (i.e., Lagrange-like multipliers)

[1] Note that given the translational invariance of covariances (i.e., $\mathrm{cov}(A_x, B_y) = \mathrm{cov}(A_x + \alpha, B_y + \beta)$ for any real values α, β), it would be equivalent to take a,b in any interval of length 2. Similarly in the case of binary outputs, instead of ± 1 one could take any two real values with a difference of 2.

[2] Note, however, that the Pearson correlators may in general be different, with $|\mathrm{r}(A'_x, B'_y)| \leq |\mathrm{r}(A_x, B_y)|$. This is why one does not obtain the same local bound on rCHSH when considering binary outputs or a larger set of possible outputs; see Appendix D 2.

[3] For simplicity we may identify a distribution P with its vectorial representation $\vec{P}$.

[4] An alternative, more compact parametrization for $\vec{P}$ may for instance be given by the 8 correlators $\langle A_x B_y\rangle$, $\langle A_x\rangle$ and $\langle B_y\rangle$; see Appendix A 2.

λ_k and μ, and define the following Lagrangian:

$$\begin{aligned} L &= \text{covCHSH}(\{q_k\}_k) + \sum_k \lambda_k\, q_k + \mu\,(1 - \sum_k q_k) \\ &= \sum_k q_k C_k - \sum_{i,j} q_i q_j Z_{ij} + \sum_k \lambda_k\, q_k + \mu\,(1 - \sum_k q_k). \end{aligned} \tag{A2}$$

Given the regularity of the problem, with the constraints being affine functions of the variables, the (local) maxima must then necessarily fulfill the KKT conditions [26]:

$$\begin{aligned} &\frac{\partial L}{\partial q_k} = C_{kk} - \sum_i q_i (C_{ik} + C_{ki}) + \lambda_k - \mu = 0 \quad \forall k, \\ &\frac{\partial L}{\partial \mu} = 1 - \sum_k q_k = 0, \\ &\lambda_k q_k = 0,\ \lambda_k \geq 0,\ q_k \geq 0. \end{aligned} \tag{A3}$$

c. *Reduction to linear systems of equations and resolution*

Because of the constraint $\lambda_k q_k = 0$, the above system of equations is nonlinear. One can however reduce it to a number of different linear systems of equations by imposing that a given subset of weigths q_k take nonzero values, and by considering separately the case of each possible subset. For those q_k in the given subset under consideration, the last line of Eq. (A3) simply becomes $\lambda_k = 0, q_k > 0$, while for the other q_k's, it becomes $q_k = 0, \lambda_k \geq 0$—with in both cases, $\lambda_k = -C_{kk} + \sum_i q_i (C_{ik} + C_{ki}) + \mu$ according to the first line of Eq. (A3). For a given number d of nonzero weights, the system (A3) then reduces to $d+1$ linear equality constraints and 16 (also linear) inequality constraints (with d strict and $16-d$ nonstrict inequalities) for the remaining $d+1$ nontrivial variables q_k and μ.

Note that if only $d = 1$ weight q_k is nonzero (and therefore equal to 1), then the distribution $P(a,b|x,y)$ is deterministic, which implies that all covariances—and hence covCHSH—are zero. Furthermore, as already mentioned previously, due to Carathéodory's theorem it is sufficient to only consider subsets with at most $d = 9$ (out of 16) nonzero q_k's—indeed, if a maximum is reached by some decomposition involving more than 9 deterministic distributions, then it is also reached by another decomposition giving the same distribution, but involving no more than 9 deterministic distributions. We used Mathematica to solve, for each value of d between 2 and 9, the $\binom{16}{d}$ different linear systems of equations corresponding to all the different subsets containing d nonzero weights (which makes a total of $\sum_{d=2}^{9} \binom{16}{d} = 50\,626$ linear systems to consider). More specifically, we first found the solutions that satisfy all equality constraints and then checked which ones also satisfy the inequality constraints. To obtain the local bound of our covariance Bell inequality (6), there then remains to check what is the largest possible value of covCHSH for all these feasible solutions.

P =	$\frac{3}{7}\,P^{\text{det}}_{k_1}$	+	$\frac{2}{7}\,P^{\text{det}}_{k_2}$	+	$\frac{2}{7}\,P^{\text{det}}_{k_3}$
$\left\{\begin{matrix} \\ \\ \end{matrix}\right.$	$\frac{3}{7}(++/++)$	+	$\frac{2}{7}(-+/--)$	+	$\frac{2}{7}(--/-+)$
	$\frac{3}{7}(--/--)$	+	$\frac{2}{7}(+-/++)$	+	$\frac{2}{7}(++/+-)$
$\left\{\begin{matrix} \\ \\ \end{matrix}\right.$	$\frac{3}{7}(+-/-+)$	+	$\frac{2}{7}(++/+-)$	+	$\frac{2}{7}(-+/--)$
	$\frac{3}{7}(-+/+-)$	+	$\frac{2}{7}(--/-+)$	+	$\frac{2}{7}(+-/++)$
$\left\{\begin{matrix} \\ \\ \end{matrix}\right.$	$\frac{3}{7}(++/+-)$	+	$\frac{2}{7}(+-/-+)$	+	$\frac{2}{7}(--/--)$
	$\frac{3}{7}(--/-+)$	+	$\frac{2}{7}(-+/+-)$	+	$\frac{2}{7}(++/++)$
$\left\{\begin{matrix} \\ \\ \end{matrix}\right.$	$\frac{3}{7}(+-/++)$	+	$\frac{2}{7}(-+/+-)$	+	$\frac{2}{7}(--/--)$
	$\frac{3}{7}(-+/--)$	+	$\frac{2}{7}(+-/-+)$	+	$\frac{2}{7}(++/++)$

Table I. Optimal decompositions of the form $P = \frac{3}{7}P^{\text{det}}_{k_1} + \frac{2}{7}P^{\text{det}}_{k_2} + \frac{2}{7}P^{\text{det}}_{k_3}$ obtained as solutions of the KKT conditions of Eq. (A3), giving covCHSH $= \frac{16}{7}$. The first 4 distributions give covCHSH$' = \frac{16}{49}$ and thus project onto the point P_{Opt} on Fig. 1 (while the last 4 give covCHSH$' = -\frac{16}{49}$; note also that there exist 8 other similar decompositions that give covCHSH $= -\frac{16}{7}$ and covCHSH$' = \pm\frac{16}{49}$). The distributions in the table are grouped by pairs (brackets on the left), in which the two decompositions are obtained from one another by flipping all outputs.

For $d = 2$, all $\binom{16}{2} = 120$ systems of equations have solutions that satisfy all equality constraints (out of which, 56 have unique solutions and 64 are still underdetermined), but only 4 out of them have (unique) solutions that also satisfy the inequality constraints. Those 4 solutions all give covCHSH $= 2$, which is the maximal value that one can reach by mixing only 2 local deterministic distributions.

With $d = 3$ nonzero weights q_k, again all $\binom{16}{3} = 560$ systems of equations have solutions that satisfy all equality constraints (432 with unique solutions and 128 still underdetermined), but only 8 of them have (unique) solutions that also satisfy the inequality constraints. Those solutions are listed in Table I; they all have the similar form $P = \frac{3}{7}P^{\text{det}}_{k_1} + \frac{2}{7}P^{\text{det}}_{k_2} + \frac{2}{7}P^{\text{det}}_{k_3}$, and all give the value covCHSH $= \frac{16}{7}$.

