Uncorrelated problem-specific samples of quantum states from zero-mean Wishart distributions

Rui Han\textsuperscript{1}, Weijun Li\textsuperscript{1,‡}, Shrobona Bagchi\textsuperscript{1,§}, Hui Khoon Ng\textsuperscript{2,1,3} and Berthold-Georg Englert\textsuperscript{1,3,4}

\textsuperscript{1}Centre for Quantum Technologies, National University of Singapore, Singapore
\textsuperscript{2}Yale-NUS College, Singapore
\textsuperscript{3}MajuLab, CNRS-UCA-SU-NUS-NTU International Joint Research Unit, Singapore
\textsuperscript{4}Department of Physics, National University of Singapore, Singapore

E-mail: han.rui@quantumlah.org

Abstract. Random samples of quantum states are an important resource for various tasks in quantum information science, and samples in accordance with a problem-specific distribution can be indispensable ingredients. Some algorithms generate random samples by a lottery that follows certain rules and yield samples from the set of distributions that the lottery can access. Other algorithms, which use random walks in the state space, can be tailored to any distribution, at the price of autocorrelations in the sample and with restrictions to low-dimensional systems in practical implementations. In this paper, we present a two-step algorithm for sampling from the quantum state space that overcomes some of these limitations.

We first produce a CPU-cheap large proposal sample, of uncorrelated entries, by drawing from the family of complex Wishart distributions, and then reject or accept the entries in the proposal sample such that the accepted sample is strictly in accordance with the target distribution. We establish the explicit form of the induced Wishart distribution for quantum states. This enables us to generate a proposal sample that mimics the target distribution and, therefore, the efficiency of the algorithm, measured by the acceptance rate, can be many orders of magnitude larger than that for a uniform sample as the proposal.

We demonstrate that this sampling algorithm is very efficient for one-qubit and two-qubit states, and reasonably efficient for three-qubit states, while it suffers from the “curse of dimensionality” when sampling from structured distributions of four-qubit states.

Keywords: sampling algorithm, quantum state space, random matrix, Wishart distribution.

Posted on the arXiv on 16 June 2021.

\‡ Now at Trinity Hall, Cambridge University, Cambridge CB3 0D2, UK \(\not\in\) EU
\§ Now at Raymond and Beverly Sackler School of Physics and Astronomy, Tel-Aviv University, Tel-Aviv 69978, Israel
1. Introduction

Random states — states with random properties — play a central role in statistical mechanics because one simply cannot account for the huge number of dynamical degrees of freedom in systems composed of very many constituents. In quantum mechanics — in particular, in quantum information applications — random states also appear naturally even if there are relatively few degrees of freedom. On the one hand, there is always the noisy and uncontrollable interaction of the quantum system under study with its environment; on the other hand, our knowledge about the system is always uncertain because the state parameters cannot be known with perfect precision.

There are many instances in quantum information science where a random sample of quantum states, pure or mixed, is desirable: in the evaluation of entanglement [1–4]; for the numerical testing of a noise model for state preparation or gate operation [5, 6]; when optimizing a function on the state space [4]; for determining the size and credibility of a region in the state space [7–9]; when determining the minimum output entropy of a quantum channel [4]; and in many other applications. While direct sampling, in which an easy-to-generate raw sample is converted into a sample from the desired distribution by an appropriate transformation, is an option for some particular distributions, it does not offer a flexible general strategy.

The seemingly simple requirement that the quantum state is represented by a positive unit-trace Hilbert-space operator translates into complicated conditions obeyed by the variables that parameterize the state, and these constraints must be enforced in the sampling algorithm. In one standard approach, the probability distribution of the quantum states is defined by the sampling algorithm — a lottery for pure system-and-ancilla states, for example, followed by tracing out the ancilla [4, 10, 11] — and this can yield large high-quality samples for the distributions that are accessible in this way. This includes uniform distributions for the Haar measure, the Hilbert–Schmidt measure, the Bures measure, or the measure induced by the partial trace [11–13]. The major drawback is that the exact probability distribution for random states constructed this way are typically not known and one has to rely on the error-prone numerical estimation for an approximated distribution.

Another standard approach generates samples by a random walk in the quantum state space (or, rather, in the parameter space for it). These Monte Carlo (MC) algorithms can accommodate any target distribution by flexible rules for the walk. The well-known Markov Chain MC algorithm can generate samples in the probability simplex, followed by an accept-reject step that enforces the constraints [15, 16]. There is also the Hamiltonian MC algorithm, in which the constraints are obeyed by design, at the price of a rather more complicated parameterization [17, 18]. These random-walk algorithms yield samples with autocorrelations, and the effective sample size can be a

1 The sample used in [14] for demonstrating that local hidden variables are a lost cause is of this lottery-defined kind, too.

2 The samples of quantum channels in [5] are generated in this way.
small fraction. Moreover, we encounter implementation problems when applying the MC algorithm to higher-dimensional system.

In view of these observations, two directions suggest themselves for exploration. One could begin with easy-to-generate and flexible raw samples drawn from a proposal distribution, richer in structure than the uniform samples mentioned above, and then perform suitable acceptance-rejection resampling (or rejection sampling in standard terminology) to obtain a sample in accordance with the target distribution. Or one could process large uncorrelated\(^3\) raw samples by tailored random walks that eventually enforce the constraints. In this work, we explore the first suggestion by adapting the complex Wishart distributions\(^4\) from the statistics literature to suitable proposal distributions of quantum states. The second suggestion is the subject matter of a separate paper where we use a sequentially constraint MC algorithm [26].

The general idea of the first suggestion — generate a sample drawn from the target distribution by rejecting or accepting entries from a larger CPU-cheap raw sample — has many potential implementations, and the target distribution will inform the choice we make for the proposal distribution from which we draw the raw sample. We demonstrate the practical feasibility of the method at the example of target distributions as they arise in the context of quantum state estimation (section 3), and we use raw samples generated from properly adapted Wishart distributions (section 2). The target and proposal distributions used for illustration have single narrow peaks, but that is no limitation as convex sums of several Wishart distributions are proposal distributions for target distributions with more peaks.

More specifically, we introduce the complex Wishart distribution of a quantum state in section 2 and derive its explicit probability distribution, followed by a look at its important properties, such as the peak location and peak shape; then we discuss how to get a useful proposal distribution from it. Section 3 deals with the exemplary target distributions that one meets in quantum state estimation. In section 4, then, we show how rejection sampling converts a sample drawn from the proposal distribution into a sample drawn from the target distribution. We verify the sample so generated by the procedure described in section 5, with technical details delegated to the appendix. After commenting on the limitations that follow from the “curse of dimensionality” in section 6, we present various samples of quantum states in section 7 that illustrate our sampling method. The examples demonstrate that the method is reliable and competitive but does not escape the curse of dimensionality. We close with a summary and outlook in section 8.

\(^3\) The term “uncorrelated” is synonymous to “independent and identically distributed” (i.i.d.), which is standard terminology in statistics.

\(^4\) So named in recognition of John Wishart’s [19] seminal work of 1928 [20]. For important follow-up developments and extensive discussions of Wishart distributions as well as, more generally, of random matrices and positive random matrices as tools for multivariate statistical analysis with real-life applications in, for example, biology, physics, social science, and meteorology, see [21–25].
2. Wishart distributions of quantum states

2.1. The quantum state space

A state $\rho$ of a $m$-dimensional quantum system, represented by a positive-semidefinite unit-trace $m \times m$ matrix, is of the form

$$\rho = \frac{1}{m} 1_m + \sum_{l=1}^{m^2-1} \varrho_l B_l,$$

where $1_m$ is the $m \times m$ unit matrix, the set \{ $B_1, B_2, \ldots, B_{m^2-1}$ \} is an orthonormal basis in the $(m^2 - 1)$-dimensional real vector space of traceless hermitian $m \times m$ matrices,

$$B_l = B_l^\dagger, \quad \text{tr}(B_l) = 0, \quad \text{tr}(B_l B_l^\dagger) = \delta_{ll'},$$

and the $\varrho_l$s are the coordinates for the vector associated with $\rho$. The volume element in this euclidean space,

$$(d\rho) = \prod_{l=1}^{m^2-1} d\varrho_l,$$

is invariant under basis changes; it is independent of the choice made for the $B_l$s. Since $\text{tr}((\rho - \rho')^2) = \sum_l (\varrho_l - \varrho_l')^2$, this is the volume element induced by the Hilbert-Schmidt distance. If we use the first $m - 1$ diagonal matrix elements $\rho_{jj}$ of $\rho$ as well as the real and imaginary parts of the off-diagonal matrix elements, $\rho_{jk} = \rho_{jk}^{(r)} + i\rho_{jk}^{(i)} = \rho_{kj}^*$ for $j < k$, as integration variables, which are coordinates in a basis that is not orthonormal, we have

$$(d\rho) = \left(2^{m(m-1)/2}\right) [d\rho] \quad \text{with} \quad [d\rho] = \prod_{j=1}^{m-1} d\rho_{jj} \prod_{k=j+1}^{m} d\rho_{jk}^{(r)} d\rho_{jk}^{(i)},$$

where the prefactor is the Jacobian determinant. The euclidean space contains quantum states ($\rho \ge 0$) and also unphysical $\rho$s that have negative eigenvalues ($\rho \not\ge 0$). The total volume of the convex quantum state space is stated in (17).

In preparation of the following discussion of sampling $\rho$ from a Wishart distribution, we note that the $m \times m$ matrix for a quantum state can be written as

$$\rho = \frac{\Psi\Psi^\dagger}{\text{tr}(\Psi\Psi^\dagger)} \quad \text{with} \quad \Psi = (\psi_1 \psi_2 \cdots \psi_n),$$

where $\Psi$ is a $m \times n$ matrix composed of $n$ $m$-component columns. If we view the columns of $\Psi$ as representations of pure states, then

$$\rho = \frac{\psi_1^\dagger \psi_1 + \psi_2^\dagger \psi_2 + \cdots + \psi_n^\dagger \psi_n}{\psi_1^\dagger \psi_1 + \psi_2^\dagger \psi_2 + \cdots + \psi_n^\dagger \psi_n}$$

is a mixed state blended from $n$ pure states, as if a pure system-and-ancilla state were marginalized by tracing out the ancilla. The rank of $\rho$ cannot exceed $\min\{m,n\}$ and is usually equal to this minimum when $\Psi$ is chosen “at random,” that is: drawn from a distribution on the state space, as discussed in the subsequent sections. For the purposes of this paper, we take $n \ge m$ for granted.
2.2. The Wishart distributions

2.2.1. Gaussian distribution of complex matrices. Now, to give meaning to the phrase “Ψ is chosen at random,” we regard the real and imaginary parts of the matrix elements \( \Psi_{jk} = \alpha_{jk} + i\beta_{jk} \) of \( \Psi \) as coordinates in a \( 2mn \)-dimensional euclidean space with the volume element

\[
(d\Psi) = \prod_{j=1}^{m} \prod_{k=1}^{n} d\alpha_{jk} d\beta_{jk},
\]

and then have a multivariate, zero-mean, gaussian distribution on this matrix space specified by the probability element

\[
(d\Psi) G(\Psi | \Sigma) = (d\Psi) (2\pi)^{-mn} \det(\Sigma)^{-n} e^{-\frac{1}{2} \text{tr}(\Psi^\dagger \Sigma^{-1} \Psi)},
\]

where \( \Sigma \) is a positive \( m \times m \) matrix — the covariance matrix for each column in \( \Psi \). Denoting the random variable by \( \Psi \) and one of its values by \( \Psi \), we say that \( \Psi \) is drawn from the zero-mean normal distribution of \( m \times n \) matrices with the \( m \times m \) covariance matrix \( \Sigma \) and, adopting the conventional notation from the statistics literature (see [27], for example), write \( \Psi \sim \mathcal{N}_{mn}(0, \mathbf{1}_n \otimes \Sigma) \).

Since, in analogy with (6), we can read the \( n \times n \) trace as

\[
\text{tr}(\Psi^\dagger \Sigma^{-1} \Psi) = \sum_{k=1}^{n} \psi_k^\dagger \Sigma^{-1} \psi_k,
\]

the exponential function in (8) is the product of \( n \) exponential functions, one factor for each column of \( \Psi \). Having noted this factorization, we can state in which sense \( \Sigma \) is the covariance matrix: For each \( m \)-component random column \( \psi \) we have the expected values \( \mathbb{E}(\psi) = 0 \) for the mean and \( \mathbb{E}\left( \left( \psi - \mathbb{E}(\psi) \right) \left( \psi - \mathbb{E}(\psi) \right)^\dagger \right) = \mathbb{E}(\psi \psi^\dagger) = \Sigma \) for the covariance. It follows that \( \mathbb{E}(\Psi \Psi^\dagger) = n\Sigma \).

