


# Fast and Robust Quantum State Tomography from Few Basis Measurements

Daniel Stilck França ✉ 🏠 

QMATH, Department of Mathematical Sciences, University of Copenhagen, Denmark

Fernando G.S L. Brandão ✉ 🏠 

AWS Center for Quantum Computing, Pasadena, CA, USA

Institute for Quantum Information and Matter,

California Institute of Technology, Pasadena, CA, USA

Richard Kueng ✉ 🏠 

Institute for Integrated Circuits, Johannes Kepler University Linz, Austria

---

## Abstract

Quantum state tomography is a powerful but resource-intensive, general solution for numerous quantum information processing tasks. This motivates the design of robust tomography procedures that use relevant resources as sparingly as possible. Important cost factors include the number of state copies and measurement settings, as well as classical postprocessing time and memory. In this work, we present and analyze an online tomography algorithm designed to optimize all the aforementioned resources at the cost of a worse dependence on accuracy. The protocol is the first to give provably optimal performance in terms of rank and dimension for state copies, measurement settings and memory. Classical runtime is also reduced substantially and numerical experiments demonstrate a favorable comparison with other state-of-the-art techniques. Further improvements are possible by executing the algorithm on a quantum computer, giving a quantum speedup for quantum state tomography.

**2012 ACM Subject Classification** Theory of computation → Quantum computation theory; Hardware → Quantum technologies; Theory of computation → Quantum information theory; Mathematics of computing → Probabilistic inference problems

**Keywords and phrases** quantum tomography, low-rank tomography, Gibbs states, random measurements

**Digital Object Identifier** 10.4230/LIPIcs.TQC.2021.7

**Related Version** *Full Version:* <https://arxiv.org/abs/2009.08216>

**Funding** *Daniel Stilck França:* D.S.F. acknowledges financial support from VILLUM FONDEN via the QMATH Centre of Excellence (Grant no. 10059).

*Fernando G.S L. Brandão:* F.B. acknowledges funding from the US National Science Foundation (PHY1733907). The Institute for Quantum Information and Matter is an NSF Physics Frontiers Center.

*Richard Kueng:* R.K. acknowledges funding from the US National Science Foundation (PHY1733907). The Institute for Quantum Information and Matter is an NSF Physics Frontiers Center.

**Acknowledgements** We thank Chris Ferrie, David Gross, Thomas Grurl, Cécilia Lancien, Robert König, Oliver H. Schwarze and Joel Tropp for valuable input and helpful discussions.

## 1 Motivation

Quantum state tomography is the task of reconstructing a classical description of a quantum state from experimental data. This problem has a long and rich history [5] and remains a useful subroutine for building, calibrating and controlling quantum information processing devices. Over the last decade, unprecedented advances in the experimental control of



© Daniel Stilck França, Fernando G.S L. Brandão, and Richard Kueng;  
licensed under Creative Commons License CC-BY 4.0

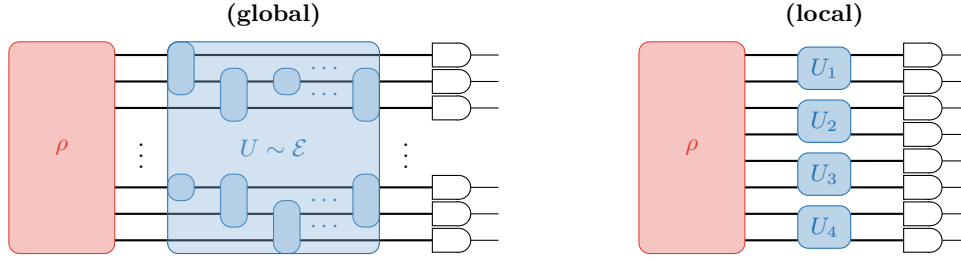
16th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2021).

Editor: Min-Hsiu Hsieh; Article No. 7; pp. 7:1–7:13



Leibniz International Proceedings in Informatics

Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany



■ **Figure 1** *Basis measurement primitive.* Global measurements (left) require implementing a global unitary that affects all  $n$  qudits prior to measuring in the computational basis. A  $k$ -local measurement primitive only allows for unitaries that affect groups of  $k$  (geometrically) local qudits; see the left-hand side for a visualization with  $n = 8$  and  $k = 2$ .

quantum architectures have pushed traditional estimation techniques to the limit of their capabilities. This is mainly due to a fundamental curse of dimension: the dimension of state space grows exponentially in the number of qudits, i.e. a quantum system comprised of  $n$   $d$ -dimensional qudits is characterized by a density matrix  $\rho$  of size  $D = d^n$ . The impact of this scaling behavior is further amplified by the probabilistic nature of quantum mechanics (“wave-function collapse”). Information about the state is only accessible via measuring the system. An informative quantum measurement is destructive and only yields probabilistic outcomes. Hence, many identically prepared samples of the quantum state are required to estimate even a single parameter of the underlying state. Characterizing the full state of a quantum system necessitates accurate estimation of many such parameters. Storing and processing the measurement data also requires substantial amounts of classical memory and computing power – another important practical bottleneck. To summarize: the curse of dimension and wave-function collapse have severe implications that necessitate the design of extremely resource-efficient protocols.

In this work, we focus on reconstructing the complete density matrix  $\rho$  from single-copy measurements. This is an actual restriction, as it excludes some of the most powerful tomography techniques known to this date [34, 19]. While very efficient in terms of state copies, these procedures are extremely demanding in terms of quantum hardware – an actual implementation would require exponentially long quantum circuits that act collectively on all the copies of the unknown state stored in a quantum memory.

