

Learning pure quantum states in any dimension (almost) without regret

Josep Lumbreras and Marco Tomamichel

Introduction. Quantum state tomography is usually formulated as the task of producing an accurate classical description of an unknown quantum state from many independently prepared copies. In the pure-state multi-armed quantum bandit problem (PSMAQB), introduced in [2, 4], tomography is instead studied as a sequential learning problem in which the measurements themselves are part of the objective. At round t , the learner receives a fresh copy of an unknown pure state $\rho = |\psi\rangle\langle\psi|$, chooses a rank-one projector $A_t = |\phi_t\rangle\langle\phi_t|$, and measures $\{A_t, \mathbb{I}_d - A_t\}$, observing a random outcome $X_t \in \{0, 1\}$ with $\mathbb{E}[X_t | A_t] = \text{Tr}(\rho A_t)$. The cumulative regret is

$$\text{Regret}(T) = \sum_{t=1}^T (1 - \text{Tr}(\rho A_t)).$$

which for pure states the fidelity is $F(\rho, A_t) = \text{Tr}(\rho A_t)$. Thus each regret term is exactly the infidelity between the unknown state and the measurement direction chosen at that round. In this model, low regret means that the learner rapidly aligns its measurements with ρ , while still obtaining enough information to keep improving its estimate. This also gives regret a physical interpretation: if $q_t = \text{Tr}(\rho A_t)$, the expected post-measurement infidelity of the consumed copy is $2q_t(1 - q_t)$, and hence the cumulative expected disturbance is at most $2 \text{Regret}(T)$. The same additive objective also appears in state-agnostic work extraction, where it can be interpreted as cumulative dissipated energy caused by suboptimal control directions [3]. In [5] it was shown that for qubits one can achieve $\text{Regret}(T) = O(\log^2 T)$ and online infidelity $1 - \text{Tr}(\rho A_t) = \tilde{O}(1/t)$ with nearly matching lower bounds.

The question is whether this gentle-learning phenomenon, known for qubits, survives in arbitrary finite dimension. We show that it does. The main point is not that qudits admit a larger Bloch vector, but rather that the pure-state manifold has enough local structure to support a low-disturbance learning strategy. Our algorithm works directly on this manifold: it learns only tangent directions, uses symmetric physical measurements to remove curvature terms, and carries information between epochs without comparing tangent vectors based at different states.

Why qudits are not just larger qubits. For qubits, the PSMAQB problem can be reduced to a linear bandit on the Bloch sphere with vanishing-variance noise [4, 6]. This reduction is special to dimension two. For general d , the pure states form the projective manifold $\mathcal{M} \simeq \mathbb{CP}^{d-1}$, $\dim_{\mathbb{R}} \mathcal{M} = 2(d-1)$, embedded in the affine space of trace-one Hermitian matrices, whose real dimension is $d^2 - 1$. Thus, at a pure state $C \in \mathcal{M}$, many ambient directions are normal to the pure-state manifold. A generalized-Bloch-vector approach treats these directions as ordinary linear degrees of freedom, but near the optimal projector valid pure-state, rank-one measurements can vary to first order only along the tangent space. The information relevant for reducing regret is therefore tangent information, not information in the extra normal directions.

This geometric mismatch is the main obstacle beyond qubits. Our algorithm overcomes it by repeatedly freezing a local coordinate system around its current estimate, building a linear model only in the tangent space of $\mathcal{M}$, and then moving the estimate back along the manifold. In this way, the proof replaces the global Bloch-sphere picture by an intrinsic local-linear picture on $\mathbb{CP}^{d-1}$.

Contributions. We make three contributions. First, we identify the geometric obstruction that prevents the qubit Bloch-sphere argument from extending directly to qudits. Second, we introduce an intrinsic tangent-space measurement scheme whose symmetric retractions give exact local linear observations despite the curvature of $\mathbb{CP}^{d-1}$. Third, we combine this local model with a hot-started,

variance-adaptive robust estimator, obtaining polylogarithmic cumulative regret and near-optimal online infidelity for every fixed dimension.

Main result. For every $d \geq 2$, low-regret pure-state tomography extends from qubits to arbitrary qudits. There is an adaptive protocol using fresh copies of an unknown $\rho \in \mathcal{S}_d^*$ and two-outcome rank-one projective measurements such that, for all T and $t \leq T$,

$$\mathbb{E}[\text{Regret}(T)] = \mathcal{O}(d^3 \log^2 T), \quad \mathbb{E}[1 - F(\rho, \hat{\rho}_t)] = \mathcal{O}\left(\min\left\{1, \frac{d^3 \log T}{t}\right\}\right).$$

Thus, for fixed d , the protocol achieves essentially optimal $1/t$ online infidelity while incurring only polylogarithmic cumulative disturbance.

