

Bayesian Weak Commutativity Does Not Guarantee Attainability of the Nagaoka–Hayashi Bound

Qihang Wang^{1,*}

¹*School of Mathematical Sciences, Peking University, Beijing 100871, China*

Bayesian weak commutativity has been suggested as the most likely condition for reaching the Nagaoka–Hayashi (NH) bound in multiparameter estimation. We show that it is not sufficient. For a full-rank four-level ensemble obeying it exactly, the effects required for NH equality cannot sum to the identity. We find the optimal risk exactly: it lies strictly above the bound and needs seven outcomes, while projective readout on the four levels adds over seven times the NH gap. A robust normalization witness bounds the gap from below across a family including truncated oscillator ladders.

Introduction—When a quantum probe must reveal several parameters at once, such as the components of a vector field [1], the optimal measurements for the individual parameters are generally incompatible [2–5]. With collective measurements on many copies, the quantum Cramér–Rao bound is asymptotically attainable exactly when the symmetric logarithmic derivatives commute on average in the probe state, the weak-commutativity condition [4]. For single-copy, locally unbiased estimation, the Nagaoka–Hayashi (NH) bound extends Nagaoka’s two-parameter bound [6, 7] to more parameters [8, 9]; it is a semidefinite program [9] with a Bayesian counterpart [10]. In the local setting, a vanishing averaged commutator does not guarantee that NH is reached: a rank-three four-level example misses the bound [11].

Bayesian estimation asks a different question: the parameters follow a known prior, estimates may be biased, and one minimizes the prior-averaged squared error [10, 12–14]. Albarelli, Branford, and Rubio [14] showed that the Bayesian NH bound relaxes one requirement, namely that the first and second moments of the estimates come from a single normalized measurement. They proved that the bound is reached for two-parameter qubit models and suggested Bayesian weak commutativity as the most likely general condition for this; all their examples satisfy it and appear to reach the bound. Neither the local counterexample nor recent related results settle this: a different Bayesian bound attainable for qubits [15], numerical NH saturation for a qubit channel [16], and bounds on the number of outcomes an optimal measurement needs [17].

In this Letter, we show that the condition is not sufficient. A receiver is given one copy of a four-level system, prepared in one of four known full-rank states, and estimates two real coordinates; we minimize the prior-averaged squared error over all positive operator-valued measures (POVMs) and estimators. Weak commutativity holds exactly, but every effect required for NH equality has the same fixed ratio between two level populations, and the identity does not, so these effects can never form a complete measurement. A normalization witness, an observable that is nonpositive on all such effects but has

positive trace, captures this mismatch, and its robust form makes the risk gap quantitative. We find the exact optimal risk and a seven-outcome measurement attaining it, and prove that fewer outcomes cannot. Projective measurements (PVMs) on the same four levels have a strictly larger exact minimum, attained with four outcomes.

The obstruction is not tied to one ladder: across a family that contains truncated oscillator and pair-annihilation ladders, a robust witness shows that at least 1/3420 of the risk reduction predicted by NH is unattainable, and on the oscillator slice we find the optimal measurement exactly.

The estimation task—A prior π selects a dimensionless label $\theta \in \mathbb{R}^2$ and a state ρ_θ on $\mathbb{C}^4$. For unweighted squared error, only three prior moments matter: the average state $S = \mathbb{E}_\pi \rho_\theta$, the label–state correlations $B_j = \mathbb{E}_\pi(\theta_j \rho_\theta)$, and $\lambda = \mathbb{E}_\pi \|\theta\|^2$, which is the risk of always reporting zero. A measurement is a normalized, countably additive POVM P on an arbitrary outcome space $\mathcal{O}$, followed by a measurable report $a : \mathcal{O} \rightarrow \mathbb{R}^2$. Its risk is

$$\mathcal{L}(P, a) = \mathbb{E}_\pi \int_{\mathcal{O}} \|a(\omega) - \theta\|^2 \text{Tr}[\rho_\theta P(d\omega)]. \quad (1)$$

We write $\mathcal{L}_{\min}$ for the infimum over all measurements and reports. The symmetric posterior-mean operators S_j are defined by $SS_j + S_jS = 2B_j$ [12], and Bayesian weak commutativity is the condition

$$\text{Tr}(S[S_1, S_2]) = 0. \quad (2)$$

It constrains only this average; the two operators themselves need not commute. For Hermitian V_1, V_2 , the two-parameter NH bound reads

$$\mathcal{L}_{\text{NH}} = \lambda + \inf_{V_1, V_2} \left\{ \text{Tr}(S(V_1^2 + V_2^2)) - 2 \sum_j \text{Tr} B_j V_j + \|iS^{1/2}[V_1, V_2]S^{1/2}\|_1 \right\}. \quad (3)$$

The last term is a trace norm, and, unlike in the local bound, the pair V_1, V_2 is not constrained by unbiasedness. Equations (3) and (2) are Eqs. (45) and (75) of Ref. [14], whose Sec. 6 suggests the latter as the most likely attainability condition.

A four-level counterexample—Let $C = |1\rangle\langle 2| + 2|2\rangle\langle 3| + 3|3\rangle\langle 4|$ be a weighted ladder operator, and let $X = (C + C^\dagger)/2$ and $Y = (C - C^\dagger)/(2i)$ be its two finite-level quadratures. At the maximally mixed average $S = I/4$ the trace in Eq. (2) vanishes for every ensemble, so we make the populations unequal. With the population ratio $k = S_{33}/S_{11} = 1000001/1000000$, define

$$\begin{aligned} p &= \left[2 + 2k - \frac{4(k-1)}{9(1+k)^2} \right]^{-1}, \\ q &= kp, \quad t = q - \frac{4p(k-1)}{9(1+k)^2}, \\ S_k &= \text{diag}(p, p, q, t). \end{aligned} \quad (4)$$

Unless stated otherwise, k is fixed at this rational value. Set $U = \text{diag}(-1, -1, -1, 1)$, $K_{\pm,k} = S_k(I \pm U)/2$, and $B_k = CK_{+,k} + K_{-,k}C$. The nonzero entries of B_k , just above the diagonal, are $\beta_1 = p$, $\beta_2 = 2p$, and $\beta_3 = 3(q+t)$; we write $B_k = B_{1,k} + iB_{2,k}$ with Hermitian $B_{j,k}$. The four equiprobable preparations are

$$\theta_{s,j} = 10se_j, \quad \rho_{s,j} = S_k + sB_{j,k}/5, \quad s = \pm 1, j = 1, 2. \quad (5)$$

They have average state S_k , first moments $B_{j,k}$, and $\lambda = 100$. Diagonal dominance gives $\rho_{s,j} \succeq I_4/20$, so every preparation has full rank. The posterior-mean operators have upper entries h_j and $-ih_j$, with $h_1 = 1/2$, $h_2 = 2p/(p+q)$, and $h_3 = 3$. Thus

$$\text{Tr}(S_k[S_{1,k}, S_{2,k}]) = 2i[(p-q)h_2^2 + 9(q-t)] = 0. \quad (6)$$

The choice of t makes the two terms cancel; since the populations are unequal, this is a genuine constraint rather than an identity. We use the phase symmetry of the moments only to construct measurements, never to restrict competing ones.

This single task has three distinct precision levels:

$$\mathcal{L}_{\text{NH},k} < \mathcal{L}_{\text{min},k} < \mathcal{L}_{\text{PVM},k}. \quad (7)$$

Here $\mathcal{L}_{\text{PVM},k}$ is the infimum over projective measurements on the original $\mathbb{C}^4$ with arbitrary reporting rules. The optimum over general measurements and its distance from NH are

$$\mathcal{L}_{\text{min},k} = \lambda - \sum_i c_{i,k}, \quad \Gamma_k = \text{Tr } T_{0,k} - \sum_i c_{i,k} > 1/32, \quad (8)$$

where $T_{0,k} = \text{diag}(0, p, 4p+9t, 9q)$ and the algebraic numbers $c_{i,k}$ are defined by the construction below and in Appendix D. Their certified enclosures give $\mathcal{L}_{\text{min},k} \approx 94.2859248008006880$ and $\Gamma_k \approx 0.0359263355222758$, and the exact projective minimum of Eq. (15) is $\mathcal{L}_{\text{PVM},k} \approx 94.5557442485432417$. The gap Γ_k is thus about 0.6% of the risk reduction predicted by NH, $\lambda - \mathcal{L}_{\text{NH},k} = \text{Tr } T_{0,k} \approx 5.75$, whereas the extra cost of projective readout, $\mathcal{L}_{\text{PVM},k} - \mathcal{L}_{\text{min},k} \approx 0.270$, is more than seven times

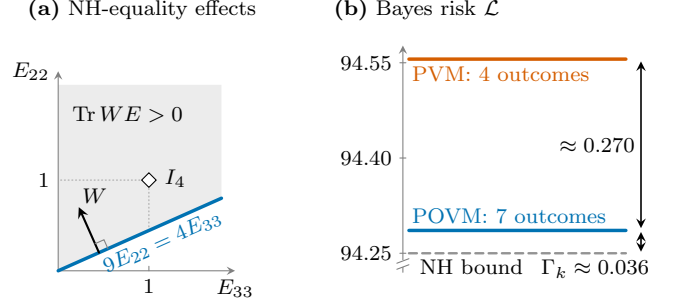


FIG. 1. Why weak commutativity is not enough, for the ensemble of Eq. (5). (a) Two diagonal entries of an effect E . Every effect allowed by NH equality, and every sum of them, lies on the ray $9E_{22} = 4E_{33}$, where $\text{Tr } WE = 0$ for the witness $W = \text{diag}(0, 9, -4, 0)$ (arrow, normal to the ray). The identity lies in the shaded region $\text{Tr } WE > 0$ ($\text{Tr } W = 5$) and cannot be reached. (b) NH bound (dashed) and exact optima over all POVMs and over PVMs on $\mathbb{C}^4$, rounded, on a truncated axis (no error bars); arrows mark Γ_k and the projective excess. Generative AI assisted the figure design and TikZ source.

larger [Fig. 1(b)]. Table I in Appendix E summarizes the three benchmarks.

Why the NH bound cannot be reached—For a Hermitian contraction R , put $K_{\pm}(R) = S_k^{1/2}(I \pm R)S_k^{1/2}/2$ and $q_R(Z) = \text{Tr}[ZK_+(R)Z^\dagger + Z^\dagger K_-(R)Z]$. The quadratic and trace-norm terms of Eq. (3) together equal $\sup_R q_R(Z)$, with $Z = V_1 + iV_2$. At $Z = C$ the supremum is attained at $R = U$, for which $K_{\pm}(U) = K_{\pm,k}$. Keep $R = U$ fixed and perturb $Z = C + \Delta$. By the definition of B_k , the terms linear in Δ cancel, leaving the nonnegative $q_U(\Delta)$. Taking the supremum over R can only raise the objective, and leaves it unchanged at C . Hence $Z = C$ is a global NH minimizer, the optimal NH operators are the quadratures, $V_1 = X$ and $V_2 = Y$, and $\mathcal{L}_{\text{NH},k} = \lambda - \text{Tr } T_{0,k}$.

To compare with actual measurements, group the outcomes by their complex report $z = a_1 + ia_2$; this defines a POVM M on the complex plane. With

$$F_k(z) = |z|^2 S_k - \bar{z} B_k - z B_k^\dagger + T_{0,k},$$

expanding the risk gives

$$\mathcal{L} - \mathcal{L}_{\text{NH},k} = \int_{\mathbb{C}} \text{Tr}[F_k(z)M(dz)]. \quad (9)$$

This is the standard Bayesian operator-duality comparison [14, 18]; local conic bounds [19] use a different objective. Because S_k has full rank, Eq. (9) holds for every finite-risk measurement, including continuous or singular ones with unbounded reports, and it does not require the first moments of M to match the NH optimizer.

Since $F_k(z) = (z - C)K_{+,k}(\bar{z} - C^\dagger) + (\bar{z} - C^\dagger)K_{-,k}(z - C) \succeq 0$, every report adds nonnegative excess risk, and

reaching NH requires all of it to vanish: the density of M with respect to the scalar measure $\mu = \text{Tr } M$ must have its range in the kernel of $F_k(z)$ for μ -almost every report z (Appendix B). A kernel vector satisfies $z\psi_1 = \psi_2$, $z\psi_2 = 2\psi_3$, $z\psi_3 = 3\psi_4$, and $\bar{z}\psi_4 = 3\psi_3$: the first three components of $C\psi = z\psi$ and the last component of $C^\dagger\psi = \bar{z}\psi$ hold. Nonzero solutions exist only at $z = 0$ and on the circle $|z| = 3$, and all obey $9|\psi_2|^2 = 4|\psi_3|^2$. Every effect built from them inherits this balance, whereas the identity has equal populations on both levels; summing such effects to the identity would require $9 = 4$, however many outcomes are used. No POVM, finite or continuous, attains NH.

Normalization witnesses—This population mismatch can be detected by a single observable, much as an entanglement witness detects entanglement [20]. Consider any comparison identity of the form (9), valid for every finite-risk measurement, with $F(z) \geq 0$. We call a Hermitian W a normalization witness if $\text{Tr } W > 0$ while $\psi^\dagger W \psi \leq 0$ for every kernel vector ψ of every $F(z)$. If a POVM attained NH, its density M_z would lie in these kernels, giving the contradiction $\text{Tr } W = \int \text{Tr}(WM_z) \mu(dz) \leq 0$. In our example $W = \text{diag}(0, 9, -4, 0)$ measures the population imbalance: $\psi^\dagger W \psi = 0$ on every kernel vector, whereas $\text{Tr } W = 5$ [Fig. 1(a)].

Conversely, if no witness exists, NH is attained. Suppose F is continuous and $F(z) \succ 0$ for large $|z|$, as here. The projectors onto unit kernel vectors then form a compact set of unit-trace matrices, which generates a closed convex cone. Without a witness, convex separation places the identity in this cone, and Carathéodory's theorem yields an NH-attaining POVM with at most 16 rank-one effects, the real dimension of 4×4 Hermitian matrices. Thus NH is attained if and only if no witness exists. For rank-one effects, membership in the cone is a weighted tight-frame condition [21].

A witness can also bound the size of the gap. Call W robust if $\text{Tr } W > 0$ and

$$\psi^\dagger W \psi \leq \kappa \|\psi\| [\psi^\dagger F(z) \psi]^{1/2} \quad \text{for all } z, \psi. \quad (10)$$

Applying Eq. (10) to the eigenvectors of each M_z , and the Cauchy-Schwarz inequality first to $\text{Tr}(WM_z)$ and then to the μ integral, whose total mass is $\text{Tr } I_4 = 4$, gives

$$\mathcal{L} - \mathcal{L}_{\text{NH}} \geq \frac{(\text{Tr } W)^2}{4\kappa^2} \quad (11)$$

for every measurement.

The optimal measurement—To find the exact optimum we sharpen the comparison by a constant matrix. Appendix D constructs an algebraic real matrix $R_\diamond$ with $R_\diamond + R_\diamond^T = -(B_k + B_k^\dagger)$ for which $R_\diamond S_k^{-1} R_\diamond^T$ is diagonal; its diagonal entries are the $c_{i,k}$. Put $A_\diamond = T_{0,k} - \text{diag}(c_{i,k})$ and $D(w) = \text{diag}(1, w, w^2, w^3)$. For every $r \geq 0$ and

$$|w| = 1,$$

$$F_k(rw) - A_\diamond = D(w)(rS_k + R_\diamond)S_k^{-1} \times (rS_k + R_\diamond)^T D(w)^\dagger \succeq 0. \quad (12)$$

Integrating against any POVM and using normalization gives $\int \text{Tr}[A_\diamond M] = \text{Tr } A_\diamond$, even though $A_\diamond$ need not be positive. Equation (12) therefore proves the lower bound in Eq. (8) with no restriction on the measurement or the estimator. Ancillas and sequential instruments on the received copy enter through their effective POVMs; extra encoded copies or postselection would change the task.

The same factorization shows how to attain the bound. The matrix $F_k(z) - A_\diamond$ is singular only on two circles, with radii $r_- < r_+$, so an optimal measurement reports only points on them. At $z = r_s w$ its kernel is spanned by $D(w)\nu_s$, where ν_s is real with positive components and $\nu_{s,1} = 1$. Positive algebraic weights α_s satisfy

$$\alpha_- + \alpha_+ = 1, \quad \sum_{s=\pm} \alpha_s \nu_{s,i}^2 = 1 \quad (1 \leq i \leq 4).$$

The optimal reported measurements are exactly those of the form

$$M_s(dw) = D(w)\nu_s \nu_s^T D(w)^\dagger \mu_s(dw), \quad (13)$$

on $z = r_s w$, $|w| = 1$, where the positive measures μ_s have masses α_s . Their first two Fourier moments vanish, and their third moments obey

$$\nu_{-,4} \int w^3 \mu_-(dw) + \nu_{+,4} \int w^3 \mu_+(dw) = 0.$$

The weights fix the diagonal of $\int M = I_4$ and the phase moments cancel its off-diagonal entries (Appendix E). Since each effect lies in the kernel at its own report, the lower bound is attained and is the exact optimum over all POVMs.

Why seven outcomes are needed—The phases on each circle must be rich enough to cancel the off-diagonal entries. Because the first and second moments vanish, the Gram matrix of the functions $1, w, w^2$ under each normalized circle measure is I_3 . It has rank three, so a finite optimal measurement uses at least three outcomes on each circle. With exactly three, the phases must form an equilateral triple with equal weights, whose normalized third moment has modulus one. Two such triples could satisfy the third-moment condition only if $\alpha_- \nu_{-,4} = \alpha_+ \nu_{+,4}$, and the certified algebraic solution violates this equality. Six outcomes are therefore not enough.

Seven are. Place an equilateral triple on the smaller circle. The larger circle then needs a normalized third moment of modulus $\alpha_- \nu_{-,4} / (\alpha_+ \nu_{+,4}) < 1$ together with vanishing first and second moments, and a positive measure on two complex-conjugate pairs of phases provides it (Appendix E and Fig. 2). This gives a seven-outcome optimum, while the counting argument excludes every

smaller finite measurement, including randomized reports. We count nonzero effects; measurements supported on several orbits also occur in covariant estimation [22, 23].

