Rank-dependent lower bounds for quantum chi-squared tomography

Abstract

We prove copy lower bounds for estimating a rank-at-most-r quantum state in dimension d under two chi-squared losses. For Bures chi-squared divergence, error at most ε with success probability at least $2/3$ requires $\Omega(\sqrt{r}\, d^{3/2}/\varepsilon)$ copies with collective measurements and $\Omega(r^{3/2}d^{3/2}/\varepsilon)$ with adaptive one-copy measurements. The bounds hold uniformly for $d \geq 2$, $1 \leq r \leq d$, and $0 < \varepsilon \leq \varepsilon_0$, where $\varepsilon_0 > 0$ is universal, and match the upper bounds of Flammia and O'Donnell (Quantum, 2024) up to logarithmic factors. For right-inverse chi-squared divergence, the corresponding rates are $\Omega(d^2/\varepsilon^2)$ and $\Omega(rd^2/\varepsilon^2)$; we give upper bounds matching exactly in the collective model and up to a logarithmic factor in the one-copy model at constant confidence. The lower bounds quantify how much output mass must be allocated outside an uncertain support. We bound the volume of support subspaces compatible with any successful estimate, then use a Sobolev inequality to relate posterior concentration on these sets to the Fisher information available from measurements. For pure states, the usual covariant measurement with a regularized output attains both collective rates, and we evaluate the exact minimax expected right-inverse chi-squared loss.

Note. This paper was generated entirely by AI, including MiMo, using an automated research pipeline developed by Chenghua Liu and Hanyu Li.

1 Introduction

Low rank reduces the number of parameters of a quantum state, but it need not reduce the cost of estimation under a loss sensitive to small output eigenvalues. Chi-squared tomography illustrates this distinction. An estimate supported on the wrong subspace can have infinite loss even when the estimated and true subspaces are close. A learner must therefore allocate output mass to directions that remain uncertain. We study the statistical cost of this allocation.

Flammia and O'Donnell [1] gave Bures chi-squared estimators using $\widetilde{O}(\sqrt{r}\, d^{3/2}/\varepsilon)$ copies with collective measurements and $\widetilde{O}(r^{3/2}d^{3/2}/\varepsilon)$ with adaptive one-copy measurements, for rank-at-most-r states in dimension d and divergence error ε. They asked whether the collective rate admits a matching rank-dependent lower bound. We prove lower bounds matching both rates up to logarithmic factors. We also analyze right-inverse chi-squared loss, which agrees with Bures chi-squared loss on commuting states but has a different accuracy dependence: its collective copy complexity is $\Theta(d^2/\varepsilon^2)$, and its adaptive one-copy complexity is $\widetilde{\Theta}(rd^2/\varepsilon^2)$, at constant confidence and sufficiently small ε.

Our hard instances are states maximally mixed on an unknown subspace. For a fixed output estimate, small loss restricts both its tail mass and the orientation of the true support. We show that these restrictions force its set of successful support subspaces to have small volume. For the Bures loss, a convexity argument shows that a scalar tail spectrum maximizes the relevant volume; for the right-inverse loss, the constraint is an ellipsoid whose volume can be evaluated directly. A Sobolev inequality then bounds the probability mass that a posterior can place in a set of that volume. Combining this bound with the Fisher information of the experiment controls success probability without an unbiasedness assumption. On the same family, adaptive one-copy measurements have a Fisher-information trace budget smaller by a factor of order r than collective measurements.

These bounds separate quantum chi-squared tomography from more familiar tomographic tasks. Haah, Harrow, Ji, Wu, and Yu [3] gave collective infidelity estimators with copy complexity $O((dr/\eta)\log(d/\eta))$ for error η. O'Donnell and Wright [8, 9] developed collective estimators based on Schur–Weyl sampling. Their spectral estimate $\widehat{\alpha}$ satisfies $\mathbb{E}\sum_{i:\alpha_i>0}(\widehat{\alpha}_i-\alpha_i)^2/\alpha_i \le d^2/n$ for the true spectrum α [9, Theorem 1.7]. This guarantee has the true spectrum in the denominator, whereas our loss has the estimate in the denominator. Moreover, estimating the spectrum does not resolve its eigenvectors. Our flat-spectrum instances isolate the latter uncertainty.

Projector states also underlie the trace-distance lower bounds of Scharnhorst, Spilecki, and Wright [11]. Their proof includes a reduction from trace-distance to Bures-distance projector tomography; Bures distance is distinct from the Bures chi-squared divergence studied here. Nayak and Zhou [7] recently proved that trace-norm tomography with adaptive measurements on at most t fresh copies at a time has copy complexity $\Theta((dr/\eta^2)\max\{1, r/\sqrt{t}\})$ at sufficiently small constant accuracy. Their lower bound uses the same flat-support family $\rho_X = P_X/r$ in local graph coordinates. In the single-copy case, their block-positivity calculation gives the same Fisher-information trace scale $O((d-r)/r)$ as Lemma 3.5; they also use a smoothly truncated Gaussian prior and an adaptive Fisher-information chain rule, but convert the resulting information bound to trace-norm risk via the van Trees inequality. Thus we do not regard support rotations, the single-copy Fisher budget, or adaptive Fisher accounting as specific to the present work. Related multiparameter Fisher-information tradeoffs appear in Gill and Massar [2] and Zhou and Chen [13].

The loss-specific ingredient here is instead the geometry of the set of supports on which a fixed output has small chi-squared loss, together with its conversion to posterior mass by a Sobolev inequality. Using only the trace-norm comparison at error $\eta = \sqrt{\varepsilon}$ gives the scales dr/ε collectively and dr^2/ε for one-copy measurements. The Bures lower bounds below are larger by $\sqrt{d/r}$, while the right-inverse bounds have a different accuracy exponent. Concurrent work of Keskin, Luo, Majid, and Radzihovsky [6] proves optimal bounded-joint-measurement lower bounds for full-rank trace-norm tomography using a metric Fano inequality and a log-Sobolev comparison between mutual and Fisher information. That result is methodologically adjacent but does not address the low-rank chi-squared losses considered here.

We complement the lower bounds with upper-bound reductions that track subnormalized blocks and all copies consumed in preparing them. The Bures construction follows the spectral-block approach of [1, Section 3.3]; the right-inverse construction uses isotropic regularization. The collective reduction uses the normalized Frobenius estimator of O'Donnell and Wright [8, Theorem 1.2]. For pure states, we analyze a regularized output of the known covariant measurement [4]. We also evaluate its exact minimax expected right-inverse loss, distinguishing the choice of measurement from the Bayes-optimal output rule [12].

1.1 Model, losses, and results

Let $\mathcal{D}_d$ be the density matrices on $\mathbb{C}^d$, and let $\mathcal{D}_{d,r} = \{\rho \in \mathcal{D}_d : \operatorname{rank}\rho \le r\}$, where $d \ge 2$ and $1 \le r \le d$ are integers. We write $\|\cdot\|_{\mathrm{F}}$, $\|\cdot\|_{\mathrm{op}}$, and $\|\cdot\|_1$ for the Frobenius, operator, and trace norms. Positive definiteness is denoted by $A > 0$; $A \succeq B$ denotes the positive-semidefinite order. An identity matrix carries a dimension subscript when needed.

For $\sigma > 0$, set $\Delta = \rho - \sigma$ and $\mathcal{J}_\sigma(X) = (\sigma X + X\sigma)/2$. The Bures (or SLD) and right-inverse losses are

$$\chi^2_{\mathrm{B}}(\rho\,\|\,\sigma) = \operatorname{Tr}[\Delta\mathcal{J}_\sigma^{-1}(\Delta)], \tag{1}$$

$$\chi^2_{\mathrm{R}}(\rho\,\|\,\sigma) = \operatorname{Tr}(\Delta^2\sigma^{-1}) = \operatorname{Tr}(\rho^2\sigma^{-1}) - 1. \tag{2}$$

For singular σ, evaluate the formulas on its support if $\operatorname{supp}\rho \subseteq \operatorname{supp}\sigma$, and set the loss to infinity otherwise. In an eigenbasis $\sigma = \operatorname{diag}(s_1, \ldots, s_d)$ with $\sigma > 0$,

$$\chi^2_{\mathrm{B}}(\rho \,\|\, \sigma) = \sum_{i,j} \frac{2|\Delta_{ij}|^2}{s_i + s_j}, \tag{3}$$

$$\chi^2_{\mathrm{R}}(\rho \,\|\, \sigma) = \sum_{i,j} \frac{|\Delta_{ij}|^2}{s_j} = \sum_{i,j} \frac{s_i + s_j}{2 s_i s_j} |\Delta_{ij}|^2. \tag{4}$$

Both reduce to $\sum_i (p_i - q_i)^2/q_i$ on commuting states. These are two members of the broader family of quantum chi-squared divergences [10, 14]. A third choice is

$$\chi^2_{\mathrm{G}}(\rho \,\|\, \sigma) = \operatorname{Tr}[(\sigma^{-1/4} \rho \sigma^{-1/4})^2] - 1 = \sum_{i,j} \frac{|\Delta_{ij}|^2}{\sqrt{s_i s_j}},$$

with the same support convention. The scalar mean inequalities give

$$\chi^2_{\mathrm{B}} \le \chi^2_{\mathrm{G}} \le \chi^2_{\mathrm{R}}. \tag{5}$$

In particular, our Bures lower bounds also apply to χ^2_{G}; the right-inverse rates concern χ^2_{R} specifically.

Definition 1.1 (Learning model)**.** A learner receives n independent copies of an unknown $\rho \in \mathcal{D}_{d,r}$ and returns an arbitrary density matrix $\widehat{\sigma} \in \mathcal{D}_d$. Its output need not have rank at most r. For a loss L, the constant-confidence requirement is

$$\inf_{\rho \in \mathcal{D}_{d,r}} \mathbb{P}_\rho\{L(\rho, \widehat{\sigma}) \le \varepsilon\} \ge \frac{2}{3}. \tag{6}$$

Collective measurements permit any positive operator-valued measure (POVM) on $\rho^{\otimes n}$. Adaptive one-copy measurements permit a POVM on each next copy chosen from previous classical outcomes and internal randomness, but no quantum memory coupling different copies. All consumed copies, including rejected projection outcomes, count toward n. Measurements and estimators are measurable.

