Efficient Bayesian credible-region certification for quantum-state tomography

Changhun Oh, Yong Siah Teo,[*] and Hyunseok Jeong

Department of Physics and Astronomy, Seoul National University, 08826 Seoul, Korea

Standard Bayesian credible-region theory for constructing an error region on the unique estimator of an unknown state in general quantum-state tomography to calculate its size and credibility relies on heavy Monte Carlo sampling of the state space followed by filtering to obtain the correct region sample. This conventional methodology typically gives negligible yield for very small error regions originating from large data sets. In this article, we discuss at length the in-region sampling theory for computing both size and credibility from region-average quantities that avoids this general problem altogether. Among the many possible numerical choices, we study the performance and properties of accelerated hit-and-run Monte Carlo algorithm for in-region sampling and provide its complexity estimates for quantum states. Finally with our in-region concept, by alternatively quantifying the region capacity with the region-average distance between two states in the region (measured for instance with either the Hilbert-Schmidt, trace-class, or Bures distance), we derive approximation formulas to analytically estimate both region capacity and credibility without Monte Carlo computation.

I. INTRODUCTION

All physical-quantity estimates obtained from collected data should be accompanied by "error bars" to accurately convey all properties of the physical system of interest. This applies to quantum-state tomography [1–5], which is an important preliminary step for implementing all quantum cryptography and computation protocols [6–8] reliably.

Bootstrapping procedures [9,10] are among some of the most widely used techniques for assigning error bars to reconstructed quantum states. Recently, it was pointed out in [11] that such assignments lack rigorous statistical foundations and may produce error bars that are too small for reliable conclusions. The rather more justified approach falls under the study of hypothesis testing [12]. Two grand schools of thought exist for this purpose. In the context of quantum-state reconstruction of an unknown state ρ, one may treat ρ as "absolute" (the frequentist school) and attempt to extract this knowledge from collected data. This suggests the constructions of *confidence regions* [13–15], which are error regions for the state estimator $\widehat{\rho}$ from all plausible data sets, including those unmeasured in the experiment. An accurate $\widehat{\rho}$ for ρ would then entail a collection of typically small confidence regions with high probability that ρ for each plausible data set lies in the corresponding region.

Given that only one data set (the measured one) is really available to the observer, we shall focus on the apt Bayesian school of thought that instead regards this data set as "absolute" and constructs *credible regions* [16,17] as the error regions for $\widehat{\rho}$ beginning with some prior distribution $p(\rho)$ of ρ. A fairly accurate estimator $\widehat{\rho}$ for some unknown ρ naturally implies a credible region (generated from the measured data set) of a small size with a large probability that this ρ is inside the region—a high credibility [18]. In order to obtain a reasonably small error region (be it that of credible or confidence type), one may either resort to adaptive strategies [19] and optimize additional properties of the region, or simply increase the data set collected in quantum tomography.

The complicated quantum-state–space boundary [20,21] renders any analytical attempt at calculating size and credibility for any credible region futile, leaving numerical Monte Carlo (MC) methods [22,23] as the only viable option. As the size of the credible region is defined as its volume fraction with the quantum-state space, one needs an extremely large sample of the state space to finally end up having a reasonable sample for the region. Despite the optimistic advantages that some of these MC schemes may have in generating samples of arbitrary distributions, one deleterious issue for such an MC-filtering strategy becomes apparent when the data set is large, which is the common situation in any tomography experiment. The resulting credible region eventually becomes too tiny relative to the quantum-state space for MC filtering to produce any effective yield to properly compute the size and credibility.

To solve this problem, we introduced the in-region sampling technique [24] to feasibly compute all credible-region properties by simply sampling an appropriate quantity over the region itself. This follows from the logic that a change in the region-average quantity encodes the change in both region size and credibility. In our theory, we prove the central lemma stating that the size (and credibility) of any credible region are related to a class of region-average quantities through a first-order differential equation that is solvable numerically. As an example study, we discuss the computation of region-average quantities using the accelerated hit-and-run algorithm, its correlation properties, and estimate its complexity from geometrical considerations of the quantum-state space. The region-average formalism encourages the formulation of the region capacity (a different way of stating "how big" a region is) by

---

[*]ys_teo@snu.ac.kr

investigating the average relative distance between two points in the region. This region-average distance may be induced by any of the common measures used in quantum information, and we shall explicitly consider the Hilbert-Schmidt, trace-class, and Bures distance measures as popular examples. It turns out that this perspective, together with the results in [18], offers closed-form analytical approximation formulas for an alternative rapid approximate Bayesian error certification with no Monte Carlo methods necessary.

This article is organized as follows: A preliminary introduction to the basic notions of standard Bayesian credible-region theory shall ensue in Sec. II, and the stage for discussions with large data is set in Sec. III. Next, we present our in-region sampling theory for size and credibility that works for any kind of data and prior in Sec. IV. Afterwards, we describe how region-average quantities can be numerically computed and estimate the computational complexities in Sec. V. Section VI then proceeds to quantify the region capacity in terms of region-average distances induced by all the three aforementioned distance measures other than the Bayesian size. Finally, for fast analytical Bayesian error estimates, we derive asymptotic formulas for all important region-average quantities in Sec. VII based on the perspective of distance-induced size. All numerical results and computation correlation properties are then presented and discussed in Sec. VIII with explicit examples in quantum tomography.

II. STANDARD BAYESIAN CREDIBLE-REGION THEORY

Before a quantum-state tomography experiment commences, the observer might have some (justifiable) preconception about the unknown quantum state $\rho \geqslant 0$ ($\mathrm{tr}\{\rho\} = 1$) of Hilbert-space dimension D. Such preconception is usually not uniquely privileged, and therefore weighted with some *prior probability distribution* $p(\rho)$. After the experiment, the observer collects a set of data $\mathbb{D}$ that are *informationally complete* (IC) such that a unique estimator $\widehat{\rho}$ for ρ is acquired. In quantum theory, the measurements are modeled as a *probability-operator measurement* (POM) consisting of a set of M positive operators $\Pi_j \geqslant 0$ that sum to the identity. Associated to every such experiment is the *likelihood function* $L = L(\mathbb{D}|\widehat{\rho})$, with which the observer obtains a *posterior probability distribution* (knowledge after the fact) that is a function of L.

It was formerly established in [16] that for this measured data set $\mathbb{D}$, if $\widehat{\rho}$ is taken to be the estimator that maximizes L—the maximum-likelihood (ML) estimator—then a Bayesian credible region (CR) $\mathscr{R}$ can be constructed around $\widehat{\rho}_{\mathrm{ML}}$, which turns out to have a constant likelihood boundary $\partial\mathscr{R}$ within the quantum-state space $\mathscr{R}_0$. For this CR, which is a subregion of $\mathscr{R}_0$, we can specify its *size* and *credibility*, the latter of which is the probability that $\rho \in \mathscr{R}$. Such a region is optimal in the sense that it gives the largest credibility for a given size, or equivalently possesses the smallest size for a given credibility.

In this article, we shall be interested in reconstructing the ($d = D^2 - 1$)-dimensional real vectorial parameter $\boldsymbol{r} \leftrightarrow \rho$ that characterizes ρ. More technically, this equivalent parametrization is achieved with a Hermitian operator basis

$\{1/\sqrt{D}, \Omega_j\}_{j=1}^d$ that contains d trace-orthonormal traceless operators Ω_j ($\mathrm{tr}\{\Omega_j\Omega_k\} = \delta_{j,k}$), by which $\boldsymbol{r} = \mathrm{tr}\{\rho\,\boldsymbol{\Omega}\}$ is defined from the column $\boldsymbol{\Omega}$ of Ω_js. Formally, in terms of the multivariate parameter $\boldsymbol{r}$, the size and credibility of $\mathscr{R} = \mathscr{R}_\lambda$ for some $0 \leqslant \lambda \leqslant 1$ are respectively given by [16]

$$S_\lambda \equiv \int_{\mathscr{R}_\lambda} (d\,\boldsymbol{r}') = \int_{\mathscr{R}_0} (d\,\boldsymbol{r}')\,\eta(L - \lambda L_{\max}),$$

$$C_\lambda \equiv \frac{1}{L(\mathbb{D})}\int_{\mathscr{R}_\lambda}(d\,\boldsymbol{r}')\,L = \frac{1}{L(\mathbb{D})}\int_{\mathscr{R}_0}(d\,\boldsymbol{r}')\,\eta(L - \lambda L_{\max})\,L,$$

$$(1)$$

where the volume measure $(d\,\boldsymbol{r})$ incorporates the prior distribution $p(\boldsymbol{r})$ for $\boldsymbol{r}$, η is the Heaviside function, $L(\mathbb{D}) = \int_{\mathscr{R}_0}(d\,\boldsymbol{r}')\,L(\mathbb{D}|\boldsymbol{r}')$. The important variable $0 \leqslant \lambda \leqslant 1$ specifies the shape and size of $\mathscr{R}_\lambda$, from which the limits $\mathscr{R}_{\lambda=0} = \mathscr{R}_0$ and $\mathscr{R}_{\lambda=1} = \{\widehat{\boldsymbol{r}}_{\mathrm{ML}}\}$ are immediate. Here, we note that the probability parametrization was adopted in [16]. Under the condition that each datum is collected independently, the inherent statistics of $\mathbb{D}$ is multinomial and the log likelihood reads $\ln L = \sum_j n_j \ln p'_j$ with the collected relative frequencies $\sum_j n_j = N$ that make up N measured data copies, and $p'_j = \mathrm{tr}\{\rho'\Pi_j\}$ for any state ρ'.

We can gain a clear physical picture of both size and credibility: they respectively quantify the prior and posterior content $\mathscr{R}$, hence the symbol S_λ for the former. Owing to the dual nature of size S_λ and credibility C_λ, it can be shown, in fact, that

$$C_\lambda = \frac{\lambda S_\lambda + \int_\lambda^1 d\,\lambda' S_{\lambda'}}{\int_0^1 d\,\lambda' S_{\lambda'}}. \tag{2}$$

Put differently, C_λ may be straightforwardly computable through single-parameter integrations in λ so long as S_λ is known up to some arbitrary constant multiple.

Nonetheless, the complicated boundary $\partial\mathscr{R}_0$ of the quantum-state space makes the computation of S_λ extremely challenging even numerically. The innate definition of S_λ, namely the volume fraction of $\mathscr{R}_\lambda$ to $\mathscr{R}_0$, requires, first, the generation of a sufficiently large sample of $\mathscr{R}_0$, followed by the filtering of all its sampled points that lie inside $\mathscr{R}_\lambda$ for any λ. There exist various Monte Carlo (MC) methods to sample $\mathscr{R}_0$ [22]. Ultimately, this MC-filtering strategy exhibits one major disadvantage: in the limit of large data sample ($N \gg 1$), $\mathscr{R}_\lambda$ would become so small relative to $\mathscr{R}_0$ that the MC-filtering strategy needs a sufficiently large number of random MC sample points from $\mathscr{R}_0$ to produce any useful yield. The scaling of MC sample size needed to maintain a fixed yield, which was estimated to be $O(N^{d/2})$ [18], thus outgrows the feasible computational yield rate very quickly. The bottom line: a much more feasible numerical strategy to perform Bayesian error certification is necessary in this practical data limit.

III. LARGE-DATA CONDITION

Before presenting an alternative operational theory, unless otherwise stated, we shall consider $N \gg 1$ as the putative limit in pragmatic tomography experiments. We emphasize here that N only has to be sufficiently large for the statistical central
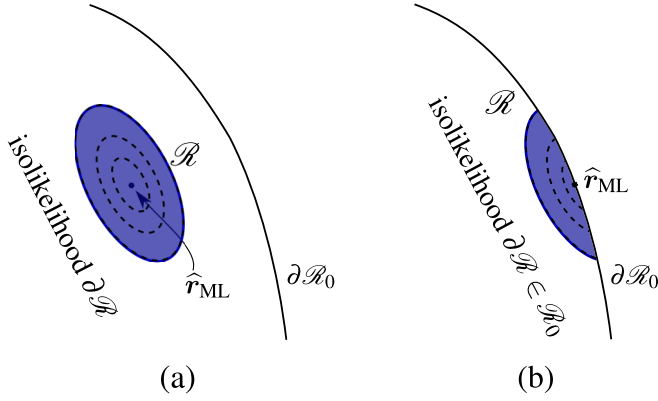
FIG. 1. Credible region $\mathscr{R}$ in (a) case A and (b) case B, consolidating the two very general situations that could happen in the limit $N \gg 1$ [24]. Case A corresponds to a state estimator $\widehat{\rho}_{\mathrm{ML}} \leftrightarrow \widehat{\boldsymbol{r}}_{\mathrm{ML}}$ that is full rank, whereas case B implies that the estimator is rank deficient.

limit theorem to dictate a Gaussian form for L. In this limit, there can only be one of two cases: either $\mathscr{R}$ is completely inside $\mathscr{R}_0$ that contains a full-rank $\widehat{\rho}_{\mathrm{ML}}$ (case A) or partially truncated by the state-space boundary $\partial \mathscr{R}_0$ of $\mathscr{R}_0$ that houses a rank-deficient $\widehat{\rho}_{\mathrm{ML}}$ (case B) (see Fig. 1).

Case A arises when the unknown state ρ is away from $\partial \mathscr{R}_0$, so that a sufficiently large N would produce untruncated regions for λ values corresponding to desirably large $C_\lambda < 1$. This case offers a simple geometrical description for $\mathscr{R}$. Upon invoking the Taylor expansion of

$$\ln L(\mathbb{D}|\boldsymbol{r}') \approx \ln L_{\max} - \frac{1}{2}(\boldsymbol{r}' - \widehat{\boldsymbol{r}}_{\mathrm{ML}}) \cdot \boldsymbol{F}_{\mathrm{ML}} \cdot (\boldsymbol{r}' - \widehat{\boldsymbol{r}}_{\mathrm{ML}}) \quad (3)$$

about the interior $\widehat{\boldsymbol{r}}_{\mathrm{ML}}$ up to the second order, we find that the likelihood L is essentially a Gaussian function centered at $\widehat{\boldsymbol{r}}_{\mathrm{ML}}$ of height $L_{\max}$, with its covariance profile shaped by $\boldsymbol{F}_{\mathrm{ML}}$, that is the *Fisher information* evaluated at $\widehat{\boldsymbol{r}}_{\mathrm{ML}}$. The CRs $\mathscr{R}_\lambda$ that go with this Gaussian likelihood are, hence, simple hyperellipsoids $\mathscr{E}_\lambda$ described by the inequality $(\boldsymbol{r}' - \widehat{\boldsymbol{r}}_{\mathrm{ML}}) \cdot \boldsymbol{F}_{\mathrm{ML}} \cdot (\boldsymbol{r}' - \widehat{\boldsymbol{r}}_{\mathrm{ML}}) \leqslant -2\ln\lambda$.

If $\boldsymbol{r}$ is located in $\partial \mathscr{R}_0$, then as N increases, the ML estimator $\widehat{\boldsymbol{r}}_{\mathrm{ML}}$ would eventually approach $\boldsymbol{r}$ and there is a high probability that $\widehat{\boldsymbol{r}}_{\mathrm{ML}} \in \partial \mathscr{R}_0$ before this happens. For sufficiently large N, we have case B where $\partial \mathscr{R} \cap \partial \mathscr{R}_0$ is not disjointed and falls on the side of $\widehat{\boldsymbol{r}}_{\mathrm{ML}}$. To asymptotically cope with such a situation, we may again expand

$$
\begin{aligned}
\ln L(\mathbb{D}|\boldsymbol{r}') &\approx \ln L_{\max} + (\boldsymbol{r}' - \widehat{\boldsymbol{r}}_{\mathrm{ML}}) \cdot \boldsymbol{g}_{\mathrm{ML}} \\
&\quad - \frac{1}{2}(\boldsymbol{r}' - \widehat{\boldsymbol{r}}_{\mathrm{ML}}) \cdot \boldsymbol{F}_{\mathrm{ML}} \cdot (\boldsymbol{r}' - \widehat{\boldsymbol{r}}_{\mathrm{ML}}) \\
&= \ln L'_{\max} - \frac{1}{2}(\boldsymbol{r}' - \boldsymbol{r}_{\mathrm{c}}) \cdot \boldsymbol{F}_{\mathrm{ML}} \cdot (\boldsymbol{r}' - \boldsymbol{r}_{\mathrm{c}}) \quad (4)
\end{aligned}
$$

about the boundary $\widehat{\boldsymbol{r}}_{\mathrm{ML}}$, where this time L is a Gaussian function centered at $\boldsymbol{r}_{\mathrm{c}} = \widehat{\boldsymbol{r}}_{\mathrm{ML}} + \boldsymbol{F}_{\mathrm{ML}}^{-1} \cdot \boldsymbol{g}_{\mathrm{ML}}$ with $\boldsymbol{g}_{\mathrm{ML}} = \partial \ln L(\mathbb{D}|\boldsymbol{r}')/\partial \boldsymbol{r}'|_{\boldsymbol{r}'=\widehat{\boldsymbol{r}}_{\mathrm{ML}}}$, and possesses a height $L'_{\max} = L_{\max} \exp(\boldsymbol{g}_{\mathrm{ML}} \cdot \boldsymbol{F}_{\mathrm{ML}}^{-1} \cdot \boldsymbol{g}_{\mathrm{ML}}/2) > L_{\max}$. The covariance profile of this Gaussian function is still governed by $\boldsymbol{F}_{\mathrm{ML}}$, which produces hyperellipsoids $\mathscr{E}'_\lambda$ described according to
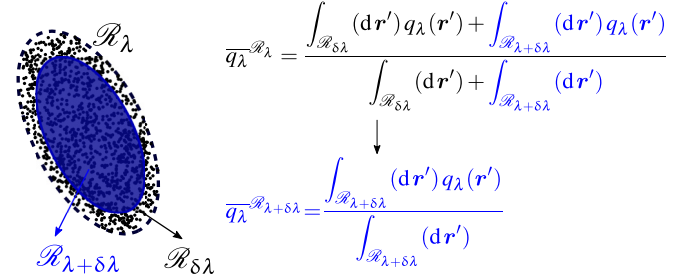

FIG. 2. An infinitesimal change in λ causes a transformation in $\mathscr{R}_\lambda$ that subsequently excludes all points in the hyperannulus $\mathscr{R}_{\delta\lambda}$ from the region average $\overline{q_\lambda}^{\mathscr{R}_{\lambda+\delta\lambda}}$.