The cost of projective readout—Orthogonal readout adds an algebraic constraint. For projections P_i with complex reports z_i , the operator $Z = \sum_i z_i P_i$ is normal, $ZZ^\dagger = Z^\dagger Z$, and every normal Z arises this way. Its excess risk is a weighted quadratic penalty for the distance between Z and the non-normal ladder C ,

$$\mathcal{L} - \mathcal{L}_{\text{NH},k} = q_U(Z - C).$$

Writing $Z = \begin{pmatrix} A & b \\ c^\dagger & d \end{pmatrix}$, normality requires $AA^\dagger - A^\dagger A + bb^\dagger = cc^\dagger$, which has rank at most one. At the target ladder the left side equals $\text{diag}(1, 3, 5)$, of rank three. Bounding the weighted perturbation needed to remove this mismatch gives

$$\mathcal{L}_{\text{PVM},k} \geq \mathcal{L}_{\text{NH},k} + \frac{112p}{165} > \mathcal{L}_{\text{min},k} + \frac{25}{192}. \quad (14)$$

The proof in Appendix F covers all complex bases, coarse-grained projectors, reporting rules, and parameter-independent classical mixtures of single-copy settings. Related tools include normal completions [24, 25], the normal Procrustes problem [26], and Bayesian readout optimization [1].

The exact projective optimum has a simple form. With the entries β_i of B_k , put $D_0 = p + q + t$, $D_2 = 2p + q$, and $\delta = t - p$. The mixing ratio h_* is the unique root on the indicated branch:

$$(\beta_1 D_0 - D_2 \beta_3 h)^2 (1 + h^2) = \beta_2^2 \delta^2 h^2, \\ D_2 \beta_3 h > \beta_1 D_0.$$

Define $x_* = h_*/\sqrt{1 + h_*^2}$, $y_* = 1/\sqrt{1 + h_*^2}$, and $\omega = e^{2\pi i/3}$. An optimal basis is then

$$v_j = \frac{(x_*, \omega^j, \omega^{2j}, y_*)^T}{\sqrt{3}}, \quad j = 0, 1, 2, \\ v_3 = (y_*, 0, 0, -x_*)^T.$$

Writing $\tau_* = p + q + px_*^2 + ty_*^2$ and $A_* = \beta_2 + \beta_1 x_* + \beta_3 y_*$, its optimal reports are $z_j = (A_*/\tau_*)\omega^j$ and $z_3 = 0$. We prove

$$\mathcal{L}_{\text{PVM},k} = 100 - \frac{A_*^2}{\tau_*}. \quad (15)$$

To show that no basis does better, write $s_i = \text{Tr}(S_k P_i)$ for the prior-averaged probability of outcome i of any orthonormal basis and $w_i = |\text{Tr}(B_k P_i)|^2$; with optimal reports the risk is $100 - \sum_i w_i/s_i$. Since every $s_i \geq p$, we replace $1/s$ on $s \geq p$ by a quadratic majorant tangent to it at $s = \tau_*/3$, the value at the three signal outcomes of the attaining basis. The projectors enter the resulting polynomial through $R = \sum_i P_i \otimes P_i$, which obeys $R^2 = R$, $(R^{\text{T}_2})^2 = R^{\text{T}_2}$, and $(R \otimes I)(I \otimes R) = \sum_i P_i^{\otimes 3}$,

with T_2 the partial transpose on the second factor. From these identities, Appendix H builds a computer-assisted positive algebraic identity that bounds the gain for every complex basis; its coefficients, Gram matrices, and rational-arithmetic verifiers in the ancillary files are part of the proof. Appendices F and G give separate analytic and moment bounds.

Beyond a single ladder—To test whether the failure depends on the transition ratio $1 : 2 : 3$, we vary the ladder C_u , with entries just above the diagonal

$$\ell_1 = \sqrt{(1 - 2u)/3}, \quad \ell_2 = \sqrt{(2 - 2u)/3}, \\ \ell_3 = 1, \quad 0 \leq u \leq \frac{1}{3}.$$

Appendix J constructs four equiprobable states for each u , population ratio $k = S_{33}/S_{11} \in [1, 2]$, and contrast $\eta \in (0, 1/6]$. The labels are $\pm e_1, \pm e_2$, every state satisfies $\rho \succeq I_4/12$, and the average state is nonscalar for $k > 1$. Weak commutativity holds throughout this family, but

$$\mathcal{L}_{\text{min}}(u, k, \eta) - \mathcal{L}_{\text{NH}}(u, k, \eta) \geq \gamma \eta^2 > 0 \quad (16)$$

with $\gamma = 1/16416$. The effects required for NH equality obey $E_{22} = \ell_2^2 E_{33}$, so $W_u = \text{diag}(0, 1, -\ell_2^2, 0)$ is a normalization witness with $\text{Tr} W_u = 1 - \ell_2^2 = (1 + 2u)/3$. Appendix J proves its robust form (10) with $\kappa^2 = 19/p$, $p = S_{11}$, for reports rescaled by $2/\eta$. Equation (11), times the factor $\eta^2/4$ from this rescaling, gives the coefficient $p(1 - \ell_2^2)^2/304 \geq \gamma$. At least $1/3420$ of the risk reduction predicted by NH is unattainable by any measurement; both constants are conservative.

At $u = 0$, C_u is the truncation (compression) of the oscillator annihilator $\hat{a}/\sqrt{3}$ to the first four Fock levels. At $u = 1/(m+3)$, with integer $m \geq 0$, it is the normalized compression of the pair annihilator $\hat{a}\hat{b}$, a standard optical $\text{SU}(1,1)$ generator [27], to $\{|n+m, n\rangle : 0 \leq n \leq 3\}$. The correspondence concerns the ladders only: the preparations are mixtures of number states and balanced adjacent-level superpositions, the label selects the mixture rather than a displacement or an interferometric phase, and full rank refers to the four-level space. On the oscillator slice $u = 0$, a two-radius, eight-effect POVM attains the exact minimum over all measurements for every $k \in [1, 2]$ and $\eta \in (0, 1/6]$ (Appendix J). The minimal number of outcomes and the projective optimum are established only for the fixed model above.

Preparation noise and other priors—The gaps survive two kinds of preparation noise. Under the unheralded replacement channel $\rho \mapsto \zeta \rho + (1 - \zeta)S_k$, $0 < \zeta \leq 1$, the optimal POVM and PVM risks and the NH bound all obey $\mathcal{L}(\zeta) = 100 + \zeta^2[\mathcal{L}(1) - 100]$; weak commutativity, the optimal effects (with reports scaled by ζ), and the seven-outcome minimum are preserved. Replacing each state of Eq. (5) by any density matrix within operator-norm distance 2^{-16} , with the same labels, prior, and loss, leaves $\tilde{\mathcal{L}}_{\text{min}} - \tilde{\mathcal{L}}_{\text{NH}} > 1/64$ and $\tilde{\mathcal{L}}_{\text{PVM}} - \tilde{\mathcal{L}}_{\text{min}} > 1/10$. Even the unmodified seven-outcome readout stays

more than $1/10$ below every PVM reoptimized for the perturbed states, although such errors need not preserve weak commutativity or exact optimality (Appendix I).

Nor is the finite prior essential. The affine family $\rho_\theta = S_k + (\theta_1 B_{1,k} + \theta_2 B_{2,k})/300$ with a uniform prior on $[-30, 30]^2$ has the same $S_k, B_{j,k}$ and $\rho_\theta \succeq I_4/20$; only λ changes, to 600, adding 500 to every risk. The scalar average $S = I/4$ is treated in Appendices A–C. An earlier real-entry example over $\mathbb{Q}(\sqrt{2})$ [28], as corrected, shows only that NH-equality effects cannot be normalized, not a gap for all POVMs.

Conclusion—Bayesian weak commutativity does not guarantee that the NH bound can be reached. In our full-rank four-level ensemble the effects that would saturate NH cannot sum to the identity; the true optimum needs seven outcomes, and projective readout on $\mathbb{C}^4$ costs more than seven times the NH gap. Normalization witnesses turn this obstruction into a test: for a continuous comparison matrix that is positive definite at large reports, NH is attained exactly when no witness exists, and a robust witness bounds the gap from below, here across a ladder family containing truncated oscillators. These results concern ideal single-copy readout of finite-level systems, not hardware or interferometric advantages. Open questions include exact risks at other ladder shapes, the family’s projective penalty, and witnesses for phase or displacement encodings.

The author acknowledges Gewu Intelligence Lab for providing the collaborative research environment in which this project was developed.

Use of generative AI—OpenAI GPT-5.6 Sol, Anthropic Claude Fable 5, Grok 4.6, OpenAI GPT-6.1 Sol, and OpenAI GPT-6 Astra contributed to the development and preparation of this manuscript. Their combined work included model and example construction, analytical derivations, proof and certificate development, exact-arithmetic checking, literature analysis, physical interpretation, manuscript organization, LaTeX drafting, and revision. Anthropic Claude Opus 5.5 revised the final text line by line, checked the citations against primary records, and redrew both figures. The author directed the work, verified its content, and takes responsibility for it; every computational step can be rechecked with the exact-arithmetic replay checkers provided as ancillary files.

Data availability—All exact matrices, certificates, and rational-arithmetic replay checkers are included as ancillary files. No experimental data were used.

* 2501110049@stu.pku.edu.cn

[1] R. Kautbruegger, A. Shankar, D. V. Vasilyev, and P. Zoller, Optimal and variational multiparameter quantum metrology and vector-field sensing, *PRX Quantum* **4**, 020333 (2023), arXiv:2302.07785.

- [2] C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic Press, New York, 1976).
- [3] A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory*, 2nd ed. (Edizioni della Normale, Pisa, 2011).
- [4] S. Ragy, M. Jarzyna, and R. Demkowicz-Dobrzański, Compatibility in multiparameter quantum metrology, *Phys. Rev. A* **94**, 052108 (2016), arXiv:1608.02634.
- [5] R. Demkowicz-Dobrzański, W. Górecki, and M. Guţă, Multi-parameter estimation beyond quantum Fisher information, *J. Phys. A* **53**, 363001 (2020), arXiv:2001.11742.
- [6] H. Nagaoka, A new approach to Cramér–Rao bounds for quantum state estimation, in *Asymptotic Theory of Quantum Statistical Inference: Selected Papers*, edited by M. Hayashi (World Scientific, Singapore, 2005) pp. 100–112, reprinted from IEICE Tech. Rep. IT89-42 (1989).
- [7] H. Nagaoka, A generalization of the simultaneous diagonalization of Hermitian matrices and its relation to quantum estimation theory, *Trans. Jpn. Soc. Ind. Appl. Math.* **1**, 305 (1991), in Japanese.
- [8] M. Hayashi, On simultaneous measurement of noncommutative observables, in *Development of Infinite-Dimensional Noncommutative Analysis*, RIMS Kôkyûroku, Vol. 1099, edited by A. Hora (Research Institute for Mathematical Sciences, Kyoto University, Kyoto, 1999) pp. 96–118, in Japanese.
- [9] L. O. Conlon, J. Suzuki, P. K. Lam, and S. M. Assad, Efficient computation of the Nagaoka–Hayashi bound for multiparameter estimation with separable measurements, *npj Quantum Inf.* **7**, 110 (2021).
- [10] J. Suzuki, Bayesian Nagaoka–Hayashi bound for multiparameter quantum-state estimation problem, *IEICE Trans. Fundam. Electron. Commun. Comput. Sci.* **E107.A**, 510 (2024).
- [11] L. O. Conlon, J. Suzuki, P. K. Lam, and S. M. Assad, Role of the extended Hilbert space in the attainability of the quantum Cramér–Rao bound for multiparameter estimation, *Phys. Lett. A* **542**, 130445 (2025), arXiv:2404.01520.
- [12] S. Personick, Application of quantum estimation theory to analog communication over quantum channels, *IEEE Trans. Inf. Theory* **17**, 240 (1971).
- [13] J. Rubio and J. Dunningham, Bayesian multiparameter quantum metrology with limited data, *Phys. Rev. A* **101**, 032114 (2020), arXiv:1906.04123.
- [14] F. Albarelli, D. Branford, and J. Rubio, Measurement incompatibility in Bayesian multiparameter quantum estimation, *Quantum* **10**, 2192 (2026), arXiv:2511.16645.
- [15] K.-H. Zhao, K. Yamagata, and J. Suzuki, Bayesian Gill–Massar bound: An attainable lower bound for qubit parameter estimation (2026), arXiv:2607.07031.
- [16] L. Bia and C. N. Gagatsos, Saturating the Bayesian Nagaoka–Hayashi bound within numerical precision for the depolarization $SU(2)$ rotation channel (2026), *phys. Rev. A* (to be published), doi:10.1103/p7vp-vyxf, arXiv:2607.15398.
- [17] K. Yamagata, Sufficient support size of measurements for quantum estimation (2026), arXiv:2604.21323.
- [18] A. S. Holevo, Statistical decision theory for quantum systems, *J. Multivariate Anal.* **3**, 337 (1973).
- [19] M. Hayashi and Y. Ouyang, Tight Cramér–Rao type bounds for multiparameter quantum metrology through conic programming, *Quantum* **7**, 1094 (2023).
- [20] B. M. Terhal, Bell inequalities and the separability criterion, *Phys. Lett. A* **271**, 319 (2000), arXiv:quant-ph/9911057.

- [21] Y. C. Eldar and G. D. Forney, Jr., Optimal tight frames and quantum measurement, *IEEE Trans. Inf. Theory* **48**, 599 (2002), arXiv:quant-ph/0106070.
- [22] G. Chiribella and G. M. D'Ariano, Extremal covariant measurements, *J. Math. Phys.* **47**, 092107 (2006), arXiv:quant-ph/0603168.
- [23] J. Calsamiglia, M. Aspachs, R. Muñoz-Tapia, and E. Bagan, Phase-covariant quantum benchmarks, *Phys. Rev. A* **79**, 050301(R) (2009), arXiv:0807.5126.
- [24] D. S. Kaliuzhnyi-Verbovetskyi, I. M. Spitkovsky, and H. J. Woerdeman, Matrices with normal defect one, *Oper. Matrices* **3**, 401 (2009), arXiv:0903.0090.
- [25] R. D. Wasson and H. J. Woerdeman, The normal defect of some classes of matrices, *Linear Algebra Appl.* **438**, 3530 (2013), arXiv:1211.4145.
- [26] K. Bierly, The normal Procrustes problem: A Riemannian optimization approach (2026), arXiv:2608.19513.
- [27] N. Fabre, A. B. Klimov, G. Leuchs, and L. L. Sánchez-Soto, Local sampling of the $SU(1,1)$ Wigner function, *AVS Quantum Sci.* **5**, 014404 (2023).
- [28] Q. Wang, Nonorthogonal Nagaoka equality supports under Bayesian weak commutativity (2026), Zenodo preprint, doi:10.5281/zenodo.22083724.
- [29] C. K. D. Law and J. Z. Bernád, Bayesian approach to simultaneous quantum multiparameter estimation with finite data (2026), arXiv:2608.10230.
- [30] D. Martínez, L. Pereira, K. Sawada, P. González, J. Cariñe, M. Muñoz, A. Delgado, E. S. Gómez, S. P. Walborn, and G. Lima, Experimental qudit state estimation via point tomography, *Phys. Rev. Res.* **7**, 033119 (2025), arXiv:2412.14915.
- [31] K. Glashoff and M. M. Bronstein, Matrix commutators: their asymptotic metric properties and relation to approximate joint diagonalization, *Linear Algebra Appl.* **439**, 2503 (2013), arXiv:1305.2135.
- [32] V. Noferini and M. Zhukov, Closest normal matrix found again using Riemannian optimization (2026), arXiv:2608.28545.
- [33] J. B. Lasserre, Global optimization with polynomials and the problem of moments, *SIAM J. Optim.* **11**, 796 (2001).

Appendix A: A scalar-average illustration

We begin with a maximally mixed average state, for which weak commutativity holds automatically. This simpler model introduces the comparison matrix used to solve the nonscalar model in the main text. Let the only nonzero entries of a weighted ladder C be

$$C_{12} = 1, \quad C_{23} = 2, \quad C_{34} = 3. \quad (\text{A1})$$

Its Hermitian components $X = (C + C^\dagger)/2$ and $Y = (C - C^\dagger)/(2i)$ are finite-level quadratures, not canonical oscillator position and momentum. Put

$$\begin{aligned} U &= \text{diag}(-1, -1, -1, 1), \\ K_\pm &= (I_4 \pm U)/8, \quad B = CK_+ + K_-C, \\ B_1 &= (B + B^\dagger)/2, \quad B_2 = (B - B^\dagger)/(2i). \end{aligned} \quad (\text{A2})$$

Thus B has superdiagonal entries $1/4, 1/2, 3/2$. For $s = \pm 1$ and $j = 1, 2$, prepare one copy of

$$p_{s,j} = \frac{1}{4}, \quad \theta_{s,j} = 10se_j, \quad \rho_{s,j} = I_4/4 + sB_j/5. \quad (\text{A3})$$

The classical label is not given to the receiver. Both B_j are traceless and Hermitian; diagonal dominance gives $\rho_{s,j} \succeq I_4/20$. The average and first moments are exactly $S = I_4/4, B_j$, with $\lambda = 100$. The unique posterior means are $S_j = 4B_j$. Weak commutativity follows from $\text{Tr}[S_1, S_2] = 0$, although the operators themselves need not commute. Indeed, a scalar average state makes this trace test automatic: the obstruction concerns what that test fails to detect.

Theorem 1 (Exact Bayesian risk and strict gap). *For the ensemble (A3) and the loss (1),*

$$\mathcal{L}_{\text{NH}} = \frac{377}{4}, \quad \mathcal{L}_{\min} = 100 - 4f(u_*, v_*) = \frac{377}{4} + \Gamma_*. \quad (\text{A4})$$

The algebraic pair and rational function f are defined below and in Appendix C. The exact value satisfies $0.03592634498992 < \Gamma_* < 0.03592634498993$. An explicit eight-effect POVM attains the minimum. The matching lower bound holds for every POVM and measurable estimator, with no bound on outcome number or estimate range.

The terminating decimals bound an algebraic value; they are not its exact specification. For comparison, the Fourier projectors onto $(1, w, w^2, w^3)^T/2$, $w \in \{1, i, -1, -i\}$, with reports $9w/4$, have risk $1519/16$ and are therefore not optimal.

Minimizing the NH expression

For a Hermitian contraction $-I \preceq R \preceq I$, define $K_\pm(R) = S^{1/2}(I \pm R)S^{1/2}/2$ and

$$q_R(Z) = \text{Tr}[ZK_+(R)Z^\dagger + Z^\dagger K_-(R)Z].$$

The quadratic plus trace norm in (3) is $\sup_R q_R(Z)$, where $Z = V_1 + iV_2$. For (A1),

$$iS^{1/2}[X, Y]S^{1/2} = \text{diag}(-1, -3, -5, 9)/8,$$

so U is the active sign and (A2) is its stationarity condition. For every perturbation E , the active branch minus the linear term changes by $q_U(E) \geq 0$. The supremum is no smaller than that branch and agrees with it at C . This proves a *global* NH minimum, without uniqueness. In particular,

$$T_0 = CK_+C^\dagger + C^\dagger K_-C = \text{diag}(0, 1/4, 13/4, 9/4), \quad (\text{A5})$$

and $\mathcal{L}_{\text{NH}} = \lambda - \text{Tr } T_0 = 377/4$.

