PAPER • OPEN ACCESS

Practical adaptive quantum tomography

To cite this article: Christopher Granade *et al* 2017 *New J. Phys.* **19** 113017

Related content

- Practical Bayesian tomography
  Christopher Granade, Joshua Combes and D G Cory

- High posterior density ellipsoids of quantum states
  Christopher Ferrie

- Optimal, reliable estimation of quantum states
  Robin Blume-Kohout

Recent citations

- Experimental quantum verification in the presence of temporally correlated noise
  S. Mavadia *et al*

- Minimax estimation of qubit states with Bures risk
  Anirudh Acharya and Mdlin Gu

New Journal of Physics

The open access journal at the forefront of physics

CrossMark

PAPER

Practical adaptive quantum tomography*

Christopher Granade ⬥ , Christopher Ferrie and Steven T Flammia

Centre for Engineered Quantum Systems, School of Physics, University of Sydney, Sydney, NSW, Australia

E-mail: cgranade@cgranade.com

Abstract

We introduce a fast and accurate heuristic for adaptive tomography that addresses many of the limitations of prior methods. Previous approaches were either too computationally intensive or tailored to handle special cases such as single qubits or pure states. By contrast, our approach combines the efficiency of online optimization with generally applicable and well-motivated data-processing techniques. We numerically demonstrate these advantages in several scenarios including mixed states, higher-dimensional systems, and restricted measurements.

Quantum information processing (QIP) promises advantages in a wide range of different contexts, including machine learning [2–4], chemistry simulation [5–7], and number theory [8, 9]. As such, the experimental effort to build useful QIP devices has exploded in recent years. In the course of this effort, *quantum tomography* is a valuable tool for diagnosing and debugging small quantum devices, and has subsequently seen a variety of different advances. In particular, Bayesian approaches to tomography which are especially well suited to utilizing prior information and adapting to changing experimental conditions have developed significantly in recent years [10–13], presenting a useful experimental tool [14–16].

In this paper we demonstrate the efficiency and accuracy of an adaptive tomography protocol that we call PAQT: practical adaptive quantum tomography. PAQT intelligently selects new measurements based on the outcomes of previous ones [10, 17–21]. Adaptivity has been experimentally demonstrated [14, 15, 22–25], but is not currently standard practice. Though adaptivity increases accuracy, the computational costs incurred outweigh that of simply repeating standard measurements many times. The PAQT approach employs a simple heuristic that can be efficiently computed between measurements, even with embedded hardware [26–29]. The algorithm we propose is therefore compatible with modern experimental design and avoids an important limitation of previous approaches.

We base our algorithm off of self-guided quantum tomography (SGQT), which treats adaptive tomography as a direct optimization problem rather than a new optimization problem between each measurement [30]. Though this affords an efficient and easy to implement adaptive heuristic, SGQT is not without its limitations. It requires assuming that the target state is pure, and it does not return rich region estimates for a state. What PAQT achieves is to effectively combine SGQT with conventional and easily-implemented tomographic estimators, such as the Bayesian particle filter or least-squares fit (LSF) estimators. Under this approach, an experimentalist can collect data using SGQT (even if its assumptions are not met), and then post-process this data using particle filtering or LSF.

The benefit of PAQT is two-fold. (1) From the point of view of traditional tomography, it gives an adaptive tomography protocol requiring only modest computational resources, as the bulk of the computational cost is offloaded to post-processing. (2) From the point of view of simulation-based optimization tomography (such as SGQT), it effectively augments the output with region estimation providing a statistically robust quantification of uncertainty. Thus, while we do not explicitly demonstrate that the improved scaling of Ferrie [30] remains in the more general case considered here, PAQT does provide a practical and efficient procedure for performing adaptive quantum tomography with rigorous statistical principles.

---

* http://cgranade.com complete data and source code for this work are available online [1], and can be previewed at https://goo.gl/koiWxR.

The outline of the paper is as follows. In section 1 we define and review the problem of tomography as well as three standard solutions: least squares, maximum likelihood and Bayesian mean estimation. In section 2, we review the approaches to measurement adaptive tomography including the recently introduced self-guided technique. In section 3, introduce PAQT by combining SGQT with adaptive Bayesian tomography and detail the results of our numerical experiments. Section 4 concludes with a discussion.

1. The tomographic problem

In quantum state tomography, we are interested in reconstructing a quantum state from a collection of *informationally complete* measurements made on that state [31–33]. That is, a set of measurements is chosen such that if one learns their frequencies given a quantum system of interest, the frequencies for any other measurement of that system can then be predicted. If the system of interest is a qubit, for instance, then knowing the expectations of the observables $\{\sigma_x, \sigma_y, \sigma_z\}$ allows for predicting the distribution over outcomes for any other measurement. The empirical reconstruction of quantum states from measurements of informationally complete observables has been reviewed by D'Ariano *et al* [34], and reviewed in the case of continuous variables by Lvovsky and Raymer [35]. Here, we will focus on the case of state tomography in finite-dimensional systems.

That a quantum state can be empirically determined in principle, however, leaves the question of how to estimate a state in practice, given finite experimental resources. For instance, given data from an informationally complete set of observables, one could use a linear reconstruction, a maximum likelihood estimator [36–38], or a Bayesian mean estimator [10–13, 39] to report a state. We will detail each such approach below, and describe their relative strengths and weaknesses.

Before proceeding, we note that though we consider the general case of tomography in this work, substantial progress has been made by considering considering important special cases under which a state can be much more easily characterized. In particular, permutationally invariant tomography reconstructs the part of a multiqubit density matrix which is invariant under exchange of the qubits [40]. Compressed sensing allows for the efficient recovery of *low-rank* quantum states [41, 42], and has been applied experimentally in systems as large as six qubits [43]. Similarly, MPS [44] and PEPS [45] tomography use the MPS and PEPS ansatzes to improve exponentially on naïve methods for states that are well-approximated by common tensor network ansatzes [46]. Though we do not explore the possibility in this work, we expect that heuristic approaches should also offer similar advantages to tomographic estimation in these cases.

1.1. Problem set up

First, consider an orthonormal basis for traceless Hermitian operators $\{B_j\}_{j=1}^{d^2-1}$—the Pauli basis, for example. That is, for all i, j, $B_j^{\dagger} = B_j$ and $\mathrm{Tr}(B_k B_j) = \delta_{kj}$ and $\mathrm{Tr}(B_j) = 0$. Then, any state ρ can be written

$$\rho = \frac{1}{d} + \sum_{j=1}^{d^2-1} \theta_j B_j, \tag{1}$$

for some vector of parameters $(\boldsymbol{\theta})_j = \theta_j$. Importantly, these parameters are constrained since $\rho \geqslant 0$. This poses a problem for many approaches, but there are well-motivated methods which produce a valid quantum state starting from a non-physical matrix [47].

Let us assume two-outcome test measurements are made such that each measurement outcome is either 1 or 0 and represented by the pair $\{P_k, \mathbb{1} - P_k\}$. The Born rule dictates that the probability to get 1, say, is $\mathrm{Pr}(1|\rho, P_k) = \mathrm{Tr}(\rho P_k)$. Since the operators $\{B_j\}$ form a basis, we can write

$$P_k = \frac{1}{d} + \sum_{j=1}^{d^2-1} p_{kj} B_j, \tag{2}$$

and the Born rule vectorizes to

$$\mathrm{Pr}(1|\rho, P_k) = \mathrm{Tr}(\rho P_k) = \frac{1}{d} + \boldsymbol{p}_k^{\mathrm{T}} \boldsymbol{\theta}, \tag{3}$$

where $(\boldsymbol{p}_k)_j = p_{kj}$. Denote $f_k = \mathrm{Pr}(1|\rho, P_k)$ and $(\boldsymbol{f})_k = f_k$. Also define the matrix $\boldsymbol{X}$ with entries $(\boldsymbol{X})_{kj} = p_{kj}$. Then the above condenses to

$$\boldsymbol{f} = \frac{1}{d} + \boldsymbol{X}\boldsymbol{\theta}. \tag{4}$$

If we perform at least d^2 such measurements such that the set $\{P_k\}_{k=1}^{d^2}$ is linearly independent, then the probabilities $\boldsymbol{f}$ are sufficient to determine ρ uniquely. That is, the linear system in (4) has a solution set with a single valid quantum state. In practice we do not have access to $\boldsymbol{f}$, but only samples drawn from the distribution

that it defines. Suppose N_k measurements of $\{P_k, \mathbb{1} - P_k\}$ yielded n_k 1s and $N_k - n_k$ 0s. Then, the empirical frequencies are

$$\hat{f}_k = \frac{n_k}{N_k}. \tag{5}$$

The task of tomography is to assign a quantum state $\boldsymbol{\theta}$ to each data set $\hat{\boldsymbol{f}}$.

1.2. Linear inversion tomography

Next, we will outline the traditional approach to solving the tomography problem. While we do not recommend this approach, it usually provides reasonable answers and is at least implicitly the starting point for more sophisticated approaches.

