Convex Optimization over Classes of Multiparticle Entanglement

Jiangwei Shang[1,2*] and Otfried Gühne[1†]
[1]Naturwissenschaftlich-Technische Fakultät, Universität Siegen, Walter-Flex-Straße 3, 57068 Siegen, Germany
[2]Beijing Key Laboratory of Nanophotonics and Ultrafine Optoelectronic Systems, School of Physics, Beijing Institute of Technology, Beijing 100081, China

(Received 18 July 2017; revised manuscript received 18 December 2017; published 1 February 2018)

A well-known strategy to characterize multiparticle entanglement utilizes the notion of stochastic local operations and classical communication (SLOCC), but characterizing the resulting entanglement classes is difficult. Given a multiparticle quantum state, we first show that Gilbert's algorithm can be adapted to prove separability or membership in a certain entanglement class. We then present two algorithms for convex optimization over SLOCC classes. The first algorithm uses a simple gradient approach, while the other one employs the accelerated projected-gradient method. For demonstration, the algorithms are applied to the likelihood-ratio test using experimental data on bound entanglement of a noisy four-photon Smolin state [Phys. Rev. Lett. **105**, 130501 (2010)].

DOI: 10.1103/PhysRevLett.120.050506

Introduction.—Entanglement is a fundamental phenomenon in quantum mechanics and is often considered to be a useful resource for tasks like quantum metrology or quantum cryptography. Consequently, the question of whether a given quantum state of two particles is entangled or separable is relevant for several fields in physics [1,2]. So far, much effort has been devoted to devise methods to certify that a given state is entangled; prominent examples are Bell inequalities and entanglement witnesses [2,3]. Contrary to that, methods to prove separability, or, equivalently, to disprove the presence of entanglement, are rare: For some instances, explicit solutions are known [4–7], but, in general, one has to rely on numerical procedures with a restricted applicability [8–11]. In order to analyze experiments, one often needs to quantify the extent to which the observed data can be explained by separable states, e.g., in a likelihood-ratio test [12]. In this case, one even needs to *optimize* over separable states, and this is nearly impossible with current analysis tools. The analysis of separability becomes even more complicated in the multiparticle case, because then different classes of entanglement exist [13,14].

In this Letter, we present methods to analyze separability for multiparticle quantum states. First, we show that an adaption of the so-called Gilbert's algorithm [15] can be used to prove separability or membership in a certain entanglement class, and the resulting algorithm outperforms known methods significantly. Second, we demonstrate that a combination of this method with gradient methods can be used to perform optimization over separability classes, allowing, for instance, the computation of likelihood ratios. We demonstrate the practical usefulness of our approach with many examples and data from recent experiments.

Notions of entanglement.—A pure state $|\psi\rangle$ of two particles is separable, if it is a product state $|\psi\rangle = |a\rangle \otimes |b\rangle$; otherwise, it is entangled. Concerning mixed states, a state is separable if it can be written as a convex combination of product states, that is,

$$\rho = \sum_i p_i |a_i\rangle\langle a_i| \otimes |b_i\rangle\langle b_i|. \tag{1}$$

Here, the p_i form a probability distribution, so they are positive and sum up to one. A state that is not separable is called entangled. While the entanglement of pure states is straightforward to characterize, the same question for mixed states is a hard problem [1,2].

For more than two particles, different classes of entanglement exist. To give a first example, let us consider $n > 2$ particles and fix a number $2 \leq k \leq n$. A pure state is then k-separable if it can be written as a tensor product of k local states:

$$|\Phi_{S_k}\rangle = \bigotimes_{i=1}^{k} |\phi_i\rangle, \tag{2}$$

where the $|\phi_i\rangle$ are k states on subsets of the n parties. More specifically, the state are called biseparable for $k = 2$, triseparable for $k = 3$, and up to fully separable for $k = n$. A state that is not k-separable contains entanglement; for instance, it is genuinely multiparticle entangled if it is not biseparable. For mixed states, one can extend this definition as before via convex combinations. In this case, one also considers mixtures of k-separable states that are separable with respect to different partitions. It is known that the characterization of mixed separable states is *NP* hard, if the number of particles increases [16,17].