As it turns out, this value of $\frac{16}{7}$ obtained by suitably mixing 3 deterministic local distributions cannot be increased by mixing more distributions. For $d = 4$ for instance, we find that out of the $\binom{16}{4} = 1\,820$ systems of equations, only 1 516 have solutions that satisfy all equality constraints, out of which only 14 have solutions that also satisfy the inequality constraints. 6 of these (still underdetermined) solutions give a local maximum of covCHSH $= 2$, while the other 8 (unique) solutions—all of the form $P = \frac{3}{8}P^{\text{det}}_{k_1} + \frac{3}{8}P^{\text{det}}_{k_2} + \frac{1}{8}P^{\text{det}}_{k_3} + \frac{1}{8}P^{\text{det}}_{k_4}$ (one solution being obtained for instance by taking $P^{\text{det}}_{k_1} = (++/++), P^{\text{det}}_{k_2} = (-+/--), P^{\text{det}}_{k_3} = (+-/-+)$ and $P^{\text{det}}_{k_4} = (--/-+)$)—give covCHSH $= \frac{9}{4}$. The results for all values of d from 2 to 9 are summarized in Table II. For $d = 6$ and 8, we also get local maxima of covCHSH $= 2$, while for $d = 5, 7$ and 9, no solutions are

d	$\binom{16}{d}$	# consistent systems of equalities	# consistent systems incl. inequalities	local max covCHSH
2	120	120	4	2
3	560	560	8	$\frac{16}{7}$
4	1 820	1 516	14	2 / $\frac{9}{4}$
5	4 368	3 376	0	–
6	8 008	1 896	4	2
7	11 440	688	0	–
8	12 870	154	1	2
9	11 440	16	0	–

Table II. For any number d between 2 and 9 (1st column), the 2nd column of the table gives the number of subsets of deterministic distributions P_k with nonzero weigths q_k, which defines the number of linear systems—obtained from the KKT conditions (A3)—to consider in the proof. The numbers of systems which have solutions when only considering the equality constraints, or when also including the inequality constraints, are given in the next two columns. The last column gives the local maxima of covCHSH obtained from these feasible solutions of the KKT conditions.

found for any of the corresponding systems of KKT conditions. As one can see, the maximal value of covCHSH obtained for all cases is thus $\frac{16}{7}$; this concludes the proof that this is indeed the value of the local bound in our covariance Bell inequality (6).

It should be clarified that although Table II gives local maxima of 2 or $\frac{9}{4}$ for $d \geq 4$, this does not mean that those are the maximum values of covCHSH when mixing $d \geq 4$ deterministic local distributions. Indeed, the table gives local maxima obtained in a set delimited by strict inequalities $q_k > 0$, while the suprema over these sets may be obtained for some $q_k \to 0$. In fact, covCHSH can get arbitrarily close to $\frac{16}{7}$ with a mixture of $d \geq 4$ deterministic local distributions by mixing an optimal decomposition with 3 distributions from Table I, with a tiny amount of $d-3$ other distributions.

Another observation of interest is that all solutions of the KKT conditions that give a local maximum value of covCHSH $= 2$ (for $d = 2, 4, 6, 8$, as listed in the table) are of the general form $P = q_1'\big[(++/++) + (--/--)\big] + q_2'\big[(++/+-) + (--/-+)\big] + q_3'\big[(+-/++) + (-+/--)\big] + q_4'\big[(+-/-+) + (-+/+-)\big]$ with $q_1', q_2', q_3', q_4' \geq 0$ and $2(q_1' + q_2' + q_3' + q_4') = 1$: for $d = 2$, three of the coefficients q_k' are zero (which, for each remaining nonzero q_k', gives a unique solution to the corresponding system of equations); for $d = 4$, two of the coefficients q_k' are zero; for $d = 6$, one of the coefficients q_k' is zero; and for $d = 8$, all coefficients q_k' are nonzero (which indeed makes the solutions in these last 3 cases underdetermined). As one can see, in this general decomposition each deterministic distribution comes with its "opposite", in which all outputs are flipped, with the same weight. This implies in particular that all marginal expectation values $\langle A_x \rangle$ and $\langle B_y \rangle$ are zero, that the covariances $\mathrm{cov}(A_x, B_y)$ are equal to the expectation values $\langle A_x B_y \rangle$, and therefore that the expression of covCHSH simply reduces to that of CHSH (i.e., the nonlinear part vanishes)—which is indeed bounded by 2 for local distributions. This highlights the crucial role played by the nonlinear terms $\langle A_x \rangle \langle B_y \rangle$ in covCHSH: these are precisely the terms that allow the local bound of covCHSH to be greater than that of CHSH; and in order to reach a value of covCHSH larger than 2 with a local distribution, one needs at least one pair of settings x, y for which $\langle A_x \rangle \langle B_y \rangle \neq 0$.

2. Optimizing over the expectation values $\langle A_x B_y \rangle$, $\langle A_x \rangle$ and $\langle B_y \rangle$

We now present a second possible approach for the proof of Eq. (6), which builds on the observation that the local bound $\frac{16}{7}$ on covCHSH follows from just imposing the local bound 2 on CHSH (which must be respected by any local distribution), and that probabilities are nonnegative. More precisely, as previously we note that one can restrict to binary outputs $a, b = \pm 1$; as it turns out, it will then suffice to impose, together with CHSH ≤ 2, that $P(a,b|x,y) \geq 0$ for all x, y, a, b satisfying $ab = (-1)^{xy+1}$.

Instead of writing covCHSH as a function of the weights of deterministic distributions in a local decomposition, here we will view it directly as a function of the expectation values $\langle A_x B_y \rangle$, $\langle A_x \rangle$ and $\langle B_y \rangle$. Our goal is thus to maximize the value of covCHSH under the above constraints—which can also all be expressed in terms of those expectation values, by writing in particular $P(a,b|x,y) = \frac{1}{4}(1 + a\langle A_x \rangle + b\langle B_y \rangle + ab\langle A_x B_y \rangle)$.

