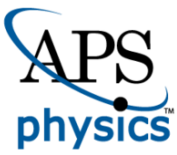

Self-consistent quantum tomography with regularization

Takanori Sugiyama, Shinpei Imori, and Fuyuhiko Tanaka

Phys. Rev. A **103**, 062615 — Published 24 June 2021

DOI: 10.1103/PhysRevA.103.062615

Self-Consistent Quantum Tomography with Regularization

Takanori Sugiyama*,[1,*] Shinpei Imori,[2,†] and Fuyuhiko Tanaka[3,4,‡]

[1] *Research Center for Advanced Science and Technology, The University of Tokyo, 4–6–1 Komaba Meguro-ku, Tokyo Japan, 153–8904.*
[2] *Graduate School of Advanced Science and Engineering, Hiroshima University, 1–3–1 Kagamiyama Higashi-Hiroshima-shi, Hiroshima Japan, 739-8526.*
[3] *Department of Systems Innovation, Graduate School of Engineering Science, Osaka University, 1–3 Machikaneyama-chou Toyonaka-shi, Osaka Japan, 560–8531.*
[4] *Quantum Information and Quantum Biology Division, Institute for Open and Transdisciplinary Research Initiatives, Osaka University, 1–3 Machikaneyama-chou Toyonaka-shi, Osaka Japan, 560–8531.*

(Dated: June 8, 2021)

Quantum tomography is a class of characterization methods frequently used in current experiments, but its standard protocols suffer from unreliability originated from pre-knowledge assumptions. Self-consistent quantum tomography is an approach to avoid the problem, which treats every quantum operations in a characterization experiment as unknown objects to be characterized. As compensation for the beneficence, it leads to a problem that its characterization results cannot be determined uniquely only from experimental data due to the existence of experimentally undetectable gauge degrees of freedom, and we need to introduce a criterion to fix the gauge. Here, we propose to use a regularization technique to fix the gauge. First, we derive a sufficient condition on a characterization experiment to obtain all information of objects to be characterized except for the gauge. Second, we propose a self-consistent data-processing method with regularization and physicality constraints. A careless use of regularization can lead non-negligible bias on the characterization result. As a solution for the concern, we propose a concrete way to tune the strength of the regularization, and mathematically prove that the method provides characterization results that converge to the gauge-equivalence class of the quantum operations of interest at the limit of data going to infinity. The asymptotic convergence guarantees the reliability of the method. We also derive the asymptotic convergence rate, which would be optimal. These theoretical results hold for any finite dimensional quantum systems. Finally, as its first numerical implementation, we show numerical results on 1-qubit system, which confirm the theoretical results and prove that the method proposed is practical.

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

I. INTRODUCTION

As error rates of elementary quantum operations implemented in recent experiments approach a fault-tolerant threshold of a surface code [1], it becomes more important to develop more reliable methods for characterizing their actions to validate and to further improve their accuracies. Standard randomized benchmarking (RB) [2–6] and the relatives [7–14] are efficient methods specified for estimating an accuracy parameter like the average gate fidelity, except for a tomographic RB protocol [15] for multi-parameter estimation. Although they are frequently used in current experiments, recent numerical work revealed that a non-negligible bias can exist in the estimation results in realistic experimental settings [16–18]. Standard quantum tomography (QT) [19–26] are methods for estimating full information of state preparations, measurements, or gates. They are also popular in experiments but have two disadvantages, exponentially growing costs of implementation and inevitable biases caused by unknown imperfections in experiments. When we restrict the use of QT to small subsystems like a few qubits, the high implementation costs do not pose a problem. If the biases stem from finiteness of data size, we can make their effects as small as necessary by increasing the size. However, in a realistic scenario, the biases in estimation results of RB and QT can survive even at the limit of data size going to infinity, regardless of what kind of data-processing is used. Hence, the possible low reliability of QT and RB can become crucial because the purpose of quantum characterization is to reliably characterize super-accurate operations beyond the fault-tolerance threshold.

Self-consistent quantum tomography (SCQT) [27–29] is an approach towards overcoming the low reliability of standard QT. In the SCQT approach, all quantum operations used in a characterization experiment are treated as unknown objects to be estimated, in contrast to standard QTs that model some of them as known. This makes it possible to avoid biases caused from our pre-knowledge discrepancy between the true unknown objects and assumed models. On the other hand, the approach causes a problem that we cannot uniquely determine the set of

* sugiyama@qc.rcast.u-tokyo.ac.jp
† imori@hiroshima-u.ac.jp
‡ ftanaka@sigmath.es.osaka-u.ac.jp

	GST	RSCQT
Data-fitting optimization	Nonlinear	Nonlinear
Gauge fixing method	Additional gauge optimization	Regularization at data-fitting
Physicality constraints	Not fully taken into account	Fully taken into account
Long gate sequences	Implemented	Not implemented

TABLE I. Comparison of GST and RSCQT. GST is a current representative tomographic method in the self-consistent approach. RSCQT is the method proposed in the paper. Both suffer from nonlinearity of data-fitting, which can cause numerical instability. There are two possible advantages of RSCQT compared to GST. One is that RSCQT does not need additional gauge-optimization, which reduces numerical costs of data-processing. The other is that RSCQT fully takes physicality constraints into account, which is suitable to the accuracy validation step. On the other hand, long gate sequences, which have been implemented in GST to amplify effects of tiny physical errors, have not been implemented in RSCQT yet.

quantum operations only from experimental data even if we have infinite amount of data. This is because there exist experimentally undetectable gauge degrees of freedom [28]. In order to obtain estimates of quantum operations in the setting of SCQT, we have to choose how to fix the gauge. Gate-set tomography (GST) [29] is a current representative method in SCQT, and a software package for performing GST, named pyGSTi, is provided [30]. For the gauge-fixing, GST uses an optimization with respect to a norm over the gauge degrees of freedom [31]. GST has superior features, e.g., it is self-consistent and free from the pre-knowledge errors, there is a method for testing the existence of time-dependent errors with data for the GST experiment, and so on. However, it has at least two problems. First, the data-processing procedure in GST is very complicated, and it becomes hard to theoretically evaluate the estimation error caused by finiteness of data. Second, the optimization is a nonlinear problem, and its numerical implementation suffers from high numerical cost, low numerical stability, and hardness of taking into account physicality constraints. Actually, the current version of pyGSTi can ensure physicality of gates, but physicality of state preparation and measurement (SPAM) are not guaranteed [32]. Such a gauge-fixing method with possibly unphysical results is not suitable for use in the validation step.

Here, we propose a new method based on SCQT with regularization that is used for fixing the gauge degrees of freedom. We call the method Regularized Self-Consistent Quantum Tomography (RSCQT). A careless use of a regularization can lead an non-negligible bias on its characterization result. We propose a method for tuning the regularization in order to avoid the bias problem and prove the validity mathematically and theoretically. Comparisons of RSCQT and GST are summarized in Table I for readers familiar with GST. In Sec. II, settings and notation are explained. Details of settings and notation are explained in Appendices A and B. Sec. III includes three theoretical results. First, we introduce the SCQT method with regularization and mathematically prove a sufficient condition on a characterization experiment to obtain all information of objects to be characterized except for the gauge degrees of freedom. Second, we prove its asymptotic convergence and derive the convergence rate, which are valid for any finite dimensional systems and have been proven for the first time in SCQT methods. Third, we propose a method to extract information of Lindbladian from a characterization result of a gate. Proofs of theorems in Sec. III are given in Appendix C, D, and E. Two statistical techniques, regularization and cross validation, are used in the method proposed, and their brief explanations are given in Appendices F and G, respectively. We performed numerical experiments for 1-qubit system, and the numerical results are reported in Sec. IV. Details of the numerical experiments are described in Appendix H. Sec. V is devoted for discussions. We conclude the main text in Sec. VI.

II. SETTING AND NOTATION

We consider a characterization problem of quantum operations on a finite-dimensional quantum system in the SCQT approach. Let d denote the dimension of the system. The dimension considered in Sec. III is arbitrary finite, and $d = 2$ in Sec. IV. The purpose of SCQT is to know mathematical representations of a set of unknown state preparations, measurements, and gates that are implemented in a QIP protocol. We use notations ρ, $\mathbf{\Pi}$, and $\mathcal{G}$ for a density matrix for a state preparation, a positive operator-valued measure (POVM) for a measurement, and a linear trace-preserving and completely-positive (TPCP) map for a gate. In the SCQT approach, every states, measurements, and gates used in its characterization experiment are treated as unknown objects to be estimated, under assumptions that (i) d is finite and known, (ii) the numbers of outcomes of measurements are known, (iii) operations are independent of each other, (iv) the actions of operations are identical at any timing during the experiment. For simplicity, we consider cases that the set of quantum operations to be estimated, $\boldsymbol{s}$, consists of single state preparation, single measurement, and multiple gates. Generalizations of theoretical results in Sec. III to cases of multiple state preparations and measurements are straightforward. Let n_{g} denote the number of gates in $\boldsymbol{s}$.

A set of quantum operations can be parametrized with a real Euclidean vector. We identify the set $\boldsymbol{s}$ and the parametrization vector. Details of the parametrization are explained in Appendix A. Let $\mathcal{S}$ denote the physical

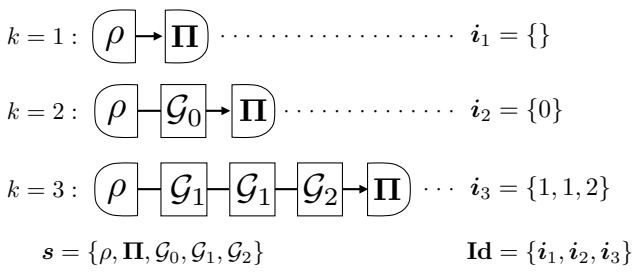

FIG. 1. Example of an experimental schedule. A set of quantum operations $\boldsymbol{s}$ consists of ρ, $\mathbf{\Pi}$, $\mathcal{G}_0$, $\mathcal{G}_1$, and $\mathcal{G}_2$. In this case, the experimental schedule $\mathbf{Id}$ consists of three index sequences, $\boldsymbol{i}_1$, $\boldsymbol{i}_2$, and $\boldsymbol{i}_3$.

region in the Euclidean space. Let $\boldsymbol{s}^{\text{target}} \in \mathcal{S}$ denote the ideal, noiseless, and known set of quantum operations that we aim to implement in a lab. An implemented set, say $\boldsymbol{s}^{\text{true}} \in \mathcal{S}$, is unknown, noisy, and different from $\boldsymbol{s}^{\text{target}}$ because of imperfections on experimental devices.

We perform a set of experiments for estimating $\boldsymbol{s}^{\text{true}}$, which consists of many different combinations of a state preparation, gate sequences, and a measurement. In the SCQT approach, the range of possible choices of the combinations is extremely wide. For example, GST chooses complicated combinations [31]. In this paper, the conditions required on the experiments for estimating full information of $\boldsymbol{s}^{\text{true}}$ are presented at Sec. III A. Concrete combinations used in the numerical experiments reported in Sec. IV are shown in Table III in Appendix H.

Each combination of operations is specified with an index sequence of gates, say $\boldsymbol{i}$, and a set of index sequences is denoted by $\mathbf{Id}$, called experimental schedule. An example of an experimental schedule is shown in Fig. 1. Let $\boldsymbol{p}^{\boldsymbol{i}}(\boldsymbol{s})$ denote the probability distribution of measurement outcomes of the $\boldsymbol{i}$-th experiment with a set of quantum operations $\boldsymbol{s} \in \mathcal{S}$. We define $\boldsymbol{p}(\mathbf{Id}, \boldsymbol{s}) := \{\boldsymbol{p}^{\boldsymbol{i}}(\boldsymbol{s})\}_{\boldsymbol{i}\in\mathbf{Id}}$. We repeat the experiment N times for each $\boldsymbol{i} \in \mathbf{Id}$. Let $\boldsymbol{f}_N^{\boldsymbol{i}}$ denote an empirical distribution calculated from data obtained in the N repetitions of a sequence $\boldsymbol{i}$ and $\boldsymbol{f}_N(\mathbf{Id}) := \{\boldsymbol{f}_N^{\boldsymbol{i}}\}_{\boldsymbol{i}\in\mathbf{Id}}$. The total amount of data is $N|\mathbf{Id}|$.

For any $\boldsymbol{s}^{\text{true}}$, there exist sets of quantum operations $\tilde{\boldsymbol{s}} \in \mathcal{S}$ satisfying $\boldsymbol{p}^{\boldsymbol{i}}(\boldsymbol{s}^{\text{true}}) = \boldsymbol{p}^{\boldsymbol{i}}(\tilde{\boldsymbol{s}})$ for any $\boldsymbol{i}$ in arbitrary $\mathbf{Id}$ [29]. We call such $\tilde{\boldsymbol{s}}$ gauge-equivalent to $\boldsymbol{s}^{\text{true}}$. Let $[\boldsymbol{s}^{\text{true}}]$ denote the gauge-equivalence class of $\boldsymbol{s}^{\text{true}}$, i.e., the set of all $\tilde{\boldsymbol{s}}$ gauge-equivalent to $\boldsymbol{s}^{\text{true}}$. Any difference in the gauge degrees of freedom is superficial and experimentally undetectable. Therefore, in the SCQT approach, we have to choose how to fix the gauge to obtain an estimate of $\boldsymbol{s}^{\text{true}}$ from experimental data. In Sec. III B, we propose a SCQT method that uses a regularization to fix the gauge. Details of the gauge degrees of freedom are explained in Appendix B.

III. THEORETICAL RESULTS

In this section, we show our theoretical results. In Sec. III A, we introduce a concept of informational completeness for the self-consistent approach, which is an expansion of informational completeness in the standard quantum tomography [26]. We prove that the expanded informational completeness is a sufficient condition on experiments for estimating all parameters of a set $\boldsymbol{s}^{\text{true}}$ except for the gauge degrees of freedom (Theorem 1). The details of the proof of Theorem 1 are shown in Appendix C. In Sec. III B, we propose a data-processing method, called an estimator in statistics, with regularization for estimating $\boldsymbol{s}^{\text{true}}$. We prove that, by tuning a regularization parameter appropriately, an estimate sequence of the estimator $\boldsymbol{s}_N^{\text{est}}$ converges into $[\boldsymbol{s}^{\text{true}}]$ at the limit of N going to infinity, assuming that an experiment satisfies the informational completeness. We also prove that $\boldsymbol{p}(\mathbf{Id}, \boldsymbol{s}_N^{\text{est}})$ converges to $\boldsymbol{p}(\mathbf{Id}, \boldsymbol{s}^{\text{true}})$ with convergence rate equivalent to or faster than $\boldsymbol{f}_N(\mathbf{Id})$ does, which would be optimal. The details of the proof of the asymptotic convergence and derivation of the convergence rate are shown in Appendix D. These results guarantee the reliability of the proposed method for sufficiently large data. In Sec. III C, we give formulae for extracting information of dynamics generators such as Hamiltonian and dissipator from the estimates of a gate in $\boldsymbol{s}_N^{\text{est}}$. Details of the derivation of the formula are shown in Appendix E. The formulae would be useful for improving accuracy of a gate in QIP experiments.

A. Informational completeness and gauge-equivalence

We derive a sufficient condition on an experimental schedule $\mathbf{Id}$ to self-consistently characterize quantum operations. Under the condition, we can know full information of $\boldsymbol{s}^{\text{true}}$ except for the gauge degrees of freedom.

We introduce informational completeness in the context of SCQT. Let $\mathbf{Id} = \{(i_{\text{g}1_k}, \ldots, i_{\text{g}L_k}) : \ k = 1, \ldots\}$ denote a set of index vectors, where k is an index for gate sequences and i_{g} is an index for gates. We call an experimental schedule $\mathbf{Id}$ *state-informationally complete* if a set of density matrix

$$\{\rho^{\boldsymbol{i}} := \mathcal{G}_{i_{\text{g}L_k}} \circ \cdots \circ \mathcal{G}_{i_{\text{g}1_k}}(\rho)\}_{\boldsymbol{i}\in\mathbf{Id}} \tag{1}$$

is a (possibly over-complete) basis of $d \times d$ matrix space. We call an experimental schedule $\mathbf{Id}$ *POVM-informationally complete* if a set of POVMs

$$\{\mathbf{\Pi}^{\boldsymbol{i}} := \mathcal{G}^{\dagger}_{i_{\text{g}1_k}} \circ \cdots \circ \mathcal{G}^{\dagger}_{i_{\text{g}L_k}}(\mathbf{\Pi})\}_{\boldsymbol{i}\in\mathbf{Id}} \tag{2}$$

is a (possibly over-complete) basis of the space. Let $\boldsymbol{i} \cup \boldsymbol{i}'$ denote the direct union of two index vectors, i.e., $\boldsymbol{i} \cup \boldsymbol{i}' = (i_1, \ldots, i_L, i'_1, \ldots, i'_{L'})$. We call $\mathbf{Id}$ *self-consistently informationally complete* (SCIC) if it includes

$$\left\{\boldsymbol{i}_{\text{s}} \cup \boldsymbol{i}_{\text{p}} \;\middle|\; \boldsymbol{i}_{\text{s}} \in \mathbf{Id}_{\text{s}}, \boldsymbol{i}_{\text{p}} \in \mathbf{Id}_{\text{p}}\right\} \tag{3}$$

and

$$\left\{ \boldsymbol{i}_{\rm s} \cup i_{\rm g} \cup \boldsymbol{i}_{\rm p} \;\middle|\; \boldsymbol{i}_{\rm s} \in \mathbf{Id}_{\rm s}, i_{\rm g} \in \{1,\dots,n_{\rm g}\}, \boldsymbol{i}_{\rm p} \in \mathbf{Id}_{\rm p} \right\} \quad (4)$$

as subsets where $\mathbf{Id}_{\rm s}$ and $\mathbf{Id}_{\rm p}$ are state-informationally and POVM-informationally complete sets of gate index sequences, respectively. Eq. (4) means that when $\boldsymbol{i}$ is SCIC, it includes quantum process tomography (QPT) experiment for all gates in $\boldsymbol{s}$, and Eq. (3) means that it includes quantum state tomography (QST) experiment for state preparations and POVM tomography (POVMT) experiment for measurements used in the QPT experiment. The SCIC condition implies that **Id** includes QST,POVMT, and QPT experiments allowing duplication of index sequences. Hence, we expect that we can obtain full information of $\boldsymbol{s}$ except for the gauge degrees of freedom from experimental data of a SCIC **Id**.

Theorem 1 *Suppose that assumptions (i)–(iv) in Sec. II hold,* **Id** *is SCIC, and inverse maps* $\mathcal{G}_{i_{\rm g}}^{-1}$ *exists for* $i_{\rm g} = 1,\dots,n_{\rm g}$. *Then, for any* $\boldsymbol{s}$, $\tilde{\boldsymbol{s}} \in \mathcal{S}$, *the following two statements are equivalent:*

1. $\tilde{\boldsymbol{s}} \in [\boldsymbol{s}]$.
2. $\boldsymbol{p}(\mathbf{Id}, \tilde{\boldsymbol{s}}) = \boldsymbol{p}(\mathbf{Id}, \boldsymbol{s})$.

Proof of Theorem 1 is given at Appendix C. Note that the inverse maps mentioned in Theorem 1 are not required to be TPCP. The inverse map always exists if a gate is implemented with dynamics obeying a time-dependent GKLS master equation [33–35], the time period is finite, and the dissipator of the dynamics is bounded [36] (see Appendix E. 1). These conditions are considered as natural in usual settings of QIP experiments, and the condition on the existence of the inverse is a natural assumption in experiments.