$(\boldsymbol{r}' - \boldsymbol{r}_{\mathrm{c}}) \cdot \boldsymbol{F}_{\mathrm{ML}} \cdot (\boldsymbol{r}' - \boldsymbol{r}_{\mathrm{c}}) \leqslant -2\ln\lambda'$ for an "effective λ" λ' defined by $2\ln(\lambda/\lambda') = \boldsymbol{g}_{\mathrm{ML}}^{\mathrm{T}} \boldsymbol{F}_{\mathrm{ML}}^{-1} \boldsymbol{g}_{\mathrm{ML}}$. The CR $\mathscr{R}_\lambda$ is then asymptotically $\mathscr{E}'_\lambda \cap \mathscr{R}_0$.

We point out that there is an intermediate case in which $\mathscr{R}_\lambda = \mathscr{E}_\lambda$, centered at $\widehat{\boldsymbol{r}}_{\mathrm{ML}} \notin \partial \mathscr{R}_0$, is truncated by $\partial \mathscr{R}_0$. Such a situation can happen when N is not sufficiently large, and tends to either case A or case B as N grows. On a separate note, Refs. [18,19] explicitly study also this intermediate case.

IV. IN-REGION SAMPLING THEORY

Suppose we have a CR $\mathscr{R}_\lambda$, with which we define the average quantity

$$u_\lambda \equiv \overline{q_\lambda(\boldsymbol{r}')}^{\mathscr{R}_\lambda} = \frac{\int_{\mathscr{R}_\lambda}(d\boldsymbol{r}')\, q_\lambda(\boldsymbol{r}')}{\int_{\mathscr{R}_\lambda}(d\boldsymbol{r}')} = \frac{1}{K_{\mathrm{smp}}}\sum_{l=1}^{K_{\mathrm{smp}}} q_{\lambda,l} \quad (5)$$

for some function q_λ, which is approximately equivalent to the discrete-sum average of $q_{\lambda,l}$ values over a sufficiently large number K_{smp} of region points. If we probe the response of u_λ with an incremental change $\lambda \to \lambda + \delta\lambda$ in λ as in Fig. 2, the result is the total change

$$\delta u_\lambda = \left(\frac{1}{S_\lambda} - \frac{\delta S_\lambda}{S_\lambda^2}\right)\int_{\mathscr{R}_{\delta\lambda}}(d\boldsymbol{r}')\, q_\lambda - \frac{\delta S_\lambda}{S_\lambda^2}\int_{\mathscr{R}_\lambda}(d\boldsymbol{r}')\, q_\lambda \quad (6)$$

after limiting all small changes to the first order, which reveals that a small increment δu_λ can be explained by a change δS_λ in size that is accompanied by the (in)exclusion of the annular sum $\int_{\mathscr{R}_{\delta\lambda}}(d\boldsymbol{r}')\, q_\lambda$. Put simply, tracking the change in u_λ allows us to infer how much S_λ has changed.

To better utilize this intuition, we first take the derivative of $u_\lambda S_\lambda$, which gives

$$
\begin{aligned}
\frac{\partial u_\lambda S_\lambda}{\partial \lambda} &= -L_{\max}\int_{\mathscr{R}_0}(d\boldsymbol{r}')\,\delta(L - \lambda L_{\max})\, q_\lambda(\boldsymbol{r}') \\
&\quad + \int_{\mathscr{R}_\lambda}(d\boldsymbol{r}')\,\frac{\partial q_\lambda(\boldsymbol{r}')}{\partial \lambda} \quad (7)
\end{aligned}
$$

after invoking the derivative identity $d\eta(x)/dx = \delta(x)$ between $\eta(x)$ and the Dirac delta function $\delta(x)$. Next, we may impose the following functional form: $q_\lambda(\boldsymbol{r}') \equiv f(L) - f(\lambda L_{\max})$ for q_λ, where $f(L)$ is some arbitrary function of L. This simplifies Eq. (7) to

$$\frac{\partial}{\partial \lambda}(u_\lambda S_\lambda) = -S_\lambda \frac{\partial}{\partial \lambda} f(\lambda L_{\max}). \quad (8)$$

We now have a first-order differential equation that describes the dynamics of S_λ according to the parametric region average u_λ. With the initial condition $S_{\lambda=0} = 1$, the entire functional form of S_λ can then be recovered with Eq. (8). This completes the constructive proof of our so-called *Region-average computation (RAC) lemma*: For any prior $(d\,\mathbf{r}')$ and measurement data $\mathbb{D}$, the prior content S_λ (up to a multiplicative factor) and hence the credibility C_λ are all inferable from u_λ defined in Eq. (5) with $q_\lambda(\mathbf{r}') \equiv f(L) - f(\lambda L_{\max})$.

To proceed, we first perform the substitution $y_\lambda = u_\lambda S_\lambda$ to yield another differential equation

$$\frac{\partial y_\lambda}{\partial \lambda} = -\frac{y_\lambda}{u_\lambda} \frac{\partial}{\partial \lambda} f(\lambda L_{\max}). \tag{9}$$

The solution to y_λ can then be obtained numerically through Euler's method [25]. In practice, we may start from $S_{\lambda \approx 0} \equiv 1$ and iterate

$$y_{\lambda_{j+1}} = y_{\lambda_j} - \frac{y_{\lambda_j}}{u_{\lambda_j}} \frac{\partial}{\partial \lambda_j} f(\lambda_j L_{\max}) \tag{10}$$

for a sequence of discretized $\lambda \to \lambda_j$ values ranging from 0 to 1. For feasible computation of u_λ, we shall choose $f(L) = \ln L$.

V. REGION-AVERAGE NUMERICAL COMPUTATION

A. Hit-and-run algorithm

The hit-and-run algorithm is a direct convex-body MC sampling scheme that generates random sample points in the body according to some predefined distribution. This algorithm is thus suited for sampling $\mathscr{R}_\lambda$ according to some prior distribution $p(\mathbf{r})$ for the unknown $\mathbf{r}$.

The sampling principles behind an efficient hit-and-run computation begin with defining the smallest possible convex set $\mathscr{B} \supseteq \mathscr{R}_\lambda$ that houses $\mathscr{R}_\lambda$ and has an easy-access geometry. Starting from a known point in $\mathscr{R}_\lambda$, say the ML estimator $\widehat{\mathbf{r}}_{\mathrm{ML}}$, a random line segment passing through this point is generated, with its end points fixed at $\partial \mathscr{B}$ that are quickly computable because of its simple geometry. Following which, sampling commences by repeatedly picking a random point along this segment until it lies in $\mathscr{R}_\lambda$. This point is next taken to be the new reference point through which another line segment is generated to find a new random point in $\mathscr{R}_\lambda$, until a set of K_{smp} points is gathered.

We can make use of the straightforward hyperellipsoidal characteristics inherent from the central limit theorem to construct $\mathscr{B}$. For case A, where $\mathscr{R}_\lambda = \mathscr{E}_\lambda$, $\mathscr{B}$ can just be taken to be $\mathscr{E}_\lambda$ characterized by $\widehat{\mathbf{F}}_{\mathrm{ML}} = \mathbf{F}_{\mathrm{ML}}/(-2\ln\lambda)$ from the earlier discussions in Sec. III. We now turn to the more interesting and practically ubiquitous case B, where the large-N arguments of Sec. III imply that we may fix $\mathscr{B} = \mathscr{E}'_\lambda$, the profile of which is governed by $\widehat{\mathbf{F}'}_{\mathrm{ML}} = \mathbf{F}_{\mathrm{ML}}/(-2\ln\lambda')$. For this case, if $\mathscr{B}$ is much larger than $\mathscr{R}_\lambda$, sampling the latter would incur a significant amount of wastage. Fortunately there exists an accelerated version of hit and run [26] that adaptively shrinks the end points of the line segment to reduce the search space each step.

To check if the random point chosen from the line segment is in $\mathscr{R}_\lambda$, we recall that $\mathscr{R}_\lambda$ is equivalently the continuous set of unit-trace operators that are both positive and sat-

isfy the hyperellipsoidal constraint defined by the inequality $L(\mathbb{D}|\mathbf{r}') > \lambda L_{\max}$ under the $N \gg 1$ limit. Since all points in $\mathscr{B}$ essentially fulfill the latter constraint, the primary task is to check if the random point corresponds to a legitimate quantum state. It is known that the Cholesky decomposition [27,28], a routine that factorizes $\rho' = A^\dagger A$ for any positive operator ρ', is an efficient and numerically stable way for the job, with a computational complexity of $O(D^3)$. A routine failure implies that the operator corresponding to the selected point is not positive.

The complete pseudocoded algorithm for the accelerated hit and run, tailored for an arbitrary prior distribution $p(\mathbf{r})$, is stated below [26,29,30]:

Accelerated hit and run for sampling $\mathscr{R}_\lambda$

Beginning with $k = 1$ and $\mathbf{r}_{\mathrm{ref}} = \widehat{\mathbf{r}}_{\mathrm{ML}}$ of $N \gg 1$:

1. Generate a random line segment characterized by $\mathbf{y} = \mathbf{r}_{\mathrm{ref}} + \mu\,\mathbf{e_v}$, where $\mathbf{e_v} = \mathbf{v}/|\mathbf{v}|$ and $\mathbf{v}$ follows the standard Gaussian distribution (mean 0 and variance 1 for each column entry). Its end points are parametrized by $\mu_\pm = [-b \pm \sqrt{b^2 - a(c-1)}]/a$, where $\boldsymbol{\Delta} = \mathbf{r}_{\mathrm{ref}} - \mathbf{r}_{\mathrm{c}}$, $a = \mathbf{e_v}^{\mathsf{T}} A\,\mathbf{e_v}$, $b = \boldsymbol{\Delta}^{\mathsf{T}} A\,\mathbf{e_v}$, $c = \boldsymbol{\Delta}^{\mathsf{T}} A\,\boldsymbol{\Delta}$, and $A = \widehat{\mathbf{F}}_{\mathrm{ML}}$ or $\widehat{\mathbf{F}'}_{\mathrm{ML}}$.
2. Define $\beta_1 \equiv \mu_{\min} = \min\{\mu_+, \mu_-\}$ and $\beta_2 \equiv \mu_{\max} = \max\{\mu_+, \mu_-\}$.
3. Pick a random number $\beta_1 \leqslant \beta \leqslant \beta_2$ according to the marginal probability distribution $p(\mathbf{r}_{\mathrm{ref}} + \beta\,\mathbf{e_v})/\int d\beta' p(\mathbf{r}_{\mathrm{ref}} + \beta'\,\mathbf{e_v})$ truncated in the interval $[\beta_1, \beta_2]$ and obtain $\mathbf{r}_{\mathrm{test}} = \mathbf{r}_{\mathrm{ref}} + \beta\,\mathbf{e_v}$.
4. Check whether $\rho_{\mathrm{test}} \leftrightarrow \mathbf{r}_{\mathrm{test}}$ is positive.
   - If so, define $\mathbf{r}_{\mathrm{ref}} = \mathbf{r}_{\mathrm{test}}$, raise k by 1, and go to step 1.
   - If not, set $\beta_1 = \beta$ if $\beta < 0$, or $\beta_2 = \beta$ if $\beta > 0$, and repeat steps 3 and 4.
5. End routine if $k > K_{\mathrm{smp}}$, the total number of sample points desired.

To further speed up the algorithm for case B, one can assign $\mathscr{B}$ to be the hyperellipsoidal cap composed by a hyperplane that is tangent to the iso-Gaussian level curve of $\mathscr{E}'_\lambda$ at $\widehat{\mathbf{r}}_{\mathrm{ML}}$ and the part of $\mathscr{E}'_\lambda$ below it (refer to Sec. VII C). Numerical experience shows that this speedup is negligible in the presence of the end-point adaptation mechanism of accelerated hit and run.

We end this introduction of hit and run by noting that a hyperellipsoidal $\mathscr{B}$ is constructed based on the large-N limit, where the boundary $\partial \mathscr{R}_\lambda \cap \mathrm{int}\{\mathscr{R}_0\}$ of the physical region is well approximated by this hyperellipsoid. The highly skeptical may insist that, perhaps, for a finite N, even if $N \gg 1$, there might still be cases where a part of this boundary protrudes $\mathscr{B}$. To be on the safer side, one may choose a hyperellipsoidal $\mathscr{B}$ of a reasonably larger size (say doubled) than the one given by the central limit theorem. This will almost surely contain the physical error region with a much smaller failure probability. The pertinent question is can we verify that $\mathscr{B}$ contains $\mathscr{R}$ with arbitrary precision? The answer unfortunately is negative both in theory and practice. This is because a positive answer would entail a complete knowledge about $\partial \mathscr{R}$, obtaining which is either computationally not feasible in general, or an NP-hard problem in some context [11].
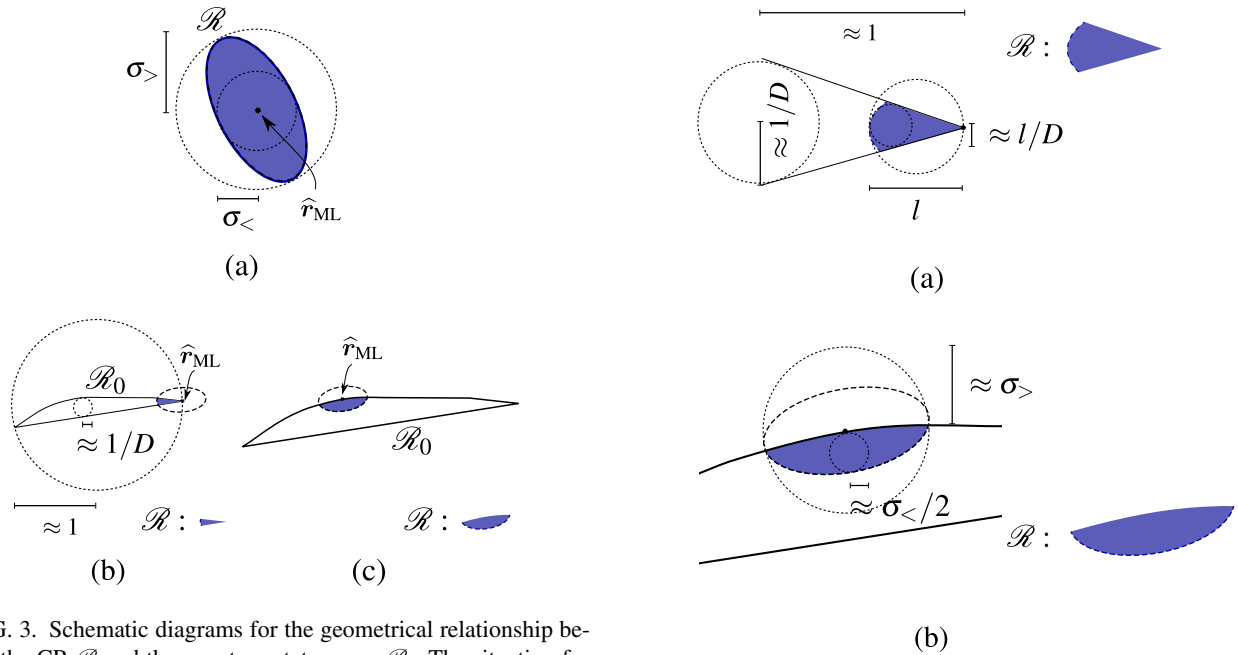
(a)

(b)   (c)

FIG. 3. Schematic diagrams for the geometrical relationship between the CR $\mathscr{R}$ and the quantum-state space $\mathscr{R}_0$. The situation for (a) case A is completely known and so complexity estimation for hit and run is a simple matter. To acquire conservative complexity estimates for case B, two special types of such CR may exist: either the CR (b) lies on an extremely sharp corner of $\mathscr{R}_0$ in at least one of its dimensions (type I) in whichever orientation, or (c) on one of its edges that is almost flat (type II) in all its dimensions, with the longest $\mathscr{E}_\lambda'$ axis oriented along the flat surface.

B. Numerical complexity estimations

After suppressing dependencies on logarithmic factors and error parameters, it was argued that the number of hit-and-run steps needed to gather enough sample points and form an ensemble described by $p(\boldsymbol{r})$ in hit and run is $O(d^2 R_{\text{out}}^2/R_{\text{in}}^2) = O(D^4 R_{\text{out}}^2/R_{\text{in}}^2)$ [31,32] in the limit $D \gg 2$, where R_{out} is the radius of the smallest outer sphere that contains $\mathscr{R}_\lambda$ and R_{in} is that of the largest inner sphere that can be inscribed in $\mathscr{R}_\lambda$. Together with the floating-point-operations complexity $O(D^3)$ in a typical Cholesky decomposition algorithm [28], we have an estimate for the complexity $\text{cmpl} = O(D^7 R_{\text{out}}^2/R_{\text{in}}^2)$ for the entire hit-and-run scheme.

