


Quantum process tomography via completely positive and trace-preserving projection

George C. Knee,^{1,*} Eliot Bolduc,² Jonathan Leach,² and Erik M. Gauger²

¹*Department of Physics, University of Warwick, Coventry CV4 7AL, United Kingdom*

²*SUPA, Institute of Photonics and Quantum Sciences, Heriot-Watt University, David Brewster Building, Edinburgh EH14 4AS, United Kingdom*

 (Received 27 April 2018; published 28 December 2018)

We present an algorithm for projecting superoperators onto the set of completely positive, trace-preserving maps. When combined with gradient descent of a cost function, the procedure results in an algorithm for quantum process tomography: finding the quantum process that best fits a set of sufficient observations. We compare the performance of our algorithm to the diluted iterative algorithm as well as second-order solvers interfaced with the popular CVX package for MATLAB, and find it to be significantly faster and more accurate while guaranteeing a physical estimate.

DOI: [10.1103/PhysRevA.98.062336](https://doi.org/10.1103/PhysRevA.98.062336)

I. INTRODUCTION

As experimental quantum information science progresses, researchers are increasingly turning to the characterization of quantum *processes* [see Fig. 1(a)] rather than the quantum *states* which they act upon. For example, the fault-tolerance theorem [1] (which underpins the feasibility of large-scale quantum computers) places a requirement on the accuracy of operations performed on quantum bits. In fact, the theorem relies on a worst-case analysis with respect to state preparation (before the operation) and with respect to measurement (after the operation) in judging their accuracy. Fortunately, to evaluate such metrics one does not need to actually prepare and measure the worst case: a complete description of the process can be inferred from a suitable set of preparations and measurements and subsequently any property of interest may then be calculated. Such inference is known as quantum process tomography (QPT). QPT has been performed in many different physical settings [2–5]: the challenge is producing an estimate of the process in a reasonable amount of time which matches the data as closely as possible and is consistent with a quantum-mechanical model of the experiment. We present algorithms for reconstructing the process subject to appropriate constraints, via a subroutine that implements a composite projection onto the set of quantum channels [see Fig. 1(b)].

In Sec. II, we describe the maximum-likelihood approach to process tomography, along with the Choi representation of processes and individual projections onto important constraint sets. In Sec. III we describe our main results, including the composite projection onto quantum channels, and two algorithms that exploit it to estimate a quantum process from data. Our algorithms are benchmarked against existing approaches in Sec. IV, and conclusions are drawn in Sec V. Further details are given in the Appendices.

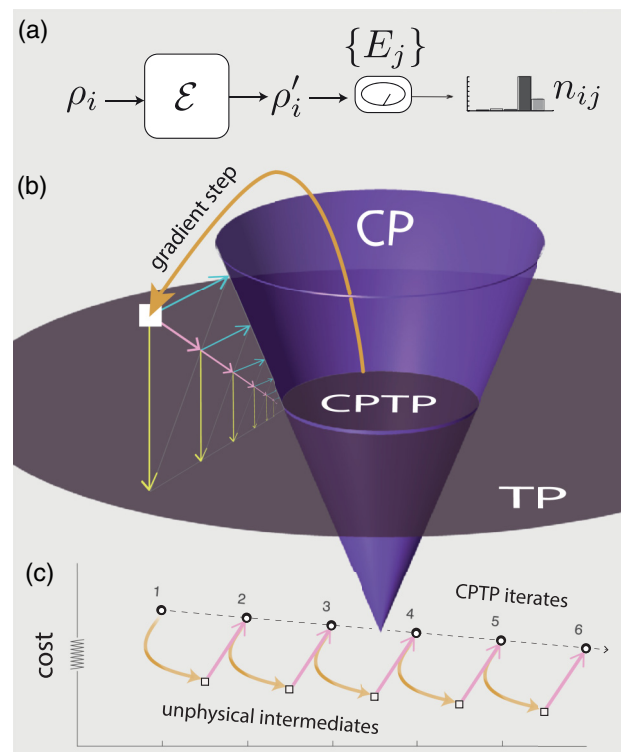


FIG. 1. (a) An unknown process \mathcal{E} maps a known input state ρ_i into an unknown output state ρ'_i . When measured with a set of positive operators $\{E_j\}$, ρ'_i produces a measurement histogram n_{ij} , leading to a cost (e.g., negative likelihood) for each possible estimate for \mathcal{E} . (b) Finding the best fit to the data is a constrained optimization problem that we solve with projected gradient descent. The desired solution set is CPTP: the set of completely positive and trace-preserving maps (the intersection of a cone with a hyperplane). Gradient descent can cause an iterate to exit this set, but it is repeatedly projected back. The schematic shows a single iteration of the algorithm: an unphysical candidate (\square) is iteratively moved to, e.g., the average (pink) of the projection onto CP (cyan) and TP (yellow arrows). (c) The overall procedure reduces the cost function while ensuring iterates are in CPTP.

*gk@physics.org

II. PROCESS TOMOGRAPHY

A quantum process is a linear “superoperator” which acts on an input density operator ρ to produce an output density operator ρ' :

$$\mathcal{E}(\rho) = \rho'. \quad (1)$$

Here ρ and ρ' are order-2 tensors (i.e., operators), which we assume (for simplicity) to act on Hilbert spaces of equal dimension d . \mathcal{E} is an order-4 tensor specified by $d^4 - d^2$ parameters, where d is itself exponential in the number of subsystems. Assumption-free QPT is therefore necessarily very expensive: cheaper alternatives are made possible by relying on prior knowledge (e.g., compressive sensing [6]) or by estimating a single summary statistic (as in randomized benchmarking [7]), but such alternatives may be insufficient to decide whether \mathcal{E} meets fault-tolerance requirements [8,9]. It is therefore important to reduce the costs associated with assumption-free QPT while maintaining accuracy.

A. Estimation of processes

The probability of observing an outcome corresponding to E_j (a positive measurement operator), when the quantum process has transformed some input state ρ_i , is

$$p_{ij} = \text{Tr}[\mathcal{E}(\rho_i)E_j]. \quad (2)$$

For each input state, N_i samples are drawn and the number of outcomes, n_{ij} , that correspond to each outcome j are recorded so that $\sum_j n_{ij} = N_i$. In order to produce an estimate for a quantum process from such a set of observations, a cost function (or measure of “un-fitness”) $f(\mathcal{E})$ is chosen and then minimized over a set of candidate processes. The minimization is handled by, for example, gradient descent steps of the cost function. Maximum-likelihood (ML) estimation is a principled and prominent choice, and one that we make here in order to showcase our general method. The likelihood of \mathcal{E} given the “data” n_{ij} is $\mathcal{L} = \prod_{ij} p_{ij}^{n_{ij}}$ up to an irrelevant constant. According to the principle of ML, we seek \mathcal{E}^* , the quantum process that maximizes \mathcal{L} . This approach is justified heuristically by considering \mathcal{L} as a measure of agreement between the model p_{ij} and the data n_{ij} , assuming a multinomial statistical model. ML enjoys the key property of asymptotic efficiency: as the number of trials, N_i , becomes large, under mild conditions [10] the variance of the ML estimator is lower than any unbiased estimator, saturating the Cramér-Rao bound [11]. Alternative estimators may be preferable for finite N_i [12] but will generally perform worse asymptotically in terms of their precision and accuracy. In the limit where n_{ij} are all large, the likelihood may be approximated by a Gaussian function and the problem reduces to minimizing the cost function $f(\mathcal{E}) \rightarrow \sum_{ij} (p_{ij} - n_{ij})^2$ [13], which may be solved by linear inversion [14]. This is an unjustified simplification in general, but may be transformed into an effective heuristic, as we show below. Regardless of the choice of cost function, our approach is flexible in that it is not tied (as some other methods are [13,15]) to a particular choice of (preparation and/or) measurement operators. Proceeding with the (convex)

cost function

$$f(\mathcal{E}) = -\ln \mathcal{L} = -\sum_{ij} n_{ij} \ln p_{ij}, \quad (3)$$

we introduce below an algorithm minimizing f subject to appropriate constraints. The minimizer of f is also the maximizer of \mathcal{L} due to the monotonicity of the logarithm. Unconstrained tomographic procedures usually produce unphysical quantum states or processes [14], often due to an implicit and idealized assumption of zero noise. A physical estimate is paramount if one is to use it in any further theoretical analysis, such as calculating purity, fidelity, entanglement, expectation values, and so on. We now introduce the constraints on \mathcal{E} and our approach to enforcing them, which uses a certain representation of \mathcal{E} .