2.2.2. Wishart distribution of hermitian matrices. The numerator in (5) or (6) is a hermitian \( m \times m \) matrix \( R = \Psi \Psi^\dagger \) with matrix elements \( R_{jk} = x_{jk} + iy_{jk} = R_{kj}^* \) with which we associate the \( m^2 \)-dimensional euclidean space with the volume element

\[
(dR) = \prod_{j=1}^{m} dx_{jj} \prod_{k=j+1}^{m} dx_{jk} dy_{jk}.
\]

As stated in Theorem 5.1 in [21], the induced probability distribution on this \( R \) space has the probability element

\[
(dR) W(R | \Sigma) = (dR) \int (d\Psi) G(\Psi | \Sigma) \delta(R - \Psi \Psi^\dagger)
\]

\[
= (dR) \frac{\det(R)^{n-m}}{2^{mn} \Gamma_m(n) \det(\Sigma)^n} e^{-\frac{1}{2} \text{tr}(\Sigma^{-1} R)}
\]

5 More general gaussian distributions have \( \text{tr}((\Sigma^{-1}_2 (\Psi - \Phi)^\dagger \Sigma^{-1}_1 (\Psi - \Phi)) \) in the exponent, with two covariance matrices \( \Sigma_1 \) and \( \Sigma_2 \) and a displacement matrix \( \Phi \), and then \( \Psi \sim \mathcal{N}_{mn}(\Phi, \Sigma_2 \otimes \Sigma_1) \). We only need the basic gaussian of (8) here, and explore the option of \( \Phi \neq 0 \) in a companion paper [28].
where
\[ \Gamma_m(n) = \prod_{j=0}^{m-1} \pi^j \Gamma(n-j) = \pi^{\frac{m(m-1)}{2}} \prod_{j=1}^{m} (n-j)! \] (12)
is a multivariate gamma function. The random variable \( R \) is drawn from the centered complex Wishart (CCW) distribution of hermitian \( m \times m \) matrices \([20–22]\), \( R \sim \text{CCW}_m(n, \Sigma) \), that derives from \( \mathcal{N}_{mn}(0, 1_n \otimes \Sigma) \).

2.2.3. The quantum Wishart distribution. The conversion of this distribution for \( R \) to a distribution for \( \rho = R / \text{tr}(R) \) requires an integration over \( \tau = \text{tr}(R) \) after expressing the \( R_{jk} \)s in terms of \( \tau \) and the \( \rho_{jk} \)s, omitting \( \rho_{mm} \) which does not appear in (4), that is
\[ R_{jk} = \rho_{jk} \tau \quad \text{for} \quad (j, k) \neq (m, m), \]
\[ R_{mm} = \left( 1 - \sum_{j=1}^{m-1} \rho_{jj} \right) \tau, \] (13)
so that the volume elements in \( R \) space and \( \rho \) space are related by
\[ (dR) = [d\rho] d\tau \tau^{m^2-1}. \] (14)
Accordingly, upon evaluating the \( \tau \) integral in
\[ (d\rho) g(\rho) = [d\rho] \int_0^\infty d\tau \tau^{m^2-1} W(R = \tau\rho|\Sigma) \] (15)
we arrive at the following Theorem: \[ \]
For \( \Psi \sim \mathcal{N}_{mn}(0, 1_n \otimes \Sigma) \) with \( n \geq m \), the \( m \times m \) matrices
\[ \rho = \frac{\Psi \Psi^\dagger}{\text{tr}(\Psi \Psi^\dagger)} \] are drawn from the Wishart distribution of \( m \)-dimensional quantum states, \( \rho \sim \mathcal{W}_m^Q(n, \Sigma) \), which is specified by the probability element
\[ (d\rho) g(\rho) = [d\rho] \frac{\Gamma(mn)}{\Gamma_m(n)} \frac{\det(\rho)^{n-m}}{\det(\Sigma)^n \text{tr}(\Sigma^{-1}\rho)^{mn}}. \] (16)
By construction, \( g(\rho) = 0 \) for all unphysical \( \rho \)s. For the sake of notational simplicity, we leave the dependence of \( g(\rho) \) on \( m, n, \) and \( \Sigma \) implicit, just as we do not indicate the dependence of \( G(\Psi|\Sigma) \) on \( m \) and \( n \) in (8).

We observe that, when drawing a \( \Psi \) from the distribution \( G(\Psi|\Sigma) \) to arrive at a quantum state \( \rho \) drawn from the distribution \( g(\rho) \), it does not matter if we replace \( \Sigma \) by \( \lambda \Sigma \) with a positive scaling factor \( \lambda \) because the product \( \det(\lambda \Sigma) \text{tr}((\lambda \Sigma)^{-1}\rho)^m \) does not depend on \( \lambda \). We write \( \Sigma' = \Sigma \) for two \( \Sigma \)s that are equivalent in this sense.

When \( \Sigma = 1_m \), the probability distribution \( g(\rho) \) is isotropic in the sense of \( g(U\rho U^\dagger) = g(\rho) \) for all unitary \( m \times m \) matrices \( U \). In the case of a generic \( \Sigma > 0 \), the distribution is invariant under \( \rho \rightarrow U\rho U^\dagger \) if \( U \) commutes with \( \Sigma \), and only then.

---

6. When \( n < m \), all \( R \)s are rank deficient and the anti-Wishart distribution is obtained instead \([29,30]\). We do not consider samples of rank-deficient matrices here.

7. The case of \( \Sigma = 1_m \), when \( g(\rho) \propto \det(\rho)^{n-m} \) accounts in full for the \( \rho \) dependence, is well known; see, for example, (15.58) in \([11]\) or (3.5) in \([31]\).
2.2.4. Uniformly distributed quantum states. A particular case is that of \( n = m \) and \( \Sigma = \mathbf{1}_m \), when \( g(\rho) \) is constant and equal to the reciprocal of the volume occupied by the quantum states in the \((m^2 - 1)\)-dimensional euclidean space,

\[
\int_{(\rho \geq 0)} (d\rho) = \left(2^{m(m-1)/2}\Gamma(m)\Gamma(m^2)\right)^{-1/2}.
\]

Here, the \( \rho \)'s are uniformly distributed over the state space — uniform in the sense of the Hilbert–Schmidt distance. This provides a computationally cheap and reliable way of generating an uncorrelated, uniform sample of quantum states of this kind.

2.3. Efficient sampling

In addition to unitary transformations of \( \rho \), we can also consider general hermitian conjugations, which invites the following question: If \( \rho \) is drawn from the distribution \( g(\rho) \), what is the corresponding distribution for \( \rho' = A\rho A^\dagger \)\( \text{tr}(A\rho A^\dagger) \),\( (18) \)

where \( A \) is a nonsingular \( m \times m \) matrix? Clearly, the sample of \( \rho \)'s can be generated by the replacement \( \Psi \rightarrow \Psi' = A\Psi \) for each \( \Psi \) drawn from \( G(\Psi|\Sigma) \). We write \( A \) in its polar form,

\[
A = HU,
\]

where \( U \) is a unitary matrix and \( H = (AA^\dagger)^{1/2} \) is a positive matrix, and note that

\[
(d\Psi') = (d\Psi) \det(H)^{2n} \quad (20)
\]

for the respective volume elements in (7). Then,

\[
(d\Psi) G(\Psi|\Sigma) = (d\Psi') \det(H)^{-2n}G(A^{-1}\Psi'|\Sigma) = (d\Psi') G(\Psi'|A\Sigma A^\dagger),
\]

so that \( \Psi' \) is drawn from the multivariate gaussian distribution with the covariance matrix \( A\Sigma A^\dagger \). Here, then, is the answer to the question above: \( \rho' \) is drawn from \( g(\rho') \) with \( \Sigma \) replaced by \( A\Sigma A^\dagger \) in (16), that is \( \rho' \sim W_m^{(Q)}(n, A\Sigma A^\dagger) \). Note that \( W_m^{(Q)}(n, A\Sigma A^\dagger) = W_m^{(Q)}(n, \Sigma) \) when \( A\Sigma A^\dagger = \Sigma \), that is: \( A = \lambda^{1/2} \Sigma^{1/2} U \Sigma^{-1/2} \) with \( \lambda \) a positive number and \( U \) a unitary matrix. This includes the case, discussed above, of a unitary \( A \) that commutes with \( \Sigma \).

This observation is of practical importance. We can sample \( \Psi \) from \( G(\Psi|\mathbf{1}_m) \) — drawing both the real and the imaginary parts of all matrix elements \( \Psi_{jk} \) from the one-dimensional gaussian distribution with zero mean and unit variance, which can be done very efficiently — and then put \( \rho = A\Psi\Psi^\dagger A^\dagger / \text{tr}(A^\dagger A\Psi\Psi^\dagger) \) with \( A \) such that \( \Sigma = AA^\dagger \); the choice \( A = \Sigma^{1/2} \) suggests itself. This yields a random sample of uncorrelated \( \rho \)'s drawn from \( g(\rho) \).

\( ^8 \) It appears that, for many authors, uniformity in the Hilbert–Schmidt distance is the default meaning of “picking a quantum state at random;” an example is the marginalization performed in [32].

\( ^9 \) This is the statement of Theorem 1.2.6 in [23].
2.4. Peak location

In (16), we have
\[ g(\rho) \propto \frac{\det(\rho_\Sigma)^n}{\det(\rho)^m} \quad \text{with} \quad \rho_\Sigma = \frac{\Sigma^{-\frac{1}{2}} \rho \Sigma^{-\frac{1}{2}}}{\text{tr}(\Sigma^{-1}\rho)}, \]
so that \( g(\rho) \) is peaked at the \( \rho \) that compromises between maximizing \( \det(\rho_\Sigma) \) and minimizing \( \det(\rho) \). Since the response of \( \det(\rho) \) to an infinitesimal change of \( \rho \) is
\[ \delta \det(\rho) = \det(\rho) \text{tr}(\rho^{-1} \delta \rho), \]
the stationarity requirement \( g(\rho_{\text{peak}} + \delta \rho) = g(\rho_{\text{peak}}) \) reads
\[ \text{tr}\left( \left[ (n-m)\rho_{\text{peak}}^{-1} - \frac{mn\Sigma^{-1}}{\text{tr}(\Sigma^{-1}\rho_{\text{peak}})} \right] \delta \rho \right) = 0 \]
for all permissible \( \delta \rho \)s. It follows that we need\(^{10}\)
\[ \Sigma = \left( \rho_{\text{peak}}^{-1} + \frac{m^2}{n-m} \mathbf{1}_m \right)^{-1} \]
if we want \( g(\rho) \) to be largest at \( \rho = \rho_{\text{peak}} \),
\[ \max_{\rho} \{ g(\rho) \} = g(\rho_{\text{peak}}). \]

Clearly, for every full-rank \( \rho_{\text{peak}} \), there are probability densities \( g(\rho) \) that have their maximum there, one such \( g(\rho) \) for each \( n \) value larger than \( m \).

The covariance matrix (25) applies for \( n > m \) and a full-rank \( \rho_{\text{peak}} \), as \( g(\rho) = 0 \) for a rank-deficient \( \rho \) when \( n > m \). If \( n = m \), we get \( \Sigma = \mathbf{1}_m \) for all full-rank \( \rho_{\text{peak}} \)s, which is the \( g(\rho) = \text{constant} \) situation noted above.

For \( n = m \) and \( \Sigma \neq \mathbf{1}_m \), \( g(\rho) \) is largest when \( \text{tr}(\Sigma^{-1}\rho) \) is smallest and, therefore, the eigenvalue-subspace of \( \Sigma \) for its largest eigenvalue comprises all \( \rho_{\text{peak}} \)s. While this enables us to design \( g(\rho) \) such that it is maximal for a rank-deficient \( \rho_{\text{peak}} \), it is of little practical use; more about this in section 3.2.

2.5. Peak shape

For \( \rho \)s near \( \rho_{\text{peak}} \),
\[ \rho = \rho_{\text{peak}} + \epsilon = \rho_{\text{peak}} + \sum_{l=1}^{m^2-1} \varepsilon_l B_l, \]
where the traceless hermitian \( m \times m \) matrix \( \epsilon \) and its real coordinates \( \varepsilon_l \) are small on the relevant scales, we have
\[ \log \frac{g(\rho_{\text{peak}} + \epsilon)}{g(\rho_{\text{peak}})} \approx -\frac{1}{2} \left( n-m \right) \left[ \text{tr}\left( \left( \rho_{\text{peak}}^{-1} \epsilon \right)^2 \right) - \frac{n-m}{mn} \text{tr}(\rho_{\text{peak}}^{-1} \epsilon)^2 \right] \]
\[ = -\frac{1}{2} \sum_{l,l'} \varepsilon_l G_{ll'} \varepsilon_{l'} \]
\[ \text{(28)} \]
\(^{10}\) For a hermitian \( m \times m \) matrix \( A_\rho \) that is a functional of \( \rho \), the requirement that \( \text{tr}(A_\rho \delta \rho) = 0 \) holds for all permissible \( \delta \rho \)s implies \( A_\rho \rho = \rho A_\rho = \rho \text{tr}(A_\rho \rho) \).
with
\[
G_{W} = (n - m) \text{tr}(\rho_{\text{peak}}^{-1}B_{l}\rho_{\text{peak}}^{-1}B_{l}) - \frac{(n - m)^{2}}{mn} \text{tr}(\rho_{\text{peak}}^{-1}B_{l}) \text{tr}(\rho_{\text{peak}}^{-1}B_{l})
\]  
(29)

upon discarding terms proportional to the third and higher powers of \(\epsilon\). For the gaussian approximation of \(g(\rho)\) in the vicinity of \(\rho_{\text{peak}}\) in (28), the inverse of the matrix \(G\) is the covariance matrix in the coordinate space. It follows that the distribution \(g(\rho)\) will be particularly narrow in the directions associated with the smallest eigenvalues of \(\rho_{\text{peak}}\).