Theorem 1 indicates that the experimental indistinguishability implies the gauge-equivalence when the set of gate index sequences is SCIC. By taking contraposition of Theorem 1, we have

$$\tilde{\boldsymbol{s}} \notin [\boldsymbol{s}] \Leftrightarrow \boldsymbol{p}(\mathbf{Id}, \tilde{\boldsymbol{s}}) \neq \boldsymbol{p}(\mathbf{Id}, \boldsymbol{s}). \quad (5)$$

This means that we can distinguish gauge-inequivalent sets of quantum operations from probability distributions of experiments satisfying the SCIC condition. Therefore the SCIC condition is a sufficient condition.

B. Asymptotically Gauge-Equivalent Estimator

We propose an estimator with regularization. The estimator is formulated with three parts, loss function, regularization function, and regularization parameter. Here we specify the classes of loss and regularization functions into squared errors for simplicity. Results in this subsection hold for much wider classes, which is mentioned in Sec. V A.

Suppose that we choose an experimental schedule **Id** satisfying the SCIC condition and obtain experimental data after N repetition. Our purpose is to obtain an estimate $s_N^{\rm est}$ from the data. In order to do that, we introduce loss regularization functions as

$$\mathrm{L}\big(\boldsymbol{p}\,(\mathbf{Id}, \boldsymbol{s}), \boldsymbol{f}_N(\mathbf{Id})\,\big) := \frac{1}{|\mathbf{Id}|} \sum_{\boldsymbol{i} \in \mathbf{Id}} \frac{1}{2} \left\| \boldsymbol{p}^{\boldsymbol{i}}(\boldsymbol{s}) - \boldsymbol{f}_N^{\boldsymbol{i}} \right\|_2^2, \quad (6)$$

and

$$\mathrm{R}(\boldsymbol{s}, \boldsymbol{s}') := \frac{1}{2} \left\| \rho - \rho' \right\|_2^2 + \frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} \frac{1}{2} \left\| \Pi_x - \Pi'_x \right\|_2^2 + \sum_{i_{\rm g}=1}^{n_{\rm g}} \frac{1}{2d^2} \left\| \mathrm{HS}\left(\mathcal{G}_{i_{\rm g}}\right) - \mathrm{HS}\left(\mathcal{G}'_{i_{\rm g}}\right) \right\|_2^2, \quad (7)$$

where $\mathrm{HS}(\mathcal{G})$ denote a Hilbert-Schmidt matrix representation of a TPCP map $\mathcal{G}$.

We propose the following estimator:

$$\boldsymbol{s}_N^{\rm est} := \operatorname*{argmin}_{\boldsymbol{s} \in \mathcal{S}} \left\{ \mathrm{L}\big(\boldsymbol{p}(\mathbf{Id}, \boldsymbol{s}), \boldsymbol{f}_N(\mathbf{Id})\big) + r_N \mathrm{R}(\boldsymbol{s}, \boldsymbol{s}^{\rm target}) \right\}, \quad (8)$$

where r_N is a positive number, called regularization parameter. It is user-tunable and can depend not only on N, but also on data. The regularization term in Eq. (8) takes a role for fixing the gauge as the estimate becomes close to $\boldsymbol{s}'$, which is a new way to use regularization. The set $\boldsymbol{s}'$ is a user-specified set of quantum operations. Its choice is arbitrary and up to the user. We propose to use the target set as the regularization point, i.e., $\boldsymbol{s}' = \boldsymbol{s}^{\rm target}$ in Eq. (8). We discuss the choice of $\boldsymbol{s}'$ in Sec. V A.

We call the estimator defined by Eq. (8) a *regularized self-consistent (RSC)* estimator, and we call a quantum tomographic protocol with the RSC estimator *regularized self-consistent quantum tomography (RSCQT)*. We have to select the value of r_N carefully. For example, if we select r_N so large that the effect of the loss function in the minimization of Eq. (8) becomes negligible, the RSC estimate $\boldsymbol{s}_N^{\rm est}$ approaches $\boldsymbol{s}^{\rm target}$. Then $\boldsymbol{p}(\mathbf{Id}, \boldsymbol{s}_N^{\rm est})$ cannot reproduce $\boldsymbol{f}_N(\mathbf{Id})$ precisely for finite N.

The following theorem gives a guideline to select a valid value of r_N. We use a mathematical notation, $\lesssim$, in such a way that $f(N) \lesssim g(N)$ indicates that, for a positive constant a, $f(N) \leq a g(N)$ holds for any sufficiently large N. An abbreviation, a.s., stands for *almost surely* in probability theory. A rigorous definition of the notation is given in Appendix D 1.

Theorem 2 (Asymptotic gauge-equivalence) *Suppose that assumptions (i)–(iv) in Sec II hold. If we select a regularization parameter satisfying*

$$\lim_{N \to \infty} r_N = 0 \text{ a.s.}, \quad (9)$$

then the sequence of the probability distributions, $\{\boldsymbol{p}(\mathbf{Id}, \boldsymbol{s}_N^{est})\}$, *converges to the true one* $\boldsymbol{p}(\mathbf{Id}, \boldsymbol{s}^{true})$ *almost surely, i.e., the equality,*

$$\lim_{N \to \infty} \sqrt{\mathrm{L}(\boldsymbol{p}(\mathbf{Id}, \boldsymbol{s}_N^{est}), \boldsymbol{p}(\mathbf{Id}, \boldsymbol{s}^{true}))} = 0 \text{ a.s.}, \quad (10)$$

holds. If we select the regularization parameter satisfying

$$r_N \lesssim 1/N \text{ a.s.}, \tag{11}$$

then inequalities,

$$\sqrt{\mathrm{L}(\boldsymbol{p}(\mathbf{Id}, \boldsymbol{s}_N^{est}), \boldsymbol{p}(\mathbf{Id}, \boldsymbol{s}^{true}))}$$
$$\lesssim \sqrt{\mathrm{L}(\boldsymbol{p}(\mathbf{Id}, \boldsymbol{s}^{true}), \boldsymbol{f}_N(\mathbf{Id}))} \tag{12}$$
$$\lesssim \sqrt{\frac{\ln \ln N}{N}} \text{ a.s.}, \tag{13}$$

hold. If Eq. (9) is satisfied and **Id** *is SCIC, then the sequence of RSC estimates* $\{\boldsymbol{s}_N^{est}\}$ *converges to* $[\boldsymbol{s}^{true}]$ *almost surely, i.e., the equality,*

$$\lim_{N\to\infty} \min\left\{ R(\boldsymbol{s}_N^{est}, \tilde{\boldsymbol{s}}) \;\middle|\; \tilde{\boldsymbol{s}} \in [\boldsymbol{s}^{true}] \right\} = 0 \text{ a.s.}, \tag{14}$$

holds.

The details of the proof are given in Appendix D. Here we sketch them.

- Proof of Eqs. (10), (12), and (13): We combine a property of $\boldsymbol{s}_N^{\mathrm{est}}$ as a minimizer with the strong law of large numbers, the central limit theorem, and the strong law of iterated logarithm [37] in order to prove them.

- Proof of Eq. (14): First, we derive an inequality that any points in $\mathcal{S}$ outside ϵ-neighborhood of $[\boldsymbol{s}^{\mathrm{true}}]$ satisfy. A main mathematical tool at the derivation is the strong law of large numbers. Second, we prove that, for any small $\epsilon > 0$, by taking a sufficiently large N, $\boldsymbol{s}_N^{\mathrm{est}}$ does not satisfy the inequality. This indicates that $\boldsymbol{s}_N^{\mathrm{est}}$ is in the ϵ-neighborhood and converges to $[\boldsymbol{s}^{\mathrm{true}}]$.

At the construction of the proof of Eq. (14), we used known results from mathematical statistics as reference. If we neglect the existence of the gauge degrees of freedom in the setting of SCQT, the RSC estimator defined by Eq. (8) can be categorized into an abstract and general class of statistical estimators, called minimum contrast estimator. In statistical parameter estimation, some sufficient conditions for a minimum contrast estimator to asymptotically converge to the true parameter are known [38]. These results are not directly applicable to the RSC estimator in the setting of SCQT because there exist the gauge degrees of freedom. Nevertheless, our setting has many properties that are easy to mathematically handle, such as finite dimensional parameter space, multinomial probability distributions, and smooth parametrization of the probability distributions. We modified the known results to make them applicable to the setting of SCQT. Simultaneously, with the good properties of the setting of SCQT and the specific form of the RSC estimates $\boldsymbol{s}_N^{\mathrm{est}}$, we simplified the modified sufficient conditions. Detailed explanations on differences between our results and known results of regularization are given in Appendix F.

Suppose that **Id** is SCIC and r_N is chosen to satisfy Eq. (9). Theorem 2 indicates the asymptotic gauge-equivalence of $\boldsymbol{s}_N^{\mathrm{est}}$, i.e., the convergence of $\boldsymbol{s}_N^{\mathrm{est}}$ to $[\boldsymbol{s}^{\mathrm{true}}]$, and this guarantees the high reliability of the RSC estimates $\boldsymbol{s}_N^{\mathrm{est}}$ for sufficiently large N. The estimates are physical, because the minimization range is restricted into the physical region $\mathcal{S}$. Hence, $\boldsymbol{s}_N^{\mathrm{est}}$ is self-consistent, stringently physical, and asymptotically gauge-equivalent, in theory. Additionally, Eq. (12) guarantees that, if we choose $r_N = c/N$ where c is a positive constant independent of N, the asymptotic convergence rate of $\boldsymbol{p}(\mathbf{Id}, \boldsymbol{s}_N^{\mathrm{est}})$ to $\boldsymbol{p}(\mathbf{Id}, \boldsymbol{s}^{\mathrm{true}})$ becomes equivalent to or better than that of $\boldsymbol{f}_N(\mathbf{Id})$. This means that $\boldsymbol{s}_N^{\mathrm{est}}$ can reproduce $\boldsymbol{p}(\mathbf{Id}, \boldsymbol{s}^{\mathrm{true}})$ at least as precise as $\boldsymbol{f}_N(\mathbf{Id})$ can. We conjecture that the asymptotic convergence rates of $\boldsymbol{p}(\mathbf{Id}, \boldsymbol{s}_N^{\mathrm{est}})$ and $\boldsymbol{f}_N(\mathbf{Id})$ are equivalent because that of $\boldsymbol{f}_N(\mathbf{Id})$ would be optimal. There is still arbitrariness of selection of c for tuning the value of r_N. In practice, even if c is independent of N, too large c can lead a large bias on the RSC estimates for finite N. We can avoid to select such unreasonably large c by combining the estimator with cross validation [39, 40], which is a standard method for selecting a regularization parameter in statistics and its brief explanation is given at Appendix G. We show the performance of the combination for the case of 1-qubit system in Sec. IV.

C. Dynamics Generator Analysis

Here we propose a method for extracting information of dynamics generators from an estimate of a gate. Suppose that the dynamics of a quantum state during the gate operation obeys the time-dependent GKLS equation [33–35].

$$\frac{d\rho}{dt} = \mathcal{L}_t(\rho) := -i[H(t), \rho] + \{J(t), \rho\} + \sum_{\alpha,\beta=1}^{d^2-1} K_{\alpha\beta}(t) B_\alpha \rho B_\beta^\dagger, \tag{15}$$

where $\boldsymbol{B} := \{B_\alpha\}_{\alpha=0}^{d^2-1}$ is an orthonormal Hermitian matrix basis satisfying $B_0 = I/\sqrt{d}$, $H(t) = \sum_{\alpha=1}^{d^2-1} H_\alpha(t) B_\alpha$, $J(t) = \sum_{\alpha=0}^{d^2-1} J_\alpha(t) B_\alpha$, $H_\alpha(t) \in \mathbb{R}$, $J_\alpha(t) \in \mathbb{R}$, and $K_{\alpha\beta}(t) \in \mathbb{C}$. When a gate $\mathcal{G}$ is implemented under Eq. (15) from $t = 0$ to $t = T$, the HS representation of the map is formally expressed as

$$\mathrm{HS}(\mathcal{G}) = \mathcal{T} \exp\left[\int_0^T dt\ \mathrm{HS}(\mathcal{L}_t) \right], \tag{16}$$

where $\mathcal{T}$ stands for the chronological operator. If $H(t)$, $J(t)$, and $K(t)$ are bounded for any $t \in [0, T]$ and T is finite, $\mathrm{HS}(\mathcal{G})^{-1}$ exists (see Appendix E. 1 for the proof), and there exists a matrix $\mathrm{L}^{\mathrm{acc}}$ satisfying

$$\mathrm{HS}(\mathcal{G}) = \exp(\mathrm{L}^{\mathrm{acc}}). \tag{17}$$

We call $\mathrm{L}^{\mathrm{acc}}$ the accumulated dynamics generator of the gate $\mathcal{G}$. It satisfies

$$\mathrm{L}^{\mathrm{acc}} = \ln \mathrm{HS}(\mathcal{G}). \tag{18}$$

Let us define $\mathcal{L}^{\mathrm{acc}}$ as a linear map satisfying

$$\mathrm{HS}(\mathcal{L}^{\mathrm{acc}}) = \mathrm{L}^{\mathrm{acc}}. \tag{19}$$

From the completeness of the matrix basis $\boldsymbol{B}$, the action of the map $\mathcal{L}^{\mathrm{acc}}$ can be represented in the following form.

$$\mathcal{L}^{\mathrm{acc}}(\rho) = -i[H^{\mathrm{acc}}, \rho] + \{J^{\mathrm{acc}}, \rho\} + \sum_{\alpha,\beta=1}^{d^2-1} K^{\mathrm{acc}}_{\alpha\beta} B_\alpha \rho B_\beta^\dagger, \tag{20}$$

where

$$H^{\mathrm{acc}} = \sum_{\alpha=1}^{d^2-1} H^{\mathrm{acc}}_\alpha B_\alpha, \tag{21}$$

$$J^{\mathrm{acc}} = \sum_{\alpha=0}^{d^2-1} J^{\mathrm{acc}}_\alpha B_\alpha. \tag{22}$$

The matrix H^{acc} represents the accumulated action of the original Hamiltonian $H(t)$ from $t = 0$ to $t = T$, and K^{acc} represents the accumulated action of the original dissipator $K(t)$ from $t = 0$ to $t = T$. When the original generators are time-independent, i.e., $H(t) = H$, $J(t) = J$, and $K(t) = K$, the accumulated generators $\{H^{\mathrm{acc}}, J^{\mathrm{acc}}, K^{\mathrm{acc}}\}$ are simply $H^{\mathrm{acc}} = T \cdot H$, $J^{\mathrm{acc}} = T \cdot J$, and $K^{\mathrm{acc}} = T \cdot K$.

Let $\mathrm{L}^{\mathrm{acc,cb}}$ denote the HS representation of $\mathcal{L}^{\mathrm{acc}}$ with respect to the computational basis. The coefficients of the accumulated generators can be calculated from $\mathcal{L}^{\mathrm{acc}}$ as follows:

$$H^{\mathrm{acc}}_\alpha = \frac{i}{2d} \mathrm{Tr}\left[\mathrm{L}^{\mathrm{acc,cb}}(B_\alpha \otimes I - I \otimes \overline{B_\alpha})\right], \quad \alpha = 1, \ldots, d^2 - 1, \tag{23}$$

$$J^{\mathrm{acc}}_\alpha = \frac{1}{2d(1+\delta_{0\alpha})} \mathrm{Tr}\left[\mathrm{L}^{\mathrm{acc,cb}}(B_\alpha \otimes I + I \otimes \overline{B_\alpha})\right], \quad \alpha = 0, \ldots, d^2 - 1, \tag{24}$$

$$K^{\mathrm{acc}}_{\alpha\beta} = \mathrm{Tr}\left[\mathrm{L}^{\mathrm{acc,cb}}\left(B_\alpha \otimes \overline{B_\beta}\right)\right], \quad \alpha, \beta = 1, \ldots, d^2 - 1, \tag{25}$$

where $\overline{B_\alpha}$ is the complex conjugate of the matrix basis element with respect to the computational basis representation. The derivations of Eqs. (23), (24), and (25) are shown in Appendix E. 2. After performing a self-consistent tomographic experiments and data-prosessing, we have an estimate $\mathrm{G}^{\mathrm{est}} := \mathrm{HS}(\mathcal{G}^{\mathrm{est}})$ of a gate $\mathrm{G} = \mathrm{HS}(\mathcal{G})$. Then we can obtain an estimate of the accumulated generators in the following procedure:

Step 1. Choose the computational basis as the representation basis of HS. Then we have $\mathrm{G}^{\mathrm{est}} = \mathrm{HS}^{\mathrm{cb}}(\mathcal{G}^{\mathrm{est}})$.

Step 2. Calculate the matrix logarithm, $\ln \mathrm{G}^{\mathrm{est}} =: (\mathrm{L}^{\mathrm{acc,cb}})^{\mathrm{est}}$.

Step 3. Substitute $(\mathrm{L}^{\mathrm{acc,cb}})^{\mathrm{est}}$ into $\mathrm{L}^{\mathrm{acc,cb}}$ in the R.H.S. of Eqs. (23), (24), and (25).

By following steps above, we can extract information of the accumulated generators H^{acc}, J^{acc}, and K^{acc}, but we cannot know the information about the original generators $H(t)$, $J(t)$, and $K(t)$ at each $t \in [0, T]$. This is because in general quantum tomography treat a gate as a black box, and a tomographic result gives the information of an input-output relation during the time period.

IV. NUMERICAL RESULTS

Theorem 2 in Sec. III guarantees the high reliability of the RSC estimator for asymptotically large N on any finite-dimensional quantum system. In practice, it is important to investigate its performances for finite N. The investigation must be done by numerical experiments because calculations of an estimation error require information of the true set $\boldsymbol{s}^{\mathrm{true}}$, and it's not available in real experiments. Numerical implementation of the RSC estimator includes a constraint nonlinear optimization problem, which is the main challenge and is different from standard QT. As the first step, we numerically implemented the RSC estimator for 1-qubit systems and performed numerical experiments

We investigated the performance of the RSC estimator for several settings and parameter regions of error models. The observed results are both of positive and negative. The positive part is that they are consistent with Eqs. (10), (12), and (13), and these indicate its high reliability on its prediction performance of probability distributions with finite data as well, even though there exist effects of bias originated from regularization. The negative part is that Hamiltonian estimated with the dynamics generator analysis includes an effect of a gauge transformation and can differ from the true value, which is not a specific feature of the RSC estimator, but would be a common feature of the SCQT approach. In this section, we briefly explain the setting and results of the numerical simulations. Details of the simulation are described in Appendix H.

A. Setting

The system simulated is a two-level system ($d = 2$). The target set is chosen as $\boldsymbol{s}^{\mathrm{target}} =$

$\{\rho^{\rm target}, \mathbf{\Pi}^{\rm target}, \mathcal{G}_0^{\rm target}, \mathcal{G}_1^{\rm target}, \mathcal{G}_2^{\rm target}\}$ such that

$$\rho^{\rm target} = |0\rangle\langle 0|, \tag{26}$$
$$\mathbf{\Pi}^{\rm target} = \{|0\rangle\langle 0|, |1\rangle\langle 1|\}, \tag{27}$$
$$\mathcal{G}_0^{\rm target}(\rho) = \rho, \tag{28}$$
$$\mathcal{G}_1^{\rm target}(\rho) = e^{-i\frac{\pi}{4}\sigma_1}\rho e^{i\frac{\pi}{4}\sigma_1}, \tag{29}$$
$$\mathcal{G}_2^{\rm target}(\rho) = e^{-i\frac{\pi}{4}\sigma_2}\rho e^{i\frac{\pi}{4}\sigma_2}, \tag{30}$$

i.e., the target state is the ground state, the target POVM is the projective measurement along with the Z-axis, and the target gates are the identity, $\frac{\pi}{2}$-rotation along with the X-axis, and $\frac{\pi}{2}$-rotation along with the Y-axis. The true set $\boldsymbol{s}^{\rm true}$ is chosen as a set that is close to the target set but it includes coherent errors and decoherence. In the realistic model, the state and POVM are affected by a depolarizing error, and the gates are generated by a rectangular pulse with decoherence obeying a GKLS master equation [41]. We chose an experimental schedule **Id** consisting of 45 operation sequences, which is SCIC. Details of the model are described in Appendix H 1.

