Improved local models and new Bell inequalities via Frank-Wolfe algorithms

Sébastien Designolle, Gabriele Iommazzo, Mathieu Besançon, Sebastian Knebel, Patrick Gelß, and Sebastian Pokutta

Zuse-Institut Berlin, Takustraße 7, 14195 Berlin, Germany

(Received 20 March 2023; accepted 13 September 2023; published 18 October 2023)

In Bell scenarios with two outcomes per party, we algorithmically consider the two sides of the membership problem for the local polytope: Constructing local models and deriving separating hyperplanes, that is, Bell inequalities. We take advantage of the recent developments in so-called Frank-Wolfe algorithms to significantly increase the convergence rate of existing methods. First, we study the threshold value for the nonlocality of two-qubit Werner states under projective measurements. Here, we improve on both the upper and lower bounds present in the literature. Importantly, our bounds are entirely analytical; moreover, they yield refined bounds on the value of the Grothendieck constant of order three: $1.4367 \leqslant K_G(3) \leqslant 1.4546$. Second, we demonstrate the efficiency of our approach in multipartite Bell scenarios, and present local models for all projective measurements with visibilities noticeably higher than the entanglement threshold. We make our entire code accessible as a JULIA library called BellPolytopes.jl.

DOI: 10.1103/PhysRevResearch.5.043059

I. INTRODUCTION

Long after the establishment of quantum mechanics, Bell uncovered in 1964 the concept of nonlocality [1]. Arguably one of the most striking features of the theory, this property makes it possible to distinguish correlations that can be obtained by classical or quantum means and since then has been extensively studied [2]. Of particular interest is the question of the relation between this notion and the one of entanglement: Although entanglement is clearly necessary to observe any bipartite nonlocality, asking whether it is sufficient is a delicate question. For pure states, this is indeed the case [3], but this is not true for general states: In 1989, Werner exhibited mixed states that are entangled but nonetheless local [4].

More precisely, for a specific one-parameter family of states, by constructing an explicit local model recovering the correlations observed, he showed that the nonlocality threshold is different from the entanglement threshold. However, although he computed the latter exactly, his proof only provided a bound on the former, sufficient to assess the above-mentioned phenomenon, but far from the actual value. In the two-qubit case, the nonlocality witnessed by the Clauser-Horne-Shimony-Holt (CHSH) inequality for this family of states gave a bound in the opposite direction [5]. The large interval between these two bounds remained untouched for almost two decades, until Acín, Gisin, and Toner [6] realized that, owing to a connection already seen by Tsirelson [7] to an equivalent mathematical problem, an improved bound already existed [8], substantially reducing the gap. Soon after, Ref. [9] improved on the CHSH bound.

More recent works [10,11] have taken numerical approaches, by relying on an optimization algorithm by Gilbert [12], and on the known fact that the set of classical correlations is a polytope whose vertices are deterministic strategies (see, e.g., Ref. [2]). These works employ Gilbert's algorithm to approximate a quantum point, by minimizing the distance to that point from this so-called local polytope. This amounts to optimizing linear approximations of a quadratic distance function, given by the local gradient, to iteratively move towards one of the polytope vertices. The algorithm can converge to a facet without the need to compute the corresponding hyperplane. New bounds have then been attained in Refs. [10,13,14] by combining this algorithm with other techniques.

In this paper, we tackle the general membership problem for the local polytope via methods from the field of

TABLE I. Successive refinements of the bounds on v_c^{Wer}, the nonlocality threshold of the two-qubit Werner states under projective measurements. Using m measurements to simulate all projective ones is denoted by $m \sim \infty$.

	v_c^{Wer}	Reference	No. of inputs	Year
Upper bounds	0.7071	CHSH [5]	2	1969
	0.7056	Vértesi [9]	465	2008
	0.7054	Hua *et al.* [21]	∞	2015
	0.7012	Brierley *et al.* [10]	42	2016
	0.6964	Diviánszky *et al.* [13]	90	2017
	0.6961	This work { Eq. (4)	97	2023
Lower bounds	0.6875	This work { Eq. (3)	$406 \sim \infty$	2023
	0.6829	Hirsch *et al.* [14]	$625 \sim \infty$	2017
	0.6595	Acín *et al.* [6] using Krivine [8]	∞	2006 1979
	0.5	Werner [4]	∞	1989