We break for a quick look at one-qubit and multi-qubit examples and then pick up the story in section 2.8.

2.6. Example: Wishart samples for a qubit

The qubit case \((m = 2)\) is the one case that we can visualize. The state space is the unit Bloch ball [33] with cartesian coordinates \(x, y, z\) introduced in the standard way,
\[
\rho = \frac{1}{2} \begin{pmatrix} 1 + z & x - iy \\ x + iy & 1 - z \end{pmatrix} = \frac{1}{2} (1_{2} + x\sigma_{x} + y\sigma_{y} + z\sigma_{z}) = \frac{1}{2} (1_{2} + (\sigma \cdot \sigma)),
\]
\[
(d\rho) = \frac{1}{\sqrt{8}} dx\,dy\,dz, \quad \det(\rho) = \frac{1}{4} (1 - x^{2} - y^{2} - z^{2}),
\]  
(30)

where \(\sigma_{x}, \sigma_{y}, \sigma_{z}\) are the standard Pauli matrices, the cartesian components of the vector matrix \(\sigma\), and \(x, y, z\) are their respective expectation values, such as \(x = \langle \sigma_{x} \rangle = \text{tr}(\rho\sigma_{x})\), with \(x^{2} + y^{2} + z^{2} \leq 1\). Any choice of right-handed coordinate axes is fine (proper rotations in the three-dimensional euclidean space are unitary transformations of the \(2 \times 2\) matrices) and, by convention, we choose the coordinate axes such that \(\Sigma\) is diagonal,
\[
\Sigma = \begin{pmatrix} e^{\vartheta} & 0 \\ 0 & e^{-\vartheta} \end{pmatrix}, \quad \det(\Sigma) \text{tr}(\Sigma^{-1}\rho)^{2} = (\cosh \vartheta - z \sinh \vartheta)^{2}.
\]  
(31)

Then we have
\[
(d\rho) g(\rho) = dx\,dy\,dz\, \frac{1}{2\pi} \frac{n - 1}{4^{n-1}} \frac{\Gamma(2n)}{\Gamma(n)^{2}} (\cosh \vartheta - z \sinh \vartheta)^{2n-2} \frac{1}{2^{n-1}} \frac{\Gamma(2n)}{\Gamma(n)^{2}} (1 - z^{2})^{n-1} \frac{1}{2^{n-1}} \frac{\Gamma(2n)}{\Gamma(n)^{2}} (\cosh \vartheta - z \sinh \vartheta)^{2n},
\]  
(32)

where \(s, \phi\) are polar coordinates in the unit disk and \((x, y) = \sqrt{1 - z^{2}} (s \cos \phi, s \sin \phi)\). In view of the factorization of the \(s, \phi, z\) version, the probability \((d\rho) g(\rho)\) is the product of its three marginal probabilities, which tells us that \(s, \phi, z\) are independent random variables.\(^{11}\) This \(g(\rho)\) is largest for \(\rho_{\text{peak}}\) with \((x, y, z)_{\text{peak}} = (0, 0, z_{\text{peak}})\) where \(z_{\text{peak}}\) is

\(^{11}\) Note that the substitution \(z = (z' + \tanh \vartheta)/(1 + z' \tanh \vartheta)\) yields a \(z'\) marginal \(\propto dz' (1 - z'^{2})^{n-1}\), that does not depend on \(\vartheta\). Similarly, the substitution \(x = x' / \sqrt{(\cosh \vartheta)^{2} - (x' \sinh \vartheta)^{2}}\) yields a \(\vartheta\)-independent \(x'\) marginal \(\propto dx' (1 - x'^{2})^{n-1}\), and likewise for \(y\).
given by
\[ \tanh \vartheta = \frac{(n - 2)z_{\text{peak}}}{n - 2z_{\text{peak}}^2}, \]  
(33)
whereas the \( z \)-marginal is largest for the \( z \) that solves
\[ \tanh \vartheta = \frac{(n - 1)z}{n - z^2}, \]  
(34)
which gives a value between \( z = 0 \) and \( z_{\text{peak}} \).

For the most natural choice of orthonormal basis matrices — that is \( B_1 = 2^{-\frac{1}{2}} \sigma_x \), \( B_2 = 2^{-\frac{1}{2}} \sigma_y \), and \( B_3 = 2^{-\frac{1}{2}} \sigma_z \) — we have
\[ \varepsilon = \frac{1}{\sqrt{2}} \begin{pmatrix} \varepsilon_3 & \varepsilon_1 - i\varepsilon_2 \\ \varepsilon_1 + i\varepsilon_2 & -\varepsilon_3 \end{pmatrix} \]  
(35)
here, and find
\[ \log \frac{g(\rho_{\text{peak}} + \varepsilon)}{g(\rho_{\text{peak}})} \cong -\frac{1}{2} \frac{4(n - 2)}{1 - \varepsilon_{\text{peak}}^2} \left( \varepsilon_1^2 + \varepsilon_2^2 + \frac{1}{1 - \varepsilon_{\text{peak}}^2} \varepsilon_3^2 \right) \]  
(36)
near the peak. The longitudinal variance (coordinate \( \varepsilon_3 \)) is smaller by the factor \((1 - \varepsilon_{\text{peak}}^2) / (1 + \frac{2}{n} \varepsilon_{\text{peak}}^2)\) than the transverse variance; all variances decrease when \( n \) and \( z_{\text{peak}} \) increase.

These matters are illustrated in figure 1 for \( n = 5 \) and \( z_{\text{peak}} = \frac{4}{5} \), where we have scatter plots of \( g(\rho) \) and of its \( x, z \) and \( x, y \) marginal distributions, of 10000 quantum states each. There are also histograms, from a sample of 100000 quantum states, for the one-dimensional marginal distributions in \( x \) and \( z \), which we compare with the histograms computed from the analytical expressions. The sample marginals are respectively obtained by ignoring the Bloch-ball coordinate \( y \), the coordinate \( z \), the coordinate pair \((y, z)\), or the coordinate pair \((x, y)\) of the sample entries. This deliberate ignorance corresponds to integrating \( g(\rho) \) over these variables. Note, in particular, the very good agreements of the sample histograms with the analytical ones, which is partial confirmation that the sampling algorithm of sections 2.2 and 2.3 does indeed yield a sample in accordance with the Wishart distribution in (32).

2.7. Example: Multi-qubit analog

As one multi-qubit example, we consider an analog of (31)–(33) for \( n_{\text{qb}} \) qubits,
\[ \Sigma = \cosh \vartheta \mathbf{1}_m + \sinh \vartheta \sigma_z \otimes^n_{\text{qb}}, \]
\[ \rho_{\text{peak}} = \frac{1}{m} \left( \mathbf{1}_m + z_{\text{peak}} \sigma_z \otimes^n_{\text{qb}} \right), \]  
(37)
with \( m = 2^n_{\text{qb}} \) and \( \vartheta \) related to \( z_{\text{peak}} \) through the analog of (33) with “2” replaced by \( m \). As observed in (29), different values of \( z_{\text{peak}} \) do not just give different peak locations, they also give different shapes of the distribution. We look at \( g(\rho) \) for \( \rho = \rho_{\text{peak}} + \varepsilon m^{-\frac{1}{2}} \sigma_z \otimes^n_{\text{qb}}, \) where \( \varepsilon \) is the coordinate in the longitudinal direction. In other words, \( \rho \) is of the form of \( \rho_{\text{peak}} \) in (37) with \( \varepsilon m^{-\frac{1}{2}} \) added to \( z_{\text{peak}} \).
We quantify the width of the distribution, as a function of ε, by the full width at half maximum (FWHM). The approximate value that follows from the gaussian approximation in (28),

\[
\text{FWHM} \cong \frac{2}{m} \left(1 - z_{\text{peak}}^2\right) \left[\frac{\log(4)}{(n-m)(1 + \frac{m}{n} z_{\text{peak}}^2)}\right]^{\frac{1}{2}},
\]

catches the dependence on \(z_{\text{peak}}\) and \(n\) well. It slightly overestimates FWHM for small \(z_{\text{peak}}\) values, is very close to the actual FWHM for \(z_{\text{peak}} \approx 0.5\), and slightly underestimates FWHM for large \(z_{\text{peak}}\) values. The relative error is in the range \(-4\% \cdots 2\%\) when \(n > m + 16/m\); this is illustrated in figure 2.

2.8. Linearily shifted distributions

After choosing \(\rho_{\text{peak}}\) and thus \(\Sigma\) such that the peak shape of the Wishart distribution \(g(\rho)\) fits to that of the target distribution, in the sense discussed at the end of section 3, the two peak shapes are matched. The peak locations usually do not agree, however.
The proposal distribution is, therefore, obtained by shifting the Wishart distribution \( g(\rho) \). After acquiring a sample drawn from \( W_{m}^{(Q)}(n, \Sigma) \), we turn each \( \rho' \) from the sample into \( \rho \) by the linear shift map
\[
\rho = \rho' + \Delta \rho \quad \text{with} \quad (d\rho) = (d\rho'),
\]
where \( \Delta \rho \) is a traceless hermitian \( m \times m \) matrix that we choose suitably. The distribution of the \( \rho \) sample then peaks at \( \rho = \rho_{\text{peak}} + \Delta \rho \) and is characterized by the probability element
\[
(d\rho) g_s(\rho) = [d\rho] \frac{\Gamma(mn)}{\Gamma_m(n)} \frac{\det(\rho - \Delta \rho)^{n-m}}{\det(\Sigma)^n \text{tr}(\Sigma^{-1}(\rho - \Delta \rho))^{mn}} .
\]
Since the mapping (39) linearly shifts the entire state space, there are \( \rho \)s with no preimage \( \rho' \) in the state space, so that \( g_s(\rho) = 0 \) in the corresponding part of the state space. There are also unphysical \( \rho \)s with \( g_s(\rho) > 0 \); they will be discarded during the rejection sampling discussed in section 4. The matter is illustrated in figure 3. While the fraction of physical \( \rho \)s in the shifted distribution \( g_s(\rho) \) is quite large (\( \gtrsim 0.6 \)) for a single qubit even when the shift is by half of the radius of the Bloch ball, this fraction is quite small for several-qubit distributions unless the shift itself is small enough. Therefore, one needs to compromise between \( \Sigma, n, \) and \( \Delta \rho \) and exploit (25) when choosing \( \Sigma \) and \( n \) such that \( \Delta \rho \) is kept small.

The rejection sampling of section 4 requires a proposal distribution that is strictly positive and, therefore, we need to fill the void created by the lack of \( \rho \)'s for certain \( \rho \)s. For this purpose, we supplement the sample obtained by the shift (39) with \( \rho \)s drawn from the uniform distribution \( W_{m}^{(Q)}(m, 1_m) \); this has no effect on the peak shape.

**Figure 2.** Accuracy of the approximation (38). For one to seven qubits \( (m = 2^{n_{qb}} = 2, 4, 8, \ldots, 128) \), the plot reports the relative error for \( n = \lfloor m + 16/m \rfloor + 1 \) as a function of \( z_{\text{peak}} \); the error is smaller for larger \( n \) values. Note that the curves for \( n_{qb} = 4 \) and \( n_{qb} = 5 \) are indiscernible for \( z_{\text{peak}} \lesssim 0.5 \).
or the peak location. The full distribution of the proposal sample is then a convex sum
of \( g_s(\rho) \) and \( g(\rho) = \text{constant} \). When \( 0 < \kappa < 1 \) is the weight assigned to the uniformly
distributed subsample, and \( N \) is the desired sample size, we repeat the sequence
\[
\begin{align*}
&- \text{draw } \Psi \text{ from } G(\Psi|1_m) ; \\
&- \text{compute } \rho = \frac{\Sigma^{\frac{1}{2}} \Psi \Psi^{\dagger} \Sigma^{\frac{1}{2}}}{\text{tr}(\Sigma \Psi \Psi^{\dagger})} + \Delta \rho ; \\
&\text{(41)}
\end{align*}
\]
until we have \((1 - \kappa)N\) entries in the sample. Then we add \( \kappa N \) \( \rho \)'s from the uniform
distribution to complete the proposal sample.

The proposal sample produced in this way is drawn from a distribution with the
probability element
\[
(\text{d}\rho) g_{s,\kappa}(\rho) \propto [\text{d}\rho] \left(1 - \kappa\right) \frac{\Gamma(mn)}{\Gamma_m(n)} \frac{\det(\rho - \Delta \rho)^{n-m}}{\det(\Sigma)^n \text{tr}(\Sigma^{-1}(\rho - \Delta \rho))^{mn}} \\
+ [\text{d}\rho] \kappa \frac{\Gamma(m^2)}{\Gamma_m(m)} ,
\]
where we write “\( \propto \)” rather than “\( = \)” because we are missing the overall factor that
normalizes \( g_{s,\kappa}(\rho) \) to unit integral over the physical \( \rho \)'s. The proposal distributions of
this kind are nonzero for rank-deficient \( \rho \)'s and the samples can have many \( \rho \)'s near a part
of the state-space boundary if \( \rho_{\text{peak}} + \Delta \rho \) is close to the boundary. This is a useful feature
when we need to mimic a target distribution that peaks near or on the boundary, and
the proposal sample should have many points near the boundary where the unshifted
\( g(\rho) \) vanishes when \( n > m \); see section 7 for examples. Such target distributions assign
very little probability to the void region of the shift, and then we may not need many
\( \rho \)'s from the uniform distribution in the proposal sample.
The rejection sampling discussed in section 4 has a high yield only if the proposal
distribution matches the target distribution well and, therefore, a judicious choice of the
parameters that define the sample is crucial for the overall efficiency of the sampling
algorithm. In practice, we choose \( n, \Sigma, \Delta \rho, \kappa \) and generate a small sample, just large
enough for a good estimate of the overall yield. After trying out several choices, we
generate the actual large sample for the \( n, \Sigma, \Delta \rho, \kappa \) with the largest yield. Then we
verify the sample by the procedure described in section 5.

