DECISION-SENSITIVE COMPRESSION WITH LEARNED ACTUATION

Anonymous authors
Paper under double-blind review

ABSTRACT

A compressed predictor can support accurate decisions without predicting every output accurately. This is harder to certify when the action response is also learned: a full-matrix error bound includes irrelevant output noise. We develop a finite-sample alternative for quadratic decisions. A Gaussian expansion at the fixed true optimal controller, together with calibrated low-dimensional moments, bounds the first-order decision error without a leading factor proportional to the square root of the output dimension; noise products and the plug-in Gram bias remain explicit. The resulting observable policy ball certifies arbitrary fitted linear controllers, and its optimal rank-constrained controller is ordinary plug-in compression. A local-rotation lower bound isolates the gap-dependent difficulty of selecting a representation, while a task-sampling bound separates physical calibration from generalization to fresh objectives. Experiments test irrelevant outputs, observable gap resolution, sampled tasks, and neural nuisance–capacity tradeoffs with learned actuation. Calibration improves markedly in the many-output regime, but remains conservative and is not uniformly tighter at small sample sizes. The analysis does not certify nonlinear pilots or closed-loop latent world models.

1 INTRODUCTION

A useful representation should preserve distinctions that change a decision, rather than every predictable variation. Control-oriented identification, task-aware prediction, and weighted compression already formalize this principle (Gevers, 2005; Hjalmarsson, 2009; Farahmand et al., 2017; Cheng et al., 2021). The question here is statistical: *when both forecasts and action effects are fitted from noisy observations, how much decision error can a smaller representation introduce?*

Consider appending noisy outputs that contain no state or action signal. The true decision problem has not changed. However, an operator-norm confidence ball for the entire action matrix expands with output dimension and can make certification much harder. This is a limitation of that bound, not necessarily a lower bound on the cost of learning. Our main result controls the controller's error directly. It separates a Gaussian first-order term, whose covariance depends on task-relevant quantities, from explicitly bounded products of calibration noise.

This also clarifies what is being compressed. For one quadratic objective, an optimal action statistic is sufficient; for a fixed collection of objectives, weighted policy-bank compression and task-weighted forecast compression are algebraically equivalent. We do not claim otherwise. We add a fresh-task result in which the encoder sees finitely many sampled objectives, then is tested on a new objective from the same distribution. Task coverage is then a separate source of error, even with perfect physical calibration.

We make three focused contributions: an observable learned-actuation certificate with an explicit finite-sample remainder; a spectral-selection lower bound with gap dependence; and an analysis of fresh-task sampling and feedback limitations. The analysis builds on standard concentration, low-rank approximation, and perturbation theory.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

2 DECISION LOSS, REPRESENTATION RANK, AND THE SPECTRAL GAP

Let $x \in \mathbb{R}^d$ have known second moment $\mathbb{E}xx^\top = I_d$. One open-loop plan has

$$y = Dx + Bu, \qquad J(u;x) = \|Dx + Bu\|^2 + \|u\|^2, \quad D \in \mathbb{R}^{p\times d},\ B \in \mathbb{R}^{p\times m}. \tag{1}$$

Here m counts all coordinates of the action sequence. A rank-r forecast $F = VE$ transmits Ex; the action map is not compressed. With fitted action map C, its controller is $u = -K_C F x$, where

$$H_B = I + B^\top B, \qquad K_B = H_B^{-1} B^\top, \qquad W_B = BK_B. \tag{2}$$

For any linear controller $u = -Tx$, completing the square gives

$$\mathcal{R}(T) = \|H_B^{1/2}(T - K_B D)\|_{\mathrm{F}}^2. \tag{3}$$

We also write $\mathcal{R}(F, C) = \mathcal{R}(K_C F)$. The known-model rank oracle is $\mathcal{R}_r^* = \sum_{j>r} \lambda_j(L)$, $L = D^\top W_B D$, attained by DP_r^*, where P_r^* is a top-r orthoprojector. This is weighted low-rank approximation (Eckart & Young, 1936; Cheng et al., 2021).

Proposition 1 (Deployed spectral decomposition)**.** *Put $\widehat{L} = \widehat{D}^\top W_C \widehat{D}$ and $\mathsf{E} = H_B^{1/2}(K_C\widehat{D} - K_B D)$. For every orthoprojector P,*

$$\mathcal{R}(\widehat{D}P, C) = \mathrm{tr}(L(I - P)) + \|\mathsf{E}P\|_{\mathrm{F}}^2. \tag{4}$$

For a top-r projector $\widehat{P}_r$ of $\widehat{L}$, $\epsilon = \|\widehat{L} - L\|_{\mathrm{op}}$, and positive population gap $\gamma_r = \lambda_r(L) - \lambda_{r+1}(L)$,

$$\mathcal{R}(\widehat{D}\widehat{P}_r, C) - \mathcal{R}_r^* \le \|\mathsf{E}\widehat{P}_r\|_{\mathrm{F}}^2 + \min\{2r\epsilon, 4r\epsilon^2/\gamma_r\}. \tag{5}$$

The first branch needs no gap; the spectral term is zero at ranks zero and d.

The cross term vanishes because the two errors have orthogonal right supports, P and $I - P$. The remaining reconstruction-risk argument is classical (Davis & Kahan, 1970; Wedin, 1972; Reiß & Wahl, 2020; Koltchinskii & Lounici, 2017). Under fixed dimensions and well-conditioned Gaussian reset data, calibration contributes $O(n^{-1})$ and spectral selection contributes $O(\min\{n^{-1/2}, 1/(\gamma_r n)\})$. This is a gap distinction, not a distinction between zero and positive oracle loss (Srebro et al., 2010; Hu et al., 2022; Elmachtoub et al., 2023).

Which fitted models are covered? Equation (4) applies to $F = \widehat{D}P$, including Prediction, Plug-in, and inflated-metric truncation. It does not automatically apply to arbitrary posterior or certificate-optimization outputs. The controller certificate below does. Exact empirical certainty equivalence requires $r \ge \mathrm{rank}(W_C^{1/2}\widehat{D})$; zero *population* oracle loss alone does not imply this condition. Appendix A.1 gives an explicit counterexample.

What rank does and does not measure. A factored forecast stores $r(p + d)$ coefficients rather than pd, but the $p \times m$ action map remains. For one fixed task, storing a rank-r controller costs only $r(m + d)$ coefficients. These are different deployment choices: neither the rank constraint nor a favorable decision loss proves a forecast interface more memory-efficient than its policy statistic. We report representation dimension and decision cost, not measured speedups. Multiple or freshly sampled tasks change which distinctions must be retained, but do not remove this accounting issue. The feature map and its second moment are supplied in the linear theory; learning them is a separate representation problem.

3 CALIBRATE DECISION ERROR, NOT EVERY ACTION-MATRIX ENTRY

The main theorem assumes independent isotropic Gaussian errors from ordinary least squares (OLS) regression

$$\widehat{D} = D + sG_D, \qquad C = B + bG_B, \tag{6}$$

with known s, b and independent standard Gaussian matrices. An orthogonal reset design with Gram $n\,\mathrm{diag}(I_d, cI_m)$ gives $s = \sigma/\sqrt{n}$ and $b = \sigma/\sqrt{nc}$. This assumption is narrower than general correlated or trajectory-dependent calibration.

Write $T_* = K_B D$, $R_* = D - BT_*$, $Z = \widehat{D} - D$, $\Delta = C - B$, and $\widehat{T} = K_C\widehat{D}$. An exact, not asymptotic, identity is

$$\begin{aligned} H_C(\widehat{T} - T_*) &= G_0 + Q_0, \\ G_0 = B^\top Z + \Delta^\top R_* - B^\top \Delta T_*, &\qquad Q_0 = \Delta^\top Z - \Delta^\top \Delta T_*. \end{aligned} \tag{7}$$

The expansion is at the *fixed true* controller; hence G_0 is Gaussian despite data-dependent fitting. Its exact preconditioned second moment is

$$\begin{aligned} v = \mathbb{E}\|H_B^{-1/2} G_0\|_{\mathrm{F}}^2 = s^2 d\,\mathrm{tr}(W_B) + b^2\big[&\mathrm{tr}(H_B^{-1})\|R_*\|_{\mathrm{F}}^2 \\ &+ \mathrm{tr}(W_B)\|T_*\|_{\mathrm{F}}^2 - 2\,\mathrm{tr}(T_*^\top H_B^{-1} T_*)\big]. \end{aligned} \tag{8}$$

Appending zero-signal outputs does not change this expression. Thus the leading action term, not only the forecast term, avoids a spurious $\sqrt{p}$ penalty.

Theorem 2 (Observable calibrated policy ball)**.** *Under* (6)*, the explicit construction in Appendix A.6 returns $r_0 \geq 0$ and $f \geq 1$, using only $(\widehat{D}, C, s, b, \delta)$. With probability at least $1 - \delta$,*

$$\mathcal{R}(T) \leq \mathcal{C}_{\mathrm{ball}}(T) := \left(r_0 + \sqrt{f}\,\|H_C^{1/2}(T - \widehat{T})\|_{\mathrm{F}}\right)^2 \quad \textit{for every } T \in \mathbb{R}^{m\times d}. \tag{9}$$

Here r_0 bounds uncompressed certainty-equivalent decision error. Its Gaussian part uses observable upper bounds on (8) *and its covariance operator. Its remainder retains centered noise products of scales $\sqrt{p}\,bs$ and $\sqrt{p}\,b^2$, and the plug-in Gram bias pb^2.*

The construction uses $\bar{H} = I + C^\top C - pb^2 I$ and $\bar{Y} = C^\top \widehat{D}$ to calibrate small decision moments. Bias-corrected moment estimation is a known principle (Loh & Wainwright, 2012); our calculation gives a finite-sample controller ball, including its observable covariance caps and remainder. Gaussian and matrix concentration are standard (Vershynin, 2018; Tropp, 2015). The theorem is not an assertion that all dimension dependence disappears, or that this certificate dominates simpler bounds in every regime.

Corollary 3 (The certifying compressor is ordinary Plug-in)**.** *Over all controllers of rank at most r, the minimum certificate equals*

$$\mathcal{C}_r = \left(r_0 + \sqrt{f\sum_{j>r}\lambda_j(\widehat{L})}\right)^2, \tag{10}$$

and is attained by $T_r = K_C\widehat{D}\widehat{P}_r$. It is therefore also optimal over all rank-r forecast interfaces using K_C.

