Precision-guaranteed quantum tomography

Takanori Sugiyama,1,∗ Peter S. Turner,1 and Mio Murao1,2

1Department of Physics, Graduate School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan 113-0033.
2Institute for Nano Quantum Information Electronics, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo, Japan 153-8505.
(Dated: May 7, 2014)

Quantum state tomography is the standard tool in current experiments for verifying that a state prepared in the lab is close to an ideal target state, but up to now there were no rigorous methods for evaluating the precision of the state preparation in tomographic experiments. We propose a new estimator for quantum state tomography, and prove that the (always physical) estimates will be close to the true prepared state with high probability. We derive an explicit formula for evaluating how high the probability is for an arbitrary finite-dimensional system and explicitly give the one- and two-qubit cases as examples. This formula applies for any informationally complete sets of measurements, arbitrary finite number of data sets, and general loss functions including the infidelity, the Hilbert-Schmidt, and the trace distances. Using the formula, we can evaluate not only the difference between the estimated and prepared states, but also the difference between the prepared and target states. This is the first result directly applicable to the problem of evaluating the precision of estimation and preparation in quantum tomographic experiments.

PACS numbers: 03.65.Wj, 03.67.-a, 02.50.Tt, 06.20.Dk

I. INTRODUCTION

The development of Science has always been supported by the development of precise and accurate techniques of measurement and control. Measurement outcomes are affected by statistical and systematic noise, and evaluating the measurement precision under these errors is a fundamental aspect of those techniques. In Physics, now more than ever high-precision experiments are required for testing whether a theoretical model is suitable for describing Nature. A popular figure of merit for this precision is the standard deviation, but in more demanding experiments a different figure of merit, called a confidence level, is also used. For example, in the search for the standard model Higgs boson, the ATLAS group at the LHC reported an experimental result that narrowed the range of the possible Higgs boson mass at the 95% confidence level [1]. This shows the confidence level to be a compelling benchmark for justifying whether an experimental result is reliable or not.

Quantum information is another field where highly precise measurement and control are necessary. It has been shown theoretically that by using “quantumness”, we can perform more efficient computation [2] and more secure cryptography [3] compared to existing protocols. In the practical implementation of these new protocols, highly precise preparation and control of specific quantum states are required. Quantum tomography is a standard tool in current quantum information experiments for verifying a successful realization of states and operations [4]. Let us consider the case of state preparation, where \( \rho_a \) denotes a target state that we are trying to prepare in the lab. In real experiments, the true prepared state \( \rho \) does not coincide with \( \rho_a \) because of imperfections. We wish to evaluate the precision of this preparation, that is, the difference between \( \rho_a \) and \( \rho \) - however we do not know \( \rho \). Instead, we perform quantum state tomography; let \( \rho^\text{est}_N \) denote an estimate of the state made from \( N \) tomographic data sets. To date the best we have been able to do is to evaluate the difference between \( \rho_a \) and \( \rho^\text{est}_N \) (see Fig. 1), but even if the difference is small, it does not guarantee that the prepared state \( \rho \) is close to the target state \( \rho_a \), because \( \rho^\text{est}_N \) is given probabilistically and can deviate from \( \rho \) when \( N \) is finite. In this context, we refer to the difference between \( \rho^\text{est}_N \) and \( \rho \) the precision of the estimation.

There are many proposals for evaluating the precision
of estimation \[\rho\] and preparation \[\rho_0\]. An approach using confidence regions is one currently popular example. Unlike standard quantum tomography, in this approach the estimate is not a point but a region in state space. In \[\rho, \rho_0\], confidence region estimators for quantum state estimation were proposed, and their volume of the region were analyzed. The confidence level can be used for evaluating the precision of region estimates, but these cannot be directly applied for evaluating the precision of point estimates in tomographic experiments and state preparation. In [5, 6], two state estimators were proposed that use random sampling of Pauli measurements. Called compressed sensing, the authors proved that the estimates are close to the true state with high probability. However, the formulae derived for evaluating the difference between the estimates and the true state include indeterminant coefficients, and the value of the difference cannot be calculated. Therefore the compressed sensing results are not directly applicable to the evaluation of the estimation precision of tomographic experiments. In [5, 6], a method for estimating the difference between \(\rho\), and \(\rho_0\) as evaluated by the fidelity was proposed. Called direct fidelity estimation, it avoids point estimates altogether, and by performing random Pauli measurements it allows the precision of state preparation to be evaluated, (it assumes that the target state \(\rho_0\) is pure). However, the method is not capable of evaluating the estimation precision of point estimates in quantum tomography. Thus, this remains a crucial problem in the current theory of quantum tomography.

Here we give a solution to this problem. We propose a new point estimator for quantum state tomography in finite-dimensional systems, and prove that the estimated states are within a distance threshold \(\delta\) from the prepared state with high probability. We derive an explicit formula for evaluating how high the probability is in arbitrary finite-dimensional systems. This formula applies for any informationally complete set of measurements, for an arbitrary finite amount of data, and for general loss functions including the infidelity, the Hilbert-Schmidt, and the trace distances. Importantly, for a given experimental setup we can calculate the value of the formula without knowing the true prepared state, and so the formula can be used to evaluate the precision of state preparation. To our knowledge this is the first result directly applicable to evaluating the precisions of both estimation and preparation in quantum tomography. We demonstrate the technique for the example of one- and two-qubit state tomography.

II. PRELIMINARIES

We consider a finite \(d(<\infty)\) dimensional quantum system, with Hilbert space \(\mathcal{H}\). A state of the system is described by a density matrix, which is a positive-semidefinite and trace-one matrix, the space of which we denote by \(S(\mathcal{H})\). Let \(\rho\) denote the density matrix describing the true prepared state on \(\mathcal{H}\). It is unknown, and we make no further assumptions on \(\rho\). Suppose that \(N\) identical copies of the unknown true state, \(\rho^0 \otimes N\), are available, and we can perform a measurement on each copy. Our aim is to estimate the unknown true state \(\rho\) from measurement results. Let \(\lambda = (\lambda_1, \ldots, \lambda_{d-1})\) denote a set of Hermitian matrices satisfying (i) (Tracelessness) \(\text{Tr}[\lambda_1] = 0\) and (ii) (Orthogonality) \(\text{Tr}[\lambda_1 \lambda_j] = 2\delta_{aj}\).

Using this set, a density matrix can be parametrized as

\[
\rho(s) = \frac{1}{d} I + \frac{1}{2} s \cdot \lambda,
\]

where \(I\) is the identity matrix and \(s\) is a vector in \(\mathbb{R}^{d^2-1}\). The matrix and vector are uniquely related as \(s_\alpha = \text{Tr}[\rho \lambda_\alpha]\). Positive-semidefiniteness of \(\rho\) requires constraints on the parameter space. Let \(S\) denote the set of parameters corresponding to density matrices, and \(S\) is a convex subset of \(\mathbb{R}^{d^2-1}\). Estimation of \(\rho \in S(\mathcal{H})\) is equivalent to that of \(s \in S\).