3. The target distributions

The target distribution — the probability distribution on the quantum state space from
which we want to sample — can be of any form. An important use of quantum samples
is in Bayesian analysis, where integrals over the high-dimensional quantum state space
are central quantities, which can often only be computed by Monte Carlo integration.
The samples are then drawn from the respective prior distribution, or the posterior
distribution after accounting for the knowledge provided by the experimental data. The
target distributions that we use for illustration are of this kind. The sampling algorithm
can, of course, also be used for other target distributions.

3.1. Target distributions that arise in quantum state estimation

To be specific, the examples used for illustration in section 7 refer to the scenario of
quantum state estimation where many uncorrelated copies of the system are measured
by an apparatus and one of the \( K \) outcomes is found for each copy. The sequence of
detection events constitute the data \( D \) and the probability of observing the actual data,
if the system is prepared in the state \( \rho \), is the likelihood

\[
L(D|\rho) = \prod_{k=1}^{K} p_{\nu_k}^{\nu_k},
\]

where \( p_k \) is the probability of observing the \( k \)th outcome and \( \nu_k \) is the count of \( k \)th
outcomes in the sequence; the counts themselves, \( \nu = (\nu_1, \nu_2, \nu_3, \ldots, \nu_K) \), are a minimal
statistic. The dependence on \( \rho \) is given by the Born rule,

\[
p_k = \text{tr}(\rho \Pi_k),
\]

with the probability operator \( \Pi_k \) for the \( k \)th outcome. The \( \Pi_k \)s are nonnegative and
add up to the identity,

\[
\Pi_k \geq 0, \quad \sum_{k=1}^{K} \Pi_k = 1,
\]

but are not restricted otherwise. The permissible probabilities \( p = (p_1, p_2, \ldots, p_K) \) are
those consistent with \( \rho \geq 0 \) and \( \text{tr}(\rho) = 1 \), they are a convex subset in the probability
simplex — identified by \( p_k \geq 0 \) and \( \sum_{k=1}^{K} p_k = 1 \) — and it can be CPU-time expensive
to check if a certain \( p \) is permissible.
The target distribution $f(\rho)$ is the posterior distribution, the product of the prior probability density $w_0(\rho)$ and the likelihood,
\[
f(\rho) \propto L(D|\rho)w_0(\rho),
\] (46)
where we do not write the proportionality constant that ensures proper normalization. Usually, it is expedient to use a conjugate prior, that is: $w_0(\rho)$ is of the product form in (43) with prechosen values for the $\nu_k$s, which need not be integers, and then the posterior is also of this form. For the purposes of this paper, therefore, we can just use a flat prior, $w_0(\rho) = \text{constant}$, for which the target distribution is the likelihood in (43), possibly with noninteger values for the $\nu_k$s, properly normalized to unit total probability when integrated over the $\rho$ space. With the volume element $(d\rho)$ of (3) and (4), the probability element of the target distribution is then
\[
(d\rho) f(\rho) \propto (d\rho) \prod_{k=1}^{K} \text{tr}(\Pi_k\rho)^{\nu_k}
\] (47)
for the physical values of the integration variables — those for which $\rho \geq 0$ in (1), the convex set of quantum states — and $f(\rho) = 0$ for all unphysical $\rho$s. Accordingly, the normalization integral
\[
\int (d\rho) f(\rho) = \int (d\rho) f(\rho) = 1
\] (48)
has no contribution from unphysical $\rho$s.

The particular form of the target distribution, proportional to the likelihood in (43), invites us to regard it as a function on the probability simplex, a polynomial on the physical subset of permissible probabilities and vanishing outside. Indeed, the sampling algorithms in [16, 18] are random walks in the probability simplex, and these algorithms have the practical limitations mentioned above: Either one needs to check if a candidate entry for the sample is permissible, which has high CPU-time costs; or one needs a complicated parameterization of the probability space and then faces issues with the large Jacobian matrix, its determinant, and its derivatives. Therefore, alternative algorithms will be useful, such as schemes that directly generate samples from the quantum state space rather than the probability space associated with the $\Pi_k$s. It is the aim of this paper to contribute such an alternative algorithm.

We emphasize a crucial difference between the target distribution in (47) and the proposal distribution in (42): While both have analytical expressions, which will be important in section 4, the proposal distribution is defined by its sampling algorithm, whereas we do not know an analogous sampling algorithm for the target distribution. Therefore, we cannot sample from the target distribution in a direct way and must resort to processing samples drawn from the proposal distribution.
3.2. Peak location

The target distribution \( f(\rho) \propto L(D|\rho) \) is peaked at \( \rho_{\text{ML}} \), given by

\[
\max_\rho \{ L(D|\rho) \} = L(D|\rho_{\text{ML}}),
\]

which is to say that \( \rho_{\text{ML}} \) is the maximum-likelihood estimator for the data \( D \) \([34, 35]\). While it is possible that \( L(D|\rho) \) is maximal for a multidimensional set of \( \rho_{\text{ML}} \)s, there is a unique \( \rho_{\text{ML}} \) for each of the target distributions that we use for illustration. If the relative frequencies \( \tilde{\nu}_k = \nu_k / \sum_{k'} \nu_{k'} \) are permissible probabilities, then \( \tilde{\nu}_k = \text{tr}(\Pi_k \rho_{\text{ML}}) \) identifies \( \rho_{\text{ML}} \). Otherwise one needs to determine \( \rho_{\text{ML}} \) numerically, perhaps by the fast algorithm of [36], and one may find a rank-deficient \( \rho_{\text{ML}} \) on the boundary of the state space. When \( \rho_{\text{ML}} \) is on, or near to, the boundary it is usually more difficult to sample from the state space in accordance with the target distribution.\(^{12}\)

Now, harking back to the final paragraph in section 2.4, we note that the option of matching a rank-deficient \( \rho_{\text{ML}} \) by the \( \rho_{\text{peak}} \) of a \( n = m \) Wishart distribution does not work well usually, because \( \mathcal{W}_m^{(Q)}(m, \Sigma) \) is then maximal for all \( \rho \) in the range of \( \rho_{\text{ML}} \). Therefore, we cannot get a good match when the rank of \( \rho_{\text{ML}} \) is larger than one.

When \( \rho_{\text{ML}} \) has full rank, the choice \( \Delta \rho = \rho_{\text{ML}} - \rho_{\text{peak}} \) suggests itself for the shift in section 2.8. This is not a viable option, however, when \( \rho_{\text{ML}} \) is rank deficient, a typical situation when the \( \nu_k \)s in (43) are small numbers. Rather, we choose \( \Delta \rho \) such that \( g_s(\rho) \) peaks for a full-rank \( \rho \) that is close to \( \rho_{\text{ML}} \), as this gives a better over-all yield; see the example in figure 8 in section 7.

3.3. Peak shape

Let us briefly consider the problem of maximizing the target distribution \( f(\rho) \) over all \( \rho \)s of the form (1) \textit{without} enforcing \( \rho \geq 0 \), which amounts to regarding \( f(\rho) \) as a function of the coordinates \( g_l \) and finding the maximum on the coordinate space. While the permissible \( \rho \)s in (49) are positive unit-trace \( m \times m \) matrices, we now maximize over all hermitian unit-trace matrices, including those with negative eigenvalues. This maximum is reached for \( \rho = \hat{\rho} \). Whenever \( \hat{\rho} \) is in the quantum state space, \( \hat{\rho} \geq 0 \), we have \( \hat{\rho} = \rho_{\text{ML}} \); and \( \rho_{\text{ML}} \) sits on the boundary of the state space whenever \( \hat{\rho} \) is outside the state space, \( \hat{\rho} \not\geq 0 \).\(^{13}\)

As noted above, we do not match the peak location of \( g_s(\rho) \) with that of \( f(\rho) \) when \( \rho_{\text{ML}} \neq \hat{\rho} \) is rank-deficient, whereas the shift \( \Delta \rho = \hat{\rho} - \rho_{\text{peak}} \) suggests itself when \( \hat{\rho} \) is a quantum state. It is then worth trying to match the peak shape of the quantum Wishart distribution \( g(\rho) \) with that of the target distribution \( f(\rho) \). The expressions

\(^{12}\) One way of checking if certain probabilities \( p \) are permissible, is to regard them as relative frequencies of mock data and determine \( \rho_{\text{ML}} \) for these data. The probabilities are physical if \( p_k = \text{tr}(\Pi_k \rho_{\text{ML}}) \) for all \( k \), and only then.

\(^{13}\) The mock probabilities \( \tilde{\rho}_k = \text{tr}(\Pi_k \hat{\rho}) \) are always in the probability simplex; they are in the convex set of the physical probabilities only when \( \hat{\rho} = \rho_{\text{ML}} \).
corresponding to (28) and (29), now for $\rho$ near $\hat{\rho} = \rho_{\text{ML}}$ in the target distribution, are
\[
\log \frac{f(\hat{\rho} + \epsilon)}{f(\hat{\rho})} \approx -\frac{1}{2} \sum_{k=1}^{K} \frac{\nu_k}{\text{tr}(\Pi_k \hat{\rho})^2} \text{tr}(\Pi_k \epsilon)^2 = -\frac{1}{2} \sum_{l \neq l'} \epsilon_l F_{ll'} \epsilon_{l'}
\] (50)

with
\[
F_{ll'} = \sum_k \frac{\text{tr}(B_l \Pi_k)}{\nu_k \text{tr}(\Pi_k \hat{\rho})^2} \text{tr}(\Pi_k B_l').
\] (51)

When choosing the proposal distribution, one opts for $\rho_{\text{peak}}$ and $n$ such that the matrix $G$ resembles the matrix $F$. It is usually not possible to get a perfect match because $G$ derives from a Wishart distribution and is, therefore, subject to the symmetries discussed after (16), whereas $F$ is not constrained in this way.

In practice, we resort to looking at the one-dimensional “longitudinal” slice of $g(\rho)$ along the line from the completely mixed state $m^{-1}1_m$ to $\rho_{\text{peak}}$, that is $\epsilon \propto \rho_{\text{peak}} - m^{-1}1_m$, and choose the parameters such that this single-parameter distribution is a bit wider than that of the corresponding longitudinal slice of the target distribution $f(\rho)$, obtained for $\epsilon \propto \hat{\rho} - m^{-1}1_m$. In example of section 2.7, this slice is parameterized by the coordinate increment $\epsilon$. The matter will be illustrated by more examples in section 7; see figure 8.

Regarding the shift $\Delta \rho$ when $\hat{\rho} \neq \rho_{\text{ML}}$, we note that $\Delta \rho = r(\rho_{\text{ML}} - \rho_{\text{peak}})$ with $r \lesssim 1$ is a good first try for the trial-and-error search in the last paragraph of section 2.8. In this situation, it is more important to match well the distributions on and near the boundary of the state space than at $\rho_{\text{ML}}$, and we adjust $n$, $\Sigma$, $\kappa$, and $\Delta \rho$ in (41) for a higher overall yield.

4. Rejection sampling

We convert the proposal sample, drawn from the distribution $g_{s,\kappa}(\rho)$, into a sample drawn from the target distribution $f(\rho)$ by rejection sampling, a procedure introduced by John von Neumann in 1951 [37]. It consists of a simple accept-reject step: A $\rho$ from the proposal sample is entered into the target sample with a probability proportional to the ratio $f(\rho)/g_{s,\kappa}(\rho)$. Since we want to have the largest possible yield, while the acceptance probability $P_{\text{acc}}(\rho)$ cannot exceed unity, we choose
\[
P_{\text{acc}}(\rho) = \frac{1}{C} \frac{f(\rho)}{g_{s,\kappa}(\rho)} \quad \text{with} \quad C = \max_{\rho} \left\{ \frac{f(\rho)}{g_{s,\kappa}(\rho)} \right\},
\] (52)

where the maximum is evaluated over all $\rho$s in the proposal sample. The unphysical $\rho$s in the proposal sample, obtained when the shift in the second step of the state lottery (41) puts $\rho$ beyond the boundary of the state space, are always rejected because $f(\rho) = 0$ for every unphysical $\rho$.

It is crucial here that we have the analytical expressions for $f(\rho)$ and $g_{s,\kappa}(\rho)$ in (47) and (42), and it does not matter that we often do not know the normalization factor

\[^{14}\text{This illustrates the rule that one should “sample from a density } g \text{ with thicker tails than } f\text{” [27].}\]
needed in (48) or missing in (42);\textsuperscript{15} it is permissible to replace $C$ by an upper bound on the maximal ratio, at the price of a lower overall yield. Note that, unless we have an independent way of computing an upper bound on $C$,$\textsuperscript{16}$ the rejection sampling cannot be done while we are entering $\rho$s into the proposal sample; we must first compose the whole proposal sample, then determine the value of $C$, and finally perform the rejection sampling.