We also adopt a measurement primitive that mimics the layout of modern quantum information processing devices. Apply a unitary  $U$  to the unknown state  $\rho \mapsto U\rho U^\dagger$  and perform measurements in the computational basis  $\{|i\rangle : i = 1, \dots, D\}$ . Fixing  $U$  and repeating this procedure many times allows for estimating the associated outcome distribution:

$$[p_U(\rho)]_i = \langle i|U\rho U^\dagger|i\rangle \quad \text{for } i = 1, \dots, D. \quad (1)$$

This outcome distribution characterizes the diagonal elements of  $U\rho U^\dagger$ . In general, access to a single diagonal is insufficient to determine  $\rho$  unambiguously. Instead, multiple repetitions of this basic measurement primitive are necessary. We refer to Fig. 1 for an illustration. Different ensembles  $\mathcal{E}$  of accessible unitary transformations give rise to different basis measurement primitives. When employed to perform state tomography – i.e. reconstruct an unknown state  $\rho$  up to accuracy  $\epsilon$  in trace distance – the following fundamental scaling laws apply to *any* (single-copy) basis measurement primitive and *any* tomographic procedure:

■ **Table 1** *Resource scaling for state tomography protocols based on global measurements (single copy)*: Here,  $D$  denotes the Hilbert space dimension,  $r$  is the rank of the target state and  $\epsilon$  is the desired precision (in trace distance). We have suppressed constants, as well as logarithmic dependencies in  $D$  and  $r$ . The first row summarizes known fundamental lower bounds, while the label “unknown” indicates a lack of rigorous theory support.

	meas. primitive	basis settings	state copies	runtime	memory
lower bounds	arbitrary	$\geq r$	$\geq Dr^2\epsilon^{-2}$	$\geq Dr^2\epsilon^{-2}$	$\geq Dr$
CS [40]	Haar	$r$	unknown	$D^4$	$D^3$
CS [27]	Clifford	$D^{2/3}r$	unknown	$D^4$	$D^3$
PLS [18]	2-design	$D$	$Dr^2\epsilon^{-2}$	$D^3$	$D^2$
this work	4-design	$r\epsilon^{-2}$	$Dr^2\epsilon^{-4}$	$D^2r^{5/2}\epsilon^{-5}$	$Dr\epsilon^{-2}$
this work	Clifford	$r^3\epsilon^{-2}$	$Dr^4\epsilon^{-4}$	$D^2r^6\epsilon^{-5}$	$Dr^2\epsilon^{-2}$

- (i) The *number of basis measurement settings*  $M$  must scale at least linearly with the (effective) target rank  $r = \text{rank}(\rho)$ :  $M = \Omega(r)$ . This corresponds to estimating a total of  $DM = \Omega(rD)$  parameters [21, 25].
- (ii) The *sampling rate*  $N$ , i.e. the number of independent state copies required to obtain sufficient data, must depend on rank, dimension and desired accuracy:  $N = \Omega(Dr^2/\epsilon^2)$  [19].
- (iii) The *classical storage*  $S$  is bounded by dimension times target rank:  $S = \Omega(rD)$ .

Constraint iii. follows from a simple parameter counting argument – specifying a general  $D \times D$ -matrix with rank  $r$  requires (order)  $rD$  parameters – while i. and ii. reflect fundamental limitations that have only been identified comparatively recently. These bounds cover three of the four most relevant cost parameters. For the last one we are not aware of a nontrivial rigorous lower bound:

- (iv) The *classical runtime* associated with processing the measurement data to produce an estimated state  $\sigma_*$  should be as fast as possible.

The last decade has seen the development of several procedures that provably optimize some of these four cost factors up to logarithmic factors in the ambient dimension. We refer to Table 1 for a detailed tabulation of resource requirements. For now, we content ourselves with emphasizing that existing procedures have been designed to either minimize the number of measurement settings (compressed sensing approaches [17, 32, 28]) or the required number of samples per measurement (least-squares approaches [37, 18]). Neither of these approaches seems to be well-suited for optimizing classical postprocessing memory and time. Finally, we point out that currently available quantum technologies are not perfect [35]. Practical tomography procedures should be *robust* with respect to imperfections, most notably state preparation and measurement errors.

## 2 Overview of results

In this work, we develop a robust algorithm for almost resource-optimal quantum state tomography from (single-copy) basis measurements that comes with rigorous convergence guarantees. The theoretical results are closely related to quantum state distinguishability [23, 22, 3, 33] and strongest for global measurement primitives (Fig. 1, left) that are sufficiently generic. In the regime of low target rank  $r$ , the proposed method improves upon state-of-the-art techniques at the cost of a worse dependence on target accuracy  $\epsilon$ . The actual numbers are summarized in Table 1. The required number of basis measurement setting matches

results from compressed sensing [17, 32, 28] – a technique that has been specifically designed to optimize this cost function – while the required number of state copies is comparable to projected least squares [37, 18] – which is known to be (almost) optimal in this regard. Classical runtime and memory cost are also reduced substantially. We also obtain rigorous results for  $k$ -local measurement primitives (Fig. 1, right), but the obtained theoretical numbers only become competitive if the locality parameter  $k$  is sufficiently large. We believe that this shortcoming is an artifact of poor constants and refer to App. B.4 of the extended version [9] for details.

## 2.1 Algorithm and theoretical runtime guarantee

The tomography algorithm – which we call *Hamiltonian updates* – is based on a variant of the versatile mirror-descent meta-algorithm [38, 10], see also [7]. Mirror descent and its cousin, matrix multiplicative weights, have led to considerable progress in algorithm design across several disciplines. Prominent examples include fast semidefinite programming solvers [20, 4, 31, 39, 8, 6, 7], quantum prediction techniques like shadow tomography [1], the online learning methods of [2] and the tomography protocol of [41]. The algorithm design is summarized in Algorithm 1. The key idea is to maintain and iteratively update a guess for the unknown state. The sequence of guess states is parametrized by Hamiltonians

$$\sigma_t = \frac{\exp(-H_t)}{\text{tr}(\exp(-H_t))} \quad \text{for } t = 0, 1, 2, \dots \quad (\text{Gibbs / thermal state})$$

and initialized to an infinite temperature state  $\sigma_0 = \mathbb{I}/D$  (maximum entropy principle). At each subsequent iteration, we choose a unitary rotation  $U \sim \mathcal{E}$  at random from a fixed ensemble, estimate the outcome distribution (1) of the rotated target state  $U\rho U^\dagger$  and compare it to the predicted outcome distribution  $U\sigma_t U^\dagger$  of the current guess state. If the two outcome distributions differ by more than mere statistical fluctuations,  $\sigma_t$  is an inadequate guess for  $\rho$ .

