Gradient-Descent Quantum Process Tomography by Learning Kraus Operators

Citation for the original published paper (version of record):
Ahmed, S., Quijandria Diaz, I., Frisk Kockum, A. (2023). Gradient-Descent Quantum Process Tomography by Learning Kraus Operators. Physical Review Letters, 130(15).
http://dx.doi.org/10.1103/PhysRevLett.130.150402

Gradient-Descent Quantum Process Tomography by Learning Kraus Operators

Shahnawaz Ahmed,[1,*] Fernando Quijandría,[1,2] and Anton Frisk Kockum[1,†]
[1]Department of Microtechnology and Nanoscience, Chalmers University of Technology, 412 96 Gothenburg, Sweden
[2]Quantum Machines Unit, Okinawa Institute of Science and Technology Graduate University, Onna-son, Okinawa 904-0495, Japan

(Received 11 August 2022; revised 17 February 2023; accepted 16 March 2023; published 14 April 2023)

We perform quantum process tomography (QPT) for both discrete- and continuous-variable quantum systems by learning a process representation using Kraus operators. The Kraus form ensures that the reconstructed process is completely positive. To make the process trace preserving, we use a constrained gradient-descent (GD) approach on the so-called Stiefel manifold during optimization to obtain the Kraus operators. Our ansatz uses a few Kraus operators to avoid direct estimation of large process matrices, e.g., the Choi matrix, for low-rank quantum processes. The GD-QPT matches the performance of both compressed-sensing (CS) and projected least-squares (PLS) QPT in benchmarks with two-qubit random processes, but shines by combining the best features of these two methods. Similar to CS (but unlike PLS), GD-QPT can reconstruct a process from just a small number of random measurements, and similar to PLS (but unlike CS) it also works for larger system sizes, up to at least five qubits. We envisage that the data-driven approach of GD-QPT can become a practical tool that greatly reduces the cost and computational effort for QPT in intermediate-scale quantum systems.

DOI: 10.1103/PhysRevLett.130.150402

Introduction.—The characterization of noisy quantum operations is of broad interest in emerging quantum technologies [1–27]. Examples include sampling complicated quantum circuits [18,25] and creating interesting quantum states [28–30] and gates [31,32]. Efficient characterization of such operations would help understanding and tackling noise in current quantum devices.

A quantum operation can be represented as a completely positive (CP) and trace-preserving (TP) linear map $\mathcal{E}$ mapping a state ρ to another state ρ' in a (possibly) different Hilbert space, i.e., $\rho' = \mathcal{E}(\rho)$. Quantum process tomography (QPT) [33–36] estimates $\mathcal{E}$ from experimental data. Examples of QPT techniques are maximum-likelihood estimation [37–40], Bayesian estimation [41], projected gradient descent (GD) [42], projected least-squares (PLS) [43], compressed-sensing (CS) methods using convex optimization [44–47], and variational QPT [48].

A challenge in QPT is the exponentially growing size of the process representation, e.g., the Choi representation [49], a complex-valued $4^n \times 4^n$ matrix for n qubits. This makes it difficult both to estimate a quantum process from noisy data, and to interpret the results [50]. However, for realistic cases, low-rank approximations require much less data for QPT [45]. In quantum state tomography (QST), such approximations are sufficient to guide toward actionable and interpretable information [51,52]. Inspired by machine-learning techniques like neural-network QST [53–64], efficiently representing a process and learning it from data has recently shown promise for QPT [58]. Other approaches, e.g., shadow tomography [65,66], avoid any process representation to circumvent the dimensionality issue, but come with their own limitations for realistic problems [67,68].

In this Letter, we tackle QPT using a simple GD-based optimization to learn Kraus operators, which can represent any quantum process as a CP linear map [49,69]. The number of Kraus terms can be flexibly adapted for a low-rank reconstruction. We formulate QPT as a learning task similar to the training of neural networks from data. The CP condition is ensured by construction [69]; an efficient gradient-based optimization using retraction [70–73] implements orthonormality constraints to always satisfy the TP condition during optimization.

Our GD-QPT is an example of Riemannian optimization on the Stiefel manifold [74,75], which is also of interest in machine learning [72], and has been applied to problems in quantum physics [73,76,77]. Similar ideas led to the quantum-inspired partial-trace-regression technique [78]. We show how the simple GD approach can reconstruct both continuous- and discrete-variable (CV and DV) quantum processes with different measurement schemes, large Hilbert-space dimensions, and limited data.

We benchmark GD-QPT against CS and PLS. In CS, QPT is formulated as a convex optimization problem using the Choi representation [44,45,47]. Using convex

programming, CS handles CPTP constraints easily, guarantees a global optimum, and shines in the regime of limited data, where standard QPT would have an underdetermined system of equations for a general full-rank process [45,79]. However, CS can be computationally expensive in practice, limiting its applicability for even three-qubit processes, which may require several hours of computation [45]. A five-qubit process, or a CV process with a larger Hilbert-space cutoff than used in existing methods [39,40], may thus become impractical.

Projection-based methods [42] like PLS [43] obtain a process estimate and then project it to the nearest CPTP estimate. The PLS method is fast: it can reconstruct processes for 5–7 qubits in reasonable time [43]. However, projection-based techniques may require an initial analytical least-squares estimate relying on an informationally complete set of measurements, along with costly projection steps involving eigendecompositions of large Choi matrices or iterative subroutines. Our approach avoids these problems, but can still handle relatively large Hilbert-space dimensions.

We show that the simple GD-QPT technique yields similar performances as CS and PLS on benchmarks using random processes with Gaussian noise in the data. We assess the performance of GD-QPT and CS against the amount of data, showing that GD-QPT combines the best of two worlds. Like CS, GD-QPT can work without informationally complete data, but it can still, like PLS, tackle larger problems.

Extending GD-QPT, we also construct neural-network QPT (NN-QPT) [80], where the Kraus operators are the output of a neural network, similar to previous works [83,84] for QST. We find no advantage of this approach over GD-QPT. Our GD approach thus introduces a flexible QPT technique, demonstrating that simple gradient-based optimization combined with appropriate regularization and efficient process representation is an effective tool for quantum process characterization.