A lower bound for every measurement

Group the outcomes of P by their reported estimates a . This gives a POVM M on the complex estimate plane $z = a_1 + ia_2$. The positive comparison matrix is

$$F(z) = |z|^2 S - \bar{z}B - zB^\dagger + T_0, \quad (\text{A6})$$

and direct risk expansion yields

$$\mathcal{L}(P, a) - \mathcal{L}_{\text{NH}} = \int \text{Tr}[M(\text{d}z)F(z)]. \quad (\text{A7})$$

There is no requirement that the measurement's first moments equal X, Y . Finite risk suffices for integrability; infinite-risk procedures obey the bound trivially. Operator-dual formulations go back to Ref. [18] and are used for the Bayesian NH bound in Ref. [14], alongside distinct conic formulations of local bounds [19].

A simple rational witness is the real matrix

$$G = \begin{pmatrix} 0 & -1/16 & 0 & 2/47 \\ -3/16 & 0 & 1/94 & 0 \\ 0 & -24/47 & 0 & -3/4 \\ -2/47 & 0 & -3/4 & 0 \end{pmatrix}. \quad (\text{A8})$$

It satisfies $G + G^T = -(B + B^\dagger)$, and GG^T is diagonal. Consequently

$$\begin{aligned} A &= T_0 - 4GG^T \\ &= \frac{1}{141376} \text{diag}(-3233, 15399, -6080, -1024), \\ \text{Tr } A &= \gamma = \frac{2531}{70688}. \end{aligned} \quad (\text{A9})$$

For $z = rw$, $r \geq 0$, $|w| = 1$, put $D(w) = \text{diag}(1, w, w^2, w^3)$ and $Q(r) = rI_4/2 + 2G$. The following identity holds over the entire estimate plane:

$$F(rw) - A = D(w)Q(r)Q(r)^T D(w)^\dagger \succeq 0. \quad (\text{A10})$$

At the origin take $r = 0, w = 1$. This is a polynomial coefficient identity, not a sampled positivity test. Although A is indefinite, POVM normalization gives

$$\int \text{Tr}[M(\text{d}z)A] = \text{Tr } A.$$

Integrating (A10) in (A7) proves the simple rational lower bound $\mathcal{L}_{\min} - \mathcal{L}_{\text{NH}} \geq \gamma$. The number γ is a certified lower bound, not the exact excess risk. Its scale can be compared with the NH-allowed gain $\lambda - \mathcal{L}_{\text{NH}} = 23/4$, giving a missed fraction of at least $\gamma/(23/4)$. This is a relaxation error, not a sensing advantage or an incompatibility measure invariant under arbitrary changes of loss. Parameter-independent ancillas, classical randomization, and sequential instruments on the same received copy induce effective POVMs and are included. Every abort outcome must remain part of the normalized experiment; either conditional postselection or the use of extra parameter-encoded copies changes the task.

A matching optimal measurement

The rational comparison can be improved until it is met by a measurement. Write $(w_0, m_0, t_0) = (1/4, 1/2, 3/2)$, and define

$$\begin{aligned} d &= v(t_0 + v) - u(w_0 + u), \\ a &= um_0(t_0 + v)/d, \quad b = um_0(w_0 + u)/d, \\ G(u, v) &= \begin{pmatrix} 0 & u & 0 & a \\ -w_0 - u & 0 & b & 0 \\ 0 & -m_0 - b & 0 & v \\ -a & 0 & -t_0 - v & 0 \end{pmatrix}. \end{aligned} \quad (\text{A11})$$

Put $f(u, v) = \|G(u, v)\|_F^2$. The selected pair (u_*, v_*) is the unique stationary point of f in the rational rectangle

$$\begin{aligned} -0.06554650 &< u < -0.06554649, \\ -0.75260404 &< v < -0.75260402. \end{aligned}$$

Appendix C fixes its algebraic branch and proves existence, nonvanishing denominators and the exact normalization identities. Set $G_* = G(u_*, v_*)$, $c_i = 4(G_* G_*^T)_{ii}$, $A_* = T_0 - \text{diag}(c_i)$, and $\Gamma_* = \text{Tr } A_* = 23/4 - 4f(u_*, v_*)$. The same coefficient identity gives, for every $z = rw$,

$$F(rw) - A_* = D(w)(rI/2 + 2G_*)(rI/2 + 2G_*)^T D(w)^\dagger \succeq 0. \quad (\text{A12})$$

Hence every experiment has risk at least $100 - 4f(u_*, v_*)$.

There are two positive radii $r_- < r_+$ at which the radial matrix $F(r) - A_*$ is singular. Its real kernel vectors ν_- , ν_+ are fixed with first component one, and positive weights α_- , α_+ satisfy

$$\alpha_- + \alpha_+ = 1, \quad \alpha_- \nu_{-,i}^2 + \alpha_+ \nu_{+,i}^2 = 1 \quad (1 \leq i \leq 4).$$

All these quantities are explicitly algebraic in Appendix C. For each sign s and fourth root of unity w , use

$$M_{s,w} = \frac{\alpha_s}{4} D(w) \nu_s \nu_s^T D(w)^\dagger, \quad z_{s,w} = r_s w. \quad (\text{A13})$$

Summing the four phases removes every off-diagonal entry, and the displayed balance gives $\sum_{s,w} M_{s,w} = I_4$. Each effect is supported in $\ker[F(z_{s,w}) - A_*]$, so its comparison residual vanishes. This protocol meets the unrestricted lower bound and proves Eq. (A4). Global optimality follows from this primal–dual equality, not from stationarity or optimization within a covariant ansatz. The two-seed construction uses established finite-group covariance [22]; neither uniqueness nor minimal outcome number is claimed.

What obstructs NH equality?— The complete kernel locus provides a complementary interpretation. A vector $\psi \in \ker F(z)$ obeys

$$z\psi_1 = \psi_2, \quad z\psi_2 = 2\psi_3, \quad z\psi_3 = 3\psi_4, \quad \bar{z}\psi_4 = 3\psi_3.$$

These are the NH equality kernels $\ker F(z)$, not the optimal-cost kernels $\ker[F(z) - A_*]$. At zero the former kernel is $\text{span}(e_1)$; otherwise it is one-dimensional and exists exactly on $|z| = 3$. Every such vector satisfies $9|\psi_2|^2 = 4|\psi_3|^2$. All effects confined to these kernels inherit $9M_{22} = 4M_{33}$, whereas a resolution of the identity would require $9 = 4$. This excludes equality even for an overcomplete or continuous POVM. Kernel exclusion alone need not imply a gap to the infimum; the uniform factor (A10) supplies that stronger conclusion.

A continuous prior with the same moments

The same moments admit a smooth, identifiable realization. For a uniform prior on the connected square $[-30, 30]^2$, let

$$p(\theta) = \frac{1}{3600}, \quad \rho_\theta = \frac{I_4}{4} + \frac{\theta_1 B_1 + \theta_2 B_2}{300}. \quad (\text{A14})$$

Its covariance is $300I_2$, so its state moments remain S, B_j . Diagonal dominance gives $\rho_\theta \succeq I_4/20$ throughout the square. The state map is affine and injective in the two real coordinates, and is smooth on an open neighborhood of the support. Here $\lambda = 600$, $\mathcal{L}_{\text{NH}} = 2377/4$, and $\mathcal{L}_{\text{min}} = 600 - 4f(u_*, v_*) = 2377/4 + \Gamma_*$. The two parameters encode linearly independent real and imaginary coherence directions. The prior is uniform with a boundary, not globally smooth on $\mathbb{R}^2$.

Appendix B: Global optimum, arbitrary outcomes, and extensions

We first prove that the NH minimum and the measurement lower bound hold without restricting the outcome space. Here $S, B_j, C, U, T_0, F, G, A$ refer to the scalar weighted-ladder model in Appendix A. Appendix D treats the nonscalar model used in the main text.

The NH minimum and measurements with arbitrary outcomes

The NH expression is a supremum of quadratic-minus-linear functions. We identify one that is minimized at C and equals the supremum there. For $Z = V_1 + iV_2$, expansion gives

$$q_R(Z) = \text{Tr } S(V_1^2 + V_2^2) + \text{Tr} \left[R i S^{1/2} [V_1, V_2] S^{1/2} \right].$$

The dual representation of the Hermitian trace norm proves (3). Each q_R is a positive-semidefinite quadratic because its two weights are positive. At C , the U branch is active. The equation $B = CK_+ + K_-C$ cancels the terms linear in an arbitrary complex perturbation E :

$$q_U(C + E) - 2 \text{Re Tr}[B^\dagger(C + E)] = q_U(C) - 2 \text{Re Tr}(B^\dagger C) + q_U(E).$$

The supremum over R is at least this branch at every $C + E$, and agrees with it at C . Thus C minimizes the full NH expression, not just a stationary branch. Stationarity and quadratic homogeneity give $q_U(C) = \text{Re Tr}(B^\dagger C) = \text{Tr } T_0$.

The positive dual matrix has the signed decomposition

$$F(z) = (zI - C)K_+(\bar{z}I - C^\dagger) + (\bar{z}I - C^\dagger)K_-(zI - C) \succeq 0.$$

This yields (A6). To compare with a general measurement, group its outcomes by the reported estimate. A randomized reporting rule can likewise be included in the POVM through its classical Markov kernel. The finite scalar measure $\mu(E) = \text{Tr } M(E)$ has total mass four; write $M(dz) = M_z \mu(dz)$, with $M_z \succeq 0$ almost everywhere. Finite risk and the finite prior second moment imply a finite predictive estimate second moment by $|z|^2 \leq 2|z - \theta|^2 + 2|\theta|^2$. For every full-rank average $S \succeq s_* I$, $s_* > 0$, the predictive measure dominates $s_* \mu$, controlling its estimate second moment. In the simplest model $s_* = 1/4$; Cauchy-Schwarz controls the first moment. Expansion of squared loss is therefore legitimate and proves (A7).

For the global factor, use $D(w)BD(w)^\dagger = \bar{w}B$. The diagonal matrices S, T_0, A, GG^T commute with $D(w)$, and

$$Q(r)Q(r)^T = r^2 I/4 + r(G + G^T) + 4GG^T = r^2 I/4 - r(B + B^\dagger) + T_0 - A.$$

This proves (A10), including the origin. Positivity implies $\text{Tr}[M_z(F(z) - A)] \geq 0$ almost everywhere. The fixed matrix A is bounded and integrable, even though it is indefinite. Normalization gives its integral as $\text{Tr } A$. The resulting lower bound does not require a finite outcome set or the existence of a minimizing measurement. Nor does it impose a first-moment constraint on the measurement.

No density relative to Lebesgue measure is assumed: the argument covers atomic, continuous and singular measures, including uncountable supports. A parameter-independent ancilla and normalized sequential instruments on the same received copy induce a normalized effective POVM. In adjoint form, their positive outcome maps give $P(E) = \Phi_E^*(I)$, and trace preservation of their total instrument gives $P(\mathcal{O}) = I$. Adaptive feedforward and classical randomization are therefore included. All failure outcomes remain normalized. Conditional postselected loss or another parameter-encoded resource defines a different experiment.

A nonoptimal Fourier benchmark

For the four fourth roots of unity w , the vectors $\psi_w = (1, w, w^2, w^3)^T/2$ are orthonormal, by the finite geometric sum of $(w\bar{v})^j$, $0 \leq j \leq 3$. Hence $\sum_w \psi_w \psi_w^\dagger = I$. They satisfy $\psi_w^\dagger B \psi_w = 9w/16$, $\psi_w^\dagger S \psi_w = 1/4$, and $\psi_w^\dagger T_0 \psi_w = 23/16$. Reports $z_w = 9w/4$ give

$$\sum_w \psi_w^\dagger F(z_w) \psi_w = 11/16.$$

This proves the nonsharp excess-risk upper bound $11/16$ for every realization with the same S, B, T_0 . The finite-ensemble risk is $1519/16$; the square-prior risk is $\lambda - \text{Tr } T_0 + 11/16$. Neither the hidden label nor a parameter-dependent measurement is used. Appendix C supplies a strictly better attaining protocol; these projectors are not optimal.

Replacing the maximally mixed average

For Eq. (4), the only nonzero entries of B_k are

$$B_{k12} = p, \quad B_{k23} = 2p, \quad B_{k34} = 3(q+t), \quad T_{0,k} = \text{diag}(0, p, 4p+9t, 9q).$$

Thus $B_{j,k}$ are traceless and Hermitian. The posterior means have zero diagonal, respective upper entries h_j and $-ih_j$, and conjugate lower entries. Their Sylvester equations hold entrywise. Their commutator gives Eq. (6); the first edge contributes zero since the first two average eigenvalues are both p . The other contributions cancel by the chosen t , not by an automatic trace identity for nonscalar averages.

The signed commutator at the unchanged C is diagonal, with entries $-p/2, -3p/2, -5q/2, 9t/2$. All p, q, t are positive, so its active sign remains U . The weights $K_{\pm,k}$ are positive and obey $B_k = CK_{+,k} + K_{-,k}C$. The active-branch proof applies to every complex perturbation, not only weighted ladders, yielding the unrestricted global NH minimum $\lambda - \text{Tr } T_{0,k}$.

The four equally weighted preparations $S_k \pm B_{j,k}/5$ at $\pm 10e_j$ have moments $S_k, B_{j,k}$. The square realization follows from covariance $300I_2$. For the specified rational k , exact diagonal dominance gives $\rho_\theta \succeq I/20$ in both cases, throughout the square in the continuous case. The independent real and imaginary coherence directions preserve identifiability and smoothness on a neighborhood of the support.

To cover every reporting rule, let K be the closed convex hull of the prior support. Euclidean metric projection Π_K satisfies $\|\Pi_K(a) - \theta\| \leq \|a - \theta\|$ for every $\theta \in K$, by its variational inequality. Replacing each report by $\Pi_K(a)$, also in a randomized rule, never increases risk. Thus every original strategy has risk at least that of a strategy reporting in K . There is no conditioning or truncation of the POVM.

The finite-prior hull has $|z| \leq 10$; the square has $|z|^2 \leq 1800$ and $|z| \leq 45$. Let $d_S = \|S_k - S\|_{\text{op}}$, $d_B = \|B_k - B\|_{\text{op}}$, and $d_T = \|T_{0,k} - T_0\|_{\text{op}}$. The first and last differences are diagonal, and the middle is a weighted shift whose norm is its largest entry modulus. Exact rational comparisons give

$$d_S \leq 1/7000000, \quad d_B \leq 3/4000000, \quad d_T \leq 5/4000000.$$

Uniformly on the projected region,

$$\|F_k(z) - F(z)\|_{\text{op}} \leq |z|^2 d_S + 2|z| d_B + d_T.$$

Take $e = 100d_S + 20d_B + d_T$ for the atoms or $e = 1800d_S + 90d_B + d_T$ for the square. Since $F(z) \succeq A$, we obtain $F_k(z) \succeq A - eI$. Normalized integration over all outcomes gives

$$\mathcal{L} - \mathcal{L}_{\text{NH},k} \geq \text{Tr}(A - eI) = \gamma - 4e \geq 1/32.$$

The rational replay checks these outward inequalities. The same bound holds before projection and for the unrestricted infimum. Continuity gives sufficiently nearby k within this tuned weak-commutativity family, not arbitrary preparation stability.

A separate scalar depolarization path

For the scalar-average realization and $0 < \eta \leq 1$, $S_\eta = S$, $B_\eta = \eta B$, and $C_\eta = \eta C$. The active sign remains U , so $T_{0,\eta} = \eta^2 T_0$ and $\mathcal{L}_{\text{NH},\eta} = \lambda - \eta^2 \text{Tr } T_0$. The global identity

$$F_\eta(z) = \eta^2 F(z/\eta) \succeq \eta^2 A$$

gives the elementary gap at least $\eta^2 \gamma$. The Fourier effects and reports $\eta 9w/4$ give excess $\eta^2 11/16$. Weak commutativity persists since the average remains scalar; full rank persists under convex mixing with $I/4$. At $\eta = 0$, no outcome carries parameter information and the centered prior makes the zero report optimal: $\mathcal{L}_{\min} = \mathcal{L}_{\text{NH}} = \lambda$. Appendix C sharpens this to exact excess $\eta^2 \Gamma_*$. This is a statement about the scalar path only.

The complete equality-kernel locus of the scalar example

The four kernel equations in Appendix A follow from the two positive residual terms. At $z = 0$ they give only e_1 . At $z \neq 0$, a nonzero solution has $\psi_3 \neq 0$; the last two equations force $|z| = 3$, and the first three determine all components from ψ_1 . Conversely those components solve the equations on that circle. The second equation then gives the population identity. If equality held, the nonnegative dual integral would vanish, so M_z would be kernel-supported almost everywhere. Integrating $9(M_z)_{22} = 4(M_z)_{33}$ would contradict normalization.

An absolutely continuous prior

Each square is $10se_j + [-1, 1]^2$; they are disjoint and each has area four. The density $1/16$ therefore assigns each square probability $1/4$. Write $\theta = 10se_j + \eta$ on that square. The conditional mean of η is zero, with $\mathbb{E}\eta_1^2 = \mathbb{E}\eta_2^2 = 1/3$. Since the state is constant there, the zeroth and first state moments are unchanged, and $\lambda = 100 + 2/3 = 302/3$. Equation (3) consequently gives $\mathcal{L}_{\text{NH}} = 1139/12$. The same dual matrix and comparison certify the same lower bound γ on excess risk. Appendix C gives its exact value Γ_* . The extra within-square variance is not encoded in the state; it adds equally to all achievable risks and to the NH bound. Outside the prior support the state can be defined arbitrarily as a valid density matrix. No differentiability or injectivity is asserted.

A smooth identifiable family on a connected square

For (A14), symmetry gives $\mathbb{E}\theta_j = 0$ and $\mathbb{E}\theta_j\theta_k = 300\delta_{jk}$. Linearity and tracelessness of B_j therefore give exactly $\mathbb{E}\rho_\theta = I/4$, $\mathbb{E}\theta_j\rho_\theta = B_j$, and $\text{Tr}\rho_\theta = 1$. The maximum absolute off-diagonal row sums of B_1 and B_2 are both one. The triangle inequality and $|\theta_j| \leq 30$ bound each off-diagonal row sum of ρ_θ by $1/5$. Its diagonal is $1/4$, proving the stated uniform positive margin. That strict margin persists on an open neighborhood of the square; the affine state map is smooth there. Since B_1 is real with a nonzero 12 entry and B_2 is purely imaginary with a nonzero 12 entry, these two Hermitian directions are linearly independent over $\mathbb{R}$, proving injectivity. The prior second moment is 600; all other quantities in (A6) are unchanged. Hence the NH value is $2377/4$, and the same global factor yields the same excess-risk bound. The prior is uniform on its compact square, not a claimed globally smooth positive density on all of $\mathbb{R}^2$.