constrained convex optimization, where it is known as the approximate Carathéodory problem [15,16]. Specifically, we rephrase the distance algorithm previously credited to Gilbert as the original Frank-Wolfe (FW) algorithm [17,18] (see Refs. [19,20] for recent reviews) to leverage the improvements brought to this algorithm over the last decade. We improve on the bounds for the nonlocality threshold of the two-qubit Werner states under projective measurements (see Table I), reducing the difference between the bounds by about 40%. Moreover, we investigate multipartite scenarios and establish new bounds for the nonlocality thresholds of the tripartite Greenberger-Horne-Zeilinger (GHZ) and W states [22]. These bounds considerably reduce the corresponding intervals and show that the GHZ state is strictly more robust than the W state when considering nonlocality under projective measurements.

II. PRELIMINARIES

Consider a bipartite scenario in which two parties, Alice and Bob, upon receiving inputs x and y chosen in $\{1\ldots m\}$, provide outputs a and b being ± 1, respectively. Here, we are only interested in the correlation matrix arising from this process, namely, the $m\times m$ real matrix whose (x,y) entries are the expectation values of $\langle ab\rangle$ with inputs x and y. Our setup indeed makes marginals, i.e., expectation values $\langle a\rangle$ and $\langle b\rangle$, irrelevant as they always vanish. This will no longer be the case in multipartite scenarios, as discussed below.

Classical correlation matrices lie in the convex hull of the deterministic strategies, which are rank-one matrices $\mathbf{d}_{\vec{a},\vec{b}}$ with entries $a_x b_y$, where $\vec{a}=(a_1,\ldots,a_m)$ and $\vec{b}=(b_1,\ldots,b_m)$ have ± 1 components. Since $\mathbf{d}_{-\vec{a},-\vec{b}}=\mathbf{d}_{\vec{a},\vec{b}}$, there are 2^{2m-1} distinct deterministic strategies, which define the local correlation polytope $\mathcal{L}_m$ [2].

Given a shared quantum state ρ, i.e., a positive semidefinite Hermitian matrix with trace one, and traceless dichotomic observables A_x and B_y, i.e., Hermitian matrices of trace zero and squaring to the identity, one can construct a correlation matrix by letting Alice and Bob measure their half of the shared state with their observables. By the Born rule, the resulting matrix then has (x,y) entries of the form $\mathrm{Tr}[(A_x\otimes B_y)\rho]$ [23].

The central problem we consider in this paper is the membership problem for the local polytope $\mathcal{L}_m$, which is twofold. On the one hand, given a correlation matrix inside $\mathcal{L}_m$, we seek to decompose it in terms of deterministic strategies, that is, to find a local model. On the other hand, given a (quantum) correlation matrix outside $\mathcal{L}_m$, we want to produce an explicit separating hyperplane to witness its nonlocality, that is, a Bell inequality.

We focus on a family of two-qubit Werner states [4],

$$\rho_v^{\mathrm{Wer}}=v\,|\psi_-\rangle\langle\psi_-|+(1-v)\frac{\mathbb{1}}{4},\qquad(1)$$

where $|\psi_-\rangle=(|01\rangle-|10\rangle)/\sqrt{2}$ is the two-qubit antisymmetric (or singlet) state. Fixing the so-called visibility v in Eq. (1) and applying qubit observables of the form $A_x=\vec{a}_x\cdot\vec{\sigma}$ and $B_y=\vec{b}_y\cdot\vec{\sigma}$, where $\vec{a}_x$ and $\vec{b}_y$ are real vectors on the unit 2-sphere (Bloch vectors) and $\vec{\sigma}=(\sigma_X,\sigma_Y,\sigma_Z)$ contains Pauli matrices, yields a correlation matrix whose (x,y) entries are $-v\,\vec{a}_x\cdot\vec{b}_y$. When the number of inputs m goes to infinity, we denote the different outputs directly by this Bloch vector: The observables are then $A_{\hat{x}}=\hat{x}\cdot\vec{\sigma}$ and $B_{\hat{y}}=\hat{y}\cdot\vec{\sigma}$, where the hat emphasizes the infinite scenario.

Question 1. With Werner states in Eq. (1), which critical visibility v_c^{Wer} is the threshold between a local behavior and a nonlocal one under projective measurements?

A. Previous works

Question 1 has gained attention after the publication of Ref. [6], where it is linked with the computation of a mathematical constant. Increasingly more accurate bounds have been obtained since then, as outlined in Table I.