As remarked in section 2.8, the proper choice of the parameters that specify the proposal distribution $g_{s,\kappa}(\rho)$ is crucial for a good overall efficiency. Here is one more aspect that one should keep in mind: If $\kappa$ is too small, the ratio $f(\rho)/g_{s,\kappa}(\rho)$ will be largest where both distributions are small — in the tails of the target distribution and where $g_{s}(\rho)$ vanishes, in particular when $\rho$ has no preimage $\rho'$ in (39) or when the preimage is rank deficient.$\textsuperscript{17}$ When this happens, the rejection sampling has a low yield. Therefore, we choose $\kappa$ large enough to avoid this situation. This is an element in the overall-yield optimization mentioned in the final paragraph of section 2.8.

5. Sample verification

We exploit some tools, which were introduced in [38] for the quantification of the errors in quantum state estimation, for a verification of the target sample. In this section, $f(\rho)$ refers to the $\rho$ dependence of the target distribution in (47) without assuming the proper normalization of (48), and $\rho_{\text{ML}}$ is the quantum state for which $f(\rho)$ is maximal,

$$f(\rho_{\text{ML}}) = \max_{\rho} \{f(\rho)\}.$$  \hspace{1cm} (53)

Then, the sets of quantum states defined by

$$\mathcal{R}_\lambda = \{\rho|f(\rho) \geq \lambda f(\rho_{\text{ML}})\} \quad \text{with} \quad 0 \leq \lambda \leq 1$$  \hspace{1cm} (54)

are nested regions in the state space with $\mathcal{R}_\lambda \subset \mathcal{R}_{\lambda'}$ if $\lambda > \lambda'$; $\mathcal{R}_{\lambda=1}$ contains only $\rho_{\text{ML}}$ and $\mathcal{R}_{\lambda=0}$ is the whole state space. It is easy to check whether a particular $\rho$ is in $\mathcal{R}_\lambda$ or not.

For lack of a better alternative, we borrow the terminology from [38] and call

$$s_\lambda = \frac{\int_{\mathcal{R}_\lambda} (d\rho)}{\int_{\mathcal{R}_0} (d\rho)} \quad \text{and} \quad c_\lambda = \frac{\int_{\mathcal{R}_\lambda} (d\rho) f(\rho)}{\int_{\mathcal{R}_0} (d\rho) f(\rho)}$$  \hspace{1cm} (55)

the size and the credibility of $\mathcal{R}_\lambda$, respectively, and we evaluate both integrals by a Monte Carlo integration.$\textsuperscript{18}$ By counting how many $\rho$s from a uniform sample are in $\mathcal{R}_\lambda$,

$\textsuperscript{15}$ Similarly, in the context of (52) we can put aside factors that the two terms in (42) have in common, such as the proportionality factor between $(d\rho)$ and $[d\rho]$ in (4).

$\textsuperscript{16}$ Such cases require symmetries in the target distribution that match those of the proposal distribution. While it may be permissible to choose a Bayesian prior accordingly, the data-driven posterior will lack the symmetries.

$\textsuperscript{17}$ The occurrence of rank-deficient preimages is an issue even when $\Delta\rho = 0$ and there is no shift in (41), but $\kappa > 0$ is still required.

$\textsuperscript{18}$ The term was coined by Stanislaw Ulam [39]. The method dates back to 1949 [40]; for a comprehensive textbook exposition, see [41], for example.
we get a value for $s_\lambda$ as the fractional count; likewise, by counting how many $\rho$s from the target sample are in $R_\lambda$, we get a value for $c_\lambda$.

The accuracy of the $s_\lambda$ and $c_\lambda$ values is determined by the sampling error and that is small if the sample is large (10$^6$ entries, say) and of good quality; see section 6. Since the uniform sample has no quality issues, it provides a reference for the target sample through the identity

$$c_\lambda = \frac{\lambda s_\lambda + \int_{\lambda}^{1} d\lambda' s_{\lambda'}}{\int_{0}^{1} d\lambda' s_{\lambda'}},$$

which links the credibility as a function of $\lambda$ to the size. Note that the $s_\lambda$ values for the target distribution $f(\rho)$ as well as the $c_\lambda$ values from (56) can be computed before any sampling from the target distribution is performed.

Good agreement between the reliable credibility values provided by (56) and those obtained by the Monte Carlo integration of (55) confirms that the target sample is of good quality. More specifically, the quality assessment proceeds from regarding $c_\lambda$ from (56) as exact and the values obtained from the target sample as estimates,

$$\hat{c}_\lambda = \frac{1}{N_{tgt}} \sum_{k=1}^{N_{tgt}} \chi(f(\rho_k) > \lambda f(\rho_{ML})) \quad \text{with} \quad \chi(A) = \begin{cases} 1 & \text{if } A \text{ true} \\ 0 & \text{if } A \text{ false} \end{cases},$$

where the sum is over all $\rho_k$s in the target sample, which has $N_{tgt}$ entries. This is an unbiased estimator, $E(\hat{c}_\lambda) = c_\lambda$, with the variance $E(\hat{c}_\lambda^2) - c_\lambda^2 = c_\lambda (1 - c_\lambda) / N_{tgt}$. For the comparison of $\hat{c}_\lambda$ with $c_\lambda$, then, we use the mean squared error

$$Q = \int_{0}^{1} d\lambda (\hat{c}_\lambda - c_\lambda)^2,$$

which has the expected value

$$E(Q) = \int_{0}^{1} d\lambda \ E((\hat{c}_\lambda - c_\lambda)^2) = \frac{1}{N_{tgt}} \int_{0}^{1} d\lambda c_\lambda (1 - c_\lambda)$$

and the variance

$$E(Q^2) - E(Q)^2 = \frac{2}{N_{tgt}^2} \int_{0}^{1} d\lambda \int_{0}^{1} d\lambda' c_\lambda (1 - c_\lambda)^2$$

$$+ \frac{1}{N_{tgt}^3} \int_{0}^{1} d\lambda \int_{0}^{1} d\lambda' c_\lambda (1 - c_\lambda)(1 - 4c_\lambda - 2c_\lambda + 6c_\lambda c_\lambda),$$

where $c_\leq = \min\{c_\lambda, c_{\lambda'}\}$ and $c_\geq = \max\{c_\lambda, c_{\lambda'}\}$. We consider the target sample to be of good quality if its $Q$ value differs from $E(Q)$ by less than two standard deviations, and of very good quality if the difference is less than one standard deviation. More sophisticated tests are possible, among them the Kolmogorov–Smirnov test [42,43]; see section 15.4 in [27], for instance. We are not exploring such other possibilities here.

---

19 Sampling errors are discussed in section VI A 2 in [14], for example.

20 When estimating $c_\lambda$ from a large uniform sample, we are accepting a small negative bias that is, however, of no consequence here. The finite difference between the estimated and the actual values of $c_\lambda$ do matter, however. See the Appendix for details.

21 More sophisticated tests are possible, among them the Kolmogorov–Smirnov test [42,43]; see section 15.4 in [27], for instance. We are not exploring such other possibilities here.
$s_\lambda$ values that, in turn, we estimate from a large uniform sample, and the difference between the approximate and the actual $c_\lambda$ adds an additional term to $E(Q)$ of (59); see (A.17) in the appendix and figure 5 in section 7.

6. The curse of dimensionality

As discussed in the preceding section, the fractional counts that give us approximate values for $c_\lambda$ are unbiased estimators for the actual values with standard deviations of $[c_\lambda(1 - c_\lambda)/N_{tgt}]^{1/2}$. Likewise the approximate values for $s_\lambda$ have standard deviations of $[s_\lambda(1 - s_\lambda)/N_{ufm}]^{1/2}$, where $N_{ufm}$ is the total number of $\rho$s in the uniform sample; we take for granted that $N_{ufm}$ is large enough that the sampling error in $s_\lambda$, and the propagated error in $c_\lambda$ of (56) can be ignored. We now emphasize that the accuracy of these estimates is determined by the sample size and is independent of the dimension of the quantum state space. In this regard, then, these estimates do not fall prey to the “curse of dimensionality” that Richard Bellman observed [44]. But we cannot really escape from the curse, which is known to affect all major sampling methods, including the rejection sampling of section 4, and also occurs in other domains, such as optimization, function approximation, numerical integration, and machine learning [46].

Let us first consider storage requirements. We need 8 bytes of memory for one double precision number, which requires 24 bytes for one single-qubit state, 120 bytes for one two-qubit state, 504 bytes for one three-qubit state, 2040 bytes = 2.04 kB for one four-qubit state, and so forth, picking up a factor $\approx 4$ for each additional qubit: growth proportional to $m^2$, the dimension of the state space. This linear-in-$m^2$ increase in memory per quantum state is, however, paired with an exponential decrease of the overall acceptance probability during the rejection sampling. Suppose we have a pretty good proposal distribution that is a 90% match in every dimension, then — roughly speaking — the overall match is $0.9^3 = 0.729$ for a single qubit, $0.9^{15} = 0.206$ for a qubit pair, $0.9^{63} = 1.31 \times 10^{-3}$ for three-qubit states, and a dismal $0.9^{255} = 2.15 \times 10^{-12}$ for four-qubit states. Accordingly, a target sample of $10^5$ states requires a proposal sample stored in 3.3 MB, 58 MB, 38 GB, and 95 EB, respectively. Even if cleverly chosen proposal distributions can gain a few powers of 10 on these rough estimates, it is clear that the curse of dimensionality prevents us from generating useful target samples for many-qubit systems by the rejection sampling of section 4.

Similarly, the CPU-cost increases with the dimension quite dramatically. There is the positivity check that ensures $f(\rho) = 0$ if $\rho \not\geq 0$, which has a CPU-cost proportional $m^3$ for a $m \times m$ matrix, and finding the value of $C$ in (52) has a CPU-cost proportional to the size of the proposal sample if we generously assume that the evaluation of $g_{n,n}(\rho)$ and $f(\rho)$ is CPU-cheap. Clearly, the CPU-cost is forbiddingly large for many-qubit systems with $m = 2^n_{\text{qb}}$. We can, however, avoid the need for checking that $\rho \geq 0$ by choosing $\Delta \rho = 0$ in (42), while accepting the price of a smaller yield of the rejection sampling or more steps of the evolution algorithm.

22 All samples for higher-dimensional spaces that we are aware of, the samples in the repository [45] among them, are specified by the sampling algorithm, not by a target distribution.
The examples in section 7 demonstrate that, when using the modified Wishart distribution of (42) as the proposal distribution, we manage to sample efficiently one-, two-, and three-qubit states. For four-qubit states, with our limited computational power, we can only sample a certain class of simple target distributions by the rejection sampling. Thus, although the sampling algorithm introduced here is superior to other methods — in particular, we obtain uncorrelated samples — yet better methods for sampling from higher-dimensional state spaces are in demand. In a separate paper we combine Wishart distributions with sequentially constrained Monte Carlo sampling [47,48] and so manage to sample higher-dimensional quantum systems rather efficiently [26].

7. Results

First, through examples of sampling qubit states, we will illustrate in detail how to make use of the Wishart distribution for sampling quantum states with a given target distribution. Following that, we present examples of sampling two-qubit states. Although algorithms for reliably sampling such low-dimensional quantum states exist (for example, the HMC sampling method works well for system with dimension $m < 6$ [16, 18]), our method has the advantage of producing non-correlated samples. Next, we apply our method to the sampling of three- and four-qubit systems which are notoriously difficult for other sampling algorithms, and the method introduced here, while applicable, suffers from a high rejection rate.

7.1. Qubits

Assume that qubits are measured by a tetrahedral POM with the probabilities [49]

$$
\begin{align*}
p_1 &= \frac{1}{4} \left( 1 + \frac{x - y - z}{\sqrt{3}} \right), \\
p_2 &= \frac{1}{4} \left( 1 + \frac{y - z - x}{\sqrt{3}} \right), \\
p_3 &= \frac{1}{4} \left( 1 + \frac{z - x - y}{\sqrt{3}} \right), \\
p_4 &= \frac{1}{4} \left( 1 + \frac{x + y + z}{\sqrt{3}} \right),
\end{align*}
$$

(61)

where $\rho$ is parameterized as in (30). Accordingly, we have $K = 4$ in (43), (45), and (47). For the observed counts of detection events $\nu = (\nu_1, \nu_2, \nu_3, \nu_4)$, the probability element of the target distribution is

$$
(d\rho) f(\rho) \propto dx \, dy \, dz \, \eta(1 - x^2 - y^2 - z^2) \, p_1^{\nu_1} \, p_2^{\nu_2} \, p_3^{\nu_3} \, p_4^{\nu_4}
$$

(62)

where the step function selects the permissible $x, y, z$ values, those of the unit Bloch ball. As noted above, we do not need to calculate the normalization factor, as it plays no role in (52).

