DECISION-WEIGHTED PREDICTION REDIRECTS WHAT A CAPACITY-LIMITED ENCODER RETAINS

Anonymous authors
Paper under double-blind review

ABSTRACT

Capacity-limited predictive encoders are standard in model-based control: each compresses a controller's belief into a code too small to hold it, and no guarantee says the code keeps the directions the control cost depends on. The unweighted objective ranks belief directions by predictive energy, which the control cost does not enter. We prove, using standard spectral arguments, that one reweighting of the target redirects the same capacity onto the directions the cost weights most. The weight is a fixed matrix, the factor that makes prediction error measure control cost. That statement is read at an exact optimum with no tie in the cost's ranking, so our finitely trained runs illustrate it rather than test it. Such a factor exists exactly when the predictive map discards no decision-relevant direction; a counterexample shows the reweighting cannot be dropped. Two companion results drop encoder linearity, show decoder linearity cannot be dropped, and make the condition checkable in advance; every theorem is machine-checked in Lean. On synthetic linear-Gaussian systems the weighted target lowers decision-relevant error wherever the capacity constraint binds.

1 INTRODUCTION

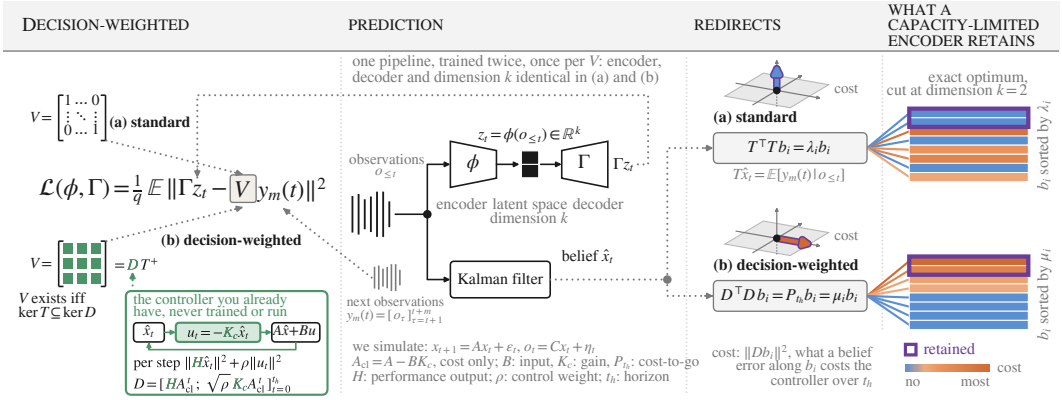


Figure 1: (a) Standard predictive training and the two belief directions a k=2 latent keeps. (b) Decision-weighted, target times $V = DT^{+}$ ($^{+}$ pseudoinverse): a different pair (Eq. 1; Thm. 2; §3). In the sketches the grey plane is the cost-relevant subspace and the arrow the direction retained.

A controller that cannot observe its state acts on a belief about it, and with high-dimensional observations a latent world model is the standard way to carry that belief (Hafner et al., 2019; 2020; Hansen et al., 2024). Such a model compresses when its code is narrower than the belief: a bounded-width recurrent state-space model, which runs in the control loop, can carry a code of dimension k below the belief-state dimension d.

Which k directions the code keeps is fixed by the prediction loss, while the control cost depends on only some belief directions. Nothing ties the two together: a code can predict well and still drop the directions the controller needed, an error only the control cost reveals.

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.

§3 identifies which belief directions an unweighted quadratic objective keeps, and which a weight built from the control cost keeps instead: the decision-relevant ones the cost weights most (Theorems 1 and 2; Figure 1). We assume linear predictive, decision and encoder maps, a given linear control law on the belief, finite-horizon quadratic cost, and an exact optimum of the population risk, the expected loss rather than a sample average. We also take the decision operator and the cost as known, where those models would have to estimate them; that case is open (§7). Linear-Gaussian generators make the predictive operator, the decision-irrelevant subspace and the capacity-k optimum exact, isolating the optimum's subspace from optimisation error and loss scale. That setting stands in for the models above rather than instantiating them: it keeps the capacity constraint and the link from belief directions to cost, and drops the nonlinearity. Every proof is machine-checked in Lean, axiom dependence recorded once for all five theorems, the leakage certificate among them: a bound on a near-optimal code's leakage into the decision-irrelevant subspace. **Relation to the literature.** Clauses 1–4 of Theorem 1 restate Ky Fan's maximum principle and Eckart–Young–Mirsky (Fan, 1949; Eckart & Young, 1936; Mirsky, 1960) for a predictive map's Gram operator (§3). The Gram operator is the map composed with its adjoint, and its eigenvalues rank directions by how much of each the map keeps. We did not find those combined with the decision weighting of §3, and claim no novelty for the mathematics. **Contributions.** §3 gives the condition that holds exactly when a weight on the prediction error reproduces the control cost. It then proves that a capacity-k code under that weight keeps only decision-relevant directions, up to the cost's rank, and gives an instance whose unweighted optimum keeps a decision-irrelevant one (Theorem 2). Clause 1 is Douglas's range-inclusion theorem (Douglas, 1966), machine-checked in belief space; what we add is its use as a realizability condition on a capacity-limited encoder. §A.3 removes the linear-encoder assumption and §A.4 decides before training whether the weight exists (Theorems 3 and 4). A training objective built from them is measured in §5 on linear-quadratic-Gaussian (LQG) belief estimation and on windowed belief identification, recovering the belief from past observations.

In LQG belief estimation (20-seed means), reweighting lowered decision-relevant error by 0.138 and 0.072 at $k{=}4$ in the two partially observed settings (§5). This error is the normalised, decision-weighted error of the code's estimate against the true state (Equation 3), above the filter's floors 0.711 and 0.660. At $k{=}2$ in the $d{=}16$ setting, decision-irrelevant energy, the share of what the code keeps that the cost ignores, falls from 0.707 to 0.00460. At $k{=}1$ the unweighted code does not beat $k{=}2$. The retention claim concerns an exact optimum; our runs illustrate rather than test it.

2 Related Work

Latent world models, joint-embedding architectures and decision-aware learning. Latent world models learn pixel dynamics (Hafner et al., 2019), fit returns (Oh et al., 2017), or embed a bisimulation metric (Zhang et al., 2021); the term *objective mismatch* refers to the likelihood-control gap (Lambert et al., 2020). Without disturbances, linear bisimulation collapses an invariant subspace in the output's kernel (Pappas, 2003), here inside $\ker D$, which the weighted code drops and the unweighted can keep (Theorem 2). Like JEPA (Assran et al., 2023), we predict in embedding space; Theorem 3 covers such losses with a fixed target and linear decoder, and Theorem 1 is the fixed-target counterpart of the linear theory of self-prediction (Tang et al., 2023; Voelcker et al., 2024; Littwin et al., 2024). Value-aware model learning replaces the likelihood by a value-prediction error (Farahmand et al., 2017); value equivalence equates models with matching Bellman updates (Grimm et al., 2020); control-aware (Nikishin et al., 2022) and decision-focused (Elmachtoub & Grigas, 2022) learning change the objective, and value-gradient weighting (Voelcker et al., 2022) reweights it, over a fixed model class; identification for control weights prediction error by the control objective, with sufficient conditions over frequency (Gevers, 1993; Zang et al., 1995). None says exactly when a weight through T reproduces the cost; Theorem 2 applies Douglas's classical condition there, and a witness shows reweighting is needed. Per-sample scalar weights (Lambert et al., 2020) leave the applied operator's rank at T's, above D's, so they cannot realize D (Appendix H.4). **Linear encoders, cost-weighted reduction and joint-embedding objectives.** The linear-autoencoder optimum is the discarded tail of the spectrum, also for nonlinear encoders (Bourlard & Kamp, 1988), and Theorem 1 is population reduced-rank regression (Izenman, 1975); Theorem 3 restates both for a general linear target, which Theorem 2 uses. Component-cost analysis ranks states by quadratic-cost contribution (Skelton, 1980); balanced truncation drops the least reachable and least visible states (Moore, 1981); frequency-weighted and goal-oriented reduction select by a performance band or a quantity of interest (Enns, 1984; Carlberg & Farhat, 2011); LQG

balancing ranks by closed-loop invariants (Jonckheere & Silverman, 1983), and Equation 1 weights an encoder by cost. Nearer still, optimal-projection reduced-order LQG (Hyland & Bernstein, 1984; Bernstein & Hyland, 1985) and rate-limited control (Tanaka et al., 2018) do constrain the estimator, by order or by information rate, but solve for the best one; we fix the objective and ask which directions its rank-k optimum keeps. Cost-driven LQG representation learning (Tian et al., 2023) recovers the cost-relevant state at $k = d$; Theorem 2 names the subspace retained at the population risk's exact optimum, bounding no control cost. Whitening (Ermolov et al., 2021) forces identity code covariance; VICReg (Bardes et al., 2022) penalises low variance, view disagreement and covariance: second moments constrained without a weighted target, which we add.

Appendix H.11 gives the full discussion, the information bottleneck included.

3 Theory and method

Setting. Vectors are columns; * is the adjoint, the transpose $^\top$ on the real matrices here. The state $x_t \in \mathbb{R}^d$ and the observation $o_t \in \mathbb{R}^{n_o}$ at time t follow the linear-Gaussian system $x_{t+1} = Ax_t + \epsilon_t$, $o_t = Cx_t + \eta_t$, with state matrix A, observation matrix C, isotropic Gaussian process noise ϵ_t and observation noise η_t. The Kalman belief mean is $\hat{x}_t = \mathbb{E}[x_t \mid o_{\le t}] \in \mathbb{R}^d$, given the observations $o_{\le t}$ up to time t. A fixed linear control law $u_t = -K_c\hat{x}_t$, with gain K_c given rather than optimised and input matrix B, closes the loop: $A_{\mathrm{cl}} = A - BK_c$. The per-step cost is $\|H\hat{x}_t\|^2 + \rho\|u_t\|^2$, with performance output H, the matrix whose rows read the state coordinates the cost depends on, and scalar control weight $\rho > 0$. The trajectories we simulate follow A with no action applied; B and K_c enter only through the closed loop $u_t = -K_c\hat{x}_t$ of the cost below. In the systems we generate (Appendix B), A is block diagonal with 2×2 blocks, independent two-dimensional subsystems. The cost reads the first coordinate of n_c of them, the *cost blocks*, and the control drives the second. So H and K_c are distinct $n_c \times d$ matrices whose rows each read one cost block's first coordinate: H is the cost's read-out and K_c the gain. We call n_c the *cost-block count*. **The decision operator.** Over a cost horizon t_h, the *cost-to-go matrix* $P_{t_h} = \sum_{t=0}^{t_h} (A_{\mathrm{cl}}^t)^\top \big(H^\top H + \rho\, K_c^\top K_c\big) A_{\mathrm{cl}}^t$ gives, for a belief error $v \in \mathbb{R}^d$, the difference between the belief and an estimate of it, the cost $v^\top P_{t_h} v$ the closed loop accumulates over steps 0 to t_h from v, noise aside. The *decision operator* is the matrix $D = \big[\, HA_{\mathrm{cl}}^t \,;\, \sqrt{\rho}\, K_c A_{\mathrm{cl}}^t \,\big]_{t=0}^{t_h}$ with d columns, the semicolon separating the two stacked blocks. Its Gram matrix is the cost-to-go, $D^*D = P_{t_h}$ (checked in Lean), so $\|Dv\|^2 = v^\top P_{t_h} v$. Write $\mu_0 \ge \cdots \ge \mu_{d-1}$ for the eigenvalues of P_{t_h}, indexed by i; they rank belief directions, unit vectors in $\mathbb{R}^d$, by cost. We call v *decision-irrelevant* when $Dv = 0$; these directions form the kernel $\ker D$, the zero-cost directions. These facts are classical (Anderson & Moore, 1990); related criteria in (Subramanian et al., 2022). **The predictive operator.** The target $y_m(t) = (o_{t+1}^\top, \ldots, o_{t+m}^\top)^\top \in \mathbb{R}^{mn_o}$ stacks the next m observations. The *predictive operator* $T : \mathbb{R}^d \to \mathbb{R}^{mn_o}$ is the linear map from belief to expected target, $T\hat{x}_t = \mathbb{E}[y_m(t) \mid o_{\le t}]$. Write $\lambda_0 \ge \lambda_1 \ge \cdots$ for the eigenvalues of T^*T and $b_0, b_1, \ldots$ for a matching orthonormal eigenbasis. The *predictive energy* of a belief direction u is $\|Tu\|^2$, the squared length of its prediction; an unweighted objective ranks belief directions by predictive energy, which the cost never enters. The code is $z \in \mathbb{R}^k$.

Theorem 1 (informal). Let the predictive operator T be linear on finite-dimensional real inner-product spaces, and let the encoder and decoder be linear with code dimension at most k; call such a pair a *capacity-k pair* and write X for its composite. Read every claim at the exact optimum of the population risk, an expectation over the distribution rather than a sample, and at identity belief covariance. What the clauses below say, at identity covariance: the eigenvalues of T^*T say how much of each belief direction the prediction preserves, and a code with k slots minimises prediction error by retaining the k the prediction preserves most; the cost never enters. **(1–2)** No orthonormal set of at most k belief directions carries more predictive energy than $\sum_{i<k} \lambda_i$, and $b_0, \ldots, b_{k-1}$ carry exactly that. **(3–4)** No capacity-k pair achieves risk below the discarded tail $\sum_{i \ge k} \lambda_i$, and the pair that encodes the k coordinates $\langle b_i, x\rangle$ and decodes through T attains it. **(5)** Given a spectral gap at the cut-off, $\lambda_k < \lambda_{k-1}$, with $0 < k < d$, every orthonormal set of at most k directions attaining the bound of clause 1 lies in the dominant k-dimensional eigenspace of T^*T, so the retained subspace is determined. **(6)** Under a belief covariance $\Sigma = RR^*$ (R a covariance factor), the risk becomes $\|(T - X)R\|_{\mathrm{HS}}^2$ and clauses 1–5 hold for the operator TR, whose Gram operator is $R^*(T^*T)R$. The formal statement is §A.1.

Theorem 2 (informal). Add a linear decision operator D to the setting of Theorem 1, with $\mu_0 \geq \mu_1 \geq \cdots$ the eigenvalues of D^*D. Clauses 1 and 3 below work in two steps: first a weight that makes prediction error measure cost, then the argument that picked the best-predicted directions picks the ones the cost ranks highest. **(1)** A positive semidefinite weight W with $T^*WT = D^*D$ exists exactly when $\ker T \subseteq \ker D$, *realizability*: T discards no direction D keeps. **(2)** If some v has $Tv = 0$ but $Dv \neq 0$, no linear W, positive semidefinite or not, gives $T^*WT = D^*D$. **(3)** Given such a W, $\mu_k < \mu_{k-1}$ and $\mu_{k-1} > 0$, every direction u the reweighted objective retains has $\|Du\|^2 \geq \mu_{k-1}$ and is orthogonal to $\ker D$, and every capacity-k pair with composite X attaining the reweighted optimum has $\ker D \subseteq \ker X$. **(4)** If an inexact weight gives a linear T' with $\|T'x\|^2 \leq \|Dx\|^2 + \varepsilon\|x\|^2$ for all x, with real slack ε (not the noise ϵ_t), and its spectrum ν has $\nu_k < \nu_{k-1}$, the directions u that T' retains have $\|Du\|^2 \geq \nu_{k-1} - \varepsilon$. **(5)** Reweighting can be necessary: in a two-dimensional *witness*, an explicit instance, whose predictive operator has zero kernel, the unweighted capacity-one optimum retains a direction D sends to zero, while the reweighted capacity-one optimum is orthogonal to it; Figure 4 draws it. The formal statement is §A.2.

The objective. We reweight the predictive target by a factor built from the control cost, so a capacity-limited encoder retains the directions the cost weights most. We train an encoder $\phi : \mathbb{R}^{wn_o} \to \mathbb{R}^k$, $z_t = \phi(o_{t-w+1}, \ldots, o_t)$, from the last w observations to the code, and a linear decoder $\Gamma : \mathbb{R}^k \to \mathbb{R}^q$, to minimise the mean over q output coordinates of a weighted prediction error,

$$\mathcal{L}(\phi, \Gamma) = \tfrac{1}{q}\, \mathbb{E}\, \big\|\Gamma z_t - V y_m(t)\big\|^2, \tag{1}$$

where the weight $V : \mathbb{R}^{mn_o} \to \mathbb{R}^q$ multiplies the target, not the residual, q is its row count, and w is the window length of Appendix B. The *unweighted* variant takes $V = I$, the identity on $\mathbb{R}^{mn_o}$, so $q = mn_o$. The *decision-weighted* variant takes $V = D\,T^+$, with T^+ the Moore–Penrose pseudoinverse of T, so q is the row count of D and $W = V^*V$ is the weight of Theorem 2 clause 1. Write $\tilde{T} = V \circ T : \mathbb{R}^d \to \mathbb{R}^q$ for the weighted predictive operator. Under realizability, $\ker T \subseteq \ker D$, $\tilde{T} = D$ exactly, so a belief error v has weighted prediction error $\|\tilde{T}v\|^2 = v^\top P_{t_h} v$, its cost-to-go: the loss, minimised over (ϕ, Γ), scores an error by what it costs, not by how large it is.

The leakage certificate (Theorem 5, informal). Theorem 5 is Theorem 2's relaxed companion: it keeps the weight equation $T^*WT = D^*D$ and relaxes attainment of the capacity-k optimum by a *tolerance* $\delta \geq 0$, the arm's suboptimality, by which its risk exceeds that optimum. Given δ, it returns the *certificate*

$$\eta \;=\; \frac{\delta}{\mu_{k-1} - \mu_k}, \tag{2}$$

a slack over a gap, which bounds, when $\mu_k < \mu_{k-1}$, the *frame weight*: the squared projection of the k retained unit directions onto the covariance-weighted eigendirections (Theorem 1, clause 6) below the cut-off, each retained direction contributing at most one. We read η as informative when $\eta < 1$, where it excludes a whole retained direction lying there, and weak otherwise. The theorem takes δ as given; our runs estimate it, so η is a per-run diagnostic, not a sample-complexity bound (Appendix D.8 gives an unused route). Appendix H.7 defines δ and μ (on an unweighted arm, the spectrum of TR, not DR), gives two reduced forms $\eta_{\mathrm{norm}}, \eta_{\mathrm{gf}}$ and the shrinkage they satisfy, and records that no theorem licenses either divisor against the frame weight of Theorem 5 clause 2.

Relation to the literature. The value-aware line (Farahmand et al., 2017; Farahmand, 2018; Grimm et al., 2020; Voelcker et al., 2022; Nikishin et al., 2022; Voelcker et al., 2025) prescribes a decision weight but no realizability condition. What we add is the condition's use here, its effect on the retained subspace, and a witness that containment fails without reweighting; the condition's algebra is classical. Theorem 2 clause 4 rests on the same spectral gap that classical perturbation bounds place in a denominator (Davis & Kahan, 1970). One arm violates realizability deliberately. Theorem 2 is the result the experiments are built around. **Note on assumptions.** Theorem 1's (A1) finite dimensions and (A2) linear predictive operator hold by construction; (A3) holds of the encoder in linear-encoder arms only, and Theorem 3 keeps the linear decoder and proves decoder linearity cannot be dropped (Appendix D.5). Theorem 4 makes realizability checkable in advance; we compute spectral gaps numerically. Theorems 1 and 2 read at that optimum, which finite training does

not attain. We report decision-relevant error

$$\frac{\mathbb{E}\big[(x-\tilde{x})^\top P_{t_h}(x-\tilde{x})\big]}{\mathbb{E}\big[x^\top P_{t_h} x\big]}, \tag{3}$$

invariant to every invertible affine recoding $z \mapsto Sz + c$ of the code, with $S \in \mathbb{R}^{k\times k}$ invertible and $c \in \mathbb{R}^k$. Here x is the true state and $\tilde{x}$ the least-squares affine fit of the belief mean from z, so the numerator carries the filter's stationary error as well as the encoder's. Δ is a per-seed difference of a metric between two arms, paired by seed. Our reweighted arms illustrate the containment $\ker D \subseteq \ker X$ at finite training but do not test it. We machine-verify Theorems 1 to 5 in Lean 4, the fifth being the leakage certificate above, with no `sorry`, opaque symbol or axiom beyond its three standard ones (`propext`, `Classical.choice`, `Quot.sound`). **Models and optimisation.** The weighting family uses linear and gated recurrent unit (GRU) encoders, the belief-identification family JEPA losses (unregularised, whitened, or VICReg, Bardes et al., 2022), contrastive and reconstruction losses. We train with Adam (Kingma & Ba, 2015) and evaluate the final checkpoint, not a selected one, on held-out trajectories by the metrics of §4; Appendix B gives widths and optimiser settings.

4 Experiments

4.1 Synthetic linear-Gaussian systems with known ground truth

We generate linear-Gaussian state-space systems with a fixed linear policy and quadratic cost, matching the generative assumptions of Theorems 1 and 2 exactly. We use three settings: d=16 with n_o=4 (belief rank 8), d=8 with n_o=3 (rank 6), and a fully observed d=8 system. The certificate sweep adds a fourth, larger cell at d=32, n_o=8, reported in Appendix H.7. The two partially observed settings range k over $\{1, 2, 3, 4, d\}$; the fully observed one runs three arms at $k \in \{2, 4\}$. Each system has cost horizon t_h=3 and two cost blocks, except in arms varying them; the generator matrices give P_{t_h} in closed form, the finite-horizon closed-loop cost Gramian of §3, an oracle quantity. We fix one generator rather than resampling per arm; Appendix B gives its transition and observation matrices, noise covariances and per-arm training budgets. At d=16, n_o=4, replacing its hand-chosen transition spectrum with twenty pre-registered draws leaves the drop in decision-irrelevant energy Holm-significant on all 20 at both binding capacities; on seven, where the two exact optima coincide at k=2, that drop is the optimisation path (Appendix H.17). On a nonlinear plant the weight still lowers decision-relevant error in all four cells and their four linear controls, degrading as σ_{obs} grows (Appendix H.18). **What we measure.** Our primary metric is decision-relevant error, Equation 3, read as an oriented delta so that positive means lower error: the P_{t_h}-weighted error of the affinely aligned prediction against the true state, over the state's own P_{t_h}-weighted cost-to-go, not a fraction above optimal cost. Clause 3 of Theorem 2 states a containment of kernels, not a cost ratio, so we measure containment directly. With $\hat{S}$ the range of the least-squares fit of the belief to the code (Appendix B), we report the fraction of retained energy in the decision-irrelevant subspace $\ker D$, *decision-irrelevant energy* below, and the minimum principal angle between $\hat{S}$ and $(\ker D)^\perp$. Both need the generator matrices, so neither is observable on real data; given a value function, a practitioner could approximate the error delta. **Baseline and its deltas.** The baseline is our own encoder with the reweighting removed: same architecture, same generator, k=2, t_h=3, two cost blocks. Arms differ from it in weighting, in one of k, encoder, t_h or cost blocks, or in both; Appendix B states the full configuration. The baseline fails in the way clause 5 of Theorem 2 predicts in advance: its unweighted capacity-limited optimum can be entirely decision-irrelevant. We vary the weighting (unweighted, reweighted, and a realizability-violating arm that moves a cost block onto an unobserved one), the encoder (linear or GRU), the cost horizon t_h, and the cost-block count. Weighting-matched contrasts hold the encoder fixed; the baseline family also pairs both GRU arms with the linear baseline. Every delta is formed within seed and reported with a BCa (bias-corrected and accelerated) bootstrap interval and Cohen's d_z, the paired mean delta over its standard deviation. Deltas here are candidate against baseline; the weighting-matched contrasts of §5 are candidate against candidate on the same 20 seeds. Arms at t_h=0, the cost-block arms and the violating arm are marked †: they change P_{t_h}, and so the decision problem, making them incomparable with the headline deltas. Holm–Bonferroni corrects within each setting, not across them: a *correction family* of K=15 candidates in each partially observed setting, of which 11 share the metric's denominator, and one of 3 under full observation.

A trained encoder satisfies neither the exact-optimum nor the population-risk antecedent, so these runs illustrate the theorem rather than test it. They can carry the certificate: we log η of Equation 2 on the trained encoder beside the quantity it bounds, how much of the retained subspace lies along the covariance-weighted eigendirections below k. A second route computes the same slack from the prediction residual, not the true belief. That removes one of the two places the belief is needed, not both; Appendix H.7 reports how far the routes disagree. We choose t_h and the cost-block count to bracket $\eta{=}1$ in earlier runs and k, in closed form, to keep the gap strict; their below-1 rate is a property of that design. Within this *weighting family* we compare only weighting variants of our objective; an input-side scalar and a value-equivalence target (Grimm et al., 2020) enter below.

One confound affects all reweighted arms above: $P_{t_h}^{1/2}$, its positive semidefinite square root, rescales the residual, so a shared step size gives the reweighted and unweighted arms different effective gradient magnitudes. We control it at $d{=}16, n_o{=}4$ and $k{=}2$, giving each arm its own step size over $\{1, 3, 9, 30\} \times 10^{-4}$ on the same 20 seeds. The two arms optimise at opposite ends of that grid, the arrangement least favourable to a shared value, and the contrast holds: positive at all four shared step sizes, and 0.136 with each arm at its own best against 0.123 at the shared 3×10^{-4} used elsewhere. The grid has a reweighted arm at $k{=}2$ alone; $k{=}4$ and the other settings have only Appendix H.4's trace-normalised control; and we chose each arm's best step size after seeing the grid rather than before, though all four weighting-matched comparisons are positive, so no choice changes the sign. Appendix H.5 gives the grid, the four matched deltas and the drift check. **Comparators.** On one generator and one seed set, the same family measures our output-side operator weight against two comparators: an input-side per-sample scalar $w(\hat{x}){=}\|P_{t_h}\hat{x}\|$, with a per-batch-normalised variant and a constant weight as a floor, and a value-equivalence arm regressing the decision target $D\hat{x}_t$ (Grimm et al., 2020). The value-equivalence arm is the best on decision-relevant error in every setting (Table 16); it shares the reweighted arm's *ceiling* (the metric at its loss's exact optimum, dashed in Figure 3) and retains nearly the same subspace, so it lies closer to the same ceiling, not at a better one. The per-sample scalar weights leave the applied operator's rank at T's, above D's, so they cannot realize D (Appendix H.4). Computed in closed form on the same generators, the ceiling of an axis-aligned (diagonal) weight recovers 76–92% of the operator weight's attainable gain, so our margin over the best axis-aligned weight this argument can price is the remaining 8–24%, not a qualitative separation; Table 17 gives the three ceilings this is read off. The comparison favours the diagonal by construction, since these generators place their cost blocks on the coordinate axes, which is the best case for an axis-aligned weight. Appendix H.4 carries the companion family of arms and every comparator with its interval, effect size and ceiling; its trace-normalised control removes the operator's scale and leaves the retained subspace where it was, so on this generator the effect is redirection, not step size. The weighting theory presumes the window carries the belief the weight acts on, so a second, the *belief-identification family*, asks whether a predictive objective recovers that belief at all. It asks whether windowed joint-embedding predictive architectures (LeCun, 2022; Assran et al., 2023) (JEPA) recover the history-inferable belief component at $d{=}16, n_o{=}8$ and $d{=}4, n_o{=}2$, measured by affine R^2 on unobserved factors,

$$R^2_{\perp} \;=\; \max_{\Theta,\,\theta_0} \; R^2(\Theta z_t + \theta_0,\; \Pi_{\perp}\hat{x}_t)\,,$$

where $\Pi_{\perp}$ projects onto the not-instantaneously-observable component of the Kalman belief mean. We draw data from the weighting family's process, so the generative-side premises again hold exactly. We report no mean correlation coefficient: these objectives fix the latent only up to the invertible linear map the maximisation ranges over, so component-wise identifiability is unattainable here. We take contrastive as the baseline and compare it against a recurrent autoencoder trained on reconstruction, a non-sequential encoder, a naive hybrid adding the contrastive term to a reconstruction loss, and three JEPA variants named by their regulariser (none, VICReg, whitening). Figure 8 traces training for five arms: the contrastive baseline, the three JEPA variants and the naive hybrid, which tracks the baseline almost exactly throughout.

4.2 Protocol

Every configuration uses 20 seeds, paired by seed, one seed setting both the data and the initialisation. We test oriented deltas two-sided, read direction from the sign of Δ, and correct within each baseline correction family by Holm–Bonferroni; the weighting-matched contrasts form their own family (Table 19). Appendix H.10 gives the family sizes, membership rule and interval and test conventions.

5 RESULTS

Decision-relevant error and the retained subspace. Decision-relevant error (↓) has a floor above zero, as no encoder removes the filter's stationary error: 0.711, 0.660 and 0.499 in the three settings (baselines in Table 21). Arms changing P_{t_h} move the floor, so we read each against its own. We also report affine R^2 after optimal alignment to the Kalman belief (↑, not observable on real data); retained-subspace quantities are unaligned. Deltas below are point estimates, Holm–Bonferroni-corrected within family; the table cited at each carries its interval and effect size. At the shared cost horizon, reweighted k=4 against the k=2 baseline lowers decision-relevant error by 0.205 at d=16, n_o=4 and 0.211 at d=8, n_o=3, and reweighted k=2 by 0.123 in both settings, all four significant (Table 21; Figure 5 draws the same contrast). The value-equivalence arm of §4 is lower still at the same ceiling (Table 16): its target $D\hat{x}_t$ is the conditional mean of our $Vy_m(t)$ and less noisy, but it needs the Kalman belief and so the noise covariances, while V needs only the system matrices and the cost. The reweighted k=4 Δs in Table 21 move weighting and capacity together, so we also form per-seed contrasts varying weighting alone at matched capacity, encoder and horizon: at k=4 they are 0.138 $[0.127, 0.150]$ and 0.072 $[0.062, 0.084]$. Such contrasts exist at every capacity and for the recurrent encoder (0.108); Table 19 carries all 20 with intervals and effect sizes, 19 separated from zero after correction; four are one contrast at four step sizes. Capacity does not dominate: at d=16, n_o=4 the two full-capacity arms sit 0.040 and 0.039 above the floor, and the matched 0.138 is the 0.173 between the two arms' exact optima less the 0.035 of optimisation gap the reweighted arm gives back (the split is derived in Appendix H.20). On the retained subspace at d=16, n_o=4, the reweighted arm lowers decision-irrelevant energy from the unweighted baseline's 0.707 to 0.00460 at k=2 (Figure 7) and to 0.038 at k=4, and the minimum principal angle from $42.6°$ to $2.2°$ and $0.6°$ (Tables 23 and 24); the k=4 energy means are carried by single high seeds, and the angle effect is weaker at d=8, n_o=3. Figure 2 plots each recovered latent against the decision direction it should track: on direction 0 the unweighted arm's scatter is a cloud and the reweighted arm's a band along the identity line. Averaged over 20 seeds, the unweighted arm spends its two-dimensional code on a direction the cost ignores, reaching $R^2 = 0.814$ on direction 4 where $\mu_4 = 0$ while holding 0.152 on the top decision direction, against 0.684 for the reweighted arm, which puts below 0.005 into every zero-weight direction (Figure 9). We read decision-irrelevant energy as the finite-sample surrogate for the containment $\ker D \subseteq \ker X$ that clause 3 of Theorem 2 concludes, since it vanishes exactly when containment holds, and the minimum angle as a diagnostic: small is necessary there, not sufficient. In a rank-2 variant of this cell, a stated error in the weight raises the energy from 0.040 to $0.101/0.172/0.244$ at realised error $0.024/0.078/0.231$ (Appendix H.14).

Horizon, capacity, the certificate and belief recovery. The short-horizon arm (t_h=0) enlarges $\ker P_{t_h}$, so improvement and confound share a cause; we report 0.270 and 0.296 unnormalised (Table 21, †). An unweighted t_h=0 partner separates them at d=16, n_o=4 (0.262), leaving d=8, n_o=3 confounded. Table 14 and Figure 3 cross weighting with capacity. Theorem 1 assumes a strict spectral gap below full capacity, so it does not cover k=d; there the constraint no longer binds and the effect nearly vanishes, the two marginal deltas nearly coinciding (0.218 reweighted against 0.217 unweighted). The matched contrast is 0.0012 at d=16, n_o=4 and 0.0004 at d=8, n_o=3 (Table 19), the second not separated from zero at n=20 and the first two orders of magnitude below the k=4 matched contrast; neither establishes independence of capacity and weighting. The spectral argument is classical (§3). Under full observation, reweighted k=4 reduces the error most, by 0.372, above reweighted k=2 (0.242) and unweighted k=4 (0.144): a weighting-matched contrast of 0.229 (Table 19). The unweighted capacity-1 encoder does not improve on the baseline at either setting: significantly *worse* at d=8, n_o=3 ($\Delta = -0.0503$), not separated from it at d=16, n_o=4; the capacity-2 recurrent encoder is at most marginal (Δ=0.012). The realizability-violating variant ($\ker T \not\subseteq \ker D$; one of two cost blocks moved onto an unobserved one, floor 0.850) sits 0.0774 below the baseline (†), from the block T still sees. On a realizable cell, grading the weight from none to realizing moves retention; no trend in recovery is detected (Appendix D.4). With no threshold fixed in advance for a failure of clause 2 of Theorem 2 on a trained run, this arm cannot separate realizability from any mechanism that merely grows with the weight. The leakage certificate η holds on all 1180 trained-encoder runs and is below 1 on 196 (17%); it exceeds the frame weight it bounds by a median 16 to 33×, which Appendix H.20 traces mostly to one derivation step, not estimator noise. Dividing by $k\mu_{k-1}$ takes the below-1 count to 779 of the 1180 runs (66%), but no theorem licenses that divisor: a diagnostic rate, not a second certificate. Our runs estimate Theorem 5's δ; 40 of the 1180 are GRU runs, outside its linear encoder (§3). Whitening is the best JEPA regulariser

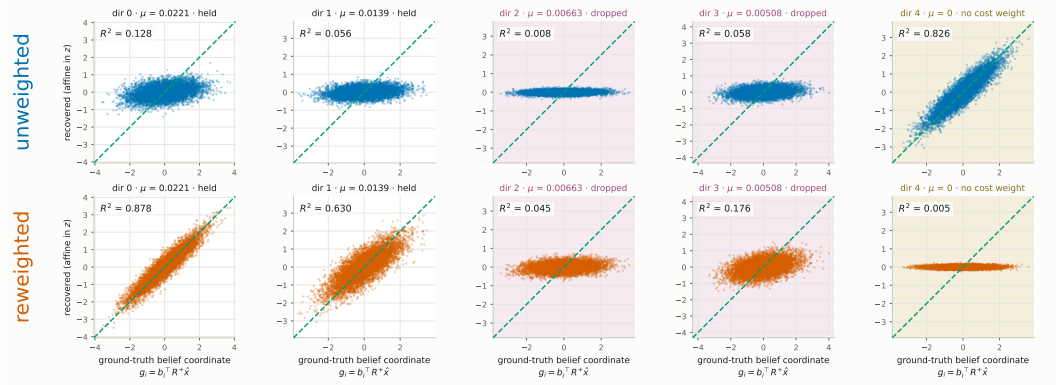
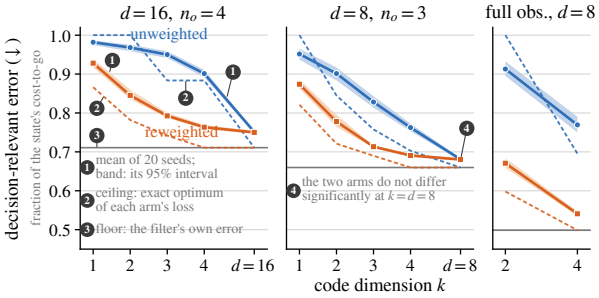


Figure 2: Recovered latents against ground-truth decision directions at d=16, n_o=4, capacity k=2: one point per held-out evaluation anchor, from a single seed. The dashed green line is the ground truth, where a perfect affine readout would lie. Rows are the unweighted and the reweighted arm, columns the four directions the cost weights, in order of μ_i, the eigenvalues of $R^* P_{t_h} R$ with $\Sigma = RR^*$, not of P_{t_h} (Theorem 1, clause 6), and then one it does not weight at all. Of the four, the weighted capacity-2 optimum retains two: the two on the pink background are dropped there, so their low R^2 is what the theorem predicts and not a shortfall. The weighted arm scores higher on them on 16 and 14 of 20 seeds: its code lies inside the span of all four weighted directions, the unweighted code largely outside. The fifth, on the cream background, carries no cost weight, so nothing is asked of the encoder there; the unweighted arm recovers it better than either direction the cost ranks highest, while the weighted arm puts almost nothing there. Figure 9 repeats these panels with affine R^2 per direction over all 20 seeds, where each arm's split between directions varies from seed to seed.

at d=16, n_o=8 (+0.370), VICReg at d=4, n_o=2 (+0.255), where seed variance is much larger and intervals overlap (Table 20; training curves in Figure 8). The non-sequential MLP encoder is significantly worse than the baseline; its extreme d_z records reproducibility, not effect magnitude (Figure 10).



Fitted A, C (H.15)		*share of gain kept*
rank given, 1 trajectory		0.976
rank estimated, 1 trajectory		0.500
rank estimated, 16–128		0.881 (at 64) to 1.000
pseudo-inverse, 1, 8, 128		−0.23
Nonlinear plant (H.18)	*unw.*	*rew.*
noise $\sigma_{\text{obs}} = 0.3$	0.595	0.091
1.0	0.676	0.298
2.0	0.776	0.558
3.0	0.868	0.746
Random A (H.17)		*draws lowered*
$k = 2$ and $k = 4$		20/20
$k = 16$, full capacity		0/20

Figure 3: Left: decision-relevant error against k in the LQG settings. Where the unweighted arm sits below its own ceiling, its loss's exact optimum retains only decision-irrelevant directions (ceiling 1.0, Table 16), so reaching it would raise the plotted error. Right: one idealisation relaxed per block, appendix section in parentheses: a weight from state-fitted maps, as the share of the exact weight's gain it keeps (at 64 the rank rule over-picks on two seeds); a pendulum plant with no action applied and its own floor, weighted by its linearisation, as decision-relevant error by observation noise; and twenty drawn transition matrices, as the draws where reweighting significantly lowers decision-irrelevant energy.

6 Discussion

Reweighting lowered decision-relevant error under partial and full observation in our linear-Gaussian family. This points to where the weighted target puts the loss, not to information available at decision time. Our runs test no theorem: each fixes an exact optimum or a population risk we only estimate. They carry the leakage certificate of §3 (Theorem 5) instead, computable on a trained encoder and never violated against the frame weight it bounds. The certificate has two forms. The frame form concerns the directions the encoder retains; the pair form concerns the composite X and

needs a bound on $\|Xv\|/\|v\|$ logged in two cells only, so we instantiate the frame form. There the pair form is loose by a median 85 to $778\times$ (Appendix H.19). We take from value equivalence only its realizability requirement, that the model reproduce the decision target (Appendix H.4). That input reweighting cannot realize D holds only on nondegenerate configurations, those whose reweighting keeps the belief covariance's rank: no such input reweighting makes the reweighted Gram operator equal the decision Gramian $P_{t_h} = D^*D$ here. The qualifier matters: a constructed *degenerate* reweighting does make them equal, and moves the exact-optimum ceiling (Appendix D.9). Whether a *trained* reweighting ever reaches that case is untested, not refuted: none of ours reweights the input except by a per-sample scalar, which keeps the belief covariance's rank in all three settings. A direct per-seed, per-coordinate test of the retained subspace against $\ker D$ is open (Appendix H.13).

Sequential encoding separates the arms: the non-sequential encoder's affine R^2 on unobserved factors is 0.00174 and 0.00051 in the two settings, against 0.31 and 0.50 for the baseline, and it loses on every seed. Factors identified only through accumulated effect across timesteps stay unrecoverable without history.

7 Limitations

Our scope is linear-Gaussian. We report no result on a continuous-control benchmark such as the DeepMind Control Suite (Tassa et al., 2018); we ran four such sweeps (Appendix D.2), of per-sample scalar weights through nonlinear decoders rather than our weight, and no number here draws on them. What we measure is containment, against the matrix P_{t_h} that scores a belief error by the cost-to-go it implies. That matrix exists only where the cost is quadratic in a known state with known dynamics; our generator supplies it exactly, while a benchmark task gives only a scalar reward from a simulator. On such a task there is nothing to measure containment against. Theorem 3 drops encoder linearity but shows decoder linearity cannot be dropped (clause 5), while keeping a linear decision map, finite support and an exact optimum. The guarantee therefore does not reach the nonlinear decoders and dynamics of standard latent world models. On one cell at one width the weight's gain survives an MLP decoder (Appendix D.6). The nearest test we could run is a nonlinear plant, and the weight still works there (Figure 3; Appendix H.18). An estimated weight is still the open case, but a narrower one: a stated error in the weight degrades containment, on both tests at the two larger sizes (Appendix H.14). On one plant, cost and controller granted, a weight fitted from one trajectory of measured states keeps 97.6% of the exact weight's benefit given the rank of T, and from 128 is not separated from it with the rank cut at a spectral gap; training on a naive pseudo-inverse does worse than no weight (Figure 3; Appendix H.15). Training on the exact weight assumes T registers every direction D weights well above the observation noise: $V = DT^+$ divides by T's singular values, so the weighted target would otherwise be mostly noise.

8 Conclusion

We reweighted a capacity-limited predictive objective toward decision-relevant directions. The guarantees hold for linear maps at a population risk's exact optimum, under a fixed linear feedback law and finite horizon, on Lean's three standard axioms alone (§3). Clauses 1–4 of Theorem 1 restate Eckart–Young–Mirsky and Ky Fan's maximum principle (Eckart & Young, 1936; Mirsky, 1960; Fan, 1949), and we reduce the finite-horizon cost to a standard observability condition (Kwakernaak & Sivan, 1972) (Appendix A.2). What we add is putting the classical condition $\ker T \subseteq \ker D$ to work on a capacity-limited encoder, with a witness showing reweighting necessary there.

In our linear-Gaussian systems, matched contrasts lower decision-relevant error by 0.138 and 0.072 under partial observation at the shared cost horizon and by 0.229 under full observation, each paired across 20 seeds (Table 19). Against the shared baseline, our unweighted encoder at k=2, reweighting and the larger code together reduce the error by 0.205, 0.211 and 0.372, covering 80%, 88% and 90% of the range to the metric's floor. Appendix H.4 compares arms modelled on the value-aware literature (Grimm et al., 2020): a value-equivalence arm scores better from inside the reweighted arm's ceiling, and the scalar weights cannot realize D. We train finitely, so our runs illustrate Theorems 1 and 2, not confirm them. Containment need not hold under approximate optimisation; the Lean statements fix what one would have to generalise.

REPRODUCIBILITY STATEMENT

Code. Each experiment is a self-contained package with its own training entry point and YAML configuration; every sweep reported here writes a manifest, one JSONL log per seed, and the statistical analysis for every comparison. Seed i of a configuration is a fixed offset from a base seed, so the same configuration reruns identically and configurations within a sweep are paired seed by seed. Every comparison in the paper is analysed the same way: a paired t-test on the per-seed differences and a Wilcoxon signed-rank test, with Cohen's d_z as the effect size. Intervals are 95% BCa bootstrap intervals over 10000 resamples. Where a sweep compares several configurations against one baseline, p-values carry a Holm-Bonferroni correction across that family. A Shapiro-Wilk test is computed for each comparison so the parametric test's assumption can be checked. An anonymized copy of the code and configurations behind every number in this paper is attached as supplementary material; the public release follows with the paper.