The statistics of a quantum measurement are described by a positive operator-valued measure (POVM), which is a set of positive-semidefinite matrices that sum to the identity. In standard setting of quantum tomography we choose a combination of measurements. Let \(\Pi = \{\Pi^{(j)}\}_{j=1}^J\) denote a finite set of POVMs. Suppose that for estimating \(\rho\) we independently perform a measurement described by a POVM \(\Pi^{(j)} = \{\Pi^{(j)}_m\}_{m=1}^{M^{(j)}}\) a number \(n^{(j)}\) of times \((j = 1, \ldots, J)\). The total number of measurement trials is \(\sum_{j=1}^J n^{(j)} = N\). Let us define \(r^{(j)} := N/n^{(j)}\). Elements of the POVM can also be parametrized as

\[
\Pi^{(j)}_m = a^{(j)}_{m,0} I + a^{(j)}_m \cdot \lambda,
\]

where \(a^{(j)}_{m,0} = \text{Tr}[\Pi^{(j)}_m]/d\), and \(a^{(j)}_{m,\beta} = \text{Tr}[\Pi^{(j)}_m \lambda_\beta]/2\).

When we perform the measurement on a system in a state described by \(\rho\), the probability that we observe an outcome \(m\) is given by

\[
p(m; \Pi^{(j)}, \rho) = \text{Tr}[\Pi^{(j)}_m \rho] = a^{(j)}_{m,0} + a^{(j)}_m \cdot s.
\]

A set of POVMs \(\Pi := \{\Pi^{(j)}\}_{j=1}^J\) is called informationally complete (IC) if it spans the vector space of Hermitian matrices on \(\mathcal{H}\) [12]. Such a set allows for the reconstruction of an arbitrary quantum state, and we will assume that our \(\Pi\) is always IC.

Let \(n^{(j)}_m\) denote the number of appearances of outcome \(m\) in the data from the \(n^{(j)}\) measurement trials of \(\Pi^{(j)}\) \((m = 1, \ldots, M^{(j)})\); then \(f^{(j)}_m := n^{(j)}_m/n^{(j)}\) is the relative frequency. A map from a data set to the space of interest – in this case the space of quantum states – is called an estimator, and an estimation result is called an estimate. One of the simplest is a linear estimator, \(\rho^0\) [13], defined as a matrix \(\sigma\) satisfying

\[
f^{(j)}_m = p(m; \Pi^{(j)}, \sigma), \ j = 1, \ldots, J, \ m = 1, \ldots, M^{(j)}.
\]
The idea is to use the relative frequencies instead of the unknown true probability distributions. This might seem natural to physicists, but there are two problems. The first problem is that when $\bar{\Pi}$ is over-complete, equation (4) might have no solutions, i.e., we happen to obtain a data set from which we cannot calculate the estimate. The second problem is that even if there exists a solution of equation (4), the solution can be unphysical, i.e., lie outside of $S$. For these two reasons linear estimators are rarely used in tomographic experiments anymore. The current standard is a maximum-likelihood (ML) estimator, which is defined as the point in $S$ maximizing the likelihood function [14]. By definition, such estimates are always physical. The asymptotic ($N \sim \infty$) behavior of the confidence level of a ML estimator is analyzed in [17], but there have been no such results for finite data sets.

A loss function is a measure for evaluating the difference between two states. We analyze the following three loss functions:

\[
\Delta_{\text{HS}}(\rho', \rho) := \frac{1}{\sqrt{2}} \text{Tr} \left[ (\rho' - \rho)^2 \right]^{1/2}, \quad (5)
\]

\[
\Delta_{\text{T}}(\rho', \rho) := \frac{1}{2} \text{Tr} \left[ |\rho' - \rho| \right], \quad (6)
\]

\[
\Delta_{\text{IF}}(\rho', \rho) := 1 - \text{Tr} \left[ \sqrt{\sqrt{\rho'} \rho \sqrt{\rho'}} \right]^2. \quad (7)
\]

called the Hilbert-Schmidt distance, the trace distance, and the infidelity, respectively. In current quantum tomography experiments the trace distance and the infidelity are most often used.

### III. RESULTS

Instead of ML estimator, we propose a new estimator $\rho^{\text{ENM}}$. In order to define the new estimator, we introduce a different estimator $\rho^{\text{LLS}}$. Let us define $p(\sigma)$ and $f_N$ as vectors with $(j, m)$-th element $p(m|\Pi^{(j)}, \sigma)$ and $f_{m(j)}$, respectively. The estimate of $\rho^{\text{LLS}}$ is defined as

\[
\rho_N^{\text{LLS}} := \arg\min_{\sigma, \rho} \|p(\sigma) - f_N\|_2. \quad (8)
\]

The range of the minimization in equation (8) is restricted by the Hermiticity and trace-one condition, but the positive-semidefiniteness ($\sigma \geq 0$) is not required. Therefore the estimates can be unphysical. $\rho_N^{\text{LLS}}$ is a linear least squares (LLS) estimator in statistics [16]. Let us define $a_0$ as a vector with $(j, m)$-th element $a_{m, 0}^{(j)}$ and $A$ as a matrix with $(j, m), \alpha$-th element $a_{m, \alpha}^{(j)}$ ($\alpha = 1, \ldots, d^2 - 1$). When $\Pi$ is IC, $A$ is full-rank, and the left-inverse matrix exists and is given by $A_L^{-1} = (A^T A)^{-1} A^T$. Then the minimization in equation (8) has the analytical solution, and the LLS estimate of the Bloch vector, $s_N^{\text{LLS}}$, is given as

\[
s_N^{\text{LLS}} = A_L^{-1}(f_N - a_0). \quad (9)
\]

The LLS estimate of density matrix is calculated by $\rho_N^{\text{LLS}} = \rho(s_N^{\text{LLS}})$. Using the LLS estimator, we define the estimate of a new estimator, $\rho^{\text{ENM}}$, as

\[
\rho_N^{\text{ENM}} := \arg\min_{\rho' \in \mathcal{S}(\mathcal{H})} \|\rho' - \rho_N^{\text{LLS}}\|_2. \quad (10)
\]

We call $\rho^{\text{ENM}}$ an extended norm-minimization (ENM) estimator [17]. The estimates are again always physical by definition.

The following theorem establishes that the ENM estimates are close to the true prepared state with high probability.

**Theorem 1 (Confidence level of ENM estimator)**

Suppose that $\Pi$ is IC. Then for arbitrary true density matrix $\rho \in \mathcal{S}(\mathcal{H})$, set of positive integers $n^{(j)}$ satisfying \( \sum_{j=1}^{J} n^{(j)} = N \), and positive number $\delta$,

\[
\Delta(\rho_N^{\text{ENM}}, \rho) \leq \delta \quad (11)
\]

holds with probability at least

\[
CL := 1 - 2 \sum_{\alpha=1}^{d^2-1} \exp \left[ -\frac{b}{c_\alpha} \delta^2 N \right]. \quad (12)
\]

where $b$ is determined by our choice of the loss function as

\[
b := \begin{cases} 
8/(d^2 - 1) & \text{if } \Delta = \Delta_{\text{HS}} \\
16/d(d^2 - 1) & \text{if } \Delta = \Delta_{\text{T}} \\
4/d(d^2 - 1) & \text{if } \Delta = \Delta_{\text{IF}}
\end{cases}. \quad (13)
\]

and $c_\alpha$ are determined by our choice of the measurement setting as

\[
c_\alpha := \sum_{j=1}^{J} p^{(j)} \left( \max_{m} \left[ A_{L}^{-1} a_{\alpha, (j, m)} \right] - \min_{m} \left[ A_{L}^{-1} a_{\alpha, (j, m)} \right] \right)^2. \quad (14)
\]

We call $CL$ in equation (12) the confidence level of $\rho_N^{\text{ENM}}$ at the (user-specified) error threshold $\delta$.