We then update the guess state  $\sigma_t \mapsto \sigma_{t+1}$  by including a small energy penalty in the associated Hamiltonian that penalizes the observed mismatch and repeat. Heuristically, it is reasonable to expect that this update rule makes progress as long as each newly selected basis provides actionable advice, i.e. discrepancies in the outcome distributions. As we prove in App. A of the extended version [9] that we indeed make progress in relative entropy. Things get more interesting when this is not the case. Predicted and estimated outcome distribution can be very close for two reasons (i): the current iterate  $\sigma_t$  is close to the unknown target  $\rho$  (*convergence*); (ii) the current basis measurement cannot properly distinguish between  $\sigma_t$  and  $\rho$ , even though they are still far apart (*false positive*). It is imperative to protect against wrongfully terminating the procedure due to the occurrence of a false positive. Hamiltonian Updates (Algorithm 1) suppresses the likelihood of wrongfully terminating by checking closeness in (up to)  $L$  additional random bases. The required size of such a control loop depends on the measurement primitive. Broadly speaking, generic measurement ensembles – like Haar-random unitary transformations – are very unlikely to produce false positives; while highly structured ensembles – like mutually unbiased bases – can be much more susceptible. The following relation introduces two ensemble-dependent summary parameters that capture this effect:

$$\Pr_{U \sim \mathcal{E}} [\|p_U(\rho) - p_U(\sigma_t)\|_{\ell_1} \geq \theta_{\mathcal{E}}(\rho, \sigma_t) \|\rho - \sigma_t\|_2] \geq \tau_{\mathcal{E}}(\rho, \sigma_t). \quad (2)$$

The parameter  $\theta_{\mathcal{E}}(\rho, \sigma_t)$  relates an observed discrepancy in outcome distributions (measured in  $\ell_1$  distance) to the Frobenius distance in state space. As detailed below, it captures the minimal progress we can expect from a successful update  $\sigma_t \mapsto \sigma_{t+1}$ . The second parameter

---

**Algorithm 1** *Hamiltonian Updates for quantum state tomography.*


---

**Input:** error tolerance  $\epsilon$ , number of loops  $L$ .  
**Initialize:**  $t = 0$ ,  $H_t = 0$ , CONVERGENCE=FALSE  
**while** CONVERGENCE=FALSE **do**  
    compute  $\sigma_t = \exp(-H_t)/\text{tr}(\exp(-H_t))$  ▷ current guess for the state  $\rho$   
    select random basis measurement  $\{U|i\rangle\langle i|U^\dagger\}$   
    compute outcome statistics  $[p_i]$  of  $\sigma_t$  ▷ classical computation  
    estimate outcome statistics  $[q_i]$  of  $\rho$  ▷ quantum measurement  
    **check** if  $[p_i]$  and  $[q_i]$  are  $\epsilon$ -close in  $\ell_1$  distance  
    **if** NO **then** set  $P = \sum_{p_i > q_i} |i\rangle\langle i|$  ▷ collect outcomes for which  $p_i > q_i$   
    Set  $\eta = \frac{1}{8}\|p - q\|_{\ell_1}$   
     $H_{t+1} \leftarrow H_t + \eta U^\dagger P U$  ▷ energy penalty for mismatch (in this basis)  
    update  $\sigma_{t+1} = \exp(-H_{t+1})/\text{tr}(\exp(-H_{t+1}))$   
     $t \leftarrow t + 1$  ▷ update counter of number of iterations  
    **else if** YES **then** ▷ current guess may be close to  $\rho$   
    check  $L$  additional random bases ▷ suppress likelihood of false positives  
    **if**  $\ell_1$  distance is always  $< \epsilon$  **then** ▷ current guess is likely to be close  
        **set** CONVERGENCE=TRUE  
    **end if**  
    **end if**  
**end while**  
**Output:**  $H_t$

---

$\tau_{\mathcal{E}}(\rho, \sigma_t)$  lower bounds the probability of observing an outcome discrepancy that appropriately reflects the current stage of convergence. This parameter controls the size of the control loop. It is desirable to choose both parameters as large as possible, but there is a trade-off (making  $\theta_{\mathcal{E}}(\rho, \sigma_t)$  larger necessarily diminishes  $\tau_{\mathcal{E}}(\rho, \sigma_t)$ ) and both depend heavily on the measurement ensemble. One of our main theoretical contributions is a rigorous convergence guarantee for Hamiltonian updates (Algorithm 1) that only depends on the ambient dimension  $D$ , the target rank  $r = \text{rank}(\rho)$ , as well as the worst-case ensemble parameters

$$\theta_{\mathcal{E}}(\rho) = \max_{\sigma \text{ state}} \theta_{\mathcal{E}}(\rho, \sigma) \quad \text{and} \quad \tau_{\mathcal{E}}(\rho) = \max_{\sigma \text{ state}} \tau_{\mathcal{E}}(\rho, \sigma). \quad (3)$$

► **Theorem 1 (informal statement).** *Fix a measurement primitive  $\mathcal{E}$ , a desired accuracy  $\epsilon$  and let  $\rho$  be a rank- $r$  target state. With high probability, Algorithm 1 requires at most  $T = \mathcal{O}(r \log(D)/(\theta_{\mathcal{E}}(\rho)\epsilon)^2)$  steps – each with a control loop of size  $L = \mathcal{O}(\log(T)/\tau_{\mathcal{E}}(\rho))$  – to produce an output  $\sigma_*$  that obeys  $\|\rho - \sigma_*\|_1 \leq \epsilon$ .*

This convergence guarantee is also stable with respect to imperfect implementations. In particular, we only need to estimate measurement outcome statistics to a certain degree of accuracy:  $\mathcal{O}(Dr/(\theta_{\mathcal{E}}(\rho)\epsilon)^2)$  measurement repetitions suffice for each basis. This implies that the total number of measurement settings and state copies are bounded by

$$M = TL \simeq \mathcal{O}(r \log(D)/(\tau_{\mathcal{E}}(\rho)\theta_{\mathcal{E}}(\rho)^2\epsilon^2)) \quad (\text{measurement settings}), \quad (4)$$

$$N \simeq \mathcal{O}(Dr^2 \log(D)/(\tau_{\mathcal{E}}(\rho)\theta_{\mathcal{E}}(\rho)^4\epsilon^4)) \quad (\text{sample complexity}). \quad (5)$$

To increase readability, we have suppressed the logarithmic contribution in  $T$ .