To characterize multiparticle entanglement further, a popular strategy uses the notion of stochastic local operations and classical communication (SLOCC) [13,14]. In mathematical terms, a SLOCC operation can be represented

as $A_{\mathrm{SLOCC}} = \bigotimes_i A_i$, where A_i is a matrix describing the local operation acting on the ith party. Under SLOCC operations, a pure state $|\phi\rangle$ can be mapped to another state $|\phi'\rangle$ iff

$$|\phi'\rangle \propto A_{\mathrm{SLOCC}}|\phi\rangle, \tag{3}$$

and $|\phi\rangle$ and $|\phi'\rangle$ are called SLOCC equivalent if the local operations A_i are invertible [13]. Remarkably, for three qubits this classification gives two inequivalent families of genuine multiparticle states: the Greenberger-Horne-Zeilinger class and the W class [13]. Again, one can define the corresponding convex sets for mixed states as in Eq. (1) [14]. We denote such a SLOCC entanglement class by $\mathcal{C}$.

Membership in SLOCC classes.—How can one determine whether a given quantum state is separable or belongs to any specific SLOCC entanglement class $\mathcal{C}$? In principle, the algorithm introduced in Refs. [10,11] is applicable, but no convergence can be guaranteed, and the algorithm fails, in general, for rank-deficient states.

Recently, Brierley, Navascués, and Vértesi [18] presented a scheme for the problem of convex separation based on the so-called Gilbert's algorithm [15]. This scheme is shown to outperform existing linear programing methods for certain large-scale problems in the quantum information theory. For instance, nonlocality in bipartite scenarios can be certified with up to 42 measurement settings; new upper bounds are obtained for the visibility of certain states, as well as the steerability limit of Werner states. Basically, given any quantum state ρ, Gilbert's algorithm searches for a state $\rho^{\mathcal{C}} \in \mathcal{C}$ which approximates the minimal distance between ρ and the convex set $\mathcal{C}$. We denote such an operation by applying Gilbert's algorithm as $\rho^{\mathcal{C}} \equiv \mathcal{S}(\rho)$ for later use; see also Appendix A [19] for detailed discussions about this algorithm.

In any case, Gilbert's algorithm searches for an approximation of a given state ρ within the convex set $\mathcal{C}$. If a good approximation is found, this does not mean that the state ρ is within the set, as still it may be outside, but close to the boundary. Nevertheless, using some facts from the entanglement theory, we can modify the algorithm.

Proposition 1.—Gilbert's algorithm can be adapted to prove separability or membership in a certain SLOCC entanglement class $\mathcal{C}$.

Proof.—The proof relies on two facts about $\mathcal{C}$: (i) convexity and (ii) highly mixed states belong to $\mathcal{C}$. If the state ρ to be checked can be written as a convex combination of the state found by Gilbert's algorithm and any state within the highly mixed region, then $\rho \in \mathcal{C}$. See Appendix B [19] for the complete proof. ∎

By making use of Proposition 1, we tested different types of entanglement for various multiparticle quantum systems in Appendix C [19]. It clearly shows that, in most of the cases, Proposition 1 gives much better results compared to those obtained from previous known methods, whereas, in the following, we use Gilbert's algorithm as a tool to ensure constraints.

Convex optimization.—Denote by $\mathcal{F}(\rho)$ a strictly concave (or convex) function defined over the quantum state space $\mathcal{Q}$, such that $\mathcal{F}(\rho)$ has a single maximum (or minimum). Many functions in quantum information science meet this requirement, for instance, the log-likelihood function, the von Neumann entropy, etc. The statistical operator $\rho \in \mathcal{Q}$ has to satisfy two constraints, namely,

$$\rho \geq 0 \quad \text{and} \quad \mathrm{tr}(\rho) = 1. \tag{4}$$

We also assume that $\mathcal{F}(\rho)$ is differentiable (except perhaps at a few isolated points) with gradient $\nabla\mathcal{F}(\rho) \equiv G(\rho)$. The objective is to maximize $\mathcal{F}(\rho)$ over a specific SLOCC entanglement class $\mathcal{C} \subseteq \mathcal{Q}$, i.e., a convex subset of the state space. Explicitly, we have

$$\text{maximize} \quad \mathcal{F}(\rho) \tag{5a}$$
$$\text{subject to} \quad \rho \in \mathcal{C}. \tag{5b}$$

We denote the solution of this optimization by $\hat{\rho}_m^{\mathcal{C}}$.