Along with Theorem 2, we select the regularization parameter as $r_N = c/N$, where c is a constant positive value. The selection of c is up to the user. In order to check the effect of the selection on the performance, we set c in a wide range, $10^{-1}, 1, 10, 10^2, 10^3$, and we combined the RSC estimator with a k-fold cross validation, which selects a reasonable value of c from the set of candidate values. The procedure of the data-processing at the k-fold cross validation is explained in Appendix G. The computational cost for k-fold cross validation procedures becomes larger as k becomes larger. In order to keep the computational cost as small as possible, we set $k = 3$. We performed a Monte Carlo simulation with $N = 10^2$ to 10^6. Statistics like expectations, variances, and standard deviations are calculated with 500 iterations.

B. Numerical Result 1: Loss and Regularization

At the first analysis, we investigate behaviors of quantities related to the loss function L in Eq. (6) and regularization function R in Eq. (7) in order to test how the performance of the RSC estimator for finite N differs from the asymptotic behaviours in Theorem 2. The results are visualized in Fig. 2, which includes four panels. Horizontal axes of the panels are the amount of data N.

Panel (a) of Fig. 2 is for the root-mean square of estimation error from $\boldsymbol{s}^{\rm true}$ w.r.t. the probability distributions, i.e., $\sqrt{\mathbb{E}[\mathrm{L}(\boldsymbol{p}(\mathbf{Id}, \mathbf{s}), \mathbf{p}(\mathbf{Id}, \mathbf{s}^{\rm true}))]}$. The black solid line is for $\boldsymbol{s} = \boldsymbol{s}^{\rm target}$, and it quantifies the discrepancy between $\boldsymbol{s}^{\rm target}$ and $\boldsymbol{s}^{\rm true}$ in the space of probability distributions, which is independent of N. The blue solid line is for empirical distributions, $\sqrt{\mathbb{E}[\mathrm{L}(\boldsymbol{f}_N(\mathbf{Id}), \boldsymbol{p}(\mathbf{Id}, \boldsymbol{s}^{\rm true}))]}$, which scales as $1/\sqrt{N}$. The red solid line is for $\mathbf{s} = \boldsymbol{s}_N^{\rm est}$ with the cross validation, and the other dotted lines are for $\boldsymbol{s} = \boldsymbol{s}_N^{\rm est}$ with fixed c. Line style and color of $\boldsymbol{s}_N^{\rm est}$ are common in all panels of Figs. 2, 3, and 4. Lines of $\boldsymbol{s}_N^{\rm est}$ converge to zero as N increases, and they are, except for the line of the largest c, almost parallel with and below the line of $\boldsymbol{f}_N$. The convergence in the space of probability distributions means the convergence to the gauge-equivalence class $[\boldsymbol{s}^{\rm true}]$ because of the SCIC of the experimental schedule and Theorem 1. This is consistent with Eqs. (10) and (13). The panel also shows that the cross validation selects the almost best value of c in the candidates on average.

It is interesting that the estimation errors of $\boldsymbol{s}_N^{\rm est}$ are smaller than that of $\boldsymbol{f}_N$ (except for $c = 10^3$), and the gap remains up to asymptotically large amount of data, at least $N = 10^6$. We observed the same tendency at the other error models as well. This means that, if we choose a reasonable value of c, the RSC estimator has the predictability of the true probability distributions higher than experimental data itself. There are four possible origins of the gap: (i) inequality constraints of physicality, (ii) equality constraints of physicality, (iii) regularization, and (iv) gauge degrees of freedom. It is known that, in the standard QT, the inequality constraints contribute to reduce an estimation error [42]. Such effect exists as well in the SCQT, but it would not be the main origin of the gap because the effect is expected to decrease as N increases. The equality constraints must be one of the main origin, because it reduces the degrees of freedom of $\boldsymbol{s}$ and $\boldsymbol{p}(\mathbf{Id}, \boldsymbol{s})$ while there are no restrictions on $\boldsymbol{f}_N$. Regularization can also be the main origin, because it tends to reduce the variance of the estimator, although it introduce a bias. Whatever the origin, the figure indicates the RSC estimator's high predictability of probability distributions.

Panel (b) of Fig. 2 is for the goodness of fit to data, i.e., $\sqrt{\mathbb{E}[\mathrm{L}(\boldsymbol{p}(\mathbf{Id}, \boldsymbol{s}), \boldsymbol{f}_N)]}$. The blue line is for $\boldsymbol{s} = \boldsymbol{s}^{\rm true}$, and note that it's equivalent to the blue line at Panel (a) due to the symmetry of $\sqrt{\mathrm{L}}$ w.r.t. the first and second variables. For $\boldsymbol{s} = \boldsymbol{s}_N^{\rm est}$, the lines correspond to the first term of R.H.S. of Eq. (8) after its minimization process. The lines are almost similar except for the largest c. This means that the selection of the regularization parameter c does not affect on the goodness of fit of the RSC estimator to data if c is not too large.

Panel (c) of Fig. 2 is for the estimation error in the space of quantum operations. The vertical axis is the root-mean-squared error from $\boldsymbol{s}^{\rm true}$, $\sqrt{\mathbb{E}[\mathrm{R}(\boldsymbol{s}, \boldsymbol{s}^{\rm true})]}$, for $\boldsymbol{s} = \boldsymbol{s}^{\rm target}$ and $\boldsymbol{s} = \boldsymbol{s}_N^{\rm est}$. In Panels (c) and (d), we have performed a gauge transformation on $\boldsymbol{s}_N^{\rm est}$ that diagonalizes the POVM in order to adjust a reference frame for comparing to $\boldsymbol{s}^{\rm true}$. The lines of $\boldsymbol{s}_N^{\rm est}$ are below that of $\boldsymbol{s}^{\rm target}$, but they do not converge to zero. Panels (a) and (c) indicate that the RSC estimates converge to a point in the gauge-equivalence class $[\boldsymbol{s}^{\rm true}]$ and the point is different from $\boldsymbol{s}^{\rm true}$. This is as expected because there is an arbitrary choice of gauge-fixing method and the gauge-fixing with the squared 2-norm distance to the target set, $\mathrm{R}(\boldsymbol{s}, \boldsymbol{s}^{\rm target})$, does not lead the estimates to the true set in general. We investigate the discrepancy of $\boldsymbol{s}_N^{\rm est}$ and $\boldsymbol{s}^{\rm true}$ later at the explanation of Figs. 3 and 4.

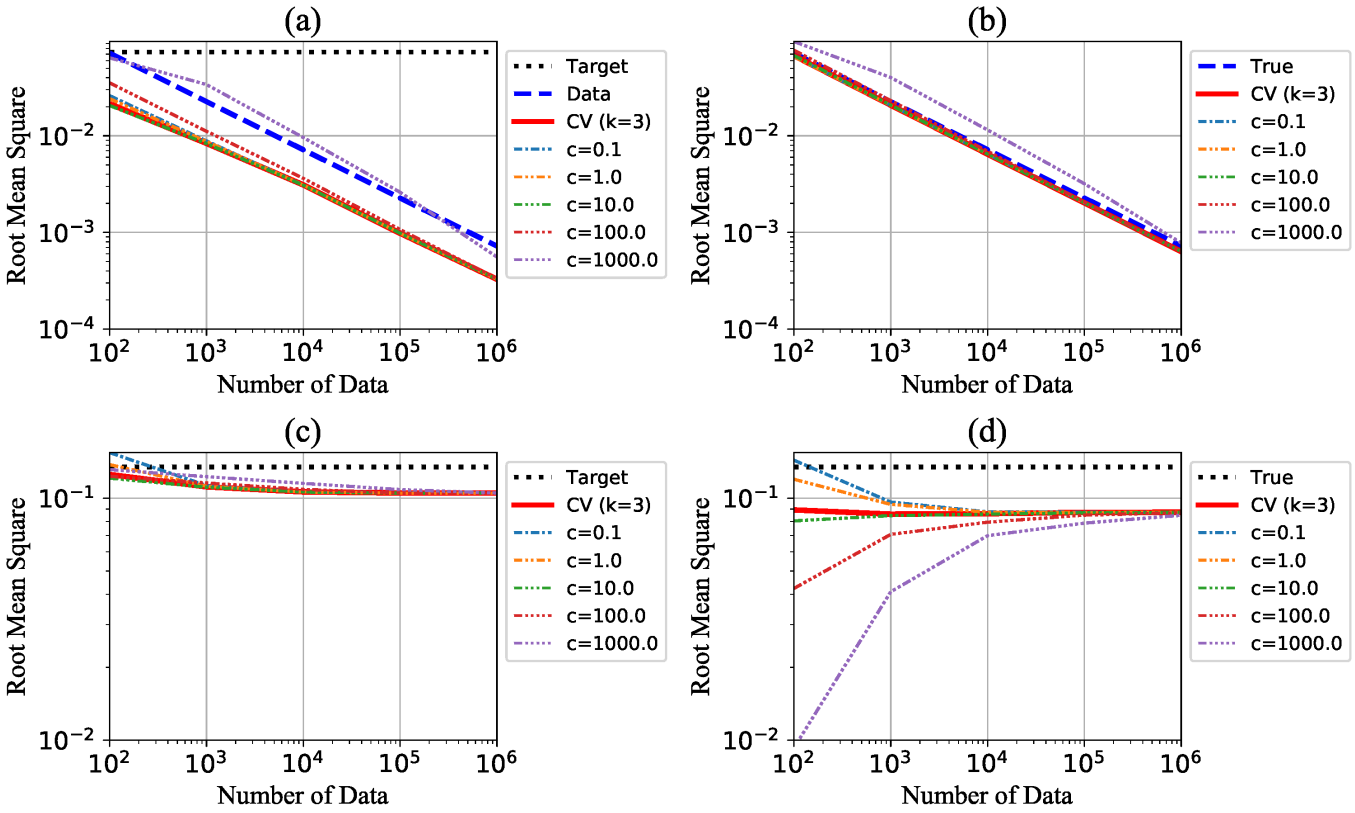

FIG. 2. Plots of root mean squares (RMS) of loss and regularization functions for the RSC estimator against number of data N when the error is generated by the GKLS master equation. Panel (a) is for the estimation error in the space of the probability distributions. Panel (b) is for the goodness of fit to data. Panel (c) is for the estimation error in the space of quantum operations. Panel (d) is for the regularization term to the target set. All of horizontal and vertical axes are log-scale. See main texts for the details.

Panel (d) of Fig. 2 is for the root mean of the regularization term without the regularization parameter, $\sqrt{\mathbb{E}[\mathrm{R}(\boldsymbol{s}, \boldsymbol{s}^{\mathrm{target}})]}$, for $\boldsymbol{s} = \boldsymbol{s}^{\mathrm{true}}$ and $\boldsymbol{s} = \boldsymbol{s}_N^{\mathrm{est}}$. Note that the black lines at Panels (c) and (d) are equivalent due to the symmetry of R. For regions of small N, $\boldsymbol{s}_N^{\mathrm{est}}$ tends to be closer to $\boldsymbol{s}^{\mathrm{target}}$ as c becomes larger. This is as expected because larger c makes its estimate closer to $\boldsymbol{s}^{\mathrm{target}}$.

The observed behaviours of $\boldsymbol{s}_N^{\mathrm{est}}$ in Fig. 2 indicate that the use of regularization for fixing the gauge degrees of freedom does not cause effective biases on estimation of probability distributions and the discrepancy originated from different choice of the regularization parameter c becomes negligible as the amount of data increases.

C. Numerical Result 2: Dynamics Generator Analysis

As shown in the previous subsection, the RSC estimator gives estimates converge to the gauge-equivalence set $[\boldsymbol{s}^{\mathrm{true}}]$, but the convergence point is different from $\boldsymbol{s}^{\mathrm{true}}$. This is due to the existence of the gauge degrees of freedom. We investigate more details of the discrepancy with the dynamics generator analysis proposed in Sec. III C.

Figure 3 shows the root mean squared errors (RMSE) of the RSC estimator to $\boldsymbol{s}^{\mathrm{true}}$. We have performed a gauge transformation on $\boldsymbol{s}_N^{\mathrm{est}}$ that diagonalizes their POVM in order to adjust a reference frame to $\boldsymbol{s}^{\mathrm{true}}$. There are six panels. Three panels at the upper row are for the RMSE of the hamiltonian part of the Lindbladian of gate-0, -1, and -2, respectively. Other panels at the lower row are for the RMSE of the dissipator part of Lindbladian of each gate. At the all panels, the black solid lines are for $\boldsymbol{s}^{\mathrm{target}}$, red solid lines are for $\boldsymbol{s}_N^{\mathrm{est}}$ with the cross validation, and the other dashed lines are for $\boldsymbol{s}_N^{\mathrm{est}}$ with a fixed c. These six panels indicate that the main sources of the non-convergence to $\boldsymbol{s}^{\mathrm{true}}$ are the hamiltonian parts of gate-1 and gate-2 (Panels (1-H) and (2-H)).

More details of the estimation errors of Hamiltonian are shown in Fig. 4, which is for the RMSE of the Pauli X-, Y-, and Z-components of the estimated Hamiltonian. There are six panels. Three panels at the upper row are for the RMSE of the components of gate-1, respectively. Other panels at the lower row are for the RMSE of them

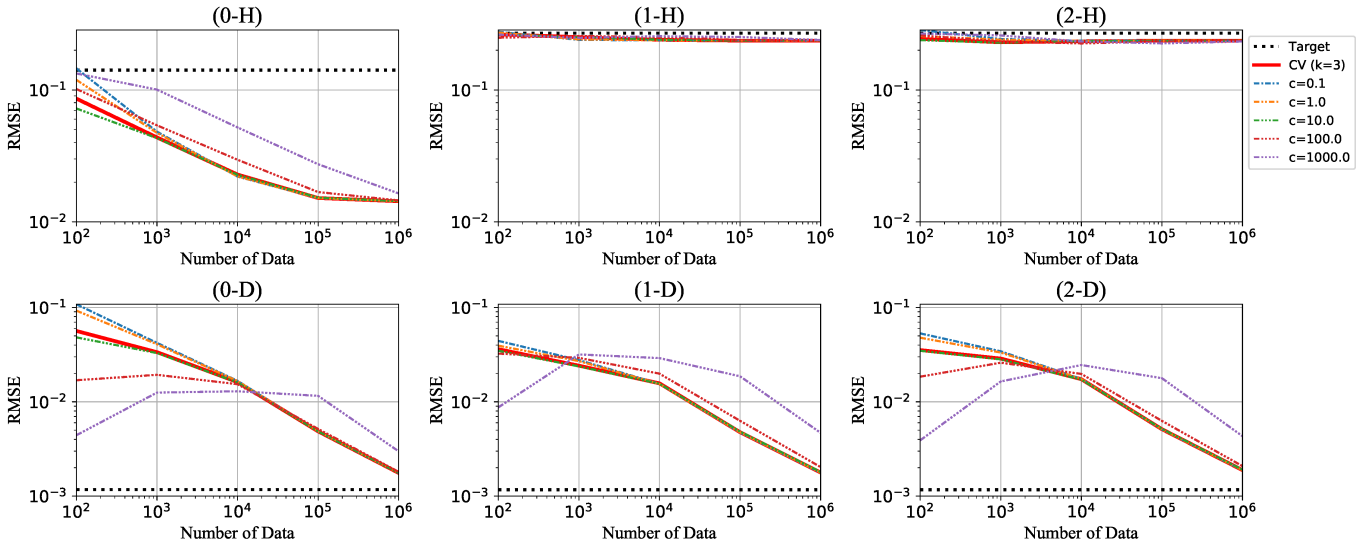

FIG. 3. Root mean squared error (RMSE) of the estimated Lindbladian against the number of data N when the error is generated by the GKLS master equation. The number and letter at each panel label correspond to the gate index (0, 1, 2) and part of Lindbladian (H for Hamiltonian, and D for Dissipator). All of horizontal and vertical axes are log-scale. See the main texts for the details.

of gate-2. Line style and color are same as in Fig. 3. Behaviors of the lines for $\boldsymbol{s}_N^{\rm est}$ in the six panels can be classified into three types: (i) they decrease almost monotonically (Panels (1-Z) and (2-Z)), (ii) they are below the black line but converge to a finite value (Panels (1-X) and (2-Y)), and (iii) they are the same order of the black line (Panels (1-Y) and (2-X)). This classification has a symmetry on X and Y for gate-1 ($\pi/2$-rotation along with X-axis) and gate-2 ($\pi/2$-rotation along with Y-axis).

These behavior can be explained by the commutation relation between the ideal Hamiltonian of each gate and the generator of the gauge transformation as follows. Let $\mathrm{G}_i = \exp(\mathrm{L}_i)$ denote the matrix representation of the i-th gate of the convergence point of the RSC estimates, where L_i denote their Lindbladian. The convergence point is gauge-equivalent to $\boldsymbol{s}^{\rm true}$, there exists a gauge transformation between them. Let A denote the matrix representation of the gauge transformation. Because of the invertibility of A, there exists a matrix a satisfying $A = e^a$. Let $\mathrm{L}_i^{\rm target}$ denote the ideal Lindbladian of the i-th gate, which consists of the Hamiltonian part only and Δ_i denote the discrepancy between L_i and $\mathrm{L}_i^{\rm target}$, i.e., $\Delta_i = \mathrm{L}_i - \mathrm{L}_i^{\rm target}$. Then

$$\begin{aligned} \mathrm{G}_i &= A\mathrm{G}_i^{\rm true}A^{-1} = \exp\left(e^a \mathrm{L}_i e^{-a}\right) \\ &\approx \exp\left(\mathrm{L}_i + [a, \mathrm{L}_i^{\rm target}]\right), \end{aligned} \tag{31}$$

where we assumed $\|a\| \ll 1$, used Tailer expansion, and neglected higher order terms. Eq. (31) indicates that the gauge transformation changes the Lindbladian from L_i to $\mathrm{L}_i + [a, \mathrm{L}_i^{\rm target}]$ (approximately). The Lindbladians contain parts of Hamiltonian and Dissipators. For simplicity, let us focus on the Hamiltonian part. In the dynamics generator analysis, we have performed the gauge transformation that makes POVM diagonal. At the reference frame of diagonal POVM, the remaining gauge degrees of freedom in the Hamiltonian part is the rotation along with Z-axis, because rotations along with other axes makes POVM non-diagonal. Then the hamiltonian part of a contains Z-component only. On the other hand, the target Lindbladian contains only X-component for gate-1 and only Y-component for gate-2. Hence, the commutator $[a, \mathrm{L}_i^{\rm target}]$ leads to Y-component for gate-1 and to X-component for gate-2.

At the discussion above, we ignored the higher order terms of the Taylor expansion. When $\|a\|$ is not so small, such higher order terms become non-negligible. Actually we observed such cases in our numerical simulations with another setting. The results of the dynamics generator analysis shown here indicate that the RSC estimates is affected by uncontrollable gauge transformation and the estimated Hamiltonian can be different from the true value in some non-negligible amount. The discrepancy makes the RSC estimates not useful for experimentalists to perform further improvement of their gate operations. Although we investigated the performance of the RSC estimator only, we believe that this defect is common for all characterization methods in the SCQT approach, because the gauge degrees of freedom remain in any way.