Theory. All 5 theorems the paper states are machine-verified in Lean 4. Each is stated with its assumptions labelled, and the appendix gives the corresponding proof, the name of the verified declaration, and the axioms it depends on. No development declares an axiom or an opaque symbol beyond Lean's standard three (propositional extensionality, the axiom of choice, and soundness of quotients). Where the underlying mathematics is classical, the appendix says so and identifies which parts are and are not a contribution here. The verification covers the theorem statements only: no run reported here satisfies a premise about an exact optimum, so the experiments illustrate the theorems rather than establishing them.

Datasets. Every number in the paper comes from 38 sweeps across 17 experiments, where an experiment is one self-contained codebase and a sweep is one grid of configurations run over it. Each configuration was run with 20 seeds, paired across configurations within a sweep. 469 candidate configurations were compared in total, each against its sweep's own baseline – 507 trained configurations in all, counting one baseline per sweep; the appendix gives the parameters of each experiment family, and the code carries each sweep's own configuration. No external dataset is used: the data are generated by the simulators and generative models described in the experimental setup, so a reproduction needs the code and the seeds and nothing else.

Compute. The results in this paper cost 18.9 days of wall-clock time on NVIDIA RTX 6000 Ada Generation GPUs, measured from the recorded event logs across 38 of 38 sweeps, covering 10140 completed runs. These compute requirements are elapsed wall-clock time: the event logs record when each sweep started and finished and which device each run used, not how long each device was held. Runs were executed concurrently across up to 2 devices of one machine, so the figure above is elapsed time and not a device-hour total. One 7-hour control-suite sweep ran on a second machine, with NVIDIA Quadro RTX 8000 GPUs. Of these totals, the four control-suite sweeps of Appendix D.2, 440 runs and 40.2 hours, feed no number in the main text. All runs used PyTorch 2.4.1, CUDA 12.1, Python 3.11.9 inside a pinned container image, with PyTorch's deterministic algorithms enabled in every run, on hosts whose CUDA driver reported 13.2.

AI USE STATEMENT

In this work, we used generative AI tools for the following tasks: help develop theoretical models or conceptual frameworks; formulate mathematical claims; provide critical ingredients for proving mathematical claims; assist in the writing of proofs; propose or refine hypotheses; design or provide feedback on research methodology or experiments; implement methods; support qualitative and thematic data analysis; and interpret results. We have not used generative AI tools for the following tasks: generate synthetic data sets. The following tasks requiring disclosure are not applicable to this work: assist with translation; and clean and reformat dataset. Additionally, we used generative AI tools for the following tasks, where disclosure is recommended: create or modify scientific figures; create or edit software code; creation of artifacts; draft parts of a research paper; summarize or analyse existing literature; discover research topics or identify gaps; brainstorming; sourcing/searching for information; edit a research paper to improve readability; identify relevant literature; format references; suggest a structure for a research paper; and propose a title or keywords. The work was done with coding agents built on several large language models: Anthropic's Claude Opus 4.6, 4.8,

5 and 5.5, Claude Fable 5 and 5.1, and Claude Sonnet 5; Google's Gemini 3.1 Pro; and OpenAI's GPT-5.6 Sol, GPT-5.6 Terra and GPT-6 Astra. Claude models carried most of the work at every stage, including the Lean proofs, the figures, and the drafting and revision of the paper. Gemini wrote and analysed the belief-identification experiment, and GPT-5.6 Sol and Terra supported literature review and the implementation and auditing of experiments. Several of these systems, GPT-6 Astra among them, also gave feedback on drafts.

We have reviewed all AI-assisted work reported here and take responsibility for the final content of this work, including text, claims and artifacts produced with the aid of generative AI. The theorem statements are reproduced from Lean sources whose proofs are machine-checked, and the reported statistics were checked against the recorded experiment results.

REFERENCES

Romina Abachi, Mohammad Ghavamzadeh, and Amir-massoud Farahmand. Policy-aware model learning for policy gradient methods, 2020. arXiv preprint arXiv:2003.00030.

Alexander A. Alemi, Ian Fischer, Joshua V. Dillon, and Kevin Murphy. Deep variational information bottleneck. In *International Conference on Learning Representations*, 2017.

Brian D. O. Anderson and John B. Moore. *Optimal control: linear quadratic methods*. Prentice Hall, Englewood Cliffs, NJ, 1990.

Mahmoud Assran, Quentin Duval, Ishan Misra, Piotr Bojanowski, Pascal Vincent, Michael G. Rabbat, Yann LeCun, and Nicolas Ballas. Self-supervised learning from images with a joint-embedding predictive architecture. In *Computer Vision and Pattern Recognition*, pp. 15619–15629, 2023. doi: 10.1109/CVPR52729.2023.01499.

Alex Ayoub, Zeyu Jia, Csaba Szepesvári, Mengdi Wang, and Lin F. Yang. Model-based reinforcement learning with value-targeted regression. In *International Conference on Machine Learning*, volume 119, pp. 463–474, 2020.

Jerzy K. Baksalary. Nonnegative definite and positive definite solutions to the matrix equation $AXA^* = B$. *Linear and Multilinear Algebra*, 16(1–4):133–139, 1984. doi: 10.1080/03081088408817616.

Pierre Baldi and Kurt Hornik. Neural networks and principal component analysis: Learning from examples without local minima. *Neural Networks*, 2(1):53–58, 1989. doi: 10.1016/0893-6080(89)90014-2.

Adrien Bardes, Jean Ponce, and Yann LeCun. VICReg: Variance-invariance-covariance regularization for self-supervised learning. In *International Conference on Learning Representations*, 2022.

Adrien Bardes, Quentin Garrido, Jean Ponce, Xinlei Chen, Michael Rabbat, Yann LeCun, Mido Assran, and Nicolas Ballas. Revisiting feature prediction for learning visual representations from video. *Transactions on Machine Learning Research*, 2024. URL `https://openreview.net/forum?id=QaCCuDfBk2`.

Peter Benner, Pawan Goyal, and Igor Pontes Duff. Gramians, energy functionals, and balanced truncation for linear dynamical systems with quadratic outputs. *IEEE Transactions on Automatic Control*, 67(2):886–893, 2022. doi: 10.1109/TAC.2021.3086319.

Dennis S. Bernstein and David C. Hyland. The optimal projection equations for reduced-order state estimation. *IEEE Transactions on Automatic Control*, 30(6):583–585, 1985. doi: 10.1109/TAC.1985.1104001.

Byron Boots, Sajid M. Siddiqi, and Geoffrey J. Gordon. Closing the learning-planning loop with predictive state representations. *The International Journal of Robotics Research*, 30(7):954–966, 2011. doi: 10.1177/0278364911404092.

H. Bourlard and Y. Kamp. Auto-association by multilayer perceptrons and singular value decomposition. *Biological Cybernetics*, 59(4–5):291–294, 1988. doi: 10.1007/BF00332918.

Kevin Carlberg and Charbel Farhat. A low-cost, goal-oriented 'compact proper orthogonal decomposition' basis for model reduction of static systems. *International Journal for Numerical Methods in Engineering*, 86(3):381–402, 2011. doi: 10.1002/nme.3074.

Pablo Samuel Castro. Scalable methods for computing state similarity in deterministic Markov decision processes. In *AAAI Conference on Artificial Intelligence*, volume 34, pp. 10069–10076, 2020. doi: 10.1609/aaai.v34i06.6564.

Gal Chechik, Amir Globerson, Naftali Tishby, and Yair Weiss. Information bottleneck for Gaussian variables. *Journal of Machine Learning Research*, 6:165–188, 2005.

Ting Chen, Calvin Luo, and Lala Li. Intriguing properties of contrastive losses. In *Neural Information Processing Systems*, volume 34, pp. 11834–11845, 2021.

Felix Creutzig, Amir Globerson, and Naftali Tishby. Past-future information bottleneck in dynamical systems. *Physical Review E*, 79(4):041925, 2009. doi: 10.1103/PhysRevE.79.041925.

Chandler Davis and W. M. Kahan. The rotation of eigenvectors by a perturbation. III. *SIAM Journal on Numerical Analysis*, 7(1):1–46, 1970. 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. doi: 10.1007/s10208-019-09426-y.

Priya Donti, Brandon Amos, and J. Zico Kolter. Task-based end-to-end model learning in stochastic optimization. In *Neural Information Processing Systems*, volume 30, 2017.

Pierluca D'Oro, Alberto Maria Metelli, Andrea Tirinzoni, Matteo Papini, and Marcello Restelli. Gradient-aware model-based policy search. In *AAAI Conference on Artificial Intelligence*, volume 34, pp. 3801–3808, 2020. doi: 10.1609/aaai.v34i04.5791.

R. G. Douglas. On majorization, factorization, and range inclusion of operators on Hilbert space. *Proceedings of the American Mathematical Society*, 17(2):413–415, 1966. doi: 10.1090/S0002-9939-1966-0203464-1.

Carl Eckart and Gale Young. The approximation of one matrix by another of lower rank. *Psychometrika*, 1(3):211–218, 1936. doi: 10.1007/BF02288367.

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

Dale F. Enns. Model reduction with balanced realizations: An error bound and a frequency weighted generalization. In *The 23rd IEEE Conference on Decision and Control*, pp. 127–132, 1984. doi: 10.1109/CDC.1984.272286.

Aleksandr Ermolov, Aliaksandr Siarohin, Enver Sangineto, and Nicu Sebe. Whitening for self-supervised representation learning. In *International Conference on Machine Learning*, volume 139, pp. 3015–3024, 2021. URL https://proceedings.mlr.press/v139/ermolov21a.html.

Benjamin Eysenbach, Alexander Khazatsky, Sergey Levine, and Ruslan Salakhutdinov. Mismatched no more: Joint model-policy optimization for model-based RL. In *Neural Information Processing Systems*, volume 35, pp. 23230–23243, 2022. doi: 10.52202/068431-1688.

Ky Fan. On a theorem of Weyl concerning eigenvalues of linear transformations I. *Proceedings of the National Academy of Sciences*, 35(11):652–655, 1949. doi: 10.1073/pnas.35.11.652.

Amir-massoud Farahmand. Iterative value-aware model learning. In *Neural Information Processing Systems*, volume 31, 2018.

Amir-massoud Farahmand, André Barreto, and Daniel Nikovski. Value-aware loss function for model-based reinforcement learning. In *International Conference on Artificial Intelligence and Statistics*, volume 54, pp. 1486–1494, 2017.

Norm Ferns, Prakash Panangaden, and Doina Precup. Metrics for finite Markov decision processes. In *Conference on Uncertainty in Artificial Intelligence*, pp. 162–169, 2004. URL https://dl.acm.org/doi/10.5555/1036843.1036863.

Xiang Fu, Ge Yang, Pulkit Agrawal, and Tommi S. Jaakkola. Learning task informed abstractions. In *International Conference on Machine Learning*, volume 139, pp. 3480–3491, 2021. URL http://proceedings.mlr.press/v139/fu21b.html.

Carles Gelada, Saurabh Kumar, Jacob Buckman, Ofir Nachum, and Marc G. Bellemare. DeepMDP: Learning continuous latent space models for representation learning. In *International Conference on Machine Learning*, volume 97, pp. 2170–2179, 2019.

Michel Gevers. Towards a joint design of identification and control? In H. L. Trentelman and J. C. Willems (eds.), *Essays on Control: Perspectives in the Theory and its Applications*, pp. 111–151. Birkhäuser, Boston, MA, 1993. doi: 10.1007/978-1-4612-0313-1_5.

Christopher Grimm, André Barreto, Satinder Singh, and David Silver. The value equivalence principle for model-based reinforcement learning. In *Neural Information Processing Systems*, volume 33, pp. 5541–5552, 2020.

Danijar Hafner, Timothy Lillicrap, Ian Fischer, Ruben Villegas, David Ha, Honglak Lee, and James Davidson. Learning latent dynamics for planning from pixels. In *International Conference on Machine Learning*, volume 97, pp. 2555–2565, 2019.

Danijar Hafner, Timothy Lillicrap, Jimmy Ba, and Mohammad Norouzi. Dream to control: Learning behaviors by latent imagination. In *International Conference on Learning Representations*, 2020.

Danijar Hafner, Timothy P. Lillicrap, Mohammad Norouzi, and Jimmy Ba. Mastering Atari with discrete world models. In *International Conference on Learning Representations*, 2021.

Danijar Hafner, Jurgis Pasukonis, Jimmy Ba, and Timothy Lillicrap. Mastering diverse control tasks through world models. *Nature*, 640(8059):647–653, 2025. doi: 10.1038/s41586-025-08744-2.

Nicklas Hansen, Hao Su, and Xiaolong Wang. Temporal difference learning for model predictive control. In *International Conference on Machine Learning*, volume 162, pp. 8387–8406, 2022. URL https://proceedings.mlr.press/v162/hansen22a.html.

Nicklas Hansen, Hao Su, and Xiaolong Wang. TD-MPC2: Scalable, robust world models for continuous control. In *International Conference on Learning Representations*, pp. 47376–47405, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/cf73d57b6dcda32b293df7c2d5341f49-Abstract-Conference.html.

Yingbo Hua, Maziar Nikpour, and Petre Stoica. Optimal reduced-rank estimation and filtering. *IEEE Transactions on Signal Processing*, 49(3):457–469, 2001. doi: 10.1109/78.905856.

David C. Hyland and Dennis S. Bernstein. The optimal projection equations for fixed-order dynamic compensation. *IEEE Transactions on Automatic Control*, 29(11):1034–1037, 1984. doi: 10.1109/TAC.1984.1103418.

David C. Hyland and Dennis S. Bernstein. The optimal projection equations for model reduction and the relationships among the methods of Wilson, Skelton, and Moore. *IEEE Transactions on Automatic Control*, 30(12):1201–1211, 1985. doi: 10.1109/TAC.1985.1103865.

Alan Julian Izenman. Reduced-rank regression for the multivariate linear model. *Journal of Multivariate Analysis*, 5(2):248–264, 1975. doi: 10.1016/0047-259X(75)90042-1.

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. doi: 10.1109/TAC.1983.1103159.

C. G. Khatri and Sujit Kumar Mitra. Hermitian and nonnegative definite solutions of linear matrix equations. *SIAM Journal on Applied Mathematics*, 31(4):579–585, 1976. doi: 10.1137/0131050.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations*, 2015.

Huibert Kwakernaak and Raphael Sivan. *Linear Optimal Control Systems*. Wiley-Interscience, New York, 1972.

Alex Lamb, Riashat Islam, Yonathan Efroni, Aniket Rajiv Didolkar, Dipendra Misra, Dylan J. Foster, Lekan P. Molu, Rajan Chari, Akshay Krishnamurthy, and John Langford. Guaranteed discovery of control-endogenous latent states with multi-step inverse models. *Transactions on Machine Learning Research*, 2023. URL https://openreview.net/forum?id=TNocbXm5MZ.

Nathan O. Lambert, Brandon Amos, Omry Yadan, and Roberto Calandra. Objective mismatch in model-based reinforcement learning. In *Conference on Learning for Dynamics and Control*, volume 120, pp. 761–770, 2020. URL http://proceedings.mlr.press/v120/lambert20a.html.

Wallace E. Larimore. Canonical variate analysis in identification, filtering, and adaptive control. In *29th IEEE Conference on Decision and Control*, pp. 596–604, 1990. doi: 10.1109/CDC.1990.203665.

Yann LeCun. A path towards autonomous machine intelligence. OpenReview preprint, 2022. URL https://openreview.net/forum?id=BZ5a1r-kVsf. Version 0.9.2, 27 June 2022.

Kimin Lee, Younggyo Seo, Seunghyun Lee, Honglak Lee, and Jinwoo Shin. Context-aware dynamics model for generalization in model-based reinforcement learning. In *International Conference on Machine Learning*, volume 119, pp. 5757–5766, 2020.

Michael L. Littman, Richard S. Sutton, and Satinder Singh. Predictive representations of state. In *Neural Information Processing Systems*, volume 14, 2001. URL https://proceedings.neurips.cc/paper_files/paper/2001/hash/1e4d36177d71bbb3558e43af9577d70e-Abstract.html.

Etai Littwin, Omid Saremi, Madhu Advani, Vimal Thilak, Preetum Nakkiran, Chen Huang, and Joshua Susskind. How JEPA avoids noisy features: The implicit bias of deep linear self distillation networks. In *Neural Information Processing Systems*, volume 37, pp. 91300–91336, 2024. doi: 10.52202/079017-2897.

Jayanta Mandi, James Kotary, Senne Berden, Maxime Mulamba, Víctor Bucarey, Tias Guns, and Ferdinando Fioretto. Decision-focused learning: Foundations, state of the art, benchmark and future opportunities. *Journal of Artificial Intelligence Research*, 80:1623–1701, 2024. doi: 10.1613/jair.1.15320.

Horia Mania, Stephen Tu, and Benjamin Recht. Certainty equivalence is efficient for linear quadratic control. In *Neural Information Processing Systems*, volume 32, 2019.

L. Mirsky. Symmetric gauge functions and unitarily invariant norms. *The Quarterly Journal of Mathematics*, 11(1):50–59, 1960. doi: 10.1093/qmath/11.1.50.

Bruce C. Moore. Principal component analysis in linear systems: Controllability, observability, and model reduction. *IEEE Transactions on Automatic Control*, 26(1):17–32, 1981. doi: 10.1109/TAC.1981.1102568.

Anusha Nagabandi, Ignasi Clavera, Simin Liu, Ronald S. Fearing, Pieter Abbeel, Sergey Levine, and Chelsea Finn. Learning to adapt in dynamic, real-world environments through meta-reinforcement learning. In *International Conference on Learning Representations*, 2019.

Suraj Nair, Silvio Savarese, and Chelsea Finn. Goal-aware prediction: Learning to model what matters. In *International Conference on Machine Learning*, volume 119, pp. 7207–7219, 2020. URL https://proceedings.mlr.press/v119/nair20a.html.

Evgenii Nikishin, Romina Abachi, Rishabh Agarwal, and Pierre-Luc Bacon. Control-oriented model-based reinforcement learning with implicit differentiation. In *AAAI Conference on Artificial Intelligence*, volume 36, pp. 7886–7894, 2022. doi: 10.1609/aaai.v36i7.20758.

Junhyuk Oh, Satinder Singh, and Honglak Lee. Value prediction network. In *Neural Information Processing Systems*, volume 30, 2017.

George J. Pappas. Bisimilar linear systems. *Automatica*, 39(12):2035–2047, 2003. doi: 10.1016/j.automatica.2003.07.003.

R. Penrose. A generalized inverse for matrices. *Mathematical Proceedings of the Cambridge Philosophical Society*, 51(3):406–413, 1955. doi: 10.1017/S0305004100030401.

Pascal Poupart and Craig Boutilier. Value-directed compression of POMDPs. In *Neural Information Processing Systems*, volume 15, pp. 1547–1554, 2002.

Roland Pulch and Akil Narayan. Balanced truncation for model order reduction of linear dynamical systems with quadratic outputs. *SIAM Journal on Scientific Computing*, 41(4):A2270–A2295, 2019. doi: 10.1137/17M1148797.

Gregory C. Reinsel and Raja P. Velu. *Multivariate Reduced-Rank Regression: Theory and Applications*, volume 136 of *Lecture Notes in Statistics*. Springer, 1998. doi: 10.1007/978-1-4757-2853-8.

Daniel E. Rivera, I. F. Pollard, and C. E. García. Control-relevant prefiltering: a systematic design approach and case study. *IEEE Transactions on Automatic Control*, 37(7):964–974, 1992. doi: 10.1109/9.148348.

Joshua Robinson, Li Sun, Ke Yu, Kayhan Batmanghelich, Stefanie Jegelka, and Suvrit Sra. Can contrastive learning avoid shortcut solutions? In *Neural Information Processing Systems*, volume 34, pp. 4974–4986, 2021.

Julian Schrittwieser, Ioannis Antonoglou, Thomas Hubert, Karen Simonyan, Laurent Sifre, Simon Schmitt, Arthur Guez, Edward Lockhart, Demis Hassabis, Thore Graepel, Timothy P. Lillicrap, and David Silver. Mastering Atari, Go, chess and shogi by planning with a learned model. *Nature*, 588(7839):604–609, 2020. doi: 10.1038/s41586-020-03051-4.

Ehsan Shafieepoorfard, Maxim Raginsky, and Sean P. Meyn. Rationally inattentive control of Markov processes. *SIAM Journal on Control and Optimization*, 54(2):987–1016, 2016. doi: 10.1137/15M1008476.

Robert E. Skelton. Cost decomposition of linear systems with application to model reduction. *International Journal of Control*, 32(6):1031–1055, 1980. doi: 10.1080/00207178008910039.

Alessio Spantini, Tiangang Cui, Karen Willcox, Luis Tenorio, and Youssef Marzouk. Goal-oriented optimal approximations of Bayesian linear inverse problems. *SIAM Journal on Scientific Computing*, 39(5):S167–S196, 2017. doi: 10.1137/16M1082123.

Jayakumar Subramanian, Amit Sinha, Raihan Seraj, and Aditya Mahajan. Approximate information state for approximate planning and reinforcement learning in partially observed systems. *Journal of Machine Learning Research*, 23(12):1–83, 2022.

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. doi: 10.1109/TAC.2017.2709618.

Yunhao Tang, Zhaohan Daniel Guo, Pierre Harvey Richemond, Bernardo Ávila Pires, Yash Chandak, Rémi Munos, Mark Rowland, Mohammad Gheshlaghi Azar, Charline Le Lan, Clare Lyle, András György, Shantanu Thakoor, Will Dabney, Bilal Piot, Daniele Calandriello, and Michal Valko. Understanding self-predictive learning for reinforcement learning. In *International Conference on Machine Learning*, volume 202, pp. 33632–33656, 2023. URL `https://proceedings.mlr.press/v202/tang23d.html`.

Yuval Tassa, Yotam Doron, Alistair Muldal, Tom Erez, Yazhe Li, Diego de Las Casas, David Budden, Abbas Abdolmaleki, Josh Merel, Andrew Lefrancq, Timothy Lillicrap, and Martin Riedmiller. DeepMind Control Suite, 2018. arXiv preprint arXiv:1801.00690.

Yi Tian, Kaiqing Zhang, Russ Tedrake, and Suvrit Sra. Can direct latent model learning solve linear quadratic Gaussian control? In *Learning for Dynamics and Control Conference*, volume 211, pp. 51–63, 2023. URL https://proceedings.mlr.press/v211/tian23a.html.

Naftali Tishby, Fernando C. Pereira, and William Bialek. The information bottleneck method. In *Proceedings of the 37th Annual Allerton Conference on Communication, Control and Computing*, pp. 368–377, 1999. arXiv:physics/0004057.

Tobias Uelwer, Jan Robine, Stefan Sylvius Wagner, Marc Höftmann, Eric Upschulte, Sebastian Konietzny, Maike Behrendt, and Stefan Harmeling. A survey on self-supervised methods for visual representation learning. *Machine Learning*, 114(4):111, 2025. doi: 10.1007/s10994-024-06708-7.

Arjan van der Schaft. Equivalence of dynamical systems by bisimulation. *IEEE Transactions on Automatic Control*, 49(12):2160–2172, 2004. doi: 10.1109/TAC.2004.838497.

Peter Van Overschee and Bart De Moor. A unifying theorem for three subspace system identification algorithms. *Automatica*, 31(12):1853–1864, 1995. doi: 10.1016/0005-1098(95)00072-0.

Claas Voelcker, Victor Liao, Animesh Garg, and Amir-massoud Farahmand. Value gradient weighted model-based reinforcement learning. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=4-D6CZkRXxI.

Claas A. Voelcker, Tyler Kastner, Igor Gilitschenski, and Amir-massoud Farahmand. When does self-prediction help? Understanding auxiliary tasks in reinforcement learning. *Reinforcement Learning Journal*, 4:1567–1597, 2024. URL https://rlj.cs.umass.edu/2024/papers/Paper197.html.

Claas A. Voelcker, Anastasiia Pedan, Arash Ahmadian, Romina Abachi, Igor Gilitschenski, and Amir-massoud Farahmand. Calibrated value-aware model learning with probabilistic environment models. In *International Conference on Machine Learning*, volume 267, pp. 61745–61768, 2025. URL https://proceedings.mlr.press/v267/voelcker25a.html.

Tongzhou Wang, Simon Du, Antonio Torralba, Phillip Isola, Amy Zhang, and Yuandong Tian. Denoised MDPs: Learning world models better than the world itself. In *International Conference on Machine Learning*, volume 162, pp. 22591–22612, 2022. URL https://proceedings.mlr.press/v162/wang22c.html.

Ran Wei, Nathan Lambert, Anthony D. McDonald, Alfredo Garcia, and Roberto Calandra. A unified view on solving objective mismatch in model-based reinforcement learning. *Transactions on Machine Learning Research*, 2024. URL https://openreview.net/forum?id=tQVZgvXhZb.

Bryan Wilder, Bistra Dilkina, and Milind Tambe. Melding the data-decisions pipeline: Decision-focused learning for combinatorial optimization. In *AAAI Conference on Artificial Intelligence*, volume 33, pp. 1658–1665, 2019. doi: 10.1609/aaai.v33i01.33011658.

Ran Yan, Shuaian Wang, and Kjetil Fagerholt. A semi-"smart predict then optimize" (semi-SPO) method for efficient ship inspection. *Transportation Research Part B: Methodological*, 142:100–125, 2020. doi: 10.1016/j.trb.2020.09.014.

Ajmal Yousuff and Robert E. Skelton. Controller reduction by component cost analysis. *IEEE Transactions on Automatic Control*, 29(6):520–530, 1984. doi: 10.1109/TAC.1984.1103571.

Ajmal Yousuff and Robert E. Skelton. Covariance equivalent realizations with application to model reduction of large-scale systems. In C. T. Leondes (ed.), *Control and Dynamic Systems*, volume 22, pp. 273–348. Academic Press, 1985. doi: 10.1016/B978-0-12-012722-1.50014-2.

Yang Yue, Bingyi Kang, Zhongwen Xu, Gao Huang, and Shuicheng Yan. Value-consistent representation learning for data-efficient reinforcement learning. In *AAAI Conference on Artificial Intelligence*, volume 37, pp. 11069–11077, 2023. doi: 10.1609/aaai.v37i9.26311.

Zhuquan Zang, Robert R. Bitmead, and Michel Gevers. Iterative weighted least-squares identification and weighted LQG control design. *Automatica*, 31(11):1577–1594, 1995. doi: 10.1016/0005-1098(95)00082-8.

Amy Zhang, Rowan Thomas McAllister, Roberto Calandra, Yarin Gal, and Sergey Levine. Learning invariant representations for reinforcement learning without reconstruction. In *International Conference on Learning Representations*, 2021.

Supplementary Materials

CONTENTS

A Proofs **20**

A.1 Rank-Constrained Retention: Which Belief Directions a Capacity-Limited Predictive Objective Retains . . . 20

A.2 Decision-Weighted Retention: Reweighting the Predictive Target Moves the Retained Subspace onto the Decision-Relevant One . . . 23

A.3 Nonlinear-Encoder Retention: Decoder Linearity Cannot Be Dropped, Encoder Linearity Can . . . 27

A.4 Realizability Test: A Decidable Pre-Flight Check for Theorem 2's Reweighting Condition, with a Quantitative Obstruction Dimension . . . 31

A.5 Approximately Optimal Retention: What a δ-Suboptimal Capacity-Limited Encoder Retains . . . 34

B Hyperparameters **37**

C Full Statistical Results **38**

C.1 Statistical Protocol . . . 38

C.2 The d=16, n_o=8 belief-identification setting . . . 39

C.3 The d=4, n_o=2 belief-identification setting . . . 40

C.4 The d=16, n_o=4 capacity setting . . . 40

C.5 The d=8, n_o=3 capacity setting . . . 40

C.6 The fully observed d=8 setting . . . 41

D Training, Compute and Premise Checks **41**

D.1 Software Framework . . . 41

D.2 Compute Budget . . . 41

D.3 Seed Management . . . 42

D.4 A continuous premise-violation axis . . . 42

D.5 Why the linear-decoder hypothesis is kept . . . 42

D.6 Whether the linear-decoder hypothesis binds on a trained run . . . 42

D.7 A deeper recurrent encoder . . . 44

D.8 The certificate's tolerance need not be supplied . . . 45

D.9 A degenerate input reweighting that realizes the decision Gramian . . . 46

E Per-Configuration Results **46**

E.1 The d=16, n_o=8 belief-identification setting . . . 46

E.2 The d=4, n_o=2 belief-identification setting . . . 46

E.3 The d=16, n_o=4 capacity setting . . . 47

E.4 The d=8, n_o=3 capacity setting 47

F Additional Figures and Tables **47**

G Extended Method Details **48**

G.1 Windowed JEPA belief identification 48

H Extended Analysis **49**

H.1 Retention geometry beyond the headline metrics 49

H.2 Decision-relevant error: horizon interacts with reweighting 50

H.3 Belief identification: what training keeps 50

H.4 Separating weighting from effective step size, and the literature's comparators . . . 51

H.5 The per-arm step-size grid 53

H.6 The weighting-matched correction family 54

H.7 The certificate without an oracle, and how far it disagrees 55

H.8 A belief-free route to the certificate's numerator, and what it costs 57

H.9 Figures for the results 57

H.10 Protocol 60

H.11 Extended Related Work 60

H.12 Notation 64

H.13 Measuring the Containment, With Uncertainty 67

H.14 How Fast Containment Degrades When the Decision Weight Is Wrong 69

H.15 A decision weight estimated from finite trajectories 71

H.16 A Nonlinear Transition Law, Where No Closed-Form Belief Exists 72

H.17 The result on randomly drawn transition matrices 73

H.18 A nonlinear transition law, read against the measured linearisation gap 74

H.19 The pair form of the approximate-retention certificate, instantiated 75

H.20 Results in Full 76

A PROOFS

Every theorem stated in the body is formalised and machine-checked in Lean 4. The proofs below are the prose form of those developments; the Lean source is the authority, and each subsection names the declaration it corresponds to. The supplementary Lean development holds one module per theorem, `RetentionProofs.X.Main` with `X` one of `RankConstrainedRetention` (Theorem 1), `DecisionWeightedRetention` (Theorem 2), `NonlinearEncoderRetention` (Theorem 3), `RealizabilityTest` (Theorem 4) and `ApproxOptimalRetention` (Theorem 5), with `LQGBeliefDecisionRelevance` for the observability paragraph of Appendix A.2; every line number below refers to that module in the supplementary development.

All five theorems depend on no axiom beyond Lean's three standard ones (propositional extensionality, the axiom of choice, and soundness of quotients): Theorems 1–5. Proposition 1 is an elementary consequence stated with its argument in Appendix H.7 and is not part of the Lean development.

The mathematics of Theorem 1, Theorem 2 and Theorem 4 is in part classical; each of those subsections states which parts, and the contribution there is the machine-checked formalisation and the specialisation to this setting rather than the mathematics.

A.1 RANK-CONSTRAINED RETENTION: WHICH BELIEF DIRECTIONS A CAPACITY-LIMITED PREDICTIVE OBJECTIVE RETAINS

Theorem 1 (Rank-Constrained Retention: Which Belief Directions a Capacity-Limited Predictive Objective Retains)**.** *Assume (A1) $\mathcal{E}$ and $\mathcal{F}$ are finite-dimensional real inner product spaces; (A2) T is linear; (A3) the encoder and decoder are linear; (A4) the statements are about the exact optimum of the population risk; (A5) for clause 5, a strict spectral gap $\lambda_k < \lambda_{k-1}$ with $0 < k < d$; (A6) for clause 6, a factorisation $\Sigma = RR^*$. Write $d = \dim \mathcal{E}$, $G = T^*T$ for the predictive Gramian, $\lambda_0 \geq \cdots \geq \lambda_{d-1}$ for its eigenvalues and $b_0, \ldots, b_{d-1}$ for an orthonormal eigenbasis, so that $Gb_i = \lambda_i b_i$; and call $X = M \circ W$ a* capacity-k pair *when $W : \mathcal{E} \to \mathcal{Z}$ and $M : \mathcal{Z} \to \mathcal{F}$ are linear with $\dim \mathcal{Z} \leq k$. Then:*

1. (Retention bound.) *For every orthonormal family $\{u_j\}_{j \in s}$ in $\mathcal{E}$ indexed by a set s with at most k members, $\sum_{j \in s} \|Tu_j\|^2 \leq \sum_{i<k} \lambda_i$.*
2. (Attainment.) $\displaystyle\sum_{i<k} \|Tb_i\|^2 = \sum_{i<k} \lambda_i$.
3. (Risk lower bound.) *Writing $\|\cdot\|_{HS}$ for the Hilbert–Schmidt norm, every capacity-k pair satisfies $\|T - M \circ W\|_{HS}^2 \geq \sum_{i \geq k} \lambda_i$.*
4. (Optimality of the dominant subspace.) *The encoder $x \mapsto (\langle b_0, x\rangle, \ldots, \langle b_{k-1}, x\rangle)$ with the decoder $c \mapsto \sum_{j<k} c_j \, Tb_j$ attains $\|T - M \circ W\|_{HS}^2 = \sum_{i \geq k} \lambda_i$, which with clause 3 is optimal at capacity k.*
5. (Determinacy under a spectral gap.) *Suppose $0 < k < d$ and $\lambda_k < \lambda_{k-1}$. Then every orthonormal frame of at most k vectors attaining clause 1's bound satisfies $\langle b_i, u_j\rangle = 0$ for every $i \geq k$ and every j: the retained subspace is the dominant k-dimensional invariant subspace of G.*
6. (Covariance weighting.) *Let the stationary belief covariance be $\Sigma = RR^*$ for a linear $R : \mathcal{E} \to \mathcal{E}$. Then $(TR)^*(TR) = R^*(T^*T)R$, and clauses 1–5 hold verbatim for the operator $T \circ R$ in place of T, with G replaced by $R^*(T^*T)R$; in particular every capacity-k pair X satisfies $\|(T - X)R\|_{HS}^2 \geq \sum_{i \geq k} \lambda_i$, with λ_i the eigenvalues of $R^*(T^*T)R$.*

Relation to the literature. Clauses 1–4 are Ky Fan's maximum principle (Fan, 1949) and Eckart–Young–Mirsky (Eckart & Young, 1936; Mirsky, 1960) specialised to T^*T; clauses 3–6 are the population, identity-weight case of reduced-rank regression (Izenman, 1975; Reinsel & Velu, 1998), whose response-weighted form (Hua et al., 2001) gives the retained subspace of Theorem 2 clause 3 once the weight is known; we claim no priority over that mathematics. Our contribution is the machine-checked specialisation to belief space. Capacity-driven feature suppression is studied em-

pirically in (Chen et al., 2021) and (Robinson et al., 2021), but neither line states which directions survive.

Formal statement. Every clause is formalised in Lean, and the formalisation is authoritative for the statement. No single declaration carries the whole conjunction, so we name one per clause, by line of this theorem's development: clause 1 is `retained_le_top` (line 335); clause 2 is `retained_dominant_eq` (line 348); clauses 3 and 4 are the two halves of `capacity_optimum` (line 590), which bundles `residual_lower_bound` (line 429) with `residual_attained` (line 542); clause 5 is `retained_dominant_of_gap` (line 362); and clause 6 is `gram_whitened` (line 618) with `covariance_residual_lower_bound` (line 628) and `covariance_residual_attained` (line 640). The remark at the end of step 10 of the proof, that the attaining pair of $T \circ R$ factors through R, is proved in prose only.

Axiom dependence. The Lean development of Theorem 1 declares no axiom and no opaque symbol; every audited declaration depends only on Lean's standard axioms (propext, Classical.choice, Quot.sound).

Proof. The proof proceeds as follows, in 10 numbered steps.

1. **A fractional knapsack bound.** Let λ be antitone (non-increasing) and nonnegative on an index set of size d, and let weights w_i satisfy $0 \le w_i \le 1$ and $\sum_i w_i \le k$. Set $\tau = \lambda_k$ when $k < d$, and $\tau = 0$ when $k = d$; in both cases $\tau \ge 0$, $\lambda_i \ge \tau$ for $i < k$, and $\lambda_i \le \tau$ for $i \ge k$. For $i < k$ the deficit $w_i - 1$ is nonpositive, so $\lambda_i(w_i - 1) \le \tau(w_i - 1)$; for $i \ge k$ the weight is nonnegative, so $\lambda_i w_i \le \tau w_i$. Adding over all i, and using that the dominant block has exactly k elements,

$$\sum_i \lambda_i w_i \;\le\; \sum_{i<k} \lambda_i + \tau\Big(\sum_i w_i - k\Big) \;\le\; \sum_{i<k} \lambda_i,$$

   the last step because $\tau \ge 0$ and $\sum_i w_i \le k$.

2. **The predictive energy of a direction, expanded in the eigenbasis.** For any $x \in \mathcal{E}$, $\|Tx\|^2 = \langle Tx, Tx\rangle = \langle T^*Tx, x\rangle = \langle Gx, x\rangle$. Expanding both arguments in the eigenbasis and using $Gb_i = \lambda_i b_i$ together with the symmetry of G,

$$\|Tx\|^2 \;=\; \sum_i \lambda_i \,\langle b_i, x\rangle^2.$$

   Restricting to a line $x = s\,v$ gives the directional profile $\varphi(s) = s^2\|Tv\|^2$, so the identity above is its subspace-level form.

3. **The weight profile of an orthonormal frame.** Let $\{u_j\}_{j\in s}$ be orthonormal with $|s| \le k$. Summing step 2 over $j \in s$ and exchanging the order of summation,

$$\sum_{j\in s}\|Tu_j\|^2 \;=\; \sum_i \lambda_i\, w_i, \qquad w_i \;=\; \sum_{j\in s}\langle b_i, u_j\rangle^2.$$

   The weights are nonnegative, being sums of squares. They are at most one: Bessel's inequality for the vector b_i against the orthonormal family $\{u_j\}$ gives $w_i \le \|b_i\|^2 = 1$. Their total is the number of frame vectors: Parseval for each u_j in the orthonormal basis $\{b_i\}$ gives $\sum_i \langle b_i, u_j\rangle^2 = \|u_j\|^2 = 1$, hence $\sum_i w_i = |s| \le k$.

4. **Clause 1.** Steps 1 and 3 combine: the retained energy is a weighted sum of the spectrum with a weight profile satisfying the knapsack hypotheses, so it is at most the sum of the k largest eigenvalues.

5. **Clause 2.** Taking $x = b_i$ in step 2 and using orthonormality, $\|Tb_i\|^2 = \lambda_i$. Summing over $i < k$ gives the claim, so the bound of clause 1 is tight and is attained on the dominant eigendirections.

6. **The Hilbert–Schmidt energy does not depend on the basis.** Let $Y : \mathcal{E} \to \mathcal{F}$ be linear, let $\{e_i\}$ be an orthonormal basis of $\mathcal{E}$ and $\{f_j\}$ one of $\mathcal{F}$. Expanding each $\|Ye_i\|^2$ by Parseval in $\{f_j\}$, exchanging the two sums, and using $\langle f_j, Ye_i\rangle = \langle Y^* f_j, e_i\rangle$ followed by Parseval in $\{e_i\}$,
$$\sum_i \|Ye_i\|^2 \;=\; \sum_j \|Y^* f_j\|^2.$$
The right-hand side does not mention $\{e_i\}$, so any two orthonormal bases of $\mathcal{E}$ give the same value. Applying this with $Y = T$ and the eigenbasis, together with step 5, gives $\|T\|_{HS}^2 = \sum_i \lambda_i$.

7. **Clause 3.** Let $X = M \circ W$ be a capacity-k pair and $K = \ker X$. Since X factors through $\mathcal{Z}$, its range has dimension at most $\dim \mathcal{Z} \le k$, so rank-nullity gives $p := \dim K \ge d - k$. Choose an orthonormal basis of K, place it at the indices $i < p$, and extend it to an orthonormal basis $\{e_i\}$ of $\mathcal{E}$; such an extension exists because $\mathcal{E}$ is a finite-dimensional inner product space. For $i < p$ we have $Xe_i = 0$, so $(T - X)e_i = Te_i$. Evaluating the Hilbert–Schmidt energy in this basis, which step 6 permits, and discarding the nonnegative terms with $i \ge p$,
$$\|T - X\|_{HS}^2 \;\ge\; \sum_{i<p} \|Te_i\|^2 \;=\; \sum_i \lambda_i \;-\; \sum_{i\ge p} \|Te_i\|^2.$$
The vectors $\{e_i\}_{i\ge p}$ form an orthonormal frame of $d - p \le k$ members, so clause 1 bounds the subtracted sum by $\sum_{i<k} \lambda_i$. Therefore $\|T - X\|_{HS}^2 \ge \sum_i \lambda_i - \sum_{i<k} \lambda_i = \sum_{i\ge k} \lambda_i$.

8. **Clause 4.** With the encoder and decoder of the statement, the composite is $x \mapsto T(\Pi_k x)$, where Π_k is the orthogonal projection onto $\operatorname{span}\{b_0, \dots, b_{k-1}\}$: indeed $\sum_{j<k} \langle b_j, x\rangle\, Tb_j = T\big(\sum_{j<k} \langle b_j, x\rangle b_j\big)$ by linearity. Evaluating the Hilbert–Schmidt energy of $T - T\Pi_k$ in the eigenbasis, which step 6 permits, the residual is $Tb_i - Tb_i = 0$ for $i < k$ and Tb_i for $i \ge k$, whose squared norm is λ_i by step 5. Hence the risk is exactly $\sum_{i\ge k} \lambda_i$, matching the lower bound of clause 3, so that value is optimal and the dominant eigenspace attains it.

9. **Clause 5.** Sharpen step 1 with two thresholds. Put $g = \lambda_{k-1}$ and $\tau = \lambda_k$, so $\lambda_i \ge g$ for $i < k$ and $\lambda_i \le \tau$ for $i \ge k$. On the dominant block $\lambda_i - \lambda_i w_i \ge g(1 - w_i)$, so
$$\sum_{i<k} \lambda_i w_i \;\le\; \sum_{i<k} \lambda_i - g\Big(k - \sum_{i<k} w_i\Big) \;\le\; \sum_{i<k} \lambda_i - g \sum_{i\ge k} w_i,$$
the last step because $\sum_i w_i \le k$ forces $k - \sum_{i<k} w_i \ge \sum_{i\ge k} w_i$ and $g \ge 0$. Off the block $\sum_{i\ge k} \lambda_i w_i \le \tau \sum_{i\ge k} w_i$. Adding,
$$\sum_i \lambda_i w_i \;\le\; \sum_{i<k} \lambda_i \;-\; (g - \tau) \sum_{i\ge k} w_i.$$
If the frame attains the bound of clause 1, the left side equals $\sum_{i<k} \lambda_i$, so $(g - \tau) \sum_{i\ge k} w_i \le 0$. The gap hypothesis gives $g - \tau > 0$, hence $\sum_{i\ge k} w_i \le 0$; as the weights are nonnegative each w_i with $i \ge k$ vanishes. Each such w_i is a sum of squares, so $\langle b_i, u_j\rangle = 0$ for every $i \ge k$ and every j. The frame therefore lies in the span of $b_0, \dots, b_{k-1}$; being orthonormal with k members inside a k-dimensional space when $|s| = k$, it spans exactly that subspace. Fewer than k vectors cannot attain the bound: an attaining frame with $|s| < k$ would lie in that span by the same argument, so step 3 would give it retained energy $\sum_{i<k} \lambda_i w_i$ with $\sum_{i<k} w_i = |s| < k$ and each $w_i \le 1$, short of $\sum_{i<k} \lambda_i$ by at least $\lambda_{k-1}(k - |s|) > 0$, since the gap gives $\lambda_{k-1} > \lambda_k \ge 0$.