On the one hand, to obtain an upper bound it is sufficient to consider a scenario with a finite number of measurements and to exhibit a Bell inequality and a quantum strategy with a good robustness to noise with respect to this inequality.

On the other hand, methods to provide a lower bound cannot be as direct, since a membership proof is required for the *infinite* scenario with all projective measurements. To go from a finite number of measurements to an infinite one, Refs. [24,25] suggest to simulate, up to an approximation factor, the infinite set of all measurements by means of a finite number of them, then to algorithmically construct a local model in this finite case, and eventually to convert the membership proof obtained there to a certificate valid for all projective measurements.

The approximation amounts to choosing m measurements used both by Alice and Bob and to computing the radius η of the largest sphere that fits in the polyhedron defined by the vertices $\vec{a}_x$ and $-\vec{a}_x$. Then any shrunk direction $\eta\hat{x}$ can be written as a convex mixture of the vectors $\vec{a}_x$, i.e., $\eta\hat{x}=\sum_x p_x^{\hat{x}}\vec{a}_x$; similarly, $\eta\hat{y}=\sum_y q_y^{\hat{y}}\vec{b}_y$. Now if we can decompose the correlation matrix with entries $-v_0\,\vec{a}_x\cdot\vec{b}_y$ in terms of deterministic strategies, then a decomposition for the infinite scenario with visibility $\eta^2 v_0$ arises from $-\eta^2 v_0\,\hat{x}\cdot\hat{y}=\sum_{x,y}p_x^{\hat{x}}q_y^{\hat{y}}(-v_0\,\vec{a}_x\cdot\vec{b}_y)$.

Reference [14] uses a polyhedron with $m=625$ measurements and finds a way to make the numerical decomposition analytical at the expense of a factor ν_1 discussed below. They eventually obtain $v_c^{\mathrm{Wer}}\geqslant\eta^2\nu_1 v_0\approx 0.6829$ where

$$\eta\geqslant\cos\left(\frac{\pi}{50}\right)^2\approx 0.9961,\qquad \nu_1=0.999,\qquad\text{and } v_0=0.689.$$

In this paper, we improve on all three factors, for each of them with a different theoretical reason: We first explain how to obtain polyhedra with a better shrinking factor η, then argue that our algorithm makes it possible to choose an initial visibility v_0 closer to the critical one, and refine the last step to have $\nu_2>\nu_1$ closer to 1.

B. Choosing the measurements

The first step is to select the m measurements that Alice and Bob perform. The more measurements we consider, the better the approximation of the set of all projective measurements is. However, the optimization problem on the resulting correlation polytope is also more difficult to solve, since the dimension of the corresponding space grows quadratically with m. Moreover, since we want the final result to be analytical, this η should have a closed form.

Algorithm 1. Gilbert's algorithm [12]

1: **for** $t = 0 \dots T-1$ **do**
2: $\omega_t = \arg\min_{\vec{a},\vec{b}} \langle \mathbf{x}_t - v_0\mathbf{p}, \mathbf{d}_{\vec{a},\vec{b}} \rangle$
3: $\gamma_t = \arg\min_{\gamma\in[0,1]} \|\gamma\mathbf{x}_t + (1-\gamma)\mathbf{d}_{\omega_t} - v_0\mathbf{p}\|_2^2$
4: $\mathbf{x}_{t+1} = \gamma_t\mathbf{x}_t + (1-\gamma_t)\mathbf{d}_{\omega_t}$
5: **end for**

In Ref. [14], this last necessity leads the authors to introduce a family of measurements corresponding to quite regular polyhedra and whose shrinking factors η enjoy a relatively tight analytical lower bound. These shrinking factors, however, are not competitive compared to polyhedra with a similar number of measurements.

Here, we take a different approach to improve the quality of the shrinking factor while not losing the analyticity. For this, we start by getting symmetric polyhedra with very good shrinking factors [26]. Then we take rational approximations of these polyhedra; importantly, we ensure that the rational points are also on the unit sphere. Eventually we can compute all faces analytically and hence obtain η^2 as a rational. The construction of this rational approximation is in Supplemental Material Sec. I [27].

