Using complete measurement statistics for optimal device-independent randomness evaluation

2014 New J. Phys. 16 013035

(http://iopscience.iop.org/1367-2630/16/1/013035)

New Journal of Physics
The open access journal at the forefront of physics
Deutsche Physikalische Gesellschaft DPG | IOP Institute of Physics

Using complete measurement statistics for optimal device-independent randomness evaluation

O Nieto-Silleras, S Pironio[1] and J Silman
Laboratoire d'Information Quantique, Université Libre de Bruxelles (ULB), Bruxelles, Belgium
E-mail: stefano.pironio@ulb.ac.be

Received 2 October 2013, revised 12 December 2013
Accepted for publication 20 December 2013
Published 21 January 2014
New Journal of Physics **16** (2014) 013035
doi:10.1088/1367-2630/16/1/013035

Abstract
The majority of recent works investigating the link between non-locality and randomness, e.g. in the context of device-independent cryptography, do so with respect to some specific Bell inequality, usually the CHSH inequality. However, the joint probabilities characterizing the measurement outcomes of a Bell test are richer than just the degree of violation of a single Bell inequality. In this work we show how to take this extra information into account in a systematic manner in order to optimally evaluate the randomness that can be certified from non-local correlations. We further show that taking into account the complete set of outcome probabilities is equivalent to optimizing over all possible Bell inequalities, thereby allowing us to determine the optimal Bell inequality for certifying the maximal amount of randomness from a given set of non-local correlations.

1. Introduction

In the context of any non-signaling theory, and in particular in the context of quantum theory, outcomes of measurements on separate systems leading to a Bell violation cannot be completely pre-determined, i.e. the violation of a Bell inequality guarantees the presence of genuine randomness. This link between non-locality [1] and randomness is interesting on the fundamental level [2, 3], but is also the main ingredient behind device-independent randomness

[1] Author to whom any correspondence should be addressed.

generation (DIRG) [4–8], randomness amplification [9, 10] and device-independent quantum key distribution (DIQKD) [11–17].

At the basis of such developments lies a quantitative relation between the amount of randomness that is necessarily produced in a Bell experiment and the degree of violation of a certain Bell inequality, such as the CHSH inequality [5, 18], the chained inequality [9, 11, 19, 20] or a Mermin-type inequality [4, 10, 21]. However, the set of data obtained in a Bell experiment is much richer than just the value of the violation of some Bell inequality. For example, in a CHSH experiment there are eight independent probabilities that determine the single number corresponding to the amount of CHSH violation. Moreover, in [3] it was shown that there exist two-input two-output Bell inequalities that can allow for the certification of more randomness than the CHSH inequality. Similar examples have been provided in [22]. Such results imply that taking into account extra data beyond the value of a single Bell violation can be useful, but leave open the questions of just how useful and how to do so in a systematic manner.

These questions are especially relevant now that the detection loophole has been closed (albeit re-opening the locality loophole) with entangled photons [23, 24], opening the door for high rate DIRG. Nevertheless, there is still work to be done on the theoretical level before we can realize this goal efficiently. In particular, low detection efficiencies ($\sim$0.75) necessitate using states of low entanglement (for efficiencies below $\simeq 0.82$ the CHSH inequality cannot be violated using maximally entangled two-qubit states [25]), for which the CHSH inequality is not optimal with respect to randomness certification [3].

In this work we show how to evaluate the randomness produced in a Bell test, or, more specifically, how to obtain the device-independent guessing probability (DIGP) by systematically taking into account the complete non-local behavior, rather than just the violation of some pre-specified Bell inequality. We also show that for any set of non-local correlations, there exists a Bell inequality that is optimal for certifying the maximal amount of randomness given these correlations. Regarding this, we note that while the protocols in [5–7, 14, 15, 17] are general in the sense that they are not formulated with respect to some specific Bell inequality, they do not tell us the optimal Bell inequality to use given the measurement data. We then show how the optimal value of the DIGP and the associated optimal Bell inequality can be computed using the semidefinite programming (SDP) hierarchy introduced in [26]. Finally, we study three numerical examples illustrating the advantage in taking into account the complete non-local behavior, as opposed to taking into account only the violation of a specific Bell inequality.

2. Background: the device-independent guessing probability

We consider the following setting. Alice has access to a pair of quantum devices, or boxes, $\mathcal{A}$ and $\mathcal{B}$, which she can prevent from communicating at will, and whose internal state may be correlated with a system in the possession of an adversary Eve (or equivalently to the environment). The joint state of the boxes and Eve's system is described by a quantum state $\rho_{\mathcal{AB}\mathcal{E}} \in \mathcal{H}_{\mathcal{A}} \otimes \mathcal{H}_{\mathcal{B}} \otimes \mathcal{H}_{\mathcal{E}}$. Alice introduces inputs x and y, each chosen at random from the finite set $\{1, \ldots, n\}$ into boxes $\mathcal{A}$ and $\mathcal{B}$ and obtains outputs a and b, respectively, each taking one of the values $\{1, \ldots, d\}$. This process is described by a pair of POVMs with elements $\{M_{a|x}\}$ and $\{M_{b|y}\}$, each acting on $\mathcal{H}_{\mathcal{A}}$ and $\mathcal{H}_{\mathcal{B}}$, respectively. The joint probability that the outputs a and b are obtained given the inputs x and y is $p_{\mathcal{AB}}(ab|xy) = \mathrm{tr}(\rho_{\mathcal{AB}} M_{a|x} \otimes M_{b|y})$, where $\rho_{\mathcal{AB}} = \mathrm{tr}_{\mathcal{E}}(\rho_{\mathcal{AB}\mathcal{E}})$. There are a total of d^2n^2 such joint probabilities, which we view as the

components of a vector $\mathbf{p} = \{p_{\mathcal{AB}}(ab|xy)\} \in \mathbb{R}^{d^2n^2}$. We refer to this vector as the (non-local) *behavior* characterizing Alice's devices.