B. Choi representation

It is well known that quantum processes, considered as superoperators with dimension $d \times d \times d \times d$, may be represented as $(d^2 \times d^2)$ -dimensional operators on a Hilbert space $\mathcal{H}_{\text{in}} \otimes \mathcal{H}_{\text{out}}$. For example Choi operators, defined as

$$C_{\mathcal{E}} = \sum_{ij} |i\rangle\langle j| \otimes \mathcal{E}(|i\rangle\langle j|) \quad (4)$$

for an orthonormal set of kets $\{|i\rangle\}$ which form a basis of both the input and output Hilbert spaces. This relation implies that \mathcal{E} may be inferred by applying it to one part of a d^2 -dimensional system prepared in a *single* maximally entangled state, and performing quantum state tomography on the output, a technique known as ancilla-assisted QPT [16,17]. Here we continue with the standard approach which uses multiple input states in d dimensions and considers the complete data set as a whole (rather than simply performing state tomography on each output state independently [15]). One can easily verify that $\mathcal{E}(\rho) = \text{Tr}_{\mathcal{H}_{\text{in}}}[(\rho^T \otimes \mathbb{I})C_{\mathcal{E}}]$ which implies that the forward model can be rewritten as

$$p_{ij} = \text{Tr}[(\rho_i^T \otimes E_j)C_{\mathcal{E}}]. \quad (5)$$

Now the cost function has a new argument $f(\mathcal{E}) \rightarrow f(C_{\mathcal{E}})$. The gradient of f with respect to this new argument follows from operator calculus:

$$\nabla f(C_{\mathcal{E}}) = \frac{\partial f(C_{\mathcal{E}})}{\partial C_{\mathcal{E}}} = -\sum_{ij} \frac{n_{ij}}{p_{ij}} \rho_i \otimes E_j^T. \quad (6)$$

These expressions can be vectorized to enable faster evaluation on a computer (see Appendix A).

C. Constraints and projections

Just as some $d \times d$ operators fail to qualify as proper density matrices (quantum states), \mathcal{E} must satisfy certain constraints in order for it to represent a proper quantum process. First, it should be completely positive (CP). This means that it must preserve the positive semidefiniteness of an arbitrary input state when acting on only a part of a larger space of arbitrary dimension: $\rho \geq 0 \Rightarrow [\mathcal{E} \otimes \mathbb{I}_{d'}](\rho) \geq 0 \forall d'$. A theorem due to Choi [18] states that the complete positivity of \mathcal{E} is equivalent to the positive semidefiniteness

of $C_{\mathcal{E}}$ [19]:

$$\mathcal{E} \in \text{CP} \Leftrightarrow C_{\mathcal{E}} \succeq 0. \quad (7)$$

Since CP is a convex set, we can find the closest CP map to any given Hermitian superoperator by using a projection of its Choi matrix onto the set of positive-semidefinite operators. This is achieved via

$$\mathcal{CP}(C_{\mathcal{E}}) = V \max(D, 0) V^{\dagger} \quad (8)$$

with $C_{\mathcal{E}} = V D V^{\dagger}$, i.e., through removing negative terms in the eigendecomposition of $C_{\mathcal{E}}$. \mathcal{CP} is then a projection orthogonal under the Frobenius norm [20]. Second, it is common to assume the process satisfies the property of trace preservation (TP): $\text{Tr}(\mathcal{E}(\rho)) = \text{Tr}(\rho) \forall \rho$. In the Choi operator picture, this is equivalent to the condition

$$\text{Tr}_{\mathcal{H}_{\text{out}}}(C_{\mathcal{E}}) = \mathbb{I}, \quad (9)$$

and the TP set is therefore defined by an affine constraint [21]. In Appendices B and C, we (respectively) explain why the TP constraint makes QPT a nontrivial extension of state tomography, and discuss relaxing the constraint to trace nonincreasing processes (which are probabilistic). We recast Eq. (9) as $M \text{vec}[C_{\mathcal{E}}] = \text{vec}[\mathbb{I}]$, defining

$$M = \sum_k \mathbb{I} \otimes \langle k | \otimes \mathbb{I} \otimes | k \rangle. \quad (10)$$

The $\text{vec}[\cdot]$ operation reshapes its argument into a vector, and we denote the inverse operation by $\text{vec}^{-1}[\cdot]$. The orthogonal projection onto TP is the solution to

$$\begin{aligned} \min_C \frac{1}{2} \|\text{vec}[C] - \text{vec}[C_{\mathcal{E}}]\|^2 \quad \text{such that} \\ M \text{vec}[C] = \text{vec}[\mathbb{I}] \end{aligned} \quad (11)$$

given by

$$\begin{aligned} \mathcal{TP}(C_{\mathcal{E}}) &= \text{vec}^{-1}[\text{vec}[C_{\mathcal{E}}] - M^{\dagger}(M M^{\dagger})^{-1} \\ &\quad \times (M \text{vec}[C_{\mathcal{E}}] - \text{vec}[\mathbb{I}])] \\ &= \text{vec}^{-1}[\text{vec}[C_{\mathcal{E}}] - \frac{1}{d} M^{\dagger} M \text{vec}[C_{\mathcal{E}}] \\ &\quad + \frac{1}{d} M^{\dagger} \text{vec}[\mathbb{I}]]. \end{aligned} \quad (12)$$

In the second line, we used $M M^{\dagger} = d \mathbb{I} \otimes \mathbb{I}$; it is useful to note that $M^{\dagger} M = \sum_{ij} \mathbb{I} \otimes |j\rangle\langle i| \otimes \mathbb{I} \otimes |j\rangle\langle i|$.

III. COMPOSITE PROJECTION

It is not obvious that the ability to separately project into the CP and TP sets of superoperators enables projection into the intersection set CPTP. While \mathcal{CP} and \mathcal{TP} are separately projections, their sequential action will not generally result in a matrix belonging to CPTP (unlike the equivalent sequential actions for projection onto quantum states [22]). This may be intuited with the help of the geometrical picture in Fig. 1(b). Repeated averaged projections, however, defined (at iteration k of the outer loop of our algorithm) by

$$H^0 = C_{\mathcal{E}}^k, \quad H^{l+1} = \frac{1}{2}(\mathcal{TP}(H^l) + \mathcal{CP}(H^l)), \quad (13)$$

will converge (in linear time [23]) to a point in CPTP [see Fig. 1(b)]. Dykstra's alternating projection algorithm is a

superior alternative, which we choose because it achieves projection $\mathcal{CPTP}(C_{\mathcal{E}})$ onto the closest point in the intersection of the sets [24]. A similar method was used in Ref. [25] to solve a feasibility problem where input and output states are given and an exact but nonunique solution is sought. By contrast the ML approach (which we adopt) finds the best possible fit to some given, noisy data and treats the more realistic situation where an exact solution cannot be expected. Pseudocode for this subroutine (showing our use of a robust stopping criterion due to Birgin and Raydan [24]) is given below.

Algorithm 1 \mathcal{CPTP} projection subroutine

```

1: Input:  $C$ 
2: Output:  $C_{\perp} \in \text{CPTP}$  such that
    $\|C - C_{\perp}\|_2^2 \leq \|C - B\|_2^2 \quad \forall B \in \text{CPTP}$ 
3: Set:  $p_0 = q_0 = y_0 = 0, k = 0, x_0 = \text{vec}[C]$ 
4: while  $\|p_{k-1} - p_k\|^2 + \|q_{k-1} - q_k\|^2 + |2p_{k-1}^{\dagger}(x_k - x_{k-1})| +$ 
    $|2q_{k-1}^{\dagger}(y_k - y_{k-1})| > 10^{-4}$  do
5:    $y_k = \text{vec}[\mathcal{TP}[\text{vec}^{-1}[x_k + p_k]]]$ 
6:    $p_{k+1} = x_k + p_k - y_k$ 
7:    $x_{k+1} = \text{vec}[\mathcal{CP}[\text{vec}^{-1}[y_k + q_k]]]$ 
8:    $q_{k+1} = y_k + q_k - x_{k+1}$ 
9:    $k = k + 1;$ 
10: Return  $C_{\perp} = \text{vec}^{-1}[x_{k+1}]$ 

```

A. Proposed algorithm: pgdB

Returning to the main problem of QPT, we are now ready to apply the principle of projected gradient descent. Such an approach has recently been shown to offer speed benefits when applied to quantum state tomography [22,26,27]. Applying it to QPT implies taking a single gradient descent step before running an alternating projection [28] under \mathcal{CP} and \mathcal{TP} until this ‘‘inner loop’’ converges, before taking another gradient descent step and repeating. Formally the ML-QPT problem is written

$$\begin{aligned} \underset{C_{\mathcal{E}} \in \mathbb{C}^{d^2 \times d^2}}{\text{minimize}} \quad & f = - \sum_{ij} n_{ij} \ln(\text{Tr}([\rho_i^T \otimes E_j] C_{\mathcal{E}})) \\ \text{subject to} \quad & C_{\mathcal{E}} \succeq 0 \quad (\text{CP constraint}), \\ & \text{Tr}_{\mathcal{H}_{\text{out}}}(C_{\mathcal{E}}) = \mathbb{I}_{d \times d} \quad (\text{TP constraint}), \end{aligned} \quad (14)$$

which our proposed algorithm solves via the iterative update rule:

$$\begin{aligned} C_{\mathcal{E}}^0 &= \mathbb{I}_{d^2 \times d^2} / d, \\ C_{\mathcal{E}}^{k+1} &= \mathcal{CPTP} \left(C_{\mathcal{E}}^k - \frac{1}{\mu} \nabla f(C_{\mathcal{E}}^k) \right). \end{aligned} \quad (15)$$

Here μ is a step-size metaparameter. At each iteration a step is taken in the locally downhill direction of f , with the aim of increasing the likelihood, before a projection ensures that the constraints are satisfied (up to any desired tolerance) [see Fig. 1(c)]. Pseudocode of the full algorithm pgdB is given below and shows our implementation of backtracking (Armijo line search) to improve convergence times [20,29,30].