V. DISCUSSION

In Sec. V A, we explain the suitability of the use of a regularization for the gauge-fixing in the SCQT approach. We discuss implementation costs in Sec. V B. Discussions and Appendixes in [43] would be useful for

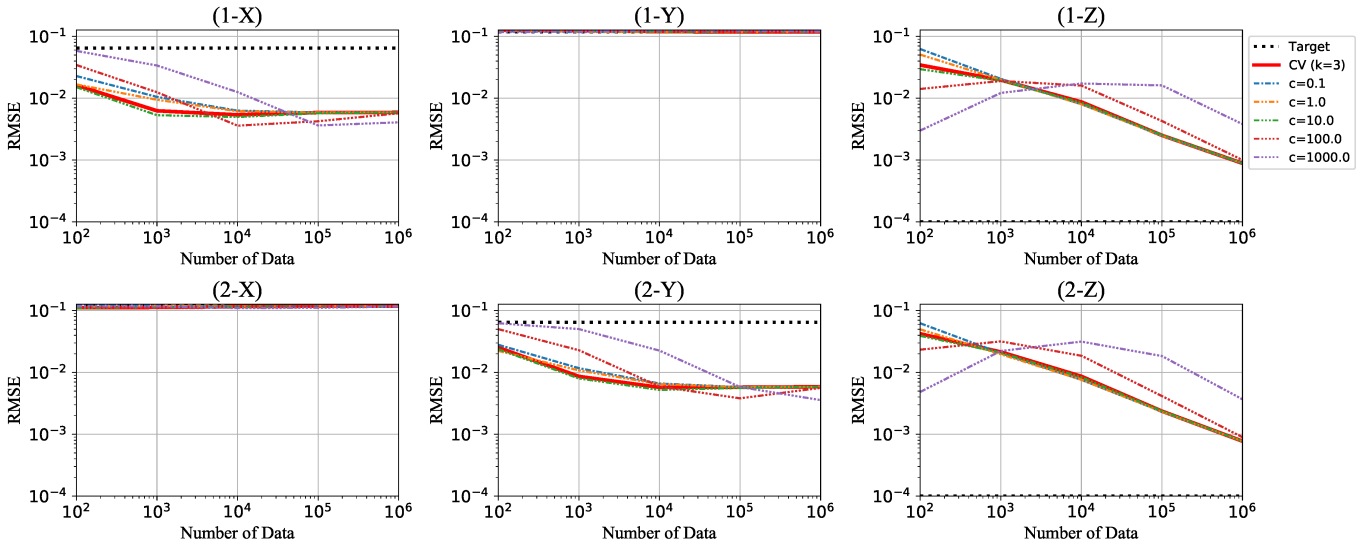

FIG. 4. Root mean squared error (RMSE) of the estimated Hamiltonian components against the number of data N when the error is generated by the GKLS master equation. The number and letter at each panel label correspond to the gate index (0, 1, 2) and Hamiltonian component (X, Y, Z). All of horizontal and vertical axes are log-scale. See the main texts for the details.

considering relations between the SCQT approach with regularization and steps of accuracy validation and improvement for quantum computing.

A. Choice of Regularization

The main purpose of this paper is to propose a reliable tomographic estimator. We require the estimator to return a physical estimate that can reproduce experimental data and predict results of a QIP experiment in the future precisely. A physical argument that minimizes the loss function, i.e., $\mathrm{argmin}_{\boldsymbol{s}\in\mathcal{S}} \mathrm{L}(\boldsymbol{p}(\mathbf{Id},\boldsymbol{s}), \boldsymbol{f}_N(\mathbf{Id}))$ might look suitable for the request. However, since there exist gauge degrees of freedom, the argument is not unique. In order to obtain an estimate from multiple candidates, we have to fix the gauge. It is desirable to choose a gauge-fixing method suitable for validation and improvement after characterization. A typical task at the validation and improvement steps is to estimate a difference between $\boldsymbol{s}^{\mathrm{true}}$ and $\boldsymbol{s}^{\mathrm{target}}$, say $D(\boldsymbol{s}^{\mathrm{true}}, \boldsymbol{s}^{\mathrm{target}})$ by evaluating the difference between $\boldsymbol{s}_N^{\mathrm{est}}$ and $\boldsymbol{s}^{\mathrm{target}}$, $D(\boldsymbol{s}_N^{\mathrm{est}}, \boldsymbol{s}^{\mathrm{target}})$. Suppose that there are two gauge-fixing methods A and B. Their respective estimates, obtained from experimental data, are denoted as $\boldsymbol{s}_{\mathrm{A},N}^{\mathrm{est}}$ and $\boldsymbol{s}_{\mathrm{B},N}^{\mathrm{est}}$. If $\boldsymbol{p}(\mathbf{Id}, \boldsymbol{s}_{\mathrm{A},N}^{\mathrm{est}})$ is as close to $\boldsymbol{f}_N(\mathbf{Id})$ as $\boldsymbol{p}(\mathbf{Id}, \boldsymbol{s}_{\mathrm{B},N}^{\mathrm{est}})$ is and $D(\boldsymbol{s}_{\mathrm{A},N}^{\mathrm{est}}, \boldsymbol{s}^{\mathrm{target}}) < D(\boldsymbol{s}_{\mathrm{B},N}^{\mathrm{est}}, \boldsymbol{s}^{\mathrm{target}})$, we consider method A better because the difference $D(\boldsymbol{s}_{\mathrm{B},N}^{\mathrm{est}}, \boldsymbol{s}^{\mathrm{target}}) - D(\boldsymbol{s}_{\mathrm{A},N}^{\mathrm{est}}, \boldsymbol{s}^{\mathrm{target}})$ is mainly caused by the difference of gauge degrees of freedom that are experimentally indistinguishable. In order to reduce such fake effect on estimates, we fix the gauge such that estimates are as close to the target $\boldsymbol{s}^{\mathrm{target}}$ as they can describe experimental data precisely.

In Eq. (8), we choose the squared 2-norm as the regularization. This is for the simplicity of mathematical and numerical treatments. We can replace the 2-norms in the loss function and in the regularization with any other norms. The estimator with other norms is also asymptotically gauge-equivalent because any norms can be upper-bounded by the 2-norm in finite dimensional complex spaces [44]. In quantum information theory, some norms like the trace-norm and diamond-norm have operational meanings [45, 46]. A regularization using such norms might be more suitable from the perspective of validation after characterization, but numerical treatments of such norms become harder and their computational costs at the minimization increase.

As numerically shown in Sec. IV, the existence of the gauge-degrees of freedom can cause discrepancy between true and estimated Hamiltonians of the RSC estimator, which is a common feature of the SCQT approach. The possible discrepancy makes the performance of the current form of the RSC estimator at the use for further improvement of accuracy low. One possible direction toward improving the performance is to exploit prior information on the experiment with the regularization term.

B. Implementation Costs

The RSC estimator proposed here has superior properties such as asymptotic gauge-equivalence and probably optimal convergence rate. On the other hand, it suffers from one disadvantage that the cost of experiments and data processing scales exponentially with respect to the number of subsystems. The exponential scaling is common in tomographic methods, where the experimental

cost of SCQT is about the same as that of GST and is higher than that of standard QT. The numerical cost of RSC estimator depends on the choice of the loss and regularization functions and optimization algorithm, but in general it is higher than that of standard QT because the number of parameters to be estimated is much higher. Comparison to GST or pyGSTi is a bit obscure, because they use an approximated likelihood function as a loss function, which is different from our choice, and the physicality constraints are not fully taken into account at the optimization [30–32], whereas the constraints are taken into account in the RSC estimator.

Suppose that, for GST and RSCQT, we have chosen the same experimental setting, the same loss function, and the same optimization algorithm with fully taking into the physicality constraints. Then only difference between them is how to fix the gauge. GST performs the gauge fixing data-processing separately after the optimization of the loss function, and in the approach we need to perform an additional optimization over the gauge degrees of freedom. On the other hand, in the RSCQT approach, the gauge-fixing is simultaneously performed during the optimization of the objective function. The objective function consists of loss and regularization functions. If we choose the squared 2-norm as the regularization, the numerical cost of the objective function is dominated by that of the loss function because the nonlinearity of the loss function is much higher than a quadratic function in general. The regularization function is a function of the set of quantum operations $\boldsymbol{s}$, and the gauge degrees of freedom does not appear explicitly there. In the RSCQT approach, we can avoid the optimization over the gauge degrees of freedom. The gauge optimization problem contains the matrix inverse (see Eq. (B5)) with the physicality constraints. The highly nonlinear constraint optimization can become numerically unstable and hard to be solve. This might be a reason that the current version of pyGSTi cannot take into account the full physicality constraints. Therefore, because of the difference of the gauge-fixing methods, for a fixed regularization parameter, the computational cost of the RSC estimator would be lower than that of GST. From the perspective of numerical stability, the RSCQT approach would be superior than GST as well. When we combine a k-fold cross validation with the RSC estimator, we have to perform the optimization of the objective function many times. In that case, it is unclear which computational cost is lower, which depends on how hard the optimization of the gauge degrees of freedom in GST with full physicality constraints is.

In quantum computation based on the circuit model, a computational process is constructed with combinations of 1-qubit state preparations, 1-qubit measurements, 1-qubit gates, and 2-qubit gates [45]. If we restrict the use of the RSC estimator to such small subsystems, the exponential increase of the implementation costs mentioned above poses no problem. Let n_Q denote the number of qubits in a device. In cases, where qubits are aligned at each node on a 2-dimensional square-grid lattice, the total number of possible locations of 1-qubit and nearest neighbor 2-qubit operations increases linearly with respect to n_Q. Even if there is concern about crosstalk errors and we need to evaluate nearest k-qubit subsystems, the scaling of the cost of characterization with the RSC estimator still remains linear with respect to n_Q, where k is assumed to be small and independent of n_Q. Therefore if we focus on reliable characterization of elementary quantum operations on the physical layer, the high implementation cost of RSC estimator would not be fatal disadvantage. Naturally, lower computational cost is better, and therefore it is important to develop more stable, more accurate, and faster numerical algorithms for solving the minimization in Eq. (8). A task to be tackled in the near future is a numerical implementation of the RSC estimator for a 2-qubit system.

VI. CONCLUSION

In this paper, we considered a quantum characterization problem based on the self-consistent quantum tomographic approach. First, we derived a sufficient condition on experimental designs which enables us to access all information of a set of unknown state preparations, measurements, and gates except for the gauge degrees of freedom. Second, we proposed a self-consistent estimator with regularization and physicality constraints. We theoretically proved that, by appropriately tuning the strength of the regularization, the sequence of estimates converges to the gauge-equivalence class of the prepared true set of operations at the limit of the data size going to infinity. This guarantees the high reliability of the estimation results for sufficiently large amount of data. We also theoretically derived the rate of the asymptotic convergence, which is expected to be optimal. These are the first mathematically rigorous proofs of asymptotic behaviors of a self-consistent quantum tomography method. Additionally, we also proposed how to extract information of dynamics generators such as Hamiltonian and dissipator from a tomographic estimation result of a gate. These theoretical results hold for any finite-dimensional systems. Third, we implemented the self-consistent estimator for 1-qubit system, performed numerical experiments, and numerically analyzed its performances for finite amount of data. The numerical results compatible with the theoretical results, and it is numerically shown that the proposed estimator has predictability of the true probability distributions higher than that of empirical distributions, even though there exists a bias originated from the regularization. The numerical results also showed that the existence of the gauge degrees of freedom makes it difficult to directly use the estimation result for the accuracy improvement step, which would be a common feature of the self-consistent approach. In order to make quantum technologies, e.g., quantum computation, quantum communication, and quantum sens-

ing more practical, it is indispensable to develop more reliable characterization method for accuracy validation and accuracy improvement of elementary quantum operations. Theoretical and numerical results indicate that the method is suitable for the reliable accuracy validation and needs additional ingenuity for contributing to the accuracy improvement.

ACKNOWLEDGEMENTS

TS would like to thank Yasunobu Nakamura, Yutaka Tabuchi, and Shuhei Tamate for their helpful discussions on realistic error models in superconducting quantum circuits, Yasunari Suzuki, Haruhisa Nagata, and Kazuki Fukui for their useful advice on numerical simulation, and Rekishu Yamazaki, Dany Lachance-Quirion, and Florian R. Schlederer for their helpful comments on the manuscript. This work was supported by JSPS KAKENHI Grant Numbers JP24700273, JP16K13775, JST ERATO Grant Number JPMJER1601, and "Funds for the Development of Human Resources in Science and Technology" under MEXT, through the "Home for Innovative Researchers and Academic Knowledge Users (HIRAKU)" consortium.

Appendix A: Parametrization of Quantum Operations

In this section, we explain real vector parametrization of state preparation, measurement (POVM), and gate. Let $\mathcal{H}$ denote a quantum system of interest. The dimension of $\mathcal{H}$ is finite, denoted by d.

Let ρ denote a density matrix on $\mathcal{H}$, which is a $d \times d$ complex matrix that is trace-one and positive semidefinite, i.e.,

$$\mathrm{Tr}[\rho] = 1 \ \& \ \rho \succeq 0. \tag{A1}$$

Let $\boldsymbol{B} = \{B_\alpha\}_{\alpha=0}^{d^2-1}$ denote a $d \times d$ Hermitian orthonormal matrix basis with $B_0 = I/\sqrt{d}$. From the completeness and orthogonality of the basis, we can uniquely expand any density matrix by $\boldsymbol{B}$ as

$$\rho = \sum_{\alpha=0}^{d^2-1} \rho_\alpha B_\alpha. \tag{A2}$$

From the Hermiticity of $\boldsymbol{B}$ and ρ, the expansion coefficients ρ_α are real. The trace-one condition leads to $\rho_0 = 1/\sqrt{d}$. The other (d^2-1) real numbers parametrize the density matrix. The positive-semidefiniteness condition restricts the possible range of $(\rho_1, \ldots, \rho_{d^2-1})$ into a compact convex region in $\mathbb{R}^{d^2-1}$.

There are two objects of description of a quantum measurement as a quantum operation, a probability distribution of measurement outcome and state transformations with respect to an obtained outcome. A positive operator-valued measure (POVM) can treat former only, and a quantum measurement process can treat both of them. In the main text, we do not consider a quantum state after measurement, and here we explain POVM. Let us assume that the set of possible outcomes of a measurement is discrete and finite. Then a POVM $\boldsymbol{\Pi} = \{\Pi_x\}_{x=0}^{m-1}$ is a discrete and finite set of $d \times d$ Hermitian matrices that is sum-identity and positive-semidefinite each, i.e.,

$$\sum_{x=0}^{m-1} \Pi_x = I \ \& \ \Pi_x \succeq 0. \tag{A3}$$

From the sum-identity condition, one of m elements of the POVM is fixed, e.g., as $\Pi_{m-1} = I - \sum_{x=0}^{m-2} \Pi_x$. We expand the $(m-1)$ matrices as a density matrix.

$$\Pi_x = \sum_{\alpha=0}^{d^2-1} \Pi_{x,\alpha} B_\alpha, \ x = 0, \ldots, m-2. \tag{A4}$$

From the Hermiticity of $\boldsymbol{B}$ and each Π_x, the $(m-1) \times d^2$ expansion coefficients $\Pi_{x,\alpha}$ are real and parametrize the POVM. The positive-semidefiniteness condition restricts the possible range of the parameters into a compact convex region in $\mathbb{R}^{(m-1)d^2}$.

A quantum gate transforms a state ρ to another state ρ', and the action is described by a linear map $\mathcal{G} : \rho \mapsto \rho' = \mathcal{G}(\rho)$ that is completely positive (CP) and trace-preserving (TP). An action of a linear map can be represented by a matrix. Let us choose a matrix representation along with the real vector representation of a state $|\rho\rangle\!\rangle := (\rho_0, \ldots, \rho_{d^2-1})^T \in \mathbb{R}^{d^2}$ with respect to the basis $\boldsymbol{B}$. Let G denote the matrix representation of $\mathcal{G}$. Then it is a $d^2 \times d^2$ real matrix. The TP condition leads to equations,

$$\mathrm{G}_{0\beta} = \delta_{0\beta} \ (\beta = 0, \ldots, d^2-1). \tag{A5}$$

The other $(d^2-1) \times d^2$ numbers $G_{\alpha\beta}$ $(\alpha = 1, \ldots, d^2-1, \beta = 0, \ldots, d^2-1)$ are real and parametrize the gate. The CP condition restricts the possible range of the parameters into a compact convex region in $\mathbb{R}^{(d^2-1)d^2}$. We introduce another matrix representation of a linear map, called Choi-Jamiołkowski matrix, $\mathrm{CJ}(\mathcal{G}) \in \mathbb{C}^{d^2 \times d^2}$, in order to treat the CP condition mathmatically. Let us define a vector $\mathrm{vec}(I) := \sum_{i=0}^{d} |i\rangle|i\rangle \in \mathbb{C}^{d^2}$. With the vector, the CJ matrix is defined as

$$\mathrm{CJ}(\mathcal{G}) := (\mathcal{G} \otimes \mathcal{I})\mathrm{vec}(I)\mathrm{vec}(I)^\dagger, \tag{A6}$$

where $\mathcal{I}$ is the identity map on $\mathbb{C}^{d \times d}$. The CP condition on $\mathcal{G}$ is equivalent to the matrix inequality,

$$\mathrm{CJ}(\mathcal{G}) \succeq 0. \tag{A7}$$

An explicit form of the relation between CJ and HS matrices is given at Eq. (H8).

For a given set of quantum operations consisting of n_s states, n_p POVMs, and n_g gates, i.e., $\{\rho_0, \ldots, \rho_{n_\mathrm{s}-1}, \ \boldsymbol{\Pi}_0, \ldots, \boldsymbol{\Pi}_{n_\mathrm{p}}, \ \mathcal{G}_0, \ldots, \mathcal{G}_{n_\mathrm{g}-1}\}$, the real

vector $\boldsymbol{s}$ for the set is the set of parameters for each operation, i.e.,

$$\begin{aligned}\boldsymbol{s} = (&\rho_{0,1},\dots,\rho_{0,d^2-1},\ \dots,\rho_{n_{\rm s}-1,1},\dots,\rho_{n_{\rm s}-1,d^2-1},\\ &\Pi_{0,0},\dots,\Pi_{0,d^2-1},\dots,\Pi_{m-2,0},\dots,\Pi_{m-2,d^2-1},\\ &G_{0,10},\dots,G_{0,d^2-1d^2-1},\dots,G_{n_{\rm g}-1,10},\dots,\\ &G_{n_{\rm g}-1,d^2-1d^2-1})^T. \end{aligned} \tag{A8}$$

Theoretical results in Sec. III hold for any $n_{\rm s}$, $n_{\rm p}$, and $n_{\rm g}$. The parametrization in Eq. (A8) is used in numerical experiments for a 1-qubit system explained in Sec. IV, which were performed for the case of $n_{\rm s}=1$, $n_{\rm p}=1$, and $n_{\rm g}=3$.

Appendix B: Gauge degrees of freedom

In this section, we explain a mathematical treatment of the gauge degrees of freedom. For a given set of quantum operations $\boldsymbol{s}$, a gauge transformation $\mathcal{A}$ is an invertible map from a set of quantum operations to another set that satisfies $\mathcal{A}(\boldsymbol{s})\in[\boldsymbol{s}]$, in which a set of quantum operations and a parameter characterizing the set are identified as explained in Sec. II. Let $\mathcal{A}_{\rm s}$, $\mathcal{A}_{\rm p}$, and $\mathcal{A}_{\rm g}$ denote the maps corresponding to the action of $\mathcal{A}$ on a state, POVM, and gate, respectively, e.g., for $\boldsymbol{s}=\{\rho,\boldsymbol{\Pi},\mathcal{G}\}$, $\mathcal{A}(\boldsymbol{s})=\{\mathcal{A}_{\rm s}(\rho),\mathcal{A}_{\rm p}(\boldsymbol{\Pi}),\mathcal{A}_{\rm g}(\mathcal{G})\}$. By definition, gauge-equivalent sets give an equivalent probability distribution for a given experimental setting. The functionalities of $\mathcal{A}_{\rm s}$, $\mathcal{A}_{\rm p}$, and $\mathcal{A}_{\rm g}$ are limited into linear because of the linearity of the Born's rule.