C. Frank-Wolfe algorithms

After selecting a polyhedron as outlined above, we can construct the correlation matrix $\mathbf{p}$ with entries $\mathbf{p}_{x,y} = -\vec{a}_x \cdot \vec{b}_y$, where $\vec{a}_x$ and $\vec{b}_y$ are pairs of antipodal points in the chosen polyhedron; this corresponds to setting $v = 1$ in Eq. (1). To obtain the distance between the local polytope $\mathcal{L}_m$ and a point $v_0\mathbf{p}$ on the line between $\mathbf{0}$ and $\mathbf{p}$, we can choose a local point $\mathbf{x}_0$ and run Algorithm 1 [10,11]. There, $\|\mathbf{y}\|_2$ denotes the 2-norm of the vectorized matrix $\mathbf{y}$.

As the number of deterministic strategies $\mathbf{d}_{\vec{a},\vec{b}}$ to explore in line 2 is exponential in m (here, 2^{2m-1}), a heuristic approach is performed, similarly to Refs. [10,11,14]; we refer to Supplemental Material Sec. II [27]. Algorithm 1 can only supply a reliable membership proof when $v_0\mathbf{p}$ belongs to the local polytope. In this case, the decomposition that the algorithm produces is valid regardless of the potential suboptimality due to the heuristic.

Although Ref. [10] credits Gilbert [12] for this algorithm, this instance coincides with the original FW algorithm [17,18] where the function to minimize is $f(\mathbf{x}) = \frac{1}{2}\|\mathbf{x} - v_0\mathbf{p}\|_2^2$ so that $\nabla f(\mathbf{x}) = \mathbf{x} - v_0\mathbf{p}$. This projection-free first-order algorithm has seen a regained interest in the last decade [28], and several lines of improvement have been proposed to upgrade its convergence in various settings. Of particular importance is the mitigation of the so-called zigzagging behavior [29]: When optimizing over polytopes, if the optimum is located near or on a facet, the original algorithm alternately selects the vertices defining this facet and moves towards them. This leads to ever-smaller improvements in the objective function value, as the gradient becomes more orthogonal to the steps performed. This forces [14] picking v_0 such that the starting point $v_0\mathbf{p}$ lies sufficiently deep inside the local polytope to ensure convergence in a reasonable time. Here, we employ a refined version of this algorithm which stores a subset of the vertices of $\mathcal{L}_m$ to speed up the computations [30]. This variant has two major benefits: The zigzagging behavior is reduced and the final decomposition is sparser. We illustrate the zigzagging and describe this improved algorithm in Supplemental Material Sec. III [27].

We implemented our approach in JULIA [31], based on the library FrankWolfe.jl [32]. Our code is freely accessible as a JULIA library entitled BellPolytopes.jl [33]. Importantly, it is not restricted to the case considered above but can tackle scenarios with any numbers of parties and inputs (identical for all parties), but only two outputs. We give a few results on the multipartite case below.

D. Analytical decomposition

After choosing an initial visibility v_0, we run our algorithm until the last iterate $\mathbf{x}_T$ satisfies $\|\mathbf{x}_T - v_0\mathbf{p}\|_2 \leqslant \epsilon$ for a chosen precision ϵ. Since we need an *exact* decomposition of a point $v\mathbf{p}$ to be able to certify that $v \leqslant v_c^{\mathrm{Wer}}$, such a numerical proximity is, however, insufficient. Reference [14] gives the following solution to overcome this difficulty: Fix a factor ν_1 close to 1, write $\nu_1 v_0\mathbf{p} = \nu_1\mathbf{x}_T + (1-\nu_1)\mathbf{y}$ by suitably defining $\mathbf{y}$, and exhibit a local decomposition for $\mathbf{y}$ by using the fact that its entries are, by construction, small.

More precisely, Ref. [14] shows that any point $\mathbf{y}$ such that $\|\mathbf{y}\|_1 \leqslant 1$ has a local model. In Supplemental Material Sec. IV [27] we tighten this result and demonstrate that this even holds for $\|\mathbf{y}\|_2 \leqslant 1$. This seemingly small change has in fact practical consequences as it is less restrictive on the quality of the algorithm output: For instance, the result from Ref. [14] jumps from 0.6829 to 0.6836. Our factor ν_2 reads