The treatment of case A is straightforward as we have the complete information about $\mathscr{R}_\lambda \approx \mathscr{E}_\lambda$ in the large-N limit. If we denote $\sigma_>$ and $\sigma_<$ to respectively be the largest and smallest eigenvalue of $\widetilde{\boldsymbol{F}}_{\text{ML}}^{-1/2}$, then the corresponding outer and inner radii are $R_{\text{out}} = \sigma_>$ and $R_{\text{in}} = \sigma_<$ [see Fig. 3(a)], so that $\text{cmpl}_A = O[D^7 \text{cond}(\boldsymbol{F}_{\text{ML}}^{-1})]$ involves the conditional number $\text{cond}(\boldsymbol{F}_{\text{ML}}^{-1}) = \sigma_>^2/\sigma_<^2$.

The analysis for case B requires extra care given the complicated state-space boundary $\partial\mathscr{R}_0$. While complete and precise details of $\mathscr{R}_0$ are absent so far, from [21], we know that in the Euclidean space, the largest inner sphere inscribable in $\mathscr{R}_0$ has a radius that approaches $1/D$ for $D \gg 2$, and that the smallest outer sphere that contains $\mathscr{R}_0$ has a radius going to 1 in the same dimension limit. The overall shape of $\mathscr{R}_0$ is therefore a "squashed" convex body for large D, such that *at least one* of its dimensions drops appreciably to zero. To estimate the complexity for case B, we consider CRs of two
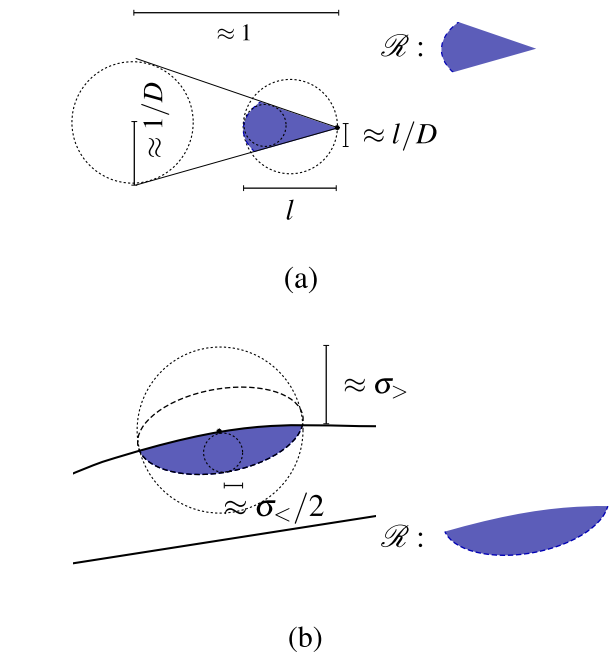

(a)

(b)

FIG. 4. Schematic diagrams for (a) type-I and (b) type-II Bayesian regions. Type-I regions have complexities that are strongly influenced by the cornered geometry (greatly exaggerated for visual aid), whereas type-II regions have complexities that strongly depend on the eigenvalue aspect ratios of $\boldsymbol{F}_{\text{ML}}$. All other intermediate CR types give rise to complexities affected by the geometries of both $\partial\mathscr{R}_0$ and $\boldsymbol{F}_{\text{ML}}$.

tractable types: a type-I CR is located at an extremely sharp corner of $\mathscr{R}_0$ that is made from at least one of its rapidly shrinking dimensions, as shown in Fig. 3(b), whereas a type-II CR is situated at an extremely flat boundary of $\mathscr{R}_0$ where all of its dimensions remain approximately constant within the CR as in Fig. 3(c). For a conservative estimate of cmpl, we consider an $\mathscr{R}$ such that the longest axis of $\mathscr{E}_\lambda'$ is aligned with the flat surface. All other types of case-B CRs may be viewed as intermediate situations of these two and have no analytical complexity estimates known to us. The data-copy number $N \gg 1$ is assumed to be sufficiently large such that $\boldsymbol{g}_{\text{ML}} \approx \boldsymbol{0}$ and $\boldsymbol{r}_{\text{c}} \approx \widehat{\boldsymbol{r}}_{\text{ML}}$.

To estimate cmpl for a type-I CR, we assume that the corner is extremely sharp in one particular dimension such that the curvature of $\partial\mathscr{R}_0$ extending out from $\widehat{\boldsymbol{r}}_{\text{ML}}$ is almost flat. Then following Fig. 4(a), the concept of similar figures give $R_{\text{out}}/R_{\text{in}} \approx D$, which is independent of $\boldsymbol{F}_{\text{ML}}$ for extremely sharp corners, and $\text{cmpl}_{B,I} = O(D^9)$. The complexity for type-II CRs may be estimated with the help of Fig. 4(b), where $R_{\text{out}}/R_{\text{in}} \approx 2\,\text{cond}(\boldsymbol{F}_{\text{ML}}^{-1/2})$ is now independent of $\partial\mathscr{R}_0$ due to its extremely mild edge features, leading us to $\text{cmpl}_{B,II} = O[D^7 \text{cond}(\boldsymbol{F}_{\text{ML}}^{-1})] = \text{cmpl}_A$.

C. Other numerical methods

Other numerical methods apart from hit and run may also be used to perform in-region sampling, each of which has its own merits and shortcomings [22,33–35]. Classical rejection and importance sampling methods are two straightforward

ways to acquire samples distributed according to some desired prior distribution. For large D and N, these methods rapidly become infeasible due to the decreasing ratio of the CR volume to the full Hermitian sampling volume that includes many more unphysical operators that are not quantum states. To cope with this low-yield difficulty, another Markov-chain method known as the Metropolis-Hastings MC scheme may also be used to do in-region sampling. Such a scheme also suffers from high correlations that are generally dependent on the starting point of a Markov-chain iteration. Hamiltonian MC methods are yet another promising class of algorithms that permit larger sample-point hopping that gives a sample with weak correlations. The scalability of such methods are, however, still a work in progress.

VI. DISTANCE-INDUCED REGION CAPACITY

A. Operational definition

The theory of in-region sampling seamlessly paves the way to other creative ways of defining the capacity of a region. Doing so permits us to talk about "how big" a CR is without referring to $\mathscr{R}_0$ entirely. To begin, one could measure the region capacity in terms of the average distance between any two points inside $\mathscr{R}$, which is a separate idea from the prior content. Intuitively, the smaller this average distance, the smaller the region and vice versa. Using this simple prescription, we propose the region-average quantity

$$S_{\mathscr{D},\lambda} \equiv \overline{\mathscr{D}(\boldsymbol{r}', \widehat{\boldsymbol{r}}_{\mathrm{ML}})}^{\mathscr{R}_\lambda} = \frac{\int_{\mathscr{R}_\lambda} (d\,\boldsymbol{r}')\,\mathscr{D}(\boldsymbol{r}', \widehat{\boldsymbol{r}}_{\mathrm{ML}})}{\int_{\mathscr{R}_\lambda} (d\,\boldsymbol{r}')} \quad (11)$$

to measure the capacity of $\mathscr{R}_\lambda$, where $\mathscr{D}(\boldsymbol{r}', \widehat{\boldsymbol{r}}_{\mathrm{ML}})$ is some prechosen distance metric. Notice that the ML estimator $\widehat{\boldsymbol{r}}_{\mathrm{ML}}$ is selected to be the reference point from which distances are measured without loss of generality.

To concretize all results, we shall look at three distance measures for states that enjoy a good reputation in quantum-information studies. We first mention the Hilbert-Schmidt (HS) distance

$$\mathscr{D}_{\mathrm{HS}} = \mathrm{tr}\{(\rho' - \widehat{\rho}_{\mathrm{ML}})^2\} = (\boldsymbol{r}' - \widehat{\boldsymbol{r}}_{\mathrm{ML}})^2, \quad (12)$$

which is equivalent to the squared l_2 norm of $\boldsymbol{r}' - \widehat{\boldsymbol{r}}_{\mathrm{ML}}$. Closely related to the HS distance is the trace-class distance

$$\mathscr{D}_{\mathrm{tr}} = \mathrm{tr}\{\sqrt{(\rho' - \widehat{\rho}_{\mathrm{ML}})^2}\} \quad (13)$$

defined by the operator absolute value $|A| = \sqrt{A^\dagger A}$. To introduce the third measure, we start by quoting the expression of quantum fidelity [36],

$$\mathscr{F} = \mathrm{tr}\{\sqrt{\sqrt{\widehat{\rho}_{\mathrm{ML}}}\rho'\sqrt{\widehat{\rho}_{\mathrm{ML}}}}\}^2, \quad (14)$$

between ρ' and $\widehat{\rho}_{\mathrm{ML}}$, to which we can define the Bures distance [37,38] $\mathscr{D}_{\mathrm{B}} = 2(1 - \sqrt{\mathscr{F}})$. In the limit of large N, where the fidelity $\mathscr{F} \approx 1 - \epsilon$ differs from 1 by a small amount, $\mathscr{D}_{\mathrm{B}}$ is also approximately the infidelity $1 - \mathscr{F}$.

B. Monotonic behavior of $S_{\mathscr{D},\lambda}$ for $N \gg 1$

Here, we show that, at least for sufficiently large N, $S_{\mathscr{D},\lambda}$, defined by any of these three distance measures, behave cor-
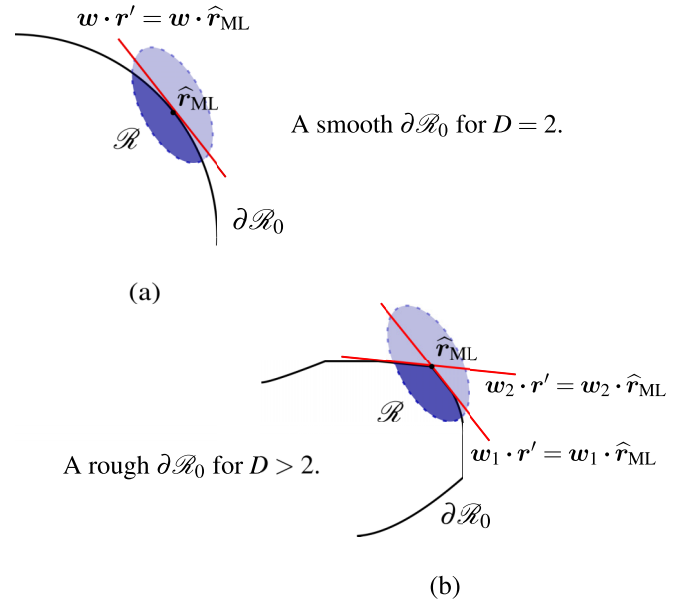

FIG. 5. The different types of state-space boundaries $\partial\mathscr{R}_0$ [24]. Excluding the exceptional single-qubit system that exhibits a smooth spherical surface (a), all higher-dimensional systems result in $\partial\mathscr{R}_0$ that is not smooth, with corners and edges. In the large-N limit, an ML estimator at the corner, for instance, may be well approximated by a collection of hyperplanes relative to $\mathscr{R}$ since every point in $\partial\mathscr{R}_0$ is a well-defined quantum state.

rectly as a capacity function in the sense that $S_{\mathscr{D},\lambda}$ should not increase as λ increases. We first look at the more complicated case B, and argue that since the $\mathscr{R}_0$ generally has only corners and edges with no other mathematically pathological features, a set of hyperplanes can then be used to model any particular boundary feature on which $\widehat{\boldsymbol{r}}_{\mathrm{ML}}$ resides (see Fig. 5). This results in the asymptotic form

$$S_{\mathscr{D},\lambda} \to \frac{\int (d\,\boldsymbol{r}'')\,\mathscr{D}\,\eta(1 - \boldsymbol{r}''^{\mathrm{T}}\boldsymbol{F}_{\mathrm{ML}}\boldsymbol{r}''/(-2\ln\lambda))\prod_j \eta(\boldsymbol{w}_j^{\mathrm{T}}\boldsymbol{r}'')}{\int (d\,\boldsymbol{r}'')\,\eta(1 - \boldsymbol{r}''^{\mathrm{T}}\boldsymbol{F}_{\mathrm{ML}}\boldsymbol{r}''/(-2\ln\lambda))\prod_j \eta(\boldsymbol{w}_j^{\mathrm{T}}\boldsymbol{r}'')}$$

$$(15)$$

after the substitution $\boldsymbol{r}'' = \boldsymbol{r}' - \widehat{\boldsymbol{r}}_{\mathrm{ML}}$.

At this stage, we shall consider the asymptotic expressions of the distance measures. The HS distance $\mathscr{D}_{\mathrm{HS}}$ takes on the simplest (quadratic) form out of all three, which very straightforwardly gives the asymptotic dependence $S_{\mathrm{HS},\lambda} \to -\ln\lambda$ provided the sufficient condition $(d\,\alpha\,\boldsymbol{r}') = g(\alpha)(d\,\boldsymbol{r}')$, which includes the *uniform primitive prior* $(d\,\boldsymbol{r}') = (d\,\boldsymbol{r}')_{\mathrm{unif}} \equiv \prod_j d\,r_j$. It is not difficult to see that the same λ dependence applies to case A by taking $\boldsymbol{w}_j = \boldsymbol{0}$, so that $S_{\mathrm{HS},\lambda}$ is monotonically decreasing with increasing λ. Next, according to Appendix A, in the limit of large D, $S_{\mathrm{tr}} \sim \sqrt{S_{\mathrm{HS}}}$, which is also clearly monotonic as well owing to S_{HS}'s monotonicity. For S_{B} and $S_{\mathscr{F}}$, one can perform a Taylor expansion on them about $\widehat{\boldsymbol{r}}_{\mathrm{ML}}$ (see Sec. VII) and realizes that both functions asymptotically depend on the dyadic $(\boldsymbol{r}' - \widehat{\boldsymbol{r}}_{\mathrm{ML}})(\boldsymbol{r}' - \widehat{\boldsymbol{r}}_{\mathrm{ML}})^{\mathrm{T}}$, so that both region-average distances are also asymptotically decreasing with λ.

VII. APPROXIMATION FORMULAS FOR $S_{\mathscr{D},\lambda}$ AND u_λ

The prior content S_λ discussed alongside C_λ in Secs. II–V quantifies the size of $\mathscr{R}_\lambda$ relative to $\mathscr{R}_0$. In our earlier article [18], analytical approximation formulas for S_λ were proposed in the large-N limit, all of which are scaled with the volume $V_{\mathscr{R}_0}$ of $\mathscr{R}_0$. As is also shown later in the section, this volume dependence is associated with the extension of every $\mathscr{R}_0$ integral,

$$\int_{\mathscr{R}_0} (d\,\boldsymbol{r}') \, \eta(L - \lambda L_{\max}) \ldots \to \frac{1}{V_{\mathscr{R}_0}} \int_{\text{all space}} \prod_j d\,r'_j \ldots, \tag{16}$$

to the entire $\boldsymbol{r}'$ space ascribed with the uniform primitive prior, which is a reasonable step to obtain analytical results under the central limit theorem since L is narrow enough to reside within the confines of $\mathscr{R}_\lambda$ under this limit. Therefore, the valid usage of these theoretical expressions hinges on the availability of $V_{\mathscr{R}_0}$. In quantum-state tomography where we have no complete theoretical information about $\mathscr{R}_0$, $V_{\mathscr{R}_0}$ is known only for certain priors and state parametrizations [20,21,39,40].

On the other hand, it is obvious that $V_{\mathscr{R}_0}$ is canceled out for any region-average quantity after such integral extensions. This allows one to derive operational asymptotic formulas for averages like $S_{\mathscr{D},\lambda}$ and u_λ regardless of $\mathscr{R}_0$ in whichever parametrization. As a calculable standard in this section, we continue to derive expressions in terms of the uniform primitive prior and $\boldsymbol{r}$, although the subsequent instructions may also work for other manageable priors with which $S_{\mathscr{D},\lambda}$ behaves as a proper region-capacity function. We first address the different $\mathscr{D}$ measures in the large-N limit.

A. Various $\mathscr{D}$ measures

1. Hilbert-Schmidt and trace-class measures

The HS distance measure $\mathscr{D}_{\text{HS}}(\boldsymbol{r}', \widehat{\boldsymbol{r}}_{\text{ML}})$ takes the very simple quadratic form in (12) under any circumstance, whereas the trace-class distance $\mathscr{D}_{\text{tr}}$ has no easy functional form in terms of $\boldsymbol{r}'$ for $D > 2$. Nevertheless in the limits $N \gg 1$ *and* $D \gg 2$, based on the principles of random matrix theory detailed in Appendix A, it is deduced that the asymptotic expression

$$S_{\text{tr}} \approx \frac{8\sqrt{D\,S_{\text{HS}}}}{3\pi} \tag{17}$$

relating the final $\mathscr{R}$ averages S_{HS} and S_{tr} is approximately valid for both cases A and B, which incidentally takes the same form found in [41] that was calculated for statistical-fluctuation studies.

2. Bures measure

The Bures distance measure $\mathscr{D}_{\text{B}}$ also has no tractable functional form in $\boldsymbol{r}'$ for general D. To find the asymptotic link with $\boldsymbol{r}'$ this time, it is technically more convenient to inspect the behavior of $\mathscr{F}$ around $\widehat{\rho}_{\text{ML}} \leftrightarrow \widehat{\boldsymbol{r}}_{\text{ML}}$ as $N \gg 1$.