We refer to a specific state $\rho_{\mathcal{ABE}}$ and sets of measurement operators $\{M_{a|x}\}$ and $\{M_{b|y}\}$, yielding the behavior $\mathbf{p}$, as a quantum realization Q of $\mathbf{p}$. We denote by $\mathcal{Q}$ the convex set of all types of behavior $\mathbf{p} \in \mathbb{R}^{d^2n^2}$ that admit a valid quantum realization Q. In the following, it will be useful to consider measurements on unnormalized quantum states $\tilde{\rho}_{\mathcal{AB}}$ (i.e. $\mathrm{tr}\,(\tilde{\rho}_{\mathcal{AB}}) \geqslant 0$). We denote the corresponding behavior by $\tilde{\mathbf{p}}$ and define the norm as $\mathrm{tr}(\tilde{\mathbf{p}}) = \mathrm{tr}(\tilde{\rho}_{\mathcal{AB}})$. We denote by $\tilde{\mathcal{Q}}$ the corresponding set of unnormalized quantum behavior, which is a convex cone.

In general, different quantum realizations Q are possible for given behavior $\mathbf{p}$. Our aim is to quantify the randomness generated by the boxes from $\mathbf{p}$ alone, independently of the possible underlying quantum realizations Q compatible with $\mathbf{p}$. To simplify the notation, we describe in the following how to quantify the *local* randomness associated with box $\mathcal{A}$'s output a when a certain input $x = x^*$ is used. The *global* randomness associated with both boxes' outputs a and b for a given pair of inputs $x = x^*$ and $y = y^*$ can be treated analogously.

To begin, let us fix a specific quantum realization Q compatible with $\mathbf{p}$. This quantum realization defines an initial state $\rho_{\mathcal{ABE}}$ and sets of projectors $\{M_{a|x}\}$ and $\{M_{b|y}\}$[2]. After Alice's measurement the correlations between her classical output a and the quantum information held by Eve are described by the classical–quantum state $\sum_a p_{\mathcal{A}}(a|x^*)|a\rangle\langle a| \otimes \rho_{\mathcal{E}}^{ax^*}$, where $\rho_{\mathcal{E}}^{ax^*}$ is the reduced state of Eve given that Alice performed measurement x^* and obtained outcome a. The randomness of box $\mathcal{A}$'s output given this side information can be quantified by the guessing probability [3, 27]: the average probability that Eve correctly guesses box $\mathcal{A}$'s output using an optimal strategy. Such an optimal strategy is described by a d-element POVM $\{M_{a|z}\}$ that Eve performs on her system; if she obtains the output a, which happens with probability $p_{\mathcal{E}}(a|z,\ a',\ x^*,\ Q) = \mathrm{tr}(\rho_{\mathcal{E}}^{a'x^*} M_{a|z})$ when her system is in the reduced state $\rho_{\mathcal{E}}^{a'x^*}$, she guesses that box $\mathcal{A}$'s output was a. Optimizing over all possible measurements, her average probability of guessing correctly is thus given by

$$G(A|E,\ x^*,\ Q) = \max_{\{M_{a|z}\}} \sum_a p_{\mathcal{A}}(a|x^*,\ Q)\, p_{\mathcal{E}}(a|z,\ a,\ x^*,\ Q)\,. \tag{1}$$

The above expression defines the guessing probability, which is related to the quantum min-entropy $H_{\min}(A|E,\ x^*,\ Q)$ through $G(A|E,\ x^*,\ Q) = 2^{-H_{\min}(A|E,\ x^*,\ Q)}$ [27].[3] Note that in the above definition we made the dependence on Q explicit to stress that we are considering a given quantum realization Q. Since our aim is to obtain a bound on the randomness of the outputs that depends only on $\mathbf{p}$, but not on a specific quantum realization Q of $\mathbf{p}$, we must further maximize $G(A|E,\ x^*,\ Q)$ over all Q compatible with $\mathbf{p}$:

$$G(A|E,\ x^*) = \max_{Q,\,\{M_{a|z}\}} \sum_a p_{\mathcal{A}}(a|x^*,\ Q)\, p_{\mathcal{E}}(a|z,\ a,\ x^*,\ Q)\,. \tag{2}$$

This defines the DIGP, the quantity which interests us in this work.

[2] We can always restrict to projectors by increasing the dimension of the Hilbert space. No loss of generality will be incurred by this, since we will be working in device-independent settings.

[3] The guessing probability or equivalently the min-entropy is an operational measure of randomness: if $\rho_{\mathcal{KE}} = \sum_{k=1}^{d} p(k)|k\rangle\langle k| \otimes \rho_{\mathcal{E}}^k$ is a cq-state with guessing probability $G(K|E) \leqslant 2^{-t}$, then a randomness extractor can be used to map $k \in \{1, \ldots, d\}$ to a t-bit string $K' \in \{1, \ldots, g\}$ that is close to being uniformly random and uncorrelated to the adversary, that is $\rho_{\mathcal{K'E}}$ is close in trace-distance to the state $\sum_{k'=1}^{2^{-t}} 2^{-t}|k'\rangle\langle k'| \otimes \sigma_{\mathcal{E}}$ [27].

