Minimax Quantum Tomography: Estimators and Relative Entropy Bounds

Christopher Ferrie and Robin Blume-Kohout

Phys. Rev. Lett. **116**, 090407 — Published 4 March 2016

DOI: 10.1103/PhysRevLett.116.090407

Minimax quantum tomography: estimators and relative entropy bounds

Christopher Ferrie[1,2] and Robin Blume-Kohout[3]

[1] *Center for Quantum Information and Control, University of New Mexico, Albuquerque, New Mexico 87131-0001*

[2] *Centre for Engineered Quantum Systems, School of Physics, The University of Sydney, Sydney, NSW, Australia*

[3] *Center for Computing Research (Sandia National Laboratories), Albuquerque, New Mexico 87185*

(Dated: February 3, 2016)

A *minimax* estimator has the minimum possible error ("risk") in the worst case. We construct the first minimax estimators for quantum state tomography with relative entropy risk. The minimax risk of non-adaptive tomography scales as $O(1/\sqrt{N})$, in contrast to that of classical probability estimation which is $O(1/N)$, where N is the number of copies of the quantum state used. We trace this deficiency to *sampling mismatch*: future observations that determine risk may come from a different sample space than the past data that determine the estimate. This makes minimax estimators very biased, and we propose a computationally tractable alternative with similar behavior in the worst case, but superior accuracy on most states.

Quantum information processing relies on physical systems that store and process quantum information, usually in the form of qubits. Testing and characterizing qubit devices is the business of quantum tomography [1], and *quantum state tomography* in particular is used to estimate the quantum state (density matrix) ρ produced by an initialization procedure. Tomography comprises two steps: (1) *data gathering*, accomplished by measuring a "quorum" of different observables on N samples of ρ; and (2) an *estimator* that maps the data to a final estimate $\hat{\rho}$. The goal, of course, is an accurate estimate—we want a $\hat{\rho}$ "close" to the true state ρ, minimizing some error metric $d(\rho : \hat{\rho})$.

We define an optimal estimator to be one which acheives the highest accuracy in the worst case. One might expect tomographers to choose an estimator that is optimal (or at least near-optimal). Surprisingly, this is not done. Although several estimators are known and used (linear inversion [2], maximum likelihood [3], Bayesian mean [4], hedged maximum likelihood [5], L_1-regularization [6], BLUE [27]), none of them is known to have optimal pointwise accuracy [29] for finite N. Until now, it wasn't even possible to evaluate whether any of these estimators is "good enough", because the bounds on achievable pointwise accuracy weren't known either.

We address this situation in the present Letter by constructing *minimax* estimators (depicted in Fig. 1; see detailed explanation after Eq. 7) with absolutely optimal performance. These estimators are unwieldy, but (i) their performance yields tight upper bounds on accuracy, effectively delineating what "good enough" means, and (ii) their construction provides quite a lot of insight into the structure of the problem. Armed with these results, we show that *hedged maximum likelihood* (HML) is remarkably close to optimal, and outperforms minimax for most states (though of course its worst-case risk is higher). We also identify the value for the hedging parameter β that appears in HML which leads to the minimax solution within that class.

Prerequisites: Defining "accuracy" requires making several choices. For example, an optimal estimator for one error metric $d(\rho : \hat{\rho})$ is generally not optimal for a different metric $d'(\rho : \hat{\rho})$. Here [7], we quantify inaccuracy by the *quantum relative entropy*,

$$d(\rho : \hat{\rho}) = \mathrm{Tr}\left[\rho(\log\rho - \log\hat{\rho})\right]. \quad (1)$$

Relative entropy [26] is a well-motivated measure of *predictive* (and information-theoretic) inaccuracy [4]. It is a uniquely well-motivated error metric [32]; critically, it is Fisher-adjusted (i.e., agrees locally with the unique metric of statistical distinguishability [31]). Non-Fisher-adjusted metrics are ill-motivated and yield arbitrary results. Analysis of a different Fisher-adjusted metric (e.g. infidelity) would produce results qualitatively similar to those we derive here.