Let X denote a $d\times d$ complex matrix with the form $X=\sum_{\alpha=0}^{d^2-1}X_\alpha B_\alpha$, where $\boldsymbol{B}=\{B_\alpha\}$ is an orthonormal matrix basis. We introduce a vectorization of X with respect to the basis, $|X\rangle\!\rangle:=\sum_{\alpha=0}^{d^2-1}X_\alpha \boldsymbol{e}_\alpha$, where $\{\mathbb{e}_\alpha\}_{\alpha=0}^{d^2-1}$ is an orthonormal basis on $\mathbb{C}^{d^2}$. The vectorization keeps the value of the Hilbert-Schmidt inner product of two matrices, i.e.,

$$\mathrm{Tr}[Y^\dagger X]=\langle\!\langle Y|X\rangle\!\rangle. \tag{B1}$$

An action of a linear map on ρ can be represented by a matrix on the vectorization $|\rho\rangle\!\rangle$, as in the case of a gate explained in Appendix. A. The matrix representation of a linear map, say $\mathcal{F}$, is called the Hilbert-Schmidt (HS) representation, and we use a notation, $\mathrm{HS}(\mathcal{F})$ for the representation. Let A denote the matrix representation of $\mathcal{A}_{\rm s}$, i.e.,

$$|\mathcal{A}_{\rm s}(\rho)\rangle\!\rangle = A|\rho\rangle\!\rangle. \tag{B2}$$

Then the Born's rule enacts the action of $\mathcal{A}_{\rm p}$, and $\mathcal{A}_{\rm g}$ as follows.

$$\langle\!\langle\mathcal{A}_{\rm p}(\Pi_x)| = \langle\!\langle\Pi_x|A^{-1}, \tag{B3}$$

$$\mathrm{HS}(\mathcal{A}_{\rm g}(\mathcal{G})) = A\,\mathrm{HS}(\mathcal{G})A^{-1}. \tag{B4}$$

Therefore, the action of a gauge transformation $\mathcal{A}$ on a state, POVM, and gate is characterized by a $d^2\times d^2$ matrix as

$$\mathcal{A}:\ \begin{array}{ccc}|\rho\rangle\!\rangle & \mapsto & A|\rho\rangle\!\rangle\\ \langle\!\langle\Pi_x| & \mapsto & \langle\!\langle\Pi_x|A^{-1}\\ G & \mapsto & AGA^{-1}\end{array} \tag{B5}$$

In general, a gauge transformation $\mathcal{A}$ is an invertible linear map, and a transformed set $\mathcal{A}(\boldsymbol{s})$ is not guaranteed to be physical even if the original set $\boldsymbol{s}$ is physical. If we require Hermiticity of the transformed state $\mathcal{A}_{\rm s}(\rho)$ as well, $\mathcal{A}_{\rm s}$ is an Hermiticity-preserving map. When we choose $\boldsymbol{B}$ as each B_α is Hermitian, the HS representation of an Hermiticity-preserving (HP) map is a real matrix. If we require the trace-oneness of the transformed state, the $\mathcal{A}_{\rm s}$ is a trace-preserving map. If we choose $\boldsymbol{B}$ such that $B_0=I/\sqrt{d}$, the HS representation of a trace-preserving map satisfies $A_{0\beta}=\delta_{0\beta}$ for $\beta=0,\dots,d^2-1$. In that case, the representation can be written as

$$A=\begin{bmatrix}1 & \boldsymbol{0}^T\\ \mathbf{b} & C\end{bmatrix}, \tag{B6}$$

where $\boldsymbol{b}$ is a d^2-1 dimensional vector, C is a $(d^2-1)\times(d^2-1)$ matrix, and T denotes the transposition with respect to the indexing of the HS representation. $\boldsymbol{0}^T$ denotes the d^2-1 dimensional zero vector transposed. The inverse A^{-1} has the following form

$$A^{-1}=\begin{bmatrix}1 & \boldsymbol{0}^T\\ -C^{-1}\mathbf{b} & C^{-1}\end{bmatrix}, \tag{B7}$$

Let us choose an Hermitian orthonormal matrix basis $\boldsymbol{B}$ satisfying $B_0=I/\sqrt{d}$. Then the vectorization of a density matrix ρ is represented as

$$|\rho\rangle\!\rangle=\frac{1}{\sqrt{d}}\begin{bmatrix}1\\ \boldsymbol{v}\end{bmatrix}. \tag{B8}$$

The parameter vector $\boldsymbol{v}\in\mathbb{R}^{d^2-1}$ is a generalized Bloch vector. The transformed vectorized density matrix via a TPHP gauge transformation is

$$A|\rho\rangle\!\rangle=\frac{1}{\sqrt{d}}\begin{bmatrix}1\\ C\boldsymbol{v}+\boldsymbol{b}\end{bmatrix}. \tag{B9}$$

From the singular value decomposition of C, the matrix C contains actions of rotation and rescaling of $\boldsymbol{v}$. The vector $\boldsymbol{b}$ acts as the origin shift. Therefore the actions of a gauge transformation are categorized into rotation, rescaling, and origin shift. The rescaling and origin shift can cause an unphysical $\mathcal{A}(\boldsymbol{s})$. The number of degrees of freedom characterizing a TPHP gauge transformation is d^4-d^2.

Appendix C: Proof of Theorem 1

In this section, we give a proof of Theorem 1. First, we mention two lemmas on vector bases as a preparation for the proof. Let dim denote a finite positive integer. We will set $\dim=d^2$ in the proof.

Lemma 1 *Let* $\{\boldsymbol{a}_i\}_{i=1}^{\dim}$ *and* $\{\boldsymbol{b}_i\}_{i=1}^{\dim}$ *denote bases of a* dim*-dimensional complex vector space* $\mathbb{C}^{\dim}$*. Then there exists a unique invertible* $\dim \times \dim$ *matrix* C *satisfying*

$$\boldsymbol{b}_i = C\boldsymbol{a}_i,\ i = 1, \ldots, \dim. \tag{C1}$$

Proof (Lemma 1): Let $\{\mathbb{e}_i\}_{i=1}^{d}$ denote an orthonormal basis of $\mathbb{C}^{\dim}$. There exist unique invertible matrices A and B satisfying

$$\boldsymbol{a}_i = A\mathbb{e}_i,\ \boldsymbol{b}_i = B\mathbb{e}_i,\ \forall i = 1, \ldots, \dim. \tag{C2}$$

Then

$$\boldsymbol{b}_i = BA^{-1}\boldsymbol{a}_i \tag{C3}$$

holds, and $C = BA^{-1}$. From the uniqueness and invertibility of A and B, C is also unique and invertible. □

Lemma 2 *Let* $\{\boldsymbol{a}_i\}_{i=1}^{\dim}$ *and* $\{\boldsymbol{b}_i\}_{i=1}^{\dim}$ *denote bases of* $\mathbb{C}^{dim}$*. If matrices* $X, Y \in \mathbb{C}^{\dim \times \dim}$ *satisfy*

$$\boldsymbol{b}_j \cdot X\boldsymbol{a}_i = \boldsymbol{b}_j \cdot Y\boldsymbol{a}_i, \forall i, j = 1, \ldots, \dim, \tag{C4}$$

then $X = Y$ *holds.*

Proof (Lemma 2): As introduced in the proof of Lemma 1, there exist unique invertible matrices A and B satisfying

$$\boldsymbol{a}_i = A\mathbb{e}_i,\ \boldsymbol{b}_i = B\mathbb{e}_i,\ \forall i = 1, \ldots, \dim. \tag{C5}$$

Then

$$\mathbb{e}_j \cdot B^\dagger XA\mathbb{e}_i = \boldsymbol{b}_j \cdot X\boldsymbol{a}_i = \boldsymbol{b}_j \cdot Y\boldsymbol{a}_i = \mathbb{e}_j \cdot B^\dagger YA\mathbb{e}_i \tag{C6}$$

Therefore we have $B^\dagger XA = B^\dagger YA$. From the invertibility of A and B, $X = Y$ holds. □

Second, we introduce a lemma on informationally complete sets of states and POVMs with gauge-equivalence. In the following, we set $\dim = d^2$. Let $|\rho\rangle\!\rangle$, $|\mathbf{\Pi}\rangle\!\rangle = \{|\Pi_x\rangle\!\rangle\}_{x\in\mathcal{X}}$, and G denote a vectorized representation of a density matrix ρ, the same representation of a POVM $\mathbf{\Pi}$, and a HS representation of a TPCP map $\mathcal{G}$ [26, 47] as explained in Appendix A. The vectors $|\rho\rangle\!\rangle$ and $|\Pi_x\rangle\!\rangle$ are in $\mathbb{C}^{\dim}$ and the matrix G is in $\mathbb{C}^{\dim\times\dim}$. Then generalized Born's rule can be rewritten with the vector representation as

$$p(x|\rho, \mathcal{G}, \mathbf{\Pi}) = \mathrm{Tr}\left[\Pi_x \mathcal{G}(\rho)\right] = \langle\!\langle \Pi_x | G | \rho \rangle\!\rangle. \tag{C7}$$

Note that

$$\langle\!\langle \mathcal{G}^\dagger(\Pi_x)| = \langle\!\langle \Pi_x | G \tag{C8}$$

holds.

Lemma 3 *Suppose that* $\{|\rho^i\rangle\!\rangle\}_{i=1}^{N_s}$ *and* $\{|\tilde{\rho}^i\rangle\!\rangle\}_{i=1}^{N_s}$ *are state-informationally complete and* $\{|\mathbf{\Pi}^j\rangle\!\rangle\}_{j=1}^{N_p}$ *and* $\{|\tilde{\mathbf{\Pi}}^j\rangle\!\rangle\}_{j=1}^{N_p}$ *are POVM-informationally complete. If*

$$\langle\!\langle \Pi_x^j | \rho^i \rangle\!\rangle = \langle\!\langle \tilde{\Pi}_x^j | \tilde{\rho}^i \rangle\!\rangle, \tag{C9}$$

holds for any i*,* j*, and* x*, then there exists a unique invertible matrix* A *satisfying*

$$|\tilde{\rho}^i\rangle\!\rangle = A|\rho^i\rangle\!\rangle, \tag{C10}$$

$$\langle\!\langle \tilde{\Pi}_x^j| = \langle\!\langle \Pi_x^j | A^{-1}, \tag{C11}$$

for any i*,* j*, and* x*.*

Proof (Lemma 3): We divide each set into a linear independent subset subscripted with 1 and the residual subset subscripted with 2.

$$\{|\rho^i\rangle\!\rangle\}_{i=1}^{N_s} = \mathcal{S}_1 \cap \mathcal{S}_2, \tag{C12}$$

$$\mathcal{S}_1 := \{|\rho^i\rangle\!\rangle\}_{i=1}^{\dim}, \tag{C13}$$

$$\mathcal{S}_2 := \{|\rho^i\rangle\!\rangle\}_{i=\dim+1}^{N_s}, \tag{C14}$$

$$\{|\tilde{\rho}^i\rangle\!\rangle\}_{i=1}^{N_s} = \tilde{\mathcal{S}}_1 \cap \tilde{\mathcal{S}}_2, \tag{C15}$$

$$\tilde{\mathcal{S}}_1 := \{|\tilde{\rho}^i\rangle\!\rangle\}_{i=1}^{\dim}, \tag{C16}$$

$$\tilde{\mathcal{S}}_2 := \{|\tilde{\rho}^i\rangle\!\rangle\}_{i=\dim+1}^{N_s}, \tag{C17}$$

$$\{|\mathbf{\Pi}^j\rangle\!\rangle\}_{j=1}^{N_p} = \mathcal{P}_1 \cap \mathcal{P}_2, \tag{C18}$$

$$\mathcal{P}_1 := \{|\Pi_{x_j}^j\rangle\!\rangle\}_{j=1}^{\dim}, \tag{C19}$$

$$\mathcal{P}_2 := \{|\Pi_{x_j}^j\rangle\!\rangle\}_{j=\dim+1}^{|\mathcal{X}|\cdot N_p}, \tag{C20}$$

$$\{|\tilde{\mathbf{\Pi}}^j\rangle\!\rangle\}_{j=1}^{N_p} = \tilde{\mathcal{P}}_1 \cap \tilde{\mathcal{P}}_2, \tag{C21}$$

$$\tilde{\mathcal{P}}_1 := \{|\tilde{\Pi}_{x_j}^j\rangle\!\rangle\}_{j=1}^{\dim}, \tag{C22}$$

$$\tilde{\mathcal{P}}_2 := \{|\tilde{\Pi}_{x_j}^j\rangle\!\rangle\}_{j=\dim+1}^{|\mathcal{X}|\cdot N_p}. \tag{C23}$$

During the division process, if necessary, we relabel the indices i, j, and x so that $\mathcal{S}_1$ and $\mathcal{P}_1$ are bases of $\mathbb{C}^{\dim}$. From Lemma 1, there exist unique invertible matrices A and B satisfying

$$|\tilde{\rho}^i\rangle\!\rangle = A|\rho^i\rangle\!\rangle, \forall i = 1, \ldots, d, \tag{C24}$$

$$\langle\!\langle \tilde{\Pi}_{x_j}^j| = \langle\!\langle \tilde{\Pi}_{x_j}^j | B, \forall j = 1, \ldots, \dim. \tag{C25}$$

Then

$$\langle\!\langle \Pi_{x_j}^j | \rho^i \rangle\!\rangle = \langle\!\langle \tilde{\Pi}_x^j | \tilde{\rho}^i \rangle\!\rangle = \langle\!\langle \Pi_{x_j}^j | BA | \rho^i \rangle\!\rangle \tag{C26}$$

holds for $i, j = 1, \ldots, d$. From Lemma 2, we have $B = A^{-1}$. Therefore it is proven that there exists a unique matrix A satisfying Eqs. (C10) and (C11) for the linear independent subsets $\mathcal{S}_1$, $\tilde{\mathcal{S}}_1$, $\mathcal{P}_1$, and $\tilde{\mathcal{P}}_1$.

Suppose that $\dim < k \leq N_{\mathrm{s}}$ in the case of $\dim < N_{\mathrm{s}}$. We can span any residual vectors in $\mathcal{S}_2$ and $\tilde{\mathcal{S}}_2$ by $\mathcal{S}_1$ and $\tilde{\mathcal{S}}_1$ as

$$|\rho^k\rangle\!\rangle = \sum_{i=1}^{\dim} c_{ki}|\rho^i\rangle\!\rangle,\ |\tilde{\rho}^k\rangle\!\rangle = A\left(\sum_{i=1}^{\dim} \tilde{c}_{ki}|\rho^i\rangle\!\rangle\right). \tag{C27}$$

Then from Eq. (C9),

$$\begin{aligned}\langle\!\langle \Pi_{x_j}^j| \left(\sum_{i=1}^{\dim} c_{ki}|\rho^i\rangle\!\rangle\right) &= \langle\!\langle \Pi_{x_j}^j | \rho^k \rangle\!\rangle = \langle\!\langle \tilde{\Pi}_{x_j}^j | \tilde{\rho}^k \rangle\!\rangle \\ &= \langle\!\langle \Pi_{x_j}^j | BA \left(\sum_{i=1}^{\dim} \tilde{c}_{ki}|\rho^i\rangle\!\rangle\right) = \langle\!\langle \Pi_{x_j}^j| \left(\sum_{i=1}^{\dim} \tilde{c}_{ki}|\rho^i\rangle\!\rangle\right)\end{aligned} \tag{C28}$$

holds for $j = 1, \ldots, \dim$. Then

$$|\tilde{\rho}^k\rangle\!\rangle = A|\rho^k\rangle\!\rangle \tag{C29}$$

holds for $k = \dim +1, \ldots, N_{\rm s}$.

In the same way as the state vector, we can prove

$$\langle\!\langle\tilde{\Pi}^j_{x_j}| = \langle\!\langle\Pi^j_{x_j}|A^{-1},\ \forall j = \dim +1, \ldots, |\mathcal{X}| \cdot N_{\rm p}. \tag{C30}$$

□

In the proof above, we assumed that numbers of possible outcomes are common for all $\mathbf{\Pi}^j$ for simplicity. A generalization to cases that each POVM has different number of elements is straightforward.

Now we are ready for proving Theorem 1.

Proof (Theorem 1): When $\tilde{\boldsymbol{s}} \in [\boldsymbol{s}]$, $\boldsymbol{p}(\mathbf{Id}, \tilde{\boldsymbol{s}}) = \boldsymbol{p}(\mathbf{Id}, \boldsymbol{s})$ holds by definition of the gauge-equivalence. Here we prove the opposite direction, i.e., when $\mathbf{Id}$ is SCIC and each gate in $\boldsymbol{s}$ has the inverse, which can be unphysical, then $\boldsymbol{p}(\mathbf{Id}, \tilde{\boldsymbol{s}}) = \boldsymbol{p}(\mathbf{Id}, \boldsymbol{s})$ implies $\tilde{\boldsymbol{s}} \in [\boldsymbol{s}]$. When $\mathbf{Id}$ is SCIC, it includes a set of index sequences satisfying Eq. (3). $\mathbf{Id}_{\rm s}$ and $\mathbf{Id}_{\rm p}$ included in $\mathbf{Id}$ are state- and POVM-informationally complete, respectively. From Lemma 3, equations

$$\langle\!\langle\tilde{\Pi}^{\boldsymbol{i}_{\rm p}}_x|\tilde{\rho}^{\boldsymbol{i}_{\rm s}}\rangle\!\rangle = \langle\!\langle\Pi^{\boldsymbol{i}_{\rm p}}_x|\rho^{\boldsymbol{i}_{\rm s}}\rangle\!\rangle \tag{C31}$$

imply that there exists a unique matrix A such that

$$|\tilde{\rho}^{\boldsymbol{i}_{\rm s}}\rangle\!\rangle = A|\rho^{\boldsymbol{i}_{\rm s}}\rangle\!\rangle,\ \forall \boldsymbol{i}_{\rm s} \in \mathbf{Id}_{\rm s}, \tag{C32}$$

$$\langle\!\langle\tilde{\Pi}^{\boldsymbol{i}_{\rm p}}_x| = \langle\!\langle\Pi^{\boldsymbol{i}_{\rm p}}_x|A^{-1},\ \forall \boldsymbol{i}_{\rm p} \in \mathbf{Id}_{\rm p},\ x \in \mathcal{X}. \tag{C33}$$

The SCIC $\mathbf{Id}$ also includes a set of index sequences satisfying Eq. (4). Then

$$\begin{aligned}\langle\!\langle\Pi^{\boldsymbol{i}_{\rm p}}_x|{\rm G}_{i_{\rm g}}|\rho^{\boldsymbol{i}_{\rm s}}\rangle\!\rangle &= \langle\!\langle\tilde{\Pi}^{\boldsymbol{i}_{\rm p}}_x|\tilde{\rm G}_{i_{\rm g}}|\tilde{\rho}^{\boldsymbol{i}_{\rm s}}\rangle\!\rangle \\ &= \langle\!\langle\Pi^{\boldsymbol{i}_{\rm p}}_x|A^{-1}\tilde{\rm G}_{i_{\rm g}}A|\rho^{\boldsymbol{i}_{\rm s}}\rangle\!\rangle\end{aligned} \tag{C34}$$

hold for $\boldsymbol{i}_{\rm s} \in \mathbf{Id}_{\rm s}$, $i_{\rm g} = 1, \ldots, n_{\rm g}$, $\boldsymbol{i}_{\rm p} \in \mathbf{Id}_{\rm p}$, and $x \in \mathcal{X}$. From Lemma 2, we have

$${\rm G}_{i_{\rm g}} = A^{-1}\tilde{\rm G}_{i_{\rm g}}A \Leftrightarrow \tilde{\rm G}_{i_{\rm g}} = A{\rm G}_{i_{\rm g}}A^{-1}, \tag{C35}$$

for $i_{\rm g} = 1, \ldots, n_{\rm g}$. With Eqs. (C35), (C32) and (C33) can be rewritten as

$$A{\rm G}^{\boldsymbol{i}_{\rm s}}A^{-1}|\tilde{\rho}\rangle\!\rangle = A{\rm G}^{\boldsymbol{i}_{\rm s}}|\rho\rangle\!\rangle, \tag{C36}$$

$$\langle\!\langle\tilde{\Pi}_x|A{\rm G}^{\boldsymbol{i}_{\rm p}}A^{-1} = \langle\!\langle\Pi_x|{\rm G}^{\boldsymbol{i}_{\rm p}}A^{-1}, \tag{C37}$$

where ${\rm G}^{\boldsymbol{i}_{\rm s}}$ and ${\rm G}^{\boldsymbol{i}_{\rm p}}$ are HS representations of gates constructed by applying $\mathcal{G}_{i_{\rm g}}$ along with $\boldsymbol{i}_{\rm s}$ and $\boldsymbol{i}_{\rm p}$, respectively. From the invertibility of A and ${\rm G}_{i_{\rm g}}$, we obtain

$$|\tilde{\rho}\rangle\!\rangle = A|\rho\rangle\!\rangle,\ \langle\!\langle\tilde{\Pi}_x| = \langle\!\langle\Pi_x|A^{-1}. \tag{C38}$$

Let $\boldsymbol{i}$ denote an arbitrary gate index sequences whose each element is in $\{1, \ldots, n_{\rm g}\}$. The length of $\boldsymbol{i}$ is arbitrary, and $\boldsymbol{i}$ itself is not necessarily in $\mathbf{Id}$. Eqs. (C38) and (C35) lead

$$\boldsymbol{p}^{\boldsymbol{i}}(\tilde{\boldsymbol{s}}) = \boldsymbol{p}^{\boldsymbol{i}}(\boldsymbol{s}). \tag{C39}$$

Therefore $\tilde{\boldsymbol{s}} \in [\boldsymbol{s}]$. □

In the proof above, the numbers of states and POVMs are assumed to be 1. A generalization of the proof to cases of multiple states and POVMs is straightforward.