Algorithm 2 pgdB

```

1:  $k = 0$ 
2: Initial estimate:  $C_{\mathcal{E}}^0 = \mathbb{I}_{d^2 \times d^2} / d$ 
3: Set metaparameters:  $\mu = 3/(2d^2)$ ,  $\gamma = 0.3$ 
4: while  $f(C_{\mathcal{E}}^k) - f(C_{\mathcal{E}}^{k+1}) > 1 \times 10^{-10}$  do
5:    $D^k = \mathcal{CPTP}[C_{\mathcal{E}}^k - \frac{1}{\mu} \nabla f(C_{\mathcal{E}}^k)] - C_{\mathcal{E}}^k$ 
6:    $\alpha = 1$ 
7:   while  $f(C_{\mathcal{E}}^k + \alpha D^k) > f(C_{\mathcal{E}}^k) + \gamma \alpha \langle D^k, \nabla f(C_{\mathcal{E}}^k) \rangle$  do
8:      $\alpha = 0.5\alpha$ 
9:    $C_{\mathcal{E}}^{k+1} = C_{\mathcal{E}}^k + \alpha D^k$ 
10:   $k = k + 1$ 
11: Return  $C_{\mathcal{E}_{\text{est}}} = C_{\mathcal{E}}^{k+1}$ 

```

It is important to point out that, although by our choice of f the considered problem is convex (meaning there are no local optima which are not global optima), there may nevertheless be regions (where $p_{ij} = 0$ for some i, j) where f has an undefined gradient and where the likelihood is stationary [31]. These subspaces correspond to those processes which take ρ_i to a state orthogonal to E_j . The probability of a randomly chosen process lying within these subdimensional regions is low. However, an iteration of our algorithm might encounter such a region, causing the algorithm to stall: this is especially the case since projections tend to result in rank deficient matrices. Our solution to this issue is presented in Appendix D.

Note that our approach can be adapted to solve other constrained optimization problems over quantum processes: whenever the cost function f is convex and continuously differentiable the projected gradient descent procedure is guaranteed to converge to an optimal solution [32].

B. Linear inversion with a final projection (LIFP)

To show that our CPTP projection has applications in QPT other than for ML, we present a further algorithm based on linear inversion of a vectorized form of Eq. (5), $\text{vec}[p_{ij}] = \text{Avec}[C_{\mathcal{E}}]$, where the rows of A are composed of $\text{vec}[\rho_i \otimes E_j^T]^T$. To allow for a comparison with the manifestly physical approach of the ML methods, we upgrade naïve linear inversion so as to force it to also produce physical estimates—simply by applying our composite CPTP projection just once. We call the resultant estimator linear inversion with a final projection (LIFP):

$$C_{\mathcal{E}}^{\text{LIFP}} := \mathcal{CPTP}(\text{vec}^{-1}[A^+ n]), \quad (16)$$

where $A^+ = A^\dagger(AA^\dagger)^{-1}$ is the pseudoinverse of A [14] and $n = \text{vec}[n_{ij}]$. This approach is in the same spirit as the “quick and dirty” state tomography approach used by Kaznady and James [33], suggested as an alternative to full-blown optimization. For full details, see Appendix E.

IV. NUMERICAL BENCHMARKING

Unlike other approaches that require an orthonormal operator basis [13], ours is flexible enough to deal with any informationally complete set of preparations and measurements. To test the performance of our proposed algorithm, we considered the reconstruction of qudit processes, with the d^2

preparations $\rho_i = |\psi_i\rangle\langle\psi_i|$ projecting onto the following pure states:

$$|\psi_i\rangle = \begin{cases} |j\rangle, & j = 1, \dots, d \\ \frac{|j\rangle + |k\rangle}{\sqrt{2}}, & j = 1, \dots, d, k > j \\ \frac{|j\rangle + \sqrt{-1}|k\rangle}{\sqrt{2}}, & j = 1, \dots, d, k > j, \end{cases} \quad (17)$$

and a positive operator valued measure (POVM) formed with the $2d^2$ elements

$$E_j = \begin{cases} \rho_i / d^2, & i = 1, \dots, d^2 \\ \mathbb{I} / d^2 - \rho_i / d^2, & i = 1, \dots, d^2, \end{cases} \quad (18)$$

which resolve to the identity $\sum_j E_j = \mathbb{I}$. This choice of preparations and measurements is a very practical one, when compared with the use of symmetric informationally complete POVMs (SIC-POVMs) [34] or ancillas [16,17] which require much greater experimental effort. Appendix F shows how the choice of measurement and preparation operators affects the conditioning of the optimization problem. We ignore systematic errors in state preparation and measurement (so-called SPAM errors): gate set tomography [9,35,36] is a recently proposed solution for a self-consistent characterization in the presence of such imperfections.

We generated random, purity $>90\%$ CPTP maps $C_{\mathcal{E}_{\text{true}}}$ in $d = 2, \dots, 8$, simulating multinomial random data with sample size $N_i = N$ ranging from 10 to 10^9 (see Appendix H). Throughout our simulations we normalize the data such that $\sum_j n_{ij} = 1$. Although this of little consequence for ML and other optimization approaches (the cost function is merely scaled), it is an important step for linear inversion and LIFP, since the latter approach is based on treating the data as probabilities—hence the need to normalize them into frequencies rather than raw counts.

We also simulated the infinite-data case by setting the frequencies equal to the theoretical probabilities $N \rightarrow \infty \Rightarrow n_{ij} = p_{ij}$. Although not a realistic situation this is a useful device to compare various iterative QPT algorithms as the limit of many observations is approached, where the reconstruction error is dominated by numerical errors rather than noise in the data. We then ran our algorithm (as well as benchmark algorithms) to find C_{est} . We recorded running times and accuracies for each. The figure of merit for accuracy was taken to be the J distance, defined as

$$J(\mathcal{E}_{\text{est}}, \mathcal{E}_{\text{true}}) = \|\mathcal{C}_{\mathcal{E}_{\text{est}}} - \mathcal{C}_{\mathcal{E}_{\text{true}}}\|_{\text{tr}} / 2d. \quad (19)$$

Here $\|\cdot\|_{\text{tr}}$ is the trace norm, or sum of singular values, and $J(\mathcal{E}_{\text{est}}, \mathcal{E}_{\text{true}}) \in [0, 1]$ is related to the average probability of distinguishing the two processes [37]. Simulations were run using MATLAB R2017a and CVX v2.1 on an Intel Core i7-4790 3.6 GHz with 8 MB L3 cache.

A. Benchmark 1: Diluted iterations

The state-of-the-art algorithm for ML-QPT is the technique of diluted iterations (DIA), devised by Fiurasek and Hradil in 2001 [38], who adapted an algorithm for quantum state tomography known as diluted $R\rho R$ [39]. The main idea is to exploit an extremal equation obeyed by the ML-CPTP map:

$$C_{\mathcal{E}} = \Lambda^{-1}(C_{\mathcal{E}}) \nabla f(C_{\mathcal{E}}) C_{\mathcal{E}} \nabla f(C_{\mathcal{E}}) \Lambda^{-1}(C_{\mathcal{E}}), \quad (20)$$

where $\Lambda(C_\mathcal{E}) = (\text{Tr}_{\mathcal{H}_{\text{out}}}(C_\mathcal{E}))^{1/2} \otimes \mathbb{I}$ is derived by incorporating the TP constraint through a Lagrange multiplier. This equation is treated as an iterative update rule. The algorithm has been applied to real data, mostly with Fock-space truncated optical (i.e., continuous variable) systems, according to the formulation by Anis and Lvovsky [40]. Cooper *et al.* reported that the algorithm ran for 6.5 h (with a machine precision stopping rule) to reconstruct a $d = 6$ conditional state engineering process [41]. Fedorov *et al.* performed QPT of a beam splitter, revealing the Hong-Ou-Mandel effect [42]. The algorithm typically requires *diluting* with a parameter $0 < \epsilon \leq 1$ to prevent f from oscillating:

$$\begin{aligned} R^k &= \epsilon \nabla f(C_\mathcal{E}^k) + (1 - \epsilon)\mathbb{I}, \\ \Lambda^k &= (\text{Tr}_{\mathcal{H}_{\text{out}}}(R^k C_\mathcal{E}^k R^k))^{1/2} \otimes \mathbb{I}, \\ C_\mathcal{E}^{k+1} &= (\Lambda^k)^{-1} R^k C_\mathcal{E}^k R^k (\Lambda^k)^{-1}. \end{aligned} \quad (21)$$

In our application of DIA, we used a crude optimization of ϵ at each iteration to prevent overshoots. Pseudocode is shown below:

Algorithm 3 DIA

```

1:  $k = 0$ 
2: Initial estimate:  $C_0 = \mathbb{I}_{d^2 \times d^2} / d$ 
3: while  $f(C_\mathcal{E}^k) - f(C_\mathcal{E}^{k+1}) > 1 \times 10^{-10}$  do
4:    $\epsilon = 1$ 
5:    $R^k = \epsilon \nabla f(C_\mathcal{E}^k) + (1 - \epsilon)\mathbb{I}$ 
6:    $\Lambda^k = (\text{Tr}_{\mathcal{H}_{\text{out}}}(R^k C_\mathcal{E}^k R^k))^{1/2} \otimes \mathbb{I}$ 
7:   while  $f((\Lambda^k)^{-1} R^k C_\mathcal{E}^k R^k (\Lambda^k)^{-1}) > f(C_\mathcal{E}^k)$  do
8:      $\epsilon = \epsilon / 2$ ;
9:      $R^k = \epsilon \nabla f(C_\mathcal{E}^k) + (1 - \epsilon)\mathbb{I}$ 
10:     $\Lambda^k = (\text{Tr}_{\mathcal{H}_{\text{out}}}(R^k C_\mathcal{E}^k R^k))^{1/2} \otimes \mathbb{I}$ 
11:     $C_\mathcal{E}^{k+1} = (\Lambda^k)^{-1} R^k C_\mathcal{E}^k R^k (\Lambda^k)^{-1}$ 
12:     $k = k + 1$ 
13: Return  $C_{\text{est}} = C_\mathcal{E}^{k+1}$ 

```

B. Benchmark 2: MOSEK

The problem (14) is straightforward to enter into the CVX modeling environment [43], which can then call one of a number of general-purpose solvers. CVX solves problems featuring logarithmic cost functions with a successive approximation method. We found SEDUMI failed to find a solution and that SDPT3 succeeded but was slow. We therefore settled on the commercial solver MOSEK: it usually produced good results but sometimes failed in higher dimensions. We estimate the failure probability to be as high as 15% (see Appendix G).

V. RESULTS AND CONCLUSIONS

Our results are summarized in Fig. 2: the CPTP-projection-based algorithms are not only faster than existing approaches, but significantly more accurate. This is especially the case as d increases but is true even for $d = 4$ (corresponding, for example, to a two-qubit gate such as the controlled-NOT).

We confirmed that due to the iterative nature of pgdB, higher accuracies can be achieved by adjusting the stopping criterion and other metaparameters, at the expense of a longer algorithm runtime. LIFP is a quick and effective method

suitable to situations with low noise and low d , and sacrificing a small amount of accuracy, typically less than an order of magnitude, with respect to the iterative methods. Appendix H includes results corresponding to reconstruction of full rank processes, which show qualitatively similar trends.

Our results indicate that CPTP projection is a valuable and versatile tool for QPT, holding great promise for applications where time and accuracy are both important [44].

ACKNOWLEDGMENTS

G.C.K. was supported by the Royal Commission for the Exhibition of 1851 and thanks Yanbao Zhang for discussions. E.M.G. thanks the Royal Society of Edinburgh and the Scottish government for support. We thank Yen-Huan Li for useful feedback on this document.

APPENDIX A: VECTORIZING THE PROBLEM

Our forward model may be vectorized as $p := \text{vec}[p_{ij}] = \text{Avec}[C_\mathcal{E}]$ upon defining

$$A = \begin{pmatrix} \text{vec}[\rho_1^* \otimes E_1^\dagger]^\dagger \\ \vdots \\ \text{vec}[\rho_1^* \otimes E_N^\dagger]^\dagger \\ \vdots \\ \text{vec}[\rho_N^* \otimes E_N^\dagger]^\dagger \end{pmatrix} = \begin{pmatrix} \text{vec}[\rho_1 \otimes E_1^T]^T \\ \vdots \\ \text{vec}[\rho_1 \otimes E_N^T]^T \\ \vdots \\ \text{vec}[\rho_N \otimes E_N^T]^T \end{pmatrix}. \quad (\text{A1})$$

Now $f(C_\mathcal{E}) = -n^T \ln(p)$ and $\nabla f(C_\mathcal{E}) = -A^\dagger \eta$ with $\eta_{ij} = n_{ij} / p_{ij}$ and $\eta = \text{vec}[\eta_{ij}]$ and so on. These facts follow from elementary matrix calculus along with the identity $\text{Tr}(A^\dagger B) \equiv \text{vec}[A]^\dagger \text{vec}[B]$. Shang *et al.* present a method to speed up calculation of the gradient when A has a tensor product structure [27]: unfortunately this does not help in our case because we have only a single tensor product $\rho_i^T \otimes E_j$ and multiple tensor products are needed to show an advantage.

In a slight abuse of notation, above we have denoted the vectorization of a matrix M by $\text{vec}[M_{ij}]$, where M_{ij} are the matrix elements of M .

APPENDIX B: TRACE PRESERVATION CONSTRAINT

Counter to a common misconception, the celebrated ‘‘Choi-Jamiolkowski isomorphism’’ [45] does not imply a 1:1 correspondence between CPTP maps and density operators $C_\mathcal{E} / \text{Tr}(C_\mathcal{E})$ in a higher-dimensional space. As discussed in the main paper, some such density operators do not correspond to CPTP maps [45] (see, for example, Refs [46,47]). As an example of why a stronger condition is important, consider the Choi matrix $C_\square = \text{diag}(0.1, 0.1, 0.1, 1.7)$. This is proportional to a perfectly valid, classically correlated two-qubit state. It is manifestly positive, and has trace equal to 2. Note, however, that $\text{Tr}_2(C_\square) = \text{diag}(0.2, 1.8) \neq \mathbb{I}$. This means that it does not represent a trace-preserving map. Note that the given example actually increases the trace of some states and decreases the trace of others so it is neither in TP nor in the set of trace nonincreasing processes (TNI, discussed below).

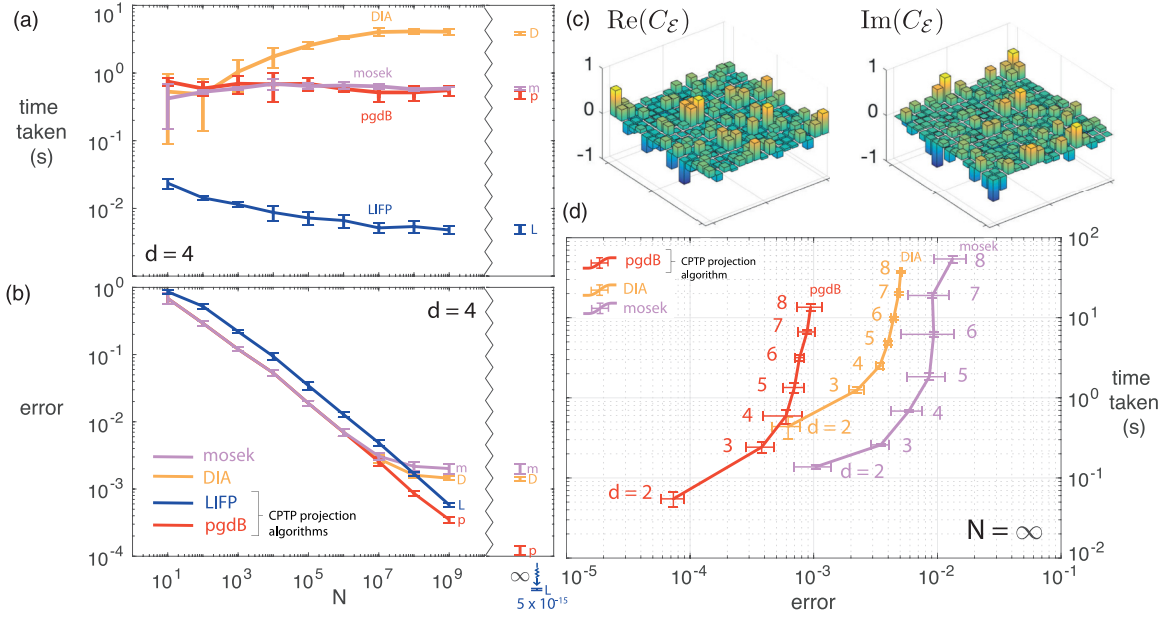


FIG. 2. (a) Our proposed algorithm pgdB (p, dark red) is significantly faster than the state-of-the-art ML-QPT algorithm DIA (D, light orange) when $N \gtrsim 10^3$. Our other proposed algorithm LIFP (L, dark blue) is faster still, due to its noniterative nature. (b) Reconstruction error (J distance) decreases as the number of observations, N , increases. All of the iterative ML algorithms produce the same quality estimate until $N \gtrsim 10^6$: as statistical noise is made smaller the accuracy of the algorithms deviates, with our proposed algorithms producing much better estimates. We could not simulate arbitrarily high N multinomial noise, but can effectively consider the $N \rightarrow \infty$ case by setting $n_{ij} = p_{ij}$ ($d = 4$ case shown). LIFP does not find the ML solution and, therefore, sacrifices accuracy in all but the highly idealized case of infinite counts. (c) Real and imaginary parts of a typical, purity $>90\%$, $d = 4$ Choi matrix returned by our algorithm. (d) Accuracy and run time for all iterative ML algorithms as d increases, with negligible statistical noise: pgdB is faster and more accurate.