We sketch the proof here, with the details shown in Appendix. The proof consists of two steps. First, we consider the Hilbert-Schmidt distance case and analyze the probability that we obtain estimates deviating from the true density matrix by more than the error threshold $\delta$. We call this probability the estimation error probability

\[
\Delta(\rho_N^{\text{ENM}}, \rho) \leq \delta \quad (11)
\]

with respect to $\Delta = \Delta_{\text{HS}}$ at the error threshold $\delta$. By using known results in convex analysis, we prove that the estimation error probability of $\rho_N^{\text{ENM}}$ is smaller than that of $\rho_N^{\text{LLS}}$. Second, we derive an upper-bound on the estimation error probability of $\rho_N^{\text{LLS}}$. In the derivation, we reduce the analysis of the probability for multi-parameter estimation to that of the probability for one-parameter estimation and derive a $\rho$-independent upper-bound of the probability with the Hoeffding’s tail inequality [18]. From the result of the first step, the derived upper-bound is also an upper-bound of that of $\rho_N^{\text{ENM}}$. The confidence level of $\Delta_{\text{HS}}$ is given by one minus the upper-bound of the estimation error probability. The confidence levels of $\Delta_{\text{T}}$ and $\Delta_{\text{IF}}$ are derived by combining that of $\Delta_{\text{HS}}$ with inequalities between these loss functions.
IV. ANALYSIS

The most important point in Theorem 1 is that equation (12) is independent of the true prepared state \( \rho \). Therefore we can use it to evaluate \( \Delta(\rho_s, \rho) \) without knowing \( \rho \). Suppose that we choose a loss function \( \Delta \) that satisfies properties of a mathematical distance. Then from the triangle inequality and Theorem 1, we have \( \Delta(\rho, \rho) \leq \Delta(\rho_s, \rho_N^{\text{ENM}}) + \Delta(\rho_N^{\text{ENM}}, \rho) \), \( \leq \Delta(\rho_s, \rho_N^{\text{ENM}}) + \delta \). Equation (13) holds at the confidence level in equation (12). We can calculate the value of the R.H.S. of equation (13) without knowing the true prepared state \( \rho \) and use it to evaluate the size of \( \Delta(\rho_s, \rho) \). In tomographic experiments, the infidelity, (or the fidelity \( F(\rho', \rho) := 1 - \Delta^\text{T}(\rho', \rho) \)), is a popular loss function used for evaluating \( \Delta(\rho_s, \rho_N^{\text{ENM}}) \). The infidelity is not a distance, it does not satisfy the triangle inequality, but it is related to the trace distance by \( \Delta^\text{IF}(\rho, \rho) \leq 2\Delta^\text{T}(\rho, \rho) \). Thus we obtain \( \Delta^\text{IF}(\rho, \rho) \leq 2\{\Delta^\text{T}(\rho_s, \rho_N^{\text{ENM}}) + \delta\} \), \( F(\rho, \rho) \leq 1 - 2\{\Delta^\text{T}(\rho_s, \rho_N^{\text{ENM}}) + \delta\} \), \( \leq 2\{\Delta^\text{T}(\rho_s, \rho_N^{\text{ENM}}) + \delta\} \).

Let us consider quantum state tomography of a k-qubit system and suppose that we make the three Pauli measurements with detection efficiency \( \eta \) on each qubit. There are \( 3^k \) different tensor products of Pauli matrices (\( J = 3^k \)), and suppose that we observe each equally \( n := N/3^k \) times. Let us choose \( A \) to be the set of tensor products of Pauli and identity matrices with the normalization factor \( 1/\sqrt{2^{k-1}} \). From the relation \( s_\alpha = \text{Tr}[\rho \lambda_\alpha] \), we obtain \( c_\alpha = 2^{3-k} \cdot 3^k \cdot \eta^{2(k-1)} \), where \( l = 0, \ldots, k-1 \) is the number of the identity matrices appearing in \( \lambda_\alpha \). From the information above, we can derive the explicit form of \( CL \) for k-qubit state tomography. When we choose \( \Delta = \Delta^\text{T} \), we have

\[
CL(k) = 1 - 2 \sum_{l=0}^{k-1} \frac{3^{k-l} \binom{k}{l}}{\eta^{2(k-l)} \eta^2 N} \leq 2\{\Delta^\text{T}(\rho_s, \rho_N^{\text{ENM}}) + \delta\} 
\]

The details of the derivation of equation (13) are explained in Appendix. Figure 2 shows plots of equation (13) for the one-qubit \( k = 1 \) and two-qubit \( k = 2 \) cases in panels (a) and (b), respectively. The error threshold is \( \delta = 0.07 \) and detection efficiency is \( \eta = 0.9, 0.8 \). Both panels indicate that smaller detection efficiency requires a larger number of prepared states. The plots tell us what value of \( N \) sufficient for guaranteeing a fixed confidence level. For example, if we want to guarantee 99% confidence level for \( \delta = 0.07 \) in one-qubit state tomography with \( \eta = 0.9 \), panel (a) indicates that \( N = 7,500 \) is sufficient for that.

In [20], an efficient ML estimator algorithm is proposed for quantum state tomography using an IC set of projective measurements with Gaussian noise whose variance is known, and numerical results for k-qubit \( k = 1, \ldots, 9 \) state tomography indicate that the computational cost would be significantly lower than that of standard ML algorithms. In general, a ML estimator is different from the ENM estimator, but in the setting considered in [20] the ML estimator is a specific case of an ENM estimator, which is defined for general IC measurements. Despite this generality, we find that their efficient algorithm can be modified and used for our ENM estimates [21]. Additionally, our result (Theorem 1) shows that the ENM estimator can be used without assuming projective measurements or Gaussian noise.
It is natural to ask if instead of performing two sequential optimizations as in the ENM case one performs a single constrained optimization. This is well-known in classical statistics and was applied to a quantum estimation problem in [22]. Define a constrained least squares estimator

$$\rho_N^{\text{CLS}} := \arg\min_{\rho' \in \mathcal{S}(\mathcal{H})} \| p(\rho') - f_N \|_2, \quad (20)$$

which always exists and is always physical. Using nearly the same proof as in Theorem 1, we can derive a confidence level for $\rho_N^{\text{CLS}}$. The result is equivalent or smaller than Eq. (12)—the details and a comparison to $\rho^{\text{ENM}}$ are shown in Appendix E. Although in some cases the confidence levels coincide, in order to calculate CLS estimates we need to solve the quadratic optimization problem under inequality constraints, which the ENM case avoids.

V. SUMMARY

We considered quantum state tomography in arbitrary finite dimensional systems, proposing a new point estimator and deriving an explicit formula evaluating the precision of estimation for an arbitrary finite number of measurement trials. We applied the idea using as an example $k$-qubit state tomography with detection errors, and showed plots for the one- and two-qubit cases. We also show how the formula can be used for evaluating the precision of state preparation, i.e., the difference between the state that we want to prepare and the state that we actually prepared. To the best of our knowledge this is the first result that makes it possible to evaluate the precision of estimation and preparation without knowing the prepared state, and we hope it finds application in the analysis of experimental data.