3. The device-independent guessing probability as a conic linear program

We have expressed the guessing probability as an average over Eve's probabilities conditioned on box $\mathcal{A}$'s outputs, but we can also express it, using Bayes' rule, as an average over Alice's probabilities conditioned on Eve's outcomes:

$$G(A|E,\ x^*) = \max_{Q,\,\{M_{a|z}\}} \sum_a p_{\mathcal{E}}(a|z,\ Q)\, p_{\mathcal{A}}(a|x^*,\ a,\ z,\ Q)\,. \tag{3}$$

Here $p_{\mathcal{E}}(a|z,\ Q)$ is the probability that Eve obtains the outcome a and $p_{\mathcal{A}}(a'|x^*,\ a,\ z,\ Q)$ is the probability that box $\mathcal{A}$ outputs a' conditioned on that event. More generally, conditioning on Eve's outcomes defines a family of types of behavior $\mathbf{p}^{azQ}$ for boxes $\mathcal{A}$ and $\mathcal{B}$, or more conveniently of unnormalized behavior $\tilde{\mathbf{p}}^{azQ} = p_{\mathcal{E}}(a|z,\ Q)\mathbf{p}^{azQ} \in \tilde{\mathcal{Q}}$. Note that averaging over these types of behavior yields back the given behavior characterizing the boxes: $\sum_a \tilde{\mathbf{p}}^{azQ} = \mathbf{p}$. Every choice of Q and $\{M_{a|z}\}$ defines a family of types of quantum behavior satisfying this property. Conversely, it is not difficult to see that any set of types of behavior $\tilde{\mathbf{p}}^a \in \tilde{\mathcal{Q}}$ satisfying $\sum_a \tilde{\mathbf{p}}^a = \mathbf{p}$ can be interpreted as describing the conditional joint output probabilities of boxes $\mathcal{A}$ and $\mathcal{B}$ for some quantum realization Q and POVM $\{M_{a|z}\}$ performed by Eve. In terms of unnormalized behavior, we can write equation (3) as $G(A|E,\ x^*) = \max_{Q,\,\{M_{a|z}\}} \sum_a \tilde{p}_{\mathcal{A}}(a|x^*,\ a,\ z,\ Q)$ and thus the DIGP associated with $\mathbf{p}$ is the solution to the following optimization problem

$$G(A|E,\ x^*) = \max_{\{\tilde{\mathbf{p}}^a\}} \sum_a \tilde{p}^a(a|x^*) \qquad \text{s.t.} \quad \sum_a \tilde{\mathbf{p}}^a = \mathbf{p}\,,\ \ \tilde{\mathbf{p}}^a \in \tilde{\mathcal{Q}}\,,\ \ a = 1, \ldots,\ d, \tag{4}$$

where the $\tilde{\mathbf{p}}^a$s are the optimization variables. This is a typical instance of a conic linear program [28], i.e. the optimization of a linear objective function ($\sum_a \tilde{\mathrm{p}}^a(a|x^*)$) subject to linear constraints ($\sum_a \tilde{\mathbf{p}}^a = \mathbf{p}$) and to the constraint that the optimization variables belong to a convex cone (the constraints $\tilde{\mathbf{p}}^a \in \tilde{\mathcal{Q}}$, since $\tilde{\mathcal{Q}}$ is a closed convex cone).

The program (4) has a simple physical interpretation. Any feasible point corresponds to a possible quantum decomposition $\mathbf{p} = \sum_a \tilde{\mathbf{p}}^a$ of the behavior $\mathbf{p}$. From the point of view of an adversary, such a decomposition can be understood as a strategy where with probability $\mathrm{tr}\,(\tilde{\mathbf{p}}^a)$ the adversary guesses that box $\mathcal{A}$'s output was a and prepares the quantum behavior $\mathbf{p}^a = \tilde{\mathbf{p}}^a/\mathrm{tr}\,(\tilde{\mathbf{p}}^a)$. The probability of correctly guessing box $\mathcal{A}$'s output in this strategy is $\sum_a \tilde{p}^a(a|x^*)$. The program (4) simply searches for the optimal quantum strategy that maximizes this expression.

4. Dual formulation and optimal Bell expressions

Every conic linear program admits a dual formulation (see, e.g., [28]), which in the case of equation (4) is readily seen to be

$$D(A|E, x^*) = \min_{\mathbf{f}}\ \mathbf{f}\cdot\mathbf{p} \quad \text{s.t.} \quad p'(a|x^*) \leqslant_{\mathcal{Q}} \mathbf{f}\cdot\mathbf{p}'\,, \quad a = 1, \ldots, d\,. \tag{5}$$

In the above problem the optimization variable is the vector $\mathbf{f} \in \mathbb{R}^{d^2n^2}$. It can be interpreted as defining a Bell expression whose expectation value is $\mathbf{f}\cdot\mathbf{p} = \sum_{a,\,b,\,x,\,y} f_{abxy}\, p_{\mathcal{AB}}(ab|xy)$. That is, it defines a linear form in the behavior $\mathbf{p}$. The constraint $p'(a|x^*) \leqslant_{\mathcal{Q}} \mathbf{f}\cdot\mathbf{p}'$ means that

$p'(a|x^*) \leqslant \mathbf{f}\cdot\mathbf{p}'$ should hold for all $\mathbf{p}' \in \mathcal{Q}$. Whenever $\mathbf{f}$ satisfies this constraint, the expectation value $\mathbf{f}\cdot\mathbf{p}$ provides an upper bound on the guessing probability since

$$\begin{aligned} G(A|E,\ x^*) &= \max_{\{\tilde{\mathbf{p}}^a\}} \sum_a \tilde{p}^a(a|x^*) \\ &\leqslant \max_{\{\tilde{\mathbf{p}}^a\}} \sum_a \mathbf{f}\cdot\tilde{\mathbf{p}}^a \\ &= \max_{\{\tilde{\mathbf{p}}^a\}} \mathbf{f}\cdot\left(\sum_a \tilde{\mathbf{p}}^a\right) = \mathbf{f}\cdot\mathbf{p}. \end{aligned} \tag{6}$$