Appendix D: Proof of Theorem 2

In this section, we give a proof of Theorem 2. In order to clarify the roles of each properties of the RSC estimator in the setting of SCQT and each condition mentioned in Theorem 2, we split the proof into two parts. First, we prove theorems for a general setting of statistical parameter estimation with the following conditions:

Conditions

- C.1 The parameter space $\mathcal{S}$ is a compact subset of a finite-dimensional Euclidean space, and the parametrization of $\boldsymbol{p}(\mathbf{Id}, \boldsymbol{s})$ is continuous.
- C.2 The regularization function ${\rm R}(\boldsymbol{s}, \boldsymbol{s}')$ is positive and bounded.
- C.3 The regularization parameter r_N is positive and satisfies $\lim_{N\to\infty} r_N = 0$ a.s..
- C.4 The regularization parameter r_N is positive and satisfies $r_N \lesssim {\rm L}(\boldsymbol{p}(\mathbf{Id}, \boldsymbol{s}^{\rm true}), \boldsymbol{f}_N(\mathbf{Id}))$.
- C.5 For a given $\boldsymbol{s}^{\rm true} \in \mathcal{S}$, a point $\boldsymbol{s} \in \mathcal{S}$ satisfying $\boldsymbol{p}^{\boldsymbol{i}}(\boldsymbol{s}) = \boldsymbol{p}^{\boldsymbol{i}}(\boldsymbol{s}^{\rm true})$ is uniquely determined up to the gauge-equivalence. In other words, $\boldsymbol{s} \notin [\boldsymbol{s}^{\rm true}] \Leftrightarrow \exists \boldsymbol{i} \in \mathbf{Id}$ such that $\left\|\boldsymbol{p}^{\boldsymbol{i}}(\boldsymbol{s}) - \boldsymbol{p}^{\boldsymbol{i}}(\boldsymbol{s}^{\rm true})\right\|_2 > 0$.

Second, we show that the theorems for the general setting are applicable to the RSC estimator in the setting of SCQT. In Appendix D 1, we give a rigorous definition of the asymptotic notation, $\lesssim$. The definition is used in proofs in this section. In Appendix D 2, we introduce a lemma about bounds of the asymptotic convergence rate of the empirical distributions to the true probability distributions. The lemma is used in Appendix D 4. In Sec. D 3, we prove Eq. (10). In Appendix D 4, we prove Eqs. (12) and (13). In Appendix D 5, we prove Eq. (14). Main tools used in the proofs are the property of $\boldsymbol{s}^{\rm est}_N$ as the minimizer of the objective function, the triangle inequality of norms, the strong law of large numbers, the central limit theorem, and the strong law of iterated logarithm.

1. Definition of the Asymptotic Notation

We give a rigorous definition of the asymptotic notation, "$\lesssim$", introduced in the main text and used in Theorem 2. Suppose that $f(N)$ and $g(N)$ are positive functions of the data size $N > 0$. Then the notation is defined as

$$f(N) \lesssim g(N) \overset{\rm def}{\Longleftrightarrow} \limsup_{N\to\infty} \frac{f(N)}{g(N)} < \infty. \tag{D1}$$

Equivalently, $f(N) \lesssim g(N)$ holds if and only if there exists a positive real numbers a and N_0 such that

$$f(N) \leq ag(N),\ \forall\ N \geq N_0. \tag{D2}$$

This is equivalent to the big O notation, $f(N) \in O(g(N))$, in computer science.

2. Asymptotic convergence rate of empirical distributions

We introduce a lemma for proving Eqs. (12) and (13).

Lemma 4 *The asymptotic convergence rate of* $\boldsymbol{f}_N(\mathbf{Id})$ *to* $\boldsymbol{p}(\mathbf{Id}, \boldsymbol{s}^{true})$ *is bounded as*

$$\frac{1}{\sqrt{N}} \lesssim \sqrt{\mathrm{L}\big(\boldsymbol{p}(\mathbf{Id}, \boldsymbol{s}^{true}), \boldsymbol{f}_N(\mathbf{Id})\big)} \lesssim \sqrt{\frac{\ln \ln N}{N}} \text{ a.s..} \quad \text{(D3)}$$

Proof (Lemma 4) : First, we prove the left inequality of Eq. (D3) by contradiction to the central limit theorem. We assume

$$\mathrm{L}\left(\boldsymbol{p}(\mathbf{Id}, \boldsymbol{s}^{\mathrm{true}}), \boldsymbol{f}_N(\mathbf{Id})\right) < \frac{C}{N} \text{ a.s.}, \quad \text{(D4)}$$

for arbitrary positive constant C and sufficiently large N. Then, due to the dominant convergence theorem, we obtain

$$\mathbb{E}\left[\mathrm{L}(\boldsymbol{p}(\mathbf{Id}, \boldsymbol{s}^{\mathrm{true}}), \boldsymbol{f}_N(\mathbf{Id}))\right] < \frac{C}{N}, \quad \text{(D5)}$$

for arbitrary positive constant C and sufficiently large N, where $\mathbb{E}$ denote the expectation with respect to the observed measurement outcomes. On the other hand, the central limit theorem [37] leads to

$$\mathbb{E}\left[\mathrm{L}\big(\boldsymbol{p}(\mathbf{Id}, \boldsymbol{s}^{\mathrm{true}}), \boldsymbol{f}_N(\mathbf{Id})\big)\right] \propto \frac{1}{N}. \quad \text{(D6)}$$

Eq. (D5) contradicts Eq. (D6). Therefore there exists a positive number a such that

$$\frac{a}{N} \leq \mathrm{L}\big(\boldsymbol{p}(\mathbf{Id}, \boldsymbol{s}^{\mathrm{true}}), \boldsymbol{f}_N(\mathbf{Id})\big) \text{ a.s.}, \quad \text{(D7)}$$

for any sufficiently large N, and we obtain the first inequality to be proved.

The right inequality of Eq. (D3) is the strong law of iterated logarithm [37] itself. □

3. Proof of Eq. (10)

We prove Eq. (10) in Theorem 2. First, we prove an equivalent statement under Conditions C.1, C.2, and C.3.

Theorem 3 *If C.1, C.2, and C.3 are satisfied, then*

$$\lim_{N\to\infty} \sqrt{\mathrm{L}(\boldsymbol{p}(\mathbf{Id}, \boldsymbol{s}_N^{est}), \boldsymbol{p}(\mathbf{Id}, \boldsymbol{s}^{true}))} = 0 \text{ a.s.} \quad \text{(D8)}$$

holds.

Proof (Theorem 3):
Under Condition C.1, there exists an argument minimizing the objective function $F_N(\boldsymbol{s})$ over $\mathcal{S}$. Then, we have

$$\begin{aligned} &\mathrm{L}\left(\boldsymbol{p}(\mathbf{Id}, \boldsymbol{s}_N^{\mathrm{est}}), \boldsymbol{f}_N(\mathbf{Id})\right) \leq \min_{\boldsymbol{s}\in\mathcal{S}} F_N(\boldsymbol{s}) \\ &\leq F_N(\boldsymbol{s}^{\mathrm{true}}) \to 0 \text{ as } N \to \infty \text{ a.s..} \end{aligned} \quad \text{(D9)}$$

Here we used the strong law of large numbers [37] and Conditions C.2 and C.3.

By using the triangle inequality of $\sqrt{\mathrm{L}}$, we have

$$\begin{aligned} &\sqrt{\mathrm{L}(\boldsymbol{p}(\mathbf{Id}, \boldsymbol{s}_N^{\mathrm{est}}), \boldsymbol{p}(\mathbf{Id}, \boldsymbol{s}^{\mathrm{true}}))} \\ &\leq \sqrt{\mathrm{L}(\boldsymbol{p}(\mathbf{Id}, \boldsymbol{s}_N^{\mathrm{est}}), \boldsymbol{f}_N(\mathbf{Id}))} \\ &\quad + \sqrt{\mathrm{L}(\boldsymbol{p}(\mathbf{Id}, \boldsymbol{s}^{\mathrm{true}}), \boldsymbol{f}_N(\mathbf{Id}))} \quad \text{(D10)} \\ &\to 0 \text{ as } N \to \infty \text{ a.s..} \quad \text{(D11)} \end{aligned}$$

□

Let us move on to the proof of Eq. (10). In the setting of SCQT, we can choose a continuous parametrization of probability distributions, and the continuous parameter space can be compact subset of a finite-dimensional Euclidean space. Hence, Condition C.1 is satisfied. Condition C.2, the positivity and boundedness of R, is satisfied in the RSC estimator. Condition C.3 is Eq. (9) itself. Therefore, Theorem 3 is applicable to the RSC estimator in the setting of SCQT, and Eq. (10) holds. □

4. Proof of Eqs. (12) and (13)

We prove Eqs. (12) and (13) in Theorem 2. First, we prove an equivalent statement under condition C.1, C.2, and C.4.

Theorem 4 *If C.1, C.2, and C.4 are satisfied, then*

$$\begin{aligned} &\sqrt{\mathrm{L}(\boldsymbol{p}(\mathbf{Id}, \boldsymbol{s}_N^{est}), \boldsymbol{p}(\mathbf{Id}, \boldsymbol{s}^{true}))} \\ &\lesssim \sqrt{\mathrm{L}(\boldsymbol{p}(\mathbf{Id}, \boldsymbol{s}^{true}), \boldsymbol{f}_N(\mathbf{Id}))} \text{ a.s.} \end{aligned} \quad \text{(D12)}$$

holds.

Proof (Theorem 4):
Under Conditions C.1 and C.2 with a property of minimizer of the estimator, we have

$$\begin{aligned} \mathrm{L}(\boldsymbol{p}(\mathbf{Id}, \boldsymbol{s}_N^{\mathrm{est}}), \boldsymbol{f}_N(\mathbf{Id})) &\leq \mathrm{L}(\boldsymbol{p}(\mathbf{Id}, \boldsymbol{s}^{\mathrm{true}}), \boldsymbol{f}_N(\mathbf{Id})) \\ &\quad + r_N R(\boldsymbol{s}^{\mathrm{true}}, \boldsymbol{s}^{\mathrm{target}}). \end{aligned} \quad \text{(D13)}$$

By combining Condition C.2 and C.4 with Eq. (D13), we obtain

$$\begin{aligned} &\mathrm{L}(\boldsymbol{p}(\mathbf{Id}, \boldsymbol{s}_N^{\mathrm{est}}), \boldsymbol{f}_N(\mathbf{Id})) \\ &\lesssim \left\{1 + R(\boldsymbol{s}^{\mathrm{true}}, \boldsymbol{s}^{\mathrm{target}})\right\} \mathrm{L}(\boldsymbol{p}(\mathbf{Id}, \boldsymbol{s}^{\mathrm{true}}), \boldsymbol{f}_N(\mathbf{Id})) \quad \text{(D14)} \\ &\lesssim \mathrm{L}(\boldsymbol{p}(\mathbf{Id}, \boldsymbol{s}^{\mathrm{true}}), \boldsymbol{f}_N(\mathbf{Id})) \text{ a.s..} \quad \text{(D15)} \end{aligned}$$

By using the triangle inequality of $\sqrt{\mathrm{L}}$ and Eq. (D15), we have

$$\sqrt{\mathrm{L}(\boldsymbol{p}(\mathbf{Id},\boldsymbol{s}_N^{\mathrm{est}}),\boldsymbol{p}(\mathbf{Id},\boldsymbol{s}^{\mathrm{true}}))} \lesssim \sqrt{\mathrm{L}(\boldsymbol{p}(\mathbf{Id},\boldsymbol{s}^{\mathrm{true}}),\boldsymbol{f}_N(\mathbf{Id}))}\ \text{a.s.}. \tag{D16}$$

□

Let us move on to the prove Eqs. (12) and (13) in Theorem 2. Conditions C.1 and C.2 are satisfied in the setting of SCQT as explained in the end of Appendix D 3. If we select a regularization parameter satisfying $r_N \lesssim 1/N$ a.s. (Eq. (11)), inequalities

$$r_N \lesssim \frac{1}{N} \lesssim \mathrm{L}\big(\boldsymbol{p}(\mathbf{Id},\boldsymbol{s}^{\mathrm{true}}),\boldsymbol{f}_N(\mathbf{Id})\big)\ \text{a.s.} \tag{D17}$$

hold because of the left inequality in Lemma 4, and Condition C.4 is satisfied. Therefore Theorem 4 is applicable to the RSC estimator in the setting of SCQT, and we obtain Eq. (12). Eq. (13) is given by combining the right inequality in Lemma 4 and Theorem 4. □

5. Proof of Eq. (14)

We prove that a sequence of estimates $\{\boldsymbol{s}_N^{\mathrm{est}}\}$ converges to the gauge-equivalence class $[\boldsymbol{s}^{\mathrm{true}}]$ almost surely at the limit of N going to infinity. To prove that, we modify the proof of Theorem 4.4 in [38] to make it applicable to the RSC estimator in the setting of SCQT, which is an ill-posed problem caused by the existence of the gauge degrees of freedom. We define

$$\mathrm{R}(\boldsymbol{s},[\boldsymbol{s}^{\mathrm{true}}]) := \min\left\{\ \mathrm{R}(\boldsymbol{s},\boldsymbol{s}')\ \middle|\ \boldsymbol{s}'\in[\boldsymbol{s}^{\mathrm{true}}]\ \right\} \tag{D18}$$

as a (squared) distance between $\boldsymbol{s}$ and the gauge-equivalence class $[\boldsymbol{s}^{\mathrm{true}}]$.

Theorem 5 *If Conditions C.1, C.2, C.3, and C.5 are satisfied, the sequence of RSC estimates $\{\boldsymbol{s}_N^{est}\}$ converges to $[\boldsymbol{s}^{true}]$ almost surely, i.e.,*

$$\lim_{N\to\infty}\mathrm{R}(\boldsymbol{s}_N^{est},[\boldsymbol{s}^{true}]) = 0\ \text{a.s.}. \tag{D19}$$

Proof (Theorem 5): First, we derive an inequality that any points in $\mathcal{S}$ outside the ϵ-neighborhood of $[\boldsymbol{s}^{\mathrm{true}}]$ satisfy. For a given $\epsilon>0$, we define

$$\eta_\epsilon := \min_{\boldsymbol{s}\in\mathcal{S}}\left\{\sqrt{\mathrm{L}(\boldsymbol{p}(\mathbf{Id},\boldsymbol{s}^{\mathrm{true}}),\boldsymbol{p}(\mathbf{Id},\boldsymbol{s})}\ ;\ \mathrm{R}(\boldsymbol{s},[\boldsymbol{s}^{\mathrm{true}}])\ge\epsilon\right\}. \tag{D20}$$

Since $\boldsymbol{p}^{\boldsymbol{i}}(\boldsymbol{s})$ are continuous functions over the compact set $\mathcal{S}$ (Condition C.1), the minimal value η_ϵ exists. From Condition C.5, $\eta_\epsilon>0$ holds.

