Entanglement Verification with Finite Data

Robin Blume-Kohout,[1] Jun O. S. Yin,[2] and S. J. van Enk[2,3]

[1]*Perimeter Institute for Theoretical Physics, Waterloo ON N2L2Y5, Canada*
[2]*Department of Physics and Oregon Center for Optics, University of Oregon, Eugene, Oregon 97403, USA*
[3]*Institute for Quantum Information, California Institute of Technology, Pasadena, California 91125, USA*

Suppose an experimentalist wishes to verify that his apparatus produces entangled quantum states. A finite amount of data cannot conclusively demonstrate entanglement, so drawing conclusions from real-world data requires statistical reasoning. We propose a reliable method to quantify the weight of evidence for (or against) entanglement, based on a likelihood ratio test. Our method is universal in that it can be applied to any sort of measurements. We demonstrate the method by applying it to two simulated experiments on two qubits. The first measures a single entanglement witness, while the second performs a tomographically complete measurement.

Entanglement is an essential resource for quantum information processing, and producing and verifying entangled states is considered a benchmark for quantum experiments (for a sample from the most recent experiments on a wide variety of physical systems, see Ref. [1]). Several methods for verifying entanglement have been developed (for overviews, see Refs. [2,3]). A bipartite state is entangled if it is not separable, and data $\mathcal{D}$ demonstrate entanglement if there is no separable state that could have generated them. As the number of data $N \rightarrow \infty$, the data are unambiguous, but for finite N, only probabilistic conclusions can be drawn. In this Letter, we quantify exactly what can be concluded from finite or small data sets, using a simple and efficient likelihood ratio test.

We demonstrate the method using two simulated experiments on two-qubit systems [4]. The first measures just one observable, an entanglement witness [5]. The other performs a tomographically complete measurement. The first case can be reduced to a single-parameter problem and can be treated analytically; the second case explores the intricacies of the full multidimensional state space and must be treated numerically. In both cases, we use likelihood ratios to draw direct conclusions about entanglement, rather than estimating the quantum state as an intermediate step. A related technique for testing violation of local realism, and based on empirical relative entropy instead of the likelihood ratio, was proposed by van Dam *et al.* [6] and applied by Zhang *et al.* [7].

Likelihood ratios.—Data $\mathcal{D}$ could have been generated by any one of many i.i.d. (independently and identically distributed) states $\rho^{\otimes N}$. Each state ρ represents a theory about the system, and the relative plausibility of different states is measured by their likelihood $\mathcal{L}(\rho)$. A state's likelihood is simply the probability of the observed data given that state,

$$\mathcal{L}(\rho) \equiv \Pr(\mathcal{D}|\rho), \qquad (1)$$

and states with higher likelihood are more plausible. If the most likely state is separable, the data clearly do not support entanglement. If it is entangled, then we need to ask how convincing the data are—specifically, whether some separable state is almost as plausible. To judge whether there is (even just one) separable state that fits the data, we compare the likelihoods of (i) the most likely separable state, and
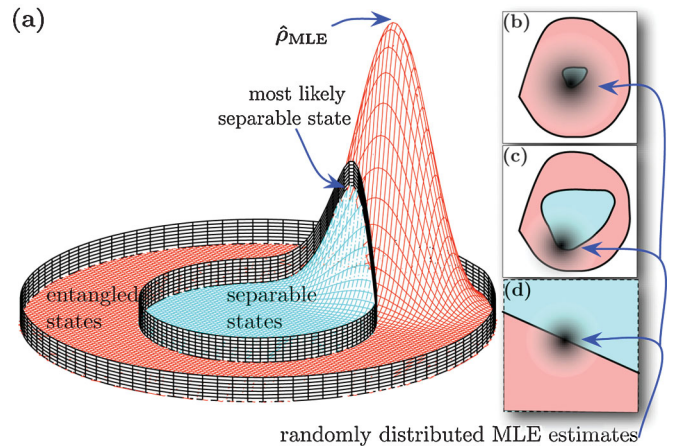