Because the POM is symmetric — technically speaking, it is a 2-design [50] — in the simplest, if untypical, scenario where each outcome is observed equally often, the target distribution is peaked at the completely mixed state $\rho_{\text{peak}} = \frac{1}{2} \mathbf{1}_2$; $f(\rho)$ is not isotropic, however, as the tetrahedron has privileged directions in the Bloch ball. We use the proposal distribution of (42) with $m = 2$, $\Sigma = \mathbf{1}_2$, $\Delta \rho = 0$, and $\kappa > 0$, that
**Figure 4.** Line densities of the qubit target distribution $f(\rho)$ in (62) (black) and the proposal distribution $g_{n,\kappa}(\rho)$ in (42) with $m = 2$ (blue) for $\rho(t) = \frac{1}{2}(1_2 + te \cdot \sigma)$ with a unit vector $e$. The target distributions (solid black) are rescaled by a factor of $\max_t\{f(\rho(t))/g(\rho(t))\}$ for each line, with the arrows pointing to where the blue and solid-black curves coincide; the dashed black curve in the middle-right plot shows the target distribution rescaled for matched peaks. In each plot, the value of $p_e$ is the acceptance rate of (63), the ratio of the areas under the solid black and blue curves.

**Top row and middle row:** The target distribution is for $\nu = \{25, 25, 25, 25\}$, and the isotropic proposal distributions have $\Sigma = 1_2$ and $\Delta \rho = 0$ as well as $\kappa = 0.1$ and $n = 14$ or 18, respectively. The unit vectors $e$ are chosen at random. The overall acceptance rates are $P_{\text{acc}} = 60.8\%$ (top row) and $P_{\text{acc}} = 33\%$ (middle row). **Bottom row:** The target distribution is for $\nu = \{25, 20, 25, 45\}$, the proposal distribution has $\Sigma = 1_2$, $\kappa = 0.2$, $n = 13$, and $\Delta \rho \neq 0$ such that the peaks of the target and proposal distributions are at the same location; the overall acceptance rate is $P_{\text{acc}} = 28.6\%$. In the left plot, the unit vector $e$ points from the center of the Bloch ball to the peak location, whereas we have a randomly chosen $e$ in the right plot; note the smaller range of values in the right plot.

is: an isotropic Wishart distribution with an admixture of the uniform distribution; the uniform-distribution component ensures $g(\rho) > f(\rho)$ in the tails. The yield — the acceptance rate of the rejection sampling — depends on the parameter $n$ that controls the width of the peak. For example, a rather high acceptance rate of $P_{\text{acc}} = 60.8\%$ is obtained with $\kappa = 0.1$ and $n = 14$ for $\nu = \{25, 25, 25, 25\}$; by contrast, we have $P_{\text{acc}} = 33\%$ for $n = 18$, which is thus a worse choice for $n$ here. The top and middle rows in figure 4 refer to these $n$ values.

In figure 4 we exploit our knowledge of the exact forms of the target distribution $f(\rho)$ and the proposal distribution $g(\rho)$ for a comparison of the distributions along diameters of the Bloch ball. More specifically, the qubit states considered are $\rho(t) = \frac{1}{2}(1_2 + te \cdot \sigma)$ with a unit vector $e$ and $-1 \leq t \leq 1$. Accordingly, the acceptance rate for this one-
parameter family of states is
\[ p_e = \max_t \left\{ \frac{f(\rho(t))}{g(\rho(t))} \right\}^{-1} \int_{-1}^1 dt f(\rho(t)) \int_{-1}^1 dt g(\rho(t)) \].

(63)

We note in passing that suitable analogous expressions work also for higher dimensional systems.

As the top and middle rows in figure 4 confirm, we have a better overall yield for \( n = 14 \) than for \( n = 18 \); for \( n = 18 \), where the width of the Wishart distribution well matches that of the target distribution, the almost perfect yields for some directions \( \mathbf{e} \), as in the middle-left example, does not compensate for the low yield for other diameters, as in the middle-right example. The middle-right plot also indicates why \( n = 18 \) is a poor choice, as the maximum of \( f(\rho(t))/g(\rho(t)) \) occurs in the tails of the distributions, not near the peak of \( f(\rho(t)) \); recall footnote 14 in this context.

In the typical experimental scenario, the counts of measurement clicks are not balanced among the POM outcomes, the target distribution does not peak at the center of the Bloch ball, and is not approximately isotropic. We use \( \nu = \{10, 20, 25, 45\} \) to illustrate this situation in the bottom row of figure 4; here the peak of \( f(\rho) \) is at distance 0.8832 from the center of the Bloch sphere. We can match the peak of the Wishart distribution by either changing the covariance matrix \( \Sigma \) in accordance with (25), or by a suitable shift \( \Delta \rho \), or by a combination of both. We find that \( \Sigma = 1_2 \) in conjunction with \( \Delta \rho = \rho_{\text{peak}} \) is very effective as this provides a high acceptance rate. For the event counts stated above and this shift, an acceptance rate of \( P_{\text{acc}} = 28.6\% \) is achieved by using a Wishart sample with \( \kappa = 0.2 \) and \( n = 13 \) — this is, in fact, a rather high acceptance rate in view of the many states in the Wishart sample that the shift renders unphysical.

We use a larger admixture of the uniform sample here as we need to fill in the void left behind after the shift by \( \Delta \rho \). The bottom row in figure 4 shows the distribution along the diameter through the peak location (left) and along another, randomly chosen, diameter (right).

To verify the sample we evaluate the size and credibility of the bounded likelihood regions as described in section 5. The size \( s_\lambda \) is estimated from a uniform sample with ten million entries. The regarded-as-exact \( c_\lambda \) values then obtained from (56) are used for reference (dashed black curves in figure 5, top row). We numerically evaluate \( c_\lambda \) using (57) for \( N_{\text{tgt}} = 500, 1000, 5000, \) and \( 10000 \) and compare it with the regarded-as-exact values in the top row in figure 5. The middle row in figure 5, with data for target sample sizes of \( N_{\text{tgt}} = 500, 1000, 2000, 5000, \) and \( 10000 \), shows that our method of sampling is reliable and the standard deviation of \( c_\lambda \) is proportional to \( N_{\text{tgt}}^{-\frac{1}{2}} \) as expected. The bottom plot in figure 5 confirms (A.17) in the appendix, which corrects (59) by accounting for the difference between the regarded-as-exact \( c_\lambda \) values and the actual ones. We note further that the standard deviations of the histograms are approximately equal to their mean values (indicated by circles), so that the criterion stated after (60) tells us that all samples with \( Q \) values [cf. (58)] less than twice the mean value are of very good quality.
Figure 5. Verification for qubit samples. **Top row:** Samples generated for data $\nu = \{25, 25, 25, 25\}$ (left) and $\nu = \{10, 20, 25, 45\}$ (right). The size $s_\lambda$ and the regarded-as-exact credibility $c_\lambda$ of (56) are evaluated with the aid of uniformly distributed samples of ($N_{uFm} = 10^7$). The estimated credibility $\hat{c}_\lambda$ is obtained from target samples with $N_{tgt} = 500, 1000, 5000, \text{or} 10000$ entries. **Middle row:** Ratio of the expected and the exact value of $\hat{c}_\lambda$ (left, for a single target sample); standard deviation of $\hat{c}_\lambda$ (right, averaged over 100 target samples); for $\nu = \{10, 20, 25, 45\}$ and target samples with $N_{tgt} = 500, 1000, 2000, 5000, \text{or} 10000$ entries each. There are values for $\lambda = 0, 0.1, 0.2, \ldots, 1.0$, connected by straight lines that guide the eye. The smooth dashed yellow curve on the right shows the square root of the variance for $N_{tgt} = 2000$; see after (57). **Bottom:** Histograms of $N_{tgt}Q$ values for 1000 samples of $N_{tgt} = 10^4, 10^2, 10^3, 10^4, \text{and} 10^5$ for $\nu = \{10, 20, 25, 45\}$. The purple vertical lines indicate 0, 10, 20, or 30 counts in the histogram bins. The circles show the mean values for each histogram, to which the blue curve is fitted in accordance with (A.17), with 0.15567 (dashed horizontal line) and $5.245 \times 10^{-7}$ for the values of the two integrals.
7.2. Qubit pairs

For simulated data for measurements on qubit pairs, we use the probabilities of a double tetrahedral measurement with the sixteen probability operators

\[ \Pi_k = \Pi_l^{(1)} \otimes \Pi_m^{(2)} \] \quad \text{with } l, m = 1, 2, 3, 4, \quad (64) \]

where \( \Pi_l^{(1)} \) and \( \Pi_m^{(2)} \) make up the single-qubit tetrahedron POMs for the probabilities in (61), and the index \( k = 4(l-1) + m \) covers the integers in the range \( 1 \leq k \leq 16 = K \). For measurement data \( \nu = \{\nu_1, \nu_2, \ldots, \nu_{16}\} \), the target distribution has the form of (47),

\[ (d\rho) f(\rho) \propto (d\rho) \prod_{k=1}^{16} \text{tr}(\Pi_k \rho)^{\nu_k}. \]

First, we compare the performance of a proposal sample from the uniform distribution with one from the Wishart distribution for the centered, symmetric target distributions that refer to fictitious measurements with the same number of events for each outcome, that is \( \nu_1 = \nu_2 = \cdots = \nu_{16} = \nu \); see figure 6. The acceptance rate \( P_{\text{acc}} \) obtained from using a uniformly distributed proposal sample decreases exponentially as the measurement count \( \nu \) increases and the target distribution becomes correspondingly narrower. By contrast, when we admix a sample drawn from the Wishart distribution, the optimal acceptance rate decreases to about 0.1% at around \( \nu = 20 \) and then it increases slowly to about 1% for \( \nu = 100 \) and remains at around this value for even larger \( \nu \)s. For example, an acceptance rate of \( P_{\text{acc}} = 0.48\% \) is obtained by adding 20% of \( W_4^{(Q)}(6, 1_4) \) for \( \nu = 10, \) \( P_{\text{acc}} = 0.091\% \) is obtained by adding 50% of \( W_4^{(Q)}(8, 1_4) \) for \( \nu = 20, \) and \( P_{\text{acc}} = 0.64\% \) is obtained by adding 90% of \( W_4^{(Q)}(35, 1_4) \) for \( \nu = 100. \)

Whereas, the acceptance rates for the uniform proposal samples are \( P_{\text{acc}} = 5.4 \times 10^{-4}, \) \( 2.3 \times 10^{-5}, \) and \( 1.0 \times 10^{-8} \) for \( \nu = 10, 20 \) and 100, respectively. The use of a proposal sample from the Wishart distribution clearly gives a much larger acceptance rate for such a peaked target distribution than the uniform proposal sample.
While the use of Wishart distribution can increase the acceptance rate, to reach the optimal acceptance rate one needs to find the most suitable Wishart distribution to use. Fortunately, this optimization is not difficult because one observes that the optimal number of columns $n_{\text{opt}}$ scales linearly with the total number of counts; see figure 7. Moreover, we find that this proportionality of $n_{\text{opt}}$ and $N$ holds quite generally when one samples non-centered target distributions. When $N$ is large, the acceptance rate can be rather robust against the choice of $n$ because the peak width depends only weakly on this parameter; see section 2.5.

Next, let’s check the performance of the method for sampling non-centered target distributions. For illustration, we use a target distribution given by a particular random data obtained from 100 measurements with $\nu = \{10, 4, 6, 4, 7, 6, 5, 6, 10, 6, 5, 6, 8, 6\}$. The peak of the target distribution in the probability simplex corresponds to a non-physical state, and the peak in the physical space is given by the maximum-likelihood estimator $\rho_{\text{ML}}$ which is a rank-3 state with eigenvalues $\{0.5033, 0.3377, 0.1589, 0\}$.

As discussed in section 2.8, we can adjust the peak location of the proposal distribution by a suitable shift. We generate the proposal distribution by the two steps in (41). First, we chose a state $\rho_{\text{peak}}' = x_1 \rho_{\text{ML}} + \frac{1}{4}(1-x_1)1_4$ to be the peak of the Wishart distribution $W_4^{(Q)}(n, \Sigma)$ and find the corresponding $\Sigma$ to produce a preliminary proposal sample. Then, we shift the sample states by $\Delta \rho = x_2 (\rho_{\text{ML}} - \frac{1}{4}1_4)$ as in (39), after which we have a sample that is peaked at $\rho_{\text{peak}} = (x_1 + x_2) \rho_{\text{ML}} + \frac{1}{4}(1-x_1-x_2)1_4$; to this we admix a fraction $\kappa$ from the uniform sample and arrive at a proposal sample in accordance with the distribution $g_{s,\kappa}(\rho)$ of (42). When $x_1 + x_2 = 1$, the peak of the proposal distribution coincides with the maximum-likelihood estimator. However, we do not need the peaks to coincide exactly, instead, they just need to be close enough to give a good acceptance rate.