As mentioned, it is hard to test whether a given state belongs to a SLOCC entanglement class, which makes the optimization defined above even harder. For this problem, we offer two iterative schemes, where the constraint in Eq. (5b) is guaranteed by Gilbert's algorithm. Specifically, each iterative step involves two operations, namely, one gradient operation (the update) followed by one Gilbert's operation (constraints enforced). For the gradient, we have two different approaches.

The direct-gradient (DG) scheme.—Let us first consider the case when $\mathcal{C} = \mathcal{Q}$, and then the constraint in Eq. (5b) is identical to that of Eq. (4), which can be ensured if one writes $\rho = A^\dagger A/\mathrm{tr}(A^\dagger A)$. In the unconstrained A space, the small variation of $\mathcal{F}(\rho)$ is given by

$$\delta\mathcal{F}(\rho) \equiv \delta\mathcal{F}(A) = \mathrm{tr}\left(\delta A \frac{[G - \mathrm{tr}(G\rho)]A^\dagger}{\mathrm{tr}(A^\dagger A)} + \mathrm{H.c.}\right), \tag{6}$$

to linear order in δA. If we choose $\delta A = \epsilon A[G - \mathrm{tr}(G\rho)]$ with ϵ being positive, then $\delta\mathcal{F}(\rho)$ is always positive and, hence, walking upwards. Thus, by following the gradient, we have the update for ρ in the DG scheme as

$$\begin{aligned}\rho_{k+1} &= \frac{1}{\mathcal{N}}\{\mathbb{1} + \epsilon[G_k - \mathrm{tr}(G_k\rho)]\}\rho_k\{\mathbb{1} + \epsilon[G_k - \mathrm{tr}(G_k\rho)]\} \\ &\equiv \mathrm{DG}(\rho_k, G_k, \epsilon)\end{aligned} \tag{7}$$

with $\mathcal{N}$ being the normalization constant.

Once the iteration is finished, the algorithm returns the optimal quantum state $\hat{\rho}_m$ with the corresponding optimal function value $\mathcal{F}_m = \mathcal{F}(\hat{\rho}_m)$ over the whole quantum state space. When $\mathcal{C}$ is strictly smaller than $\mathcal{Q}$, the state after the update in Eq. (7) may easily be outside of $\mathcal{C}$. Whenever this happens, we use Gilbert's algorithm to project ρ_k back to $\mathcal{C}$, i.e., $\rho_k \to \rho_k^{\mathcal{C}} = \mathcal{S}(\rho_k)$. Note that we also assume $\hat{\rho}_m \notin \mathcal{C}$; otherwise, $\hat{\rho}_m^{\mathcal{C}} \equiv \hat{\rho}_m$, and then the optimization in Eq. (5) is solved. With all the ingredients at hand, the DG algorithm proceeds as follows:

Algorithm: DG

Given $\epsilon > 0$ and $0 < \beta < 1$.
Choose any $\rho_0^{\mathcal{C}} \in \mathcal{C}$, and $\mathcal{F}_0 = \mathcal{F}(\rho_0^{\mathcal{C}})$.
for $k = 1, \ldots,$ *do*
  Update $\rho_k = \mathrm{DG}[\rho_{k-1}^{\mathcal{C}}, G(\rho_{k-1}^{\mathcal{C}}), \epsilon]$.
  Calculate $\rho_k^{\mathcal{C}} = \mathcal{S}(\rho_k)$, and $\mathcal{F}_k = \mathcal{F}(\rho_k^{\mathcal{C}})$.
  Termination criterion!
  if $\mathcal{F}_k < \mathcal{F}_{k-1}$, *then* (no update)
    Reset $\epsilon = \beta\epsilon$ and $\rho_k^{\mathcal{C}} = \rho_{k-1}^{\mathcal{C}}$.
  end if
end for