Weighted singular-value truncation proves the corollary. The smallest r with $\mathcal{C}_r \leq \tau$ is certified at tolerance τ; if $r_0^2 > \tau$, no rank passes. Data-dependent selection is the immediate consequence of the uniform event, not a separate contribution (Angelopoulos et al., 2022).

4 WHAT LIMITED DATA CANNOT RESOLVE

4.1 A LOWER BOUND ON SPECTRAL SELECTION ITSELF

A lower bound arising only from controller estimation says little about the price of choosing an encoder. We therefore grant the learner the true controller *after* it freezes a rank-one encoder. For $D = I_2$, let

$$B_\theta = R_\theta \operatorname{diag}\left(\sqrt{\frac{1/4+g}{3/4-g}}, \frac{1}{\sqrt{3}}\right) R_\theta^\top, \quad 0 < g \le 1/4. \tag{11}$$

Its decision Gram has gap exactly g, and its leading direction rotates with θ.

Theorem 4 (Gap-dependent spectral-selection lower bound)**.** *Observe $C = B + \sigma G/\sqrt{n}$. For $0 < \gamma \le 1/4$ and $\sigma/\sqrt{n} \le 1/4$, over the bounded family* (11) *with gaps* at least γ*, every rank-one encoder selection rule satisfies*

$$\sup_B \mathbb{E}[\textit{selection excess}] \ge \frac{1}{256} \min\left\{\frac{\sigma}{\sqrt{n}}, \frac{\sigma^2}{\gamma n}\right\}. \tag{12}$$

This holds even with the true decoder/controller revealed after selection.

The proof uses two rotations $\pm\theta$ with $g = \max(\gamma, \sigma/\sqrt{n})$ and $\theta = \frac{1}{8}\min(1, \sigma/(g\sqrt{n}))$. It is a standard two-point method with a different construction, not a new testing technique. For an *exactly prescribed* gap g, the corresponding lower bound is $c\min\{g, \sigma^2/(ng)\}$; confusing these two parameter classes would overstate the result. Appendix A.9 gives all constants.

4.2 GENERALIZATION TO FRESH OBJECTIVES

A task τ supplies a transform S_τ and cost $\|S_\tau(Dx + Bu)\|^2 + \|u\|^2$. Tasks are independent and identically distributed (IID) and independent of physical calibration. An encoder is fitted using K sampled tasks; a fresh task is revealed *after* encoding. Define

$$L_\tau = D^\top S_\tau^\top W_{S_\tau B} S_\tau D, \qquad L_{\mathcal{T}} = \mathbb{E}_\tau L_\tau, \qquad \widehat{L}_K = K^{-1}\sum_{i=1}^K \widehat{L}_{\tau_i}.$$

Theorem 5 (Task sampling and physical calibration)**.** *Assume $0 \preceq L_\tau \preceq MI$, and a physical confidence event gives $\sup_\tau \|\widehat{L}_\tau - L_\tau\|_{\mathrm{op}} \le \epsilon_{\mathrm{cal}}$ and $\mathbb{E}_\tau\|\mathsf{E}_\tau\|_{\mathrm{F}}^2 \le a_{\mathrm{cal}}$. Put*

$$a_K = M\left(\sqrt{\frac{2\log(2d/\delta_t)}{K}} + \frac{2\log(2d/\delta_t)}{3K}\right), \quad e = \epsilon_{\mathrm{cal}} + a_K. \tag{13}$$

Except with the physical failure probability plus δ_t, a top-r encoder of $\widehat{L}_K$ has fresh-task excess above its population rank oracle at most $a_{\mathrm{cal}} + \min\{2re, 4re^2/\gamma_{\mathcal{T}}\}$ when the population gap is positive; the first branch needs no gap.

This is matrix Bernstein plus the deployed decomposition, with explicit physical-event sufficient conditions in Appendix A.10. For any finite training bank, $\widehat{L}_K$ is exactly the Gram matrix of its Hessian-weighted certainty-equivalent policies. Thus the method is also principal component analysis (PCA) of the policy bank. The fresh-task theorem quantifies generalization; it does not distinguish two algebraically identical estimators.

5 EXPERIMENTS: LEARNED ACTUATION AND IRRELEVANT OUTPUTS

Design and statistical units. All experiments are synthetic. The main calibration suites use the exact Gaussian OLS sufficient-statistic law in (6); the neural pilots do not satisfy the linear OLS assumptions. Every reported linear certificate has nominal failure probability $\delta = 0.05$. Policy-ball calibration allocates twelve tail units. The full-action stationarity comparison allocates three tail units to action, full forecast, and projected forecast errors; a deployment-only variant uses two. Task-concentration diagnostics separately use $\delta_t = 0.05$ and privileged calibration quantities, and are not reported as deployable 95% bounds.

For the dimension study, $d = 8$, $m = r = 2$, $n = 4096$, and $\sigma = 0.1$. A fixed four-output system is padded by outputs with zero forecast and action signal but nonzero observation noise. The fitted action map is unrestricted. We use 160 paired seeds per output dimension and excitation level, 1,600 fits in total. The original four signal rows use the same noise across dimensions. All methods at a fit see exactly the same estimates; no extra validation samples are given to the policy-ball certificate.

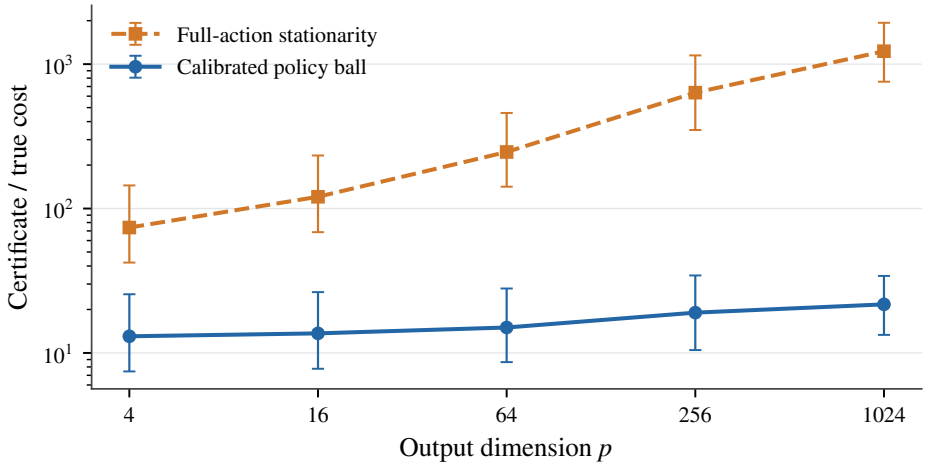


Figure 1: Learned actuation, excitation $c = 0.3$. Median certificate/true-cost ratio; bars show 10th–90th percentiles over 160 fits, not confidence intervals. Policy-ball calibration removes most of the artificial growth. It does not make the remaining factors close to one.

Table 1: Means for true cost and $\|H_B^{1/2}(K_C - K_B)F\|_{\mathrm{F}}$; median ratios. The model and fitted controller are unchanged between certificate columns.

p	True cost	True action deviation	Stationarity ratio	Policy-ball ratio
4	4.7×10^{-5}	3.10×10^{-3}	73.9	13.0
16	4.7×10^{-5}	3.10×10^{-3}	120.6	13.7
64	4.7×10^{-5}	3.11×10^{-3}	246.5	15.0
256	4.8×10^{-5}	3.34×10^{-3}	634.7	19.0
1024	7.1×10^{-5}	6.04×10^{-3}	1229.0	21.7

The true cost is almost flat through $p = 256$, confirming that the full-action penalty mainly reflected a loose bound. At $p = 1024$, the pb^2 plug-in Gram bias is no longer negligible. We retain it in the theorem rather than extrapolating dimension independence indefinitely. A debiased-moment controller is recorded as a comparator, not asserted to be universally preferable. With stronger excitation $c = 1$, policy-ball median ratios range from 8.1 to 13.0, versus 44.4 to 931.6 for the stationarity bound.

Reallocating the comparison-only confidence budget improves the stationarity bound by about 0.9–4.5% in this suite. That small change cannot explain the order-of-magnitude improvement at large p.

6 RATES AND WHETHER THE DATA ACTUALLY RESOLVE A GAP

Pure spectral-selection experiments. We evaluate the two rotated worlds from Theorem 4, choose the leading projector using the noisy action estimate, and then grant the true controller. The experiment therefore measures encoder-selection error alone. There are 400 draws at each of four minimum gaps and five sample counts: 8,000 rows. The true gap is $\max(\gamma, \sigma/\sqrt{n})$; it is not held fixed in cells where the classwise worst case moves with sample size.

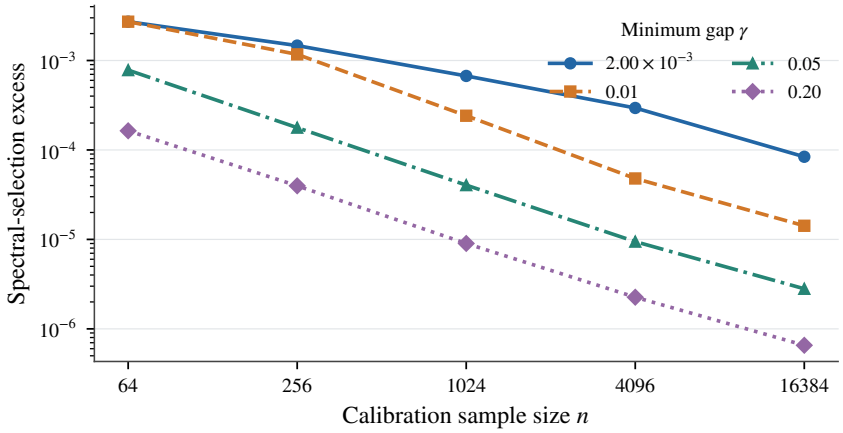


Figure 2: Spectral-selection excess under the local-rotation construction. Fixed separated instances eventually follow the inverse-sample-size branch; smaller allowed gaps delay that transition. Numerical observations do not establish minimax optimality; Theorem 4 supplies the lower bound.

A population gap is not an observable certificate. For two positive-tail systems at rank two, we record how often $\underline{\gamma}_r = \widehat{\gamma}_r - 2\epsilon_{\text{conf}}$ is positive. The diagonal system has oracle tail about 0.06; the independently oriented system has tail about 0.06. There are 160 draws per cell and 2,240 fits. The projected stationarity analysis and the small-moment radius give the following fractions.

Table 2: Fraction of datasets with a positive gap lower confidence bound. Each fraction summarizes one condition; the conditions are not independent trials.