Write $N_{\mathrm{coll}}(\mathcal{D}_{d,r}, L, \varepsilon)$ and $N_{\mathrm{one}}(\mathcal{D}_{d,r}, L, \varepsilon)$ for the minimum copy counts satisfying Equation (6) in the two models. The parameter ε bounds the divergence itself; the convention $\chi^2 \le \varepsilon^2$ is obtained by substitution. The notation O and Ω hides universal constants, and $\widetilde{O}$ and $\widetilde{\Theta}$ additionally suppress logarithmic factors.

Theorem 1.2 (Main lower bounds)**.** *There are universal constants $c, \varepsilon_0 > 0$ such that, for all integers $d \ge 2$, $1 \le r \le d$, and $0 < \varepsilon \le \varepsilon_0$,*

$$N_{\mathrm{coll}}(\mathcal{D}_{d,r}, \chi^2_{\mathrm{B}}, \varepsilon) \ge c \frac{\sqrt{r}\, d^{3/2}}{\varepsilon}, \qquad N_{\mathrm{one}}(\mathcal{D}_{d,r}, \chi^2_{\mathrm{B}}, \varepsilon) \ge c \frac{r^{3/2} d^{3/2}}{\varepsilon}, \tag{7}$$

$$N_{\mathrm{coll}}(\mathcal{D}_{d,r}, \chi^2_{\mathrm{R}}, \varepsilon) \ge c \frac{d^2}{\varepsilon^2}, \qquad N_{\mathrm{one}}(\mathcal{D}_{d,r}, \chi^2_{\mathrm{R}}, \varepsilon) \ge c \frac{r d^2}{\varepsilon^2}. \tag{8}$$

For $1 \le r \le d/2$, the same bounds hold on the smaller family $\rho = P/r$, where P ranges over rank-r orthogonal projectors.

For pure states, the Bures bound is already $\Omega(d^{3/2}/\varepsilon)$, rather than the d/ε scale suggested by a local parameter count. Pure states also suffice for the collective right-inverse bound $\Omega(d^2/\varepsilon^2)$. Allowing a low-rank input does not remove the need for a higher-rank output that assigns mass to uncertain directions.

Theorem 1.3 (Upper bounds)**.** *For $0 < \varepsilon \le 1$ and $0 < \delta < 1/2$, let $L_0 = \lceil \log_2(16d/\varepsilon) \rceil$. There are learners that succeed with probability at least $1 - \delta$ uniformly on $\mathcal{D}_{d,r}$, with the following copy bounds:*

$$\textit{Bures, collective:} \quad O\left(\frac{\sqrt{r}\, d^{3/2}}{\varepsilon} L_0^2 \log \frac{L_0}{\delta}\right), \tag{9}$$

$$\textit{Bures, adaptive one-copy:} \quad O\left(\frac{r^{3/2} d^{3/2}}{\varepsilon} L_0^2 \log \frac{2dL_0}{\delta}\right), \tag{10}$$

$$\textit{Right-inverse, collective:} \quad O\left(\frac{d^2}{\varepsilon^2} \log \frac{1}{\delta}\right), \tag{11}$$

$$\textit{Right-inverse, adaptive one-copy:} \quad O\left(\frac{rd^2}{\varepsilon^2} \log \frac{2d}{\delta}\right). \tag{12}$$

For pure inputs, collective measurements achieve $O(d^{3/2}/\varepsilon)$ for χ^2_{B} and $O(d^2/\varepsilon^2)$ for χ^2_{R} at constant confidence, without logarithmic factors.

The Bures constructions recover the rates of [1]; we include their Frobenius-to-chi-squared reduction with explicit error and copy accounting. The pure-state construction also explains the different accuracy exponents: the Bures tail contribution is quadratic in the residual mass outside the estimated direction, whereas the right-inverse tail contribution is linear.

1.2 Comparison inequalities

We use the squared Bures distance

$$\mathsf{b}(\rho, \sigma)^2 = 2 - 2 \operatorname{Tr} \sqrt{\sigma^{1/2} \rho \sigma^{1/2}} = \min_{U \text{ unitary}} \left\| \rho^{1/2} - \sigma^{1/2} U \right\|_{\mathrm{F}}^2 .$$

The amplitude representation gives the triangle inequality: align an amplitude of the intermediate state with one endpoint, and an amplitude of the other endpoint with that intermediate amplitude, then apply the Frobenius triangle inequality.

Lemma 1.4 (Two comparison inequalities)**.** *For density matrices,*

$$\mathsf{b}(\rho, \sigma)^2 \le \chi^2_{\mathrm{B}}(\rho \,\|\, \sigma), \qquad \|\rho - \sigma\|_1^2 \le \chi^2_{\mathrm{B}}(\rho \,\|\, \sigma). \tag{13}$$

Moreover χ^2_{B} is the largest classical chi-squared divergence obtainable by measuring the pair of states with the same POVM.

Proof. Assume first $\sigma > 0$ and put $H = \mathcal{J}_\sigma^{-1}(\Delta)$. For a POVM element E_y, set $q_y = \operatorname{Tr}(E_y \sigma)$ and $d_y = \operatorname{Tr}(E_y \Delta) = \operatorname{Re} \operatorname{Tr}(\sigma E_y H)$. Hilbert–Schmidt Cauchy–Schwarz yields

$$\frac{d_y^2}{q_y} \le \operatorname{Tr}(\sigma H E_y H).$$

Summing gives classical chi-squared at most $\operatorname{Tr}(\sigma H^2) = \chi^2_{\mathrm{B}}$. Equality holds for the projective measurement in an eigenbasis of H.

To compare with Bures distance, diagonalize the positive semidefinite matrix

$$T = \sigma^{-1/2} (\sigma^{1/2} \rho \sigma^{1/2})^{1/2} \sigma^{-1/2}.$$

It satisfies $T\sigma T = \rho$. In its eigenbasis, with eigenvalues t_i, the two measured distributions obey $p_i = t_i^2 q_i$ and

$$\sum_i \sqrt{p_i q_i} = \mathrm{Tr}(\sigma T) = \mathrm{Tr}\sqrt{\sigma^{1/2}\rho\sigma^{1/2}}.$$

Their squared Hellinger distance is therefore b^2, and $\sum_i(\sqrt{p_i} - \sqrt{q_i})^2 \le \sum_i (p_i - q_i)^2/q_i$.

Finally, weighted Cauchy–Schwarz gives, for Hermitian K,

$$|\,\mathrm{Tr}(K\Delta)|^2 \le \chi^2_{\mathrm{B}}\ \mathrm{Tr}(K\mathcal{J}_\sigma(K)) = \chi^2_{\mathrm{B}}\ \mathrm{Tr}(\sigma K^2).$$

Choose $K = \mathrm{sign}(\Delta)$ to obtain the trace-norm bound. If σ is singular and contains the support of ρ, restrict the argument to its support. Otherwise both quantum losses are infinite, and measuring the projection onto $\ker\sigma$ also gives infinite classical chi-squared divergence. □

The trace-norm inequality gives a first reduction from chi-squared to trace-distance tomography. To obtain the additional $\sqrt{d/r}$ factor in the Bures bounds, and the additional $1/\varepsilon$ in the right-inverse bounds, we must also use the restriction on the estimate's tail spectrum. The next section measures its effect on the volume of a successful set.

2 Volumes of chi-squared loss balls

Fix integers $1 \le r \le s$, put $d = r + s$, and write

$$q = rs, \qquad p = 2rs.$$

The hard family consists of $\rho_P = P/r$, with P a rank-r orthogonal projector. Let μ be the invariant probability measure on these projectors. Complex $s \times r$ matrices have real dimension p; their Lebesgue measure uses the real and imaginary parts of their entries.

2.1 Grassmannian coordinates

Lemma 2.1 (Grassmann coordinates and a matrix ball)**.** *Let*

$$V_{s,r} = \mathrm{vol}\{Z \in \mathbb{C}^{s\times r} : Z^*Z < \mathrm{I}_r\}.$$

Then

$$V_{s,r} = \pi^q \prod_{j=1}^{r} \frac{(j-1)!}{(s+j-1)!}, \qquad \left(\frac{\pi}{2s}\right)^q \le V_{s,r} \le \left(\frac{\pi e}{s}\right)^q. \tag{14}$$

In the graph chart

$$U_X = \begin{pmatrix} \mathrm{I}_r \\ X \end{pmatrix} (\mathrm{I}_r + X^*X)^{-1/2}, \qquad P_X = U_X U_X^*, \tag{15}$$

Haar measure has density

$$\frac{d\mu}{dX} = V_{s,r}^{-1} \det(\mathrm{I}_r + X^*X)^{-d}. \tag{16}$$

If U is a Haar orthonormal r-frame in $\mathbb{C}^{r+s}$ and Z denotes its last s rows, then Z is uniform on the matrix ball in Equation (14)*.*

Proof. One way to derive the chart density is to take a complex Gaussian matrix $G = \binom{Y}{Z}$ with Y square of size r. Its column space is Haar and, almost surely, its graph coordinate is $X = ZY^{-1}$. Substituting $Z = XY$ has real Jacobian $|\det Y|^{2s}$. Changing $Y = (\mathrm{I} + X^*X)^{-1/2}T$ in the remaining Gaussian integral leaves the factor $\det(\mathrm{I} + X^*X)^{-(s+r)}$.