An estimator's *pointwise risk* is a function of the true state ρ and is given by the average of $d(\rho : \hat{\rho})$ over all possible data sets D of finite size N:

$$\overline{d}(\rho) = \sum_D Pr(D|\rho) d(\rho : \hat{\rho}(D)). \quad (2)$$

In the minimax paradigm, we quantify an estimator's accuracy by its *worst-case* risk, $\overline{d}_{\max} = \max_\rho \overline{d}(\rho)$. The *minimax risk* of the estimation problem is the minimum achievable risk (minimized over all possible estimators), and a *minimax estimator* is one that achieves this bound.

In most inference problems, the sample space of possible observations (data) is fixed by the problem. Not so in quantum tomography. Quantum systems can be measured in many different and incomparable ways. This is the single most significant difference between quantum and classical estimation. This freedom is often removed in quantum problems by choosing the best or worst possible measurement (e.g., as in the definition of quantum relative entropy as the classical relative entropy of the most difficult-to-predict measurement). This is usually not done in tomography, because the measurements which

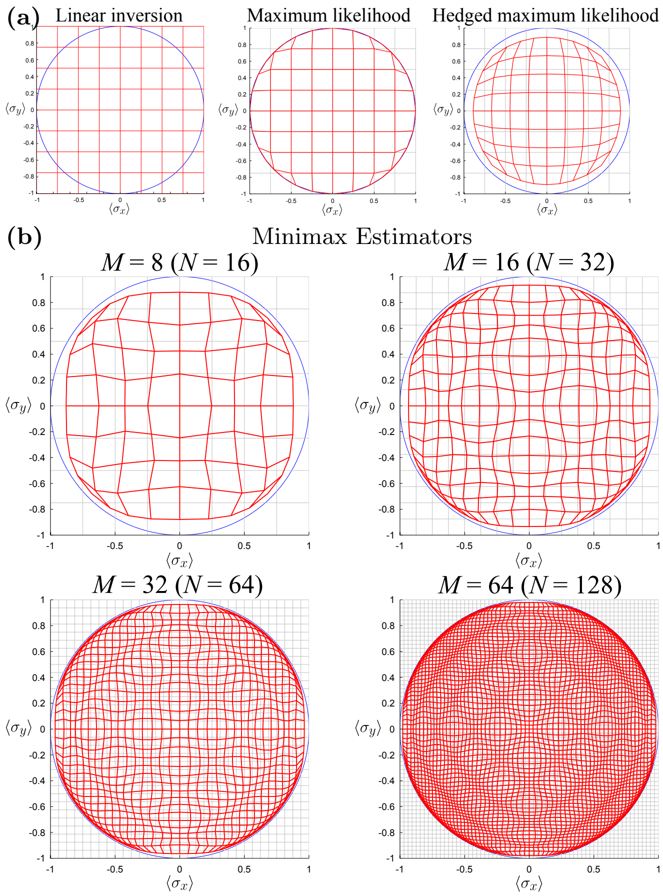

FIG. 1: **Estimators for Pauli measurements on a rebit**, depicted as distortions of the "linear inversion grid" (see text after Eq. 7). **(a)** Three standard estimators for $M = 8$ measurements of X and Y. Vertices of the red grid correspond to estimated states. Linear inversion estimates extend outside the "Bloch disk" of physical states. MLE's estimates are non-negative; HML's are strictly positive. **(b)** Minimax estimators for $M = 8, 16, 32, 64$ measurements of X and Y on a rebit. "Ripples" indicate local bias toward support points of the least favorable prior [32].

have the lowest expected risk are far too difficult. In this letter, we follow the majority of experiments and analyze tomography based on Pauli measurements on a single qubit. However, we also prove analytic lower bounds on minimax risk that apply to *any* non-adaptive measurement and any d-dimensional quantum system. In some parts of our analysis, we use a *rebit* – a quantum system with a 2-dimensional *real* Hilbert space, whose state space corresponds to the equatorial plane of the Bloch sphere – as an easier-to-analyze proxy for a qubit.