## 2.2 Connections to quantum state distinguishability

The bounds for  $M$  in Eq. (4) and  $N$  in Eq. (5) are characterized by worst-case ensemble parameters (2). These are intimately related to quantum state distinguishability: how good is a fixed measurement primitive  $\mathcal{E}$  at distinguishing state  $\rho$  from state  $\sigma$  in the single-shot limit? Ambainis and Emerson [3] showed that the optimal probability of successful discrimination is given by  $p_{\text{succ}} = \frac{1}{2} + \frac{1}{4} \mathbb{E}_{U \sim \mathcal{E}} \|p_U(\rho) - p_U(\sigma)\|_{\ell_1}$  and achieved by the maximum likelihood rule, see also [33]. It is possible to relate this bias to the Frobenius distance in state space:

$$\mathbb{E}_{U \sim \mathcal{E}} \|p_U(\rho) - p_U(\sigma)\|_{\ell_1} \geq \lambda_{\mathcal{E}}(\rho, \sigma) \|\rho - \sigma\|_2.$$

The proportionality constant  $\lambda_{\mathcal{E}}(\rho, \sigma)$  measures how well the measurement primitive is equipped to distinguish  $\rho$  from  $\sigma$ . It is closely related to the ensemble parameters defined in Eq. (2) and has been the subject of considerable attention in the community. Tight bounds have been derived for a variety of measurement primitives, such as Haar random unitaries and approximate 4-designs [3, 33], random Clifford unitaries [29] and  $k$ -local (approximate) 4-designs [30]. Simple probabilistic arguments allow for converting these assertion into lower bounds on both  $\theta_{\mathcal{E}}(\rho)$  and  $\tau_{\mathcal{E}}(\rho)$ . Inserting these bounds into Eq. (4) and Eq. (5) then implies the measurement and sample complexity assertions advertised in Table 1. We refer to Appendix B of the extended version [9] for a detailed case-by-case analysis and content ourselves here with an overview. We start with the strongest measurement primitive: Haar random unitaries and approximate 4-designs achieve  $\theta_{\mathcal{E}}(\rho), \tau_{\mathcal{E}}(\rho) = \text{const}$  for any target state. Hence,  $M = \mathcal{O}(r \log(D)/\epsilon^2)$  basis settings and  $N = \mathcal{O}(Dr^2 \log(D)/\epsilon^4)$  state copies suffice. Clifford random measurements achieve  $\theta_{\mathcal{E}}(\rho) \sim r^{-\frac{1}{2}}, \tau_{\mathcal{E}}(\rho) \sim r^{-2}$ . That is, they only have a worse dependency on the rank, but perform as well as Haar measurements in terms of the ambient dimension. On the other hand, more local measurement settings defined by unitaries acting on at most  $k$  qubits have  $\theta_{\mathcal{E}}(\rho) \sim \exp(-\mathcal{O}(n/k)), \tau_{\mathcal{E}}(\rho) \sim \exp(-\mathcal{O}(n/k))$ , showing an (exponentially) worse dependency on the number of qudits when compared to Haar measurements. Empirical studies below do, however, suggest a much more favorable performance in practice.

This scaling highlights both a core strength and a core weakness of Hamiltonian updates. In terms of dimension  $D$  and rank  $r$ , these numbers saturate fundamental lower bounds on any tomographic procedure up to a logarithmic factor. However, the number of measurement settings also depends inverse quadratically on the accuracy. Furthermore, the accuracy enters as  $\epsilon^{-4}$ , not  $\epsilon^{-2}$  in the sample complexity. Thus, high accuracy solutions do not only require many samples, but also many basis measurement settings. This drawback is a consequence of a “curse of mirror descent (or multiplicative weights)”. These meta-algorithms are very efficient in terms of problem dimension, but scale comparatively poorly in accuracy [4]. However, inverse polynomial scaling in accuracy  $\epsilon$  is an unavoidable feature of quantum state tomography. Hence, tomography is a reasonable setting to apply algorithms that trade dimensional dependency for accuracy. Moreover, for most applications, it suffices to recover the state up to precision  $\epsilon = \mathcal{O}(\text{polylog}(D)^{-1})$ .

## 3 Summary and comparison to relevant existing work

We propose a variant of mirror descent [38, 10] to obtain resource-efficient algorithms for quantum state tomography. In recent years, mirror descent and its cousins have been extensively used to obtain fast SDP solvers [20, 4, 31, 39, 8, 6, 7], to develop prediction algorithms like shadow tomography [1], the online learning methods of [2] and the tomography protocol of [41]. Key advantages are resource efficiency, as well as intrinsic resilience towards

noise. Empirical studies summarized in Fig. 2 confirm these theoretical assertions. A downside is, however, that the number of iterations may depend on the desired target accuracy  $\epsilon$ . We focus on obtaining a  $\epsilon$ -approximation in trace distance of a  $D$ -dimensional state  $\rho$  from (random) basis measurements on i.i.d. copies (*global classical description*). Our goal is to optimize the different resources required for that task. These include the number of state copies (sample complexity), the cost for processing measurement data (classical postprocessing), as well as the associated memory cost. The multipronged resource efficiency of our results becomes particularly pronounced if the underlying target state has (approximately) low-rank  $r \ll D$ . This is a natural assumption in most applications, but can also be relaxed to states with low Rényi entropy, see App. G of the extended version [9].