The normalizing integral is elementary. Integrate the rows one at a time, using the determinant lemma and

$$\int_{\mathbb{C}^r} (1 + \|z\|^2)^{-\alpha}\, dz = \pi^r \frac{\Gamma(\alpha - r)}{\Gamma(\alpha)}, \qquad \alpha > r.$$

The resulting product is

$$\int_{\mathbb{C}^{s\times r}} \det(\mathrm{I} + X^*X)^{-d}\, dX = \pi^q \prod_{\ell=1}^{s} \frac{(\ell-1)!}{(r+\ell-1)!} = \pi^q \prod_{j=1}^{r} \frac{(j-1)!}{(s+j-1)!}.$$

For completeness, the change of variables $F(X) = X(\mathrm{I}+X^*X)^{-1/2}$ maps onto the operator-norm unit ball and has Jacobian $\det(\mathrm{I} + X^*X)^{-d}$. To check this, let t_i be the squared singular values of X; those of $F(X)$ are $u_i = t_i/(1 + t_i)$. In complex rectangular singular-value coordinates the radial factor is $\prod_i t_i^{s-r} \prod_{i<j}(t_i - t_j)^2 \prod_i dt_i$. The substitution contributes, for each i, the exponent $-(s - r) - 2(r - 1) - 2 = -d$. Angular coordinates are unchanged. Thus $F(X)$ is uniform on the matrix ball under Equation (16). A Haar frame over the graph subspace is $U_X W$ with a Haar $W \in U(r)$; right-unitary invariance shows its bottom block has the same law as $F(X)$.

Finally,

$$s! \le \frac{(s+j-1)!}{(j-1)!} \le (2s)^s, \qquad s! \ge (s/e)^s,$$

which proves the two volume bounds. □

We next fix an estimate and determine which support subspaces can have small loss relative to it. The first restriction is on the mass outside its leading eigenspace.

Fix a positive definite estimate σ and let Q be a projector onto its largest r eigenvalues. In this basis write

$$\sigma = B \oplus A, \qquad B \in \mathbb{C}^{r\times r}, \quad A \in \mathbb{C}^{s\times s}, \quad a = \operatorname{Tr} A.$$

Lemma 2.2 (Tail mass and subspace distance)**.** *If $0 < \varepsilon \le 1/4$ and $\chi^2_{\mathrm{B}}(P/r \,\|\, \sigma) \le \varepsilon$, then*

$$a \le \varepsilon, \qquad t := \frac{\operatorname{Tr}(P(\mathrm{I} - Q))}{r} \le 4\varepsilon. \tag{17}$$

The same holds under a $\chi^2_{\mathrm{R}} \le \varepsilon$ guarantee.

Proof. Let f denote root fidelity. For a flat rank-r state,

$$f(P/r, \sigma) = \frac{1}{\sqrt{r}} \operatorname{Tr} \sqrt{P\sigma P|_{\mathrm{ran}P}} \le \sqrt{\operatorname{Tr}(P\sigma)}.$$

By Lemma 1.4, $f \ge 1 - \varepsilon/2$, so $\operatorname{Tr}(P\sigma) \ge 1 - \varepsilon$. The variational principle for the sum of the top r eigenvalues gives $\operatorname{Tr}(Q\sigma) \ge \operatorname{Tr}(P\sigma)$ and hence $a \le \varepsilon$.

The eigenvalues of a compression $P\sigma P$ are individually at most the corresponding top eigenvalues of σ. Therefore Q/r minimizes $\mathsf{b}(R/r, \sigma)$ over all rank-r projectors R. The triangle inequality implies $\mathsf{b}(P/r, Q/r) \le 2\sqrt{\varepsilon}$. If $\theta_1, \ldots, \theta_r$ are the principal angles between the two subspaces, then

$$t = \frac{1}{r}\sum_i \sin^2\theta_i \le \frac{2}{r}\sum_i (1 - \cos\theta_i) = \mathsf{b}(P/r, Q/r)^2 \le 4\varepsilon.$$

Use $\chi^2_{\mathrm{B}} \le \chi^2_{\mathrm{R}}$ for the last assertion. □

2.2 Tail geometry and successful volumes

For $A > 0$ and Hermitian Y, define

$$Q_A(Y) = \operatorname{Tr}[Y \mathcal{J}_A^{-1}(Y)], \qquad H_A(Z) = Q_A(ZZ^*).$$

Lemma 2.3 (Uniform tail eigenvalues maximize a volume)**.** *For fixed $K > 0$ and $\operatorname{Tr} A = a$, the Lebesgue volume of $\{Z \in \mathbb{C}^{s\times r} : H_A(Z) \le K\}$ is at most its value at $A = (a/s)\mathrm{I}_s$.*

Proof. The variational identity

$$Q_A(Y) = \sup_{H=H^*} \{2\operatorname{Tr}(YH) - \operatorname{Tr}(AH^2)\} \tag{18}$$

shows joint convexity in (A, Y). Also

$$\mathcal{J}_A^{-1}(Y) = 2\int_0^\infty e^{-uA} Y e^{-uA}\, du$$

is a positive map. Thus $Q_A(Y)$ is increasing in the positive-semidefinite order when $Y \ge 0$: if $D \ge 0$, then $Q_A(Y + D) - Q_A(Y) = Q_A(D) + 2\operatorname{Tr}[D\mathcal{J}_A^{-1}(Y)] \ge 0$. It follows that

$$H_A(Z) = \inf_{Y \succeq ZZ^*} Q_A(Y).$$

The constraint $Y \succeq ZZ^*$ is the convex Schur-complement constraint $\left(\begin{smallmatrix} Y & Z \\ Z^* & \mathrm{I}_r \end{smallmatrix}\right) \succeq 0$. Consequently $H_A(Z)$ is jointly convex in (A, Z).

Let $\mathcal{K}_A = \{Z : H_A(Z) \le K\}$. These are bounded convex bodies: positivity of $\mathcal{J}_A^{-1}$ as a quadratic form and the quartic homogeneity in Z imply coercivity, and a neighborhood of zero lies inside each body. Joint convexity gives $(1-u)\mathcal{K}_A + u\mathcal{K}_{A'} \subseteq \mathcal{K}_{(1-u)A+uA'}$. The Brunn–Minkowski inequality in real dimension $2rs$ therefore says that $\operatorname{vol}(\mathcal{K}_A)^{1/(2rs)}$ is concave in A. Unitary conjugation of A, accompanied by left multiplication of Z, preserves volume. Diagonalize A and average its conjugates by all coordinate permutations; their average is $(a/s)\mathrm{I}_s$. Concavity gives the claimed volume bound. □

The preceding symmetrization removes the dependence on individual tail eigenvalues. This gives a uniform success-volume bound for every output estimate, which is the geometric input to the statistical argument.

Theorem 2.4 (Haar volumes of loss balls)**.** *For $0 < \varepsilon \le 1/4$ and any density matrix σ,*

$$\mu\{P : \chi^2_{\mathrm{B}}(P/r \,\|\, \sigma) \le \varepsilon\} \le \left(6e\,\varepsilon\sqrt{r/s}\right)^q, \tag{19}$$

$$\mu\{P : \chi^2_{\mathrm{R}}(P/r \,\|\, \sigma) \le \varepsilon\} \le \left(18e\, r\varepsilon^2/s\right)^q. \tag{20}$$

Proof. If σ is singular, finite loss requires $\operatorname{ran} P$ to lie in one fixed proper subspace. This is a Haar-null set. We may assume $\sigma > 0$ and work in the $Q \oplus Q^\perp$ basis above. Take a Haar frame $U = \binom{Y}{Z}$ for P. By Lemma 2.1, Z is uniform on the operator-norm unit ball, and the true bottom block is ZZ^*/r.

For the Bures loss, the contribution of this bottom block is

$$\frac{H_A(Z)}{r^2} - 2t + a.$$

All entrywise contributions outside it are nonnegative. By Lemma 2.2, a successful P therefore satisfies

$$H_A(Z) \le 9r^2\varepsilon, \qquad a \le \varepsilon. \tag{21}$$

Discarding the restriction $\|Z\|_{\mathrm{op}} \le 1$ only enlarges the numerator volume. By Lemma 2.3, it is enough to consider $A = (a/s)\mathrm{I}$. In that case

$$H_A(Z) = \frac{s}{a}\operatorname{Tr}[(ZZ^*)^2].$$

Cauchy–Schwarz on the at most r nonzero eigenvalues of ZZ^* gives $\|Z\|_{\mathrm{F}}^4 \le r\operatorname{Tr}[(ZZ^*)^2]$. Thus Equation (21) implies

$$\|Z\|_{\mathrm{F}}^2 \le T := 3r^{3/2}\varepsilon/\sqrt{s}.$$

The Frobenius ball of squared radius T has volume $\pi^q T^q/q!$. Divide by $V_{s,r} \ge (\pi/(2s))^q$ and use $q! \ge (q/e)^q$ to obtain $(2esT/q)^q = (6e\varepsilon\sqrt{r/s})^q$.