Process representations.—The Kraus-operator representation of $\mathcal{E}$ is k complex-valued matrices $\{K_l\}$ of dimension $N \times N$, acting on a density matrix ρ as $\mathcal{E}(\rho) = \sum_{l=1}^{k} K_l \rho {K_l}^\dagger = \rho'$. Here N is the Hilbert-space dimension; for n qubits $N = 2^n$. The Kraus representation guarantees that the process is *CP* [69,78,85]. The TP condition translates to satisfying $\sum_{l=1}^{k} K_l^\dagger K_l = \mathbb{I}$.

The Choi representation [49,86] of $\mathcal{E}$ is an $N^2 \times N^2$ complex-valued matrix Φ that can be written using Kraus operators as $\Phi = \sum_{l=1}^{k} |K_l\rangle\langle K_l|$ with $|K_l\rangle = (\mathbb{I} \otimes K_l) \sum_i |i\rangle \otimes |i\rangle$. The Choi matrix is thus a linear operator acting on the tensor product of the input and output Hilbert spaces $H_{\text{in}} \otimes H_{\text{out}}$. The action of Φ on ρ is given by the partial trace $\rho' = \text{Tr}_{H_{\text{in}}}[(\rho^T \otimes \mathbb{I})\Phi]$. For Φ to be CPTP, it should be positive semidefinite [42,49] and satisfy $\text{Tr}_{H_{\text{out}}}(\Phi) = \mathbb{I}$.

We consider the Kraus operator form because it allows us to control the size of the process representation. The Choi rank r of a process is the minimum number of Kraus operators necessary to represent the process. The maximum rank is $r = N^2$, but in realistic cases, process matrices can have low ranks $r \ll N^2$ (for a unitary process, $r = 1$). Our approach gives the flexibility to choose the rank $r = k$ (the number of Kraus operators) of the process ansatz, allowing us to obtain low-rank approximations without constructing the full Choi matrix. Most previous QPT methods preferred the Choi-matrix representation because it made CPTP constraints easier to handle [42] and the problem could be cast in a linear form [45].

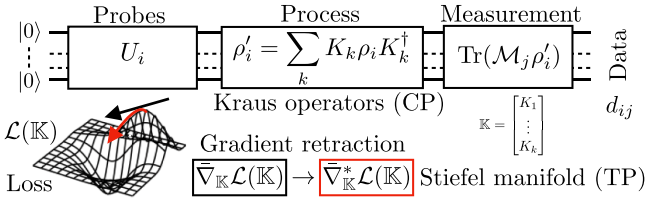

FIG. 1. We estimate a quantum process $\mathcal{E}$ from data d_{ij} (outcomes of measurements $\mathcal{M}_j$ on states given by $\mathcal{E}$ acting on probes ρ_i). Gradient-based optimization of a loss function $\mathcal{L}$ yields $\mathbb{K}$, a set of Kraus operators representing $\mathcal{E}$. The Kraus form enforces complete positivity; a gradient-retraction technique restricts $\mathbb{K}$ to the Stiefel manifold, enforcing trace preservation.

Learning processes.—We illustrate the GD-QPT idea in Fig. 1. In QPT experiments, data d_{ij} are estimates (computed by averaging single-shot outcomes) of expectation values of measurements $\mathcal{M}_j$ on output states $\rho_i' = \mathcal{E}(\rho_i)$ for probe (input) states ρ_i. Noise sampled from a zero-mean Gaussian distribution $\mathcal{N}(0, \epsilon)$ with standard deviation ϵ represents statistical errors. The process-reconstruction problem can be cast as a learning task: minimizing a loss function $\mathcal{L}$ quantifying the discrepancy between the data d_{ij} and our process estimate. We use

$$\mathcal{L}(\mathbb{K}) = \sum_{ij} \left[d_{ij} - \text{Tr}\left[\mathcal{M}_j \left(\sum_k K_k \rho_i {K_k}^\dagger \right) \right] \right]^2 + \lambda \|\mathbb{K}\|_1, \tag{1}$$

combining least-squares-error loss with L1 regularization; $\mathbb{K} = [K_1, \ldots, K_k]$ is a $kN \times N$ matrix constructed by stacking the k Kraus operators representing the process, and the matrix norm $\|A\|_1 = \max_j \sum_i |A_{ij}|$ is induced by the L1 norm [87,88], with $\lambda \geq 0$ the strength of the regularization. We fix $\lambda = 10^{-3}$ in this Letter, but it can be optimized further as a hyperparameter.

We can use any other objective function, e.g., the likelihood function, which, assuming Gaussian noise, becomes equivalent to minimizing the least-squares loss [89]. We choose the latter for its simplicity and wide applicability beyond data representing probabilities (e.g., Wigner functions that can have negative values).

The loss function is minimized with GD by updating $\mathbb{K}$ along the negative (conjugate, since the Kraus operators are complex [90]) gradient $\nabla_{\mathbb{K}}\mathcal{L}(\mathbb{K})$. However, simple GD might lead to estimates that violate the TP constraint (equivalently, the orthonormality condition $\mathbb{K}^\dagger\mathbb{K} = \mathbb{I}$). To counter such violations, one could add a loss term that penalizes them, e.g., $\|\sum_l K_l^\dagger K_l - \mathbb{I}\|_1$. However, this penalty does not strictly enforce the TP condition.

In the Choi representation, we can linearize the problem to implement CS-QPT as

$$\Phi_{\mathrm{CS}} = \arg\ \min \|\Phi'\|_1 \ \text{s.t.}\Phi' \geq 0, \qquad \|S\vec{\Phi}' - \vec{d}\|_2 \leq \delta, \tag{2}$$

where δ is the noise level set as a threshold. The matrix S, constructed using the probes and measurement operators $\{\rho_i, \mathcal{M}_j\}$ [42], is similar to the sensing matrix in QST [91]. The data is a vector $\vec{d}$ with an appropriate flattening $\vec{\Phi}'$ of the Choi matrix. The TP condition is implemented by the constraint $\mathrm{Tr}_{H_{\mathrm{out}}}(\Phi') - \mathbb{I} = 0$. We use the splitting conic solver [92] to solve the convex optimization task for CS in Python with CVXPY [93,94] following Qiskit [95] to implement the CPTP constraints.