For example, we achieve an acceptance rate of $P_{\text{acc}} > 0.5\%$ with a proposal distribution obtained for $\kappa = 0.6, n = 5, x_1 = 0.75, \text{and } x_2 = 0.15$; see figure 8. For the verification of the sample, we evaluate the credibility $c_{\lambda}$ for different sample sizes and compare it to the regarded-as-exact value computed from $s_{\lambda}$; see figure 9. The result
shows that our method is reliable for sampling two-qubit systems and 10000 sample points are enough for the estimation of physical quantities, such as the credibility, with an error less than 0.1%. For this particular example, the acceptance rate for a uniformly distributed proposal distribution is about 0.06% which is smaller by about a factor of ten. Thus, using the Wishart distribution does improve the acceptance rate but the improvement is not significant enough when applied to a target distribution that is not narrowly peaked, such as the one for $N = 100$ measurements. The gain in efficiency is much more remarkable when the target distribution is more peaked. For instance, if the target distribution is given by $N = 1000$ measurements with the same frequency of clicks for each POM outcome, an acceptance rate of larger than $10^{-5}$ is achieved with a

$$\nu = \{10, 4, 6, 4, 7, 6, 5, 6, 5, 6, 10, 6, 5, 6, 8, 6\}$$

Figure 9. The credibility $c_\lambda$ evaluated for different sample sizes. $10^6$ sample states from the uniform distribution are used to evaluate the size and the ‘exact’ values of the credibility given by the dashed curve.
Table 1. CPU time taken for generating a sample of $10^5$ states with different sampling strategies. All the tasks are CPU-parallelized and run on the same regular desktop. (a) Two-qubit sample with $\nu_k = 10$ for $k = 1, \ldots, 16$; (b) two-qubit sample with $\nu = \{10, 4, 6, 4, 7, 6, 5, 6, 5, 6, 10, 6, 5, 6, 8, 6\}$; (c) two-qubit sample with each $\nu_k$ equal to ten times the value of example (b); (d) three-qubit sample with $\nu_k = 10$ for $k = 1, \ldots, 64$; (e) and (f) three- and four-qubit samples with a randomly generated count sequence of 3000 experiments.$^{23}$

Remarks: $\star$ the acceptance rate is too low, less than $1 \times 10^8$; $\dagger$ the acceptance rate is about $1.2 \times 10^{-5}$; $\dagger\dagger$ the acceptance rate is 21 samples from $1.5 \times 10^8$ proposal samples.

| sampling strategies          | two qubits     | three qubits   | four |
|------------------------------|----------------|---------------|------|
| uniform $[\kappa = 1 \text{ in (42)}]$ | 60 min 30 min | $\star$ $\star$ $\star$ |      |
| Wishart-uniform $[0 < \kappa < 1]$ | 7 min 5 min 31 h | 100 h$^{\dagger}$ 13000 h$^{\dagger\dagger}$ | $\star$ |
| HMC [38]                     | 80 min 80 min | – – – – – – – – |      |
| SCMC+Wishart [26]            | 2.5 min 2.5 min 2.5 min | 9 min 60 min 60 h |      |

Proposal distribution drawn from $W_4^{(Q)}(10, \Sigma)$ with $x_1 = 0.8$ and $x_2 = 0.15$. This is at least $10^3$ times more efficient than using a uniform proposal distribution which fails to produce any accepted sample point for as many as $10^8$ proposal states.

When sampling from the two-qubit state space, the CPU time taken when using the Wishart distribution for the proposal is significantly lower than that for other methods. Table 1 shows that for sampling a centered distribution with $N = 160$ and a non-centered distribution with $N = 100$ it is a few times more efficient to use the Wishart distribution with an admixture of the uniform distribution than solely the uniform distribution; see the top-two entries in columns (a) and (b). The advantage of using the Wishart distribution can become much more prominent for sampling more peaked distributions with larger $N$; see column (c). Our method is also more efficient than the Hamiltonian Monte Carlo (HMC) method that is discussed in [38]; see the third row. While the efficiency of the HMC method does not depend on $N$, which can give it an advantage over the sampling from a uniform distributions for large $N$, the current implementation of the HMC algorithm is only reliable for systems with low dimension, owing to issues with the stability of the algorithm; note also that HMC yields correlated samples even if

$^{23}$ The simulated data for the three-qubit example are

$\nu = \{36, 13, 64, 71, 14, 16, 7, 15, 60, 10, 84, 63, 64, 9, 55, 71, 8, 12, 10, 16, 16, 48, 67, 62, 9, 64, 75, 63, 10, 74, 60, 73, 65, 14, 62, 66, 9, 57, 76, 53, 82, 78, 128, 22, 61, 44, 25, 27, 56, 12, 52, 66, 14, 76, 56, 78, 45, 47, 22, 27, 66, 68, 25, 102\}$

with $N = 3000$. The 64 counts refer to the three-qubit analog of (64). For these data, $\rho_{ML}$ is rank deficient.
the correlations are usually weaker than those in samples from other Markov chain MC methods. In the fourth row in table 1 we list the CPU time for generating samples with the sequentially constrained MC (SCMC) algorithm that we describe in [26]; it appears to outperform all other methods. If we consider the one-qubit and two-qubit situations of columns (a) and (b), however, the “Wishart-uniform method” of this paper is just as practical and requires much less effort in writing and debugging computer code; also, the “SCMC+ Wishart method” of reference [26] builds on the foundations laid by the “Wishart-uniform method.”

7.3. Three and four qubits

The sampling method introduced here works for systems of any dimension — at least, there are no reasons of principle why it shouldn’t. In practice, however, the ‘curse of dimensionality’ can become a serious obstacle, due to limitations in both CPU time and memory; see section 6.

The storage aspects discussed in section 6 is much more a concern for three-qubit systems than for one-qubit or two-qubit systems. When sampling three-qubit states, we use up to a maximum number of $2.4 \times 10^8$ proposal states in each run of the algorithm, which takes up about 120 GB of storage. The first example we investigated is a centered distribution for $\nu = \{\nu_k = 10 \text{ for } k = 1, \ldots, 64\}$. No useful sample can be produced by drawing from a uniform distribution as its acceptance rate is smaller than $10^{-8}/2.4$. On the other hand, when sampling with a proposal distribution composed 20% of the Wishart distribution $W_{8}^{(Q)}(9, 1_8)$ and 80% of the uniform distribution — $\kappa = 0.8$ in (42) — the acceptance rate is approximately $1.2 \times 10^{-5}$. This allows us to obtain a sample size of $10^5$ within 100 hours.

Reliable samples for non-centered and/or narrower distributions can also be generated from the Wishart-plus-uniform proposal distribution (42). For example, for the target distribution that corresponds to the simulated data in footnote 23, we obtain $P_{\text{acc}} \approx 1.4 \times 10^{-7}$ using 40% of $W_{8}^{(Q)}(80, 1_8)$, with shift parameters $x_1 = 0.6$ and $x_2 = 0.35$, plus 60% of the uniform distribution. It takes about 2.75 hours to produce 21 sample states from $1.5 \times 10^8$ proposal states. Thus, to generate $10^5$ samples it would take about 13000 hours, as entered in column (e) of table 1.

For a four-qubit system, storage is even more of an issue; we can only store about $6 \times 10^7$ states in 120 GB of memory. We tried to sample a target distribution for $N = 3000$ simulated counts from a Wishart distribution but failed to produce a reliable sample. This suggests that the acceptance rate is well below $1.6 \times 10^{-8}$ and, therefore, to sample states of such high dimension efficiently, yet other methods are to be sought. One option is the adaptation of the SCMC sampler [47, 48] to the sampling of quantum states. We managed to produce uncorrelated samples of high-dimensional quantum states in this way. The “SCMC+ Wishart algorithm” is, however, beyond the scope of this paper; we deal with it in [26].
8. Summary and outlook

We established the probability distribution of the unit-trace, positive square matrices that represent quantum states, generated from the gaussian distribution for nonsquare matrices and the induced intermediate Wishart distribution for positive matrices. These distributions of quantum-state matrices are shifted suitably and supplemented with an admixture of the uniform distribution so that they can serve as tailored proposal distributions for the target distribution from which we want to sample.

The target sample is generated from a proposal sample by an accept-reject step. Since the proposal sample is uncorrelated (or i.i.d.), so is the target sample. This is a notable advantage over random walk algorithms, including that of Hamiltonian Monte Carlo, which generate correlated samples with a correspondingly reduced effective sample size.

The sampling algorithm works very well for samples from the one-qubit and two-qubit state spaces, whereas the efficiency in generating three-qubit samples is low, owing to the curse of dimensionality. We could not generate a useful sample of four-qubit states drawn from a structured target distribution. (Drawing from the uniform distribution or other highly symmetric distributions is efficient.)

The discussion here is limited to the Wishart distributions that derive from zero-mean gaussian distributions. We explore the options offered by gaussian distributions with a nonzero mean in a companion paper [28] but the increase in flexibility does not overcome the curse of dimensionality.

The general strategy explored in this paper — equip yourself with a large, CPU-cheap, uncorrelated proposal sample and turn it into a smaller target sample by rejection sampling — does not require a proposal distribution of the Wishart-uniform kind for its implementation. Other proposal distributions are possible and could very well yield substantially larger acceptance rates. While we have no particular suggestions to make, we trust that others will till this field.

There is, fortunately, solid evidence that the sequentially constrained Monte Carlo sampling algorithm succeeds in overcoming the curse of dimensionality to some extent. Rather than accepting or rejecting the quantum states in the proposal sample, the sample as a whole is processed and turned into a proper target sample step by step; this method also benefits from using a tailored Wishart distributions for the proposal. We present this approach in a separate paper [26].

Acknowledgments

S.B. is supported by a PBC postdoctoral fellowship at Tel-Aviv University. The Centre for Quantum Technologies is a Research Centre of Excellence funded by the Ministry of Education and the National Research Foundation of Singapore.
Appendix

In this appendix, we elaborate on footnote 20. As the analog of (57), we have

\[
\hat{s}_\lambda = \frac{1}{N_{ufm}} \sum_{k=1}^{N_{ufm}} \chi(f(\rho_k) > \lambda f(\rho_{ML})) = \frac{1}{N_{ufm}} \sum_{k=1}^{N_{ufm}} \chi(\lambda < \lambda_k)
\] (A.1)

when estimating \( s_\lambda \) from the \( N_{ufm} \) entries \( \rho_k \) in the uniform sample, with

\[
\lambda_k = \frac{f(\rho_k)}{f(\rho_{ML})}, \quad 0 \leq \lambda_k \leq 1.
\] (A.2)

The \( \lambda_k \)'s are independent (uncorrelated) random variables, each with the probability element

\[
\mathbb{E}(\lambda' < \lambda_k < \lambda' + d\lambda') = d\lambda' \left( -\frac{\partial s_{\lambda'}}{\partial \lambda'} \right),
\] (A.3)

so that, for each \( \lambda_k \),

\[
\mathbb{E}(\chi(\lambda < \lambda_k)) = \int_\lambda^1 d\lambda' \left( -\frac{\partial s_{\lambda'}}{\partial \lambda'} \right) = s_\lambda.
\] (A.4)

It follows that \( \hat{s}_\lambda \) is unbiased, \( \mathbb{E}(\hat{s}_\lambda) = s_\lambda \), and its variance is proportional to \( N_{ufm}^{-1} \),

\[
\mathbb{E}(\hat{s}_\lambda^2) - \mathbb{E}(\hat{s}_\lambda)^2 = \frac{1}{N_{ufm}} s_\lambda (1 - s_\lambda).
\] (A.5)

Further, the numerator in (56) is

\[
\lambda s_\lambda + \int_\lambda^1 d\lambda' s_{\lambda'} = \int_\lambda^1 d\lambda' \left( -\frac{\partial s_{\lambda'}}{\partial \lambda'} \right) \lambda' = \int_0^1 d\lambda' \left( -\frac{\partial s_{\lambda'}}{\partial \lambda'} \right) \chi(\lambda < \lambda') \lambda'
\] = \( \mathbb{E}(\chi(\lambda < \lambda_k) \lambda_k) \),
\] (A.6)

and

\[
\lambda \hat{s}_\lambda + \int_\lambda^1 d\lambda' \hat{s}_{\lambda'} = \frac{1}{N_{ufm}} \sum_{k=1}^{N_{ufm}} \chi(\lambda < \lambda_k) \lambda_k,
\] (A.7)

is the corresponding unbiased estimator for the numerator; for \( \lambda = 0 \), this is the estimator for the denominator,

\[
\int_0^1 d\lambda' \hat{s}_{\lambda'} = \frac{1}{N_{ufm}} \sum_{k=1}^{N_{ufm}} \lambda_k.
\] (A.8)

