entropy

MDPI

Article

Discrimination of Non-Local Correlations

Alberto Montina * and Stefan Wolf *

Facoltà di Informatica, Università della Svizzera italiana, 6900 Lugano, Switzerland
* Correspondence: montia@usi.ch (A.M.); wolfs@usi.ch (S.W.)

Received: 15 December 2018; Accepted: 17 January 2019; Published: 23 January 2019

Abstract: In view of the importance of quantum non-locality in cryptography, quantum computation, and communication complexity, it is crucial to decide whether a given correlation exhibits non-locality or not. As proved by Pitowski, this problem is NP-complete, and is thus computationally intractable unless NP is equal to P. In this paper, we first prove that the Euclidean distance of given correlations from the local polytope can be computed in polynomial time with arbitrary fixed error, granted the access to a certain oracle; namely, given a fixed error, we derive two upper bounds on the running time. The first bound is linear in the number of measurements. The second bound scales with the number of measurements to the sixth power. The former holds only for a very high number of measurements, and is never observed in the performed numerical tests. We, then, introduce a simple algorithm for simulating the oracle. In all of the considered numerical tests, the simulation of the oracle contributes with a multiplicative factor to the overall running time and, thus, does not affect the sixth-power law of the oracle-assisted algorithm.

Keywords: local polytope; quantum nonlocality; communication complexity; optimization

1. Introduction

Non-local correlations, displayed by certain entangled quantum systems, mark a clear departure from the classical framework made up of well-defined, locally interacting quantities [1]. Besides their importance in foundation of quantum theory, non-local correlations have gained interest as information-processing resources in cryptography [2–8], randomness amplification [9,10], quantum computation, and communication complexity [11]. In view of their importance, a relevant problem—hereafter called the *non-locality problem*—is to find a criterion for deciding if observed correlations are actually non-local. Such a criterion is, for example, provided by the *Bell inequalities* [12]. However, a result by Pitowski [12] suggests that the problem of discriminating between local and non-local correlations is generally intractable. Pitowski proved that deciding membership to the correlation polytope is NP-complete, and is therefore intractable unless NP is equal to P. This result also implies that the opposite problem, deciding whether given correlations are outside the polytope, is not even in NP, unless NP=co-NP—which is believed to be false.

In this paper, we present an algorithm whose numerical tests suggest a polynomial running time for all the considered quantum-correlation problems. More precisely, the algorithm computes the distance from the local polytope. First, we prove that the time cost of computing the distance with an arbitrary fixed error grows polynomially in the size of the problem input (number of measurements and outcomes), granted the access to a certain oracle. Namely, given a fixed error, we derive two upper bounds on the running time. The first bound is linear in the number of measurements. The second bound scales with the number of measurements to the sixth power. The former holds only for a very high number of measurements, and is never observed in the performed numerical tests. Thus, the problem of computing the distance is reduced to determining an efficient simulation of the oracle. Then, we introduce a simple algorithm that simulates the oracle. The algorithm is probabilistic and provides the right answer in a subset of randomized inputs. Thus, to have a correct answer with sufficiently high probability, the simulation of the oracle has to be performed with a suitably high number of initial random inputs. In the numerical tests, the number of random initial trials has pragmatically been chosen such that the simulation of the oracle contributes to the overall running time with a multiplicative factor and, thus, does not affect the sixth-power law of the oracle-assisted algorithm. In all of the performed numerical tests, the overall algorithm always computes the distance within the desired accuracy. The scaling of the running time observed in the tests is compatible with the sixth-power law, derived theoretically.

Similar results have independently been published in [13], almost simultaneously to a first version of this paper [14]. The algorithm in [13] is a modification of Gilbert's algorithm for minimizing quadratic forms in a convex set. In its original form, the algorithm uses the following strategy for generating a sequence of points, which converge to the minimizer: Given a point P_n of the sequence, a procedure of linear optimization generates another point Q_n, such that the next point P_{n+1} of the sequence is computed as a convex combination of P_n and Q_n. If the convex set is a polytope, the points $Q_1, \dots$ turn out to be vertices of the polytope. The modified algorithm, introduced in [13], keeps track of the previous vertices $Q_{n-m}, Q_{n-m+1}, \dots, Q_n$, m being some fixed parameter, and computes the next point P_{n+1} as convex combination of these points and P_n. In our algorithm, we compute P_{n+1} as a convex combination of a suitable set of previously computed vertices, without using the point P_n (Section 5). This difference does not result in substantial computational differences. However, our approach has the advantage of keeping track of the minimal number of vertices required for a convex representation of the optimizer. In particular, in the case of local correlations, the algorithm immediately gives a minimal convex representation of them. This representation provides a certificate, which another party can use for directly proving locality. As another minor difference, our algorithm actually computes the distance from what we will call the *local cone*. This allows us to eliminate a normalization constraint from the optimization problem.

The paper is organized as follows. In Section 2, we introduce our general scenario. For the sake of simplicity, we will discuss only the two-party case, but the results can be extended to the general case of many parties. After introducing the local polytope in Section 3, we formulate the non-locality problem as a minimization problem; namely, the problem of computing the distance from the local polytope (Section 4). In Section 5, the algorithm is introduced. The convergence and the computational cost are then discussed in Section 6. After introducing the algorithm for solving the oracle, we finally discuss the numerical results in Section 7.

2. Nonsignaling Box

In a Bell scenario, two quantum systems are prepared in an entangled state and delivered to two spatially separate parties; say, Alice and Bob. These parties each perform a measurement on their system and get an outcome. In general, Alice and Bob are allowed to choose among their respective sets of possible measurements. We assume that the sets are finite, but arbitrarily large. Let us denote the measurements performed by Alice and Bob by the indices $a \in \{1, \dots, A\}$ and $b \in \{1, \dots, B\}$, respectively. After the measurements, Alice gets an outcome $r \in \mathcal{R}$ and Bob an outcome $s \in \mathcal{S}$, where $\mathcal{R}$ and $\mathcal{S}$ are two sets with cardinality R and S, respectively. The overall scenario is described by the joint conditional probability $P(r,s|a,b)$ of getting (r,s), given (a,b). Since the parties are spatially separate, causality and relativity imply that this distribution satisfies the nonsignaling conditions

$$\begin{gathered} P(r|a,b) = P(r|a,\bar{b}) \;\; \forall r,a,b,\bar{b}, \text{ and} \\ P(s|a,b) = P(s|\bar{a},b) \;\; \forall s,b,a,\bar{a}, \end{gathered} \tag{1}$$

where $P(r|a,b) \equiv \sum_s P(r,s|a,b)$ and $P(s|a,b) \equiv \sum_r P(r,s|a,b)$ are the marginal conditional probabilities of r and s, respectively. In the following discussion, we consider a more general scenario than quantum correlations, and we just assume that $P(r,s|a,b)$ satisfies the nonsignaling conditions. The abstract machine producing the correlated variables r and s from the inputs a and b will be called the *nonsignaling box* (briefly, NS-box).

3. Local Polytope

The correlations between the outcomes r and s, associated with the measurements a and b, are *local* if and only if the conditional probability $P(r,s|a,b)$ can be written in the form

$$P(r,s|a,b)=\sum_x P^A(r|a,x)P^B(s|b,x)P^S(x), \tag{2}$$

where P^A, P^B, and P^S are suitable probability distributions. It is always possible to write the conditional probabilities P^A and P^B as convex combination of local deterministic processes, that is,

$$\begin{gathered} P^A(r|a,x)=\textstyle\sum_{\mathbf{r}} P^A_{\text{det}}(r|\mathbf{r},a)\rho^A(\mathbf{r}|x), \text{ and} \\ P^B(s|b,x)=\textstyle\sum_{\mathbf{s}} P^B_{\text{det}}(s|\mathbf{s},b)\rho^B(\mathbf{s}|x), \end{gathered} \tag{3}$$

where $\mathbf{r}\equiv(r_1,\dots,r_A)$, $\mathbf{s}\equiv(s_1,\dots,s_B)$, $P^A_{\text{det}}(r|\mathbf{r},a)=\delta_{r_a,r}$, and $P^B_{\text{det}}(s|\mathbf{s},b)=\delta_{s_b,s}$. Using this decomposition, Equation (2) takes the form of a convex combination of local deterministic distributions. That is,

$$\begin{aligned} P(r,s|a,b) &= \sum_{\mathbf{r},\mathbf{s}} P^A_{\text{det}}(r|\mathbf{r},a)P^B_{\text{det}}(s|\mathbf{s},b)P^{AB}(\mathbf{r},\mathbf{s}) \\ &= \sum_{\mathbf{r},\mathbf{s}} \delta_{r,r_a}\delta_{s,s_b}P^{AB}(\mathbf{r},\mathbf{s}) \\ &= \sum_{\mathbf{r},r_a=r}\ \sum_{\mathbf{s},s_b=s} P^{AB}(\mathbf{r},\mathbf{s}), \end{aligned} \tag{4}$$

where $P^{AB}(\mathbf{r},\mathbf{s})\equiv\sum_x\rho^A(\mathbf{r}|x)\rho^B(\mathbf{s}|x)P^S(x)$ and $\delta_{i,j}$ is the Kronecker delta. Equation (5) is known as Fine's theorem [15]. Thus, a local distribution can always be written as convex combination of local deterministic distributions. Clearly, the converse is also true and a convex combination of local deterministic distributions is local. Therefore, the set of local distributions is a polytope, called a *local polytope*. As the deterministic probability distributions $P^A_{\text{det}}(r|\mathbf{r},a)P^B_{\text{det}}(s|\mathbf{s},b)$ are not convex combinations of other distributions, they all define the vertices of the local polytope. Thus, there are R^AS^B vertices, each one specified by the sequences $\mathbf{r}$ and $\mathbf{s}$. Let us denote the map from $(\mathbf{r},\mathbf{s})$ to the associated vertex by $\vec{V}$. That is, $\vec{V}$ maps the sequences to a deterministic local distribution,

$$\vec{V}(\mathbf{r},\mathbf{s})\equiv P_{det}:(r,s,a,b)\mapsto\delta_{r,r_a}\delta_{s,s_b}. \tag{5}$$

Since the elements of the local polytope are normalized distributions and satisfy the nonsignaling conditions (1), the $RSAB$ parameters defining $P(r,s|a,b)$ are not independent and the polytope lives in a lower-dimensional subspace. The dimension of this subspace and, more generally, of the subspace of NS-boxes, is equal to [16]

$$d_{NS}\equiv AB(R-1)(S-1)+A(R-1)+B(S-1). \tag{6}$$

By the Minkowski–Weyl theorem, the local polytope can be represented as the intersection of finitely many half-spaces. A half-space is defined by an inequality

$$\sum_{r,s,a,b} P(r,s|a,b)B(r,s;a,b)\le L. \tag{7}$$

In the case of the local polytope, these inequalities are called *Bell inequalities*. Given the coefficients $B(r,s;a,b)$, we can choose L such that the inequality is as restrictive as possible. This is attained by imposing that at least one vertex of the local polytope is at the boundary of the half-space; that is, by taking

$$L = \max_{\mathbf{r},\mathbf{s}} \sum_{a,b} B(r_a, s_b; a, b). \tag{8}$$

The oracle, which is central in this work, and introduced later in Section 4, returns the value L from the coefficients $B(r, s; a, b)$.