The following arguments hold almost surely. From the strong law of large numbers [37], for any $\boldsymbol{i}\in\mathbf{Id}$

$$\lim_{N\to\infty}\boldsymbol{f}_N^{\boldsymbol{i}} = \boldsymbol{p}^{\boldsymbol{i}}(\boldsymbol{s}^{\mathrm{true}})\ \text{a.s.}. \tag{D21}$$

From Eq. (D21) and Condition C.3, for every $\epsilon>0$, there exists a constant $N(\epsilon)$ such that

$$N\ge N(\epsilon)\Rightarrow\left\{\begin{array}{l}\sqrt{\mathrm{L}(\boldsymbol{f}_N(\mathbf{Id}),\boldsymbol{p}(\mathbf{Id},\boldsymbol{s}^{\mathrm{true}}))}<\frac{\eta_\epsilon}{4}\\ \sqrt{r_N\mathrm{R}\left(\boldsymbol{s}^{\mathrm{true}},\boldsymbol{s}^{\mathrm{target}}\right)}<\frac{\eta_\epsilon}{4}\end{array}\right.. \tag{D22}$$

Then, for every $\boldsymbol{s}$ satisfying $\mathrm{R}(\boldsymbol{s},[\boldsymbol{s}^{\mathrm{true}}])\ge\epsilon$, we have

$$\begin{aligned}&\sqrt{\mathrm{L}\big(\boldsymbol{f}_N(\mathbf{Id}),\boldsymbol{p}(\mathbf{Id},\boldsymbol{s})\big)}\\ &\ge\sqrt{\mathrm{L}(\boldsymbol{p}(\mathbf{Id},\boldsymbol{s}),\boldsymbol{p}(\mathbf{Id},\boldsymbol{s}^{\mathrm{true}}))}\\ &\quad-\sqrt{\mathrm{L}(\boldsymbol{f}_N(\mathbf{Id}),\boldsymbol{p}(\mathbf{Id},\boldsymbol{s}^{\mathrm{true}}))} &\text{(D23)}\\ &\ge\frac{3}{4}\eta_\epsilon, &\text{(D24)}\end{aligned}$$

where we used the triangle inequality for the 2-norm ($\sqrt{\mathrm{L}}$). Therefore, an inequality

$$\begin{aligned}&\min_{\boldsymbol{s}\in\mathcal{S}}\left\{\,\mathrm{L}(\boldsymbol{p}(\mathbf{Id},\boldsymbol{s}),\boldsymbol{f}_N(\mathbf{Id}))\ \middle|\ \mathrm{R}(\boldsymbol{s},[\boldsymbol{s}^{\mathrm{true}}])\ge\epsilon\,\right\}\\ &>\frac{9}{16}\eta_\epsilon^2 &\text{(D25)}\end{aligned}$$

holds. Then we obtain

$$\begin{aligned}&\min_{\boldsymbol{s}\in\mathcal{S}}\left\{F_N(\boldsymbol{s})\ \middle|\ \mathrm{R}(\boldsymbol{s},[\boldsymbol{s}^{\mathrm{true}}])\ge\epsilon\right\} &\text{(D26)}\\ &\ge\min_{\boldsymbol{s}\in\mathcal{S}}\left\{\,\mathrm{L}\big(\boldsymbol{p}(\mathbf{Id},\boldsymbol{s}),\boldsymbol{f}_N(\mathbf{Id})\big)\ \middle|\ \mathrm{R}\left(\boldsymbol{s},[\boldsymbol{s}^{\mathrm{true}}]\right)\ge\epsilon\right\} &\text{(D27)}\\ &>\frac{9}{16}\eta_\epsilon^2. &\text{(D28)}\end{aligned}$$

Next, we show that $\boldsymbol{s}_N^{\mathrm{est}}$ does not satisfy Eq. (D28). Since $\boldsymbol{s}_N^{\mathrm{est}}$ is the argument minimizing $F_N(\boldsymbol{s})$ over $\boldsymbol{s}\in\mathcal{S}$, $F_N(\boldsymbol{s}_N^{\mathrm{est}})\le F_N(\boldsymbol{s})$ holds for any $\boldsymbol{s}\in\mathcal{S}$. Then, from Eq. (D22), we have

$$F_N(\boldsymbol{s}_N^{\mathrm{est}})\le F_N(\boldsymbol{s}^{\mathrm{true}})<\frac{2}{16}\eta_\epsilon^2<\frac{9}{16}\eta_\epsilon^2. \tag{D29}$$

Hence, $\boldsymbol{s}_N^{\mathrm{est}}$ does not satisfy Eq. (D28), and it means that $\boldsymbol{s}_N^{\mathrm{est}}$ is in the ϵ-neighborhood of $[\boldsymbol{s}^{\mathrm{true}}]$. Thus, we obtain

$$N\ge N(\epsilon)\Rightarrow\mathrm{R}(\boldsymbol{s}_N^{\mathrm{est}},[\boldsymbol{s}^{\mathrm{true}}])<\epsilon. \tag{D30}$$

Since ϵ is an arbitrary positive number, we obtain

$$\lim_{N\to\infty}\mathrm{R}(\boldsymbol{s}_N^{\mathrm{est}},[\boldsymbol{s}^{\mathrm{true}}])=0\ \text{a.s.}. \tag{D31}$$

□

Let us move on to the proof of Eq. (14). Condition C.1 and C.2 are satisfied in the setting of SCQT as explained in the ends of Appendix D 3 and D 4. Condition C.3 is Eq. (9) itself. When **Id** is SCIC, Condition C.5 is satisfied (Theorem 1). Therefore Theorem 5 is applicable to the RSC estimator in the setting of SCQT, and it leads to Eq. (14) in Theorem 2. □

Appendix E: Proofs on Dynamics Generator Analysis

In this section, we prove the existence of the matrix logarithm of the HS representation of a gate and

Eqs. (23), (24), and (25) in Sec. III C, if its dynamics obeys the time-dependent GKLS equation under conditions of finite energy and finite time period. In Sec. E 3, we explain a relation between the dynamics generator analysis proposed and similar known results.

1. Proof of the Existence of the Matrix Logarithm

We give the proof of the existence of the matrix logarithm, which is assumed in the dynamics generator analysis proposed in Sec. III C. In the vectorized state representation, the time-dependent version of the GKLS equation is rewritten as

$$\frac{d}{dt}|\rho(t)\rangle\!\rangle = \mathrm{HS}(\mathcal{L}_t)|\rho(t)\rangle\!\rangle, \tag{E1}$$

$$\Leftrightarrow \frac{d}{dt}\mathrm{HS}(\mathcal{G}_t)|\rho(0)\rangle\!\rangle = \mathrm{HS}(\mathcal{L}_t)\mathrm{HS}(\mathcal{G}_t)|\rho(0)\rangle\!\rangle, \tag{E2}$$

where we used $|\rho(t)\rangle\!\rangle = \mathrm{HS}(\mathcal{G}_t)|\rho(0)\rangle\!\rangle$, and $\mathcal{G}_t$ is defined as a map corresponding to the gate implemented with the dynamics during the time period $[0,t]$, and $\mathcal{G}_T$ corresponds to $\mathcal{G}$ in Sec. III C. Eq. (E2) holds for arbitrary $\rho(0)$, and it implies

$$\frac{d}{dt}\mathrm{HS}(\mathcal{G}_t) = \mathrm{HS}(\mathcal{L}_t)\mathrm{HS}(\mathcal{G}_t). \tag{E3}$$

Therefore $\mathrm{HS}(\mathcal{G}_t)$ is a solution of the homogeneous first-order linear differential equation. The general theory of differential equations guarantees the unique existence of the solution, and the following equality holds (Problems 4a in Sec. 6.5, pp.507-508 in [36]),

$$\det \mathrm{HS}(\mathcal{G}_t) = \exp\left[\int_0^t dt' \,\mathrm{Tr}\left\{\mathrm{HS}(\mathcal{L}_{t'})\right\}\right]. \tag{E4}$$

When $H(t')$, $J(t')$, and $K(t')$ in Eq. (15) are bounded for any $t' \in [0,t]$ with finite t, $\int_0^t dt' \mathrm{Tr}\left\{\mathrm{HS}(\mathcal{L}_{t'})\right\} > -\infty$ and $\det \mathrm{HS}(\mathcal{G}_t) > 0$ hold. This implies that $\mathrm{HS}(\mathcal{G}_t)$ is invertible. Every invertible matrix can be written as the exponential of a complex matrix (Exercises 2.9 and 2.10 in [48]). Then, for every $\mathrm{HS}(\mathcal{G}_t)$, there exists a matrix $X(t)$ that satisfies

$$\mathrm{HS}(\mathcal{G}_t) = \exp\left[X(t)\right]. \tag{E5}$$

□

Note that the trace in the R.H.S. of Eq. (E4) can be rewritten as

$$\mathrm{Tr}\left\{\mathrm{HS}(\mathcal{L}_t)\right\} = \mathrm{Tr}\left\{\mathrm{HS}^{\mathrm{cb}}(\mathcal{L}_t)\right\} = 2dJ_0(t), \tag{E6}$$

where the superscript cb stands for the computational basis, and this means that the Hamiltonian part does not affect on the invertibility. When the dynamics is trace-preserving, J and K are related as

$$J(t) = -\frac{1}{2}\sum_{\alpha,\beta=1}^{d^2-1} K_{\alpha\beta}(t) B_\beta^\dagger B_\alpha, \tag{E7}$$

and

$$J_0(t) = \mathrm{Tr}\left\{B_0^\dagger J(t)\right\} = -\frac{1}{2\sqrt{d}}\mathrm{Tr}\left\{K(t)\right\}. \tag{E8}$$

Therefore $K(t)$ affects on the invertibility through $J_0(t)$. When $K(t)$ is positive semidefinite, the dynamics becomes completely positive, and $\mathrm{Tr}\left\{K(t)\right\} \geq 0$ and $J_0(t) \leq 0$ hold. Then $\mathrm{Tr}\left\{\mathrm{HS}(\mathcal{L}_t)\right\} \geq 0 > -\infty$ holds and the inverse exists.

2. Proof of Eqs. (23), (24), and (25)

For simplicity of notation, we omit the superscript, acc, below. Eq. (20) can be rewritten as

$$\begin{aligned}\mathcal{L}(\rho) = &-i\sum_{\alpha=1}^{d^2-1} H_\alpha\left(B_\alpha\rho - \rho B_\alpha\right) + \sum_{\alpha=0}^{d^2-1} J_\alpha\left(B_\alpha\rho + \rho B_\alpha\right) \\ &+ \sum_{\alpha,\beta=1}^{d^2-1} K_{\alpha\beta} B_\alpha \rho B_\beta^\dagger.\end{aligned} \tag{E9}$$

In the matrix vectorization, or the HS representation w.r.t. the computational basis in the row major order, $|X\rangle\!\rangle := \sum_{i,j} X_{ij}|i\rangle|j\rangle$, an equality, $|ABC\rangle\!\rangle = A\otimes C^T|B\rangle\!\rangle$, holds. Then

$$\begin{aligned}\mathrm{HS}^{\mathrm{cb}}(\mathcal{L})|\rho\rangle\!\rangle = \Bigg\{ &-i\sum_{\alpha=1}^{d^2-1} H_\alpha\left(B_\alpha\otimes I - I\otimes\overline{B_\alpha}\right) \\ &+ \sum_{\alpha=0}^{d^2-1} J_\alpha\left(B_\alpha\otimes I + I\otimes\overline{B_\alpha}\right) \\ &+ \sum_{\alpha,\beta=1}^{d^2-1} K_{\alpha\beta} B_\alpha\otimes\overline{B_\beta}\Bigg\}|\rho\rangle\!\rangle\end{aligned} \tag{E10}$$

hold for any ρ, and we have

$$\begin{aligned}\mathrm{L}^{\mathrm{cb}} := \mathrm{HS}^{\mathrm{cb}}(\mathcal{L}) = &-i\sum_{\alpha=1}^{d^2-1} H_\alpha\left(B_\alpha\otimes I - I\otimes\overline{B_\alpha}\right) \\ &+ \sum_{\alpha=0}^{d^2-1} J_\alpha\left(B_\alpha\otimes I + I\otimes\overline{B_\alpha}\right) \\ &+ \sum_{\alpha,\beta=1}^{d^2-1} K_{\alpha\beta} B_\alpha\otimes\overline{B_\beta}.\end{aligned} \tag{E11}$$

By combining Eq. (E11) with the orthonormality and Hermiticity of $\boldsymbol{B}$, we obtain Eqs. (23), (24), and (25).
□

3. Related Work on Dynamics Generator Analysis

Here we discuss relation of the dynamics generator analysis proposed in Sec. III C to known methods.

In recent experiments on superconducting quantum circuits [49, 50], experimentalists try to estimate the accumulated Hamiltonian $H^{\rm acc}$ in which experiments and data-processing procedures are different from the methods proposed here. They report that calibration methods for gates using the estimated information worked well. The data-processing procedures depend on specific models of target accumulated Hamiltonians and do not take into account the effects of decoherence during the gate operations. On the other hand, the method we propose is very general, there are no assumptions on the model of the accumulated generators, and in the data-processing both effects of Hamiltonian and decoherence are taken into account. Therefore, our method can give us more accurate information of the accumulated generators, which would be useful for calibration.

An error generator, defined as $\ln\left\{(\mathrm{G}^{\rm target})^{-1}\mathrm{G}\right\}$, is estimated with results of gate-set tomography in [31]. When a target gate is unitary, any gate can be decomposed into the form of $\mathcal{G} = \mathcal{G}^{\rm target} \circ \mathcal{E}$, where $\mathcal{E} := (\mathcal{G}^{\rm target})^{-1} \circ \mathcal{G}$. This leads to

$$\mathrm{E} := \mathrm{HS}(\mathcal{E}) = (\mathrm{G}^{\rm target})^{-1}\mathrm{G}. \tag{E12}$$

The error generator, $\ln \mathrm{E}$, can be considered as a representation of errors on the accumulated generators. However, in general $\ln \mathrm{E}$ and $\ln \mathrm{G}^{\rm target} =: (\mathrm{L}^{\rm acc})^{\rm target}$ are not commutable, and $\mathrm{L}^{\rm acc} \neq (\mathrm{L}^{\rm acc})^{\rm target} + \ln \mathrm{E}$ because

$$\begin{aligned} \mathrm{G} = \exp(\mathrm{L}^{\rm acc}) &= \exp((\mathrm{L}^{\rm acc})^{\rm target}) \exp(\ln \mathrm{E}) \\ &\neq \exp((\mathrm{L}^{\rm acc})^{\rm target} + \ln \mathrm{E}) \end{aligned} \tag{E13}$$

Therefore the error generator $\ln E$ does not represent the direct discrepancy of the accumulated generators. On the other hand, if we define $\Delta\mathrm{L} := \ln \mathrm{G} - \ln \mathrm{G}^{\rm target} = \mathrm{L}^{\rm acc} - (\mathrm{L}^{\rm acc})^{\rm target}$, $\mathrm{L}^{\rm acc} = (\mathrm{L}^{\rm acc})^{\rm target} + \Delta\mathrm{L}$ holds by definition. We consider $\Delta\mathrm{L}$ or $\{H_\alpha^{\rm acc} - (H_\alpha^{\rm acc})^{\rm target}\}_{\alpha=1}^{d^2-1}$ more suitable for the use in a calibration process. In theory of quantum information, especially in quantum error correction, an error model on a quantum gate is typically introduced as $\mathcal{G} = \mathcal{E}' \circ \mathcal{G}^{\rm target}$. Note that the timing of the error's action is different from $\mathcal{E}$. If the purpose of analysis is to know information of $\mathcal{E}'$, for comparison to numerical simulation of a quantum error correction code, for example, $\mathrm{E}' := \mathrm{G}(\mathrm{G}^{\rm target})^{-1}$ would be an appropriate quantity to analyze.

Appendix F: Regularization

In this section, we briefly explain conventional purpose of using a regularization in applied mathematics. After that, we describe differences between regularization in RSCQT and that in conventional settings, from the viewpoints of motivation and three mathematical properties.

Regularization is an attractive way to stably solve inverse problems [51, 52], i.e., the solution is not much changed when the observed data are slightly fluctuated. Much attention has been paid to the regularization in many other fields of mathematical sciences including integral equation [53], signal processing [54], statistics and machine learning [55–57]. It is also used in variants of standard QT [58–60]. A purpose to use regularization depends on each research field. For example, it is used to avoid over-fitting to observed data in machine learning, to make a solution stable or to solve ill-posed problems in inverse problems, and among others. Moreover, in order to improve interpretability of models in regression problems by increasing sparsity of estimates or to obtain a smooth function in non-parametric estimation, the regularization also plays an important role in statistics. From the viewpoint of Bayesian analysis, a regularization can be regarded as exploiting prior information with respect to model parameters under some conditions.

Our purpose of introducing a regularization into the setting of SCQT is to fix the gauge degrees of freedom, suitable for the improvement and validation steps. This is originated from the role of quantum characterization in quantum information processing and our choice of SCQT approach. This is new and quite different from the conventional purposes of regularization mentioned above, although the use of $\boldsymbol{s}^{\rm target}$ in the regularization function can be regarded as a use of prior information of the target set of quantum operations that we aim to implement. Additionally, a mathematical framework of RSCQT has at least the following three attributes: (i) non-uniqueness of the true solution of the original (unregularized) problem, (ii) constraint parameter space, (iii) non-linear parameterization, of which difficulties make our problem more complicated than the previous studies. Hereafter, we briefly explain these three attributes.

First, there exist the gauge degrees of freedom, which is originated from the self-consistent approach and Born's rule, a fundamental principle of quantum theory. We cannot determine parameters of interests only from experimental data, and such estimation problem can be categorized into an ill-posed problem in the inverse problem. A conventional approach to the ill-posed problem in applied mathematics is to neglect or remove such unaccessible degrees of freedom. On the other hand, we cannot neglect or remove the gauge degrees of freedom because at the validation and improvement steps after characterization each mathematical representation of state, measurement, and gate is necessary. In order to separate the representations from each other, we need to fix the gauge somehow.

Second, the region of possible parameters is constrained, which is originated from the requirement of physicality on estimates of quantum operations. When an accuracy of quantum operations is high, the true set lies close to the boundary of the physical region. If we require physicality on estimates, we have to take the boundary into account at the data-processing. In standard QT, the boundary affects on the performance of estimators [42] for finite data. In RSCQT, the dimension of the parameter space is much larger than that of standard QT, and the analysis of the boundary effect becomes much

harder.

Third, the parametrization of probability distributions is non-linear, originated from the SCQT approach. A non-linear function in a loss function is often analyzed in the inverse problems [61]. The asymptotic convergence for non-linear Tikhonov regularization was derived in [62] under an assumption that the true solution of the original (unregularized) problem is unique, and [63] showed its convergence rate under a similar assumption. Since their proofs are shown by exploiting the uniqueness of the original solution, it is non-trivial to extend their results to our framework with the gauge degrees of freedom, in which the original solution is not unique.

A previous study, which takes three attributes, (i), (ii) and (iii), exists [64]. However, the study only considers a case that the regularization parameter, r_N, is fixed, and does not show the asymptotic convergence to equivalence class of the true parameter nor the convergence rate of estimator. As a matter of fact, when we fix the regularization parameter, a bias caused by a regularization remains even at the limit of data size going to infinity, and the asymptotic convergence does not hold. On the other hand, we proved that the RSC estimator has the asymptotic convergence if we select $r_N \lesssim 1/N$, and we derived its convergence rate.

Appendix G: Cross Validation

Cross validation is a standard method for selecting a regularization parameter in statistics and machine learning [39, 40]. In the numerical experiments reported in Sec. IV, we combined the RSC estimator with k-fold cross validation ($k = 3$). Roughly speaking, the k-fold cross validation selects a regularization parameter from the perspective of prediction. If we calculate both of the goodness of fit and the estimate from common data, an over-fitting to the data occurs, and the performance of predicting the true probability distributions or the goodness of fit to different data can become worse. The over-fitting problem is caused by the statistical dependence of the data for calculating the estimate and goodness of fit. In order to avoid the problem, the cross validation divides the data into two parts. One is for calculating an estimate, which is called *learning data*. The other is for calculating the goodness of fit, which is called *test data*. This division makes learning data and test data statistically independent. In order to reduce an effect of the way of division, divisions are differently performed k times. A goodness of a regularization parameter is evaluated by an average value of the goodness of fit over k divisions. We explain the details of the procedure of k-fold cross validation below.

Suppose that we performed experiments with a SCIC set of experimental schedules **Id** and obtained experimental data D_N with an amount of data N. For a given coefficient c, a regularization parameter in the RSC estimator is calculated from the coefficient and amount of data as $r_N = c/N$. Let $\boldsymbol{c} = \{c_1, \ldots, c_{n_r}\}$ denote a set of candidates of regularization parameter coefficients. Let k denote a positive integer larger than or equal to 2. The k-fold cross validation selects a value from $\boldsymbol{c}$ for the RSC estimator along with the following procedure.

Step 1. Data Division
We randomly divide the data into k distinct parts as equally as possible. Let $D_{N,1}, \ldots, D_{N,k}$ denote the k parts of D_N ($D_N = \cup_{j=1}^{k} D_{N,j}$). We introduce a notation for complement sets $\overline{D_{N,j}} := D_N \backslash D_{N,j}$, $j = 1, \ldots, k$. At the j-th division $D_N = D_{N,j} \cup \overline{D_{N,j}}$, $D_{N,j}$ is the test data and $\overline{D_{N,j}}$ is the learning data. Let N_j and $\overline{N_j}$ denote the amounts of data for $D_{N,j}$ and $\overline{D_{N,j}}$, respectively. Roughly speaking, $N_j \approx N/k$ and $\overline{N_j} \approx N - N/k$ hold.

Step 2. Calculation of Empirical Distributions
We calculate empirical distributions from each $D_{N,j}$ and $\overline{D_{N,j}}$. Let $\boldsymbol{f}_{N_j}(\mathbf{Id})$ and $\boldsymbol{f}_{\overline{N_j}}(\mathbf{Id})$ denote the set of empirical distributions calculated from $D_{N,j}$ and $\overline{D_{N,j}}$, respectively. For simplicity of notation, we omit **Id** from the notation of the set of empirical distributions below in this section.