ACKNOWLEDGMENTS

The authors thank the Referee for drawing their attention to [22]. TS thanks Steven T. Flammia, David Gross, and Fuyuhiko Tanaka for helpful comments on least squares estimators. This work was supported by the JSPS Research Fellowships for Young Scientists (22-7564), the JSPS Postdoctoral Fellowships for Research Abroad (H25-32), and the Project for Developing Innovation Systems of the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan.

[1] ATLAS Collaboration, Combined search for the Standard Model Higgs boson in $pp$ collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector. Phys. Rev. D 86, 032003 (2012).
[2] Shor, P. W. Algorithms for quantum computation: discrete logarithms and factoring, in Proceedings, 35th Annual Symposium on Foundations of Computer Science, IEEE Press (1994).
[3] Bennett, C. H. & Brassard, G. Quantum cryptography: Public key distribution and coin tossing, in Proceedings of IEEE International Conference on Computers, Systems and Signal Processing, 175-179, IEEE Press (1984).
[4] Paris, M. & Reháček, J., eds. Quantum State Estimation. Lecture Notes in Physics (Springer, Berlin, 2004).
[5] Blume-Kohout, R. Robust error bars for quantum tomography. arXiv:1202.5270 [quant-ph] (2012).
[6] Christandl, M. & Renner, R. Reliable quantum state tomography. Phys. Rev. Lett. 109, 120403 (2012).
[7] Flammia, S. T., Gross, D., Liu, Y. -K. & Eisert, J. Quantum tomography via compressed sensing: error bounds, sample complexity and efficient estimators. New J. Phys. 14, 095022 (2012).
[8] Flammia, S. T. & Liu, Y. -K. Direct fidelity estimation from few Pauli measurements. Phys. Rev. Lett. 106, 230501 (2011).
[9] Da Silva, M. P., Landon-Cardinal, O. & Poulin, D. Practical characterization of quantum devices without tomography. Phys. Rev. Lett. 107, 210404 (2011).
[10] Kimura, G. The Bloch vector for $N$-level systems. Phys. Lett. A 314, 339-349 (2003).
[11] Byrd, M. S. & Khaneja N. Characterization of the positivity of the density matrix in terms of the coherence vector representation. Phys. Rev. A 68, 062322 (2003).
[12] Prugovečki, E. Information-theoretical aspects of quantum measurement. Int. J. Theor. Phys. 16, 321-331 (1977).
[13] Fano, U. Description of states in quantum mechanics by density matrix and operator techniques. Rev. Mod. Phys. 29, 74-93 (1957).
[14] Hradil, Z., Quantum-state estimation. Phys. Rev. A 55, R1561-R1564 (1997).
[15] Sugiyama, T., Turner, P. S. & Murao, M. Error probability analysis in quantum tomography: A tool for evaluating experiments. Phys. Rev. A 83, 012105 (2011).
[16] Rao, C. R., Toutenburg, H., Shalabh & Heumann, C. Linear Models and Generalizations: Least Squares and Alternatives. Springer Series in Statistics (Springer, Berlin, 2008).
[17] The ENM estimator is categorized into a weighted least squares estimator with nonlinear constraints.
[18] Hoeffding, W. Probability inequalities for sums of bounded random variables. J. Am. Stat. Assoc. 58, 13-30 (1963).
[19] Bengtsson, I. & Życzkowski, K. Geometry of Quantum States. (Cambridge Univ. Press, Cambridge, 2006).
[20] Smolin, J. A., Gambetta, J. M. & Smith, G. Efficient method for computing the maximum-likelihood quantum state from measurements with additive gaussian noise. Phys. Rev. Lett. 108, 070502 (2012).
[21] These calculations involve numerical errors, and strictly speaking the calculated estimate is different from the exact ENM estimate. We analyze the effect of numerical (and systematic) errors on error thresholds and confidence levels and give a solution to this problem in Ap-
Appendix

In Appendix A, we give the proof of Theorem 1, explain the generalization to quantum process tomography, and discuss the possible improvement of Theorem 1. In Appendix B, we derive the confidence level for k-qubit state tomography using the Pauli measurements with detection losses. In Appendix C and D, we explain a way of evaluating the effect of systematics and numerical errors on an error threshold and confidence level. In Appendix E, we derive a confidence level of the CLS estimator and compare the performance to that of the ENM estimator.

Appendix A: Proof of Theorem 1

In order to prove Theorem 1 in the manuscript, we introduce four lemmas. Lemmas 1 and 3 are used in proving Lemmas 2 and 4, respectively.

First, we introduce a lemma about projections in Euclidean space, which is known in convex analysis and so we omit the proof. This lemma is used for proving that the ENM estimator gives more precise estimates than those of the LLS estimator.

Lemma 1 (Non-expandability of projection [23])
Let \( K \) denote a positive integer and \( V \) denote a non-empty closed convex subset of \( \mathbb{R}^K \). In convex analysis, a vector \( v := \arg\min_{v' \in V} \|v' - w\|_2 \) is called the projection of \( w \in \mathbb{R}^K \) into \( V \), and it is known that for any \( v'' \in V \),
\[
    \|v - v''\|_2 \leq \|w - v''\|_2 \quad \text{(A1)}
\]
holds.

Second, by using Lemma 1 we prove that the ENM estimator gives more precise estimates than those of the LLS estimator. Let us define \( \mathbb{P} \left[ \Delta(\rho^N_{\text{ENM}}, \rho) > \delta \right] \) as the probability that we obtain estimates deviating from the true density matrix by more than \( \delta \) with respect to a loss function \( \Delta \). We call \( \mathbb{P} \left[ \Delta(\rho^N_{\text{ENM}}, \rho) > \delta \right] \) the estimation error probability of estimator \( \rho^N_{\text{ENM}} \) with respect to loss function \( \Delta \). Then the following lemma holds:

Lemma 2 Suppose that \( \mathcal{H} \) is IC. For arbitrary \( \rho \in \mathcal{S}(\mathcal{H}) \), set of positive integers \( n^{(j)} \) satisfying \( \sum_{j=1}^J n^{(j)} = N \), and positive number \( \delta \),
\[
    \mathbb{P} \left[ \Delta_{\text{HS}}(\rho^N_{\text{ENM}}, \rho) > \delta \right] \leq \mathbb{P} \left[ \Delta_{\text{HS}}(\rho^N_{\text{LLS}}, \rho) > \delta \right] \quad \text{(A2)}
\]
holds.