A minimal representation of a polytope is given by the set of facets of the polytope. A half-space $\sum_{r,s,a,b} P(r,s|a,b)B(r,s;a,b) \leq L$ specifies a facet if the associated hyperplane $\sum_{r,s,a,b} P(r,s|a,b)B(r,s;a,b) = L$ intersects the boundary of the polytope in a set with dimension equal to the dimension of the polytope minus one. A distribution $P(r,s|a,b)$ is local if and only if every facet inequality is not violated. Deciding whether some inequality is violated is generally believed to be intractable, due to a result by Pitowski [12], but to test the membership of a distribution to the local polytope can be done in polynomial time, once the vertices—of which the distribution is a convex combination—are known. Thus, deciding membership to the local polytope is an NP problem. Furthermore, the problem is NP-complete [12].

4. Distance from the Local Polytope

The non-locality problem can be reduced to a convex optimization problem, such as the computation of the nonlocal capacity, introduced in [17], and the distance from the local polytope, which can be reduced to a linear program if the L^1 norm is employed [18]. Here, we define the distance of a distribution $P(r,s|a,b)$ from the local polytope as the Euclidean distance between $P(r,s|a,b)$ and the closest local distribution. As mentioned in Section 3 (see Equation (5)), and stated by Fine's theorem [15], a conditional distribution $\rho(r,s|a,b)$ is local if and only if there is a non-negative function $\chi(\mathbf{r},\mathbf{s})$ such that

$$\rho(r,s|a,b) = \sum_{\mathbf{r},r_a=r} \sum_{\mathbf{s},s_b=s} \chi(\mathbf{r},\mathbf{s}). \tag{9}$$

That is, a conditional distribution $\rho(r,s|a,b)$ is local if it is the marginal of a multivariate probability distribution χ of the outcomes of all the possible measurements, provided that χ does not depend on the measurements a and b.

The distributions $P(r,s|a,b)$ and $\rho(r,s|a,b)$ can be represented as vectors in a $RSAB$-dimensional space. Let us denote them by $\vec{P}$ and $\vec{\rho}$, respectively. Given a positive-definite matrix $\hat{M}$ defining the metrics in the vector space, the computation of the distance from the local polytope is equivalent to the minimization of a functional of the form

$$F[\chi] = \frac{1}{2}\left(\vec{P} - \vec{\rho}\right)^T \hat{M} \left(\vec{P} - \vec{\rho}\right) \tag{10}$$

with respect to χ, under the constraints that χ is non-negative and normalized. Namely, the distance is the square root of the minimum of $2F$. Hereafter, we choose the metrics so that the functional takes the form

$$F[\chi] \equiv \frac{1}{2} \sum_{r,s,a,b} \left[P(r,s|a,b) - \rho(r,s|a,b)\right]^2 W(a,b), \tag{11}$$

where $W(a,b)$ is some probability distribution. The normalization $\sum_{a,b} W(a,b) = 1$ guarantees that the distance does not diverge in the limit of infinite measurements performed on a given entangled state. In particular, we will consider the case with

$$W(a,b) \equiv \frac{1}{AB}. \tag{12}$$

Another choice would be to take the distribution $W(a,b)$ maximizing the functional, so that the computation of the distance would be a minimax problem. This case has some interesting advantages, but is more sophisticated and will not be considered here. Since we are interested in a quantity that is equal to zero if and only if $P(r,s|a,b)$ is local, we can simplify the problem of computing the distance by dropping the normalization constraint on χ. Indeed, if the distance is equal to zero, ρ, and thus χ, are necessarily normalized. Conversely, if the distance is different from zero for every normalized local distribution, it is so also for every unnormalized local distribution. Thus, the discrimination between local and non-local correlation is equivalent to the following minimization problem.

Problem 1.

$$\min_\chi F[\chi]$$

$$\textit{subject to the constraints}$$

$$\chi(\mathbf{r},\mathbf{s}) \ge 0.$$

Let us denote the solution of this problem and the corresponding optimal value by χ^{min} and F^{min}, respectively. The associated (unnormalized) local distribution is denoted by $\rho^{min}(r,s|a,b)$. The square root of $2F^{min}$ is the minimal distance of $P(r,s|a,b)$ from the cone defined as the union of all the lines connecting the zero distribution $\rho(r,s|a,b)=0$ and an arbitrary point of the local polytope. Let us call this set the *local cone*. Hereafter, we will consider the problem of computing the distance from the local cone, but the results can be easily extended to the case of the local polytope, so that we will use “local cone” and “local polytope” as synonyms in the following discussion. Note that there are generally infinite minimizers χ^{min}, since χ lives in a R^AS^B-dimensional space, whereas the functional F depends on χ through $\rho(r,s|a,b)$, which lives in a $(d_{NS}+1)$-dimensional space. In other words, since the local polytope has R^AS^B vertices, but the dimension of the polytope is d_{NS}, a (normalized) distribution ρ has generally infinite representations as convex combination of the vertices, unless ρ is on a face whose dimension plus 1 is equal to the number of vertices defining the face.

At first glance, the computational complexity of this problem seems intrinsically exponential, as the number of real variables defining χ is equal to R^AS^B. However, the dimension of the local polytope is d_{NS} and grows polynomially in the number of measurements and outcomes. Thus, by Carathéodory’s theorem, a (normalized) local distribution can always be represented as the convex combination of a number of vertices smaller than $d_{NS}+2$. This implies that there is a minimizer χ^{min} of F whose support contains a number of elements not greater than $d_{NS}+1$. Therefore, the minimizer can be represented by a number of variables growing polynomially in the input size. The main problem is to find a small set of vertices that are suitable for representing the closest local distribution $\rho^{min}(r,s|a,b)$. In the following, we will show that the computation of the distance from the local cone with an arbitrary fixed accuracy has polynomial complexity, granted the access to the following oracle.

Oracle Max: Given a function $g(r,s;a,b)$, the oracle returns the sequences $\mathbf{r}$ and $\mathbf{s}$ maximizing the function

$$G(\mathbf{r},\mathbf{s}) \equiv \sum_{a,b} g(r_a,s_b;a,b)W(a,b) \tag{13}$$

and the corresponding maximal value.

Thus, Problem 1 is reduced to determining an efficient simulation of the oracle. Let us consider the case of binary outcomes, with r and s taking values ±1 ($R=S=2$). The function $G(\mathbf{r},\mathbf{s})$ takes the form

$$G(\mathbf{r},\mathbf{s}) = \sum_{a,b} J_{ab} r_a s_b + \sum_a A_a r_a + \sum_b B_b s_b + G_0, \tag{14}$$

whose minimization falls into the class of spin-glass problems, which are notoriously computationally hard to handle. This suggests that the oracle is generally an intractable problem. Nonetheless, the oracle has a particular structure that can make the problem easier to be solved, in some instances. This will be discussed later, in Sections 6.3 and 7. There, we will show that the oracle can be simulated efficiently in many relevant cases, by using a simple block-maximization strategy. Assuming for the moment that we have access to the oracle, let us introduce the algorithm solving Problem 1.

5. Computing the Distance

The distance from the local polytope can be computed efficiently, once we have a set Ω of vertices that is small enough and suitable for representing the closest distribution $\rho^{min}(r,s|a,b)$. The algorithm introduced in this paper solves Problem 1 by iteratively generating a sequence of sets Ω. At each step, the minimal distance is first computed over the convex hull of the given vertices. Then, the oracle is consulted. If the set does not contain the right vertices, the oracle returns a strictly positive maximal value and a vertex, which is added to the set Ω (after possibly removing vertices with zero weight). The optimization Problem 1 is solved once the oracle returns zero, which guarantees that all the optimality conditions of the problem are satisfied. Before discussing the algorithm, let us derive these conditions.

5.1. Necessary and Sufficient Conditions for Optimality

Problem 1 is a convex optimization problem whose constraints satisfy Slater's condition, requiring the existence of an interior point of the feasible region. This is the case, as a positive χ strictly satisfies all the inequality constraints. Thus, the four Karush–Kuhn–Tucker (KKT) conditions are necessary and sufficient conditions for optimality. Let us briefly summarize these conditions. Given an objective function $F(\vec{x})$ of the variables $\vec{x}$ and equality constraints $G_{k=1,\dots,n_c}(\vec{x})=0$, it is well known that the function F is stationary at $\vec{x}$ if the gradient of the Lagrangian $\mathcal{L}(\vec{x})\equiv F(\vec{x})-\sum_{k=1}^{n_c}\eta_k G_k(\vec{x})$ is equal to zero, for some value of the Lagrange multipliers η_k. This is the first KKT condition. The second condition is the feasibility of the constraints; that is, the stationary point $\vec{x}$ must satisfy the constraints $G_k(\vec{x})=0$. These two conditions are necessary and sufficient, as there are only equality constraints. If there are also inequalities, two additional conditions on the associated Lagrange multipliers are required. Given inequality constraints $H_k(\vec{x})\ge 0$, with associated Lagrange multipliers λ_k, the third condition is the non-negativity of the multipliers; that is, $\lambda_k\ge 0$. This condition says that the constraint acts only in one direction, like a floor acts on objects through an upward force, but not with a downward force. The last condition states that the Lagrange multiplier λ_k can differ from zero only if the constraint is active; that is, if $H_k(\vec{x})=0$. This is like stating that a floor acts on a body only if they are touching (contact force). This condition can concisely be written as $\lambda_k H_k(\vec{x})=0$.