APPENDIX C: PROJECTION ONTO THE SET OF TRACE NONINCREASING PROCESSES

It will be possible to relax our algorithmic projections in such a way as to search in larger spaces, i.e., supersets of TP. This may be important for certain applications: in fact, whenever there is loss or some other nondeterministic process such as a measurement, CPTP maps (or “channels”) are replaced by CPTNI maps (or “operations”) which are *trace nonincreasing*. The trace of the output density matrix corresponds to the probability of success of the map. CP maps admit a Kraus representation [18]:

$$\mathcal{E}(\rho) = \sum_i K_i \rho K_i^\dagger. \quad (\text{C1})$$

The difference in trace between an input and an output state is given by

$$\begin{aligned} \text{Tr}(\mathcal{E}(\rho)) - \text{Tr}(\rho) &= \text{Tr}\left(\sum_i K_i \rho K_i^\dagger\right) - \text{Tr}(\rho) \\ &= \text{Tr}([\mathcal{Y} - \mathbb{I}]\rho), \end{aligned} \quad (\text{C2})$$

where we used the cyclic property of the trace and defined $\mathcal{Y} = \sum_i K_i^\dagger K_i$. Since ρ is an arbitrary positive operator, it is clear that trace preservation is equivalent to

$$\mathcal{Y} = \mathbb{I} \quad (\text{TP}). \quad (\text{C3})$$

In fact, it is always possible to diagonalize \mathcal{Y} with a similarity transformation \mathcal{V} . Let λ_i be the eigenvalues of \mathcal{Y} [48]. Then

$$\text{Tr}(\mathcal{E}(\rho)) - \text{Tr}(\rho) = \text{Tr}(\text{diag}[\lambda_1 - 1, \lambda_2 - 1, \dots]\mathcal{V}^\dagger \rho \mathcal{V}). \quad (\text{C4})$$

So a general CP map will change the trace of an input state depending on the eigenvalues of \mathcal{Y} , and on the projection of the state onto the eigenvectors of \mathcal{Y} . We can consider the case where the process has a uniform and known success probability p :

$$\mathcal{Y} = p\mathbb{I} \quad (\text{US}_p). \quad (\text{C5})$$

The projection onto this set is easy: simply alter the TP projection given in Eq. (12) by taking $\text{vec}[\mathbb{I}] \rightarrow p\text{vec}[\mathbb{I}]$. If we demand that the trace is nonincreasing for an arbitrary input state, we have

$$\mathcal{Y} \leq \mathbb{I} \quad (\text{TNI}). \quad (\text{C6})$$

A commonly used approach to reconstruct TNI operations involves introducing of a fictitious “shelving” state (thereby extending the space $d \rightarrow d + 1$ [40]). One then performs a TP reconstruction in the larger space before projecting out the fictitious state. We present an alternative here that avoids (the possibly very expensive downside of) having to increase the size of the Hilbert space. By the definition of the Choi matrix,

$$\text{Tr}_{\mathcal{H}_{\text{out}}}(C_{\mathcal{E}}) = \mathcal{Y}. \quad (\text{C7})$$

It is straightforward to project onto TNI (which is a convex superset of TP). The correlations χ in the Choi matrix are

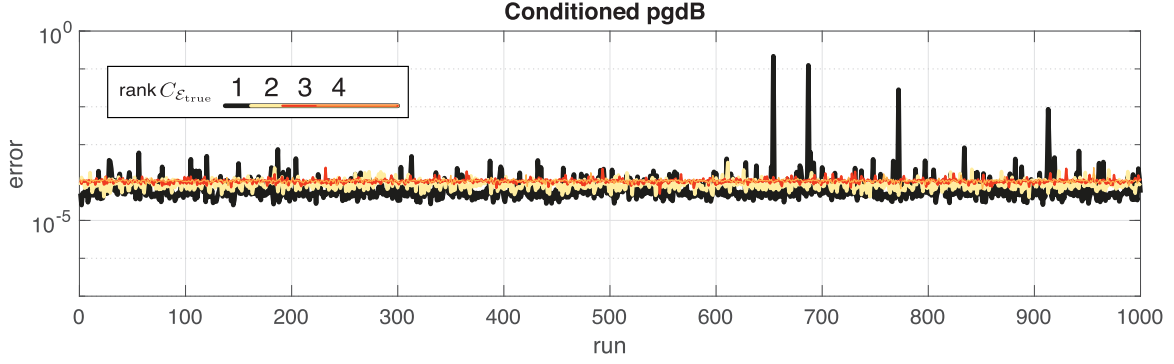


FIG. 3. Our conditioning step means the pgdB algorithm is robust to reconstruction of rank-deficient processes, even with a multinomial likelihood function. The algorithm never stalls but in rare cases the accuracy is adversely affected. Here $d = 2$ and $\epsilon = 1 \times 10^{-16}$.

implicitly defined through

$$C_{\mathcal{E}} = \text{Tr}_{\mathcal{H}_{\text{out}}}(C_{\mathcal{E}}) \otimes \mathbb{I} + \mathbb{I} \otimes \text{Tr}_{\mathcal{H}_{\text{in}}}(C_{\mathcal{E}}) + \chi. \quad (\text{C8})$$

We can achieve the TNI projection via

$$\mathcal{TN}\mathcal{I}(C_{\mathcal{E}}) = C_{\mathcal{E}} + \frac{1}{d} [\mathcal{G}(\text{Tr}_{\mathcal{H}_{\text{out}}}(C_{\mathcal{E}})) - \text{Tr}_{\mathcal{H}_{\text{out}}}(C_{\mathcal{E}})] \otimes \mathbb{I}, \quad (\text{C9})$$

where we introduced \mathcal{G} , defined and computed via

$$\mathcal{G}(X) = \arg \min_{\text{Tr}_{\mathcal{H}_{\text{out}}}(B) \leq \mathbb{I}} \|\text{Tr}_{\mathcal{H}_{\text{out}}}(B) \otimes \mathbb{I} - X \otimes \mathbb{I}\| \quad (\text{C10})$$

$$= \arg \min_{\text{Tr}_{\mathcal{H}_{\text{out}}}(B) \leq \mathbb{I}} \|\text{Tr}_{\mathcal{H}_{\text{out}}}(B) - X\| \cdot \|\mathbb{I}\| \quad (\text{C11})$$

$$= \arg \min_{\text{Tr}_{\mathcal{H}_{\text{out}}}(B) \leq \mathbb{I}} \|\text{Tr}_{\mathcal{H}_{\text{out}}}(B) - X\| \quad (\text{C12})$$

$$= \arg \min_{E \leq \mathbb{I}} \|E - X\| \quad (\text{C13})$$

$$= V \min(D, 1) V^{\dagger} \quad (\text{C14})$$

with $X = V D V^{\dagger}$ [49]. To see that we have the orthogonal projection, consider that for any $B \in \text{TNI}$ we may write $B = \text{Tr}_{\mathcal{H}_{\text{out}}}(B) \otimes \mathbb{I} + \mathbb{I} \otimes \text{Tr}_{\mathcal{H}_{\text{in}}}(B) + \chi_B$. Now

$$\|B - \mathcal{E}\| = \|\text{Tr}_{\mathcal{H}_{\text{out}}}(B - C_{\mathcal{E}}) \otimes \mathbb{I} + \mathbb{I} \otimes \text{Tr}_{\mathcal{H}_{\text{in}}}(B - C_{\mathcal{E}}) + \chi_B - \chi\| \quad (\text{C15})$$

$$\leq \|\text{Tr}_{\mathcal{H}_{\text{out}}}(B - C_{\mathcal{E}}) \otimes \mathbb{I}\| + \|\mathbb{I} \otimes \text{Tr}_{\mathcal{H}_{\text{in}}}(B - C_{\mathcal{E}})\| + \|\chi_B - \chi\| \quad (\text{C16})$$

$$\leq \|\text{Tr}_{\mathcal{H}_{\text{out}}}(B - C_{\mathcal{E}}) \otimes \mathbb{I}\| \leq \min_{\text{Tr}_{\mathcal{H}_{\text{out}}}(B) \leq \mathbb{I}} \|\text{Tr}_{\mathcal{H}_{\text{out}}}(B) \otimes \mathbb{I} - \text{Tr}_{\mathcal{H}_{\text{out}}}(C_{\mathcal{E}}) \otimes \mathbb{I}\| \quad (\text{C17})$$

$$= \min_{\text{Tr}_{\mathcal{H}_{\text{out}}}(B) \leq \mathbb{I}} \|\text{Tr}_{\mathcal{H}_{\text{out}}}(B) - \text{Tr}_{\mathcal{H}_{\text{out}}}(C_{\mathcal{E}})\| \cdot \|\mathbb{I}\| \quad (\text{C18})$$

$$= \min_{\text{Tr}_{\mathcal{H}_{\text{out}}}(B) \leq \mathbb{I}} \frac{1}{d} \|\text{Tr}_{\mathcal{H}_{\text{out}}}(B) - \text{Tr}_{\mathcal{H}_{\text{out}}}(C_{\mathcal{E}})\| \quad (\text{C19})$$

$$= \|\mathcal{TN}\mathcal{I}(C_{\mathcal{E}}) - C_{\mathcal{E}}\|,$$

and therefore $\mathcal{TN}\mathcal{I}(C_{\mathcal{E}})$ is the closest element in TNI to $C_{\mathcal{E}}$. We verified this numerically using CVX. Using the techniques shown in the main text, this result enables tomography of

quantum operations belonging to CPTNI without expanding the reconstruction space.