Proof (Lemma 2) The Hilbert-Schmidt distance and \( \ell_2 \)-distance on \( \mathbb{R}^{d-1} \) in the Bloch representation are related by \( \Delta_{\text{HS}}(\rho', \rho) = \frac{1}{2} \|s' - s\|_2 \). The definition of \( \rho^N_{\text{ENM}} \) is rewritten in the vector representation as
\[
    s^N_{\text{ENM}} := \arg\min_{s' \in S} \|s' - s^N_{\text{LLS}}\|_2. \quad \text{(A3)}
\]
Equation (A3) indicates that \( s^N_{\text{ENM}} \) is the projection of \( s^N_{\text{LLS}} \) into the convex subset \( S \), and by substituting \( K = K \),
\[d^2 - 1, \ \mathbf{v} = \mathbf{s}_{N,\text{ENM}}, \ \mathbf{w} = \mathbf{s}_{N,\text{LLS}}\], and \(\mathbf{v''} = \mathbf{s}\) into equation (A11), we obtain \(\|\mathbf{s}_{N,\text{ENM}} - \mathbf{s}\|_2 \leq \|\mathbf{s}_{N,\text{LLS}} - \mathbf{s}\|_2\). Then we have
\[
\|\mathbf{s}_{N,\text{ENM}} - \mathbf{s}\|_2 > 2\delta \implies \|\mathbf{s}_{N,\text{LLS}} - \mathbf{s}\|_2 > 2\delta, \quad (A4)
\]
and equation (A2) holds. ■

Lemma 2 indicates that the estimation error probability of \(\rho_{\text{ENM}}\) with respect to \(\Delta_{\text{HS}}\) is always smaller than or equivalent to that of \(\rho_{\text{LLS}}\).

Third, we introduce a lemma about an error probabilities of a sum of independent random variables, which is known in statistics and we omit the proof.

**Lemma 3 (Hoeffding’s tail probability inequality)**

Let \(X_1, \ldots, X_N\) be independent bounded random variables such that \(X_i\) takes value in \([t_i, u_i]\) with probability one. Let us define \(S_N := \sum_{i=1}^N X_i\). Then for any \(\delta > 0\) we have

\[
\mathbb{P}[|S_N - \mathbb{E}[S_N]| > \delta] \leq 2 \exp \left( - \frac{2\delta^2}{\sum_{i=1}^N (u_i - t_i)^2} \right) \quad (A5)
\]

where \(\mathbb{E}[S_N]\) is the expectation of the random variable \(S_N\).

Lemma 3 is used for deriving an upper-bound on an estimation error probability of \(\mathbf{s}_{\text{LLS}}\).

Fourth, by using Lemma 3, we derive an upper-bound on the estimation error probability of \(\mathbf{s}_{\text{LLS}}\) with respect to the \(\ell_\infty\)-distance.

**Lemma 4** For arbitrary \(\mathbf{s} \in S\), set of positive integers \(n^{(j)}\) satisfying \(\sum_{j=1}^J n^{(j)} = N\), and positive number \(\delta\),

\[
\mathbb{P}[\|\mathbf{s}_{N,\text{LLS}} - \mathbf{s}\|_\infty > \delta] \leq 2 \sum_{\alpha=1}^{d^2-1} \exp \left( - \frac{2}{c_\alpha} \delta^2 N \right) \quad (A6)
\]

holds where

\[
c_\alpha := \sum_{j=1}^J r^{(j)} \left\{ \max_m \left[ \mathbf{A}_{\alpha,\text{LLS}}^{-1} \right]_{n^{(j)},m} - \min_m \left[ \mathbf{A}_{\alpha,\text{LLS}}^{-1} \right]_{n^{(j)},m} \right\}^2 \quad (A7)
\]

**Proof** (Lemma 4) From the explicit form of the LLS estimate, \(\mathbf{s}_{N,\text{LLS}} = \mathbf{A}_{\alpha,\text{LLS}}^{-1} (\mathbf{f}_N - \mathbf{a}_0)\), we have

\[
\mathbf{s}_{N,\text{LLS}} - \mathbf{s} = \mathbf{A}_{\alpha,\text{LLS}}^{-1} (\mathbf{f}_N - \mathbf{p}(\rho)) \quad (A8)
\]

Let \(n^{(j)}\) denote the \(q\)-th outcome in \(n^{(j)}\) trials of the measurement \(\Pi^{(j)}\) \((j = 1, \ldots, J, q = 1, \ldots, n^{(j)})\). Then the \(\alpha\)-th element of the first term in the R. H. S. of equation (A8) is rewritten as

\[
\left[ \mathbf{A}_{\alpha,\text{LLS}}^{-1} \right]_{\alpha,\text{LLS}} = \sum_{j,m} \left[ \mathbf{A}_{\alpha,\text{LLS}}^{-1} \right]_{\alpha,\text{LLS}} \frac{n^{(j)}_{m}}{n^{(j)}} \quad (A9)
\]

The R. H. S. of equation (A10) is a sum of independent (and non-identical) random variables, and the expectation coincides with the second term in the R. H. S. of equation (A8). Each random variable in the sum takes value in

\[
\left[ \frac{1}{n^{(j)}} \min_m \left[ \mathbf{A}_{\alpha,\text{LLS}}^{-1} \right]_{\alpha,\text{LLS}} \cdot \max_m \left[ \mathbf{A}_{\alpha,\text{LLS}}^{-1} \right]_{\alpha,\text{LLS}} \right] \quad (A11)
\]

By applying Lemma 3 to these random variables, we obtain

\[
\mathbb{P}[\|\mathbf{s}_{N,\text{LLS}} - \mathbf{s}\|_\infty > \delta] \leq 2 \exp \left( - \frac{2\delta^2}{c_\alpha} \right) \quad (A12)
\]

where

\[
c'_{\alpha} := \sum_{j,q} \left\{ \frac{\max_m \left[ \mathbf{A}_{\alpha,\text{LLS}}^{-1} \right]_{\alpha,\text{LLS}}}{n^{(j)}} - \frac{\min_m \left[ \mathbf{A}_{\alpha,\text{LLS}}^{-1} \right]_{\alpha,\text{LLS}}}{n^{(j)}} \right\}^2
\]

\[
= \frac{c_{\alpha}}{N} \quad (A13)
\]

By definition of the \(\ell_\infty\)-distance, \(\|\mathbf{s}_{N,\text{LLS}} - \mathbf{s}\|_\infty > \delta\) indicates that there exists at least one \(\alpha \in \{1, \ldots, d^2 - 1\}\) such that \(\|\mathbf{s}_{N,\text{LLS}} - \mathbf{s}\|_\infty > \delta\). Then we have

\[
\mathbb{P}[\|\mathbf{s}_{N,\text{LLS}} - \mathbf{s}\|_\infty > \delta] \leq \sum_{\alpha=1}^{d^2-1} \mathbb{P}[\|\mathbf{s}_{N,\text{LLS}} - \mathbf{s}\|_\infty > \delta] \quad (A14)
\]

By combining equations (A12), (A13), and (A14), we obtain equation (A6). ■

Finally, we prove Theorem 1 by combining Lemmas 2 and 3.

**Proof** (Theorem 1) Suppose that for two loss functions \(\Delta\) and \(\Delta'\), there exists a constant \(c > 0\) such that \(\Delta(\sigma, \sigma') \leq c \cdot \Delta'(\sigma, \sigma')\) holds for any Hermitian matrices \(\sigma\) and \(\sigma'\). Let \(\rho_{\text{est}}\) denote an estimator. Then for any \(\rho \in S(\mathcal{H})\), \(\Delta(\rho_{\text{est}}^\mathcal{H}, \rho) > \delta\) implies \(\Delta'(\rho_{\text{est}}^\mathcal{H}, \rho) > \delta/c\), and for arbitrary \(\rho \in S(\mathcal{H})\), estimator \(\rho_{\text{est}}^\mathcal{H}\), set of positive integers \(n^{(j)}\) satisfying \(\sum_{j=1}^J n^{(j)} = N\), and positive number \(\delta\),

\[
\mathbb{P}[\Delta(\rho_{\text{est}}^\mathcal{H}, \rho) > \delta] \leq \mathbb{P}[\Delta'(\rho_{\text{est}}^\mathcal{H}, \rho) > \delta/c] \quad (A15)
\]

holds. The Hilbert-Schmidt distance and \(\ell_\infty\)-distance satisfy the following inequality \(\|\mathbf{s} - \mathbf{s}'\|_\infty \leq \sqrt{\frac{2d^2 - 1}{2}} \|\mathbf{s} - \mathbf{s}'\|_\infty\).