Main ideas of the protocol. The protocol is organized into epochs, i.e. blocks of measurement rounds. At the beginning of epoch m , the algorithm has a current pure-state estimate $C_m = |\psi_m\rangle\langle\psi_m| \in \mathcal{M}$. Rather than learning in the full ambient space, it works locally in the tangent space $T_{C_m}\mathcal{M}$, whose real dimension is $d_{\text{tan}} = 2(d-1)$.

The local quantity to be learned is the tangent component of the displacement from the current estimate to the unknown state,

$$\Delta_*^{(m)} := \mathcal{P}_{T_{C_m}}(\rho - C_m) \in T_{C_m}\mathcal{M},$$

where $\mathcal{P}_{T_{C_m}}$ denotes Frobenius-orthogonal projection onto the tangent space. This is the basic geometric reduction: during an epoch, the estimator targets only the first-order degrees of freedom along which a pure state can actually move.

The next difficulty is that a tangent vector is not itself a quantum state that can be measured. To turn a tangent direction into a valid measurement, the algorithm moves a small distance away from C_m along the pure-state manifold. Concretely, if $V = \frac{1}{\sqrt{2}}(|\varphi\rangle\langle\psi_m| + |\psi_m\rangle\langle\varphi|) \in T_{C_m}\mathcal{M}$, $\langle\varphi|\psi_m\rangle = 0$ is a unit tangent direction, we define

$$\text{Retract}_{C_m}(\pm\tau V) := |\gamma_V(\pm\tau)\rangle\langle\gamma_V(\pm\tau)|, \quad |\gamma_V(\pm\tau)\rangle = \cos\left(\frac{\tau}{\sqrt{2}}\right)|\psi_m\rangle \pm \sin\left(\frac{\tau}{\sqrt{2}}\right)|\varphi\rangle.$$

Thus $\text{Retract}_{C_m}(\pm\tau V)$ is the rank-one projector onto the pure state obtained by moving from $|\psi_m\rangle$ in the tangent direction V . In density-matrix form, its tangent component is $\mathcal{P}_{T_{C_m}}(\text{Retract}_{C_m}(\pm\tau V)) = \pm \frac{1}{\sqrt{2}} \sin(\sqrt{2}\tau) V$.

At step s of epoch m , the protocol chooses a Frobenius-orthonormal basis $\{v_{m,i}\}_{i=1}^{d_{\text{tan}}}$ of $T_{C_m}\mathcal{M}$, where $d_{\text{tan}} = 2(d-1)$, and measures the symmetric pair $A_{m,s,i}^\pm = \text{Retract}_{C_m}(\pm r_{m,s} v_{m,i})$, where $r_{m,s}$ is a small displacement. The associated tangent observation direction is therefore

$$O_{m,s,i} := \frac{A_{m,s,i}^+ - A_{m,s,i}^-}{2} = \frac{1}{\sqrt{2}} \sin(\sqrt{2}r_{m,s}) v_{m,i}.$$

Then, if X^+ and X^- are the two binary outcomes obtained by measuring $A_{m,s,i}^+$ and $A_{m,s,i}^-$ on independent copies, and $Y = (X^+ - X^-)/2$, we have

$$Y = \text{Tr}(\Delta_*^{(m)} O_{m,s,i}) + \varepsilon, \quad \mathbb{E}[\varepsilon \mid \text{past}] = 0.$$

The symmetry of the two measurements is what makes this model exactly linear: the even, curvature-dependent terms in $A_{m,s,i}^+$ and $A_{m,s,i}^-$ cancel, while the odd tangent term remains. This is the local linear model that replaces the simple Bloch-sphere geometry available for qubits.

It remains to estimate the tangent displacement and transfer information across epochs. The idea is to build a least squares estimator but the tangent space changes when the center moves

from C_m to C_{m+1} , so the algorithm does not transport a full covariance matrix between different base points. Instead, it transfers only one scalar precision parameter. Within epoch m , the linear observations are summarized by a *tangent design operator*. For s completed measurement steps, and $X \in T_{C_m}\mathcal{M}$, define the tangent design operator $\mathcal{V}_{m,s}^{\text{tan}}$ as

$$\mathcal{V}_{m,s}^{\text{tan}}(X) = \mu_{m-1}X + \sum_{r=1}^s \sum_{i=1}^{d_{\text{tan}}} \omega_m \text{Tr}(O_{m,r,i}X)O_{m,r,i},$$

for a sequence of positive weights $\{\omega_m\}_{m=1}^M$. Thus $\mathcal{V}_{m,s}^{\text{tan}}$ is the regularized covariance, or design, of the tangent directions that have been measured so far. It quantifies how accurately different tangent components can be estimated. The first term is the hot start. At the beginning of epoch m we set $\mathcal{V}_{m,0}^{\text{tan}} = \mu_{m-1}I_{T_{C_m}\mathcal{M}}$, $\mu_{m-1} = \lambda_{\min}(\mathcal{V}_{m-1,T_{m-1}}^{\text{tan}})$, where $I_{T_{C_m}\mathcal{M}}$ is the identity superoperator in $T_{C_m}\mathcal{M}$. Hence the new epoch inherits the amount of information accumulated in the previous epoch only through the scalar precision μ_{m-1} , but constructs its new design in the new tangent space $T_{C_m}\mathcal{M}$. This avoids transporting a whole covariance matrix between tangent spaces based at different pure states.