Gradient descent on the Stiefel manifold.—The orthonormal condition on $\mathbb{K}$ defines the so-called Stiefel manifold [70,72,74]. It is possible to restrict the gradients such that we never leave this manifold during the optimization [70–72,74,75]; this is an example of Riemannian optimization on a manifold [75]. Several works have addressed this problem using a retraction technique that is an approximation to the exponential map [96]. The retraction restricts the updated $\mathbb{K}$ to the Stiefel manifold while minimizing the loss (see Fig. 1).

Let $G' = \nabla_{\mathbb{K}}\mathcal{L}(\mathbb{K})$. In each update step, we normalize the gradients with the L2 norm: $G = G'/\|G'\|_2$. With stacked matrices $A = [G \quad \mathbb{K}]$ and $B = [\mathbb{K} \quad -G]$, the TP condition is implemented by a retraction updating the gradients as [80]

$$\nabla^*_{\mathbb{K}}\mathcal{L}(\mathbb{K}) = A\left(\mathbb{I} + \frac{\eta}{2}B^\dagger A\right)^{-1} B^\dagger\mathbb{K}, \tag{3}$$

where η is the learning rate. We iteratively minimize the loss $\mathcal{L}(\mathbb{K})$ with gradient-based updates $\mathbb{K}' = \mathbb{K} - \eta\nabla^*_{\mathbb{K}}\mathcal{L}(\mathbb{K})$ keeping $\mathbb{K}$ in the Stiefel manifold. The retraction formula is based on the Cayley transform and the Sherman-Morrison-Woodbury formula [74].

Gradient-based optimization is quite successful as a heuristic with convergence guarantees in specific cases [97], but our optimization contains nonconvex constraints, making it difficult to guarantee finding a global minimum. We use stochastic gradient descent, where iterative updates are noisy due to batching data for gradient computation; this enables avoiding local minima in some cases [98,99]. Actually, saddle points are a bigger issue than local minima in high-dimensional nonconvex optimization [100], but gradient-based optimization is effective in finding global minima for such problems [101] with provable guarantees for, e.g., overparametrized neural networks [102].

The starting estimate for $\mathbb{K}$ is random unitary matrices with appropriate normalization guaranteeing that they describe a CPTP process. Our learning rate decays by a factor 0.999 in each step with $\eta(0) = 0.1$, fixed throughout the Letter for consistent benchmarking. Better results may be possible by hyperparameter search or choosing η from the Armijo-Wolfe conditions, guaranteeing that each update step decreases the objective function [70,80,103,104].

Results and benchmarking.—We first consider QPT for a CV quantum operation—a selective number-dependent arbitrary phase (SNAP) gate [80,105,106] and a displacement operation—using coherent states as probes [39]. SNAP gates have recently been used experimentally to create interesting CV quantum states such as Gottesman-Kitaev-Preskill states and the cubic phase state [30]. The parameters of the SNAP + displacement operation are given in the Supplemental Material [80]. In such CV problems, choosing an appropriate Hilbert-space cutoff to correctly describe the state at hand is fundamental [39,40]. Here, we consider a cutoff of 32, which, to the best of our knowledge, is the largest dimension explored for single-mode CV QPT [36,39,107,108].

In Fig. 2, we show the Wigner functions for one instance of a coherent probe state $\rho_i = |\alpha_i\rangle\langle\alpha_i|$, target state $\rho'_i = \mathcal{E}(\rho_i)$ after the process, and sampled data d_{ij}. The data correspond to measurements of the displaced parity operator $\Pi(\beta)$ on ρ'_i. Both α_i and β are continuous, but sampled in a grid. We deliberately choose a coarse grid to highlight that we do not require full Wigner tomography for each probe state during QPT. In an experiment, probes and measurements have to be appropriately chosen depending on the process and the limitations of the experimental setup.

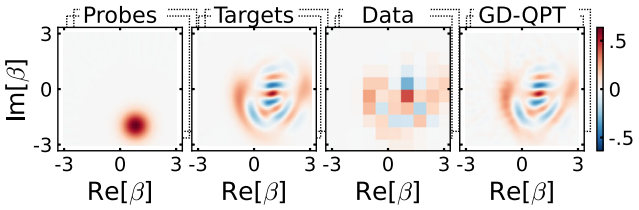

FIG. 2. Reconstruction of a CV quantum process using GD-QPT with a Hilbert-space cutoff of 32. The probes are coherent states $|\alpha_i\rangle$ in a 10×10 grid with $\mathrm{Re}[\alpha_i]$, $\mathrm{Im}[\alpha_i] \in [-2.5, 2.5]$. The data are measured values of the displaced parity $\Pi(\beta_j)$ in a 10×10 grid with $\mathrm{Re}[\beta_j]$, $\mathrm{Im}[\beta_j] \in [-3, 3]$. The grid is coarse, demonstrating that we only need few measurements per probe in our reconstruction. We sample data d'_{ij} from our process estimate in a finer grid to show the agreement with the true process.

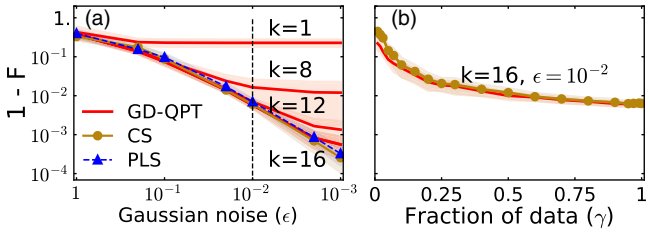

FIG. 3. Benchmarking GD-QPT (red) against PLS (blue) and CS (gold) for random two-qubit ($n = 2$) full-rank ($r = 16$) processes. Mean infidelities (a) as the noise ϵ is decreased for 30 random processes; (b) for reconstructions of these processes using a fraction γ of the data used in (a) for $\epsilon = 10^{-2}$ (randomly selecting a $\sqrt{\gamma}6^n \times \sqrt{\gamma}6^n$ subset of the total $6^n \times 6^n$ Pauli probes and measurements). Shading shows 1 standard deviation.