In particular, given a fixed Bell expression, such as the CHSH expression $\mathbf{c}$, one can determine coefficients α and β (effectively defining a new linear form $\mathbf{f} = \alpha\,\mathbf{c} + \beta$) such that $p(a|x^*) \leqslant_{\mathcal{Q}} \alpha\,\mathbf{c}\cdot\mathbf{p} + \beta$ and thus $G(A|E, x^*) \leqslant \alpha\,\mathbf{c}\cdot\mathbf{p} + \beta$. Such bounds on the DIGP are the ones that are used in most works related to DIRG or DIQKD, see e.g. [5–8, 29] and [14–17], respectively. The program (5) goes further since it does not assume a fixed Bell expression, but determines the linear form that yields the lowest upper-bound $D(A|E,\ x^*)$ on the DIGP for given behavior $\mathbf{p}$.

The fact that the dual optimal solution $D(A|E,\ x^*) \geqslant G(A|E,\ x^*)$ yields an upper bound on the primal optimal solution is a general result that holds between any primal and dual conic linear program pairs. Provided that one of the two programs admits a strictly feasible solution, it further holds that there is no gap between the primal and dual optimal solutions, i.e. $G(A|E,\ x^*) = D(A|E,\ x^*)$. This is the case here since the form $\mathbf{f}$, defined by $f_{abxy} = 1$ for all a, b, x, and y, satisfies $\mathbf{f}\cdot\mathbf{p} = n^2$, and consequently $p(a|x^*) <_{\mathcal{Q}} \mathbf{f}\cdot\mathbf{p}$, and so represents a strictly feasible point of the dual problem.

The programs (4) and (5) are equivalent but have different interpretations. As we have explained above, the feasible points of the primal program correspond to explicit strategies for the adversary. Any such strategy yields a lower bound on the DIGP. The primal program (4) searches for the optimal strategy that maximizes the guessing probability. On the other hand, any feasible point of the dual program corresponds to a Bell expression, which certifies that a certain amount of randomness is present in the given behavior $\mathbf{p}$, and yields an upper bound on the DIGP. The dual program (5) searches for the Bell expression which certifies the maximal amount of randomness. The duality theorem of conic linear programming tells us that the optimal solutions of both programs are identical, and thus that for every type of behavior $\mathbf{p}$ there exists a Bell expression, which certifies the full amount of randomness present in the correlations.

5. Semidefinite programming relaxations

The above conic linear programming formulations of the DIGP are in general difficult to implement exactly. However, they can be relaxed using the SDP method introduced in [26, 30]. This method introduces a hierarchy of convex sets $\tilde{\mathcal{Q}}_1 \supseteq \tilde{\mathcal{Q}}_2 \supseteq \cdots \supseteq \tilde{\mathcal{Q}}$, which approximate the quantum set $\tilde{\mathcal{Q}}$ from the outside[4]. The hierarchy of programs

$$G_k(A|E,\ x^*) = \max_{\{\tilde{\mathbf{p}}^a\}} \sum_a \tilde{p}^a(a|x^*) \qquad \text{s.t.} \quad \sum_a \tilde{\mathbf{p}}^a = \mathbf{p}\,,\ \ \tilde{\mathbf{p}}^a \in \tilde{\mathcal{Q}}_k\,,\ \ a = 1,\ldots,d \tag{7}$$

[4] The hierarchy as presented in [26, 30] applies to normalized behavior $\mathbf{p} \in \mathcal{Q}$, but it can be trivially adapted to the unnormalized behavior $\tilde{\mathbf{p}} \in \tilde{\mathcal{Q}}$ by removing the normalization constraint, e.g. $\Gamma_{11} = 1$ in the notation of [26, 30].

therefore provides a sequence of relaxations to equation (4), which yields upper bounds $G_1(A|E,\ x^*) \geqslant G_2(A|E,\ x^*) \geqslant \cdots \geqslant G(A|E,\ x^*)$ on the DIGP. In this approach behavior $\tilde{\mathbf{p}}$ belongs to $\tilde{\mathcal{Q}}_k$ if and only if there exists a positive semidefinite matrix $\Gamma_k \succcurlyeq 0$ satisfying a series of linear constraints of the form $\mathrm{tr}\,(G\,\Gamma_k) = \mathbf{h}\cdot\tilde{\mathbf{p}}$ (see [30, 33] for details). Since the objective function and the first set of constraints in equation (7) are also linear, the problems (7) can be cast as SDP problems for which efficient algorithms are available.

This SDP hierarchy can also be understood from the perspective of the dual problem equation (5). To see this, we note that the constraint $p'(a|x^*) \leqslant_{\mathcal{Q}} \mathbf{f}\cdot\mathbf{p}'$ in equation (5) is equivalent to $\langle\psi|\mathcal{F}_a|\psi\rangle \geqslant 0$ for all possible quantum states $|\psi\rangle$ and all possible $\mathcal{F}_a$ of the form $\mathcal{F}_a = \sum_{abxy} f_{abxy} M_{a|x}\otimes M_{b|y} - M_{a|x^*}\otimes\mathbb{I}$, where $\{M_{a|x}\}$ and $\{M_{b|y}\}$ are valid sets of measurement operators. This in turn is equivalent to $\mathcal{F}_a \succeq 0$ for all $\mathcal{F}_a = \sum_{abxy} f_{abxy} M_{a|x}\otimes M_{b|y} - M_{a|x^*}\otimes\mathbb{I}$. We say that $\mathcal{F}_a$ admits a sum of squares (SOS) decomposition of degree $2k$, and write $\mathcal{F}_a = \mathrm{SOS}_k$ if there exists a set $\{\mathcal{S}_a^i\}$ of polynomials of degree k in the operators $\{M_{a|x}\otimes\mathbb{I},\ \mathbb{I}\otimes M_{b|y}\}$ such that $\mathcal{F}_a = \sum_i \mathcal{S}_a^{i\,\dagger}\mathcal{S}_a^i$ holds for any sets of valid measurement operators $\{M_{a|x}\}$ and $\{M_{b|y}\}$. If this is the case, it clearly follows that $\mathcal{F}_a = \sum_i \mathcal{S}_a^{i\,\dagger}\mathcal{S}_a^i \succeq 0$. Therefore, the series of problems