FIG. 1 (color). General schema of a likelihood ratio test. The separable states $\mathcal{S}$ (blue) are a convex subset of all states, surrounded by entangled states (red). Data from an experiment on a state ρ yield a quasiconvex likelihood function [(a)] with a unique maximum ($\hat{\rho}_{\mathrm{MLE}}$). $\hat{\rho}_{\mathrm{MLE}}$ is randomly distributed around ρ, at a typical length scale $\delta = O(1/\sqrt{N})$. If $\hat{\rho}_{\mathrm{MLE}}$ is separable then there is no evidence for entanglement, but if it is entangled (as shown), then the relative likelihoods of $\hat{\rho}_{\mathrm{MLE}}$ and the most likely separable state determine the weight of evidence. Data are "convincing" if they are very unlikely to have been produced by a borderline separable state. Typical likelihood ratios for such states depend on the shape of $\mathcal{S}$. In (b)–(d) we show three possible cases: in (b) $\mathcal{S}$ is smaller than δ and behaves like a point; in (c) it is of size δ and its behavior is hard to characterize; in (d) it is much bigger than δ and behaves like a half-space.

(ii) the most likely of all states. Letting $\mathcal{S}$ be the set of separable states, we define

$$\Lambda \equiv \frac{\max_{\rho \in \mathcal{S}} \mathcal{L}(\rho)}{\max_{\text{all } \rho} \mathcal{L}(\rho)}. \qquad (2)$$

Λ is a likelihood ratio, and

$$\lambda = -2\log\Lambda \qquad (3)$$

represents the weight of evidence in favor of entanglement [8]. To demonstrate entanglement convincingly, an experiment must yield a sufficiently large value for λ.

A likelihood ratio does not assign a probability to "ρ is entangled". Instead, it yields a confidence level. We can determine what values of λ typically result from measurements on $\rho^{\otimes N}$, and how their distribution depends on whether ρ is entangled or separable. If we measure $\lambda = \lambda_{\text{exp}}$, and no separable state produces $\lambda \geq \lambda_{\text{exp}}$ with probability higher than ϵ, then we have demonstrated entanglement at the $1 - \epsilon$ confidence level. If an experimentalist plans (before taking data) to calculate λ and report "ρ is entangled" only when the data imply $1 - \epsilon$ confidence, then the probability that he erroneously reports entanglement [9] is at most ϵ.

So, ρ may be (i) entangled, (ii) separable, or (iii) on the boundary. Boundary states are still separable, and they are the hardest separable states to rule out. To demonstrate entanglement at the $1 - \epsilon$ confidence level, we must show that there is no boundary state for which $\Pr(\lambda \geq \lambda_{\text{exp}}) \geq \epsilon$. It is difficult to make rigorous probabilistic statements about λ for small N. But as $N \to \infty$, the following analysis becomes exact, and is generally thought to be reliable for $N \gtrsim 30$ [10].

The distribution of λ.—The set of quantum states ρ is a convex subset of the affine space of trace-1 $d \times d$ Hermitian operators, $\mathbb{R}^{d^2-1}$. An entanglement-verification measurement is represented by a POVM (positive operator-valued measure) $\mathcal{M} = \{E_i, \ldots, E_m\}$, in which each operator E_k represents an event that occurs with probability $p_k = \text{Tr}E_k\rho$ (Born's rule), and each ρ defines a probability distribution $\vec{p} = \{p_1, \ldots, p_m\}$. Data in which E_k appeared n_k times define empirical frequencies $\vec{f} = \{f_1, \ldots, f_m\}$, where $f_k \equiv \frac{n_k}{N}$. Both $\vec{p}$ and $\vec{f}$ can be represented as elements of an m-simplex embedded in a vector space $\mathbb{R}^{m-1}$. The probabilities in $\vec{p}$ may be linearly dependent (e.g., if $E_j + E_k = \mathbb{1}$, then $p_j + p_k = 1$ for all ρ), and at most $d^2 - 1$ of them can be independent (because ρ contains only $d^2 - 1$ parameters). We define $\dim(\mathcal{M})$ as the number of independent probabilities.

So Born's rule defines an affine mapping from the operator space containing quantum states into the probability space for measurement $\mathcal{M}$. If $\dim(\mathcal{M}) < d^2 - 1$, then the mapping from states to $\vec{p}$ vectors is many to one, and the experiment is completely insensitive to some parameters of ρ. Ignoring these irrelevant parameters makes ρ an (effectively) $\dim(\mathcal{M})$-dimensional parameter. Separable states form a convex subset of all states (see Fig. 1).

These sets' images in probability space are also nested convex sets (although if $\dim(\mathcal{M}) < d^2 - 1$, then some entangled states will be indistinguishable from separable ones in this experiment).

