ANISOTROPIC REPRESENTATIONS IMPROVE PLANNING IN JEPA WORLD MODELS

Anonymous authors
Paper under double-blind review

ABSTRACT

Latent world models learn action-conditioned dynamics in representation space and often score candidate actions by Euclidean distance to a goal representation. Joint training typically regularizes the representation to prevent collapse, but the resulting representation geometry also determines how terminal errors are weighted during planning. We show that accurate prediction and noncollapsed representations do not guarantee a task-aligned latent planning cost: isotropic Gaussian regularization can induce a geometry that ranks feasible outcomes differently from the task cost. To address this mismatch, we introduce AnisoWM with ΛReg, which replaces the fixed isotropic Gaussian target with a learnable diagonal covariance under fixed-trace and anisotropy constraints. The prediction objective, predictor architecture, and Euclidean planner remain unchanged; the target is used only during training. Our analysis characterizes the prediction-driven allocation of target variance, its dependence on the training distribution, and the conditions under which the induced metric reduces planning regret. Across four visual control environments, AnisoWM improves planning success over LeWorldModel in all four. Its latent planning cost also shows better agreement with task outcomes.

1 INTRODUCTION

JEPA-based world models (Maes et al., 2026; Assran et al., 2025; Zhou et al., 2025) predict how an agent's state will evolve under candidate actions in a learned representation space. In visual goal planning, the planner rolls out candidate actions in this space and scores the predicted outcome by its distance to the goal representation, often using Euclidean distance. As a result, the encoder does more than provide features for prediction: it also determines the geometry of the planning cost, and hence how different terminal errors are weighted.

LeWorldModel (LeWM) (Maes et al., 2026) jointly learns an encoder and an action-conditioned predictor from visual observations. Given a current observation and a visual goal, the encoder maps them into latent representations, while the predictor rolls out candidate action sequences in latent space. LeWM scores an action sequence U by

$$J_z(U) = \|\hat{z}_H(U) - z_g\|^2, \tag{1}$$

where H is the planning horizon, $\hat{z}_H(U)$ is the predicted terminal representation, and z_g is the goal representation. Since the planner minimizes $J_z(U)$, its ranking of candidate outcomes should agree with the task cost.

However, LeWM does not explicitly optimize the representation geometry for alignment with the task cost. The objective combines prediction loss with Sketched Isotropic Gaussian Regularization (SIGReg), introduced in LeJEPA (Balestriero & LeCun, 2025), to prevent representation collapse. SIGReg encourages the learned features to follow an isotropic Gaussian distribution. The isotropic target is motivated by theoretical criteria for downstream representation quality under linear and nonlinear probing, rather than by whether the resulting Euclidean distances are suitable for planning. This mismatch motivates our central question: does the latent geometry learned by jointly optimizing the predictor and SIGReg rank candidate outcomes in the same order as the task cost?

We answer this question by analyzing how joint prediction–SIGReg training determines the geometry used for planning (Sec. 3). In a linear Gaussian setting, as process noise vanishes, the joint objective

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.

selects an approximately whitened representation, so Euclidean latent distance induces inverse-state-covariance weighting in state space. This places greater emphasis on low-variance directions and can change the ordering of feasible outcomes relative to the task cost, leaving positive planning regret even with exact conditional-mean prediction in representation space. A nonlinear MLP toy experiment exhibits the same prediction–planning separation. Together, these results motivate learning how variance is allocated across latent directions rather than fixing every direction to the same target variance.

We therefore introduce AnisoWM with ΛReg (Sec. 4), which learns a diagonal Gaussian target jointly with the encoder and predictor. We keep the total target variance fixed and bound its anisotropy by a condition-number constraint κ, while the allocation across latent directions is learned through joint prediction–regularization training. The prediction objective, predictor architecture, and Euclidean planner remain unchanged. Our analysis shows that the learned target can counteract the inverse-covariance weighting induced by isotropic regularization, while also showing that excessive anisotropy can increase planning regret.

We evaluate AnisoWM on four visual goal-planning environments. AnisoWM improves planning success over LeWM in all four environments. Its latent costs also show better agreement with task outcomes, providing an empirical counterpart to the ordering mismatch highlighted by our analysis. Under the same anisotropy bound, training learns different target spectra across environments, and increasing the allowed anisotropy does not monotonically improve planning. Together, these results suggest that the learned variance allocation plays an important role in the observed planning improvements.