The initial step size ϵ can be chosen rather arbitrarily, which does not affect the final output too much. Though the DG algorithm is very simple and straightforward to use, it suffers two problems: slow convergence and low precision. These are due to the fact that the iterations in DG are actually performed in the unconstrained A space. When ρ is close to the boundary of the state space, it eventually becomes rank deficient with at least one small eigenvalue. The highly asymmetric spectrum would cause the gradient to be locally ill defined [28]. To avoid these problems, the state-of-the-art optimization method is to walk directly in the ρ space.

The accelerated projected-gradient (APG) scheme.—The APG approach [29–31] is generally applicable in all kinds of constrained problems where the constrains are enforced by a projection operation [32–36]. In the current scenario, we have to make sure that the update for ρ at each iterative step stays in $\mathcal{C}$ all the time, for which we use Gilbert's algorithm. Rather different from common gradient approaches, the update of the target ρ in APG is based on another state σ, such that each update gets some "momentum" from the previous step in order to find the optimal direction for the update. The momentum is controlled by θ in the algorithm, which will be reset to 1 whenever it causes the current step to point too far from the DG direction. Upon convergence, ρ and σ will eventually merge to the same point. For more technical details about the APG algorithm, e.g., the "restart" and "accelerate" operations, we refer to Ref. [28]. Then the APG algorithm proceeds as follows:

Algorithm: APG

Given $\epsilon > 0$ and $0 < \beta < 1$.
Choose any $\rho_0^{\mathcal{C}} \in \mathcal{C}$, and $\mathcal{F}_0 = \mathcal{F}(\rho_0^{\mathcal{C}})$;
  $\sigma_0 = \rho_0^{\mathcal{C}}$, and $\theta_0 = 1$.
for $k = 1, \ldots,$ *do*
  Update $\rho_k^{\mathcal{C}} = \mathcal{S}[\sigma_{k-1} + \epsilon G(\sigma_{k-1})]$, $\mathcal{F}_k = \mathcal{F}(\rho_k^{\mathcal{C}})$.
  Termination criterion!
  if $\mathcal{F}_k < \mathcal{F}_{k-1}$, *then* (restart)
    Reset $\epsilon = \beta\epsilon$, $\rho_k^{\mathcal{C}} = \rho_{k-1}^{\mathcal{C}}$, $\sigma_k = \rho_k^{\mathcal{C}}$, and $\theta_k = 1$.
  else (accelerate)
    Set $\theta_k = \frac{1}{2}(1 + \sqrt{1 + 4\theta_{k-1}^2})$;
    Update $\sigma_k = \rho_k^{\mathcal{C}} + [(\theta_{k-1} - 1)/\theta_k](\rho_k^{\mathcal{C}} - \rho_{k-1}^{\mathcal{C}})$.
  end if
end for

The MATLAB codes for the DG and APG algorithms, with accompanying documentation and implementations, are available online [37]. To guarantee the validity of these two algorithms, we have the following theorem.

Theorem 1.—Suppose Gilbert's algorithm is precise, i.e., the operation $\rho^{\mathcal{C}} \equiv \mathcal{S}(\rho)$ always returns the closest $\rho^{\mathcal{C}}$ with respect to ρ in the Hilbert-Schmidt norm. Then, if the iteration reaches a fixed point by the DG algorithm or the APG algorithm, this point is the solution to the optimization in Eq. (5).