10. **Clause 6.** The adjoint of a composite reverses the order, so $(TR)^*(TR) = R^* T^* TR = R^*(T^*T)R$. Moreover $(T - M \circ W) \circ R = (T \circ R) - M \circ (W \circ R)$, and $W \circ R$ is again a linear map into $\mathcal{Z}$, so $M \circ (W \circ R)$ is a capacity-k pair for the predictive operator $T \circ R$. Clauses 1 through 5 applied to $T \circ R$ therefore give the covariance-weighted statements, with the spectrum and eigenbasis those of $R^*(T^*T)R$; clauses 1 and 2 are statements about a linear operator and transport to $T \circ R$ unchanged, as clauses 3, 4 and 5 do. *Remark, proved*

in prose only. Attainment survives a singular R. Writing λ_i and b_i here for the spectrum and eigenbasis of $R^*(T^*T)R$, the value $\sum_{i\geq k}\lambda_i$ is attained by the dominant pair of $T \circ R$ (`covariance_residual_attained`, line 640), whose encoder reads $\langle b_i, x\rangle$ for $i < k$, and that pair has the form $X \circ R$. For $\lambda_i > 0$, $b_i = \lambda_i^{-1} R^*(T^*TRb_i)$ lies in $\operatorname{range}(R^*)$, so $\langle b_i, x\rangle = \langle \lambda_i^{-1} T^*TRb_i, Rx\rangle$ depends on x only through Rx; for $\lambda_i = 0$ the decoder sends that coordinate to $TRb_i = 0$, so it can be dropped. The composite is therefore $X \circ R$ for a linear X of rank at most k, so the bound of clause 6 is also attained by a capacity-k pair on $\mathcal{E}$. This factorisation is not part of the formal statement: the Lean declaration states the value for pairs of $T \circ R$.

□

A.2 Decision-Weighted Retention: Reweighting the Predictive Target Moves the Retained Subspace onto the Decision-Relevant One

Theorem 2 (Decision-Weighted Retention: Reweighting Moves the Retained Subspace)**.** *Assume (A1) $\mathcal{E}$, $\mathcal{F}$, $\mathcal{G}$ are finite-dimensional real inner product spaces; (A2) T, D, the encoder and the decoder are linear; (A3) the decision criterion is the one of §3, under which v is decision-irrelevant exactly when $Dv = 0$; (A4) realizability, $\ker T \subseteq \ker D$, for clause 3 only, clause 1 being the equivalence that decides it and assuming nothing; (A5) a strict gap $\mu_k < \mu_{k-1}$ of the decision Gramian $P = D^*D$, whose eigenvalues are $\mu_0 \geq \mu_1 \geq \cdots$, for clause 3, clause 4 carrying its own gap hypothesis on T'^*T' in the clause itself; (A6) capacity within the decision rank, $\mu_{k-1} > 0$, for clause 3; (A7) every statement is about the exact optimum of the population risk. Then:*

1. (Realizability.) *A factor $V : \mathcal{F} \to \mathcal{G}$ with $V \circ T = D$ exists, hence a positive semidefinite $W = V^*V$ with $T^*WT = P$, if and only if $\ker T \subseteq \ker D$.*

2. (Necessity of realizability.) *If some $v \in \mathcal{E}$ satisfies $Tv = 0$ and $Dv \neq 0$, then $T^*WT \neq P$ for **every** linear $W : \mathcal{F} \to \mathcal{F}$, positive semidefinite or not.*

3. (Exact transport.) *Let $V \circ T = D$. Let $\{u_j\}_{j\in s}$ be an orthonormal family of at most k belief directions attaining Theorem 1 clause 1's bound for the reweighted operator. Then:*
   (a) each u_j satisfies $\|Du_j\|^2 \geq \mu_{k-1} > 0$, and $\langle u_j, v\rangle = 0$ whenever $Dv = 0$;
   (b) every capacity-k pair $X = M \circ W_{\mathrm{enc}}$ whose reweighted risk $\|VT - X\|_{HS}^2$ attains its optimum $\sum_{i\geq k}\mu_i$ satisfies $\ker D \subseteq \ker X$.

4. (Approximate transport.) *Let T' be linear on $\mathcal{E}$ with $\|T'x\|^2 \leq \|Dx\|^2 + \varepsilon\|x\|^2$ for all x and real slack ε, let ν_i be the descending eigenvalues of T'^*T', and suppose $\nu_k < \nu_{k-1}$. Then every orthonormal family $\{u_j\}_{j\in s}$ of at most k directions attaining the retention bound for T' satisfies $\|Du_j\|^2 \geq \nu_{k-1} - \varepsilon$.*

5. (Necessity of the reweighting.) *There exist a two-dimensional belief space, a predictive operator T_c with $\ker T_c = 0$, and a decision operator D such that for every c with $c^2 > 1$: $T_c^*T_c$ has eigenvalues $(c^2, 1)$ and D^*D has $(1, 0)$; the dominant eigendirection of $T_c^*T_c$ is sent to zero by D; and every capacity-one family attaining the retention bound for D is orthogonal to it.*

Proof sketch. Realizability defines V by $VTx = Dx$, so $W = V^*V$ gives $T^*WT = P$, and no W does once some v has $Tv = 0 \neq Dv$. The retention bound for the reweighted operator, with Gram operator P, puts the retained subspace inside $(\ker D)^\perp$. The full proof follows. □

Formal statement. Every clause is formalised in Lean, and the formalisation is authoritative for the statement. `redirection` (line 466 of this theorem's development) is the packaged form, and it carries the forward half of clause 1 together with clause 3(b) under the hypothesis $\ker T \subseteq \ker D$; the other clauses are separate declarations in the same file. By clause: clause 1 is `reweighting_iff` (line 141), which proves the equivalence with no realizability hypothesis, with `gram_reweighted` (line 126) for the weight $W = V^*V$; clause 2 is `no_weight_realizes_of_blind` (line 164); clause 3(a) is `retained_decision_energy`

(line 259) for the energy floor and `retained_mem_ker_orthogonal` (line 318) for the orthogonality, and clause 3(b) is `optimal_encoder_annihilates_irrelevant` (line 345); clause 4 is `approx_retained_decision_energy` (line 234); and clause 5 is the witness of `wit_dominant_decision_irrelevant` (line 687), `wit_decision_spectrum` (line 699) and `wit_reweighted_retained_relevant` (line 744), whose reweighting exists by `wit_reweighting_exists` (line 603).

Axiom dependence. The Lean development of Theorem 2 declares no axiom and no opaque symbol; every audited declaration depends only on Lean's standard axioms (propext, Classical.choice, Quot.sound).

Belief covariance. Like the body, the theorem is stated at identity belief covariance, and it needs no separate clause for a general one. Its hypotheses concern linear maps T and D, so for a stationary belief covariance $\Sigma = RR^*$, with $R = \Sigma^{1/2}$, it applies to $T \circ R$ and $D \circ R$, exactly as clause 6 of Theorem 1 does, with the gap and capacity hypotheses read on the eigenvalues of R^*D^*DR; the capacity hypothesis then restricts k to at most $\operatorname{rank}(\Sigma)$. Realizability carries over, since $V \circ T = D$ gives $V \circ (T \circ R) = D \circ R$. With a zero-mean belief the covariance-weighted reweighted risk is $\|(VT - X)R\|_{HS}^2$ up to a constant no code changes, and clause 6 of Theorem 1, read for D, with the remark at the end of its proof, makes XR an optimal capacity-k pair for $D \circ R$ whenever X is optimal for that risk. Clause 3(b) then gives $\ker(DR) \subseteq \ker(XR)$: every optimal pair's composite sends to zero each direction of $\operatorname{range}(\Sigma)$ that the cost ignores, and in whitened coordinates the directions it retains are ranked by R^*D^*DR rather than by D^*D. This is the whitened belief space $\operatorname{range}(\Sigma_{\hat{x}})$ on which the experiments instantiate the theorems (Appendix A.3), of rank 8 at d=16, n_o=4. The theorems thus hold for any belief covariance, by this paragraph and clause 6 of Theorem 1, and do not need the closed-loop one; the runs use the covariance of the open-loop law the trajectories are simulated under (§3). For the window encoder of Equation 1, future noise is independent of the window, so $\mathbb{E}[y_m(t) \mid o_{t-w+1}, \ldots, o_t] = T\hat{x}_t^w$ with $\hat{x}_t^w = \mathbb{E}[x_t \mid o_{t-w+1}, \ldots, o_t]$, and q times the loss is a constant plus the same risk with $\hat{x}_t^w$ and its covariance in place of $\hat{x}_t$ and $\Sigma_{\hat{x}}$. Our exact optima use the Kalman covariance, so the difference between the two beliefs is part of what Table 21's caption calls the error the encoder's finite window adds.

What clause 4 gives. Clause 4's floor $\nu_{k-1} - \varepsilon$ is in the spectrum of T'^*T', not of P, and its slack hypothesis is one-sided: it bounds $\|T'x\|^2$ above by $\|Dx\|^2 + \varepsilon\|x\|^2$ and not below, so a T' much smaller than D meets it with a small ν_{k-1}. The clause is empty when $\nu_{k-1} \le \varepsilon$. When $\nu_{k-1} > \varepsilon$ it gives each retained direction positive decision energy, so none lies in $\ker D$, but unlike clause 3(a) it gives no orthogonality to $\ker D$. A two-sided version is not part of the Lean development, and we claim none.

The finite-horizon cost as an observability condition. With P_{t_h} as in §3 and $\rho > 0$, a belief direction v leaves the horizon-t_h cost of the fixed linear law unchanged when added, at any scale, to any belief exactly when $P_{t_h}v = 0$, that is when $v \in \ker D$; and $P_{t_h}v = 0$ exactly when $HA_{\mathrm{cl}}^t v = 0$ and $K_c A_{\mathrm{cl}}^t v = 0$ for every $t \le t_h$. So $\ker D$ is the unobservable subspace of the pair $\left(A_{\mathrm{cl}}, [H; \sqrt{\rho}\, K_c]\right)$ truncated at horizon t_h, a finite rank condition, and by the Cayley–Hamilton theorem it is the whole unobservable subspace once $t_h \ge d - 1$ (Kwakernaak & Sivan, 1972). Both equivalences are machine-checked in the supplementary Lean development `LQGBeliefDecisionRelevance`, as `decisionIrrelevant_iff_costToGo_mulVec_zero` (line 318) and `decisionIrrelevant_iff_stacked_kernel` (line 365), with no `sorry`, axiom or opaque symbol. The result is classical and is not one of the five theorems the paper states.

The cost-to-go as the price of a belief error. Scoring a belief error v by $v^\top P_{t_h} v$ is a modelling choice: it is the cost the fixed closed loop accumulates over the horizon from v treated as a state. Under certainty-equivalent control $u_t = -K_c\hat{x}_t$, acting on a wrong belief changes the applied control by $-K_c v$, so the cost of acting on it is a quadratic form in $K_c v$. That form vanishes on $\ker K_c$, and $v^\top P_{t_h} v$ in general does not, since it also charges what the state $A_{\mathrm{cl}}^t v$ costs at later steps; the two are different quadratic forms. The theorems hold for any linear D, so for this weight as defined. The choice fixes which directions count as decision-irrelevant, $\ker D = \ker P_{t_h}$, and how the metric is read, not whether the results hold.

Relation to the literature. Value-equivalent and decision-aware model learning (Farahmand et al., 2017; Farahmand, 2018; Ayoub et al., 2020; Grimm et al., 2020; Abachi et al., 2020; Nikishin et al., 2022) make the model objective depend on how a planner uses the model; the value-gradient weighted model, VaGraM (Voelcker et al., 2022), applies a per-sample diagonal value-gradient metric to the vector prediction error, which weights each coordinate of the target differently for each sample rather than scaling a data point by one scalar. Identification for control weights the identification criterion by the control objective, with sufficient conditions for transfer functions over frequency (Gevers, 1993; Zang et al., 1995; Rivera et al., 1992). Given the weight, the retained subspace of clause 3 is that of response-weighted reduced-rank estimation (Hua et al., 2001), and goal-oriented low-rank approximation places its directions orthogonal to a goal operator's null space (Spantini et al., 2017); value-directed compression names the reward-generated belief subspace whose linear-Gaussian analogue is $(\ker D)^{\perp}$ (Poupart & Boutilier, 2002); cost-driven LQG representation learning recovers the cost-relevant state at the state's dimension under cost observability (Tian et al., 2023). In control and model reduction, cost decomposition, component cost analysis, LQG and frequency-weighted balancing, and goal-oriented or quadratic-output reduction likewise select reduced models by a control cost, a performance band, or a quantity of interest (Skelton, 1980; Moore, 1981; Jonckheere & Silverman, 1983; Enns, 1984; Yousuff & Skelton, 1984; Carlberg & Farhat, 2011; Pulch & Narayan, 2019; Benner et al., 2022). Those results reduce plants, compensators, controllers or descriptor systems and analyse a transfer-function or output error; the object here is a capacity-limited predictive encoder and the conclusion is which subspace it retains.

The algebra of clauses 1 and 2 is classical, and we claim none of it. That a factor V with $V \circ T = D$ exists exactly when $\operatorname{range}(D^*) \subseteq \operatorname{range}(T^*)$, equivalently $\ker T \subseteq \ker D$, is Douglas's range-inclusion theorem (Douglas, 1966), which also supplies the positive semidefinite weight directly as $W = V^*V$. That *some* linear W, positive semidefinite or not, solves $T^*WT = D^*D$ under the same condition is Penrose's consistency criterion $AA^+CB^+B = C$ for $AXB = C$ (Penrose, 1955). Necessary and sufficient conditions and general expressions for Hermitian and nonnegative definite solutions of these equations are due to Khatri & Mitra (1976), with the $AXA^* = B$ case restated by Baksalary (1984). Clauses 3 and 4 rest on Ky Fan (1949) and Eckart–Young (1936) through Theorem 1.

What the theorem adds is the join: the realizability criterion carried into the retention statement for a rank-constrained *encoder*, so that every exact weighted optimum lies in the decision-relevant subspace, together with an unweighted instance in which it does not. We claim that scoped combination and its machine-checked proof, not operator factorisation, cost-sensitive reduction, or low-rank approximation separately. We are not aware of prior work stating the encoder-side transport or the unweighted witness. The closest analogues we know are the tabular result that reconstruction with a temporal-difference loss misses the value when the eigenvectors spanning the reward lie outside the top-k singular span (Voelcker et al., 2024) and the data-informed but goal-irrelevant directions of Spantini et al. (2017).

Proof. The proof proceeds as follows, in 11 numbered steps.

1. **Decision-irrelevance in factored form.** By the defining property of the adjoint, $\langle Px, x\rangle = \langle D^*Dx, x\rangle = \|Dx\|^2$. Hence $Pv = 0$ implies $\|Dv\|^2 = 0$ and so $Dv = 0$; conversely $Dv = 0$ gives $Pv = 0$. Therefore $\ker P = \ker D$, and the criterion for a belief direction to be decision-irrelevant under a linear policy and quadratic cost is the vanishing of the decision operator.

2. **Realizability reduces to a factorisation.** Suppose $V : \mathcal{F} \to \mathcal{G}$ satisfies $V \circ T = D$. Then $W = V^*V$ is positive semidefinite, because $\langle Wy, y\rangle = \|Vy\|^2 \geq 0$, and by the adjoint of a composite

$$T^*WT \;=\; T^*V^*VT \;=\; (VT)^*(VT) \;=\; D^*D \;=\; P.$$

   So the analytic requirement "a positive semidefinite weight whose reweighted Gram operator is the decision Gramian" is met as soon as the algebraic requirement "the decision operator factors through the predictive operator" is met.

3. **The factorisation exists exactly under the kernel condition.** If $V \circ T = D$ and $Tv = 0$, then $Dv = V(Tv) = 0$; hence $\ker T \subseteq \ker D$. Conversely, assume $\ker T \subseteq \ker D$.

The first isomorphism theorem gives a linear isomorphism $\mathcal{E}/\ker T \cong \mathrm{range}(T)$, and D descends to a linear map on $\mathcal{E}/\ker T$ because it sends to zero $\ker T$. Composing the inverse isomorphism with that descent produces a linear $L_0 : \mathrm{range}(T) \to \mathcal{G}$ with $L_0(Tx) = Dx$ for every x. Extending L_0 from the subspace $\mathrm{range}(T)$ to all of $\mathcal{F}$ (any linear extension serves, and this is where the behaviour of W off the range of T is chosen) yields V with $V \circ T = D$.

4. **No weight helps when a decision-relevant direction is predictively invisible.** Suppose $Tv = 0$ and $Dv \neq 0$. For any linear W on $\mathcal{F}$, $(T^*WT)v = T^*W(Tv) = 0$, so $\langle (T^*WT)v, v\rangle = 0$, while step 1 gives $\langle Pv, v\rangle = \|Dv\|^2 > 0$. Hence $T^*WT \neq P$. Positive semidefiniteness of W is never used: the obstruction is that the predictive operator has already discarded the direction, and an output-side weight acts only on what the predictive operator produced.

5. **Transport is the retention theorem applied to the decision operator.** When $V \circ T = D$, the reweighted predictive operator *is* the decision operator, so the reweighted risk is the risk of D and every clause of Theorem 1 applies with the predictive spectrum replaced by the decision spectrum. In particular the capacity-k optimal value of the reweighted risk is $\sum_{i \geq k} \mu_i$, attained by the encoder reading the k dominant eigendirections of the decision Gramian, so the attainment hypotheses used below are satisfiable.

6. **A direction confined to the dominant eigenspace carries at least the boundary eigenvalue.** Let T' be a linear map with Gram spectrum ν_i and eigenbasis b_i, and let u be a unit vector with $\langle b_i, u\rangle = 0$ for every $i \geq k$. Theorem 1's expansion $\|T'u\|^2 = \sum_i \nu_i \langle b_i, u\rangle^2$ then has only the terms $i < k$, each with $\nu_i \geq \nu_{k-1}$ because the spectrum is antitone, and Parseval gives $\sum_{i<k} \langle b_i, u\rangle^2 = \|u\|^2 = 1$. Hence $\|T'u\|^2 \geq \nu_{k-1}$.

7. **Approximate transport.** Let $\{u_j\}_{j \in s}$ attain the retention bound for T', with $|s| \leq k$ and $\nu_k < \nu_{k-1}$. Theorem 1's determinacy clause places every u_j in the dominant eigenspace of T'^*T', so step 6 gives $\|T'u_j\|^2 \geq \nu_{k-1}$. The closeness hypothesis at the unit vector u_j gives $\|T'u_j\|^2 \leq \|Du_j\|^2 + \varepsilon$, whence $\|Du_j\|^2 \geq \nu_{k-1} - \varepsilon$. This proves clause 4, and clause 3's energy statement is the case $T' = D$, $\varepsilon = 0$.

8. **Retained directions are orthogonal to decision-irrelevant ones.** If $\mu_i > 0$ and $Dv = 0$, then
$$\mu_i \langle b_i, v\rangle \;=\; \langle Pb_i, v\rangle \;=\; \langle Db_i, Dv\rangle \;=\; 0,$$
so $\langle b_i, v\rangle = 0$: a decision-irrelevant direction has no component along any eigendirection of positive decision energy. Now expand in the eigenbasis, $\langle u_j, v\rangle = \sum_i \langle u_j, b_i\rangle\langle b_i, v\rangle$. The terms with $i \geq k$ vanish because determinacy places u_j in the dominant eigenspace; the terms with $i < k$ vanish by the previous sentence, since $\mu_i \geq \mu_{k-1} > 0$ there. Hence $u_j \perp \ker D$, that is, u_j lies in the decision-relevant subspace.

9. **From attaining families to optimal encoders.** Let $X = M \circ W_{\mathrm{enc}}$ be a capacity-k pair whose reweighted risk is at most $\sum_{i \geq k} \mu_i$. Rank-nullity and $\mathrm{rank}(X) \leq \dim \mathcal{Z} \leq k$ give $r = \dim \ker X \geq d - k$, with $d = \dim \mathcal{E}$ as in Theorem 1. Take an orthonormal basis of $\ker X$, place it at the first r indices, and extend it to an orthonormal basis $\{e_i\}$ of the belief space. On the kernel block the residual is the full decision energy of D, so
$$\sum_{i<r} \|De_i\|^2 \;\leq\; \|D - X\|_{HS}^2 \;\leq\; \sum_{i \geq k} \mu_i,$$
and since $\sum_{i<r} \|De_i\|^2 + \sum_{i \geq r} \|De_i\|^2 = \sum_i \mu_i$, the complementary family $\{e_i\}_{i \geq r}$, orthonormal, of size $d - r \leq k$, satisfies $\sum_{i \geq r} \|De_i\|^2 \geq \sum_{i<k} \mu_i$: it attains the retention bound. Step 8 applies to it, so $\langle e_i, v\rangle = 0$ for every $i \geq r$ and every decision-irrelevant v. Expanding $v = \sum_i \langle e_i, v\rangle e_i$ and applying X, every term vanishes: on the kernel block because $Xe_i = 0$, off it because $\langle e_i, v\rangle = 0$. Hence $Xv = 0$, i.e. $\ker D \subseteq \ker X$.

10. **The witness.** Figure 4 draws it. Let $\mathcal{E}$ be two-dimensional with orthonormal basis $\{b_0, b_1\}$, fix c with $c^2 > 1$, and define
$$T_c x \;=\; c\langle b_0, x\rangle b_0 + \langle b_1, x\rangle b_1, \qquad Dx \;=\; \langle b_1, x\rangle b_1.$$

Writing $\alpha = \langle b_0, x\rangle$ and $\beta = \langle b_1, x\rangle$, orthonormality gives $\|T_c x\|^2 = c^2\alpha^2 + \beta^2$, $\|Dx\|^2 = \beta^2$ and $\|x\|^2 = \alpha^2 + \beta^2$. Hence $\|T_c x\|^2 \le c^2\|x\|^2$, with equality only when $(c^2 - 1)\beta^2 = 0$, i.e. only when $\beta = 0$, i.e. only when $Dx = 0$: every maximiser of the unweighted predictive energy is decision-irrelevant. The leading eigenvalue of a Gram operator is the maximum of the energy over unit vectors (the lower bound is the retention bound at $k = 1$ applied to the one-element family $\{b_0\}$, and the upper bound is the display above applied to the leading eigenvector), so the leading eigenvalue of $T_c^* T_c$ is c^2; since the eigenvalues sum to the Hilbert–Schmidt energy $c^2 + 1$, the spectrum is $(c^2, 1)$. The gap $1 < c^2$ therefore holds, determinacy applies, and the retained direction of the unweighted capacity-one optimum is the leading eigendirection, which the display shows is sent to zero by D. The same computation for D gives decision spectrum $(1, 0)$, so the hypotheses of clause 3 hold at $k = 1$ and every capacity-one family attaining the reweighted bound is orthogonal to $\ker D$. Finally $\ker T_c = 0$, so realizability holds and a reweighting exists: the dissociation is not caused by the failure of clause 1's hypothesis.

11. **The magnitude of the dissociation.** In the witness, the capacity-one optimum of the unweighted risk retains decision energy 0 out of an available maximum of 1, at every c with $c^2 > 1$, while the reweighted optimum retains all of it. The fraction of total predictive energy carried by the wrongly retained direction is $c^2/(c^2 + 1)$, which approaches 1 as c grows: the unweighted objective's preference for the decision-irrelevant direction can be made arbitrarily pronounced.

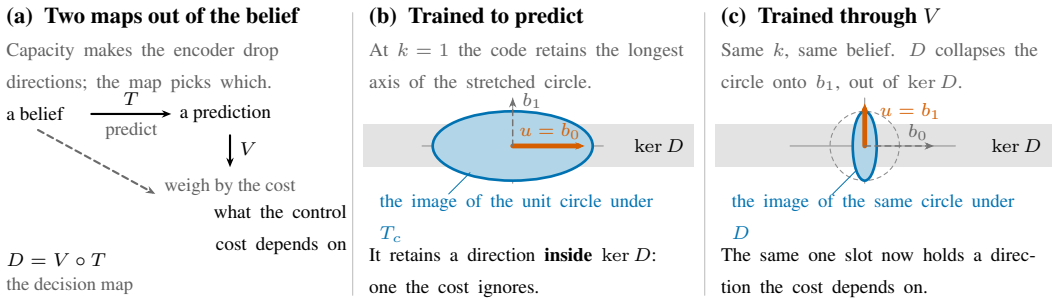
□

The encoder itself (a prose step, not in the Lean development). Clause 3(b) is stated for the composite $X = M \circ W_{\mathrm{enc}}$, and so is its Lean form: `optimal_encoder_annihilates_irrelevant` concludes $(M \circ W)v = 0$. At the optimum the containment also holds for the encoder, so no optimal code carries a decision-irrelevant direction that its decoder then discards. In step 9 the attaining family $\{e_i\}_{i \ge r}$ has $d - r \le k$ members, and clause 1 of Theorem 1, read for D at capacity $d - r$, bounds what it carries by $\sum_{i<d-r} \mu_i$. If $d - r < k$, that is less than $\sum_{i<k} \mu_i$, since $\mu_i \ge \mu_{k-1} > 0$ for every $i < k$ by (A6). So $d - r = k$ and $\mathrm{rank}(X) = k$. As $\mathrm{rank}(X) \le \mathrm{rank}(W_{\mathrm{enc}}) \le \dim \mathcal{Z} \le k$, the encoder maps onto a k-dimensional code on which M is injective, hence $\ker W_{\mathrm{enc}} = \ker X \supseteq \ker D$. For a general belief covariance the same argument, applied to $W_{\mathrm{enc}} \circ R$ with (A6) read on the eigenvalues of $R^* D^* D R$, gives $\ker(DR) \subseteq \ker(W_{\mathrm{enc}} R)$. This step is not machine-checked.

A.3 Nonlinear-Encoder Retention: Decoder Linearity Cannot Be Dropped, Encoder Linearity Can

Theorem 3 (Nonlinear-Encoder Retention: Decoder Linearity Cannot Be Dropped, Encoder Linearity Can)**.** *Assume (A1) $\mathcal{E}$, $\mathcal{G}$ and $\mathcal{Z}$ are finite-dimensional real inner product spaces, $\mathcal{Z}$ carrying an inner product only so that the capacity constraint $\dim \mathcal{Z} \le k$ is stated in Theorem 1's language; (A2) D is linear; (A3) the decoder M is linear and the encoder W_{enc} is an arbitrary function, for clauses 1 to 4; clause 5 is the sharpness witness and exhibits a nonlinear decoder; (A4) π is a finitely supported measure on $\mathcal{E}$ with nonnegative weights; (A5) strictly positive weights, for clause 4; (A6) the atoms span $\mathcal{E}$, for clause 4(c); (A7) a strict gap $\mu_k < \mu_{k-1}$ and $\mu_{k-1} > 0$, for clause 4(c); (A8) every statement is about the exact optimum of the population risk. Write π's atoms as a_j with weights p_j, and $\mathcal{R}_\pi(F) = \sum_j p_j \|Da_j - F(a_j)\|^2$ for the risk of a map F under π. Write m for the number of atoms (local to this theorem, not the offset count m of the Notation), $R : \mathbb{R}^m \to \mathcal{E}$ for the **atom factor**, the linear map with $Re_j = \sqrt{p_j}\, a_j$ on the standard basis e_j of $\mathbb{R}^m$, and $\mu_0 \ge \mu_1 \ge \cdots \ge \mu_{m-1}$ for the eigenvalues of $(DR)^*(DR)$ on $\mathbb{R}^m$ in decreasing order, with $k \le m$. These μ_i, and not the eigenvalues of D^*D, are the ones clause 3, clause 4 and (A7) use. Then:*

1. ***(Risk correspondence.)** For every **linear** $Y : \mathcal{E} \to \mathcal{G}$, $\mathcal{R}_\pi(Y) = \|(D - Y)R\|_{HS}^2$: Theorem 1 clause 6's covariance-weighted risk written out atom by atom, not a substituted functional.*



Every symbol above

k code dimension: how many numbers the encoder may keep. Here $k = 1$.

T predict. V weigh by the cost. $D = V \circ T$ decide. $\ker D$ what the cost ignores.

b_0, b_1 the two axes of this example. u the direction the code retains.

T_c the example's predictive map, $T_c x = c\langle b_0, x\rangle b_0 + \langle b_1, x\rangle b_1$ with $c^2 > 1$.

Scope. Every quantity drawn here is exact, and every claim is about the exact optimum of a population risk.

Figure 4: The witness of clause 5 of Theorem 2, drawn. (a) The predictive map T, the weight V, and the decision map $D = V \circ T$ they compose to. In (b) and (c) the blue shape is the image of the unit circle and the vermillion arrow is the direction u the one-slot code retains; the grey band is $\ker D$, the directions the cost ignores. (b) Under the unweighted objective the image is the ellipse T_c makes, stretched by c along b_0, and the code spends its one slot on the ellipse's longest axis, which lies inside $\ker D$. (c) Weighted through V the image is the one D makes: D retains only the b_1 component, so the circle collapses onto b_1 with no extent left along $\ker D$, and the same one slot now holds a direction the cost depends on. The strip below the panels defines every symbol drawn.

2. ***(Linear-decoder collapse.)*** *For every latent space $\mathcal{Z}$, every **linear** $M : \mathcal{Z} \to \mathcal{G}$ and every **function** $W_{\text{enc}} : \mathcal{E} \to \mathcal{Z}$, put $S = \operatorname{range} M$ and let $X^\star = \Pi_S \circ D$. Then $X^\star$ is linear, $\operatorname{rank} X^\star \le \dim \mathcal{Z}$, and for **every** $x \in \mathcal{E}$ $\|Dx - X^\star x\| \le \|Dx - M(W_{\text{enc}}(x))\|$, with equality if and only if $M(W_{\text{enc}}(x)) = X^\star x$. Being pointwise, the inequality does not depend on π.*

3. ***(No nonlinear-encoder gain.)*** *Every capacity-k pair with a linear decoder and an **arbitrary** encoder therefore satisfies $\sum_{i \ge k} \mu_i \le \mathcal{R}_\pi(M \circ W_{\text{enc}})$, the bound Theorem 1 clause 3 proves for **linear** encoders. A linear pair attains it, so the capacity-k optima over arbitrary and over linear encoders coincide.*

4. ***(Retention.)*** *Assume $p_j > 0$ for every j, and let $M \circ W_{\text{enc}}$ attain clause 3's bound. Then: **(a)** $M(W_{\text{enc}}(a_j)) = X^\star a_j$ at every atom, so on the support of π the optimal composite coincides with a linear capacity-k map. **(b)** if $Da_j = Da_{j'}$ then $M(W_{\text{enc}}(a_j)) = M(W_{\text{enc}}(a_{j'}))$: the composite depends on an atom only through its decision image, and separates no two atoms the decision operator identifies. **(c)** if in addition the atoms span $\mathcal{E}$, $\mu_k < \mu_{k-1}$ and $\mu_{k-1} > 0$, then $X^\star v = 0$ for every decision-irrelevant v, which also follows at once from $X^\star = \Pi_S \circ D$: the gap and spanning hypotheses, which the Lean declaration carries, are not needed for it. With (a), the optimal composite sends every atom a_j with $Da_j = 0$ to zero.*

5. ***(Sharpness: the decoder's linearity carries the result.)*** *There are a three-dimensional $\mathcal{E}$ with orthonormal basis b_0, b_1, b_2, the rank-2 decision operator $Dx = \langle b_0, x\rangle b_0 + \langle b_1, x\rangle b_1$ with $\ker D = \operatorname{span}(b_2)$, the four unit-weight atoms $b_0,\ b_1,\ b_1 + b_2,\ b_1 - b_2$, and a capacity-1 pair with a **linear** encoder $W_{\text{enc}} : \mathcal{E} \to \mathbb{R}$ and a **nonlinear** decoder $M : \mathbb{R} \to \mathcal{E}$, such that $\mathcal{R}_\pi(M \circ W_{\text{enc}}) = 1/4$ while every capacity-1 pair with a **linear** decoder has $\mathcal{R}_\pi \ge 1$. With clauses 2–4 this locates the operative half of Theorem 1 assumption 3 and Theorem 2 assumption 2: the conclusions survive an arbitrary encoder and fail under a nonlinear decoder.*

The encoder class of Theorem 3 includes our recurrent encoders, and one of its hypotheses can be measured on them. Reading π as the finite held-out anchor family every logged risk averages over, the spanning clause is measured while finite support is asserted. Under that reading $\mathcal{R}_\pi$ in every clause, clause 4 included, is the empirical risk, the mean over one record's anchors, and not the population risk of the generator; (A8)'s population risk is then the risk under that finite measure. An anchor is one atom a_j of that family, and the anchor rank of a configuration is the dimension of the span of its anchors, so (A6) asks exactly that this rank equal $\dim \mathcal{E}$. The theorems are instantiated on the whitened belief space $\mathcal{E} = \operatorname{range}(\Sigma_{\hat{x}})$ rather than on all of $\mathbb{R}^d$, so the rank is asked against $\operatorname{rank}(\Sigma_{\hat{x}})$, which is 8 at d=16, n_o=4, 6 at d=8, n_o=3 and 8 under full observation, not against d. It holds on all 2200 records (110 configurations $\times$ 20 seeds), each formed from 7300 anchors per record. The anchors are the Kalman belief means at every step at which the encoder emits a code on the held-out evaluation trajectories, collected in trajectory order rather than subsampled, so the count per record is fixed by the length and number of those trajectories and not by a sampling rule; each of a configuration's 20 seeds draws its own 7300. The rank is read off the singular values of the matrix whose columns are those anchors, a singular value counting as nonzero when it exceeds 10^{-10} times the largest. The smallest-to-largest ratio stayed at or above 0.40 on every record, so 0.40 is the observed margin rather than a criterion, and the verdict does not turn on where the 10^{-10} threshold is placed.

Formal statement. Every clause is formalised in Lean, and the formalisation is authoritative for the statement. `nonlinear_atomRisk_lower_bound` (line 316 of this theorem's development) is clause 3's lower bound and nothing else. By clause: clause 1 is `atomRisk_eq_hsSq` (line 225); clause 2 is `collapse_pointwise` (line 144) with `collapse_eq_iff` (line 158) and `linSurrogate_finrank_le` (line 165); clause 3 is `nonlinear_atomRisk_lower_bound` (line 316) with its attainment `linear_optimum_attained` (line 329); clause 4 is `nonlinear_retention` (line 469), which bundles (a) `nonlinear_optimal_eq_linSurrogate` (line 377), (b) `nonlinear_optimal_factors_through_decision` (line 416) and (c) `nonlinear_optimal_annihilates_irrelevant` (line 432); and clause 5 is `wit_sharpness` (line 706), with the value $1/4$ from `wit_nonlinear_risk` (line 667).

Axiom dependence. The Lean development of Theorem 3 declares no axiom and no opaque symbol; every audited declaration depends only on Lean's standard axioms (propext, Classical.choice, Quot.sound).

Proof. The proof proceeds as follows, in 10 numbered steps.

1. **Pythagoras at a subspace.** Let $S \leq \mathcal{G}$ be a subspace, $y \in \mathcal{G}$ and $s \in S$. Write $y - s = (y - \Pi_S y) + (\Pi_S y - s)$. The first summand lies in $S^\perp$ and the second in S, so their inner product vanishes and expanding the squared norm gives
$$\|y - s\|^2 \;=\; \|y - \Pi_S y\|^2 + \|\Pi_S y - s\|^2.$$
Two consequences are used below: $\|y - \Pi_S y\| \leq \|y - s\|$, since the discarded term is nonnegative; and if the two squared errors are equal then $\|\Pi_S y - s\|^2 = 0$, so $s = \Pi_S y$.

2. **Linear-decoder collapse.** Let $M : \mathcal{Z} \to \mathcal{G}$ be linear and $W_{\mathrm{enc}} : \mathcal{E} \to \mathcal{Z}$ any function. Whatever the encoder does, $M(W_{\mathrm{enc}}(x))$ lies in $S = \operatorname{range} M$. Applying step 1 with $y = Dx$ and $s = M(W_{\mathrm{enc}}(x))$ gives the pointwise inequality of clause 2 together with its equality case. The surrogate $X^\star = \Pi_S \circ D$ is a composite of linear maps, hence linear, and its range is contained in S, so its rank is at most $\dim S \leq \dim \mathcal{Z}$. This is where the encoder's nonlinearity is discarded, and it does not reappear: a linear decoder confines the composite to a subspace of dimension at most $\dim \mathcal{Z}$, and the best point of a subspace is the orthogonal projection, which is a linear function of Dx.

3. **The risk correspondence.** Take m and the atom factor $R : \mathbb{R}^m \to \mathcal{E}$ as in the statement. Evaluating the atom factor on the standard orthonormal basis of $\mathbb{R}^m$ gives $Re_j = \sqrt{p_j}\, a_j$. Since the Hilbert–Schmidt energy of a linear map is independent of the orthonormal basis

used to compute it, for any linear $Y : \mathcal{E} \to \mathcal{G}$

$$\|YR\|_{HS}^2 \;=\; \sum_j \|Y(\sqrt{p_j}\, a_j)\|^2 \;=\; \sum_j p_j \|Y a_j\|^2,$$

using $(\sqrt{p_j})^2 = p_j$ for $p_j \geq 0$. Taking $Y = D - X$ for a linear X gives clause 1: the L^2 risk of a linear map under π is the covariance-weighted Hilbert–Schmidt risk with the atom factor as covariance factor.

4. **Monotonicity of the risk under the collapse.** Step 2 is an inequality between nonnegative quantities at each atom. Squaring, multiplying by $p_j \geq 0$ and summing gives

$$\mathcal{R}_\pi(X^\star) \;\leq\; \mathcal{R}_\pi(M \circ W_{\mathrm{enc}}).$$

5. **The lower bound.** Take $\mathcal{Z}' = \operatorname{range} M$, a subspace of $\mathcal{G}$ with $\dim \mathcal{Z}' \leq \dim \mathcal{Z} \leq k$; let M' be its inclusion into $\mathcal{G}$ and let $W' = P_{\mathcal{Z}'} \circ D \circ R$, viewed as a map into $\mathcal{Z}'$. Then $M' \circ W' = X^\star \circ R$ by the definition of the orthogonal projection, so by step 3

$$\mathcal{R}_\pi(X^\star) \;=\; \|(D - X^\star)R\|_{HS}^2 \;=\; \|DR - M' \circ W'\|_{HS}^2.$$

The rank-constrained retention theorem's lower bound, applied to the operator DR on $\mathbb{R}^m$ with the linear capacity-k pair (M', W'), bounds the right side below by $\sum_{i \geq k} \mu_i$. Chaining with step 4 gives clause 3's lower bound for the pair with a nonlinear encoder.

6. **Attainment.** The rank-constrained retention theorem also supplies a linear capacity-k pair (M_0, W_0) for the operator DR meeting the bound exactly. Let $S_0 = \operatorname{range} M_0$ and apply step 2 inside the Hilbert–Schmidt sum, at each vector of an orthonormal basis of $\mathbb{R}^m$:

$$\|DR - (\Pi_{S_0} \circ D)R\|_{HS}^2 \;\leq\; \|DR - M_0 \circ W_0\|_{HS}^2 \;=\; \sum_{i \geq k} \mu_i.$$

Step 5 gives the reverse inequality for $\Pi_{S_0} \circ D$, so that linear map on $\mathcal{E}$ attains the optimum. Together with step 5 this proves that the optimum over arbitrary encoders and the optimum over linear encoders coincide. The collapse of step 2 is used twice: once to bound the nonlinear pairs from below, and once to carry the attaining pair from the whitened domain back to $\mathcal{E}$.

7. **The optimal composite is linear on the support.** Suppose $\mathcal{R}_\pi(M \circ W_{\mathrm{enc}}) \leq \sum_{i \geq k} \mu_i$. Steps 4 and 5 give

$$\sum_{i \geq k} \mu_i \;\leq\; \mathcal{R}_\pi(X^\star) \;\leq\; \mathcal{R}_\pi(M \circ W_{\mathrm{enc}}) \;\leq\; \sum_{i \geq k} \mu_i,$$

so all four quantities are equal. The difference of the two risks is a sum of terms $p_j\big(\|Da_j - M(W_{\mathrm{enc}}(a_j))\|^2 - \|Da_j - X^\star a_j\|^2\big)$, each nonnegative by step 2 and summing to zero, hence each zero. With $p_j > 0$ the two squared errors agree at every atom, and the equality case of step 1 gives $M(W_{\mathrm{enc}}(a_j)) = X^\star a_j$. This is clause 4(a).

8. **Factoring through the decision image.** By step 7 the composite equals $X^\star = \Pi_S \circ D$ at every atom, and $\Pi_S \circ D$ depends on its argument only through D of it. So $Da_j = Da_{j'}$ forces $M(W_{\mathrm{enc}}(a_j)) = \Pi_S(Da_j) = \Pi_S(Da_{j'}) = M(W_{\mathrm{enc}}(a_{j'}))$. This is clause 4(b), and it uses neither the spectral gap nor the spanning hypothesis.

9. **The decision-irrelevant subspace is sent to zero.** Apply the decision-weighted retention theorem's encoder form to the decision operator DR on $\mathbb{R}^m$, with the linear capacity-k pair (M', W') of step 5, whose risk is at most $\sum_{i \geq k} \mu_i$ by steps 4 and 5. Its conclusion is that $(DR)u = 0$ implies $(M' \circ W')u = X^\star(Ru) = 0$. If the atoms span $\mathcal{E}$ then R is surjective, because each a_j is R applied to $p_j^{-1/2} e_j$; so a decision-irrelevant $v \in \mathcal{E}$ equals Ru for some u, and $D(Ru) = Dv = 0$ places u in the kernel of DR. Hence $X^\star v = 0$, which is clause 4(c).

10. **The sharpness witness.** Fix an orthonormal basis b_0, b_1, b_2 of a three-dimensional $\mathcal{E}$, take $\mathcal{G} = \mathcal{E}$ and D the orthogonal projection onto $\operatorname{span}(b_0, b_1)$, so $\ker D = \operatorname{span}(b_2)$ and $\operatorname{rank} D = 2$. Take the four atoms $a = (b_0,\ b_1,\ b_1 + b_2,\ b_1 - b_2)$ with unit weights, and $k = 1$. *The floor for linear decoders.* Let $\dim \mathcal{Z} \le 1$ and let $M : \mathcal{Z} \to \mathcal{E}$ be linear. Every element of $\mathcal{Z}$ is a scalar multiple of a single z_0, so every decoder value is a multiple of $w = Mz_0$. Discarding the last two atoms, whose contributions are nonnegative, leaves $\|b_0 - c_0 w\|^2 + \|b_1 - c_1 w\|^2$ for some reals c_0, c_1. Writing $\alpha = \langle b_0, w\rangle$, $\beta = \langle b_1, w\rangle$ and $N = \|w\|^2$ and expanding, this equals $2 - 2c_0\alpha + c_0^2 N - 2c_1\beta + c_1^2 N$, so the claim $\mathcal{R}_\pi \ge 1$ reduces to $c_0^2 N - 2c_0\alpha + c_1^2 N - 2c_1\beta + 1 \ge 0$. Multiplying by N,

$$N\big(c_0^2 N - 2c_0\alpha + c_1^2 N - 2c_1\beta + 1\big) = (c_0 N - \alpha)^2 + (c_1 N - \beta)^2 + \big(N - \alpha^2 - \beta^2\big),$$

which is nonnegative because Bessel's inequality for the orthonormal pair (b_0, b_1) gives $\alpha^2 + \beta^2 \le N$. For $N > 0$ the claim follows on dividing; for $N = 0$ the vector w vanishes, so $\alpha = \beta = 0$ and the risk is 2. *The pair with a nonlinear decoder.* Take $\mathcal{Z} = \mathbb{R}$, the linear encoder

$$W_{\mathrm{enc}}(x) = \langle b_0, x\rangle + 2\langle b_1, x\rangle + 4\langle b_2, x\rangle,$$

whose values at the four atoms are $1, 2, 6, -2$, and the decoder

$$M(t) = \begin{cases} b_0 & t = 1, \\ \frac{1}{2} b_1 & t = -2, \\ b_1 & \text{otherwise.} \end{cases}$$

The decision images of the four atoms are b_0, b_1, b_1, b_1, so the composite reproduces the decision image exactly at the first three atoms and returns $\frac{1}{2} b_1$ at the fourth. The risk is therefore $\|b_1 - \frac{1}{2} b_1\|^2 = \frac{1}{4}$, which is below the floor 1. Both halves of clause 5 follow.