Strict separation in every higher Hilbert-space dimension

Each prior realization above is centered, with $\mathbb{E}_\pi\theta = 0$. For $N > 4$, choose $0 < q < 1$ and a parameter-independent full-rank density matrix τ on $\mathbb{C}^{N-4}$. Set $\tilde{\rho}_\theta = q\rho_\theta \oplus (1-q)\tau$ and $\tilde{A} = qA \oplus 0$. Then let $\tilde{S} = qS \oplus (1-q)\tau$, $\tilde{B}_j = qB_j \oplus 0$, $\tilde{C} = C \oplus 0$, and $\tilde{R} = U \oplus 0$. The latter is an active Hermitian contraction, and its zero block splits the spectator population equally between the two weights, $\tilde{K}_\pm = qK_\pm \oplus (1-q)\tau/2$. Their stationarity equation holds blockwise, hence the lifted NH minimum is $\lambda - q\text{Tr}T_0$. Direct multiplication gives $\tilde{F}(z) = qF(z) \oplus (1-q)|z|^2\tau$, and

$$\tilde{F}(z) - \tilde{A} = q[F(z) - A] \oplus (1-q)|z|^2\tau \succeq 0.$$

Its comparison trace is $q\gamma > 0$, proving the strict gap for all POVMs, including ones that couple the blocks. The posterior means are $S_j \oplus 0$, with weak commutator $q\text{Tr}S[S_1, S_2] = 0$, and every preparation remains full rank. For example, $q = 4/(N+1)$ and $\tau = I_{N-4}/(N-4)$ give average-state eigenvalues $1/(N+1)$ and $(N-3)/[(N+1)(N-4)]$, which are distinct. Thus the higher-dimensional conclusion is not confined to scalar average states.

Appendix C: An algebraic optimum and an attaining measurement

We determine the optimum for the scalar-average model by matching a measurement lower bound with an explicit measurement. The construction starts with a small rational system whose relevant root is specified exactly below. The nonscalar model is treated separately. All terminating decimals used as interval endpoints denote rational numbers.

Exact branch and certificate

Use $G(u, v), d, a, b, f$ from Eq. (A11). The two possibly nonzero off-diagonal row products of G are $u(-m_0 - b) + av$ and $a(w_0 + u) - b(t_0 + v)$; both vanish identically. Thus GG^T is diagonal and $G + G^T = -(B + B^\dagger)$ whenever $d \neq 0$. The following fixes the pair without reference to a numerical optimizer.

Lemma 1 (Algebraic branch certificate). *In the open rectangle*

$$\mathcal{R} = (-.06554650, -.06554649) \times (-.75260404, -.75260402)$$

there is exactly one solution of $\partial_u f = \partial_v f = 0$. At that solution all rational functions below are defined, and the radius and normalization predicates stated in this section hold exactly.

Here is an explicit algebraic representation and its verification. Let $H(\tau) = \sum_{j=0}^{19} h_j \tau^{19-j}$, with descending coefficients

$$(h_0, \dots, h_{19}) = (131072, -1245184, 11370496, -64700416, 322347008, \\ -1258181632, 4206531584, -11721978624, 27448818560, \\ -53716943712, 88531411546, -122973008797, 149294842212, \\ -170720691580, 204085237692, -217657066176, 156759120176, \\ -65808870640, 14355966240, -1244703600).$$

There is exactly one root in $(.26218596, .26218600)$. The certificate isolates that root more narrowly by

$$\tau_* = \text{the root of } H \text{ in } (n/10^{85}, (n+1)/10^{85}),$$

where

$$n = 262185980717760078672813858628124444009858668228893460 \\ 0183539745432024070490218894321 \quad (\text{concatenation of the two displayed integer strings}).$$

Set $u_* = -\tau_*/4$ and $v_* = N(u_*)/D(u_*)$. The complete integer coefficient arrays of the degree-18 polynomials N, D , in descending powers of u , are given by `v_numerator_u_desc` and `v_denominator_u_desc` in the ancillary file `P3349-optimal-certificate.json`. That file also gives H and the two narrow interval endpoints as exact rational strings. Together these four fields specify the branch fully.

A Sturm sequence over the rationals gives root count one in both intervals. Substitution in the rational system gives $\partial_u f = \partial_v f = 0$ by polynomial reduction modulo H . The resulting rational univariate representation (RUR) expresses the pair in terms of one algebraic root. Interval evaluation maps the root into $\mathcal{R}$, with $D(u_*) \neq 0$ and $d(u_*, v_*) \neq 0$. The Hessian satisfies $f_{uu} > 0$ and $f_{uu}f_{vv} - f_{uv}^2 > 0$ throughout the closed rectangle. The denominator d stays away from zero there, so f is smooth and strictly convex there. This proves local uniqueness of the stationary pair, not global minimization of f outside the rectangle.

The ancillary checker `P3349-optimal-checker.py` reproduces these exact predicates. Its module `optimal_exact_arithmetic.py` implements rational polynomial division, Sturm variations, interval arithmetic, and first and second derivative jets. Inverses in the quotient algebra are verified by the Euclidean algorithm and rejected if they are nonunits; no unproved irreducibility of H is needed. Rational intervals reject every denominator containing zero. Thus the branch, rather than a rounded stationary point, supplies the inputs.

The two report radii and effect directions

Equality in the measurement lower bound requires the effects to lie in the kernel of the comparison matrix. We now find these directions and their allowed reports. All quantities refer to the selected pair above. Reordering G_* by indices $(1, 3, 2, 4)$ gives $\begin{pmatrix} 0 & R \\ T & 0 \end{pmatrix}$, with

$$R = \begin{pmatrix} u_* & a_* \\ -m_0 - b_* & v_* \end{pmatrix}, \quad T = \begin{pmatrix} -w_0 - u_* & b_* \\ -a_* & -t_0 - v_* \end{pmatrix}.$$

Define

$$\sigma = 16 \operatorname{Tr}(RT), \quad \kappa = 256 \det(RT), \quad h(x) = x^2 - \sigma x + \kappa.$$

Exact interval evaluation proves $\sigma > 0$, $\kappa > 0$, $\sigma^2 - 4\kappa > 0$. Let $x_- < x_+$ be the two roots, and $r_s = \sqrt{x_s} > 0$. The block determinant identity is

$$\det(rI/2 + 2G_*) = h(r^2)/16.$$

Put $c_i = 4(G_* G_*^T)_{ii}$. For $x = r^2$ equal to either root, define

$$\begin{aligned}\nu_1(x) &= 1, \\ \nu_2(x) &= \frac{x/4 + c_1}{rw_0}, \\ \nu_3(x) &= \frac{(x/4 + c_2)(x/4 + c_1) - xw_0^2}{xw_0m_0}, \\ \nu_4(x) &= \frac{(x/4 + c_3)\nu_3(x) - m_0(x/4 + c_1)/w_0}{rt_0}.\end{aligned}$$

The radial matrix $N(r) = F(r) - A_*$ is tridiagonal, with diagonal $x/4 + c_i$ and off-diagonal entries $-r(w_0, m_0, t_0)$. The first three row equations give precisely these recurrences; the final one reduces to zero modulo $h(x)$. Explicitly, with $n_2 = x/4 + c_1$ and $n_4 = (x/4 + c_3)\nu_3 - m_0n_2/w_0$, that reduction is

$$(x/4 + c_4)n_4 - xt_0^2\nu_3 = 0 \pmod{h(x)}.$$

Consequently $N(r_s)\nu(x_s) = 0$. These are left-kernel vectors of $rI/2 + 2G_*$, not an assumed right kernel. Write $\nu_s = \nu(x_s)$.

Choosing weights so the effects sum to the identity

Since $\kappa > 0$, reduction modulo h uses $x^{-1} = (\sigma - x)/\kappa$. The squared components reduce to $\nu_i(x)^2 = a_i^{\text{sq}} + b_i^{\text{sq}}x$. For the second component,

$$a_2^{\text{sq}} = 8c_1 + \frac{16c_1^2\sigma}{\kappa}, \quad b_2^{\text{sq}} = 1 - \frac{16c_1^2}{\kappa}.$$

Interval evaluation excludes zero from b_2^{sq} . Set $y = (1 - a_2^{\text{sq}})/b_2^{\text{sq}}$. Exact reduction gives the two additional identities

$$a_3^{\text{sq}} + b_3^{\text{sq}}y = 1, \quad a_4^{\text{sq}} + b_4^{\text{sq}}y = 1.$$

For clarity, these coefficients are not the Bayesian moments B_1, B_2 . An exact rational enclosure gives $h(y) < 0$, hence $x_- < y < x_+$. The weights

$$\alpha_+ = \frac{y - x_-}{x_+ - x_-}, \quad \alpha_- = \frac{x_+ - y}{x_+ - x_-}$$

are strictly positive and satisfy $\alpha_-x_- + \alpha_+x_+ = y$. All four squared-component balances in Appendix A follow. The zero reductions for the kernel and the three normalization identities, and the strict interval predicates, are explicitly tested by the ancillary checker. They do not rely on floating-point equality.

For $i \neq j$, the phase sum in Eq. (A13) contains $\sum_{w^4=1} w^{i-j} = 0$, since $1 \leq |i - j| \leq 3$. Its diagonal sum is one by those balances. Thus the eight nonzero rank-one effects resolve the identity exactly. Their supporting vectors lie in $\ker[F(r_sw) - A_*]$, as follows by phase conjugation of the radial kernel. The NH equality kernels $\ker F(z)$ are a different locus.

The minimum risk and ensembles sharing it

Integrating Eq. (A12) with the arbitrary-outcome argument of the previous appendix gives

$$\mathcal{L} \geq \lambda - \sum_i c_i = \lambda - 4f(u_*, v_*).$$

For the eight effects the residual trace vanishes term by term, giving equality. Thus this is a global optimum over all POVMs and estimators, not only over two-seed strategies. Exact rational interval evaluation proves

$$.03592634498992 < \Gamma_* < .03592634498993, \quad 94.28592634498992 < \mathcal{L}_{\min} < 94.28592634498993$$

for the four-label ensemble. The exact risk is the algebraic expression $100 - 4f(u_*, v_*)$; the displayed decimals are only outward bounds.

Every finite-second-moment realization with the same S, B_j has $\mathcal{L}_{\min} = \lambda - 4f(u_*, v_*)$ and the same attaining protocol. This includes the connected square and the four-square density realization. For the separate scalar depolarization path, keep the effects and multiply all reports by $\eta > 0$; then

$$\mathcal{L}_{\min, \eta} = \lambda - 4\eta^2 f(u_*, v_*), \quad \mathcal{L}_{\min, \eta} - \mathcal{L}_{\text{NH}, \eta} = \eta^2 \Gamma_*.$$

At $\eta = 0$ the zero report is optimal.

For the centered spectator lift, use comparison $qA_* \oplus 0$. The attaining effects are $M_{s,w} \oplus 0$ and one further effect $0 \oplus I_{N-4}$ reported at zero. The core effects are not multiplied by the mixing weight q : state probabilities already supply that factor. They normalize to I_N and meet the full-matrix lower bound, yielding

$$\mathcal{L}_{\min, N} = \lambda - 4qf(u_*, v_*), \quad \mathcal{L}_{\min, N} - \mathcal{L}_{\text{NH}, N} = q\Gamma_*.$$

Coupled-block competing measurements obey the same comparison. This scalar calculation does not determine the nonscalar four-level optimum, which is treated separately in Appendix D. No optimizer uniqueness or outcome minimality is inferred.

Appendix D: Exact optimum of the nonscalar four-level model

We now adapt the scalar construction to the nonscalar average used in the main text. Throughout, $k = 1000001/1000000$ is fixed as in Eq. (4); no claim is made for an arbitrary average state. Write $s_1 = s_2 = p, s_3 = q, s_4 = t$, and let the three upper entries of B_k be $w_0 = p, m_0 = 2p, n_0 = 3(q + t)$. The finite and square-prior models have the same $S_k, B_k, T_{0,k}$; only λ changes. We retain the dimensionless identity loss.

A lower comparison at every estimate

The factorization must now account for the unequal diagonal entries of S_k . We choose the following matrix so that its weighted row products vanish off the diagonal. For real u, v , put

$$\begin{aligned} d_k &= s_1 s_2 v(n_0 + v) - s_3 s_4 u(w_0 + u), \\ a &= \frac{u m_0 s_1 s_4 (n_0 + v)}{d_k}, \quad b = \frac{u m_0 s_3 s_4 (w_0 + u)}{d_k}, \\ R(u, v) &= \begin{pmatrix} 0 & u & 0 & a \\ -w_0 - u & 0 & b & 0 \\ 0 & -m_0 - b & 0 & v \\ -a & 0 & -n_0 - v & 0 \end{pmatrix}. \end{aligned} \tag{D1}$$

Whenever $d_k \neq 0$, one has $R + R^T = -(B_k + B_k^\dagger)$. Only the 13 and 24 entries of $RS_k^{-1}R^T$ can be off-diagonal. Substitution gives

$$-\frac{u(m_0 + b)}{s_2} + \frac{av}{s_4} = 0, \quad \frac{a(w_0 + u)}{s_1} - \frac{b(n_0 + v)}{s_3} = 0.$$

Define $c_i = (RS_k^{-1}R^T)_{ii}$ and $A_k(u, v) = T_{0,k} - \text{diag}(c_i)$. Expanding the weighted Gram product proves Eq. (12) with $R(u, v)$ in place of $R_\diamond$, for all $z = rw$, including $z = 0$. The phase identity is unchanged because $S_k, T_{0,k}$ and the diagonal Gram matrix commute with $D(w)$. Consequently every normalized POVM and measurable estimator has risk at least $\lambda - \sum_i c_i$. The Radon–Nikodym and full-rank integrability argument above applies verbatim with S_k in place of $I/4$.

The allowed reports and effect normalization

Set $H = RS_k^{-1}$, and reorder its indices as 1, 3, 2, 4, so that $H = \begin{pmatrix} 0 & K \\ L & 0 \end{pmatrix}$. Define

$$\sigma = \text{Tr}(KL), \quad \kappa = \det(KL), \quad h(x) = x^2 - \sigma x + \kappa.$$

Block elimination, or the characteristic polynomial of H , gives

$$\det(rS_k + R) = \det(S_k)h(r^2).$$

If $\sigma > 0, \kappa > 0, \sigma^2 > 4\kappa$, the two roots $x_- < x_+$ of h are positive. Put $r_s = \sqrt{x_s}$. The real tridiagonal matrix $N(r) = F_k(r) - A_k(u, v)$ has diagonal $s_i r^2 + c_i$ and neighboring entries $-r(w_0, m_0, n_0)$. At $x = x_s$ define

$$\begin{aligned} \nu_1 &= 1, & \nu_2 &= \frac{s_1 x + c_1}{r_s w_0}, \\ \nu_3 &= \frac{(s_2 x + c_2)(s_1 x + c_1) - x w_0^2}{x w_0 m_0}, & \nu_4 &= \frac{(s_3 x + c_3)\nu_3 - m_0(s_1 x + c_1)/w_0}{r_s n_0}. \end{aligned}$$

The first three row equations fix these components. The determinant identity $\det N(r) = \det(S_k)h(r^2)^2$, together with the nonzero neighboring entries, then forces the last row equation. Equivalently, $(r_s S_k + R)^T \nu_s = 0$: these are left-null vectors of the Gram factor, not right-null vectors.

Reduce each square modulo h , writing $\nu_i(x)^2 = \eta_i + \xi_i x$ at its roots, with $x^{-1} = (\sigma - x)/\kappa$. In particular,

$$\eta_2 = \frac{2s_1 c_1 + c_1^2 \sigma / \kappa}{w_0^2}, \quad \xi_2 = \frac{s_1^2 - c_1^2 / \kappa}{w_0^2}.$$

For completeness, the other rational coefficients can be evaluated without square roots. Put

$$\begin{aligned} A_3 &= \frac{s_1 c_2 + s_2 c_1 - w_0^2 + c_1 c_2 \sigma / \kappa}{w_0 m_0}, & B_3 &= \frac{s_1 s_2 - c_1 c_2 / \kappa}{w_0 m_0}, \\ A_4 &= c_3 A_3 - s_3 B_3 \kappa - m_0 c_1 / w_0, & B_4 &= s_3 A_3 + c_3 B_3 \sigma - m_0 s_1 / w_0. \end{aligned}$$

Then

$$\begin{aligned} \eta_3 &= A_3^2 - B_3^2 \kappa, & \xi_3 &= 2A_3 B_3 + B_3^2 \sigma, \\ \eta_4 &= \frac{(A_4^2 - B_4^2 \kappa) \sigma / \kappa + 2A_4 B_4 + B_4^2 \sigma}{n_0^2}, & \xi_4 &= -\frac{A_4^2 - B_4^2 \kappa}{\kappa n_0^2}. \end{aligned}$$

Where $\xi_2 \neq 0$, let $y = (1 - \eta_2)/\xi_2$. The required branch is a root of the two rational equations

$$g_3(u, v) = \eta_3 + \xi_3 y - 1 = 0, \quad g_4(u, v) = \eta_4 + \xi_4 y - 1 = 0. \quad (\text{D2})$$

If also $h(y) < 0$, the weights

$$\alpha_+ = \frac{y - x_-}{x_+ - x_-}, \quad \alpha_- = 1 - \alpha_+$$

are strictly positive. They satisfy $\alpha_- \nu_{-,i}^2 + \alpha_+ \nu_{+,i}^2 = 1$ for every i ; the first component uses $\alpha_- + \alpha_+ = 1$, the second the definition of y , and the remaining two Eq. (D2). Thus the eight rank-one effects

$$M_{s,w} = \frac{\alpha_s}{4} D(w) \nu_s \nu_s^T D(w)^\dagger, \quad z_{s,w} = r_s w, \quad w \in \{1, i, -1, -i\},$$

are positive and sum to I_4 . Each annihilates its comparison residual. Their risk is therefore exactly $\lambda - \sum_i c_i$, matching the unrestricted lower bound. Covariance constructs one attaining measurement; it does not restrict the lower-bound quantifier.

Specifying the algebraic solution and bounding its risk

Let the rational center be the terminating decimals

$$\begin{aligned} b_u &= -0.06554647804483970189007943245440558, \\ b_v &= -0.75260434123165148908923759763752985. \end{aligned}$$

In Eq. (D1) select the unique root $(u_\diamond, v_\diamond)$ of Eq. (D2) in the closed box $\mathcal{B} = b + [-10^{-30}, 10^{-30}]^2$. This box, not the rounded approximations to a measurement, defines the branch. We now justify existence and uniqueness.