A Taylor expansion about $\widehat{\rho}_{\text{ML}}$ as guided in Appendix B, we have

$$\mathscr{F}_{\text{A}} \approx 1 - \tfrac{1}{2}(\boldsymbol{r}' - \widehat{\boldsymbol{r}}_{\text{ML}})^{\mathrm{T}} \boldsymbol{Q}_D \, (\boldsymbol{r}' - \widehat{\boldsymbol{r}}_{\text{ML}}) \tag{18}$$

for case A and

$$\mathscr{F}_{\text{B}} \approx 1 + (\boldsymbol{r}' - \widehat{\boldsymbol{r}}_{\text{ML}})^{\mathrm{T}} \text{tr}\{P_r \boldsymbol{\Omega}\}$$
$$+ \tfrac{1}{2}(\boldsymbol{r}' - \widehat{\boldsymbol{r}}_{\text{ML}})^{\mathrm{T}} \left(\tfrac{1}{2}\text{tr}\{P_r \boldsymbol{\Omega}\}\text{tr}\{P_r \boldsymbol{\Omega}^{\mathrm{T}}\} - \boldsymbol{Q}_r\right)(\boldsymbol{r}' - \widehat{\boldsymbol{r}}_{\text{ML}}) \tag{19}$$

for case B, where P_r is the projector onto the support of $\widehat{\rho}_{\text{ML}}$ having the rank-deficient spectral decomposition $\widehat{\rho}_{\text{ML}} = \sum_{j=1}^r |\lambda_j\rangle \lambda_j \langle\lambda_j|$, and

$$\boldsymbol{Q}_r = \sum_{j=1}^r \sum_{k=1}^r \frac{\langle \lambda_j \boldsymbol{\Omega} | \lambda_k \rangle \langle \lambda_k | \boldsymbol{\Omega}^{\mathrm{T}} | \lambda_j \rangle}{\lambda_j + \lambda_k}. \tag{20}$$

B. Case A: Hyperellipsoidal theory

The presentation in Sec. VII A reduces the necessary ingredients for large-N (or D) analytical estimations of $S_{\mathscr{D},\lambda}$ to just the scalar $\int_{\mathscr{R}_\lambda}(d\,\boldsymbol{r}')$, column $\int_{\mathscr{R}_\lambda}(d\,\boldsymbol{r}')\,\boldsymbol{\Delta}'_{\text{ML}}$, and dyadic $\int_{\mathscr{R}_\lambda}(d\,\boldsymbol{r}')\,\boldsymbol{\Delta}'_{\text{ML}}\,\boldsymbol{\Delta}'^{\mathrm{T}}_{\text{ML}}$, where $\boldsymbol{\Delta}'_{\text{ML}} = \boldsymbol{r}' - \widehat{\boldsymbol{r}}_{\text{ML}}$.

When $\mathscr{R}_\lambda \approx \mathscr{E}_\lambda$, these three integrals take on simple analytical forms. We start with

$$\int_{\mathscr{R}_\lambda} (d\,\boldsymbol{r}') = \int_{\mathscr{R}_0} (d\,\boldsymbol{r}')\, \eta\big(1 - \boldsymbol{\Delta}'^{\mathrm{T}}_{\text{ML}} \boldsymbol{F}_{\text{ML}} \boldsymbol{\Delta}'_{\text{ML}}/(-2\ln\lambda)\big) \tag{21}$$

and transform $\boldsymbol{r}' \to \boldsymbol{r}'' = \boldsymbol{D}^{1/2}\boldsymbol{O}^{\mathrm{T}}\boldsymbol{\Delta}'_{\text{ML}}$ to the translated diagonal coordinate variables of $\boldsymbol{F}_{\text{ML}}/(-2\ln\lambda) = \boldsymbol{O}\boldsymbol{D}\boldsymbol{O}^{\mathrm{T}}$, so that in the large-$N$ limit and uniform primitive prior, we may relax the boundary of $\mathscr{R}_0$ and write

$$\int_{\mathscr{R}_\lambda} (d\,\boldsymbol{r}') \to \frac{\det\{\boldsymbol{D}^{-1/2}\}}{V_{\mathscr{R}_0}} \int (d\,\boldsymbol{r}'')_{\text{unif}}\, \eta(1 - \boldsymbol{r}''^2)$$
$$= \frac{V_d}{V_{\mathscr{R}_0}}\, (-2\ln\lambda)^{d/2}\, \det\{\boldsymbol{F}_{\text{ML}}\}^{-1/2}, \tag{22}$$

which is a function of the volume $V_d = \pi^{d/2}/(d/2)!$ of the d-dimensional unit hyperball, the inverse of $\boldsymbol{F}_{\text{ML}}$ that characterizes $\mathscr{E}_\lambda$ together with the logarithm of λ.

In this case, the integral column is zero since the integrand after variable transformation becomes odd in $\boldsymbol{r}''$, and we are thus left with

$$\int_{\mathscr{R}_\lambda} (d\,\boldsymbol{r}')\, \boldsymbol{\Delta}'_{\text{ML}} \boldsymbol{\Delta}'^{\mathrm{T}}_{\text{ML}} \to \frac{\det\{\boldsymbol{D}^{-1/2}\}}{V_{\mathscr{R}_0}}\, \boldsymbol{O}\boldsymbol{D}^{-1/2}\boldsymbol{I}\boldsymbol{D}^{-1/2}\boldsymbol{O}^{\mathrm{T}}, \tag{23}$$

and

$$\boldsymbol{I} = \int (d\,\boldsymbol{r}'')_{\text{unif}}\, \eta(1 - \boldsymbol{r}''^2)\, \boldsymbol{r}''\boldsymbol{r}''^{\mathrm{T}}$$
$$= \int_{\text{unit sphere}} (d\,\boldsymbol{r}'')\, \boldsymbol{r}''\boldsymbol{r}''^{\mathrm{T}}$$
$$= \int_0^1 d\,r''\, r''^{d+1} \int (d\,\{\text{solid angle}\})\, \boldsymbol{e}''\boldsymbol{e}''^{\mathrm{T}} = \frac{V_d}{d+2}\mathbf{1}, \tag{24}$$

where the last equality is explained by the orthogonal invariant of the $(d-1)$-dimensional solid-angle measure over the unit
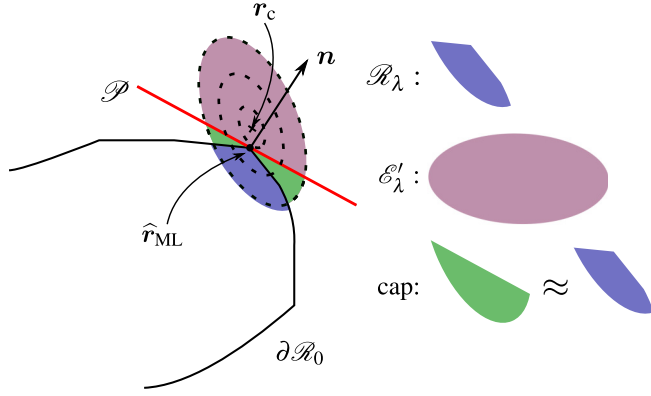
FIG. 6. Modeling the boundary $\partial\mathcal{R}_\lambda \cap \partial\mathcal{R}_0$ [24]: A hyperplane $\mathcal{P}$ (red solid line) is introduced in a manner that its normal $\boldsymbol{n}$ is orthogonal to the level curve at $\widehat{\boldsymbol{r}}_{\mathrm{ML}}$ to form a cap that approximates $\mathcal{R}_\lambda$.

columns $\boldsymbol{e}$, and so

$$\int_{\mathcal{R}_\lambda} (d\,\boldsymbol{r}')\,\boldsymbol{\Delta}'_{\mathrm{ML}}\,\boldsymbol{\Delta}'^{\,\mathrm{T}}_{\mathrm{ML}} \rightarrow \frac{V_d}{V_{\mathcal{R}_0}}\,(-2\ln\lambda)^{d/2+1}\,\det\{\boldsymbol{F}_{\mathrm{ML}}\}^{-1/2}\,\boldsymbol{F}^{-1}_{\mathrm{ML}}. \tag{25}$$

With all these components, the relevant asymptotic formulas concerning all three distance measures are

$$S^{(\mathrm{A})}_{\mathrm{HS},\lambda} \approx \mathrm{Tr}\{\boldsymbol{F}^{-1}_{\mathrm{ML}}\}\frac{-\ln\lambda}{d/2+1},$$

$$S^{(\mathrm{A})}_{\mathrm{tr},\lambda} : \text{ as in (17)},$$

$$S^{(\mathrm{A})}_{\mathrm{B},\lambda} \approx \mathrm{Tr}\{\boldsymbol{F}^{-1}_{\mathrm{ML}}\boldsymbol{Q}_D\}\frac{-\ln\lambda}{d+2}. \tag{26}$$

Here Tr now addresses the dyadic character, as opposed to tr, and we witness the manifestation of logarithmic divergences from both the relaxation of $\partial\mathcal{R}_0$ and Gaussian approximation of L.

Next, to analytically calculate u_λ using $f(L) = \ln L$ with which C_λ can be found, we note that, due to the Gaussian form of L,

$$u_\lambda = -\ln\lambda - \frac{\int_{\mathcal{R}_\lambda}(d\,\boldsymbol{r}')\,\boldsymbol{\Delta}'^{\,\mathrm{T}}_{\mathrm{ML}}\,\boldsymbol{F}_{\mathrm{ML}}\,\boldsymbol{\Delta}'_{\mathrm{ML}}}{2\int_{\mathcal{R}_\lambda}(d\,\boldsymbol{r}')} \tag{27}$$

is a dyadic trace function of $\int_{\mathcal{R}_\lambda}(d\,\boldsymbol{r}')\,\boldsymbol{\Delta}'_{\mathrm{ML}}\,\boldsymbol{\Delta}'^{\,\mathrm{T}}_{\mathrm{ML}}$, so that we may use the right-hand side of (25) and put down

$$u_{\mathrm{A},\lambda} = -\frac{2}{d+2}\ln\lambda \tag{28}$$

after some basic trace and logarithmic manipulations. It is clear that $d/(d+2) \leqslant u_{\mathrm{A},\lambda} \leqslant 1$ is bounded.

C. Case B: Hyperellipsoidal-cap theory

In case B, although the geometry of $\mathcal{R}_\lambda \approx \mathcal{E}'_\lambda \cap \mathcal{R}_0$ is now much trickier to deal with, the central limit theorem proposed in Sec. III allows us to approximate $\mathcal{R}_\lambda$ by a regular analytical region.

As shown in Fig. 6, one can introduce a hyperplane $\mathcal{P}$, described by $\boldsymbol{n} \cdot (\boldsymbol{r}' - \widehat{\boldsymbol{r}}_{\mathrm{ML}}) = 0$ ($\boldsymbol{n} \propto \boldsymbol{g}_{\mathrm{ML}}$) that is tangent to the level curve of the Gaussian function in (4) at $\widehat{\boldsymbol{r}}_{\mathrm{ML}}$. The

hyperspherical cap formed by $\mathcal{P}$ and $\mathcal{E}'_\lambda$ hence asymptotically contains $\mathcal{R}_\lambda$, where we have essentially modeled the highly nontrivial $\partial\mathcal{R}_\lambda \cap \partial\mathcal{R}_0$ as $\mathcal{P}$. This model implies the estimated assignment

$$\int_{\mathcal{R}_\lambda}(d\,\boldsymbol{r}')\ldots$$

$$= \frac{1}{V_{\mathcal{R}_0}}\int(d\,\boldsymbol{r}')_{\mathrm{unif}}\,\eta(1 - (\boldsymbol{r}' - \boldsymbol{r}_{\mathrm{c}})^{\mathrm{T}}\boldsymbol{F}_{\mathrm{ML}}\,(\boldsymbol{r}' - \boldsymbol{r}_{\mathrm{c}})/(-2\ln\lambda'))$$

$$\times \eta(\boldsymbol{n}\cdot(\widehat{\boldsymbol{r}}_{\mathrm{ML}} - \boldsymbol{r}'))\ldots. \tag{29}$$

The change of variable $\boldsymbol{r}' \rightarrow \boldsymbol{r}'' = \boldsymbol{D}'^{1/2}\boldsymbol{O}'^{\,\mathrm{T}}(\boldsymbol{r}' - \boldsymbol{r}_{\mathrm{c}})$ with respect to the diagonal coordinates of $\boldsymbol{F}_{\mathrm{ML}}/(-2\ln\lambda') = \boldsymbol{O}'\boldsymbol{D}'\boldsymbol{O}'^{\,\mathrm{T}}$ leads to

$$\int_{\mathcal{R}_\lambda}(d\,\boldsymbol{r}')\,q_\lambda(\boldsymbol{r}') \approx \frac{\det\{\boldsymbol{D}'^{-1/2}\}}{V_{\mathcal{R}_0}}\int(d\,\boldsymbol{r}'')_{\mathrm{cap}}\,q_\lambda(\boldsymbol{O}'\boldsymbol{D}'^{-1/2}\,\boldsymbol{r}'') \tag{30}$$

for any function q, which is parametrized by the cap element $(d\,\boldsymbol{r}'')_{\mathrm{cap}} = (d\,\boldsymbol{r}'')_{\mathrm{unif}}\,\eta(1 - \boldsymbol{r}''^2)\,\eta(a - \boldsymbol{b}^{\mathrm{T}}\boldsymbol{r}'')$, $a = \boldsymbol{n}^{\mathrm{T}}(\widehat{\boldsymbol{r}}_{\mathrm{ML}} - \boldsymbol{r}_{\mathrm{c}})$, and $\boldsymbol{b} = \boldsymbol{D}'^{-1/2}\boldsymbol{O}'^{\,\mathrm{T}}\boldsymbol{n}$. One can check that

$$l \equiv \frac{a}{|\boldsymbol{b}|} = \frac{\boldsymbol{g}_{\mathrm{ML}}\cdot(\widehat{\boldsymbol{r}}_{\mathrm{ML}} - \boldsymbol{r}_{\mathrm{c}})}{|\boldsymbol{D}'^{-1/2}\boldsymbol{O}'^{\,\mathrm{T}}\boldsymbol{g}_{\mathrm{ML}}|}$$

$$= \sqrt{\frac{(\widehat{\boldsymbol{r}}_{\mathrm{ML}} - \boldsymbol{r}_{\mathrm{c}})^{\mathrm{T}}\boldsymbol{F}_{\mathrm{ML}}\,(\widehat{\boldsymbol{r}}_{\mathrm{ML}} - \boldsymbol{r}_{\mathrm{c}})}{(-2\ln\lambda')}} \leqslant 1. \tag{31}$$

In other words, we have

$$\overline{q_\lambda(\boldsymbol{r}')}^{\,\mathcal{R}_\lambda} \approx \frac{\int(d\,\boldsymbol{r}'')_{\mathrm{cap}}\,q_\lambda(\boldsymbol{O}'\boldsymbol{D}'^{-1/2}\,\boldsymbol{r}'')}{\int(d\,\boldsymbol{r}'')_{\mathrm{cap}}}, \tag{32}$$

and that for any q_λ belonging to either one of the three distance measures or $\ln L - \ln(\lambda L_{\max})$, as reasoned in Sec. VII C, the building blocks of $\overline{q_\lambda(\boldsymbol{r}')}^{\,\mathcal{R}_\lambda}$ are only $\int(d\,\boldsymbol{r}'')_{\mathrm{cap}}$, $\int(d\,\boldsymbol{r}'')_{\mathrm{cap}}\,\boldsymbol{r}''$, and $\int(d\,\boldsymbol{r}'')_{\mathrm{cap}}\,\boldsymbol{r}''\,\boldsymbol{r}''^{\,\mathrm{T}}$. These integrations are all carried out in Appendix C.

In combining all results gathered from Appendixes B and C, we denote $\mathcal{N}_{d,l,x} = V_d\,\mathrm{I}_{(1-l)/2}((d+x)/2, (d+x)/2)$, which depends on the incomplete Euler's beta function $\mathrm{I}(.,.)$, and organize two new auxiliary quantities

$$\boldsymbol{m} = \left[-\frac{V_{d-1}}{l(d+1)}(1 - l^2)^{(d+1)/2} + \mathcal{N}_{d,l,1}\right]\boldsymbol{F}^{-1}_{\mathrm{ML}}\boldsymbol{g}_{\mathrm{ML}},$$

$$\boldsymbol{M} = \frac{-\ln\lambda'}{d+2}\,\mathcal{N}_{d,l,3}\,\boldsymbol{F}^{-1}_{\mathrm{ML}} + \frac{1}{2}\,\boldsymbol{m}\,\boldsymbol{g}^{\mathrm{T}}_{\mathrm{ML}}\boldsymbol{F}^{-1}_{\mathrm{ML}}. \tag{33}$$

This helps to clean the respective formulas

$$S^{(\mathrm{B})}_{\mathrm{HS},\lambda} \approx \frac{\mathrm{Tr}\{2\boldsymbol{M}\}}{\mathcal{N}_{d,l,1}},$$

$$S^{(\mathrm{B})}_{\mathrm{tr},\lambda} : \text{ as in(17)},$$

$$S^{(\mathrm{B})}_{\mathrm{B},\lambda} \approx \frac{\mathrm{tr}\{P_r\,\boldsymbol{\Omega}^{\mathrm{T}}\boldsymbol{m}\} + \mathrm{Tr}\{\boldsymbol{M}\,\boldsymbol{Q}_r\}}{\mathcal{N}_{d,l,1}} \approx \frac{\mathrm{tr}\{P_r\,\boldsymbol{\Omega}^{\mathrm{T}}\boldsymbol{m}\}}{\mathcal{N}_{d,l,1}} \tag{34}$$

for the distance-induced capacity functions and

$$u_{\mathrm{B},\lambda} = [-\ln\lambda' + \mathrm{Tr}\{\boldsymbol{g}_{\mathrm{ML}}\boldsymbol{m}^{\mathrm{T}} - \boldsymbol{F}_{\mathrm{ML}}\boldsymbol{M}\}/\mathcal{N}_{d,l,1}]$$

$$\times \ln(\lambda\,L_{\max})/\ln(\lambda'\,L_{\max}). \tag{35}$$
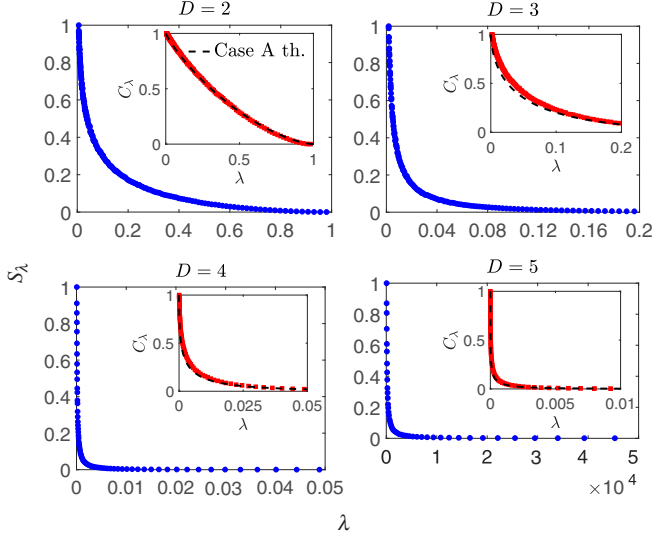
FIG. 7. Graphs of size S_λ and credibility C_λ for case A with $2 \leqslant D \leqslant 5$. M, the number of POM outcomes, is set to D^3 and the POM is chosen to be a random square-root measurement as a simulation example for each D. Here $N/M = 500$. The dashed curve in every inset is computed with the case-A large-N formula for C_λ in Eq. (7) of [18]. A randomly chosen rank-D true state ρ is used in each panel. A total of 20 000 points are collected during in-region sampling of u_λ per λ.