We begin by setting the empirical frequencies equal to the (rescaled) theoretical probabilities $\hat{\boldsymbol{f}} = \boldsymbol{f}$. After all, $\mathrm{Tr}(\rho P_k)$ is literally the expectation value of the observable P_k. In any case, if we let $\boldsymbol{Y} = \hat{\boldsymbol{f}} - 1/d$, the new system of equations

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\theta}, \tag{6}$$

may not have a solution if more than d^2 different measurements have been made. The traditional approach is to use the *least squares* estimator

$$\hat{\boldsymbol{\theta}}_{\mathrm{LS}} = \underset{\boldsymbol{\theta}}{\mathrm{argmin}} \, \|\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\theta}\|_2^2, \tag{7}$$

which has the exact solution

$$\hat{\boldsymbol{\theta}}_{\mathrm{LS}} = (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y}. \tag{8}$$

This solution is not guaranteed to produce a positive semidefinite estimate. One can resort to performing *constrained* least squares (which is 'not that hard' since one probably has access to a black box implementation of this using a canned scientific software library) or one can use a two-step approach [47] that outputs the 'closest' physical state to a given matrix. There is no consensus on which should be preferred and we make no recommendations here. In our simulations, we have set all negative eigenvalues to zero, as we observe that in practice, measurements designed by self-guided tomography tends to only rarely yield $\hat{\boldsymbol{\theta}}_{\mathrm{LS}}$ corresponding to $\rho \ngeq 0$.

1.3. Maximum likelihood tomography

The linear least squares approach is folklore as old as the problem of tomography, but has been stated explicitly by Qi *et al* [48]. It usually arises when using a Gaussian approximation to the likelihood function in maximum likelihood estimation (MLE) (see, for example, Kaznady and James [49]). The *likelihood function* is the probability distribution of the data *given* a state $\boldsymbol{\theta}$, thought of as a function of $\boldsymbol{\theta}$. Since each measurement is an independent binomial trial, the likelihood function is quite simple:

$$\mathrm{Pr}(\hat{\boldsymbol{f}}|\boldsymbol{\theta}, \boldsymbol{X}) = \prod_k \binom{N_k}{N_k \hat{f}_k} \left(\frac{1}{d} + \boldsymbol{p}_k^{\mathrm{T}}\boldsymbol{\theta}\right)^{N_k \hat{f}_k} \left(1 - \frac{1}{d} - \boldsymbol{p}_k^{\mathrm{T}}\boldsymbol{\theta}\right)^{N_k(1-\hat{f}_k)}. \tag{9}$$

One of the oldest techniques in classical statistical estimation is MLE, which prescribes the estimate

$$\hat{\boldsymbol{\theta}}_{\mathrm{MLE}} = \underset{\boldsymbol{\theta}}{\mathrm{argmin}} \, \mathrm{Pr}(\hat{\boldsymbol{f}}|\boldsymbol{\theta}, \boldsymbol{X}). \tag{10}$$

This does not have a closed form in general. To make some traction, we can approximate the likelihood function by a Gaussian (perhaps with appeal to the central limit theorem). A Gaussian is defined by its mean and variance, so we need only those from the actual distribution to make the approximation. These are simple enough to derive from the properties of the binomial distribution:

$$\mathbb{E}[\hat{\boldsymbol{f}}] = \frac{1}{d} + \boldsymbol{X}\boldsymbol{\theta}, \tag{11}$$

$$\mathbb{V}[\hat{\boldsymbol{f}}]_{kj} = \delta_{kj} \frac{\left(\frac{1}{d} + \boldsymbol{p}_k^{\mathrm{T}}\boldsymbol{\theta}\right)\left(1 - \frac{1}{d} - \boldsymbol{p}_k^{\mathrm{T}}\boldsymbol{\theta}\right)}{N_k}. \tag{12}$$

The location of the maximum of a function is the same as that of the log of the function. The logarithm of the Gaussian approximation to the likelihood function (ignoring terms which do not depend on $\boldsymbol{\theta}$) is

$$-\frac{1}{2}\sum_k \frac{(Y_k - \boldsymbol{p}_k^{\mathrm{T}}\boldsymbol{\theta})^2 N_k}{\left(\frac{1}{d} + \boldsymbol{p}_k^{\mathrm{T}}\boldsymbol{\theta}\right)\left(1 - \frac{1}{d} - \boldsymbol{p}_k^{\mathrm{T}}\boldsymbol{\theta}\right)}. \tag{13}$$

We make one more approximation, which is again replacing the probabilities with their empirical frequencies[1] such that the maximum likelihood problem then becomes

$$\hat{\boldsymbol{\theta}}_{\mathrm{MLE}} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \|\boldsymbol{Y}' - \boldsymbol{X}'\boldsymbol{\theta}\|_2^2, \tag{14}$$

where we have *weighted* $\boldsymbol{Y}$ and $\boldsymbol{X}$ by the variance:

$$Y_k' = \sqrt{\frac{N_k}{\hat{f}_k(1 - \hat{f}_k)}}\, Y_k, \quad X_k' = \sqrt{\frac{N_k}{\hat{f}_k(1 - \hat{f}_k)}}\, X_k. \tag{15}$$

Notably, this approach fails if $\hat{f}_k = 0$ or 1 for any k, as the variance in these cases approaches zero, so that $Y'_k \to \infty$. To solve this, we hedge the empirical frequencies by $\beta = 0.5$, so that we use $\hat{f}_k = (n_k + 0.5)/(N_k + 1)$ when computing the MLE [50].

1.4. Bayesian tomography

As opposed to the frequentist techniques noted above, the Bayesian approach centers around Bayes' rule, which prescribes how to update a prior distribution $\Pr(\boldsymbol{\theta})$ to a posterior distribution $\Pr(\boldsymbol{\theta}|\hat{f}, \boldsymbol{X})$ that is conditioned on the observed frequencies $\hat{f}$. Concretely,

$$\Pr(\boldsymbol{\theta}|\hat{f}, \boldsymbol{X}) = \frac{\Pr(\hat{f}|\boldsymbol{\theta}, \boldsymbol{X})\Pr(\boldsymbol{\theta})}{\Pr(\hat{f}|\boldsymbol{X})}, \tag{16}$$

where $\Pr(\hat{f}|\boldsymbol{\theta}, \boldsymbol{X})$ is the likelihood function of (9), and where $\Pr(\hat{f}|\boldsymbol{X})$ is a pesky normalization that we will deal with implicitly when doing numerical calculation. When Bayes' rule is used iteratively, the posterior for one experiment becomes the prior for the next. In words, this equation is a prescription of the full distribution of knowledge about the quantum state *given* the data that was actually observed. What can we do with this?

First, we can produce a single 'point' estimate of $\boldsymbol{\theta}$ via the posterior mean:

$$\hat{\boldsymbol{\theta}}_{\mathrm{BME}} = \mathbb{E}_{\boldsymbol{\theta}|\hat{f}, \boldsymbol{X}}[\boldsymbol{\theta}], \tag{17}$$

where BME stands for Bayesian mean estimator. The mean estimator is not the only option, though it is optimal for certain figures of merit [39], or at least near-optimal [51]. Second, the posterior distribution naturally encodes 'error bars' by way of the posterior covariance tensor [13, 39]. Finally, the data can be processed *online* in the sense that new data can be incorporated into the distribution without the need to reanalyze all previous data at the same time. This lends itself naturally to *adaptive* tomography, discussed in the next section.

In practice, however, exactly implementing Bayesian mean estimation is quite difficult, as the expectation value in (17) may not be analytically tractable outside of important special cases. We will therefore follow the approach of Huszár and Houlsby [10] and use the particle filtering algorithm [52] to numerically implement Bayesian estimation. This approach has since been used by Ferrie [11, 12] and by Granade *et al* [13] to develop useful applications of Bayesian tomography, by Stenberg *et al* [53] to learn coherent states, and has been successfully applied outside of tomography to efficiently learn Hamiltonians using classical [54] and quantum resources [55]. For our purposes here, we are primarily interested in the property that once a datum has been incorporated into a particle filter, it may be discarded, such that we do not incur computational costs that grow faster than the amount of data. Utilizing this advantage, together with the results of Beskos *et al* [56] we conjecture that Bayesian tomography with particle filtering requires computational costs scaling as $O(d^4 N p)$, where $\rho \in \mathbb{C}^{d \times d}$, N is the number of measurements, and p is the number of particles, as explained below. Naïvely, one might expect that $p \in O(\exp d)$ is required, but the result of Beskos *et al* [56] shows that p can always be chosen to be subexponential in d. Moreover, p can be chosen independently of d in cases where we use p to control the estimation accuracy rather than the problem dimension.

Particle filtering proceeds by approximating the prior and posterior distributions at each step of Bayesian inference as a weighted sum of δ functions,

$$\Pr(\boldsymbol{\theta}) \approx \sum_i w_i \delta(\boldsymbol{\theta} - \boldsymbol{\theta}_i), \tag{18}$$

where $\{w_i\}$ are the weights of the *particles* located at $\{\boldsymbol{\theta}_i\}$. Upon observing a datum $\hat{f}_k$, the weights are then updated by calling the likelihood function for each particle,

---

[1] A discussion of the consistency of this replacement can be found in [78].

$$w_i \mapsto w_i \times \Pr(\hat{f}_k | \boldsymbol{\theta}_i) \big/ \mathcal{N}, \tag{19}$$

where $\mathcal{N}$ is the normalization factor in Bayes' rule (16), which can be found implicitly by demanding that $\sum_i w_i = 1$. The BME is then found by taking a sum over the particles representing the current posterior,

$$\hat{\boldsymbol{\theta}}_{\mathrm{BME}} = \sum_i w_i \boldsymbol{\theta}_i. \tag{20}$$

Numerical stability in particle filtering is provided by the use of a *resampling algorithm* which replaces the particles by a new set of particles that more effectively represents the same posterior. We will use the Liu and West resampling algorithm [57], which mixes the current posterior with a Gaussian distribution of the same mean and covariance. The resampling is controlled by a parameter $a \in [0, 1]$, with smaller a corresponding to 'more Gaussian' posteriors.