Minimax risk: The first main result of this Letter is a lower bound on the asymptotic ($N \to \infty$) minimax relative entropy risk of Pauli tomography on qubits and rebits,

$$\bar{d}_{\max} \geq \frac{e^{-\frac{1}{2}}}{4} \frac{\sqrt{\mathfrak{D} - 1}}{\sqrt{N}}, \qquad (3)$$

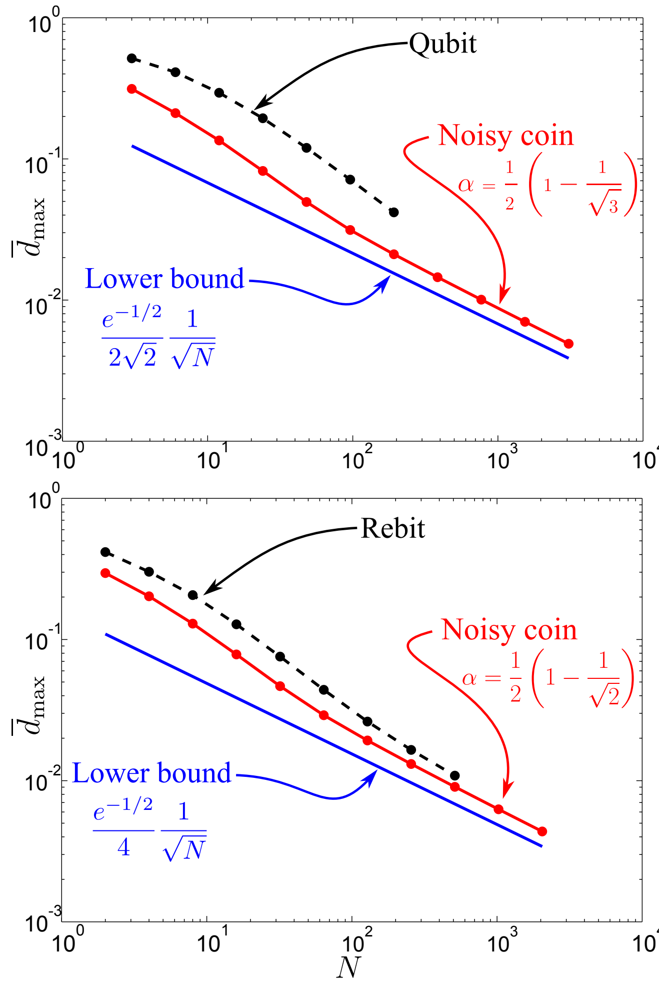

FIG. 2: **Numerical minimax risk for qubits, rebits, and noisy coins**. Black curves show the risk of numerically constructed minimax estimators for **(a)** a qubit and **(b)** a rebit, as a function of the number of samples (N), up to the maximum that was numerically feasible. Red curves illustrate the numerically-computed risk of "noisy coin" systems whose noise levels are chosen to match the effective "noise" of the qubit and the rebit (respectively). Blue lines show the the lower bound given in Eq. (6).

where $\mathfrak{D} = 2$ for rebits and $\mathfrak{D} = 3$ for qubits. Its $O(1/\sqrt{N})$ scaling contrasts sharply with the minimax risk of estimating a *classical* bit, which is almost exactly $0.5/N$ [10, 11]. We derive this bound below by mapping the minimax risk of qubit and rebit state tomography to a classical "noisy coin" model. In Figure 2, we compare these bounds to numerical calculations of the minimax risk, for small N, of qubits, rebits, and noisy coins.

A d-dimensional quantum state is analogous in many ways to a classical d-outcome probability distribution. However, its minimax risk scales differently because of a phenomenon instrinsic to quantum tomography (though not uniquely quantum) that we call *sampling mismatch*: the sample space for the observed events is neither unique nor isomorphic to the underlying state space. For example, the possible statistics for the three 2-outcome Pauli measurements on a qubit naturally define a cube, whereas the possible *quantum* states form a sphere (the Bloch ball).