From equations (A15) and (A16), we obtain

\[
\mathbb{P}[\Delta(\rho_{\text{est}}^\mathcal{H}, \rho) > \delta] \leq \mathbb{P}[\|\mathbf{s}_{N,\text{LLS}} - \mathbf{s}\|_\infty > \sqrt{\frac{2d^2 - 1}{2}}] \quad (A17)
\]

By combining equations (A2), (A17), and (A6), we obtain

\[
\mathbb{P}[\Delta(\rho_{\text{est}}^\mathcal{H}, \rho) > \delta] \leq 2 \sum_{\alpha=1}^{d^2-1} \exp \left( - \frac{8}{d^2 - 1} \delta^2 N \right) \quad (A18)
\]
The trace distance and the infidelity satisfy the following inequalities [19, 24]:

\[
\Delta^T(\rho', \rho) \leq \sqrt{\frac{d}{2}} \Delta_{HS}(\rho', \rho),
\]

\[
\Delta^I(\rho', \rho) \leq \sqrt{2d} \Delta_{HS}(\rho', \rho),
\]

and from equation (A15) we obtain Theorem 1. ■

Like quantum state tomography, quantum process tomography, which estimates quantum processes, is an important tool in quantum information experiments. A quantum process is described by a linear, trace-preserving, and completely positive map. By using the Choi-Jamiolkowski isomorphism [25, 26], this map can be represented by a density matrix. Therefore our result of quantum state tomography can be generalized to quantum process tomography straightforwardly.

In the proof of Lemma 2, we proved that the estimation error probability of \(\rho_{ENM}^N\) is always smaller than that of \(\rho_{LS}^N\) and the confidence level in Theorem 1 is about \(\rho_{LS}^N\). Therefore a property of the projection in \(\rho_{ENM}^N\) is used in the proof, but the effect is not included in the confidence level in Theorem 1. The effect of the projection becomes larger as the prepared state is closer to the boundary of the state space. Usually it is expected that states close to the boundary would not affect on the value of the confidence level because a confidence level is derived by analyzing the state that is most difficult to estimate in all possible prepared states and it would be around the completely mixed state, which is far from the boundary. Therefore in many cases we could expect that the projection in \(\rho_{ENM}^N\) does not affect the confidence level. However, if we choose a loss function that put a high weight on states close to the boundary, like infidelity, the above expectation might not be true. The projection is a nonlinear map and the detailed analysis of the effect on the confidence level is a theoretically challenging problem.

In the proofs explained above, we used five inequalities for deriving the confidence level for the trace distance. The equalities of these inequalities do not hold simultaneously. This means that we under-evaluate the confidence levels of \(\rho_{LS}^N\) and \(\rho_{ENM}^N\) and there must be possible improvement of Theorem 1. An improved confidence level guarantees a higher reliability of the estimation results for same data. This is an open problem that is important for theorists and experimentalists both.

One interesting and natural question is how close the confidence level in Theorem 1 is from the optimal value. For this comparison, we need to calculate the optimal value of the confidence level. It seems to be very hard to analytically solve this problem, and a numerical analysis using a Monte Carlo method might be considered as a practical approach. In the analysis of the estimation error probability, however, we need to analyze rare events, \(\{\Delta(\rho_{ENM}^N, \rho) > \delta\}\), whose probability decreases exponentially fast with respect to \(N\), and Monte Carlo methods require high computational cost in order to accurately evaluate the probability of such very rare events. Furthermore, in order to calculate the optimal confidence level, we need to perform the maximization of the estimation error probability over all possible states. This optimization makes the computational cost higher. Therefore the comparison of our result to the optimal value is both a theoretically and also probably a numerically challenging open problem. On the other hand, the difficulty of the optimality analysis implies an advantage of our approach and result. That is, the confidence level in Theorem 1 might be far from the optimal value that is very hard to compute, but in exchange for optimality the confidence level in Theorem 1 is very easy to calculate.

Appendix B: Confidence level for \(k\)-qubit state tomography

We derive the confidence level for a \(k\)-qubit state tomography explained in the manuscript.

Suppose that we prepare \(N\) identical copies of \(\rho \in \mathcal{S}((\mathbb{C}^2)^\otimes k)\) and make the three Pauli measurements with detection efficiency \(\eta\) on each qubit. The POVMs describing the ideal Pauli measurements on each qubit, \(\Pi^{(i)} = \{\Pi^{(i)}_{+1}, \Pi^{(i)}_{-1}\}\), are given as

\[
\Pi^{(i)}_{\pm1} := \frac{1}{2} (I \pm e_i \cdot \sigma),
\]

\[
\Pi^{(i)}_{0} := \frac{1}{2} (I \pm e_i \cdot \sigma) = \frac{1}{2} (I \pm e_i \cdot \sigma),
\]

where \(i = 1, 2, 3,\)

\[
e_1 := \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad e_2 := \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad e_3 := \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix},
\]

and

\[
\sigma_1 := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 := \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.
\]

When the measurements have detection loss, the corresponding POVMs, \(\Pi^{(i)}_{\eta} = \{\Pi^{(i)}_{+1}, \Pi^{(i)}_{-1}, \Pi^{(i)}_{0}\}\), are given as

\[
\Pi^{(i)}_{\eta \pm1} := \frac{\eta}{2} (I \pm e_i \cdot \sigma),
\]

\[
\Pi^{(i)}_{\eta 0} := (1 - \eta) I,
\]

where \(\eta\) is the detection efficiency and takes the value from 0 to 1. The outcome “0” means no detection at all. The outcome “+1” means a detection at the +1 detction efficiency.

We use the notation \(\Pi^{(i)}_{\eta \pm1} := \bigotimes_{q=1}^k \Pi^{(q)\eta}_{\pm1}\) for the imperfect Pauli measurements on each qubit, and the POVM on \(k\)-qubit state is given as

\[
\Pi^{(i)}_{\eta} := \bigotimes_{q=1}^k \Pi^{(q)\eta}_{(q)\eta},
\]

where \(i = \{i_q\}_{q=1}^k\) and \(i_q = 1, 2, 3\). The label of the different POVMs, \(i\), corresponds to \(j\) in the manuscript. Suppose that we perform each measurement described by \(\Pi^{(i)}_{\eta}\) equally and \(n := N/3^k\) times.
Let us choose $\lambda$ to be the set of tensor products of Pauli and identity matrices with the normalization factor $1/\sqrt{2^{k-1}}$, i.e.,

$$\lambda_{\beta} := \frac{1}{\sqrt{2^{k-1}}} \otimes_{q=1}^{k} \sigma_{\beta_q},$$

where $\beta := \{\beta_q\}_{q=1}^{k}$ and $\beta_q = 0, 1, 2, 3$. We eliminate from $\beta$ the case that all $\beta_q$ are 0. The label of the matrices, $\beta$, corresponds to $\alpha$ in the manuscript. Using this $\lambda$, any density matrices are represented as

$$\rho = \frac{1}{2\pi} I + \frac{1}{2} \lambda \cdot s,$$  

where

$$s_{\beta} = \text{Tr} [\rho \lambda_{\beta}]$$

and $\eta = 1$ or $-1$.