Let us characterize the optimal solution of Problem 1 through the four KKT conditions.

- First KKT condition (*stationarity condition*): The gradient of the Lagrangian is equal to zero. The Lagrangian of Problem 1 is

$$\mathcal{L}=F[\chi]-\sum_{\mathbf{r},\mathbf{s}}\lambda(\mathbf{r},\mathbf{s})\chi(\mathbf{r},\mathbf{s}), \tag{15}$$

  where $\lambda(\mathbf{r},\mathbf{s})$ are the Lagrange multipliers associated with the inequality constraints.
- Second KKT condition (*feasibility of the constraints*): The function χ is non-negative, $\chi(\mathbf{r},\mathbf{s})\ge 0$.
- Third condition (*dual feasibility*): The Lagrange multipliers λ are non-negative; that is,

$$\lambda(\mathbf{r},\mathbf{s})\ge 0. \tag{16}$$

- Fourth condition (*complementary slackness*): If $\chi(\mathbf{r},\mathbf{s})\neq 0$, then the multiplier $\lambda(\mathbf{r},\mathbf{s})$ is equal to zero; that is,

$$\lambda(\mathbf{r},\mathbf{s})\chi(\mathbf{r},\mathbf{s})=0. \tag{17}$$

The stationarity condition on the gradient of the Lagrangian gives the equality

$$\sum_{a,b}W(a,b)\left[P(r_a,s_b|a,b)-\rho(r_a,s_b|a,b)\right]+\lambda(\mathbf{r},\mathbf{s})=0. \tag{18}$$

Eliminating λ, this equality and the dual feasibility yield the inequality

$$\sum_{a,b} W(a,b)\left[P(r_a,s_b|a,b)-\rho(r_a,s_b|a,b)\right]\le 0. \tag{19}$$

From Equation (18), we have that the complementary slackness is equivalent to the following condition,

$$\begin{gathered}\chi(\mathbf{r},\mathbf{s})\ne 0\Rightarrow\\ \sum_{a,b} W(a,b)\left[P(r_a,s_b|a,b)-\rho(r_a,s_b|a,b)\right]=0;\end{gathered} \tag{20}$$

that is, the left-hand side of the last inequality is equal to zero if $(\mathbf{r},\mathbf{s})$ is in the support of χ. The slackness condition (20), the primal constraint and Equation (19) provide necessary and sufficient conditions for optimality. Let us introduce the function

$$g(r,s;a,b)\equiv P(r,s|a,b)-\rho(r,s|a,b), \tag{21}$$

which is the opposite of the gradient of F with respect to ρ, up to the factor $W(a,b)$. Summarizing, the conditions are

$$\sum_{a,b} W(a,b)g(r_a,s_b;a,b)\le 0, \tag{22}$$

$$\chi(\mathbf{r},\mathbf{s})\ne 0\Rightarrow\sum_{a,b} W(a,b)g(r_a,s_b;a,b)=0, \tag{23}$$

$$\chi(\mathbf{r},\mathbf{s})\ge 0. \tag{24}$$

The second condition can be rewritten in the more concise form

$$\sum_{r,s,a,b}\rho(r,s|a,b)g(r,s|a,b)W(a,b)=0. \tag{25}$$

Indeed, using Equations (22) and (24), it is easy to show that condition (23) is satisfied if and only if

$$\sum_{\mathbf{r},\mathbf{s}}\chi(\mathbf{r},\mathbf{s})\sum_{a,b}W(a,b)g(r_a,s_b;a,b)=0,$$

which gives equality (25), by definition of ρ (Equation (9)).

Condition (22) can be checked, by consulting the oracle with $g(r,s;a,b)$ as the query. If the oracle returns a non-positive maximal value, then the condition is satisfied. Actually, at the optimal point, the returned value turns out to be equal to zero, as implied by the other optimality conditions.

Similar optimality conditions hold if we force χ to be equal to zero outside some set Ω. Let us introduce the following minimization problem.

Problem 2.

$$\begin{gathered}\min_\chi F[\chi]\\ \textit{subject to the constraints}\\ \chi(\mathbf{r},\mathbf{s})\ge 0,\\ \chi(\mathbf{r},\mathbf{s})=0\quad\forall(\mathbf{r},\mathbf{s})\notin\Omega.\end{gathered}$$

The optimal value of this problem gives an upper bound on the optimal value of Problem 1. The two problems are equivalent if the support of a minimizer χ^{min} of Problem 1 is in Ω. The necessary and sufficient conditions for optimality of Problem 2 are the same as of Problem 1, with the only difference that condition (22) has to hold only in the set Ω. That is, the condition is replaced by the weaker condition

$$(\mathbf{r},\mathbf{s})\in\Omega\Rightarrow\sum_{a,b}W(a,b)g(r_a,s_b;a,b)\le 0. \tag{26}$$

Thus, an optimizer of Problem 2 is solution of Problem 1 if the value returned by the oracle with query $g = P - \rho$ is equal to zero.

Hereafter, the minimizer and the minimal value of Problem 2 will be denoted by χ_Ω^{min} and F_Ω^{min}, respectively. The associated optimal local distribution $\rho(r,s|a,b)$, defined by Equation (9), will be denoted by $\rho_\Omega^{min}(r,s|a,b)$.

5.2. Overview of the Algorithm

Problem 1 can be solved iteratively by finding the solution of Problem 2 over a sequence of sets Ω. The sets are built according to the answer of the oracle, which is consulted at each step of the iteration. The procedure stops when a desired accuracy is reached or Ω contains the support of a minimizer χ^{min}, and the solution of Problem 2 is also the solution of Problem 1. Let us outline the algorithm. Suppose that we choose the initial Ω as a set of sequences $(\mathbf{r},\mathbf{s})$ associated to n_0 linearly independent vertices (n_0 being possibly equal to 1). Let us denote this set by Ω_0. We solve Problem 2 with $\Omega = \Omega_0$ and get the optimal value $F_0^{min} \equiv F_{\Omega_0}^{min}$ with minimizer $\chi_0^{min} \equiv \chi_{\Omega_0}^{min}$. Let us denote the corresponding (unnormalized) local distribution by $\rho_0^{min} \equiv \rho_{\Omega_0}^{min}$. That is,

$$\rho_0^{min}(r,s|a,b) \equiv \sum_{\mathbf{r},r_a=r} \sum_{\mathbf{s},s_b=s} \chi_0^{min}(\mathbf{r},\mathbf{s}). \tag{27}$$

Since the cardinality of Ω_0 is not greater than $d_{NS}+1$ and the problem is a convex quadratic optimization problem, the corresponding computational complexity is polynomial. Generally, a numerical algorithm provides an optimizer, up to some arbitrarily small but finite error. In Section 5.5, we will provide a bound on the accuracy required for the solution of Problem 2. For now, let us assume that Problem 2 is solved exactly. If the support of χ^{min} is in Ω_0, F_0^{min} is equal to the optimal value of Problem 1, and we have computed the distance from the local polytope. We can verify if this is the case by checking the first optimality condition (22), as the conditions (23) and (24) are trivially satisfied by the optimizer of Problem 2 for every $(\mathbf{r},\mathbf{s})$. The check is made by consulting the oracle with the function $P(r,s|a,b) - \rho_0^{min}(r,s|a,b)$ as the query. If the oracle returns a maximal value equal to zero, then we have the solution of Problem 1. Note that if the optimal value of Problem 2 is equal to zero, then also the optimal value of the main problem is equal to zero and the conditional distribution $P(r,s|a,b)$ is local. In this case, we have no need of consulting the oracle.

If the optimal value of Problem 2 is different from zero and the oracle returns a maximal value strictly positive, then the minimizer of Problem 2 satisfies all the optimality conditions of Problem 1, except Equation (22) for some $(\mathbf{r},\mathbf{s}) \notin \Omega$. The next step is to add the pair of sequences $(\mathbf{r},\mathbf{s})$ returned by the oracle to the set Ω and solve Problem 2 with the new set. Let us denote the new set and the corresponding optimal value by Ω_1 and $F_1^{min} \equiv F_{\Omega_1}^{min}$, respectively. Once we have solved Problem 2 with $\Omega = \Omega_1$, we consult again the oracle to check if we have obtained the solution of Problem 1. If we have not, we add the pair of sequences $(\mathbf{r},\mathbf{s})$ given by the oracle to the set Ω and we solve Problem 2 with the new set, say Ω_2. We continue until we get the solution of Problem 1 or its optimal value up to some desired accuracy. This procedure generates a sequence of sets $\Omega_{n=1,2,\dots}$ and values $F_{n=1,2,\dots}^{min}$. The latter sequence is strictly decreasing, that is, $F_{n+1}^{min} < F_n^{min}$ until Ω_n contains the support of χ^{min} and the oracle returns zero as maximal value. Let us show that. Suppose that χ_n^{min} is the optimizer of Problem 2 with $\Omega = \Omega_n$ and $(\mathbf{r}',\mathbf{s}')$ is the new element in the set Ω_{n+1}. Let us denote by $\rho_n^{min}(r,s|a,b)$ the local distribution associated with χ_n^{min}, that is,

$$\rho_n^{min}(r,s|a,b) \equiv \sum_{\mathbf{r},r_a=r} \sum_{\mathbf{s},s_b=s} \chi_n^{min}(\mathbf{r},\mathbf{s}). \tag{28}$$