$$\nu_2 = \frac{1}{1 + \|\mathbf{x}_T - v_0\mathbf{p}\|_2}. \tag{2}$$

E. Lower bound

Having outlined all steps of the proof, we can now give the computational settings used to attain our lower bound on v_c^{Wer}. We choose $m = 406$ measurements yielding a shrinking factor of $\eta \approx 0.9968$. We tested several initial visibilities v_0 and selected $v_0 = 0.692$; for this value, the objective function steadily decreases. Using our algorithm, we end up with 78 747 deterministic strategies (out of all 2^{811} of them) reproducing the correlation matrix $v_0\mathbf{p}$ up to $\epsilon = 2\times10^{-4}$, corresponding to a factor $\nu_2 \approx 0.9998$ in Eq. (2). Finally, we recompute the decomposition with rational weights (as in Ref. [14]) to get an analytical expression of ν_2. Combining all the steps, we obtain the final analytical lower bound

$$v_c^{\mathrm{Wer}} \geqslant v_{\mathrm{low}} = \eta^2\nu_2 v_0 \approx 0.6875, \tag{3}$$

whose analytical value is in the supplemental JULIA file.

The computation ran on a 64-core Intel® Xeon® Gold 6338 machine with 512 GB of RAM and took about 1 month. This runtime is long because we want to test our methodology extensively; this alone is not responsible for the improvement over Ref. [14]. Owing to our theoretical improvements, we can indeed reproduce the bound therein in about 20 h and with only 181 measurements.

F. Upper bound

One can extract a separating hyperplane from the result of FW algorithms, specifically, by taking the gradient at an

TABLE II. Summarized history of the successive refinements of the bounds on the nonlocality threshold for three-qubit GHZ (left) and W (right) states under projective measurements. Importantly, although the 16 measurements used for the GHZ state are exactly the same as in Ref. [10] (a regular polygon on the XY plane), we can reach a more robust Bell inequality owing to the improved algorithm we are using. The 16 measurements used for the W state correspond to a pentakis dodecahedron. Remarkably, $v_c^{\mathrm{GHZ}_3} < v_c^{W_3}$ arises as a consequence of our bounds.

	$v_c^{\mathrm{GHZ}_3}$	Reference	No. of inputs	Year
Upper	0.5	GHZ [22]	2	1989
Upper	0.4961	Vértesi and Pál [38]	5	2011
Upper	0.4932	Brierley *et al.* [10]	16	2016
Upper	0.4916	This work	16	2023
Lower	0.4688	This work	$61\sim\infty$	2023
Lower	0.232	Cavalcanti *et al.* [24]	$12\sim\infty$	2016
Lower	0.2	Dür and Cirac [39]	Entanglement threshold	2000

	$v_c^{W_3}$	Reference	No. of inputs	Year
Upper	0.6442	Sen(De) *et al.* [40]	2	2003
Upper	0.6007	Gruca *et al.* [41]	5	2010
Upper	0.5956	Pandit *et al.* [42]	6	2022
Upper	0.5482	This work	16	2023
Lower	0.4917	This work	$61\sim\infty$	2023
Lower	0.228	Cavalcanti *et al.* [24]	$12\sim\infty$	2016
Lower	0.2096	Szalay [43]	Entanglement threshold	2011

approximately optimal solution. This property is already used in Refs. [10,13] to construct Bell inequalities with a high resistance to noise. The difficulty to improve on these works lies in the computation of the local value of the Bell inequality provided by the algorithm [34–36]. Interestingly, however, this problem can be converted into a quadratic unconstrained binary optimization (QUBO) instance, a class of problems which has seen some recent improvements (see Ref. [37] and references therein).

We had access to a version of the solver from Ref. [37]. With $m = 97$ measurements, we ran our algorithm starting from $v_0 = 0.6964$ and fed the QUBO solver with the resulting hyperplane to obtain, in about 30 min, the bound

$$v_c^{\mathrm{Wer}} \leqslant v_{\mathrm{up}} \approx 0.6961, \tag{4}$$

whose analytical value is in the supplemental JULIA file together with the corresponding Bell inequality. The formulation of the local bound computation as a QUBO is in Supplemental Material Sec. II [27]. Importantly, this bound is also analytical as the Bell inequality used has integer entries, so that the decisions made in the QUBO solver used are exact.

G. Multipartite case

The entire procedure naturally generalizes to multipartite scenarios. One important difference, however, is that marginals no longer vanish; hence, we must take them into account and reproduce them in the local model. Computationally, it is also harder to compute a good direction in the larger correlation space, hence we are restricted to a smaller number of measurements. We summarize our results in the tripartite case in Table II. Notably, we show that the three-qubit GHZ state is more robust to noise than the three-qubit W state for nonlocality under projective measurements. We refer to Supplemental Material Sec. V [27] for details.