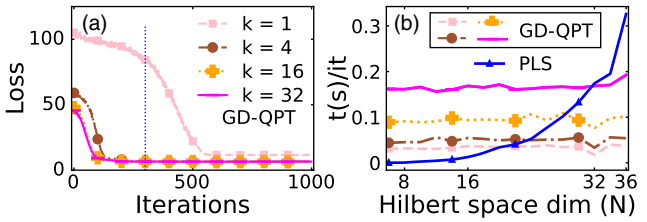

FIG. 4. Comparing computational time for GD-QPT and PLS (blue). (a) Loss in GD-QPT after each update step for random 5-qubit processes of rank $r = 3$. One step for GD-QPT takes a batch of 256 data points from $6^5 \times 6^5$ measurements, computes the loss and its gradient, and performs an update on the Stiefel manifold. In Ref. [43], convergence of the *CP* projection required ~300 iterations when $n = 5$. (b) Time taken per GD-QPT iteration and the most expensive step in PLS, the *CP* projection, as a function of Hilbert-space dimension with random processes [80].

We calculate the reconstruction fidelity $F(\Phi, \Phi') = \mathrm{tr}(\sqrt{\sqrt{\Phi}\Phi'\sqrt{\Phi}})$ [109], one of several distance measures for processes [110,111], using the Choi form with the appropriate normalization, $\Phi^{\text{GD-QPT}}/N$. The average value of F for 30 random choices of the Kraus operators with a noise $\epsilon = 10^{-2}$ is > 0.97. Each reconstruction converges within 50 iterations, taking tens of seconds on a laptop with 32 GB memory and 2.9 GHz 6-core Intel Core i9 processor. We used $k = 3$ Kraus terms, but actually only needed $k = 1$ since SNAP and displacement operations are unitary.

To study the effects of the number of Kraus operators in our ansatz, noise, and amount of data, we turn to reconstructing random DV quantum processes (processes acting on n qubits). First, in Fig. 3(a), we quantify the impact of Gaussian measurement noise (related to the number of measurement samples [43]) and the number of Kraus operators. We follow the direct QPT approach of Ref. [43], where the 6^n probes and 6^n measurements are tensor products of the eigenstates to the Pauli matrices $\{\sigma_x, \sigma_y, \sigma_z\}$. We compare our results against a hyperplane-intersection projection method [43] and a CS implementation using convex programming [45].

We find that all three methods perform similarly as a function of the measurement noise ϵ, assuming the full set of $k = 4^n$ Kraus operators for a full-rank process. Reducing the number of Kraus terms in our ansatz, the fidelity saturates at lower values for lower k's, which is expected as our approximation then is not expressive enough to represent the full-rank process. In a practical setting, we may only be able to reach a certain fidelity with a finite number of measurement shots (nonzero ϵ). Interestingly, using more Kraus terms in such situations may not help.

In most realistic cases, where we might be interested in implementing quantum gates which are unitary or near-unitary processes, CS-QPT methods can work with very little data [45]. Nevertheless, we benchmark for full-rank processes in a two-qubit system to demonstrate the general applicability of our approach (Ref. [80] contains results for low-rank processes). In Fig. 3(b), we compare GD-QPT and CS performance using a random subset of probes and measurements (a fraction γ of the total). For *two-qubit* processes in this informationally incomplete regime, GD-QPT achieves similar fidelity as CS. However, the advantage of GD-QPT is that by using few Kraus operators, we can reconstruct processes in a larger Hilbert space (5-qubit DV systems and CV processes with a Hilbert-space cutoff of 32).

The PLS method was omitted from Fig. 3(b) since it was not clear how to adapt it to noninformationally complete data. However, PLS can be used to reconstruct processes with more qubits, where CS struggles. The dimension of the matrix S in CS is $6^{2n} \times 4^{2n}$ for the probabilities and the flattened Choi matrix. Therefore, running convex optimization programs even for reconstructing three-qubit processes with CS is challenging, requiring several hours of computational time [45]. In contrast, GD-QPT can easily tackle five-qubit processes, similar to PLS. Further, due to the restricted number of Kraus operators, GD-QPT iterations are faster than PLS for larger Hilbert spaces.

In Fig. 4, we compare the number of iterations for the convergence of GD-QPT and PLS for random 5-qubit processes. While GD-QPT converges in a similar number of iterations as PLS, it is faster per iteration due to the smaller number of Kraus terms considered. The most expensive step in the PLS technique, the *CP* projection, requires a diagonalization involving the eigendecomposition of the $4^n \times 4^n$-dimensional Choi-matrix estimate. The time taken for each step is limited by the complexity (cubic) of this eigendecomposition. In comparison, the most expensive step in GD-QPT is the retraction, which involves inverting smaller matrices of dimensions $k2^n \times 2^n$, where the number of Kraus terms $k \ll 4^n$. We provide the data and code for all results along with our implementation of GD-QPT, PLS, and CS in Ref. [112].

Conclusion and outlook.—In this Letter, we introduced a simple yet powerful technique for QPT using gradient-based learning of Kraus operators—GD-QPT. Our approach can

reconstruct both CV and DV processes for Hilbert spaces of dimension at least 32. We benchmarked GD-QPT against both the recently proposed PLS algorithm and CS. Using randomly generated processes, we showed that for low-rank processes, estimating Kraus operators directly gives fidelities similar to PLS and CS for the same amount of Gaussian noise in the data.

Further, we showed that our approach performs similarly to CS in the regime of informationally incomplete data, yet works for a larger number of qubits than CS. In this regard, we achieve the goal outlined in Ref. [45], where the convex-programming-based CS-QPT technique suffered from numerical time and memory complexity issues. Our simple approach alleviates some of the numerical issues by considering QPT as a learning problem with a limited number of Kraus operators as a model. Using data, we learn the Kraus operators through gradient-based optimization on the Stiefel manifold.