We caution the reader once more regarding the actions of tr and Tr at the right-hand side of $S_{\mathrm{B},\lambda}^{(\mathrm{B})}$ in (34).

For consistency, we end this section by noting that Eqs. (34) and (35) cover Eqs. (26) and (27) because case A implies that $\lambda' = \lambda$ ($\boldsymbol{g}_{\mathrm{ML}} = \boldsymbol{0} = \boldsymbol{m}$), such that $l = 0$ then gives $\mathcal{N}_{d,0,x} = V_d$ and $\boldsymbol{M} = (-\ln \lambda) \boldsymbol{F}_{\mathrm{ML}}^{-1}/(d + 2)$.

VIII. RESULTS AND DISCUSSIONS

A. Region reconstruction

We first present, under the uniform primitive prior $p(\boldsymbol{r}) \propto 1$, the computation results of S_λ and C_λ from u_λ in Figs. 7 and 8 for quantum systems of various dimensions D. To be more technically precise about our use of Euler's method described in Sec. IV, we first solve (9) for y_λ by iterating (10) starting with a numerically small λ value to $\lambda \approx 1$ using the function $f(\lambda) = \ln L$.

The behavior of S_λ shows the expected decreasing trend not only in λ, but also in overall magnitude as D increases. This indicates that the (log) likelihood is turning into a delta-function peak. For larger D or N, the computational accuracy of S_λ and C_λ using Euler's numerical method may be maintained by exploring many more λ values near zero, such as via logarithmic scaling of λ, as all curves possess sharp gradient changes in this λ range.

In Fig. 9, both simulated data and theoretical curves of all three capacity functions S_{HS}, S_{tr}, and S_{B} are plotted against the credibility C for case A. In this case, there exist no other factors that could spoil the perfect hyperellipsoidal geometry of $\mathscr{R}_\lambda$. As such, the analytical curves fit almost perfectly with the simulated points. We note that even the average trace-class distance S_{tr}, which is approximated with (17) through the
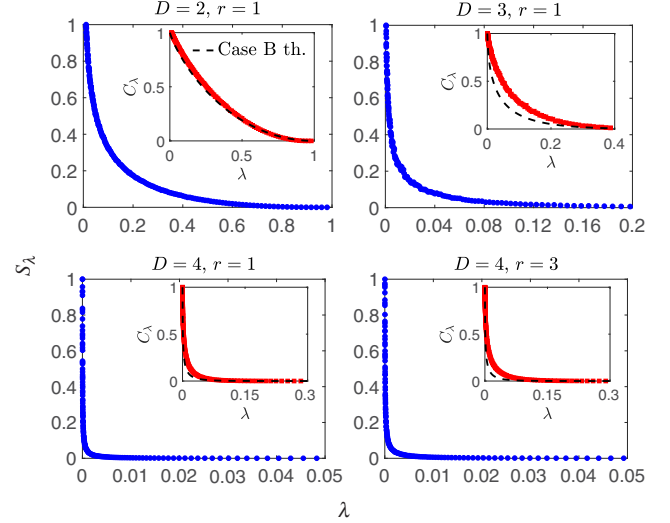

FIG. 8. Graphs of computed S_λ and C_λ Case B against λ, with $2 \leqslant D \leqslant 4$, which have, otherwise, the same specifications as Fig. 7. The rank r of $\widehat{\rho}_{\mathrm{ML}}$, which characterizes a pure state, for each panel is explicitly stated. The dashed curves in this figure are generated from the case B formulas in Eq. (14) of [18].

theory of random matrices, performs very well relative to the simulated data points.

In case B, we can start to see discrepancies between theory and simulation from Fig. 10 especially for larger D. Such deviations are inevitable as the hyperellipsoidal-cap estimation of the actual CR $\mathscr{R}_\lambda$ proposed in Sec. VII C introduces additional space outside $\mathscr{R}_0$ that is certainly not contained
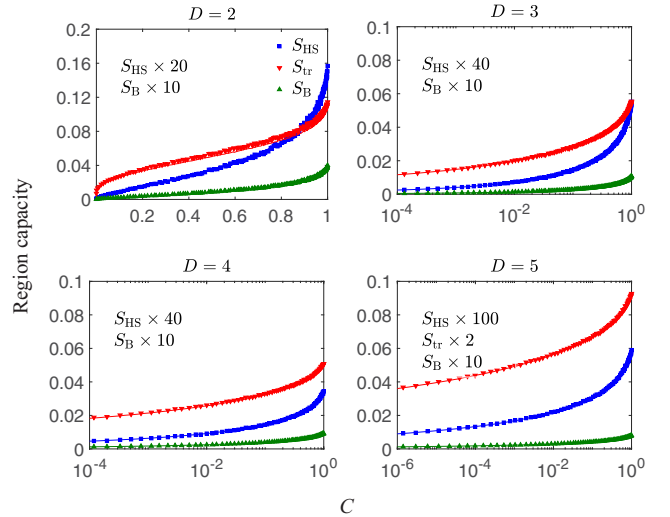

FIG. 9. Graphs of the distance-induced region-capacity function against the credibility C for case A with $2 \leqslant D \leqslant 5$ for all the full-rank ML estimators that produced Fig. 7. The measurement POM is, again, a set of $M = D^3$ square-root measurement outcomes for each D that measures $N/M = 500$. *All* horizontal axes represent C, and vertical axes $S_{\mathscr{D}}$. The solid analytical curves are calculated using Eq. (26). All $S_{\mathscr{D}}$s are magnified—according to the magnification factors stated in the panels—so that all graphs and markers can be visibly coplotted inside each panel.
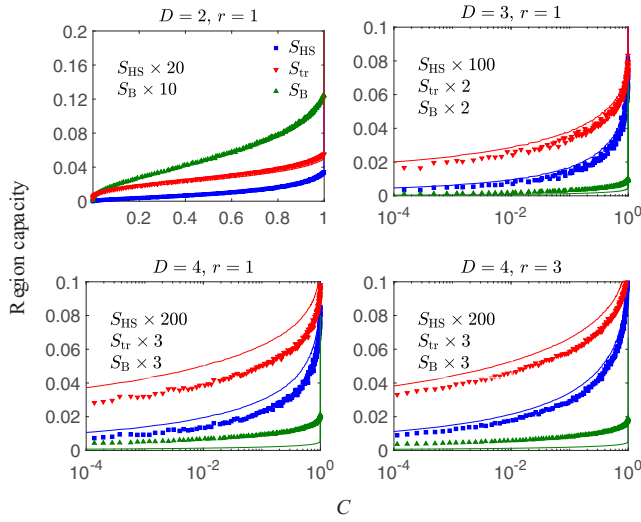
FIG. 10. Graphs of the region capacity against C for case B for all the rank-deficient ML estimators involved in Fig. 8. The solid curves, which originate from Eq. (34), still lie reasonably close to the simulated markers for these low-dimensional examples. All specifications otherwise follow those of Fig. 9. The geometrical difference between the actual rank-deficient CR boundary and a hyperplane manifests as deviations from theoretical predictions. High-rank ML estimators, nonetheless, generally give a better theoretical predictions in contrast to low-rank estimators.
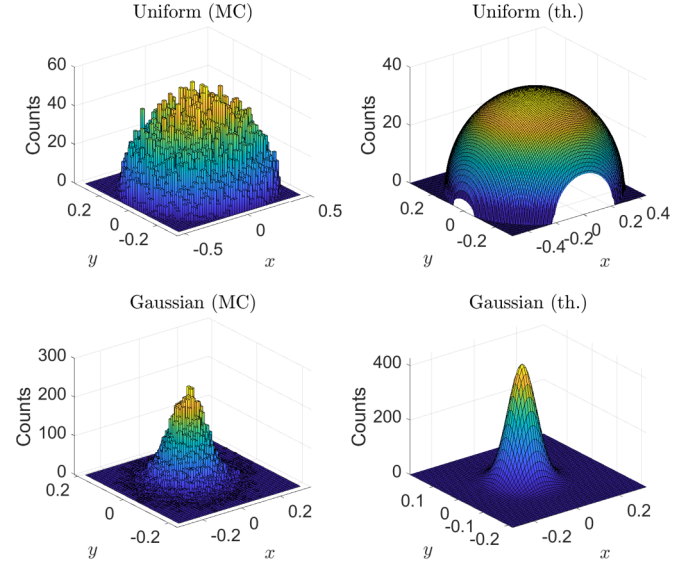

FIG. 11. Example comparisons between hit-and-run simulated and theoretical distributions made in case A, for $D = 2$, random square-root measurement ($M = 4$), and $N = 50$. Their common co-ordinate system is centered at $\widehat{\boldsymbol{r}}_{\mathrm{ML}}$ and rotated in the frame of the error region. Both uniform and Gaussian (of covariance $10\boldsymbol{F}_{\mathrm{ML}}$) prior distributions considered here are projected onto the error region, which is approximated as a hyperellipsoid for calculating the theoretical distributions (see Appendix D for their explicit probability-density expressions).

in $\mathscr{R}_\lambda$. More specifically, for very large D, if the rank-deficient ML estimator $\widehat{\rho}_{\mathrm{ML}}$ is located at an extremely sharp state-space corner, which is labeled as the type-I situation in Sec. V B, this additional space would be exceedingly large relative to the physical CR, which incurs a proportionately large theory-simulation mismatch. On the other hand, if the rank-deficient $\widehat{\rho}_{\mathrm{ML}}$ lies on a relatively flat part of the state-space boundary (the type-II scenario), then this overestimated space, and hence the mismatch, would be much smaller. ML estimators of ranks 1 and 3, which are considered in Fig. 10, are prime examples of the respective type-I and -II situations. Regardless, the asymptotic formulas in Sec. VII C may still be used for an order-of-magnitude estimation of $S_\mathscr{D}$ and C.

We note that in-region sampling is not restricted to just the uniform distribution, so long as the MC method employed is sufficiently general. Such is the case for hit and run. For a calibration check of the general hit-and-run algorithm in Sec. V, we generate and compare both uniform and Gaussian distributions with their respective theoretically derived counterparts for the single-qubit case in Fig. 11.

Next, as a real demonstration, we consider another natural prior that is asymptotically conjugate to the likelihood function, that is the Gaussian form $p(\boldsymbol{r}) \propto \exp\left(-(\boldsymbol{r} - \widehat{\boldsymbol{r}}_{\mathrm{ML}}) \cdot \boldsymbol{F}_{\mathrm{ML}} \cdot (\boldsymbol{r} - \widehat{\boldsymbol{r}}_{\mathrm{ML}})/(2N)\right)$ having a much broader covariance $N\boldsymbol{F}_{\mathrm{ML}}^{-1}$ defined by the Fisher information for *one copy* . This prior is a logical choice given that our knowledge about $\boldsymbol{r}$ is updated to $\widehat{\boldsymbol{r}}_{\mathrm{ML}}$ after the measurement, which should be used as the most recent prior information for future Bayesian analyses. The corresponding marginal distribution needed to sample along line segments in hit and run is therefore the one-dimensional Gaussian distribution of mean $[\mathbf{e}_{\boldsymbol{v}} \cdot \boldsymbol{F}_{\mathrm{ML}} \cdot (\widehat{\boldsymbol{r}}_{\mathrm{ML}} - \boldsymbol{r}_{\mathrm{ref}})]/\mathbf{e}_{\boldsymbol{v}} \cdot \boldsymbol{F}_{\mathrm{ML}} \cdot \mathbf{e}_{\boldsymbol{v}}$ and variance $1/\mathbf{e}_{\boldsymbol{v}} \cdot \boldsymbol{F}_{\mathrm{ML}} \cdot \mathbf{e}_{\boldsymbol{v}}$. With

these specific distributions, the size and credibility plots as well as the region capacity behaviors for $D = 8$ and 16 are given in Figs. 12 and 13.

B. Correlation properties of hit and run

We recall that, at least under the uniform primitive prior $(d\,\boldsymbol{r}'')_{\mathrm{unif}}$, hit and run converges efficiently to the correct uniform distribution in $O(d^2 R_{\mathrm{out}}^2/R_{\mathrm{in}}^2)$ as discussed in Sec. V B. Furthermore, it was argued [42], as a consequence of the above expression, that given an initial point that has a shortest distance l from the boundary $\partial\mathscr{R}_\lambda$, hit and run eventually mixes sample points into the uniform distribution after $O(d^3 R_{\mathrm{out}}^2/R_{\mathrm{in}}^2 \ln(R_{\mathrm{out}}/l))$ steps.

This reveals a technical caveat for almost all Markov-chain random-walk algorithms: a random walk starting from a sharp corner point of a convex body requires very many steps to approach the stationary target distribution. It is generally much harder to scout the entire convex region from such a corner than from an interior point, since the Markov chain terminates as soon as a next admissible point is obtained, which is probabilistically near the initial corner point around which the admissible region is tight. Such a situation is essentially *status quo* for high-dimensional state reconstruction where the state space $\mathscr{R}_0$ is filled with plenty of extremely sharp corners.

Doing hit and run from an interior point is therefore a primary objective for general CR analysis. Even without the full knowledge about the CR, it is still possible to numerically compute a point that is sufficiently interior for this purpose. The idea is to first find multiple random states on $\partial\mathscr{R}_\lambda \cap \mathrm{int}(\mathscr{R}_0)$, the boundary of $\mathscr{R}_\lambda$ in the interior of $\mathscr{R}_0$, and next

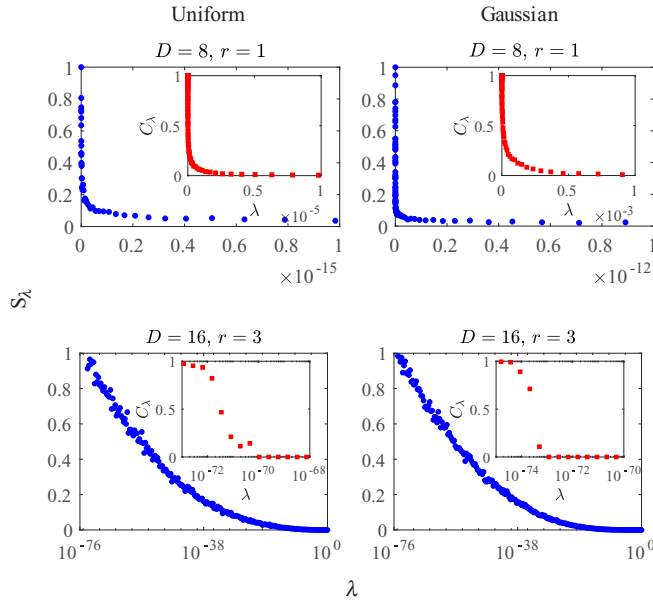FIG. 12. Graphs of S_λ and C_λ for $D = 8$ and 16 under uniform and Gaussian prior distributions for case B as explained in the main text. The measurement configurations are set to $N/M = 5000$ and $N/M = 50$ respectively for $D = 8$ and 16, where $M = D^3$. All plot markers are computed with 20 000 points generated during in-region sampling of u_λ per λ. All case-A credibility curves (not shown in this figure) match the theoretical results from Ref. [18]. On the other hand, the case-B theoretical curves for C_λ can now be very different from the actual ones because of the complicated state-space boundary.
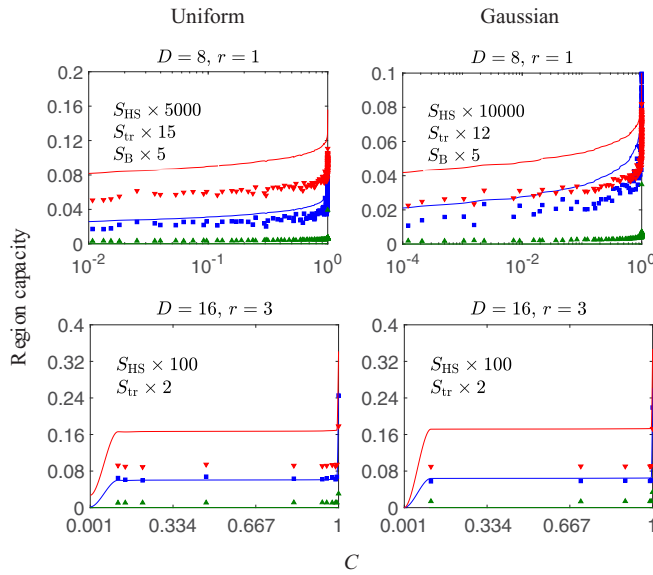

FIG. 13. Graphs of the region capacity against C for all the rank-deficient ML estimators referred by Fig. 12. The theoretical curves for S_{HS}, S_{tr}, and S_B are represented by the upper red, middle blue, and bottom green solid curves respectively, whereas the simulated marker colors are as specified in Fig. 10. The deviations from theoretical approximations for the region capacity, which is based on hyperplanar geometry, are apparently relatively more robust to high-dimensional state-space boundary features as compared to S and C. More points are concentrated around large C.

average all these states to obtain an interior state of $\mathscr{R}_\lambda$. This is evidently equivalent to the minimization of the convex function $[(\boldsymbol{x}' - \boldsymbol{r}_c) \cdot \boldsymbol{F}_{ML} \cdot (\boldsymbol{x}' - \boldsymbol{r}_c)/(-2\ln\lambda') - 1]^2$ with respect to $\rho' \geqslant 0$ for which $\boldsymbol{x}' = \text{tr}\{\rho'\boldsymbol{\Omega}\}$ multiple times in the large-N limit. Fortunately, this can be carried out extremely quickly by using the superfast accelerated projected gradient routine [43] (see Appendix E).