For the right-inverse loss, the projector identity $P^2 = P$ gives

$$1 + \chi^2_{\mathrm{R}}(P/r \,\|\, \sigma) = \frac{\operatorname{Tr}(P_{11}B^{-1})}{r^2} + \frac{\operatorname{Tr}(Z^*A^{-1}Z)}{r^2}.$$

Matrix Cauchy–Schwarz and the principal angles give

$$\frac{\operatorname{Tr}(P_{11}B^{-1})}{r^2} \ge \frac{(r^{-1}\sum_i \cos\theta_i)^2}{1-a} \ge (1-t)^2 \ge 1 - 2t.$$

Thus a successful P satisfies $\operatorname{Tr}(Z^*A^{-1}Z) \le 9r^2\varepsilon$. The volume of this ellipsoid is

$$\frac{\pi^q}{q!}(9r^2\varepsilon)^q \det(A)^r \le \frac{\pi^q}{q!}\left(9r^2\varepsilon^2/s\right)^q,$$

where $\det(A) \le (a/s)^s \le (\varepsilon/s)^s$. Dividing by $V_{s,r}$ proves Equation (20). □

Corollary 2.5 (Euclidean volumes in a fixed local chart)**.** *Let $0 < \varepsilon \le 1/4$ and restrict X in Equation (15) to $\|X\|_{\mathrm{op}} \le 1/2$. For every fixed estimate σ, let v_{B} and v_{R} be the Lebesgue volumes of the two successful parameter sets in this chart. Then*

$$v_{\mathrm{B}}^{1/q} \le C\frac{\varepsilon\sqrt{r/s}}{s}, \qquad v_{\mathrm{R}}^{1/q} \le C\frac{r\varepsilon^2}{s^2}. \tag{22}$$

Proof. On this chart, Equation (16) is at least $V_{s,r}^{-1}(5/4)^{-dr}$. Therefore a local Lebesgue volume is at most $V_{s,r}(5/4)^{dr}$ times the corresponding Haar probability. Take qth roots, use $dr/q = d/s \le 2$, and apply Equation (14) and Theorem 2.4. □

3 Posterior concentration and the measurement lower bounds

Each possible output succeeds on a set whose volume was bounded in Corollary 2.5. A successful experiment must therefore produce posteriors concentrated on small sets. We first quantify the information needed for that concentration, then compare it with the information a measurement can extract from the flat-state family.

3.1 Posterior mass and Fisher information

For a density f on $\mathbb{R}^p$, write

$$\mathcal{I}(f) = \int f(x) \left\|\nabla \log f(x)\right\|^2 dx = 4 \int \left\|\nabla \sqrt{f}(x)\right\|^2 dx.$$

Weak derivatives suffice.

Lemma 3.1 (Fisher information and volume)**.** *Let $p \geq 3$, let $\sqrt{f} \in H^1(\mathbb{R}^p)$ vanish outside a compact set, and let S have Lebesgue volume v. Then*

$$\int_S f \leq \frac{v^{2/p}}{p} \mathcal{I}(f). \tag{23}$$

An absolute constant in place of 1 *would give the same conclusions below.*

Proof. The L^1 Sobolev inequality, obtained from the Euclidean isoperimetric inequality by coarea (see, e.g., [5]), is

$$\|g\|_{p/(p-1)} \leq \frac{1}{p\omega_p^{1/p}} \|\nabla g\|_1 ,$$

where ω_p is the volume of the unit ball in $\mathbb{R}^p$. Apply it to $g = |h|^{2(p-1)/(p-2)}$ and use Cauchy–Schwarz to obtain

$$\|h\|_{2p/(p-2)}^2 \leq \frac{[2(p-1)/(p-2)]^2}{p^2\omega_p^{2/p}} \|\nabla h\|_2^2 \leq \frac{4}{p} \|\nabla h\|_2^2 .$$

For the last inequality, the cube of side $2/\sqrt{p}$ lies in the unit ball, so $\omega_p^{2/p} \geq 4/p$, and $2(p-1)/(p-2) \leq 4$. Hölder's inequality with $h = \sqrt{f}$ gives

$$\int_S f \leq v^{2/p} \left\|\sqrt{f}\right\|_{2p/(p-2)}^2 ,$$

which proves the claim. Approximation extends the argument to H^1. □

Let π be a prior with compactly supported square root $h = \sqrt{\pi} \in H^1(\mathbb{R}^p)$. Write the likelihood of a classical outcome y as $\ell_\theta(y)$ relative to a parameter-independent measure ν, and put $g_y(\theta) = \sqrt{\ell_\theta(y)}$. We use the following regularity conditions: g_y is locally Lipschitz in θ for ν-almost every y, the local squared gradient energies are integrable in y, and differentiation can be passed under $\int \ell_\theta(y)\, d\nu(y) = 1$. We verify these conditions for our quantum experiments below. Define

$$F_\theta = 4 \int \nabla g_y(\theta) \nabla g_y(\theta)^{\mathsf{T}}\, d\nu(y).$$

This is the usual classical Fisher information matrix where the likelihood is positive, interpreted through weak derivatives at zeros.

Lemma 3.2 (Posterior information identity)**.** *Under the preceding conditions, assume $\mathbb{E}_\pi \operatorname{Tr} F_\theta < \infty$. Then*

$$\mathbb{E}_Y \mathcal{I}(\pi(\cdot \mid Y)) = \mathcal{I}(\pi) + \mathbb{E}_{\theta\sim\pi} \operatorname{Tr} F_\theta. \tag{24}$$

In particular, almost every posterior has a compactly supported H^1 square root whenever the right-hand side is finite.

Proof. Let $m(y) = \int \pi(\theta)\ell_\theta(y)\,d\theta$ be the marginal density. For $m(y) > 0$, the posterior square root is $hg_y/\sqrt{m(y)}$. Expanding its gradient and averaging gives

$$\begin{aligned}\mathbb{E}_Y \mathcal{I}(\pi(\cdot \mid Y)) &= 4 \iint \|g_y \nabla h + h \nabla g_y\|^2 \, d\theta \, d\nu(y) \\ &= 4 \int \|\nabla h\|^2 \, d\theta + 4 \int \pi(\theta) \int \|\nabla g_y\|^2 \, d\nu(y) \, d\theta.\end{aligned}$$

The cross term vanishes because $2 \int g_y \nabla g_y \, d\nu(y) = \int \nabla \ell_\theta(y) \, d\nu(y) = 0$. Its absolute integrability follows by Cauchy–Schwarz from the two energies on the right. Multiplication by the locally Lipschitz function g_y preserves the H^1 property on the compact support of h. Outcomes with $m(y) = 0$ make no contribution. □

Corollary 3.3 (Bayesian success bound)**.** *Under the preceding conditions, let $p \geq 3$. Suppose that every estimate's successful parameter set within the prior support has Lebesgue volume at most v. Then*

$$\mathbb{P}\{\textit{success}\} \leq \frac{v^{2/p}}{p} \left(\mathcal{I}(\pi) + \mathbb{E}_\pi \operatorname{Tr} F_\theta\right). \tag{25}$$

Proof. Apply Lemma 3.1 to each posterior and the success set specified by its output, then average and use Equation (24). Include any internal randomization in Y. □

For the Grassmannian family, the prior must stay inside the chart where the volume comparison is uniform. A truncated Gaussian achieves this with Fisher information of order rs^2.

Lemma 3.4 (Local prior)**.** *On $\mathbb{C}^{s\times r} \cong \mathbb{R}^{2rs}$, $r \leq s$, there is a prior supported on $\|X\|_{\mathrm{op}} \leq 1/2$ such that*

$$\mathcal{I}(\pi) \leq C_0 r s^2. \tag{26}$$

Its square root is compactly supported and belongs to H^1.

Proof. We give a construction to keep the dimension dependence explicit. Let $a_0 = 1/2$ and let the $p = 2rs$ real coordinates of X be independent centered Gaussians of variance $\tau^2 = a_0^2/(1024s)$, with density ϕ. Let $g(u)$ be 1 for $u \leq 1/2$, linear from 1 to 0 on $[1/2, 1]$, and zero for $u \geq 1$. Set

$$\pi(X) = Z_\pi^{-1} \phi(X) g(\|X\|_{\mathrm{op}} / a_0)^2.$$

Using 1/4-nets of the complex unit spheres of sizes at most 9^{2s} and 9^{2r}, respectively,

$$\mathbb{P}\{\|X\|_{\mathrm{op}} > a_0/2\} \leq 9^{2(s+r)} \exp[-a_0^2/(32\tau^2)] < \frac{1}{2}.$$

Here $\|X\|_{\mathrm{op}}$ is at most twice the largest $|u^* X v|$ over the two nets, and each such complex Gaussian has the usual radial Gaussian tail. Consequently $Z_\pi \geq 1/2$.

The operator norm is 1-Lipschitz in Frobenius norm, so the squared norm of the weak gradient of the cutoff is at most $4/a_0^2$. Differentiating $\sqrt{\pi}$ and using $\|u+v\|^2 \leq 2\|u\|^2 + 2\|v\|^2$ gives

$$\mathcal{I}(\pi) \leq \frac{4p}{\tau^2} + \frac{64}{a_0^2} \leq C_0 r s^2.$$

For example, $C_0 = 65536$ suffices. The cutoff vanishes at the boundary, so extension by zero has the stated H^1 property. □

3.2 Information budgets of quantum measurements

For the graph family Equation (15), use also the orthonormal complementary frame

$$V_X = \begin{pmatrix} -X^* \\ \mathrm{I}_s \end{pmatrix} (\mathrm{I}_s + XX^*)^{-1/2}.$$

A coordinate variation dX induces

$$d\rho_X = \frac{V_X K U_X^* + U_X K^* V_X^*}{r}, \qquad K = (\mathrm{I}_s + XX^*)^{-1/2} dX (\mathrm{I}_r + X^*X)^{-1/2}. \tag{27}$$

The map $dX \mapsto K$ is a contraction for the real Frobenius norm.

Lemma 3.5 (Collective and adaptive information budgets)**.** *For the $p = 2rs$ real coordinates of X:*

$$\operatorname{Tr} F_X \le \frac{4np}{r} = 8ns \qquad \textit{for any collective measurement on } n \textit{ copies}, \tag{28}$$

$$\operatorname{Tr} F_X \le \frac{4ns}{r} \qquad \textit{for any adaptive one-copy protocol.} \tag{29}$$

Proof. We write sums over POVM outcomes; for a continuous POVM these are integrals against a dominating scalar measure. At $\rho = P/r$, a tangent K has symmetric logarithmic derivative (SLD) $H = 2(VKU^* + UK^*V^*)$ and quantum Fisher information

$$\operatorname{Tr}(\rho H^2) = \frac{4}{r} \|K\|_{\mathrm{F}}^2 .$$

The same weighted Cauchy–Schwarz argument as in Lemma 1.4 bounds any measured Fisher information in this direction by that quantity. On n copies the SLD is the sum of its single-copy versions. The cross terms vanish because $\operatorname{Tr}(\rho H) = 0$, so the bound multiplies by n. Summing over an orthonormal real coordinate basis and using the contraction in Equation (27) proves Equation (28).