Future work could replace the Kraus representation with other efficient (approximate) representations, e.g., tensor networks [58,113]. A more expressive ansatz does not necessarily perform better under noisy data [cf. Fig. 3(a)]; the relation between sample complexity of measurements and ansatz expressivity should be explored. Also, benchmarking against techniques such as shadow tomography [65,66] could reveal if GD-QPT can strike a balance between tackling QPT for larger quantum systems and having an explicit, interpretable process representation. Other issues to explore are fast uncertainty estimation techniques [114,115], aleatoric (statistical) and epistemic (lack of data) uncertainties [116] for error bars [117–119], analysis of convergence and sample efficiency of the optimization, and hyperparameter optimization (learning rate and regularization strength) for reliable results.

The Python-based packages JAX [120] and OPTAX [121] were used for automatic gradient calculation and optimization. QUTIP [122,123] was used for visualization and generating data. The CS algorithm was implemented using CVXPY [93,94]. We acknowledge useful discussions with Ingrid Strandberg, Christopher Warren, Axel Eriksson, Mikael Kervinen, Roeland Wiersema, Juan Carrasquilla, and Nathan Killoran. We acknowledge support from the Knut and Alice Wallenberg Foundation through the Wallenberg Centre for Quantum Technology (WACQT).

*shahnawaz.ahmed95@gmail.com
†anton.frisk.kockum@chalmers.se

[1] R. P. Feynman, Simulating physics with computers, Int. J. Theor. Phys. **21**, 467 (1982).
[2] S. Lloyd, Enhanced sensitivity of photodetection via quantum illumination, Science **321**, 1463 (2008).
[3] H. J. Kimble, The quantum internet, Nature (London) **453**, 1023 (2008).
[4] J. Q. You and F. Nori, Atomic physics and quantum optics using superconducting circuits, Nature (London) **474**, 589 (2011).
[5] M. Bal, C. Deng, J.-L. Orgiazzi, F. Ong, and A. Lupascu, Ultrasensitive magnetic field detection using a single artificial atom, Nat. Commun. **3**, 1324 (2012).
[6] I. M. Georgescu, S. Ashhab, and F. Nori, Quantum simulation, Rev. Mod. Phys. **86**, 153 (2014).
[7] A. Montanaro, Quantum algorithms: an overview, Quantum Inf. **2**, 15023 (2016).
[8] X. Gu, A. F. Kockum, A. Miranowicz, Y.-x. Liu, and F. Nori, Microwave photonics with superconducting quantum circuits, Phys. Rep. **718–719**, 1 (2017).
[9] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets, Nature (London) **549**, 242 (2017).
[10] G. Wendin, Quantum information processing with superconducting circuits: A review, Rep. Prog. Phys. **80**, 106001 (2017).
[11] C. L. Degen, F. Reinhard, and P. Cappellaro, Quantum sensing, Rev. Mod. Phys. **89**, 035002 (2017).
[12] S. Pirandola, B. R. Bardhan, T. Gehring, C. Weedbrook, and S. Lloyd, Advances in photonic quantum sensing, Nat. Photonics **12**, 724 (2018).
[13] J. Preskill, Quantum Computing in the NISQ era and beyond, Quantum **2**, 79 (2018).
[14] A. M. Childs, D. Maslov, Y. Nam, N. J. Ross, and Y. Su, Toward the first quantum simulation with quantum speedup, Proc. Natl. Acad. Sci. U.S.A. **115**, 9456 (2018).
[15] J. Argüello-Luengo, A. González-Tudela, T. Shi, P. Zoller, and J. I. Cirac, Analogue quantum chemistry simulation, Nature (London) **574**, 215 (2019).
[16] T. Gefen, A. Rotem, and A. Retzker, Overcoming resolution limits with quantum sensing, Nat. Commun. **10**, 4992 (2019).
[17] M. Korobko, Y. Ma, Y. Chen, and R. Schnabel, Quantum expander for gravitational-wave observatories, Light Sci. Appl. **8**, 118 (2019).
[18] F. Arute *et al.*, Quantum supremacy using a programmable superconducting processor, Nature (London) **574**, 505 (2019).
[19] H. Ma, M. Govoni, and G. Galli, Quantum simulations of materials on near-term quantum computers, npj Comput. Mater. **6**, 85 (2020).
[20] S. McArdle, S. Endo, A. Aspuru-Guzik, S. C. Benjamin, and X. Yuan, Quantum computational chemistry, Rev. Mod. Phys. **92**, 015003 (2020).
[21] B. Bauer, S. Bravyi, M. Motta, and G. Kin-Lic Chan, Quantum algorithms for quantum chemistry and quantum materials science, Chem. Rev. **120**, 12685 (2020).
[22] S. Barzanjeh, S. Pirandola, D. Vitali, and J. M. Fink, Microwave quantum illumination using a digital receiver, Sci. Adv. **6**, eabb0451 (2020).
[23] J. Wang, F. Sciarrino, A. Laing, and M. G. Thompson, Integrated photonic quantum technologies, Nat. Photonics **14**, 273 (2020).
[24] J. Yin *et al.*, Entanglement-based secure quantum cryptography over 1,120 kilometres, Nature (London) **582**, 501 (2020).