Sampling mismatch can be reproduced in a simple classical model called the "noisy coin" [12]. It is a classical system with a 2-outcome sample space (i.e., a coin flip) where each observation is erroneous with known probability α. Sampling mismatch arises when we attempt to assign probabilities to future *noiseless* observations using data from *noisy* measurements. The noisy coin's minimax risk is $O(1/\sqrt{N})$, because nearly-pure states are hard to estimate accurately from noisy statistics. The corresponding minimax estimators are strongly biased toward nearly-pure states (see [12] for details). We are going to use a variant of the noisy coin model to bound the risk of tomography.

We define "tomography" thus: N samples (copies) of a single-qubit state ρ will be prepared; each sample will be measured independently (not jointly together with other samples) in a predefined fashion (not adaptively). The kth sample is measured in an arbitrary basis, and this measurement can be described by a POVM (positive operator-valued measure) $\mathcal{M}_k = \{\Pi_k, 1\!\!1 - \Pi_k\}$ whose outcomes have probabilities $\{q, 1-q\}$ with $q = \mathrm{Tr}\Pi_k\rho$. Based on the N measurement results, we report a state $\hat{\rho}$, and seek to minimize relative entropy cost.

Now, suppose that before analyzing the data (but after choosing the measurements!) we are told the eigenbasis of ρ. This helps us (only ρ's spectrum must be estimated), so the risk of spectrum estimation is a strict *lower* bound on the risk of full tomography[30].

We define $\{|0\rangle, |1\rangle\}$ to be the eigenstates of ρ, and write

$$\rho = p\,|0\rangle\langle 0| + (1-p)\,|1\rangle\langle 1|\,. \quad (4)$$

Now, we need only estimate $p \in [0,1]$. This parameter manifold is identical to that of a coin. Furthermore, the quantum relative entropy between two diagonal density matrices is identical to the classical relative entropy between the corresponding distributions. So, since ρ's eigenbasis is known, estimating ρ is identical to estimating the bias of a coin. However, unless the eigenbases of ρ and the Π_k happen to coincide, the measurement data obtained from the N samples of ρ are not "noiseless". Even if $p = 0$ (i.e., ρ is pure), the data remain somewhat random. The probability of observing Π_k is not p, but

$$\begin{aligned} q &= p\,\langle 0|\,\Pi_k\,|0\rangle + (1-p)\,\langle 1|\,\Pi_k\,|1\rangle \\ &= p(1-2\alpha_k) + \alpha_k \end{aligned}$$

where the *effective noise* in sample k is

$$\alpha_k = \langle 1|\,\Pi_k\,|1\rangle^2\,. \quad (5)$$

We can model this situation perfectly by a noisy coin (as in Ref. [12]) where each observation fails with a different error probability. The error probability for the kth sample is α_k. In [32], we bound this estimation problem's minimax risk by

$$\overline{d}_{\max} \geq \frac{e^{-\frac{1}{2}}}{2\sqrt{\bar{\beta}}}\frac{1}{\sqrt{N}}, \quad (6)$$

where $\bar{\beta}$ is the average *resolution* provided by the N noisy samples:

$$\bar{\beta} = \frac{1}{N}\sum_{k=1}^{N}\beta_k = \frac{1}{N}\sum_{k=1}^{N}\frac{(1-2\alpha_k)^2}{\alpha_k(1-\alpha_k)}. \quad (7)$$

For any fixed measurement strategy – e.g., the standard one where $N/3$ samples are measured in the X, Y, Z bases – the *maximum* risk occurs when we choose the eigenbasis of ρ to maximize $\bar{\beta}$ in Eq. 7. This "least favorable" basis is the one that lies as far as possible from all measured bases. For a rebit, it lies halfway between the X and Z bases, and $\alpha_k = \frac{1}{2}(1 - 1/\sqrt{2})$. For a qubit, it is the geometric mean of the X, Y, and Z bases, and $\alpha_k = \frac{1}{2}(1 - 1/\sqrt{3})$. Inserting these values for α_k yields the final bound given in Eq. 3.