	Diagonal system		Rotated system	
n	Stationarity	Moment radius	Stationarity	Moment radius
64	0	0	0	0
256	0	0	0	0
1024	0	0	1	0
4096	0	0	1	1
16384	1	0	1	1
65536	1	1	1	1
262144	1	1	1	1

The observable gap bounds remain conservative, and the moment version is *not* better in these low-output-dimensional examples. At the largest sample count, the diagonal moment-based excess bound averages 0.02 for an achieved excess of about 10^{-5}. Appendix A.8 describes the true-gap/same-radius diagnostic and the exact-error diagnostic. This distinction prevents a rate theorem based on a population gap from being mistaken for a sharp usable certificate.

7 FRESH-TASK TRANSFER AND FITTING BASELINES

Objectives not seen when fitting the encoder. Eight features and one physical action are tested under continuous objectives $S_\tau = q^\top$, with $q = \mathrm{diag}(a)g/\|\mathrm{diag}(a)g\|$, $g \sim N(0, I_8)$, and $a = (1.6, 1.4, 1.2, 1, 0.9, 0.8, 0.7, 0.6)$. Encoders use $K = 2, 8, 32, 128, 512, 2048$ sampled tasks. Physical calibration uses $n = 128, 1024, 8192$, 48 paired seed blocks per sample count. New-task expectations are approximated using 32,768 independent task draws; evaluation Monte Carlo errors are assessed separately from variation across calibration fits.

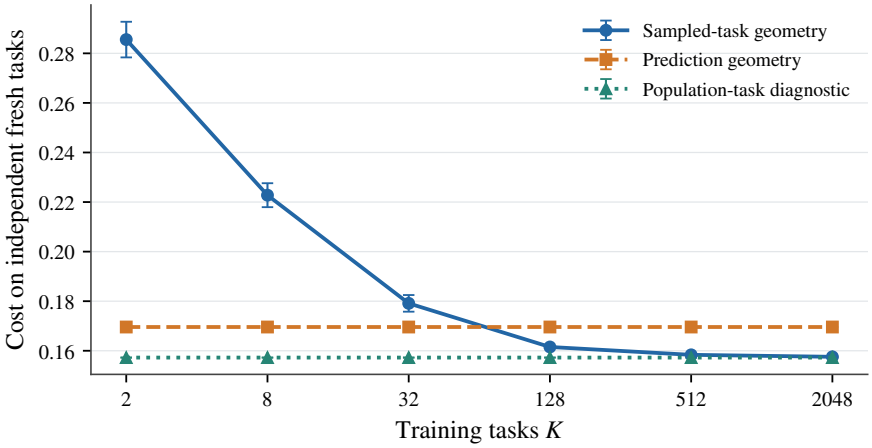


Figure 3: Fresh-task cost, rank two, $n = 8192$. Bars are one standard error over 48 paired fits. The population-task diagnostic uses the large evaluation task bank and is privileged. Finite sampled-task compression exactly equals compression of its Hessian-weighted policy bank.

At $K = 2$, sampled-task cost is 0.29, worse than Prediction's 0.17. With $K = 128$ it is 0.16, and with $K = 2048$ it is 0.16, close to the population-task diagnostic 0.16. Thus a decision-sensitive representation can overfit an inadequately sampled objective distribution. The task-sampling theorem describes this additional resource requirement. Its displayed concentration constants are conservative: the $K = 2048$ diagnostic bound is about 1.23, versus measured excess 3.20×10^{-4}. This diagnostic does not give a tight deployable budget.

A better certificate does not imply a better fitted controller. The policy-ball certificate is minimized by Plug-in (Corollary 3), but minimizing a confidence bound does not establish that the fitted controller minimizes true decision cost. For example, posterior expected-loss optimization gives a shrunk controller, whereas fixing K_C restricts the posterior comparison to representations under the same deployed controller. Appendix B gives both controller conventions. Table 3 compares certificates for identical Plug-in fits across independent forecast/action orientations.

Table 3: Independent-orientation study: 16 systems, 12 draws per sample count, same fits in all columns. Both bounds use nominal $\delta = 0.05$ separately. The policy ball is worse at the smaller sample size.

n	Rank	True Plug-in cost	Stationarity	Policy ball
256	2	0.11	1.97	2.80
1024	2	0.10	0.72	0.55
1024	3	5.57×10^{-3}	0.27	0.17

8 NEURAL CAPACITY TRADEOFFS WITH LEARNED ACTION EFFECTS

This experiment is a controlled objective-mismatch test, not a neural extension of Theorem 2. Eight raw coordinates are uniform on $[-1, 1]^8$. Four outputs are $0.4\tanh(1.5x_{1:4} + 0.5x_{5:8})$; four nuisance outputs are $a\tanh(2x_{5:8})$, with $a \in \{0.2, 0.5, 1\}$. We vary the nuisance amplitude rather than fixing the strongest mismatch. Student widths are two, four, five, six, and seven, all below the raw input dimension.

The action map is now *learned*. Its true value is $B = [\mathrm{diag}(1, 0.8, 0.55, 0.35); 0]$, but training uses only noisy action probes and the resulting W_C. Each pilot uses 1,024 forecast labels; each action estimate uses 32 or 512 separate probes, Gaussian noise scale 0.3, and excitation 0.3. Students distill the learned pilot using 4,096 unlabeled states. Costs are evaluated on 16,384 independent states. No confidence certificate is claimed for these networks.

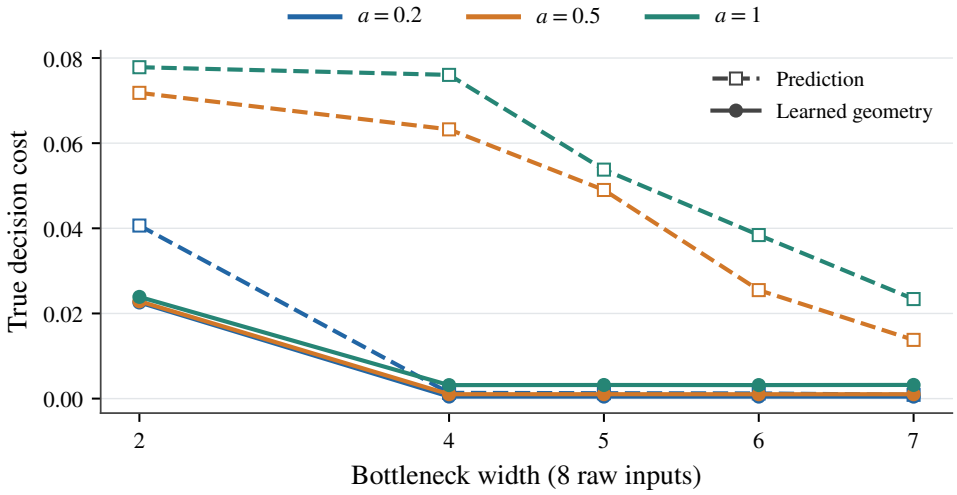


Figure 4: Learned action geometry, 512 action probes. Six paired seed blocks; curves show mean cost. The advantage of task weighting depends on nuisance amplitude and width. Table 4 includes the uncompressed pilot for selected conditions.

Table 4: Actual bottlenecks with learned actuation. All three cost columns are in units of 10^{-3}. The pilot has no bottleneck. Methods share pilot data, initialization, and evaluation states within each comparison. These rows are illustrative cells.

a	Action probes	Width	Prediction	Learned geometry	Pilot
0.2	32	4	4.47	4.72	4.98
0.2	512	4	1.35	0.48	0.44
0.5	512	6	25.48	1.02	1.12
1.0	512	2	77.84	23.89	3.51
1.0	512	7	23.38	3.21	3.51

There are 288 distinct trained networks and 396 evaluated controller/network combinations. These are *six independent seed blocks*, reused across amplitudes, widths, and rescaled action-noise levels; there are 18 pilot-training conditions, not 288 independent calibration events. With few action probes and small nuisance amplitude, task weighting can lose. With larger amplitude, Prediction uses limited capacity on high-variance nuisance outputs; its disadvantage shrinks as width increases. This is the expected objective-mismatch mechanism (Lambert et al., 2020; Voelcker et al., 2022), not an independent novelty claim.

9 INTERPRETATION AND LIMITATIONS

Riccati weighting changes the geometry. For the infinite-horizon linear-quadratic regulator (LQR), let P_* solve the Riccati equation, K_* be optimal, and $H_* = R + B^\top P_* B$. A stabilizing feedback K has occupancy $\Sigma_K = \sum_{t\geq 0}(A - BK)^t \Sigma_0 ((A - BK)^t)^\top$. The classical performance-difference identity is (Fazel et al., 2018)

$$J(K) - J(K_*) = \operatorname{tr}\big((K - K_*)^\top H_* (K - K_*)\Sigma_K\big). \tag{14}$$

For $K = \widehat{K}P$, retained and discarded Euclidean subspaces are generally *not* orthogonal in Σ_K. A cross term remains, and Σ_K depends on the chosen controller. Thus replacing W_B by a Riccati weight does not directly turn Proposition 1 into a closed-loop theorem. Appendix A.11 proves (14), exhibits the cross term, and states a sufficient occupancy-comparability condition for a fixed-reference upper bound. That condition is not estimated or certified by our open-loop procedure.

This issue affects how the controller operates. Repeated feedback changes which states are visited, so it changes the input weighting used to judge compression. Freezing a reference occupancy can recover a weighted projection analysis only after establishing that the fitted controller is stable and its true occupancy is comparable. Neither condition follows from a small open-loop forecast loss.

Paired comparison with reduced feedback control. The paired feedback reanalysis reuses calibration draws from eight generated systems. It uses one mean per system, averaging its eight calibration draws; these system means, rather than individual draws, are the independent statistical units. At $n = 1024$, rank two, Plug-in's 40-step replanning cost ratio is 1.01 and linear-quadratic-Gaussian (LQG) reduction's is 1.00. Their paired difference is $(3.47 \pm 2.51) \times 10^{-3}$ (one standard error); an exploratory 95% Student-t interval is $[-2.46, 9.40] \times 10^{-3}$. These eight systems do not resolve a closed-loop winner, nor establish equality. Ratios are relative to true-model finite-horizon model predictive control (MPC), not globally optimal infinite-horizon control. LQG reduction also compresses dynamics and actuation, unlike our forecast interface (Jonckheere & Silverman, 1983; Moore, 1981; Rowley, 2005).