[25] H. S. Zhong *et al.*, Quantum computational advantage using photons, Science **370**, 1460 (2020).
[26] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, and P. J. Coles, Variational quantum algorithms, Nat. Rev. Phys. **3**, 625 (2021).
[27] L. S. Madsen, F. Laudenbach, M. F. Askarani, F. Rortais, T. Vincent, J. F. F. Bulmer, F. M. Miatto, L. Neuhaus, L. G. Helt, M. J. Collins, A. E. Lita, T. Gerrits, S. W. Nam, V. D. Vaidya, M. Menotti, I. Dhand, Z. Vernon, N. Quesada, and J. Lavoie, Quantum computational advantage with a programmable photonic processor, Nature (London) **606**, 75 (2022).
[28] A. Grimm, N. E. Frattini, S. Puri, S. O. Mundhada, S. Touzard, M. Mirrahimi, S. M. Girvin, S. Shankar, and M. H. Devoret, Stabilization and operation of a Kerr-cat qubit, Nature (London) **584**, 205 (2020).
[29] P. Campagne-Ibarcq, A. Eickbusch, S. Touzard, E. Zalys-Geller, N. E. Frattini, V. V. Sivak, P. Reinhold, S. Puri, S. Shankar, R. J. Schoelkopf, L. Frunzio, M. Mirrahimi, and M. H. Devoret, Quantum error correction of a qubit encoded in grid states of an oscillator, Nature (London) **584**, 368 (2020).
[30] M. Kudra, M. Kervinen, I. Strandberg, S. Ahmed, M. Scigliuzzo, A. Osman, D. P. Lozano, M. O. Tholén, R. Borgani, D. B. Haviland, G. Ferrini, J. Bylander, A. F. Kockum, F. Quijandría, P. Delsing, and S. Gasparinetti, Robust preparation of wigner-negative states with optimized SNAP-displacement sequences, PRX Quantum **3**, 030301 (2022).
[31] Y. Kim, A. Morvan, L. B. Nguyen, R. K. Naik, C. Jünger, L. Chen, J. M. Kreikebaum, D. I. Santiago, and I. Siddiqi, High-fidelity three-qubit iToffoli gate for fixed-frequency superconducting qubits, Nat. Phys. **18**, 783 (2022).
[32] C. W. Warren, J. Fernández-Pendás, S. Ahmed, T. Abad, A. Bengtsson, J. Biznárová, K. Debnath, X. Gu, C. Križan, A. Osman, A. F. Roudsari, P. Delsing, G. Johansson, A. F. Kockum, G. Tancredi, and J. Bylander, Extensive characterization of a family of efficient three-qubit gates at the coherence limit, arXiv:2207.02938.
[33] J. F. Poyatos, J. I. Cirac, and P. Zoller, Complete Characterization of a Quantum Process: The Two-Bit Quantum Gate, Phys. Rev. Lett. **78**, 390 (1997).
[34] 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).
[35] G. M. D'Ariano and P. Lo Presti, Quantum Tomography for Measuring Experimentally the Matrix Elements of an Arbitrary Quantum Operation, Phys. Rev. Lett. **86**, 4195 (2001).
[36] 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).
[37] J. Fiurášek and Z. Hradil, Maximum-likelihood estimation of quantum processes, Phys. Rev. A **63**, 020101(R) (2001).
[38] M. F. Sacchi, Maximum-likelihood reconstruction of completely positive maps, Phys. Rev. A **63**, 054104 (2001).
[39] S. Rahimi-Keshari, A. Scherer, A. Mann, A. T. Rezakhani, A. I. Lvovsky, and B. C. Sanders, Quantum process tomography with coherent states, New J. Phys. **13**, 013006 (2011).
[40] A. Anis and A. I. Lvovsky, Maximum-likelihood coherent-state quantum process tomography, New J. Phys. **14**, 105021 (2012).
[41] K. Schultz, Exponential families for Bayesian quantum process tomography, Phys. Rev. A **100**, 062316 (2019).
[42] G. C. Knee, E. Bolduc, J. Leach, and E. M. Gauger, Quantum process tomography via completely positive and trace-preserving projection, Phys. Rev. A **98**, 062336 (2018).
[43] T. Surawy-Stepney, J. Kahn, R. Kueng, and M. Guta, Projected least-squares quantum process tomography, Quantum **6**, 844 (2022).
[44] C. H. Baldwin, A. Kalev, and I. H. Deutsch, Quantum process tomography of unitary and near-unitary maps, Phys. Rev. A **90**, 012110 (2014).
[45] A. V. Rodionov, A. Veitia, R. Barends, J. Kelly, D. Sank, J. Wenner, J. M. Martinis, R. L. Kosut, and A. N. Korotkov, Compressed sensing quantum process tomography for superconducting quantum gates, Phys. Rev. B **90**, 144504 (2014).
[46] L. Banchi, J. Pereira, S. Lloyd, and S. Pirandola, Convex optimization of programmable quantum computers, npj Quantum Inf. **6**, 42 (2020).
[47] Y. S. Teo, G. I. Struchalin, E. V. Kovlakov, D. Ahn, H. Jeong, S. S. Straupe, S. P. Kulik, G. Leuchs, and L. L. Sánchez-Soto, Objective compressive quantum process tomography, Phys. Rev. A **101**, 022334 (2020).
[48] S. Xue, Y. Liu, Y. Wang, P. Zhu, C. Guo, and J. Wu, Variational quantum process tomography of unitaries, Phys. Rev. A **105**, 032427 (2022).
[49] M. D. Choi, Completely positive linear maps on complex matrices, Linear Algebra Appl. **10**, 285 (1975).
[50] A. G. Kofman and A. N. Korotkov, Two-qubit decoherence mechanisms revealed via quantum process tomography, Phys. Rev. A **80**, 042103 (2009).
[51] S. T. Flammia, D. Gross, Y. K. Liu, and J. Eisert, Quantum tomography via compressed sensing: Error bounds, sample complexity and efficient estimators, New J. Phys. **14**, 095022 (2012).
[52] C. A. Riofrío, D. Gross, S. T. Flammia, T. Monz, D. Nigg, R. Blatt, and J. Eisert, Experimental quantum compressed sensing for a seven-qubit system, Nat. Commun. **8**, 15305 (2017).
[53] G. Carleo and M. Troyer, Solving the quantum many-body problem with artificial neural networks, Science **355**, 602 (2017).
[54] G. Carleo, Y. Nomura, and M. Imada, Constructing exact representations of quantum many-body systems with deep neural networks, Nat. Commun. **9**, 5322 (2018).
[55] G. Torlai, G. Mazzola, J. Carrasquilla, M. Troyer, R. Melko, and G. Carleo, Neural-network quantum state tomography, Nat. Phys. **14**, 447 (2018).
[56] J. Carrasquilla, G. Torlai, R. G. Melko, and L. Aolita, Reconstructing quantum states with generative models, Nat. Mach. Intell. **1**, 155 (2019).
[57] A. Melkani, C. Gneiting, and F. Nori, Eigenstate extraction with neural-network tomography, Phys. Rev. A **102**, 022412 (2020).

[58] G. Torlai, C. J. Wood, A. Acharya, G. Carleo, J. Carrasquilla, and L. Aolita, Quantum process tomography with unsupervised learning and tensor networks, arXiv:2006.02424.