This argument applies (qualitatively) to tomography on any finite-dimensional system with any discrete POVM. As long as no samples are measured in a basis that diagonalizes ρ, the minimax risk scales as $O(1/\sqrt{N})$ (although the prefactor will vary). However, if any non-vanishing fraction of the N samples are measured in a basis that diagonalizes ρ, then Eq. 6 no longer applies. Thus, continuous POVMs such as the unitarily invariant Haar-uniform rank-1 POVM (a.k.a. the uniform POVM), require a slightly different argument. In [32], we prove that even in this case, the minimax risk is lower bounded by $O\left((N\log N)^{-1/2}\right)$.

Estimators: To confirm the bound given by Eq. 3 and explore minimax risk at small N, we use numerics to find minimax estimators. An estimator is a map from the set of all possible datasets into the set of density matrices. The possible outcomes of the measurement(s) performed are represented by a set of positive operators $\{E_k\}$, and the data themselves by a set of raw counts $D = \{n_k\}$. For qubit Pauli tomography, the data comprise $M = N/3$ samples each of σ_x, σ_y, and σ_z measurements; for rebits, they comprise $M = N/2$ samples each of σ_x and σ_y measurements [13].

We used numerical optimization (over the set of possible estimators) to find minimax estimators. The algorithms are described in [32]. In Figure 1, we depict the resulting estimators, and compare them to three canonical estimators:

1. **Linear inversion** ($\hat{\rho}_{\mathrm{LI}}$): The first tomographic estimator, it is obtained by equating each probability $\Pr(k|\hat{\rho}_{\mathrm{LI}}) = \mathrm{Tr}E_k\hat{\rho}_{\mathrm{LI}}$ to its observed frequency $\frac{n_k}{M}$.

2. **Maximum likelihood** ($\hat{\rho}_{\rm ML}$): MLE assigns the density matrix that maximizes the probability of the observed data (the likelihood), $\mathcal{L}(\rho) = Pr(D|\rho) = \prod_k [\mathrm{Tr}(E_k\rho)^{n_k}]$.

3. **Hedged maximum likelihood** ($\hat{\rho}_{\mathrm{HML},\beta}$): The HML estimator maximizes the product of $\mathcal{L}(\rho)$ and a "hedging function" $h(\rho) = \det(\rho)^\beta$. This function is strictly convex and vanishes for rank-deficient states, so the HML estimate is always full-rank.

To simplify visualization, we depict *rebit* estimators, which are qualitatively similar to qubit estimators and easier to depict. A rebit estimator is a map from datasets to Bloch vectors, as $\hat{\rho} : \{0, \ldots, M\}^2 \to \mathbb{R}^2$. We use the linear inversion estimator as a reference. As a linear map from the 2-dimensional space of datasets ($\{0 \ldots M\}^2$) and the 2-dimensional space of rebit states (the unit disc in $\mathbb{R}^2$), the linear inversion estimator is represented by a uniform grid on the "Bloch square" (Fig. 1a). Every *other* estimator is represented as a distortion of this grid. The vertices of the grid are estimates $\hat{\rho}$, and the position of such a vertex within the grid indicates what dataset it came from.

Minimax estimators for $N = 16, 32, 64$ and 128 (total) Pauli measurements on a rebit are shown in Figure 1b. The most striking feature of these estimators is a pronounced "ripple" phenomenon. This is not a numerical artifact. Instead, it represents a consistent bias toward certain discrete points within the state space (support points of the least favorable prior – see Fig. 1 in [32]), which can be identified in Figure 1 as regions where the grid lines cluster together. The minimax estimator demonstrates this bias because these points are, in a particular sense, the most difficult to estimate accurately.