□

A.4 Realizability Test: A Decidable Pre-Flight Check for Theorem 2's Reweighting Condition, with a Quantitative Obstruction Dimension

Theorem 4 (Realizability Test: A Decidable Pre-Flight Check for Theorem 2's Reweighting Condition, with a Quantitative Obstruction Dimension)**.** *Write* $\mathrm{Obs}(T, D) = (\ker T \cap \ker D)^\perp \cap \ker T$ *for the* ***obstruction subspace*** *and* $\omega(T, D) = \dim \mathrm{Obs}(T, D)$ *for the* ***obstruction dimension****, the quantity the title names. Assume (A1)* $\mathcal{E}$*,* $\mathcal{F}$*,* $\mathcal{G}$ *are finite-dimensional real inner product spaces; (A2)* T *and* D *are linear; (A3)* T *and* D *are the* population *operators of the generator, not estimates; (A4) clause 5 additionally assumes the decision criterion of §3 and the stacked factorisation* $D^*D = P_{t_h}$ *of Equation 8; (A5) the conclusion is about* existence *of a realizing weight, not about benefit. Then:*

1. ***(The test.)*** $\ker T \subseteq \ker D$ *if and only if* $\mathrm{Obs}(T, D) = \{0\}$*, if and only if* $\omega(T, D) = 0$*. Theorem 2 clause 1's factorisation question is therefore decided by* ω*.*

2. ***(Rank form.)*** $\dim(\ker T \cap \ker D) + \omega(T, D) = \dim \ker T$*, hence* $\omega(T, D) = \dim \ker T - \dim(\ker T \cap \ker D)$ *and the test is the comparison* $\dim(\ker T \cap \ker D) = \dim \ker T$*. No subspace has to be constructed to* decide *the test, only to exhibit a basis of the obstruction.*

3. ***(The dichotomy.)*** *There exists a linear* $W : \mathcal{F} \to \mathcal{F}$*, positive semidefinite or not, with* $T^*WT = P$ ***if and only if*** $\omega(T, D) = 0$*. When* $\omega(T, D) = 0$ *the weight may be taken positive semidefinite, as* $W = V^*V$ *for a factor* V *with* $V \circ T = D$*; when* $\omega(T, D) \ge 1$*, every linear* W *fails.*

4. ***(*ω ***counts unreachable directions.)*** *Every nonzero* $v \in \mathrm{Obs}(T, D)$ *satisfies* $Tv = 0$ *and* $Dv \ne 0$*, hence carries strictly positive decision energy* $\langle Pv, v\rangle = \|Dv\|^2 > 0$*, and* $\mathrm{Obs}(T, D) \ne \{0\}$ *exactly when some* v *has* $Tv = 0$ *and* $Dv \ne 0$*, which is precisely the hypothesis of Theorem 2 clause 2.*

5. ***(Instantiation on the linear-quadratic generator.)*** *For the stacked decision operator built from* $(A_{\mathrm{cl}}, H, K_c, R_f)$ *at horizon* t_h*, with* $R = R_f^\top R_f$ *here a control weight and not the covariance factor (*$R_f = \sqrt{\rho}\, I$ *in our generator),* $\ker D$ *may be replaced throughout by*

$\ker P_{t_h}$*, since* $D^*D = P_{t_h}$*:* $\mathrm{Obs}(T, D) = (\ker T \cap \ker P_{t_h})^\perp \cap \ker T$*, the test is the nullity comparison* $\dim(\ker T \cap \ker P_{t_h}) = \dim \ker T$*, and the obstruction is non-trivial exactly when some belief direction is sent to zero by the predictive operator and is* not *decision-irrelevant. Both halves of clause 3 are restated at that concrete operator.*

6. ***(Why the complement is taken inside*** $\ker T$***.)*** *The candidate* $\mathrm{Obs}'(T, D) = \ker T \cap (\ker D)^\perp$ *is always contained in* $\mathrm{Obs}(T, D)$*, but it does not decide the question: there are a two-dimensional* $\mathcal{E}$ *and operators* T*,* D *with* $\mathrm{Obs}'(T, D) = \{0\}$*,* $\ker T \not\subseteq \ker D$*,* $\omega(T, D) \geq 1$*, and no linear* W *satisfying* $T^*WT = P$*. A test built on* Obs' *would report "realizable" there.*

Clause 2's arithmetic holds exactly on all 720 per-seed records, on every one of which the realizability indicator equals the indicator that $\omega = 0$; this neither confirms nor refutes the theorem, since the operators are computed in floating point from the generator's matrices rather than exactly. The rank tolerance is fixed in the code: a singular value of T or D below 10^{-10} times the largest counts as zero when forming the two kernels, and a principal-angle cosine between them above $1 - 10^{-8}$ counts as a shared direction when forming $\ker T \cap \ker D$.

Formal statement. Every clause is formalised in Lean, and the formalisation is authoritative for the statement. `generator_dichotomy` (line 452 of this theorem's development) is clause 5's dichotomy at the concrete operator. By clause: clause 1 is `obsSpace_eq_bot_iff` (line 117) with `obsDim_eq_zero_iff` (line 133) and `realizable_iff_obsDim_eq_zero` (line 139); clause 2 is `finrank_inf_add_obsDim` (line 150) with `obsDim_eq_sub` (line 158); clause 3 is `weight_exists_iff` (line 263), whose halves are `exists_psd_weight_of_obsDim_eq_zero` (line 240) and `no_weight_of_obsDim_ne_zero` (line 251); clause 4 is `obs_decision_ne_zero` (line 187) with `obs_decision_energy_pos` (line 201) and `obsSpace_ne_bot_iff` (line 211); clause 5 is `generator_dichotomy` (line 452) with `obsSpace_decLin` (line 411), `realizable_iff_nullity_costToGo` (line 423) and `obsSpace_ne_bot_iff_exists_relevant_blind` (line 436); and clause 6 is `naiveObs_le_obsSpace` (line 286) with `naive_test_unsound` (line 383).

Axiom dependence. The Lean development of Theorem 4 declares no axiom and no opaque symbol; every audited declaration depends only on Lean's standard axioms (propext, Classical.choice, Quot.sound).

Relation to the literature. The algebra of this test is classical and we claim none of it. The dichotomy of clause 3 is two published results: the positive semidefinite half is Douglas's range-inclusion theorem (Douglas, 1966), which gives the factor V with $V \circ T = D$ under $\ker T \subseteq \ker D$ and the weight as $W = V^*V$; the unrestricted-weight half is Penrose's consistency criterion $AA^+CB^+B = C$ for $AXB = C$ (Penrose, 1955). Khatri & Mitra (1976) state the same criterion as a matrix equation and characterise its Hermitian and nonnegative definite solutions. Clauses 1 and 2 are an orthogonal decomposition plus rank-nullity, which is textbook material. The condition being tested comes from the decision-aware model-learning line (Grimm et al., 2020; Farahmand et al., 2017; Farahmand, 2018; Nikishin et al., 2022; Voelcker et al., 2022) and, earlier, from identification for control (Gevers, 1993; Zang et al., 1995; Rivera et al., 1992); both weight a model-learning or identification criterion by the control objective without giving a necessary and sufficient condition for such a weight to exist.

What we put forward is not the algebra but its use: that $\omega(T, D) = \dim \ker T - \dim(\ker T \cap \ker D)$ is worth computing before fitting, that it counts the decision-relevant directions the predictive operator has already discarded, and that a tempting shortcut, testing $\ker T \cap (\ker D)^\perp$, is unsound; clause 6 exhibits a two-dimensional instance where it reports "realizable" and no weight exists. We are not aware of prior work applying this criterion in reinforcement learning, representation learning or world-model learning, so what we claim here is the application rather than the mathematics.

Proof. The proof proceeds as follows, in 13 numbered steps.

1. **Notation.** Write $S = \ker T$, $U = \ker D$, $N = S \cap U$ and $\mathrm{Obs} = N^{\perp} \cap S$, so $N \subseteq S$ and, by definition, $\mathrm{Obs} \subseteq S$; and write $\omega = \dim \mathrm{Obs}$, as in the statement.

2. **The orthogonal decomposition of $\ker T$ along N.** In a finite-dimensional inner product space, a subspace contained in another spans it together with the part of the larger one orthogonal to it: from $N \subseteq S$ one gets $N + (N^{\perp} \cap S) = S$, i.e. $N + \mathrm{Obs} = S$. Since $\mathrm{Obs} \cap N \subseteq N^{\perp} \cap N = \{0\}$, the sum is direct: $S = N \oplus \mathrm{Obs}$.

3. **Clause 1, sufficiency.** Suppose $S \subseteq U$. Then $N = S \cap U = S$, so $\mathrm{Obs} = S^{\perp} \cap S = \{0\}$, because a subspace meets its own orthogonal complement only in the origin.

4. **Clause 1, necessity.** Suppose $\mathrm{Obs} = \{0\}$. By step 2, $S = N + \{0\} = N = S \cap U \subseteq U$, which is $\ker T \subseteq \ker D$. Together with step 3 this proves the first equivalence of clause 1. The second, $\mathrm{Obs} = \{0\}$ if and only if $\omega = 0$, is the statement that a subspace of a finite-dimensional space is trivial exactly when its dimension is zero.

5. **Clause 2.** Applying to the pair $N \subseteq S$ the identity that the dimension of a subspace and the dimension of the part of the ambient subspace orthogonal to it add to the dimension of the ambient subspace, one gets $\dim N + \dim(N^{\perp} \cap S) = \dim S$, which is $\dim(\ker T \cap \ker D) + \omega = \dim \ker T$. Subtracting gives the difference form. The identity used is the one whose proof is step 2 together with the fact that the dimensions of a direct sum add; no hypothesis beyond finite dimension of $\ker T$ enters.

6. **Clause 4, elementwise.** Let $v \in \mathrm{Obs}$. Then $v \in S$, so $Tv = 0$. Suppose in addition $v \neq 0$ and, for contradiction, $Dv = 0$. Then $v \in S \cap U = N$; but $v \in N^{\perp}$ as well, so $\langle v, v \rangle = 0$ and $v = 0$, contradicting $v \neq 0$. Hence $Dv \neq 0$, and since the decision Gramian's quadratic form at v equals $\|Dv\|^2$, that energy is strictly positive.

7. **Clause 4, existence form.** If $\mathrm{Obs} \neq \{0\}$, pick $v \in \mathrm{Obs}$ with $v \neq 0$; step 6 gives $Tv = 0$ and $Dv \neq 0$. Conversely, if some v has $Tv = 0$ and $Dv \neq 0$ then $v \in S$ and $v \notin U$, so $S \not\subseteq U$, and clause 1 gives $\mathrm{Obs} \neq \{0\}$.

8. **Clause 3, the positive half.** Suppose $\omega = 0$. By clause 1, $\ker T \subseteq \ker D$, and by the realizability half of Theorem 2 clause 1 there is a linear $V : \mathcal{F} \to \mathcal{G}$ with $V \circ T = D$. Put $W = V^{*}V$. Then $\langle Wy, y \rangle = \|Vy\|^2 \geq 0$, so W is positive semidefinite, and $T^{*}WT = (VT)^{*}(VT) = D^{*}D = P$ by the adjoint of a composite. Hence a realizing weight exists, and one may take it positive semidefinite.

9. **Clause 3, the negative half.** Suppose $\omega \geq 1$. By clauses 1 and 4 there is a v with $Tv = 0$ and $Dv \neq 0$. For any linear W whatsoever, $(T^{*}WT)v = T^{*}W(Tv) = 0$, so $\langle (T^{*}WT)v, v \rangle = 0$, whereas $\langle Pv, v \rangle = \|Dv\|^2 > 0$. Hence $T^{*}WT \neq P$. This is the necessity half of Theorem 2 clause 2, and positive semidefiniteness of W is not used.

10. **Clause 3, as an equivalence.** Steps 8 and 9 give both directions: a realizing weight exists if and only if $\omega = 0$, and the positive semidefinite refinement is free. The content added over Theorem 2 is only the bookkeeping that joins its two clauses into one two-valued test, and the identification of the deciding quantity as a difference of two nullities by clause 2.

11. **Clause 5.** The stacked-factorisation identity $D^{*}D = P_{t_h}$ of Equation 8 gives, for the stacked decision operator built from $(A_{\mathrm{cl}}, H, K_c, R_f)$ at horizon t_h, that the Gramian of that operator is the cost-to-go matrix P_{t_h} and hence that the two operators have the same kernel, $\ker D = \ker P_{t_h}$. Substituting this equality into the definition of $\mathrm{Obs}(T, D)$ gives the displayed form of clause 5. Substituting it into clause 4's existence form, and reading "$P_{t_h} v = 0$" as decision-irrelevance by the characterisation above, transported through that same identity, gives the second sentence of clause 5: the obstruction is non-trivial exactly when some predictively invisible belief direction is not decision-irrelevant.

12. **Clause 6, the containment.** If $v \in \ker T \cap (\ker D)^{\perp}$ then $v \in S$, and v is orthogonal to every element of U, hence in particular to every element of $N = S \cap U \subseteq U$. So $v \in N^{\perp} \cap S = \mathrm{Obs}$, giving $\mathrm{Obs}' \subseteq \mathrm{Obs}$.

13. **Clause 6, the separating instance.** Let $\mathcal{E}$ be two-dimensional with orthonormal basis b_0, b_1, and set
$$Tx = \big(\langle b_0, x\rangle - \langle b_1, x\rangle\big)\, b_0, \qquad Dx = \langle b_1, x\rangle\, b_1.$$
Both are linear. First, $b_0 \in \ker D$ because $\langle b_1, b_0\rangle = 0$. Now let $v \in \ker T \cap (\ker D)^{\perp}$. From $v \perp \ker D$ and $b_0 \in \ker D$, $\langle b_0, v\rangle = 0$. From $Tv = 0$ and $b_0 \neq 0$, $\langle b_0, v\rangle = \langle b_1, v\rangle$, so $\langle b_1, v\rangle = 0$ too, and Parseval gives $\|v\|^2 = \langle b_0, v\rangle^2 + \langle b_1, v\rangle^2 = 0$, so $v = 0$. Hence $\mathrm{Obs}' = \{0\}$. Second, put $v_* = b_0 + b_1$. Then $\langle b_0, v_*\rangle = \langle b_1, v_*\rangle = 1$, so $Tv_* = 0$, while $Dv_* = b_1 \neq 0$. By step 7, $\mathrm{Obs} \neq \{0\}$, so $\omega \geq 1$, and by step 9 no linear W satisfies $T^*WT = P$ on this instance. The two subspaces therefore differ, and only $\mathrm{Obs}(T, D)$ as defined above decides the question.

□

A.5 Approximately Optimal Retention: What a δ-Suboptimal Capacity-Limited Encoder Retains

Theorem 5 (Approximately Optimal Retention: What a δ-Suboptimal Capacity-Limited Encoder Retains)**.** *Let $\mathcal{E}$ be a real inner product space with $\dim \mathcal{E} = n$, let $D : \mathcal{E} \to \mathcal{G}$ be a linear **decision operator** into a real inner product space $\mathcal{G}$, and let $\mu_0 \geq \mu_1 \geq \cdots \geq \mu_{n-1} \geq 0$ be the eigenvalues of the decision Gramian $P = D^*D$ with orthonormal eigenbasis $b_0, \ldots, b_{n-1}$ (the notation of Theorem 1).*

Assume (A1) $\mathcal{E}$ and $\mathcal{G}$ are finite-dimensional real inner product spaces; (A2) D is linear, and the encoder and decoder are linear; (A3) $0 < k < n$; (A4) Strict decision spectral gap at the capacity boundary, $\mu_k < \mu_{k-1}$; (A5) $\mu_{k-1} > 0$ for clauses 3 and 5; (A6) In clause 5, an operator bound C_X with $\|Xx\| \leq C_X\|x\|$; (A7) δ is given, not derived (the runs take δ from Equation 4 of Appendix H.7 and measure C_X in Appendix H.19). For the code dimension k of (A3) and the slack δ, write

$$\eta \;=\; \frac{\delta}{\mu_{k-1} - \mu_k}$$

*for the **leakage bound**, whose denominator is positive by (A4). Then:*

1. δ-relaxed knapsack bound. *Let $\lambda : \{0, \ldots, n-1\} \to \mathbb{R}$ be antitone and nonnegative and let $w_i \in [0, 1]$ satisfy $\sum_i w_i \leq k$. If $\lambda_k < \lambda_{k-1}$ and*

$$\Big(\sum_{i<k} \lambda_i\Big) - \delta \;\leq\; \sum_i \lambda_i w_i,$$

then $\sum_{i\geq k} w_i \leq \delta/(\lambda_{k-1} - \lambda_k)$. At $\delta = 0$ every w_i with $i \geq k$ is zero, which is Theorem 1's exact statement.

2. Frame form. *Let $\{u_j\}_{j\in s}$ be an orthonormal family in $\mathcal{E}$ with $|s| \leq k$ satisfying the δ-relaxed retention condition*

$$\Big(\sum_{i<k} \mu_i\Big) - \delta \;\leq\; \sum_{j\in s} \|Du_j\|^2.$$

Then, writing $w_i = \sum_{j\in s} \langle b_i, u_j\rangle^2$ for the weight the family places on the i-th decision eigendirection (the notation of Theorem 1),

$$\sum_{i\geq k} w_i \;\leq\; \eta.$$

3. Leakage into the decision-irrelevant subspace. *Under the hypotheses of clause 2 together with $\mu_{k-1} > 0$, every $v \in \mathcal{E}$ with $Dv = 0$ satisfies*

$$\sum_{j\in s}\langle u_j, v\rangle^2 \;\le\; \eta\,\|v\|^2.$$

At $\delta = 0$ each inner product is exactly zero, which is Theorem 2's exact statement.

4. Per-direction decision energy floor. *Under the hypotheses of clause 2, for every $j \in s$,*

$$\|Du_j\|^2 \;\ge\; \mu_{k-1}\,(1-\eta).$$

At $\delta = 0$ this reads $\|Du_j\|^2 \ge \mu_{k-1}$, which is Theorem 2's exact statement.

5. Pair form. *Let $X = M \circ W_{\mathrm{enc}}$ be a capacity-k encoder/decoder pair ($W_{\mathrm{enc}} : \mathcal{E} \to \mathcal{Z}$ and $M : \mathcal{Z} \to \mathcal{G}$ linear with $\dim \mathcal{Z} \le k$) whose Hilbert–Schmidt risk is δ-suboptimal,*

$$\|D - X\|_{HS}^2 \;\le\; \Big(\sum_{i\ge k}\mu_i\Big) + \delta,$$

and let C_X satisfy $\|Xx\| \le C_X\|x\|$ for all x. Then, assuming in addition $\mu_{k-1} > 0$, every v with $Dv = 0$ satisfies

$$\|Xv\|^2 \;\le\; C_X^2\,\eta\,\|v\|^2.$$

At $\delta = 0$ the conclusion is $Xv = 0$, which is Theorem 2's exact statement.

Formal statement. Every clause is formalised in Lean, and the formalisation is authoritative for the statement. `approx_optimal_encoder_leakage` (line 521 of this theorem's development) is clause 5's pair form and nothing else. By clause: clause 1 is `knapsack_delta_leakage` (line 146); clause 2 is `retained_leakage_le` (line 214); clause 3 is `retained_irrelevant_leakage` (line 255); clause 4 is `retained_decision_energy_floor` (line 357); and clause 5 is `approx_optimal_encoder_leakage` (line 521) with its exact-optimum degeneration `approx_optimal_encoder_annihilates_irrelevant` (line 586).

Axiom dependence. The Lean development of Theorem 5 declares no axiom and no opaque symbol; every audited declaration depends only on Lean's standard axioms (propext, Classical.choice, Quot.sound).

Belief covariance and the operator in play. The theorem holds for any linear D, and the runs apply it with D replaced by $T' \circ R$, the operator the arm regresses onto (Equation 4) composed with the belief-covariance factor, as clause 6 of Theorem 1 does. Its μ_i are then the eigenvalues of $R^*(T'^*T')R$, and the frame weight is taken in whitened coordinates. On a reweighted arm under realizability $T' \circ R = D \circ R$, so clause 3 bounds leakage into $\ker(DR)$; on an unweighted arm $T' = T$, and the bound concerns the discarded block of TR's spectrum, not $\ker D$.

Proof. The proof proceeds as follows, in 8 numbered steps.

1. **The knapsack inequality with its loss term made explicit.** Let λ be antitone and nonnegative, let $w_i \in [0,1]$ with $\sum_i w_i \le k$, and write S for the dominant block, $g = \lambda_{k-1}$ and $\tau = \lambda_k$. For $i \in S$ antitonicity gives $\lambda_i \ge g$, and multiplying the nonpositive quantity $w_i - 1$ by $\lambda_i \ge g \ge 0$ yields the termwise bound $g(1-w_i) \le \lambda_i - \lambda_i w_i$. Summing over the k indices of S,
$$\sum_{i\in S}\lambda_i w_i \;\le\; \sum_{i\in S}\lambda_i \;-\; g\Big(k - \sum_{i\in S} w_i\Big).$$
For $i \notin S$ antitonicity gives $\lambda_i \le \tau$, and $w_i \ge 0$, so $\sum_{i\notin S}\lambda_i w_i \le \tau\sum_{i\notin S} w_i$. The budget constraint $\sum_{i\in S} w_i + \sum_{i\notin S} w_i \le k$ gives $\sum_{i\notin S} w_i \le k - \sum_{i\in S} w_i$, which converts the

first display into $\sum_{i\in S}\lambda_i w_i \le \sum_{i\in S}\lambda_i - g\sum_{i\notin S} w_i$. Adding the two halves,

$$\sum_i \lambda_i w_i \;\le\; \sum_{i\in S}\lambda_i \;-\; (g-\tau)\sum_{i\notin S} w_i.$$

This inequality uses no attainment hypothesis. It is a statement about real numbers; Assumptions 1–3 play no role in it.

2. **Proof of clause 1.** Chain step 1 with the relaxed attainment hypothesis $\sum_{i\in S}\lambda_i - \delta \le \sum_i \lambda_i w_i$. The term $\sum_{i\in S}\lambda_i$ cancels and leaves $(g-\tau)\sum_{i\notin S} w_i \le \delta$. By Assumption 4 the factor $g-\tau$ is strictly positive, so dividing by it gives $\sum_{i\notin S} w_i \le \delta/(g-\tau)$, which is clause 1. Setting $\delta = 0$ makes the right-hand side zero; since the weights are nonnegative and their sum over the discarded block is at most zero, each of them is zero. That is the exact statement, so the relaxation is consistent with it and is conservative: at $\delta = 0$ it asserts neither more nor less.

3. **Proof of clause 2.** Apply step 2 with $\lambda_i = \mu_i$ and w_i the frame weight of $\{u_j\}_{j\in s}$. The hypotheses hold: μ is antitone and nonnegative by construction; $w_i \in [0,1]$ by Bessel and $\sum_i w_i \le k$ by Parseval and $|s| \le k$; and that same frame-weight identity rewrites the retention hypothesis $\sum_{i<k}\mu_i - \delta \le \sum_{j\in s}\|Du_j\|^2$ as $\sum_{i<k}\mu_i - \delta \le \sum_i \mu_i w_i$, which is the hypothesis step 2 needs. The conclusion is $\sum_{i\ge k} w_i \le \eta$.

4. **A decision-irrelevant direction lives in the discarded block.** Let $Dv = 0$. If $\mu_i > 0$ then $b_i \perp v$: applying the Gramian, $\langle Pb_i, v\rangle = \langle Db_i, Dv\rangle = 0$, while $Pb_i = \mu_i b_i$ gives $\langle Pb_i, v\rangle = \mu_i\langle b_i, v\rangle$, so $\mu_i\langle b_i, v\rangle = 0$ and $\langle b_i, v\rangle = 0$. For $i < k$, Assumption 5 and antitonicity give $\mu_i \ge \mu_{k-1} > 0$, so this applies to every dominant eigendirection. Expanding in the eigenbasis,

$$\langle u_j, v\rangle \;=\; \sum_i \langle u_j, b_i\rangle\langle b_i, v\rangle \;=\; \sum_{i\ge k}\langle u_j, b_i\rangle\langle b_i, v\rangle.$$

5. **Proof of clause 3.** Apply the Cauchy–Schwarz inequality to the sum in step 4:

$$\langle u_j, v\rangle^2 \;\le\; \Big(\sum_{i\ge k}\langle b_i, u_j\rangle^2\Big)\Big(\sum_{i\ge k}\langle b_i, v\rangle^2\Big).$$

Bessel's inequality bounds the second factor by $\|v\|^2$. Summing over $j \in s$ and exchanging the order of summation turns the first factor into $\sum_{i\ge k} w_i$, which step 3 bounds by η. Hence $\sum_{j\in s}\langle u_j, v\rangle^2 \le \eta\|v\|^2$. At $\delta = 0$ the right-hand side is zero, and since each summand is a square, each inner product is zero: the retained span is exactly orthogonal to every decision-irrelevant direction, which is the exact statement.

6. **Proof of clause 4.** Fix $j \in s$ and write $c_i = \langle b_i, u_j\rangle^2$. Expanding the decision energy in the eigenbasis gives $\|Du_j\|^2 = \sum_i \mu_i c_i$, and Parseval with $\|u_j\| = 1$ gives $\sum_i c_i = 1$. Since j is one of the nonnegative summands defining w_i, we have $c_i \le w_i$, so step 3 gives $\sum_{i\ge k} c_i \le \eta$ and therefore $\sum_{i<k} c_i \ge 1-\eta$. Discarding the nonnegative terms of the discarded block and bounding $\mu_i \ge \mu_{k-1}$ on the dominant block,

$$\|Du_j\|^2 \;\ge\; \sum_{i<k}\mu_i c_i \;\ge\; \mu_{k-1}\sum_{i<k} c_i \;\ge\; \mu_{k-1}(1-\eta),$$

the last step using $\mu_{k-1} \ge 0$. At $\delta = 0$ the floor is μ_{k-1}, the exact statement.

7. **A δ-suboptimal pair supplies a δ-suboptimal frame.** Let $X = M\circ W_{\mathrm{enc}}$ be as in clause 5, let $K = \ker X$ and $p = \dim K$. The rank of X is at most $\dim \mathcal{Z} \le k$, so rank-nullity gives $n - p \le k$. Take an orthonormal basis of K, place it at the first p indices, and extend it to an orthonormal basis $e_0, \dots, e_{n-1}$ of $\mathcal{E}$. On the kernel block $Xe_i = 0$, so $(D-X)e_i = De_i$ there, and discarding the nonnegative remaining terms of the Hilbert–Schmidt sum gives

$$\sum_{i<p}\|De_i\|^2 \;\le\; \|D-X\|_{HS}^2 \;\le\; \sum_{i\ge k}\mu_i + \delta.$$

By basis independence of the Hilbert–Schmidt energy, $\sum_i \|De_i\|^2 = \sum_i \mu_i = \sum_{i<k} \mu_i + \sum_{i\geq k} \mu_i$. Subtracting,

$$\sum_{i\geq p} \|De_i\|^2 \;\geq\; \sum_{i<k} \mu_i - \delta.$$

The family $\{e_i\}_{i\geq p}$ is orthonormal and has $n-p \leq k$ members, so it satisfies the hypothesis of clauses 2 and 3 with the same δ.

8. **Proof of clause 5.** Expand $v = \sum_i \langle e_i, v\rangle e_i$ in the basis of step 7 and set $v' = \sum_{i\geq p} \langle e_i, v\rangle e_i$. Because X sends to zero every e_i with $i < p$, applying X to both expansions gives $Xv = Xv'$. Orthonormality gives $\|v'\|^2 = \sum_{i\geq p} \langle e_i, v\rangle^2$, and step 5 applied to the frame of step 7 bounds that by $\eta\|v\|^2$. Assumption 6 then gives $\|Xv\| = \|Xv'\| \leq C_X\|v'\|$, and squaring both sides,

$$\|Xv\|^2 \;\leq\; C_X^2\|v'\|^2 \;\leq\; C_X^2\,\eta\,\|v\|^2.$$

At $\delta = 0$ the leakage bound is zero, so $\|Xv\|^2 \leq 0$ and $Xv = 0$: the exact statement that an optimal capacity-k pair sends to zero every decision-irrelevant direction.

□

B HYPERPARAMETERS

Every value below is the configuration of the capacity family as it was run, read from that experiment's `lqg_capacity.yaml` and the per-cell sweep overrides. Three cells are reported: d=8, n_o=3; d=16, n_o=4; and a fully observed d=8 cell. Where a setting differs between cells, all three values are given.

The generator. The LQG system is constructed deterministically from these settings: it does not depend on the seed, and every arm within a cell sees the same A, C, Σ_ε and Σ_η (Appendix H.17 alone replaces this A with seeded draws). The state matrix A is block diagonal; block b is a damped rotation with angle $\theta_b = 0.3 + 0.15b$ and radius $r_b = \mathrm{clip}(0.9 - 0.03b,\ 0.55,\ 0.95)$, and any leftover one-dimensional block is the scalar 0.6. Under partial observation C has one row per observed coordinate, each selecting the first coordinate of a distinct block; under full observation $C = I_d$. The cost applies to `system.n_cost_blocks` of the blocks, starting at `system.cost_block_start`: the performance output H and the gain K_c both read the first coordinate of each of those blocks, and B routes that block's control into the block's second coordinate, so both the read and the write coordinates of a block's control lie inside its own block and $A_{\mathrm{cl}} = A - BK_c$ stays block diagonal in the same partition. Trajectories are simulated open loop, $x_{t+1} = Ax_t + \varepsilon_t$; the control law enters only P_{t_h} and D. The noise covariances are isotropic, $\Sigma_\varepsilon = \sigma^2_{\mathrm{proc}} I$ and $\Sigma_\eta = \sigma^2_{\mathrm{obs}} I$. The control weight ρ of §3 is `system.control_weight` below.

Table 1: Hyperparameters of the capacity family. "Varies" marks a setting the sweep changes between arms rather than holding fixed.

Setting	Value	Scope
Generator		
system.state_dim	8 / 16 / 8	per cell
system.obs_dim	3 / 4 / 8	per cell
system.observation_mode	partial / partial / full	per cell
system.sigma_proc	0.1	all cells
system.sigma_obs	0.3	all cells
system.control_gain	0.2	all cells
system.control_weight (ρ)	1.0	all cells
system.horizon_T	3	all cells; *varies*
system.cost_block_start	1 / 2 / 1	per cell; *varies*
system.n_cost_blocks	2	all cells; *varies*
Data		
data.n_traj	100	all cells

Setting	Value	Scope
data.traj_length	200	all cells
data.n_traj_eval	50	all cells
data.window_length	32	all cells
data.pred_horizon	4	all cells
data.burn_in	50	all cells
Encoder		
model.encoder_type	linear	baseline; *varies*
model.head_type	linear	body arms; `mlp` in App. D.6
model.z_dim (k)	2	baseline; *varies*
model.hidden_dim	128	all arms
model.gru_layers	1	body arms; 2 in App. D.7
Objective		
objective.reweight	false	baseline; *varies*
Optimisation		
optimiser	Adam	all arms
train.lr	3×10^{-4}	all arms
train.batch_size	256	all arms
train.epochs	100	all arms
train.steps_per_epoch	50	all arms
train.grad_clip	100.0	all arms
train.num_seeds	20	all arms
train.deterministic_torch	true	all arms
Evaluation		
eval.batch_size	4096	all arms
eval.retained_rtol	1×10^{-6}	all arms

The per-arm training budget is therefore $100 \times 50 = 5000$ optimiser steps at batch size 256, identical for every arm in every cell; no arm is trained longer than another. The body's sweeps vary six knobs, and the value given above is the one the baseline arm of that cell carries. Three are the knobs the paper's contrasts are built on: `model.z_dim`, over $\{1, 2, 3, 4, d\}$ in the two partially observed cells and $\{2, 4\}$ under full observation; `model.encoder_type`, over $\{$`linear`, `gru`$\}$ in the two partially observed cells and `linear` only under full observation; and `objective.reweight`, over $\{$`false`, `true`$\}$ in all three. Three more change the decision problem itself and are varied only by the arms named for them, and only in the two partially observed cells: `system.horizon_T` takes 0 on the short-horizon arms and 3 elsewhere; `system.n_cost_blocks` takes 1 and 3 on the `Cost1` and `Cost3` arms and 2 elsewhere; and `system.cost_block_start` moves with it, and also on the realizability-violating arm, whose cost block range reaches one block past the observed ones. That arm trains on the same $V = DT^{+}$ as every reweighted arm, built from its own D; since $\ker T \not\subseteq \ker D$ there, $VT = D\Pi$ with Π the orthogonal projector onto $(\ker T)^{\perp}$, so its target reproduces only the part of D that T registers, and it is scored against its own P_{t_h}. The arm tables in Appendix E and the arm glossary of Table 22 name the setting of each. Two side sweeps add keys of their own: `objective.reweight_alpha` (Appendix D.4) and `objective.input_map` with its scale (Appendix D.9). The affine readout from the code to the belief, and the retained subspace $\hat{S}$ read off it, are fitted and scored on the same 7,300 evaluation anchors; with at most 17 regressors the in-sample optimism is at most about 0.0023 and nearly cancels in paired deltas.

C Full Statistical Results

C.1 Statistical Protocol

All comparisons follow the same paired-seed protocol. For each candidate, deltas are computed as $\Delta_i = x_i^{\text{cand}} - x_i^{\text{base}}$ per paired seed, then *oriented* so that positive Δ always indicates improvement (i.e., sign-flipped for metrics where lower is better). Two two-sided significance tests are applied to the oriented deltas:

1. **One-sample t-test** ($H_1 : \mu_\Delta \neq 0$), reported as p_t.

2. **Wilcoxon signed-rank test** (H_1: median $\Delta \neq 0$, `zero_method='wilcox'`), reported as p_w. Computed for every comparison as a non-parametric robustness check, and weighed more where Shapiro–Wilk rejects normality of the deltas ($\alpha = 0.05$).

Multiple comparison correction. Within each ablation, all candidates compared against the same baseline form one *correction family*. For families with $K > 1$ candidates, Holm–Bonferroni step-down correction is applied to both p_t and p_w across the K hypotheses. Tables that report exact per-comparison p-values give *raw* (uncorrected) values alongside a "Sig." column that indicates whether each comparison survives Holm–Bonferroni correction at $\alpha = 0.05$ (✓ = significantly better than the baseline, *worse* = significantly worse, — = does not survive Holm–Bonferroni at $\alpha = 0.05$). The intervals beside it are uncorrected, so a row can carry — and still have an interval that excludes zero; four rows of Tables 3, 4 and 5 do. Tables with a p_{corr} heading instead print the Holm–Bonferroni-corrected value directly. Corrected p-values below 10^{-3} are written as an upper bound rather than an exact value. Most tables use "< 0.001" and some use a tighter threshold such as "$< 10^{-4}$". All such strings are upper bounds on the corrected p-value, all satisfy the decision rule below, and all may be read as "< 0.001" for its purpose. The number of candidates K per correction family is stated for transparency.

The *Win* column. *Win* is the win rate: the proportion of the n paired seeds on which the oriented delta is strictly positive, that is on which the candidate beat the baseline on that seed. It is a count over seeds and carries no test. Every table that reports it heads the column *Win*.

How to read d_z in this paper. Paired Cohen's d_z is the same quantity in every table here, computed the same way from the same per-seed deltas. In the linear-Gaussian capacity family and the belief-identification family the optimisation is near-deterministic given the objective and the seed (the nonlinear pendulum of Appendix H.16, whose contrastive deltas are bimodal, is the exception), so the seed-to-seed spread of the deltas is small and d_z is correspondingly large: values in the tens record how reproducible a contrast is, not how large an effect a new setting should expect. We therefore do not rank arms by d_z anywhere, and a reader should not read it as a generalisable effect magnitude. Tables 20 and 21 head the column d_z (repro.) as a reminder of this; the number is the same number their per-configuration and statistical counterparts print unqualified.

Primary decision rule. The one-sample t-test (p_t) serves as the primary significance criterion. The Wilcoxon signed-rank test (p_w) is reported as a non-parametric robustness check. When the two tests disagree (i.e., one is significant at $\alpha = 0.05$ and the other is not), the result is considered *borderline* and is not treated as a reliable finding in the main text.

Test directionality. We use two-sided tests throughout. A one-sided test in the hypothesised direction cannot report a harmful effect: it returns $p \to 1$ however large that effect is, so under a one-sided convention the sign of Δ carries information the p-value does not. Because the oriented delta is already signed, we read direction from Δ and d_z and ask the test only whether the effect is distinguishable from zero. A two-sided value is not in general twice the one-sided one: the doubling identity holds only when the effect runs in the hypothesised direction and is $2(1 - p)$ otherwise, so a reader cannot recover one convention from the other by arithmetic. The Wilcoxon statistic is $\min(W^+, W^-)$ under a two-sided alternative, and the analysis records W^+, the sum of the signed ranks of the positive oriented deltas, for every comparison it runs, so the direction of the signed ranks is recoverable from the per-contrast record; no table here prints it.

Confidence intervals. Bootstrap 95% CIs (BCa method, 10,000 resamples) are computed on the oriented mean delta.

Effect size. Cohen's d_z (paired) is reported where applicable.

C.2 The d=16, n_o=8 belief-identification setting

Metric: affine R^2 on unobserved factors (`affine_r2_unobserved`), $n = 20$ paired seeds, 6 candidates in the correction family.

We write `Autoencoder` for the recurrent autoencoder arm (`dvae` in the supplementary code): a deterministic GRU encoder whose decoder reconstructs the current observation, with no sampling and no KL term. Its row appears in every table below. We use "Naive Hybrid" to name the arm that trains a single encoder on the sum of the reconstruction and predictive losses, with no whitening

applied to the target representation. Its row in Table 2 is labelled "Naive Hybrid D16 D8", where the suffix records the latent and observation dimensions of the configuration.

Table 2: Statistical comparisons for d=16, n_o=8. The naive hybrid arm combines the reconstruction and predictive objectives without whitening.

Candidate	Base.	Cand.	Δ	Win	d_z	p_t	p_w	95% CI	Sig.
Naive Hybrid D16 D8	0.311	0.307	-0.00324	0.35	-0.39	0.0991	0.114	[-0.00710, 1.83×10^{-5}]	—
JEPA Whiten D16 D8	0.311	0.681	0.370	1.00	9.61	2.1×10^{-20}	1.9×10^{-6}	[0.352, 0.385]	✓
JEPA None D16 D8	0.311	0.526	0.215	1.00	5.69	3.9×10^{-16}	1.9×10^{-6}	[0.198, 0.231]	✓
Autoencoder D16 D8	0.311	0.247	-0.0637	0.00	-2.90	7.1×10^{-11}	1.9×10^{-6}	[-0.0727, -0.0539]	*worse*
JEPA VICReg D16 D8	0.311	0.620	0.309	1.00	8.16	4.6×10^{-19}	1.9×10^{-6}	[0.290, 0.323]	✓
MLP Encoder D16 D8	0.311	1.74×10^{-3}	-0.309	0.00	-34.91	5.3×10^{-31}	1.9×10^{-6}	[-0.313, -0.305]	*worse*

C.3 The d=4, n_o=2 belief-identification setting

Metric: affine R^2 on unobserved factors (`affine_r2_unobserved`), $n = 20$ paired seeds, 6 candidates in the correction family.

Table 3: Statistical comparisons for d=4, n_o=2.

Candidate	Base.	Cand.	Δ	Win	d_z	p_t	p_w	95% CI	Sig.
Naive Hybrid D4 D2	0.495	0.485	-0.0108	0.45	-0.075	0.740	0.927	[-0.0913, 0.0353]	—
JEPA VICReg D4 D2	0.495	0.750	0.255	0.90	1.22	2.9×10^{-5}	2.7×10^{-5}	[0.180, 0.364]	✓
Autoencoder D4 D2	0.495	0.195	-0.301	0.05	-2.49	8.8×10^{-10}	3.8×10^{-6}	[-0.341, -0.233]	*worse*
JEPA Whiten D4 D2	0.495	0.688	0.193	0.75	0.73	0.00408	0.00831	[0.0954, 0.320]	✓
MLP Encoder D4 D2	0.495	5.11×10^{-4}	-0.495	0.00	-4.33	5.8×10^{-14}	1.9×10^{-6}	[-0.523, -0.389]	*worse*
JEPA None D4 D2	0.495	0.565	0.0694	0.75	0.35	0.133	0.0696	[0.00344, 0.176]	—

C.4 The d=16, n_o=4 capacity setting

Metric: decision-relevant error, $n = 20$ paired seeds, 15 candidates in the correction family.

Table 4: Statistical comparisons for d=16, n_o=4.

Candidate	Base.	Cand.	Δ	Win	d_z	p_t	p_w	95% CI	Sig.
A1 Unweighted K4	0.968	0.901	0.0669	1.00	2.08	1.7×10^{-8}	1.9×10^{-6}	[0.0528, 0.0804]	✓
A2 Reweighted K1	0.968	0.928	0.0401	0.90	1.20	3.4×10^{-5}	6.3×10^{-5}	[0.0258, 0.0546]	✓
A2 Reweighted K3	0.968	0.793	0.176	1.00	9.30	4.0×10^{-20}	1.9×10^{-6}	[0.167, 0.183]	✓
A2 Reweighted K2 Cost1	0.968	0.791	0.177	1.00	7.58	1.9×10^{-18}	1.9×10^{-6}	[0.165, 0.186]	✓
A2 Reweighted K2	0.968	0.845	0.123	1.00	3.44	3.5×10^{-12}	1.9×10^{-6}	[0.105, 0.136]	✓
A2 Reweighted K2 Cost3	0.968	0.844	0.124	1.00	4.88	6.4×10^{-15}	1.9×10^{-6}	[0.113, 0.135]	✓
A1 Unweighted K3	0.968	0.950	0.0181	0.70	0.71	0.0051	0.0073	[0.00614, 0.0282]	✓
A2 Reweighted K2 T0	0.968	0.698	0.270	1.00	8.60	1.7×10^{-19}	1.9×10^{-6}	[0.248, 0.279]	✓
B GRU Reweighted K2	0.968	0.849	0.119	1.00	4.41	4.1×10^{-14}	1.9×10^{-6}	[0.108, 0.131]	✓
A2 Reweighted K4	0.968	0.763	0.205	1.00	9.92	1.2×10^{-20}	1.9×10^{-6}	[0.194, 0.212]	✓
A4 Unweighted K Eq D	0.968	0.751	0.217	1.00	11.74	5.0×10^{-22}	1.9×10^{-6}	[0.207, 0.224]	✓
A4 Reweighted K Eq D	0.968	0.750	0.218	1.00	11.39	8.7×10^{-22}	1.9×10^{-6}	[0.208, 0.225]	✓
B GRU Unweighted K2	0.968	0.956	0.0117	0.60	0.38	0.105	0.143	[-0.00214, 0.0238]	—
A3 Violating Partial K2	0.968	0.891	0.0774	1.00	3.63	1.4×10^{-12}	1.9×10^{-6}	[0.0657, 0.0847]	✓
A1 Unweighted K1	0.968	0.982	-0.0137	0.35	-0.50	0.0363	0.0637	[-0.0261, -0.00265]	—

C.5 The d=8, n_o=3 capacity setting

Metric: decision-relevant error, $n = 20$ paired seeds, 15 candidates in the correction family.