$$G_k(A|E,x^*) = \min_{\mathbf{f}}\ \mathbf{f}\cdot\mathbf{p} \qquad \text{s.t.} \quad \mathcal{F}_a = \mathrm{SOS}_k\,, \quad a = 1,\ldots,d\,. \tag{8}$$

represents a sequence of relaxations of the dual problem (5) yielding upper bounds $G_1(A|E,x^*) \geqslant G_2(A|E,x^*) \geqslant \cdots \geqslant G(A|E,x^*)$ on the DIGP.

It is well known that an SOS constraint of the form $\mathcal{F}_a = \mathrm{SOS}_k$ can be represented as an SDP constraint [31] and thus that the relaxations (8) are SDP problems. Such SDP relaxations turn out to be nothing but the dual formulation of the SDP relaxations (7) [30, 32] (see [33] for more details on the relation between the primal and dual of the SDP hierarchy).

Even though the primal and dual SDP relaxations (7) and (8) are equivalent, like the original programs, they have different interpretations. Feasible points of the primal programs correspond to decompositions of $\mathbf{p}$ in terms of supra-quantum behavior in $\mathcal{Q}_k$. They can be understood as characterizing the strategies available to an adversary who is able to prepare supra-quantum behavior. Such strategies are not necessarily always available in a purely quantum setting and thus the associated values $G_k(A|E,x^*)$ represent upper bounds on the DIGP. The dual programs, on the other hand, return explicit Bell expressions certifying that the DIGP cannot be higher than a certain value $G_k(A|E,x^*)$. Such bounds are valid—and optimal—for any strategy in $\mathcal{Q}_k$ and thus are also valid—though not necessarily optimal—for any quantum strategy in $\mathcal{Q}$. In other words, the SDP relaxations (7) and (8) not only give a bound on the DIGP, but also return explicit Bell expressions that can be used in any analysis based on a quantitative relation between the amount of Bell violation and randomness, such as in [5–10, 14, 16, 17].

6. Numerical examples

In this section we present three numerical examples demonstrating the advantage in taking into account the complete non-local behavior.

In the first two examples, we consider a two-input two-output Bell scenario. We introduce the eight parameters $\langle A_x\rangle = \sum_{a=\pm1} a\ p_{\mathcal{A}}(a|x)$, $\langle B_y\rangle = \sum_{b=\pm1} b\ p_{\mathcal{B}}(b|y)$, $\langle A_x B_y\rangle = \sum_{a,b=\pm1} ab\ p_{\mathcal{AB}}(ab|xy)$, where $x,\ y = 1, 2$, knowledge of which is equivalent to knowledge of the complete set of probabilities $p_{\mathcal{AB}}(ab|xy)$.

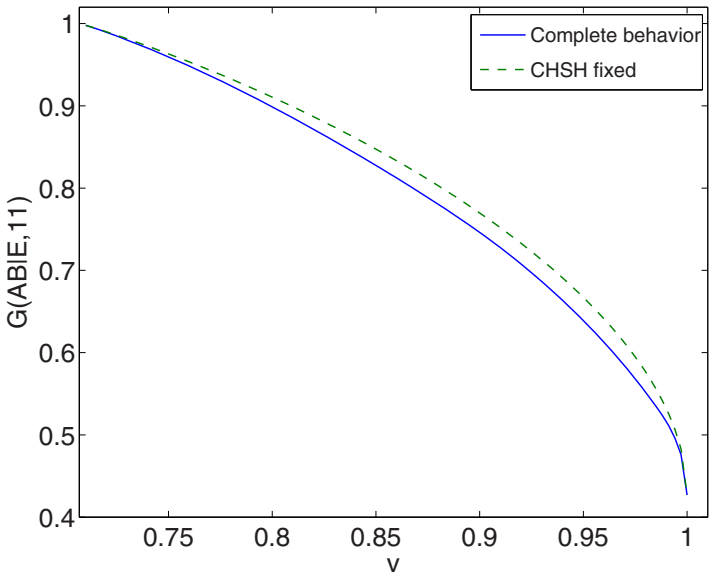

Figure 1. Global randomness $G(A, B|E, 1, 1)$ as a function of the visibility v for optimally violating CHSH correlations in the presence of white noise. The dashed curve was obtained by taking into account only the CHSH value (i.e. $2\sqrt{2}v$), while the solid curve was obtained by taking into account the full non-local behavior. Both curves were obtained using the second order relaxation of the SDP hierarchy and are actually optimal up to the numerical precision of 10^{-6} used (we have verified optimality by finding explicit states and measurements saturating the bounds given by the SDP programs). Except when $v = 1$, i.e. when there is no noise, we see that there is a small advantage in taking into account the full non-local behavior.