Equation (B11) indicates that the parameter $s_{\beta}$ is the expectation of a tensor product of ideal Pauli and identity matrices.

Therefore we have

$$B_{\beta(i,m)} = \pm \frac{1}{\sqrt{2^{k-1}}} \frac{1}{\eta^{k-l} \sqrt{3}!},$$

if $i_q = 1, 2, 3$ and $m_q = \pm 1, 0$ for $q = 1, \ldots, l$ and $i_q = \beta_q$ and $m_q = \pm 1$ for $q = l + 1, \ldots, k$. Otherwise $B_{\beta(i,m)} = 0$. Then for each $\beta$ and $i$,

$$\max_{m} B_{\beta(i,m)} - \min_{m} B_{\beta(i,m)}$$

holds, and we obtain

$$c_{\beta} = \sum_{i} 3^k \left( \max_{m} B_{\beta(i,m)} - \min_{m} B_{\beta(i,m)} \right)^2$$

(B18)

$$= \sum_{i_1, \ldots, i_l} 3^k \left( \frac{2}{\sqrt{2^{k-1}} \eta^{k-l} \sqrt{3}!} \right)^2$$

(B19)

$$= \frac{3^{k-l}}{2^{k-3} \eta^{2(k-l)}}.$$  

(B20)

From the above discussion, we can see that $c_{\beta}$ takes same value for different $\lambda_{\beta}$ with the same $l$. The confidence level is calculated as

$$CL(k) = 1 - 2 \sum_{\beta} \exp \left[ -\frac{b}{c_{\beta}} (\frac{2}{3})^2 N \right]$$

(B21)
When we choose the trace distance as the loss function, we have
\[
 b = \frac{16}{d(d^2 - 1)} = \frac{1}{2^{k-1} \cdot (2^{2k} - 1)},
\]
and
\[
 CL(k) = 1 - 2 \sum_{l=0}^{k-1} 3^{k-l} \binom{k}{l} \exp \left[ -\frac{2}{9} \eta^2 (k-l) \delta^2 N \right].
\]
(B23)

In one-qubit \((k = 1)\) and two-qubit \((k = 2)\) cases, we have
\[
 CL(k = 1) = 1 - 6 \exp \left[ -\frac{2}{9} \eta^2 \delta^2 N \right],
\]
(B24)
\[
 CL(k = 2) = 1 - 18 \exp \left[ -\frac{2}{135} \eta^4 \delta^2 N \right] - 12 \exp \left[ -\frac{2}{45} \eta^2 \delta^2 N \right].
\]
(B25)

As in the above discussion, when the directions of each Pauli measurement are perfectly orthogonal, it is easy to derive \(c_{β}\). When the directions are not orthogonal, we need to calculate \(A_L^{-1} = (A_T A)^{-1} A^T\). Then, it becomes more difficult to analyze \(c_{β}\), and we might need to calculate them numerically.

**Appendix C: Effect of systematic errors**

Theorems 1 is valid for any informationally complete POVMs and is applicable for cases in which a systematic error exists. However, we must know exactly the mathematical representation of the systematic error in order to strictly verify a value of the confidence level. This assumption can be unrealistic in some experiments. In this section, we will weaken the assumption to a more realistic condition and give a formula of confidence levels in such a case.

Let \(\tilde{\Pi}\) denote a set of POVMs exactly describing the measurement used, and let \(\tilde{\Pi}' (\neq \tilde{\Pi})\) denote a set of POVMs that we mistake as the correct set of POVMs. We assume that \(\tilde{\Pi}\) and \(\tilde{\Pi}'\) are both informationally complete. Suppose that we do not know \(\tilde{\Pi}\), but we know that \(\tilde{\Pi}\) is in a known set \(\mathcal{M}\). For example, consider the case where an experimentalist wants to perform a projective measurement of \(σ_1\). If they can guarantee that their actual measurement is prepared within 0.5 degrees from the \(x\)-axis, and if their detection efficiency is 0.9, then \(\mathcal{M}\) is the set of all POVMs whose measurement direction and detection efficiency are within 0.5 degrees of the \(x\)-axis and 0.9, respectively.

For given relative frequencies \(f_N\), the correct and mistaken LLS estimates are
\[
 s_{N}^{\text{LLS}} = A_L^{-1}(\tilde{\Pi}) \{ f_N - a_0 \}, \quad (C1)
\]
\[
 s_{N}^{\text{LLS'}} = A_L^{-1}(\tilde{\Pi}') \{ f_N - a_0' \}. \quad (C2)
\]
Then the actual and mistaken ENM estimates are
\[
 s_{N}^{\text{ENM}} = \arg \min_{s' \in \tilde{\Pi}} \| s' - s_{N}^{\text{LLS}} \|_2 , \quad (C3)
\]
\[
 s_{N}^{\text{ENM'}} = \arg \min_{s' \in \tilde{\Pi}'} \| s' - s_{N}^{\text{LLS'}} \|_2. \quad (C4)
\]
Let \(ρ_{N}^{\text{ENM}}\) and \(ρ_{N}^{\text{ENM'}}\) denote the corresponding density matrix estimates. Let us define the size of the systematic error as
\[
 ξ := \max_{\tilde{n} \in \mathcal{M}} \Delta(ρ_{N}^{\text{ENM}}, ρ_{N}^{\text{ENM'}}). \quad (C5)
\]
This is a function of \(Δ, f_N, \tilde{\Pi}',\) and \(\mathcal{M}\). Then for any \(ρ \in \mathcal{S}(\mathcal{H})\) and \(\tilde{\Pi} \in \mathcal{M},\)
\[
 Δ(ρ_*, ρ) \leq Δ(ρ_*, ρ_{N}^{\text{ENM}}) + Δ(ρ_{N}^{\text{ENM}}, ρ_{N}^{\text{ENM'}}) + Δ(ρ_{N}^{\text{ENM}}, ρ) \leq Δ(ρ_*, ρ_{N}^{\text{ENM'}}) + ξ + δ \quad (C6)
\]
holds with probability at least
\[
 \min_{\tilde{n} \in \mathcal{M}} CL = 1 - 2 \max_{\tilde{n} \in \mathcal{M}} \sum_{s=1}^{d^2-1} \exp \left[ -\frac{b}{c_a} \delta^2 N \right]. \quad (C7)
\]
Using equations \((C6)\) and \((C7)\), we can evaluate the precision of state preparation, \(Δ(ρ_*, ρ)\), without knowing the true state \(ρ\) and true sets of POVMs \(\tilde{\Pi}\).

**Appendix D: Effect of numerical errors**

In this section, we analyze the effect of numerical errors and explain a method for evaluating the precision of the state preparation in the cases that numerical errors exist.