Table 5: Statistical comparisons for d=8, n_o=3.

Candidate	Base.	Cand.	Δ	Win	d_z	p_t	p_w	95% CI	Sig.
A1 Unweighted K4	0.901	0.762	0.139	1.00	3.40	4.3×10^{-12}	1.9×10^{-6}	[0.123, 0.158]	✓
A2 Reweighted K1	0.901	0.874	0.0274	0.70	0.49	0.0425	0.0532	[0.00308, 0.0510]	—
A2 Reweighted K3	0.901	0.713	0.188	1.00	5.13	2.6×10^{-15}	1.9×10^{-6}	[0.172, 0.203]	✓
A2 Reweighted K2 Cost1	0.901	0.723	0.178	1.00	4.65	1.5×10^{-14}	1.9×10^{-6}	[0.161, 0.193]	✓
A2 Reweighted K2	0.901	0.778	0.123	0.95	2.17	8.7×10^{-9}	5.7×10^{-6}	[0.0931, 0.144]	✓
A2 Reweighted K2 Cost3	0.901	0.762	0.140	1.00	3.46	3.1×10^{-12}	1.9×10^{-6}	[0.122, 0.156]	✓
A1 Unweighted K3	0.901	0.828	0.0732	0.95	1.35	8.7×10^{-6}	3.8×10^{-6}	[0.0507, 0.0966]	✓
A2 Reweighted K2 T0	0.901	0.605	0.296	1.00	7.78	1.1×10^{-18}	1.9×10^{-6}	[0.280, 0.312]	✓
B GRU Reweighted K2	0.901	0.795	0.107	1.00	1.83	1.2×10^{-7}	1.9×10^{-6}	[0.0823, 0.132]	✓
A2 Reweighted K4	0.901	0.691	0.211	1.00	5.01	4.0×10^{-15}	1.9×10^{-6}	[0.191, 0.228]	✓
A4 Unweighted K Eq D	0.901	0.681	0.221	1.00	6.08	1.1×10^{-16}	1.9×10^{-6}	[0.205, 0.236]	✓
A4 Reweighted K Eq D	0.901	0.680	0.221	1.00	6.00	1.4×10^{-16}	1.9×10^{-6}	[0.205, 0.236]	✓
B GRU Unweighted K2	0.901	0.887	0.0143	0.65	0.43	0.0717	0.0897	[6.45×10^{-4}, 0.0292]	—
A3 Violating Partial K2	0.901	0.852	0.0498	0.90	1.27	1.8×10^{-5}	3.6×10^{-5}	[0.0333, 0.0665]	✓
A1 Unweighted K1	0.901	0.952	-0.0503	0.10	-1.15	5.8×10^{-5}	0.0001	[-0.0676, -0.0305]	*worse*

C.6 The fully observed d=8 setting

Metric: decision-relevant error, $n = 20$ paired seeds, 3 candidates in the correction family.

Table 6: Statistical comparisons for the fully observed d=8 system.

Candidate	Base.	Cand.	Δ	Win	d_z	p_t	p_w	95% CI	Sig.
A5 Full Unweighted K4	0.913	0.769	0.144	1.00	2.47	1.0×10^{-9}	1.9×10^{-6}	[0.120, 0.170]	✓
A5 Full Reweighted K2	0.913	0.671	0.242	1.00	4.56	2.2×10^{-14}	1.9×10^{-6}	[0.220, 0.266]	✓
A5 Full Reweighted K4	0.913	0.541	0.372	1.00	8.79	1.2×10^{-19}	1.9×10^{-6}	[0.355, 0.391]	✓

D Training, Compute and Premise Checks

D.1 Software Framework

All experiments were run using PyTorch 2.4.1, CUDA 12.1, Python 3.11.9 inside the pinned container image.

D.2 Compute Budget

All training runs but one sweep were performed on NVIDIA RTX 6000 Ada Generation GPUs of one machine; one 7-hour control-suite sweep ran on a second machine, with NVIDIA Quadro RTX 8000 GPUs. The Reproducibility Statement gives the totals from the recorded event logs: 10140 completed runs over 38 sweeps, 469 candidate configurations against 507 trained configurations counting one baseline per sweep, at 20 seeds each, and 18.9 days of elapsed wall-clock time. That figure is elapsed time and not a device-hour total: runs were executed concurrently across up to two devices of the first machine, and the event logs record when each sweep started and finished and, on that machine, which device each run used, not how long each device was held. The counts the next paragraph subtracts from are these.

Two control-suite experiments inside these counts. Four of the 38 sweeps are DeepMind Control Suite runs that no number in the body draws on: a value-gradient-weighted latent world model on cartpole-swingup, cheetah-run and walker-walk, trained over a grid of body-mass and joint-damping multipliers and scored by return on held-out settings (three weighted and two unweighted arms, crossed with a learned or an oracle context encoder, and a baseline; 20 seeds each), and a criticality-weighted one on walker-walk, weighting by ensemble disagreement, TD error or advantage (three arms and a baseline, 20 seeds). They are 18 of the 469 candidate configurations, 440 of the 10140 runs and 40.2 hours of the 18.9 days, and the supplementary package carries their code in its `vagram_dmc` and `ensemble_criticality_dmc` directories. Both weight the observation loss by a per-sample scalar, the class that on our linear-Gaussian generator cannot realize D

(§4), and both decode through a multilayer perceptron, outside Theorem 3's linear-decoder hypothesis. On walker-walk the weighted arms without a learned context encoder do not separate from their baselines at 20 seeds (value-gradient $d_z = 0.20$, corrected $p = 0.37$; ensemble disagreement $d_z = -0.05$, corrected $p = 1.0$), and the value-gradient arm with one does ($d_z = 1.04$, corrected $p = 8.6 \times 10^{-4}$), though it changes the encoder as well as the weight. On cartpole-swingup no value-gradient arm separates from its baseline (corrected p from 0.20 to 0.33; with the learned context encoder $d_z = -0.44$), and on cheetah-run only the arm with the learned context encoder comes near ($d_z = 0.64$, corrected $p = 0.051$, corrected Wilcoxon $p = 0.0035$, borderline under the decision rule of Appendix C.1; the other two, corrected $p = 1.0$). The nulls are a failure to detect on those configurations, not evidence against the operator weight, which no arm there instantiates.

D.3 Seed Management

Seed i of a configuration is `base_seed` $+ i$, so a configuration reruns identically and configurations within a sweep are paired seed by seed.

D.4 A continuous premise-violation axis

The weight of §5 is binary: absent or realizing. Interpolating it as $V_\alpha = (1-\alpha)I + \alpha V$, defined because V is square in this cell, makes the defect in the weight equation $T^*WT = D^*D$ continuous: computed in closed form on the generator before any training, clause 4's ε falls strictly monotonically from 2.18 at $\alpha = 0$ to 3.7×10^{-16} at $\alpha = 1$. The condition $\ker T \subseteq \ker D$ holds at every α, so this axis does not pass through the realizability-violating arm. Ten arms over $\alpha \in \{0, 0.25, 0.5, 0.75, 1\}$ and $k \in \{2, 4\}$, 20 paired seeds, at d=16, n_o=4; the $\alpha \in \{0, 1\}$ endpoints reproduce the corresponding binary arms to all reported digits.

Along that axis the two things the binary weight moves together come apart. The decision-irrelevant energy of the retained subspace, its mean squared projection onto $\ker D$, falls monotonically with α, from 0.707 to 0.005 at k=2 and from 0.570 to 0.038 at k=4 (Spearman $\rho = -0.94$ in both, $p < 10^{-47}$, treating the seeds at each α as independent). The fraction of each arm's *own* attainable reduction that training captures shows no detected trend: $\rho = -0.05$ ($p = 0.69$) at k=2 over $\alpha \geq 0.5$, where the gap to the ceiling is not near zero (it is 0 at $\alpha = 0$ and 0.022 at $\alpha = 0.25$), and $+0.12$ ($p = 0.23$) at k=4 over all five, again per seed ($n = 60$ and 100). So on this generator the weight's distance from realizing controls which directions survive, and does not detectably control how much of what is attainable the encoder reaches.

D.5 Why the linear-decoder hypothesis is kept

Theorem 3 is stated for linear decoders, and that hypothesis cannot be dropped. The Lean development carries a constructive witness: a three-dimensional belief space with orthonormal basis b_0, b_1, b_2, decision operator the orthogonal projection onto $\mathrm{span}(b_0, b_1)$, four unit-weight atoms b_0, b_1, b_1+b_2, b_1-b_2, and capacity $k = 1$. On it, every capacity-1 pair with a linear decoder and an arbitrary encoder has risk at least 1, while one pair with a *linear* encoder and a *nonlinear* decoder attains risk exactly $1/4$. The separation is possible because the decision operator identifies the last two atoms while the composite does not. The witness is machine-checked on the same footing as the theorems, with no `sorry`, axiom or opaque symbol.

D.6 Whether the linear-decoder hypothesis binds on a trained run

The witness above says the hypothesis cannot be dropped from Theorem 3. Whether it binds on the runs this paper reports is a separate question, and one sweep that changes a single setting answers it for one cell. Starting from the d=16, n_o=4 setting of §4, we move a single knob, the decoder family, from the linear map M of Theorem 1 to a two-layer network of width 128 with a ReLU, and change nothing else: same generator, same linear encoder on every arm, cost on blocks 2 and 3 at the shared cost horizon t_h=3, $k \in \{2, 4\}$, six arms at 20 paired seeds each for 120 runs. The baseline is the unweighted k=2 arm with a linear head. Three of the six arms duplicate configurations the body already reports, and they reproduce them seed by seed on all $9{,}660$ scalars, so the nonlinear-head arms are read against the same numbers §5 gives.

Every nonlinear-head arm lies outside the hypotheses of Theorem 1, Theorem 2 and Theorem 3, which is the point of running it: no row below tests, confirms or fails any of them. The linear-head arms are not theorem tests either, those theorems being exact-optimum statements about a population risk that no finite-training run satisfies.

Table 7: Nonlinear decoder at d=16, n_o=4, against the shared unweighted linear-head k=2 baseline (mean 0.968). Primary metric is decision-relevant error, lower is better, so positive Δ is a reduction in decision-relevant error. Intervals are BCa; p_{corr} is the Holm–Bonferroni correction of the two-sided paired t-test over the five candidates, read as in Table 21, so the small p_{corr} on the negative row records a significant *increase* in error. d_z carries the reading of Appendix C.1.

Arm	Head	k	Mean	Δ	95% BCa	d_z	p_{corr}	Win
Reweighted	linear	4	0.763	$+0.205$	$[0.194, 0.212]$	9.92	4.7×10^{-20}	20/20
Reweighted	mlp	4	0.764	$+0.204$	$[0.196, 0.211]$	11.83	2.1×10^{-21}	20/20
Reweighted	mlp	2	0.814	$+0.154$	$[0.145, 0.160]$	8.71	4.1×10^{-19}	20/20
Reweighted	linear	2	0.845	$+0.123$	$[0.105, 0.136]$	3.44	7.1×10^{-12}	20/20
Unweighted	mlp	2	0.998	-0.0302	$[-0.0399, -0.0232]$	-1.58	9.7×10^{-7}	0/20

That table shares one unweighted linear-head baseline, so it confounds the head with the reweighting. The contrast the sweep is for is the head at the same objective and the same capacity. Both members of each pair are paired against the same baseline seed by seed, so the per-seed difference follows by subtraction. Oriented so that positive means the nonlinear head is better, the mean difference is -0.0302, BCa $[-0.0398, -0.0229]$ ($d_z = -1.58$, $0/20$ seeds) on the unweighted k=2 pair, $+0.0308$, BCa $[+0.0222, +0.0446]$ ($d_z = +1.22$, $20/20$) on the reweighted k=2 pair, and -0.000633, BCa $[-0.0062, +0.0051]$ ($d_z = -0.05$, $10/20$) on the reweighted k=4 pair, the capacity the body's headline uses. These three are post-hoc descriptive figures computed from the per-seed paired records; we ran no additional significance test on them and claim no p-value for them. At k=4 fifteen of the sixteen secondary metrics agree with the primary: every oriented mean has $|d_z| \leq 0.22$, every bootstrap interval contains zero, and every win rate falls between $7/20$ and $14/20$. The exception is affine R^2, which the nonlinear head lowers by 0.071 ($d_z = -1.28$, $0/20$ seeds): the code is a less linear function of the belief, while decision-relevant error and retention do not move. On those, that is a failure to detect a head effect on this configuration, not evidence that the head is inert in general.

At k=2 the head does move the numbers, and the objective rather than the head sets the sign. The mechanism is visible in δ, each arm's own suboptimality slack against the operator it regresses onto: the nonlinear head cuts it from 0.0285 to 0.00812 on the unweighted k=2 pair and from 0.0224 to 0.00751 on the reweighted one, against 0.0152 to 0.0149 at k=4. So at k=2 the nonlinear head carries each arm most of the way onto *its own* exact capacity-2 optimum, and the two objectives have different optima. In this cell the unweighted capacity-2 optimum is entirely decision-irrelevant, which is the situation Theorem 2 clause 5 exhibits, so converging harder onto it converges harder onto a decision-useless code: the unweighted nonlinear arm's decision-irrelevant energy is 0.981 against the linear baseline's 0.707, and its decision-relevant error 0.998 against an irreducible floor of 0.712. The reweighted optimum is decision-relevant, so the same better optimisation moves the error the other way, and the reweighting contrast at k=2 is therefore *larger* once both arms carry the nonlinear decoder, $+0.184$ ($d_z = 23.85$) against $+0.123$ ($d_z = 3.44$) with linear heads.

The reason the head changes the path and not the target is specific to this configuration. Every arm here uses a linear encoder on a linear-Gaussian generator, so the window, the belief and the code are jointly Gaussian and the population-optimal head at any fixed encoder is already linear. The network cannot lower the population risk the encoder is trained against; what it changes is the optimisation path through a bilinear encoder-decoder problem, and δ stayed strictly positive on all 120 runs, with a minimum of 0.00511. This would not transfer to a nonlinear encoder or to a non-Gaussian generator, neither of which this sweep runs.

One quantity does degrade with the premise. Theorem 5's surrogate certificate $\eta = \delta/(\mu_{k-1} - \mu_k)$ of Equation 2, with δ from the belief route of §H.7, is compared against the measured decision-irrelevant energy on each run; it exceeds it on 101 of these 120 runs, and all 19 crossings fall on the single arm that combines a nonlinear decoder with an unweighted target, against one crossing in the

640 runs of the three parent sweeps that carry a defined bound (linear encoder, $k < d$). Both moves are needed: the head takes the mean bound from 2.946 to 0.840 while the measured energy rises from 0.707 to 0.981. This is not a counterexample to Theorem 5, whose antecedent a finite-training run does not satisfy and whose conclusion concerns leakage into the discarded block of the operator the arm regresses onto, which on an unweighted arm is not the decision operator.

What this licenses, and what it does not. At d=16, n_o=4, with cost on blocks 2 and 3, at $k \in \{2, 4\}$, at one network width of 128, at this training budget, on this linear-Gaussian generator and at 20 seeds per arm, decoder linearity is not an empirically binding premise for the retention effect the body reports: the effect survives the premise's removal unchanged at the headline capacity, and where the nonlinear head does change the outcome it does so by reaching each arm's own linear-theory optimum more closely rather than by reaching a different one. We ran one cell, two capacities and one width, so nothing here bears on other cells, on wider or deeper decoders, on longer budgets or on generators outside this family.

Theorem 3 is untouched by this: clause 5 is a constructed witness, machine-checked with no `sorry`, axiom or opaque symbol, on which every capacity-1 pair with a linear decoder has risk at least 1 while a nonlinear-decoder pair attains $1/4$; an empirical null on one generator cannot bear on a statement that exhibits a specific $\mathcal{E}$, D and atom family. The theorem says the gap is possible, and the sweep says this generator at these capacities does not realize it, which is the paragraph above this one rather than a qualification of the theorem.

D.7 A DEEPER RECURRENT ENCODER

Clause 3 of Theorem 3 drops encoder linearity, and the recurrent arms the body reports exercise that at one GRU layer of width 128. One sweep moves the encoder depth one step, to two layers, and changes nothing else: same generator at d=16, n_o=4, cost on blocks 2 and 3 at the shared cost horizon t_h=3, capacity k=2 on every arm, GRU width 128 and a linear decision head wherever each exists, five arms at 20 paired seeds each for 100 runs. The baseline is the linear-encoder unweighted k=2 arm of §4. Two of the five arms are the depth-one recurrent arms the body already reports, rerun here so that the depth-two arms have a reference paired by seed inside the same sweep; they reproduce those arms identically on all 680 per-seed scalars the two sweeps record for them, over 17 metrics. That agreement is observed here, not guaranteed: PyTorch's deterministic mode ran in warn-only form, which lets an operation without a deterministic implementation run anyway. None of the five arms is a theorem test, Theorem 1 and Theorem 2 being exact-optimum statements about a population risk that no finite-training run satisfies.

Table 8: Encoder depth at d=16, n_o=4, k=2, against the shared linear-encoder unweighted k=2 baseline (mean 0.968). Primary metric is decision-relevant error, lower is better, so positive Δ is a reduction in decision-relevant error. Intervals are BCa; p_{corr} is the Holm–Bonferroni correction of the two-sided paired t-test over the four candidates of this sweep. d_z carries the reading of Appendix C.1.

Encoder	Objective	Mean	Δ	95% BCa	d_z	p_{corr}	Win
GRU, 1 layer	reweighted	0.849	$+0.119$	$[0.108, 0.131]$	4.41	1.6×10^{-13}	20/20
GRU, 2 layers	reweighted	0.860	$+0.108$	$[0.0919, 0.122]$	3.09	7.0×10^{-11}	20/20
GRU, 1 layer	unweighted	0.956	$+0.0117$	$[-0.00214, 0.0238]$	0.38	0.21	12/20
GRU, 2 layers	unweighted	0.972	-0.00376	$[-0.0175, 0.00786]$	-0.13	0.57	11/20

The reduction the reweighting produces survives the second layer. Against the shared linear-encoder baseline it is $+0.119$ at depth one and $+0.108$ at depth two, both on 20 of 20 seeds. Against the unweighted arm at the same depth, which holds capacity, encoder and horizon fixed, it is 0.108 at depth one and 0.112 at depth two, again $20/20$ on each; the depth-one value is `MW-06` of Table 19, the row these arms reproduce. Neither unweighted arm separates from the baseline, which is a failure to detect on those two configurations and not evidence that depth removes the effect. The depth contrast itself, read at a fixed objective, is small and unseparated: two layers cost 0.0154 of decision-relevant error more than one on the unweighted arms (d_z=0.46, higher on 13 of 20 seeds) and 0.0115 more on the reweighted arms (d_z=0.25, $13/20$). Like the nonlinear-head pair of Appendix D.6, the within-depth and within-objective figures in this paragraph are post-hoc and

descriptive, computed by subtraction from the per-seed paired records; we ran no additional test on them, claim no p-value for them, and do not add them to the $K{=}20$ weighting-matched family of Table 19, which carries the contrasts for which the paper does claim a corrected p. None of this paragraph reads a fitted belief-to-latent map, so none of the measurement caveat below applies to it.

Table 9: The containment metric and the fidelity of the fit it is read from, at the same four arms and the same baseline (means 0.707 and 0.926). Decision-irrelevant energy is read off the range of an ordinary least-squares fit of the belief to the latent code; affine R^2 is the aggregate R^2 of that same fit. Δ is oriented so that positive is better on each metric's own direction, lower energy and a higher R^2; p_{corr} is Holm–Bonferroni over the four candidates within each metric separately.

		Decision-irrelevant energy (↓)				Affine R^2 (↑)			
Encoder	**Objective**	**Mean**	Δ	d_z	p_{corr}	**Mean**	Δ	d_z	p_{corr}
GRU, 1 layer	reweighted	0.00738	$+0.700$	5.54	2.5×10^{-15}	0.918	-0.00821	-0.17	0.45
GRU, 2 layers	reweighted	0.0161	$+0.691$	5.40	3.0×10^{-15}	0.811	-0.115	-1.12	3.0×10^{-4}
GRU, 1 layer	unweighted	0.628	$+0.0789$	0.48	0.091	0.946	$+0.020$	0.71	0.010
GRU, 2 layers	unweighted	0.724	-0.0166	-0.08	0.71	0.879	-0.0478	-0.81	5.7×10^{-3}

Read on its own, the containment column says that the retained subspace holds more decision-irrelevant variance at depth two than at depth one: 0.0161 against 0.00738 on the reweighted arms and 0.724 against 0.628 on the unweighted ones. We do not attribute either movement to depth, because the instrument moves with it. The retained subspace is the range of an ordinary least-squares fit of the belief to the latent code, which is exact in population for a linear encoder and a linearisation for a recurrent one, and a deeper encoder is a less linear function of the observation window. Affine R^2 is the aggregate R^2 of that same regression, run on the same arrays, and it falls with depth under both objectives: 0.811 against 0.918 reweighted ($\Delta{=}-0.115$ against the baseline, $d_z{=}-1.12$, $p_{\mathrm{corr}}{=}3.0 \times 10^{-4}$) and 0.879 against 0.946 unweighted ($\Delta{=}-0.0478$, $d_z{=}-0.81$, $p_{\mathrm{corr}}{=}5.7 \times 10^{-3}$). Both columns move in the direction a loss of fit alone would produce, and this sweep cannot separate the two, so the depth comparison on containment measures the instrument as well as the encoder. The same caution applies within depth, where the reweighted arm fits less well than the unweighted one by 0.0282 at depth one and 0.0671 at depth two. Two limits of the column are worth stating with it. It bounds how much of the code the linearisation explains and not the angular error of the fitted range, which is the object the containment metrics are read from, so a high value would not by itself exclude a mis-oriented range. And it is not comparable across capacities; every arm here is at $k{=}2$, which is what makes it comparable across these four.

What this licenses, and what it does not. At $d{=}16$, $n_o{=}4$, with cost on blocks 2 and 3 at $t_h{=}3$, at $k{=}2$, at GRU width 128, over one step of depth, at this training budget and at 20 seeds per arm, the error reduction the reweighting produces is present at two recurrent layers as well as at one, on 20 of 20 seeds at both depths. We ran one cell, one capacity, one width and one step of depth, so nothing here bears on other cells, on wider or deeper encoders, on longer budgets or on generators outside this family; and the containment movement is reported rather than explained, for the reason above. No certificate is reported on any arm of this sweep but the baseline. Theorem 5's (A2) asks for a linear encoder, which the four recurrent arms are not (§5), and the floor $\delta \geq 0$ that its ratio rests on is proved from joint Gaussianity of the code and the belief, which a linear encoder on this generator supplies and a recurrent one is not shown to. As it happens δ stayed strictly positive on all 100 runs, with a minimum of 0.017, which is an observation about these runs and not the guarantee the linear arms carry.

D.8 The certificate's tolerance need not be supplied

The δ-relaxed leakage certificate $\eta = \delta/(\mu_{k-1} - \mu_k)$ takes δ as an input, which is why no sample-complexity claim follows from it. That is a property of how we report it rather than of the quantity: δ is bounded above by $\|y - f\|^2$ less the sum, over the indices $i \geq k$, of the eigenvalues of the regression target's second moment taken in descending order; we give that spectrum no letter here, so that ν keeps the meaning the Notation gives it. A run could compute both terms from its own evaluation sample without privileged access to the generator; ours do not, and we do not report that bound. Substituting it costs the top-k block of the noise spectrum, so whether the resulting certificate

is still informative at the capacities we run is untested here, and §3's reading, that our runs illustrate rather than test the theorems, is unaffected.

D.9 A DEGENERATE INPUT REWEIGHTING THAT REALIZES THE DECISION GRAMIAN

The observation of §6 is restricted to nondegenerate configurations, and the restriction is doing work. Take $R_w = (L^\top)^{-1}P^{1/2}$, with L the Cholesky factor of $Q^\top(T^*T)Q$, the predictive Gram on $\mathrm{range}(\Sigma)$ in an orthonormal basis Q, and $P = Q^\top P_{t_h} Q$. It satisfies $R_w^*(T^*T)R_w = P_{t_h}$ on $\mathrm{range}(\Sigma)$ to a relative residual of 1.7×10^{-15}, at $\mathrm{rank}(\Sigma_w) = 4 < \mathrm{rank}(\Sigma) = 6$, with $\Sigma_w = R_w R_w^*$ the reweighted belief second moment, degenerate by construction, and realizing. A capacity-limited linear encoder trained to regress the operator this induces, TG with $G\Sigma G^* = \Sigma_w$ and the belief distribution unchanged, rescaled to the baseline's target power (d=8, n_o=3, k=2, 20 seeds) takes decision-irrelevant energy from 0.472 to 0.002 and lowers its own exact-optimum ceiling by 0.111, whereas a per-sample scalar weight leaves the applied operator's rank at that of T and so cannot realize D. Read descriptively across these constructed operators, degeneracy is what moves the retained subspace off $\ker P_{t_h}$, and realizability is what converts that into lower decision-relevant error.

This is an existence demonstration and not a method. The operator is built from the generator's own P_{t_h} and applied operator, so it is not self-supervised, and it is not evidence about any *trained* input reweighting, which is why the observation carries the nondegenerate qualifier.

E PER-CONFIGURATION RESULTS

Tables below show per-configuration results: the baseline and candidate means over seeds, the oriented paired Δ, the win rate against the baseline and the number of paired seeds. Per-seed dispersion is not printed here; the standard deviation of the candidate is given, for the arms the two headline tables carry, in Tables 20 and 21, and the paired quantities (interval, d_z, p_t, p_w) are in Tables 2 through 6.

E.1 THE d=16, n_o=8 BELIEF-IDENTIFICATION SETTING

Table 10: Per-configuration results for d=16, n_o=8. Baseline and Candidate columns report unobserved-factor affine R^2 (higher is better); Δ is oriented so that positive values indicate a higher R^2 than the baseline.

Configuration	Baseline	Candidate	Δ	Win	n
JEPA Whiten D16 D8	0.311	0.681	0.370	1.00	20
JEPA VICReg D16 D8	0.311	0.620	0.309	1.00	20
JEPA None D16 D8	0.311	0.526	0.215	1.00	20
Naive Hybrid D16 D8	0.311	0.307	-0.00324	0.35	20
Autoencoder D16 D8	0.311	0.247	-0.0637	0.00	20
MLP Encoder D16 D8	0.311	1.74×10^{-3}	-0.309	0.00	20

E.2 THE d=4, n_o=2 BELIEF-IDENTIFICATION SETTING

Table 11: Per-configuration results for d=4, n_o=2. Baseline and Candidate columns report unobserved-factor affine R^2 (higher is better); Δ is oriented so that positive values indicate a higher R^2 than the baseline.

Configuration	Baseline	Candidate	Δ	Win	n
JEPA VICReg D4 D2	0.495	0.750	0.255	0.90	20
JEPA Whiten D4 D2	0.495	0.688	0.193	0.75	20
JEPA None D4 D2	0.495	0.565	0.0694	0.75	20
Naive Hybrid D4 D2	0.495	0.485	-0.0108	0.45	20
Autoencoder D4 D2	0.495	0.195	-0.301	0.05	20
MLP Encoder D4 D2	0.495	5.11×10^{-4}	-0.495	0.00	20

E.3 THE d=16, n_o=4 CAPACITY SETTING

This section reports per-configuration results on decision-relevant error for the d=16, n_o=4 setting, one row per configuration.

Table 12: Per-configuration results for d=16, n_o=4. Baseline and Candidate columns show decision-relevant error (lower is better); positive Δ indicates improvement (lower error).

Configuration	Baseline	Candidate	Δ	Win	n
A2 Reweighted K2 T0	0.968	0.698	0.270	1.00	20
A4 Reweighted K Eq D	0.968	0.750	0.218	1.00	20
A4 Unweighted K Eq D	0.968	0.751	0.217	1.00	20
A2 Reweighted K4	0.968	0.763	0.205	1.00	20
A2 Reweighted K2 Cost1	0.968	0.791	0.177	1.00	20
A2 Reweighted K3	0.968	0.793	0.176	1.00	20
A2 Reweighted K2 Cost3	0.968	0.844	0.124	1.00	20
A2 Reweighted K2	0.968	0.845	0.123	1.00	20
B GRU Reweighted K2	0.968	0.849	0.119	1.00	20
A3 Violating Partial K2	0.968	0.891	0.0774	1.00	20
A1 Unweighted K4	0.968	0.901	0.0669	1.00	20
A2 Reweighted K1	0.968	0.928	0.0401	0.90	20
A1 Unweighted K3	0.968	0.950	0.0181	0.70	20
B GRU Unweighted K2	0.968	0.956	0.0117	0.60	20
A1 Unweighted K1	0.968	0.982	-0.0137	0.35	20

E.4 THE d=8, n_o=3 CAPACITY SETTING

The table below gives the per-configuration records behind §5's aggregates.

Table 13: Per-configuration results for d=8, n_o=3. Baseline and Candidate columns show decision-relevant error (lower is better); positive Δ indicates improvement (lower error).

Configuration	Baseline	Candidate	Δ	Win	n
A2 Reweighted K2 T0	0.901	0.605	0.296	1.00	20
A4 Reweighted K Eq D	0.901	0.680	0.221	1.00	20
A4 Unweighted K Eq D	0.901	0.681	0.221	1.00	20
A2 Reweighted K4	0.901	0.691	0.211	1.00	20
A2 Reweighted K3	0.901	0.713	0.188	1.00	20
A2 Reweighted K2 Cost1	0.901	0.723	0.178	1.00	20
A2 Reweighted K2 Cost3	0.901	0.762	0.140	1.00	20
A1 Unweighted K4	0.901	0.762	0.139	1.00	20
A2 Reweighted K2	0.901	0.778	0.123	0.95	20
B GRU Reweighted K2	0.901	0.795	0.107	1.00	20
A1 Unweighted K3	0.901	0.828	0.0732	0.95	20
A3 Violating Partial K2	0.901	0.852	0.0498	0.90	20
A2 Reweighted K1	0.901	0.874	0.0274	0.70	20
B GRU Unweighted K2	0.901	0.887	0.0143	0.65	20
A1 Unweighted K1	0.901	0.952	-0.0503	0.10	20

F ADDITIONAL FIGURES AND TABLES

Figure 5 and Table 14 support §5.

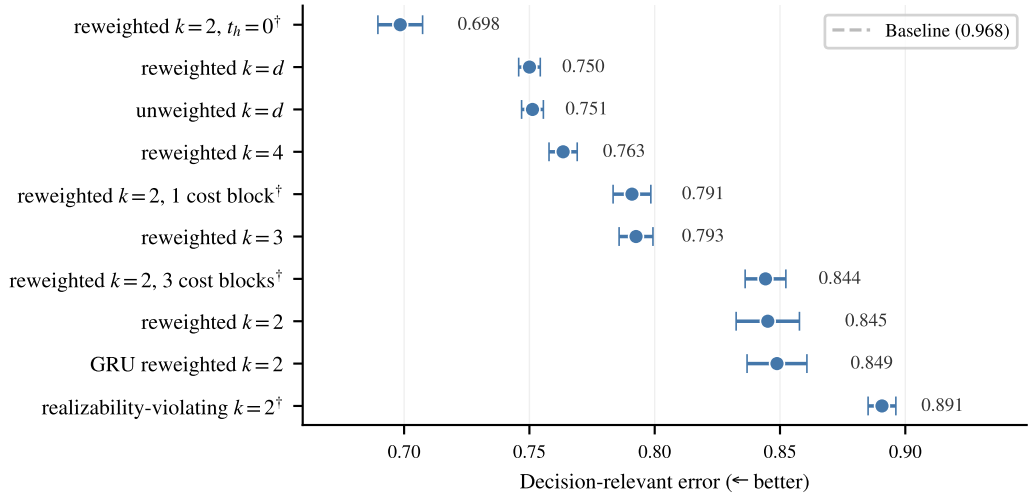


Figure 5: Reweighting lowers decision-relevant error at d=16, n_o=4; the realizability-violating arm, which also moves a cost block onto an unobserved one, sits 0.0774 below the baseline, each scored on its own P_{t_h} (†). The x-axis is decision-relevant error on the same positive scale as Tables 21 and the per-configuration tables, lower being better; rows are the ten best of the fifteen candidates in this comparison family, ordered by improvement over the reweighting-removed baseline, whose value the legend gives; at 0.968 it lies beyond the right edge of the axis, so no rule is drawn for it. Dots are means over 20 seeds and horizontal bars are 95% Student-t intervals of each arm's mean, computed with the population standard deviation, and the number beside each row is that arm's mean decision-relevant error. Rows marked † change P_{t_h}, and so the decision problem. The 15 candidates form one Holm–Bonferroni correction family.

Table 14: Weighting crossed with code dimension k at d=16, n_o=4, 20 seeds per row, same baseline and correction family as Table 21. Δ is oriented so that positive means lower error. Boldface marks the largest reduction at each capacity. The axis has no k=2 row because the unweighted k=2 arm *is* this family's shared baseline, so its Δ is zero by construction and there is no unweighted–reweighted pair to print at that capacity; the reweighted k=2 arm's Δ against that baseline is 0.123, in Table 21.

Capacity	Weighting	Δ (↑)	95% CI	d_z	p_{corr}
k=1	Unweighted	−0.014	[−0.026, −0.003]	−0.50	0.073
k=1	Reweighted	**0.040**	[0.026, 0.055]	1.20	< 0.001
k=3	Unweighted	0.018	[0.006, 0.028]	0.71	0.015
k=3	Reweighted	**0.176**	[0.167, 0.183]	9.30	< 0.001
k=4	Unweighted	0.067	[0.053, 0.080]	2.08	< 0.001
k=4	Reweighted	**0.205**	[0.194, 0.212]	9.92	< 0.001
k=d	Unweighted	0.217	[0.207, 0.224]	11.74	< 0.001
k=d	Reweighted	**0.218**	[0.208, 0.225]	11.39	< 0.001

G Extended Method Details

G.1 Windowed JEPA belief identification

The synthetic generator is the linear-Gaussian system of Appendix B, whose observation matrix reads coordinate 0 of each of the first n_o two-dimensional state blocks. Two cells are run: state dimension 4 with observation dimension 2, and state dimension 16 with observation dimension 8. Nothing else differs between them in the data or the architecture. The losses do scale with the code dimension: the prediction term is a mean over the code's coordinates, so against a unit-weight regulariser it is four times weaker at d=16 than at d=4, and the whitening and VICReg penalties scale differently with d; the ranking of the regularisers across the two cells may partly reflect this.

Both use process noise standard deviation 0.1 and observation noise standard deviation 0.3, and in both the code dimension tracks the state dimension, so $z_t \in \mathbb{R}^4$ in the first cell and $z_t \in \mathbb{R}^{16}$ in the second. In each, half the state coordinates are never observed instantaneously and are recoverable only through temporal aggregation. Training uses 100 trajectories of length 200, and evaluation a disjoint set of 50 trajectories generated with held-out noise draws. Windows of length 16 are extracted by dense striding over each trajectory, so consecutive windows overlap; prediction targets are taken at horizons $\{1, 3, 5\}$ from the window endpoint; horizons enter the training objectives only, and the metric is read at the anchor.

The encoder is a single-layer GRU with hidden width 128 read out at the final window step, except in the non-sequential ablation, whose encoder is a multilayer perceptron with two hidden layers of width 128 applied to the window's last observation alone. Hidden activations are ReLU throughout, and no normalisation layers appear inside any encoder, decoder or predictor.

The arms carrying a contrastive term roll a latent transition operator forward one step at a time and apply the loss at each horizon in $\{1, 3, 5\}$. That operator is affine (an unconstrained $\mathbb{R}^{d\times d}$ matrix plus a bias) so that the identified subspace is comparable to the analytic LQG construction. The contrastive baseline's objective is InfoNCE with the negative squared Euclidean distance as similarity, at temperature 1.0, with in-batch negatives and the positive in the denominator, and nothing besides: it has no decoder and no reconstruction term. The reconstruction decoder, a one-hidden-layer multilayer perceptron of width 128 mapping the code back to the window's last observation, belongs to the recurrent autoencoder, a deterministic GRU encoder with this decoder and no sampling or KL term, which trains on reconstruction alone, and to the naive hybrid, which sums the contrastive and reconstruction terms at unit weight each. Negatives are drawn from all other windows in the minibatch, including temporally adjacent windows from the same trajectory.

The three windowed JEPA arms share none of that machinery: no decoder, no reconstruction term, no contrastive term and no latent transition matrix. Each keeps two GRU encoders, an online one that takes gradients and a target one, initialised independently and updated only as an exponential moving average of the online weights at coefficient 0.99. A separate one-hidden-layer multilayer perceptron of width 128, one per horizon and trained jointly with the encoder, predicts the target encoder's embedding of the horizon-h window from the online encoder's embedding of the anchor window; the loss is the mean over the three horizons of the squared error between the two after both have been L_2-normalised, so it compares embedding direction alone. The arms differ only in the anti-collapse regulariser added to that loss, which acts on the online embeddings: none at all, or VICReg's variance hinge at floor 1.0 together with its off-diagonal covariance penalty, both at unit weight, or a whitening penalty at unit weight driving the online embedding covariance towards the identity. The whitening arm applies no transform to the embedding; the whitening is a penalty on the loss.

Optimisation uses Adam at learning rate 3×10^{-4} with default moments, batch size 256, 50 steps per epoch for 100 epochs, and gradient-norm clipping at 100.0 (a loose guard against rare early-training spikes rather than an active constraint). Every run sets deterministic PyTorch kernels; the 20 seeds control network initialisation, minibatch ordering, and trajectory sampling jointly, and win rates are computed over seed-matched pairs.

Affine R^2 on unobserved factors is measured by fitting ordinary least squares, with an intercept, from frozen encoder outputs to the unobserved component $\Pi_\perp \hat{x}_t$ of the Kalman belief mean. The probe is fitted and scored on the same 50 evaluation trajectories, refit per seed, and its R^2 is pooled over the unobserved coordinates; the encoder receives no gradient from it. Fitting and scoring on one set adds an in-sample optimism of about p/M, with p regressors and $M = 9{,}750$ anchors: below 0.002 here, and nearly cancelling in paired deltas. It is the whole of the non-sequential encoder's 0.00174 and 0.00051, whose population value is zero, since that encoder sees only o_t.

H Extended Analysis

H.1 Retention geometry beyond the headline metrics

The main paper reports the retained decision-irrelevant energy and the principal angle to the decision-relevant subspace. Both are functions of the same eigenbasis, and their joint behaviour

is informative: at capacity $k = 2$ with $d = 16$ and four observed coordinates, decision-irrelevant energy falls from 0.707 to 0.00460 while the principal angle falls from $42.60°$ to $2.158°$. The near-total collapse of both quantities is consistent with reweighting not merely rescaling retained directions but reordering the eigenvalue ranking that the capacity constraint truncates. The fully observed setting behaves identically in kind ($0.609 \rightarrow 0.00350$), which is consistent with the mechanism being a property of the quadratic objective rather than of partial observation.

H.2 Decision-relevant error: horizon interacts with reweighting

Table 15 gives the two short-horizon arms and the fully observed k=4 arm, whose deltas are nominal.

Table 15: Decision-relevant error by cost horizon. Short-horizon arms (t_h=0) change P_{t_h} and are therefore not comparable with the shared-horizon rows of Table 21; Δ is oriented so that positive means lower error.

Setting	Variant	Δ error	Win	95% CI	p (corrected)
d=16, n_o=4	`A2_reweighted_k2_T0`	0.270	1.00	$[0.248, 0.279]$	$< 10^{-4}$
d=8, n_o=3	`A2_reweighted_k2_T0`	0.296	1.00	$[0.280, 0.312]$	$< 10^{-4}$
Full obs.	`A5_full_reweighted_k4`	0.372	1.00	$[0.355, 0.391]$	$< 10^{-4}$

A win rate of 1.00 over 20 seeds with CI widths below 0.04 says the reduction is consistent across seeds; one seed sets the data, the batch order and the initialisation together, so these runs cannot apportion seed variance between them. The two partially observed arms, both `A2_reweighted_k2_T0`, use a cost horizon of zero, and each row's Δ is measured against a baseline at horizon 3, so on its own it moves the horizon and the weighting together. The full-observation arm keeps t_h=3 but moves k from 2 to 4, so its Δ moves weighting and capacity together; the matched k=4 contrast is `MW-14` (0.229).

Separating the two factors requires an unweighted arm at the same horizon, and a further sweep supplies it at d=16, n_o=4. `A1_unweighted_k2_T0` (the t_h=0 arm with the reweighting removed and every other knob held) lowers decision-relevant error by 0.0073 ($[0.0017, 0.0126]$, d_z=0.57, win rate 0.80) against the same shared baseline, against 0.270 for its reweighted partner. The horizon change alone therefore accounts for about 3% of the reweighted short-horizon arm's nominal reduction, and the weighting-matched contrast at t_h=0 is 0.262 ($[0.237, 0.273]$, d_z=7.22, win rate 1.00, p_{corr}<0.001 over K=20).

The same sweep supplies unweighted partners for the two cost-block arms, which makes the cost-block ablation a weighting-matched contrast rather than a confound of block count with weighting. At d=16, n_o=4, with one, two and three cost blocks, the matched contrast is 0.190 ($[0.183, 0.198]$), 0.123 ($[0.105, 0.136]$) and 0.0901 ($[0.0811, 0.100]$), all with p_{corr}<0.001. Each is well posed against its own partner, but the three pose different decision problems (Table 19), so we read their order only descriptively: the fewer directions carry cost, the more the reweighting recovers. The unweighted arms alone do not order that way: the one-block arm is 0.0128 *worse* than the baseline and the three-block arm 0.0338 better, so the order does not come from the unweighted arms. Both series are d=16, n_o=4 only; we did not run these partners at d=8, n_o=3 or under full observation.

H.3 Belief identification: what training keeps

Whitening at d=16, n_o=8 raises affine R^2 on unobserved factors by 0.370 ($[0.352, 0.385]$, win rate 1.00), and VICReg at d=4, n_o=2 by 0.255 ($[0.180, 0.364]$, win rate 0.90). At d=16, n_o=8 the regularised JEPA objectives lose least of the unobserved-component information an early window encoder already carries, ending at 0.681 (whitening) and 0.620 (VICReg) against first-epoch means of 0.632 to 0.688; the contrastive baseline and the hybrid lose most of it, ending at 0.311 and 0.307 (Figure 8), and the reconstruction arm, not drawn there, ends at 0.247 (Table 2). Whitening at d=16 is the only arm that ends above its first-epoch level.