With $g = (g_3, g_4)^T$, calculate the exact rational matrix $C_0 = Jg(b)^{-1}$. Rational interval evaluation on the whole box certifies that all evaluated denominators exclude zero and that

$$\|I - C_0 Jg(\mathcal{B})\|_\infty < 2 \cdot 10^{-23} < 1, \quad b - C_0 g(b) + (I - C_0 Jg(\mathcal{B}))(\mathcal{B} - b) \subset \text{int } \mathcal{B}.$$

The norm bound is the maximum absolute row-sum enclosure. The map $T(x) = x - C_0 g(x)$ is consequently a contraction of the closed box into itself. The mean-value formula proves the displayed image inclusion for every point, so the contraction theorem gives exactly one fixed point. Since C_0 is invertible, that point solves $g = 0$. The same bound makes Jg nonsingular on the box. Clearing its nonzero rational-function denominators produces polynomial equations over $\mathbb{Q}$ with a nonsingular isolated solution. Thus the selected coordinates, radii, weights and risk are algebraic. No assertion of minimal polynomial degree is needed.

The whole-box certificate also verifies $\sigma > 0, \kappa > 0, \sigma^2 - 4\kappa > 0, \xi_2 \neq 0, h(y) < 0$, and the strict risk bounds. With $R_\diamond = R(u_\diamond, v_\diamond)$, they give

$$\begin{aligned} 94.28592480080068802537609221 &< \mathcal{L}_{\min, k} < 94.28592480080068802537609222, \\ 0.03592633552227579562708351 &< \Gamma_k < 0.03592633552227579562708352. \end{aligned}$$

The first line uses $\lambda = 100$. For the square prior, $\lambda = 600$ and the minimum risk is larger by 500; Γ_k is unchanged. In both cases the exact definition is $\mathcal{L}_{\min, k} = \lambda - \sum_i c_i$, with the branch just specified.

The ancillary certificate `nonscalar-root-certificate.json` contains only k , the rational center and box radius. `weighted_exact_checker.py` uses exact fractions, forward differentiation and outward-rounded 180-bit dyadic intervals. `independent_symbolic_checker.py` instead uses the explicit component formulas above, symbolic differentiation and outward-rounded decimal intervals. Each independently verifies contraction, strict self-inclusion, positive radii and mixture weights, and the risk enclosure. Neither treats a numerical residual as proof or imports the numerical proposal generator. A wrong model parameter is rejected by both.

Appendix E: All optimal reported measurements and the seven-outcome requirement

Fix the nonscalar model $k = 1000001/1000000$ of Eq. (4), with the algebraic branch selected in Appendix D. All quantities in this section refer to that branch. The result applies to both its four-atom and square priors. They have the same $S_k, B_k, A_\diamond$; changing λ only adds a constant to every risk. No scalar-model outcome-minimality assertion is made.

TABLE I. Precision benchmarks for the four-state prior in Eq. (5), with $\lambda = 100$ and dimensionless squared error. The PVM excess is relative to the general optimum and includes parameter-independent mixtures of fixed projective settings on the original $\mathbb{C}^4$. Projective implementations of POVMs on larger spaces are not restricted.

Benchmark	Established result
NH relaxation	Strictly below achievable risk, gap ≈ 0.036
General POVM	Exact minimum; seven outcomes are minimal
Original-space PVM	Exact minimum, four outcomes; excess ≈ 0.270 , analytically $> 25/192$

Characterizing all optimal reported measurements

The lower bound determines where an optimal measurement can report and which effect directions it can use. Normalization then determines the allowed weights. We make these requirements explicit before counting the outcomes. For $s \in \{-, +\}$, retain the positive radii r_s , real left-null vectors ν_s normalized by $\nu_{s,1} = 1$, and positive masses α_s . Write $\mathbb{T} = \{w \in \mathbb{C} : |w| = 1\}$. The radial comparison residual $N_s = F_k(r_s) - A_\diamond$ is positive semidefinite and singular. Its nonzero neighboring entries make the tridiagonal recurrence determine every null-vector component from the first; hence its kernel is one-dimensional. Moreover, $tI - N_s$ is nonnegative and irreducible for sufficiently large t . Since $N_s \succeq 0$, its null vector is the Perron vector of $tI - N_s$. Thus every component of ν_s is strictly positive.

Define the two coefficient determinants

$$\begin{aligned} \Delta_1 &= \nu_{-,2}\nu_{+,2}(\nu_{+,3} - \nu_{-,3}), \\ \Delta_2 &= \nu_{-,3}\nu_{+,2}\nu_{+,4} - \nu_{+,3}\nu_{-,2}\nu_{-,4}. \end{aligned} \tag{E1}$$

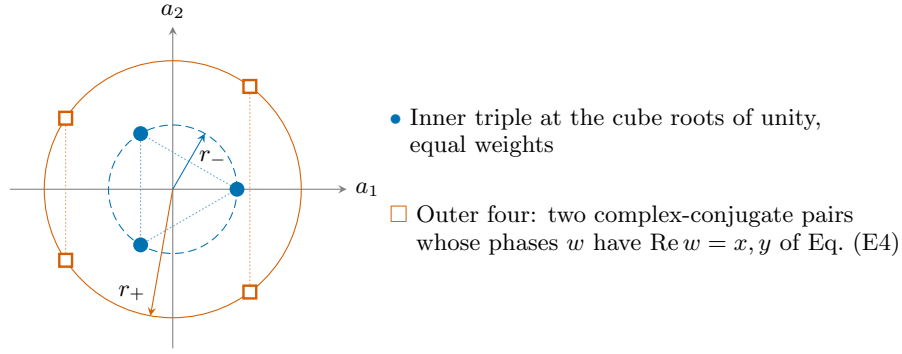


FIG. 2. Reports of a seven-outcome optimal measurement in the plane $z = a_1 + ia_2$. Filled circles: equilateral triple on the inner circle $|z| = r_-$, at the cube roots of unity with equal weights. Open squares: four reports on the outer circle $|z| = r_+$, forming two complex-conjugate pairs (dotted chords) whose cosines x, y obey $xy = -1/2$ and $x + y = \rho/2$ [Eq. (E4)]. The radii and the value of ρ are drawn schematically, and marker sizes do not represent probabilities. The effects are $D(w)\nu_s\nu_s^T D(w)^\dagger$ times the positive weights fixed by Eq. (E3); seven arbitrary points need not satisfy those equations. The plot shows report coordinates, not Hilbert-space effect directions. Generative AI assisted the figure design and TikZ source.

Exact interval evaluation at the selected branch gives

$$\nu_{+,2}^2 \neq \nu_{-,2}^2, \quad \Delta_1 > 0, \quad \Delta_2 > 0, \quad A_+ > A_- > 0, \quad A_s = \alpha_s \nu_{s,4}.$$

These are branch-specific facts, not assumed genericity.

Theorem 2 (Optimal reported-measurement characterization). *A POVM on the estimate plane attains $\mathcal{L}_{\min,k}$ if and only if it is concentrated on the two circles $z = r_s w$ and there has the form*

$$M_s(dw) = D(w)\nu_s\nu_s^T D(w)^\dagger \mu_s(dw), \quad w \in \mathbb{T}, \quad (\text{E2})$$

where μ_s are positive measures satisfying

$$\begin{aligned} \mu_s(\mathbb{T}) &= \alpha_s, & m_{1,s} &= m_{2,s} = 0, \\ \nu_{-,4}m_{3,-} + \nu_{+,4}m_{3,+} &= 0, & m_{j,s} &= \int_{\mathbb{T}} w^j \mu_s(dw). \end{aligned} \quad (\text{E3})$$

This includes continuous and singular measures.

Proof. The full-plane determinant is $\det(F_k(z) - A_\diamond) = \det(S_k)h(|z|^2)^2$. It vanishes only on the two circles. The arbitrary-outcome risk identity gives, for every finite-risk reported POVM,

$$\mathcal{L} - \mathcal{L}_{\min,k} = \int_{\mathbb{C}} \text{Tr}[(F_k(z) - A_\diamond)M(dz)].$$

Its Radon–Nikodym integrand is nonnegative. At equality it vanishes almost everywhere. For positive matrices Q, E , $\text{Tr}(QE) = \|Q^{1/2}E^{1/2}\|_{\text{HS}}^2 = 0$ forces $\text{ran } E \subset \ker Q$. Thus the measure is zero off the two circles, and on each circle its density has the rank-one form (E2). Its 11 entry defines μ_s , since $\nu_{s,1} = 1$. This argument constrains the pushed-forward reported measurement; it does not uniquely label raw refinements.

Let $q_s = \mu_s(\mathbb{T})$. The 11 and 22 entries of $\int M = I_4$ are

$$q_- + q_+ = 1, \quad q_- \nu_{-,2}^2 + q_+ \nu_{+,2}^2 = 1.$$

Their unique solution is $q_s = \alpha_s$. The other populations then follow from the previously proved normalization equations. For Fourier order one, entries 12 and 23 give a homogeneous two-by-two system with determinant Δ_1 ; for order two, entries 13 and 24 give determinant Δ_2 . Consequently both first and second moments vanish separately. Entry 14 is the third-moment equation in (E3), up to complex conjugation. The unused entry 34 vanishes with the first moments. Conversely, these conditions give every diagonal and off-diagonal entry of I_4 . Each effect also annihilates its comparison residual, so it attains the global minimum. $\square$

Why six outcomes cannot attain the minimum

Normalize the circle measures as $\sigma_s = \mu_s/\alpha_s$ and put $f_{j,s} = m_{j,s}/\alpha_s$. Their first two moments vanish, so the Gram matrix of $1, w, w^2$ in $L^2(\sigma_s)$ is I_3 . For n_s nonzero atoms its rank is at most n_s ; therefore each circle in a finite-outcome optimizer requires at least three atoms. If $n_s = 3$, the weighted square Vandermonde matrix V obeys $VV^\dagger = I_3$, and is unitary. Column norms force all three weights to be $1/3$. Orthogonality of distinct columns gives $1 + \tau + \tau^2 = 0$ for their phase ratios. The support is therefore a rotated equilateral triple, and $|f_{3,s}| = 1$.

An optimizer with six or fewer nonzero reports would consequently have exactly three on each circle. The remaining equation $A_- f_{3,-} + A_+ f_{3,+} = 0$ would require $A_- = A_+$, contrary to the strict branch-specific inequality. This excludes arbitrary finite report locations, not just covariant or equally weighted ansatzes.

The same lower bound on outcome number applies to finite raw POVMs with randomized reports. For any nonzero effect E , its probability $\text{Tr}(S_k E)$ is positive because $S_k > 0$. Conditional Jensen reduces randomized squared-error reporting to its mean; the risk difference is $\text{Tr}(S_k E)$ times the report variance. At a finite global minimum this variance must be zero at every nonzero effect, or the risk could be strictly improved. Thus optimal raw reports are deterministic almost surely. Zero effects may be removed, and duplicate labels do not reduce the required number of distinct reports.

An algebraic three-plus-four measurement

Put $\rho = -A_-/A_+ \in (-1, 0)$. On the smaller circle use the three cube roots of unity, each with scalar weight $\alpha_-/3$. They give $(f_{1,-}, f_{2,-}, f_{3,-}) = (0, 0, 1)$. For the larger circle set

$$x = \frac{\rho + \sqrt{\rho^2 + 8}}{4}, \quad y = \frac{\rho - \sqrt{\rho^2 + 8}}{4}. \quad (\text{E4})$$

Then $x + y = \rho/2$, $xy = -1/2$, $0 < x < 1$, and $-1 < y < 0$. Choose its four distinct unit phases

$$x \pm i\sqrt{1-x^2}, \quad y \pm i\sqrt{1-y^2}.$$

Assign each phase in the first pair scalar weight $\alpha_+ q_x/2$, and each in the second pair weight $\alpha_+ q_y/2$, where

$$q_x = -\frac{y}{x-y} > 0, \quad q_y = \frac{x}{x-y} > 0, \quad q_x + q_y = 1.$$

Conjugate pairing makes the moments real. The identities $xy = -1/2$ and $x + y = \rho/2$ give

$$\begin{aligned} q_x x + q_y y &= 0, \\ q_x(2x^2 - 1) + q_y(2y^2 - 1) &= 0, \\ q_x(4x^3 - 3x) + q_y(4y^3 - 3y) &= \rho. \end{aligned}$$

Hence the third moments also cancel, $A_- + A_+ \rho = 0$. For each selected phase, take its scalar weight $\beta_{s,w}$ and use

$$E_{s,w} = \beta_{s,w} D(w) \nu_s \nu_s^T D(w)^\dagger, \quad z_{s,w} = r_s w.$$

These seven nonzero rank-one effects sum exactly to I_4 and attain $\lambda - \sum_i c_{i,k}$. Combined with the lower bound, the minimum number of nonzero outcomes of an optimal finite POVM is exactly seven. The eight-effect construction remains optimal but is not minimal. The seven-atom selection is not a $U(1)$ -covariant measurement; the phase symmetry of the risk moments has not restricted the optimization domain.

A strict penalty for all at-most-six-outcome strategies

Let K be the compact convex hull of the chosen prior support, either the finite-prior diamond or the square. Projecting any report onto K cannot increase its squared distance from any parameter in that support. Together with conditional Jensen this reduces all six-outcome strategies to deterministic reports in K . Include zero effects and define

$$\mathcal{L}_6 = \min_{\substack{E_i \succeq 0, \sum_{i=1}^6 E_i = I_4 \\ z_i \in K, i=1, \dots, 6}} \mathcal{L}(E, z).$$

The feasible set is compact: $0 \preceq E_i \preceq I_4$, the normalization constraint is closed, and K is compact. The risk is a continuous polynomial in these finite-dimensional variables. A minimizer therefore exists. Equality $\mathcal{L}_6 = \mathcal{L}_{\min,k}$ would contradict seven-outcome minimality. Consequently

$$\delta_6 = \mathcal{L}_6 - \mathcal{L}_{\min,k} > 0. \quad (\text{E5})$$

This holds separately for each of the two priors; no numerical value or explicit lower bound for δ_6 is asserted.

Every projective measurement on the original $\mathbb{C}^4$ has at most four nonzero projections and obeys this same strict penalty. Parameter-independent classical randomization over fixed, single-copy settings with at most six outcomes cannot bypass it: each conditional setting has risk at least $\mathcal{L}_{\min,k} + \delta_6$, so their average does too. Discarding the setting flag cannot improve the optimal squared-error risk. All outcomes are normalized and retained, with no postselection. The statement is not an outcome bound per step of an adaptive instrument, a restriction on an auxiliary-space projective realization of a general POVM, or a hardware/ancilla minimum.

Projective strategies for finite-data Bayesian estimation are studied by Law and Bernád [29]; their construction is restricted to PVMs and does not claim unrestricted global optimality. Seven-outcome four-level measurements also occur in point tomography [30], where the target is local pure-state reconstruction under infidelity. Here the result instead concerns a full-rank, two-coordinate, single-copy prior-integrated quadratic-risk problem. Neither the number seven alone nor the use of moment ranks is a new general quantum-estimation principle.

Exact predicates and reproducibility

`seven_predicate_checker.py` verifies the selected root and the strict component, population-rank, harmonic-rank, amplitude and phase-weight predicates using weighted block formulas and outward dyadic intervals. `independent_seven_checker.py` uses explicit component formulas, decimal intervals, a fresh independent root replay, and seven exact symbolic identities for the conjugate pairs and cube roots. Both reject a changed model parameter. Their rational square-root enclosures do not use floating-point signs. These finite checks support the branch-specific hypotheses above; the measure-theoretic characterization, rank equality case and compactness argument are supplied analytically in this section.

Appendix F: A quantitative penalty for projective readout

This appendix gives an elementary explanation of the projective penalty and an earlier feasible benchmark. Appendix G improves the lower bound; Appendix H proves the exact optimum with a different mixing ratio.

We retain the fixed nonscalar model, its two priors and the identity quadratic loss. This section concerns projective measurements on the original $\mathbb{C}^4$, not a projective realization of a general POVM on a larger space. Bayesian comparisons between general and restricted projective readout are established [1]. Normal completions [24, 25] and quantitative joint-diagonalization bounds [31] provide related matrix methods. The cost below is a degenerate weighted quadratic form, rather than the full Frobenius off-diagonal energy of a Hermitian pair.

Proposition 1 (Certified projective-risk bracket). *Let $\mathcal{L}_{\text{PVM},k}$ be the infimum over all PVMs on the original four-level system and all reporting rules. For either fixed prior,*

$$\mathcal{L}_{\text{NH},k} + \frac{112p}{165} \leq \mathcal{L}_{\text{PVM},k} \leq \lambda - \frac{\mathcal{A}_k^2}{\tau_k}, \quad (\text{F1})$$

where

$$\tau_k = p + q + \frac{p + 36t}{37}, \quad \mathcal{A}_k = 2p + \frac{p + 18(q + t)}{\sqrt{37}}.$$

In particular,

$$\mathcal{L}_{\text{PVM},k} > \mathcal{L}_{\min,k} + \frac{25}{192}. \quad (\text{F2})$$

The lower bounds include parameter-independent classical mixtures of single-copy original-space PVM settings. The upper endpoint is the exact risk of the explicit four-result PVM given below, not the exact restricted optimum. Neither endpoint is asserted sharp.

How projectivity constrains the risk

Projectivity forces the matrix of complex reports to be normal. We first express the risk as a weighted quadratic cost, then show why normality prevents that cost from vanishing. For deterministic complex reports z_i at orthogonal projections P_i , put $Z = \sum_i z_i P_i$. This matrix is normal, including for degenerate projections or coincident reports. Write

$$K_+ = \text{diag}(0, 0, 0, t), \quad K_- = \text{diag}(p, p, q, 0),$$

and retain the ladder C , $B_k = CK_+ + K_-C$, and $T_{0,k}$. For $E = Z - C$ define

$$q_U(E) = \text{Tr}(EK_+E^\dagger + E^\dagger K_-E).$$

Expanding its cross terms gives the unrestricted identity

$$\begin{aligned} \text{Tr}(S_k Z^\dagger Z) - 2\Re \text{Tr}(B_k^\dagger Z) + \text{Tr} T_{0,k} - q_U(Z - C) \\ = \text{Tr}[K_-(Z^\dagger Z - ZZ^\dagger)]. \end{aligned}$$

Normality makes the right side zero. The PVM risk is $\lambda + \text{Tr}(S_k Z^\dagger Z) - 2\Re \text{Tr}(B_k^\dagger Z)$; hence

$$\mathcal{L}(P, z) - \mathcal{L}_{\text{NH},k} = q_U(Z - C). \quad (\text{F3})$$

No first-moment constraint on the measurement is imposed.