Suppose that N copies of a state ρ_0 are measured, yielding a likelihood function $\mathcal{L}(\rho)$. $\mathcal{L}(\rho)$ has a unique global maximum $\hat{\rho}_{\text{MLE}}$. As $N \to \infty$, the distribution of $\hat{\rho}_{\text{MLE}}$ approaches a Gaussian around ρ_0 with covariance tensor Δ. $\mathcal{L}(\rho)$ itself is a Gaussian function with the same covariance matrix Δ (see note [11]). This defines a characteristic length scale $\delta = |\Delta|_2$ that scales as $\delta = O(1/\sqrt{N})$. We can use Δ to define a stretched Euclidean metric

$$d(\rho_1, \rho_2) = \sqrt{\text{Tr}[(\rho_1 - \rho_2)\Delta^{-1}(\rho_1 - \rho_2)]}. \qquad (4)$$

Using this metric, $\hat{\rho}_{\text{MLE}}$ is univariate Gaussian distributed around ρ_0, and $\log\mathcal{L}(\rho) = -\frac{d(\rho,\hat{\rho}_{\text{MLE}})^2}{2}$. Thus, λ is determined entirely by $d(\hat{\rho}_{\text{MLE}}, \mathcal{S})$, the distance from $\hat{\rho}_{\text{MLE}}$ to the separable set $\mathcal{S}$. If ρ_0 is demonstrably entangled, then λ will grow proportional to N—but if it is indistinguishable from a separable state, then λ will converge almost certainly to zero (see Fig. 2).

When ρ_0 is on the boundary, λ neither grows with N nor converges to zero, but continues to fluctuate as $N \to \infty$. Its distribution is controlled by the shape and radius of $\mathcal{S}$. If $\mathcal{S}$ is small with respect to δ, it behaves like a point
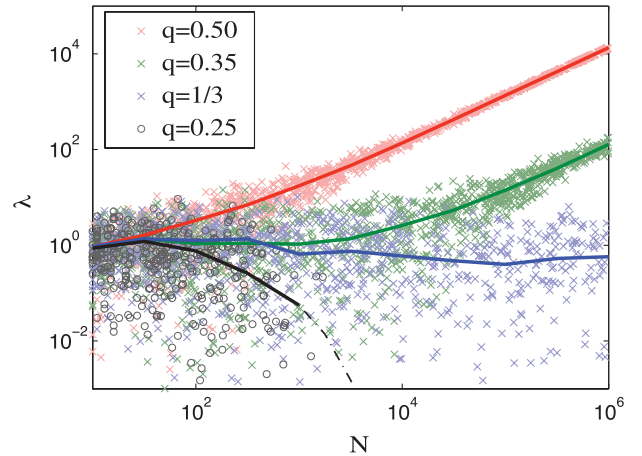

FIG. 2 (color). Log-likelihood ratios (λ) behave dramatically differently for different states. 1000 independent simulated tomographically complete experiments were performed, on four different Werner states—separable ($q = 0.25$), barely separable ($q = 1/3$), slightly entangled ($q = 0.35$), and highly entangled ($q = 0.50$). λ is shown for each trial (points), and averaged over all 1000 trials (solid lines). For small N the experiment cannot reliably distinguish them. As N grows, it resolves shorter distances in the state space. For entangled states, typical values of λ increase linearly with N, whereas the separable state almost certainly yields $\lambda = 0$ [not visible in these plots; for $\rho_{q=0.25}$ (black), all trials with more than $N \sim 10^3$ measurements yielded $\lambda = 0$, and the average (dashed line) plunges off the graph]. For barely separable states, λ behaves as a semi-χ_k^2 variable with $k = 1$ as $N \to \infty$ (see Fig. 3).

[see Fig. 1(b)]. Then $d(\hat{\rho}_{MLE}, \mathcal{S}) \approx d(\hat{\rho}_{MLE}, \rho_0)$, $\lambda = -2\log[\mathcal{L}_{max}/\mathcal{L}(\rho_0)] = d(\rho, \hat{\rho}_{MLE})^2$, and so λ is a χ^2 random variable with $\dim(\mathcal{M})$ degrees of freedom (also known as a $\chi^2_{\dim(\mathcal{M})}$ variable). On the other hand, if $\mathcal{S}$ is much larger than δ, then it behaves like a half-space (see Fig. 1(d) and Ref. [12]). If $\mathcal{S}$ were a k-dimensional hyperplane, λ would be a $\chi^2_{\dim(\mathcal{M})-k}$ variable. A half-space behaves like a hyperplane of dimension $[\dim(\mathcal{M}) - 1]$, except with probability $\frac{1}{2}$, $\hat{\rho}_{MLE}$ is separable. Thus, λ is what we will call a semi-χ^2_1 variable: it equals zero with probability $\frac{1}{2}$, and is χ^2_1 distributed otherwise.