III. OBSERVATIONS

Our bounds in Eqs. (3) and (4) have two immediate consequences, already described in Ref. [14] and which we only summarize here.

First, our local model for projective qubit measurements can be extended to qubit positive operator-valued measures (POVMs), as the latter can be simulated by the former up to a factor of $\sqrt{2/3}$ (see Lemma 2 in Ref. [14] or Ref. [44]). Therefore, the nonlocality threshold for POVMs admits the lower bound $2/3 \cdot v_{\mathrm{low}} \approx 0.4583$.

Second, there is a formal correspondence between the construction of local hidden variable models for two-qubit Werner states and the Grothendieck constant of order three $K_G(3)$. Our results in Eqs. (3) and (4) directly translate into the following analytical bounds:

$$1.4367 \approx \frac{1}{v_{\mathrm{up}}} \leqslant K_G(3) \leqslant \frac{1}{v_{\mathrm{low}}} \approx 1.4546, \tag{5}$$

whose exact values are in the supplemental JULIA file.

Other applications directly benefit from the improvement of the bounds on v_c^{Wer}, such as quantum key distributions [45] or prepare-and-measure scenarios [46].

IV. CONCLUSION

In this paper, we construct local models and Bell inequalities by using FW algorithms in local polytopes with binary outcomes and arbitrarily many inputs and parties. Our main application is to improve the bounds on the nonlocality threshold of the two-qubit Werner states, hence on the Grothendieck constant of order three. We also investigate multipartite states and find new bounds for GHZ and W states, far above their entanglement thresholds. This opens a practical way to a better understanding of the nonlocality properties of these states. To facilitate the reuse of our method, we provide a JULIA library with our implementation [33].

A natural extension would be to increase the number of outcomes of the scenario, the algorithm working exactly the same way in the probability space. In the qubit case, the range of the nonlocality threshold of two-qubit Werner states under POVMs indeed remains wide open and a good approximation of the set of general measurements may help reduce this gap. In higher dimensions, this extension would also require a suitable approximation of the set of projective measurements, a difficulty that was already mentioned in Ref. [25]. Following our approach here to construct good polyhedra in the Bloch sphere, we expect symmetric measurements such as those in Ref. [47] to provide good seeds for the exploration of this direction.

More generally, the progress made in the constrained convex optimization community and leveraged in this paper could benefit all existing applications of FW algorithms, e.g., for entanglement detection [48,49], and could also help finding different utilizations, for instance, for large-scale semidefinite programming problems.

ACKNOWLEDGMENTS

The authors are grateful to Flavien Hirsch for his assistance in drawing the connection between FW algorithms and Gilbert's algorithm, to Daniel Rehfeldt for allowing us to utilize his QUBO solver [37], to Tamás Vértesi for providing us with useful references and with data we used to verify our algorithm, and to Mateus Araújo, Nicolas Brunner, Máté Farkas, Antonio Frangioni, and Leo Liberti for their contribution in the discussions. Research reported in this paper was supported through the Research Campus Modal funded by the German Federal Ministry of Education and Research (fund numbers 05M14ZAM, 05M20ZBM), the Einstein Research Unit "Near-term quantum computational devices and quantum processors" of the Berlin University Alliance, and the QuantERA II Programme that has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 101017733.