The general optimization framework is already the normal Procrustes problem [26]: set $X_{\text{in}} = S_k^{1/2}$, $Y_{\text{in}} = B_k S_k^{-1/2}$ and minimize $\|ZX_{\text{in}} - Y_{\text{in}}\|_F^2$ over normal Z . For normal matrices this squared norm equals

$$\mathcal{L}(P, z) - \lambda + \text{Tr}(B_k S_k^{-1} B_k^\dagger).$$

Eliminating reports in an orthonormal basis v_i gives the established gain $\sum_i |v_i^\dagger B_k v_i|^2 / (v_i^\dagger S_k v_i)$ [26]. For scalar S , the related nearest-normal Frobenius formulation is also established [32]. Full rank bounds Z , and thus its eigenvalue reports, on bounded-cost sublevel sets; it does not assume arbitrary reporting rules are bounded. These frameworks do not by themselves certify a global minimum. The result of this appendix is an elementary lower bound and a feasible benchmark.

Use the block notation

$$Z = \begin{pmatrix} A & b \\ c^\dagger & d \end{pmatrix}, \quad A_0 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix}, \quad b_0 = (0, 0, 3)^T,$$

and set $D = A - A_0$, $f = b - b_0$, and $Q = q_U(Z - C)$. The top-left block of normality yields

$$H(A, b) := AA^\dagger - A^\dagger A + bb^\dagger = cc^\dagger, \quad H_0 = H(A_0, b_0) = \text{diag}(1, 3, 5).$$

Here $p < q, t$. The weighted entry expansion gives

$$Q = \sum_{i=1}^3 s_i \|(Z - C)_{i,\cdot}\|^2 + t \|(Z - C)_{\cdot,4}\|^2 \geq p \|D\|_F^2 + 2p \|f\|^2.$$

The unpenalized c -row is not bounded by this degenerate cost.

A direction-dependent rank obstruction

Choose a rank-two orthogonal projection P with $Pc = 0$. If $c = 0$, any such projection works. Write $P = I_3 - vv^\dagger$, $\|v\| = 1$, and $u_i = |v_i|^2$, so $\sum_i u_i = 1$. Then $\text{Tr}(PH) = 0$ and

$$\text{Tr}(PH_0) = 4 + 4u_1 + 2u_2.$$

Retaining this direction dependence is useful. In the same direction,

$$\begin{aligned}\| [P, A_0] \|_{\mathbb{F}}^2 &= v^\dagger (A_0 A_0^\dagger + A_0^\dagger A_0) v - 2|v^\dagger A_0 v|^2 \\ &\leq u_1 + 5u_2 + 4u_3, \quad \|Pb_0\|^2 = 9(1 - u_3).\end{aligned}$$

The linear part of $\text{Tr}[P(H - H_0)]$ is

$$2\Re \text{Tr}(D^\dagger [P, A_0]) + 2\Re(f^\dagger Pb_0).$$

Weighted Cauchy–Schwarz bounds it below by

$$-2\sqrt{\mathcal{X}(v)\vartheta}, \quad \vartheta = Q/p, \quad \mathcal{X}(v) = 4 + \frac{3}{2}u_1 + \frac{11}{2}u_2.$$

The quadratic part satisfies

$$\text{Tr}[P(DD^\dagger - D^\dagger D + ff^\dagger)] \geq -\|D\|_{\mathbb{F}}^2 \geq -\vartheta.$$

Consequently

$$0 \geq 4 + 4u_1 + 2u_2 - 2\sqrt{\mathcal{X}(v)\vartheta} - \vartheta. \quad (\text{F4})$$

Young’s inequality with $\epsilon = 4/11$ gives $2\sqrt{\mathcal{X}\vartheta} \leq (4/11)\mathcal{X} + (11/4)\vartheta$. Thus

$$0 \geq \frac{28}{11} + \frac{38}{11}u_1 - \frac{15}{4}\vartheta, \quad Q \geq \frac{112p}{165}.$$

This uniform nonstrict lower bound also bounds the infimum. The exact model gives $p > 55/224$, hence $112p/165 > 1/6$. Appendix D gives $\Gamma_k < 7/192$. Subtracting this excess above NH proves (F2) because $1/6 - 7/192 = 25/192$. In particular, this refines the simpler $3/64$ separation without solving the restricted optimization.

For every nonzero PVM effect, $\text{Tr}(S_k P_i) > 0$. Conditional Jensen replaces a randomized report by its mean without increasing risk; infinite-risk procedures obey the bound trivially. Parameter-independent choice of a fixed single-copy PVM setting preserves the uniform barrier on averaging. Forgetting that setting cannot improve Bayes risk. Auxiliary-space projections, a per-step outcome restriction on an adaptive instrument, and postselection are not the comparison class. This penalty does not quantify the larger class of all at-most-six-outcome general POVMs in Appendix E.

An exact four-result projective benchmark

Let $\omega = e^{2\pi i/3}$ and choose the orthonormal basis

$$\xi = \frac{e_1 + 6e_4}{\sqrt{37}}, \quad \eta = \frac{6e_1 - e_4}{\sqrt{37}}, \quad \psi_j = \frac{\xi + \omega^j e_2 + \omega^{2j} e_3}{\sqrt{3}} \quad (j = 0, 1, 2),$$

consisting of the three ψ_j and η . Their rank-one projectors form a normalized four-result PVM. The predictive probabilities and complex first moments are

$$\psi_j^\dagger S_k \psi_j = \tau_k/3, \quad \psi_j^\dagger B_k \psi_j = \mathcal{A}_k \omega^j / 3.$$

The fourth probability is $(36p + t)/37 > 0$, and its first moment is zero. Use posterior-mean reports $\mathcal{A}_k \omega^j / \tau_k$ and 0, respectively. The fourth outcome is retained, not postselected away. The attained risk is exactly

$$\mathcal{L}_{\text{trial},k} = \lambda - \mathcal{A}_k^2 / \tau_k \in \mathbb{Q}(\sqrt{37}). \quad (\text{F5})$$

Measurement entries also contain the cube root of unity; the risk’s quadratic field is not a restriction to real measurement entries. The convenient mixing ratio 1 : 6 is a strictly suboptimal benchmark; the exact optimum is given in Appendix H.

For $\lambda = 100$, the exact lower endpoint is

$$\mathcal{L}_{\text{NH},k} + \frac{112p}{165} = \frac{74780508814454731109009}{792001166000594000099} > 94.419695354840734919600705,$$

and two independent rational enclosures give

$$94.555744248543274360535112 < \mathcal{L}_{\text{trial},k} < 94.555744248543274360535113.$$

Thus $94.41969535 < \mathcal{L}_{\text{PVM},k} < 94.55574425$. The four Fourier vectors $(1, w, w^2, w^3)^T/2$, $w \in \{1, i, -1, -i\}$, with posterior reports $3(1-p)w$, instead have risk $\lambda - 9(1-p)^2 > \mathcal{L}_{\text{trial},k}$. The same projectors and reports apply to both priors; changing λ adds 500 to their total risks. This is a certified bracket, not an exact PVM optimum.

Appendix G sharpens this lower endpoint using a separate exact moment certificate. The elementary barrier and explicit upper protocol in this section remain unchanged. Appendix H matches a stronger global bound with an attaining measurement.

`pvm_bracket_checker.py` implements exact arithmetic in $\mathbb{Q}(\sqrt{37}, \omega)$, constructs all four projectors and reports, and checks the risk, Young coefficients, model floor and outward enclosures. `independent_pvm_bracket_checker.py` independently constructs the orthonormal basis with symbolic radicals and isolates $\sqrt{37}$ by rational polynomial intervals, using a fresh independent root replay. The earlier `pvm_penalty_checker.py` and its independently implemented formal-polynomial counterpart also retain checks of the normal-risk and perturbation identities. The universal projection, Cauchy, Young and reporting arguments above are analytical. No local nonconvex numerical minimum is used as proof.

Appendix G: A weighted global projective bound from projector moments

For the same fixed nonscalar model and finite prior, we obtain

$$94.5557442477 < \mathcal{L}_{\text{PVM},k} < 94.555744248544. \quad (\text{G1})$$

The lower endpoint is an exact computer-assisted certificate for every complex orthonormal basis. The upper endpoint is an outward enclosure of the explicit four-result protocol in Appendix F. The interval is narrow, but neither its endpoints nor numerical agreement establish an exact optimal projective measurement. Appendix H supplies the separate matching lower bound and attaining basis.

Necessary moments for every projective measurement

Refining orthogonal projections into a rank-one basis and optimizing reports cannot increase squared-error risk. It therefore suffices to lower-bound every orthonormal basis v_i of $\mathbb{C}^4$. With $P_i = v_i v_i^\dagger$, set

$$R = \sum_i P_i \otimes P_i, \quad Q = R^{\top_2}, \quad H = (B_k \otimes B_k^\dagger + B_k^\dagger \otimes B_k)/2.$$

Here $\top_2$ denotes partial transpose on the second tensor factor, in the fixed preparation basis. Both R and $Q = \sum_i P_i \otimes \bar{P}_i$ are rank-four orthogonal projections; R is supported on the symmetric subspace and $\text{Tr}_2 R = I_4$. In particular,

$$R \succeq 0, \quad Q \succeq 0, \quad R^2 = R, \quad Q^2 = Q. \quad (\text{G2})$$

These are necessary conditions, not a convex characterization of bases.

Put $s_i = v_i^\dagger S_k v_i$ and $b_i = v_i^\dagger B_k v_i$. The s_i are the predictive probabilities of the outcomes. Optimizing the reported estimates gives

$$\mathcal{L}(P) = \lambda - \sum_i \frac{|b_i|^2}{s_i}.$$

Since $pI \preceq S_k \preceq qI$, every s_i lies in $[p, q]$. Replacing all denominators by p loses their dependence on the measurement direction. Instead use the reciprocal secant

$$\frac{1}{s} \leq \frac{p+q-s}{pq}, \quad \frac{p+q-s}{pq} - \frac{1}{s} = \frac{(s-p)(q-s)}{pq s} \geq 0. \quad (\text{G3})$$

Define $G_2 = \sum_i |b_i|^2 = \text{Tr}(HR)$ and $G_3 = \sum_i |b_i|^2 s_i$. Then

$$\mathcal{L}(P) \geq \lambda - \frac{(p+q)G_2 - G_3}{pq}. \quad (\text{G4})$$

The additional term retains the state-dependent weight.

Use the normalized symmetric basis

$$|aa\rangle, \quad (|ab\rangle + |ba\rangle)/\sqrt{2} \quad (0 \leq a < b \leq 3),$$

in lexicographic order of (a, b) , and write R for the ten-by-ten matrix of R in this basis. The basis vector indexed by (a, b) has phase charge $a + b$ under $D(w) \otimes D(w)$. Average polynomial moments over this phase action and complex conjugation. The real ladder makes the weighted objective invariant. This averages moments of genuine bases, not bases themselves: neither $(\mathbb{E}Q)^2 = \mathbb{E}Q$ nor a symmetric optimizing basis is assumed.

Take all monomials of degree at most two in R_{AB} with total charge zero. Identify commuting factors and conjugate monomials, and let x be their real averaged values, including $x_\emptyset = 1$. There are 349 such coordinates. Because R is a norm-one projection, $|R_{AB}| \leq 1$, hence

$$|x_j| \leq 1. \tag{G5}$$

For each nonnegative charge, the moment matrix of the coordinates 1, R_{AB} of that charge is positive semidefinite. The constant occurs only in the charge-zero sector. Negative-charge blocks are conjugate copies. Together with the charge blocks of $\mathbb{E}R$ and $\mathbb{E}Q$, these give 21 positive-semidefinite matrices. Their sizes are, in order,

$$(17, 14, 12, 8, 5, 2, 1), \quad (1, 1, 2, 2, 2, 1, 1), \quad (1, 2, 3, 4, 3, 2, 1).$$

The first family consists of averaged Gram matrices; the other two are averages of positive matrices. Complex bases are not excluded.

The affine equations initially comprise the constant normalization, all entries of $\text{Tr}_2 R = I$ multiplied by monomials of degree zero or one, and all entries of $R^2 - R = 0$ and $Q^2 - Q = 0$, followed by the stated averaging. Zero rows are omitted, but dependent rows are retained. These give 277 rows. The reshape $Q_{ab,cd} = R_{ad,cb}$ fixes the coefficients.

The weighted denominator as a three-factor contraction

Orthogonality gives $P_i P_j = \delta_{ij} P_i$, and therefore

$$T_3 = R_{12} R_{23} = \sum_i P_i^{\otimes 3}, \quad G_3 = \text{Tr}[(B_k \otimes B_k^\dagger \otimes S_k) T_3]. \tag{G6}$$

Here $R_{12} = R \otimes I$ and $R_{23} = I \otimes R$. This is algebraic bookkeeping for a single-copy basis, not a three-copy measurement protocol. In the ordinary tensor-product basis,

$$(T_3)_{ijk,abc} = \sum_{f=0}^3 R_{ij,af} R_{fk,bc}. \tag{G7}$$

Thus G_3 is quadratic in the entries of R and uses the same 349 coordinates.

The physical T_3 is positive and has symmetric support. Impose its row and column permutation identities, by equating each entry to the entry with both triples sorted, and its Hermitian identities, before moment averaging. Include all resulting nonzero rows. Together with the preceding equations these give $Ax = b$ with 731 rows. From each charge sector of the 20 sorted triples $0 \leq i \leq j \leq k \leq 3$, also retain the corresponding principal submatrix of $\mathbb{E}T_3$. These are ordinary tensor-coordinate principal submatrices, not matrices in a normalized three-particle symmetric basis. Their sizes are

$$(1, 1, 2, 3, 3, 3, 3, 2, 1, 1).$$

Symmetrizing their entries gives ten real positive-semidefinite blocks after conjugation averaging. In total there are 31 blocks $M_\ell(x) = \sum_j C_{\ell j} x_j \geq 0$. All coefficients lie in $\mathbb{Q}(\sqrt{2})$. These necessary moment conditions use established finite polynomial-optimization machinery [33]; no completeness or flat-extraction assertion is used.

Bounding the coefficient residual in the dual identity

Set $c^T x = -[(p+q)G_2 - G_3]/(pq)$. The ancillary file `pvm-weighted-certificate.json` specifies $A, b, c, C_{\ell j}$, a rational vector y , and 31 rational symmetric matrices Z_ℓ . Exact positive LDL pivots prove $Z_\ell \succ 0$, with 105 pivots in total. Define

$$r_j = c_j + (A^T y)_j - \sum_{\ell} \text{Tr}(C_{\ell j} Z_\ell).$$

For every physical moment vector,

$$\begin{aligned} c^T x &= \sum_{\ell} \text{Tr}[Z_\ell M_\ell(x)] - b^T y + r^T x \\ &\geq -b^T y - \sum_j |r_j|. \end{aligned} \tag{G8}$$

The last step uses Eq. (G5). No numerical solver tolerance is discarded. A rational interval $[\ell, u]$, checked by $0 < \ell < u$ and $\ell^2 < 2 < u^2$, encloses $\sqrt{2}$. Outward evaluation of every $r_j \in \mathbb{Q}(\sqrt{2})$ gives an exact rational upper bound $\varepsilon \geq \sum_j |r_j|$. For the supplied certificate, $\varepsilon < 8 \cdot 10^{-10}$, and

$$b^T y + \varepsilon < 100 - 94.5557442477.$$

Equations (G4) and (G8) prove the strict lower endpoint of Eq. (G1). The objective already includes the factor $1/(pq)$; no further denominator division is taken in this last inequality.

The companion `pvm_weighted_checker.py` verifies the fixed certificate's exact PSD and residual arithmetic. The independently implemented `independent_pvm_weighted_checker.py` reconstructs every monomial, constraint, matrix block and weighted objective coefficient from these definitions, verifies the certificate by rational Schur elimination, and rejects corrupted model, coefficient and positivity data. Neither implementation needs a numerical optimization package. The earlier, coarser $1/p$ certificate remains reproducible with `pvm-moment-certificate.json` and its two companion checkers.

The secant does not make the denominator exact. Its total gain excess obeys the uniform bound

$$0 \leq \frac{(p+q)G_2 - G_3}{pq} - \sum_i \frac{|b_i|^2}{s_i} \leq \frac{(q-p)^2}{4p^2q} \|B_k\|_F^2 < 3 \cdot 10^{-12}.$$

Indeed, Eq. (G3) bounds each defect, and $\sum_i |b_i|^2 \leq \|B_k\|_F^2$ by orthogonal projection onto the basis-diagonal subspace. This error estimate does not imply equality: the three signal outcomes of the explicit protocol have $p < s_i < q$ and $b_i \neq 0$, so their secant defects are strictly positive. Thus this secant certificate cannot itself be a matching proof. Appendix H instead uses a reciprocal majorant with exact contact at the attaining basis and a stronger positive moment identity.

The bounds extend to coarse projectors, randomized reports and parameter-independent classical mixtures by the refinement and Jensen arguments in Appendix F. The square prior adds 500 to both endpoints and changes no measurement or risk difference. No auxiliary-space projectivity restriction follows.

Appendix H: The exact projective optimum

We now match a lower bound over every complex orthonormal basis with an explicit four-result measurement. Throughout this appendix, $k = 1000001/1000000$, the preparation space is the original $\mathbb{C}^4$, and the prior and loss are those of Eq. (5). The tensor operators below encode one readout basis; they do not supply additional encoded copies.

An attaining basis and its exact risk

Use $\beta_1, \beta_2, \beta_3, D_0, D_2, \delta$ as defined in the main text. The equation

$$(\beta_1 D_0 - D_2 \beta_3 h)^2 (1 + h^2) - \beta_2^2 \delta^2 h^2 = 0, \quad h > \frac{\beta_1 D_0}{D_2 \beta_3}, \tag{H1}$$

selects a unique positive root h_* . Indeed, on this interval the equation is equivalent to $(D_2\beta_3 - \beta_1 D_0/h)\sqrt{1+h^2} = \beta_2\delta$. Its left side increases strictly from zero to infinity, whereas the right side is positive. The ancillary `candidate.json` also specifies a rational isolating interval for this root.

Set $x_* = h_*/\sqrt{1+h_*^2}$ and $y_* = 1/\sqrt{1+h_*^2}$. The four vectors displayed in the main text are orthonormal: $x_*^2 + y_*^2 = 1$ and $1 + \omega + \omega^2 = 0$ give their norms and inner products directly. For $P_j = v_j v_j^\dagger$, $j = 0, 1, 2$,

$$\text{Tr}(S_k P_j) = \frac{\tau_*}{3}, \quad \text{Tr}(B_k P_j) = \frac{A_* \omega^j}{3},$$

whereas $\text{Tr}(B_k P_3) = 0$. Completing the square separately at each outcome gives the optimal reports stated above and the attained risk

$$\mathcal{L}_* = 100 - g_*, \quad g_* = A_*^2/\tau_*. \quad (\text{H2})$$

No stationarity or restricted-family argument is used as a global lower bound in this step.