Improving on Minimax: The minimax criterion is an elegant concept, but a dangerous one. In its single-minded quest to improve the *maximum* risk, it has no concern for the pointwise risk at states that are "easier" to estimate. In such regions, it may incur extreme bias and inaccuracy, for the sole purpose of achieving a tiny reduction in the maximum risk. For quantum tomography, this effect become extreme. While $O(1/N)$ risk can be achieved on all full-rank states, the risk is unavoidably $O(1/\sqrt{N})$ near the boundary. Our numerical experiments confirm that the minimax estimator's pointwise risk is $O(1/\sqrt{N})$ everywhere, whereas other estimators easily achieve $O(1/N)$ risk in the interior of the Bloch sphere (Fig. 3b). If ρ really was selected adversarially, then minimax would be a wise strategy. But in realistic cases, we would prefer an estimator that achieved $O(1/N)$ scaling where possible, even at the cost of slightly worse *worst-case* behavior.

A good estimator should achieve $O(1/N)$ risk in the interior, while coming as close as possible to minimax performance near the boundary. The maximum likelihood estimator (MLE) is disqualified because its point-

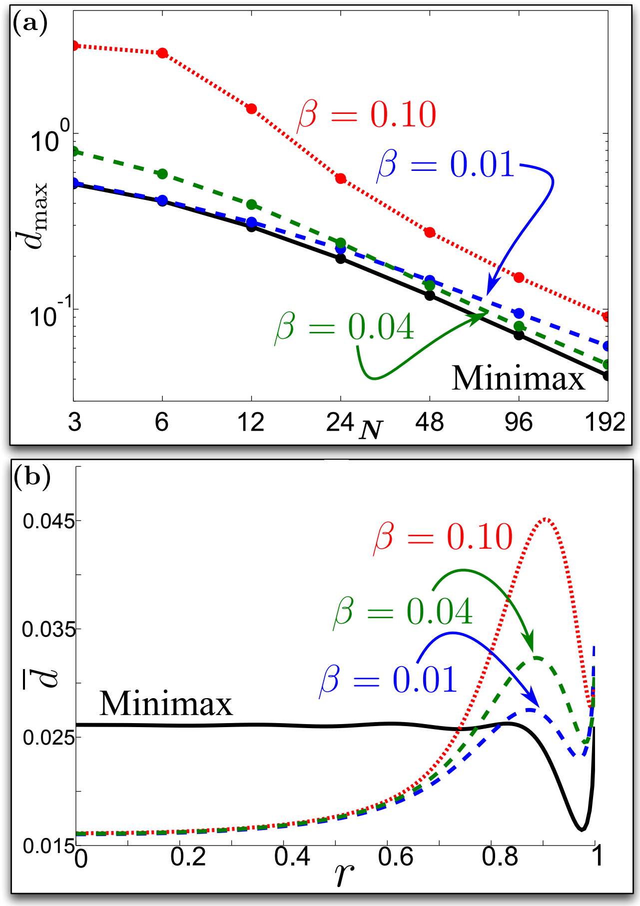

FIG. 3: **Maximum and pointwise risk of minimax and HML estimators**. Plot **(a)** shows the maximum risk, for *qubit* tomography, of the minimax estimator and three different HML estimators ($\beta = 0.01, 0.04, 0.10$) for $N \leq 192$ samples distributed equally among the 3 Pauli bases. Plot **(b)** shows the pointwise risk, along the axis oriented at 45 degrees to both X and Y, of the same estimators for $N = 128$ samples for a *rebit* (this minimax estimator is depicted in Fig. 1b). The two local maxima of $\bar{d}(\rho)$ are at $r = 1$ and $r \approx 1 - \frac{1}{\sqrt{N}}$. Choosing $\beta \approx 0.04$ balances these risks, and is therefore minimax among HML estimators. This HML estimator comes very close to matching the worst-case performance of the minimax estimator, and outperforms it dramatically in the interior of the state space.

wise expected risk is uniformly infinite (it has nonzero probability of returning a rank-deficient estimate for every ρ, so $\bar{d}(\rho) = \infty$). However, *hedged maximum likelihood* (HML) does not have this behavior. Introduced in Ref. [5] as a full-rank alternative to MLE, HML generalizes classical "add-β" estimators. Like them, it never assigns zero probabilities, and has a adjustable parameter β that governs how much it avoids zero eigenvalues. Classical "add-β" estimators are very nearly minimax (for $\beta \approx 1/2$), which suggests that HML estimators might have similar near-optimality properties.