[59] A. W. R. Smith, J. Gray, and M. S. Kim, Efficient quantum state sample tomography with basis-dependent neural networks, PRX Quantum **2**, 020348 (2021).

[60] S. Lohani, B. T. Kirby, M. Brodsky, O. Danaci, and R. T. Glasser, Machine learning assisted quantum state estimation, Mach. Learn. **1**, 035007 (2020).

[61] M. Neugebauer, L. Fischer, A. Jäger, S. Czischek, S. Jochim, M. Weidemüller, and M. Gärttner, Neural-network quantum state tomography in a two-qubit experiment, Phys. Rev. A **102**, 042604 (2020).

[62] A. M. Palmieri, E. Kovlakov, F. Bianchi, D. Yudin, S. Straupe, J. D. Biamonte, and S. Kulik, Experimental neural network enhanced quantum tomography, Quantum Inf. **6**, 20 (2020).

[63] L. Banchi, E. Grant, A. Rocchetto, and S. Severini, Modelling non-Markovian quantum processes with recurrent neural networks, New J. Phys. **20**, 123030 (2018).

[64] L. Che, C. Wei, Y. Huang, D. Zhao, S. Xue, X. Nie, J. Li, D. Lu, and T. Xin, Learning quantum Hamiltonians from single-qubit measurements, Phys. Rev. Res. **3**, 023246 (2021).

[65] R. Levy, D. Luo, and B. K. Clark, Classical shadows for quantum process tomography on near-term quantum computers, arXiv:2110.02965.

[66] J. Kunjummen, M. C. Tran, D. Carney, and J. M. Taylor, Shadow process tomography of quantum channels, arXiv:2110.03629.

[67] S. Hillmich, C. Hadfield, R. Raymond, A. Mezzacapo, and R. Wille, Decision diagrams for quantum measurements with shallow circuits, *Proceedings—2021 IEEE International Conference on Quantum Computing and Engineering, QCE* (IEEE, 2021), 24–34, 10.1109/QCE52317.2021.00018.

[68] H.-Y. Huang, and R. Kueng, and J. Preskill, Efficient Estimation of Pauli Observables by Derandomization, Phys. Rev. Lett. **127**, 030503 (2021).

[69] D. W. Leung, Choi's proof as a recipe for quantum process tomography, J. Math. Phys. (N.Y.) **44**, 528 (2003).

[70] Z. Wen and W. Yin, A feasible method for optimization with orthogonality constraints, Math. Program. **142**, 397 (2013).

[71] B. Jiang and Y. H. Dai, A framework of constraint preserving update schemes for optimization on Stiefel manifold, Math. Program. **153**, 535 (2015).

[72] J. Li, L. Fuxin, and S. Todorovic, Efficient Riemannian optimization on the Stiefel manifold via the Cayley transform, International Conference on Learning Representations, arXiv:2002.01113.

[73] S. Adhikary, S. Srinivasan, G. Gordon, and B. Boots, Expressiveness and learning of hidden quantum Markov models, in *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, edited by S. Chiappa and R. Calandra (PMLR, 2020), Vol. 108, pp. 4151–4161.

[74] H. D. Tagare, *Notes on optimization on Stiefel manifolds* (Yale University, New Haven, 2011).

[75] N. Boumal, *An Introduction to Optimization on Smooth Manifolds* (Cambridge University Press, Cambridge, England, 2023).

[76] R. Wiersema and N. Killoran, Optimizing quantum circuits with Riemannian gradient-flow, arXiv:2202.06976.

[77] I. A. Luchnikov, M. E. Krechetov, and S. N. Filippov, Riemannian geometry and automatic differentiation for optimization problems of quantum physics and quantum technologies, New J. Phys. **23**, 073006 (2021).

[78] H. Kadri, S. Ayache, R. Huusari, A. Rakotomamonjy, and L. Ralaivola, Partial trace regression and low-rank kraus decomposition, *37th International Conference on Machine Learning, ICML 2020* (2020), Vol. PartF168147-7, p. 4998, arXiv:2007.00935.

[79] M. Mohseni, A. T. Rezakhani, and D. A. Lidar, Quantum-process tomography: Resource analysis of different strategies, Phys. Rev. A **77**, 032322 (2008).

[80] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.130.150402 for details, which includes Refs. [81,82].

[81] W. W. Hager, Updating the inverse of a matrix, SIAM Rev. **31**, 221 (1989).

[82] S. Vaswani, A. Mishkin, I. Laradji, M. Schmidt, G. Gidel, and S. Lacoste-Julien, Painless stochastic gradient: Interpolation, line-search, and convergence rates, Adv. Neural Inf. Proc. Syst. **32** (2019).

[83] S. Ahmed, C. Sánchez Muñoz, F. Nori, and A. F. Kockum, Quantum State Tomography with Conditional Generative Adversarial Networks, Phys. Rev. Lett. **127**, 140502 (2021).

[84] S. Ahmed, C. Sánchez Muñoz, F. Nori, and A. F. Kockum, Classification and reconstruction of optical quantum states with deep neural networks, Phys. Rev. Res. **3**, 033278 (2021).

[85] R. Bhatia, *Positive Definite Matrices* (Princeton University Press, Princeton, NJ, 2009).

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

[87] F. Santosa and W. W. Symes, Linear inversion of band-limited reflection seismograms, SIAM J. Sci. Stat. Comput. **7**, 1307 (1986).

[88] R. Tibshirani, Regression shrinkage and selection via the lasso, J. R. Stat. Soc. Ser. B **58**, 267 (1996).

[89] 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).

[90] A. Hjorungnes, D. Gesbert, and D. P. Palomar, Unified theory of complex-valued matrix differentiation, in *2007 IEEE International Conference on Acoustics, Speech and Signal Processing—ICASSP '07* (2007), Vol. 3, pp. III–345–III–348, 10.1109/ICASSP.2007.366543.

[91] C. Shen, R. W. Heeres, P. Reinhold, L. Jiang, Y.-K. Liu, R. J. Schoelkopf, and L. Jiang, Optimized tomography of continuous variable systems using excitation counting, Phys. Rev. A **94**, 052327 (2016).