Proof.—We prove this theorem by assuming contradictions: the convexity properties of $\mathcal{F}(\rho)$ and $\mathcal{C}$ are not compatible with two different solutions. For the complete proof, see Appendix D [19]. ∎

The likelihood-ratio test.—In real-world experiments, resources are limited; thus, the data obtained are always finite. Drawing conclusions from a finite amount of data requires statistical reasoning. In Ref. [12], a universal method for quantifying the weight of evidence for (or against) entanglement with finite data was introduced. However, being boiled down to an optimization over specific convex sets of entanglement classes, this method is generally not doable. Here, we show that this problem can be tackled by using our algorithms.

In a typical quantum tomographic scenario [38], N independently and identically prepared copies of the quantum state ρ are measured by a positive operator-valued measure $\{\Pi_k\}_{k=1}^K$, with $\Pi_k \geq 0 \forall k$ and $\sum_{k=1}^K \Pi_k = \mathbb{1}$. The data $D = \{n_1, n_2, \ldots, n_K\}$ consist of a sequence of detector clicks, the probability of getting which is given by the likelihood function

$$\mathcal{L}(D|\rho) = \prod_k p_k^{n_k} = \left(\prod_k [\mathrm{tr}(\rho\Pi_k)]^{f_k}\right)^N, \qquad (8)$$

where $p_k = \mathrm{tr}(\rho\Pi_k)$ (Born rule) is the probability for outcome Π_k and $f_k = n_k/N$ denotes the relative frequency. Note that $\mathcal{L}(D|\rho)$ is not strictly concave, but the normalized log-likelihood $\mathcal{F}(\rho) \equiv (1/N)\ln\mathcal{L}(D|\rho)$ is.

The likelihood ratio in Ref. [12] is defined as

$$\Lambda \equiv \frac{\max_{\rho\in\mathcal{C}}\mathcal{L}(D|\rho)}{\max_{\mathrm{all}\rho}\mathcal{L}(D|\rho)}, \quad \text{and} \quad \lambda = -2\ln(\Lambda) \qquad (9)$$

represents the weight of evidence in favor of entanglement. Hence, to demonstrate entanglement convincingly, a large value of λ is demanded. Moreover, for states lying close to the boundary of $\mathcal{C}$, it has been shown in Ref. [12] that λ follows a semi-χ_1^2 distribution for large enough N. By having this, one can perform hypothesis testing to demonstrate entanglement and then construct confidence levels. Suppose we get ρ_{exp} with the corresponding λ_{exp} in an experiment; the p-value for the null hypothesis that $\rho_{\mathrm{exp}} \in \mathcal{C}$ is given by the probability $\Pr(\lambda > \lambda_{\mathrm{exp}}) \equiv \epsilon$. Therefore, with the $(1 - \epsilon)$ confidence level, the null hypothesis has to be rejected, indicating that the state is entangled.

The likelihood ratio defined in Eq. (9) involves two optimizations over two different convex sets. The maximization in the denominator is well known as the maximum-likelihood estimation [39], for which various algorithms exist; while the maximization in the numerator fits exactly into our problem. Let us denote the solutions to these two maximizations by $\hat{\rho}_m =: \arg\max_{\text{all}\rho}\mathcal{F}(\rho)$ and $\hat{\rho}_m^{\mathcal{C}} =: \arg\max_{\rho\in\mathcal{C}}\mathcal{F}(\rho)$, respectively, and then Eq. (9) can be rewritten as $\lambda = 2N[\mathcal{F}(\hat{\rho}_m) - \mathcal{F}(\hat{\rho}_m^{\mathcal{C}})]$.

Four-qubit W state with white noise.—For the first application, let us consider the four-qubit W state $|W_4\rangle = (|0001\rangle + |0010\rangle + |0100\rangle + |1000\rangle)/2$ mixed with white noise:

$$\rho_{W_4}(q) = q|W_4\rangle\langle W_4| + \frac{1-q}{16}\mathbb{1}. \tag{10}$$

By employing Proposition 1, we find $\rho_{W_4}(q)$ is biseparable for $q \leq 0.4555$ and fully separable for $q \leq 0.09$; see Table I in Appendix C [19].