APPENDIX D: STALLING

As mentioned in the main text, it is possible for pgdB to stall if a situation with $p_{ij} = 0$ is encountered, and in this section we discuss a simple solution allowing the algorithm to continue. Importantly, our proposed solution alerts the user to when the stalling fix has been applied, turning such a rare occurrence into a heralded event. It is important to note that other potential solutions also exist, but a comprehensive study of their merits and drawbacks (including bounding any systemic inaccuracies that arise) are beyond the scope of the present work.

One straightforward solution to keeping the algorithm running when a $p_{ij} = 0$ situation arises is to modify the cost function and gradient at each iteration with the heralded conditioning step:

$$p_{ij} = \max(p_{ij}, \epsilon) \quad (\text{D1})$$

for ϵ some small parameter. It is clearly possible that the algorithm never encounters $p_{ij} < \epsilon$; in that case our conditional modification does not occur and we proceed toward the ML solution. In the case that the modification is triggered, the algorithm is capable of raising a flag, warning the user to interpret the resultant solution accordingly; it would also be possible for the flag to trigger a restart of the algorithm with a random initial guess. In theory, the ML solution cannot have any $p_{ij} = 0$ (since that is a zero of the likelihood, which is positive semidefinite). The exception is when $p_{ij} = n_{ij} = 0$, trivial cases excluded by taking the convention $0^0 = 1$. Therefore,

$$p_{ij} = \begin{cases} p_{ij}, & p_{ij} > 0 \\ \epsilon, & p_{ij} = 0 \end{cases} \quad (\text{D2})$$

is sufficient to keep the algorithm going without altering the turning points of the objective function. In practice, machines with finite precision cannot calculate the logarithm of very small numbers, necessitating the stronger conditioning step given in Eq. (D1). The implication is that, while the ML solution cannot be at $p_{ij} = 0$, it can have $p_{ij} < \epsilon$. Therefore, with low probability the conditioning step can affect the accuracy of the reconstructed solution (but this possibility is heralded). The probability of this occurring is related to the

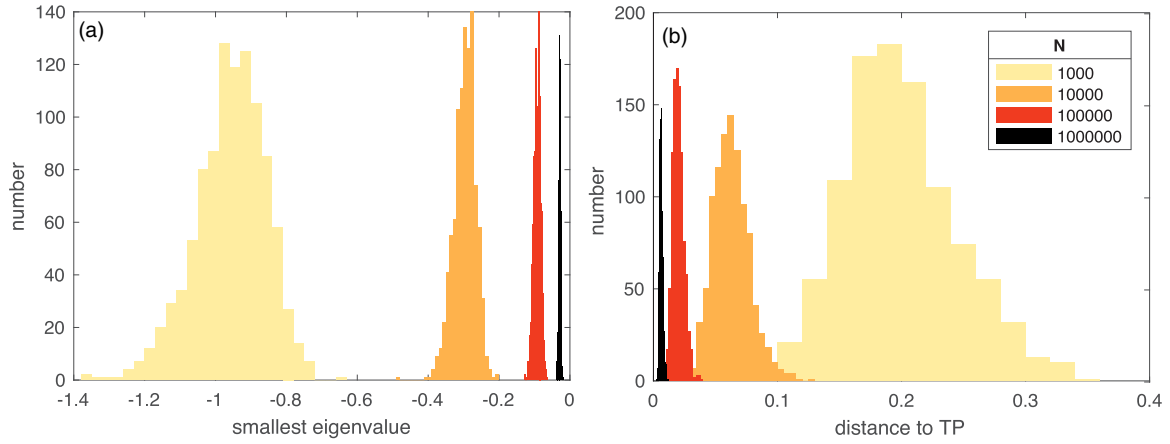


FIG. 4. Distribution of (a) the smallest eigenvalue of C_{ξ}^{LI} and (b) distance $\|\text{Tr}_2(C_{\xi}^{LI}) - \mathbb{I}\|_2$ to the TP set under various degrees of multinomial shot noise in $d = 4$ with an ensemble of full-rank CPTP maps. As the number of clicks, N , increases, the typical magnitude of the smallest eigenvalue decreases, as does the distance to TP. Unphysical estimates (e.g., where the smallest eigenvalue is negative) occur in almost all cases (1000 sampled for each value of N).

Kraus rank of the process (see Fig. 3, which shows that the conditioning step prevents the stalling problem with no impact on the reconstruction error in most cases).

APPENDIX E: LINEAR INVERSION WITH A FINAL PROJECTION ONTO CPTP

The most straightforward approach to inverting the forward model is arguably the approach of linear inversion. Here, the

estimate is

$$C_{\xi}^{LI} := \text{vec}^{-1}[A^{+}n], \quad (\text{E1})$$

where $A^{+} = A^{\dagger}(AA^{\dagger})^{-1}$ is the pseudoinverse of A [14], and A is defined in Eq. (A1). This estimator minimizes the distance in Frobenius norm between the data and the forward model—a familiar notion of “least-squares” fitting. This notion of discrepancy is distinct from the negative likelihood used in the main paper. Note, however, that in the special case where the likelihood is a Gaussian function, or indeed when there is no noise at all, the two notions of discrepancy coincide [13]. Linear inversion is a noniterative approach so it is likely to be fast in low dimensions, although the number of floating point operations will be $O(d^{12})$ and will therefore not scale as well as pgdB [which is $O(d^8)$]. Furthermore, without the explicit expression of the CP and TP constraints, there is no guarantee of a physical estimate.

We implemented this estimator using the MATLAB backslash operator \backslash , which avoids explicit calculation of the

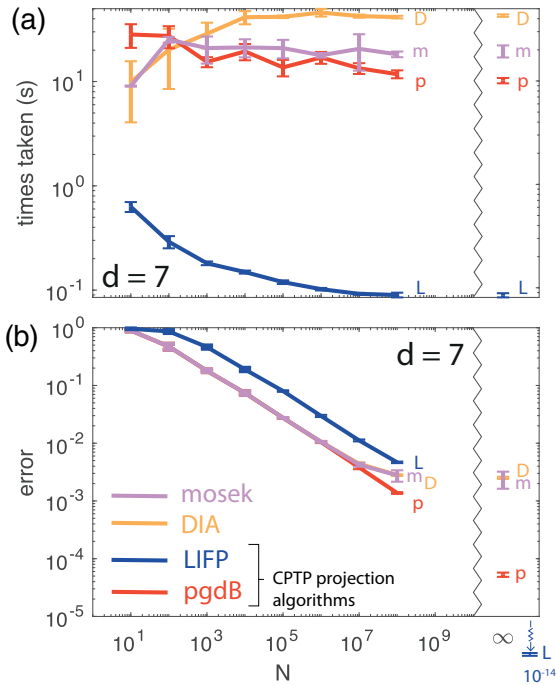


FIG. 5. Performance (in terms of (a) time and (b) accuracy) of linear inversion with a final projection (LIFP) compared to the ML techniques, for $d = 7$ quasipure processes. When there is no noise whatsoever $N \rightarrow \infty$, the final projection is not necessary, and the error is very low (consistent with the product of the machine precision with the condition number of A). In all other cases shown, a significant drop in precision may be seen.

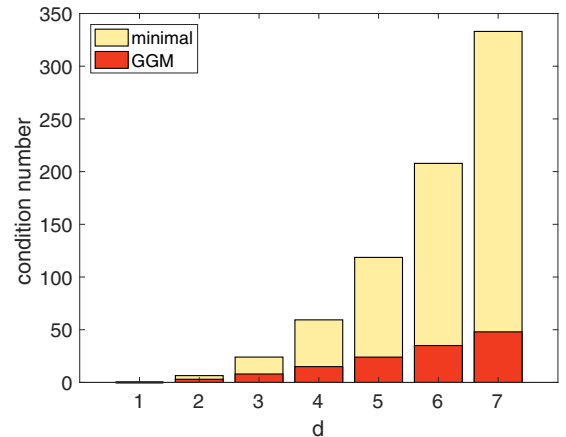


FIG. 6. Condition number of A as a function of d for our chosen “minimal” set of preparations and measurements (blue) and for a set constructed from all eigenvectors of the generalized Gell-Mann matrices (GGM, red).

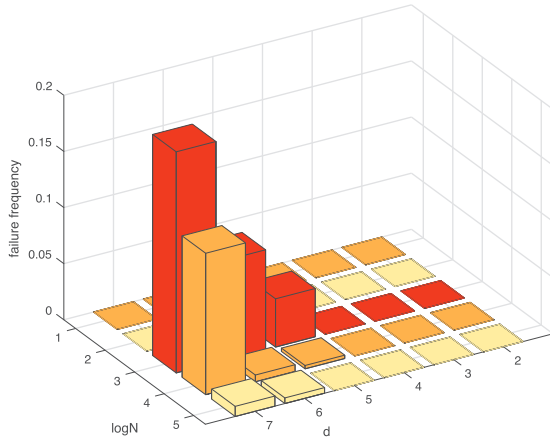


FIG. 7. Failure frequency of CVX and the MOSEK solver, depending on multinomial statistical noise $\log_{10} N$ and Hilbert-space dimension d . There were 1000 points sampled for each combination of N and d .

pseudoinverse and instead uses a method of Gaussian elimination. We projected the estimate onto the space of Hermitian matrices $C_{\mathcal{E}}^{LI} \rightarrow (C_{\mathcal{E}}^{LI} + (C_{\mathcal{E}}^{LI})^\dagger)/2$ in order to investigate their subsequently real eigenspectrum. We found that estimates were unphysical with overwhelming probability (see Fig. 4). The severity of the problem increases with shot noise. It is therefore not possible to use the standard analytical tools of infidelity or J distance to meaningfully compare the

accuracy of linear inversion with ML. It is worth noting that the ML approach provides estimates with smallest eigenvalue no less than some negative quantity which can be set arbitrarily small.