Our main contributions can be summarized as follows:

- We show that the expected prediction objective with isotropic SIGReg can select a planning-misaligned latent geometry despite non-collapse and exact conditional-mean prediction.
- We introduce AnisoWM with ΛReg, which replaces the fixed isotropic Gaussian target with a constrained anisotropic target whose variance allocation is learned jointly with the encoder and predictor. This learned variance allocation reshapes the latent geometry, and we theoretically show that it can better align latent distances with task costs and thereby reduce planning regret.
- We empirically verify that AnisoWM improves LeWM in the success rate of planning across four visual control environments, as well as the agreement between latent costs and recorded task outcomes.

2 Related Work

Latent world models. Latent world models learn predictive dynamics in representation space for planning and control. PlaNet (Hafner et al., 2019) learns latent dynamics directly from pixels, while TD-MPC (Hansen et al., 2022) learns task-oriented latent dynamics for model-predictive control. More recent visual world models predict over learned or pretrained visual representations, including DINO-WM (Zhou et al., 2025) and V-JEPA-based approaches (Assran et al., 2025). LeWorldModel (LeWM) (Maes et al., 2026) jointly trains an encoder and action-conditioned predictor with SIGReg and plans using Euclidean distance between predicted and goal representations. Related work modifies this pipeline in different ways: Fast-LeWM (Gao & Xu, 2026) changes the predictive structure to reduce rollout cost and accumulated error, while RC-aux (Li et al., 2026b) augments training with multi-horizon prediction and budget-conditioned reachability supervision.

Representation regularization. Self-supervised objectives commonly constrain feature statistics to prevent collapse and redundancy. Barlow Twins (Zbontar et al., 2021), whitening-based methods (Ermolov et al., 2021), and VICReg (Bardes et al., 2022) impose second-order constraints on learned representations. LeJEPA (Balestriero & LeCun, 2025) derives an isotropic Gaussian target from probing-risk criteria and introduces SIGReg to encourage representations to match that target. Alternative representation distributions include radial Gaussianization in Radial-VCReg (Kuang et al., 2026a) and sparse nonnegative targets in Rectified LpJEPA (Kuang et al., 2026c) and LpWM (Kuang et al., 2026b). HamJEPA (Alvarez, 2026) studies anisotropic Gaussian geometry derived from a prescribed structured geometry, whereas TC-LeWM (Liu et al., 2026) changes which features are regularized by SIGReg through temporal centering.

Geometry for planning. A broader line of work studies representations whose geometry reflects control-relevant state similarity. Bisimulation-based methods (Ferns et al., 2004; Zhang et al., 2021) and DeepMDP (Gelada et al., 2019) connect latent distances and dynamics to behavioral equivalence. More recent work focuses directly on latent planning. Temporal Straightening (Wang et al., 2026b) reduces trajectory curvature for gradient-based planning, while SCALE (Hu et al., 2026) calibrates LeWM distances against a task-relevant state space. TRM (Li et al., 2026a) learns a horizon-aware trajectory-reachability metric that replaces or augments the terminal planning cost, and Decision-Metric Alignment (Wang et al., 2026a) introduces latent–outcome ranking diagnostics together with action-conditioned objectives for improving planning geometry. Related identifiability results (Klindt et al., 2026) characterize conditions under which predictive learning recovers state up to transformations that preserve Euclidean geometry. AnisoWM instead addresses planning geometry through the Gaussian representation regularizer while retaining the Euclidean planner.

3 Analysis: Isotropic Regularization and Planning Cost

Throughout this section, we use *metric matrix* to denote a positive-definite matrix $M \succ 0$ that weights the discrepancy between two physical states $x, x_g \in \mathbb{R}^D$ through the quadratic cost

$$(x - x_g)^\top M (x - x_g).$$

Since planning depends only on cost rankings, M and cM for any $c > 0$ are equivalent for our purposes. To show that isotropic regularization can induce planning regret, we first identify the state-space metric induced by isotropic feature covariance, then show that the expected joint prediction–SIGReg objective selects this metric in arbitrary dimension. Finally, we connect the selected metric to finite-horizon planning regret. Proofs and the explicit finite-horizon construction are given in Appendices B and E.