Relation to adjacent work. Identification for control asks how model errors affect a specified application (Hjalmarsson, 2005); task-optimal exploration and finite-data LQR give related statistical viewpoints (Wagenmaker et al., 2021; Dean et al., 2020; Mania et al., 2019). Our method concerns a rank-constrained forecast/action statistic fitted after calibration, not a new general identification principle. End-to-end decision losses and differentiable planning provide different fitting tools (Donti et al., 2017; Elmachtoub & Grigas, 2022; Amos et al., 2018; Agrawal et al., 2019; Shin et al., 2024); Covariance balancing reduction using adjoint snapshots (CoBRAS) studies state/gradient-based reduction (Otto et al., 2023). Our results do not replace these methods or show that Plug-in is always best.

What the theory does not cover. The main certificate assumes known feature second moments, correctly specified linear forecasts, independent isotropic Gaussian OLS blocks, and known noise scales. It is not a certificate for the neural sweep. Unknown noise scales require accounting for uncertainty in the bias-centering term; simply substituting an upper noise estimate is insufficient (Appendix A.12). General dependent trajectories, estimated covariance shape, unknown whitening, action constraints, and closed-loop stability remain outside the established guarantees. The action map remains uncompressed. A finite policy bank and a forecast interface can be information-equivalent; fresh-task sampling does not prove superiority over the bank.

Conclusion. Decision calibration should not inherit the full-matrix error of every noisy output. An explicit Gaussian controller expansion and small-moment bounds substantially reduce that conservatism while retaining second-order costs. Gap-dependent encoder selection and fresh-task sampling describe distinct statistical limits. These are focused quadratic results with controlled tests, not a general solution to latent world-model compression.

AI USE STATEMENT

In this work, we used generative AI tools to draft sections of the paper and polish writing. Additionally, we used generative AI tools for editing code. We have reviewed all AI-assisted work. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REFERENCES

Akshay Agrawal, Brandon Amos, Shane Barratt, Stephen Boyd, Steven Diamond, and Zico Kolter. Differentiable Convex Optimization Layers. Advances in Neural Information Processing Systems 32, 2019. URL https://arxiv.org/abs/1910.12430.

Alexander A. Alemi, Ian Fischer, Joshua V. Dillon, and Kevin Murphy. Deep Variational Information Bottleneck. International Conference on Learning Representations, 2017. URL https://arxiv.org/abs/1612.00410.

Brandon Amos, Ivan Dario Jimenez Rodriguez, Jacob Sacks, Byron Boots, and J. Zico Kolter. Differentiable MPC for End-to-end Planning and Control. Advances in Neural Information Processing Systems 31, 2018. URL https://arxiv.org/abs/1810.13400.

Anastasios N. Angelopoulos, Stephen Bates, Emmanuel J. Candès, Michael I. Jordan, and Lihua Lei. Learn then Test: Calibrating Predictive Algorithms to Achieve Risk Control. arXiv:2110.01052, version 5, 2022. URL https://arxiv.org/abs/2110.01052v5.

Dilip Arumugam and Benjamin Van Roy. Deciding What to Model: Value-Equivalent Sampling for Reinforcement Learning. Advances in Neural Information Processing Systems 35, 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/3b18d368150474ac6fc9bb665d3eb3da-Abstract-Conference.html.

Gal Chechik, Amir Globerson, Naftali Tishby, and Yair Weiss. Information Bottleneck for Gaussian Variables. Journal of Machine Learning Research, 6:165–188, 2005. URL https://jmlr.org/papers/v6/chechik05a.html.

Jiangnan Cheng, Marco Pavone, Sachin Katti, Sandeep Chinchali, and Ao Tang. Data Sharing and Compression for Cooperative Networked Control. Advances in Neural Information Processing Systems 34, 2021. URL https://arxiv.org/abs/2109.14675.

Chandler Davis and W. M. Kahan. The Rotation of Eigenvectors by a Perturbation. III. SIAM Journal on Numerical Analysis, 7(1):1–46, 1970. URL https://epubs.siam.org/doi/10.1137/0707001.

Sarah Dean, Horia Mania, Nikolai Matni, Benjamin Recht, and Stephen Tu. On the Sample Complexity of the Linear Quadratic Regulator. Foundations of Computational Mathematics, 20(4):633–679, 2020. URL https://doi.org/10.1007/s10208-019-09426-y.

Priya L. Donti, Brandon Amos, and J. Zico Kolter. Task-based End-to-end Model Learning in Stochastic Optimization. Advances in Neural Information Processing Systems 30, 2017. URL https://arxiv.org/abs/1703.04529.

Carl Eckart and Gale Young. The Approximation of One Matrix by Another of Lower Rank. Psychometrika, 1:211–218, 1936. URL https://link.springer.com/article/10.1007/BF02288367.

Adam N. Elmachtoub and Paul Grigas. Smart "Predict, then Optimize". Management Science, 68(1):9–26, 2022. URL https://doi.org/10.1287/mnsc.2020.3922.

Adam N. Elmachtoub, Henry Lam, Haofeng Zhang, and Yunfan Zhao. Estimate-Then-Optimize versus Integrated-Estimation-Optimization versus Sample Average Approximation: A Stochastic Dominance Perspective. arXiv:2304.06833, 2023. URL https://arxiv.org/abs/2304.06833.

Amir-Massoud Farahmand, André Barreto, and Daniel Nikovski. Value-Aware Loss Function for Model-based Reinforcement Learning. Proceedings of Machine Learning Research, 54:1486–1494, 2017. URL https://proceedings.mlr.press/v54/farahmand17a.html.

Maryam Fazel, Rong Ge, Sham M. Kakade, and Mehran Mesbahi. Global Convergence of Policy Gradient Methods for the Linear Quadratic Regulator. Proceedings of Machine Learning Research, 80:1466–1475, 2018. URL https://proceedings.mlr.press/v80/fazel18a.html.

Carles Gelada, Saurabh Kumar, Jacob Buckman, Ofir Nachum, and Marc G. Bellemare. DeepMDP: Learning Continuous Latent Space Models for Representation Learning. International Conference on Machine Learning, 2019. URL https://arxiv.org/abs/1906.02736.

Michel Gevers. Identification for Control: From the Early Achievements to the Revival of Experiment Design. European Journal of Control, 11(4–5):335–352, 2005. URL https://doi.org/10.3166/ejc.11.335-352.

Christopher Grimm, André Barreto, Satinder Singh, and David Silver. The Value Equivalence Principle for Model-Based Reinforcement Learning. Advances in Neural Information Processing Systems 33, 2020. URL https://arxiv.org/abs/2011.03506.

David Ha and Jürgen Schmidhuber. World Models. arXiv:1803.10122, 2018. URL https://arxiv.org/abs/1803.10122.

Danijar Hafner, Timothy Lillicrap, Ian Fischer, Ruben Villegas, David Ha, Honglak Lee, and James Davidson. Learning Latent Dynamics for Planning from Pixels. International Conference on Machine Learning, 2019. URL https://arxiv.org/abs/1811.04551.

Danijar Hafner, Jurgis Pasukonis, Jimmy Ba, and Timothy Lillicrap. Mastering Diverse Control Tasks through World Models. Nature, 640:647–653, 2025. URL https://doi.org/10.1038/s41586-025-08744-2.

Nicklas Hansen, Hao Su, and Xiaolong Wang. TD-MPC2: Scalable, Robust World Models for Continuous Control. International Conference on Learning Representations, 2024. URL https://arxiv.org/abs/2310.16828.

Håkan Hjalmarsson. From Experiment Design to Closed-loop Control. Automatica, 41(3):393–438, 2005. URL https://doi.org/10.1016/j.automatica.2004.11.021.

Håkan Hjalmarsson. System Identification of Complex and Structured Systems. European Journal of Control, 15(3–4):275–310, 2009. URL https://doi.org/10.3166/ejc.15.275-310.

Yichun Hu, Nathan Kallus, and Xiaojie Mao. Fast Rates for Contextual Linear Optimization. Management Science, 68(6):4236–4245, 2022. URL https://doi.org/10.1287/mnsc.2022.4383.

Edmond A. Jonckheere and Leonard M. Silverman. A New Set of Invariants for Linear Systems—Application to Reduced Order Compensator Design. IEEE Transactions on Automatic Control, 28(10):953–964, 1983. URL https://doi.org/10.1109/TAC.1983.1103159.

Vladimir Koltchinskii and Karim Lounici. Normal Approximation and Concentration of Spectral Projectors of Sample Covariance. Annals of Statistics, 45(1):121–157, 2017. URL https://doi.org/10.1214/16-AOS1437.

Nathan Lambert, Brandon Amos, Omry Yadan, and Roberto Calandra. Objective Mismatch in Model-based Reinforcement Learning. Proceedings of Machine Learning Research, 120:761–770, 2020. URL https://proceedings.mlr.press/v120/lambert20a.html.

Po-Ling Loh and Martin J. Wainwright. High-dimensional Regression with Noisy and Missing Data: Provable Guarantees with Nonconvexity. Annals of Statistics, 40(3):1637–1664, 2012. URL https://doi.org/10.1214/12-AOS1018.

Horia Mania, Stephen Tu, and Benjamin Recht. Certainty Equivalence is Efficient for Linear Quadratic Control. Advances in Neural Information Processing Systems 32, 2019. URL https://arxiv.org/abs/1902.07826.

Bruce C. Moore. Principal Component Analysis in Linear Systems: Controllability, Observability, and Model Reduction. IEEE Transactions on Automatic Control, 26(1):17–32, 1981. URL https://doi.org/10.1109/TAC.1981.1102568.

Samuel E. Otto, Alberto Padovan, and Clarence W. Rowley. Model Reduction for Nonlinear Systems by Balanced Truncation of State and Gradient Covariance. SIAM Journal on Scientific Computing, 45(5):A2325–A2355, 2023. URL https://doi.org/10.1137/22M1513228.

Markus Reiß and Martin Wahl. Nonasymptotic Upper Bounds for the Reconstruction Error of PCA. The Annals of Statistics, 48(2):1098–1123, 2020. URL https://doi.org/10.1214/19-AOS1839.

Clarence W. Rowley. Model Reduction for Fluids, Using Balanced Proper Orthogonal Decomposition. International Journal of Bifurcation and Chaos, 15(3):997–1013, 2005. URL https://doi.org/10.1142/S0218127405012429.

Jaeuk Shin, Giho Kim, Howon Lee, Joonho Han, and Insoon Yang. On Task-Relevant Loss Functions in Meta-Reinforcement Learning. Proceedings of Machine Learning Research, 242:1174–1186, 2024. URL https://proceedings.mlr.press/v242/shin24a.html.

Max Simchowitz, Horia Mania, Stephen Tu, Michael I. Jordan, and Benjamin Recht. Learning Without Mixing: Towards a Sharp Analysis of Linear System Identification. Proceedings of Machine Learning Research, 75:439–473, 2018. URL https://proceedings.mlr.press/v75/simchowitz18a.html.