For a single-copy POVM element, use its blocks in the $U \oplus V$ basis:

$$E_y = \begin{pmatrix} A_y & C_y^* \\ C_y & B_y \end{pmatrix} \succeq 0, \qquad p_y = \operatorname{Tr}(A_y)/r.$$

In the aligned real tangent coordinates K, the sum of squares of the probability derivatives is $4\|C_y\|_{\mathrm{F}}^2 / r^2$. Positivity gives $|(C_y)_{ij}|^2 \le (A_y)_{jj}(B_y)_{ii}$ and thus $\|C_y\|_{\mathrm{F}}^2 \le \operatorname{Tr}(A_y)\operatorname{Tr}(B_y)$. Consequently the Fisher trace in these coordinates is at most

$$\sum_y \frac{4\|C_y\|_{\mathrm{F}}^2}{r \operatorname{Tr}(A_y)} \le \frac{4}{r} \sum_y \operatorname{Tr}(B_y) = \frac{4s}{r}.$$

Terms with $\operatorname{Tr} A_y = 0$ have $C_y = 0$ and contribute zero. Pullback by the contractive map in Equation (27) cannot increase the Fisher trace.

For adaptation, condition on the previous outcomes. Each conditional likelihood has the same bound. Its score has conditional mean zero, so scores at distinct times are orthogonal in expectation. Their Fisher traces add to at most $4ns/r$. □

We now justify the likelihood regularity used in Lemma 3.2, including continuous outcomes and zero probabilities. The full transcript of either measurement model, including internal randomness, is the outcome of a fixed POVM E on n copies. In finite dimension it has an operator-valued

density E_y relative to the finite scalar measure $\nu(B) = \operatorname{Tr} E(B)$, with $E_y \succeq 0$ and $\operatorname{Tr} E_y = 1$ almost everywhere. Hence

$$g_y(X) = \sqrt{\operatorname{Tr}(E_y \rho_X^{\otimes n})} = \left\| E_y^{1/2} (\rho_X^{1/2})^{\otimes n} \right\|_{\mathrm{F}}.$$

Here $\rho_X^{1/2} = P_X/\sqrt{r}$ is smooth in X. The norm of a smooth matrix-valued function is locally Lipschitz. On every compact chart, its Lipschitz constants are bounded uniformly in y, since $\|E_y\|_{\mathrm{op}} \le 1$. The same bounds and finiteness of ν justify integration of weak derivatives and differentiation of $\int \ell_X(y)\, d\nu(y) = 1$. A weak gradient vanishes almost everywhere on the zero set of a locally Lipschitz function, so the zero-likelihood terms agree with the convention in the budget proof. Thus the information budgets bound exactly the square-root energies used in Equation (24). No positivity assumption on the likelihoods is needed.

3.3 Proof of the rank-dependent lower bounds

Assume first $rs \ge 2$, so $p \ge 4$. Apply Corollary 3.3 with the prior in Lemma 3.4, the volumes in Corollary 2.5, and the budgets in Lemma 3.5. For any learner on the flat rank-r family, its prior-average success probability is at most the corresponding expression below:

$$\chi^2_{\mathrm{B}},\ \text{collective}: \quad C\varepsilon\sqrt{r/s}\left(1 + \frac{n}{rs}\right), \tag{30}$$

$$\chi^2_{\mathrm{B}},\ \text{one-copy}: \quad C\varepsilon\sqrt{r/s}\left(1 + \frac{n}{r^2 s}\right), \tag{31}$$

$$\chi^2_{\mathrm{R}},\ \text{collective}: \quad C\frac{r\varepsilon^2}{s}\left(1 + \frac{n}{rs}\right), \tag{32}$$

$$\chi^2_{\mathrm{R}},\ \text{one-copy}: \quad C\frac{r\varepsilon^2}{s}\left(1 + \frac{n}{r^2 s}\right). \tag{33}$$

The constants absorb only numerical quantities. Since $r \le s$, a small universal choice of ε_0 makes the term independent of n at most $1/3$. A uniform success guarantee of $2/3$ therefore forces

	collective	adaptive one-copy
χ^2_{B}	$c\sqrt{r}\, s^{3/2}/\varepsilon$	$cr^{3/2}s^{3/2}/\varepsilon$
χ^2_{R}	cs^2/ε^2	crs^2/ε^2.

(34)

The omitted case $rs = 1$ is $d = 2, r = 1$; the direct pure-state argument in Section 4 proves the same bounds without a Sobolev inequality in dimension two.

To prove Theorem 1.2 for $\mathcal{D}_{d,r}$, choose $k = \min\{r, \lfloor d/2 \rfloor\}$ and use flat rank-k inputs. Then $d - k$ is comparable to d. If $r > d/2$, k is also comparable to r; otherwise $k = r$. Substituting in Equation (34) gives all four bounds in Theorem 1.2. This completes the lower-bound proof.

Thus the separation between the two measurement models arises from their information budgets on the same geometric family: the respective traces are $O(ns)$ and $O(ns/r)$. Conditioning on previous outcomes preserves the one-copy budget, so classical adaptation retains the factor-r separation.

4 Pure states: matching collective bounds

For a pure input, the symmetric subspace gives a direct bound on posterior concentration. This supplies the case $d = 2, r = 1$, where the preceding Sobolev inequality does not apply, and leads to a matching covariant estimator in every dimension.

Let $m = d-1$ and let ψ be Haar on pure states. The n-copy input lies in the symmetric subspace, of dimension

$$D_n = \binom{n+d-1}{d-1} = \binom{n+m}{m}.$$

If every fixed estimate has a successful pure-state set of Haar measure at most v, then any POVM has Haar-average success probability at most $D_n v$. Indeed, restricted to the symmetric subspace,

$$\langle \psi^{\otimes n} | E_y | \psi^{\otimes n} \rangle \le \operatorname{Tr} E_y,$$

so integration over the success set of outcome y, followed by summation, gives at most $v \sum_y \operatorname{Tr} E_y = D_n v$. The same proof uses integrals for continuous POVMs.

Apply Theorem 2.4 with $r = 1, s = m$, and use $D_n \le [e(n+m)/m]^m$. Success probability at least $2/3$ implies, for a small universal ε,

$$n = \Omega(m^{3/2}/\varepsilon) \quad \text{for } \chi^2_{\mathrm{B}}, \qquad n = \Omega(m^2/\varepsilon^2) \quad \text{for } \chi^2_{\mathrm{R}}. \tag{35}$$

This proof works also when $m = 1$.

The same symmetry supplies an estimator attaining these rates. Its output is a measured direction with a small isotropic component on the orthogonal complement.

Use the covariant POVM of Hayashi [4] on the symmetric subspace

$$E(dv) = D_n |v\rangle\langle v|^{\otimes n} \, d\mu(v).$$

The Haar moment identity $\int |v\rangle\langle v|^{\otimes n} d\mu(v) = P_{\mathrm{sym}}/D_n$ verifies its normalization. On input $|\psi\rangle\langle\psi|^{\otimes n}$, the outcome density is $D_n |\langle v, \psi\rangle|^{2n}$. If $t = 1 - |\langle v, \psi\rangle|^2$, then

$$t \sim \mathrm{Beta}(m, n+1), \qquad \mathbb{E}t = \frac{m}{n+d}. \tag{36}$$

This follows immediately from the Haar density mt^{m-1} before the likelihood factor $(1-t)^n$ is applied.

For a parameter $0 < a < 1$, output

$$\sigma_v = (1-a)|v\rangle\langle v| + \frac{a}{m}(\mathrm{I} - |v\rangle\langle v|). \tag{37}$$

Direct substitution in the two loss formulas gives

$$\chi^2_{\mathrm{B}}(|\psi\rangle\langle\psi| \,\|\, \sigma_v) = \frac{(a-t)^2}{1-a} + \frac{mt^2}{a} - 2t + a + \frac{4t(1-t)}{1-a+a/m}, \tag{38}$$

$$\chi^2_{\mathrm{R}}(|\psi\rangle\langle\psi| \,\|\, \sigma_v) = \frac{1-t}{1-a} + \frac{mt}{a} - 1. \tag{39}$$

The factor m in the tail terms is the statistical cost of spreading output mass over uncertain orthogonal directions.

For $0 < \varepsilon \le 1$, take $a = \varepsilon/16$ for Bures chi-squared and $n \ge 48m^{3/2}/\varepsilon$. Markov's inequality and Equation (36) give $t \le a/\sqrt{m}$ with probability at least $2/3$. On that event,

$$\chi^2_{\mathrm{B}} \le 2a^2 + a + a + 8t \le 11a < \varepsilon.$$

For the right-inverse loss, take $a = \varepsilon/4$ and $n \ge 24m^2/\varepsilon^2$. With probability at least $2/3$, $t \le \varepsilon^2/(8m)$; hence

$$\chi^2_{\mathrm{R}} \le 2a + mt/a \le \varepsilon.$$

Together with Equation (35), these estimators prove the pure-state claims of Theorem 1.3.

5 Upper-bound constructions

A mixed state can have several spectral scales. Following the block approach of [1, Section 3.3], we refine nested principal blocks and freeze eigenvalues once their scale has been resolved. The analysis charges each matrix entry at the stage when one of its indices is frozen; it does not require the global Frobenius error to decrease after every refinement. We first establish this accounting and then give one-copy and collective submatrix estimators. For right-inverse loss, isotropic regularization provides a separate reduction. Appendix C records the elementary block identities and counterexamples explaining why the spectral weights and final normalization must be tracked.