The optimal value F_{n+1}^{min} of Problem 2 is bounded from above by the value taken by the function $F[\chi]$ for every feasible χ, in particular, for

$$\chi(\mathbf{r},\mathbf{s};\alpha) = \chi_n^{min}(\mathbf{r},\mathbf{s}) + \alpha\delta_{\mathbf{r},\mathbf{r}'}\delta_{\mathbf{s},\mathbf{s}'}, \tag{29}$$

with α positive. Let us set α equal to the value minimizing F; that is,

$$\alpha \equiv \alpha_n = \sum_{ab} W(a,b)[P(r'_a, s'_b|a,b) - \rho_n^{min}(r'_a, s'_b|a,b)], \tag{30}$$

which is equal to the value returned by the oracle. It is strictly positive, as the oracle returned a positive value—provided that Ω_n does not contain the support of χ^{min}. Hence, $\chi(\mathbf{r},\mathbf{s};\alpha_n)$ is a feasible point and, thus, the corresponding value taken by F,

$$F|_{\alpha=\alpha_n} = F_n^{min} - \frac{1}{2}\alpha_n^2, \tag{31}$$

is an upper bound on F_{n+1}^{min}. Hence,

$$F_{n+1}^{min} \le F_n^{min} - \frac{1}{2}\alpha_n^2, \tag{32}$$

that is, F_{n+1}^{min} is strictly smaller than F_n^{min}.

This procedure generates a sequence F_n^{min} that converges to the optimal value of Problem 1, as shown in Section 6. For any given accuracy, the computational cost of the procedure is polynomial, provided that we have access to the oracle.

To avoid growth of the cardinality of Ω beyond $d_{NS}+1$ during the iteration and, thus, the introduction of redundant vertices, we have to be sure that the sets $\Omega_0, \Omega_1, \dots$ contain points $(\mathbf{r},\mathbf{s})$ associated to linearly independent vertices $\vec{V}(\mathbf{r},\mathbf{s})$ of the local polytope. This is guaranteed by the following procedure of cleaning up. First, after the computation of χ_n^{min} at step n, we remove the elements in Ω_n where $\chi_n^{min}(\mathbf{r},\mathbf{s})$ is equal to zero (this can be checked even if the exact χ_n^{min} is not known, as discussed later in Section 5.6). Let us denote the resulting set by Ω_n^{clean}. Then, the set Ω_{n+1} is built by adding the point given by the oracle to the set Ω_n^{clean}. Let us denote by $\mathcal{V}$ the set of vertices associated to the elements in the support of χ_n^{min}. The cleaning up ensures that the optimizer ρ_n^{min} is in the interior of the convex hull of $\mathcal{V}$, up to a normalization constant, and the new vertex returned by the oracle is linearly independent of the ones in $\mathcal{V}$. Indeed, we have seen that the introduction of such a vertex allows us to lower the optimal value of Problem 2. This would not be possible if the added vertex was linearly dependent on the vertices in $\mathcal{V}$, as the (normalized) optimizer ρ_n^{min} of Problem 2 is in the interior of the convex hull of $\mathcal{V}$.

This is formalized in Lemma 1.

Lemma 1. *Let $(\mathbf{r}',\mathbf{s}')$ be a sequence such that*

$$\sum_{a,b} g(r'_a, s'_b; a,b)W(a,b) \ne 0. \tag{33}$$

If Ω is a set such that

$$(\mathbf{r},\mathbf{s}) = \Omega \Rightarrow \sum_{a,b} g(r_a, s_b; a,b)W(a,b) = 0, \tag{34}$$

then the vertex $\vec{V}(\mathbf{r}',\mathbf{s}')$ is linearly independent of the vertices associated to the sequences in Ω.

Proof. The proof is by contradiction. Suppose that the vector $\vec{V}(\mathbf{r}',\mathbf{s}')$ is linearly dependent with the vectors $\vec{V}(\mathbf{r},\mathbf{s})$ with $(\mathbf{r},\mathbf{s})\in\Omega$, then there is a real function $t(\mathbf{r},\mathbf{s})$ such that

$$\vec{V}(\mathbf{r}',\mathbf{s}')=\sum_{(\mathbf{r},\mathbf{s})\in\Omega} t(\mathbf{r},\mathbf{s})\vec{V}(\mathbf{r},\mathbf{s}). \tag{35}$$

By definition of $\vec{V}$, this equation implies that $\sum_{\mathbf{r},\mathbf{s}} t(\mathbf{r},\mathbf{s})\delta_{r,r_a}\delta_{s,s_b}=\delta_{r,r'_a}\delta_{s,s'_b}$. From this equation and Equation (34), we have

$$\sum_{r,s}\delta_{r,r'_a}\delta_{s,s'_b}\sum_{a,b} g(r,s;a,b)W(a,b)=0. \tag{36}$$

Summing over r and s, we get a contradiction with Equation (33). □

This lemma and the optimality conditions (22) and (23) imply that the sets $\Omega_0,\Omega_1,\dots,$ built through the previously discussed procedure of cleaning up, always contain points associated to independent vertices and, thus, never contain more than $d_{NS}+1$ elements. Indeed, the set Ω_n^{clean} contains points $(\mathbf{r},\mathbf{s})$ where the minimizer χ_n^{min} is different from zero, for which

$$\sum_{a,b}\left[P(r_a,s_b|a,b)-\rho_n^{min}(r_a,s_b|a,b)\right]W(a,b)=0,$$

as implied by condition (23). Furthermore, given the sequence $(\mathbf{r}',\mathbf{s}')$ returned by the oracle, condition (22) implies that

$$\sum_{a,b}\left[P(r'_a,s'_b|a,b)-\rho_n^{min}(r'_a,s'_b|a,b)\right]W(a,b)>0$$

until the set Ω_n contains the support of χ^{min} and the iteration generating the sequence of sets Ω is terminated.

The procedure of cleaning up is not strictly necessary for having a polynomial running time, but it can speed up the algorithm. Furthermore, the procedure guarantees that the distribution $\rho(r,s|a,b)$ approaching the minimizer during the iterative computation is always represented as the convex combination of a minimal number of vertices. Thus, we have a minimal representation of the distribution at each stage of the iteration.

5.3. The Algorithm

In short, the algorithm for computing the distance from the local polytope with given accuracy is as follows.

Algorithm 1. *Input:* $P(r,s|a,b)$

```
1. Set $(\mathbf{r}',\mathbf{s}')$ equal to the sequences given by the oracle with $P(r,s|a,b)$ as query.
2. Set $\Omega=\{(\mathbf{r}',\mathbf{s}')\}$.
3. Compute the optimizers $\chi(\mathbf{r},\mathbf{s})$ and $\rho(r,s|a,b)$ of Problem 2. The associated $F$ provides an upper bound of the optimal value $F^{min}$.
4. Consult the oracle with $g(r,s;a,b)=P(r,s|a,b)-\rho(r,s|a,b)$ as query. Set $(\mathbf{r}',\mathbf{s}')$ and $\alpha$ are equal to the sequences returned by the oracle and the associated maximal value, respectively. That is,
       $(\mathbf{r}',\mathbf{s}')=argmax_{(\mathbf{r},\mathbf{s})}\sum_{a,b} g(r_a,s_b;a,b)W(a,b),$
       $\alpha=\sum_{a,b} g(r'_a,s'_b|a,b)W(a,b),$
5. Compute a lower bound on the $F^{min}$ from $\rho$ and $\alpha$ (see following discussion and Section 6.1). The difference between the upper and lower bounds provides an upper bound on the reached accuracy.
```

6. *If a given accuracy is reached, stop.*
7. *Remove from Ω the points where χ is zero and add $(\mathbf{r}', \mathbf{s}')$.*
8. *Go back to Step 3.*

The algorithm stops at Step 6 when a desired accuracy is reached. To estimate the accuracy, we need to compute a lower bound on the optimal value F^{min}. To guarantee that the algorithm eventually stops, the lower bound has to converge to the optimal value as the algorithm approaches the solution of Problem 1. We also need a stopping criterion for the numerical routine solving the optimization problem in Step 3. Let us first discuss the stopping criterion for Algorithm 1.

5.4. Stopping Criterion for Algorithm 1

The lower bound on F^{min}, denoted by $F^{(-)}$, is computed by using the dual form of Problem 1. As shown in Section 6.1, any local distribution ρ induces the lower bound

$$F^{(-)} = \frac{1}{2} \sum_{rsab} \left\{ P^2(r,s|a,b) - \left[\rho(r,s|a,b) + \alpha \right]^2 \right\} W(a,b), \tag{37}$$

where α is the maximal value returned by the oracle with $g(r,s;a,b) = P(r,s|a,b) - \rho(r,s|a,b)$ as query. An upper bound on F^{min} is obviously

$$F^{(+)} = F[\chi]. \tag{38}$$

In the limit of ρ equal to the local distribution minimizing F, the lower bound is equal to the optimal value F^{min}. This can be shown by using the optimality conditions. Indeed, conditions (22) and (25) imply the limits

$$\lim_{\chi \to \chi^{min}} \alpha = 0, \tag{39}$$

$$\lim_{\chi \to \chi^{min}} \sum_{r,s,a,b} \rho(r,s|a,b) g(r,s;a,b) W(a,b) = 0, \tag{40}$$

which imply $F^{(-)} \to F^{min}$ as χ approaches the minimizer. This is made even more evident, by computing the difference between the upper bound and the lower bound. Indeed, given the local distribution $\rho(r,s|a,b)$ computed at Step 3 and the corresponding α returned by the oracle at Step 4, the difference is

$$\begin{gathered} F^{(+)} - F^{(-)} \equiv \Delta F = \tfrac{RS}{2} \alpha^2 + \\ \textstyle\sum_{rsab} \rho(r,s|a,b) \left[\alpha - g(r,s;a,b) \right] W(a,b), \end{gathered} \tag{41}$$

which evidently goes to zero as χ goes to χ^{min}. Thus, the upper bound ΔF on the accuracy computed in Step 5 goes to zero as $\rho(r,s|a,b)$ approaches the solution. This guarantees that the algorithm stops sooner or later at Step 6, provided that χ converges to the solution. If Problem 2 is solved exactly at Step 3, then the distribution $\rho(r,s|a,b)$ satisfies condition (25), and the upper bound on the reached accuracy takes the form

$$F^{(+)} - F^{(-)} = \frac{RS}{2} \alpha^2 + \alpha \sum_{rsab} \rho(r,s|a,b) W(a,b). \tag{42}$$

Even if Condition (25) is not satisfied, we can suitably normalize $\chi(\mathbf{r}, \mathbf{s})$ so that the condition is satisfied.