As $N \to \infty$, $\delta \to 0$ and the latter case applies. For small N, however, the real situation is somewhere in between [see Fig. 1(c)]. $\mathcal{S}$ may be small, and its boundary may be sharply curved, increasing λ. In the absence of a detailed understanding of $\mathcal{S}$'s shape, case (1) provides the best rigorous upper bound on λ. Its cumulative distribution is upper bounded by that of a $\chi^2_{\dim(\mathcal{M})}$ variable—i.e., $\Pr(\lambda > x)$ is no greater than it would be if λ was a $\chi^2_{\dim(\mathcal{M})}$ variable. As $N \to \infty$, the more optimistic semi-χ^2_1 ansatz is valid—but only if we know that N is "large enough".

A χ^2_k variable has expected value k, and higher values are exponentially suppressed. So $\lambda \gg \dim(\mathcal{M})$ is sufficient to demonstrate entanglement at a high confidence level. This implies a tradeoff between an experiment's power (ability to identify many entangled states) and its efficiency (ability to do so rapidly). Powerful experiments have large dimension—e.g., a tomographically complete measurement can identify any entangled state, but has $\dim(\mathcal{M}) = d^2 - 1$. This comes at a price; experiments with large dimension are potentially much more prone to spurious large values of λ, so more data is required to achieve conclusive results [$\lambda \gg \dim(\mathcal{M})$]. Conversely, an entanglement witness (see below) is targeted at a particular state, but it can rapidly and conclusively demonstrate entanglement.

Implementation.—Computing λ involves maximizing $\mathcal{L}(\rho)$ over two convex sets (the set of all states, and the set $\mathcal{S}$ of separable states). $\mathcal{L}(\rho)$ is log-convex, so in principle this is a convex program.

Testing separability is NP-hard, so efficient minimization over $\rho \in \mathcal{S}$ is impossible in general. But for two qubits, the positive partial transpose (PPT) criterion perfectly characterizes entanglement, and λ can be calculated easily (see examples below). For larger systems, $\mathcal{S}$ can be bounded by simpler convex sets, as $\mathcal{S}_- \subset \mathcal{S} \subset \mathcal{S}_+$, (e.g., $\mathcal{S}_+ = $ PPT states, and $\mathcal{S}_- = $ convex combinations of specific product states). Maximizing $\mathcal{L}(\rho)$ over $\mathcal{S}_+$ and $\mathcal{S}_-$ yields bounds on $\max_{\rho \in \mathcal{S}} \mathcal{L}(\rho)$, which may (depending on how wisely the bounding sets were chosen) be tight enough to confirm or deny entanglement.

Examples.—To demonstrate the likelihood ratio test, we simulate two different experiments on two qubits. We imagine an experimentalist trying to produce the singlet state $|\Psi\rangle$, and producing instead a Werner state [13], $\rho_q = q\Pi_{singlet} + (1-q)\mathbb{1}/4$, where $\Pi_{singlet} = |\Psi\rangle\langle\Psi|$. Werner states are separable when $q \le 1/3$, and entangled otherwise. The experimentalist's repeated preparations are assumed to be i.i.d. [14].

Witness data.—The simplest way to test for entanglement is to repeatedly measure a single entanglement witness [2,5]. An optimal witness for Werner states is $W = \mathbb{1}/2 - \Pi_{singlet}$. Measuring W yields one of two outcomes—"yes" or "no"—corresponding to POVM elements $\{\Pi_{singlet}, \mathbb{1} - \Pi_{singlet}\}$. The probability of a yes outcome is given by Born's rule as $p = \text{Tr}\rho\Pi_{singlet}$, so p completely characterizes a state ρ for the purposes of this experiment. The data from N measurements are fully characterized by the frequency of yes results, $f = n_{yes}/N$. As $N \to \infty$, $f > \frac{1}{2}$ represents definitive proof that $\langle W \rangle < 0$, and therefore that ρ is entangled. For finite N, $f \le \frac{1}{2}$ means that a separable state fits as well as any other, so there is no case for entanglement. When $f > \frac{1}{2}$, our likelihood ratio quantifies the weight of the evidence for entanglement. The likelihood function depends only on p, as