5.1 Shell errors and normalization

Suppose nested active subspaces are refined successively. At stage j, a diagonalized estimate $\widehat{R}_j$ of the active true block R_j satisfies

$$\left\|R_j - \widehat{R}_j\right\|_{\mathrm{F}}^2 \le \eta_j.$$

Freeze the coordinates whose estimated eigenvalues are at least λ_j, and allow later unitaries only on the remaining active subspace. The final positive diagonal matrix A retains these frozen eigenvalues. The active coordinates can always be placed first, fixing a common coordinate convention for the successive blocks.

Lemma 5.1 (Shell accounting)**.** *For each stage j, assign to it all matrix entries with one index in the newly frozen shell and both indices in the active subspace at that stage. The sum of $|\rho_{uv} - A_{uv}|^2$ over these ordered entries is at most η_j. Their contribution to*

$$E_{\mathrm{B}}(\rho, A) := \mathrm{Tr}\big[(\rho - A)\mathcal{J}_A^{-1}(\rho - A)\big]$$

is therefore at most $2\eta_j/\lambda_j$. Every entry outside the final active block is assigned exactly once.

Proof. Immediately after diagonalization, the estimate has zero cross blocks between the frozen and remaining coordinates. It agrees with A on the frozen block. Every later unitary preserves these zero cross blocks and rotates the true cross block unitarily. Its Frobenius norm, and the frozen-block error, are unchanged. Their squared norms are part of the original error η_j. For an assigned entry at least one output eigenvalue is at least λ_j, so its weight $2/(a_u + a_v)$ is at most $2/\lambda_j$. □

The final active block is handled separately in the construction below. The shell estimate uses the Bures denominator $a_u + a_v$: one frozen index bounds its reciprocal. For right-inverse loss, both indices matter, and we will instead use an isotropic regularization.

After all blocks have been estimated, the following identity controls the normalization of the output.

Lemma 5.2 (Exact normalization formula)**.** *Let ρ have trace one, let $A > 0$, and set $t = \mathrm{Tr}\, A$. For either $E_{\mathrm{B}}(\rho, A)$ as above or $E_{\mathrm{R}}(\rho, A) = \mathrm{Tr}[(\rho - A)^2 A^{-1}]$,*

$$\chi^2(\rho \,\|\, A/t) = tE(\rho, A) - (t-1)^2, \qquad (t-1)^2 \le tE(\rho, A). \tag{40}$$

In particular, $E \le 1/2$ implies $t \le 2$ and $\chi^2(\rho \,\|\, A/t) \le 2E$.

Proof. For the Bures case, use $\mathcal{J}_A^{-1}(A) = \mathrm{I}$ and $\mathcal{J}_{A/t}^{-1} = t\mathcal{J}_A^{-1}$, then expand the square. For the right-inverse case, expand $E = \mathrm{Tr}(\rho^2 A^{-1}) - 2 + t$. Weighted Cauchy–Schwarz applied to $\mathrm{Tr}(\rho - A) = 1 - t$ gives $(t-1)^2 \le tE$. Solving this quadratic inequality proves the stated bound on t. □

5.2 A one-copy submatrix estimator

A covariant one-copy measurement gives an unbiased matrix observation for a subnormalized block. Matrix Bernstein concentration [15, Theorem 1.4] and rank truncation then provide the Frobenius estimate needed at each scale.

Let $R \succeq 0$ be a $k \times k$ principal block of the current true state, let $w = \operatorname{Tr} R$, and suppose $\operatorname{rank} R \le r$ and $w \le W \le 1$, where W is known. On one full copy, project onto this block. If projection fails, record $Y = 0$. If it succeeds, perform the covariant rank-one POVM on $\mathbb{C}^k$ and, for outcome v, record

$$Y = (k+1)|v\rangle\langle v| - \mathrm{I}_k. \tag{41}$$

This is implementable by choosing a Haar orthonormal basis and measuring in that basis. The unconditional probability density of v is $k\langle v|R|v\rangle$.

The second Haar moment gives

$$\mathbb{E}Y = R, \qquad \mathbb{E}Y^2 = kw\mathrm{I}_k + (k-1)R, \qquad \|Y - R\|_{\mathrm{op}} \le 2k. \tag{42}$$

In particular, $\|\mathbb{E}(Y-R)^2\|_{\mathrm{op}} \le 2kW$. For the average $\overline{Y}$ of N independent observations, applying matrix Bernstein to both signs of the centered observations yields

$$\mathbb{P}\{\left\|\overline{Y} - R\right\|_{\mathrm{op}} > u\} \le 2k \exp\left[-\frac{Nu^2}{4kW + (4/3)ku}\right]. \tag{43}$$

Keep the largest r positive eigenvalues of $\overline{Y}$, or all of them if $r \ge k$, and set the others to zero. Call the result $\widehat{R}$. If $\left\|\overline{Y} - R\right\|_{\mathrm{op}} \le u$, Weyl's inequality shows $\left\|\widehat{R} - \overline{Y}\right\|_{\mathrm{op}} \le u$, and hence

$$\left\|\widehat{R} - R\right\|_{\mathrm{F}}^2 \le 2r \left\|\widehat{R} - R\right\|_{\mathrm{op}}^2 \le 8ru^2.$$

Thus we have proved the following submatrix routine, with all projection failures counted as consumed copies.

Lemma 5.3 (One-copy submatrix routine)**.** *Let $0 < \beta < 1/2$ and $\eta > 0$. If $W^2 \le \eta$, the zero estimate suffices. Otherwise the construction above outputs a positive semidefinite estimate with squared Frobenius error at most η, with probability at least $1 - \beta$, using*

$$O\left[\left(\frac{rkW}{\eta} + k\sqrt{\frac{r}{\eta}}\right)\log\frac{2k}{\beta}\right] \tag{44}$$

copies. The rank bound can be replaced by $\min\{r, k\}$. In the nontrivial case $W^2 > \eta$, the displayed bound is at least a positive constant, so rounding the copy count to an integer does not change its order.

For a trace-one state, a normalized output satisfies the same bound. Put $r' = \min\{r, k\}$ and let $\mu_1 \ge \cdots \ge \mu_k$ be the eigenvalues of $\overline{Y}$. Keep the first r' and replace them by $\max\{\mu_i - h, 0\}$, where h makes their sum one; set the rest to zero. If the operator error is at most u, Weyl's inequality gives $|\mu_i - \lambda_i(R)| \le u$. At $h = u$ the retained sum is at most one, and at $h = -u$ it is at least one, since $\operatorname{rank} R \le r'$. Thus a suitable $h \in [-u, u]$ exists. Every retained eigenvalue changes by at most u; every discarded eigenvalue has absolute value at most u. The normalized estimate is therefore within operator distance $2u$ of R and has rank at most r', giving the same squared Frobenius bound $8r'u^2$.

5.3 Multiscale refinement for Bures loss

Let $0 < \varepsilon \le 1$, set

$$e = \varepsilon/16, \qquad L = \lceil \log_2(d/e) \rceil, \qquad \lambda_j = 2^{-j}, \qquad \eta_j = \frac{e\lambda_j}{8L} \quad (1 \le j \le L). \tag{45}$$

Initially the whole space is active. At stage j:

1. Estimate the current active block R_j using Lemma 5.3, with error η_j and failure probability δ/L.
2. Diagonalize the estimate within the active subspace. Freeze all eigenvalues at least λ_j, retaining them as output entries, and put the remaining coordinates first for the next stage.
3. If no coordinates remain, stop. After stage L, assign every remaining coordinate the output value λ_L.

Let $A > 0$ be the resulting diagonal matrix in the final accumulated basis. Return $A/\operatorname{Tr} A$, undoing that accumulated basis change for the original coordinate system.

The mass bound used by the routine is $W_1 = 1$ and, for $j \ge 2$,

$$W_j = \min\{1,\ r\lambda_{j-1} + \sqrt{r\eta_{j-1}}\}. \tag{46}$$

To verify it, condition on all previous estimates being accurate. On the new active subspace, the previous estimate C has operator norm below λ_{j-1} and $\|R_j - C\|_{\mathrm{F}} \le \sqrt{\eta_{j-1}}$. Let S project onto the support of R_j. Then

$$\operatorname{Tr} R_j = \operatorname{Tr}(SR_j) \le \operatorname{Tr}(SC) + \sqrt{\operatorname{rank} S}\,\|R_j - C\|_{\mathrm{F}} \le r\lambda_{j-1} + \sqrt{r\eta_{j-1}}.$$

Thus the bound is valid conditionally at every stage. Fresh copies at each stage and a union bound on the first failed stage give simultaneous accuracy with probability at least $1 - \delta$.

To bound the loss of the final estimate, condition on the simultaneous accuracy event. By Lemma 5.1, all frozen shells contribute at most

$$\sum_{j=1}^{L} \frac{2\eta_j}{\lambda_j} = e/4$$

to $E_{\mathrm{B}}(\rho, A)$. If a final active block remains, its last estimate C has eigenvalues in $[0, \lambda_L)$ and is within Frobenius-squared error η_L of the true block R. Therefore

$$\frac{\|R - \lambda_L \mathrm{I}\|_{\mathrm{F}}^2}{\lambda_L} \le \frac{2\eta_L}{\lambda_L} + 2d\lambda_L \le \frac{e}{4L} + 2e.$$

There is no tail term if all coordinates were frozen. In either case $E_{\mathrm{B}}(\rho, A) \le (5/2)e$. By Lemma 5.2, normalization gives $\chi^2_{\mathrm{B}}(\rho \,\|\, A/\operatorname{Tr} A) \le 5e < \varepsilon$.

For the copy count, the decreasing mass of the active block compensates for the increasing accuracy required at later stages. For $j \ge 2$,

$$W_j \le 2r\lambda_j + \sqrt{re\lambda_j/(4L)}, \qquad \lambda_j \ge \lambda_L > e/(2d).$$

Substitution in Equation (44), with $k \le d$, bounds a stage's nonlogarithmic cost by

$$C\left(\frac{r^2 dL}{e} + \frac{r^{3/2} d\sqrt{L}}{\sqrt{e\lambda_j}} + \frac{d\sqrt{rL}}{\sqrt{e\lambda_j}}\right) \le C\frac{r^{3/2} d^{3/2} L}{e}.$$

The first stage satisfies the same upper bound. Summing L stages proves Equation (10).