In the simulation, we choose the noise level $q = 0.9$ and then employ the standard Pauli tomographic scheme where each qubit is measured in the basis of the three Pauli operators. Without the loss of generality, we set $\{f_k = p_k\}$ such that the maximum-likelihood estimator *is* the true state. Hereafter, we calculate instead the normalized log-likelihood ratios, i.e., λ/N. Figure 1 shows the results for testing biseparability as well as full separability for this case. As expected, the λ/N value obtained for biseparability is much smaller than that for full separability, as the fully separable states consist of a strictly smaller subset of the biseparable region. Moreover, the APG algorithm usually has a better precision-resolvent capability than DG does. For more simulated examples, see Appendix E [19].

Experimental bound entanglement.—The four-party Smolin state [40] is

$$\rho_S = \frac{1}{4}\sum_{\mu=1}^{4} |\Psi^\mu\rangle\langle\Psi^\mu|_{AB} \otimes |\Psi^\mu\rangle\langle\Psi^\mu|_{CD}, \tag{11}$$

where the subscripts label the parties and $|\Psi^\mu\rangle$ are the two-qubit Bell states. By adding white noise, we have $\rho_S(q) = q\rho_S + (1-q)\mathbb{1}/16$, which is fully separable for $q \leq 1/3$ and bound entangled for $q > 1/3$ [41]. In Ref. [42], a family of noisy four-photon Smolin states was generated by spontaneous parametric down-conversion. By varying the noise level, bound entanglement was successfully demonstrated for $q = 0.51$. Here, we reanalyze their experimental data using the likelihood-ratio test.

To demonstrate bound entanglement, one has to show that the state has a positive partial transpose (PPT) [43] but is nevertheless entangled. For this, optimizations over two different convex sets, namely, sets of the fully separable states as well as the PPT states, have to be performed; see the results in Fig. 2. At noise level $q = 0.51$, we get $\lambda \approx 2.42 \times 10^3$ with the p-value ≈ 0 for the null hypothesis that the state is separable. Thus, the null

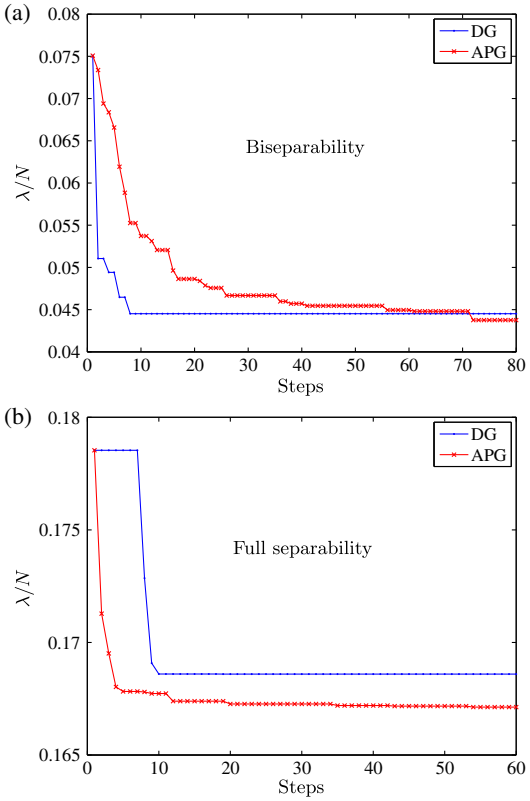

FIG. 1. The normalized log-likelihood ratios λ/N at each iterative step by DG and APG, respectively, for the four-qubit W state with white noise: (a) Biseparability; (b) Full separability. The plateaus in the plots imply the process where the algorithms are searching for suitable step sizes for the next update. As shown, the APG algorithm usually returns more accurate solutions than DG does.