In the following, we assume that this condition is satisfied.

5.5. Stopping Criterion for Problem 2 (Optimization at Step 3 of Algorithm 1)

In Algorithm 1, Step 3 is completed when the solution of Problem 2 with a given set Ω is found. Optimization algorithms iteratively find a solution $\rho_\Omega^{min}(r,s|a,b)$ up to some accuracy. We can stop when the error is of the order of the machine precision. Here, we will discuss a more effective stopping criterion. This criterion should preserve the two main features previously described:

- The sequence $F_0^{min}, F_1^{min}, \dots$ of the exact optimal values of Problem 2, with $\Omega = \Omega_0, \Omega_1, \dots,$ is monotonically decreasing.
- The sets $\Omega_0, \Omega_1, \dots$ contain points associated with linearly independent vertices of the local polytope, implying that the cardinality of Ω_n is never greater than $d_{NS}+1$.

To guarantee that the first feature is preserved, it is sufficient to compute a lower bound on F_Ω^{min} from a given χ so that the bound approaches F_Ω^{min} as χ approaches the optimizer χ_Ω^{min}. If the lower bound with the set $\Omega = \Omega_n$ is greater than the upper bound $F_n - \alpha_n^2/2$ on F_{n+1}^{min} (see Equation (31)), then $F_{n+1}^{min} < F_n^{min}$. Denoting by $F_\Omega^{(-)}$ the lower bound on the optimal value F_Ω^{min}, the monotonicity of the sequence $F_0^{min}, F_1^{min}, \dots$ is implied by the inequality

$$F_n - \frac{1}{2}\alpha_n^2 \le F_{\Omega_n}^{(-)}. \tag{43}$$

As shown later, by using dual theory, a lower bound on F_Ω^{min} is

$$F_\Omega^{(-)} = \frac{1}{2}\sum_{rsab}\left\{P^2(r,s|a,b) - \left[\rho(r,s|a,b)+\beta\right]^2\right\} W(a,b), \tag{44}$$

where

$$\beta \equiv \max_{(\mathbf{r},\mathbf{s})\in\Omega}\sum_{ab} W(a,b)[P(r_a,s_b|a,b)-\rho(r_a,s_b|a,b)], \tag{45}$$

and $\rho(r,s|a,b)$ is an unnormalized local distribution, associated to a function $\chi(\mathbf{r},\mathbf{s})$ with support in Ω. This bound becomes equal to F_Ω^{min} in the limit of ρ equal to the minimizer of Problem 2. Equation (43) gives the condition

$$\begin{gathered}\alpha^2 > RS\beta^2 + \\ 2\textstyle\sum_{rsab}\left[\beta - g(r,s;a,b)\right]\rho(r,s|a,b)W(a,b),\end{gathered} \tag{46}$$

where $g(r,s;a,b) = P(r,s|a,b) - \rho(r,s|a,b)$ and $\rho(r,s|a,b)$ is the local distribution computed in Step 3. If this condition is satisfied by the numerical solution found in Step 3, then the series $F_0^{min}, F_1^{min}, \dots$ is monotonically decreasing. As we will see, to prove that the series converges to the minimizer of Problem 1, we need the stronger condition

$$\begin{gathered}\gamma\alpha^2 \ge RS\beta^2 + \\ 2\textstyle\sum_{rsab}\left[\beta - g(r,s;a,b)\right]\rho(r,s|a,b)W(a,b),\end{gathered} \tag{47}$$

where γ is any fixed real number in the interval $(0,1)$. A possible choice is $\gamma = 1/2$. If this inequality is satisfied in each iteration of Algorithm 1, the sequence $F_0^{min}, F_1^{min}, \dots$ satisfies the inequality

$$F_{n+1}^{min} \le F_n^{min} - \frac{1-\gamma}{2}\alpha_n^2, \tag{48}$$

which turns out to be equal to Equation (32) in the limit $\gamma \to 0$. The right-hand side of Equation (47) goes to zero as ρ approaches the optimizer, as implied by the optimality conditions of Problem 2. Thus, if the set Ω does not contain all the points where χ^{min} is different from zero, then the inequality is surely satisfied at some point of the iteration solving Problem 2, as α tends to a strictly positive number. When the inequality is satisfied, the minimization at Step 3 of Algorithm 1 is terminated. If Ω is the support of χ^{min}, the inequality will never be satisfied and the minimization at Step 3 will terminate when the desired accuracy on F^{min} is reached.

5.6. Cleaning Up (Step 7)

As previously said, we should also guarantee that the sets Ω_n contain only points associated with linearly independent vertices. This is granted if the procedure in Step 7 of Algorithm 1 successfully removes the points where the exact minimizer χ_n^{min} is equal to zero. How can we find the support of the minimizer from the approximate numerical solution computed in Step 3? Using dual theory, it is possible to prove the following.

Theorem 1. *Let $\chi(\mathbf{r},\mathbf{s})$ be a non-negative function with support in Ω and $\rho(r,s|a,b)$ be the associated unnormalized local distribution. Then, the inequality*

$$\begin{array}{c} \sum_{a,b}\rho_\Omega^{min}(r_a,s_b|a,b)W(a,b)\ge\sum_{a,b}\rho(r_a,s_b|a,b)W(a,b) \\ -\left[2\left(F^{(+)}-F_\Omega^{(-)}\right)\right]^{1/2} \end{array} \tag{49}$$

holds.

A direct consequence of this theorem and the slackness condition (23) for optimality is the following.

Corollary 1. *Let $\chi(\mathbf{r},\mathbf{s})$ be a non-negative function with support in Ω and $\rho(r,s|a,b)$ the associated unnormalized local distribution. If the inequality*

$$\begin{array}{c} \sum_{ab} g(r_a,s_b;a,b)\le\{RS\beta^2+ \\ 2\sum_{rsab}\left[\beta-g(r,s;a,b)\right]\rho(r,s|a,b)W(a,b)\}^{1/2} \end{array} \tag{50}$$

holds, with $g(r,s|a,b)=P(r,s|a,b)-\rho(r,s|a,b)$, then $\chi_\Omega^{min}(\mathbf{r},\mathbf{s})$ is equal to zero.

Condition (50) is sufficient for having $\chi_\Omega^{min}(\mathbf{r},\mathbf{s})$ equal to zero, but it is not necessary. A necessary condition can be derived by computing the lowest eigenvalue of the Hessian of the objective function $F[\chi]$. Both the necessary and sufficient conditions allow us to determine the support of the minimizer χ_Ω^{min} once the distribution χ is enough close to χ_Ω^{min}. Thus, the minimization in Step 3 should not stop until each sequence $(\mathbf{r},\mathbf{s})$ satisfies the sufficient condition or does not satisfy the necessary condition, otherwise the cleaning up could miss some points where the minimizer is equal to zero. However, numerical experiments show that the use of these conditions is not necessary, and the number of elements in the sets Ω_n is generally bounded by $d_{NS}+1$, provided that Problem 2 is solved by the algorithm described in the following section.

5.7. Solving Problem 2

There are standard methods for solving Problem 2, and numerical libraries are available. The interior point method [19] provides a quadratic convergence to the solution, meaning that the number of digits of accuracy is almost doubled at each iteration step, once χ is sufficiently close to the minimizer. The algorithm uses the Newton method and needs to solve a set of linear equations. Since this can be computationally demanding in terms of memory, we have implemented the solver by using the conjugate gradient method, which does not use the Hessian. Furthermore, if the Hessian turns out to have a small condition number, the conjugate gradient method can be much more efficient than the Newton method, especially if we do not need to solve Problem 2 with high accuracy. This is the case in the initial stage of the computation, when the set Ω is growing and does not contain all the points of the support of χ^{min}.

The conjugate gradient method iteratively performs a one-dimensional minimization, along directions that are conjugate with respect to the Hessian of the objective function [19]. The directions are computed iteratively, by setting the first direction equal to the gradient of the objective function. The conjugate gradient method is generally used with unconstrained problems, whereas Problem 2 has the inequality constraints $\chi(\mathbf{r},\mathbf{s}) \geq 0$. To adapt the method to our problem, we perform the one-dimensional minimization in the region where χ is non-negative. Whenever an inactive constraint becomes active, or vice versa, we set the search direction equal to the gradient and restart the generation of the directions from that point. Once the procedure terminates, the algorithm provides a list of active constraints with $\chi_n(\mathbf{r},\mathbf{s}) = 0$. Numerical simulations show that this list is generally complete, and corresponds to the points where the minimizer χ_n^{min} is equal to zero.