5.4 Right-inverse loss by regularization

Lemma 5.4 (Frobenius error to right-inverse chi-squared)**.** *Let $\widehat{\rho}$ be a density matrix and $\sigma = (1-\alpha)\widehat{\rho} + \alpha \mathrm{I}_d/d$, $0 < \alpha < 1$. Then*

$$\chi^2_{\mathrm{R}}(\rho \,\|\, \sigma) \le \frac{2d}{\alpha} \|\rho - \widehat{\rho}\|^2_{\mathrm{F}} + \frac{2\alpha}{1-\alpha}. \tag{47}$$

In particular, $\alpha = \varepsilon/8$ and $\|\rho - \widehat{\rho}\|^2_{\mathrm{F}} \le \varepsilon^2/(64d)$ imply $\chi^2_{\mathrm{R}} \le \varepsilon$ for $0 < \varepsilon \le 1$.

Proof. Use the squared-norm inequality in the inner product weighted by σ^{-1}, splitting $\rho - \sigma$ at $\widehat{\rho}$. Since $\sigma \succeq \alpha \mathrm{I}/d$, the first term is at most $2d \|\rho - \widehat{\rho}\|^2_{\mathrm{F}} / \alpha$. The matrices σ and $\widehat{\rho}$ commute, and their eigenvalues give

$$\chi^2_{\mathrm{R}}(\widehat{\rho} \,\|\, \sigma) = \mathrm{Tr}(\widehat{\rho}^2 \sigma^{-1}) - 1 \le \frac{1}{1-\alpha} - 1.$$

This proves Equation (47) and its stated specialization. □

Apply the normalized version of Lemma 5.3 with $k = d$, $W = 1$, and $\eta = \varepsilon^2/(64d)$. Its copy cost is

$$O\left[\left(\frac{rd^2}{\varepsilon^2} + \frac{\sqrt{r}\, d^{3/2}}{\varepsilon}\right) \log \frac{2d}{\delta}\right] = O\left(\frac{rd^2}{\varepsilon^2} \log \frac{2d}{\delta}\right).$$

Regularize using Lemma 5.4. This proves Equation (12).

For the unrestricted class $r = d$, the logarithm can be omitted at constant confidence. In the full-space measurement Equation (41),

$$\mathbb{E} \left\| \overline{Y} - \rho \right\|^2_{\mathrm{F}} = \frac{d^2 + d - 1 - \mathrm{Tr}\, \rho^2}{N} \le \frac{2d^2}{N}.$$

Projection onto the convex set of density matrices cannot increase Frobenius distance to ρ. Markov's inequality and Lemma 5.4 therefore give $O(d^3/\varepsilon^2)$ copies directly.

5.5 Collective measurements and the Frobenius reduction

O'Donnell and Wright [8, Theorem 1.2] give a collective estimator in dimension k whose output is a density matrix and satisfies

$$\mathbb{E} \|\widehat{\tau} - \tau\|^2_{\mathrm{F}} \le \frac{4k - 3}{N} \le \frac{4k}{N} \tag{48}$$

from $N \ge 1$ copies of any trace-one state τ. Their output is $U \,\mathrm{diag}(\lambda/N) U^*$, where λ is the observed partition of N. To apply this estimate to a subnormalized block, we must learn its mass and count the copies rejected by projection.

Proposition 5.5 (Collective submatrix routine)**.** *Let R be a positive semidefinite block of dimension k and mass $w \le W \le 1$. For $\eta > 0$ and $0 < \beta < 1/2$, a collective routine returns a positive semidefinite estimate with squared Frobenius error at most η, with probability at least $1 - \beta$, using*

$$O\left(\frac{kW}{\eta} \log \frac{1}{\beta}\right) \tag{49}$$

full copies. If $W^2 \le \eta$, no copies are required.

Proof. If $W^2 \le \eta$, the zero estimate already suffices. Otherwise $kW/\eta \ge 1$, so integer rounding of the copy counts below is harmless. Project N full copies onto the block. If $w = 0$, all projections fail and the zero output is exact, so assume $w > 0$. Let M succeed; then $M \sim \mathrm{Bin}(N, w)$. Conditional on $M > 0$, the retained copies have state $\tau = R/w$. Apply Equation (48) to them and output $\widehat{R} = (M/N)\widehat{\tau}$; output zero for $M = 0$. Conditioning on M and expanding at $(M/N)\tau$ gives

$$\mathbb{E}\left\|\widehat{R} - R\right\|_{\mathrm{F}}^2 \le \frac{8k\mathbb{E}M}{N^2} + 2\mathbb{E}(M/N - w)^2 \left\|\tau\right\|_{\mathrm{F}}^2 \le \frac{2(4k+1)w}{N}.$$

This accounts for the mass estimate and all rejected copies without a separate assumption that w is known. Markov's inequality supplies a Frobenius estimate with success probability at least $3/4$ at cost $O(kW/\eta)$. For amplification, run independent batches whose individual estimates are within Frobenius distance $u = \sqrt{\eta}/3$ of R with probability at least $3/4$. Choose an output whose ball of radius $2u$ contains a strict majority of the outputs, choosing the first output if none qualifies. If a majority are within distance u of R, a qualifying output exists and every qualifying ball intersects that majority. Its center is therefore within $3u = \sqrt{\eta}$ of R. A scalar Chernoff bound makes this event have probability at least $1 - \beta$ after $O(\log(1/\beta))$ batches, proving the claim. □

Completion of the collective upper bounds. Use Proposition 5.5 in the dyadic algorithm, with the same mass bounds, thresholds, and error accounting as before. For $j \ge 2$, the nonlogarithmic cost at stage j is at most

$$C\left(\frac{rdL}{e} + \frac{\sqrt{r}\,d\sqrt{L}}{\sqrt{e\lambda_j}}\right) \le C\frac{\sqrt{r}\,d^{3/2}L}{e}.$$

The first stage satisfies the same bound. Summing over at most $L = L_0$ stages, each with failure probability δ/L, proves Equation (9). Later measurements may depend on previous outcomes, which is permitted by the unrestricted collective model.

For right-inverse loss, apply Equation (48) in dimension d and use Markov's inequality to obtain squared Frobenius error $\varepsilon^2/(64d)$ with constant success probability from $O(d^2/\varepsilon^2)$ copies. The same independent-batch selection used in Proposition 5.5, applied to normalized estimates, reduces the failure probability to δ at a factor $O(\log(1/\delta))$ in cost. The selected estimate remains a density matrix. Finally apply Lemma 5.4. This proves Equation (11) and completes Theorem 1.3. □

References

[1] Steven T. Flammia and Ryan O'Donnell. Quantum chi-squared tomography and mutual information testing. *Quantum* **8**, 1381 (2024). doi:10.22331/q-2024-06-20-1381.

[2] Richard D. Gill and Serge Massar. State estimation for large ensembles. *Physical Review A* **61**, 042312 (2000). doi:10.1103/PhysRevA.61.042312.

[3] Jeongwan Haah, Aram W. Harrow, Zhengfeng Ji, Xiaodi Wu, and Nengkun Yu. Sample-optimal tomography of quantum states. *IEEE Transactions on Information Theory* **63**(9), 5628–5641 (2017). doi:10.1109/TIT.2017.2719044.

[4] Masahito Hayashi. Asymptotic estimation theory for a finite dimensional pure state model. *Journal of Physics A: Mathematical and General* **31**(20), 4633–4655 (1998). doi:10.1088/0305-4470/31/20/006. Corrigendum: **31**, 8405 (1998), doi:10.1088/0305-4470/31/41/015.

[5] Elliott H. Lieb and Michael Loss. *Analysis*, second edition. American Mathematical Society, 2001.

[6] Ufuk Keskin, Jason Luo, Mahbod Majid, and Matthew Radzihovsky. Tight lower bounds for state tomography with limited entanglement. arXiv:2609.05718 (2026). https://arxiv.org/abs/2609.05718.

[7] Ashwin Nayak and Xingyu Zhou. Optimal low-rank quantum state tomography with bounded-sample joint measurements. arXiv:2609.10514 (2026). https://arxiv.org/abs/2609.10514.

[8] Ryan O'Donnell and John Wright. Efficient quantum tomography. In *Proceedings of the 48th Annual ACM Symposium on Theory of Computing (STOC)*, 899–912 (2016). doi:10.1145/2897518.2897544.

[9] Ryan O'Donnell and John Wright. Efficient quantum tomography II. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing (STOC)* (2017). doi:10.1145/3055399.3055454. Preprint: https://arxiv.org/abs/1612.00034.

[10] Dénes Petz. Monotone metrics on matrix spaces. *Linear Algebra and its Applications* **244**, 81–96 (1996). doi:10.1016/0024-3795(94)00211-8.

[11] Thilo Scharnhorst, Jack Spilecki, and John Wright. Optimal lower bounds for quantum state tomography. arXiv:2510.07699 (2025). https://arxiv.org/abs/2510.07699.

[12] Fuyuhiko Tanaka. Generalized Bayesian predictive density operators. arXiv:quant-ph/0602072 (2006). https://arxiv.org/abs/quant-ph/0602072.

[13] Sisi Zhou and Senrui Chen. Randomized measurements for multiparameter quantum metrology. *PRX Quantum* **7**, 010314 (2026). doi:10.1103/s27y-gbrp.

[14] Kristan Temme, Michael J. Kastoryano, Mary Beth Ruskai, Michael M. Wolf, and Frank Verstraete. The χ^2-divergence and mixing times of quantum Markov processes. *Journal of Mathematical Physics* **51**, 122201 (2010).