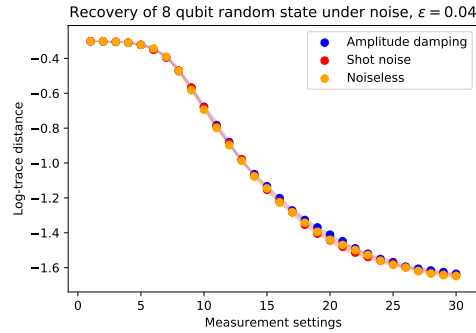
Thus, our results are similar in spirit to the tomography algorithms based on compressed sensing (CS) [17, 32, 14, 36, 28], or projected least squares (PLS) [37, 18]. These also focus on rigorous and (nearly) optimal sample complexity in the low-rank regime combined with efficient postprocessing. Table 1 summarizes the resources required for these protocols, as well as our new results. These compare favorably with existing methods. We note that for approximate 4-design measurements, both sample complexity and memory – as functions of  $D$  and  $r$  – are essentially optimal [34, 19]. Compared to existing approaches, we obtain significant savings in both runtime and memory. Moreover, as pointed out in [41], there are also qualitative advantages.

Current schemes that minimize the number of basis settings [40, 27] are only known to do so with perfect knowledge of the underlying measurement outcomes. This will never be the case in practice, due to statistical fluctuations. Thus, to the best of our knowledge, our work is the first to rigorously obtain recovery guarantees with imperfect knowledge of outcomes and basis settings that only scale logarithmically with the ambient dimension and linearly with rank (albeit with the extra  $\epsilon$  dependency).

The focus of this work differs from other recent applications of mirror descent to quantum learning [2, 1, 6]. Broadly speaking, these works focus on obtaining a classical description of the state – a shadow – that approximately reproduces a fixed set of target observables. This is a different and weaker form of recovery. Moreover, these works prioritize sample complexity, not necessarily classical postprocessing resources. Minimizing these classical resources is a core focus of this work.

Having said this, the idea of using (variants of) mirror descent for quantum state (and process) tomography is not completely new. Similar ideas were proposed in Refs. [13, 16] and have been experimentally tested [11, 24]. More recently, Youssry, Tomamichel and Ferrie proposed and analyzed state tomography based on matrix exponentiated gradient descent [41]. They focused on the practically relevant case of local (single-qubit) Pauli measurements and established convergence to the target state as the number of samples goes to infinity. They also pointed out conceptual advantages, such as online implementation and noise-robustness. The results presented here add to this promising picture. We equip (a variant of) mirror descent with rigorous performance guarantees in the non-asymptotic setting, optimize actual implementations and establish robustness in a more general setting. Moreover, our results apply to any measurement procedure that is capable of distinguishing arbitrary pairs of quantum states.

We also want to point out that the method presented here could also be implemented on a quantum computer. This would result in substantial runtime savings – a quantum speedup for quantum state tomography. Suppressing polylogarithmic terms, a runtime of order  $\tilde{O}(D^{\frac{3}{2}}r^3\epsilon^{-9})$  suffices to obtain a *classical* description of the target state. We refer to App. E of the extended version [9] for details and proofs. To the best of our knowledge, this



■ **Figure 2** *Convergence of Algorithm 1 for different noise models.* We consider Haar-random global measurements of a 8-qubit pure target state with target accuracy  $\epsilon = 0.04$ . Different colors track convergence for different noise models: (blue) amplitude damping noise with parameter  $\epsilon/4$ ; (red) white noise with standard deviation  $\epsilon/4$  that mimics one-shot noise; (orange) zero noise. All logarithms are base 10 and the shaded areas indicate 25% and 75% quartiles, estimated from 20 samples.

is the first quantum speedup for low-rank tomography beyond the results of Kerenidis and Prakash [26] which cover pure, real target states ( $r = 1$ ) exclusively and work under the stronger assumption of access to a controlled unitary that prepares the state.

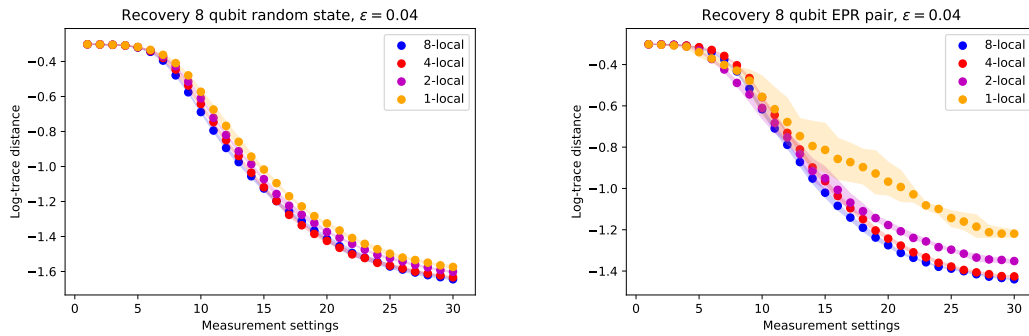
Finally, we want to emphasize that the proposed reconstruction procedure can be empowered by advantageous measurement structure. Storage-efficiency stems from the fact that we can keep track of the Hamiltonian – not the associated Gibbs state – which inherits structure from the underlying measurement procedure. Runtime savings are achieved by only exponentiating the Hamiltonian approximately and exploiting fast matrix-vector multiplication. We refer to App. D of the extended version [9] for details and content ourselves here with a vague, but instructive, analogy: View Algorithm 1 as an adaptive cool-down procedure. We start with a Gibbs state at infinite temperature and, at each step, we cool down the system in a controlled fashion that guides the thermal state towards the unknown target. Importantly, each update is small and the number of total cooling steps is also benign. Hence, we never truly leave the moderate temperature regime and avoid computational bottlenecks that typically only arise at low temperatures. In turn, the output of our algorithm is in the form of a Hamiltonian whose Gibbs state is close to the target state. A list of Gibbs state eigenvalues and corresponding eigenvectors can be obtained by block Krylov iterations, see App. F of the extended version [9]. Runtime and memory cost of this conversion procedure can never exceed those of Algorithm 1.