2. Adaptive and self-guided tomography

We have not yet addressed the issue of $\boldsymbol{X}$, the matrix defining the choice of measurements. How should this choice of measurements be done? This is an open problem, with the lack of consensus mostly due to incompatible choices of criteria for optimality. In any case, the fact that some measurements are better than others suggests that improvements can be made through *adaptive* tomography—that is, choosing new measurement settings based on information obtained from past measurement settings.

2.1. Adaptive tomography
The first to consider adaptive state tomography was Fischer *et al* [17], who did so for a single qubit assumed to be in a pure state. That is, the prior was taken to be a uniform distribution on the surface of the Bloch sphere. The adaptivity consists of maximizing the entropy of the sampling distribution and expected fidelity. The estimator was chosen to be the maximum of the posterior distribution. This was later experimentally realized for a short set of measurements by pre-computing and storing the optimal experiment choices in a look-up table [22].

Adaptive state tomography has also been investigated in the context of parameterized models and Fisher information. Barndorff-Nielsen and Gill [18] showed that the quantum Fisher information for a single parameter can be obtained asymptotically by adaptively choosing the measurement settings in a two-stage procedure. The asymptotic two-step approach seems also to have been independently discovered by Řeháček *et al* [19] and Bagan *et al* [20]. An experimental demonstration has verified a quadratic improvement in accuracy [23, 58]. These approaches, however, are of more theoretical interest as they are guaranteed only asymptotically or require the total number of measurements to be specified *a priori*.

A generic approach using the maximum likelihood estimator and measurements minimizing the expected variance also showed an improvement over standard quantum tomography [21]. This has been made more practical through use of a recursive least-squares formula in Qi *et al* [25]. Below we will see that our choice of heuristic for adaptation may lead the least squares estimator to fail due to ill-conditionedness. Our results below will suggest that the better approach is the Bayesian one.

2.2. Bayesian adaptive quantum tomography
The Bayesian method also allows for a principled approach to adaptive measurements since one has a very formal definition of *expected utility* of a measurement. Consider (16) in the case of a *hypothetical* measurement $\boldsymbol{X}$, which could produce data $\hat{f}$. Then, one can define the expected utility of the measurement as

$$U(\boldsymbol{X}) = \mathbb{E}_{\hat{f}, \boldsymbol{\theta} | \boldsymbol{X}}[L(\boldsymbol{\theta})], \tag{21}$$

where L is an arbitrary loss function.

Fischer *et al* [17] considered both the log-loss and fidelity for a single qubit. Huszár and Houlsby [10] considered the *information gain*, which has since been used to define an adaptive protocol in one- and two-qubit optical experiments [14, 15]. Most recently, the fidelity for arbitrary dimensions has been studied and numerics performed on one and two-qubits [59].

Calculating these utilities, however, poses a problem since one may be able to perform a great deal of non-optimized experiments before the calculation of the 'best' experiment can be completed. These intermediate experiments, while not optimized, still contain useful information about the state and may provide better accuracy when the cost of optimization is included. Hence the need for experiment design *heuristics* that realize the benefits of adaptivity without computing or optimizing over utility functions, providing significant improvements in efficiency.

In the context of Hamiltonian learning, for example, heuristics have been used to obtain many of the benefits of explicitly optimizing a utility, while avoiding much of the computational expense [55, 60]. Machine learning

techniques have recently been applied to the design of good heuristics for quantum characterization problems [61], but we will take a different approach and instead use *stochastic optimization* to provide an efficient heuristic.

2.3. Self-guided quantum tomography

SGQT is an adaptive tomography scheme which avoids the linear inversion problem altogether by posing the tomography problem as one of optimization rather than estimation [30]. In particular, self-guided tomography finds a pure state $|\phi\rangle$ such that the overlap $F(\phi, \rho) = \langle\phi|\rho|\phi\rangle$ is maximized for a true state ρ. If $\rho = |\psi\rangle\langle\psi|$ is a pure state, then $F(\phi, \rho)$ is maximized if and only if $|\phi\rangle = e^{i\theta}|\psi\rangle$ for a phase θ, such that an optimal solution is also an accurate estimate of the true state.

An earlier work took a similar approach by testing whether the unknown qubit state was symmetric with a reference state [62], where the reference state is chosen adaptively to maximize fidelity. However, the method is defined only for a single qubit and requires a second fully characterized and controllable qubit along with an entangled measurement.

Having phrased state estimation as an optimization problem, self-guided tomography proceeds by experimentally estimating the objective function F from empirical frequencies. This results in a stochastically evaluated objective function, such that the optimization problem is amenable to attack by stochastic optimization algorithms. We will in particular rely on the simultaneous perturbative stochastic approximation (SPSA) [63].

The SGQT estimate is precisely defined as follows. We let $|\phi\rangle$ be a parameterization of pure states of a given dimension in terms of a vector ϕ of real numbers; for instance, qubit states can be parameterized by their Bloch angles. We then begin with a random state $|\phi_0\rangle$ and iteratively produce new states $|\phi_k\rangle$ which serve the dual role of specifying the current estimate of the state and next measurements to perform. At iteration k, we perform the measurements $\{P_{k,\pm}, \mathbb{1} - P_{k,\pm}\}$, where

$$P_{k,\pm} = |\phi_{k-1} \pm \epsilon_k\Delta_k\rangle\langle\phi_{k-1} \pm \epsilon_k\Delta_k|, \tag{22}$$

and Δ_k is a random vector that is constructed by setting each entry to ± 1 with equal probability. Here ϵ_k is a step-size parameter chosen below. The outcomes of these measurements are denoted $\hat{f}_{k,\pm}$. The gradient of the fidelity is estimated from these measurements to be

$$\hat{g}_k = \frac{\hat{f}_{k,+} - \hat{f}_{k,-}}{2\epsilon_k}\Delta_k. \tag{23}$$

Using these, and an additional gain parameter α_k, the SPSA algorithm mimics standard gradient ascent, but along the random direction Δ_k:

$$|\phi_k\rangle = |\phi_{k-1} + \alpha_k\hat{g}_k\rangle. \tag{24}$$

Convergence is guaranteed [63] given the specification of Δ_k above and

$$\epsilon_k = \frac{1}{k^{1/3}}, \tag{25a}$$

$$\alpha_k = \frac{1}{k}. \tag{25b}$$

Unless otherwise noted, however, we shall use the parameters suggested by Spall [63],

$$\epsilon_k = 0.1/k^{0.101} \text{ and } \alpha_k = 10/k^{0.602}. \tag{26}$$

SPSA has also been applied in quantum information to design high-fidelity control sequences given randomized benchmarking experiments [64, 65]. In particular, Ferrie showed that self-guided tomography can rapidly learn pure states for comparatively large quantum systems [30]. To the best of our knowledge, self-guided tomography is the only adaptive tomography technique which has gone beyond two qubits, even in simulation. SGQT has also recently been demonstrated in an optical experiment [24].

SGQT is not without its limitations, however. The aim of the current work is to mitigate the following three limitations of SGQT: (1) it is restricted to pure state tomography, (2) it does not report error bars, and (3) it can not be restricted to local measurements.

3. Practical adaptive quantum tomography

From the above discussion, we find that SGQT potentially offers many advantages for experimental practicality over traditional protocols, but at the cost that it does not accurately report mixed states, and does not certify its own errors. Happily, these are precisely the advantages of the Bayesian approach, such that we can collect data using self-guided tomography, then post-process with offline estimation.
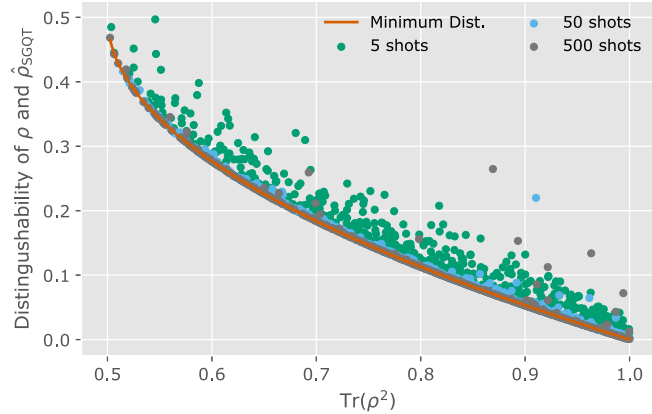
Figure 1. Distinguishability between self-guided estimated states and true states drawn from the Hilbert–Schmidt prior for a qubit, plotted versus the best achievable distinguishability for any estimator constrained to pure states $\rho = |\psi\rangle\langle\psi|$, where the distinguishability between ρ and σ is defined as the trace distance $\frac{1}{2}\|\rho - \sigma\|_1$, minimized for pure ρ and fixed σ by $\left(1 - \sqrt{2\,\mathrm{Tr}(\sigma^2) - 1}\right)/2$. The self-guided estimates are drawn from 10,000 iterations with either $n = 5, 50$ or 500 shots per measurement, such that each measurement f_k is drawn from a binomial distribution with n trials. As the number of shots per measurement increases, the self-guided estimates approach the closest states allowed by the pure state assumption, demonstrating that the self-guided procedure produces useful data even when the true state is mixed.