3.1 The metric induced by isotropic covariance

An affine encoder with linear part A induces the latent Euclidean cost

$$\|A\Delta x\|^2 = \Delta x^\top M \Delta x, \qquad M = A^\top A, \qquad \Delta x = x - x_g.$$

For invertible A, $M \succ 0$; for singular A, the same expression defines a positive-semidefinite quadratic cost. The following lemma identifies the metric under isotropic feature covariance.

Lemma 1 (Metric induced by isotropic covariance)**.** *Let A be the linear part of a square invertible affine encoder and let $\Sigma \succ 0$ denote the state covariance. If the encoded covariance is isotropic,*

$$A\Sigma A^\top = sI$$

for some $s > 0$, then

$$M = A^\top A = s\Sigma^{-1}, \qquad \|A(x - x_g)\|^2 = s(x - x_g)^\top \Sigma^{-1}(x - x_g). \tag{2}$$

Consequently, this latent Euclidean cost agrees up to positive scale with a task cost $\Delta x^\top Q \Delta x$, where $Q \succ 0$, for all residuals if and only if Q is proportional to Σ^{-1}.

Thus isotropic feature covariance does not in general induce an isotropic metric in the original state coordinates. Instead, it weights errors by inverse state covariance: directions with lower training variance receive greater weight. If the task metric and the latent metric weight directions differently, they can prefer different actions when planning requires trading off errors across directions.

3.2 The metric selected by joint prediction–SIGReg training

The preceding lemma is a geometric statement about an isotropic representation. We now ask whether the expected joint prediction–SIGReg objective actually selects this geometry. Consider independent Gaussian training tuples (x, a, x') of state, action, and next state,

$$\begin{aligned} x' &= F_\eta x + Ga + \xi, \\ x &\sim \mathcal{N}(0, \Sigma), \qquad a \sim \mathcal{N}(0, \Gamma), \qquad \xi \sim \mathcal{N}(0, \eta W), \end{aligned} \tag{3}$$

where x, a, and ξ are mutually independent, $\Sigma, W \succ 0$, $\Gamma \succ 0$, and $G \in \mathbb{R}^{D \times d_a}$.

We optimize square linear encoders $A \in \mathbb{R}^{D \times D}$, including singular encoders to avoid assuming noncollapse, together with a linear predictor p taking (Ax, a) as input. Write the prediction loss $\mathcal{L}_{\text{pred}}$ and the state-whitened process-noise covariance R as

$$\mathcal{L}_{\text{pred}}(A, p) = \tfrac{1}{2}\mathbb{E}\left\|p(Ax, a) - Ax'\right\|^2, \qquad R = \Sigma^{-1/2} W \Sigma^{-1/2},$$

and consider the isotropic-target objective,

$$\mathcal{L}_B(A, p) = \mathcal{L}_{\text{pred}}(A, p) + \lambda_B \mathcal{R}_N(A\Sigma A^\top), \qquad \lambda_B > 0. \tag{4}$$

Here $\mathcal{R}_N$ is the expected finite-batch SIGReg statistic characterized in Appendix B.2, and λ_B balances this regularization against prediction. Under the finite-statistic assumptions there, $\mathcal{R}_N$ is uniquely minimized at qI_D for some $q > 0$, so it favors a noncollapsed isotropic feature covariance. The common scale q does not affect planning-cost rankings.

Proposition 1 (Metric selected by isotropic joint training)**.** *Assume the finite-statistic conditions of Lemma 2 and fix $\lambda_B > 0$. For all sufficiently small $\eta > 0$, Eq.* (4) *attains a global minimum. Every minimizing encoder is invertible, even though singular encoders are admissible, and every minimizing predictor recovers the exact encoded conditional mean,*

$$p_A(z, a) = AF_\eta A^{-1} z + AGa.$$

All global minimizers induce the same Gram matrix

$$K_{B,\eta} = (A\Sigma^{1/2})^\top (A\Sigma^{1/2}),$$

and, uniformly over the global minimizers,

$$K_{B,\eta} \longrightarrow qI_D, \qquad M_{B,\eta} := A^\top A \longrightarrow q\Sigma^{-1} \qquad \text{as } \eta \downarrow 0. \tag{5}$$