Nathan Srebro, Karthik Sridharan, and Ambuj Tewari. Optimistic Rates for Learning with a Smooth Loss. arXiv:1009.3896, 2010. URL https://arxiv.org/abs/1009.3896.

Takashi Tanaka, Peyman Mohajerin Esfahani, and Sanjoy K. Mitter. LQG Control with Minimum Directed Information: Semidefinite Programming Approach. IEEE Transactions on Automatic Control, 63(1):37–52, 2018. URL https://doi.org/10.1109/TAC.2017.2709618.

Sekhar Tatikonda, Anant Sahai, and Sanjoy K. Mitter. Stochastic Linear Control Over a Communication Channel. IEEE Transactions on Automatic Control, 49(9):1549–1561, 2004. URL https://doi.org/10.1109/TAC.2004.834430.

Naftali Tishby, Fernando C. Pereira, and William Bialek. The Information Bottleneck Method. arXiv:physics/0004057, 2000. URL https://arxiv.org/abs/physics/0004057.

Joel A. Tropp. An Introduction to Matrix Concentration Inequalities. Foundations and Trends in Machine Learning, 8(1–2):1–230, 2015. URL https://doi.org/10.1561/2200000048.

Roman Vershynin. High-Dimensional Probability: An Introduction with Applications in Data Science. Cambridge University Press, first edition, 2018. URL https://doi.org/10.1017/9781108231596.

Claas A. Voelcker, Victor Liao, Animesh Garg, and Amir-Massoud Farahmand. Value Gradient weighted Model-Based Reinforcement Learning. International Conference on Learning Representations, 2022. URL https://arxiv.org/abs/2204.01464.

Andrew J. Wagenmaker, Max Simchowitz, and Kevin Jamieson. Task-Optimal Exploration in Linear Dynamical Systems. Proceedings of Machine Learning Research, 139:10641–10652, 2021. URL https://proceedings.mlr.press/v139/wagenmaker21a.html.

Per-Åke Wedin. Perturbation Bounds in Connection with Singular Value Decomposition. BIT Numerical Mathematics, 12:99–111, 1972. URL https://link.springer.com/article/10.1007/BF01932678.

Amy Zhang, Rowan McAllister, Roberto Calandra, Yarin Gal, and Sergey Levine. Learning Invariant Representations for Reinforcement Learning without Reconstruction. International Conference on Learning Representations, 2021. URL https://arxiv.org/abs/2006.10742.

Gaoyue Zhou, Hengkai Pan, Yann LeCun, and Lerrel Pinto. DINO-WM: World Models on Pre-trained Visual Features Enable Zero-shot Planning. Proceedings of Machine Learning Research, 267:79115–79135, 2025. URL https://proceedings.mlr.press/v267/zhou25t.html.

A PROOFS AND EXPLICIT CONSTRUCTIONS

All models in the linear results are fixed before the Gaussian calibration data are drawn. Matrix norms are operator or Frobenius norms as marked. Input second moments and positive action-cost whitening are supplied. Noise scales in the policy-ball theorem are known; unknown-scale variants require an additional event and centering correction. The word "simultaneous" concerns controllers fitted from the same calibration data, not arbitrary changes to the statistical model after observing results.

A.1 QUADRATIC LOSS, RANK ORACLE, AND EXACT EMPIRICAL POLICY SUFFICIENCY

Expanding (1),

$$J(u;x) = x^\top D^\top Dx + 2u^\top B^\top Dx + u^\top H_B u.$$

Its minimizer is $u_* = -K_B Dx$. Completion of the square and $\mathbb{E}xx^\top = I$ give (3). The identity $K_B^\top H_B K_B = W_B$ and Woodbury's formula $W_B = I - (I + BB^\top)^{-1}$ imply $0 \preceq W_B \prec I$. For known B, rank-r forecast optimization is the best rank-r approximation to $W_B^{1/2} D$ in Frobenius norm. The oracle tail is attained by DP_r^* without inverting W_B, so a singular action map is allowed. This is a direct application of weighted low-rank approximation (Eckart & Young, 1936; Cheng et al., 2021).

The oracle tail vanishes precisely when $r \geq \operatorname{rank}(W_B^{1/2} D)$. The *empirical* identity $K_C \widehat{D} \widehat{P}_r = K_C \widehat{D}$ holds precisely when the discarded empirical policy directions have zero energy; in particular, it holds if $r \geq \operatorname{rank}(W_C^{1/2} \widehat{D})$. These are different conditions. Take

$$D = \widehat{D} = \operatorname{diag}(2, 1), \quad B = \operatorname{diag}(0, 1), \quad C = \operatorname{diag}(\varepsilon, 1).$$

For $0 < \varepsilon < 1/\sqrt{7}$, the empirical top-one projector keeps the second coordinate. The population rank-one oracle loss is zero, but the full fitted controller has first coefficient $2\varepsilon/(1+\varepsilon^2)$ while the compressed controller has first coefficient zero. Thus a rank-deficient true action map with zero population oracle loss need not yield exact empirical certainty equivalence.

A.2 PROOF OF THE DEPLOYED DECOMPOSITION AND GAP BOUND

Set $A_* = H_B^{1/2} K_B D$. The deployed error for $F = \widehat{D} P$ is

$$H_B^{1/2}(K_C \widehat{D} P - K_B D) = \mathsf{E} P - A_*(I - P).$$

Its cross inner product vanishes because $\operatorname{tr}(P\mathsf{E}^\top A_*(I-P)) = \operatorname{tr}((I-P)P\mathsf{E}^\top A_*) = 0$. This proves (4). Let $q = \operatorname{tr}(L(P_r^* - \widehat{P}_r))$. Empirical spectral optimality gives

$$0 \leq q \leq \operatorname{tr}((L - \widehat{L})(P_r^* - \widehat{P}_r)) \leq \epsilon \|P_r^* - \widehat{P}_r\|_* \leq \epsilon\sqrt{2r}\|P_r^* - \widehat{P}_r\|_{\mathrm{F}}.$$

The nuclear norm is also at most $2r$, yielding $q \leq 2r\epsilon$. In an eigenbasis of L, let $s = \operatorname{tr}(P_r^*(I - \widehat{P}_r))$. The same leakage occurs in the opposite direction because both projectors have trace r, and

$$q \geq \gamma_r s = \frac{\gamma_r}{2}\|P_r^* - \widehat{P}_r\|_{\mathrm{F}}^2.$$

Combining the last two displays gives $q \leq 4r\epsilon^2/\gamma_r$; if the projector difference is zero the conclusion is immediate. This is a reconstruction-risk argument rather than a claim of a new eigenvector perturbation theorem (Davis & Kahan, 1970; Wedin, 1972; Reiß & Wahl, 2020; Koltchinskii & Lounici, 2017). No small-error or eigengap assumption is needed for the exact deployed identity.

A.3 TWO GLOBAL RESPONSE-MAP INEQUALITIES

For compatible B, C,

$$\|K_B\|_{\mathrm{op}} \leq \tfrac{1}{2}, \qquad \|K_B - K_C\|_{\mathrm{op}} \leq \|B - C\|_{\mathrm{op}}, \qquad \|W_B - W_C\|_{\mathrm{op}} \leq \min\{1, \|B-C\|_{\mathrm{op}}\}. \tag{15}$$

The singular values of K_B are $s/(1+s^2)$, proving the first assertion. For the second, form the real symmetric dilation $\mathcal{B} = \left[\begin{smallmatrix} 0 & B \\ B^\top & 0 \end{smallmatrix}\right]$. The off-diagonal block of $f(\mathcal{B}) = \mathcal{B}(I + \mathcal{B}^2)^{-1}$ is K_B or its transpose. Since

$$f(A) = \tfrac{1}{2}\{(A - iI)^{-1} + (A + iI)^{-1}\}$$

for symmetric A, the resolvent identity and $\|(A \pm iI)^{-1}\|_{\mathrm{op}} \le 1$ give $\|f(\mathcal{B}) - f(\mathcal{C})\|_{\mathrm{op}} \le \|\mathcal{B} - \mathcal{C}\|_{\mathrm{op}} = \|B - C\|_{\mathrm{op}}$. For the last assertion, differentiate $W_B = I - (I + BB^\top)^{-1}$ along a line. With $V = (I + BB^\top)^{-1}$, the derivative is $V(EB^\top + BE^\top)V$. Its norm is at most $2\|V\|_{\mathrm{op}}\|VB\|_{\mathrm{op}}\|E\|_{\mathrm{op}} \le \|E\|_{\mathrm{op}}$. Integrate along the line and use $0 \preceq W \preceq I$ for the cap at one. These bounds justify the covariance caps below without assuming shared singular vectors.

A.4 Calibrating low-dimensional decision moments

We first give a directly implementable moment bound. It is useful for the proof and for observable gap diagnostics, although the policy-ball certificate is the primary experimental method.

For any fixed matrix M and observation $X = M + aG$ with k entries, two one-sided Gaussian/chi-square tail events yield

$$\langle M, aG\rangle \ge -a\|M\|_{\mathrm{F}}\sqrt{2t}, \qquad \|aG\|_{\mathrm{F}}^2 \ge a^2(k - 2\sqrt{kt}).$$

Each fails with probability at most e^{-t}. Therefore

$$\|M\|_{\mathrm{F}} \le c_t(X, a) := a\sqrt{2t} + \sqrt{\max\{0, \|X\|_{\mathrm{F}}^2 - ka^2 + 2a^2\sqrt{kt} + 2a^2t\}}. \tag{16}$$

Indeed, expanding $\|X\|_{\mathrm{F}}^2$ gives a quadratic inequality in $\|M\|_{\mathrm{F}}$; solving it gives (16). The radicand cannot invalidate the upper bound on the stated event. These are standard Gaussian norm and chi-square bounds; see Vershynin (2018) for the concentration tools.