To allow for a comparison with the manifestly physical approach of ML, we upgraded linear inversion to force it to also produce physical estimates—simply by reapplying the composite CPTP projection derived in the main paper just once. We call the resultant estimator linear inversion with a final projection:

$$C_{\mathcal{E}}^{LIFP} := \mathcal{CPTP}(\text{vec}^{-1}[A^+n]), \quad (\text{E2})$$

once more forgoing explicit calculation of A^+ in favor of the MATLAB backslash operator \backslash . This approach is very much in the same spirit as the “quick and dirty” state tomography approach used by Kaznady and James [33]. We supplement the results in the main paper with an investigation in $d = 7$, shown in Fig. 5. The results confirm the idea that linear inversion with a final projection is a quick and effective method when the number of trials is taken to infinity, but that the loss in accuracy for finite N is exacerbated in higher dimensions. In fact, the projection is not necessary when there is zero noise. When noise is negligible, linear inversion provides a very high-precision estimate due to its noniterative nature, but our results suggest that ML approaches this situation faster as multinomial noise is decreased. Otherwise, in the physically relevant scenario of finite N a significant loss in precision is clear to see. The running time of linear inversion with a final

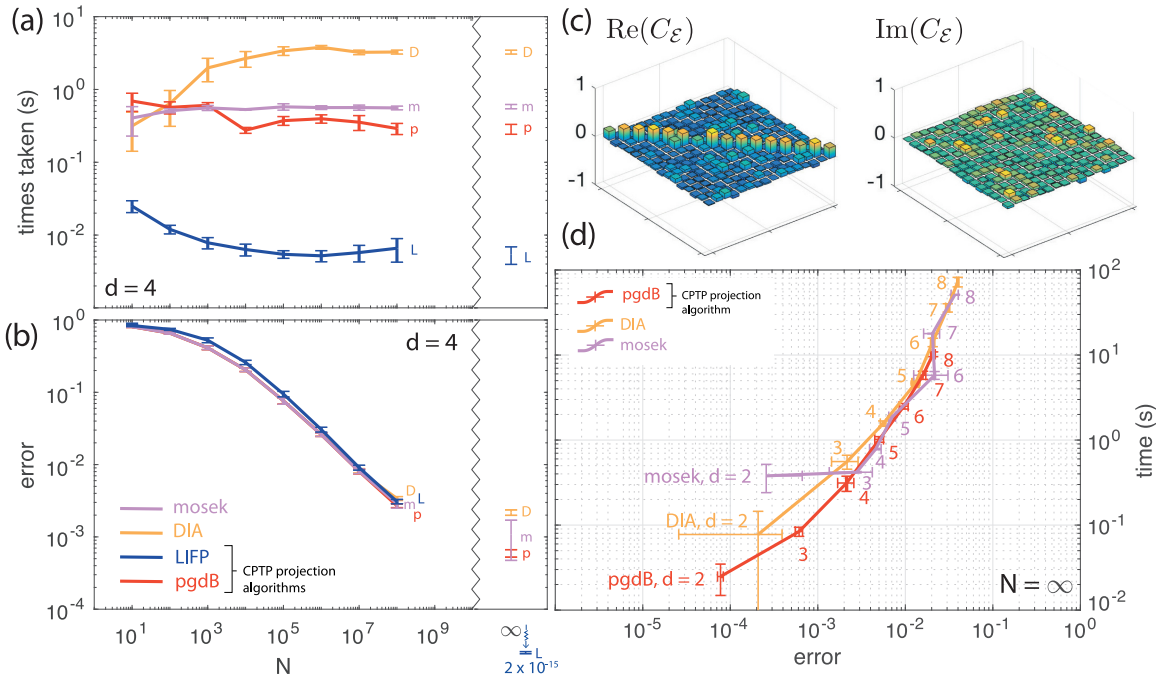


FIG. 8. Reconstruction of full-rank processes (a) Our proposed algorithm pgdB (p, dark red) is significantly faster than the state-of-the-art ML-QPT algorithm DIA (D, light orange) when $N \gtrsim 10^3$. LIFP (L, dark blue) is faster still, due to its noniterative nature. (b) Reconstruction error (J distance) decreases as the number of observations, N , increases. All algorithms produce approximately the same quality estimate. We could not simulate arbitrarily high N multinomial noise, but can effectively consider the $N \rightarrow \infty$ case by setting $n_{ij} = p_{ij}$, where pgdB becomes the most accurate ML algorithm ($d = 4$ case shown). LIFP does not find the ML solution, and therefore sacrifices accuracy in all but the highly idealized case of infinite counts. (c) Real and imaginary parts of a typical, full-rank, $d = 4$ Choi matrix returned by our algorithm. (d) Accuracy and run time for all iterative ML algorithms as d increases, with negligible statistical noise: pgdB is faster and more accurate.

projection can be favorable in low dimensions but will become much slower than the other methods as d increases.

APPENDIX F: CONDITION NUMBER OF A

Figure 6 shows the condition number of A rising with d , for the “minimal” choice of preparations and measurements made in the main text. A high condition number means a slower run time and a less accurate result in general. The condition number might be brought down by considering overcomplete preparations and measurements, or SIC-POVMs [34]. The density of A is typically around 10% or less, meaning efficiency savings for sparse operations with MATLAB.

The computational complexity of the gradient-based algorithms is dominated by the matrix-vector product required to calculate the gradient. The complexity is $O(d^4 n_{\text{combs}})$, where n_{combs} is the number of combinations of preparations and measurement outcomes. Our choice above implies $n_{\text{combs}} = 2d^4$. An overcomplete set would impact running time: for example, choosing the eigenvectors of the generalized Gell-Mann matrices [50] implies $n_{\text{combs}} = (d^3 - d)^2$, and raises the complexity from $O(d^8)$ to $O(d^{10})$.

APPENDIX G: FAILURE OF CVX

Depending on dimensionality and statistical noise, we found the experimental method of successive approximation used by CVX [43] to be unreliable. This is a known issue and to be expected due to the heuristic nature of the method.

Figure 7 shows the failure rate to be as high as 15%. The data for MOSEK used in Figs. 2, 5, and 8 are postselected on the algorithm succeeding.

APPENDIX H: GENERATION OF FULL-RANK AND QUASIPURE RANDOM CPTP MAPS

Random CPTP maps may be uniformly randomly generated according to a prescription by Bruzda *et al.* [51]:

$$Y = (\text{Tr}_{\mathcal{H}_{\text{out}}}[X X^\dagger])^{1/2}, \quad (\text{H1})$$

$$B = (\mathbb{I} \otimes Y^{-1/2}) X X^\dagger (\mathbb{I} \otimes Y^{-1/2}), \quad (\text{H2})$$

where X is a $d^2 \times M$ complex random matrix with entries distributed normally. The Kraus rank of the CPTP map is M . We generated d^2 rank-1 matrices B_i and formed their convex combination $C_{\varepsilon_{\text{true}}} = \sum_i P_i B_i$ with an exponentially decaying probability distribution P_i with $\sum_i P_i^2 = 0.9$. The resulting ensemble is CPTP (by convexity) and has purity $\text{Tr}[C_{\varepsilon}^2]/d^2$ no less than 90%. We call this the “quasipure” ensemble.

Using the first part of the above procedure we generated an alternative data set corresponding to full-rank processes. The results are shown in Fig. 8, where the lower purity of this ensemble (compared to the quasipure ensemble of Fig. 2) is apparent from the relative closeness of a typical Choi matrix to $\text{diag}(1, 1, \dots, 1)/d$. Our proposed algorithm is still superior to existing methods, although less dramatically so.