## 4 Numerical experiments

We complement our theoretical assertions with empirical test evaluations for systems comprised of up to 10 qubits. The results look promising and may establish Algorithm 1 as a practical tool for quantum state tomography. We remark that our numerical implementation has two additional details when compared with the one described in Algorithm 1. Although these modifications do not change the asymptotic runtime analysis of the algorithm, they can substantially reduce runtime and sample complexity in practice.

The first alteration we do is to recycle the last measurement data after a successful update. More precisely, after each update  $\sigma_t \rightarrow \sigma_{t+1}$ , we then check if the new iteration  $\sigma_{t+1}$  is still distinguishable from  $\rho$  under the previous measurement basis. Only if this is





■ **Figure 3** Convergence of Algorithm 1 for different measurement localities. Different colors track convergence (in logarithmic trace distance) for 8-qubit basis measurements with different localities and target accuracy  $\epsilon = 0.04$ . Individual basis measurements are subject to white noise with standard deviation  $\epsilon/4$ . (Left) Reconstruction of a generic pure target state. (Right) Reconstruction of a highly structured target state (EPR/Bell state). All logarithms are base 10 and the shaded area indicate 25% and 75% quartiles, estimated from 20 samples.

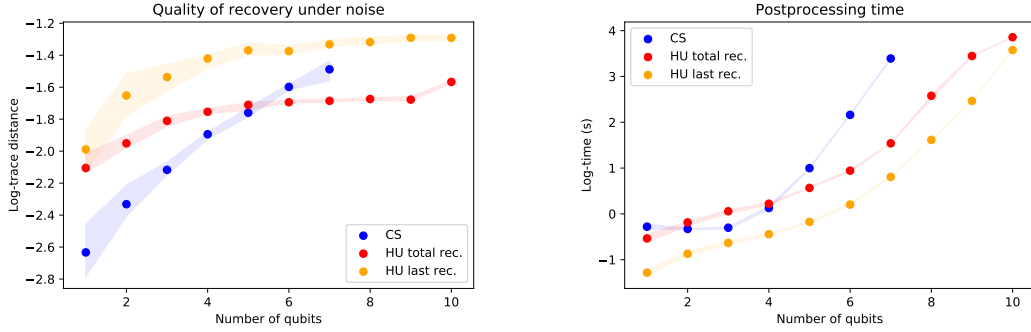
not the case, we move on to sample a new measurement setting. Otherwise, we re-use the already known measurement basis to drive another update in the same direction. We observe empirically that this minor modification has very desirable consequences. It leads to a much faster convergence throughout early stages of the algorithm and, by extension, reduces the number of required measurement settings significantly.

What is more, this recycling procedure cannot change the asymptotic scaling of the algorithm. To see this, note that the modification can only affect postprocessing complexity. Indeed, it clearly does not require us to sample more states or measurement settings. Finding another violation can only bring us closer to the state in relative entropy. And the postprocessing time can only double in the worst case. This worst case scenario happens when after updating every basis once, we have already converged in that basis and checking again does not lead to further convergence. We will refer to this variation as the *last step recycling strategy*. It is explained in detail in the appendix (Algorithm 2 of the extended version [9]).

Other variations of this basic principle come to mind. For instance, we need not stop at testing the current iteration against the previous measurement basis. We can also test it against all measurements that have already accumulated. This variation can further reduce the (total) number of basis settings required to converge. Fig. 4 confirms this intuition. However, this strategy comes at the expense of an increase in the computational complexity of the postprocessing. We refer to this strategy as the *complete recycling strategy*.

Apart from these practical improvements, we have also tested desirable fundamental properties of Algorithm 1. Chief among them is noise resilience. As advertised in Sec. 2 and proved in App. C of the extended version [9], the performance of the algorithm under arbitrary noise of bounded intensity is indistinguishable from the noiseless case. This feature is empirically confirmed by Fig. 2. For detecting a random pure state on 8 qubits, different noise sources – such as shot noise and amplitude damping – affect convergence in a very mild fashion only (*robustness*). It is also interesting to note that the convergence in trace norm appears to be polynomial for the first measurements and then switches to an exponential phase.

Another interesting figure of merit is measurement locality. The assertions that underpin Algorithm 1 do, in principle, extend to local measurement primitives. But, as detailed in App. B.4 of the extended version [9], the resulting numbers look rather pessimistic and scale unfavorably with measurement locality  $k$ . Empirical studies do paint a much more favorable picture, see Fig. 3. The two subplots address reconstruction of a typical 8-qubit target state (left), as well as a highly structured one (right). A direct comparison lends credence to a conjecture voiced in App. B.4 of the extended version [9] below: generic or typical states are easier to reconstruct with local measurements than highly structured ones.



**Figure 4** Comparison between Algorithm 1 (HU) and compressed sensing (CS) tomography. (Left) Reconstruction of a random  $n$ -qubit pure state from 15 globally random basis measurements corrupted by amplitude damping noise ( $p = 0.005$ ). Different colors track the logarithmic trace distance error achieved by either CS (blue) or variants of HU (orange and red) for  $\epsilon = 0.01$ . Shaded regions indicate the 25 – 75 percentiles over 20 independent runs. (Right) Empirical runtime for executing (naive implementations of) the three different reconstruction procedures on a conventional laptop. CVX [12] – a standard solver for semidefinite programs – could not go beyond 7 qubits.