Figures 14 and 15 supply graphical visualization of the key sampling activities that goes on for case B with $D = 4$, where a rank-1 ML estimator is obtained. It is clear that starting hit and run from a corner point (namely the ML estimator, for instance) introduces small average hopping distances for subsequent Markov chains. This can be interpreted as strong sample correlations that prevent wide coverage of the CR, contrary to performing hit and run starting from an interior point.

C. Constructions of plausible regions

The matter of inspecting S_λ for a fixed C_λ, say 0.95, is rather subjective, for very often one requires experienced judgment to decide if such a value is sufficient for subsequent prediction tasks. As advocated in [44,45], there exists a statistically meaningful treatment of the measured data set $\mathbb{D}$ based on the concept of *evidence*. It is thus fitting for us to end this article with a short review on how our in-region sampling technique studied here directly supports another interesting kind of Bayesian analysis.

By definition, we say that $\boldsymbol{r}'$ is a *plausible* candidate parameter for the true $\boldsymbol{r}$ if there is indeed evidence in favor of this supposition. That is, its normalized posterior probability $L(\mathbb{D}|\boldsymbol{r}')\,p(\boldsymbol{r}')/L(\mathbb{D})$ after the measurement is larger than its prior probability $p(\boldsymbol{r}')$ before this measurement was performed. Therefore, $\boldsymbol{r}$ is a plausible parameter if the evidence supports the prior knowledge. Under this evidence-belief framework, one can construct another type of Bayesian region—the *plausible region* (PR)—that contains *all* plausible choices of $\boldsymbol{r}$. This is really the CR $\mathscr{R} = \mathscr{R}_{\lambda=\lambda_{crit}}$ characterized by the critical value [17]

$$\lambda_{crit} = \int_0^1 d\lambda'\, S_{\lambda'}, \qquad (36)$$

for which $L(\mathbb{D}|\boldsymbol{r} \in \partial\mathscr{R}_{\lambda=\lambda_{crit}}) = L(\mathbb{D})$, or the CR that contains all plausible points and nothing else. This follows quickly from the following equality chain:

$$L(\mathbb{D}) = \int (d\,\boldsymbol{r}')\, L(\mathbb{D}|\boldsymbol{r}') = \int (d\,\boldsymbol{r}') \int_0^{L(\mathbb{D}|\boldsymbol{r}')} dx'$$

$$= L_{max} \int (d\,\boldsymbol{r}') \int_0^1 d\lambda'\, \eta[L(\mathbb{D}|\boldsymbol{r}') - \lambda' L_{max}]$$

$$= L_{max} \int_0^1 d\lambda'\, S_{\lambda'}, \qquad (37)$$

so that $L(\mathbb{D}|\boldsymbol{r} \in \partial\mathscr{R}_{\lambda=\lambda_{crit}}) \equiv \lambda_{crit} L_{max} = L(\mathbb{D})$ gives Eq. (36).

So, constructing a PR is nothing more than one additional step of computing λ_{crit} after a CR construction. In our previous work [18,19], we have supplied MC-less asymptotic approximations to the expression of λ_{crit} for the uniform primitive prior. In the current context, it clearly follows that λ_{crit} is also
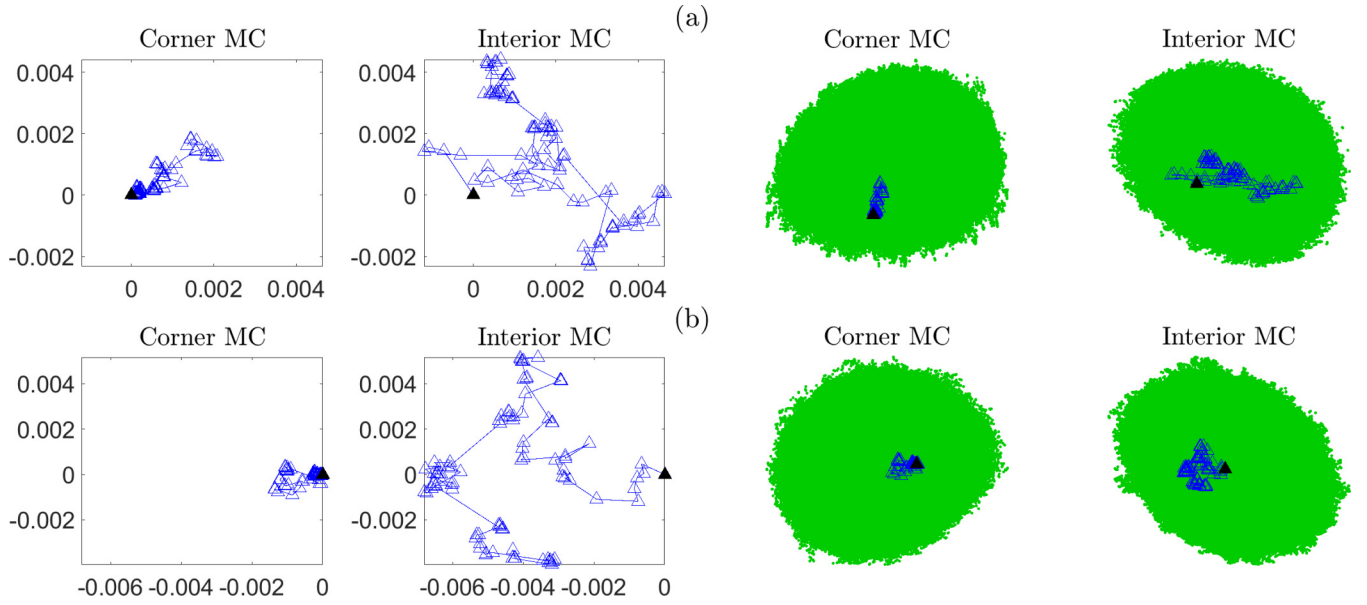
FIG. 14. Case-B correlation strengths (translated to state-space hopping distances) of the first 100 hit-and-run MC-sampled points starting from both a corner of the CR (green area made up of 10 million uniformly sampled points) and an interior point, illustrated for two-qubit systems ($D = 4$), a rank-1 $\widehat{\rho}_{\mathrm{ML}}$ obtained using $M = D^3$ and $N/M = 500$, and a uniform primitive prior in $\boldsymbol{r}$. The corner MC run starts (black shaded marker) from $\widehat{\rho}_{\mathrm{ML}}$, and the interior run starts from a fully mixed state inside the CR generated by averaging the first 1000 points of the hit-and-run algorithm beginning with $\widehat{\rho}_{\mathrm{ML}}$. In properly scaled axes, corner MC shows a stronger correlation (shorter average hopping distances) than interior MC in the respective planes of (a) shortest and (b) longest average hopping distances, as the former encounters the region boundary much more frequently than the latter. This significantly limits the span of sampled points in the entire CR.

directly computable by simply doing a Riemann summation of the full S_λ spectrum obtained through in-region sampling in accordance with Eq. (36).

IX. CONCLUSIONS

Quantum-state tomography is an important application of multidimensional parameter estimation. The construction

of Bayesian credible regions for the reconstructed quantum states after tomography is, unfortunately, a highly nontrivial problem owing to the complex constraints inherited from the state space. The standard numerical recipe of first doing a Monte Carlo sampling of the state space and next discarding points outside the credible region to compute its region qualities (size and credibility) quickly becomes infeasible when the data set collected in an experiment is relatively big, as
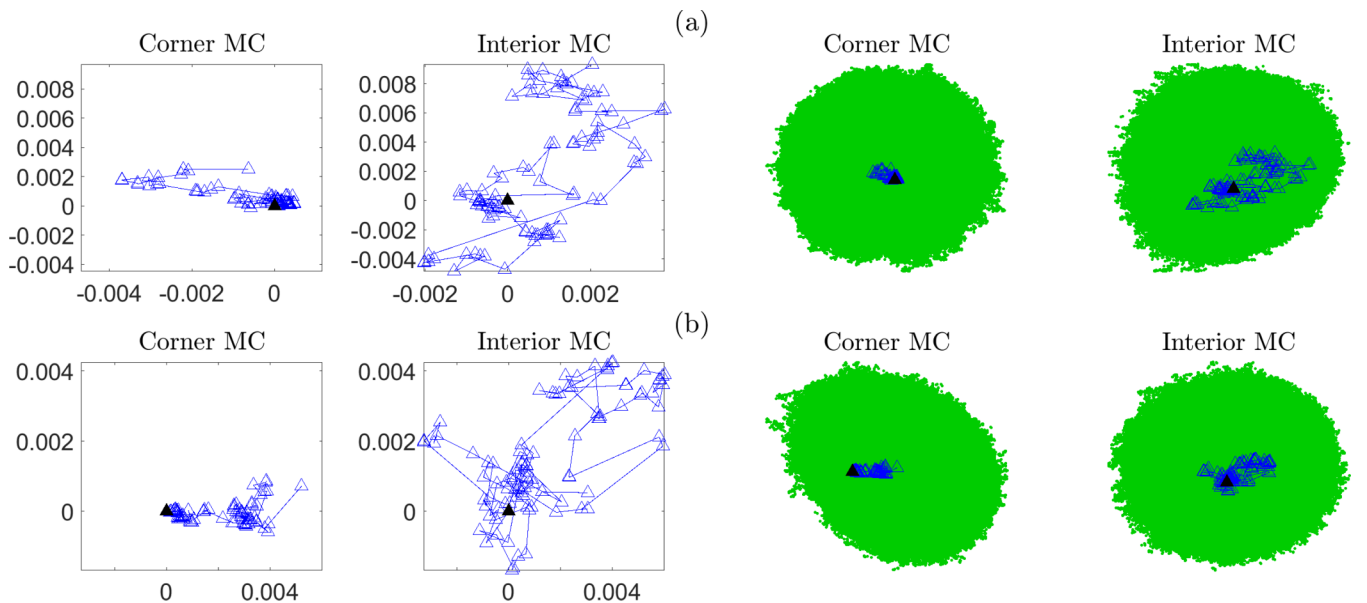

FIG. 15. Case-B correlation strengths (translated to state-space hopping distances) of the first 100 hit-and-run MC-sampled points for the Gaussian prior distribution centered at $\widehat{\rho}_{\mathrm{ML}}$. All figure specifications and conclusions are otherwise identical to those in Fig. 14.

the corresponding credible region would be very small with respect to the state space.

In this article, we performed an extensive analysis of our recent in-region sampling technique that can construct credible regions that are usually very small in practice for reasonably high credibility values and large data samples. This technique computes credible-region qualities of a small credible region for any given prior distribution by inspecting how an appropriately chosen region-average quantity changes as the shape of the region varies. This procedure transforms the general credible-region construction into a sequence of direct region sampling followed by a simple numerical solution to a single-variable differential equation. This results in no sample wastage since no points are discarded. The method of accelerated hit and run is one numerical scheme that can be used to compute region averages rather efficiently provided that sample correlation is properly mitigated with good Monte Carlo starting points. One can also estimate its numerical complexity in the context of tomography despite the complicated state-space boundary.

Furthermore, for highly complex quantum systems of extremely large dimensions, where all numerical methods eventually become practically infeasible, we derive a set of analytical formulas to perform approximate Bayesian error certification through the perspective of distance-induced region capacity measures that alternatively quantifies how large a credible region is. These formulas are now fully operational and further complement those for the conventional size function developed in previous works that require knowledge of the state-space volume.

ACKNOWLEDGMENTS

APPENDIX A: RELATIONSHIPS BETWEEN S_{HS} AND S_{tr} IN THE LARGE-N (OR D) LIMIT

Apart from $\mathscr{D}_{\mathrm{HS}}$, all other measures have no direct analogs in the $\boldsymbol{r}'$ parametrization. However in certain limits, all these measures have approximate relations with $\mathscr{D}_{\mathrm{HS}}$.

We start with making an approximate connection between S_{tr} and S_{HS} by examining the Hermitian operator $\Delta\rho' = \rho' - \widehat{\rho}_{\mathrm{ML}}$ ($\rho' \in \mathscr{R}$). In case A, the distribution of $\Delta\rho'$ in $\mathscr{R}$ has zero mean, $\overline{\Delta\rho'}^{\mathscr{R}} = 0$. This is also approximately true for the case B situation when N is sufficiently large such that $\mathscr{R}$ is small. Furthermore, the space of $\Delta\rho'$ is essentially a bounded set of Hermitian random operators. Here, we shall make the assumption that each matrix entry $\Delta\rho'_{jk}$ in the computational basis is an independent random complex number. Under this condition, the $\Delta\rho$'s form what is now known as a *Wigner ensemble* [46–49] with the second moment equal to $\overline{|\Delta\rho'_{jk}|^2}^{\mathscr{R}} = \mathrm{tr}\{(\Delta\rho')^2\} = S_{\mathrm{HS}}$. Moreover, they are known to have an independent and identically distributed eigenvalue

spectrum that follows the Wigner semicircle law,

$$\sigma(\Delta\rho'/\sqrt{D}) \sim \frac{1}{2\pi S_{\mathrm{HS}}}\sqrt{4\,S_{\mathrm{HS}} - x^2}$$
$$\text{for} \ -2\sqrt{S_{\mathrm{HS}}} \leqslant x \leqslant 2\sqrt{S_{\mathrm{HS}}}, \quad \text{(A1)}$$

in the large-D limit. The trace-class distance $\mathscr{D}_{\mathrm{tr}}$ can thus be calculated with the integral

$$\mathscr{D}_{\mathrm{tr}} \approx \frac{\sqrt{D}}{2\pi S_{\mathrm{HS}}} \int_{-2\sqrt{S_{\mathrm{HS}}}}^{2\sqrt{S_{\mathrm{HS}}}} d\,x\, |x|\, \sqrt{4\,S_{\mathrm{HS}} - x^2} = \frac{8\sqrt{D\,S_{\mathrm{HS}}}}{3\pi}, \quad \text{(A2)}$$

so that we end up with (17).

For case B, that $\overline{\Delta\rho'}^{\mathscr{R}} = 0$ is obvious, but as we have no means of analytically estimating $\overline{\Delta\rho'}^{\mathscr{R}}$, we make a further approximation that as long as $\mathscr{R}$ is sufficiently small, the offset to $\overline{\Delta\rho'}^{\mathscr{R}}$ will be proportionally small, so that (17) remains a reasonable asymptotic approximation.

APPENDIX B: FIDELITY IN THE LARGE-N LIMIT

A Taylor expansion of $\mathscr{F}$ about $\widehat{\boldsymbol{r}}_{\mathrm{ML}}$, or

$$\mathscr{F} \approx 1 + (\boldsymbol{r}' - \widehat{\boldsymbol{r}}_{\mathrm{ML}})^{\mathrm{T}} \frac{\partial \mathscr{F}_{\mathrm{ML}}}{\partial \widehat{\boldsymbol{r}}_{\mathrm{ML}}}$$
$$+ \frac{1}{2}(\boldsymbol{r}' - \widehat{\boldsymbol{r}}_{\mathrm{ML}})^{\mathrm{T}} \frac{\partial}{\partial \widehat{\boldsymbol{r}}_{\mathrm{ML}}} \frac{\partial \mathscr{F}_{\mathrm{ML}}}{\partial \widehat{\boldsymbol{r}}_{\mathrm{ML}}} (\boldsymbol{r}' - \widehat{\boldsymbol{r}}_{\mathrm{ML}}), \quad \text{(B1)}$$

reveals the large-N characteristics that is needed for analysis. The structure of (14), however, demands the operator variation of $\sqrt{A}$ for a positive (semidefinite) A. An integral representation of $\sqrt{A}$ exists [50] and can be written as

$$\sqrt{A} = \lim_{\epsilon \to 0^+} \int_0^\infty \frac{d\,t}{\pi\sqrt{t}} \frac{A}{t + A + \epsilon}, \quad \text{(B2)}$$

where the limit is understood to be applied at the very end of all calculations so that Eq. (B2) is valid even for A with zero eigenvalues.