We introduce PAQT—practical adaptive quantum tomography—an optimized numerical approach which implements the idea of merging self-guided tomography as an online experiment design heuristic into Bayesian data analysis. In principle, post-processing of self-guided tomography data could be carried out with any tomographic estimator. We *define* PAQT as the use of Bayesian estimation in particular on the data gathered through the course of a self-guided tomography experiment, owing to the rich statistical principles underlying Bayesian inference. In utilizing self-guided tomography, PAQT automatically selects experiments online and can be implemented with modest experimental hardware, including modern embedded controllers such as field-programmable gate arrays. The advantages of PAQT are that it provides the enhanced precision of adaptive tomography together with fast data processing and experiment design. The framework provides robust and easily interpretable error regions without additional overhead. Explicitly, PAQT uses the results of the measurements (22) specified by SGQT with the Bayesian mean estimator (17). As described in section 1.1, the frequencies f upon which the Bayesian estimator are conditioned are a description of the results of each measurement. Since (22) specifies which measurements are to be performed, this is a complete specification of our protocol. Although we demonstrate the algorithm for state tomography, the method is equally applicable to channel tomography and other estimation tasks and can easily accommodate other estimators.
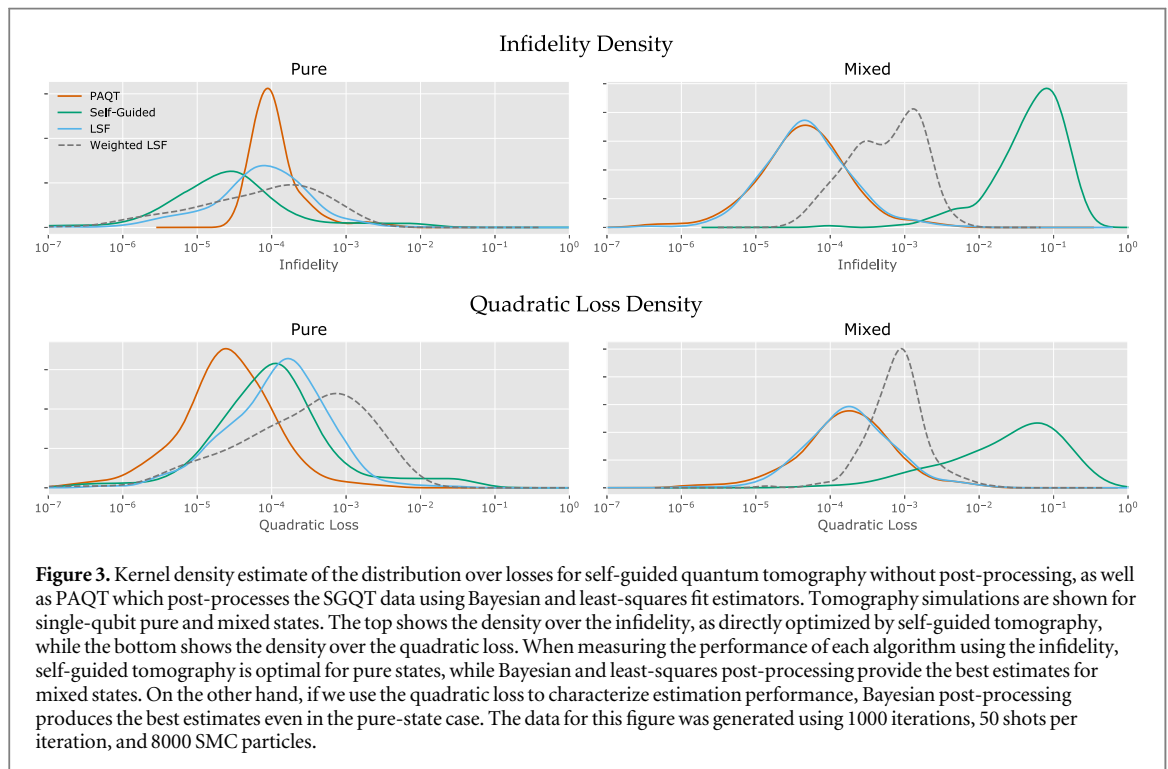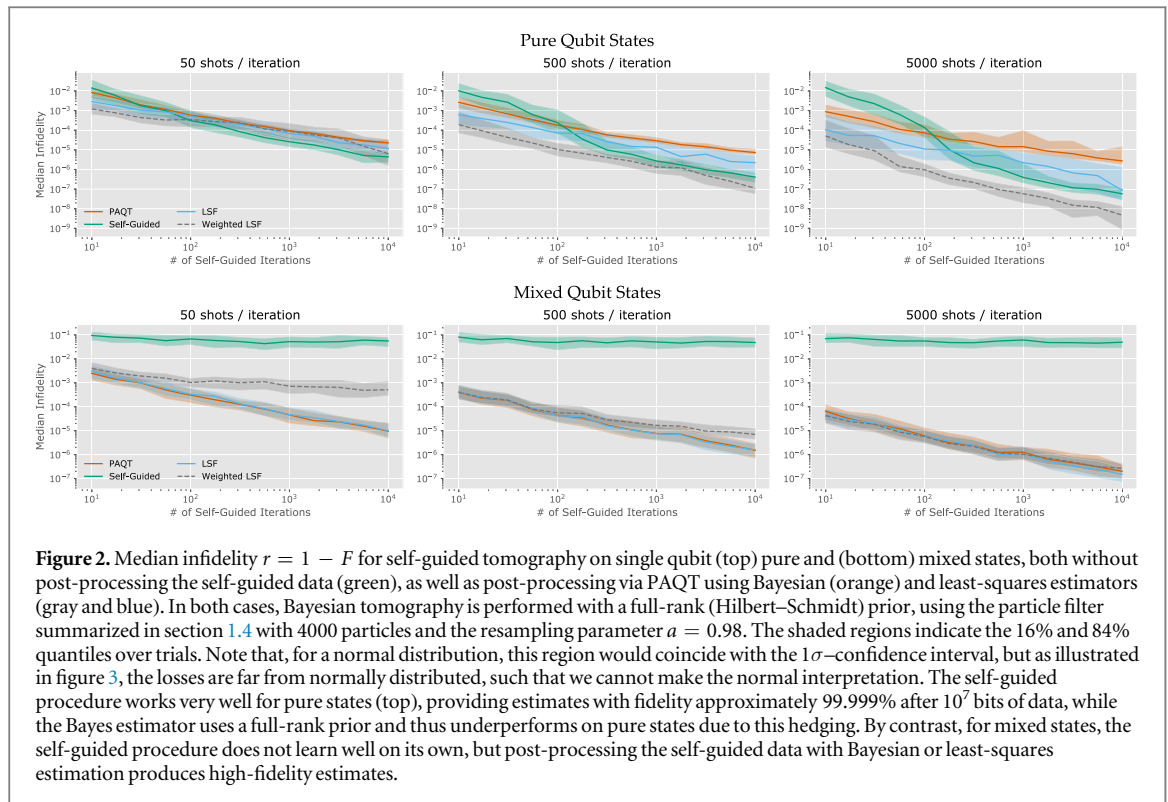
Our results use the QInfer 1.0a1 [66], QuTiP [67] 3.2.0, NumPy [68] 1.9.2, Pandas [69] 0.16.2 and SciPy 0.15.1 [70] libraries for Python 2.7 (Enthought Canopy 1.5.4) to perform the Bayesian analysis. We performed all simulations on the University of Sydney School of Physics cluster. Full source code for our simulations, and for our implementations of self-guided and least-squares tomography, as well as data summarizing all 554, 250 trials used in our numerical results can be found online in the supplementary material [1]. These trials are split over 652 different experimental conditions, such that for most plots, each point is generated from approximately 850 trials. In all numerical experiments, true states are chosen at random for each trial from the Ginibre distribution, which includes the Hilbert–Schmidt uniform and Haar uniform distributions as special cases [71]. For brevity, we will indicate these special cases as 'mixed' and 'pure', respectively. The supplemental material also includes complete details for all figures in this paper, and can be used to reproduce all numerical results shown here.

We start by noting in figure 1 that, in the case of qubits, the states estimated by self-guided tomography are almost as indistinguishable as the pure state closest to each true state in terms of the 1-norm. This makes it clear that, although self-guided tomography should not be expected to return a useful estimate if the true state is mixed, it is still heavily dependent on the true state such that we should expect self-guided tomography to collect useful data.