All quantities needed by the bound lie in the real algebraic field $K = \mathbb{Q}[h_*]$. In particular,

$$y_* = \frac{D_2\beta_3 h_* - \beta_1 D_0}{\beta_2 \delta h_*}, \quad x_* = h_* y_*.$$

The selected branch makes $y_* > 0$, and Eq. (H1) gives $x_*^2 + y_*^2 = 1$. Thus the square roots in the basis introduce no additional field for the risk or the certificate coefficients. The exact enclosure of the final optimum is

$$\begin{aligned} 94.555744248543241682944116234978 &< \mathcal{L}_{\text{PVM},k} \\ &< 94.555744248543241682944116234979. \end{aligned} \quad (\text{H3})$$

The enclosure describes the algebraic value in Eq. (15); equality of upper and lower risks follows from the proof below, not from the width of this interval.

A denominator bound with exact contact

For any rank-one PVM write $s_i = \text{Tr}(S_k P_i)$ and $w_i = |\text{Tr}(B_k P_i)|^2$. Since $S_k \succeq pI$, $s_i \geq p > 0$. For $\alpha = \tau_*/3$, define

$$H_\alpha(s) = \frac{1}{\alpha} - \frac{s - \alpha}{\alpha^2} + \frac{(s - \alpha)^2}{p\alpha^2}.$$

The exact identity

$$H_\alpha(s) - \frac{1}{s} = \frac{(s - \alpha)^2(s - p)}{p\alpha^2 s} \geq 0 \quad (\text{H4})$$

holds for every competing outcome. With

$$G_2 = \sum_i w_i, \quad D_3 = \sum_i w_i(s_i - p), \quad D_4 = \sum_i w_i(s_i - p)^2,$$

it gives

$$\sum_i \frac{w_i}{s_i} \leq \frac{G_2}{p} - \frac{2\alpha - p}{p\alpha^2} D_3 + \frac{D_4}{p\alpha^2}. \quad (\text{H5})$$

At the proposed basis the three signal probabilities equal α , and the remaining w_i vanishes. Thus this inequality has exact contact there, unlike the reciprocal secant.

Universal moments of complex projective measurements

Let $R = \sum_i P_i \otimes P_i$ and $Q = R^{\top_2}$, with the partial transpose convention of Appendix G. Both are orthogonal projections. Moreover,

$$\text{Tr}_2 R = I, \quad T_3 = R_{12}R_{23} = \sum_i P_i^{\otimes 3}, \quad T_4 = R_{12}R_{23}R_{34} = \sum_i P_i^{\otimes 4}.$$

These facts hold for arbitrary complex basis vectors. The latter tensors are positive, and their row and column indices are separately symmetric. They express D_3 and D_4 as polynomials of degree two and three in the entries of R , the matrix of R in the normalized symmetric basis of Appendix G.

Average these polynomial moments over $D(e^{i\phi}) \otimes D(e^{i\phi})$ and complex conjugation. The real ladder and diagonal S_k make all three objective components invariant. This is an average of moment vectors of genuine complex bases, not an assumption that a competitor is real or covariant. For lexicographically ordered pairs $A = (a, b)$, $a \leq b$, put

$$m_A = \begin{cases} 1, & a = b, \\ 2, & a < b, \end{cases} \quad X_{AB} = \frac{R_{AB}}{\sqrt{m_A m_B}}.$$

The ancillary model lists the real averaged monomials in these coordinates, including the empty monomial of value one. Entries of its matrices and equalities have rational coefficients.

The positive matrices M_ℓ are principal submatrices and positive diagonal congruences of the following universal constructions: charge blocks of the Gram matrix of $1, R_{AB}$; averages of R and Q ; symmetric tensor-coordinate blocks of T_3, T_4 ; and localizations of R and $I - R$ by constant or diagonal coordinate features. The additional two-by-two localizations use an arbitrary complex coordinate feature on each diagonal and the corresponding product in the off-diagonal entry. Positivity follows by restricting R or $I - R$ to those two feature-weighted coordinates, then averaging. Taking the real part preserves positivity of these real symmetric blocks. Explicitly, for complex coordinate features v, w , let $V = [\bar{v}e_a, we_b]$. Then

$$V^\dagger R V = \begin{pmatrix} |v|^2 R_{aa} & vw R_{ab} \\ \bar{v}w R_{ab} & |w|^2 R_{bb} \end{pmatrix} \succeq 0.$$

The same congruence applies to $I - R$. Neither feature must be real, and the selected coordinates need not differ.

The equations $A_r = 0$ consist of the partial-trace identities, $R^2 = R$, $Q^2 = Q$, and the separate permutation symmetries and partial traces of T_3, T_4 , together with their listed monomial localizations. Multiplication precedes averaging. In particular, these are not idempotency equations for the averaged matrices. The supplied model contains 3,198 positive matrix blocks and 17,538 equations. Its reconstruction checks every matrix entry, every equation and all three weighted objective components. Only necessary conditions for projective measurements are used; no claim that this finite relaxation characterizes all bases is required.

An exact positive identity and global matching

Each M_ℓ is positive for every physical basis, and each A_r vanishes for every such basis. The matrices N_ℓ form congruences of these positive blocks; the supplied Q_ℓ weight their traces. If M_ℓ has size d_ℓ , then N_ℓ is d_ℓ by r_ℓ and Q_ℓ is r_ℓ by r_ℓ ; the factor file specifies these dimensions. For these matrices N_ℓ and symmetric Q_ℓ over K , and multipliers $f_r \in K$, the literal coefficient identity is

$$g_* - \frac{G_2}{p} + \frac{2\alpha - p}{p\alpha^2} D_3 - \frac{D_4}{p\alpha^2} = \sum_\ell \text{Tr}[Q_\ell N_\ell^T M_\ell N_\ell] + \sum_r f_r A_r. \quad (\text{H6})$$

Every coefficient is checked exactly modulo the quartic, without a coefficient residual. Every nonempty Q_ℓ is strictly positive at the selected real root. Directed rational interval LDL factorization certifies all 4,551 pivots as positive; integer-dyadic and Fraction backends give the same result. The matrices in this positivity check are the same matrices appearing in the exact identity. A separate implementation also checks the complete identity and all 4,551 signs. It uses a four-coefficient rational quotient, Gaussian inverses, rational Sturm isolation and exact algebraic LDL, without importing the preceding interval-Schur implementation.

For a physical basis, $A_r = 0$, $M_\ell \succeq 0$, and $N_\ell^T M_\ell N_\ell \succeq 0$. Each trace on the right is therefore nonnegative. Equations (H5) and (H6) give $\sum_i w_i/s_i \leq g_*$, and hence $\mathcal{L}(P) \geq 100 - g_*$ for every complex orthonormal basis. The explicit basis attains this value, proving Eq. (15). Refinement and conditional-mean reporting extend the lower bound to coarse projectors and randomized reports. Parameter-independent classical mixtures of fixed settings cannot lower it, since their risks are averages of risks already bounded below. No restriction on projective dilations of POVMs in a larger space is inferred.

Reproducing the proof

This is a computer-assisted exact proof. The analytical constructions above establish that its constraints hold for every complex basis; exact arithmetic establishes the coefficient identity and the signs of its Gram matrices. The full certificate data and verification code are part of the proof, not an additional numerical test. The PDF states the identity's form rather than printing every coefficient. The directory `anc/pvm-global/` contains the rational model, algebraic factors, direct coefficient identity and root specification. The verifier `audit_model.py` reconstructs the universal positive operators and checks the physical identities. The independent `verify_identity.py` reconstructs the weighted target and every raw coefficient, without an equality-elimination transcript. The model audit uses the supplied operator generator; the analytic projector, tensor and feature-congruence constructions above establish why its constraints apply to every complex basis. The two modes of `verify_psd.py` certify the ordered-field signs on that same certificate. Finally, `verify_attainment.py` multiplies all sixteen pairs of the four explicit projectors, checks normalization and all four optimal reports, and encloses the achieved risk with rational arithmetic. These checks require no numerical optimization software. File hashes and the commands are supplied in `SHA256SUMS.txt` and `README.txt` in that directory. The additional `independent_exact.py` requires GMP rational arithmetic and implements the separate algebraic check just described. All enclosure receipts are bound to their arithmetic dependencies.

Appendix I: Preparation noise and stability of the precision hierarchy

This section concerns the fixed four-state ensemble of Eq. (5), with the same labels, equal prior and identity-weight quadratic loss. We suppress the subscript k . There are two different statements: a specified noise family preserves weak commutativity exactly, whereas arbitrary small preparation errors preserve the risk separations but need not preserve weak commutativity.

Unheralded replacement by the average state

Let $0 < \zeta \leq 1$ and apply the quantum channel

$$\mathcal{N}_\zeta(\rho) = \zeta \rho + (1 - \zeta)S.$$

No flag identifying the replacement branch is supplied to the receiver. The replacement state is the fixed nonscalar average S , not $I_4/4$. Since the prior is centered, the resulting moments are $S_\zeta = S$, $B_{j,\zeta} = \zeta B_j$ and $\lambda = 100$.

For any normalized POVM P , use the finite scalar measure $\mu(d\omega) = \text{Tr } P(d\omega)$ and its positive matrix density P_ω . Set $s(\omega) = \text{Tr } S P_\omega$ and $b_j(\omega) = \text{Tr } B_j P_\omega$. Full rank makes $s > 0$ almost everywhere. Completing the square in the reports gives

$$\inf_a \mathcal{L}(P, a; \zeta) = 100 - \zeta^2 \int \frac{b_1^2 + b_2^2}{s} d\mu, \quad a_j^{\text{opt}} = \zeta b_j/s.$$

The posterior means lie in the convex hull of the four labels, so these integrals are finite. This argument uses densities, not a finite-outcome assumption. It also covers randomized reports by conditional averaging. Consequently, for either the unrestricted POVM class or the original-space PVM class, including parameter-independent classical mixtures,

$$\mathcal{L}_C(\zeta) = 100 + \zeta^2[\mathcal{L}_C(1) - 100]. \quad (\text{I1})$$

The NH expression has the same scaling: substitute $V_j = \zeta W_j$ into Eq. (3), a bijection on Hermitian pairs for positive ζ . Every quadratic and trace-norm term then has the common factor ζ^2 . The symmetric posterior means become ζS_j , so their averaged commutator remains zero.

For positive ζ , maximizing the measurement gain is the same problem as before. The optimal effects are unchanged, all optimal reports scale by ζ , and the complete reported-measurement face in Appendix E is carried to its scaled copy. Seven nonzero finite outcomes remain necessary and sufficient. At $\zeta = 0$, all measurement effects can instead attain risk 100 with zero reports; seven-outcome minimality does not persist.

Arbitrary errors around the original ensemble

Let $\tilde{\rho}_a$ be four density matrices satisfying

$$\max_a \|\tilde{\rho}_a - \rho_a\|_{\text{op}} \leq \epsilon \leq 2^{-16}. \quad (\text{I2})$$

This neighborhood is around the original, unattenuated ensemble, not uniformly around all members of the replacement-noise family. Write $\delta\rho_a = \tilde{\rho}_a - \rho_a$. Then $\text{Tr } \delta\rho_a = 0$, $\|\delta\rho_a\|_1 \leq 4\epsilon$, and $\tilde{\rho}_a \succeq (1/20 - \epsilon)I_4$.

Project any report onto the closed convex hull of the four labels. Euclidean metric projection cannot increase the distance to any label, so it cannot increase either ensemble's risk and stays in the same measurement class. The hull has diameter 20. For the clipped report, the loss effect for each label satisfies

$$A_a = \int \|a(\omega) - \theta_a\|^2 P(d\omega), \quad 0 \preceq A_a \preceq 400I_4.$$

The trace-zero perturbation therefore obeys

$$|\text{Tr } \delta\rho_a A_a| = |\text{Tr } \delta\rho_a (A_a - 200I_4)| \leq 200\|\delta\rho_a\|_1 \leq 800\epsilon.$$

Averaging and taking infima over the identical set of clipped procedures gives the uniform comparison

$$|\tilde{\mathcal{L}}_C - \mathcal{L}_C| \leq 800\epsilon. \quad (\text{I3})$$

It holds for both classes above, including arbitrary outcome spaces and classical mixtures; an attained class minimum is not assumed.

For NH we need only an upper bound on its perturbed infimum. Keep the feasible trial $V_1 = X, V_2 = Y$. Direct multiplication gives

$$H = i[X, Y] = \text{diag}(-1/2, -3/2, -5/2, 9/2), \quad D = X^2 + Y^2 + |H| = \text{diag}(1, 4, 9, 9).$$

For any positive average S' , decomposing $H = H_+ - H_-$ and using the trace-norm triangle inequality gives

$$\|S'^{1/2} H S'^{1/2}\|_1 \leq \text{Tr } S' |H|.$$

At the original diagonal S equality holds, and X, Y attain the NH infimum. Thus, with $\delta S = \tilde{S} - S$ and $\delta B_j = \tilde{B}_j - B_j$,

$$\tilde{\mathcal{L}}_{\text{NH}} - \mathcal{L}_{\text{NH}} \leq \text{Tr } \delta S D - 2 \sum_j \text{Tr } \delta B_j V_j.$$

Here $\|\delta S\|_{\text{op}} \leq \epsilon$ and $\text{Tr } \delta S = 0$, whereas $\text{Tr } |D - 5I_4| = 13$. The first term is at most 13ϵ . The four fixed labels give $\|\delta B_j\|_{\text{op}} \leq 5\epsilon$. Finally $\text{Tr } X^2 = \text{Tr } Y^2 = 7$ implies $\|X\|_1, \|Y\|_1 \leq \sqrt{28} < 6$. The two linear terms cost at most 120ϵ , yielding

$$\tilde{\mathcal{L}}_{\text{NH}} \leq \mathcal{L}_{\text{NH}} + 133\epsilon. \quad (\text{I4})$$

This is a one-sided feasible-trial estimate, not a claim that the trial remains an optimizer.

Combining Eqs. (I3) and (I4) with the original strict bounds proves

$$\tilde{\mathcal{L}}_{\text{min}} - \tilde{\mathcal{L}}_{\text{NH}} > 1/32 - 933\epsilon \geq 1115/65536 > 1/64, \quad (\text{I5})$$

$$\tilde{\mathcal{L}}_{\text{PVM}} - \tilde{\mathcal{L}}_{\text{min}} > 25/192 - 1600\epsilon \geq 325/3072 > 1/10. \quad (\text{I6})$$

There is also a guarantee for the original explicit seven-outcome protocol (P_*, a_*) , with both its effects and reports unchanged. Its original optimal reports are posterior means and lie in the label hull. The same bounded-loss comparison therefore gives $\tilde{\mathcal{L}}(P_*, a_*) \leq \mathcal{L}_{\text{min}} + 800\epsilon$. Combining this with the PVM side of Eq. (I3) yields

$$\tilde{\mathcal{L}}_{\text{PVM}} - \tilde{\mathcal{L}}(P_*, a_*) > 25/192 - 1600\epsilon \geq 325/3072 > 1/10. \quad (\text{I7})$$

Thus this fixed protocol retains its advantage even when the competing PVM is reoptimized for the perturbed preparations. No knowledge of the individual errors is required to use the fixed protocol, provided the stated error bound applies. Measurement effects and reporting are ideal; errors in the readout itself are not modeled. Neither exact weak commutativity, continued global optimality of this protocol, exact risks, nor a seven-outcome minimum is asserted under arbitrary preparation errors. The tolerances are conservative mathematical bounds in the stated dimensionless loss units, not experimental calibration specifications. The ancillary robustness checkers reconstruct the exact matrix identities and constants; the measure and optimization arguments are the analytical proof above.

Appendix J: A family containing oscillator and pair-annihilation ladders

This appendix varies both relative transition strengths and average-state populations. It proves a strict gap above NH for all measurements throughout the family. The oscillator slice also admits the exact optimal readout derived below; the remaining shapes' exact risks and the family's outcome classification are not determined.

States, moments, and full rank

Take $0 \leq u \leq 1/3$, $1 \leq k \leq 2$, and $0 < \eta \leq 1/6$. The upper ladder C_u has positive entries $\ell_1, \ell_2, 1$ defined in the main text. For this family only, set

$$\begin{aligned} p &= \left[2 + 2k - \frac{\ell_2^2(k-1)}{(1+k)^2} \right]^{-1}, & q &= kp, \\ t &= q - \frac{\ell_2^2 p(k-1)}{(1+k)^2}, & S &= \text{diag}(p, p, q, t), \\ K_- &= \text{diag}(p, p, q, 0), & K_+ &= \text{diag}(0, 0, 0, t), \\ B &= C_u K_+ + K_- C_u, & B_1 &= (B + B^\dagger)/2, \quad B_2 = (B - B^\dagger)/(2i). \end{aligned} \tag{J1}$$

These quantities are distinct from the fixed model's moments. The four equiprobable preparations and labels are

$$\rho_{s,j} = S + s\eta B_j, \quad \theta_{s,j} = se_j, \quad s = \pm 1, \quad j = 1, 2. \tag{J2}$$

Their actual first moments are gB_j , where $g = \eta/2$, and their second prior moment is $\lambda = 1$.

Here $\ell_1^2 \in [1/9, 1/3]$, $\ell_2^2 \in [4/9, 2/3]$, and $\ell_2^2 - \ell_1^2 = 1/3$. The denominator defining p lies in $[4, 6]$: subtracting 4 gives $(k-1)[2 - \ell_2^2/(1+k)^2]$, while subtracting it from 6 gives $2(2-k) + \ell_2^2(k-1)/(1+k)^2$. Both are nonnegative. Also

$$t - p = p(k-1) \left[1 - \frac{\ell_2^2}{(1+k)^2} \right] \geq 0.$$

Thus $2p + q + t = 1$, $1/6 \leq p \leq 1/4$, and $p \leq t \leq q$. For $k > 1$ the average state is nonscalar; $k = 1$ is the scalar boundary $S = I_4/4$.

Write the three upper entries of B as $\beta = (\ell_1 p, \ell_2 p, q + t)$, and put $\beta_0 = \beta_4 = 0$. Every adjacent sum $\beta_{i-1} + \beta_i$ is at most 1: the two interior sums are bounded by $2p \leq 1$ and $\ell_2 p + q + t = 1 - (2 - \ell_2)p \leq 1$, respectively. Consequently $\|B_j\| \leq 1/2$, and $\rho_{s,j} \succeq (p - \eta/2)I_4 \succeq I_4/12$. An explicit preparation decomposition is

$$\rho_{s,j} = \sum_{i=1}^4 r_i |i\rangle\langle i| + \sum_{i=1}^3 \eta \beta_i |\chi_i^{s,j}\rangle\langle \chi_i^{s,j}|, \tag{J3}$$

where

$$r_i = S_{ii} - \frac{\eta}{2}(\beta_{i-1} + \beta_i) \geq \frac{1}{12}, \quad |\chi_i^{s,j}\rangle = (|i\rangle + \text{si}^{j-1}|i+1\rangle)/\sqrt{2}.$$

The weights sum to one. The random preparation record is not supplied to the receiver. These balanced adjacent-level superpositions are not Glauber coherent states.