As previously, let us introduce KKT multipliers λ and λ_{xyab} (for the 8 combinations of x, y, a, b such that $ab = (-1)^{xy+1}$), and define the Lagrangian

$$\begin{aligned} &L(\{\langle A_x B_y \rangle, \langle A_x \rangle, \langle B_y \rangle\}_{xy}, \lambda, \{\lambda_{xyab}\}_{xyab}) \\ &\quad = \mathrm{covCHSH} + \lambda\,(2 - \mathrm{CHSH}) \\ &\qquad + \sum_{x,y,a,b:\, ab=(-1)^{xy+1}} \lambda_{xyab}\, P(a,b|x,y). \end{aligned} \tag{A4}$$

The KKT necessary conditions for optimality are

$$\begin{aligned} &\frac{\partial L}{\partial \langle A_x B_y \rangle} = (-1)^{xy}(1-\lambda) + \tfrac{1}{4}\sum_{a,b} \lambda_{xyab}\, ab = 0 \qquad \forall x, y, \\ &\frac{\partial L}{\partial \langle A_x \rangle} = -\sum_{y} (-1)^{xy} \langle B_y \rangle + \tfrac{1}{4}\sum_{y,a,b} \lambda_{xyab}\, a = 0 \qquad \forall x, \\ &\frac{\partial L}{\partial \langle B_y \rangle} = -\sum_{x} (-1)^{xy} \langle A_x \rangle + \tfrac{1}{4}\sum_{x,a,b} \lambda_{xyab}\, b = 0 \qquad \forall y, \\ &\lambda\,(2 - \mathrm{CHSH}) = 0, \quad \lambda \geq 0, \quad 2 - \mathrm{CHSH} \geq 0, \\ &\lambda_{xyab}\, P(a,b|x,y) = 0, \quad \lambda_{xyab} \geq 0, \quad P(a,b|x,y) \geq 0, \\ &\qquad\qquad \forall x, y, a, b:\ ab = (-1)^{xy+1}. \end{aligned} \tag{A5}$$

	$\langle A_0B_0\rangle$	$\langle A_0B_1\rangle$	$\langle A_1B_0\rangle$	$\langle A_1B_1\rangle$	$\langle A_0\rangle$	$\langle A_1\rangle$	$\langle B_0\rangle$	$\langle B_1\rangle$
{	1	$\frac{3}{7}$	$\frac{3}{7}$	-$\frac{1}{7}$	-$\frac{1}{7}$	$\frac{3}{7}$	-$\frac{1}{7}$	$\frac{3}{7}$
	1	$\frac{3}{7}$	$\frac{3}{7}$	-$\frac{1}{7}$	$\frac{1}{7}$	-$\frac{3}{7}$	$\frac{1}{7}$	-$\frac{3}{7}$
{	$\frac{1}{7}$	$\frac{3}{7}$	$\frac{3}{7}$	-1	$\frac{3}{7}$	$\frac{1}{7}$	-$\frac{3}{7}$	-$\frac{1}{7}$
	$\frac{1}{7}$	$\frac{3}{7}$	$\frac{3}{7}$	-1	-$\frac{3}{7}$	-$\frac{1}{7}$	$\frac{3}{7}$	$\frac{1}{7}$
{	$\frac{3}{7}$	$\frac{1}{7}$	1	-$\frac{3}{7}$	$\frac{3}{7}$	-$\frac{1}{7}$	-$\frac{1}{7}$	-$\frac{3}{7}$
	$\frac{3}{7}$	$\frac{1}{7}$	1	-$\frac{3}{7}$	-$\frac{3}{7}$	$\frac{1}{7}$	$\frac{1}{7}$	$\frac{3}{7}$
{	$\frac{3}{7}$	1	$\frac{1}{7}$	-$\frac{3}{7}$	-$\frac{1}{7}$	-$\frac{3}{7}$	$\frac{3}{7}$	-$\frac{1}{7}$
	$\frac{3}{7}$	1	$\frac{1}{7}$	-$\frac{3}{7}$	$\frac{1}{7}$	$\frac{3}{7}$	-$\frac{3}{7}$	$\frac{1}{7}$

Table III. Optimal solutions of the KKT conditions of Eq. (A5) (together with $\lambda = \frac{5}{7}$ in all cases and λ_{xyab} as given in Eq. (A6)), giving covCHSH $= \frac{16}{7}$ and defining the same distributions as in Table I.

The first 3 lines define 8 linear equality constraints for the 8 multipliers λ_{xyab}, which can be solved and give

$$\lambda_{xyab} = 2\big[1 - \lambda - a(-1)^y\langle A_{\bar{x}}\rangle - b(-1)^x\langle B_{\bar{y}}\rangle\big], \quad \text{(A6)}$$

with $\bar{x} = 1 - x$ and $\bar{y} = 1 - y$. We are then left with just the last couple of lines of Eq. (A5), so that either $\lambda = 0$ or CHSH $= 2$, and for each x, y, a, b such that $ab = (-1)^{xy+1}$, either $\lambda_{xyab} = 0$ or $P(a, b|x, y) = 0$. We can then consider the $2^{1+8} = 512$ corresponding cases separately, and solve for each case the resulting linear system of 9 equations for the 9 variables $\langle A_xB_y\rangle$, $\langle A_x\rangle$, $\langle B_y\rangle$ and λ.

Using Mathematica we found feasible solutions for those equations in 320 of the 512 cases. However, one still needs to check if these solutions satisfy the remaining inequality constraints in Eq. (A5); only in 207 of those cases could we find solutions that indeed satisfy all KKT conditions. Among all these solutions, we found that the maximal value of covCHSH was $\frac{16}{7}$, which again proves that this is the local bound of our covariance Bell inequality (6). This value of $\frac{16}{7}$ was obtained for the 8 solutions listed in Table III, which indeed define the same 8 distributions as those listed (in the same order) in Table I.

3. Generalizing our proofs to other covariance Bell inequalities

The two versions of the proof of inequality (6) presented above could in principle be generalized to calculate the local bound of any covariance Bell inequality constructed from a linear Bell inequality with binary outcomes, following the recipe we suggest in the main text. However, as the number of inputs, and therefore the dimension of the relevant probability space, increase, the number of different cases to consider in the proof may become too large to be tractable in a reasonable time.

For the case of cov3322 with ternary inputs for instance, after reducing to binary outputs and taking into account normalization and non-signalling constraints, the relevant probability space is of dimension $\dim_{\vec{P}} = 15$. Following the first approach for the proof above, local distributions can be decomposed onto the 2^6 deterministic local distributions, and the number of different cases to consider (corresponding to the number of subsets of d nonzero weights q_k, with $2 \leq d \leq \dim_{\vec{P}} + 1$) is $\sum_{d=2}^{16}\binom{2^6}{d} \geq 7 \times 10^{14}$, a number far too large for all cases to be considered separately.

One may hope that following our second approach for the proof, instead, may reduce the number of cases to be considered down to a tractable one. The first thing to check here would be which constraints (locality constraints in terms of CHSH or I_{3322} inequalities, together with non-negativity of the probabilities constraints) are sufficient to impose in order to obtain the local bound on cov3322. We did not follow this approach any further, and leave as an open question, whether this would be tractable enough to provide an analytical proof of inequality (9).

Instead, to obtain the local bound of inequality (9), we resorted to numerical optimization, by optimizing over the $2^6 = 64$ weights q_k in a local decomposition of $P = \sum_k q_k P_k^{\text{det}}$. Although the optimization problem is nonconvex and involves quite a few free parameters, we found that the result of the numerical optimization was stable enough when starting from different starting points, which makes us confident that the local bound of $\frac{9}{2}$ in inequality (9) is indeed correct.