Put $q = \sqrt{2t}$, $\ell = \sqrt{m} + q$, $B_+ = c_t(C, b)$, and $D_+ = c_t(\widehat{D}, s)$. Define

$$\bar{H} = I + C^\top C - pb^2 I, \qquad \bar{Y} = C^\top \widehat{D}, \tag{17}$$
$$a_H = 2bB_+(\sqrt{m} + q) + b^2(2\sqrt{p}\,\ell + \ell^2),$$
$$a_Y = sB_+(\sqrt{d} + q) + bD_+(\sqrt{m} + q) + bs(\sqrt{p} + \ell)(\sqrt{md} + q).$$

Lemma 6 (Moment confidence event)**.** *Under* (6)*, except on probability* $10e^{-t}$*,*

$$\|\bar{H} - H_B\|_{\mathrm{op}} \le a_H, \qquad \|\bar{Y} - B^\top D\|_{\mathrm{F}} \le a_Y, \qquad \|B\|_{\mathrm{F}} \le B_+, \quad \|D\|_{\mathrm{F}} \le D_+.$$

Proof. Expand

$$\bar{H} - H_B = B^\top\Delta + \Delta^\top B + (\Delta^\top\Delta - pb^2 I), \quad \bar{Y} - B^\top D = B^\top Z + \Delta^\top D + \Delta^\top Z.$$

For fixed B, Gaussian concentration of the Frobenius norm gives $\|B^\top\Delta\|_{\mathrm{F}} \le b\|B\|_{\mathrm{F}}(\sqrt{m} + q)$ with failure at most e^{-t}. The analogous bounds for $B^\top Z$ and $\Delta^\top D$ use $(s, \sqrt{d})$ and $(b, \sqrt{m})$, respectively. The operator norm is no larger than the Frobenius norm. The Gaussian singular-value inequalities, with total failure at most $2e^{-t}$, give

$$\|\Delta^\top\Delta - pb^2 I\|_{\mathrm{op}} \le b^2(2\sqrt{p}\,\ell + \ell^2), \qquad \|\Delta\|_{\mathrm{op}} \le b(\sqrt{p} + \ell).$$

For $p < m$, the smallest singular-value lower bound may be vacuous; the displayed symmetric deviation bound is still valid because its right side then exceeds pb^2. Condition on Δ, using independence of the two OLS blocks. Gaussian concentration gives

$$\|\Delta^\top Z\|_{\mathrm{F}} \le s\{\sqrt{d}\|\Delta\|_{\mathrm{F}} + q\|\Delta\|_{\mathrm{op}}\} \le bs(\sqrt{p} + \ell)(\sqrt{md} + q),$$

with conditional failure at most e^{-t}, since $\|\Delta\|_{\mathrm{F}} \le \sqrt{m}\|\Delta\|_{\mathrm{op}}$. The two signal caps cost four events, the three linear terms cost three, the Wishart bound costs two, and the conditional product costs one: ten tail units in total. Replacing the fixed signal norms by their caps on the same event does not require independence. □

Let $h_- = \max\{1, \lambda_{\min}(\bar{H}) - a_H\}$. For every controller T,

$$\mathcal{R}(T) \le \frac{\left(\|\bar{H}T - \bar{Y}\|_{\mathrm{F}} + a_H\|T\|_{\mathrm{F}} + a_Y\right)^2}{h_-}. \tag{18}$$

To prove this, the true normal-equation residual is $H_BT - B^\top D$; its H_B^{-1} norm squared equals (3). The moment event bounds the difference between true and empirical residuals, and $H_B \succeq h_-I$. The event holds for arbitrary T, including nonlinear functions of the calibration data. It does not cover a nonlinear function of the *deployment input* unless separately represented in the specified linear feature model.

A.5 Exact first-order covariance and the finite remainder

Let $T_* = K_B D$ and $R_* = D - BT_*$. The normal equation gives $B^\top R_* = T_*$. Expanding $H_C\widehat{T} = C^\top \widehat{D}$ and subtracting $H_C T_*$ proves (7). All terms in G_0 are linear in jointly centered Gaussian errors with fixed coefficients.

Write $H = H_B$. Independence of Z and Δ separates the forecast contribution:

$$\mathbb{E}\|H^{-1/2}B^\top Z\|_{\mathrm{F}}^2 = s^2 d\,\mathrm{tr}(H^{-1}B^\top B) = s^2 d\,\mathrm{tr}(W_B).$$

The two action contributions have second moments $b^2\,\mathrm{tr}(H^{-1})\|R_*\|_{\mathrm{F}}^2$ and $b^2\,\mathrm{tr}(W_B)\|T_*\|_{\mathrm{F}}^2$. For the cross term, the Gaussian entry covariance gives $\mathbb{E}[\Delta H^{-1}B^\top\Delta] = b^2 BH^{-1}$, hence

$$\mathbb{E}\langle H^{-1/2}\Delta^\top R_*, H^{-1/2}B^\top\Delta T_*\rangle = b^2\,\mathrm{tr}(R_*^\top BH^{-1}T_*) = b^2\,\mathrm{tr}(T_*^\top H^{-1}T_*).$$

Subtracting twice this quantity proves (8). The negative sign is a consequence of the *true optimum*; it would not be justified by linearizing at an arbitrary data-dependent controller.

Let Σ_G be the covariance of the vectorized $H^{-1/2}G_0$. The operator norm of each linear Gaussian map gives

$$\|\Sigma_G\|_{\mathrm{op}} \le u := s^2\|W_B\|_{\mathrm{op}} + b^2\left(\|H^{-1/2}\|_{\mathrm{op}}\|R_*\|_{\mathrm{op}} + \sqrt{\|W_B\|_{\mathrm{op}}}\|T_*\|_{\mathrm{op}}\right)^2. \tag{19}$$

Gaussian norm concentration therefore yields

$$\|H^{-1/2}G_0\|_{\mathrm{F}} \le \sqrt{v} + \sqrt{2tu} \tag{20}$$

except with probability e^{-t}. The mean norm is bounded by $\sqrt{v}$ using Jensen's inequality.

The quadratic term is not dropped. On the moment event,

$$\|H^{-1/2}Q_0\|_{\mathrm{F}} \le \frac{bs(\sqrt{p}+\ell)(\sqrt{md}+q) + [b^2(2\sqrt{p}\,\ell+\ell^2) + pb^2]\|T_*\|_{\mathrm{F}}}{\sqrt{h_-}}. \tag{21}$$

The pb^2 term is the mean of $\Delta^\top\Delta$, not a tail bound. It can create real plug-in bias when pb^2 is not small. Bias-corrected estimators change this mean, but need not minimize frequentist decision loss on every fixed instance.

Projected moment constants, including the expectation bound. The calibration term need not be summarized by a leading pd error moment. In fact the exact second moment of the preconditioned quadratic term is

$$\begin{aligned} v_Q = \mathbb{E}\|H^{-1/2}Q_0\|_{\mathrm{F}}^2 &= pd\,b^2 s^2\,\mathrm{tr}(H^{-1}) \\ &\quad + b^4\{p(p+1)\,\mathrm{tr}(T_*^\top H^{-1}T_*) + p\,\mathrm{tr}(H^{-1})\|T_*\|_{\mathrm{F}}^2\}. \end{aligned} \tag{22}$$

For a standard Gaussian $p \times m$ matrix G, Wick's identity gives $\mathbb{E}[(G^\top G)A(G^\top G)] = p(p+1)A + p\,\mathrm{tr}(A)I$ for symmetric A; substituting $A = H^{-1}$ proves the second line. The first line follows by conditioning on Δ, and the mixed terms involving Z have zero conditional mean. The cross expectation between G_0 and Q_0 is zero by odd Gaussian moments. Since $H_C \succeq I$,

$$\mathbb{E}\|\mathsf{E}\|_{\mathrm{F}}^2 \le \|H_B\|_{\mathrm{op}}^2\,(v + v_Q). \tag{23}$$

Indeed $\mathsf{E} = H_B^{1/2}H_C^{-1}H_B^{1/2}H_B^{-1/2}(G_0+Q_0)$ and the first factor has norm at most $\|H_B\|_{\mathrm{op}}$. Equation (23) has leading forecast moment $s^2 d\,\mathrm{tr}(W_B)$, not $s^2 pd$. Output dimension occurs in the explicitly second-order products. This separates task-relevant first-order calibration from output-dependent second-order errors.

For fixed dimensions, $s, b = O(n^{-1/2})$ imply $v = O(n^{-1})$ and $v_Q = O(n^{-2})$. Gaussian fourth moments give $\mathbb{E}\|H^{-1/2}G_0\|_{\mathrm{F}}^4 \le 3v^2$. A simple bound $\|\widehat{L} - L\|_{\mathrm{op}} \le 2\|D\|_{\mathrm{op}}\|Z\|_{\mathrm{op}} + \|Z\|_{\mathrm{op}}^2 + \|D\|_{\mathrm{op}}^2\min(1, \|\Delta\|_{\mathrm{op}})$ then gives $\mathbb{E}\epsilon = O(n^{-1/2})$, $\mathbb{E}\epsilon^2 = O(n^{-1})$; these spectral constants are conservative. Combining with (5) proves the stated fixed-dimensional rates. Neither this argument nor classical optimistic-rate forms are claimed as new (Srebro et al., 2010; Hu et al., 2022; Mania et al., 2019).

A.6 Fully observable construction and proof of Theorem 2

For $b > 0$, use $t = \log(12/\delta)$, $q = \sqrt{2t}$ and the moment quantities (17) at this t. Its failure probability is at most $10e^{-t}$. The same Wishart event supplies $e_B = b(\sqrt{p} + \sqrt{m} + q)$; add one Gaussian norm event $\|Z\|_{\mathrm{F}} \le e_D = s(\sqrt{pd} + q)$ and one event (20). Total failure probability is at most $12e^{-t} = \delta$.

The following formulas contain only observations and known noise scales. Put $\eta = \min(1, e_B)$, $\widehat{T} = K_C\widehat{D}$, $\widehat{R} = \widehat{D} - C\widehat{T}$, and

$$\begin{aligned} t_F &= \|\widehat{T}\|_{\mathrm{F}} + e_B\|\widehat{D}\|_{\mathrm{F}} + \tfrac{1}{2}e_D, \\ t_o &= \min\{t_F, \|\widehat{T}\|_{\mathrm{op}} + e_B\|\widehat{D}\|_{\mathrm{op}} + \tfrac{1}{2}e_D\}, \\ r_F &= \|\widehat{R}\|_{\mathrm{F}} + e_D + \eta\|\widehat{D}\|_{\mathrm{F}}, \\ r_o &= \min\{r_F, \|\widehat{R}\|_{\mathrm{op}} + e_D + \eta\|\widehat{D}\|_{\mathrm{op}}\}. \end{aligned} \tag{24}$$

The controller difference is $T_* - \widehat{T} = (K_B - K_C)\widehat{D} - K_B Z$. Thus (15) proves $\|T_*\|_{\mathrm{F}} \le t_F$ and $\|T_*\|_{\mathrm{op}} \le t_o$. Likewise, $R_* - \widehat{R} = -(I - W_B)Z + (W_C - W_B)\widehat{D}$, proving the residual caps.