$$\mathcal{L}(\rho) = \mathcal{L}(p) = \Pr(f|p) = p^{Nf}(1-p)^{N(1-f)}$$
$$= e^{-N[-f\log p - (1-f)\log(1-p)]}, \quad (5)$$

making this a single-parameter problem. The maximum likelihood, attained at $p = f$, is $\mathcal{L}_{max} = e^{-NH(f)}$, expressed in terms of the data's empirical entropy,

$$H(f) = -f\log f - (1-f)\log(1-f). \quad (6)$$

If $f > \frac{1}{2}$, the most likely separable state has $p = \frac{1}{2}$, so that $\mathcal{L}_{sep} = 2^{-N}$, which yields

$$\lambda = -2\log\frac{\mathcal{L}_{sep}}{\mathcal{L}_{max}} = 2N[\log(2) - H(f)]. \quad (7)$$

Our numerical explorations (not shown here) confirm that for a barely separable Werner state, λ behaves as a semi-χ^2_1 variable, even for N as low as 20.

Tomographically complete data.—Many entanglement-verification experiments measure a tomographically complete set of observables on a finite-dimensional system (with a heroic example being tomography on 8 ions in an ion trap [15]). Such data identify ρ uniquely as $N \to \infty$, so one can determine with certainty whether ρ is entangled (modulo the computational difficulties in determining whether a specified ρ is separable). Analyzing finite data is more complicated than in the witness example, for the data constrain a multidimensional parameter space. *Ad hoc* techniques are unreliable, and the likelihood ratio test comes into its own.

We consider an apparatus that applies a SIC (symmetric informationally complete)-POVM [16] to each of our two qubits, independently. This measurement (not to be confused with a four-dimensional SIC-POVM) is
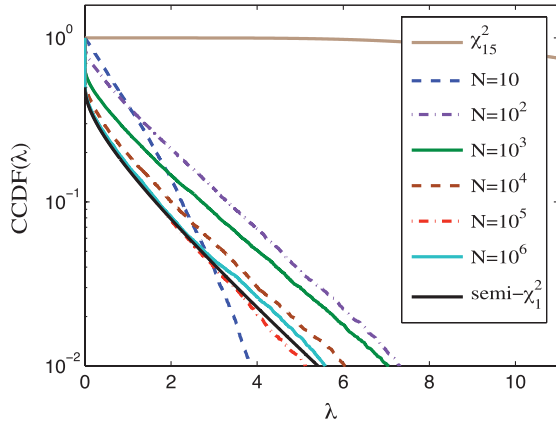
FIG. 3 (color). Distribution of λ for a SIC-POVM experiment. We show the empirical complementary cumulative distribution function of λ, $\mathrm{CCDF}(\lambda_c) = \Pr(\lambda > \lambda_c)$, for the state $\rho_{q=1/3}$ and simulated datasets of size $N = \{10, \ldots, 10^6\}$. The CCDF is used to compute confidence levels—e.g., to report entanglement at the 95% confidence level, it is necessary to observe λ such that $\mathrm{CCDF}(\lambda) < 0.05$. For this particular state, the chance of a zero λ approaches 50% as N increases. For each N, $\mathrm{CCDF}(\lambda_c)$ was based on roughly 10^4 data points from independent trials, each of which generated a value of λ from N tomographically complete measurements on $\rho_{q=1/3}$. We also show CCDFs for a semi-χ_1^2 variable and a $\chi_{\dim(\mathcal{M})}^2 = \chi_{15}^2$ variable. The semi-χ_1^2 ansatz is good for large N, but unreliable for small N (yielding too many false positives), while the χ_{15}^2 ansatz is very conservative.

tomographically complete, has $4 \times 4 = 16$ outcomes, and yields 15 independent frequencies. Unlike W, it has no special relationship to Werner states, so any entangled ρ will yield overwhelmingly convincing data as $N \to \infty$.

We repeatedly simulated $N = 10, \ldots, 10^6$ measurements on a barely-separable Werner state ($\rho_{q=1/3}$), and compared the empirical distribution of λ to those of semi-χ_1^2 and χ_{15}^2 random variables (see Fig. 3). As N gets large, λ becomes indistinguishable from a semi-χ_1^2 variable. For smaller N, this ansatz is too optimistic (and would produce excessive false positives), but the $\chi_{d^2-1}^2$ ansatz is wildly overcautious. We found that for small N, λ behaves like a semi-χ_D^2 variable, with D a bit larger than 1 (e.g., $D \approx 1.6$ for $N = 100$).