[1] J. S. Bell, On the Einstein Podolsky Rosen paradox, Phys. Phys. Fiz. **1**, 195 (1964).
[2] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, Bell nonlocality, Rev. Mod. Phys. **86**, 419 (2014).
[3] N. Gisin, Bell's inequality holds for all non-product states, Phys. Lett. A **154**, 201 (1991).
[4] R. F. Werner, Quantum states with Einstein-Podolsky-Rosen correlations admitting a hidden-variable model, Phys. Rev. A **40**, 4277 (1989).
[5] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, Proposed experiment to test local hidden-variable theories, Phys. Rev. Lett. **23**, 880 (1969).
[6] A. Acín, N. Gisin, and B. Toner, Grothendieck's constant and local models for noisy entangled quantum states, Phys. Rev. A **73**, 062105 (2006).
[7] B. S. Tsirelson, Some results and problems on quantum Bell-type inequalities, Hadronic J. Suppl. **8**, 329 (1993).
[8] J.-L. Krivine, Constantes de Grothendieck et fonctions de type positif sur les sphères, Adv. Math. **31**, 16 (1979).
[9] T. Vértesi, More efficient Bell inequalities for Werner states, Phys. Rev. A **78**, 032112 (2008).
[10] S. Brierley, M. Navascués, and T. Vértesi, Convex separation from convex optimization for large-scale problems, arXiv:1609.05011.
[11] A. Montina and S. Wolf, Discrimination of non-local correlations, Entropy **21**, 104 (2019).
[12] E. G. Gilbert, An iterative procedure for computing the minimum of a quadratic form on a convex set, J. SIAM Control **4**, 61 (1966).
[13] P. Diviánszky, E. Bene, and T. Vértesi, Qutrit witness from the Grothendieck constant of order four, Phys. Rev. A **96**, 012113 (2017).
[14] F. Hirsch, M. T. Quintino, T. Vértesi, M. Navascués, and N. Brunner, Better local hidden variable models for two-qubit Werner states and an upper bound on the Grothendieck constant $K_G(3)$, Quantum **1**, 3 (2017).
[15] V. Mirrokni, R. P. Leme, A. Vladu, and S. C.-W. Wong, Tight bounds for approximate Carathéodory and beyond, in *Proceedings of the 34th International Conference on Machine Learning*, edited by D. Precup and Y. W. Teh (PMLR, 2017), Vol. 70, p. 2440.
[16] C. W. Combettes and S. Pokutta, Revisiting the approximate Carathéodory problem via the Frank–Wolfe algorithm, Math. Program. **197**, 191 (2023).
[17] M. Frank and P. Wolfe, An algorithm for quadratic programming, Nav. Res. Logist. Q. **3**, 95 (1956).
[18] E. S. Levitin and B. T. Polyak, Constrained minimization methods, USSR Comput. Math. Math. Phys. **6**, 1 (1966).
[19] G. Braun, A. Carderera, C. W. Combettes, H. Hassani, A. Karbasi, A. Mokhtari, and S. Pokutta, Conditional gradient methods, arXiv:2211.14103.
[20] I. M. Bomze, F. Rinaldi, and D. Zeffiro, Frank–Wolfe and friends: A journey into projection-free first-order optimization methods, 4OR-Q. J. Oper. Res. **19**, 313 (2021).
[21] B. Hua, M. Li, T. Zhang, C. Zhou, X. Li-Jost, and S.-M. Fei, Towards Grothendieck constants and LHV models in quantum mechanics, J. Phys. A **48**, 065302 (2015).
[22] D. M. Greenberger, M. A. Horne, and A. Zeilinger, Going beyond Bell's theorem, in *Bell's Theorem, Quantum Theory, and Conceptions of the Universe*, edited by M. Kafatos (Springer, Dordrecht, 1989), pp. 69–72.
[23] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*, 10th ed. (Cambridge University Press, New York, 2011).
[24] D. Cavalcanti, L. Guerini, R. Rabelo, and P. Skrzypczyk, General method for constructing local hidden variable models for entangled quantum states, Phys. Rev. Lett. **117**, 190401 (2016).
[25] F. Hirsch, M. T. Quintino, T. Vértesi, M. F. Pusey, and N. Brunner, Algorithmic construction of local hidden variable models for entangled quantum states, Phys. Rev. Lett. **117**, 190402 (2016).
[26] https://levskaya.github.io/polyhedronisme; see the fork https://github.com/sebastiendesignolle/polyhedronisme which displays the shrinking factor and implements the operation S projecting vertices on the sphere.
[27] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevResearch.5.043059 for the detailed derivations and the multipartite extension, which includes Refs. [50–55].
[28] M. Jaggi, Revisiting Frank–Wolfe: Projection-free sparse convex optimization, in *Proceedings of the 30th International Conference on Machine Learning*, edited by S. Dasgupta and D. McAllester (PMLR, 2013), Vol. 28, p. 427.
[29] S. Lacoste-Julien and M. Jaggi, On the global linear convergence of Frank–Wolfe optimization variants, Adv. Neural Inf. Process. Syst. **28**, 496 (2015).
[30] K. Tsuji, K. Tanaka, and S. Pokutta, Sparser kernel herding with pairwise conditional gradients without swap steps, arXiv:2110.12650.

[31] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah, Julia: A fresh approach to numerical computing, SIAM Rev. **59**, 65 (2017).

[32] M. Besançon, A. Carderera, and S. Pokutta, FrankWolfe.jl: A high-performance and flexible toolbox for Frank–Wolfe algorithms and conditional gradients, arXiv:2104.06675.