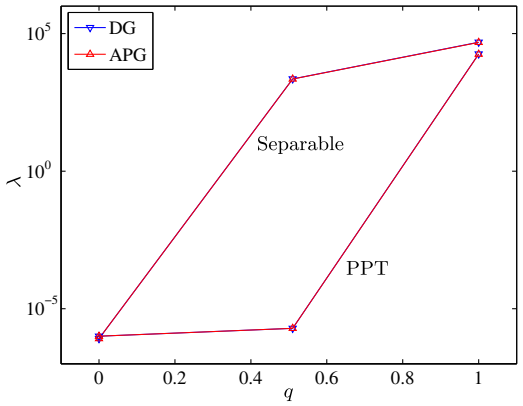

FIG. 2. Likelihood-ratio test for bound entanglement for the experimental four-photon Smolin state [42]. The top curve labeled "Separable" indicates the λ values obtained via maximizing over the fully separable states; while the bottom curve labeled "PPT" gives values obtained by maximizing over the PPT states. Hypothesis testing suggests that the state at noise level $q = 0.51$ is both entangled and PPT and, thus, bound entangled.

hypothesis has to be rejected, so the state is indeed entangled. Meanwhile, we get $\lambda \approx 1.94 \times 10^{-6}$ for the optimization over PPT states, indicating strongly that the state is PPT [44]. Therefore, the state at noise level $q = 0.51$ is both entangled and PPT and, thus, bound entangled. Similarly, one can conclude from the λ values that the state at noise level $q = 0$ is separable, while the state at $q = 1$ is genuinely entangled.

In Appendix F [19], we use simulated data to perform the likelihood-ratio test for various noise levels. By doing so, we identify the parameter range $q \sim [0.35, 0.8]$ (containing $q = 0.51$ from the real experiment), which is most likely to show bound entanglement.

Conclusions.—The characterization of multiparticle entanglement is generally hard. In this work, we show that Gilbert's algorithm can be adapted to prove a given quantum state is either separable or belongs to a SLOCC entanglement class, with the thresholds thus obtained being much better than those reported by previous known methods. Furthermore, with the help of Gilbert's algorithm, two reliable schemes are presented for the convex optimization over any defined SLOCC entanglement classes. For demonstration, we reanalyzed the experimental data on bound entanglement of the noisy four-photon Smolin states using the likelihood-ratio test. As such, we expect that our methods would become a reliable tool for experimentalists to test the entanglement property of their quantum systems with confidence.

This work has been supported by the European Research Council (Consolidator Grant No. 683107/TempoQ) and the Deutsche Forschungsgemeinschaft. We thank H. K. Ng, Z. Zhang, S. Brierley, T. Vértesi, and H. Zhu for stimulating discussions. We are also grateful to J. Lavoie for sharing the experimental data in Ref. [42], to M. Kleinmann and T. Monz for sharing the data in Ref. [10], and to H. Kampermann for sharing the codes used in Ref. [11].