Last but not least, we compare Algorithm 1 against the state of the art regarding tomography from very few basis measurements. Compressed sensing (CS) [17, 14, 27, 28] has been designed to fit a low rank solution to the observed measurement data by also minimizing the trace norm over the cone of positive semidefinite matrices ( $X \succeq 0$ ):

$$\text{minimize}_{X \succeq 0} \quad \text{tr}(X) \quad \text{subject to} \quad \sum_{i=1}^M \|\hat{p}_{U_i}(\rho) - p_{U_i}(X)\|_{\ell_2}^2 \leq \epsilon. \quad (6)$$

Fig. 4 compares Algorithm 1 with compressed sensing (CS). CS is contingent on solving a semidefinite program. We used CVX [12], a standard SDP solver, in Python. Algorithm 1 has also been implemented in Python. Open source code is available at [15]. We see that Hamiltonian Updates is more noise-resilient than CS. The rightmost plot also underscores the importance of memory improvements. A high-end desktop computer already struggles to solve SDP (6) for 8 qubits (even though the extrapolated computation time Fig. 4 still seems reasonable), while 10 qubits (and more) have not been a problem for Algorithm 1. We believe that Fig. 4 conveys both quantitative and qualitative advantages of Hamiltonian Updates over CS methods. This seems particularly noteworthy, because we compared both procedures for pure target states ( $\text{rank}(\rho) = 1$ ) – a use-case tailor-made for CS approaches. We also stress that the implementation of the algorithm used to generate this data was not optimized, there is room for further improvements.

Let us conclude with the most important take-away from Figs. 2, 3 and 4. The theoretical assertions from Sec. 2 carry over to practice. Moreover, recycling of data ensures that the number of measurement settings remains small even if we try to characterize the state up to

high precision. Our theoretical results suggest that order  $10^5$  algorithm iterations, and thus also measurement settings, might be required to obtain a  $\epsilon = 10^{-2}$ -approximation of a pure state in dimension  $D = 2^{10}$ . But our numerics demonstrate that already order  $10^1$  suffice to achieve convergence. The main theoretical drawbacks of Algorithm 1 – most notably, the poor scaling in accuracy – may be a non-issue in practical use cases. These findings establish our algorithm as a rare instance of a method that is provably (essentially) optimal and has a competitive performance in practice.

---

## References

- 1 Scott Aaronson. Shadow tomography of quantum states. In *STOC'18—Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, pages 325–338. ACM, New York, 2018. doi:10.1145/3188745.3188802.
- 2 Scott Aaronson, Xinyi Chen, Elad Hazan, Satyen Kale, and Ashwin Nayak. Online learning of quantum states. *Journal of Statistical Mechanics: Theory and Experiment*, 2019(12):124019, 2019. doi:10.1088/1742-5468/ab3988.
- 3 Andris Ambainis and Joseph Emerson. Quantum t-designs: t-wise independence in the quantum world. In *22nd Annual IEEE Conference on Computational Complexity (CCC 2007), 13-16 June 2007, San Diego, California, USA*, pages 129–140. IEEE Computer Society, 2007. doi:10.1109/CCC.2007.26.
- 4 Sanjeev Arora and Satyen Kale. A combinatorial, primal-dual approach to semidefinite programs. *J. ACM*, 63(2):12:1–12:35, 2016. doi:10.1145/2837020.
- 5 K Banaszek, M Cramer, and D Gross. Focus on quantum tomography. *New J. Phys.*, 15(12):125020, 2013. doi:10.1088/1367-2630/15/12/125020.
- 6 Fernando G. S. L. Brandão, Amir Kalev, Tongyang Li, Cedric Yen-Yu Lin, Krysta M. Svore, and Xiaodi Wu. Quantum SDP solvers: large speed-ups, optimality, and applications to quantum learning. In *46th International Colloquium on Automata, Languages, and Programming*, volume 132 of *LIPICs. Leibniz Int. Proc. Inform.*, pages Art. No. 27, 14. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2019.
- 7 Fernando G. S. L. Brandão, Richard Kueng, and Daniel Stilck França. Faster quantum and classical SDP approximations for quadratic binary optimization. *preprint arXiv:1909.04613*, 2019.
- 8 Fernando G. S. L. Brandão and Krysta M. Svore. Quantum speed-ups for solving semidefinite programs. In Chris Umans, editor, *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017, Berkeley, CA, USA, October 15-17, 2017*, pages 415–426. IEEE Computer Society, 2017. doi:10.1109/FOCS.2017.45.
- 9 Fernando G.S.L. Brandão, Richard Kueng, and Daniel Stilck França. Fast and robust quantum state tomography from few basis measurements, 2020. arXiv:2009.08216v2. arXiv:2009.08216.
- 10 Sébastien Bubeck. Convex optimization: Algorithms and complexity. *Found. Trends Mach. Learn.*, 8(3-4):231–357, 2015. doi:10.1561/22000000050.
- 11 Robert J. Chapman, Christopher Ferrie, and Alberto Peruzzo. Experimental demonstration of self-guided quantum tomography. *Phys. Rev. Lett.*, 117:040402, July 2016. doi:10.1103/PhysRevLett.117.040402.
- 12 Inc. CVX Research. CVX: Matlab software for disciplined convex programming, version 2.0. <http://cvxr.com/cvx>, 2012.
- 13 Christopher Ferrie. Self-guided quantum tomography. *Phys. Rev. Lett.*, 113:190404, November 2014. doi:10.1103/PhysRevLett.113.190404.
- 14 Steven T Flammia, David Gross, Yi-Kai Liu, and Jens Eisert. Quantum tomography via compressed sensing: error bounds, sample complexity and efficient estimators. *New J. Phys.*, 14(9):095022, 2012. doi:10.1088/1367-2630/14/9/095022.
- 15 Daniel Stilck Franca. Hamiltonian updates tomography. [https://github.com/dsfranca/hamiltonian\\_updates\\_tomography](https://github.com/dsfranca/hamiltonian_updates_tomography), 2020.