The same design also determines how far the next measurements are placed from the current center. Writing $\lambda_{m,s-1} = \lambda_{\min}(\mathcal{V}_{m,s-1}^{\text{tan}})$, the displacements in $A_{m,s,i}^{\pm}$ are chosen as $r_{m,s} = 1/\sqrt{\lambda_{m,s-1}}$. Thus the measurements concentrate around C_m as the tangent design becomes more informative.

Using the design operator $\mathcal{V}_{m,s}^{\text{tan}}$ and the combined outcomes Y , we estimate the tangent displacement from C_m toward ρ . To make the estimate robust, the measurements are organized into independent blocks, as in a *median-of-means construction*. For each block j , the corresponding *weighted least-squares estimate* is

$$\hat{\Delta}_m^{(j)} = (\mathcal{V}_{m,T_m}^{\text{tan}})^{-1} \left(\sum_{s=1}^{T_m} \sum_{i=1}^{d_{\text{tan}}} \omega_m Y_{m,s,i}^{(j)} O_{m,s,i} \right).$$

The two scalar quantities now have clear roles: μ_{m-1} carries the precision from previous epochs, while ω_m is the weight assigned to the new observations in the current epoch. We choose ω_m proportional to μ_{m-1} because the measurement outcomes become less noisy as the protocol gets closer to ρ . Indeed, if a projector A is close to ρ , then $\text{Tr}(\rho A)$ is close to one, so the variance of the binary outcome is small. Finally, the median-of-means step selects a central candidate among the block estimates $\hat{\Delta}_m^{(j)}$. This gives a robust tangent estimate without changing the geometric structure of the protocol.

After estimating the tangent displacement, the algorithm converts it back into a pure state by retracting from C_m along the estimated tangent direction. This produces the next center C_{m+1} . The analysis shows that, in the local region where the protocol operates, tangent-space estimation error translates into infidelity error on the pure-state manifold. Thus accurate estimation in $T_{C_m}\mathcal{M}$ is enough to ensure that the next center is closer to ρ .

The regret proof combines two effects: the epoch centers C_m approach ρ , and the symmetric measurement displacements $r_{m,s}$ decrease with the accumulated precision. Thus the measured projectors $A_{m,s,i}^{\pm}$ remain close to ρ . Summing the resulting infidelities over epochs gives the stated regret bound, and the same estimates give the online infidelity bound.

Impact and open questions. The main open problem is the dimensional dependence: our bounds are $\mathcal{O}(d^3 \log^2 T)$ for regret and $\mathcal{O}(d^3 \log T/t)$ for online infidelity, while the known lower bound is $\Omega(d \log T)$ [4] and adaptive tomography suggests a possible $\tilde{\Theta}(d/t)$ infidelity scaling [1]. Closing this gap would clarify whether sample-optimal tomography is compatible with negligible cumulative disturbance. The tangent-space tools developed here may also be useful for other curved or low-rank quantum learning problems.

-
- [1] S. Chen, B. Huang, J. Li, A. Liu, and M. Sellke. “When Does Adaptivity Help for Quantum State Learning?”. In *2023 IEEE 64th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 391–404, Los Alamitos, CA, USA(2023).
 - [2] J. Lumbreras, E. Haapasalo, and M. Tomamichel. “Multi-armed quantum bandits: Exploration versus exploitation when learning properties of quantum states”. *Quantum* **6**:749, (2022).
 - [3] J. Lumbreras, R. C. Huang, Y. Hu, M. Gu, and M. Tomamichel. “Quantum state-agnostic work extraction (almost) without dissipation”. arXiv preprint arXiv:2505.09456 , (2025).
 - [4] J. Lumbreras, M. Terekhov, and M. Tomamichel. “Learning pure quantum states (almost) without regret”. arXiv preprint arXiv:2406.18370 , (2024).
 - [5] J. Lumbreras and M. Tomamichel. “Linear bandits with polylogarithmic minimax regret”. arXiv preprint arXiv:2402.12042 (2024).
 - [6] J. Lumbreras and M. Tomamichel. “Linear bandits with polylogarithmic minimax regret”. In *Proceedings of Thirty Seventh Conference on Learning Theory*, volume 247 of *Proceedings of Machine Learning Research*, pages 3644–3682, (2024).
 - [7] J. Lumbreras and M. Tomamichel. “Learning Pure Quantum States in Any Dimension (Almost) Without Regret”. arXiv preprint arXiv:2605.09019 , (2026).