The attained prediction loss is

$$\mathcal{L}_{\text{pred}}(A, p) = \tfrac{\eta}{2} \operatorname{tr}(K_{B,\eta} R) \longrightarrow 0.$$

The mechanism follows by reducing the joint objective to

$$K = (A\Sigma^{1/2})^\top (A\Sigma^{1/2}).$$

For an invertible encoder, the exact conditional-mean predictor attains the irreducible prediction loss $\eta \operatorname{tr}(KR)/2$, while orthogonal invariance makes the expected SIGReg term depend only on K. The reduced objective is therefore

$$\frac{\eta}{2} \operatorname{tr}(KR) + \lambda_B \mathcal{R}_N(K).$$

Proposition 1 ensures that every global minimizer is invertible for sufficiently small η, so this reduction applies at the optima of interest. Prediction favors shrinking the encoding along high-noise directions, but as $\eta \downarrow 0$ the isotropic regularizer determines the leading-order geometry, forcing $K \to qI_D$ and hence $A^\top A \to q\Sigma^{-1}$. Thus joint training selects the inverse-covariance geometry despite noncollapse, exact encoded prediction, and vanishing prediction loss.

3.3 Finite-horizon planning separation

We next ask whether the metric mismatch can change the action sequence selected over a finite planning horizon. For a deterministic sequence $U = (a_0, \ldots, a_{H-1})$, let $J_{*,\eta,H}(U)$ denote the expected squared Euclidean terminal task cost, and define

$$\operatorname{Regret}_{\eta,H}(U) = J_{*,\eta,H}(U) - \min_V J_{*,\eta,H}(V).$$

Under the finite-statistic conditions of Lemma 2, consider any state dimension $D \geq 2$, any fixed finite horizon $H \geq 1$, any action bound $\bar{u} > 0$, and any nonscalar state covariance $\Sigma \succ 0$. Appendix E.5 constructs a fully actuated linear Gaussian control family with stationary covariance Σ, isotropic process noise, bounded per-step planning actions, and a fixed goal.

Theorem 1 (Finite-horizon planning separation). *For any fixed $\lambda_B > 0$ and sufficiently small $\eta > 0$, every global optimum of the expected prediction–SIGReg objective has a noncollapsed encoder and exact encoded conditional-mean rollouts, with training prediction loss and fixed-H encoded rollout error vanishing as $\eta \downarrow 0$. Nevertheless, every exact Euclidean latent-planning minimizer $U_{B,\eta}$ satisfies*

$$\mathrm{Regret}_{\eta,H}(U_{B,\eta}) \longrightarrow \Delta_H > 0. \tag{6}$$

As $\eta \downarrow 0$, the reachable terminal means in this construction form a Euclidean ball. The task cost selects the Euclidean projection of the goal onto this ball, whereas Proposition 1 makes the latent planner select the projection under Σ^{-1}. For a goal outside this ball with $\Sigma^{-1}x_g$ not parallel to x_g, the two projections differ, yielding the positive regret in Eq. (6) even as prediction and fixed-H rollout error vanish. The mismatch arises because an unreachable goal forces the planner to trade off terminal errors across directions, which the inverse-covariance metric weights differently from the Euclidean task cost.

3.4 Illustrating the Prediction–Planning Gap

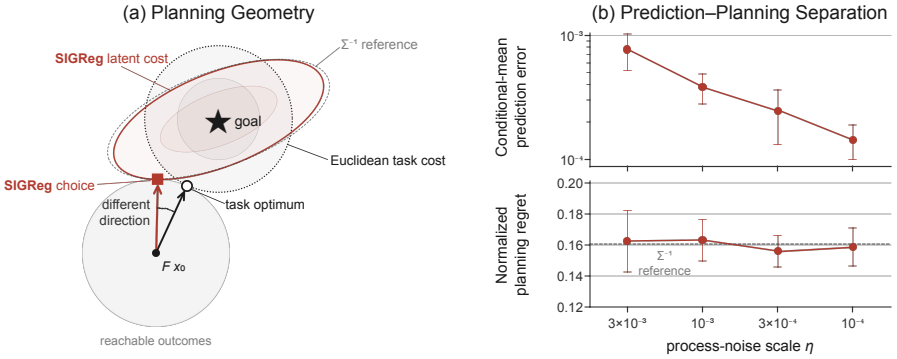

Figure 1: **Prediction–planning separation with a nonlinear encoder. (a)** The Euclidean task cost and the SIGReg latent cost select different outcomes from the same reachable set; the learned latent geometry closely follows the Σ^{-1} reference. **(b)** As process noise decreases, held-out conditional-mean prediction error decreases while normalized physical planning regret remains nearly unchanged. Points and error bars show the mean and 95% t-confidence interval over ten training seeds.