In general, the slackness condition (25) is not satisfied by the numerical solution. However, as previously pointed out, we can suitably normalize χ_n so that this condition is satisfied by $\rho_n(r,s|a,b)$. Thus, we will assume that the equality

$$\sum_{r,s,a,b} \rho_n(r,s|a,b) g_n(r,s;a,b) W(a,b) = 0 \tag{51}$$

holds with $g_n = P - \rho_n$. This also implies that

$$\begin{aligned} \alpha_n = \alpha|_{\rho=\rho_n} \geq 0 \text{ and} \\ \beta_n = \beta|_{\rho=\rho_n} \geq 0. \end{aligned} \tag{52}$$

6. Convergence Analysis and Computational Cost

Here, we provide a convergence analysis and show that the error on the distance from the local polytope is bounded above by a function decaying at least as fast as $1/n$, where n is the number of iterations. The convergence of this function to zero is sublinear, but its derivation relies on a very rough estimate of a lower bound on the optimal value χ^{min}. Actually, the iteration converges to the solution in a finite number of steps (up to the accuracy of the solver of Problem 2). Indeed, since the number of vertices is finite, also the number of their sets Ω is finite. Thus, the sequence Ω_n converges to the support of the optimizer χ^{min} in a finite number of steps, as the accuracy goes to zero.

We expect that this finite number of steps is of the order of the dimension d_{NS} of the local polytope. Interestingly, the computed bound on the number of required iterations for given error does not depend on the number of measurements. Using this bound, we show that the computational cost for any given error on the distance grows polynomially with the size of the problem input; that is, with A, B, R, and S, provided that the oracle can be simulated in polynomial time.

To prove the convergence, we need to introduce the dual form of Problem 1 (see Ref. [19] for an introduction to dual theory). The dual form of a minimization problem (primal problem) is a maximization problem, whose maximum is always smaller than or equal to the primal minimum, the difference being called the *duality gap*. However, if the constraints of the primal problem satisfy some mild conditions, such as Slater's conditions [19], then the duality gap is equal to zero. As previously said, this is the case of Problem 1.

The dual form is particularly useful for evaluating lower bounds on the optimal value of the primal problem. Indeed, the value taken by the dual objective function in a feasible point of the dual constraints provides such a bound. After introducing the dual form of Problem 1, we derive the lower bound $F^{(-)}$ on F^{min}, given by Equation (37). Then, we use this bound and Equation (48) to prove the convergence.

6.1. Dual Problem

The dual problem of Problem 1 is a maximization problem over the space of values taken by the Lagrange multipliers $\lambda(\mathbf{r},\mathbf{s})$ subject to the dual constraints $\lambda(\mathbf{r},\mathbf{s}) \geq 0$. The dual objective function is given by the minimum of the Lagrangian $\mathcal{L}$, defined by Equation (15), with respect to χ. The dual constraint is the non-negativity of the Lagrange multipliers, that is,

$$\lambda(\mathbf{r},\mathbf{s}) \geq 0. \tag{53}$$

As this minimum cannot be derived analytically, a standard strategy for getting an explicit form of the dual objective function is to enlarge the space of primal variables and, correspondingly, to increase the number of primal constraints. The minimum is then evaluated over the enlarged space. In our case, it is convenient to introduce Equation (9) and $\rho(r,s|a,b)$ as additional constraints and variables, respectively. Thus, F is made independent of χ and expressed as a function of ρ. The new optimization problem, which is equivalent to Problem 1, has Lagrangian

$$\mathcal{L} = F[\rho] - \sum_{\mathbf{r},\mathbf{s}} \lambda(\mathbf{r},\mathbf{s})\chi(\mathbf{r},\mathbf{s}) + \sum_{rsab} W(a,b) \times \eta(r,s,a,b) \left[\rho(r,s|a,b) - \sum_{\mathbf{r},\mathbf{s}} \delta_{r,r_a}\delta_{s,s_b}\chi(\mathbf{r},\mathbf{s})\right], \tag{54}$$

where $\eta(r,s,a,b)$ are the Lagrange multipliers associated with the added constraints. To find the minimum of the Lagrangian, we set its derivative, with respect to the primal variables χ and ρ, equal to zero. We get the equations

$$\textstyle\sum_{a,b} W(a,b)\eta(r_a,s_b,a,b) = -\lambda(\mathbf{r},\mathbf{s}) \tag{55}$$

$$\rho(r,s|a,b) = P(r,s|a,b) - \eta(r,s,a,b). \tag{56}$$

The first equation does not depend on the primal variables and sets a constraint on the dual variables. If this constraint is not satisfied, the dual objective function is equal to $-\infty$. Thus, its maximum is in the region where Equation (55) is satisfied. Let us add it to the dual constraint (53). The second stationarity condition, Equation (56), gives the optimal ρ. By replacing it in the Lagrangian, we get the dual objective function

$$F_{dual} = \sum_{r,s,a,b} W(a,b)\eta(r,s,a,b) \times \left[P(r,s|a,b) - \frac{\eta(r,s,a,b)}{2}\right]. \tag{57}$$

Eliminating λ, which does not appear in the objective function, the dual constraints (53) and (55) give the inequality

$$\sum_{a,b} W(a,b)\eta(r_a,s_b;a,b) \le 0. \tag{58}$$

Thus, Problem 1 is equivalent to the following.

Problem 3 (dual problem of Problem 1)**.**

$$\max_\eta F_{dual}[\eta]$$
$$\textit{subject to the constraints}$$
$$\textstyle\sum_{a,b} W(a,b)\eta(r_a,s_b;a,b) \le 0.$$

The value taken by F_{dual} at a feasible point provides a lower bound on F^{min}. Given any function $\bar\eta(r,s;a,b)$, a feasible point is

$$\eta_f(r,s;a,b) \equiv \bar\eta(r,s;a,b) - \max_{\mathbf{r},\mathbf{s}} \sum_{\bar a,\bar b} W(\bar a,\bar b)\bar\eta(r_{\bar a},s_{\bar b};\bar a,\bar b). \tag{59}$$

Indeed,

$$\begin{aligned}\textstyle\sum_{a,b} \eta_f(r_a,s_b;a,b)W(a,b) &= \textstyle\sum_{a,b} \bar\eta(r_a,s_b;a,b) \\ &- \max_{\mathbf{r}',\mathbf{s}'} \textstyle\sum_{a,b} \bar\eta(r'_a,s'_b;a,b)W(a,b) \le 0.\end{aligned} \tag{60}$$

The lower bound turns out to be the optimal value F^{min}, if the distribution $\rho(r,s|a,b)$ given by Equation (56) in terms of $\eta = \eta_f$ is solution of the primal Problem 1. This suggests the transformation

$$\bar\eta(r,s;a,b) = P(r,s|a,b) - \rho(r,s|a,b), \tag{61}$$

where $\rho(r,s|a,b)$ is some local distribution up to a normalization constant (in fact, ρ can be any real function). Every local distribution induces a lower bound on the optimal value F^{min}. This lower bound turns out to be an accurate approximation of F^{min} if ρ is close enough to the optimal local distribution. Using the last equation and Equation (59), we get the lower bound (37) from F_{dual}.

The dual problem of Problem 2 is similar to Problem 3, but the constraints have to hold for sequences $(\mathbf{r},\mathbf{s})$ in Ω.

Problem 4 (dual problem of Problem 2)**.**

$$\begin{gathered}\max_\eta F_{dual}[\eta] \\ \textit{subject to the constraints} \\ (\mathbf{r},\mathbf{s})\in\Omega \Rightarrow \textstyle\sum_{a,b} W(a,b)\eta(r_a,s_b;a,b)\le 0.\end{gathered}$$

This dual problem induces the lower bound F_Ω^{min} on the optimal value of Problem 2 (Equation (44)).

6.2. Convergence and Polynomial Cost

Let $\rho_n(r,s|a,b)$ be the local distribution computed in Step 3 of Algorithm 1. From the lower bound (37), we have

$$F^{min}\ge F_n-\frac{RS}{2}\alpha_n^2+\sum_{r,s,a,b}W(a,b)\rho_n(r,s|a,b)\left[g_n(r,s;a,b)-\alpha_n\right], \tag{62}$$

where α_n is given by Equation (30), and $g_n = P-\rho_n$. The part of the summation linear in g_n is equal to zero, by Equation (51). The remaining part, linear in α_n, is bounded from below by $-\alpha_n[1+(RS)^{1/2}]$ (α_n is positive). This can be shown by minimizing it under the constraint (51). Thus, we have that

$$F^{min}\ge F_n-\frac{RS}{2}\alpha_n^2-[1+(RS)^{1/2}]\alpha_n. \tag{63}$$

As α_n is not greater than 1, the factor α_n^2 in the right-hand side of the inequality can be replaced by α_n, so that we have

$$\alpha_n\ge 2\frac{F_n-F^{min}}{RS+2+2(RS)^{1/2}}, \tag{64}$$

which gives, with Equation (48), the following

$$F_n^{min}-F_{n+1}^{min}\ge 2(1-\gamma)\left(\frac{F_n-F^{min}}{RS+2+2(RS)^{1/2}}\right)^2. \tag{65}$$

This inequality implies that

$$F_n^{min}-F^{min}\le\frac{(RS+2+2(RS)^{1/2})^2}{2(1-\gamma)n}. \tag{66}$$

This can be proved by induction. It is easy to prove that inequality holds for $n=n_0>1$, if it holds for $n=n_0-1$. Let us prove that it holds for $n=1$. It is sufficient to prove that $F_1^{min}-F^{min}\le 1/2$. Using the identity

$$\sum_{r,s,a,b}W(a,b)\rho_1^{min}(r,s|a,b)\times\left[P(r,s|a,b)-\rho_1^{min}(r,s|a,b)\right]=0, \tag{67}$$

we have

$$\begin{aligned}&F_1^{min}-F^{min}\le F_1^{min}=\\ &\textstyle\sum_{r,s,a,b}W(a,b)\frac{\left[P(r,s|a,b)-\rho_1^{min}(r,s|a,b)\right]^2}{2}\\ &=\textstyle\sum_{r,s,a,b}W(a,b)\frac{P^2(r,s|a,b)-(\rho_1^{\min})^2(r,s|a,b)}{2}\\ &\le\textstyle\sum_{r,s,a,b}W(a,b)\frac{P^2(r,s|a,b)}{2}\le\frac{1}{2}.\end{aligned} \tag{68}$$