Their ratio,

\[
\hat{c}_\lambda^{(ufm)} = \frac{\lambda \hat{s}_\lambda + \int_\lambda^1 d\lambda' \hat{s}_{\lambda'}}{\int_0^1 d\lambda' \hat{s}_{\lambda'}} = \frac{\sum_{k=1}^{N_{ufm}} \chi(\lambda < \lambda_k) \lambda_k}{\sum_{l=1}^{N_{ufm}} \lambda_l},
\] (A.9)

is what we get from (56) for the regarded-as-exact credibility.
As we shall now demonstrate, \( \widehat{c}_\lambda^{(ufm)} \) has a negative bias. First, we note that
\[
\mathbb{E}(\widehat{c}_\lambda^{(ufm)}) = \mathbb{E} \left( \sum_k \chi(\lambda < \lambda_k) \lambda_k \int_0^\infty d\alpha \, e^{-\alpha \sum_l \lambda_l} \right) = \sum_k \int_0^\infty d\alpha \, \mathbb{E} \left( \chi(\lambda < \lambda_k) \lambda_k e^{-\alpha \lambda_k} \prod_{l(\neq k)} \mathbb{E}(e^{-\alpha \lambda_l}) \right) = N_{ufm} \int_0^\infty d\alpha \, \mathbb{E} \left( \chi(\lambda < \lambda') \lambda' e^{-\alpha \lambda'} \right) \mathbb{E}(e^{-\alpha \lambda'})^{-\frac{N_{ufm}}{2}},
\]
(A.10)
where \( \lambda' \) is a random variable with the probability element of (A.3) and the expected values refer to this random variable. Then, we integrate by parts and arrive at
\[
\mathbb{E}(\widehat{c}_\lambda^{(ufm)}) = c_\lambda + \int_0^\infty d\alpha \, \mathbb{E}(e^{-\alpha \lambda'})^{N_{ufm}} \frac{\partial}{\partial \alpha} C_\lambda(\alpha),
\]
(A.11)
where
\[
C_\lambda(\alpha) = \frac{\mathbb{E}(\chi(\lambda < \lambda') \lambda' e^{-\alpha \lambda'})}{\mathbb{E}(\lambda' e^{-\alpha \lambda'})} = \left( 1 + \frac{\mathbb{E}(\chi(\lambda > \lambda') \lambda' e^{-\alpha \lambda'})}{\mathbb{E}(\chi(\lambda < \lambda') \lambda' e^{-\alpha \lambda'})} \right)^{-1}
\]
(A.12)
is a monotonic decreasing function of \( \alpha \), and \( C_\lambda(0) = \mathbb{E}(\lambda')^{-1} \mathbb{E}(\chi(\lambda < \lambda') \lambda') = c_\lambda \) is taken into account. Accordingly, the term added to \( c_\lambda \) in (A.11) is negative, that is: \( \widehat{c}_\lambda^{(ufm)} \) has a negative bias. Indeed, when \( \alpha' > \alpha \), we observe that
\[
C_\lambda(\alpha'^{-1}) = 1 + \frac{\mathbb{E}(\chi(\lambda > \lambda') \lambda' e^{-\alpha' \lambda'}) e^{-\alpha'(\lambda - \lambda')}}{\mathbb{E}(\chi(\lambda < \lambda') \lambda' e^{-\alpha' \lambda'})} > 1 + \frac{\mathbb{E}(\chi(\lambda > \lambda') \lambda' e^{-\alpha' \lambda'})}{\mathbb{E}(\chi(\lambda < \lambda') \lambda' e^{-\alpha' \lambda'})} = C_\lambda(\alpha)^{-1},
\]
(A.13)
or \( C_\lambda(\alpha') < C_\lambda(\alpha) \).

Repeated integrations by parts express the bias in (A.11) as a sum of terms proportional to powers of \( N_{ufm}^{-1} \). The leading term is exhibited here:
\[
\int_0^\infty d\alpha \, \mathbb{E}(e^{-\alpha \lambda'})^{\frac{N_{ufm}}{2}} \frac{\partial}{\partial \alpha} C_\lambda(\alpha) = \frac{1}{N_{ufm} \mathbb{E}(\lambda'^2)} \left[ c_\lambda - \frac{\mathbb{E}(\chi(\lambda < \lambda') \lambda'^2)}{\mathbb{E}(\lambda'^2)} \right] + \cdots
\]
(A.14)
with
\[
\mathbb{E}(\chi(\lambda < \lambda') \lambda'^2) = \lambda^2 s_\lambda + 2 \int_0^1 d\lambda' \lambda' s_{\lambda'}, \quad \mathbb{E}(\lambda'^2) = 2 \int_0^1 d\lambda' \lambda' s_{\lambda'}.
\]
(A.15)

Since both the variance of \( \widehat{s}_\lambda \) and the bias in \( \widehat{c}_\lambda^{(ufm)} \) are proportional to \( N_{ufm}^{-1} \), the bias in \( \widehat{c}_\lambda^{(ufm)} \) is smaller than the standard deviation of \( \widehat{s}_\lambda \), from which \( \widehat{c}_\lambda^{(ufm)} \) is computed, by a factor of \( N_{ufm}^{-\frac{1}{2}} \). Accordingly, this bias is of no consequence for the considerations in section 5, the more so if we recall the typical circumstances of \( N_{ufm} \gg N_{igt} \gg 1 \).
In practice, we are using the estimate $\hat{c}_\lambda^{(ufm)}$ of (A.9), obtained from a large uniform sample, in the place of the actual $c_\lambda$ when computing $Q$ of (58), that is
\[
Q \to \int_0^1 d\lambda \left[ (\hat{c}_\lambda - c_\lambda) - (\hat{c}_\lambda^{(ufm)} - c_\lambda) \right]^2; \tag{A.16}
\]
the expected value of $\hat{c}_\lambda^{(ufm)} - c_\lambda$ (with respect to the uniform distribution) is the bias in (A.11). As a consequence, the expected value of (59) (with respect to the target distribution) acquires a small additional term that does not depend on the size of the target sample,
\[
\mathbb{E}(Q) \to \frac{1}{N_{tgt}} \int_0^1 d\lambda c_\lambda (1 - c_\lambda) + \int_0^1 d\lambda \left( \hat{c}_\lambda^{(ufm)} - c_\lambda \right)^2, \tag{A.17}
\]
and the variance in (60) also acquires a corresponding additional term.

References

[1] Horodecki Karol, Horodecki Michal, Horodecki Pawel, Leung Debbie and Oppenheim Jonathan 2008 Unconditional privacy over channels which cannot convey quantum information Phys. Rev. Lett. 100 110502
[2] Hamma Alioscia, Santra Siddharta and Zanardi Paolo 2012 Quantum entanglement in random physical states Phys. Rev. Lett. 109 040502
[3] Dupuis Frédéric, Fawzi Omar and Wehner Stephanie 2015 Entanglement sampling and applications IEEE Trans. Inf. Theory 61 1093–1112
[4] Collins Benoît and Nechita Ion 2016 Random matrix techniques in quantum information theory J. Math. Phys. 57 015215
[5] Sim Jun Yan, Suzuki Jun, Englert Berthold-Georg and Ng Hui Khoon 2020 User-specified random sampling of quantum channels and its applications Phys. Rev. A 101 022307
[6] Lu Yiping, Sim Jun Yan, Suzuki Jun, Englert Berthold-Georg and Ng Hui Khoon 2020 Direct estimation of minimum gate fidelity Phys. Rev. A 102 022410
[7] Li Xikun, Shang Jiangwei, Ng Hui Khoon and Englert Berthold-Georg 2016 Optimal error intervals for properties of the quantum state Phys. Rev. A 94 062112
[8] Oh Changhun, Teo Yong Siah and Jeong Hyunseok 2019 Probing Bayesian Credible Regions Intrinsically: A Feasible Error Certification for Physical Systems Phys. Rev. Lett. 123 040602
[9] Oh Changchun, Teo Yong Siah and Jeong Hyunseok 2019 Efficient Bayesian credible-region certification for quantum-state tomography Phys. Rev. A 100 012345
[10] Życzkowski Karol, Penson Karol A, Nechita Ion and Collines Benoît 2011 Generating random density matrices J. Math. Phys. 52 06201
[11] Bengtsson Ingemar and Życzkowski Karol 2017 Geometry of Quantum States second edition (Cambridge: Cambridge University Press)
[12] Braunstein Samuel Leon 1996 Geometry of quantum inference Phys. Lett. A 219 169–174
[13] Hall Michael J W 1998 Random quantum correlations and density operator distributions Phys. Lett. A 242 123–129
[14] Gu Yanwu, Li Weijun, Evans Michael and Englert Berthold-Georg 2019 Very strong evidence in favor of quantum mechanics and against local hidden variables from a Bayesian analysis Phys. Rev. A 99 026104
[15] Liu Jun S 2008 Monte Carlo Strategies in Scientific Computing (Heidelberg: Springer)
[16] Shang Jiangwei, Seah Yi-Lin, Ng Hui Khoon, Nott David John and Englert Berthold-Georg 2015 Monte Carlo sampling from the quantum state space. I New J. Phys. 17 043017
[17] Neal Radford M 2011 MCMC using Hamiltonian dynamics (Handbook of Markov Chain Monte Carlo) ed Brooks Steve, Gelman Andrew, Jones Galin L and Meng Xiao-Li (Boca Raton: Chapman and Hall) chapter 5

[18] Seah Yi-Lin, Shang Jiangwei, Ng Hui Khoon, Nott David John and Englert Berthold-Georg 2015 Monte Carlo sampling from the quantum state space. II New J. Phys. 17 043018

[19] Pearson Egon Sharpe 1957 John Wishart 1898–1956 Biometrika 44 1–8

[20] Wishart John 1928 The generalized product-moment distribution in samples from a normal multivariate population Biometrika 20 32–43

[21] Goodman N R 1963 Statistical analysis based on a certain multivariate complex Gaussian distribution (An introduction) Ann. Math. Statist. 34 152–177

[22] Srivastava Muni S 1965 On the Complex Wishart Distribution Ann. Math. Statist. 36 313–315

[23] Muirhead Robb John 1982 Aspects of Multivariate Statistical Theory (Hoboken: Wiley)

[24] Johnson Richard A and Wichern Dean W 2007 Applied multivariate statistical analysis sixth edition (London: Pearson Education)

[25] Anderson Theodore Wilbur 2003 An Introduction to Multivariate Statistical Analysis third edition (New York: Wiley)

[26] Li Weijun, Han Rui, Shang Jangwei, Ng Hui Khoon and Englert Berthold-Georg 2021 Sequentially constrained Monte Carlo sampler for quantum states (in preparation)

[27] Wasserman Larry 2004 All of Statistics: A Concise Course in Statistical Inference (New York: Springer Science+Business Media)

[28] Bagchi Shrobona, Li Weijun, Han Rui, Ng Hui Khoon and Englert Berthold-Georg 2021 Uncorrelated problem-specific samples of quantum states from nonzero-mean Wishart distributions (in preparation)

[29] Yu Yi-Kuo and Zhang Yi-Cheng 2002 On the anti-Wishart distribution Physica A 312 1–22

[30] Janik Romuald A and Nowak Maciej A 2003 Wishart and anti-Wishart random matrices J. Phys. A 36 3629–3647

[31] Žyczkowski Karol and Sommers Hans-Jürgen 2001 Induced measures in the space of mixed quantum states J. Phys. A 34 7111–7125

[32] Faist Philippe and Renner Renato 2016 Practical, Reliable Error Bars in Quantum Tomography Phys. Rev. Lett. 117 010404

[33] Bloch Felix 1946 Nuclear induction Phys. Rev. 70 460–474

[34] Hradil Zdenek 1997 Quantum-state estimation Phys. Rev. A 55 R1561–R1564

[35] Paris Matteo and Reháček Jaroslav (ed) 2004 Quantum State Estimation (Lecture Notes in Physics vol 649) (Heidelberg: Springer)

[36] Shang Jiangwei, Zhang Zhengyun and Ng Hui Khoon 2017 Superfast maximum likelihood reconstruction for quantum tomography Phys. Rev. A 95 062336

[37] von Neumann John 1951 Various Techniques Used in Connection With Random Digits J. Res. Nat. Bur. Stand. Appl. Math. Series 3 36–38

[38] Shang Jiangwei, Ng Hui Khoon, Schrawat Arun, Li Xikun and Englert Berthold-Georg 2013 Optimal error regions for quantum state estimation New J. Phys. 15 123026

[39] Metropolis Nicholas 1987 The beginning of the Monte Carlo method Los Alamos Science Special Issue 125–130

[40] Metropolis Nicholas and Ulam Stanislaw 1949 The Monte Carlo method J. Am. Stat. Assoc. 44 335–341

[41] Evans Michael and Swartz Tim 2000 Approximating integrals via Monte Carlo and deterministic methods (New York: Oxford University Press)

[42] Andrey Kolmogorov 1933 Sulla determinazione empirica di una legge di distribuzione G. Ist. Ital. Attuari 4 83–91

[43] Nicolai Šmírnov 1948 Table for estimating the goodness of fit of empirical distributions Ann. Math. Stat. 19 279–281

[44] Bellman Richard Ernest 1961 Adaptive control processes: a guided tour (Princeton: Princeton University Press)
[45] Shang Jiangwei, Seah Yi-Lin, Wang Boyu, Ng Hui Khoon, Nott David John and Englert Berthold-Georg 2016 Random samples of quantum states: Online resources eprint arXiv:1612.05180 [quant-ph] Repository at https://www.quantumlah.org/page/page.php?key=qsampling

[46] Donoho David Leigh 2000 High-dimensional data analysis: The curses and blessings of dimensionality AMS Math Challenges Lecture 1–32

[47] Del Moral Pierre, Doucet Arnaud and Jasra Ajay 2006 Sequential Monte Carlo samplers J. R. Statist. Soc. B 68 411–436

[48] Golchi Shirin and Campbell David A 2016 Sequentially Constrained Monte Carlo Comput. Stat. Data Anal. 97 98–113

[49] Řeháček Jaroslav, Englert Berthold-Georg and Kaszlikowski Dagomir 2004 Minimal qubit tomography Phys. Rev. A 70 052321

[50] Zauner Gerhard 2011 Quantum designs: Foundations of a noncommutative design theory Int. J. Quantum Inf. 9 445–507