Let $\lambda_i(\bar{H})$ be its eigenvalues. Weyl's inequality and $H_B \succeq I$ give lower and upper eigenvalue caps

$$h_i^- = \max\{1, \lambda_i(\bar{H}) - a_H\}, \qquad h_i^+ = \max\{1, \lambda_i(\bar{H}) + a_H\}.$$

Define

$$i_t = \sum_i (h_i^-)^{-1}, \quad i_o = (\min_i h_i^-)^{-1}, \quad w_t = \sum_i [1 - (h_i^+)^{-1}], \quad w_o = \max_i [1 - (h_i^+)^{-1}].$$

These upper-bound $\mathrm{tr}(H_B^{-1})$, $\|H_B^{-1}\|_{\mathrm{op}}$, $\mathrm{tr}(W_B)$ and $\|W_B\|_{\mathrm{op}}$, respectively. Discarding the nonpositive cross term in (8) gives observable covariance caps

$$\bar{v} = s^2 d w_t + b^2(i_t r_F^2 + w_t t_F^2), \qquad \bar{u} = s^2 w_o + b^2(\sqrt{i_o} r_o + \sqrt{w_o} t_o)^2. \tag{25}$$

Set

$$\begin{aligned} g_+ &= \sqrt{\bar{v}} + q\sqrt{\bar{u}}, \\ q_+ &= \frac{bs(\sqrt{p} + \ell)(\sqrt{md} + q) + [b^2(2\sqrt{p}\,\ell + \ell^2) + pb^2]t_F}{\sqrt{h_-}}, \\ f &= 1 + \frac{a_H + pb^2}{\lambda_{\min}(H_C)}, \qquad r_0 = f(g_+ + q_+). \end{aligned} \tag{26}$$

On the common event, $H_B \preceq f H_C$. Consequently $\|H_B^{1/2} H_C^{-1} H_B^{1/2}\|_{\mathrm{op}} \le f$, and (7), (20), (21) imply $\|H_B^{1/2}(\widehat{T} - T_*)\|_{\mathrm{F}} \le r_0$. For any T, the triangle inequality and $H_B \preceq f H_C$ give

$$\|H_B^{1/2}(T - T_*)\|_{\mathrm{F}} \le r_0 + \sqrt{f}\,\|H_C^{1/2}(T - \widehat{T})\|_{\mathrm{F}}.$$

Squaring proves Theorem 2. Notice that no random coefficient was treated as fixed inside a Gaussian tail argument: the Gaussian term is defined at (B, D, T_*), and its observable caps are substituted only on the common event.

When $b = 0$, $B = C$ is known and $\widehat{T} - T_* = K_B Z$. A single fixed-projection Gaussian event suffices:

$$f = 1, \qquad r_0 = s\left(\sqrt{d\,\mathrm{tr}(W_B)} + \sqrt{2\|W_B\|_{\mathrm{op}}\log(1/\delta)}\right). \tag{27}$$

For known $B = 0$, $r_0 = 0$ and every forecast deployed through K_B has zero regret and certificate. This exact special case requires no tail allocation.

What output-dimension adaptation means. For fixed signal norms and fixed d, m, the first-order expressions (8) and (19) do not grow with the number of appended zero-signal outputs. The observable caps introduce extra finite-data uncertainty, and the remainder still contains pb^2, $\sqrt{p}b^2$, and $\sqrt{p}bs$. We therefore claim removal of the *leading full-action norm penalty*, not dimension-free loss or calibration in every joint (p, n) regime. In particular, naive certainty equivalence itself can be biased by the noise Gram when pb^2 is appreciable.

A.7 Proof of the rank-certificate corollary

Let $A = H_C^{1/2}\widehat{T}$. Since H_C is invertible, minimizing the certificate over rank-r T is equivalent to minimizing $\|H_C^{1/2}T - A\|_{\mathrm{F}}$. The weighted Eckart–Young solution is $T_r = H_C^{-1/2}[A]_r$. Its right singular vectors are eigenvectors of

$$A^\top A = \widehat{D}^\top K_C^\top H_C K_C \widehat{D} = \widehat{D}^\top W_C \widehat{D} = \widehat{L},$$

so $T_r = K_C \widehat{D}\widehat{P}_r$. The discarded squared norm is $\sum_{j>r}\lambda_j(\widehat{L})$, proving (10). Every rank-r forecast gives a controller of rank at most r, and the displayed minimizer is feasible via $F_r = \widehat{D}\widehat{P}_r$. Thus it is also the global certificate minimizer over forecasts with that rank constraint. This is an optimization consequence of the policy-ball certificate and weighted low-rank approximation.

All ranks are certified on the same event. Thresholding their values is ordinary confidence-set risk control; it is not a separate selection-validity theorem. A general-purpose comparison is Learn then Test (Angelopoulos et al., 2022), whose held-out risk setting is different from bounding an unknown oracle-relative quadratic cost from shared model data. A split-and-certify baseline may freeze a training-half controller and certify it using an independent calibration half; it may not refit after validation and retain the same validation guarantee.

A.8 Observable eigengap and excess bound

The gap diagnostic uses small moments without claiming that their constants are best in low output dimension. Let $J = C^\top \widehat{D}$, $Y = B^\top D$, $h_C = \lambda_{\min}(H_C)$, $a = a_H + pb^2$, $v = a_Y$, and $j = \|J\|_{\mathrm{op}}$. Since $L = Y^\top H_B^{-1} Y$ and $\widehat{L} = J^\top H_C^{-1} J$, the moment event implies

$$\|\widehat{L} - L\|_{\mathrm{op}} \le \epsilon_M := \frac{v(2j+v)}{h_C} + \frac{a(j+v)^2}{h_C h_-}. \tag{28}$$

For the first term, replace Y by J while holding H_C^{-1} fixed; $\|Y\| \le j + v$. For the second, the resolvent identity bounds $\|H_C^{-1} - H_B^{-1}\| \le a/(h_C h_-)$. Weyl's inequality gives $\underline{\gamma}_r = \lambda_r(\widehat{L}) - \lambda_{r+1}(\widehat{L}) - 2\epsilon_M$. If it is positive, (5) can use it; otherwise it falls back to $2r\epsilon_M$. The uncompressed moment certificate (18) bounds $\|\mathsf{E}\widehat{P}_r\|_{\mathrm{F}}^2$ by bounding $\|\mathsf{E}\|_{\mathrm{F}}^2$. This gives the recorded moment-based excess diagnostic. It concerns excess over an unknown rank oracle, not an absolute-risk certificate.

The experiment separately records a version substituting the *true* gap but retaining the same radius, and one using exact calibration errors. These are privileged diagnostics. Comparing only achieved $1/n$ slopes with a positive population gap would hide how much data are needed before a deployable lower confidence bound becomes informative.

A.9 Local-rotation lower bound: proof of Theorem 4

Write $h(\lambda) = \sqrt{\lambda/(1-\lambda)}$. On $[1/4, 1/2]$, $h'(\lambda) \le 2$, so the two eigenvalues of B_θ differ by at most $2g$. For $B_+ = B_\theta$ and $B_- = B_{-\theta}$,

$$\|B_+ - B_-\|_{\mathrm{F}}^2 = 2\{h(1/4+g) - h(1/4)\}^2 \sin^2(2\theta).$$

Under Gaussian calibration, the KL divergence is this quantity times $n/(2\sigma^2)$. Choosing $\theta = \frac{1}{8}\min\{1, \sigma/(g\sqrt{n})\}$ gives

$$\mathrm{KL} \le 16 n g^2 \theta^2/\sigma^2 \le 1/4, \qquad \mathrm{TV} \le \sqrt{\mathrm{KL}/2} \le 1/\sqrt{8}.$$

Let v_+, v_- be the leading unit vectors at the two angles. For any rank-one orthoprojector P, its selection loss in world $\pm$ is $g(1 - v_\pm^\top P v_\pm)$. The sum of these losses is at least $2g\sin^2\theta$, because the largest eigenvalue of $v_+v_+^\top + v_-v_-^\top$ is $1 + \cos(2\theta)$. Integrating against the minimum of the two likelihoods yields

$$\max_{\pm} \mathbb{E}_\pm[\text{selection loss}] \ge g\sin^2\theta(1 - \mathrm{TV}).$$

This also holds for randomized rules by averaging. A rank-zero rule cannot do better. For a linear encoder of rank one and $x \sim N(0, I_2)$, granting the true model for decoding makes conditional expectation Px optimal, so arbitrary linear encoder scaling cannot evade the bound.

Since $\theta \le 1/8$, $\sin\theta \ge \theta/\sqrt{2}$. Also $1 - 1/\sqrt{8} > 1/2$. Therefore the last display is at least $\frac{1}{256}\min\{g, \sigma^2/(ng)\}$. For the class of gaps at least γ, take $g = \max\{\gamma, \sigma/\sqrt{n}\} \le 1/4$. This gives (12). For a class with *exact* gap g below the noise scale, the attainable loss itself is bounded by g, so an $n^{-1/2}$ lower bound independent of g would be false. The theorem is sharp in these two sample-size/gap branches up to constants on the stated bounded class; it does not claim sharp ambient dimension or confidence dependence.

A.10 Fresh tasks: proof and physical calibration conditions

Assume task draws are independent of the physical dataset and let P be any fitted orthoprojector. Apply the exact decomposition separately to the transformed models $(S_\tau D, S_\tau B, S_\tau \widehat{D}, S_\tau C)$ and average:

$$\mathbb{E}_\tau \mathcal{R}(S_\tau \widehat{D} P, S_\tau C) = \mathrm{tr}(L_{\mathcal{T}}(I - P)) + \mathbb{E}_\tau \|\mathsf{E}_\tau P\|_{\mathrm{F}}^2. \tag{29}$$

Here $\mathsf{E}_\tau = H_{S_\tau B}^{1/2}(K_{S_\tau C} S_\tau \widehat{D} - K_{S_\tau B} S_\tau D)$. If $0 \preceq L_\tau \preceq MI$, then each centered matrix has norm at most M and matrix variance at most $M^2 I$. Matrix Bernstein (Tropp, 2015) gives $\|K^{-1}\sum_i L_{\tau_i} - L_{\mathcal{T}}\|_{\mathrm{op}} \le a_K$ with probability at least $1 - \delta_t$, using the conservative radius in (13). On the physical event,

$$\|\widehat{L}_K - L_{\mathcal{T}}\|_{\mathrm{op}} \le \left\| K^{-1} \sum_i (\widehat{L}_{\tau_i} - L_{\tau_i}) \right\|_{\mathrm{op}} + a_K \le \epsilon_{\mathrm{cal}} + a_K.$$

Substitute this into the spectral proof from Appendix A.2 and use $\mathbb{E}_\tau \|\mathsf{E}_\tau P\|_{\mathrm{F}}^2 \le a_{\mathrm{cal}}$. A union bound over the physical event and task event proves Theorem 5. It does not require a union bound over candidate ranks.