Indeed, as we show in figure 2, PAQT effectively combines self-guided tomography with least-squares and Bayesian estimators for both pure and mixed states on a qubit. In particular, even though self-guided tomography has ceased to learn states when the true state is a mixed state, the data collected can be used by both the Bayesian and LSF tomographic estimators to return very good estimates of the state.
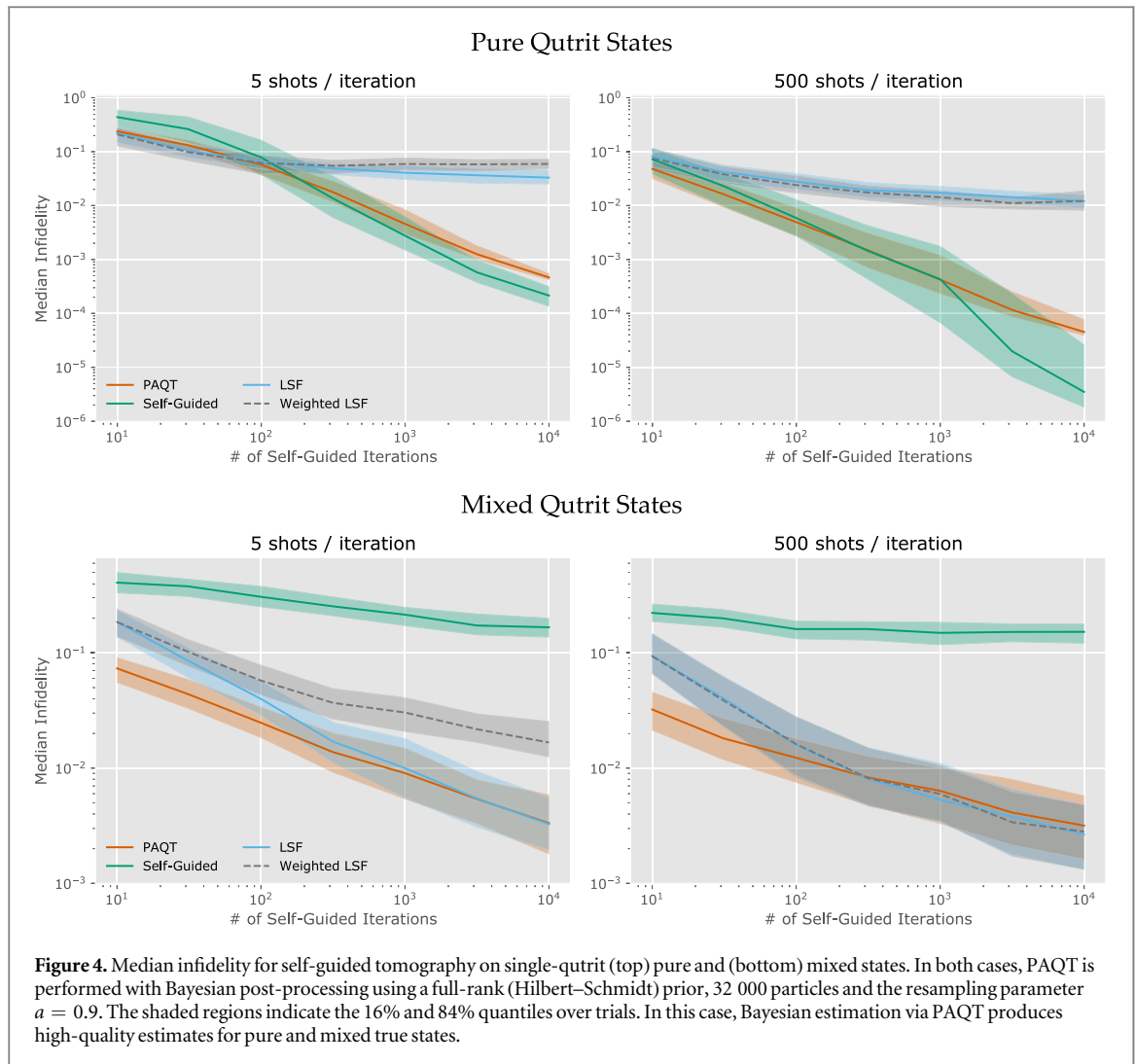
Which estimator in particular gives the lowest error depends strongly, however, on the loss function that one uses to quantify error. In figure 3, we compare the distribution over losses for the four tomographic procedures as applied to qubit pure and mixed states, and as measured by the infidelity and quadratic loss functions.

Figure 2. Median infidelity $r = 1 - F$ for self-guided tomography on single qubit (top) pure and (bottom) mixed states, both without post-processing the self-guided data (green), as well as post-processing via PAQT using Bayesian (orange) and least-squares estimators (gray and blue). In both cases, Bayesian tomography is performed with a full-rank (Hilbert–Schmidt) prior, using the particle filter summarized in section 1.4 with 4000 particles and the resampling parameter $a = 0.98$. The shaded regions indicate the 16% and 84% quantiles over trials. Note that, for a normal distribution, this region would coincide with the 1σ–confidence interval, but as illustrated in figure 3, the losses are far from normally distributed, such that we cannot make the normal interpretation. The self-guided procedure works very well for pure states (top), providing estimates with fidelity approximately 99.999% after 10^7 bits of data, while the Bayes estimator uses a full-rank prior and thus underperforms on pure states due to this hedging. By contrast, for mixed states, the self-guided procedure does not learn well on its own, but post-processing the self-guided data with Bayesian or least-squares estimation produces high-fidelity estimates.

Figure 3. Kernel density estimate of the distribution over losses for self-guided quantum tomography without post-processing, as well as PAQT which post-processes the SGQT data using Bayesian and least-squares fit estimators. Tomography simulations are shown for single-qubit pure and mixed states. The top shows the density over the infidelity, as directly optimized by self-guided tomography, while the bottom shows the density over the quadratic loss. When measuring the performance of each algorithm using the infidelity, self-guided tomography is optimal for pure states, while Bayesian and least-squares post-processing provide the best estimates for mixed states. On the other hand, if we use the quadratic loss to characterize estimation performance, Bayesian post-processing produces the best estimates even in the pure-state case. The data for this figure was generated using 1000 iterations, 50 shots per iteration, and 8000 SMC particles.

Whereas self-guided tomography directly optimizes the infidelity, we note that it performs very well according to this measure in the pure-state case. Similarly, the Bayesian mean estimator is optimal for Bregman divergences such as the quadratic loss $L(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}) := (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^\mathrm{T}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})$, so that it performs very well if we choose to quantify errors accordingly.

In figure 4, we consider self-guided tomography of pure and mixed qutrit states, showing that the benefits of using PAQT to combine SGQT with Bayesian tomography persist in this case. Notably, least-squares fitting does significantly less well for self-guided datasets on pure qutrits. Reducing the resampling parameter a to 0.9 allows the Bayesian estimator to remain robust in this case, however.
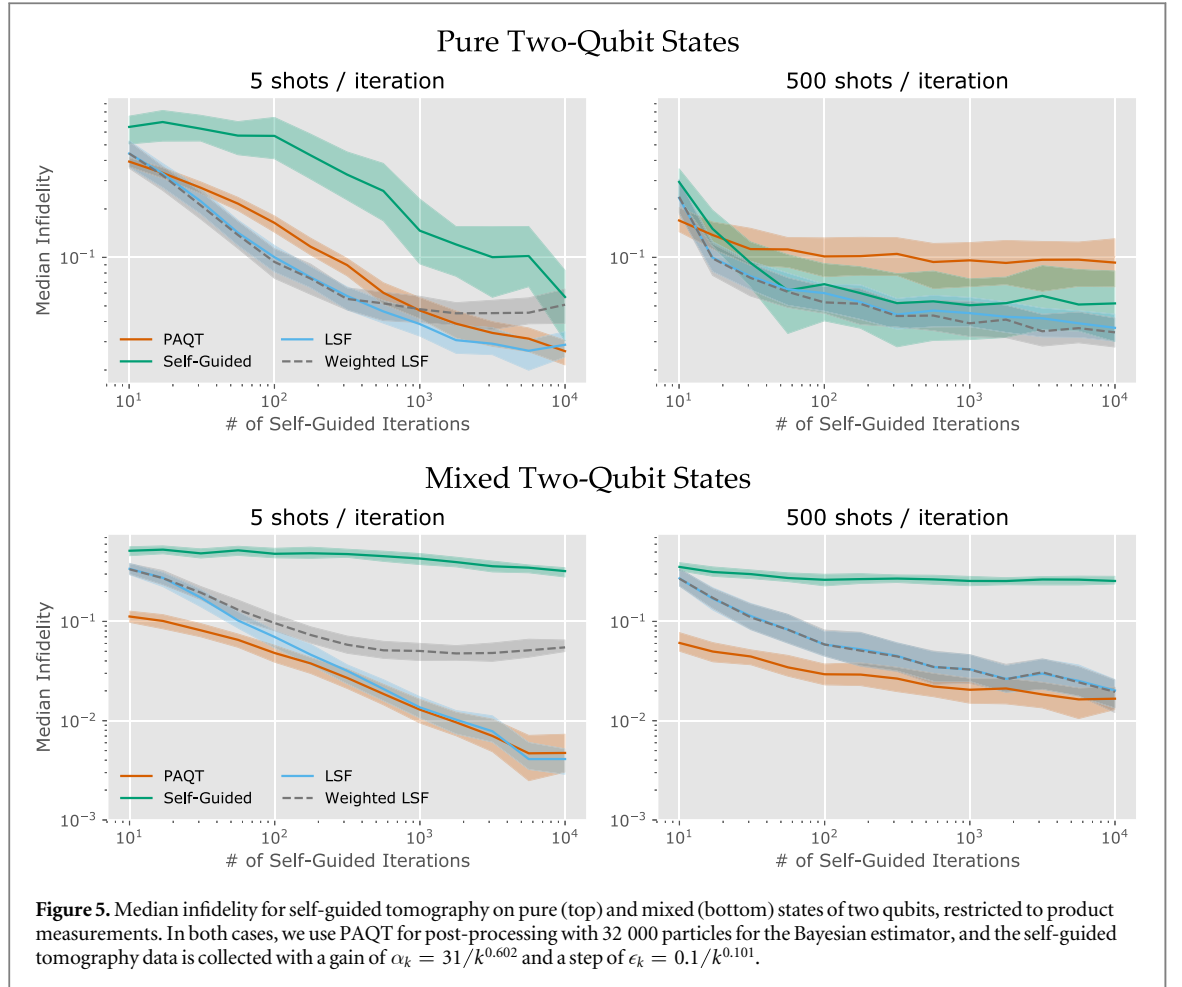
Figure 4. Median infidelity for self-guided tomography on single-qutrit (top) pure and (bottom) mixed states. In both cases, PAQT is performed with Bayesian post-processing using a full-rank (Hilbert–Schmidt) prior, 32 000 particles and the resampling parameter $a = 0.9$. The shaded regions indicate the 16% and 84% quantiles over trials. In this case, Bayesian estimation via PAQT produces high-quality estimates for pure and mixed true states.