[33] https://github.com/ZIB-IOL/BellPolytopes.jl.

[34] D. Avis, H. Imai, and T. Ito, On the relationship between convex bodies related to correlation experiments with dichotomic observables, J. Phys. A **39**, 11283 (2006).

[35] D. Avis and T. Ito, Comparison of two bounds of the quantum correlation set, in *Proceedings of the 2007 First International Conference on Quantum, Nano, and Micro Technologies (ICQNM'07)* (IEEE, New York, 2007).

[36] D. Avis, S. Moriyama, and M. Owari, From Bell inequalities to Tsirelson's theorem, IEICE Trans. Fundam. Electron. Commun. Comput. Sci. **E92-A**, 1254 (2009).

[37] D. Rehfeldt, T. Koch, and Y. Shinano, Faster exact solution of sparse MaxCut and QUBO problems, arXiv:2202.02305.

[38] T. Vértesi and K. F. Pál, Nonclassicality threshold for the three-qubit Greenberger–Horne–Zeilinger state, Phys. Rev. A **84**, 042122 (2011).

[39] W. Dür and J. I. Cirac, Classification of multiqubit mixed states: Separability and distillability properties, Phys. Rev. A **61**, 042314 (2000).

[40] A. Sen(De), U. Sen, M. Wieśniak, D. Kaszlikowski, and M. Żukowski, Multiqubit W states lead to stronger nonclassicality than Greenberger-Horne-Zeilinger states, Phys. Rev. A **68**, 062306 (2003).

[41] J. Gruca, W. Laskowski, M. Żukowski, N. Kiesel, W. Wieczorek, C. Schmid, and H. Weinfurter, Nonclassicality thresholds for multiqubit states: Numerical analysis, Phys. Rev. A **82**, 012118 (2010).

[42] M. Pandit, A. Barasiński, I. Márton, T. Vértesi, and W. Laskowski, Optimal tests of genuine multipartite nonlocality, New J. Phys. **24**, 123017 (2022).

[43] S. Szalay, Separability criteria for mixed three-qubit states, Phys. Rev. A **83**, 062337 (2011).

[44] M. Oszmaniec, L. Guerini, P. Wittek, and A. Acín, Simulating positive-operator-valued measures with projective measurements, Phys. Rev. Lett. **119**, 190501 (2017).

[45] M. Farkas, M. Balanzó-Juandó, K. Łukanowski, J. Kołodyński, and A. Acín, Bell nonlocality is not sufficient for the security of standard device-independent quantum key distribution protocols, Phys. Rev. Lett. **127**, 050503 (2021).

[46] P. Diviánszky, I. Márton, E. Bene, and T. Vértesi, Certification of qubits in the prepare-and-measure scenario with large input alphabet and connections with the Grothendieck constant, Sci. Rep. **13**, 13200 (2023).

[47] H. C. Nguyen, S. Designolle, M. Barakat, and O. Gühne, Symmetries between measurements in quantum mechanics, arXiv:2003.12553.

[48] J. Shang and O. Gühne, Convex optimization over classes of multiparticle entanglement, Phys. Rev. Lett. **120**, 050506 (2018).

[49] M. Wieśniak, Two-qutrit entanglement: 56-years old algorithm challenges machine learning, arXiv:2211.03213.

[50] J. J. Thomson, On the structure of the atom: An investigation of the stability and periods of oscillation of a number of corpuscles arranged at equal intervals around the circumference of a circle; with application of the results to the theory of atomic structure, London, Edinburgh Dublin Philos. Mag. J. Sci. **7**, 237 (1904).

[51] B. Legat, R. Deits, G. Goretkin, T. Koolen, J. Huchette, D. Oyama, and M. Forets, https://github.com/JuliaPolyhedra/Polyhedra.jl (2021).

[52] G. Braun, S. Pokutta, and D. Zink, Lazifying conditional gradient algorithms, J. Mach. Learn. Res. **20**, 2577 (2019).

[53] E. B. Bajalinov, *Linear-Fractional Programming Theory, Methods, Applications and Software* (Springer, New York, 2003).

[54] B. S. Tsirelson, Quantum analogues of the Bell inequalities. The case of two spatially separated domains, J. Sov. Math. **36**, 557 (1987).

[55] M. Żukowski and Č. Brukner, Bell's theorem for general N-qubit states, Phys. Rev. Lett. **88**, 210401 (2002).