-
- [1] D. Aharonov and M. Ben-Or, Fault-tolerant quantum computation with constant error, in *Proceedings of the Twenty-Ninth Annual ACM Symposium on Theory of Computing, STOC '97* (ACM, New York, 1997), pp. 176–188.
 - [2] J. L. O’Brien, G. J. Pryde, A. Gilchrist, D. F. V. James, N. K. Langford, T. C. Ralph, and A. G. White, Quantum Process Tomography of a Controlled-NOT Gate, *Phys. Rev. Lett.* **93**, 080502 (2004).
 - [3] L. A. Pachón, A. H. Marcus, and A. Aspuru-Guzik, Quantum process tomography by 2D fluorescence spectroscopy, *J. Chem. Phys.* **142**, 212442 (2015).
 - [4] R. C. Bialczak, M. Ansmann, M. Hofheinz, E. Lucero, M. Neeley, A. D. O’Connell, D. Sank, H. Wang, J. Wenner, M. Steffen, A. N. Cleland, and J. M. Martinis, Quantum process tomography of a universal entangling gate implemented with Josephson phase qubits, *Nat. Phys.* **6**, 409 (2010).
 - [5] M. Howard, J. Twamley, C. Wittmann, T. Gaebel, F. Jelezko, and J. Wrachtrup, Quantum process tomography and Lindblad estimation of a solid-state qubit, *New J. Phys.* **8**, 33 (2006).
 - [6] A. Shabani, R. L. Kosut, M. Mohseni, H. Rabitz, M. A. Broome, M. P. Almeida, A. Fedrizzi, and A. G. White, Efficient Measurement of Quantum Dynamics via Compressive Sensing, *Phys. Rev. Lett.* **106**, 100401 (2011).
 - [7] J. J. Wallman and S. T. Flammia, Randomized benchmarking with confidence, *New J. Phys.* **16**, 103032 (2014).
 - [8] Y. R. Sanders, J. J. Wallman, and B. C. Sanders, Bounding quantum gate error rate based on reported average fidelity, *New J. Phys.* **18**, 012002 (2016).
 - [9] R. Blume-Kohout, J. K. Gamble, E. Nielsen, K. Rudinger, J. Mizrahi, K. Fortier, and P. Maunz, Demonstration of qubit operations below a rigorous fault tolerance threshold with gate set tomography, *Nat. Commun.* **8**, 14485 (2017).
 - [10] H. E. Daniels, The asymptotic efficiency of a maximum likelihood estimator, in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability Vol. 1* (University of California Press, Berkeley, CA, 1961), pp. 151–163.
 - [11] P. Kok and B. W. Lovett, *Introduction to Optical Quantum Information Processing* (Cambridge University Press, Cambridge, UK, 2010).
 - [12] T. Sugiyama, P. S. Turner, and M. Mura, Precision-Guaranteed Quantum Tomography, *Phys. Rev. Lett.* **111**, 160406 (2013).
 - [13] J. A. Smolin, J. M. Gambetta, and G. Smith, Efficient Method for Computing the Maximum-Likelihood Quantum State from Measurements with Additive Gaussian Noise, *Phys. Rev. Lett.* **108**, 070502 (2012).
 - [14] Y. S. Teo, *Introduction to Quantum-State Estimation* (World Scientific, Singapore, 2015).
 - [15] I. L. Chuang and M. A. Nielsen, Prescription for experimental determination of the dynamics of a quantum black box, *J. Mod. Opt.* **44**, 2455 (1997).
 - [16] D. W. Leung, Choi’s proof as a recipe for quantum process tomography, *J. Math. Phys.* **44**, 528 (2003).
 - [17] J. B. Altepeter, D. Branning, E. Jeffrey, T. C. Wei, P. G. Kwiat, R. T. Thew, J. L. O’Brien, M. A. Nielsen, and A. G. White, Ancilla-Assisted Quantum Process Tomography, *Phys. Rev. Lett.* **90**, 193601 (2003).

- [18] M.-D. Choi, Completely positive linear maps on complex matrices, *Linear Algebra Appl.* **10**, 285 (1975).
- [19] A positive semidefinite matrix is denoted by the expression $C \geq 0$, meaning that C belongs to the subset of Hermitian matrices with non-negative eigenvalues.
- [20] S. Boyd and L. Vandenberghe, *Convex Optimization* (Cambridge University Press, Cambridge, UK, 2009).
- [21] J. Plesník, Finding the orthogonal projection of a point onto an affine subspace, *Linear Algebra Appl.* **422**, 455 (2007).
- [22] E. Bolduc, G. C. Knee, E. M. Gauger, and J. Leach, Projected gradient descent algorithms for quantum state tomography, *npj Quantum Inf.* **3**, 44 (2017).
- [23] R. Escalante and M. Raydan, *Alternating Projection Methods* (Society for Industrial and Applied Mathematics, Philadelphia, PA, 2011).
- [24] E. G. Birgin and M. Raydan, Dykstra's algorithm and robust stopping criteria, in *Encyclopedia of Optimization*, edited by C. A. Floudas and P. M. Pardalos (Springer, Boston, 2009), pp. 828–833.
- [25] D. Drusvyatskiy, C.-K. Li, D. C. Pelejo, Y.-L. Voronin, and H. Wolkowicz, Projection methods for quantum channel construction, *Quantum Inf. Process.* **14**, 3075 (2015).
- [26] D. S. Gonçalves, M. A. Gomes-Ruggiero, and C. Lavor, A projected gradient method for optimization over density matrices, *Optim. Methods Software* **31**, 328 (2016).
- [27] J. Shang, Z. Zhang, and H. K. Ng, Superfast maximum-likelihood reconstruction for quantum tomography, *Phys. Rev. A* **95**, 062336 (2017).
- [28] D. Henrion and J. Malick, Projection methods in conic optimization, in *Handbook on Semidefinite, Conic and Polynomial Optimization*, edited by Miguel F. Anjos and Jean B. Lasserre (Springer, Boston, 2012), pp. 565–600.
- [29] L. Armijo, Minimization of functions having Lipschitz continuous first partial derivatives, *Pacific J. Math.* **16**, 1 (1966).
- [30] D. Bertsekas, On the Goldstein-Levitin-Polyak gradient projection method, *IEEE Trans. Autom. Control* **21**, 174 (1976).
- [31] Y.-H. Li and V. Cevher (private communication).
- [32] A. Beck, *Introduction to Nonlinear Optimization: Theory, Algorithms, and Applications with MATLAB*, MOS-SIAM Series on Optimization (Society for Industrial and Applied Mathematics, Philadelphia, PA, 2014).
- [33] M. S. Kaznady and D. F. V. James, Numerical strategies for quantum tomography: Alternatives to full optimization, *Phys. Rev. A* **79**, 022109 (2009).
- [34] J. M. Renes, R. Blume-Kohout, A. J. Scott, and C. M. Caves, Symmetric informationally complete quantum measurements, *J. Math. Phys.* **45**, 2171 (2004).
- [35] S. T. Merkel, J. M. Gambetta, J. A. Smolin, S. Poletto, A. D. Córcoles, B. R. Johnson, C. A. Ryan, and M. Steffen, Self-consistent quantum process tomography, *Phys. Rev. A* **87**, 062119 (2013).
- [36] R. Blume-Kohout, J. K. Gamble, E. Nielsen, J. Mizrahi, J. D. Sterk, and P. Maunz, Robust, self-consistent, closed-form tomography of quantum logic gates on a trapped ion qubit, [arXiv:1310.4492](https://arxiv.org/abs/1310.4492).
- [37] A. Gilchrist, N. K. Langford, and M. A. Nielsen, Distance measures to compare real and ideal quantum processes, *Phys. Rev. A* **71**, 062310 (2005).
- [38] J. Fiurášek and Z. Hradil, Maximum-likelihood estimation of quantum processes, *Phys. Rev. A* **63**, 020101(R) (2001).
- [39] J. Řeháček, Z. Hradil, E. Knill, and A. I. Lvovsky, Diluted maximum-likelihood algorithm for quantum tomography, *Phys. Rev. A* **75**, 042108 (2007).
- [40] A. Anis and A. I. Lvovsky, Maximum-likelihood coherent-state quantum process tomography, *New J. Phys.* **14**, 105021 (2012).
- [41] M. Cooper, E. Slade, M. Karpiński, and B. J. Smith, Characterization of conditional state-engineering quantum processes by coherent state quantum process tomography, *New J. Phys.* **17**, 033041 (2015).
- [42] I. A. Fedorov, A. K. Fedorov, Y. V. Kurochkin, and A. I. Lvovsky, Tomography of a multimode quantum black box, *New J. Phys.* **17**, 043063 (2015).
- [43] M. Grant and S. Boyd, CVX: Matlab software for disciplined convex programming, version 2.1, <http://cvxr.com/cvx> (2014).
- [44] MATLAB code is available at <https://github.com/geoknee/CPTPprojection>
- [45] M. Jiang, S. Luo, and S. Fu, Channel-state duality, *Phys. Rev. A* **87**, 022310 (2013).
- [46] M. F. Sacchi, Maximum-likelihood reconstruction of completely positive maps, *Phys. Rev. A* **63**, 054104 (2001).
- [47] J. Fiurasek and Z. Hradil, Comment on “Maximum likelihood reconstruction of CP maps”, [arXiv:quant-ph/0101048](https://arxiv.org/abs/quant-ph/0101048).
- [48] I. Bongioanni, L. Sansoni, F. Sciarrino, G. Vallone, and P. Mataloni, Experimental quantum process tomography of non-trace-preserving maps, *Phys. Rev. A* **82**, 042307 (2010).
- [49] A. S. Lewis and J. Malick, Alternating projections on manifolds, *Math. Oper. Res.* **33**, 216 (2008).
- [50] R. A. Bertlmann and P. Krammer, Bloch vectors for qudits, *J. Phys. A: Math. Theor.* **41**, 235303 (2008).
- [51] W. Bruzda, V. Cappellini, H.-J. Sommers, and K. Życzkowski, Random quantum operations, *Phys. Lett. A* **373**, 320 (2009).