To empirically illustrate Theorem 1, we test whether the prediction–planning separation persists beyond the linear setting on a two-dimensional controlled system designed to isolate the geometric mismatch analyzed above. The training actions and process noise are isotropic, while the dynamics induce an anisotropic stationary state distribution. At evaluation, goals are chosen outside the one-step reachable set, forcing the planner to trade off residual errors across state dimensions; the system is constructed so that the Euclidean task cost and the inverse-covariance latent cost prefer different reachable outcomes. We train an MLP encoder and action-conditioned predictor with the prediction–SIGReg objective, using the nonlinear encoder to test whether the same behavior persists beyond the linear setting analyzed above. As the process-noise scale decreases by $30\times$, held-out conditional-mean prediction error decreases substantially, while mean normalized physical planning regret remains near 0.16 over ten seeds (Figure 1). The observed regret closely matches the 0.1607 inverse-covariance reference, and the learned local pullback metric is close to Σ^{-1} up to scale. Thus improved prediction does not remove the geometric action-ranking mismatch, motivating the anisotropic regularization introduced in Sec. 4.

4 Method

AnisoWM replaces SIGReg's fixed isotropic Gaussian target with a learnable diagonal covariance. This permits different target variances across latent coordinates rather than imposing the same target variance in every direction. The design follows the analysis in Sec. 3: when Euclidean distance is used for planning, the regularization target also shapes how terminal errors are weighted by the

learned representation. We therefore relax the isotropy constraint during representation learning while leaving the prediction objective, predictor architecture, and Euclidean planner unchanged.

4.1 Learnable Gaussian target

Let $f_\theta(o) \in \mathbb{R}^D$ be the encoder output and let $\Lambda = \mathrm{diag}(v_1, \dots, v_D)$ denote the covariance of a zero-mean Gaussian target. We constrain

$$\mathcal{T}_{D,\kappa} = \{\Lambda = \mathrm{diag}(v_1, \dots, v_D) \succ 0 : \mathrm{tr}\,\Lambda = D, \quad \mathrm{cond}(\Lambda) \le \kappa\}, \qquad \kappa \ge 1, \tag{7}$$

where $\mathrm{cond}(\Lambda) = \max_i v_i / \min_i v_i$. The trace fixes the overall target scale, while κ bounds the allowed anisotropy. The target is diagonal in representation coordinates, while the encoder remains free to orient those coordinates relative to the underlying state geometry. We initialize $\Lambda = I_D$; $\kappa = 1$ recovers the isotropic target.

For a feature batch $Z = (z_1, \dots, z_N)$, we define

$$\widehat{\mathcal{R}}_{N,\Lambda}(Z; \boldsymbol{\omega}) = \widehat{\mathcal{R}}_N\left(\Lambda^{-1/2} Z; \boldsymbol{\omega}\right), \tag{8}$$

where $\widehat{\mathcal{R}}_N$ is the finite-batch SIGReg statistic (Balestriero & LeCun, 2025). Because $z \sim \mathcal{N}(0, \Lambda)$ implies $\Lambda^{-1/2} z \sim \mathcal{N}(0, I_D)$, the original SIGReg reference distribution can be used unchanged. This standardization is applied only inside the regularizer; prediction and planning use the original representation z.

4.2 Joint training and planning

The predictor produces $\hat{z}_{t+1} = g_\phi(z_t, a_t)$ and is trained with mean squared prediction loss $\mathcal{L}_{\mathrm{pred}}(\theta, \phi; \mathcal{B})$. Let Z_θ denote the encoder features used by the regularizer. We minimize

$$\widehat{\mathcal{L}}_T = \mathcal{L}_{\mathrm{pred}}(\theta, \phi; \mathcal{B}) + \lambda \widehat{\mathcal{R}}_{N,\Lambda}(Z_\theta; \boldsymbol{\omega}), \qquad \Lambda \in \mathcal{T}_{D,\kappa}. \tag{9}$$

Both terms update the encoder, prediction loss updates the predictor, and Λ is updated only through the regularization term.