6.1. CHSH correlations in the presence of white noise

We first consider the randomness that can be extracted from a mixture of maximally violating CHSH correlations plus white noise, i.e. correlations of the form $v\,\mathbf{q} + (1 - v)\mathbf{r}$, where $\mathbf{q}$ are the quantum correlations yielding the maximal CHSH violation of $2\sqrt{2}$ and $\mathbf{r}$ denotes completely random correlations for which $p_{AB}(ab|xy) = 1/4$ for all a, b, x and y. As a function of the 'visibility' v the CHSH violation is thus given by $2\sqrt{2}\,v$. Naively, one would expect that in such a simple example of knowledge of the full non-local behavior is of no greater utility than knowledge of the CHSH violation alone. Surprisingly, figure 1 shows that this is not the case, although the improvement that we get by considering the full non-local behavior is modest. We have determined numerically the corresponding optimal Bell inequalities as a function of v by solving explicitly the dual programs. We find that these inequalities all have the form

$$f_{11}\langle A_1 B_1\rangle + \langle A_1 B_2\rangle + \langle A_2 B_1\rangle - f_{22}\langle A_2 B_2\rangle, \tag{9}$$

where the coefficients f_{11} and f_{22} are given in figure 2. The case $f_{11} = f_{22} = 1$ corresponds to the CHSH inequality and only arises in the case of perfect visibility ($v = 1$). This shows that in any real experiment, in which the visibility is necessarily imperfect (i.e. $v < 1$), the optimal Bell inequality for randomness certification is not always the CHSH inequality.

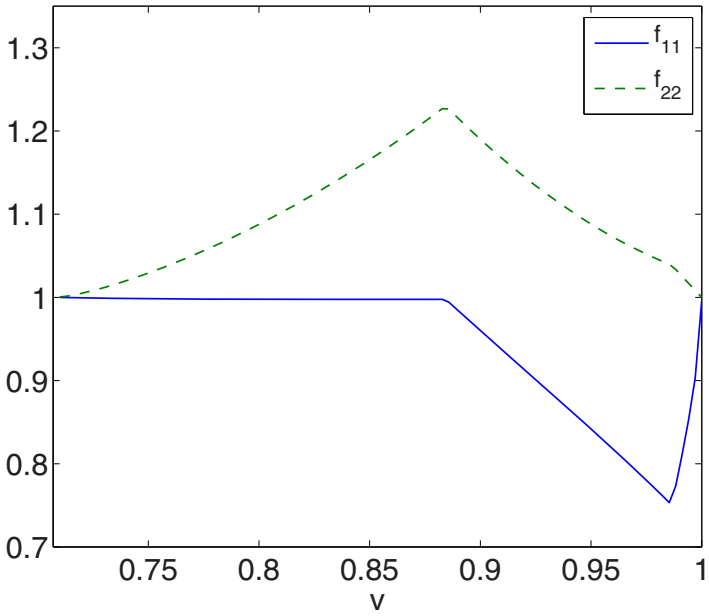

Figure 2. Coefficients of the optimal Bell inequalities equation (9) as a function of v. The CHSH inequality corresponds to the case $f_{11} = f_{22} = 1$ and is optimal only for perfect visibility $v = 1$ (and trivially $v = 1/\sqrt{2}$).

6.2. Randomness from partially entangled states

In the second example, we consider the following set of correlations

$$\begin{aligned} &\langle A_1 B_1\rangle = \langle A_1 B_2\rangle = v\cos\mu, \\ &\langle A_2 B_1\rangle = -\langle A_2 B_2\rangle = v\sin 2\theta\sin\mu\,, \\ &\langle A_1\rangle = v\cos 2\theta, \quad \langle A_2\rangle = 0, \quad \langle B_1\rangle = \langle B_2\rangle = v\cos 2\theta\cos\mu\,, \end{aligned} \tag{10}$$

where $\tan\mu = \sin 2\theta$. For $v = 1$ these correlations are obtained by measuring a partially entangled state of the form $|\Psi\rangle = \cos\theta\,|00\rangle + \sin\theta\,|11\rangle$ and give rise to a maximal violation of the I_1^β inequality [3] ($I_1^\beta = I_{\mathrm{CHSH}} + \beta\langle A_1\rangle \leqslant 2 + \beta$) with $\beta = 2\cos(2\theta)/\sqrt{1+\sin^2(2\theta)}$. A value of $v < 1$ corresponds to a mixture of these correlations with completely white noise in the respective fractions of v and $1 - v$. Figure 3 presents bounds on the global DIGP $G(A,\ B|E,\ 2,\ 1)$ corresponding to the pair of outcomes associated with the measurements A_2 and B_1 as a function of θ for $v = 0.99$. We see that taking into account complete sets of correlations can provide a very significant advantage, not only as compared with taking into account only the violation of a single Bell inequality, but also violations of two independent Bell inequalities.

It is interesting to see what the optimal Bell inequalities, obtained via the dual formulation of the SDP programs, look like. The significant advantage obtained in figure 2 by taking into account complete data suggests that the corresponding optimal Bell inequalities would be more than mere tweaks of any of the Bell inequalities that have thus far been investigated for the purposes of DIRG (essentially the I_α^β inequalities of [3]). This intuition is indeed backed up