We note, finally, that the same proof techniques as above do not work for Bell inequalities with Pearson correlators like inequality (10). The first point to be noticed is that, as emphasized before, the local sets for binary outputs $a, b \in \{+1, -1\}$ and for more outputs $a, b \in [-1, +1]$ are not the same—indeed we found different local bounds for binary and ternary outputs. Even restricting to a fixed number of inputs and trying to follow our approach with KKT multipliers, the Lagrangians we would write in either of the two proof versions would not be nicely quadratic functions of the weights q_k in a local decomposition of the form $P = \sum q_k P_k^{\text{det}}$ (as in Eq. (A2)), or of the expectation values $\langle A_xB_y\rangle$, $\langle A_x\rangle$ and $\langle B_y\rangle$ (as in Eq. (A4)), so that the KKT necessary conditions for linearity would not simplify to linear systems of equations. Instead of pursuing such an approach, we again resorted here to numerical optimizations. As before, the numerical results we obtained were stable enough to make us confident that the local bounds we give are correct.

Appendix B: Tsirelson bounds and quantum violations for covCHSH and rCHSH

In this Appendix we give the proofs that the Tsirelson bounds for both covCHSH and rCHSH are the same as for CHSH, namely $2\sqrt{2}$. The proofs presented here are

inspired by that of Landau (for rCHSH) in Ref. [30].

1. Tsirelson bound for covCHSH

First note that by Neumark's dilation theorem, any quantum correlation can be obtained by projective measurements on pure states in Hilbert spaces of large enough dimensions.

Consider a bipartite state $|\psi\rangle \in \mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$ and projective measurement operators $\hat{A}_x$ for Alice and $\hat{B}_y$ for Bob (with the random variables A_x, B_y corresponding to the results of those measurements). Let us then define the ket vectors $|\alpha_x\rangle = (\hat{A}_x \otimes \mathbb{1}_B - \langle \hat{A}_x\rangle_\psi \mathbb{1}_{AB})\,|\psi\rangle$ and $|\beta_y\rangle = (\mathbb{1}_A \otimes \hat{B}_y - \langle \hat{B}_y\rangle_\psi \mathbb{1}_{AB})\,|\psi\rangle$, where $\mathbb{1}_X$ denotes the identity operator acting on the Hilbert space $\mathcal{H}_X$ and where $\langle \hat{A}_x\rangle_\psi = \langle\psi|\,\hat{A}_x \otimes \mathbb{1}_B\,|\psi\rangle$ and $\langle \hat{B}_y\rangle_\psi = \langle\psi|\,\mathbb{1}_A \otimes \hat{B}_y\,|\psi\rangle$. With these definitions, we have

$$\begin{aligned}
&\langle\alpha_x|\beta_y\rangle \\
&= \langle\psi|\,(\hat{A}_x\otimes\mathbb{1}_B - \langle\hat{A}_x\rangle_\psi\mathbb{1}_{AB})(\mathbb{1}_A\otimes\hat{B}_y - \langle\hat{B}_y\rangle_\psi\mathbb{1}_{AB})\,|\psi\rangle \\
&= \langle\psi|\,\hat{A}_x\otimes\hat{B}_y\,|\psi\rangle - \langle\hat{A}_x\rangle_\psi\langle\hat{B}_y\rangle_\psi \\
&= \langle A_x B_y\rangle - \langle A_x\rangle\langle B_y\rangle = \mathrm{cov}(A_x, B_y)
\end{aligned} \tag{B1}$$

and similarly,

$$\begin{aligned}
\langle\alpha_x|\alpha_x\rangle &= \langle A_x^2\rangle - \langle A_x\rangle^2 = \sigma(A_x)^2, \\
\langle\beta_y|\beta_y\rangle &= \langle B_y^2\rangle - \langle B_y\rangle^2 = \sigma(B_y)^2.
\end{aligned} \tag{B2}$$

Restricting to measurement operators with eigenvalues in $[-1,+1]$,[5] this implies in particular that

$$\begin{aligned}
\|\,\langle\alpha_x|\,\| &= \sqrt{\langle\alpha_x|\alpha_x\rangle} = \sigma(A_x) \leq 1, \\
\|\,\langle\beta_y|\,\| &= \sqrt{\langle\beta_y|\beta_y\rangle} = \sigma(B_y) \leq 1.
\end{aligned} \tag{B3}$$

We thus obtain

$$\begin{aligned}
&\mathrm{covCHSH} \\
&\quad= \langle\alpha_0|\beta_0\rangle + \langle\alpha_0|\beta_1\rangle + \langle\alpha_1|\beta_0\rangle - \langle\alpha_1|\beta_1\rangle \\
&\quad\leq \|\,\langle\alpha_0|\,\|\,\|\,|\beta_0\rangle + |\beta_1\rangle\,\| + \|\,\langle\alpha_1|\,\|\,\|\,|\beta_0\rangle - |\beta_1\rangle\,\| \\
&\quad\leq \|\,|\beta_0\rangle + |\beta_1\rangle\,\| + \|\,|\beta_0\rangle - |\beta_1\rangle\,\| \\
&\quad\leq \sqrt{2\big(\|\,|\beta_0\rangle + |\beta_1\rangle\,\|^2 + \|\,|\beta_0\rangle - |\beta_1\rangle\,\|^2\big)} \\
&\quad\quad= 2\sqrt{\|\,|\beta_0\rangle\,\|^2 + \|\,|\beta_1\rangle\,\|^2} \leq 2\sqrt{2},
\end{aligned} \tag{B4}$$

which is necessarily satisfied by any quantum correlation. This bound can be reached by adequate measurements on a maximally entangled state, see Appendix B 3 below.

[5] We note that the restriction to bounded operators with eigenvalues in $[-1,+1]$ is in fact not necessary in the proof, as it is sufficient to just impose $\sigma(A_x), \sigma(B_y) \leq 1$ (for any state $|\psi\rangle$) for the Tsirelson bound on covCHSH to follow.

Let us mention here that following further the proof of Ref. [30], one can actually also refine the characterization of the quantum set in terms of covariances. Indeed, defining the ket vectors $|\gamma_0\rangle = |\alpha_0\rangle, |\gamma_1\rangle = |\alpha_1\rangle, |\gamma_2\rangle = |\beta_0\rangle, |\gamma_3\rangle = |\beta_1\rangle$ and the matrix Γ with coefficients $\Gamma_{ij} = \langle\gamma_i|\gamma_j\rangle$ (for $0 \leq i,j, \leq 3$), one necessarily has that Γ is positive semidefinite (as $\Gamma = M^\dagger M$ with $M = \sum_i |\gamma_i\rangle\langle i|$, where $\{|i\rangle\}$ is the computational basis). That is, taking Eqs. (B1)–(B2) into account, one must have

$$\Gamma = \begin{pmatrix} \sigma(A_0)^2 & \Gamma_{01} & \mathrm{cov}(A_0,B_0) & \mathrm{cov}(A_0,B_1) \\ \Gamma_{01}^* & \sigma(A_1)^2 & \mathrm{cov}(A_1,B_0) & \mathrm{cov}(A_1,B_1) \\ \mathrm{cov}(A_0,B_0) & \mathrm{cov}(A_1,B_0) & \sigma(B_0)^2 & \Gamma_{23} \\ \mathrm{cov}(A_0,B_1) & \mathrm{cov}(A_1,B_1) & \Gamma_{23}^* & \sigma(B_1)^2 \end{pmatrix} \geq 0. \tag{B5}$$

Given that the variances are upper-bounded by 1, one also has $\tilde{\Gamma} \geq 0$, with the matrix $\tilde{\Gamma}$ obtained from Γ above by replacing all diagonal terms by 1. Now, it can be shown (following the same arguments as in Ref. [30], see also Ref. [34]) that the condition that there exists (real or complex) coefficients Γ_{01} and Γ_{23} such that $\tilde{\Gamma} \geq 0$ requires

$$\begin{aligned}
&\arcsin\mathrm{cov}(A_0,B_0) + \arcsin\mathrm{cov}(A_0,B_1) \\
&\quad+ \arcsin\mathrm{cov}(A_1,B_0) - \arcsin\mathrm{cov}(A_1,B_1) \;\leq\; \pi
\end{aligned} \tag{B6}$$

(as well as the symmetrical inequalities obtained by permuting the measurement inputs and/or flipping the sign of the measurement outcomes), which must therefore necessarily be satisfied by quantum correlations, and strengthens the condition that covCHSH $\leq 2\sqrt{2}$.