Weak commutativity and the global NH minimum

The actual posterior-mean operators have upper entries gh_i and $-igh_i$, with $h = (\ell_1/2, \ell_2 p/(p+q), 1)$. They satisfy

$$\mathrm{Tr}(S[S_1, S_2]) = 2ig^2 \left[\frac{(p-q)\ell_2^2 p^2}{(p+q)^2} + q - t \right] = 0. \quad (\text{J4})$$

The cancellation follows from Eq. (J1); it is not an approximation near the scalar boundary.

For a Hermitian contraction R , use the quadratic form q_R defined in the main text, now with this S . The NH objective is $1 + \sup_R q_R(Z) - 2g\Re \mathrm{Tr}(B^\dagger Z)$. The contraction $U = \mathrm{diag}(-1, -1, -1, 1)$ is active at $Z = gC_u$: indeed

$$i[X_u, Y_u] = -\frac{1}{2} \mathrm{diag}(\ell_1^2, \ell_2^2 - \ell_1^2, 1 - \ell_2^2, -1),$$

where X_u, Y_u are the Hermitian components of C_u . The first three displayed diagonal magnitudes are strictly positive. The relation $B = C_u K_+ + K_- C_u$ gives, for every matrix E ,

$$q_U(gC_u + E) - 2g\Re \mathrm{Tr}[B^\dagger(gC_u + E)] = -g^2 \mathrm{Tr} T_0 + q_U(E),$$

where $T_0 = C_u K_+ C_u^\dagger + C_u^\dagger K_- C_u = \mathrm{diag}(0, p\ell_1^2, p\ell_2^2 + t, q)$. Since $q_U(E) \geq 0$, this supporting branch bounds the objective globally and is attained at gC_u . Therefore

$$\mathcal{L}_{\mathrm{NH}}(u, k, \eta) = 1 - g^2 \mathrm{Tr} T_0. \quad (\text{J5})$$

No uniqueness of the NH minimizer is needed.

Every complex report and every normalized POVM

For $w \in \mathbb{C}$ define $F(w) = |w|^2 S - \bar{w}B - wB^\dagger + T_0$. For every $\psi \in \mathbb{C}^4$,

$$\begin{aligned} \psi^\dagger F(w) \psi &= p|w\psi_1 - \ell_1\psi_2|^2 + p|w\psi_2 - \ell_2\psi_3|^2 \\ &\quad + q|w\psi_3 - \psi_4|^2 + t|\bar{w}\psi_4 - \psi_3|^2. \end{aligned} \quad (\text{J6})$$

This proves positivity on the entire complex report plane. If $w \neq 0$ and a nonzero ψ lies in the kernel, then $\psi_3 \neq 0$: otherwise the four residual equations force every component to vanish. The last two equations give $|w| = 1$, and the second gives $|\psi_2|^2 = \ell_2^2 |\psi_3|^2$. At $w = 0$ the kernel is $\mathrm{span}(e_1)$, which obeys the same relation. Thus every positive effect supported in any of these kernels satisfies

$$E_{22} = \ell_2^2 E_{33}. \quad (\text{J7})$$

For the actual model the residual at a report z is $F_g(z) = g^2 F(z/g)$. Push any measurement and its reporting rule forward to a normalized POVM M on $\mathbb{C}$. Risk expansion yields

$$\mathcal{L}(M) - \mathcal{L}_{\mathrm{NH}} = \int_{\mathbb{C}} \mathrm{Tr}[F_g(z)M(\mathrm{d}z)].$$

For finite risk, integrability follows from $S \succeq I_4/6$; infinite-risk procedures cannot attain NH. Relative to $\mu(E) = \mathrm{Tr} M(E)$, the POVM has a positive matrix density $A(z)$. Equality of risks would force $\mathrm{Tr}[F_g(z)A(z)] = 0$ almost everywhere, hence $\mathrm{range} A(z) \subseteq \ker F_g(z)$. Equation (J7), integrated against μ , would give $1 = \ell_2^2$, whereas $\ell_2^2 \leq 2/3$. This excludes equality for arbitrary countably additive POVMs, including continuous or singular report distributions and randomized reporting. No reality, phase covariance or finite-support assumption has entered the argument.

An explicit gap for every measurement

The population mismatch also bounds how closely a measurement can approach NH. Write $a = \ell_1$, $b = \ell_2$, and, for an arbitrary complex report w and vector ψ , put

$$\begin{aligned} e_1 &= w\psi_1 - a\psi_2, & e_2 &= w\psi_2 - b\psi_3, \\ e_3 &= w\psi_3 - \psi_4, & e_4 &= \bar{w}\psi_4 - \psi_3. \end{aligned}$$

Let $E^2 = \sum_i |e_i|^2$, $R = \psi^\dagger F(w)\psi$, and $d = |\psi_2|^2 - b^2|\psi_3|^2$. Equation (J6) gives $R \geq pE^2$. We shall prove the one-sided estimate

$$d \leq \sqrt{19} \|\psi\| E \leq \sqrt{19/p} \|\psi\| \sqrt{R}. \quad (\text{J8})$$

No bound on $|d|$ is required.

Set $\rho = |w|^2$ and $\rho_0 = 7/25$. The residuals obey

$$\bar{w}e_3 + e_4 = (\rho - 1)\psi_3, \quad a|\psi_2|^2 = \Re[(w\psi_1 - e_1)\bar{\psi}_2].$$

For $\rho \leq \rho_0$, first take absolute values in the second identity: $a|\psi_2|^2 \leq |e_1||\psi_2| + |\psi_1||w\psi_2|$. Then use $|w\psi_2| \leq b|\psi_3| + |e_2|$ and the first identity to obtain

$$d \leq \frac{1}{a} \left[|\psi_2||e_1| + |\psi_1| \left(|e_2| + \frac{b\sqrt{\rho}|e_3| + b|e_4|}{1 - \rho} \right) \right].$$

The corresponding two-row feature matrix has disjoint row supports. Its squared operator norm is

$$\frac{1 + b^2(1 + \rho)/(1 - \rho)^2}{a^2} \leq 9 + \frac{4(1 + \rho_0)}{(1 - \rho_0)^2} = \frac{1529}{81} < 19.$$

Here $a^{-2} \leq 9$, $b^2/a^2 \leq 4$, and $(1 + \rho)/(1 - \rho)^2$ increases on $[0, \rho_0]$.

For $\rho \geq \rho_0$, use instead the exact identity

$$\rho d = -b^2 \Re[(\bar{w}e_3 + e_4)\bar{\psi}_3] + \Re[e_2(b\bar{\psi}_3 + \bar{w}\bar{\psi}_2)].$$

It implies

$$d \leq (|\psi_2|, |\psi_3|) \begin{pmatrix} \rho^{-1/2} & 0 & 0 \\ b/\rho & b^2/\sqrt{\rho} & b^2/\rho \end{pmatrix} \begin{pmatrix} |e_2| \\ |e_3| \\ |e_4| \end{pmatrix}.$$

Call this nonnegative matrix $K(\rho, b)$. Its entries decrease with ρ and increase with b , so its operator norm is at most that of $K_0 = K(7/25, \sqrt{2/3})$. This comparison follows by applying the matrices to the componentwise absolute value of a vector; entrywise order is not being used as a positive-semidefinite order. The leading principal minor and determinant of $19I_2 - K_0K_0^T$ are, respectively,

$$\frac{108}{7} > 0, \quad \frac{20194}{1029} > 0.$$

Thus $\|K_0\| < \sqrt{19}$, completing Eq. (J8) on the entire complex plane.

Put $Q_b = \text{diag}(0, 1, -b^2, 0)$. Spectral decomposition and Cauchy–Schwarz extend the one-sided bound to every positive matrix A :

$$\text{Tr}(Q_b A) \leq \sqrt{19/p} \sqrt{\text{Tr} A \text{Tr}[F(w)A]}.$$

For any normalized POVM M , rescale reports by $z = gw$ and use its density with respect to $\mu = \text{Tr} M$. Normalization gives $\int \text{Tr}(Q_b A) d\mu = 1 - b^2 > 0$ and $\mu(\mathbb{C}) = 4$. Integrating first and only then squaring yields

$$(1 - b^2)^2 \leq \frac{76}{p} \int \text{Tr}[F(w)M(dw)].$$

Since $F_g(z) = g^2 F(z/g)$ and $g = \eta/2$, every measurement obeys

$$\mathcal{L} - \mathcal{L}_{\text{NH}} \geq \frac{p(1 - b^2)^2}{304} \eta^2 \geq \frac{\eta^2}{16416}. \quad (\text{J9})$$

The argument applies to unbounded finite-risk reports and continuous or singular POVMs; infinite-risk procedures satisfy the bound trivially. It does not require phase covariance or finite support.

There is also a contrast-independent relative statement. With $T = \text{Tr } T_0 = 1 - 2p + p(a^2 + b^2)$, positivity is immediate from $T = p(a^2 + b^2) + q + t$. Also $a^2 + b^2 = 1 - 4u/3 \leq 1$, so $T \leq 1 - p \leq 5/6$. Since $1 - \mathcal{L}_{\text{NH}} = g^2 T$,

$$\frac{\mathcal{L}_{\min} - \mathcal{L}_{\text{NH}}}{1 - \mathcal{L}_{\text{NH}}} \geq \frac{p(1 - b^2)^2}{76T} \geq \frac{1}{3420}. \quad (\text{J10})$$

Thus a fixed fraction of the risk reduction predicted by NH is unattainable throughout this family. The coefficients are conservative lower bounds, not optimal constants or exact minimum risks. The unscaled gap vanishes at $\eta = 0$, which is excluded.

The ancillary scripts `verify_gap.py` and `independent_gap.py` check the residual identities and rational endpoint inequalities by symbolic algebra and, independently, sparse Gaussian-rational polynomial arithmetic. A wrong conjugation is rejected; replacing 19 by 18 fails the small-radius endpoint test.

Optical instances and the meaning of the compression

Let Π_4 project onto the first four single-mode Fock states. At $u = 0$,

$$C_0 = \Pi_4 \hat{a} \Pi_4 / \sqrt{3} \Big|_{\text{range } \Pi_4}.$$

For an integer $m \geq 0$, let $\Pi_{m,4}$ project onto $\{|n+m, n\rangle : 0 \leq n \leq 3\}$. Since $\hat{a}\hat{b}|n+m, n\rangle = \sqrt{n(n+m)}|n-1+m, n-1\rangle$,

$$C_{1/(m+3)} = \frac{\Pi_{m,4} \hat{a} \hat{b} \Pi_{m,4}}{\sqrt{3(m+3)}} \Big|_{\text{range } \Pi_{m,4}}.$$

The normalized $1 : 2 : 3$ ladder occurs at $m = 0$; its matrix coefficients approach the oscillator shape as m grows. Only this discrete sequence is identified with integer number-difference sectors. Other u interpolate the relative transition strengths. The standard uncompressed $\text{SU}(1,1)$ pair realization is given in Ref. [27], Eqs. (2.1)–(2.5).

Compression changes commutation relations at the highest retained level. It does not produce a finite-dimensional unitary $\text{SU}(1,1)$ representation or canonical position and momentum. The prepared states are full rank on the encoded space, not on the entire Fock space. Likewise, an original-space PVM restriction cannot be imposed on an unrestricted optical apparatus: unused levels may implement a projective dilation of an effective POVM. No optical PVM disadvantage or experimental resource advantage is inferred from this identification.

The two checkers in `anc/oscillator-family/` verify the multivariate identities by separate symbolic-polynomial implementations. The first uses SymPy; the second uses standard-library rational sparse polynomials and imports none of the first. Their algebraic checks complement, rather than replace, the kernel and quantitative matrix-measure arguments above.

Exact optimal readout on the oscillator slice

The oscillator endpoint admits a stronger result than a gap estimate. For $u = 0$, we determine the exact minimum risk and an attaining eight-effect POVM for every $1 \leq k \leq 2$ and $0 < \eta \leq 1/6$. The lower bound covers all measurements; only the attaining construction uses four equally spaced phases at each of two radii.

Retain the moments in Eq. (J1), now with $\ell_1 = 1/\sqrt{3}$, $\ell_2 = \sqrt{2/3}$, and put

$$w_0 = p/\sqrt{3}, \quad m_0 = p\sqrt{2/3}, \quad n_0 = q + t.$$

For real ξ, ζ , define

$$\begin{aligned} d &= p^2 \zeta (n_0 + \zeta) - qt \xi (w_0 + \xi), \\ A &= \frac{\xi m_0 p t (n_0 + \zeta)}{d}, \quad D = \frac{\xi m_0 q t (w_0 + \xi)}{d}, \\ R_k(\xi, \zeta) &= \begin{pmatrix} 0 & \xi & 0 & A \\ -w_0 - \xi & 0 & D & 0 \\ 0 & -m_0 - D & 0 & \zeta \\ -A & 0 & -n_0 - \zeta & 0 \end{pmatrix}, \quad c_i = (R_k S^{-1} R_k^T)_{ii}. \end{aligned}$$

These are the weighted-Gram formulas of Appendix D with different, explicitly stated ladder entries. Their algebra does not require that appendix's fixed value of k . Indeed, $R_k + R_k^T = -(B + B^\dagger)$, and the only possible off-diagonal weighted row products are

$$-\frac{\xi(m_0 + D)}{p} + \frac{A\zeta}{t} = 0, \quad \frac{A(w_0 + \xi)}{p} - \frac{D(n_0 + \zeta)}{q} = 0.$$

Consequently, with $A_0 = T_0 - \text{diag}(c_i)$,

$$F(r\omega) - A_0 = D(\omega)(rS + R_k)S^{-1}(rS + R_k)^T D(\omega)^\dagger \succeq 0 \quad (\text{J11})$$

for every $r \geq 0$, $|\omega| = 1$, including the origin. Here $D(\omega) = \text{diag}(1, \omega, \omega^2, \omega^3)$, distinct from the scalar entry D in R_k .

We specify the exact branch through normalization, not stationarity. Reorder $R_k S^{-1}$ as 1, 3, 2, 4, obtaining off-diagonal blocks K, L , and set

$$\sigma = \text{Tr}(KL), \quad \kappa = \det(KL), \quad h(x) = x^2 - \sigma x + \kappa.$$

Apply the tridiagonal null-vector recurrence in Appendix D with these $w_0, m_0, n_0, p, q, t, c_i$. At the roots of h , write the squared components as $\nu_i(x)^2 = E_i + X_i x$, with $\nu_1 = 1$. The explicit rational formulas there give $y = (1 - E_2)/X_2$ and the two equations

$$G_3(k; \xi, \zeta) = E_3 + X_3 y - 1 = 0, \quad G_4(k; \xi, \zeta) = E_4 + X_4 y - 1 = 0. \quad (\text{J12})$$

All their denominators are nonzero on the boxes used below. On each box,

$$\sigma > 0, \quad \kappa > 0, \quad \sigma^2 > 4\kappa, \quad X_2 \neq 0, \quad h(y) < 0.$$

Thus the roots $x_- < x_+$ are positive and distinct, and $\alpha_+ = (y - x_-)/(x_+ - x_-)$, $\alpha_- = 1 - \alpha_+$ are positive. With $r_s = \sqrt{x_s}$ and the real vectors $\nu_s = \nu(x_s)$, the effects and physical reports are

$$M_{s,\omega} = \frac{\alpha_s}{4} D(\omega) \nu_s \nu_s^T D(\omega)^\dagger, \quad z_{s,\omega} = \frac{\eta}{2} r_s \omega, \quad \omega \in \{1, i, -1, -i\}. \quad (\text{J13})$$

Equation (J12) gives every diagonal entry of $\sum M$ equal to one; summing the four phases removes all off-diagonal entries. The vectors are left-null vectors: $(r_s S + R_k)^T \nu_s = 0$. They therefore annihilate the residual in Eq. (J11).

Integrating that positive factor against any normalized POVM gives the global lower bound $1 - g^2 \sum_i c_i$, where $g = \eta/2$. The displayed measurement attains it exactly. Hence

$$\mathcal{L}_{\min}(0, k, \eta) = 1 - \frac{\eta^2}{4} \sum_{i=1}^4 c_i(k). \quad (\text{J14})$$

This is a matching lower bound and measurement, not a numerical agreement or an optimum within a two-orbit ansatz. The arbitrary-outcome integration from the preceding subsections applies unchanged.

Exact branch selection on the full interval. The ancillary directory `anc/oscillator-optimal/` supplies 20 000 closed parameter intervals

$$I_j = [1 + j/20000, 1 + (j + 1)/20000], \quad 0 \leq j < 20000.$$

Their union is exactly $[1, 2]$. For each I_j , a listed rational center $b_j \in \mathbb{Q}^2$ and radius $\varepsilon = 7/10^6$ define $\mathcal{B}_j = b_j + [-\varepsilon, \varepsilon]^2$. For every $k \in I_j$, select the unique root of Eq. (J12) in $\mathcal{B}_j$. The two certificate files contain the intervals and centers; no floating-point approximation is used as the definition of the root.

Here is the existence and uniqueness argument verified on every closed box. An invertible rational preconditioner C_j obeys $\|I - C_j \partial_{(\xi, \zeta)} G(\mathcal{B}_j, I_j)\|_\infty < 1$. The centered image

$$b_j - C_j G(b_j, k) + [I - C_j \partial_{(\xi, \zeta)} G(\mathcal{B}_j, I_j)](b_j - b_j)$$

lies strictly inside $\mathcal{B}_j$, uniformly for $k \in I_j$. To bound its parameter dependence, the implementations use the mean-value identity centered at the midpoint k_j of I_j :

$$G(b_j, k) = G(b_j, k_j) + \partial_k G(b_j, I_j)(k - k_j),$$

where the last expression denotes an interval enclosure. The contraction theorem then proves a root for every real parameter in the interval, not merely at its midpoint. It also proves uniqueness within the stated root box and nonsingularity of the normalization Jacobian. Pole exclusion and all the strict radius/mixture inequalities are checked on the entire same box. At a shared endpoint either adjoining box may be used: both constructions attain a universal lower bound, so their minimum-risk values agree. No uniqueness of the optimal measurement is asserted.

`verify_oscillator.py` uses outward dyadic rational intervals and forward derivatives. `independent_oscillator.py` imports none of it, instead using decimal-rational intervals, a separately assembled weighted matrix and block determinants. The reduced null-vector square coefficients are independently derived by quadratic quotient recurrence in `matching_identities.py`. Exact interval arithmetic verifies the same full coverage and strict predicates in both implementations. For algebraic k , the isolated roots, measurement entries and risk are algebraic; Eq. (J14) also defines the exact optimum for arbitrary real $k \in [1, 2]$. The result does not determine the minimum number of outcomes or the projective optimum on this slice, nor the exact risks for all u .