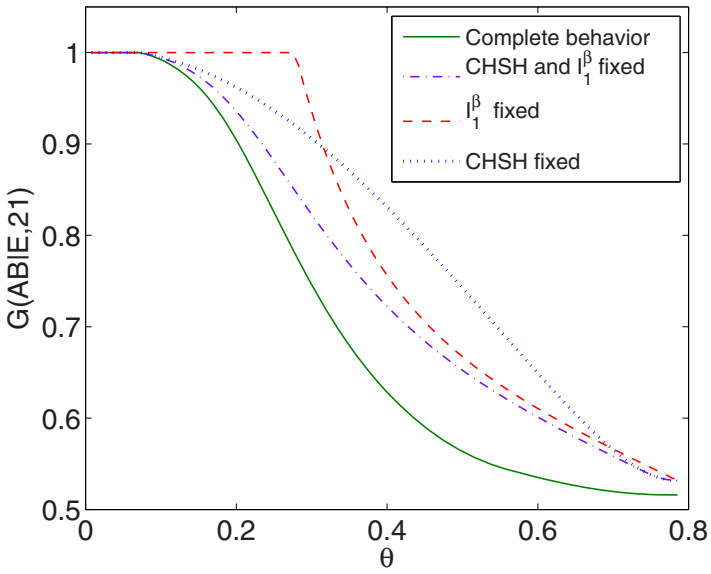

Figure 3. $G(A,\ B|E,\ 2,\ 1)$ as a function of θ computed by taking into account partial or complete non-local data for $v = 0.99$. The dashed curve was obtained by constraining only the value of the I_1^β expression, the dotted curve by constraining only the value of the CHSH expression, the dashed-dotted curve by constraining the values of both I_1^β and the CHSH expressions, and the solid curve by taking into account the values of all correlators in accordance with equation (10). These curves were obtained using the third order relaxation of the SDP hierarchy. The dashed-dotted curve is optimal up to a precision of 10^{-6}.

by the numerics. For example, for $\theta = 27\pi/200$ ($G(A,\ B|E,\ 2,\ 1) \simeq 0.609$) we obtain the Bell expression

$$2.74\,\langle A_1B_1\rangle + 2.60\,\langle A_1B_2\rangle + 2.35\,\langle A_2B_1\rangle - 3.86\,\langle A_2B_2\rangle + 1.36\,\langle A_1\rangle + 1.51\,\langle A_2\rangle - 0.390\,\langle B_1\rangle + 2.05\,\langle B_2\rangle\,, \tag{11}$$

whose local bound is 8.36.

6.3. Randomness from entangled qutrits

As the last example, we consider the two-input, three-ouput Bell-CGLMP scenario [34]. Specifically, we consider correlations which violate the CGLMP inequality and which arise by performing the measurements specified in [34] on the family of states

$$\alpha|00\rangle + \sqrt{1-2\alpha^2}|11\rangle + \alpha|22\rangle\,, \tag{12}$$

with $0 \leqslant \alpha \leqslant 1/\sqrt{2}$. For $\alpha = 0$ the state is a product state, for $\alpha = 1/\sqrt{3}$ it is a maximally entangled two-qutrit state, while for $\alpha = 1/\sqrt{2}$ it is a maximally entangled two-qubit state. For $\alpha \simeq 0.6169$ the CGLMP inequality is maximally violated [35], while no violation is obtained for $\alpha \leqslant \sqrt{3/22} \simeq 0.3693$ using the set of measurements considered. Figure 4 presents bounds on the randomness $G(A|E,\ 1)$, which can be certified in this scenario, for $\sqrt{3/22} \leqslant \alpha \leqslant 1/\sqrt{2}$, taking into account only the CGLMP violation or the full non-local behavior. Unsuprisingly, at the point of maximal violation of the CGLMP inequality, we can certify one trit of randomness,

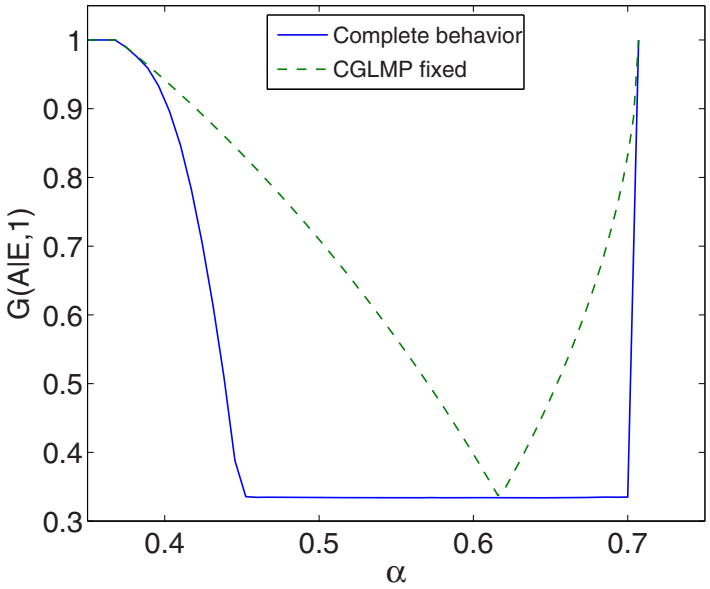

Figure 4. Local DIGP $G(A|E,\ 1)$ as a function of the parameter α defined in equation (12). The dashed curve is obtained by taking into account only the CGLMP value, and the solid one the complete behavior. Both curves were obtained using the second order relaxation of the SDP hierarchy, and the dashed one has been verified to be optimal up to a numerical precision of 10^{-5}.

i.e. $G(A|E,\ 1) = 1/3$. However, taking into account the complete behavior, a large interval of values of α yields $G(A|E,\ 1) = 1/3$, including values for which the CGLMP violation is small. These results have been obtained using the second order relaxation of the SDP hierarchy. The range of values of α for which $G(A|E,\ 1) = 1/3$ may thus turn out to be larger when going to higher order SDP relaxations or using different measurements from those specified in [34].

7. Conclusion

We have shown how the device-independent guessing probability can be evaluated by taking into account in a systematic way the complete non-local behavior characterizing a Bell test and not only the violation of a pre-specified Bell inequality. We have also shown that for any given non-local correlations, there exists an optimal Bell inequality that can certify the maximal amount of randomness compatible with such correlations. Explicit upper bounds on the device-independent guessing probability and their associated Bell inequalities can be computed by adapting the SDP hierarchy introduced in [26]. Low order relaxations, as is often the case with applications of the SDP hierarchy, usually already yield the optimal value of the guessing probability.