The first-order variation of $\mathrm{tr}\{\sqrt{A}\}^2$ produces

$$\delta\,\mathrm{tr}\{\sqrt{A}\}^2 = 2\,\mathrm{tr}\{\sqrt{A}\} \lim_{\epsilon \to 0} \int_0^\infty \frac{d\,t}{\pi\sqrt{t}} \mathrm{tr}\left\{ \delta \frac{A}{t + A + \epsilon} \right\}$$
$$= 2\,\mathrm{tr}\{\sqrt{A}\} \lim_{\epsilon \to 0} \int_0^\infty \frac{d\,t}{\pi\sqrt{t}} \mathrm{tr}\left\{ \delta A \frac{1}{t + A + \epsilon} \right.$$
$$\left. - \frac{A}{t + A + \epsilon} \delta A \frac{1}{t + A + \epsilon} \right\}$$
$$= 2\,\mathrm{tr}\{\sqrt{A}\} \lim_{\epsilon \to 0} \mathrm{tr}\left\{ \delta A \frac{A + 2\,\epsilon}{2\,(A + \epsilon)^{3/2}} \right\}. \quad \text{(B3)}$$

In terms of $\mathscr{F}$, we substitute $A \equiv \widehat{\rho}_{\mathrm{ML}}^{1/2}\, \rho'\, \widehat{\rho}_{\mathrm{ML}}^{1/2}$, and evaluate the above result with $\rho' = \widehat{\rho}_{\mathrm{ML}}$, or $A \to A_{\mathrm{ML}} = \widehat{\rho}_{\mathrm{ML}}^2$, then with $\delta A_{\mathrm{ML}} = \widehat{\rho}_{\mathrm{ML}}^{1/2}\, \delta\widehat{\boldsymbol{r}}_{\mathrm{ML}} \cdot \boldsymbol{\Omega}\, \widehat{\rho}_{\mathrm{ML}}^{1/2}$,

$$\frac{\partial \mathscr{F}_{\mathrm{ML}}}{\partial \widehat{\boldsymbol{r}}_{\mathrm{ML}}} = 2 \lim_{\epsilon \to 0} \mathrm{tr}\left\{ \widehat{\rho}_{\mathrm{ML}} \frac{\widehat{\rho}_{\mathrm{ML}}^2 + 2\,\epsilon}{2\,(\widehat{\rho}_{\mathrm{ML}}^2 + \epsilon)^{3/2}} \boldsymbol{\Omega} \right\}, \quad \text{(B4)}$$

where we remind the reader that tr acts on operators only, not on the vectorial character. For case B in which $\widehat{\rho}_{\mathrm{ML}} = \sum_{j=1}^r |\lambda_j\rangle\lambda_j\langle\lambda_j|$ is rank deficient, we get, after taking

the trace,

$$\frac{\partial \mathscr{F}_{\mathrm{ML}}}{\partial \widehat{\boldsymbol{r}}_{\mathrm{ML}}} = \mathrm{tr}\{P_r \, \boldsymbol{\Omega}\}, \tag{B5}$$

where $P_r = \sum_{j=1}^{r} |\lambda_j\rangle\langle\lambda_j|$. It is then trivial to realize that this first-order derivative is zero for case A. Qualitatively, this confirms the fact that when $\widehat{\boldsymbol{r}}_{\mathrm{ML}}$ is an interior point, $\mathscr{F}$ has a local maximum at this point as it should, while a boundary estimator evaluates to a nonzero $\mathscr{F}$ slope.

Upon denoting $\boldsymbol{W}_{\mathrm{ML}} = \widehat{\rho}_{\mathrm{ML}}^{1/2} \, \boldsymbol{\Omega} \, \widehat{\rho}_{\mathrm{ML}}^{1/2}$, the second-order variation follows from the second line of (B3):

$$\delta \frac{\partial \, \mathrm{tr}\{\sqrt{A}\}^2}{\partial \, \boldsymbol{r}'} = 2 \, \delta \Bigg[\mathrm{tr}\{\sqrt{A}\} \lim_{\epsilon \to 0} \int_0^\infty \frac{d\,t}{\pi\sqrt{t}}$$

$$\times \, \mathrm{tr}\Bigg\{ \boldsymbol{W}_{\mathrm{ML}} \frac{t+\epsilon}{(t+A+\epsilon)^2} \Bigg\} \Bigg]. \tag{B6}$$

A product-rule dissociation of (B6) comprises a $\delta \, \mathrm{tr}\{\sqrt{A}\}$ and

$$\lim_{\epsilon \to 0} \int_0^\infty \frac{d\,t}{\pi\sqrt{t}} \, \mathrm{tr}\Bigg\{ \boldsymbol{W}_{\mathrm{ML}} \delta \frac{t+\epsilon}{(t+A+\epsilon)^2} \Bigg\}$$

$$= -\lim_{\epsilon \to 0} \int_0^\infty \frac{d\,t}{\pi\sqrt{t}} \, \mathrm{tr}\Bigg\{ \boldsymbol{W}_{\mathrm{ML}} \Bigg[\frac{t+\epsilon}{t+A+\epsilon} \, \delta A \, \frac{1}{(t+A+\epsilon)^2}$$

$$+ \frac{t+\epsilon}{(t+A+\epsilon)^2} \, \delta A \, \frac{1}{t+A+\epsilon} \Bigg] \Bigg\}. \tag{B7}$$

After evaluating the variation at $\rho' = \widehat{\rho}_{\mathrm{ML}}$ and further undoing all integrations with the help of its spectral decomposition, case B yields

$$\frac{\partial}{\partial \widehat{\boldsymbol{r}}_{\mathrm{ML}}} \frac{\partial \mathscr{F}_{\mathrm{ML}}}{\partial \widehat{\boldsymbol{r}}_{\mathrm{ML}}} = \frac{1}{2} \mathrm{tr}\{P_r \, \boldsymbol{\Omega}\} \mathrm{tr}\{P_r \, \boldsymbol{\Omega}^{\mathrm{T}}\}$$

$$- \sum_{j=1}^{r} \sum_{k=1}^{r} \frac{\langle\lambda_j|\boldsymbol{\Omega}|\lambda_k\rangle \langle\lambda_k|\boldsymbol{\Omega}^{\mathrm{T}}|\lambda_j\rangle}{\lambda_j + \lambda_k}. \tag{B8}$$

The counterpart expression for case A is immediate, of course.

APPENDIX C: HYPERELLIPSOIDAL-CAP AVERAGES

Under the uniform primitive prior, calculations of the hyperellipsoidal-cap integrals

$$I_0 = \int (d\,\boldsymbol{r}'')_{\mathrm{cap}}, \tag{C1}$$

$$\boldsymbol{I}_1 = \int (d\,\boldsymbol{r}'')_{\mathrm{cap}} \, \boldsymbol{r}'', \tag{C2}$$

$$\boldsymbol{I}_2 = \int (d\,\boldsymbol{r}'')_{\mathrm{cap}} \, \boldsymbol{r}''\boldsymbol{r}''^{\mathrm{T}}, \tag{C3}$$

specified by the uniform cap-volume element $(d\,\boldsymbol{r}'')_{\mathrm{cap}} = (d\,\boldsymbol{r}')_{\mathrm{unif}} \, \eta(1 - \boldsymbol{r}''^2) \, \eta(a - \boldsymbol{b}^{\mathrm{T}}\boldsymbol{r}'')$ for $0 \leqslant a \leqslant |\boldsymbol{b}|$ and some column $\boldsymbol{b}$, include systematic manipulations of the double Heaviside functions. One route to take exploits the following integral representation:

$$\eta(x) = \int \frac{d\,t}{2\pi i} \frac{e^{ix}}{t - i\epsilon} \tag{C4}$$

with the implicit limit $\epsilon \to 0^+$. We then have, for (C1),

$$I_0 = \int \frac{d\,t}{2\pi i} \frac{e^{it}}{t - i\epsilon} \int \frac{d\,t'}{2\pi i} \frac{e^{iat'}}{t' - i\epsilon} \int (d\,\boldsymbol{r}'')_{\mathrm{unif}}$$

$$e^{-it\,\boldsymbol{r}''^2 - it'\,\boldsymbol{b}^{\mathrm{T}}\boldsymbol{r}''}$$

$$= \pi^{d/2} \int \frac{d\,t}{2\pi} \frac{e^{it}}{(it)^{d/2+1}} \int \frac{d\,t'}{2\pi} \frac{e^{iat'}}{it'} \, e^{(i/4t)t'^2\boldsymbol{b}^2}, \tag{C5}$$

upon noting the well-known d-dimensional Gaussian integral result

$$\int (d\,\boldsymbol{r}'')_{\mathrm{unif}} \, e^{-\boldsymbol{r}''^{\mathrm{T}}A\boldsymbol{r}'' + \boldsymbol{c}^{\mathrm{T}}\boldsymbol{r}''} = \frac{\pi^{d/2}}{\det\{A\}^{1/2}} \, e^{\boldsymbol{c}^{\mathrm{T}}A^{-1}\boldsymbol{c}/4} \tag{C6}$$

for any positive A. Let us first do the t' integration by invoking the useful transformation

$$\frac{1}{z^m} = \frac{1}{(m-1)!} \int_0^\infty d\,y \, y^{m-1} \, e^{-zy} \quad \text{for } m > 0 : \tag{C7}$$

$$\int \frac{d\,t'}{2\pi} \frac{e^{iat'}}{it'} \, e^{(i/4t)t'^2\boldsymbol{b}^2}$$

$$= \int_0^\infty d\,y \int \frac{d\,t'}{2\pi} \, e^{(i/4t)t'^2\boldsymbol{b}^2 + i(a-y)t'}$$

$$= \sqrt{\frac{it}{\pi\,\boldsymbol{b}^2}} \int_0^\infty d\,y \, e^{-it(a-y)^2/\boldsymbol{b}^2}. \tag{C8}$$

As a consequence,

$$I_0 = \frac{\pi^{(d-1)/2}}{|\boldsymbol{b}|} \int_0^\infty d\,y \int \frac{d\,t}{2\pi} \frac{e^{it}}{(it)^{(d+1)/2}} \, e^{-it(a-y)^2/\boldsymbol{b}^2}$$

$$= \frac{\pi^{(d-1)/2}}{\left(\frac{d-1}{2}\right)!|\boldsymbol{b}|} \int_0^\infty d\,y \int_0^\infty d\,y' \, y'^{(d-1)/2}$$

$$\times \underbrace{\int \frac{d\,t}{2\pi} \, e^{it[1-y'-(y-a)^2/\boldsymbol{b}^2]}}_{= \, \delta(1 - y' - (y-a)^2/\boldsymbol{b}^2)}$$

$$= \frac{\pi^{(d-1)/2}}{\left(\frac{d-1}{2}\right)!|\boldsymbol{b}|} \int_0^{a+|\boldsymbol{b}|} d\,y \, [1 - (y-a)^2/\boldsymbol{b}^2]^{(d-1)/2}. \tag{C9}$$

The above integral in y represents well-known special functions and to see this we further perform the substitutions

$\cos u = (y - a)/|\boldsymbol{b}|$ and $l = a/b$:

$$\int_0^{a+|\boldsymbol{b}|} d\,y\,[1 - (y - a)^2/\boldsymbol{b}^2]^{(d-1)/2}$$

$$= |\boldsymbol{b}| \int_0^{\cos^{-1} l} d\,u\,(\sin u)^d$$

$$= |\boldsymbol{b}|\,\mathrm{B}\left(\frac{1}{2}, \frac{d+1}{2}\right) \mathrm{I}_{(1-l)/2}\left(\frac{d+1}{2}, \frac{d+1}{2}\right), \quad \text{(C10)}$$

which is a product of the beta function and its normalized incomplete form

$$\mathrm{I}_{0 \leqslant a \leqslant 1}(b, c) = \frac{1}{\mathrm{B}(b, c)} \int_0^a d\,u\,u^b\,(1 - u)^c. \quad \text{(C11)}$$

The final answer reads

$$I_0 = V_d\,\mathrm{I}_{(1-l)/2}\left(\frac{d+1}{2}, \frac{d+1}{2}\right). \quad \text{(C12)}$$

For (C2),

$$\boldsymbol{I}_1 = \int \frac{d\,t}{2\pi i} \frac{e^{i\,t}}{t - i\epsilon} \int \frac{d\,t'}{2\pi i} \frac{e^{i\,a\,t'}}{t' - i\epsilon}$$

$$\times \int (d\,\boldsymbol{r}'')_{\mathrm{unif}}\,\boldsymbol{r}''\,e^{-i\,t\,\boldsymbol{r}''^2 - i\,t'\,\boldsymbol{b}^{\mathrm{T}}\boldsymbol{r}''}, \quad \text{(C13)}$$

where the $\boldsymbol{r}''$ integration

$$\int (d\,\boldsymbol{r}'')_{\mathrm{unif}}\,\boldsymbol{r}''\,e^{-i\,t\,\boldsymbol{r}''^2 - i\,t'\,\boldsymbol{b}^{\mathrm{T}}\boldsymbol{r}''}$$

$$= -\frac{1}{i\,t'} \frac{\partial}{\partial\,\boldsymbol{b}'} \int (d\,\boldsymbol{r}'')_{\mathrm{unif}}\,e^{-i\,t\,\boldsymbol{r}''^2 - i\,t'\,\boldsymbol{b}^{\mathrm{T}}\boldsymbol{r}''}$$

$$= -i\,\frac{\pi^{d/2}\,t'}{2\,(i\,t)^{d/2+1}}\,\boldsymbol{b}\,e^{(i/4t)\,t'^2\,\boldsymbol{b}^2} \quad \text{(C14)}$$

is simplified after a differentiation under the integral sign. Then

$$\boldsymbol{I}_1 = -\frac{\pi^{d/2}}{2}\,\boldsymbol{b} \int \frac{d\,t}{2\pi} \frac{e^{i\,t}}{(i\,t)^{d/2+2}} \int \frac{d\,t'}{2\pi}\,e^{(i/4t)\,t'^2\,\boldsymbol{b}^2 + i\,a\,t'}$$

$$- \frac{\pi^{(d-1)/2}}{2} \frac{\boldsymbol{b}}{|\boldsymbol{b}|} \int \frac{d\,t}{2\pi} \frac{e^{i\,t}\,(1 - l^2)}{(i\,t)^{(d+3)/2}}. \quad \text{(C15)}$$

To simplify the t integration, we again recall Eq. (C7) to finally get

$$\boldsymbol{I}_1 = -\frac{\pi^{(d-1)/2}}{2\left(\frac{d+1}{2}\right)!} \frac{\boldsymbol{b}}{|\boldsymbol{b}|}\,\left(1 - l^2\right)^{(d+1)/2}, \quad l = \frac{a}{|\boldsymbol{b}|}. \quad \text{(C16)}$$

We can at least verify the $d = 1$ for Eq. (C16) after paying attention to the convention $\boldsymbol{b} \to -b$, for $b \geqslant 0$. This corresponds to the integral

$$I_{1,d=1} = \int d\,r''\eta(1 - r''^2)\,\eta(a + b\,r'')\,r''$$

$$= \int_{-1}^1 d\,r''\,\eta(a + b\,r'')\,r''$$

$$= \int_{-1}^{-l} d\,r''\,r'' = -\frac{1}{2}(1 - l^2). \quad \text{(C17)}$$

By the same token, we may explore the dyadic integral $\boldsymbol{I}_2$ in (C1) first with (C4) to obtain

$$\boldsymbol{I}_2 = \int \frac{d\,t}{2\pi i} \frac{e^{i\,t}}{t - i\epsilon} \int \frac{d\,t'}{2\pi i} \frac{e^{i\,a\,t'}}{t' - i\epsilon}$$

$$\times \int (d\,\boldsymbol{r}'')_{\mathrm{unif}}\,\boldsymbol{r}''\boldsymbol{r}''^{\mathrm{T}}\,e^{-i\,t\,\boldsymbol{r}''^2 - i\,t'\,\boldsymbol{b}^{\mathrm{T}}\boldsymbol{r}''}, \quad \text{(C18)}$$

where the dyadic $\boldsymbol{r}''$ subintegral

$$\int (d\,\boldsymbol{r}'')_{\mathrm{unif}}\,\boldsymbol{r}''\boldsymbol{r}''^{\mathrm{T}}\,e^{-i\,t\,\boldsymbol{r}''^2 - i\,t'\,\boldsymbol{b}^{\mathrm{T}}\boldsymbol{r}''}$$

$$= -\frac{1}{i\,t} \frac{\delta}{\delta\boldsymbol{A}} \int (d\,\boldsymbol{r}'')_{\mathrm{unif}}\,e^{-it\boldsymbol{r}''^{\mathrm{T}}\boldsymbol{A}\boldsymbol{r}'' - it'\boldsymbol{b}^{\mathrm{T}}\boldsymbol{r}''}\Big|_{\boldsymbol{A}=\boldsymbol{1}}$$

$$= -\frac{\pi^{d/2}}{(i\,t)^{d/2+1}} \frac{\delta}{\delta\boldsymbol{A}} \frac{1}{\det\{\boldsymbol{A}\}^{1/2}}\,e^{(i\,t'^2/4\,t)\boldsymbol{b}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{b}}\Big|_{\boldsymbol{A}=\boldsymbol{1}} \quad \text{(C19)}$$

after an application of Eq. (C6) and a dyadic differentiation under the integral sign this time.