We also consider the case in which the optimization procedure used by self-guided tomography is restricted to an incorrect model of the system under study. In particular, in figure 5, we collect data under the restriction that the true state is a mixed or pure product state of two qubits, then draw the true state from a Haar or Hilbert–Schmidt prior on the full four-dimensional state. In this way, the self-guided algorithm is explicitly following an incorrect model for the state. We note that, despite this, the Bayesian and least-squares estimators are both able to improve on their initial uncertainty by using data collected from the product state measurements. It is also interesting to note that the protocol with fewer measurements seems to perform better, which might be counterintuitive. However, remember that for this scenario, the *model* is wrong. More data will produce an estimate that is more accurate, but with respect to the wrong model. Thus, the procedure with fewer measurements performs better in this scenario through *less accurate* (noisy) measurements.

Finally, we note that the performance of the Bayesian estimator can be dramatically improved if we postselect on diagnostic information provided by the particle filtering algorithm. In figure 6, we show the kernel-density estimated distribution over infidelity for each of the qutrit and two-qubit cases, postselecting on the smallest effective sample size observed during a tomography run. That is, we accept a tomography trial if the particle filter weights $\{w_i\}$ satisfy

$$\frac{1}{\sum\limits_i w_i^2} \geqslant n_{\text{th}} \qquad (27)$$

throughout the experiment, for some choice of threshold n_{th}. For the qutrit case, using either 32 000 or 128 000 particles, we observe that as we increase this threshold (that is, as we demand a larger effective sample size), the mean performance rapidly approaches the median performance. Thus, performing this postselection allows us to exclude the worst-case performance of the Bayesian estimator. On the other hand, when the data are not especially informative, as in the two-qubit product measurement case, the benefit of postselection is significantly less pronounced.

**Figure 5.** Median infidelity for self-guided tomography on pure (top) and mixed (bottom) states of two qubits, restricted to product measurements. In both cases, we use PAQT for post-processing with 32 000 particles for the Bayesian estimator, and the self-guided tomography data is collected with a gain of $\alpha_k = 31/k^{0.602}$ and a step of $\epsilon_k = 0.1/k^{0.101}$.

4. Discussion

Though the point of SGQT is to avoid solving a large system of linear equations, the data collected from the performed measurements still define a set of equations that can be inverted in one way or another. This is the approach of LSF and weighted LSF. However, we note that these approaches do not perform well in all but a few of the cases considered. The explanation for this observation is that the constructed linear system is in general ill-conditioned.

Given infinite precision data, SGQT measurements would trace out a straight path through state space from the initial guess to the true state, following the gradient of the fidelity. This set of measurements will not be informationally complete. Due to the stochasticity of the algorithm, for finite data, a sufficiently large number of SGQT iterations will be informationally complete, but most of the measurements will be linearly dependent. This frustrates the stability of attempting to solve the linear equations defined by (6). The standard approach to quantify the stability of a linear system is through the condition number

$$\kappa(\boldsymbol{X}) = \frac{\sigma_1(\boldsymbol{X})}{\sigma_{d^2}(\boldsymbol{X})}, \tag{28}$$

where $\sigma_1(\boldsymbol{X})$ is the largest and $\sigma_{d^2}(\boldsymbol{X})$ is the smallest singular value. Smaller condition numbers lead to more stable linear systems. We will argue and demonstrate that self-guided tomography leads to measurements which define a linear system with large condition number. Importantly, it is only the process by which data is gathered (rather than analyzed) that determines the condition number. We will therefore restrict our discussion of condition numbers to SGQT as a data gathering procedure and the resultant effect on the numerical stability of different estimation strategies. In particular, our use of the condition number is distinct from its use in assessing the utility of a set of measurements for tomographic estimation [72, 73], as we are concerned not with tomographic completeness but with numerical stability.
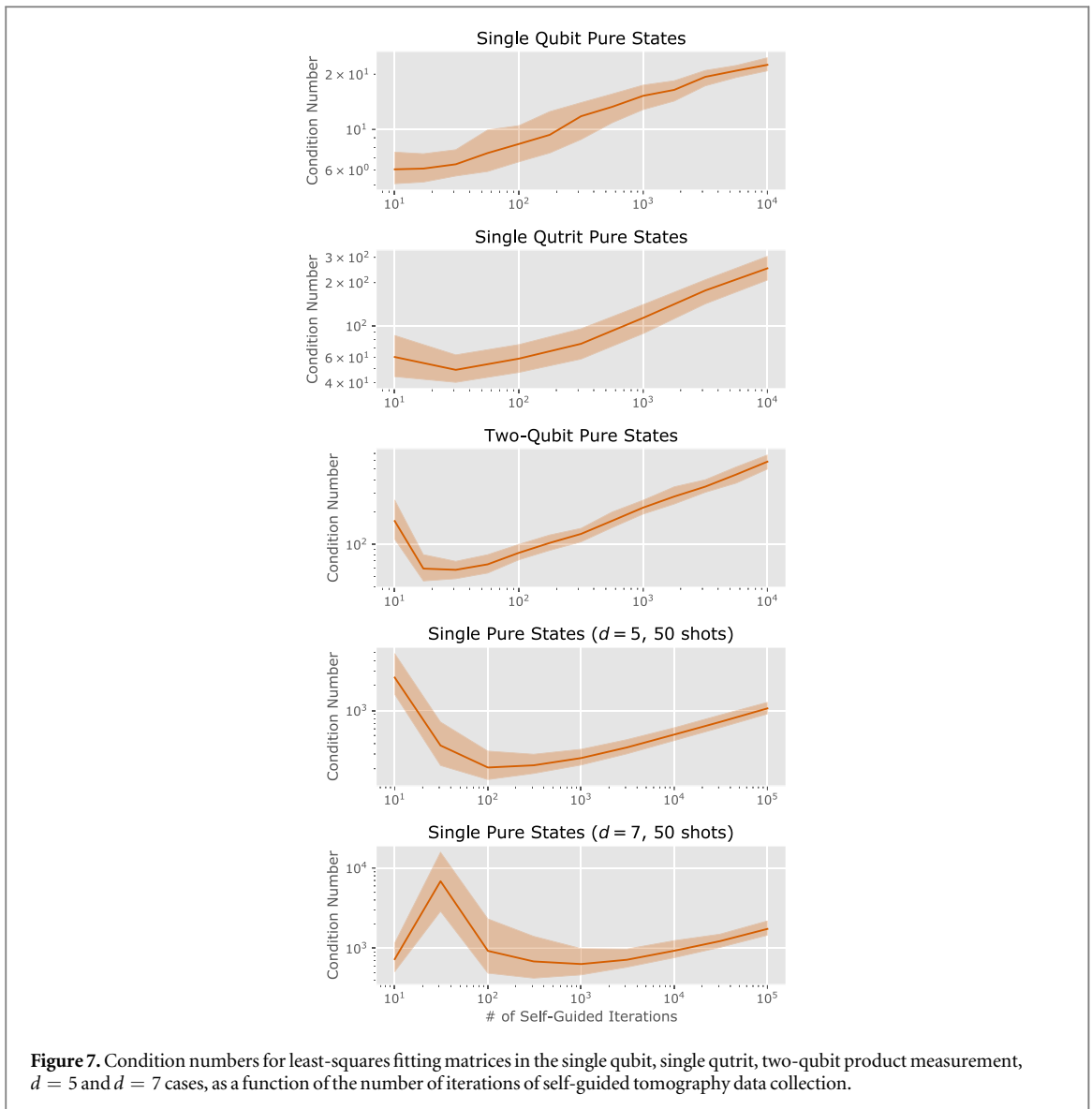
The largest singular value will be related to the total number of SGQT iterations since most of the late measurements will be nearly co-linear, clustering around the true state. The smallest singular value would be 1 in the ideal case of performing a subset of orthogonal basis measurements. However, as noted above, the system is only *barely* informationally complete—the matrix $\boldsymbol{X}$ is nearly rank-deficient (rank$<d^2 - 1$), in other words.
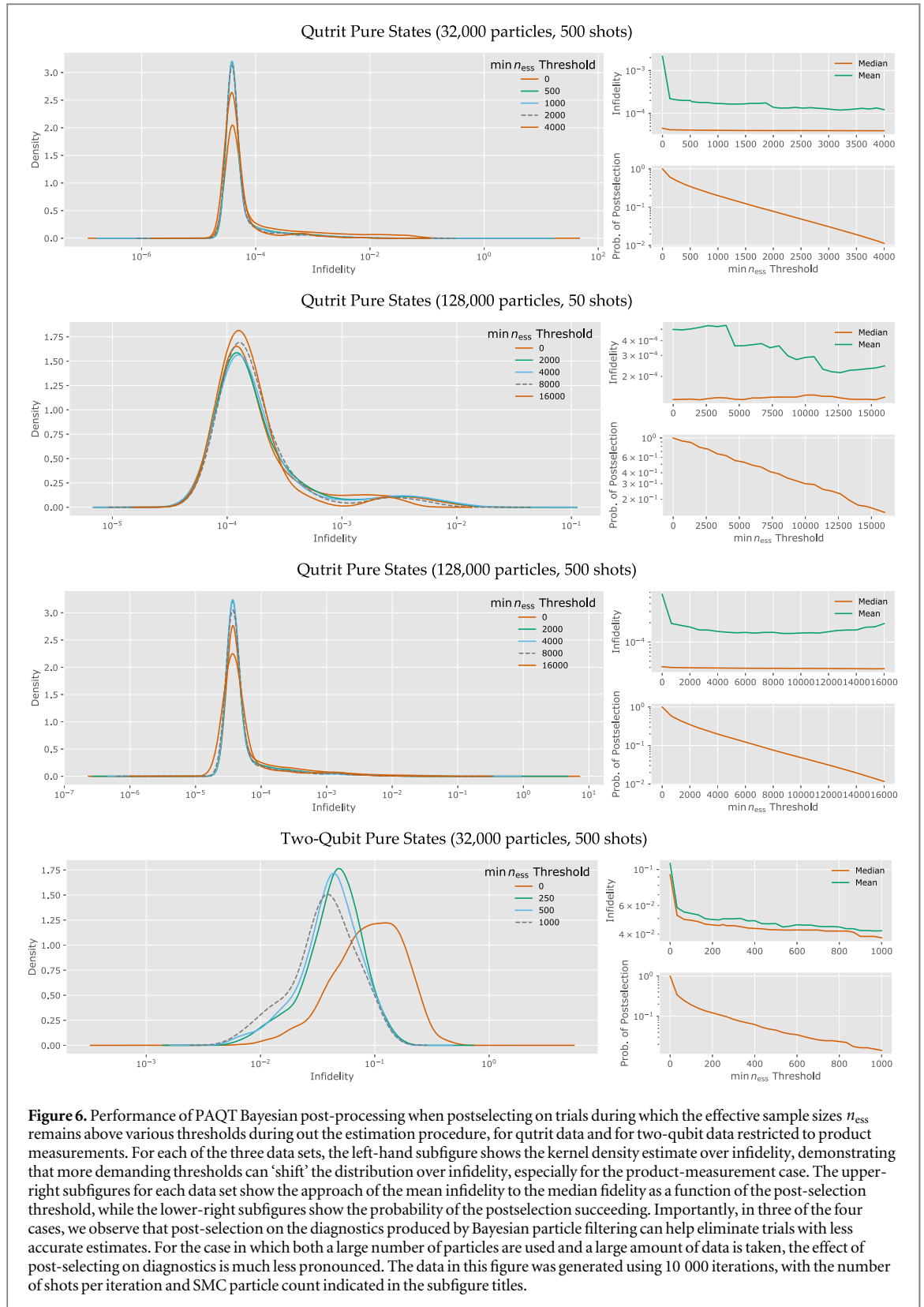
Figure 6. Performance of PAQT Bayesian post-processing when postselecting on trials during which the effective sample sizes n_{ess} remains above various thresholds during out the estimation procedure, for qutrit data and for two-qubit data restricted to product measurements. For each of the three data sets, the left-hand subfigure shows the kernel density estimate over infidelity, demonstrating that more demanding thresholds can 'shift' the distribution over infidelity, especially for the product-measurement case. The upper-right subfigures for each data set show the approach of the mean infidelity to the median fidelity as a function of the post-selection threshold, while the lower-right subfigures show the probability of the postselection succeeding. Importantly, in three of the four cases, we observe that post-selection on the diagnostics produced by Bayesian particle filtering can help eliminate trials with less accurate estimates. For the case in which both a large number of particles are used and a large amount of data is taken, the effect of post-selecting on diagnostics is much less pronounced. The data in this figure was generated using 10 000 iterations, with the number of shots per iteration and SMC particle count indicated in the subfigure titles.