All HML estimators have good behavior ($O(1/N)$

pointwise risk) in the interior, so we choose β to be the one which is minimax among HML estimators. As illustrated in Fig. 3b, an HML estimator's pointwise risk has local maxima at the boundary (pure states) and/or at a slightly depolarized state (with purity $\sim 1 - 1/\sqrt{N}$). To minimize its maximum, we choose β to equalize the risk at these two local maxima. The asymptotically minimax β for the noisy coin model was shown in Ref. [12] to be $\beta_{\text{optimal}} \approx 0.0389$, and our numerics confirm that $\beta \approx 0.04$ is minimax to within the available numerical precision for rebit tomography as well (Fig. 3b; qubit results for smaller N are not shown, but confirm that $\beta \approx 0.04$ has nearly-minimax performance).

For this value of β, HML compares favorably with minimax estimators. Its worst-case risk is very close to the minimax risk (Fig. 3a), and it dramatically outperforms minimax in the interior of the state space (Fig. 3b). So while hedging estimators do not offer strictly optimal performance by the global minimax criterion, they are (i) easy to specify and calculate, (ii) close to minimax, and (ii) more accurate than minimax estimators for almost all states ρ. We do not know why the minimax β is so different for noiseless coins (≈ 0.5) and for qubits/rebits/noisy coins (≈ 0.04), but it suggests fundamental differences between noiselessly sampled systems and those (like qubits and noisy coins) where sampling mismatch is important.

CF was supported by National Science Foundation grant number PHY-1212445, the Canadian Government through the NSERC PDF program, the IARPA MQCO program, the ARC via EQuS project number CE11001013, and by the US Army Research Office grant numbers W911NF-14-1-0098 and W911NF-14-1-0103. Sandia National Laboratories is a multi-program laboratory operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

---

[1] M. Paris and J. Rehacek, *Quantum state estimation.* Springer, New York (2004).

[2] M. A. Nielsen and I. L. Chuang, *Quantum computation and quantum information.* Cambridge University Press (2010).

[3] Z. Hradil, *Quantum-state estimation*, Physical Review A **55**, R1561 (1997).

[4] R. Blume-Kohout, *Optimal, reliable estimation of quantum states*, New Journal of Physics **12**, 043034 (2010).

[5] R. Blume-Kohout, *Hedged maximum likelihood quantum state estimation*, Physical Review Letters **105**, 200504 (2010).

[6] D. Gross, Y.-K. Liu, S. T. Flammia, S. Becker, and J. Eisert, *Quantum state tomography via compressed sensing*, Physical Review Letters **105**, 150401 (2010).

[7] Without a doubt there are other reasonable definitions that would lead to different answers. Examples are given by Refs. [19–22], who each obtain different conclusions by choosing different loss functions. We expect future work to consider other variations, and we hope that our results and discussion will help to inform and guide such future work.

[8] E. L. Lehmann and G. Casella, *Theory of point estimation*, Springer (1998).

[9] A rebit is a system with a 2-dimensional *real* Hilbert space, containing states $|\psi\rangle = \sin\theta\,|0\rangle + \cos\theta\,|1\rangle$. Alternatively, consider a qubit with the constraint $\langle\sigma_y\rangle = 0$.

[10] A. Rukhin, *Minimax estimation of the binomial parameter under entropy loss*, Statistics and Decisions **13**, 69 (1993).

[11] R. E. Krichevskiy, *Laplace's law of succession and universal encoding*, IEEE Transactions on Information Theory **44**, 296 (1998).

[12] C. Ferrie and R. Blume-Kohout, *Estimating the bias of a noisy coin*, AIP Conference Proceeding **1443**, 14 (2012).

[13] The assumption of perfectly known measurements can be relaxed [14, 15].