[92] B. O'Donoghue, E. Chu, N. Parikh, and S. Boyd, Conic optimization via operator splitting and homogeneous self-dual embedding, J. Optim. Theory Appl. **169**, 1042 (2016).

[93] S. Diamond and S. Boyd, CVXPY: A Python-embedded modeling language for convex optimization, J. Mach. Learn. Res. **17**, 1 (2016).

[94] A. Agrawal, R. Verschueren, S. Diamond, and S. Boyd, A rewriting system for convex optimization problems, J. Control Decision **5**, 42 (2018).
[95] M. S. ANIS *et al.*, *Qiskit: An open-source framework for quantum computing* (2023), 10.5281/zenodo.2573505.
[96] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds* (Princeton University Press, Princeton, NJ, 2009).
[97] L. Armijo, Minimization of functions having Lipschitz continuous first partial derivatives, Pac. J. Math. **16**, 1 (1966).
[98] D. P. Bertsekas and J. N. Tsitsiklis, Gradient convergence in gradient methods with errors, SIAM J. Optim. **10**, 627 (2000).
[99] R. Kleinberg, Y. Li, and Y. Yuan, An alternative view: When does SGD escape local minima?, in *Proceedings of the 35th International Conference on Machine Learning* (PMLR, 2018), Vol. 6, p. 4226.
[100] Y. N. Dauphin, R. Pascanu, C. Gulcehre, K. Cho, S. Ganguli, and Y. Bengio, Identifying and attacking the saddle point problem in high-dimensional non-convex optimization, in *Proceedings of the 27th International Conference on Neural Information Processing Systems—Volume 2*, NIPS'14 (MIT Press, Cambridge, MA, USA, 2014), pp. 2933–2941.
[101] I. Panageas, G. Piliouras, and X. Wang, First-order methods almost always avoid saddle points: The case of vanishing step-sizes, Adv. Neural Inf. Process. Syst. **32**, 1 (2019).
[102] S. S. Du, X. Zhai, B. Poczos, and A. Singh, Gradient descent provably optimizes over-parameterized neural networks, in *International Conference on Learning Representations* (2019).
[103] W. Ring and B. Wirth, Optimization methods on Riemannian manifolds and their application to shape space, SIAM J. Optim. **22**, 596 (2012).
[104] R. Fletcher, *Practical Methods of Optimization* (John Wiley & Sons, New York, 2013).
[105] R. W. Heeres, B. Vlastakis, E. Holland, S. Krastanov, V. V. Albert, L. Frunzio, L. Jiang, and R. J. Schoelkopf, Cavity State Manipulation Using Photon-Number Selective Phase Gates, Phys. Rev. Lett. **115**, 137002 (2015).
[106] T. Fösel, S. Krastanov, F. Marquardt, and L. Jiang, Efficient cavity control with SNAP gates, arXiv:2004.14256.
[107] R. Kumar, E. Barrios, C. Kupchak, and A. I. Lvovsky, Experimental Characterization of Bosonic Creation and Annihilation Operators, Phys. Rev. Lett. **110**, 130403 (2013).
[108] C. Kupchak, S. Rind, B. Jordaan, and E. Figueroa, Quantum process tomography of an optically-controlled kerr non-linearity, Sci. Rep. **5**, 16581 (2015).
[109] C. J. Wood, J. D. Biamonte, and D. G. Cory, Tensor networks and graphical calculus for open quantum systems, Quantum Inf. Comput. **15**, 759 (2015).
[110] A. Gilchrist, N. K. Langford, and M. A. Nielsen, Distance measures to compare real and ideal quantum processes, Phys. Rev. A **71**, 062310 (2005).
[111] I. Nechita, Z. Puchała, Ł. Pawela, and K. Zyczkowski, Almost all quantum channels are equidistant, J. Math. Phys. (N.Y.) **59**, 052201 (2018).
[112] Code and data for gradient-descent quantum process tomography is available at https://github.com/quantshah/gd-qpt.
[113] S. Srinivasan, S. Adhikary, J. Miller, B. Pokharel, G. Rabusseau, and B. Boots, Towards a trace-preserving tensor network representation of quantum channels, in *First Workshop on Quantum Tensor Networks in Machine Learning In conjunction with NeurIPS* (2021).
[114] O. Di Matteo, J. Gamble, C. Granade, K. Rudinger, and N. Wiebe, Operational, gauge-free quantum tomography, Quantum **4**, 364 (2020).
[115] E. Kiktenko, D. Norkin, and A. Fedorov, Confidence polytopes for quantum process tomography, New J. Phys. **23**, 123022 (2021).
[116] E. Hüllermeier and W. Waegeman, Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods, Mach. Learn. **110**, 457 (2021).
[117] R. Blume-Kohout, Robust error bars for quantum tomography, arXiv:1202.5270.
[118] P. Faist and R. Renner, Practical and Reliable Error Bars in Quantum Tomography, Phys. Rev. Lett. **117**, 010404 (2016).
[119] L. P. Thinh, P. Faist, J. Helsen, D. Elkouss, and S. Wehner, Practical and reliable error bars for quantum process tomography, Phys. Rev. A **99**, 052311 (2019).
[120] J. Bradbury, R. Frostig, P. Hawkins, M. J. Johnson, C. Leary, D. Maclaurin, G. Necula, A. Paszke, J. VanderPlas, S. Wanderman-Milne, and Q. Zhang, *JAX: Composable transformations of Python+NumPy programs* (2018).
[121] M. Hessel, D. Budden, F. Viola, M. Rosca, E. Sezener, and T. Hennigan, Optax: Composable gradient transformation and optimisation, in *JAX!* (2020).
[122] J. R. Johansson, P. D. Nation, and F. Nori, QuTiP: An open-source Python framework for the dynamics of open quantum systems, Comput. Phys. Commun. **183**, 1760 (2012).
[123] J. R. Johansson, P. D. Nation, and F. Nori, QuTiP: An open-source Python framework for the dynamics of open quantum systems, Comput. Phys. Commun. **184**, 1234 (2013).