2. Tsirelson bound for rCHSH

To obtain the Tsirelson bound for rCHSH, we follow a very similar approach to the one above, considering now normalized ket vectors $|\tilde{\alpha}_x\rangle = |\alpha_x\rangle/\||\alpha_x\rangle\|$ and $|\tilde{\beta}_y\rangle = |\beta_y\rangle/\||\beta_y\rangle\|$ (if $|\alpha_x\rangle$ or $|\beta_y\rangle$ are null vectors, we define $|\tilde{\alpha}_x\rangle$ or $|\tilde{\beta}_y\rangle$ to be any unit vector orthogonal to all other ones). With these, we now have

$$\langle\tilde{\alpha}_x|\tilde{\beta}_y\rangle = \mathrm{r}(A_x, B_y). \tag{B7}$$

Because of the normalization of $|\tilde{\alpha}_x\rangle, |\tilde{\beta}_y\rangle$ and $\mathrm{r}(A_x, B_y)$ (via the division by the variances $\sigma(A_x), \sigma(B_y)$), we do not need to restrict here to measurement operators with eigenvalues in $[-1,+1]$ (or such that $\sigma(A_x), \sigma(B_y) \leq 1$). With the same calculations as in Eq. (B4), we obtain, for any choice of (bounded) measurement operators $\hat{A}_x, \hat{B}_y$,

$$\begin{aligned}
\mathrm{rCHSH} &= \langle\tilde{\alpha}_0|\tilde{\beta}_0\rangle + \langle\tilde{\alpha}_0|\tilde{\beta}_1\rangle + \langle\tilde{\alpha}_1|\tilde{\beta}_0\rangle - \langle\tilde{\alpha}_1|\tilde{\beta}_1\rangle \\
&\leq \cdots \leq 2\sqrt{2}.
\end{aligned} \tag{B8}$$

This bound can again be reached quantum mechanically, see below.

As just mentioned, here no restriction is imposed on the measurement outcomes (provided only that $\sigma(A_x), \sigma(B_y) < \infty$). The bound above also holds for local distributions (which can always be realized quantum mechanically) when measurement outcomes are not necessarily restricted to be binary; as we show explicitly in Appendix D 2 b, it can also be reached locally with ternary outcomes. It is quite remarkable that the local and the Tsirelson bounds coincide in this case.

Let us finally mention again that, as it was done in Ref. [30] (see also Ref. [34]), a tighter characterization of the quantum set in terms of Pearson correlators can also be given: any quantum correlation satisfies

$$\begin{aligned} &\arcsin \mathrm{r}(A_0, B_0) + \arcsin \mathrm{r}(A_0, B_1) \\ &+ \arcsin \mathrm{r}(A_1, B_0) - \arcsin \mathrm{r}(A_1, B_1) \leq \pi \end{aligned} \tag{B9}$$

(as well as the symmetrical inequalities).

3. Explicit quantum violations of inequalities (6) and (10)

The Tsirelson bounds for covCHSH and rCHSH can be reached by the same quantum mechanical correlations as those that reach the maximal quantum value of CHSH $= 2\sqrt{2}$. For instance, consider that Alice and Bob share a two-qubit maximally entangled state $|\phi_+\rangle = \frac{1}{\sqrt{2}}(|0,0\rangle + |1,1\rangle)$ and perform the local measurements $\hat{A}_0 = \hat{\sigma}_\mathrm{Z}, \hat{A}_1 = \hat{\sigma}_\mathrm{X}, \hat{B}_0 = \frac{\hat{\sigma}_\mathrm{Z}+\hat{\sigma}_\mathrm{X}}{\sqrt{2}}$ and $\hat{B}_1 = \frac{\hat{\sigma}_\mathrm{Z}-\hat{\sigma}_\mathrm{X}}{\sqrt{2}}$, where $\hat{\sigma}_\mathrm{Z}$ and $\hat{\sigma}_\mathrm{X}$ denote the Pauli matrices. In that case the marginal expectation values $\langle A_x \rangle$ and $\langle B_y \rangle$ vanish, so that covCHSH and rCHSH simply reduce to CHSH, and we indeed obtain covCHSH = rCHSH = CHSH $= 2\sqrt{2}$.

By then rotating for instance Bob's measurement settings together around the y axis of the Bloch sphere, one obtains the full circle drawn on Fig. 1.

Appendix C: Higher quantum violations of Ineq. (6) with mixed states: a consequence of nonlinearity

In this Appendix we discuss an example illustrating the fact that the nonlinearity of the Bell inequality (6) has nontrivial consequences also in the quantum case. Specifically, consider two pure entangled states of the form

$$|\phi_\theta\rangle = \cos\tfrac{\theta}{2}\,|0,0\rangle + \sin\tfrac{\theta}{2}\,|1,1\rangle\,, \tag{C1}$$

$$|\psi_\theta\rangle = \sin\tfrac{\theta}{2}\,|0,0\rangle + \cos\tfrac{\theta}{2}\,|1,1\rangle\,, \tag{C2}$$

with $\theta \in]0, \pi/2]$, and let us then define the equal mixture of these states:

$$\rho_\theta = \frac{1}{2}(|\phi_\theta\rangle\langle\phi_\theta| + |\phi_\theta\rangle\langle\phi_\theta|)\,. \tag{C3}$$

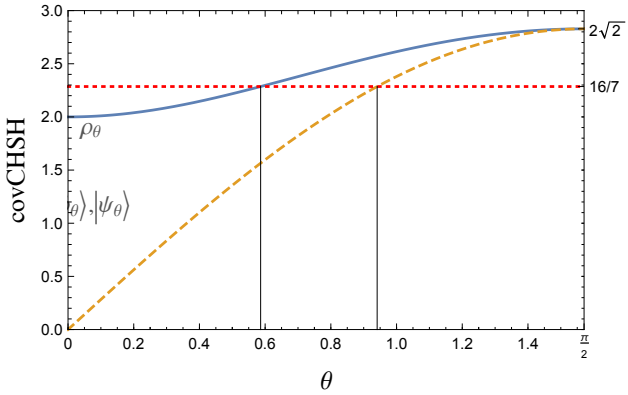

Figure 3. Maximal value of the Bell expression covCHSH for the pure entangled states $|\phi_\theta\rangle$ and $|\psi_\theta\rangle$ (dashed orange curve), and for the mixture ρ_θ (solid blue curve). The horizontal dotted red line represents the local bound of the Bell inequality (6). In the interval $\theta \in]\simeq 0.59, \simeq 0.94]$, none of the two pure states violates the inequality, while the mixture does.