The actually value of $\sigma_{d^2}(\boldsymbol{X})$, and hence, $\kappa(\boldsymbol{X})$, will vary quite a bit from run to run, but the scaling with the total number of measurements, K, will be $O(\sqrt{K})$. This is because most of the measurements will be approximately co-linear. In the exact case where $\boldsymbol{X}$ consists of $(d^2 - 1) \times (d^2 - 1)$ orthogonal submatrix and $K - 1$ repeated rows, the condition number is identically $\sqrt{K}$.

In figure 7, we plot the empirical condition number of $\boldsymbol{X}$ as a function of the total number of SGQT iterations. We see the expected behavior. The condition number starts high as there are simply not enough

Figure 7. Condition numbers for least-squares fitting matrices in the single qubit, single qutrit, two-qubit product measurement, $d = 5$ and $d = 7$ cases, as a function of the number of iterations of self-guided tomography data collection.

measurements to ensure informational completeness. Then, the condition number reaches a minimum value before rising at a rate of approximately $\sqrt{K}$ due to many nearly (but not exactly) identical measurements[2].

This effect identifies a fundamental tension between the benefit of measurement adaptivity and offline data analysis, which is why PAQT does well in spite of this tension. We note that in most cases using PAQT with a Bayesian mean estimator performs quite well and comes with many added benefits, as discussed above. In the cases where the Bayesian mean estimator does not perform well, we conjecture this is due to non-optimal choices of the particle filtering algorithm parameters rather than a fundamental problem of ill-conditionedness. This is not a problem to be swept under the rug, however, and a non-trivial optimization will need to be performed to find good operating points for the particle filtering algorithm.

A second comment concerns the standard claim in quantum state tomography work that all results obtained for states will immediately apply to quantum *process* tomography due to the isomorphism between quantum states and channels. Though this claim is broadly true, there is an important subtlety that we must consider. Under the Choi–Jamiłokowski isomorphism [74, 75], process tomography is equivalent to state tomography *with a restriction on allowable priors and measurements*. Thus, the product-state model of figure 5 is especially important in that it immediately shows that our adaptive state tomography protocol also provides a protocol for *process tomography*. Indeed, the Choi–Jamiłokowski isomorphism gives that product measurements on two copies of a quantum system are equivalent to preparing a state, evolving under an unknown map, and then measuring the output state [13, 76]. This observation has recently been utilized by Pogorelov *et al* [16] to perform adaptive quantum process tomography with Bayesian estimation implemented by particle filtering.

---

[2] For the case of $d = 7$, the condition number has some interesting transient behavior that we do not yet understand. However, it is still consistent with the asymptotic behavior described above.

With this in mind, then, our results show that self-guided state tomography is an efficient heuristic for designing quantum process tomography experiments, and may pose an interesting tradeoff for the computational cost of explicit adaptivity, in the same sense as self-guided tomography without the product measurement constraint provides a useful tradeoff for adaptive state tomography. The generalization to process tomography will be explored further in future work.

5. Conclusion

In summary, we have shown how to mitigate the drawbacks of SGQT using PAQT to provide explicit and statistically principled adaptive quantum tomographic estimates. In numerically testing PAQT, we have shown that SGQT alone is extremely efficient when the true state is pure, such that it is computationally challenging to compete with SGQT in high-dimensional problems where the pure state assumption is explicitly met. This allows us to more carefully deliniate between the advantages of each protocol, and to provide practical solutions for adaptive tomography.

However, more work needs to be done to refine the self-guided heuristic for mixed states and restricted measurement scenarios. An interesting open problem suggested by our work is to investigate if the scaling advantages of SGQT remain when using mixed states and, in the case of two or more qubits, when using product measurements. We expect that designing good heuristics for the challenging estimation problems which lie ahead for quantum technology will become an active area of research, as it has for classical machine learning problems.

Acknowledgments

ORCID iDs

Christopher Granade ⓘ https://orcid.org/0000-0001-6233-3132

References

[1] Granade C, Ferrie C and Flammia S T Practical adaptive quantum tomography: supplementary material (doi:10.6084/m9.figshare.3156700.v2)
[2] Wiebe N 2014 Using quantum computers to learn physics arXiv:1401.4507 [quant-ph]
[3] Wiebe N, Kapoor A and Svore K 2015 Quantum nearest-neighbor algorithms for machine learning *Quantum Inf. Comput.* **15** 0318–58
[4] Wiebe N and Granade C 2017 Can small quantum systems learn? *Quantum Inf. Comput.* **17** 0568–94
[5] Babbush R, McClean J, Wecker D, Aspuru-Guzik A and Wiebe N 2015 On the chemical basis of Trotter–Suzuki errors in quantum chemistry simulation *Phys. Rev.* A **91** 022311
[6] Hastings M B, Wecker D, Bauer B and Troyer M 2015 Improving quantum algorithms for quantum chemistry *Quantum Info. Comput.* **15** 1
[7] Kassal I, Whitfield J D, Perdomo-Ortiz A, Yung M-H and Aspuru-Guzik A 2011 Simulating chemistry using quantum computers *Annu. Rev. Phys. Chem.* **62** 185
[8] Shor P W 1995 Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer *SIAM J. Sci. Stat. Comput.* **26** 1484
[9] van Dam W and Sasaki Y 2012 Quantum algorithms for problems in number theory, algebraic geometry, and group theory arXiv:1206.6126 [quant-ph]
[10] Huszár F and Houlsby N M T 2012 Adaptive Bayesian quantum tomography *Phys. Rev.* A **85** 052120
[11] Ferrie C 2014 *New J. Phys.* **16** 093035
[12] Ferrie C 2014 High posterior density ellipsoids of quantum states *New J. Phys.* **16** 023006
[13] Granade C, Combes J and Cory D G 2016 Practical Bayesian tomography *New J. Phys.* **18** 033024
[14] Kravtsov K S, Straupe S S, Radchenko I V, Houlsby N M T, Huszár F and Kulik S P 2013 Experimental adaptive Bayesian tomography *Phys. Rev.* A **87** 062122
[15] Struchalin G, Pogorelov I, Straupe S, Kravtsov K, Radchenko I and Kulik S 2015 *Phys. Rev.* A **93** 012103
[16] Pogorelov I A, Struchalin G I, Straupe S S, Radchenko I V, Kravtsov K S and Kulik S P 2017 Experimental adaptive process tomography *Phys. Rev.* A **95** 012302
[17] Fischer D G, Kienle S H and Freyberger M 2000 Quantum-state estimation by self-learning measurements *Phys. Rev.* A **61** 032306

[18] Barndorff-Nielsen O E and Gill R D 2000 Fisher information in quantum statistics *J. Phys. A: Math. Gen.* **33** 4481

[19] Řeháček J, Englert B-G and Kaszlikowski D 2004 Minimal qubit tomography *Phys. Rev. A* **70** 052321