Conclusions.—Entanglement verification is easy when $N \to \infty$. In practice, N is finite and data are never conclusive. Likelihood ratios provide a simple, reliable test of significance that can be applied to any experimental data. Large values of λ are very unlikely to be generated by any separable state, but the hardest separable states to rule out are on the boundary. For such states, theory predicts (and our numerics confirm) that λ behaves like a semi-χ^2 random variable. If the underlying state is separable, $\Pr(\lambda > x)$ can be upper bounded using a $\chi_{\dim(\mathcal{M})}^2$ distribution, scaling as e^{-x} for large x. For entangled states,

λ grows linearly with N, and will thus rapidly become distinguishable from any separable state.

[1] L. Hofstetter *et al.*, Nature (London) **461**, 960 (2009); M. Ansmann *et al.*, Nature (London) **461**, 504 (2009); E. Amselem and M. Bourennane, Nature Phys. **5**, 748 (2009); P. Bohi *et al.*, Nature Phys. **5**, 592 (2009); L. DiCarlo *et al.*, Nature (London) **460**, 240 (2009); J. Janousek *et al.*, Nat. Photon. **3**, 399 (2009); J. D. Jost *et al.*, Nature (London) **459**, 683 (2009); J. C. F. Matthews *et al.*, Nat. Photon. **3**, 346 (2009); A. Fedrizzi *et al.*, Nature Phys. **5**, 389 (2009); A. Ourjoumtsev *et al.*, Nature Phys. **5**, 189 (2009); A. S. Coelho *et al.*, Science **326**, 823 (2009); Scott B. Papp *et al.*, Science **324**, 764 (2009); Ryo Okamoto *et al.*, Science **323**, 483 (2009); S. Olmschenk *et al.*, Science **323**, 486 (2009).
[2] O. Gühne and G. Toth, Phys. Rep. **474**, 1 (2009).
[3] S. J. van Enk, N. Lutkenhaus, and H. J. Kimble, Phys. Rev. A **75**, 052318 (2007).
[4] For larger systems, determining whether a given state is entangled is an NP-hard problem. In multipartite systems, different classes of entanglement exist, but their classification is still an open problem. Our likelihood ratio method applies to any case where the decision is *binary*: do the data demonstrate entanglement in a particular class or not? So in this Letter, the word separable can be generalized to "not in the desired entanglement class".
[5] M. Horodecki, P. Horodecki, and R. Horodecki, Phys. Lett. A **223**, 1 (1996); B. Terhal, Phys. Lett. A **271**, 319 (2000); P. Hyllus *et al.*, Phys. Rev. A **72**, 012321 (2005).
[6] W. van Dam, R. D. Gill, and P. D. Grunwald, IEEE Trans. Inf. Theory **51**, 2812 (2005).
[7] Y. Zhang, E. Knill, and S. Glancy, Phys. Rev. A **81**, 032117 (2010).
[8] The factor of -2 may seem arbitrary. Statisticians use this convention because λ (as defined) is in many circumstances a χ^2 random variable (see text, below).
[9] Statisticians call this a "Type I error". Erroneously *rejecting* entanglement, even though the experiment is capable of demonstrating entanglement (which is not the same as reporting separability), is a "Type II error". In entanglement verification one tries to avoid Type I errors and is merely mildly unenthusiastic about Type II errors.
[10] J. F. Geweke and K. J. Singleton, J. Am. Stat. Assoc. **75**, 133 (1980).
[11] Technically, this Gaussian ansatz is true only when ρ_0 is full rank—i.e., not on the boundary of the state set. If ρ_0 is rank deficient, then both the distribution of $\hat{\rho}_{\mathrm{MLE}}$ and $\mathcal{L}(\rho)$ itself are typically truncated by the boundary. However, the analysis remains valid (as $N \to \infty$) except if ρ_0 is simultaneously rank deficient *and* on the boundary between separable and entangled states.
[12] As long as the boundary of $\mathcal{S}$ is differentiable at ρ_0.
[13] R. F. Werner, Phys. Rev. A **40**, 4277 (1989).
[14] R. Renner, Nature Phys. **3**, 645 (2007).
[15] H. Häffner *et al.*, Nature (London) **438**, 643 (2005).
[16] J. M. Renes, R. Blume-Kohout, A. J. Scott, and C. M. Caves, J. Math. Phys. (N.Y.) **45**, 2171 (2004).