We numerically estimated the largest value of the Bell expression (5), optimizing over all local qubit measurements (including positive-operator valued measures) by Alice and Bob, for various values of the parameter θ. The result is presented in Fig. 3. We found, up to numerical precision, that for the states $|\phi_\theta\rangle$ and $|\psi_\theta\rangle$ the value of covCHSH was maximal when using the Pauli measurements $\hat{A}_0 = \hat{\sigma}_\mathrm{X}, \hat{A}_1 = \hat{\sigma}_\mathrm{Y}, \hat{B}_0 = \frac{\hat{\sigma}_\mathrm{X}-\hat{\sigma}_\mathrm{Y}}{\sqrt{2}}$ and $\hat{B}_1 = \frac{\hat{\sigma}_\mathrm{X}+\hat{\sigma}_\mathrm{Y}}{\sqrt{2}}$, giving covCHSH $= 2\sqrt{2}\sin\theta$ (orange dashed curve on Fig. 3), while the optimal measurements on the state ρ_θ were found to be $\hat{A}_0 = \hat{\sigma}_\mathrm{Z}, \hat{A}_1 = \hat{\sigma}_\mathrm{X}, \hat{B}_0 = \frac{\hat{\sigma}_\mathrm{Z}+\sin\theta\,\hat{\sigma}_\mathrm{X}}{\sqrt{1+\sin^2\theta}}$ and $\hat{B}_1 = \frac{\hat{\sigma}_\mathrm{Z}-\sin\theta\,\hat{\sigma}_\mathrm{X}}{\sqrt{1+\sin^2\theta}}$, giving covCHSH $= 2\sqrt{1+\sin^2\theta}$ (blue curve on Fig. 3). This gives violations of inequality (6) for $\theta > \arcsin\frac{16/7}{2\sqrt{2}} \simeq 0.94$ in the first case, and for $\theta > \arcsin\sqrt{(\frac{16/7}{2})^2 - 1} \simeq 0.59$ in the second case. We thus find that in the range of parameters $0.59 \lesssim \theta \lesssim 0.94$, none of the two pure states can individually violate the covariance Bell inequality (6), whereas the mixture (C3) does violate it. This effect is possible only via the nonlinearity of the covariance Bell inequality.

Note that the pure states $|\phi_\theta\rangle$ and $|\psi_\theta\rangle$ and the mixture ρ give in fact the same correlation functions $\langle A_x B_y \rangle$ for any (projective) Pauli measurements $\hat{A}_x = \hat{\sigma}_{A_x}$ and $\hat{B}_y = \hat{\sigma}_{B_y}$ for Alice and Bob. The difference resides in the marginals $\langle A_x \rangle$ and $\langle B_y \rangle$, which are opposite for $|\phi_\theta\rangle$ and $|\psi_\theta\rangle$ and are therefore unbiased for the equal mixture ρ_θ.

Let us finally mention that we also investigated quantum violations of inequality (6) using mixed entangled states of rank 3 and higher (up to dimension 6×6). We always found that the optimal violations were obtained with pure or rank-2 states. It thus seems that mixing more than 2 pure entangled states does not help increase the violations.

Appendix D: Optimal local strategies and violations for our covariance Bell inequalities (9)–(10)

In this Appendix we present some local strategies that reach the local bounds of our Bell inequalities (9) and (10), and show how these can be violated quantum mechanically and by non-signalling correlations. Recall that the local bounds were obtained numerically; however, we are confident that the numerics are trustworthy enough, and therefore that the local strategies we present below are indeed optimal.

1. Covariance Bell inequality (9)

a. *Optimal local strategies*

To obtain the local bound for cov3322 we ran a numerical optimization several times and conclude from the stability of the result that it must be $\frac{9}{2}$, as in inequality (9).

There are 4 different local distributions that give the value cov3322 $= \frac{9}{2}$, and that decompose onto three deterministic correlations with weights $\frac{3}{8}$, $\frac{3}{8}$ and $\frac{1}{4}$: namely, $P = \frac{3}{8}(+++/+++)+\frac{3}{8}(--+/--+)+\frac{1}{4}(-+-/-+-)$, $P = \frac{3}{8}(+++/++-)+\frac{3}{8}(--+/---)+\frac{1}{4}(+--/-++)$ (using here the notation $(A_0A_1A_2/B_0B_1B_2)$), and the other 2 "opposite decompositions", where all outputs are flipped.

b. *Violations of inequality (9)*

Inequality (9) can be violated by quantum correlations. For example, taking a two-qubit maximally entangled state $|\phi_+\rangle = \frac{1}{\sqrt{2}}(|0,0\rangle+|1,1\rangle)$ and suitable measurements (see below), one can obtain cov3322$=5$, which is also the highest value for I_{3322} (Eq. (8)) obtainable with qubits and projective measurements [9, 10, 29] (note that in that case, all marginal terms vanish and cov3322 effectively reduces to I_{3322}). We have performed an extensive numerical search, considering entangled states of dimension up to 5×5, and could not find any larger quantum violation than cov3322=5. It thus remains an open question, whether higher dimensional entangled states could lead to larger violations, as it is the case for the standard I_{3322} inequality [48].

Considering again the families of states $|\phi_\theta\rangle$, $|\psi_\theta\rangle$ and ρ_θ of Eqs. (C1)–(C3), the largest values of cov3322 we found (numerically, up to machine precision) are cov3322 $= 5\sin\theta$ for $|\phi_\theta\rangle$ or $|\psi_\theta\rangle$ (obtained e.g. for $\hat{A}_0 = \hat{B}_1 = \hat{\sigma}_{\rm X}$, $\hat{A}_1 = \hat{B}_2 = \cos\frac{\pi}{3}\,\hat{\sigma}_{\rm X} + \sin\frac{\pi}{3}\,\hat{\sigma}_{\rm Y}$, $\hat{A}_2 = -\hat{B}_0 = \cos\frac{2\pi}{3}\,\hat{\sigma}_{\rm X} + \sin\frac{2\pi}{3}\,\hat{\sigma}_{\rm Y}$) and cov3322 $= 4+\sin^2\theta$ for ρ_θ (obtained e.g. for $\hat{A}_0 = \hat{B}_0 = \sqrt{1-\frac{\sin^2\theta}{4}}\,\hat{\sigma}_{\rm Z} + \frac{\sin\theta}{2}\,\hat{\sigma}_{\rm X}$, $\hat{A}_1 = \hat{B}_1 = \sqrt{1-\frac{\sin^2\theta}{4}}\,\hat{\sigma}_{\rm Z} - \frac{\sin\theta}{2}\,\hat{\sigma}_{\rm X}$, $\hat{A}_2 = \hat{B}_2 = \hat{\sigma}_{\rm X}$). These violate inequality (9) for $\theta > \arcsin\frac{9/2}{5} \simeq 1.12$ and $\theta > \arcsin\sqrt{\frac{9}{2}-4} = \frac{\pi}{4}$, respectively. As it was the case with covCHSH, we thus find a range of parameters (for $\frac{\pi}{4} < \theta \lesssim 1.12$) for which neither $|\phi_\theta\rangle$ nor $|\psi_\theta\rangle$ violates inequality (9), but their equal mixture does—which is a consequence of the nonlinearity of cov3322. Similarly to covCHSH this property does not seem to extend to mixtures of more than two pure states, that is, the violation of rank-3 states (or higher) does not appear to be larger than when mixing only two pure states. It is also interesting to note that within the families under consideration, cov3322 is violated by a smaller range of states than covCHSH. This is in contrast to the relationship between the standard CHSH and I_{3322} linear Bell inequalities, where all entangled (possibly mixed) quantum states violating (a symmetry of) the former also violate the latter [10].[6]