H.4 Separating weighting from effective step size, and the literature's comparators

The companion family of arms: the operator's scale. We reuse the same generator exactly, adding the control the family above lacks: the same operator weight rescaled to unit trace, $\mathrm{tr}\, V^*V{=}1$. Rescaling V by a positive constant scales every eigenvalue of T^*V^*VT, the Gram operator of the weighted target VT, and preserves their order, so the retained-direction structure is fixed and only the target's scale moves, by factors 0.390, 0.389 and 0.577 across the three settings (the loss by their squares). Against each setting's unweighted baseline over 20 paired seeds, the reweighted arm lowers decision-relevant error by 0.123, 0.123 and 0.242 at $k{=}2$ and the trace-normalised arm by 0.099, 0.102 and 0.220; at $k{=}4$ under full observation, 0.372 against 0.362. Every interval excludes zero. Where we computed the ceiling, the closed-form attainable value at that code dimension, the two arms agree in every digit printed; it comes from the generator, not training. The trace-normalised arm reaches 99.3–99.8% of the reweighted arm's movement in retained decision-irrelevant energy. Removing the operator's scale raises the trained error by 0.010–0.024 and leaves the retained subspace where it was: on this generator the effect is redirection, not step size. The control runs at $k{=}2$ in all three settings and at $k{=}4$ in all three as well: against the unweighted $k{=}4$ arm, over 20 paired seeds, the trace-normalised arm lowers decision-relevant error by 0.119 against the reweighted arm's 0.138 at $d{=}16$, $n_o{=}4$, where the two share the closed-form optimum 0.711, and by 0.062 against 0.072 at $d{=}8$, $n_o{=}3$, all four with $p_{\mathrm{corr}}{<}0.001$ within that $k{=}4$ sweep's own six candidates, and with the trace-normalised arm reaching 98.1 and 98.3% of the reweighted arm's movement in retained decision-irrelevant energy. One limit remains: it removes the operator's scale without matching the learning rate per arm.

Comparator measurements in full. No arm here forms a belief-space model, so we instantiate only value equivalence's realizability requirement. Where we computed the ceiling, it settles the first comparison without a test. In all three settings the operator weight and the value-equivalence target share one lower ceiling per setting, 0.12–0.40 below the unweighted one. For the input-side arms we did not compute the ceiling of their own objective: a per-sample scalar changes the covariance the objective sees, so its optimum can move, and the entries printed for them are the unweighted objective's value (1.000 at $d{=}16$, $n_o{=}4$, where the unweighted capacity-2 optimum is entirely decision-irrelevant). What holds exactly is that a scalar weight leaves the applied operator's rank at that of T (8, 6 and 8, against 4 for the operator weight), so it cannot realize D. The trained scalar arms move the retained subspace by 4.9–11.8% of the operator arm's movement at $k{=}2$. Every arm here reads P_{t_h} or the Kalman belief in its weight or its target, so none is self-supervised: the operator weights need the system matrices and the cost, and the arms that read the belief also need the noise covariances. We do not call the scalar arms a reproduction of any published method, since Voelcker et al. (2022) weight by a per-sample *diagonal* metric, not by its norm. Theorem 2's realizability condition is unaffected either way. Table 16 gives every arm with its interval, effect size and ceiling, and the companion family's subspace-movement columns.

Table 16 carries the companion family of §4 in full. It reuses the capacity family's generator unchanged and adds five arms on the one axis the capacity family holds fixed. Every row is $n{=}20$ seeds paired on seed against that setting's own unweighted baseline, with the procedure of §H.10. The correction family here is this family's own candidate set (six in the two partially observed settings, thirteen under full observation), not the $K{=}15$ of Table 21, so the corrected p values in the two tables are not on one scale and we do not pool them.

The last two columns are what the table is for. *Ceiling* is the attainable value of the primary metric for that arm's objective at that code dimension, computed in closed form from the generator matrices and not from any trained model; where it is reported it is identical across all 20 seeds of an arm, which is why it carries no interval. The word *Ceiling* has the meaning of §4: it is the metric at the exact optimum of the arm's own objective, where the arm would land if training reached that optimum, and so the most an arm that reaches it improves. It is not a floor under every trained run: an arm whose objective is not the metric can score below it, as the unweighted arm does at small k in Figure 3. Each entry is that arm's own value, at that arm's objective and code dimension, so entries differ from arm to arm and from the single encoder-side floor §4 quotes per setting. The two values coincide where an arm's objective already ranks belief directions by cost and its capacity reaches $\mathrm{rank}\, P_{t_h}$, which is 4 in all three settings: the reweighted, trace-normalised and value-equivalence $k{=}4$ rows under full observation all print §4's 0.499. Entries marked u are the

Table 16: The companion family: the trace-normalised step-size control and the input-side and value-equivalence comparators, against each setting's own unweighted baseline. Δ is the oriented paired reduction in decision-relevant error, positive meaning lower error; intervals are 95% BCa over 10,000 resamples; p_{corr} is Holm–Bonferroni within this family's candidate set, applied to the two-sided paired t-test of Appendix C.1. *Ceiling* is the closed-form attainable value of the primary metric for that objective at that code dimension, constant over seeds; a superscript u marks the unweighted objective's value, printed for the input-side scalar arms, whose own optimum we did not compute. Every other arm's ceiling is the metric at its own objective's optimum and code dimension, in the sense the accompanying paragraph gives the word. *Ret. irrel.* is the paired reduction in retained decision-irrelevant energy. The row marked *Unweighted, capacity-matched* is the $k{=}4$ unweighted arm, which this table carries only under full observation (in the two partially observed settings it is the baseline of the $k{=}4$ control this subsection's first paragraph reports); the constant-weight row beside it agrees with it in every digit printed here. Table 22 keys the capacity family's arm codes and does not cover this family; the arm codes and configuration keys for every row here are in this family's sweep configurations in the supplementary material.

Setting	Arm	k	Δ	95% CI	d_z	p_{corr}	Win	Ceiling	Ret. irrel.
$d{=}16, n_o{=}4$	Reweighted (operator)	2	0.123	$[0.105, 0.136]$	3.44	$1.8{\times}10^{-11}$	1.00	0.782	0.703
	Trace-normalised operator	2	0.0991	$[0.0789, 0.115]$	2.40	$6.8{\times}10^{-9}$	1.00	0.782	0.698
	Value-equivalence target	2	0.161	$[0.144, 0.174]$	4.54	$1.4{\times}10^{-13}$	1.00	0.782	0.707
	Input-side scalar	2	0.0036	$[0.0020, 0.0070]$	0.69	0.0172	0.70	1.000^{u}	0.0342
	Input-side scalar, norm.	2	0.0036	$[0.0020, 0.0070]$	0.70	0.0172	0.75	1.000^{u}	0.0341
	Constant weight	2	$4.0{\times}10^{-7}$	$[1.1{\times}10^{-7}, 7.4{\times}10^{-7}]$	0.54	0.0263	0.65	1.000	$3.4{\times}10^{-6}$
	Unweighted baseline	—	—	—	—	—	—	1.000	—
$d{=}8, n_o{=}3$	Reweighted (operator)	2	0.123	$[0.0931, 0.143]$	2.17	$4.4{\times}10^{-8}$	0.95	0.721	0.469
	Trace-normalised operator	2	0.102	$[0.0746, 0.122]$	1.89	$3.0{\times}10^{-7}$	0.95	0.721	0.467
	Value-equivalence target	2	0.154	$[0.136, 0.172]$	3.66	$7.2{\times}10^{-12}$	1.00	0.721	0.472
	Input-side scalar	2	0.0148	$[0.0104, 0.0205]$	1.26	$6.0{\times}10^{-5}$	0.95	0.845^{u}	0.0345
	Input-side scalar, norm.	2	0.0147	$[0.0104, 0.0204]$	1.26	$6.0{\times}10^{-5}$	0.95	0.845^{u}	0.0342
	Constant weight	2	$9.2{\times}10^{-8}$	$[-4.8{\times}10^{-8}, 3.1{\times}10^{-7}]$	0.22	0.327	0.55	0.845	$5.3{\times}10^{-7}$
	Unweighted baseline	—	—	—	—	—	—	0.845	—
Full obs.	Reweighted (operator)	4	0.372	$[0.355, 0.391]$	8.79	$1.3{\times}10^{-18}$	1.00	0.499	0.599
	Trace-normalised operator	4	0.362	$[0.346, 0.380]$	9.19	$6.0{\times}10^{-19}$	1.00	0.499	0.595
	Value-equivalence target	4	0.419	$[0.402, 0.438]$	9.75	$2.1{\times}10^{-19}$	1.00	0.499	0.609
	Unweighted, capacity-matched	4	0.144	$[0.120, 0.170]$	2.47	$5.1{\times}10^{-9}$	1.00	0.696	0.0583
	Input-side scalar	4	0.189	$[0.167, 0.215]$	3.34	$4.2{\times}10^{-11}$	1.00	0.696^{u}	0.0893
	Input-side scalar, norm.	4	0.188	$[0.166, 0.215]$	3.34	$4.2{\times}10^{-11}$	1.00	0.696^{u}	0.0891
	Constant weight	4	0.144	$[0.120, 0.170]$	2.47	$5.1{\times}10^{-9}$	1.00	0.696	0.0583
	Reweighted (operator)	2	0.242	$[0.220, 0.266]$	4.56	$2.0{\times}10^{-13}$	1.00	0.598	0.605
	Trace-normalised operator	2	0.220	$[0.196, 0.246]$	3.76	$5.9{\times}10^{-12}$	1.00	0.598	0.604
	Value-equivalence target	2	0.281	$[0.258, 0.304]$	5.19	$2.1{\times}10^{-14}$	1.00	0.598	0.609
	Input-side scalar	2	0.0321	$[0.0238, 0.0417]$	1.54	$4.5{\times}10^{-6}$	1.00	1.000^{u}	0.0711
	Input-side scalar, norm.	2	0.0320	$[0.0236, 0.0415]$	1.53	$4.5{\times}10^{-6}$	1.00	1.000^{u}	0.0707
	Constant weight	2	$3.1{\times}10^{-7}$	$[1.1{\times}10^{-7}, 5.6{\times}10^{-7}]$	0.58	0.0173	0.65	1.000	$1.4{\times}10^{-6}$
	Unweighted baseline	—	—	—	—	—	—	1.000	—

unweighted objective's value, printed for the input-side scalar arms, whose own optimum we did not compute. *Ret. irrel.*, the last column, is the paired reduction in the containment surrogate of Table 23.

We draw no comparison from the entries marked u. Elsewhere the column is read only as reported in §4, and the digits it prints there are the digits printed in the table. The trained columns, Δ and its interval, carry every other claim this table supports.

Two limits on what this table settles. Its $k{=}4$ rows are under full observation only and are measured against the $k{=}2$ baseline; the $k{=}4$ step-size control in the two partially observed settings, measured against the unweighted $k{=}4$ arm and so matched to the $k{=}4$ weighting-matched contrast of Table 21, is the one this subsection's first paragraph reports. And it removes the operator's scale without matching the learning rate per arm; that second route to the same confound is measured in Appendix H.5, which gives each arm its own step size over a grid on the same 20 seeds, and not here. Separately, every arm in this table reads P_{t_h} or the Kalman belief in its weight or its target, so the comparators are informed references and not self-supervised methods; a reader should not read the value-equivalence row as a deployable baseline that beats ours.

The diagonal comparator's exact ceiling. §4 says that an axis-aligned diagonal weight recovers most of the operator weight's attainable gain. The weight priced here is the constant diagonal of squared value-gradient coordinates, taken in closed form as an expectation over the stationary belief distribution, not the per-sample diagonal of Voelcker et al. (2022): a per-sample weight has no capacity-k ceiling in closed form, so the constant one is the closest form this argument can price.

That statement is arithmetic rather than a measurement: the capacity-k optimum of a given weight is a property of the generator, so it is computed in closed form from the generator's own matrices and no trained run enters it. At k=2:

Table 17: Exact capacity-2 ceilings of three weights, in decision-relevant error (↓), computed in closed form from the generator. *Identity* is the belief-space identity weight, *Diagonal* the axis-aligned per-coordinate weight and *Operator* the full output-side operator. *Identity* is not the observation-space unweighted arm whose ceiling Table 16 reports: at d=8, n_o=3 the two read 0.865 and 0.845, because the objectives are written in different spaces. Recovery is (Identity − Diagonal)/(Identity − Operator), the fraction of the operator weight's improvement over the identity weight that the diagonal already reaches. Nothing here is trained.

Setting	Identity	Diagonal	Operator	Recovery
d=16, n_o=4	1.00	0.799	0.782	0.923
d=8, n_o=3	0.865	0.752	0.721	0.787
Full observation	1.00	0.696	0.598	0.758

The recovery column is the 76–92% of §4; the remaining 8–24% is what the operator weight adds over the best axis-aligned weight this argument can price. The comparison favours the diagonal by construction, for the reason §4 gives: this generator places its cost blocks on the coordinate axes, which is the best case for an axis-aligned weight. The same experiment also trained the per-sample diagonal and the full operator on a belief-space target; their code and configurations are in the supplementary material, and we do not compare them with the arms of Table 16, because their target reads the Kalman belief.

H.5 The per-arm step-size grid

§4 controls the step-size confound by giving every arm its own `lr` over $\{1, 3, 9, 30\} \times 10^{-4}$ at d=16, n_o=4, 20 seeds each, on the capacity family's generator unchanged. Table 18 gives the grid.

Table 18: Mean decision-relevant error at the final step (↓) over the per-arm step-size grid at d=16, n_o=4, 20 seeds each. Each arm's best is in bold.

Arm	1×10^{-4}	3×10^{-4}	9×10^{-4}	3×10^{-3}
Unweighted k=2 (the baseline arm)	**0.954**	0.968	0.981	0.998
Reweighted k=2	0.853	0.845	0.829	**0.818**
Unweighted k=1	**0.981**	0.982	0.989	0.999
Unweighted k=3	0.939	0.950	0.948	**0.921**
Unweighted k=4	0.904	0.901	0.885	**0.879**
Unweighted k=d=16	0.755	**0.751**	0.754	0.757
Unweighted GRU k=2	**0.946**	0.956	0.992	0.999

The two arms of the contrast optimise at opposite ends of a 30× range: the unweighted arm at the smallest step size, the reweighted arm at the largest, so no single shared value can be said to favour the reweighted arm. Read at each of the four shared values, the weighting-matched contrast is 0.100 ([0.079, 0.118]), 0.123 ([0.105, 0.136]), 0.152 ([0.143, 0.159]) and 0.181 ([0.177, 0.183]), every one with p_{corr}<0.001 in the family of Table 19; read with each arm at its own best value it is 0.136 ([0.126, 0.144], d_z=6.26, win rate 1.00), and under the alternative rule of taking each *seed*'s own best it is 0.134 (d_z=6.48, win rate 1.00), so the reading does not turn on which best-rule is used. The shared 3×10^{-4} that the rest of the paper uses gives the second smallest of the four, so that setting is if anything unfavourable to the reweighted arm rather than favourable to it.

Two properties of the grid are worth recording as descriptive facts about the generator rather than as results. The unweighted arms at $k \in \{1, 2\}$ and the recurrent arm get monotonically worse as the step size rises, and the reweighted arm and the unweighted k=4 arm get monotonically better; the unweighted k=3 arm improves over the range (0.939 to 0.921) but not monotonically, rising at 3×10^{-4} before it falls. The full-capacity arm is flat across the whole 30× range, its four values

Table 19: The $K{=}20$ weighting-matched contrasts. Unweighted and reweighted columns are mean decision-relevant error over 20 paired seeds (↓); Δ is the paired mean of unweighted minus reweighted, with a two-sided 95% BCa interval, paired Cohen's d_z, and the win rate over seeds. The three step-size rows read the $k{=}2$ contrast at the non-default values of the grid in §4; the $k{=}2$ row of the first block is the same contrast at the shared 3×10^{-4}, so these four rows are four readings of one contrast rather than four independent hypotheses, and the rerun reproduces it to the bit. The three cost-structure rows carry a P_{t_h} that differs from the shared baseline's; each is matched against its own partner at that same P_{t_h}, so the contrast is well posed, but those rows are not comparable with each other or with the rest of the table. The $k{=}2$ row of each setting also sits in a baseline family, because each setting's shared baseline is itself the unweighted $k{=}2$ arm; the value printed is the larger of the two corrected values. The $k{=}d{=}8$ row is the family's one member not separated from zero at $n{=}20$.

ID	Setting	Pair	Unw.	Rew.	Δ	95% CI	d_z	Win	p_{corr}
MW-01	$d{=}16, n_o{=}4$	$k{=}1$	0.982	0.928	+0.0538	$[0.0410, 0.0652]$	1.88	0.95	< 0.001
MW-02		$k{=}2$	0.968	0.845	+0.123	$[0.105, 0.136]$	3.44	1.00	< 0.001
MW-03		$k{=}3$	0.950	0.792	+0.157	$[0.145, 0.169]$	5.43	1.00	< 0.001
MW-04		$k{=}4$	0.901	0.763	+0.138	$[0.127, 0.150]$	5.22	1.00	< 0.001
MW-05		$k{=}d{=}16$	0.751	0.750	+0.0012	$[0.0005, 0.0021]$	0.64	0.75	0.019
MW-06		GRU, $k{=}2$	0.956	0.849	+0.108	$[0.0915, 0.122]$	2.96	1.00	< 0.001
MW-07	$d{=}8, n_o{=}3$	$k{=}1$	0.952	0.874	+0.0777	$[0.0602, 0.0938]$	1.96	0.95	< 0.001
MW-08		$k{=}2$	0.901	0.778	+0.123	$[0.0931, 0.143]$	2.17	0.95	< 0.001
MW-09		$k{=}3$	0.828	0.713	+0.115	$[0.0992, 0.130]$	3.15	1.00	< 0.001
MW-10		$k{=}4$	0.762	0.691	+0.0718	$[0.0620, 0.0842]$	2.76	1.00	< 0.001
MW-11		$k{=}d{=}8$	0.681	0.680	+0.0004	$[-0.0003, 0.0010]$	0.25	0.60	0.283
MW-12		GRU, $k{=}2$	0.887	0.795	+0.0923	$[0.0678, 0.120]$	1.51	0.95	< 0.001
MW-13	full obs., $d{=}8$	$k{=}2$	0.913	0.671	+0.242	$[0.220, 0.266]$	4.56	1.00	< 0.001
MW-14		$k{=}4$	0.769	0.541	+0.229	$[0.211, 0.249]$	5.07	1.00	< 0.001
MW-15	$d{=}16, n_o{=}4$	$k{=}2$, lr 1×10^{-4}	0.954	0.853	+0.100	$[0.0789, 0.118]$	2.21	1.00	< 0.001
MW-16		$k{=}2$, lr 9×10^{-4}	0.981	0.829	+0.152	$[0.143, 0.159]$	7.91	1.00	< 0.001
MW-17		$k{=}2$, lr 3×10^{-3}	0.998	0.818	+0.181	$[0.177, 0.183]$	24.04	1.00	< 0.001
MW-18		$k{=}2$, $t_h{=}0$	0.961	0.699	+0.262	$[0.237, 0.273]$	7.22	1.00	< 0.001
MW-19		$k{=}2$, 1 cost block	0.981	0.791	+0.190	$[0.183, 0.198]$	10.95	1.00	< 0.001
MW-20		$k{=}2$, 3 cost blocks	0.934	0.844	+0.0901	$[0.0811, 0.100]$	4.05	1.00	< 0.001

spanning 0.751 to 0.757, which is what an arm whose capacity constraint does not bind should look like.

The arms rerun at the shared 3×10^{-4} duplicate reported arms by design. They reproduce them to the bit: the same 20 per-seed values and the same baseline column, which is what licenses pairing an arm of this sweep against an arm of the original one, as Table 19's cost-structure rows do.

H.6 The weighting-matched correction family

Table 19 is the whole family whose membership rule and correction Appendix H.10 states. Every row is a per-seed paired difference between two arms of one setting that differ in the reweighting flag alone, oriented unweighted minus reweighted, so a positive Δ means the reweighting lowered decision-relevant error. The first fourteen rows are the family as it was fixed, before any of its p-values was computed; the last six joined in arrival order as later ablations supplied the partners it had been missing. Each row carries a permanent identifier, printed in the *ID* column and never reassigned, so that a row keeps its meaning if the family grows. p_{corr} is Holm–Bonferroni over all $K{=}20$ members at $\alpha{=}0.05$. That K counts non-independent hypotheses. Four members are readings of the same $d{=}16, n_o{=}4$, $k{=}2$ weighting-matched contrast at four step sizes, so collapsing them to one leaves 17 distinct contrasts. Correcting at $K{=}20$ rather than at the deduplicated $K{=}17$ makes every p_{corr} printed here an upper bound on the corrected value the deduplicated family would give. Wilcoxon agrees with the t-test on every member. Shapiro–Wilk rejects normality of the deltas on two members at $\alpha{=}0.05$: the $d{=}8, n_o{=}3$ $k{=}2$ row ($p{=}0.048$) and the $t_h{=}0$ row ($p{=}6.6 \times 10^{-5}$). The protocol of Appendix C.1 reads those two through the Wilcoxon signed-rank test, whose corrected value is 3.8×10^{-5} on both, so neither survives on the t-test alone. Shapiro–Wilk does not reject on the remaining eighteen, the smallest of which is $p{=}0.073$.

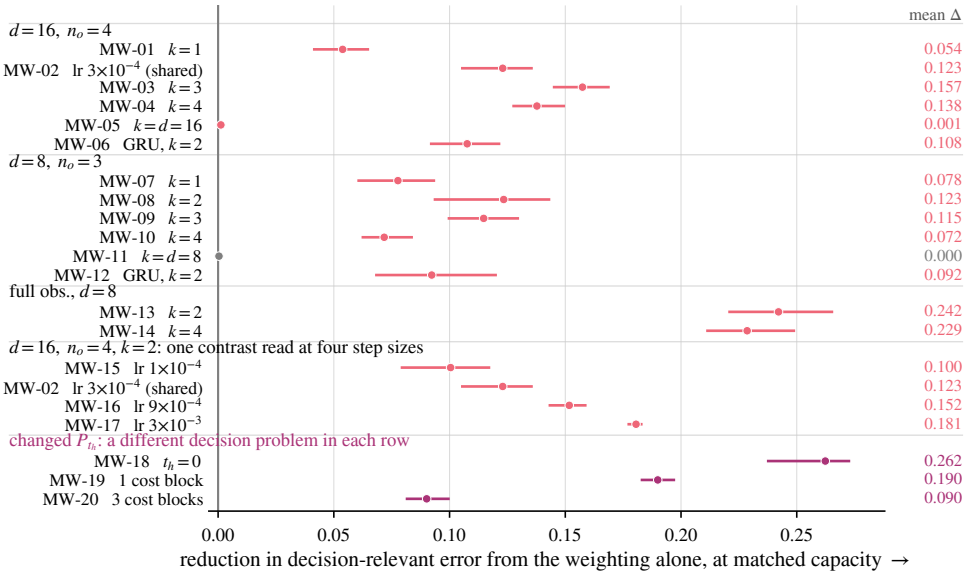


Figure 6: The $K=20$ weighting-matched contrasts of Table 19, each 20 paired seeds with a 95% BCa interval: weighting alone, at matched capacity, encoder and horizon. Grey marks the one contrast not separated from zero. lr is the step size.

H.7 The certificate without an oracle, and how far it disagrees

The certificate's slack, its reduced forms and their shrinkage. §3 states η and how we read it. Here δ is the arm's suboptimality in Hilbert–Schmidt units, squared error summed over an orthonormal basis of belief space,

$$\delta \;=\; \mathbb{E}\big\|T'\hat{x} - f\big\|^2 \;-\; \textstyle\sum_{i\geq k}\mu_i, \tag{4}$$

with f the arm's prediction, T' the operator that arm regresses onto ($\tilde{T}$ when reweighted, the bare predictive T when not), and μ_i the spectrum of $R^*(T'^*T')R$, with R the belief-covariance factor of Theorem 1 clause 6 applied to that operator, computed in closed form from the generator. An unweighted arm's μ therefore ranks by prediction and not by cost. The residual route below converts the logged residual. Every reported η takes δ from the belief route; the residual route is a comparison only. As an input, not a sample estimate, δ makes η a per-run diagnostic, not a sample-complexity bound. Two reduced forms:

$$\eta_{\text{norm}} = \frac{\delta}{k\,(\mu_{k-1}-\mu_k)} = \frac{\eta}{k}, \qquad \eta_{\text{gf}} = \frac{\delta}{\mu_{k-1}}. \tag{5}$$

Proposition 1. *Let $\delta \geq 0$, $k \geq 1$ and $\mu_{k-1} > \mu_k \geq 0$, with $\eta, \eta_{\text{norm}}, \eta_{\text{gf}}$ as above. Then $\eta_{\text{norm}} \leq \eta$ and $\eta_{\text{gf}} \leq \eta$. Equality holds in the first exactly when $k = 1$ or $\delta = 0$, and in the second exactly when $\mu_k = 0$ or $\delta = 0$.*

Proof. Both inequalities follow from $k \geq 1$ and $\mu_k \geq 0$. Since $\delta \geq 0$ and $\mu_{k-1} - \mu_k > 0$, $\eta \geq 0$, so $k \geq 1$ gives $\eta/k \leq \eta$, with equality exactly when $k = 1$ or $\eta = 0$, that is $\delta = 0$. And $\mu_k \geq 0$ gives $\mu_{k-1} \geq \mu_{k-1} - \mu_k > 0$, so $\delta/\mu_{k-1} \leq \delta/(\mu_{k-1} - \mu_k)$, with equality exactly when $\mu_k = 0$ or $\delta = 0$. □

No theorem licenses either divisor against the frame weight of Theorem 5 clause 2: they are diagnostics, not certificates.

Every η reported in §5 takes its numerator δ from the belief route, which needs the true Kalman belief and is therefore unavailable to a practitioner. The certificate comparisons also log a second route, which reaches the same suboptimality slack through the model's own prediction residual. That route removes one of the two places the belief is needed, not both. Write y for the arm's regression target, f for the arm's prediction, $T'\hat{x}$ for the conditional mean of y given the observations up to t, and

read every $\mathbb{E}$ below as a mean over the held-out evaluation anchors. The belief route is Equation 4, $\delta = \mathbb{E}\|T'\hat{x} - f\|^2 - \sum_{i\geq k}\mu_i$, and the residual route is

$$\delta_{\text{obs}} \;=\; \mathbb{E}\|y - f\|^2 \;-\; \mathbb{E}\|y - T'\hat{x}\|^2 \;-\; \textstyle\sum_{i\geq k}\mu_i.$$

Its first term is a function of the target and the model alone. Its second, the noise floor, is a population constant of the generator that we estimate as a sample mean through the true belief, and the third is the same generator spectrum both routes subtract. So the residual route removes the belief from the risk term and from nowhere else, and neither route is free of the generator: it is a control on how much of the belief the risk route needs, not a deployable diagnostic. The two are related by a finite-sample term,

$$\delta_{\text{obs}} \;=\; \delta \;+\; \kappa, \qquad \kappa \;=\; \mathbb{E}\|y - f\|^2 - \mathbb{E}\|y - T'\hat{x}\|^2 - \mathbb{E}\|T'\hat{x} - f\|^2 \;=\; 2\,\mathbb{E}\big\langle y - T'\hat{x},\; T'\hat{x} - f\big\rangle,$$

the prediction cross term, in the same covariance-weighted units as δ. Kalman orthogonality makes κ zero in population, since $y - T'\hat{x}$ is uncorrelated with every function of the observations and f is one of them; on a finite evaluation sample it is not zero. We log κ as the difference in its first form, and expanding the square makes that difference equal to the cross term exactly, for sample means as for population ones. The disagreement between the routes is therefore κ, on all 1180 runs of the four certificate settings, and κ does not vanish at this sample size. In the two eight-dimensional settings its median is 0.0012 and it does not shrink with training, while trained suboptimality falls to medians of 0.0129 and 0.0192.

It is large enough to matter. Taking $|\delta_{\text{obs}} - \delta|/\delta$ per run, the median disagreement is 11.0% at $d{=}8, n_o{=}3$, 9.3% at $d{=}16, n_o{=}4$, 7.5% under full observation and 10.2% at $d{=}32, n_o{=}8$, with maxima of 94.8%, 159.3%, 78.4% and 59.8%. It is also two-sided: the signed medians are -2.1%, -0.7%, $+1.2\%$ and -4.3%, and the residual route lands below the belief route on 55.6%, 52.5%, 45.4% and 63.5% of runs, so it is neither reliably conservative nor reliably optimistic.

The consequence for a practitioner is the one that should be reported, because it is a disagreement about the verdict and not only about the value. On the binary question the certificate exists to answer, whether $\eta < 1$ so that the bound excludes something, the two routes return different answers on 13 of 320 runs at $d{=}8, n_o{=}3$, 14 of 320 at $d{=}16, n_o{=}4$, 4 of 280 under full observation and 2 of 260 at $d{=}32, n_o{=}8$: 33 of 1180, or 2.8%. On one further run the residual route returns a negative slack ($\delta_{\text{obs}} = -0.00216$ against $\delta{=}0.00364$, with $\kappa = -0.00580$), which no slack can be; the belief route holds against the quantity it bounds on $1180/1180$ runs and the residual route on $1178/1180$: the exceptions are that run and one run at $d{=}8, n_o{=}3$ on which its value, with a positive slack, falls below the frame weight.

The $d{=}32, n_o{=}8$ cell is a fourth linear-Gaussian setting, run for this comparison and counted in the 1180-run certificate totals of §5. It is the generator of Appendix B at state dimension 32, observation dimension 8, partially observed, with the cost on two blocks starting at block 4; every other value in Table 1 is unchanged, including the 100-epoch budget, batch 256, 3×10^{-4} and 20 seeds. Thirteen configurations run there, the baseline plus twelve reweighted arms over $t_h \in \{0, 1, 3\}$, one or two cost blocks and $k \in \{1, 2, 3, 4\}$; at 20 seeds each that is the 260 runs counted above. The other three settings run the same twelve reweighted arms, which keep only the combinations with a strict covariance-weighted gap at k, beside the baseline and, at $d{=}8, n_o{=}3$ and $d{=}16, n_o{=}4$, three arms of the capacity family (the unweighted $k{=}4$ arm, the reweighted $k{=}2$ arm with one cost block and the GRU arm), or under full observation the unweighted $k{=}4$ arm: 16, 16 and 14 configurations.

Its two verdict disagreements are 0.8% of its 260 runs, and its signed median and its 63.5% share, both given above, are no more one-sided than the other three settings' and do not read as a safety margin either. We report each setting on its own terms and draw no trend in the state dimension across them.

We therefore keep the reading of §3: the residual route is a robustness comparison, not an oracle-free substitute. It agrees with the belief route to a median 7.5–11.0% and returns the same informative-or-weak verdict on 97.2% of runs, which is enough to say that an oracle-free diagnostic of this shape is not hopeless on this generator, and not enough to license reporting one in place of the belief route. The disagreement is a finite-sample effect with a named and logged cause, so it is the kind that more data would shrink; four settings at one sample size do not show that it does.

H.8 A BELIEF-FREE ROUTE TO THE CERTIFICATE'S NUMERATOR, AND WHAT IT COSTS

This subsection runs inside the same linear–Gaussian generator at d=8, n_o=3 and answers a different question, whether the tolerance δ of Equation 4 can be obtained without the Kalman belief. Appendix H.7's residual route removes the belief from the risk term but still estimates the noise floor $\mathbb{E}\|y - T'\hat{x}\|^2$ through it. A third route eliminates that floor. Since $\mathbb{E}\|y - T'\hat{x}\|^2 = \mathbb{E}\|y\|^2 - \sum_i \mu_i$,

$$\delta \;=\; \mathbb{E}\|y - f\|^2 \;-\; \mathbb{E}\|y\|^2 \;+\; \textstyle\sum_{i<k} \mu_i,$$

every term a function of the target, the model and the spectrum μ the denominator already requires. It is belief-free and not filter-free: it is Appendix D.8's upper bound with the head of the target's own second-moment spectrum replaced by μ's, which makes it an identity; evaluating μ from the observable law still needs the Bayes predictor. The belief *sample* is removable; the requirement it stands for is not.

Logging the two new quantities per run on this setting (16 arms, 20 seeds, 320 runs) leaves every key shared with Appendix H.7's runs bit-identical, timing and device fields aside, so the three routes are paired on identical trained models. Per run, $|\delta_{\mathrm{bf}} - \delta|/\delta$ has median 15.2% against the residual route's 11.0%, and the belief-free route is the closer of the two on 122 of 320 runs. The error leans positive: signed mean $+9.7 \times 10^{-4}$ against a median δ of 1.30×10^{-2}, positive on 186 of 320 runs, and averaging each arm's 20 seeds still leaves a median over the arms of 7.6% against 2.8%. Its cause is logged: the evaluation-sample orthogonality residual $2\,\mathbb{E}\langle y - T'\hat{x},\, T'\hat{x}\rangle$, zero in population, is a median 20.7% of δ. Its absolute error is flat in δ (Spearman -0.01) where the residual route's grows with it ($+0.51$), so its relative error tracks $1/\delta$ (-0.65) and reaches 54–75% on the two smallest-slack arms (median η 0.10 and 2.8). On the 68 runs with $\eta < 1$, where the certificate excludes anything, its median is 19.5%, against 13.6% on the other 252. On 3 of 320 runs it returns a negative slack, which the other two routes never do.

As an ordinal instrument it fares better. Over the 16 arm medians it reproduces the belief route's ranking of δ at Spearman 0.997 (0.994 for η), against the residual route's 0.991 (the same for η), and at Holm-corrected $p < 0.05$ it flags the same six arms below baseline that the belief route does, where the residual route flags five. It flips the informative-or-weak verdict on 20 of 320 runs against the residual route's 13.

We therefore keep reporting η from the belief route. Removing the belief sample from δ is possible and costs most accuracy on the runs where the certificate is informative. That is one setting at one evaluation sample size ($M = 7300$ anchors), and the identity being estimated is a population statement that these runs illustrate rather than test.

H.9 FIGURES FOR THE RESULTS

Figure 8 traces training for the belief-identification family; Figure 9 draws, per direction, the affine R^2 of the code at one capacity; Figure 7 draws the decision-irrelevant energy §5 reports against capacity, one dot per seed.

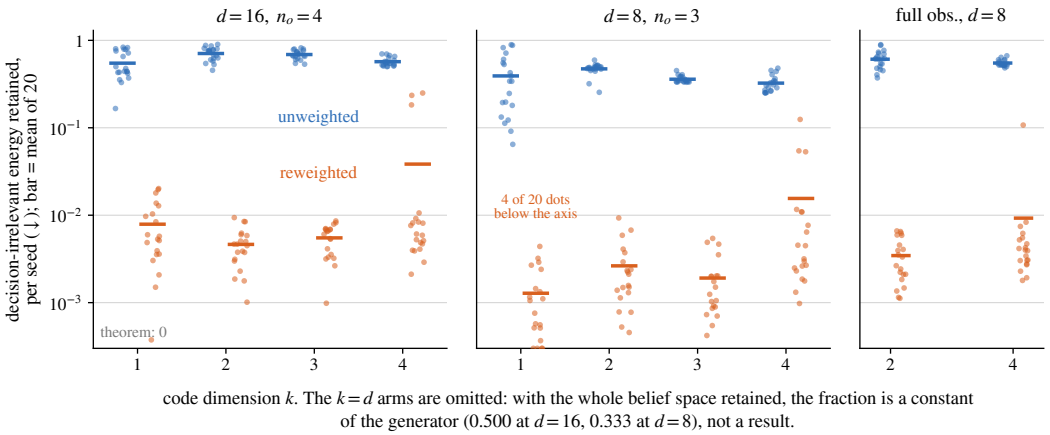


Figure 7: Decision-irrelevant energy each code retains, one dot per seed and bars the mean of 20, against code dimension k in the three settings. Theorem 2 clause 3 predicts zero for the reweighted arm at the exact optimum. The $k{=}d$ arms are omitted: with the whole belief space retained the fraction is a constant of the generator, 0.500 at $d{=}16$ and 0.333 at $d{=}8$, not a result. Four of the 400 dots fall below the axis floor and are drawn on it, all in the $d{=}8$, $n_o{=}3$ reweighted $k{=}1$ arm.

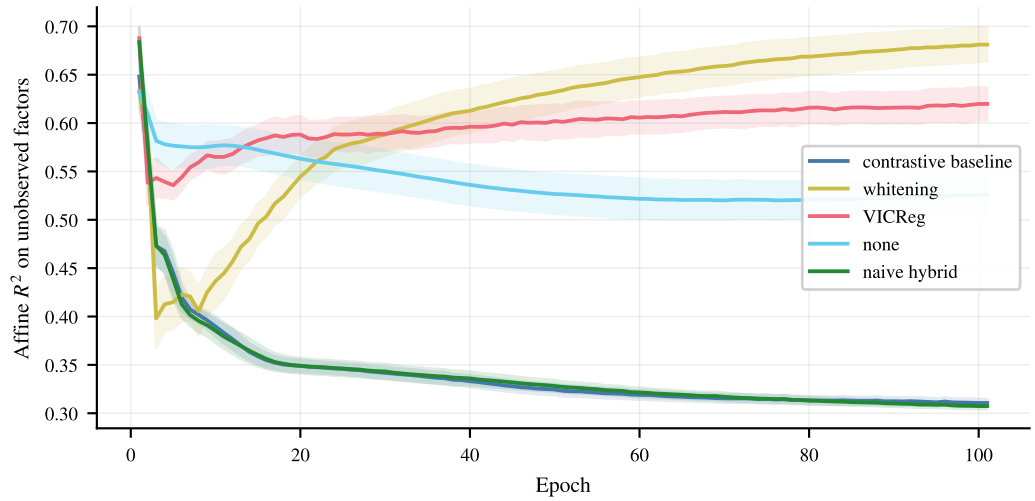


Figure 8: Training curves of affine R^2 at $d{=}16, n_o{=}8$ for five arms: the contrastive baseline, the three JEPA variants, named by their regulariser (whitening, VICReg, none), and the naive hybrid, which tracks the baseline almost exactly throughout. Lines show means across seeds; shaded regions are 95% Student-t intervals of the mean, computed with the population standard deviation. Every arm begins high: the mean over the 20 seeds at the first epoch runs from 0.632 to 0.688 across the five, against the baseline's final 0.311. The metric fits an affine probe to the code at every epoch, so a value this high early says that a barely trained encoder already passes enough of the 16-step window for the probe to read the unobserved factors, not that the objective has shaped the code; what each objective then does with that structure is what the curves separate. The contrastive baseline and the naive hybrid descend to 0.311 and 0.307, well below their own first epoch, while whitening ends highest at 0.681. We report the final step for every arm without early stopping, so the paired contrasts of Table 20 are between trained encoders throughout and none of them is read at initialisation.

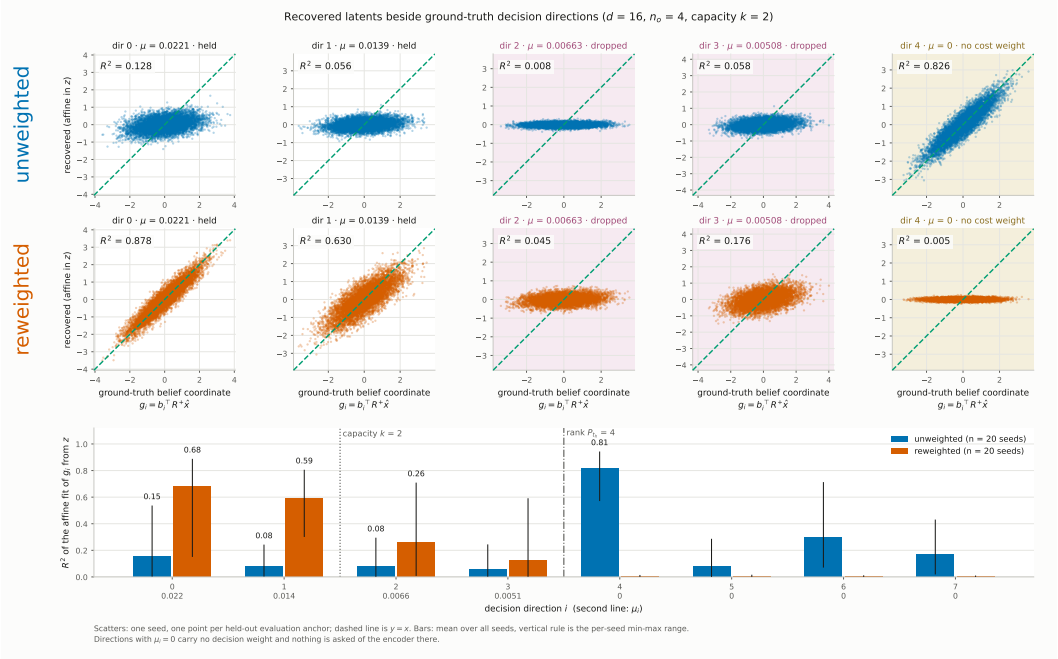


Figure 9: Which latent directions survive the capacity constraint, at d=16, n_o=4, k=2. Top: recovered against ground-truth belief coordinates for one seed, over the four cost-weighted decision directions and the first direction the cost does not weight; the dashed green line is the ground truth, the two directions on the pink background are dropped by the capacity constraint, so their low R^2 is the predicted behaviour, and the one on the cream background carries no cost weight at all. Bottom: affine R^2 per direction for both arms, mean over 20 seeds, with the per-seed min–max range; directions are ordered by their cost weight μ_i under the covariance-weighted operator (Theorem 1, clause 6), which is nonzero on exactly four of the eight and zero on the rest. The eight are every direction of the whitened belief space $\mathcal{E} = \mathrm{range}(\Sigma_{\hat{x}})$ the theorems are instantiated on, whose dimension is $\mathrm{rank}(\Sigma_{\hat{x}}) = 8$ here although d=16 (Appendix A.3); the panel omits no direction of it. The unweighted arm spends its two-dimensional code on a direction the cost ignores, reaching $R^2 = 0.814$ on direction 4 where $\mu_4 = 0$ while holding only 0.152 on the top decision direction; the reweighted arm reaches 0.684 and 0.588 on directions 0 and 1 and puts below 0.005 into every zero-weight direction. The min–max ranges carry two caveats the means hide. How much of the top decision direction an arm happens to retain is seed-dependent in both arms (0.00–0.54 unweighted and 0.15–0.89 reweighted on direction 0, whose weight μ_0 is the largest in the figure). Which zero-weight direction the unweighted arm spends its code on is seed-dependent too (0.07–0.71 on direction 6, where $\mu_6 = 0$), while the reweighted arm stays under 0.02 on every zero-weight direction. Drawn from saved encoder weights.

H.10 PROTOCOL

We use 20 seeds per configuration in every family, paired by seed, one seed setting both the data and the model initialisation, so that in the synthetic case each contrast spans 20 seeds, each fixing one dataset and one initialisation. We test oriented deltas two-sided (H_1: candidate differs from baseline) with a one-sample t-test and a Wilcoxon signed-rank test, reading the direction of any effect from the sign of Δ, and correct within each baseline family by Holm-Bonferroni ($K{=}15$ capacity, $K{=}6$ belief identification, $K{=}3$ fully observed decision-relevant error). The $K{=}20$ weighting-matched contrasts form their own family, and every further sweep is corrected within itself, against its own baseline unless its section states otherwise. The matched family's membership is the list of 20 pairs in Table 19: per-seed paired differences between two arms of one setting that differ in the reweighting flag alone, with the same state and observation dimensions, observation mode, cost blocks, cost horizon, code dimension, encoder, step size and 20 seeds, drawn from the three baseline sweeps and from the step-size and cost-structure sweeps that followed them. The list was fixed at $K{=}14$ before any of its p-values was computed, and grew to $K{=}20$ when later sweeps supplied six partners it had been missing. Later pairs of this kind, in the sweeps of Appendices D.6 and D.7, are reported in their own families, and pairs of this kind in two exploratory sweeps that no reported number uses were not admitted. Holm–Bonferroni is not incremental, so the whole family is re-corrected and re-reported at the enlarged K rather than the new members being appended to it. The family runs across all three linear-Gaussian settings jointly, not one family per setting, because the claim it supports is quoted across the three settings in one sentence. It is disjoint from the three baseline families above except for three members, the reweighted $k{=}2$ contrast of each setting, whose unweighted partner *is* that setting's shared baseline. For those three we report the larger of the two corrected values. We treat the corrected p_{corr} as primary; Wilcoxon is a non-parametric check, and we report disagreement as borderline. We compute 95% BCa bootstrap intervals with 10,000 resamples and paired Cohen's d_z as effect size; every interval we plot or tabulate uses this procedure, with three exceptions: the bars of Figure 5 and the shaded bands of Figure 8 are 95% Student-t intervals of each arm's mean, and the intervals of Appendix H.15 are paired t-intervals against the exact weight. Contrasts whose interval contains zero we report as showing no statistically significant difference, and we do not describe them as improvements. We report the metric at the final training step, without early stopping. All linear-Gaussian capacity runs and belief-identification runs train a fixed 100 epochs of 50 steps with Adam at learning rate 3×10^{-4}, batch 256 and gradient clipping at 100. The four control-suite sweeps, which no number in the main text uses, train a fixed episode budget at batch 32 and sequence length 50, and their curves are still rising at that budget. Every arm within a family inherits the same optimiser settings. The one exception is the per-arm step-size grid of Appendix H.5 (§4), which runs the arms of the $d{=}16, n_o{=}4$ setting at step sizes from $\{1, 3, 9, 30\}\times 10^{-4}$ and reports each arm at its own best as well as at the shared value. We regenerate evaluation trajectories per seed. We scope all conclusions to these three linear-Gaussian capacity settings and two belief-identification settings. The certificate-route comparison of Appendix H.7 adds a fourth linear-Gaussian setting, $d{=}32, n_o{=}8$, whose configuration that appendix gives; its runs are counted in the certificate totals of §5.