The ENM estimator \(ρ_{N}^{\text{ENM}}\) requires a nonlinear minimization, which requires the use of a numerical algorithm. Suppose that we choose an algorithm for the minimization and obtain a result \(σ_{N}^{\text{ENM}}\) for a given data set. In practice, there exists a numerical error on the result, and \(σ_{N}^{\text{ENM}}\) differs from the exact solution \(ρ_{N}^{\text{ENM}}\). We cannot obtain the exact solution, but we can guarantee the accuracy of the numerical result with accuracy-guaranteed algorithms \((27)\). Suppose that we use an algorithm for which \(Δ(σ_{N}^{\text{ENM}}, ρ_{N}^{\text{ENM}}) \leq ζ\) is guaranteed. Then
\[
 Δ(ρ_*, ρ) \leq Δ(ρ_*, σ_{N}^{\text{ENM}}) + Δ(σ_{N}^{\text{ENM}}, ρ_{N}^{\text{ENM}}) + Δ(ρ_{N}^{\text{ENM}}, ρ) \leq Δ(ρ_*, σ_{N}^{\text{ENM}}) + ζ + δ \quad (D1)
\]
holds at the confidence level in Theorem 1. The error threshold is changed from \(δ\) to \(ζ + δ\).

Usually systematic and numerical errors both exists. In such a case, by combining equations \((C6)\) and \((D1)\), we can prove that the inequality
\[
 Δ(ρ_*, ρ) \leq Δ(ρ_*, σ_{N}^{\text{ENM}}) + ζ + ξ + δ \quad (D2)
\]
holds at the confidence level in equation \((C7)\), where \(ζ\) is a numerical error threshold for \(\tilde{\Pi}'\). Therefore Theorem 1 with a modification can apply for the cases that systematic and numerical errors exist.
Appendix E: Confidence level of constrained least squares estimator

In this section, we derive a confidence level for it using almost the same argument in Appendix A. By definition, the probability distribution of \( \rho_{N}^{LLS} \) is the projection of \( f_N \) on the probability space of trace-one Hermitian matrices \( \{|p(\sigma)|\sigma = \sigma^t, \text{Tr}[\sigma] = 1\} \), and we have

\[
\|p(\rho') - f_N\|_2^2 = \|p(\rho') - p(\rho_{N}^{LLS})\|_2^2 + \|p(\rho_{N}^{LLS}) - f_N\|_2^2, \quad \forall \rho' \in \mathcal{S}(\mathcal{H})
\]

(E1)

Therefore, Eq. (20) is rewritten as

\[
\rho_{N}^{CLS} = \arg\min_{\rho' \in \mathcal{S}(\mathcal{H})}\|p(\rho') - p(\rho_{N}^{LLS})\|_2,
\]

(E2)

and \( \rho_{N}^{CLS} \) is the projection of \( \rho_{N}^{LLS} \) on \( \mathcal{S}(\mathcal{H}) \) with respect to the 2-norm on the probability space. We can see from Eqs. (11) and (22) that \( \rho_{N}^{ENM} \) and \( \rho_{N}^{CLS} \) are the projections of \( \rho_{N}^{LLS} \) with respect to difference spaces (or different norms).

Using Lemma 1, we obtain

\[
\|p(\rho_{N}^{CLS}) - p(\rho)\|_2 \leq \|p(\rho_{N}^{LLS}) - p(\rho)\|_2, \quad \forall \rho \in \mathcal{S}(\mathcal{H})
\]

(E3)

\[
\|A(s_{N}^{CLS} - s)\|_2 \leq \|A(s_{N}^{LLS} - s)\|_2, \quad \forall s \in \mathcal{S},
\]

(E4)

where \( s_{N}^{CLS} \) is the Bloch vector corresponding to \( \rho_{N}^{CLS} \). Let us define \( \|A\|_{\text{max}} \) and \( \|A\|_{\text{min}} \) as

\[
\|A\|_{\text{max}} := \max_{v \neq 0} \frac{\|Av\|_2}{\|v\|_2}, \quad \text{(E5)}
\]

\[
\|A\|_{\text{min}} := \min_{v \neq 0} \frac{\|Av\|_2}{\|v\|_2}. \quad \text{(E6)}
\]

When \( \mathbf{H} \) is informationally complete, \( A \) is full-rank and \( \|A\|_{\text{min}} > 0 \). We have

\[
\|A\|_{\text{min}} \cdot \|s_{N}^{CLS} - s\|_2 \leq \|A(s_{N}^{CLS} - s)\|_2 \quad \text{(E7)}
\]

\[
\|A\|_{\text{max}} \cdot \|s_{N}^{LLS} - s\|_2 \leq \|A\|_{\text{max}} \cdot \|s_{N}^{LLS} - s\|_2 \quad \text{(E8)}
\]

\[
\|A\|_{\text{max}} \cdot \|s_{N}^{LLS} - s\|_2 \leq \|A\|_{\text{max}} \cdot \|s_{N}^{LLS} - s\|_2. \quad \text{(E9)}
\]

We obtain

\[
\mathbb{P}[\|s_{N}^{CLS} - s\|_2 > \delta] \leq \mathbb{P}\left[\|s_{N}^{LLS} - s\|_2 > \frac{\|A\|_{\text{min}}}{\|A\|_{\text{max}}}\right]. \text{(E10)}
\]

From the same logic, we obtain the following theorem:

**Theorem 2 (Confidence level of CLS estimator)**

*Under the same conditions and notation in Theorem 1*

\[
\Delta(\rho_{N}^{CLS}, \rho) \leq \delta \quad \text{(E11)}
\]

holds with probability at least

\[
CL_{N}^{CLS} := 1 - 2 \sum_{a=1}^{d^2-1} \exp \left[ -\left( \frac{\|A\|_{\text{min}}}{\|A\|_{\text{max}}} \right)^2 \frac{b}{c \pi} \right]. \text{(E12)}
\]

Compared to the confidence level of the ENM estimator (Eq. (12)), there is an additional factor \( \left( \frac{\|A\|_{\text{min}}}{\|A\|_{\text{max}}} \right)^2 \) (\( \leq 1 \)) in the rate of exponential decrease in Eq. (E12). When \( \|A\|_{\text{max}} = \|A\|_{\text{min}} \) holds, Eq. (E12) coincides with Eq. (12). Roughly speaking, the condition \( \|A\|_{\text{max}} = \|A\|_{\text{min}} \) implies that we perform measurements extracting information of each Bloch vector element with an equivalent weight. When \( \|A\|_{\text{max}} > \|A\|_{\text{min}} \), the confidence level of \( \rho_{N}^{CLS} \) is smaller than that of \( \rho_{N}^{ENM} \). This does not mean we can immediately conclude that \( \rho_{N}^{CLS} \) is less precise than \( \rho_{N}^{ENM} \) because their confidence levels, Eqs. (12) and (E12), are probably not optimal as explained in the end of Appendix A. However, we can say that \( \rho_{N}^{CLS} \) is less precise than \( \rho_{N}^{ENM} \) insofar as Eqs. (12) and (E12) are the only confidence levels known for point estimators in quantum tomography to date. Additionally, the computational cost of \( \rho_{N}^{ENM} \) can be smaller than that of \( \rho_{N}^{CLS} \) as mentioned in the Letter. Therefore, we believe that the ENM estimator performs better than the CLS estimator and is at present our best choice.