Step 3. Calculation of Cross Validation Losses
First, we calculate multiple RSC estimates from complement data and coefficient candidates. Next, we calculate values of loss functions for cross validation. Let ℓ denote an index for the candidates of regularization parameter. For $\ell = 1, \ldots, n_r$, we repeat the following procedures:

3.1 Calculation of Estimates
We calculate each RSC estimate from complement empirical distribution $\boldsymbol{f}_{\overline{N_j}}$ and a regularization parameter $c_\ell/\overline{N_j}$ for $j = 1, \ldots, k$ along with Eq. (8). Let $\boldsymbol{s}^{\mathrm{est}}_{\overline{N_j}}(c_\ell)$ denote the estimates, in which their dependency on c_ℓ is explicitly shown in the notation for clarifying the dependence.

3.2 Calculation of Cross Validation Losses
We calculate values of the loss function in Eq. (6) between the probability distributions predicted by the estimates calculated in the previous sub-step from the leaning data and the empirical distributions calculated from the test data for $j = 1, \ldots, k$. We calculate the arithmetic mean, which is the definition of the cross validation (cv) loss for a candidate c_ℓ. Let $\mathrm{L}^{\mathrm{cv}}(c_\ell)$ denote the cv loss of c_ℓ. An explicit mathematical form of the cv loss is as follows:

$$\mathrm{L}^{\mathrm{cv}}(c_\ell) := \frac{1}{k} \sum_{j=1}^{k} \mathrm{L}\left(\boldsymbol{p}(\mathbf{Id}, \boldsymbol{s}^{\mathrm{est}}_{\overline{N_j}}(c_\ell)), \boldsymbol{f}_{N_j}\right). \quad \text{(G1)}$$

At the end of Step 3, we have a set of values of cv loss, $\{\mathrm{L}^{\mathrm{cv}}(c_\ell)\}_{\ell=1}^{n_{\mathrm{r}}}$.

Step 4. Selection of Regularization Parameter From $\boldsymbol{c}$, we choose the coefficient candidate c_ℓ that has the minimal value of the cv loss. Let c^{cv} denote the selected coefficient. It is defined as

$$c^{\mathrm{cv}} := \underset{c\in\boldsymbol{c}}{\mathrm{argmin}}\, \mathrm{L}^{\mathrm{cv}}(c). \tag{G2}$$

In the procedure of k-fold cross validation explained above, we need to perform the optimization for calculating an RSC estimate kn_{r} times. After the procedure, we obtain the selected coefficient c^{cv}. Finally, we calculate the RSC estimate $\boldsymbol{s}_N^{\mathrm{est}}(c^{\mathrm{cv}})$ with the total data D_N and the selected regularization parameter $r_N^{\mathrm{cv}} := c^{\mathrm{cv}}/N$. The estimate $\boldsymbol{s}_N^{\mathrm{est}}(c^{\mathrm{cv}})$ is the result of the RSC estimator with k-fold cross validation. In total, we need to perform the optimization $(kn_{\mathrm{r}}+1)$ times for the combination. The kn_{r} times optimizations are additional costs for using k-fold cross validation.

Appendix H: Numerical Experiments

We describe details of numerical experiments for 1-qubit system explained in Sec. IV.

1. Setting

Three quantum gates are implemented with a Hamiltonian model [65],

$$H(t) = -\frac{\Delta\omega}{2}\sigma_3 + \frac{f(t)}{2}\left\{\cos(\phi)\sigma_1 + \sin(\phi)\sigma_2\right\}, \tag{H1}$$

where $\Delta\omega$ is the frequency detuning, $f(t)$ is the pulse shape, and ϕ is the relative phase. For simplicity, we choose a rectangular pulse,

$$f(t) = \begin{cases} A & (0 \le t \le W) \\ 0 & \text{otherwise} \end{cases}. \tag{H2}$$

Each target gate in Eqs. (28), (29), and (30) corresponds to the combination of $\Delta\omega = 0$, $A\cdot W = 0$ ($\mathcal{G}_0^{\mathrm{target}}$), $\pi/2$ ($\mathcal{G}_1^{\mathrm{target}}$, $\mathcal{G}_2^{\mathrm{target}}$), and $\phi = 0$ ($\mathcal{G}_0^{\mathrm{target}}$, $\mathcal{G}_1^{\mathrm{target}}$), $\pi/2$ ($\mathcal{G}_2^{\mathrm{target}}$), respectively. In the numerical experiments, true gates were chosen so that they include the following coherent errors and decoherence. We choose the gate time as 15 ns and $W = 10$ ns with coherent errors shown in Table II. Decoherence is modeled by the following three dissipation operators [41] in the GKLS master equation.

$$\sqrt{\Gamma_+}|1\rangle\langle 0|,\ \sqrt{\Gamma_-}|0\rangle\langle 1|,\ \sqrt{\Gamma_\phi}\frac{\sigma_3}{\sqrt{2}}. \tag{H3}$$

Gate	$\Delta\omega$	$A\cdot W$	ϕ	AGIF
$\mathcal{G}_0^{\mathrm{true}}$	0.01	0	0	2.0×10^{-3}
$\mathcal{G}_1^{\mathrm{true}}$	0.01	$\pi/2+0.1$	0.1	5.5×10^{-3}
$\mathcal{G}_2^{\mathrm{true}}$	0.01	$\pi/2+0.1$	$\pi/2+0.1$	5.5×10^{-3}

TABLE II. Coherent error parameters and average gate infidelity (AGIF) for $\mathcal{G}^{\mathrm{true}}$ in the numerical experiments.

Relations between the dissipation ratios, Γ_+, Γ_-, Γ_ϕ and coherence times T_1, T_2, T_ϕ, and the thermal population p_{th} are given as

$$\Gamma_+ + \Gamma_- = \frac{1}{T_1}, \tag{H4}$$

$$\frac{1}{2}\Gamma_+ + \frac{1}{2}\Gamma_- + \Gamma_\phi = \frac{1}{T_2}, \tag{H5}$$

$$\Gamma_\phi = \frac{1}{T_\phi}, \tag{H6}$$

$$\frac{\Gamma_+ - \Gamma_-}{\Gamma_+ + \Gamma_-} = p_{\mathrm{th}}. \tag{H7}$$

In the numerical experiments, we choose $T_1 = 30$ μs, $T_2 = 20$ μs, and $p_{\mathrm{th}} = 0.01$. Values of the average gate infidelity for each gate, which include both of coherent errors and decoherence, are shown in Table II. They are shown only the first two digits and in order of 10^{-3}. The depolarizing error rates for state and POVM are 0.015 and 0.010, respectively.

The schedule of the experiments consists of sub-experiments. Every sub-experiments start with the state initialization ρ^{true} and end with the measurement $\mathbf{\Pi}^{\mathrm{true}}$. Gate sequences between ρ^{true} and $\mathbf{\Pi}^{\mathrm{true}}$ are shown in Table III. The set of sub-experiments satisfies the SCIC condition. We choose common number of repetitions, N, for each sub-experiment.

ID	Gate Sequence	ID	Gate Sequence	ID	Gate Sequence
1	G0 · G0	16	G0 · G2 · G0	31	G2 · G1 · G0
2	G0 · G1	17	G0 · G2 · G1	32	G2 · G1 · G1
3	G0 · G2	18	G0 · G2 · G2	33	G2 · G1 · G2
4	G1 · G0	19	G1 · G0 · G0	34	G2 · G2 · G0
5	G1 · G1	20	G1 · G0 · G1	35	G2 · G2 · G1
6	G1 · G2	21	G1 · G0 · G2	36	G2 · G2 · G2
7	G2 · G0	22	G1 · G1 · G0	37	G1 · G1 · G0 · G0
8	G2 · G1	23	G1 · G1 · G1	38	G1 · G1 · G0 · G1
9	G2 · G2	24	G1 · G1 · G2	39	G1 · G1 · G0 · G2
10	G0 · G0 · G0	25	G1 · G2 · G0	40	G1 · G1 · G1 · G0
11	G0 · G0 · G1	26	G1 · G2 · G1	41	G1 · G1 · G1 · G1
12	G0 · G0 · G2	27	G1 · G2 · G2	42	G1 · G1 · G1 · G2
13	G0 · G1 · G0	28	G2 · G0 · G0	43	G1 · G1 · G2 · G0
14	G0 · G1 · G1	29	G2 · G0 · G1	44	G1 · G1 · G2 · G1
15	G0 · G1 · G2	30	G2 · G0 · G2	45	G1 · G1 · G2 · G2

TABLE III. List of gate sequences used in the numerical experiments. The operation order is from left to right. G0, G1, G2 correspond to $\mathcal{G}_0^{\mathrm{true}}$, $\mathcal{G}_1^{\mathrm{true}}$, $\mathcal{G}_2^{\mathrm{true}}$, respectively.

2. Optimization Solver and Physicality Constraints

We numerically implemented the RSC estimator for 1-qubit systems with IPOPT [66]. IPOPT is implemented by C++ and provides interfaces to convert the objective function and constraints into a standard form of the solver in several programming languages. We used C++ to shorten the computation time. The information of the first and second derivatives of the objective function and constraints are optionally acceptable with an interface of IPOPT. Such optional information is helpful for making the computation time even shorter. We provided them to the interface with optional parameters. At the interface, we can specify our degree of tolerance of acceptable violation of the constraints, δ. The tolerance parameter of $\delta = 0$ means that we do not accept any violation of the constraints, and ideally $\delta = 0$ would be desired. However, we chose $\delta = 10^{-4}$ for the numerical simulations reported in this manuscript, because computational time for the optimization becomes longer as we set smaller δ. We observed unphysical estimates sometimes in the simulations and confirmed that all violations are controlled to be below δ.

We explain our numerical treatment of the physicality constraints on quantum operations. We chose the parametrization of quantum operations by real numbers explained in Appendix A. As explained there (Eqs. (A1), (A3), (A5), (A7)), the physicality constraints on quantum operations are categorized into two types, equality constraints and inequality constraints. The equality constraints have been taken into account by the parametrization itself automatically. All of the inequality constraints are represented in the form of the positive-semidefiniteness of an Hermitian matrix such as $\rho \succeq 0$, $\Pi_x \succeq 0$, and $\mathrm{CJ}(\mathcal{G}_j) \succeq 0$. The positive-semidefiniteness of an Hermitian matrix is rewritten as a set of polynomial inequalities [67, 68]. We provided the information of the polynomial inequalities for quantum operations with their first and second derivatives to the interface of IPOPT. The parametrization of a gate is based on the HS matrix, and we derived and used the following equality to represent the inequality constraint on the gate w.r.t. the HS matrix,

$$\mathrm{CJ}(\mathcal{G}) = \sum_{\alpha,\beta=0}^{d^2-1} \mathrm{HS}(\mathcal{G})_{\alpha\beta} B_\alpha \otimes \overline{B_\beta}, \tag{H8}$$

where $\boldsymbol{B} = \{B_\alpha\}_{\alpha=0}^{d^2-1}$ is the matrix basis introduced in Appendix. A, and $\overline{B_\beta}$ is the complex conjugate of the matrix B_β.

Proof (Eq. (H8)): Matrix elements of the HS matrix of a linear map $\mathcal{G}$ with respect to the basis $\boldsymbol{B}$ is given as

$$\mathrm{HS}(\mathcal{G})_{\alpha\beta} = \mathrm{Tr}\left[B_\alpha^\dagger\ \mathcal{G}(B_\beta)\right], \tag{H9}$$

for $\alpha, \beta = 0, \ldots, d^2-1$. The action of the map is represented with the CJ matrix as

$$\mathcal{G}(\rho) = \mathrm{Tr}_2\left[\left(I_1 \otimes \rho^T\right)\mathrm{CJ}(\mathcal{G})\right]. \tag{H10}$$

Then

$$\mathrm{HS}(\mathcal{G})_{\alpha\beta} = \mathrm{Tr}_{1,2}\left[\left(B_\alpha \otimes \overline{B_\beta}\right)^\dagger \mathrm{CJ}(\mathcal{G})\right]. \tag{H11}$$

□

Note that the proof holds for any orthonormal matrix basis $\boldsymbol{B}$, which is not necessarily Hermitian or $B_0 = I/\sqrt{d}$. Therefore Eq. (H8) holds not only for the (generalized) Pauli basis, but also for the other orthonormal basis including the computational basis.

[1] R. Barends et al., Nature **508**, 500 (2014).
[2] J. Emerson et al., J. Opt. B: Quantum Semiclass. Opt. **7**, S347 (2005).
[3] J. Emerson et al., Science **317**, 1893 (2007).
[4] E. Knill et al., Phys. Rev. A **77**, 012307 (2008).
[5] E. Magesan, J. M. Gambetta, and J. Emerson, Phys. Rev. Lett. **106**, 180504 (2011).
[6] E. Magesan, J. M. Gambetta, and J. Emerson, Phys. Rev. A **85**, 042311 (2012).
[7] E. Magesan et al., Phys. Rev. Lett. **109**, 080505 (2012).
[8] J. Gambetta et al., Phys. Rev. Lett. **109**, 240504 (2012).
[9] T. Chasseur and F. K. Wilhelm, Phys. Rev. A **92**, 042333 (2015).
[10] J. Wallman et al., New J. Phys. **18**, 043021 (2016).
[11] J. Wallman et al., New J. Phys. **17**, 113020 (2015).
[12] S. Sheldon, L. S. Bishop, E. Magesan, S. Filipp, J. M. Chow, and J. M. Gambetta, Phys. Rev. A **93** 012301 (2016).
[13] J. J. Wallman, M. Barnhill, J. Emerson, Phys. Rev. Lett. **115**, 060501 (2015).
[14] A. W. Cross et al., npj Quantum Inf. **2**, 16012 (2016).
[15] S. Kimmel, M. P. daSilva, C. A. Ryan, B. R. Johnson, T. Ohki, Phys. Rev. X **4**, 011050 (2014).
[16] J. M. Epstein, A. W. Cross, E. Magesan, J. M. Gambetta, Phys. Rev. A **89**, 062321 (2014).
[17] T. Proctor, K. Rudinger, K. Young, M. Sarovar, R. Blume-Kohout, Phys. Rev. Lett. **119**, 130502 (2017).
[18] J. Qi and H. K. Ng, Int. J. Quantum Information **17**, 1950031 (2019).
[19] U. Fano, Rev. Mod. Phys. **29**, 74 (1957).
[20] D. T. Smithey, M. Beck, M. G. Raymer, A. Faridani, Phys. Rev. Lett. **70**, 1244 (1993).
[21] Z. Hradil, Phys. Rev. A **55**, R1561 (1997).
[22] K. Banaszek, G. M. DAriano, M. G. A. Paris, M. F. Sacchi, Phys. Rev. A **61**, 010304(R) (1999).
[23] J. F. Poyatos, J. I. Cirac, P. Zoller, Phys. Rev. Lett. **78**, 390 (1997).
[24] I. L. Chuang and M. A. Nielsen, J. Mod. Phys. **44**, 2455 (1997).
[25] A. Luis and L. L. Sanchez-Soto, Phys. Rev. Lett. **83**, 3573 (1999).
[26] M. Paris and J. Rehacek (Eds.), *Quantum State Estimation*, Springer Berlin Heidelberg (2004).
[27] S. T. Merkel, J. M. Gambetta, J. A. Smolin, S. Poletto,

A. D. Corcoles, B. R. Johnson, C. A. Ryan, M. Steffen, Phys. Rev. A **87**, 062119 (2013).
[28] C. Stark, Phys. Rev. A **89**, 052109 (2014).
[29] R. Blume-Kohout et al., arXiv:1310.4492 [quant-ph].
[30] pyGSTi, `http://www.pygsti.info/`, DOI: 10.5281/zenodo.1209246.
[31] R. Blume-Kohout et al., Nature Commun. **8**, 14485 (2017).
[32] "Note that currently this only ensures that the gates are CPTP - not the SPAM operations.", quoted from the FAQ page of pyGSTi (Version 0.9.9.2 on 3 June, 2020). `https://github.com/pyGSTio/pyGSTi/blob/master/jupyter\_notebooks/FAQ.ipynb`
[33] H.-P. Breuer and F. Petruccione, *The Theory of Open Quantum Systems*, Oxford University Press, Oxford (2002).
[34] V. Gorini, A. Kossakowski, and E. C. G. Sudarshan, J. Math. Phys. 17, 821 (1976).
[35] G. Lindblad, Commun. Math. Phys. 48, 119 (1976).
[36] R. A. Horn and C. R. Johnson, "Topics In Matrix Analysis", Cambridge University Press (1991).
[37] A. Klenke, "Probability Theory: A Comprehensive Course", Springer-Verlag, London (2008).
[38] N. Yoshida, *Suritokeigaku* (in Japanese), Asakura, Tokyo (2006).
[39] M. Stone, J. R. Stat. Soc. Ser. B **36**, 111 (1974).
[40] S. Arlot and A. Celisse, Statistics Surveys **4**, 40 (2010).
[41] H. J. Briegel and B. G. Englert, "Quantum optical master equations: The use of damping bases", Phys. Rev. A 47, 3311 (1993).
[42] T. Sugiyama, P. S. Turner, and M. Murao, New J. Phys. **14**, 085005 (2012).
[43] T. Sugiyama, S. Imori, and F. Tanaka, arXiv:1806.02696 [quant-ph, v3].
[44] R. A. Horn and C. R. Johnson, *Matrix Analysis*, 2nd eds., Cambridge University Press, New York (2013).
[45] M. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*, Cambridge University Press, Cambridge (2000).
[46] A. Y. Kitaev, Russian Mathematical Surveys, **52**, 1191 (1997).
[47] C. J. Wood, J. D. Biamonte, and D. G. Cory, Quant. Inf. Comp. **15**, 0759-0811 (2015).
[48] B. Hall, "Lie groups, Lie algebras, and representations : an elementary introduction", 2nd Eds., Springer (2015).
[49] S. Sheldon, E. Magesan, J. M. Chow, J. M. Gambetta, Phys. Rev. A **93**, 060302(R) (2016).
[50] A. D. Patterson et al., arXiv:1905.05670[quant-ph].
[51] A. N. Tikhonov, Acad. Sci. SSSR **39**, 176 (1943).
[52] A. N. Tihonov, Soviet Math. **4**, 1035 (1963).
[53] D. L. Phillips, J. ACM **9**, 84 (1962).
[54] M. Foster, J. Soc. Indust. Appl. Math. **9**, 387 (1961).
[55] A. E. Hoerl, Chemical Engineering Progress **58** (3), 54 (1962).
[56] A. E. Hoerl and R. W. Kennard, Technometrics **12**, 55 (1970).
[57] R. Tibshirani, J. R. Stat. Soc. Ser. B **58**, 267 (1996).
[58] R. Blume-Kohout, Phys. Rev. Lett. **105**, 200504 (2010).
[59] D. Gross, Y. K. Liu, S. T. Flammia, S. Becker, J. Eisert, Phys. Rev. Lett. **105**, 150401 (2010).
[60] S. T. Flammia et al., New J. Phys. **14**, 095022 (2012).
[61] M. Benning and M. Burger, Acta Numerica **27**, 1 (2018).
[62] T. I. Seidman and C. R. Vogel, Inverse Problems **5**, 227 (1989).
[63] H. W. Engl, K. Kunisch, and A. Neubauer, Inverse Problems **5**, 523 (1989).
[64] T. A. Johansen, Automatica **33**, 441 (1997).
[65] P. Krantz, M. Kjaergaard, F. Yan, T.P. Orlando, S. Gustavsson, and W. D. Oliver, Applied Physics Reviews 6, 021318 (2019).
[66] A. Wächter and L. T. Biegler, Mathematical Programming **106**, 25 (2006).
[67] G. Kimura, Phys. Lett. A **314**, 339 (2003).
[68] M. S. Byrd and N. Khaneja, Phys. Rev. A **68**, 062322 (2003).