[20] Bagan E, Ballester M A, Gill R D, Muñoz-Tapia R and Romero-Isart O 2006 Separable measurement estimation of density matrices and its fidelity gap with collective protocols *Phys. Rev. Lett.* **97** 130501

[21] Sugiyama T, Turner P S and Murao M 2012 Adaptive experimental design for one-qubit state estimation with finite data based on a statistical update criterion *Phys. Rev. A* **85** 052107

[22] Hannemann T, Reiss D, Balzer C, Neuhauser W, Toschek P E and Wunderlich C 2002 Self-learning estimation of quantum states *Phys. Rev. A* **65** 050303

[23] Mahler D H, Rozema L A, Darabi A, Ferrie C, Blume-Kohout R and Steinberg A M 2013 Adaptive quantum state tomography improves accuracy quadratically *Phys. Rev. Lett.* **111** 183601

[24] Chapman R J, Ferrie C and Peruzzo A 2016 *Phys. Rev. Lett.* **117** 040402

[25] Qi B, Hou Z, Wang Y, Dong D, Zhong H-S, Li L, Xiang G-Y, Wiseman H M, Li C-F and Guo G-C 2017 *npj Quantum Inf.* **3** 19

[26] Stockton J, Armen M and Mabuchi H 2002 Programmable logic devices in experimental quantum optics *J. Opt. Soc. Am.* B **19** 3019

[27] Casagrande S 2014 On design and testing of a spectrometer based on an FPGA development board for use with optimal control theory and high-Q resonators *Masters Thesis* University of Waterloo, Ontario

[28] Lamb I D C, Colless J I, Hornibrook J M, Pauka S J, Waddy S J, Frechtling M K and Reilly D J 2016 An FPGA-based instrumentation platform for use at deep cryogenic temperatures *Rev. Sci. Instrum.* **87** 014701

[29] Hornibrook J M *et al* 2015 Cryogenic control architecture for large-scale quantum computing *Phys. Rev. Appl.* **3** 024010

[30] Ferrie C 2014 Self-guided quantum tomography *Phys. Rev. Lett.* **113** 190404

[31] Newton R G and Young B-l 1968 Measurability of the spin density matrix *Ann. Phys., NY* **49** 393

[32] Band W and Park J L 1970 The empirical determination of quantum states *Found. Phys.* **1** 133

[33] Band W and Park J L 1979 Quantum state determination: Quorum for a particle in one-dimension *Am. J. Phys.* **47** 188

[34] D'Ariano G M, Paris M G A and Sacchi M F 2003 *Advances in Imaging and Electron Physics* (vol 128) (Amsterdam: Elsevier) pp 205–308

[35] Lvovsky A I and Raymer M G 2009 Continuous-variable optical quantum-state tomography *Rev. Mod. Phys.* **81** 299

[36] Hradil Z 1997 Quantum-state estimation *Phys. Rev. A* **55** R1561

[37] Christandl M and Renner R 2012 Reliable quantum state tomography *Phys. Rev. Lett.* **109** 120403

[38] Faist P and Renner R 2016 *Phys. Rev. Lett.* **117** 010404

[39] Blume-Kohout R 2010 Optimal, reliable estimation of quantum states *New J. Phys.* **12** 043034

[40] Tóth G, Wieczorek W, Gross D, Krischek R, Schwemmer C and Weinfurter H 2010 Permutationally invariant quantum tomography *Phys. Rev. Lett.* **105** 250403

[41] Gross D, Liu Y-K, Flammia S T, Becker S and Eisert J 2010 Quantum state tomography via compressed sensing *Phys. Rev. Lett.* **105** 150401

[42] Flammia S T, Gross D, Liu Y-K and Eisert J 2012 Quantum tomography via compressed sensing: error bounds, sample complexity and efficient estimators *New J. Phys.* **14** 095022

[43] Schwemmer C, Tóth G, Niggebaum A, Moroder T, Gross D, Gühne O and Weinfurter H 2014 Experimental comparison of efficient tomography schemes for a six-qubit state *Phys. Rev. Lett.* **113** 040503

[44] Cramer M, Plenio M B, Flammia S T, Somma R, Gross D, Bartlett S D, Landon-Cardinal O, Poulin D and Liu Y-K 2010 Efficient quantum state tomography *Nat. Commun.* **1** 149

[45] Landon-Cardinal O and Poulin D 2012 Practical learning method for multi-scale entangled states *New J. Phys.* **14** 085004

[46] Bridgeman J C and Chubb C T 2017 Hand-waving and interpretive dance: an introductory course on tensor networks *J. Phys. A: Math. Theor.* **50** 223001

[47] Smolin J A, Gambetta J M and Smith G 2012 Efficient method for computing the maximum-likelihood quantum state from measurements with additive Gaussian noise *Phys. Rev. Lett.* **108** 070502

[48] Qi B, Hou Z, Li L, Dong D, Xiang G and Guo G 2013 Quantum state tomography via linear regression estimation *Sci. Rep.* **3** 3496

[49] Kaznady M S and James D F V 2009 Numerical strategies for quantum tomography: alternatives to full optimization *Phys. Rev. A* **79** 022109

[50] Blume-Kohout R 2010 Hedged maximum likelihood quantum state estimation *Phys. Rev. Lett.* **105** 200504

[51] Kueng R and Ferrie C 2015 Near-optimal quantum tomography: estimators and bounds *New J. Phys.* **17** 123013

[52] Doucet A and Johansen A M 2011 A tutorial on particle filtering and smoothing: fifteen years later *The Oxford Handbook of Nonlinear Filtering* ed D Crisan and B Rozovskii (Oxford: Oxford University Press)

[53] Stenberg M P V, Pack K and Wilhelm F K 2015 Adaptive identification of coherent states *Phys. Rev. A* **92** 063852

[54] Granade C E, Ferrie C, Wiebe N and Cory D G 2012 Robust online Hamiltonian learning *New J. Phys.* **14** 103013

[55] Wiebe N, Granade C, Ferrie C and Cory D G 2014 Hamiltonian learning and certification using quantum resources *Phys. Rev. Lett.* **112** 190501

[56] Beskos A, Crisan D and Jasra A 2014 On the stability of sequential Monte Carlo methods in high dimensions *Ann. Appl. Probab.* **24** 1396

[57] Liu J and West M 2001 Combined parameter and state estimation in simulation-based filtering *Sequential Monte Carlo Methods in Practice* ed D Freitas and N Gordon (New York: Springer)

[58] Hou Z, Zhu H, Xiang G-Y, Li C-F and Guo G-C 2016 Achieving quantum precision limit in adaptive qubit state tomography *Npj Quantum Inf.* **2** 16001

[59] Kalev A and Hen I 2015 Fidelity-optimized quantum state estimation *New J. Phys.* **17** 093008

[60] Ferrie C, Granade C E and Cory D G 2013 How to best sample a periodic probability distribution, or on the accuracy of Hamiltonian finding strategies *Quantum Inf. Process.* **12** 611

[61] Stenberg M P V, Köhn O and Wilhelm F K 2016 Characterization of decohering quantum systems: machine learning approach *Phys. Rev. A* **93** 012122

[62] Happ C J and Freyberger M 2008 Adaptive estimation of qubits by symmetry measurements *Phys. Rev. A* **78** 064303

[63] Spall J 1992 Multivariate stochastic approximation using a simultaneous perturbation gradient approximation *IEEE Trans. Autom. Control* **37** 332

[64] Ferrie C and Moussa O 2015 Robust and efficient *in situ* quantum control *Phys. Rev. A* **91** 052306

[65] Granade C E 2015 Characterization, verification and control for large quantum systems *PhD Thesis* University of Waterloo, Ontario

[66] Granade C *et al* 2017 *Quantum* **1** 5

[67] Johansson J R, Nation P D and Nori F 2013 QuTiP 2: a Python framework for the dynamics of open quantum systems *Comput. Phys. Commun.* **184** 1234

[68] van der Walt S, Colbert S C and Varoquaux G 2011 The NumPy array: a structure for efficient numerical computation *Comput. Sci. Eng.* **13** 22

[69] McKinney W 2010 Data structures for statistical computing in Python *Proc. 9th Python in Science Conf.* pp 51–6

[70] Jones E *et al* 2001 SciPy: open source scientific tools for Python (https://www.scipy.org)

[71] Zyczkowski K and Sommers H-J 2001 Induced measures in the space of mixed quantum states *J. Phys. A: Math. Gen.* **34** 7111

[72] Bogdanov Y I, Brida G, Bukeev I D, Genovese M, Kravtsov K S, Kulik S P, Moreva E V, Soloviev A A and Shurupov A P 2011 Statistical estimation of the quality of quantum-tomography protocols *Phys. Rev.* A **84** 042108

[73] Řeháček J, Teo Y S and Hradil Z 2015 Determining which quantum measurement performs better for state estimation *Phys. Rev.* A **92** 012108

[74] Choi M-D 1975 Completely positive linear maps on complex matrices *Linear Algebr. Appl.* **10** 285

[75] Jamiołkowski A 1972 Linear transformations which preserve trace and positive semidefiniteness of operators *Rep. Math. Phys.* **3** 275

[76] Wood C J, Biamonte J D and Cory D G 2015 Tensor networks and graphical calculus for open quantum systems, Quantum *Inf. Comput.* **15** 0579

[77] Okabe M and Ito K 2002 Color Universal Design (CUD): How to make figures and presentations that are friendly to colorblind people (http://jfly.iam.u-tokyo.ac.jp/color/)

[78] Berkson J 1956 Estimation by least squares and by maximum likelihood *Proc. 3rd Berkeley Symp. on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics* (Berkeley, CA: University of California Press)