Note, finally, that the algebraic maximum value of cov3322 $= 8$ can be attained by the non-signalling distributions such that $\langle A_0B_0\rangle = \langle A_0B_1\rangle = \langle A_0B_2\rangle = \langle A_1B_0\rangle = \langle A_1B_1\rangle = -\langle A_0B_2\rangle = \langle A_2B_0\rangle = -\langle A_2B_1\rangle = 1$ (with $\langle A_2B_2\rangle$ remaining a free parameter) and with all marginals $\langle A_x\rangle = \langle B_y\rangle = 0$. Such distributions also reach the maximal non-signalling value of $I_{3322} = 8$.

2. Bell inequality (10) for Pearson correlators

a. *Optimal local strategies with binary outputs $a, b = \pm 1$*

For the case of binary outputs (taking e.g. $a, b = \pm 1$) we find, again numerically, that the local bound for rCHSH is $\frac{5}{2}$, as in inequality (10). This bound can be reached by the 8 distributions that decompose as mixtures of the same triplets of deterministic distributions $P^{\rm det}_{k_1}$, $P^{\rm det}_{k_2}$, $P^{\rm det}_{k_3}$ as those in Table I of Appendix A (i.e., those that allow one to reach the local bound of the covariance inequality (6)), but taking here uniform mixtures, $P = \frac{1}{3}P^{\rm det}_{k_1} + \frac{1}{3}P^{\rm det}_{k_2} + \frac{1}{3}P^{\rm det}_{k_3}$. E.g., for the first one: $P = \frac{1}{3}(++/++) + \frac{1}{3}(-+/--) + \frac{1}{3}(--/-+)$.

---

[6] This can indeed be seen as follows: taking $\hat{A}_2 = \hat{B}_0 = -\mathbb{1}$, the expression of cov3322 in Eq. (8) simplifies to $I_{3322} = \langle A_0B_1\rangle + \langle A_0B_2\rangle + \langle A_1B_1\rangle - \langle A_1B_2\rangle + 2 = {\rm CHSH}_{01/12} + 2$, where ${\rm CHSH}_{01/12}$ is a symmetry of CHSH, Eq. (3). Clearly if there exist measurement operators giving ${\rm CHSH}_{01/12} > 2$, one then also gets a value $I_{3322} > 4$. Note that the same argument does not extend to covCHSH and cov3322: for the choice of $\hat{A}_2 = \hat{B}_0 = -\mathbb{1}$, we just get cov3322 $=$ covCHSH$_{01/12} \leq 4$, so that no violation of inequality (9) is possible with those measurement settings.

b. *Optimal local strategies with ternary outputs* $a, b = +1, 0, -1$

As emphasized before, in the case of Pearson correlators, considering binary outputs or a larger alphabet of possible outputs does make a difference. As shown in Appendix B 2, the value of rCHSH, for any choice of possible outputs, that can be reached locally is upper-bounded by $2\sqrt{2}$. As it turns out, this bound can be reached with ternary outputs $a, b = +1, 0, -1$, for instance by the local distribution $P = \frac{4}{9}(++/+0) + \frac{4}{9}(+-/0+) + \frac{1}{9}(00/--)$.

Note that if any output variable A_x or B_y is deterministic, its variance is zero and by convention we took $\mathrm{r}(A_x, B_y) = 0$. This implies that rCHSH defined in Eq. (10) contains at most 2 nonzero Pearson correlators, which implies that rCHSH ≤ 2. Furthermore, if the two output variables of any party are both deterministic, then rCHSH $= 0$. Nevertheless, one can also locally reach a value of rCHSH arbitrarily close to the bound $2\sqrt{2}$ by certain distributions that get arbitrarily close to some deterministic ones: e.g., $P = (1-\epsilon)(++/+0) + \frac{\epsilon}{2}(+-/-+) + \frac{\epsilon}{2}(-+/--)$ with $0 < \epsilon < 1$ gives rCHSH $= 2(1+\sqrt{1-\epsilon})/\sqrt{2-\epsilon} \to 2\sqrt{2}$ as $\epsilon \to 0$ (note that Alice still has a binary output ± 1 here; only Bob has a ternary output $+1, 0, -1$, for one of his inputs only). Hence, in this case the value of rCHSH cannot be used as a shared randomness witness.

c. *Violations of inequality* (10)

As proven in Appendix B, the Tsirelson bound for rCHSH is $2\sqrt{2}$, which can be reached by adequate (binary) projective measurements on a maximally 2-qubit entangled states. The quantum correlations thus obtained thus violate the inequality (10) (with the local bound $\frac{5}{2}$, for the case where one restricts to binary outputs), but do not violate the analogous inequality for arbitrary outcomes, when the local bound also becomes $2\sqrt{2}$ (see above).

We note that for rCHSH, the nonlinearity of the Bell expression does not seem to exhibit the same interesting features as covCHSH and cov3322 presented in the previous appendices: the maximal value of a mixture of pure states appears, from our numerical investigations, to always be attained by one of the pure states.

Finally, note that algebraic maximum of rCHSH=4 can be obtained by a PR box (just as for CHSH and covCHSH).

Appendix E: Minimal amount of shared randomness required to reach a given value of covCHSH

In this appendix we provide the details on how one can find the minimal amount of shared randomness—quantified here in terms of the Shannon entropy $H(\Lambda)$—required to locally reach a given value of covCHSH.

A general local model provides a decomposition for the probabilities $P(a, b|x, y)$ as in Eq. (1). Our aim is thus to estimate the function

$$\min H_\Lambda^{\mathrm{gen}}(c) \equiv \min_{\substack{\text{local distributions (1)} \\ \text{giving covCHSH}=c}} H(\Lambda), \tag{E1}$$

for any value of c between 0 and $\frac{16}{7}$.

Rather than considering general local decompositions directly, it is in fact useful to first concentrate on the case of binary outputs $a, b = \pm 1$, and consider decompositions onto deterministic local response functions—that is, with $P_A(a|x, \lambda), P_B(b|y, \lambda)$ equal to 0 or 1. We shall thus first estimate the function

$$\min H_\Lambda^{\mathrm{det}}(c) \equiv \min_{\substack{\text{local distributions (1)} \\ \text{with } a,b=\pm 1 \text{ and } P_A,P_B=0,1 \\ \text{giving covCHSH}=c}} H(\Lambda), \tag{E2}$$

which clearly satisfies $\min H_\Lambda^{\mathrm{det}}(c) \geq \min H_\Lambda^{\mathrm{gen}}(c)$. We will then prove that these two functions in fact coincide: to reach the minimal amount of shared randomness for a given value of covCHSH, it is indeed enough to just consider binary outputs ± 1, and decompositions onto deterministic local response functions.