This time, we choose to perform the t and t' integrals before taking the derivative, inasmuch as

$$\boldsymbol{I}_2 = -\frac{\delta}{\delta\boldsymbol{A}} \frac{\pi^{d/2}}{\det\{\boldsymbol{A}\}^{1/2}} \int \frac{d\,t}{2\pi} \frac{e^{i\,t}}{(i\,t)^{d/2+2}}$$

$$\times \int \frac{d\,t'}{2\pi} \frac{e^{i\,a\,t'}}{i\,t'}\,e^{(i\,t'^2/4\,t)\boldsymbol{b}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{b}}\Big|_{\boldsymbol{A}=\boldsymbol{1}}, \quad \text{(C20)}$$

where the usage of (C7) evaluates the t' integral

$$\int \frac{d\,t'}{2\pi} \frac{e^{i\,a\,t'}}{i\,t'}\,e^{(i\,t'^2/4\,t)\boldsymbol{b}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{b}}$$

$$= \int_0^\infty d\,y \int \frac{d\,t'}{2\pi}\,e^{(i\,t'^2/4\,t)\boldsymbol{b}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{b} + i\,t'\,(a - y)}$$

$$= \sqrt{\frac{i\,t}{\pi\,\boldsymbol{b}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{b}}} \int_0^\infty d\,y\,\exp\left(-i\,\frac{(y - a)^2\,t}{\boldsymbol{b}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{b}}\right) \quad \text{(C21)}$$

into another Gaussian integral. Its convenient feature becomes clear when substituted back into (C20):

$$\boldsymbol{I}_2 = -\frac{\delta}{\delta\boldsymbol{A}} \frac{\pi^{(d-1/2)}}{\sqrt{\det\{\boldsymbol{A}\}\boldsymbol{b}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{b}}}$$

$$\times \int \frac{d\,t}{2\pi} \frac{e^{i\,t}}{(i\,t)^{(d+3)/2}} \int_0^\infty d\,y\,\exp\left(-i\,\frac{(y - a)^2\,t}{\boldsymbol{b}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{b}}\right)\Big|_{\boldsymbol{A}=\boldsymbol{1}}$$

$$= -\frac{\delta}{\delta\boldsymbol{A}} \frac{\pi^{(d-1)/2}}{\sqrt{\det\{\boldsymbol{A}\}\boldsymbol{b}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{b}}}$$

$$\times \frac{1}{\left(\frac{d+1}{2}\right)!} \int_0^\infty d\,y \int_0^\infty d\,y'\,y'^{(d+1)/2}$$

$$\times \underbrace{\int \frac{d\,t}{2\pi}\,\exp\left[-i\,t\left(1 - y' - \frac{(y - a)^2\,t}{\boldsymbol{b}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{b}}\right)\right]}_{= \delta\left(1 - y' - \frac{(y - a)^2\,t}{\boldsymbol{b}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{b}}\right)}\Big|_{\boldsymbol{A}=\boldsymbol{1}}$$

$$= -\frac{\pi^{(d-1)/2}}{\left(\frac{d+1}{2}\right)!} \frac{\delta}{\delta\boldsymbol{A}} \frac{1}{\sqrt{\det\{\boldsymbol{A}\}\boldsymbol{b}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{b}}}$$

$$\times \int_0^{a+\sqrt{\boldsymbol{b}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{b}}} dy \left[1 - \frac{(y-a)^2}{\boldsymbol{b}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{b}} \right]^{(d+1)/2}$$

$$= -\frac{\pi^{(d-1)/2}}{\left(\frac{d+1}{2}\right)!} \frac{\delta}{\delta \boldsymbol{A}} \frac{1}{\det\{\boldsymbol{A}\}^{1/2}} \int_0^{\cos^{-1} l_{\boldsymbol{A}}} du \, (\sin u)^{d+2} \bigg|_{\boldsymbol{A}=\mathbf{1}}.$$

(C22)

The end of the tunnel becomes visible after a product-rule differentiation carried out with the basic dyadic identities

$$\delta \boldsymbol{A}^{-1} = -\boldsymbol{A}^{-1}\delta\boldsymbol{A}\,\boldsymbol{A}^{-1},$$

$$\delta \det\{\boldsymbol{A}\} = \det\{\boldsymbol{A}\}\mathrm{Tr}\{\boldsymbol{A}^{-1}\delta\boldsymbol{A}\},$$

(C23)

after which we end up with the final answer,

$$\boldsymbol{I}_2 = \frac{\pi^{(d-1)/2}}{2\left(\frac{d+1}{2}\right)!} \left[\mathrm{B}\left(\frac{1}{2}, \frac{d+3}{2}\right) \mathrm{I}_{(1-l)/2}\left(\frac{d+3}{2}, \frac{d+3}{2}\right)\mathbf{1} \right.$$

$$\left. + l(1-l^2)^{(d+1)/2} \frac{\boldsymbol{b}\,\boldsymbol{b}^{\mathrm{T}}}{\boldsymbol{b}^2} \right].$$

(C24)

The one-dimensional special case can again be extracted from Eq. (C24),

$$\boldsymbol{I}_2\bigg|_{d=1} = \frac{1}{2}\left[\mathrm{B}\left(\frac{1}{2}, 2\right)\mathrm{I}_{(1-l)/2}(2,2) + l(1-l^2) \right]$$

$$= \frac{1}{2}\left[8\int_0^{(1-l)/2} du\, u(1-u) + l(1-l^2) \right]$$

$$= \frac{1}{2}\left[8\left(\frac{l^3}{24} - \frac{l}{8} + \frac{1}{12}\right) + l(1-l^2) \right]$$

$$= \frac{1}{3}(1-l^3),$$

(C25)

and compared with the direct calculation

$$I_{2,d=1} = \int dr''\eta(1-r''^2)\,\eta(a+br'')\,r''^2$$

$$= \int_{-1}^{1} dr''\,\eta(a+br'')\,r''^2$$

$$= \int_{-1}^{-l} dr''\,r''^2 = \frac{1}{3}(1-l^3).$$

(C26)

APPENDIX D: PROJECTION OF UNIFORM AND GAUSSIAN DISTRIBUTIONS ONTO A HYPERELLIPSOID

One may begin with a rotated coordinate system (centered at $\widehat{\boldsymbol{r}}_{\mathrm{ML}}$) that diagonalizes the Fisher information $\boldsymbol{F}_{\mathrm{ML}}$, so that the projected uniform distribution onto the error region in the large-N approximation is calculated from the integral

$$p_{\mathrm{unif}} \propto \int dz\, \eta(1 - ax^2 - by^2 - cz^2)$$

(D1)

for the eigenvalues a, b, and c of $\boldsymbol{F}_{\mathrm{ML}}/(-2\ln\lambda)$. The preceding exercises of Appendix C swiftly give $p_{\mathrm{unif}} \propto \sqrt{1 - ax^2 - by^2}$.

The Gaussian distribution, with covariance chosen to be proportional to $\boldsymbol{F}_{\mathrm{ML}}$ that possesses the eigenvalues a', b', and c', is given by

$$p_{\mathrm{gauss}} \propto \int dz\, \eta(1 - ax^2 - by^2 - cz^2)\, e^{-a'x^2 - b'y^2 - c'z^2}.$$

(D2)

This can be simplified to $p_{\mathrm{gauss}} \propto e^{-a'x^2-b'y^2}\gamma(1/2, c'(1 - ax^2 - by^2)/c)$ in terms of the lower incomplete Gamma function $\gamma(.,.)$ using again results from Appendix C.

APPENDIX E: GRADIENT OPTIMIZATION FOR OBTAINING AN ERROR-REGION INTERIOR POINT

To acquire an interior point of a CR for state tomography, which is essentially a Hilbert subspace, it is sufficient to generate very many ($>D$) region boundary points and take the average of these points. In the limit of large N, we may approximate the inner boundary of the region as part of the hyperellipsoid described by $(\boldsymbol{x}' - \boldsymbol{r}_{\mathrm{c}}) \cdot \boldsymbol{A} \cdot (\boldsymbol{x}' - \boldsymbol{r}_{\mathrm{c}}) \leqslant 1$.

If we perform a variation on the relevant function

$$f = [(\boldsymbol{x}' - \boldsymbol{r}_{\mathrm{c}}) \cdot \boldsymbol{A} \cdot (\boldsymbol{x}' - \boldsymbol{r}_{\mathrm{c}}) - 1]^2,$$

(E1)

where $\boldsymbol{A} = \boldsymbol{F}_{\mathrm{ML}}/(-2\ln\lambda')$, we get

$$\delta f = 4\,[(\boldsymbol{x}' - \boldsymbol{r}_{\mathrm{c}}) \cdot \boldsymbol{A} \cdot (\boldsymbol{x}' - \boldsymbol{r}_{\mathrm{c}}) - 1]\,\delta\boldsymbol{x}' \cdot \boldsymbol{A} \cdot (\boldsymbol{x}' - \boldsymbol{r}_{\mathrm{c}}).$$

(E2)

A gradient method, such as the accelerated projected gradient method, requires the definition of the (operator) gradient defined by $\delta f/\delta\rho'$, which requires the connection $\delta\boldsymbol{x}' = \mathrm{tr}\{\delta\rho'\boldsymbol{\Omega}\}$. Naturally then, we must acquire the resulting operator,

$$\frac{\delta f}{\delta\rho'} = 4\,[(\boldsymbol{x}' - \boldsymbol{r}_{\mathrm{c}}) \cdot \boldsymbol{A} \cdot (\boldsymbol{x}' - \boldsymbol{r}_{\mathrm{c}}) - 1]\,\boldsymbol{\Omega} \cdot \boldsymbol{A} \cdot (\boldsymbol{x}' - \boldsymbol{r}_{\mathrm{c}}),$$

(E3)

where the dot products operate only on the vectorial character only.

The mechanisms that drive the accelerated projected gradient search algorithm are beyond the scope of this article. Instead we provide a simple manual to immediately modify and use the open-source MATLAB code file qse_apg.m [51]. For this purpose, we note the three important variables fval_varrho, fval_new, and gradient, which stores the function values of f evaluated with the varrho and rho_new variables, as well as the gradient operator $\delta f/\delta\rho'$ evaluated with varrho . In order to minimize f with qse_apg.m, one may simply overwrite the existing functional expressions [namely -f'.*log(probs_...) and -qmt(...)] for the three variables with the ones in Eqs. (E1) and (E3). By our numerical experience with this minimization task, it is advisable to set the parameters defaults.threshold_fval and defaults.imax respectively to eps and $>10^{-8}$ for better accuracies.

[1] D. T. Smithey, M. Beck, M. G. Raymer, and A. Faridani, Measurement of the Wigner Distribution and the Density Matrix of a Light Mode using Optical Homodyne Tomography: Application to Squeezed States and the Vacuum, Phys. Rev. Lett. **70**, 1244 (1993).

[2] I. Chuang and M. Nielsen, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, UK, 2000).

[3] J. Řeháček, Z. Hradil, E. Knill, and A. I. Lvovsky, Diluted maximum-likelihood algorithm for quantum tomography, Phys. Rev. A **75**, 042108 (2007).

[4] Y. S. Teo, H. Zhu, B.-G. Englert, J. Řeháček, and Z. Hradil, Quantum-State Reconstruction by Maximizing Likelihood and Entropy, Phys. Rev. Lett. **107**, 020404 (2011).

[5] H. Zhu, Quantum state estimation with informationally overcomplete measurements, Phys. Rev. A **90**, 012115 (2014).

[6] T. D. Ladd, F. Jelezko, R. Laflamme, Y. Nakamura, C. Monroe, and J. L. O'Brien, Quantum computers, Nature (London) **464**, 45 (2010).

[7] E. T. Campbell, B. M. Terhal, and C. Vuillot, Roads towards fault-tolerant universal quantum computation, Nature (London) **549**, 172 (2017).

[8] B. Lekitsch, S. Weidt, A. G. Fowler, K. Mølmer, S. J. Devitt, Ch. Wunderlich, and W. K. Hensinger, Blueprint for a microwave trapped ion quantum computer, Sci. Adv. **3**, e1601540 (2017).

[9] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap* (Chapman & Hall/CRC, New York, 1993).

[10] A. C. Davison and D. V. Hinkley, *Bootstrap Methods and their Application* (Cambridge University Press, Cambridge, UK, 1997).

[11] D. Suess, Ł. Rudnicki, T. O Maciel, and D. Gross, Error regions in quantum state tomography: Computational complexity caused by geometry of quantum states, New J. Phys. **19**, 093013 (2017).

[12] R. Christensen, Testing Fisher, Neyman, Pearson, and Bayes, Am. Stat. **59**, 121 (2005).

[13] M. Christandl and R. Renner, Reliable Quantum State Tomography, Phys. Rev. Lett. **109**, 120403 (2012).

[14] R. Blume-Kohout, Robust error bars for quantum tomography, arXiv:1202.5270.

[15] P. Faist and R. Renner, Practical and Reliable Error Bars in Quantum Tomography, Phys. Rev. Lett. **117**, 010404 (2016).

[16] J. Shang, H. K. Ng, A. Sehrawat, X. Li, and B.-G. Englert, Optimal error regions for quantum state estimation, New J. Phys. **15**, 123026 (2013).

[17] X. Li, J. Shang, H. K. Ng, and B.-G. Englert, Optimal error intervals for properties of the quantum state, Phys. Rev. A **94**, 062112 (2016).

[18] Y. S. Teo, C. Oh, and H. Jeong, Bayesian error regions in quantum estimation I: Analytical reasonings, New J. Phys. **20**, 093009 (2018).

[19] C. Oh, Y. S. Teo, and H. Jeong, Bayesian error regions in quantum estimation II: region accuracy and adaptive methods, New J. Phys. **20**, 093010 (2018).

[20] I. Bengtsson and K. Życzkowski, *Geometry of Quantum States: An Introduction to Quantum Entanglement* (Cambridge University Press, Cambridge, UK, 2006).

[21] I. Bengtsson, S. Weis, and K. Życzkowski, in *Geometric Methods in Physics. XXX Workshop 2011*, edited by P. Kielanowski,

S. T. Ali, A. Odzijewicz, M. Schlichenmaier, and T. Voronov, Trends in Mathematics (Springer, Basel, 2013).

[22] J. Shang, Y.-L. Seah, H. K. Ng, D. J. Nott, and B.-G. Englert, Monte carlo sampling from the quantum state space. I, New J. Phys. **17**, 043017 (2015).

[23] Y.-L. Seah, J. Shang, H. K. Ng, D. J. Nott, and B.-G. Englert, Monte carlo sampling from the quantum state space. II, New J. Phys. **17**, 043018 (2015).

[24] C. Oh, Y. S. Teo, and H. Jeong, Probing Bayesian Credible Regions Intrinsically: A Feasible Error Certification for Physical Systems, Phys. Rev. Lett. **123**, 040602 (2019).

[25] J. C. Butcher, *Numerical Methods for Ordinary Differential Equations.* (John Wiley & Sons, New York, 2003).

[26] S. Kiatsupaibul, R. L. Smith, and Z. B. Zabinsky, An analysis of a variation of hit-and-run for uniform sampling from general regions, ACM Trans. Model. Comput. Simul. **21**, 16 (2011).

[27] R. A. Horn and C. R. Johnson, *Matrix Analysis* (Cambridge University Press, New York, 1985).

[28] N. J. Higham, Cholesky factorization, WIREs Comput. Stat. **1**, 251 (2009).

[29] C. J. P. Bélisle, H. E. Romeijn, and R. L. Smith, Hit-and-run algorithms for generating multivariate distributions, Math. Oper. Res. **18**, 255 (1993).

[30] R. L. Smith, The hit-and-run sampler: a globally reaching markov chain sampler for generating arbitrary multivariate distributions, in *WSC'96 Proceedings of the 28th Conference on Winter Simulation, Coronado, California, USA, 1996*, edited by J. M. Charnes, D. J. Morrice, D. T. Brunner, and J. J. Swain (IEEE Computer Society, Washington, DC, 1996), p. 260.

[31] B. Cousins and S. Vempala, A practical volume algorithm, Math. Prog. Comput. **8**, 133 (2016).

[32] L. Lovász, Hit-and-run mixes fast, Math. Program., Ser. A **86**, 443 (1999).

[33] P. Del Moral, A. Doucet, and A. Jasra, Sequential monte carlo samplers, J. R. Stat. Soc. Ser. B **68**, 411 (2006).

[34] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E Teller, Equations of state calculations by fast computing machines, J. Chem. Phys. **21**, 1087 (1953).

[35] W. K. Hastings, Monte carlo sampling methods using markov chains and their applications, Biometrika **57**, 97 (1970).

[36] A. Uhlmann, The "transition probability" in the state space of a *-algebra, Rep. Math. Phys. **9**, 273 (1976).

[37] D. Bures, An extension of Kakutani's theorem on infinite product measures to the tensor product of semifinite w*-algebra, Trans. Am. Math. Soc. **135**, 199 (1969).

[38] S. L. Braunstein and C. M. Caves, Statistical Distance and the Geometry of Quantum States, Phys. Rev. Lett. **72**, 3439 (1994).

[39] K. Życzkowski and H.-J. Sommers, Hilbert-Schmidt volume of the set of mixed quantum states, J. Phys. A: Math. Gen. **36**, 10115 (2003).

[40] A. Andai, Volume of the quantum mechanical state space, J. Phys. A: Math. Gen. **39**, 13641 (2006).

[41] H. Zhu and B.-G. Englert, Quantum state tomography with fully symmetric measurements and product measurements, Phys. Rev. A **84**, 022327 (2011).

[42] L. Lovász and S. Vempala, Hit-and-run from a corner, SIAM J. Comput. **35**, 985 (2006).

[43] J. Shang, Z. Zhang, and H. K. Ng, Superfast maximum-likelihood reconstruction for quantum tomography, Phys. Rev. A **95**, 062336 (2017).

[44] M. Evans, Measuring statistical evidence using relative belief, Comput. Struct. Biotechnol. J. **14**, 91 (2016).

[45] L. Al-Labadi, Z. Baskurt, and M. Evans, Statistical reasoning: choosing and checking the ingredients, inferences based on a measure of statistical evidence with some applications, Entropy **20**, 289 (2018).

[46] E. Wigner, Characteristic vectors of bordered matrices with infinite dimensions, Ann. Math. **62**, 548 (1955).

[47] E. Wigner, Characteristic vectors of bordered matrices with infinite dimensions II, Ann. Math. **65**, 203 (1957).

[48] A. Pizzo, D. Renfrew, and A. Soshnikov, On finite rank deformations of Wigner matrices, Ann. I. H. Poincare-Pr. **49**, 64 (2013).

[49] M. L. Mehta, *Random Matrices* (Elsevier, Amsterdam, 2004).

[50] D. A. Bini, N. J. Higham, and B. Meini, Algorithms for the matrix pth root, Numer. Algorithms **39**, 349 (2005).

[51] Available on https://github.com/qMLE/qMLE.