[15] Joel A. Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of Computational Mathematics* **12**, 389–434 (2012). doi:10.1007/s10208-011-9099-z.

A Exact pure-state risk

For a fixed measurement, the Bayes-optimal output for right-inverse loss is the normalized square root of the posterior second moment. This is the $\alpha = -3$ specialization of Tanaka's generalized Bayesian predictive rule [12]. On pure inputs, the second moment equals the posterior mean state. We evaluate the resulting risk and optimize over all measurements, complementing the constant-success bounds of Section 4.

Theorem A.1 (Minimax expected right-inverse loss for pure states)**.** *For $n \geq 0$ copies of an unknown pure state in dimension d and arbitrary collective measurements,*

$$\inf_{\text{learners}} \sup_{\psi} \mathbb{E}_{\psi} \chi^2_{\mathrm{R}}(|\psi\rangle\langle\psi| \,\|\, \widehat{\sigma}) = \frac{(\sqrt{n+1} + d - 1)^2}{n+d} - 1. \tag{50}$$

Proof. Use the Haar prior. Restrict a POVM element E to the symmetric subspace; discard elements with zero trace. For $n \geq 1$, its posterior mean state is

$$M_E = \frac{\mathrm{I} + n\tau_E}{n+d}, \qquad \tau_E = \frac{\mathrm{Tr}_{2,\ldots,n}\, E}{\mathrm{Tr}\, E}. \tag{51}$$

To verify this, use the Haar moment identity on $n+1$ copies. On the first n symmetric factors,

$$P^{(n+1)}_{\mathrm{sym}} = \frac{1}{n+1}\left(\mathrm{I} + \sum_{j=1}^{n} F_{j,n+1}\right)(P^{(n)}_{\mathrm{sym}} \otimes \mathrm{I}),$$

where F swaps the indicated factors. Partial trace against E gives $(\mathrm{Tr}\, E)\mathrm{I} + n\,\mathrm{Tr}_{2,\ldots,n}\, E$. Finally, $D_{n+1}/D_n = (n+d)/(n+1)$ supplies the denominator in Equation (51).

Since a pure projector squares to itself, the conditional expected loss is $\mathrm{Tr}(M_E\sigma^{-1}) - 1$. Since $M_E \succeq \mathrm{I}/(n+d)$, a singular output has infinite conditional loss. Matrix Cauchy–Schwarz gives

$$\inf_{\substack{\sigma \in \mathcal{D}_d \\ \sigma > 0}} \mathrm{Tr}(M_E\sigma^{-1}) = (\mathrm{Tr}\,\sqrt{M_E})^2, \qquad \sigma_{\mathrm{opt}} = \frac{\sqrt{M_E}}{\mathrm{Tr}\,\sqrt{M_E}}.$$

Concavity of $\mathrm{Tr}\,\sqrt{\mathrm{I} + n\tau}$ implies that its minimum over density matrices τ is attained at a pure state; its value there is $\sqrt{n+1} + d - 1$. Thus every measurement has Haar Bayes risk at least the right-hand side of Equation (50).

The covariant POVM of Section 4 has $\tau_E = |v\rangle\langle v|$. Returning

$$\widehat{\sigma}_v = \frac{\sqrt{n+1}|v\rangle\langle v| + (\mathrm{I} - |v\rangle\langle v|)}{\sqrt{n+1} + d - 1}$$

achieves the lower bound for every outcome. The procedure is covariant, so its risk is the same for every true pure state. Its Bayes risk is therefore also its worst-case risk, proving minimax equality. For $n = 0$, the Haar posterior mean is I/d, so the same Cauchy–Schwarz bound gives risk at least $d - 1$. The uniform estimate attains it, agreeing with the formula. □

For n much larger than d^2, the expression is asymptotic to $2(d-1)/\sqrt{n}$. This agrees with the d^2/ε^2 copy scale, while also showing why a bare infidelity-to-chi-squared conversion can miss a dimension factor.

B Common centers and confidence bounds

Proposition B.1 (Common center of two flat states)**.** *Let P, Q be rank-r projectors with principal angles $\theta_1, \ldots, \theta_r$, in dimension at least $2r$. Then*

$$\inf_{\sigma} \max\{\chi^2_{\mathrm{R}}(P/r \,\|\, \sigma), \chi^2_{\mathrm{R}}(Q/r \,\|\, \sigma)\} = \left(\frac{1}{r}\sum_{i=1}^{r}\sqrt{1+\sin\theta_i}\right)^2 - 1 \;\geq\; (\sqrt{2}-1)\frac{\|P-Q\|_1}{r}. \tag{52}$$

The same minimum is obtained if the maximum is replaced by the arithmetic mean. A minimizing state is proportional to $\sqrt{P+Q}$ on its support.

Proof. The average loss plus one is

$$\mathrm{Tr}\left[\frac{P+Q}{2r^2}\sigma^{-1}\right].$$

Its minimum is the squared trace of the square root of $(P+Q)/(2r^2)$, by the same Cauchy–Schwarz calculation used above. In a principal-angle decomposition, the eigenvalues of $P+Q$ are $1 \pm \cos\theta_i$, including zero eigenvalues when an angle is zero. Therefore the minimum equals the displayed expression. At the proposed minimizer the two losses are equal: the principal-angle decomposition provides a unitary swapping P and Q and leaving $P+Q$ fixed. Hence the minimum of the maximum is the same as that of the average.

Use $\sqrt{1+x} \geq 1 + (\sqrt{2}-1)x$ for $0 \leq x \leq 1$, and $\|P-Q\|_1 = 2\sum_i \sin\theta_i$, to obtain the inequality. □

For two pure states this radius is exactly $\sin\theta$. This differs from the quadratic small-angle behavior of squared Bures distance, and is another direct way to see why the two chi-squared definitions cannot be interchanged.

Dependence on confidence.

The main statements use success probability 2/3. A simple two-point argument adds a necessary confidence term for $0 < \delta \leq 1/3$. Choose two pure states with angle θ and write $u = \sin\theta$. Their n-copy optimal equal-prior discrimination error is

$$\frac{1-\sqrt{1-(1-u^2)^n}}{2}.$$

If their two loss balls are disjoint, a learner failing with probability at most δ on each state would discriminate them with error at most δ. Thus

$$n \geq \frac{\log[1/(4\delta(1-\delta))]}{-\log(1-u^2)}. \tag{53}$$

For Bures chi-squared, take $u = 2\sqrt{\varepsilon}$; the trace-norm inequality in Lemma 1.4 makes the loss balls disjoint. For right-inverse chi-squared, take $u = 2\varepsilon$ and apply Proposition B.1 with $r = 1$. For small ε, Equation (53) gives, respectively,

$$\Omega\left(\frac{\log(1/\delta)}{\varepsilon}\right), \qquad \Omega\left(\frac{\log(1/\delta)}{\varepsilon^2}\right).$$

Every rank-at-most-r class contains these pure states. Combining these bounds with Theorem 1.2 adds $\log(1/\delta)$ to the corresponding dimension-dependent numerator, up to constants. These give necessary confidence terms in addition to the constant-success bounds.

C Block replacement and spectral weights

The shell argument in Section 5 avoids two pitfalls of local Frobenius refinement. First, replacing a block need not decrease the global error. Write

$$\rho = \begin{pmatrix} R & C \\ C^* & B \end{pmatrix}, \qquad \sigma_1 = D \oplus E,$$

where D and E are diagonal. After conjugation by $U \oplus \mathrm{I}$, replace UDU^* by a new diagonal block D_2 and set $\widetilde{\sigma} = D_2 \oplus E$. With $\rho' = (U \oplus \mathrm{I})\rho(U^* \oplus \mathrm{I})$, block orthogonality gives

$$\|\rho' - \widetilde{\sigma}\|_{\mathrm{F}}^2 = \|\rho - \sigma_1\|_{\mathrm{F}}^2 - \|R - D\|_{\mathrm{F}}^2 + \|URU^* - D_2\|_{\mathrm{F}}^2 . \tag{54}$$

The cross-block norm is unchanged because $\|UC\|_{\mathrm{F}} = \|C\|_{\mathrm{F}}$. Thus improvement requires comparison with the old error on the replaced block, not with the old global error. For example,

$$\rho = \mathrm{diag}(.3, .3, .2, .2), \quad \sigma_1 = \mathrm{diag}(.3, .3, .25, .15), \quad \widetilde{\sigma} = \mathrm{diag}(.32, .28, .25, .15)$$

are density matrices. The old global error is .005 and the new first-block error is .0008, but the new global error is .0058, since the old first-block error was zero.

Trace preservation is separate: $\mathrm{Tr}\,\widetilde{\sigma} - \mathrm{Tr}\,\sigma_1 = \mathrm{Tr}\,D_2 - \mathrm{Tr}\,D$. For instance, replacing the first entry .5 of $\mathrm{diag}(.5, .5)$ by .6 produces trace 1.1, even when .6 is the correct true entry. Positive replacement blocks preserve positivity but not normalization; Lemma 5.2 accounts for the final normalization.

Second, Frobenius error alone cannot uniformly bound either chi-squared loss. For fixed $0 < b < 1$, take

$$\sigma_q = \mathrm{diag}(q, 1-q), \qquad \rho_q = \mathrm{diag}(q+b, 1-q-b), \qquad 0 < q < 1-b.$$

Then

$$\|\rho_q - \sigma_q\|_{\mathrm{F}}^2 = 2b^2, \qquad \chi^2_{\mathrm{B}}(\rho_q \,\|\, \sigma_q) = \chi^2_{\mathrm{R}}(\rho_q \,\|\, \sigma_q) = b^2 \left(\frac{1}{q} + \frac{1}{1-q}\right) \longrightarrow \infty \tag{55}$$

as $q \downarrow 0$. The eigenvalue thresholds in the Bures construction, and the regularization floor in the right-inverse construction, supply the missing denominator control.