1. Local decompositions onto deterministic response functions with binary outputs ± 1

Let us first note that for binary inputs and binary outputs, there is only a finite number (16) of local deterministic strategies for Alice and Bob together. Even though the local decomposition (1) with deterministic response functions may involve more than 16 different values for λ (and even continuous values, in an appropriate limit), clearly the entropy $H(\Lambda)$ is minimized when the different values of λ that define the same deterministic strategies are grouped together. This implies that it is sufficient to consider here local decompositions of the form $P = \sum_k q_k P_k^{\mathrm{det}}$ onto the 16 different deterministic distributions P_k^{det}. The optimization problem in Eq. (E2) then reduces to optimize the 16 weights q_k under the constraints $q_k \geq 0$, $\sum_k q_k = 1$ and $\mathrm{covCHSH}(\{q_k\}_k) = c$.

We performed this optimization numerically, for various values of $c = \mathrm{covCHSH}$. Our results are shown on Fig. 2 (the function $\min H_\Lambda^{\mathrm{det}}$ coincides with $\min H(\Lambda)$ shown there). We repeated the numerical optimization several times, starting with different random starting points; the consistency of the results we obtained make us confident that we indeed reached the global minima in Eq. (E2).

What we found (up to numerical precision) is that for $c = \mathrm{covCHSH} \leq 2$, the minimum in (E2) is reached by mixing only 2 deterministic local distributions with nonzero weights: for instance, taking $P = q(++/++) + (1-q)(--/--)$ gives covCHSH $= 8q(1-q)$, $H(\Lambda) =$

$H(\{q,1-q\}) = -q\log_2 q - (1-q)\log_2(1-q)$, and leads (for $0 \le c \le 2$) to $\min H_\Lambda^{\rm det}(c) = h_2(\sqrt{1-c/2})$, with the binary entropy function $h_2(x) = -\frac{1+x}{2}\log_2\frac{1+x}{2} - \frac{1-x}{2}\log_2\frac{1-x}{2}$.

For $2 < c = \text{covCHSH} \le \frac{16}{7} \simeq 2.29$, we found that the minimum in (E2) is reached by mixing now 3 deterministic local distributions with nonzero weights: for instance, with distributions of the form $P = q_1(++/++) + q_2(-+/--) + q_3(--/-+)$. That gives covCHSH $= (1-q_1)(1+7q_1) - (q_2-q_3)^2$, $H(\Lambda) = H(\{q_1,q_2,q_3\})$. For a given value of covCHSH $= c$, that reduces to optimizing over just one parameter, say q_1, as q_2 and q_3 are then given by $q_{2,3} = \frac{1-q_1\pm\sqrt{(1-q_1)(1+7q_1)-c}}{2}$ (which requires $\frac{3-\sqrt{16-7c}}{7} \le q_1 \le \frac{3+\sqrt{16-7c}}{7}$, so that all q_k are between 0 and 1). For $c \gtrsim 2.27$ we find that the optimal is obtained for $q_1 = \frac{3+\sqrt{16-7c}}{7}$ (and $q_2 = q_3 = \frac{1-q_1}{2}$); for $c \lesssim 2.27$ we cannot give an analytical solution (however, a good approximation seems to be $q_1 \simeq \frac{1}{2} - 0.01(c-2)$).

We note that as can clearly be seen on Fig. 2, $\min H_\Lambda^{\rm det}$ is a (strictly) increasing function of $c = \text{covCHSH}$.

2. General local decompositions

For a general local decomposition of the form (1), one gets

$$\begin{aligned}&\text{covCHSH}\\ &= \int \rho(\lambda)\big[\langle A_0\rangle_\lambda\,(\langle B_0\rangle_\lambda - \langle B_0\rangle + \langle B_1\rangle_\lambda - \langle B_1\rangle)\\ &\qquad + \langle A_1\rangle_\lambda\,(\langle B_0\rangle_\lambda - \langle B_0\rangle - \langle B_1\rangle_\lambda + \langle B_1\rangle)\big]d\lambda \quad \text{(E3)}\end{aligned}$$

with $\langle A_x\rangle_\lambda = \sum_a a\,P(a|x,\lambda)$ and $\langle B_y\rangle_\lambda = \sum_b b\,P(b|y,\lambda)$ (so that $\langle A_x\rangle = \int \rho(\lambda)\langle A_x\rangle_\lambda d\lambda$ and similarly for $\langle B_y\rangle$).

Looking at this expression, it is clear that for a given distribution $\rho(\lambda)$ (and therefore a given value of $H(\Lambda)$) and some given response functions of Bob, the maximal value of covCHSH is obtained when $|\langle A_0\rangle_\lambda| = 1$, i.e., when Alice uses deterministic response functions with output ± 1 (specifically, Alice's optimal response function is $A_x(\lambda) = \text{sign}[\langle B_0\rangle_\lambda - \langle B_0\rangle + (-1)^x(\langle B_1\rangle_\lambda - \langle B_1\rangle)]$).

A similar reasoning can be followed for Bob. We then conclude that for a given value of $H(\Lambda)$, the maximal value of covCHSH is obtained when Alice and Bob use deterministic response functions with outputs $a, b = \pm 1$: essentially, local randomness (or considering other possible outputs in $[-1,+1]$) does not help increase the value of covCHSH.

Note, however, that the optimization problem in Eq. (E1) is subtly different: rather than maximizing covCHSH for a given value of $H(\Lambda)$, we want to mimimize $H(\Lambda)$ for a given value of covCHSH. To conclude here we will use the previous observation, that $\min H_\Lambda^{\rm det}(c)$ is an increasing function of $c = \text{covCHSH}$.

Consider indeed a (general) local decomposition of a distribution P with a given amount of shared randomness $H(\Lambda) = h$, and which gives some value covCHSH $= c$. From the above reasoning, it follows that there exists another distribution P' using the same shared random variable Λ (hence, with the same value $H(\Lambda) = h$) but now decomposed onto deterministic local response functions with binary outputs ± 1, that gives a value covCHSH $= c' \ge c$. This then implies that $\min H_\Lambda^{\rm det}(c') \le h$ and, because $\min H_\Lambda^{\rm det}$ is an increasing function, $\min H_\Lambda^{\rm det}(c) \le h$ as well. Hence, any general local decomposition that gives some value covCHSH $= c$ necessarily satisfies $H(\Lambda) \ge \min H_\Lambda^{\rm det}(c)$. This implies that $\min H_\Lambda^{\rm gen}(c) \ge \min H_\Lambda^{\rm det}(c)$—and therefore, $\min H_\Lambda^{\rm gen}(c) = \min H_\Lambda^{\rm det}(c)$—with $\min H_\Lambda^{\rm det}(c)$ estimated above.