[14] A. M. Branczyk, D. H. Mahler, L. A. Rozema, A. Darabi, A. M. Steinberg, D. F. V. James, *Self-calibrating Quantum State Tomography*, New Journal of Physics **14**, 085003 (2012).

[15] D. Greenbaum, *Introduction to Quantum Gate Set Tomography*, arXiv:1509.02921 (2015).

[16] D. Mahler, L. A. Rozema, A. Darabi, C. Ferrie, R. Blume-Kohout, and A. Steinberg, *Adaptive quantum state tomography improves accuracy quadratically*, Physical Review Letters **111**, 183601 (2013).

[17] E. Bagan, M. Ballester, R. Gill, R. Munoz-Tapia, and O. Romero-Isart, *Separable measurement estimation of density matrices and its fidelity gap with collective protocols*, Physical Review Letters **97**, 130501 (2005).

[18] O. Barndorff-Nielsen and R. Gill, *Fisher information in quantum statistics*, Journal of Physics A: Mathematical and General **33**, 4481 (2000).

[19] H. K. Ng and B.-G. Englert, *A simple minimax estimator for quantum states*, International Journal of Quantum Information **10** (2012).

[20] N.K. Langford, *Errors in quantum tomography: diagnosing systematic versus statistical errors*, New Journal of Physics **15**, 035003 (2013).

[21] R. Schmied, *Quantum State Tomography of a Single Qubit: Comparison of Methods*, arXiv:1407.4759.

[22] S. T. Flammia, D. Gross, Y.-K. Liu and Jens Eisert, *Quantum tomography via compressed sensing: error bounds, sample complexity and efficient estimators*, New Journal of Physics **14** 095022 (2012).

[23] M. V. Burnashev and S.-i. Amari,, *On density estimation under relative entropy loss criterion*, Problems of Information Transmission **38**, 323 (2002).

[24] P. J. Kempthorne, *Numerical specification of discrete least favorable prior distributions*, SIAM Journal on Scientific and Statistical Computing **8**, 171 (1987).

[25] M. Guţă and J. Kahn, *Local asymptotic normality for qubit states*, Physical Review A **73**, 052108 (2006).

[26] V. Vedral, *The role of relative entropy in quantum information theory*, Reviews of Modern Physics **74**, 197 (2002).

[27] H. Zhu, *Quantum state estimation with informationally overcomplete measurements*, Phys. Rev. A **90**, 012115

(2014).

[28] D. Petz and C. Sudar, *Geometries of quantum states*, Journal of Mathematical Physics **37**, 2662 (1996).

[29] Bayesian mean *is* known to provide optimal average accuracy for certain important error metrics $d(\rho : \hat{\rho})$.

[30] The risk is a sum of terms due to (i) spectral (eigenvalue) error, and (ii) eigenbasis error. The minimax risk is dominated by spectral error, because $O(\epsilon)$ spectral variations (changes in eigenvalues) can induce $O(\epsilon)$ changes in probabilities p that are themselves $O(\epsilon)$ and thus change the risk by $O(\epsilon)$ – whereas an $O(\epsilon)$ unitary variation changes p by at most $O(\epsilon p)$, and changes the risk by at most $O(\epsilon^2)$. Thus, giving up the eigenbasis error does not lower the risk substantially.

[31] To be precise, there is a unique *classical* metric of statistical distinguishability (the Fisher metric), and for quantum states there exists a limited but nontrivial family of such metrics [28]. However, all of the quantum "Fisher" metrics necessarily coincide for spectral (eigenvalue) deviations , which dominate the pointwise risk. Thus, our statement in the text (while technically flawed) is fully accurate in practice; all quantum Fisher-adjusted error metrics agree asymptotically *on the relevant errors*.

[32] See Supplemental Material [url], which includes Refs. [33,34].

[33] Audenaert, K. M. R., et al. *Discriminating States: The Quantum Chernoff Bound*, Physical Review Letters **98**, 160501 (2007).

[34] C. A. Fuchs, *Distinguishability and accessible information in quantum theory*, Ph.D. Dissertation (1996).