Thus, the error decreases at least as fast as $1/n$. Although the convergence of the upper bound is sublinear, we derived this inequality by using Equation (63), which provides a quite loose bound on the optimal value χ^{min}. Nonetheless, the constraint set by Equation (66) on the accuracy is strong enough to imply the polynomial convergence of the algorithm, provided that the oracle can be simulated in polynomial time. Indeed, the inequality implies that the number of steps required to reach a given accuracy does not grow faster than $(RS)^2$. Since the computational cost of completing each step is polynomial, the overall algorithm has polynomial cost. More precisely, each step is completed by solving a quadratic minimization problem. If we do not rely on the specific structure of the quadratic problem, its computational cost does not grow faster than $\max\{n_1^3, n_1^2 n_2, D\}$ [19], where n_1, n_2, and D are the number of variables, the number of constraints, and the cost of evaluating the first and second derivatives of the objective and constraint functions. The numbers n_1 and n_2 are equal, and D is equal to $n_1^2(A+B)$. As the number of vertices in the set Ω_n is not greater than the number of iterations (say, $\bar{n}$), we have that $n_1 \le \bar{n}$. Furthermore, the number of vertices cannot be greater than d_{NS}. Thus, the number of variables is, in the worst case,

$$n_1 = \min\{\bar{n}, ABRS\}. \tag{69}$$

As implied by Equation (66), about $(RS)^2/\epsilon$ iterations are sufficient for reaching an error not greater than ϵ. Let us set $\bar{n} = (RS)^2/\epsilon$. Denoting the computational cost of Algorithm 1 with accuracy ϵ by $\mathcal{C}_\epsilon$, we have that

$$\mathcal{C}_\epsilon \le K\bar{n}\max\{n_1^3, n_1^2(A+B)\} = K\frac{(RS)^2}{\epsilon} n_1^2 \max\{n_1, A+B\}, \tag{70}$$

where K is some constant. Let us consider the two limiting cases with $\epsilon(A+B) \ge (RS)^2$ (high number of measurements) and $\epsilon ABRS \le (RS)^2$ (high accuracy).

In the first case, we have that $A+B \ge \bar{n}$, which also implies that $n_1 = \bar{n}$ (there are at least 2 measurements per party). We have

$$\epsilon(A+B) \ge (RS)^2 \Longrightarrow \mathcal{C}_\epsilon \le K\frac{(RS)^6}{\epsilon^3}(A+B) \equiv \mathcal{B}_0 \tag{71}$$

Thus, given a fixed error, the computational cost is asymptotically linear in the number of measurements. For $\epsilon = 10^{-2}$ and $R = S = 2$, this bound holds for a number of measurements per party greater than 800. If $A = B = 800$, the computation ends in few hours in the worst case by using available personal computers, provided that the bound $\mathcal{B}_0$ is saturated in the most pessimistic scenario.

In the second case, we have that $ABRS \le \bar{n}$ and $n_1 = ABRS$. Thus,

$$\epsilon AB \le RS \Longrightarrow \mathcal{C}_\epsilon \le K\frac{(RS)^2}{\epsilon}(ABRS)^3 \equiv \mathcal{B}_1. \tag{72}$$

Thus, for a fixed error ϵ and AB smaller than RS/ϵ, the bound on the computational cost scales as the third power of the product AB; that is, the sixth power of the number of measurments, provided that $A = B$. This scaling is in good agreement with the numerical tests, as discussed later. However the tests indicate that the scaling $1/\epsilon$ and, thus, the sublinear convergence is too pessimistic. For example, for $\epsilon = 10^{-3}$, $A = B \le 40$, and $R = S = 2$, the bound gives a running time of the order of months, whereas the running time in the tests turns out to be less than one hour.

6.3. Simulation of the Oracle

We have shown that the cost of computing the distance from the local polytope grows polynomially, provided that we have access to the oracle. But what is the computational complexity of the oracle? In the case of measurements with two outcomes, we have seen that the solution of the oracle is equivalent to finding the minimal energy of a particular class of Ising spin glasses. These problems are known to be NP-hard. However, the oracle has a particular structure that can make many physically relevant instances numerically tractable. For example, the couplings of the Ising spin model are constrained by the nonsignaling conditions on $P(r,s|a,b)$ and the optimality conditions (22)–(24). Furthermore, the Hamiltonian (14) is characterized by two classes of spins, described by the variables r_k and s_k, respectively, and each element in one class is coupled only to elements in the other class. This particular structure suggests the following block-maximization algorithm for solving the oracle.

Algorithm 2. *Input:* $g(r,s;a,b)$

```
1. Generate a random sequence $\mathbf{r}$.
2. Maximize $\sum_{a,b} g(r_a,s_b;a,b)W(a,b)$ with respect to the sequence $\mathbf{s}$ (see later discussion).
3. Maximize $\sum_{a,b} g(r_a,s_b;a,b)W(a,b)$ with respect to the sequence $\mathbf{r}$.
4. Repeat from Step 2 until the block-maximizations stop making progress.
```

Numerical tests show that this algorithm, when used for computing the distance from the local polytope, stops after a few iterations. Furthermore, only a few trials of the initial random sequence **r** are required for convergence of Algorithm 1. We also note that the probability of a successful simulation of the oracle increases when χ is close to the optimal solution χ^{min}, suggesting that the optimality conditions (22)–(24) play some role in the computational complexity of the oracle. Pragmatically, we have chosen the number of trials equal to d_{NS}, such that the computational cost of simulating the oracle contributes to the overall running time with a constant multiplicative factor and, thus, the sixth-power law of the oracle-assisted algorithm is not affected.

Before discussing the numerical results, let us explain how the maximization on blocks is performed. Let us consider the maximization with respect to **r**, as the optimization with respect to **s** has an identical procedure. We have

$$\begin{gathered}\max_{\mathbf{r}} \textstyle\sum_{a,b} W(a,b) g(r_a,s_b;a,b) = \\ \textstyle\sum_a \max_r \sum_b g(r,s_b;a,b) W(a,b) \equiv \\ \textstyle\sum_a \max_r \tilde{g}(r,\mathbf{s};a).\end{gathered} \tag{73}$$

Thus, the maximum is found by maximizing the function $\tilde{g}(r,\mathbf{s};a)$, with respect to the discrete variable r for every a. Taking into account the sum over b required for generating $\tilde{g}$, the computational cost of the block-maximization is proportional to RAB. Thus, it does not grow more than linearly with respect to the size of the problem input; that is, $RSAB$.

7. Numerical Tests

In the previous sections, we introduced an algorithm that computes the distance from the local polytope in polynomial time, provided that we have access to oracle Max. Surprisingly, in every simulation performed on entangled qubits, the algorithm implementing the oracle successfully finds the solution in polynomial time. More precisely, the algorithm finds a sequence $(\mathbf{r},\mathbf{s})$ sufficiently close to the maximum to guarantee convergence of Algorithm 1 to the solution of Problem 1. Interestingly, the probability of a successful simulation of the oracle increases as χ approaches the solution. This suggests that the optimality conditions (22)–(24) play a fundamental role in the computational complexity of the oracle. To check that the algorithm successfully finds the optimizer χ^{min} up to the desired accuracy, we have solved the oracle with a brute-force search at the end of the computation, whenever this was possible in a reasonable time. All of the checks show that the solution is found within the desired accuracy.

In the tests, we considered the case of maximally entangled states, Werner states, and pure non-maximally entangled states. The numerical data are compatible with a running time scaling as the sixth power of the number of measurements. This is in accordance with the theoretical analysis, given in Section 6.2. Furthermore, the simulations show that the sublinear convergence of the upper bound $\mathcal{B}_1$ on the error is very loose, and the convergence turns out to be much faster. Let us discuss the case of entangled qubits in a pure quantum state.

7.1. Maximally Entangled State

In Figure 1, we report the time required for computing the distance from the local polytope as a function of the number of measurements, M, in log-log scale. The distance has been evaluated with accuracy equal to 10^{-3}, 10^{-4}, and 10^{-5} (red, blue, and green points, respectively). We have considered the case of planar measurements on the Bloch sphere. For the sake of comparison, we have also plotted the functions $10^{-6}M^6$ and $10^{-9}M^6$ (dashed lines). The data are compatible with the theoretical power law derived previously. They also show that the sublinear convergence of $\mathcal{B}_1$, derived in Section 6.2, is too pessimistic and the algorithm actually shows better performances. In particular, the bound $\mathcal{B}_1$ says that the running time is not greater than years for $A = B = 40$ and $\epsilon = 10^{-5}$, whereas the observed running time is actually less than one hour. Other simulations have been performed with random measurements. We generated a set of measurements corresponding to random vectors on the Bloch sphere, by considering both the planar and non-planar case. Then, we computed the distance from the local polytope for a different number of measurements. We always observed that the running time scales with the same sixth power law. For a number of measurements below 28, we have solved the oracle with a brute-force search at the end of the computation, and we have always found that Algorithm 1 successfully converged to the solution within the desired accuracy.

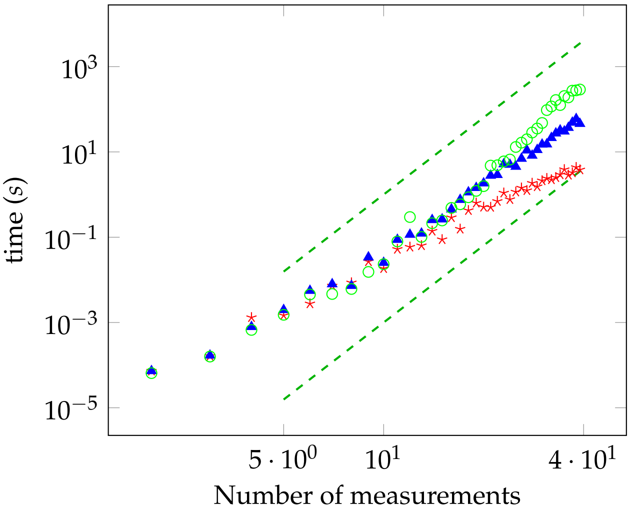

Figure 1. Time required for computing the distance from the local polytope for a maximally entangled state as a function of the number of measurements (log-log scale) with accuracy equal to 10^{-3}, 10^{-4}, and 10^{-5} (red, blue, and green points, respectively).