- 16 Christopher Granade, Christopher Ferrie, and Steven T Flammia. Practical adaptive quantum tomography. *New J. Phys.*, 19(11):113017, November 2017. doi:10.1088/1367-2630/aa8fe6.
- 17 David Gross, Yi-Kai Liu, Steven T. Flammia, Stephen Becker, and Jens Eisert. Quantum state tomography via compressed sensing. *Phys. Rev. Lett.*, 105:150401, 2010. doi:10.1103/PhysRevLett.105.150401.
- 18 Madalin Guță, Jonas Kahn, Richard Kueng, and Joel A Tropp. Fast state tomography with optimal error bounds. *J. Phys. A*, 53(20):204001, 2020. doi:10.1088/1751-8121/ab8111.
- 19 Jeongwan Haah, Aram Wettroth Harrow, Zhengfeng Ji, Xiaodi Wu, and Nengkun Yu. Sample-optimal tomography of quantum states. *IEEE Trans. Inf. Theory*, 63(9):5628–5641, 2017. doi:10.1109/TIT.2017.2719044.
- 20 Elad Hazan. *Efficient algorithms for online convex optimization and their applications*. PhD thesis, Princeton University, 2006.
- 21 Teiko Heinosaari, Luca Mazzarella, and Michael M. Wolf. Quantum tomography under prior information. *Commun. Math. Phys.*, 318(2):355–374, 2013. doi:10.1007/s00220-013-1671-8.
- 22 Carl W. Helstrom. Quantum detection and estimation theory. *J. Statist. Phys.*, 1:231–252, 1969. doi:10.1007/BF01007479.
- 23 Alexander S. Holevo. Statistical decision theory for quantum systems. *J. Multivariate Anal.*, 3:337–394, 1973. doi:10.1016/0047-259X(73)90028-6.
- 24 Zhibo Hou, Jun-Feng Tang, Christopher Ferrie, Guo-Yong Xiang, Chuan-Feng Li, and Guang-Can Guo. Experimental realization of self-guided quantum process tomography. *Phys. Rev. A*, 101:022317, February 2020. doi:10.1103/PhysRevA.101.022317.
- 25 Michael Kech and Michael M. Wolf. Constrained quantum tomography of semi-algebraic sets with applications to low-rank matrix recovery. *Inf. Inference*, 6(2):171–195, 2017. doi:10.1093/imaiai/iaw019.
- 26 Iordanis Kerenidis and Anupam Prakash. A Quantum Interior Point Method for LPs and SDPs. *ACM Transactions on Quantum Computing*, 1(1):1–32, December 2020. doi:10.1145/3406306.
- 27 Richard Kueng. Low rank matrix recovery from few orthonormal basis measurements. In *2015 International Conference on Sampling Theory and Applications (SampTA)*, pages 402–406, 2015.
- 28 Richard Kueng, Holger Rauhut, and Ulrich Terstiege. Low rank matrix recovery from rank one measurements. *Appl. Comput. Harmon. Anal.*, 42(1):88–116, 2017. doi:10.1016/j.acha.2015.07.007.
- 29 Richard Kueng, Huangjun Zhu, and David Gross. Distinguishing quantum states using Clifford orbits. *preprint arXiv:1609.08595*, 2016.
- 30 Cécilia Lancien and Andreas Winter. Distinguishing multi-partite states by local measurements. *Comm. Math. Phys.*, 323(2):555–573, 2013. doi:10.1007/s00220-013-1779-x.
- 31 James R. Lee, Prasad Raghavendra, and David Steurer. Lower bounds on the size of semidefinite programming relaxations. In Rocco A. Servedio and Ronitt Rubinfeld, editors, *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing, STOC 2015, Portland, OR, USA, June 14-17, 2015*, pages 567–576. ACM, 2015. doi:10.1145/2746539.2746599.
- 32 Yi-Kai Liu. Universal low-rank matrix recovery from Pauli measurements. In J. Shawe-Taylor, R. S. Zemel, P. L. Bartlett, F. Pereira, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 24*, pages 1638–1646. Curran Associates, Inc., 2011. URL: <http://papers.nips.cc/paper/4222-universal-low-rank-matrix-recovery-from-pauli-measurements.pdf>.
- 33 William Matthews, Stephanie Wehner, and Andreas Winter. Distinguishability of quantum states under restricted families of measurements with an application to quantum data hiding. *Comm. Math. Phys.*, 291(3):813–843, 2009. doi:10.1007/s00220-009-0890-5.
- 34 Ryan O’Donnell and John Wright. Efficient quantum tomography. In Daniel Wichs and Yishay Mansour, editors, *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2016, Cambridge, MA, USA, June 18-21, 2016*, pages 899–912. ACM, 2016. doi:10.1145/2897518.2897544.

- 35 John Preskill. Quantum Computing in the NISQ era and beyond. *Quantum*, 2:79, 2018. doi:10.22331/q-2018-08-06-79.
- 36 Carlos A. Ríofrio, David Gross, Steven T. Flammia, Thomas Monz, Daniel Nigg, Rainer Blatt, and Jens Eisert. Experimental quantum compressed sensing for a seven-qubit system. *Nat. Commun.*, 8(1), 2017. doi:10.1038/ncomms15305.
- 37 Takanori Sugiyama, Peter S. Turner, and Mio Muraō. Precision-guaranteed quantum tomography. *Phys. Rev. Lett.*, 111:160406, 2013. doi:10.1103/PhysRevLett.111.160406.
- 38 Koji Tsuda, Gunnar Rätsch, and Manfred K. Warmuth. Matrix exponentiated gradient updates for on-line learning and Bregman projection. *J. Mach. Learn. Res.*, 6:995–1018, 2005. URL: <http://jmlr.org/papers/v6/tsuda05a.html>.
- 39 Joran van Apeldoorn, András Gilyén, Sander Gribling, and Ronald de Wolf. Quantum SDP-solvers: Better upper and lower bounds. In Chris Umans, editor, *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017, Berkeley, CA, USA, October 15-17, 2017*, pages 403–414. IEEE Computer Society, 2017. doi:10.1109/FOCS.2017.44.
- 40 Vladislav Voroninski. Quantum tomography from few full-rank observables. *preprint arXiv:1309.7669*, 2013.
- 41 Akram Youssry, Christopher Ferrie, and Marco Tomamichel. Efficient online quantum state estimation using a matrix-exponentiated gradient method. *New J. Phys.*, 21(3):033006, 2019. doi:10.1088/1367-2630/ab0438.