H.11 EXTENDED RELATED WORK

Latent world models. Latent dynamics models optimise behaviour in a learned latent space, learning the dynamics from pixels and planning or training a policy inside the model (Hafner et al., 2019; 2020). The recurrent state-space model is a latent transition model learned from pixels, inside which a cross-entropy-method planner searches for actions (Hafner et al., 2019); the actor-critic version instead trains a policy by backpropagating through imagined latent rollouts (Hafner et al., 2020). Later versions scale the same recipe with discrete latents trained by straight-through gradients (Hafner et al., 2021) and with fixed hyperparameters across domains, return normalisation in the actor loss and a symexp two-hot loss for reward and value prediction (Hafner et al., 2025). Both retain a reconstruction term, so the encoder is asked to keep whatever the decoder needs. A parallel line drops reconstruction and trains the latent model against quantities used by the planner. Value prediction networks train abstract states to predict future values rather than future observations (Oh et al., 2017); MuZero's only training signal is reward, value and policy prediction, so the latent is never required to be decodable (Schrittwieser et al., 2020). TD-MPC learns a task-oriented latent model for short-horizon trajectory optimisation with a learned terminal value (Hansen et al., 2022), and its successor extends the design to a single 317M-parameter agent performing 80 tasks across

several task domains (Hansen et al., 2024). These methods choose *what* the latent must support; our question is which encoder directions survive once capacity binds, given a fixed predictive operator. The objective-mismatch framing motivates that question: a model trained to predict well is not thereby trained to control well, and the two dissociate empirically (Lambert et al., 2020). Wei et al. (2024) survey the solutions to objective mismatch and organise them into a taxonomy; our weight changes the model-learning objective, as the value-aware methods do, and adds a condition for when that weight reproduces the cost.

Joint-embedding predictive architectures. A joint-embedding predictive architecture (JEPA) predicts the representation of a target view from a context view rather than predicting the target in input space, using a context encoder, a target encoder and a predictor between their embeddings (LeCun, 2022). The image-based instantiation trains that triple over masked blocks of a single image, with the target encoder given by an exponential moving average of the context encoder (Assran et al., 2023); the video instantiation predicts masked spatio-temporal blocks of a video from the rest, and reports prediction of later frames from earlier ones only as an ablation (Bardes et al., 2024). We use the term in that sense: the windowed belief-identification models of Section 4 are JEPA models, with the predictor trained jointly with the encoder and a separately initialised target encoder, updated as an exponential moving average of it, and the encoder is the object whose retained directions we characterise. In linear models the retained directions of such objectives are known: idealised self-predictive learning performs principal-subspace PCA on the transition matrix, with no reward in the objective (Tang et al., 2023); the stationary points of a reconstruction loss span the top-k singular vectors, and reconstruction with a temporal-difference loss is guaranteed not to minimise value error when the eigenvectors spanning the reward lie outside that span (Voelcker et al., 2024); and deep linear JEPAs are biased towards features with large regression coefficients where reconstruction prioritises variance (Littwin et al., 2024). Theorem 1 is the fixed-target, partially observed counterpart of those unweighted results, and Theorem 2 changes the operator whose principal subspace is kept; the moving target of our JEPA arms lies outside Theorem 3, which assumes a fixed target map. Multi-step inverse models recover the control-endogenous state with a multi-step inverse loss, $-\log P(a_t \mid f(x_t), f(x_{t+k}); k)$, and a discrete code (Lamb et al., 2023), outside the quadratic-risk setting of Theorem 3. Context-conditioned world models infer a latent description of the dynamics from a window of recent transitions, either by adapting a learned dynamics model online from recent experience (Nagabandi et al., 2019) or by conditioning the dynamics model on a context encoder over transition histories (Lee et al., 2020). Those context encoders are trained by prediction loss alone, so what they retain is determined implicitly.

Decision-aware model learning. Value-aware model learning replaces the likelihood by the error in the model's expected next-state value, taken robustly over a value class (Farahmand et al., 2017) or at the value iterate the planner currently holds (Farahmand, 2018); with a linear value class the robust loss is the norm of the projection of the model error onto the value features, an operator weight of the kind used here rather than a weight on data points. Value-gradient weighting scales the squared prediction error by a per-sample diagonal matrix built from the value gradient (Voelcker et al., 2022). Calibrated value-aware learning shows that these losses, as normally used, are uncalibrated surrogates that need not recover the correct model and value, and gives conditions under which a latent model shares the value-aware optimum (Voelcker et al., 2025); those are conditions on a surrogate's minimiser, not on the existence of a target weight. Grimm et al. (2020) formalise value equivalence, characterising which models are interchangeable for planning; value-consistent representation learning trains a latent transition model for value consistency and uses it to shape the representation of a search-free agent (Yue et al., 2023). The idea is older outside reinforcement learning. Identification for control holds that, with a restricted-complexity model set, the control objective must determine the distribution of the plant-model error (Gevers, 1993); control-relevant prefiltering makes the estimated model retain the plant characteristics most significant for the control requirements (Rivera et al., 1992); and, at fixed model order, a data filter together with a chosen reference-excitation spectrum makes the least-squares identification criterion and a weighted closed-loop performance criterion share a minimiser (Zang et al., 1995). Those are sufficient conditions, over frequency, for two criteria to share a minimiser; Theorem 2 gives a necessary and sufficient condition for a weight through T to reproduce the decision cost exactly, which is sufficient, not necessary, for the optimum to keep the decision-relevant subspace, and it names that subspace. Task-based end-to-end learning trains a probabilistic model on the task loss of a stochastic program rather than on prediction accuracy, and uses a cost-weighted error only as a baseline (Donti et al., 2017);

smart predict-then-optimize replaces the prediction loss by a convex surrogate for decision regret (Elmachtoub & Grigas, 2022); and decision-focused learning differentiates through a continuous relaxation of a combinatorial optimisation problem (Wilder et al., 2019), one of several gradient-based classes (Mandi et al., 2024). The same coupling of prediction to a downstream objective appears in operations research, where a tree-based predictor feeds a ship-inspection optimisation (Yan et al., 2020). Related reinforcement learning analyses give regret in terms of the complexity of the model class needed to represent value backups (Ayoub et al., 2020), or take the error in the policy gradient computed through the learned model as the loss, which model error bounds (Abachi et al., 2020). Control-oriented model learning trains the model on the agent's return through the model's Bellman fixed point and bounds the resulting error in the optimal action values (Nikishin et al., 2022). Goal-aware prediction directs a dynamics model, conditioned on the goal, towards the task-relevant parts of the state, without a reward function or labels (Nair et al., 2020); a joint objective for model and policy makes updates to either increase a lower bound on expected return (Eysenbach et al., 2022); and Denoised MDPs factor out of the world model information that is not both controllable and reward-relevant (Wang et al., 2022). Each of these prescribes an objective or bounds a quantity computed from the model; none gives a condition for a target weight to reproduce the decision cost through T, or the subspace a capacity-limited optimum then keeps, which is what Theorem 2 addresses. Value-directed compression names the decision-relevant subspace of a POMDP belief: the best lossless linear compression is the smallest invariant subspace containing the reward, a Krylov subspace (Poupart & Boutilier, 2002); the linear-Gaussian analogue is $(\ker D)^{\perp}$, and Theorem 2 asks when the rank-k optimum of a fixed predictive objective lies inside it. Predictive state representations represent the state of a dynamical system by multi-step, action-conditional predictions of future observations (Littman et al., 2001), and a spectral algorithm learns them consistently from action-observation sequences (Boots et al., 2011); $T\hat{x}_t = \mathbb{E}[y_m(t) \mid o_{\leq t}]$ is a predictive-state vector of that kind, with no action to condition on, and we ask which of its directions a capacity-k optimum keeps and when a weight through T makes them the decision-relevant ones. Approximate information states bound the loss of compressing a history (Subramanian et al., 2022).

Bisimulation. Bisimulation metrics address retention through state similarity, which their encoders are then trained to reproduce. The original metrics assign a distance that is zero exactly when two states are behaviourally indistinguishable (Ferns et al., 2004), and later work makes the computation tractable in deterministic settings by replacing the Wasserstein term with a cheaper recursion (Castro, 2020). DeepMDP couples latent prediction to bisimulation-style abstraction (Gelada et al., 2019). Among bisimulation methods, deep bisimulation for control is closest in kind, since it also constrains what a latent encoder keeps, but the constraint is reward-and-transition similarity between states rather than a spectral condition on the predictive operator (Zhang et al., 2021). For linear systems a bisimulation quotient collapses a subspace inside the kernel of the output map that is controlled invariant, the disturbance or unmatched input playing the role of the control, and invariant when there is none (Pappas, 2003; van der Schaft, 2004). With the per-step cost as the output, the fixed law closing the loop and no disturbance, the directions the coarsest bisimulation collapses form the unobservable subspace of $(A_{\mathrm{cl}}, [H; \sqrt{\rho}K_c])$, which lies in $\ker D$ at every horizon and equals it once $t_h \geq d - 1$; a probabilistic bisimulation restricted to linear quotients collapses the same subspace, since additive noise that does not depend on the state leaves both conditions unchanged. Were the process noise instead treated as a disturbance entering every state direction, the coarsest quotient would collapse all of $\ker H \cap \ker K_c$, which can be larger than $\ker D$. Theorem 2 clause 3 gives $\ker D \subseteq \ker X$ at the reweighted optimum, so the weighted code drops every direction this bisimulation collapses, while the unweighted optimum can keep one (clause 5). What separates our condition from bisimulation is the other direction: a capacity-k code must also rank the directions bisimulation keeps, and Theorem 2 says which survive, by cost. Task-informed abstractions likewise select what the encoder retains, but by separating task-relevant from task-irrelevant factors (Fu et al., 2021).

Information-bottleneck objectives pose the trade-off in a different currency, trading the mutual information between input and code against the mutual information between code and target through a Lagrange multiplier (Tishby et al., 1999), with the variational version bounding both terms for deep encoders (Alemi et al., 2017). For Gaussian variables the bottleneck names its retained directions: left eigenvectors of $\Sigma_{x|y}\Sigma_x^{-1}$, the canonical-correlation basis, entering through a cascade of phase transitions as the trade-off parameter grows (Chechik et al., 2005). Applied to the input past and output future of a dynamical system, the trade-off becomes a generalised eigenvalue problem,

related to canonical correlation analysis, whose structural phase transitions add state-space dimensions (Creutzig et al., 2009); that past-future bottleneck is the information-bottleneck analogue of Theorem 1. Those directions are unchanged by any invertible reweighting of the target, since $\Sigma_{x|y}$ is; a rank constraint with a decision weight keeps the top eigenspace of D^*D instead. Our bottleneck is a rank constraint rather than a rate penalty. Our weight is an operator on the predictive target, replacing the Gram operator by T^*WT; a per-sample scalar weight (Lambert et al., 2020; D'Oro et al., 2020) reweights data points, and a constant weight is inert; the per-sample diagonal of Voelcker et al. (2022) weights each coordinate of the target differently for each sample, rather than scaling a data point by one scalar. Appendix H.4 measures both statements on this paper's generator rather than asserting them: the per-sample scalar moves the retained subspace 4.9–11.8% as far as the operator weight at k=2, with disjoint intervals, and a constant weight reproduces the unweighted arm. Input reweighting is recovered only when $\operatorname{rank} D = \operatorname{rank} T$. For one fixed linear policy, one quadratic value function, finite horizon and linear models sharing the cost and the noise covariance, M_1, M_2 are value-equivalent exactly when $M_1^\top P_{t_h} M_1 = M_2^\top P_{t_h} M_2$; matching the decision map is sufficient but not necessary. We likewise characterise only the exact optimum of a population risk.

Linear encoders and principal subspaces. In Theorem 3 we lower-bound the risk of any capacity-k pair with a linear decoder by the tail $\sum_{i \geq k} \mu_i$, so arbitrary and linear encoders share the optimum, the composite collapsing to a projection onto the decoder range. This is classical: for auto-association the hidden nonlinearities are useless and the optimum follows from singular value decomposition (Bourlard & Kamp, 1988), and in the general linear regression case a rank-p critical map is the least-squares regression matrix followed by a projection onto p eigenvectors of the covariance of its fitted values (Baldi & Hornik, 1989), which is reduced-rank regression (Izenman, 1975; Reinsel & Velu, 1998). Theorem 1 clauses 3 to 6 are its population, identity-weight form; the retained subspace of Theorem 2 clause 3, given the weight $W = V^*V$, is that of reduced-rank estimation of the weighted response Vy, which keeps the leading singular directions of VTR (Hua et al., 2001). Goal-oriented low-rank approximation of linear-Gaussian inverse problems likewise places its optimal update directions orthogonal to the null space of a goal operator, in the prior-precision inner product (Spantini et al., 2017); there the goal acts on the parameters directly, whereas our weight can act only on the predicted observations, through T, which is possible exactly when $\ker T \subseteq \ker D$. We use the decoder step in Theorem 2's weighted retention condition.

Cost-weighted model reduction. Component-cost analysis ranks states by contribution to a quadratic cost (Skelton, 1980); covariance-equivalent realisations extend it to large-scale systems (Yousuff & Skelton, 1985). Balanced truncation ranks instead by joint controllability and observability (Moore, 1981). LQG balancing ranks state components by their contribution to the closed-loop LQG behaviour of the system (Jonckheere & Silverman, 1983); frequency-weighted balancing selects by input and output weightings (Enns, 1984); goal-oriented and quadratic-output reduction select by a quantity of interest (Carlberg & Farhat, 2011; Pulch & Narayan, 2019; Benner et al., 2022); component-cost analysis also reduces controllers (Yousuff & Skelton, 1984). Equation 1 reads as a cost-weighted reduction of an encoder rather than of a plant. Sample-complexity analyses of the linear-quadratic regulator tie control suboptimality to model estimation error (Dean et al., 2020), with certainty equivalence shown to do better, quadratically in the parameter error (Mania et al., 2019); those bounds price plant error and rank no encoder direction. Nearer to our setting, cost-driven representation learning for LQG regresses multi-step costs and has finite-sample guarantees when the cumulative cost observes every state direction and the latent has the state's dimension (Tian et al., 2023); we keep observation prediction, allow $\ker D \neq 0$ and $k < d$, and characterise which directions the weighted optimum keeps. Subspace identification truncates the past-to-future predictor at a chosen order under user-defined weights that determine which part of its information is retained (Van Overschee & De Moor, 1995), solving the reduced-order past/future prediction problem by canonical variate analysis (Larimore, 1990); our V is an output-side weight of that kind derived from the decision operator, and Theorem 2 says when it reproduces the cost. LQG balancing ranks the states of a given realisation and says nothing about which directions a trained encoder keeps; in Theorem 2 we name the subspace surviving at the population risk's exact optimum, given the predictive operator and the weight, and bound no control cost.

Reduced-order and rate-limited estimation. Two lines constrain the compensator or estimator itself rather than the plant, and with the works above are prior art for the question Theorem 2 asks.

Optimal-projection theory gives first-order necessary conditions for the best fixed-order dynamic compensator (Hyland & Bernstein, 1984) and for the best reduced-order state estimator (Bernstein & Hyland, 1985), each coupled by a projection whose rank equals the order; its model-reduction form relates the component-cost and balancing methods above (Hyland & Bernstein, 1985). Rate-limited control replaces the order constraint with an information one: minimum-directed-information LQG realises the optimal policy as a virtual sensor, a Kalman filter and a certainty-equivalence controller, designed sequentially with a semidefinite program for the sensor (Tanaka et al., 2018), and rationally inattentive control bounds the mutual information a stationary randomised policy may use (Shafieepoorfard et al., 2016). Both lines characterise the *optimal* compressor under their constraint, optimal projection through first-order necessary conditions that can have several solutions, for a control or estimation criterion fixed in advance; neither starts from a predictive training objective. We ask the complementary question: given a predictive objective and a rank-k constraint, which directions does its optimum keep, and what weight makes those the decision-relevant ones. That is why a weight appears here and not there, and why our statement is about one objective's optimum rather than about the best cost attainable at a given rate or order.

Joint-embedding objectives. Whitening removes the collapse mode by decorrelating embedding coordinates (Ermolov et al., 2021; Uelwer et al., 2025); variance-invariance-covariance regularisation replaces it with separate variance and covariance penalties (Bardes et al., 2022). Both control the embedding's second-moment structure; VICReg's own ablation finds a predictor redundant with its variance term (Bardes et al., 2022), and what we add is the decision weight on the predictive target, not the predictor.

H.12 Notation

Notation. Symbols local to a single theorem, proof or experiment, together with the indexing convention used there, are defined at the point where they appear. The encoder is written ϕ in §3 and f_ϕ where its parameters ϕ are separated from it, and the loss below is optimised over (ϕ, Γ). Two operators recur: T, the population predictive operator of the generator, and D, the stacked decision operator of §3. The generator's matrices are the state matrix $A \in \mathbb{R}^{d\times d}$ and the observation matrix $C \in \mathbb{R}^{n_o\times d}$; writing n_c for the cost-block count, the performance output H and the fixed gain K_c are $n_c \times d$ and the input matrix B is $d \times n_c$. The closed loop is $A_{\mathrm{cl}} = A - BK_c$, and ρ is the scalar control weight, which enters as $\rho\, K_c^\top K_c$ rather than through a separate control weight matrix. The letter R is the covariance factor of the theorems, except inside Theorem 4's instantiation clause, which carries its own control weight $R = R_f^\top R_f$ and defines it there. The trained decoder Γ of §3 is the linear decoder the theorems write M; the implementation gives it and the linear encoder a bias, so both are affine, on data the generator draws with zero mean. The letter H names the performance output alone: $\mathcal{H}$ is the horizon set and H_1 an alternative hypothesis. The Gramians of T and D carry the spectra used in the theorems: $G = T^*T$ with eigenvalues $\lambda_0 \geq \cdots \geq \lambda_{d-1}$, constrained by assumption (A5) of the first theorem of §3 and by the "dominant k-dimensional invariant subspace of G" in its fifth clause; and $P = D^*D$, the *decision Gramian*, with eigenvalues $\mu_0 \geq \cdots \geq \mu_{d-1}$ and orthonormal eigenbasis $b_0, \ldots, b_{d-1}$, constrained by the gap condition $\mu_k < \mu_{k-1}$ in Theorem 2 (A5) and carried by Theorem 5 throughout, with two exceptions: Theorem 3's μ is the spectrum of $(DR)^*(DR)$ for its atom factor R, and the certificate's μ in Appendices H.7 and H.20 is the arm's own spectrum, that of $R^*(T'^*T')R$ for the operator T' the arm regresses onto (Equation 4). The letter b_i follows the Gramian in play: it is the eigenbasis of $G = T^*T$ in the Setting of §3, in Theorem 1 and in Figure 1(a); of $P = D^*D$ in Figure 1(b) and in Theorem 5; and of its covariance-weighted form R^*PR on the axis of Figure 2; inside a proof the step names the Gramian, and the witnesses of Theorems 2, 3 and 4 use it for a plain orthonormal basis. A third Gramian appears only inside Theorem 2 clause 4 and its proof: $G' = T'^*T'$ for the operator an arm regresses onto, T', whose eigenvalues are written $\nu_0 \geq \cdots \geq \nu_{d-1}$, and it is ν, not μ, that clause 4's own gap hypothesis $\nu_k < \nu_{k-1}$ constrains. The letter P names one object: bare P is the decision Gramian D^*D, and P_{t_h}, subscripted by a cost horizon, is the cost-to-go matrix, which equals it. Orthogonal projections carry Π, as in Π_S onto a subspace, Π_k onto the span of the first k eigendirections, and $\Pi_\perp$ of §4. We write d for $\dim \mathcal{E}$ throughout, except in Theorem 5, whose preamble sets $\dim \mathcal{E} = n$ and uses n in its clauses and proof; the two letters name one dimension. The three inner-product spaces are $\mathcal{E} = \mathbb{R}^d$, the belief space, $\mathcal{F} = \mathbb{R}^{mn_o}$, the stacked-observation target space, and $\mathcal{G} = \mathbb{R}^q$, the decision operator's output space, with $T : \mathcal{E} \to \mathcal{F}$, $V : \mathcal{F} \to \mathcal{G}$

and the composite $\tilde{T} = V \circ T : \mathcal{E} \to \mathcal{G}$, which equals D when the reweighting is realizable. The theorems add W, the positive semidefinite weight of Theorem 2 clause 1; X, the composite of a capacity-k pair; $\Sigma = RR^*$, the belief covariance and its factor; and $\mathrm{Obs}(T, D)$ with its dimension ω, the obstruction subspace of Theorem 4. Six further letters are local to where they appear: S and s name a recoding matrix, an index set, a scalar, a dominant index block, three different subspaces and a perturbation size; b is an eigenbasis vector and a state block index; N is a squared norm in Theorem 3's proof and a subspace in Theorem 4's; α is a witness coefficient, a significance level and the premise-violation dial of §D.4; L is a linear extension, a Cholesky factor and, as $\mathcal{L}$, the training loss; p is a p-value, a kernel dimension and an atom weight. Q, A and K likewise take a second meaning inside a single proof or appendix, stated there. Three pairs are separated by their form rather than by their context: ϵ_t is the process noise and ε is Theorem 2 clause 4's slack; η_t, indexed by time, is the observation noise, and η with no index, or with a named subscript as in η_{norm}, is the certificate of Equation 2 or a reduced form of it; n written with a numeral, as $n{=}20$, is a seed count.

The two families' windows. Let $\hat{x}_t \in \mathbb{R}^d$ be the Kalman belief mean of the linear-Gaussian system. The two families differ in their encoder input, and the settings in this paragraph are the *belief-identification* family's: a 16-step causal window encoded into a code $z_t \in \mathbb{R}^k$, striding densely over each trajectory so consecutive windows overlap, with prediction targets at the horizon set $\mathcal{H} = \{1, 3, 5\}$ measured from the window endpoint, so each of its objectives is applied at $m = 3$ offsets. The *capacity* family, which carries the paper's headline result, uses a 32-step window and predicts the next $m = 4$ consecutive observations; Appendix B states both alongside every other setting. Horizons enter the training objectives only; R^2 is read at the anchor. Equation 1 is the capacity family's objective alone: its target is the stack of the next $m{=}4$ observations, $y_m(t) = (o_{t+1}^\top, \ldots, o_{t+m}^\top)^\top \in \mathbb{R}^{mn_o}$, and the belief-identification family trains the contrastive and joint-embedding objectives of §4 instead. The decision-weighted variant imposes no Kronecker constraint on V, and none is available to it: the m target blocks lie in observation space, while the cost-to-go matrix P_{t_h} of §3 acts on belief space, so there is no blockwise repetition of P_{t_h} to make. It builds instead the single $q \times mn_o$ matrix $V = DT^+$, with T^+ the Moore–Penrose pseudo-inverse, which is the factor Theorem 2 clause 1 supplies. Writing $\Pi = T^+T$ for the orthogonal projector onto $(\ker T)^\perp$,

$$(VT)^*(VT) = (D\Pi)^*(D\Pi) = \Pi\, D^* D\, \Pi = \Pi\, P_{t_h}\, \Pi, \tag{6}$$

which is P_{t_h} itself, and $\tilde{T} = D$, exactly under the realizability condition $\ker T \subseteq \ker D$. The reweighting belongs to that variant alone. The cost-to-go matrix itself is, with the performance output H and the closed loop $A_{\mathrm{cl}} = A - BK_c$ of §3,

$$P_{t_h} = \sum_{t=0}^{t_h} (A_{\mathrm{cl}}^t)^\top \big(H^\top H + \rho\, K_c^\top K_c\big) A_{\mathrm{cl}}^t, \tag{7}$$

with $\rho > 0$. The stacked decision operator is $D = \big[\, HA_{\mathrm{cl}}^t \,;\, \sqrt{\rho}\, K_c A_{\mathrm{cl}}^t \,\big]_{t=0}^{t_h}$, so stacking the two blocks at each stage and summing over stages gives

$$D^*D = \sum_{t=0}^{t_h} (A_{\mathrm{cl}}^t)^\top \Big(H^\top H + \big(\sqrt{\rho}\big)^2 K_c^\top K_c\Big) A_{\mathrm{cl}}^t = \sum_{t=0}^{t_h} (A_{\mathrm{cl}}^t)^\top \big(H^\top H + \rho\, K_c^\top K_c\big) A_{\mathrm{cl}}^t = P_{t_h}, \tag{8}$$

where the middle equality is $\big(\sqrt{\rho}\big)^2 = \rho$. The identity $D^*D = P_{t_h}$ holds with the same ρ as Equation 7, matching the identity carried by the Lean development.

Here V is the output-side factor of Theorem 2, realizable only when $\ker T \subseteq \ker D$, and Γ is its linear decoder M; a trained run meets neither the exact-optimum nor the population-risk premise. The reweighted constraint fixes V only up to a left orthogonal factor: if $Q^*Q = I$ then $(QV)^*(QV) = V^*V$, so the quadratic form in Equation 1 and its optimum are invariant to the choice, which the implementation settles by taking $V = DT^+$; the trained numbers were not checked for that invariance. The weight acts on the output side, weighting the error and not the input, because the cost is incurred on the error, not on the input; and it is an operator, not a per-sample scalar, because the cost depends on the direction the error points in. We build P_{t_h} in closed form from the known generator matrices, an oracle unavailable outside the synthetic family. The short-horizon variant sets $t_h{=}0$, a single-stage rather than accumulated weight.

Table 20: Belief identification: affine R^2 on unobserved factors (↑), 20 seeds per row, computed after we fit the optimal affine alignment to the unobserved part of the Kalman belief mean; $\pm$ is the standard deviation over the 20 seeds with divisor 20. Δ is the paired improvement over the setting baseline; we correct p_{corr} with Holm–Bonferroni over $K = 6$ comparisons within each setting, and the p_{corr} column corrects the two-sided paired t-test, which is authoritative for each improvement claim once the sign of Δ establishes the direction; the Wilcoxon signed-rank test is a robustness check reported in the per-configuration tables. Setting labels give the latent dimension d and the observation dimension n_o, as in the figures and per-configuration tables; the three windowed variants are named by their regulariser, so that "Windowed, whitening", "Windowed, VICReg" and "Windowed, no regulariser" are the arms written `JEPA Whiten`, `JEPA VICReg` and `JEPA None` elsewhere, with the setting suffix (`D16 D8`, `D4 D2`) encoding (d, n_o) in that order; Autoencoder is a deterministic recurrent autoencoder trained to reconstruct the current observation, written `Autoencoder` in the per-configuration tables and the `dvae` arm in the supplementary code; the non-sequential encoder is the MLP encoder, written `MLP Encoder`; and the naive hybrid, written `Naive Hybrid`, adds the contrastive term to a reconstruction loss while keeping the reconstruction decoder. Per-configuration tables are authoritative for per-arm win rates, which are proportions over the 20 seeds. The d_z column is the paired Cohen's d_z of Appendix C.1, the same number the statistical tables print unqualified; the header repeats the reading that appendix gives it, namely seed-to-seed reproducibility of a near-deterministic optimisation rather than a generalisable effect magnitude. An em dash marks a quantity that a baseline row does not define, the baseline having no Δ, no interval and no p; every non-baseline row prints every column. Boldface marks the largest improvement per setting.

Setting	Variant	R^2 (mean $\pm$ std)	Δ (↑)	95% CI	d_z (repro.)	p_{corr}
d=16, n_o=8	Baseline	0.31 ± 0.01	—	—	—	—
d=16, n_o=8	Windowed, whitening	0.68 ± 0.04	**0.370**	$[0.352, 0.385]$	9.61	< 0.001
d=16, n_o=8	Windowed, VICReg	0.62 ± 0.04	0.309	$[0.290, 0.323]$	8.16	< 0.001
d=16, n_o=8	Windowed, no regulariser	0.53 ± 0.04	0.215	$[0.198, 0.231]$	5.69	< 0.001
d=16, n_o=8	Naive hybrid	0.31 ± 0.01	-0.003	$[-7.10\times10^{-3}, 1.83\times10^{-5}]$	-0.39	0.099
d=16, n_o=8	Autoencoder	0.25 ± 0.02	-0.064	$[-0.073, -0.054]$	-2.90	< 0.001
d=16, n_o=8	Non-sequential encoder	$(1.74 \pm 0.25)\times10^{-3}$	-0.309	$[-0.313, -0.305]$	-34.91	< 0.001
d=4, n_o=2	Baseline	0.50 ± 0.11	—	—	—	—
d=4, n_o=2	Windowed, VICReg	0.75 ± 0.15	**0.255**	$[0.180, 0.364]$	1.22	< 0.001
d=4, n_o=2	Windowed, whitening	0.69 ± 0.21	0.193	$[0.0954, 0.320]$	0.73	0.012
d=4, n_o=2	Windowed, no regulariser	0.56 ± 0.13	0.069	$[0.003, 0.176]$	0.35	0.27
d=4, n_o=2	Naive hybrid	0.48 ± 0.12	-0.011	$[-0.091, 0.035]$	-0.08	0.74
d=4, n_o=2	Autoencoder	0.19 ± 0.09	-0.301	$[-0.341, -0.233]$	-2.49	< 0.001
d=4, n_o=2	Non-sequential encoder	$(5.11 \pm 3.29)\times10^{-4}$	-0.495	$[-0.523, -0.389]$	-4.33	< 0.001

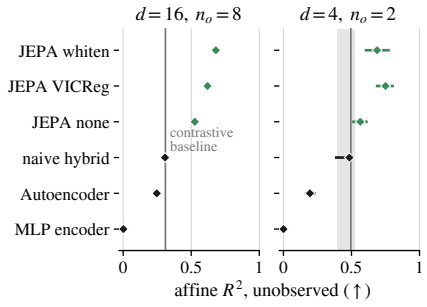


Figure 10: Affine R^2 on unobserved factors for every belief-identification arm: diamonds are means over 20 seeds and horizontal bars their 95% BCa intervals, green for the three windowed JEPA arms and black for the other three; the grey vertical line and band are the contrastive baseline's mean and its 95% BCa interval.

How to read Table 21. Setting labels give the latent dimension d and the observation dimension n_o, as in the per-configuration tables and appendix sections; k is the code dimension. We orient Δ so that positive means *lower* (better) error relative to the weak baseline of that setting, whose failure clause 5 of Theorem 2 exhibits. Rows marked † change P_{t_h}, and so the decision problem: the short-horizon rows shorten the cost horizon, and the realizability-violating row moves one of its two cost blocks onto an unobserved block. Their Δ subtracts the arm's error on its own P_{t_h} from the baseline's on the shared one, so it is not comparable with other rows of the same setting. We correct p_{corr} with Holm–Bonferroni within each setting. The p_{corr} column corrects the two-sided paired t-test, which is authoritative for each improvement claim once the sign of Δ establishes the direction; a small p_{corr} on a row whose Δ is negative therefore records a significant *increase* in error,

Table 21: Decision-relevant error in the linear-quadratic-Gaussian belief settings, 20 seeds per row; the raw means are decision-relevant error, for which lower is better (↓). The rows shown are the baseline, the two reweighted code dimensions k=2 and k=4, the unweighted arm at the matched code dimension k=4, any short-horizon or realizability-violating arm of that setting, and the full-capacity k=d arm. The scale for the Δ column is the metric's floor, the filter's own stationary error in closed form from the generator (0.711, 0.660 and 0.499 here; §5): the attainable range at d=16, n_o=4 is $0.968 - 0.711 = 0.257$, of which the reweighted k=4 arm's 0.205 covers 80%, and at d=8, n_o=3 its 0.211 covers 88% of 0.241; under full observation the range is $0.913 - 0.499 = 0.414$, of which 0.372 covers 90%. The full-capacity row is not that floor and is not a bound: it is what this architecture reaches with the capacity constraint removed, 0.750 and 0.680, still 0.039 and 0.020 above the closed-form floor. That residual is optimisation gap and the error the encoder's finite window adds, not a capacity limit. The two partially observed blocks print the unweighted arm at k=1 as well, so every block carries an unweighted arm at the code dimension of the best reweighted arm of its setting. For the matched unweighted arms of the d=16, n_o=4 and d=8, n_o=3 settings we print Δ alone. Their interval, d_z and both raw p-values are the `A1 Unweighted K4` rows of Tables 4 and 5, where the *Sig.* column stands in for p_{corr} and marks both as surviving Holm–Bonferroni. Their premise check is printed here, in place, like every other row's. Per-seed dispersion is not printed for them anywhere. Those tables are also authoritative for per-arm win rates, which are proportions over the 20 seeds. Boldface marks the largest reduction among the capacity-limited rows shown at the shared cost horizon. The full-capacity k=d rows of the same two blocks reach larger reductions, 0.218 against the bold 0.205 and 0.221 against the bold 0.211, and are printed here two rows below: they are what the architecture reaches with the capacity constraint removed, not a competing arm, which is why the bold is not theirs. Column conventions, the sign of Δ and the premise legend are stated in the accompanying paragraph on how to read this table.

Setting	Variant	Error (mean ± std)	Δ (↑)	95% CI	d_z (repro.)	p_{corr}	Premise
d=16, n_o=4	Baseline	0.968 ± 0.019	—	—	—	—	✓
d=16, n_o=4	Reweighted k=4	0.763 ± 0.012	**0.205**	[0.194, 0.212]	9.92	< 0.001	✓
d=16, n_o=4	Reweighted k=2	0.845 ± 0.027	0.123	[0.105, 0.136]	3.44	< 0.001	✓
d=16, n_o=4	Unweighted k=4	—	0.0669	—	—	—	✓
d=16, n_o=4	Unweighted k=1	0.982 ± 0.013	−0.0137	[−0.026, −0.003]	−0.50	0.073	✓
d=16, n_o=4	Reweighted k=2, short horizon†	0.698 ± 0.019	0.270†	[0.248, 0.279]	8.60	< 0.001	−
d=16, n_o=4	Realizability-violating k=2†	0.891 ± 0.012	0.0774†	[0.066, 0.085]	3.63	< 0.001	×
d=16, n_o=4	Full capacity k=d (no capacity limit)	0.750 ± 0.009	0.218	—	—	—	−
d=8, n_o=3	Baseline	0.901 ± 0.038	—	—	—	—	✓
d=8, n_o=3	Reweighted k=4	0.691 ± 0.009	**0.211**	[0.191, 0.228]	5.01	< 0.001	✓
d=8, n_o=3	Reweighted k=2	0.778 ± 0.035	0.123	[0.093, 0.144]	2.17	< 0.001	✓
d=8, n_o=3	Unweighted k=4	—	0.139	—	—	—	✓
d=8, n_o=3	Unweighted k=1	0.951 ± 0.033	−0.0503	[−0.068, −0.031]	−1.15	< 0.001	✓
d=8, n_o=3	Reweighted k=2, short horizon†	0.605 ± 0.009	0.296†	[0.280, 0.312]	7.78	< 0.001	−
d=8, n_o=3	Full capacity k=d (no capacity limit)	0.680 ± 0.007	0.221	—	—	—	−
Full obs., d=8	Baseline	0.913 ± 0.042	—	—	—	—	✓
Full obs., d=8	Reweighted k=4	0.541 ± 0.013	**0.372**	[0.355, 0.391]	8.79	< 0.001	✓
Full obs., d=8	Reweighted k=2	0.671 ± 0.029	0.242	[0.220, 0.266]	4.56	< 0.001	✓
Full obs., d=8	Unweighted k=4	0.769 ± 0.046	0.144	[0.120, 0.170]	2.47	< 0.001	✓

as on the unweighted k=1 row at d=8, n_o=3; the bootstrap interval is descriptive, and the Wilcoxon signed-rank test is a robustness check reported in the per-configuration tables. The d_z column is the paired Cohen's d_z of Appendix C.1, the same number the statistical tables print unqualified; the header repeats the reading that appendix gives it, namely seed-to-seed reproducibility of a near-deterministic optimisation rather than a generalisable effect magnitude, and we do not rank arms by it. In the premise column, ✓ marks a row whose structural premises of Theorem 2, realizability and $\ker T \subseteq \ker D$, we checked and found to hold. A × marks a violated structural premise. A − marks a row that fails the strict eigenvalue gap at the capacity boundary or the positive μ_{k-1} that the recovery clauses require, placing it outside the scope of either theorem. Four rows carry it, for two different reasons. On the two short-horizon rows the t_h=0 cost weighting leaves P_{t_h} without both. On the two full-capacity k=d rows, which run at the shared t_h=3, the capacity exceeds the decision rank, so $\mu_{k-1} = 0$ and there is no capacity boundary left for a gap to sit at. A ✓ on a baseline row records the same structural check and no comparison, those rows having no Δ, no interval and no p; and no ✓ asserts the exact-optimum or population-risk antecedents, which no trained run satisfies.

H.13 Measuring the Containment, With Uncertainty

Clause 3 of Theorem 2 concludes a containment, and the two quantities of Tables 23 and 24 are where that conclusion is measured in full, arm by arm and with intervals. Every row is n=20 seeds, paired on seed, against the shared unreweighted baseline; intervals are 95% BCa over 10,000 resamples and d_z is paired Cohen's d_z, the same procedure used for every error contrast in Table 21.

Table 22: Key to eight of the linear-Gaussian arm codes, giving the encoder, the weighting, the code dimension k, the cost horizon t_h and the number of cost blocks; the family defaults are t_h=3 and two cost blocks. Arm codes not listed here are read off the same rule: an `A` prefix marks a linear encoder and `B` the GRU encoder, the middle token gives the weighting, and the trailing `K` token the code dimension k, with the family defaults above. The remaining tokens are `T0` for a cost horizon t_h=0, `Cost1` and `Cost3` for one and three cost blocks, the quantity defined in §3, `K Eq D` for k=d, `Full` for the fully observed system, and `Partial`, in `A3 Violating Partial K2`, for a partial violation, one of its two cost blocks unobserved; an arm code carrying none of the last three belongs to the d=16, n_o=4 setting. An arm code is canonically rendered in the spaced title-case form of the first column, and any rendering that differs from it only in case or separator, such as `A5_full_reweighted_k4` for `A5 Full Reweighted K4`, names the same arm. One arm carries a descriptive name in place of its code in the error table and in the text: the realizability-violating arm, `A3 Violating Partial K2` in the first column below, is called "Realizability-violating k=2" there, the `Partial` token of its code recording that the violation is partial, one of its two cost blocks sitting on an unobserved block; the figure that plots it, Figure 5, names it the same way, as it names all its rows descriptively. The error table names its other rows descriptively rather than by code. Setting suffixes carry the convention used in the figures and the logged records: `D16 D8` and `D4 D2` give (d, n_o) in that order, while `Full Obs D8` gives d alone.

Arm code	Encoder	Weighting	k	t_h	Cost blocks
A1 Unweighted K4	Linear	Unweighted	4	3	2
A2 Reweighted K2 Cost1	Linear	Reweighted	2	3	1
A2 Reweighted K2 Cost3	Linear	Reweighted	2	3	3
A2 Reweighted K2 T0	Linear	Reweighted	2	0	2
A3 Violating Partial K2	Linear	Realizability-violating	2	3	2, one unobserved
A4 Reweighted K Eq D	Linear	Reweighted	d	3	2
A5 Full Reweighted K4	Linear	Reweighted	4	3	2
B GRU Reweighted K2	GRU	Reweighted	2	3	2

Table 23: Retained decision-irrelevant energy; the theorem predicts 0, lower is better. Baselines are 0.707, 0.472 and 0.609 for the three settings in order. The k=4 maxima are why the mean alone is not enough: one seed in each setting leaks one to two orders of magnitude above that setting's median. $^{\dagger}k < \operatorname{rank} P_{t_h}$; $^{\ddagger}k = \operatorname{rank} P_{t_h}$. $\operatorname{rank} P_{t_h} = 4$ in all three settings.

Setting	Arm	Cand.	Δ	95% BCa	d_z	Win	min/med/max
d=16, n_o=4	k=2†	0.0046	0.703	[0.650, 0.754]	5.64	1.00	0.0010/0.0046/0.0094
	k=4‡	0.0384	0.669	[0.604, 0.730]	4.41	1.00	0.0021/0.0060/0.250
d=8, n_o=3	k=2†	0.0026	0.469	[0.427, 0.492]	6.59	1.00	0.0005/0.0018/0.0093
	k=4‡	0.0155	0.456	[0.416, 0.481]	6.17	1.00	0.0010/0.0038/0.125
full obs., d=8	k=2†	0.0035	0.605	[0.545, 0.670]	4.17	1.00	0.0011/0.0030/0.0066
	k=4‡	0.0093	0.599	[0.540, 0.665]	4.11	1.00	0.0018/0.0040/0.107

The last column is min / median / max over the candidate's own 20 seeds, the median taken as the mean of the two central values. $\operatorname{rank} P_{t_h} = 4$ in all three settings, so the k=2 arms must select a proper subspace of the decision-relevant one while the k=4 arms have exactly enough capacity for containment to be attainable. These are different claims and we mark them apart in the tables rather than reporting one number for both. Angles are carried at three decimals throughout so that the near-zero candidate values remain legible in the same column as the tens-of-degrees baselines: one candidate mean is zero to three decimals and four of the six per-seed minima fall below 0.1°; the digits beyond the first are not resolved by the BCa intervals on Δ, which are more than twelve degrees wide in four of the six rows.

What the k=2 rows support, and what they do not. On energy, every one of the 20 reweighted k=2 seeds sits below 0.01 in all three settings, and the worst of them is far below the best baseline seed, so the two distributions do not overlap. On the angle the picture is not uniform: at d=8, n_o=3 the effect sizes are 0.67 and 0.75 against 2.4–2.9 elsewhere, and the k=2 arm wins on 17 of 20 seeds rather than all of them. We report decision-irrelevant energy as the finite-sample surrogate for the containment, since it vanishes exactly when the containment holds, and the minimum angle as a diagnostic: small is necessary there, not sufficient. The two do not order the capacities the same way. On the surrogate, on medians as well as means, k=2 is the better arm at every setting, while

Table 24: Minimum principal angle between the retained and decision subspaces, in degrees; the theorem predicts 0°. Baselines are 42.597, 6.923 and 37.394. Point estimates are printed to three decimals to keep the sub-degree candidate values legible beside the baselines; only the leading digit of each Δ is resolved by its interval. The two d=8, n_o=3 rows carry much smaller effect sizes than the rest, and the k=2 row there is the one arm in either table that does not win on every seed. $^\dagger k < \operatorname{rank} P_{t_h}$; $^\ddagger k = \operatorname{rank} P_{t_h} = 4$, so at k=4 the retained subspace has exactly enough dimensions for the containment to be attainable and a near-zero angle says only that one retained direction lies almost inside that four-dimensional subspace; Table 23's energy is the whole-subspace reading. The k=2 rows are the ones carrying evidence about the mechanism.