Our approach can be straightforwardly adapted to quantify randomness in purely non-signaling settings (i.e. without requiring the validity of quantum theory). The corresponding programs are simply the analogues of equations (4) and (5), where the constraints $\tilde{\mathbf{p}}^a \in \tilde{Q}$ and $p'(a|x^*) \leqslant_Q \mathbf{f}\cdot\mathbf{p}'$ are replaced by $\tilde{\mathbf{p}}^a \in \widetilde{\mathcal{NS}}$ and $p'(a|x^*) \leqslant_{\mathcal{NS}} \mathbf{f}\cdot\mathbf{p}'$, respectively, with $\mathcal{NS}$ denoting the set of non-signaling behavior. Since $\mathcal{NS}$ is entirely characterized by linear constraints (the no-signaling constraints [36] and the positivity of probabilities), these programs can be solved using linear programming.

We expect that the tools that we have presented will contribute to advancing our fundamental understanding of the relation between non-locality and randomness, and its cryptographic applications. In particular, the simple examples that we have studied (especially figures 1, 2 and 4) already yield unexpected results that motivate further investigations. Finally, it would be interesting to understand what is the optimal way to incorporate directly our method in protocols for DIRG and DIQDKD taking into account finite statistics effects.

Acknowledgments

We acknowledge financial support from the European Union under the project QCS, QALGO, DIQIP and from the FRS-FNRS under the project DIQIP. SP acknowledges support from the Brussels-Capital Region through a BB2B grant. JS is chargé de recherches du FRS-FNRS. ONS acknowledges support from the FRS-FNRS under a grant from the Fonds pour la Formation à la Recherche dans l'Industrie et l'Agriculture (FRIA). The Matlab toolboxes YALMIP [38] and SeDuMi [39] were used to solve the SDPs giving rise to the figures in section 6.

Note added. Similar results to our own have been obtained independently and in parallel by J D Bancal, L Sheridan, and V Scarani [37].

References

[1] Brunner N *et al* 2013 arXiv:1303.2849
[2] Valentini A 2002 *Phys. Lett.* A **297** 273
[3] Acín A, Massar S and Pironio S 2012 *Phys. Rev. Lett.* **108** 100402
[4] Colbeck R 2007 *PhD Dissertation* Cambridge University arXiv:0911.3814
Colbeck R and Kent A 2011 *J. Phys. A: Math Theor.* **44** 095305
[5] Pironio S *et al* 2010 *Nature* **464** 1021
[6] Fehr S, Gelles R and Schaffner C 2013 *Phys. Rev.* A **87** 012335
[7] Pironio S and Massar S 2013 *Phys. Rev.* A **87** 012336
[8] Vazirani U and Vidick T 2012 *Phil. Trans. R. Soc.* A **370** 3432
[9] Colbeck R and Renner R 2012 *Nature Phys.* **8** 450
[10] Gallego R *et al* 2012 arXiv:1210.6514
[11] Barrett J, Hardy L and Kent A 2005 *Phys. Rev. Lett.* **95** 010503
[12] Acín A *et al* 2007 *Phys. Rev. Lett.* **98** 230501
Pironio S *et al* 2009 *New J. Phys.* **11** 045021
[13] Mayers D and Yao A 2004 *Quantum Inf. Comput.* **4** 273
[14] Masanes Ll, Pironio S and Acín A 2011 *Nature Commun.* **2** 238
[15] Hänggi E and Renner R 2010 arXiv:1009.1833
[16] Vazirani B U and Vidick T 2012 arXiv:1210.1810
[17] Pironio S *et al* 2013 *Phys. Rev. X* **3** 031007
[18] Clauser J F *et al* 1969 *Phys. Rev. Lett.* **23** 880
[19] Braunstein S L and Caves C M 1990 *Ann. Phys.* **202** 22
[20] Barrett J, Kent A and Pironio S 2007 *Phys. Rev. Lett.* **97** 170409
[21] Mermin D 1990 *Phys. Rev. Lett.* **65** 1838
[22] Mironowicz P and Pawlowski M 2013 *Phys. Rev.* A **88** 032319
[23] Giustina M *et al* 2013 *Nature* **497** 227
[24] Christensen B G *et al* 2013 *Phys. Rev. Lett.* **111** 130406
[25] Eberhard P 1993 *Phys. Rev.* A **47** 747

[26] Navascués M, Pironio S and Acín A 2007 *Phys. Rev. Lett.* **98** 010401
[27] König R, Renner R and Schaffner C 2009 *IEEE Trans. Inf. Theory* **55** 4337
[28] Boyd S and Vandenberghe L 2004 *Convex Optimization* (Cambridge: Cambridge University Press)
[29] Silman J, Pironio S and Massar S 2013 *Phys. Rev. Lett.* **110** 100504
[30] Navascués M, Pironio S and Acín A 2008 *New J. Phys.* **10** 073013
[31] Helton J W 2002 *Ann. Math.* **56** 675
[32] Doherty A C *et al* 2008 *Proc. 23rd IEEE Conf. on Computational Complexity (CCC)* (Los Alamitos, CA: IEEE CS Press) p 199
[33] Pironio S, Navascués M and Acín A 2010 *SIAM J. Optim.* **20** 2157
[34] Collins D *et al* 2002 *Phys. Rev. Lett.* **88** 040404
[35] Acín A *et al* 2002 *Phys. Rev.* A **65** 052325
[36] Barrett J *et al* 2005 *Phys. Rev.* A **71** 022101
[37] Bancal J D, Sheridan L and Scarani V 2013 arXiv:1309.3894
[38] Löfberg J YALMIP: A Toolbox for Modeling and Optimization in MATLAB. Available at http://users.isy.liu.se/johanl/yalmip
[39] Sturm J F and Polik I SeDuMi: a package for conic optimization. Available at http://sedumi.ie.lehigh.edu