One explicit sufficient physical event is $\|Z\|_{\mathrm{op}} \le e_o$, $\|Z\|_{\mathrm{F}} \le e_F$, $\|\Delta\|_{\mathrm{op}} \le e_B$, together with supplied $\|D\|_{\mathrm{op}} \le D_0$ and uniformly bounded $\|S_\tau\|_{\mathrm{op}} \le s_0$. Then

$$\begin{aligned} M &= s_0^2 D_0^2, \\ \epsilon_{\mathrm{cal}} &= s_0^2\{(2D_0 + e_o)e_o + \min(1, s_0 e_B) D_0^2\}, \\ a_{\mathrm{cal}} &= \left(s_0 e_F + \tfrac{3}{2} s_0^2 e_B \|\widehat{D}\|_{\mathrm{F}}\right)^2. \end{aligned} \tag{30}$$

The forecast part of the geometry bound expands the two outer factors and uses $\|W\| \le 1$. The action part uses (15). The calibration bound uses $\|H_A^{1/2} K_A\| \le 1$ and $\|H_A^{1/2}(K_C - K_A)\| \le \frac{3}{2}\|C - A\|$. For completeness, the latter follows by differentiating K_{A+tE} in an H_A-weighted normal equation, or by the following nonlocal identity:

$$H_A(K_C - K_A) = (C - A)^\top (I - CK_C) - A^\top (C - A) K_C.$$

After multiplication by $H_A^{-1/2}$, use $\|I - CK_C\| \le 1$, $\|H_A^{-1/2} A^\top\| \le 1$, and $\|K_C\| \le 1/2$. This proves the coefficient $3/2$ directly. These physical bounds are sufficient and conservative; the fresh-task experiment reports a privileged calibration-error diagnostic rather than suggesting its Monte Carlo population estimates instantiate a deployable confidence set.

Exact policy-bank equivalence. Define the sampled bank

$$\widehat{A}_K = K^{-1/2} \begin{bmatrix} H_{S_{\tau_1} C}^{1/2} K_{S_{\tau_1} C} S_{\tau_1} \widehat{D} \\ \vdots \\ H_{S_{\tau_K} C}^{1/2} K_{S_{\tau_K} C} S_{\tau_K} \widehat{D} \end{bmatrix}.$$

Then $\widehat{A}_K^\top \widehat{A}_K = \widehat{L}_K$ exactly. Thus both procedures select the same encoder up to choices within tied eigenspaces. A forecast representation can generate a fresh task head through (S_τ, C, F); an identical encoder paired with the same model-based head has identical performance. We make no storage, computational, or statistical superiority claim over that equivalent construction.

A.11 RICCATI-WEIGHTED FEEDBACK: AN EXACT IDENTITY BUT A DIFFERENT PROBLEM

Let A, B, Q, R define a stabilizable, detectable discrete-time LQR with $Q \succeq 0$, $R \succ 0$, and stabilizing Riccati solution P_*. Let K_* be the optimal feedback, $H_* = R + B^\top P_* B$, and let K be another stabilizing feedback. The Bellman equation yields

$$x^\top Q x + (Kx)^\top R(Kx) + x^\top (A-BK)^\top P_* (A-BK)x - x^\top P_* x = x^\top (K-K_*)^\top H_* (K-K_*)x,$$

Summing along the stable trajectory telescopes the value terms. Taking expectation over initial covariance Σ_0 proves (14). Equivalently, subtract the two Lyapunov equations. This standard identity underlies LQR policy analysis (Fazel et al., 2018); it is not new here.

For $K = \widehat{K}P$, write $E = (\widehat{K} - K_*)P$ and $V = K_*(I - P)$. The excess is

$$\operatorname{tr}(E^\top H_* E \Sigma_K) + \operatorname{tr}(V^\top H_* V \Sigma_K) - 2\operatorname{tr}(E^\top H_* V \Sigma_K). \tag{31}$$

The final term need not vanish since $(I - P)\Sigma_K P$ need not be zero. A Riccati Hessian by itself does not restore the Euclidean orthogonality argument.

A useful deterministic sufficient condition is a supplied reference covariance $\Sigma_{\mathrm{ref}} \succ 0$ with $\Sigma_K \preceq \kappa \Sigma_{\mathrm{ref}}$. Then

$$J(K) - J(K_*) \le \kappa \| H_*^{1/2} (K - K_*) \Sigma_{\mathrm{ref}}^{1/2} \|_{\mathrm{F}}^2.$$

For a whitened projector controller $K = \widehat{K} \Sigma_{\mathrm{ref}}^{1/2} P \Sigma_{\mathrm{ref}}^{-1/2}$, the right side splits into retained calibration and discarded signal terms by $P(I - P) = 0$. However, establishing stability, estimating P_*, and bounding policy-dependent occupancy require additional work; the open-loop certificate does none of these automatically. With process noise, an average-cost stationary-occupancy identity is possible for stable controllers, whereas undiscounted infinite cumulative cost with persistent process noise is not finite. These objectives are different.

A.12 UNKNOWN SCALES AND CORRELATED OBSERVATIONS

The moment theorem assumes the noise variances used to subtract the Gram bias are known. For a scalar-noise stationarity certificate, a residual chi-square upper bound can replace σ after allocating a scale event. Under independent p-output reset regression, the pooled residual degrees of freedom are $p(n - d - m)$, not simply $n - d - m$.

For the bias-corrected moment certificate, simply substituting an upper scale into the centering term is not justified. A valid interval construction would use a confidence interval $b^2 \in [v_-, v_+]$, center by $p(v_- + v_+)/2$, and add $p(v_+ - v_-)/2$ to the action-moment radius. In the signal cap, use the lower variance in the negative term $-kv_-$ and the upper variance in positive tail terms; Gaussian product radii also use upper scales. An analogous interval is required for s^2. The scale event must be union-bounded with the conditional Gaussian events. This describes the additional centering uncertainty; the policy-ball theorem and reported certificates assume known scales.

A known within-episode covariance shape can be incorporated through Gaussian covariance calculations. It is not the same as estimating an unknown covariance or observing a single endogenous trajectory. Orthogonal-block independence is used essentially for the product and covariance calculations of the policy-ball theorem; dependence needs a different analysis or a carefully accounted split design. Classical identification results address related challenges (Simchowitz et al., 2018; Hjalmarsson, 2009); they do not automatically validate this particular certificate under arbitrary dependence.

B RELATED WORK AND FAIR COMPARATOR SCOPE

Control-aware representations. Weighted forecast compression is close to Cheng et al. (2021). Balanced truncation (Moore, 1981), balanced proper orthogonal decomposition (POD) (Rowley, 2005), and LQG reduction (Jonckheere & Silverman, 1983) compress dynamical state and input/output behavior. CoBRAS (Otto et al., 2023) combines state variability and gradient sensitivity. We neither claim that task weighting is new nor interpret a favorable open-loop loss as superiority over full dynamical reduction. Identification-for-control and application-oriented design (Gevers,

2005; Hjalmarsson, 2005; 2009), task-optimal exploration (Wagenmaker et al., 2021), finite-sample LQR (Dean et al., 2020), and certainty-equivalent efficiency (Mania et al., 2019) are closely related prior work.

Prediction objectives and world models. World Models, latent planning, Dreamer, Temporal Difference Model Predictive Control 2 (TD-MPC2), and DINO World Model (DINO-WM) learn richer representations and dynamics than the object studied here (Ha & Schmidhuber, 2018; Hafner et al., 2019; 2025; Hansen et al., 2024; Zhou et al., 2025). Value equivalence, deep Markov decision process (DeepMDP) models, invariant representations, and value-equivalent model sampling formalize predictive distinctions relevant to a decision (Grimm et al., 2020; Gelada et al., 2019; Zhang et al., 2021; Arumugam & Van Roy, 2022). We do not infer Markov sufficiency or visual generalization from a whitened linear feature model. Value-aware losses and objective-mismatch studies motivate the neural stress test (Farahmand et al., 2017; Lambert et al., 2020; Voelcker et al., 2022). Its varying nuisance amplitude is designed to expose a known tradeoff, not to manufacture evidence for a new learning principle.

Information and end-to-end optimization. Information bottlenecks and Gaussian bottlenecks (Tishby et al., 2000; Chechik et al., 2005; Alemi et al., 2017) constrain information rather than noiseless linear rank. Sequential information-constrained control (Tatikonda et al., 2004; Tanaka et al., 2018) studies causal communication and directed information, unlike the single-plan objective. Task-based end-to-end learning, smart predict-then-optimize, differentiable MPC, differentiable convex layers, and task-relevant model losses (Donti et al., 2017; Elmachtoub & Grigas, 2022; Amos et al., 2018; Agrawal et al., 2019; Shin et al., 2024) address fitting rather than the particular finite-data controller ball here. Optimistic rates and plug-in decision comparisons are also prior art (Srebro et al., 2010; Hu et al., 2022; Elmachtoub et al., 2023).

Posterior comparator and the controller convention. A posterior compressor minimizes posterior expected *deployed* regret holding K_C fixed. With a flat Gaussian likelihood posterior and independent OLS blocks, $\mathbb{E}[B^\top B \mid \text{data}] = C^\top C + pb^2 I$ and $\mathbb{E}[B^\top D \mid \text{data}] = C^\top \widehat{D}$. Ignoring terms independent of T, posterior expected cost is

$$\operatorname{tr}(T^\top (H_C + pb^2 I)T) - 2\operatorname{tr}(T^\top C^\top \widehat{D}).$$

Jointly optimizing the controller therefore gives

$$T_{\text{Bayes}} = (H_C + pb^2 I)^{-1} C^\top \widehat{D}, \tag{32}$$

a shrunk controller. A representation-only comparator fixes K_C so that only compression changes across methods; it is not claimed to optimize the full posterior decision problem. Its weighted reduced-rank solution is a different estimator from the projector form $\widehat{D}P$, so the spectral rate statements do not automatically apply. The policy-ball certificate does apply to any resulting linear T on its common event.

The frequentist debiased-moment comparator instead subtracts $pb^2 I$ from the observed Gram and clips eigenvalues below the known floor one. These are different statistical objectives, not contradictory prescriptions. Neither is declared universally superior. Corollary 3 gives a global optimizer for the policy-ball certificate directly.