Setting	Arm	Cand.	Δ	95% BCa	d_z	Win	min/med/max
d=16, n_o=4	k=2†	2.158	40.438	[34.035, 46.540]	2.73	1.00	1.244/1.935/4.346
	k=4‡	0.566	42.031	[35.840, 47.996]	2.93	1.00	0.010/0.403/1.664
d=8, n_o=3	k=2†	0.643	6.280	[3.317, 11.919]	0.67	0.85	0.044/0.697/1.172
	k=4‡	0.000	6.923	[3.992, 12.547]	0.75	1.00	0.000/0.000/0.000
full obs., d=8	k=2†	1.861	35.533	[29.311, 41.850]	2.43	1.00	0.803/1.763/3.079
	k=4‡	0.504	36.890	[30.838, 43.209]	2.56	1.00	0.077/0.286/1.974

on the angle k=4 is the better aligned, so the angle does not corroborate the surrogate here. Both quantities are scalar summaries of a subspace relationship. The angle does not separate a retained subspace that sits inside the decision-relevant one from one that merely has a small minimum angle to it. The energy does in exact arithmetic, since it is zero only under containment, but the measured k=2 energies of Table 23 are small, not zero, so they bound how far the retained subspace leans into $\ker D$ without showing that it does not lean at all. Deciding exact containment requires the retained subspace and $\ker D$ compared directly, per seed and per arm, in the coordinates the affine alignment of Table 24's companion analysis already fixes; the scalars in the two tables above marginalise exactly the per-coordinate structure that would decide it. We did not carry out that comparison, so the k=2 containment reading rests on those scalars, which bound the leakage and do not show it is zero.

The k=4 maxima. In all three settings the k=4 energy maximum is one to two orders of magnitude above that setting's median, and a single seed carries it in each case. We report the mean, the median and the maximum together rather than the mean alone, because the mean of these 20 values is not a summary a reader can invert.

H.14 How Fast Containment Degrades When the Decision Weight Is Wrong

Every P_{t_h} used elsewhere in this paper is handed over exactly by the generator, which leaves open what an error in that weight would do to the containment. This appendix reports a sensitivity measurement against a *stated* perturbation. We form $P_{\mathrm{raw}} = P_{t_h} + s\|P_{t_h}\|_F E$ at nominal relative size $s > 0$, with $\|\cdot\|_F$ the Frobenius norm, with E a symmetrised Gaussian of unit Frobenius norm drawn before s is applied, so all four arms share one drawn direction per seed; project it back onto the positive semidefinite cone, $P_{\mathrm{psd}} = \Pi_{\mathrm{PSD}}(P_{\mathrm{raw}})$; factor $D_{\mathrm{pert}} = \Lambda^{1/2}Q^\top$ from the symmetric eigendecomposition $P_{\mathrm{psd}} = Q\Lambda Q^\top$; and build the weight $V = D_{\mathrm{pert}}T^+$ from it. The perturbed weight enters the training objective, and through it the certificate η below; P_{t_h}, D and $\ker D$ stay the true objects, and every quantity in Table 25 is scored against the true decision problem. The cell is d=16, n_o=4, k=2, a linear encoder over a 32-step window, trained as the capacity family of Appendix B with one cost block, at block 1, four arms $\times$ 20 paired seeds, 80 runs and no failures. Its cost structure puts $\operatorname{rank} P_{t_h} = 2 = k$, so the ideal reweighted encoder retains exactly $\operatorname{range} P_{t_h}$ and the containment surrogate has a hard zero to hit. That is the sharpest containment geometry available and it is not the cost structure of the d=16, n_o=4 setting of Table 23, where $\operatorname{rank} P_{t_h} = 4$.

The projection is not cosmetic, and it is what bounds the claim. About seven of the 16 eigenvalues go negative at every scale and are clipped, the weight's rank rises from 2 to about 9, and its trace rises by 3.6, 12.1 and 36.9 percent. The realised error is therefore smaller than the nominal one: its norm is 21.3, 21.7 and 22.9 percent below the drawn perturbation's, and that share is flat in the scale rather than growing with it. (A second diagnostic, the correction's norm relative to $\|P_{t_h}\|_F$, reads 0.0183, 0.0613 and 0.186. It is on the same normalisation as s, not on the perturbation's, and

Table 25: Sensitivity of the retained subspace to a perturbed decision weight, at d=16, n_o=4, k=2, $\operatorname{rank} P_{t_h} = 2$, 20 paired seeds per arm. Means over seeds. The first row is the true weight. *Realised* is $\|P_{\mathrm{psd}} - P_{t_h}\|_F / \|P_{t_h}\|_F$ and is the size of the error actually applied; the nominal s is a label, and realised error is the axis any plot or ordering should use. The five middle columns are the projection diagnostics that make the kernel-side one-sidedness checkable: how many of the 16 eigenvalues the projection clipped to zero, the resulting rank against the true weight's 2, the trace it added, and how far the weight's top-2 eigenspace rotated. The gap at the capacity boundary stays strictly positive throughout, against the true weight's 0.507, so the capacity-k optimum stays determinate and what moves is which optimum the objective points at. *Retained* is decision-irrelevant energy against the true $\ker D$ (↓), *error* is decision-relevant error (↓), and *perf. readout* is the decision-relevant error of each run's own retained subspace with a perfect readout, scored under the true P_{t_h}; the *perturbed* objective's own capacity-k optimum scores 0.635, 0.635, 0.635 and 0.639. The paired tests and effect sizes (d_z) of *retained* and *error* against the true weight are given in the text of this appendix.

		projection diagnostics					scored against the true P_{t_h}		
s	realised	neg. eigs	rank	tr. ratio	$\angle_{\mathrm{top2}}$	gap at k	retained	error	perf. readout
0	0	0	2	1.000	0.00°	0.507	0.0402	0.669	0.640
0.03	0.0236	7.05	8.95	1.036	1.96°	0.476	0.101	0.675	0.648
0.10	0.0783	7.05	8.95	1.121	6.61°	0.408	0.172	0.682	0.656
0.30	0.231	7.10	8.90	1.369	20.96°	0.260	0.244	0.690	0.666

it is not that shrinkage.) **Inside** $\ker P_{t_h}$ **the tested error is one-sided.** There the realised weight is positive semidefinite, so on decision-irrelevant directions the perturbed weight can only *over-claim*. On the decision-relevant directions the draw keeps both signs, and on 17, 14 and 11 of 20 seeds at the three scales it lowers one of the two relevant eigenvalues. A kernel error of negative sign is not probed and cannot be, since a matrix with negative curvature is not a decision Gramian. An estimator's error would carry both signs.

Degradation is significant at the two larger errors, and the response is concave. Friedman $\chi^2 = 45.66$, $p < 10^{-6}$ over the four arms. Against the true weight's 0.0402 the retained decision-irrelevant energy is 0.101, 0.172 and 0.244, with d_z of -0.41, -0.87 and -1.34, win rates $2/20$, $1/20$ and $1/20$, corrected Wilcoxon p of 7.1×10^{-4}, 4.2×10^{-4} and 8.0×10^{-5}, and BCa intervals excluding zero at all three. The t-test agrees at the two larger scales (corrected p 0.0019 and 2.8×10^{-5}) and not at the smallest (0.080), where Shapiro–Wilk rejects normality ($p = 2.5 \times 10^{-4}$) and the Wilcoxon is the applicable test; the disagreement traces to one seed, and under the primary rule of Appendix C.1 the smallest error's effect is borderline. Containment loss per unit of realised error is 2.56, 1.68 and 0.88 as the error grows: the response is concave, the first increment of error is the most expensive per unit, and these runs show no threshold within the tested range (realised error 0.0236 to 0.231) below which the weight can be wrong at no cost, though at the smallest error that rests on the borderline result. Seed 0 is an outlier of the *true-weight* arm, which reaches 0.006 to 0.032 on 19 of 20 seeds and 0.496 there; it is the only seed on which any perturbed arm beats the baseline, and dropping it takes the three deltas to -0.0812, -0.147 and -0.221, all larger. Keeping it in understates the degradation. What makes that one seed fail under the true weight is not diagnosed.

The error half is mostly the trained subspace, not the moving optimum. Trained decision-relevant error degrades significantly at all three scales (d_z -0.49, -0.91, -1.45; corrected t-test p 0.040, 0.00128, 9.4×10^{-6}; 19 of 20 seeds at each), by 0.00617, 0.0128 and 0.0202. The capacity-k optimum of the perturbed objective, scored under the true P_{t_h}, moves by less than 10^{-4}, 0.00044 and 0.00474, so a perfect optimiser would have lost at most a quarter of the trained degradation. The rest is the trained encoder's subspace moving further from that optimum, partly offset by a readout gap, trained error less the *perf. readout* column of Table 25, that narrows from 0.0292 to 0.0274, 0.0256 and 0.0240, on 15, 17 and 17 of 20 seeds. The absolute sizes belong beside the effect sizes. At the largest error the containment surrogate rises sixfold while decision-relevant error rises 2.0 percentage points on a base of 66.9 percent, a 3.0 percent relative rise, because the capacity-2 constraint against d=16 already fixes most of the error above the floor and the perturbation only redistributes the remainder. The mechanism is visible in the belief split: R^2 on the decision-relevant belief component falls 0.846 to 0.803, 0.757 and 0.713 while R^2 on the belief overall *rises* 0.234 to

0.259 and predictive NMSE barely moves, so the code reallocates a fixed two dimensions towards directions with more generic belief variance and less decision content rather than fitting worse.

How much of the benefit survives a wrong weight. A badly wrong weight is still much better than no weight. At the largest error tested the retained decision-irrelevant energy is 0.244, against 0.707 for the unweighted k=2 baseline at d=16, n_o=4 in Table 23: about three times better, and the comparison is across cost structures ($\operatorname{rank} P_{t_h} = 4$ there against 2 here) rather than within one arm of one sweep, so we read it as an order of magnitude and not as a contrast.

What this does and does not establish. Nothing here estimates P_{t_h} from data. The perturbation size is a configuration key and its direction a seeded symmetric draw, not an estimator's error distribution, whose covariance would be set by the sample size, the identification geometry and the estimator's own bias. These runs therefore support statements of the form "a symmetric, positive-semidefinite-admissible weight error of realised relative size r moves containment by x" and support no statement that a fitted weight recovers containment or that estimation error is harmless at any sample size. Nor is this a test of Theorems 1 or 2, which are exact-optimum statements about the *true* weight: removing that weight removes the object they are stated for, so these runs neither support nor contradict them. What they bear on is clause 4 of Theorem 2, the approximate-transport clause, which is the one clause about an operator that is merely close to the decision operator and the one clause with no other experimental illustration in this paper. Its quantity, the decision energy $\|Du\|^2$ of the retained directions, is logged: on the seed means the trained encoder stays below the exact-optimum bound $\nu_{k-1} - \varepsilon$ at every scale, including the true weight, as expected of a finitely trained run, while single seeds of the perturbed arms reach it, on 4, 8 and 9 of 20. That is a measurement beside the bound, not a test of it. One premise that could have broken did not: the gap at the capacity boundary stays positive at every scale (Table 25), so Theorem 1's determinacy hypothesis survives and the degradation is a change in which optimum the objective points at. The leakage certificate is vacuous throughout, $\eta = 2.36$ at the true weight and 2.23, 2.49 and 3.73 across the scales (each computed on its arm's perturbed objective) against a quantity that is a fraction, so it neither certifies nor denies robustness to weight error here. Limits: one cell and one capacity; an error of one sign only inside $\ker P_{t_h}$; n=20 with seed 0's true-weight failure undiagnosed; a linear encoder, no planning and analytic Kalman beliefs for scoring.

H.15 A DECISION WEIGHT ESTIMATED FROM FINITE TRAJECTORIES

Every other weight in this paper is built from generator matrices handed over exactly. Here the plant maps A and C are fitted instead, by ordinary least squares on a separate open-loop calibration set in which the state is measured: $\hat{A}$ from pairs (x_t, x_{t+1}) and $\hat{C}$ from pairs (x_t, o_t), 239 transitions per trajectory, drawn independently of the encoder's training data. The cost (H, ρ), the controller (B, K_c) and the horizons stay granted. The weight $\hat{V} = \hat{D}\hat{T}^{+}$ is built from the fitted maps: $\hat{T}$ from $\hat{A}$ and $\hat{C}$, and $\hat{D}$ from $\hat{A}$ with the granted H, ρ, B and K_c. Every number is scored against the true decision problem. The generator is that of Appendix H.18 with its curvature dial at 0, where it is linear-Gaussian: d=16, n_o=4, k=2, σ_{obs}=0.3, a linear encoder over a 160-step window, 20 paired seeds per arm, with process noise 0.25 on every coordinate, trajectories of 400 steps after a burn-in of 160, Δt=0.05, damping 2.0, squared block frequencies $18 - 2b$ for block b, and extended-Kalman beliefs. Its unweighted and exact-weight arms reproduce that appendix's curvature-0 arms on every seed.

The pseudo-inverse needs a rank, and three rules are compared: the true rank of T, which is 8, granted; a parameter-free gap rule that cuts $\hat{T}$'s singular spectrum at its largest relative gap; and a plain pseudo-inverse at the default numerical tolerance, which keeps every spurious singular value. Table 26 gives the fraction of the exact weight's improvement over the unweighted arm that each estimated weight keeps. Against the exact weight, intervals are paired t-intervals and Holm runs over the 16 estimated arms.

Table 26: Fraction of the exact weight's improvement over the unweighted arm kept by an estimated weight, as a ratio of mean decision-relevant errors, 20 paired seeds per cell. The granted rule was not run at 128 trajectories: the gap rule picks the true rank on all 20 seeds there, so it is the same weight. The granted rule was not run at 16 or 32 trajectories either, for the same reason, and the plain pseudo-inverse was run at three sizes only.

Trajectories	1	2	4	8	16	32	64	128
Rank granted	0.976	0.991	0.996	0.998			1.000	
Rank by the gap rule	0.500	0.748	0.834	0.971	0.999	1.000	0.881	1.000
Plain pseudo-inverse	−0.229			−0.227				−0.230

With the rank granted, one trajectory keeps 97.6% of the improvement. Every granted size is still detectably worse than the exact weight, at one trajectory by $+0.0151$ (95% CI $[+0.0108, +0.0193]$, worse on 20 of 20 seeds), and almost all of that loss is the fitted problem's own optimum moving under the true cost (97% of it at one trajectory, 95% and 97% at two and four, and 99% or more from eight), not optimisation. Under the gap rule, 128 trajectories is the one size that neither test separates from the exact weight (paired difference $+2.7\times10^{-5}$, CI $[-1.7\times10^{-5}, +7.1\times10^{-5}]$, Holm-corrected $p_t = 0.45$, Wilcoxon 0.14). At 8 and 64 trajectories, where the rule over-picks the rank on two seeds each, the t-test does not separate it either (Holm $p_t = 0.45$), but the Wilcoxon test does (Holm $p_w = 3.1\times10^{-5}$ and 0.00814); at 16 and 32 trajectories the loss is detectable and under 0.1% of the improvement. The curve is not monotone because the rule over-picks the rank on some seeds, two of them at 64 trajectories, and every gap-rule deficit beyond the granted arm's comes from seeds where the rule picks the wrong rank, all over-picks except one under-pick at two trajectories: each spurious singular direction retained puts an amplified copy of observation noise into the target. The plain pseudo-inverse retains all of them, its target carries about 10^5 times the exact weight's irreducible noise, and training on it gives higher decision-relevant error than the unweighted arm at every size tested, by 0.144 to 0.146 (Holm-corrected $p_t = 0.0023$, worse on 15 or 16 of 20 seeds). Its population optimum is 0.005 above the true ceiling at 128 trajectories, so the loss is in training rather than in the direction of the weight; the gradient clip bound on 57 of its 60 seeds, a confound this sweep does not separate. Only the plant maps are estimated: an estimated cost or controller is not tested.

H.16 A Nonlinear Transition Law, Where No Closed-Form Belief Exists

This relaxes the synthetic setting on one axis, and only weakly: it moves the transition law off the linear-Gaussian family, and the metric it carries there turns out to have almost no headroom on the plant it moves to.

The generator is a damped pendulum under explicit Euler, $\theta_{t+1} = \theta_t + \Delta t\,\omega_t + \epsilon_t^\theta$ and $\omega_{t+1} = \omega_t + \Delta t\big(-(g/L)\sin\theta_t - b\,\omega_t\big) + \epsilon_t^\omega$, with $(\epsilon_t^\theta, \epsilon_t^\omega) \sim \mathcal{N}(0, 0.25^2 I_2)$ and the angle alone observed: d=2, n_o=1, Δt=0.05, g/L=9.81, b=2.0, observation noise 0.3. Everything else follows the belief-identification family of §4: 16-step causal windows, a width-128 GRU encoder, horizons $\{1, 3, 5\}$, 100 epochs, 20 paired seeds, five arms with the naive hybrid as baseline. Two of the five, the contrastive arm and an identify-then-decode variant, reduce to the same loss here and coincide on every seed. A nonlinearity check, fixed in the configuration before any run, passed on all 100 runs, with final-step $\mathbb{E}|\sin\theta - \theta|/\mathbb{E}|\theta|$ between 0.173 and 0.452 against a floor of 0.05 and the final-step share of trajectories with $|\theta| > \pi$ under its cap of 0.10, so the nonlinearity was exercised rather than assumed.

What replaces the ground truth. The nonlinear transition admits no closed-form posterior, so the target is a bootstrap particle filter at N=4000 in double precision. Its residual against an N=16000 reference is $\mathrm{nrmse}\ 0.0186$, against a Monte-Carlo floor of 0.0125 between independent runs. That error is common to all arms at a given seed, so it enters every paired contrast through the same target; it was measured on one benchmark set of 150 trajectories.

Result. Reconstruction leads on every seed, so the hypothesis, that a nonlinear transition lets the contrastive objective overtake reconstruction, is not borne out on this axis.

contrast	Δ	wins	d_z	p_w
Autoencoder − contrastive	$+0.496$	$20/20$	$+1.188$	1.9×10^{-6} (post hoc)
contrastive − baseline	-0.419	$2/20$	-0.944	1.07×10^{-4} (Holm)
Autoencoder − baseline	$+0.0769$	$11/20$	$+0.358$	0.0969 (Holm)

p_w is the Wilcoxon signed-rank test, Holm-corrected over the four candidates where marked, used because Shapiro–Wilk rejects normality of the deltas on every arm; the Holm-corrected t-test gives 1.85×10^{-3} and 0.126 for the two baseline rows. The autoencoder is a deterministic recurrent autoencoder trained to reconstruct the current observation.

The significant row is the contrastive arms losing ground, an effect opposite in sign to that hypothesis; reconstruction's own margin over the baseline is not significant and is carried by one collapsed baseline seed.

What limits the reading. The metric sits at 0.966 (baseline) and 0.954 (recurrent autoencoder) after a single epoch, and three of the four distinct arms end below where they started. With d=2 and one observed coordinate, the angle is read directly and the angular velocity is close to a finite difference of it across the window, so what was measured is which objective damages an already-adequate representation least. Two further gaps: this metric regresses the latent on the belief rather than the belief on the latent, and it is not restricted to the not-instantaneously-observable component, so it is not the $R^2_\perp$ of Table 20 and the levels are not comparable across the two. Relative to Theorem-style guarantees the reading is unchanged: the statement this plant was built for is a local, exact-optimum, infinite-data one, a finite-training run does not satisfy its antecedent, and the target here approximates the posterior it speaks about rather than being it.

H.17 The result on randomly drawn transition matrices

This tests whether the empirical section depends on one hand-built system, but only along one axis of it: the generator's spectrum, not its structure, and at one cell. Appendix B fixes A block diagonal with $\theta_b = 0.3 + 0.15b$ and $r_b = \mathrm{clip}(0.9 - 0.03b,\ 0.55,\ 0.95)$, two authored constants. We replace them with a pre-registered draw, its law and its twenty seeds fixed in the code before any run: $\theta_b \sim U[0.15,\ \pi - 0.15)$ and $r_b \sim U[0.55,\ 0.95)$, the endpoints of the clip the authored formula already applies. We run twenty independent draws at d=16, n_o=4: unweighted and reweighted arms at $k \in \{2, 4, 16\}$ on each, 20 paired seeds per arm, plus six arms on the authored A (2520 runs in all). These six rebuild the body's configuration; across 17 metrics, 6 arms and 20 seeds they equal the reported per-seed record of §5 on all 2040 values with no tolerance, the 40 exceptions being the leakage bound at k=d, undefined in both.

The quantity is the within-draw contrast, reweighted minus unweighted on the same draw and the same seeds, oriented so that positive means less decision-irrelevant energy retained.

Table 27: Within-draw reweighting contrast on decision-irrelevant energy, twenty drawn A. Holm–Bonferroni corrects the paired t-test over the twenty draws at each capacity; d_z carries the reading of Appendix C.1.

k	$\Delta > 0$	**Holm sig.**	**Δ min / med. / max**	**Authored A**
2	20/20	20/20	$+0.103$ / $+0.352$ / $+0.859$	$+0.703$
4	20/20	20/20	$+0.159$ / $+0.322$ / $+0.714$	$+0.532$
16	13/20	0/20	-0.000 / $+0.000$ / $+0.000$	-0.000

No draw reverses the effect at either binding capacity. Per-draw d_z runs 0.84 to 8.66 at k=2 and 1.29 to 12.30 at k=4; the authored A gives 5.64 and 5.51, and the largest Holm-corrected p over the forty contrasts is 1.9×10^{-3}. At k=d=16 nothing is discarded, every arm sits at 0.500 on every seed, and that row is the design's control rather than a third measurement. The effect's direction does not depend on the authored ladder; its size does, the median draw giving about half the authored A's delta at k=2.

That size is predictable from the generator before any training. Spearman ρ between the trained within-draw delta and the share of belief variance that is decision-relevant is -0.958 at k=2 and

Table 28: Decision-relevant error at d=16, n_o=4, k=2, 20 paired seeds per cell, means over seeds. *Curv.* is the curvature dial; 0 is the plant's own linearisation. p is a paired t-test within the cell, uncorrected, on decision-relevant error (↓). R normalises against the capacity-2 optimum of the run's *own* data, $R = (1 - \text{error})/(1 - \text{ceiling})$, the ceiling being the error at the exact optimum of the reweighted arm's loss on that data, so 1.0 is that optimum. * marks the one cell where the surrogate's optimal subspace has itself moved. d_z is of reweighted minus unweighted and is not oriented: negative is better.

σ_{obs}	curv.	unweighted	reweighted	d_z	p	R
0.3	0	0.769	0.136	−4.32	6.0×10^{-14}	0.988
0.3	1	0.595	0.091	−2.03	2.4×10^{-8}	0.976
1.0	0	0.833	0.409	−3.80	6.0×10^{-13}	0.936
1.0	1	0.676	0.298	−1.64	5.8×10^{-7}	0.902
2.0	0	0.884	0.703	−2.74	1.8×10^{-10}	0.739
2.0	1	0.776	0.558	−1.31	1.3×10^{-5}	0.753
3.0	0	0.940	0.880	−1.82	1.3×10^{-7}	0.455
3.0	1	0.868	0.746	−0.96	4.0×10^{-4}	0.776*

-0.895 at k=4, on twenty points. On the seven draws whose closed-form gap in the ceiling is below 10^{-15} at k=2, where reweighting cannot lower the error at the exact optimum, the trained delta in decision-irrelevant energy is still positive and Holm-significant on all seven ($+0.103$ to $+0.317$); what those draws measure is the optimisation path.

Scope: one cell, one draw law, block diagonal and stable throughout (spectral radius 0.821 to 0.946), with realizability holding and Theorem 4's obstruction dimension 0 on all twenty. A coupled or unstable A, another cell and a different draw law are untested. These are finite-training runs; none tests Theorem 2 or Theorem 5, which are statements about a population risk (Theorem 2 at an exact optimum, Theorem 5 at a δ-suboptimal one).

H.18 A NONLINEAR TRANSITION LAW, READ AGAINST THE MEASURED LINEARISATION GAP

This relaxes the idealised setting on one axis and leaves the others where they were. Every generator the body draws a number from is linear-Gaussian, so T, D, P_{t_h} and the capacity-k optimum are exact objects. In this one the transition law is an explicit-Euler damped pendulum on eight blocks at d=16, n_o=4, k=2, t_h=3, and those are computed from its linearisation, so each is a surrogate for a system the data is not drawn from. The observation map is still a coordinate selection and the cost still the quadratic form of §3: §6's boundary moves by one premise, not by all of them.

A curvature dial set to 0 or 1 gives a paired control: the weight $V = DT^{+}$, the generator matrices, the window, the seeds and the optimiser are identical across the pair. σ_{obs} is the gap axis, because posterior width rather than curvature decides whether a linearisation is accurate. Beliefs come from an extended Kalman filter on the sixteen arms of Table 28; the other three rerun the linear-plant reference (see Scope). Nineteen arms × 20 paired seeds, all completed.

The plant and its filter. The state is eight blocks (θ_b, ω_b), $b = 0, \ldots, 7$, each an explicit-Euler damped pendulum with curvature dial $c \in \{0, 1\}$:

$$\theta_{b,t+1} = \theta_{b,t} + \Delta t\, \omega_{b,t} + \epsilon^{\theta}_{b,t}, \qquad \omega_{b,t+1} = \omega_{b,t} + \Delta t\big(-w_b^2[\theta_{b,t} + c(\sin\theta_{b,t} - \theta_{b,t})] - 2.0\,\omega_{b,t}\big) + \epsilon^{\omega}_{b,t},$$

with Δt=0.05, $w_b^2 = 18 - 2b$ and all sixteen process-noise terms independent $\mathcal{N}(0, 0.25^2)$. At c=0 this is the linearisation at θ=0, from which T, D and P_{t_h} are computed; at c=1 it is the pendulum. The angles of blocks 0 to 3 are observed with noise $\mathcal{N}(0, \sigma_{\text{obs}}^2 I_4)$, the cost sits on blocks 2 and 3, and no control is applied to the simulated process. The extended Kalman filter linearises the transition at the current filtered mean each step and carries its covariance per trajectory, starting from mean zero and the linearisation's stationary predictor covariance, from which x_0 is drawn; at c=0 it is the Kalman filter. Each seed draws 100 training and 50 evaluation trajectories of 400 steps, encoded in 160-step windows with the first 160 anchors dropped.

The effect survives, and degrades with the gap. Reweighting lowers decision-relevant error in all eight cells, and decision-irrelevant energy falls from 0.57–0.71 to 0.0001–0.022. $R = (1 -$

error)/(1 − ceiling), normalised against the capacity-2 optimum of the run's own data (Table 28), falls monotonically along σ_{obs} with curvature off, and with it on over the three cells the invalidation check passes. Curvature's own paired cost is -0.012 and -0.034 of R at the two tightest levels ($p = 9.1 \times 10^{-5}$, 6.9×10^{-3}) and a null at 2.0 ($p = 0.62$).

Where the surrogate stops describing the data. Its optimal subspace agrees with the best linear predictor's to 1.000, 0.999 and 0.994 containment at the first three noise levels, and to 0.961 at σ_{obs}=3.0 with curvature on, against that cell's control of 0.991. That cell carries the containment number and no claim, and holds the only non-monotone step in R. The ceiling computed from the linearisation is wrong in *value* wherever curvature is on, by -0.11 to -0.46 relative and always understating a nonlinear arm, so R uses the ceiling computed from the run's own data throughout.

The belief carries the penalty, not the operator. Per seed on the reweighted arms, degradation is not predicted by how far the analytic T sits from the best linear predictor: neither its relative deviation (1.4–3.3× larger with curvature on; pooled Spearman $+0.03$, $p = 0.78$) nor the prediction error it adds over that predictor (6–10× larger; -0.09, $p = 0.41$) correlates with it, but the filter's error entering through the retained subspace does (pooled Spearman -0.357, $p = 1.1 \times 10^{-3}$).

Scope. No run here tests Theorem 1, 2 or 5: each assumes a linear predictive operator (Theorems 1 and 2 at an exact optimum, Theorem 5 at a δ-suboptimal one), and the curvature-on arms violate that antecedent deliberately. Their construction still moves a trained encoder's retained subspace over a measured range of operator mismatch. One plant, k=2, 20 seeds, capacity not swept. Two arms rerun the linear-plant reference's unweighted and reweighted configurations inside this experiment, with its Kalman filter; their 20-seed means equal the reported values at every stored digit, and a third, the unweighted configuration with the extended Kalman filter, equals them too.

H.19 The pair form of the approximate-retention certificate, instantiated

Every run below is drawn from the same linear-Gaussian family as §5, at two cells, and adds no new system. What it does instead is report, at the two cells that log it, a quantity §6 sets aside, and then report that the certificate assembled from it is quantitatively weak.

Clause 5 of Theorem 5 bounds $\|Xv\|^2 \le C_X^2\,\eta\,\|v\|^2$ for every v with $Dv = 0$, on the trained composite $X = M \circ W_{\text{enc}}$, which we measure as $X = M\hat{W}$, with $\hat{W}$ the least-squares map from belief to code on the evaluation set (for the GRU arm, a linearisation), and with norms taken in whitened coordinates. Assumption (A6) supplies C_X as a hypothesis and no clause bounds it, so it can only be measured. Of the runs of §5, only the d=32, n_o=8 cell computed it after training; the others did not save the trained weights, so we reran the d=8, n_o=3 cell with nothing else changed, 16 arms at 20 paired seeds for 320 runs, and computed C_X after training; the rerun reproduces every per-epoch value of the reported cell except timing and device fields. C_X, η and the leakage $L = \sup\{\|Xv\|^2 : \|v\| = 1,\ Dv = 0\}$ are finite on 320/320 runs, so clause 5 has a numerical right-hand side and the quantity it bounds beside it.

The bound is never violated and never close. On the 280 runs of the reweighted arms, whose applied operator has a nontrivial whitened kernel, $C_X^2\eta/L$ has median 778, minimum 5.69 and maximum 2.33×10^5. The other 40 runs, on the two unweighted arms, regress onto T, whose whitened Gram has no kernel; L is then set to 0, and these arms carry no measurement of leakage into $\ker D$. That ratio factors exactly as $(C_X^2/L) \cdot \eta$, and at d=8 the two factors are not comparable in size.

The d=32, n_o=8 cell logged the same three quantities in its original runs, 13 arms at 20 paired seeds, all with the linear window encoder, and all three are finite on 260/260 runs. On its 240 runs with a nontrivial whitened kernel the bound again holds on every run and is less loose: $C_X^2\eta/L$ has median 84.5, minimum 6.89 and maximum 7577. The other 20 runs, its unweighted baseline, regress onto T, whose whitened Gram has no kernel; L is set to 0 there and measures no leakage into $\ker D$.

Table 29: Where the looseness sits, over the runs of the reweighted arms, whose whitened kernel is nontrivial: 280 at d=8, n_o=3 and 240 at d=32, n_o=8.

cell	factor	med.	min	max
d=8	$C_X^2\eta/L$	778	5.69	2.33×10^5
	C_X^2/L	432	27.7	7909
	η	1.94	0.057	109
d=32	$C_X^2\eta/L$	84.5	6.89	7577
	C_X^2/L	18.3	6.42	87.1
	η	4.27	0.485	151

The dominant term is not the theorem's own $\delta/(\mu_{k-1} - \mu_k)$ slack. It is that (A6) asks for a bound on X over all of $\mathcal{E}$ while the conclusion is read only on $\ker D$, where at d=8 these trained encoders are about twenty times smaller in norm; clause 5 has no way to say so. At d=32 that factor is much smaller, a median 18.3 or about four times in norm, while η is larger, a median 4.27. That mismatch is structural, so a reader certifying a trained encoder this way should expect it, but its size here is a property of these encoders and this generator.

Scope: two cells. At d=8 and n_o=3, sixteen arms at $k \in \{1, 2, 3, 4\}$ and $t_h \in \{0, 1, 3\}$, 20 seeds, a linear window encoder on fifteen of them and a GRU on the sixteenth; at d=32 and n_o=8, thirteen arms over the same k and t_h, 20 seeds, all linear. The bound holds in both, and its median looseness falls from 778 to 84.5 between them, so its size is a property of the cell. The d=16 and full-observation cells were not re-run with C_X logged, so two values of d are all this measures, not a trend in d. No run here tests Theorem 5, whose δ is given rather than derived; consistency on $320/320$ and $260/260$ runs checks the instantiation, and a measurement that a conclusion is weak is not evidence against the theorem that proves it.

H.20 Results in Full

Reading the numbers. Decision-relevant error (↓) is the P_{t_h}-weighted error of the affinely aligned prediction against the true state, over the P_{t_h}-weighted cost-to-go of that state, not a fraction above optimal closed-loop cost. The metric has a floor above zero: the affine fit regresses onto the belief mean by ordinary least squares, so the filter's own stationary error $\mathbb{E}[(x - \hat{x})^\top P_{t_h}(x - \hat{x})]$ stands in the numerator whatever the encoder does. Kalman orthogonality makes the split into that floor and the encoder's own term $\mathbb{E}[(\hat{x} - \tilde{x})^\top P_{t_h}(\hat{x} - \tilde{x})]$ exact in expectation, and we log the two parts separately together with the finite-sample cross term between them. In closed form from the generator, that floor is 0.711 at d=16, n_o=4, 0.660 at d=8, n_o=3 and 0.499 under full observation, against baselines of 0.968, 0.901 and 0.913: an encoder can move 0.257, 0.241 and 0.414. Arms that change P_{t_h} move the floor, so we read each against its own. We also report affine R^2 on unobserved factors, computed after optimal affine alignment to the unobserved part of the Kalman belief mean (↑, hence not observable on real data). That alignment is part of the metric itself. The retained-subspace quantities of Tables 23 and 24, retained decision-irrelevant energy and the minimum principal angle, are not affine-aligned. Figure 9 is a separate, affine-aligned view of the same capacity family (d=16, n_o=4, k=2), drawn from saved encoder weights: it plots the affine fit of each true decision direction from the code, per direction. Latents are identified only up to an affine transform, so we define no permutation-level metric. We use n=20 seeds per arm, reported as mean ± standard deviation; win rates are proportions over seeds. We test oriented deltas; we form two-sided BCa intervals from 10,000 bootstrap resamples. Every improvement claim rests on p_{corr}, the Holm–Bonferroni-corrected two-sided paired t-test value, read together with the sign of Δ, which is what distinguishes a significant gain from a significant loss; Wilcoxon signed-rank is confirmatory. Interval and test share the two-sided convention, so an interval excluding zero on the harmful side is accompanied by a small p_{corr} and a negative Δ.

Reweighting lowers decision-relevant error. At the shared cost horizon, the reweighted k=4 encoder reduces decision-relevant error by 0.205 at d=16, n_o=4 and by 0.211 at d=8, n_o=3 (Table 21, Figure 5); both survive Holm–Bonferroni over K=15 comparisons (p_{corr}<0.001). The reweighted k=2 encoder reduces the error by 0.123 at both settings. We measure every Δ in Table 21 against the shared unreweighted baseline, so the reweighted k=4 Δs move weighting

and capacity together; the capacity-matched unweighted $k{=}4$ arm reduces the error by 0.0669 at $d{=}16, n_o{=}4$ and by 0.139 at $d{=}8, n_o{=}3$, so we also form a per-seed weighting-matched contrast between reweighted and unweighted arms at matched capacity, encoder and horizon. At $k{=}4$ it is 0.138 ($[0.127, 0.150]$, $d_z{=}5.22$, win rate 1.00) at $d{=}16, n_o{=}4$ and 0.072 ($[0.062, 0.084]$, $d_z{=}2.76$, win rate 1.00) at $d{=}8, n_o{=}3$, both with $p_{\mathrm{corr}}{<}0.001$. The 20 matched contrasts carry their own Holm–Bonferroni family, so no shared-baseline Δ changes. The family was fixed at 14 before any of its p-values was computed and grew to 20 when later sweeps supplied the missing partners: three step-size-matched pairs at $k{=}2$ (§4) and three at cost structures whose unweighted partner had not been run. Correcting over the enlarged family moves no value across a reporting threshold; 19 of the 20 survive at $\alpha{=}0.05$. Three sit in both families, because each setting's shared baseline is the unweighted $k{=}2$ arm: the reweighted $k{=}2$ Δ is already weighting-matched, carrying $[0.105, 0.136]$ and $d_z{=}3.44$ at $d{=}16, n_o{=}4$; for these we report the larger corrected value. A matched pair exists at every capacity we ran and for the recurrent encoder, where reweighting lowers the error by 0.108 ($[0.092, 0.122]$, $d_z{=}2.96$) at $d{=}16, n_o{=}4$. The capacity constraint is not what dominates this ratio. The two full-capacity arms, which remove the capacity constraint, sit at 0.751 and 0.750, 0.040 and 0.039 above the floor. The exact optimum of each arm's own objective at its capacity, also from the generator, separates the rest. At $k{=}4$, $d{=}16, n_o{=}4$ the reweighted arm's exact optimum is the floor itself, so its whole shortfall of 0.053 is optimisation gap, which here also contains the error the encoder's finite window adds: the exact optima are computed on the Kalman belief, which uses every past observation. The unweighted $k{=}4$ arm's optimum stands 0.173 above the floor with a shortfall of 0.018. The matched 0.138 is that 0.173 between the two exact optima less the 0.035 the reweighted arm gives back in optimisation gap. At $d{=}8, n_o{=}3$ the gap between the optima is 0.043 and the reweighted arm recovers a further 0.029 of optimisation gap rather than giving any back, for 0.072; under full observation the two are 0.197 and 0.032, for 0.229.

What the reweighted code keeps. On the retained subspace at $d{=}16, n_o{=}4$, the reweighted arm lowers decision-irrelevant energy from the unweighted $k{=}2$ baseline's 0.707 to 0.00460 at $k{=}2$ (drawn per seed in Figure 7) ($\Delta 0.703$, $[0.650, 0.754]$, $d_z{=}5.64$, win rate 1.00) and to 0.038 at $k{=}4$ ($\Delta 0.669$, $[0.604, 0.730]$, $d_z{=}4.41$, win rate 1.00), and the minimum principal angle from $42.6°$ to $2.2°$ ($d_z{=}2.73$) and $0.6°$ ($d_z{=}2.93$). We give both quantities for all three settings in Tables 23 and 24; the $k{=}4$ energy mean is carried by a single high seed in each (0.250, 0.125, 0.107 against medians near 0.005), and at $d{=}8, n_o{=}3$ the angle effect sizes fall to 0.67 and 0.75. Clause 3 of Theorem 2 concludes a containment, $\ker D \subseteq \ker X$; we take decision-irrelevant energy as its finite-sample surrogate, since it vanishes exactly when the containment holds, and the minimum principal angle as a diagnostic, small being necessary there and not sufficient. The two do not order the capacities the same way: on the surrogate $k{=}2$ is the better arm at every setting, on the angle $k{=}4$ is the better aligned, so the angle does not corroborate the surrogate here.

A short cost horizon changes the decision problem. The short-horizon arm sets $t_h{=}0$, altering P_{t_h}: at $d{=}16, n_o{=}4$ this raises $\dim\ker P_{t_h}$ from 12 to 14. Concentrating cost on fewer directions gives the improvement and the confound the same cause. We therefore report its nominal 0.270 and 0.296 without cross-arm normalisation (Table 21, marked †), not comparable with the numbers above. The two causes separate once the unweighted arm at $t_h{=}0$ exists: at $d{=}16, n_o{=}4$ it lowers the error by 0.0073 against the shared baseline, so shortening the horizon achieves almost nothing on its own and the weighting-matched contrast there is 0.262 ($[0.237, 0.273]$, $d_z{=}7.22$, $p_{\mathrm{corr}}{<}0.001$). We ran that partner in that setting only, so the $d{=}8, n_o{=}3$ row keeps the confound.

The certificate holds on the trained runs but stays loose. The leakage certificate η exceeds the frame weight it bounds by a median factor of 21 to $33\times$ across the three main settings; the fourth, $d{=}32$, $n_o{=}8$, is less loose ($15.6\times$). This is not estimator noise: we find the fitted and analytically derived retained subspaces agree to a median 0.55–$2.9°$. Logging, beside δ, the frame slack $\delta_{\mathrm{fr}} = \sum_{i<k}\mu_i - \sum_i \mu_i w_i$ of the fitted retained frame, with w_i the frame's weight on the i-th eigendirection, on 920 runs of the three main settings (40 of them on the GRU arm, which (A2) excludes), the ratio $\delta/\delta_{\mathrm{fr}}$, which is what step 7 of the proof of Theorem 5 (a δ-suboptimal pair supplies a δ-suboptimal frame) gives away, has median $15.4\times$ (12.1 to $18.1\times$ by setting); the remainder, median $1.4\times$, is the division by the gap. Two arithmetic corrections reduce η: normalising by k, and replacing the spectral gap $\mu_{k-1} - \mu_k$ by the boundary eigenvalue μ_{k-1}, a gap-free form of the classical perturbation bound. The trained-encoder logs are a population of 1180 runs (59 arms $\times$ 20

seeds). The quantity we hold each ratio against is the one Theorem 5 clause 2 bounds: $\sum_{i \geq k} w_i$, the weight the retained frame places on the discarded block of the applied operator's covariance-weighted spectrum, computed in the whitened coordinates that spectrum lives in. Against it, η holds on 1180 of 1180 runs. Across the 1180 the below-1 count runs from 196 (17%) for η to 779 (66%) for $\delta/(k\mu_{k-1})$, the form that applies both corrections. That is not an independent measurement, and it is not a second certificate. Both corrections replace η's divisor with one no smaller than the gap, so every value can only shrink and the count can only rise; and since no theorem licenses either divisor (§3), reducing the right-hand side leaves the left-hand side where it was. Measured against the same frame weight, the reduced forms are not bounds: $\delta/\big(k(\mu_{k-1} - \mu_k)\big)$ falls below it on 16 of the 1180 runs, by factors down to 0.61, and $\delta/(k\mu_{k-1})$ on 179, down to 0.21. We therefore report 66% as a diagnostic rate and keep the word certificate for η. On the 140 runs where the index set $\{i : \mu_i = 0\}$ that the gap-free form addresses is empty, that form has nothing to say. None of this tests Theorem 5, and not because training is finite: assumption (A7) takes δ as given rather than derived, so what our runs do not supply is the population δ, whose risk term we estimate on a held-out sample, and the linear encoder of (A2), which the recurrent arms are not. A consistency count over runs checks the instantiation.

Sequence encoding and belief recovery. Whitening is the best regulariser at d=16, n_o=8 (+0.370), while VICReg wins at d=4, n_o=2 (+0.255) (Table 20, Figure 8). At d=4, n_o=2 seed variance is much larger (baseline standard deviation 0.11 against 0.01) and the intervals overlap. The non-sequential multilayer perceptron (MLP) encoder, which sees only o_t, scores zero within the probe's floor (-0.309 and -0.495, win rate 0.00). Its d_z of -34.91 and whitening's 9.61 record seed-to-seed reproducibility, not a generalisable effect magnitude; the best arm at d=4, n_o=2 gives 1.22. Both losses are significant against the baseline (p_{corr}=3.2×10^{-30} and 3.5×10^{-13}), in the harmful direction; we read the direction from the sign of Δ rather than from the test.