*jiangwei.shang@bit.edu.cn
†otfried.guehne@uni-siegen.de

[1] O. Gühne and G. Tóth, Phys. Rep. **474**, 1 (2009).
[2] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, Rev. Mod. Phys. **81**, 865 (2009).
[3] B. M. Terhal, Phys. Lett. A **271**, 319 (2000).
[4] W. K. Wootters, Phys. Rev. Lett. **80**, 2245 (1998).
[5] A. Sanpera, R. Tarrach, and G. Vidal, Phys. Rev. A **58**, 826 (1998).
[6] R. Unanyan, H. Kampermann, and D. Bruß, J. Phys. A **40**, F483 (2007).
[7] A. Kay, Phys. Rev. A **83**, 020303(R) (2011).
[8] F. M. Spedalieri, Phys. Rev. A **76**, 032318 (2007).
[9] M. Navascues, M. Owari, and M. B. Plenio, Phys. Rev. Lett. **103**, 160404 (2009).
[10] J. T. Barreiro, P. Schindler, O. Gühne, T. Monz, M. Chwalla, C. F. Roos, M. Hennrich, and R. Blatt, Nat. Phys. **6**, 943 (2010).
[11] H. Kampermann, O. Gühne, C. Wilmott, and D. Bruß, Phys. Rev. A **86**, 032307 (2012).
[12] R. Blume-Kohout, J. O. S. Yin, and S. J. van Enk, Phys. Rev. Lett. **105**, 170501 (2010).
[13] W. Dür, G. Vidal, and J. I. Cirac, Phys. Rev. A **62**, 062314 (2000).
[14] A. Acín, D. Bruß, M. Lewenstein, and A. Sanpera, Phys. Rev. Lett. **87**, 040401 (2001).
[15] E. G. Gilbert, SIAM J. Control **4**, 61 (1966).
[16] L. Gurvits, in *Proceedings of the 35th ACM Symposium on Theory of Computation* (ACM Press, New York, 2003), pp. 10–19.
[17] S. Gharibian, Quantum Inf. Comput. **10**, 343 (2010).
[18] S. Brierley, M. Navascués, and T. Vértesi, arXiv:1609.05011.
[19] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.120.050506 for the Appendixes, which includes Refs. [20–27].
[20] L. Gurvits and H. Barnum, Phys. Rev. A **66**, 062311 (2002).
[21] P. Horodecki, Phys. Lett. A **232**, 333 (1997).
[22] Z.-H. Chen, Z.-H. Ma, O. Gühne, and S. Severini, Phys. Rev. Lett. **109**, 200503 (2012).
[23] B. Jungnitsch, T. Moroder, and O. Gühne, Phys. Rev. Lett. **106**, 190502 (2011).
[24] C. Eltschka and J. Siewert, Phys. Rev. Lett. **108**, 020502 (2012).
[25] A. C. Doherty, P. A. Parrilo, and F. M. Spedalieri, Phys. Rev. Lett. **88**, 187904 (2002).
[26] P. Hyllus, Ph.D. thesis, Universität Hannover, 2005.
[27] S. Lu, S. Huang, K. Li, J. Li, J. Chen, D. Lu, Z. Ji, Y. Shen, D. Zhou, and B. Zeng, arXiv:1705.01523.
[28] J. Shang, Z. Zhang, and H. K. Ng, Phys. Rev. A **95**, 062336 (2017).
[29] A. Beck and M. Teboulle, SIAM J. Imaging Sci. **2**, 183 (2009).
[30] B. O'Donoghue and E. Candès, Found. Comput. Math. **15**, 715 (2015).
[31] S. R. Becker, E. Candès, and M. C. Grant, Math. Program. Comput. **3**, 165 (2011).
[32] A. A. Goldstein, Bull. Am. Math. Soc. **70**, 709 (1964).
[33] E. S. Levitin and B. T. Polyak, Zh. Vychisl. Mat. Mat. Fiz. **6**, 787 (1966) [USSR Computational Mathematics and Mathematical Physics **6**, 1 (1966)].
[34] R. J. Bruck, J. Math. Anal. Appl. **61**, 159 (1977).
[35] G. B. Passty, J. Math. Anal. Appl. **72**, 383 (1979).
[36] Y. Nesterov, *Introductory Lectures on Convex Optimization: A Basic Course* (Kluwer Academic, Dordrecht, 2004).
[37] https://github.com/qCvxOpt/CvxOptSLOCC/.
[38] *Quantum State Estimation*, edited by M. Paris and J. Řeháček, Lecture Notes in Physics Vol. 649 (Springer, Heidelberg, 2004).
[39] Z. Hradil, J. Řeháček, J. Fiurášek, and M. Ježek, in Ref. [30], Chap. 3.
[40] J. A. Smolin, Phys. Rev. A **63**, 032306 (2001).
[41] R. Augusiak and P. Horodecki, Phys. Rev. A **74**, 010305 (2006).
[42] J. Lavoie, R. Kaltenbaek, M. Piani, and K. J. Resch, Phys. Rev. Lett. **105**, 130501 (2010).
[43] A. Peres, Phys. Rev. Lett. **77**, 1413 (1996).
[44] Unfortunately, we cannot do hypothesis testing to check for PPT, as the set of negative partial transpose (NPT) states is not convex. A p-value ≈ 1 for the null hypothesis that the state is PPT should not be interpreted as a p-value ≈ 0 for the null hypothesis that the state is NPT.