7.2. Non-Maximally Entangled State

In the case of the non-maximally entangled state

$$|\psi\rangle = \frac{|00\rangle + \gamma|11\rangle}{\sqrt{1+\gamma^2}}, \tag{74}$$

with $\gamma \in [0,1]$, we have considered planar measurements orthogonal to the Bloch vector $\vec{v}_z \equiv (0,0,1)$ (such that the marginal distributions are unbiased), as well as planar measurements lying in the plane containing $\vec{v}_z$ (biased marginal distributions).

In Figure 2, we report the distance from the local polytope as a function of γ with 10 measurements. The distance changes slightly for higher numbers of measurements. In the unbiased case, the distance goes to zero for γ equal to about 0.4, whereas the correlations become local for $\gamma = 0$ in the biased case.

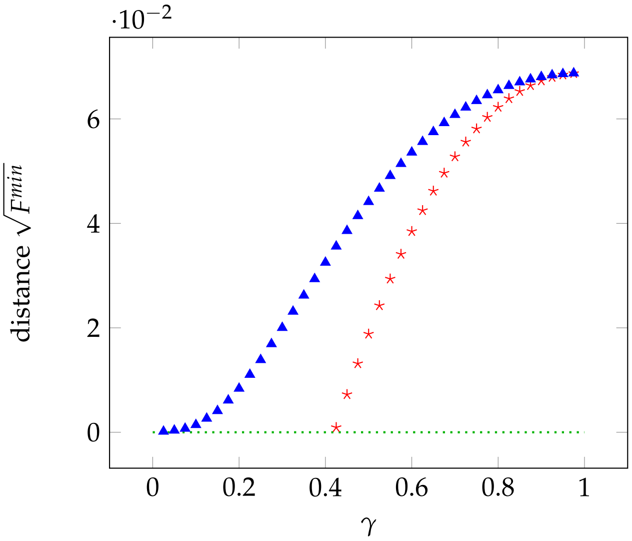

Figure 2. Distance from the local polytope as a function of γ in the unbiased case (red stars) and biased case (blue triangles).

In Figures 3 and 4, the running time as a function of the number of measurements is reported for the unbiased and biased cases, respectively. The power law is, again, in accordance with the theoretical analysis. As done for the maximally entangled case, we have checked the convergence to the solution by solving the oracle with a brute force search for a number of measurements up to 28.

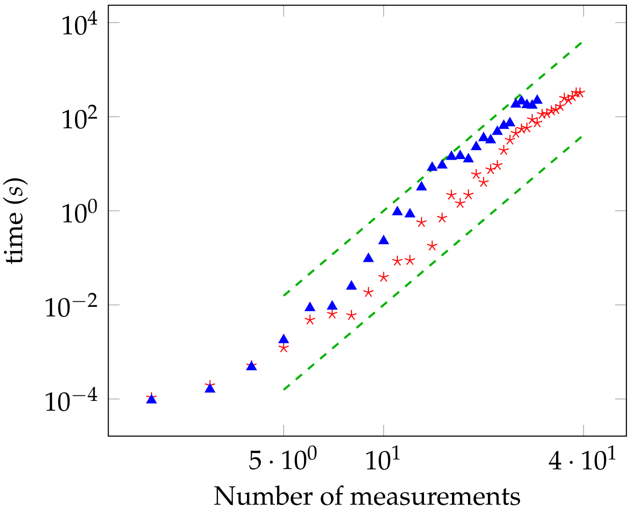

Figure 3. Time required for computing the distance from the local polytope as a function of the number of measurements (log-log scale) in the unbiased case, for $\gamma = 0.8$ (red stars) and $\gamma = 0.6$ (blue triangles). The green lines are the functions $10^{-6}M^6$ and $10^{-8}M^6$. The accuracy is 10^{-5}.

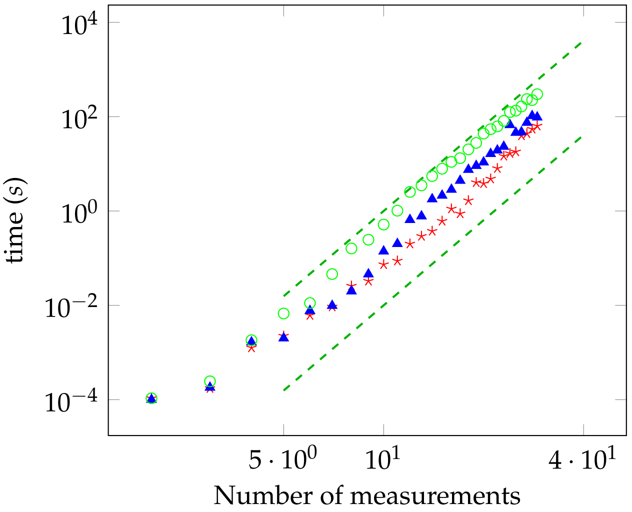

Figure 4. The same as Figure 3 in the biased case, for $\gamma = 0.8$ (red stars), $\gamma = 0.6$ (blue triangles), and $\gamma = 0.4$ (green circles).

8. Conclusions

In conclusion, we have presented an algorithm that computes the distance of a given non-signaling box to the local polytope. The running time, with given arbitrary accuracy, scaled polynomially, granted the access to an oracle determining the optimal locality bound of a Bell inequality. We also proposed an algorithm for simulating the oracle. In all of the numerical tests, the overall algorithm successfully computed the distance with the desired accuracy and a scaling of the running time, in agreement with the bound theoretically derived for the oracle-assisted algorithm. Our algorithm opens the way to tackle many unsolved problems in quantum theory, such as the non-locality of Werner states. Since the non-locality problem is NP-hard, our work and its further refinements could provide alternative algorithms to solve some instances of computationally hard problems.

Author Contributions: Conceptualization, formal analysis, investigation and software, A.M.; draft preparation, supervision, project administration, funding acquisition, A.M. and S.W.

Funding: This research was funded by Hasler Stiftung, grant number 16057, and Swiss National Science Foundation.

Acknowledgments: We wish to thank Arne Hansen for valuable comments and suggestions. This work is supported by the Swiss National Science Foundation, the NCCR QSIT, and the Hasler foundation through the project "Information-Theoretic Analysis of Experimental Qudit Correlations".

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bell, J. On the einstein podolsky rosen paradox. *Physics* **1964**, *1*, 195. [CrossRef]
2. Barrett, J.; Hardy, L.; Kent, A. No signaling and quantum key distribution. *Phys. Rev. Lett.* **2005**, *95*, 010503. [CrossRef] [PubMed]
3. Acín, A.; Gisin, N.; Masanes, L. From Bell's theorem to secure quantum key distribution. *Phys. Rev. Lett.* **2006**, *97*, 120405. [CrossRef] [PubMed]
4. Scarani, V.; Gisin, N.; Brunner, N.; Masanes, L.; Pino, S.; Acín, A. Secrecy extraction from no-signaling correlations. *Phys. Rev. A* **2006**, *74*, 042339. [CrossRef]
5. Acín, A.; Massar, S.; Pironio, S. Efficient quantum key distribution secure against no-signalling eavesdroppers. *New J. Phys.* **2006**, *8*, 126. [CrossRef]
6. Acín, A.; Brunner, N.; Gisin, N.; Massar, S.; Pironio, S.; Scarani, V. Device-independent security of quantum cryptography against collective attacks. *Phys. Rev. Lett.* **2007**, *98*, 230501. [CrossRef] [PubMed]
7. Masanes, L.; Renner, R.; Christandl, M.; Winter, A.; Barrett, J. Full security of quantum key distribution from no-signaling constraints. *IEEE Trans. Inf. Theory* **2014**, *60*, 4973–4986. [CrossRef]

8. Hänggi, E.; Renner, R.; Wolf, S. The impossibility of non-signaling privacy amplification. *Theor. Comput. Sci.* **2013**, *486*, 27–42. [CrossRef]
9. Colbeck, R.; Renner, R. Free randomness can be amplified. *Nat. Phys.* **2012**, *8*, 450. [CrossRef]
10. Gallego, R.; Masanes, L.; de la Torre, G.; Dhara, C.; Aolita, L.; Acín, A. Full randomness from arbitrarily deterministic events. *Nat. Commun.* **2013**, *4*, 2654. [CrossRef] [PubMed]
11. Buhrman, H.; Cleve, R.; Massar, S.; de Wolf, R. Nonlocality and communication complexity. *Rev. Mod. Phys.* **2010**, *82*, 665. [CrossRef]
12. Pitowski, I. *Quantum Probability—Quantum Logic*; Academic Press: London, UK, 1989.
13. Brierley, S.; Navascues, M.; Vertesi, T. Convex separation from convex optimization for large-scale problems. *arXiv* **2016**, arXiv:1609.05011.
14. Montina, A.; Wolf, S. Can non-local correlations be discriminated in polynomial time? *arXiv* **2016**, arXiv:1609.06269v1.
15. Fine, A. Hidden variables, joint probability, and the Bell inequalities. *Phys. Rev. Lett.* **1982**, *48*, 291. [CrossRef]
16. Collins, D.; Gisin, N. A relevant two qubit Bell inequality inequivalent to the CHSH inequality. *J. Phys. A Math. Theor.* **2004**, *37*, 1775. [CrossRef]
17. Montina, A.; Wolf, S. Information-based measure of nonlocality. *New J. Phys.* **2016**, *18*, 013035. [CrossRef]
18. Bernhard, C.; Bessire, B.; Montina, A.; Pfaffhauser, M.; Stefanov, A.; Wolf, S. Non-locality of experimental qutrit pairs. *J. Phys. A: Math. Theor.* **2014**, *42*, 424013. [CrossRef]
19. Boyd, S.; Vandenberghe, L. *Convex Optimization*; Cambridge University Press: Cambridge, UK, 2004.