We parameterize

$$v_i = D \frac{e^{\alpha_i}}{\sum_j e^{\alpha_j}}, \qquad \alpha_i \in \left[-\tfrac{1}{2}\log\kappa, \tfrac{1}{2}\log\kappa\right], \tag{10}$$

with zero initialization and clipping after each update. This enforces $\mathrm{tr}\,\Lambda = D$ and $\mathrm{cond}(\Lambda) \le \kappa$ and covers all of $\mathcal{T}_{D,\kappa}$ (Appendix C.3).

Planning uses autoregressive latent rollouts and is otherwise unchanged from LeWM:

$$J_z(U) = \|\hat{z}_H(U) - z_g\|^2, \qquad z_g = f_\theta(o_g).$$

The covariance Λ is used only by the training regularizer and can be discarded afterward; thus AnisoWM changes the learned representation geometry without introducing an additional planning-time metric.

4.3 Effect on planning geometry

We now characterize how the learnable target changes the state-space metric selected by joint training. Return to the Gaussian model of Section 3.2, with state covariance $\Sigma \succ 0$ and state-whitened process-noise covariance

$$R = \Sigma^{-1/2} W \Sigma^{-1/2}.$$

For a linear encoder A, write

$$K = (A\Sigma^{1/2})^\top (A\Sigma^{1/2}), \qquad A\Sigma^{1/2} = OK^{1/2}, \qquad L = O^\top \Lambda O,$$

where O is orthogonal. The matrix L expresses the diagonal target covariance in state-whitened coordinates, including the orientation selected by the encoder. Its feasible set is

$$\mathcal{S}_{D,\kappa} = \{L \succ 0 : \mathrm{tr}\,L = D,\ \mathrm{cond}(L) \le \kappa\}.$$

Although L need not be diagonal, it arises from the encoder orientation together with a diagonal Λ and does not introduce a full-covariance target parameterization.

For the expected Gaussian model, the learned-target objective is

$$\mathcal{L}_T(A, p, \Lambda) = \mathcal{L}_{\text{pred}}(A, p) + \lambda_T \mathcal{R}_N\left(\Lambda^{-1/2} A \Sigma A^\top \Lambda^{-1/2}\right), \qquad \Lambda \in \mathcal{T}_{D,\kappa}. \tag{11}$$

Proposition 2 (Metric selected by target learning)**.** *Assume the finite-statistic conditions of Lemma 2, and fix $\lambda_T > 0$ and $\kappa > 1$. For all sufficiently small $\eta > 0$, Eq. (11) attains a global minimum. Every global minimizer has an invertible encoder and an exact encoded conditional-mean predictor. Uniformly over global minimizers,*

$$K - qL \longrightarrow 0, \qquad A^\top A - q\Sigma^{-1/2} L \Sigma^{-1/2} \longrightarrow 0. \tag{12}$$

Every accumulation point of L minimizes

$$\operatorname{tr}(LR) \qquad \textit{over} \qquad L \in \mathcal{S}_{D,\kappa}. \tag{13}$$

If the minimizer is unique, then L converges to it uniformly over global training optima, and the attained prediction loss tends to zero.

The proof is given in Appendix E.3.

Prediction-driven metric selection. Under the isotropic target, $L = I_D$, and Proposition 1 recovers the inverse-covariance metric $q\Sigma^{-1}$. Learning the target enlarges the limiting family to

$$q\Sigma^{-1/2} L \Sigma^{-1/2}, \qquad L \in \mathcal{S}_{D,\kappa}.$$

Along global minimizers, $K = qL + o(1)$, so the prediction term is

$$\frac{\eta}{2} \operatorname{tr}(KR) = \frac{\eta q}{2} \operatorname{tr}(LR) + o(\eta),$$

yielding the selection rule in Eq. (13). Thus κ bounds the admissible anisotropy, while predictive training selects how that anisotropy is allocated.

Task alignment and finite-horizon compensation. For a task metric $Q \succ 0$, the limiting representation metric is proportional to Q when

$$L_Q = \frac{D\Sigma^{1/2} Q \Sigma^{1/2}}{\operatorname{tr}(\Sigma Q)}. \tag{14}$$

Exact alignment therefore requires both $L_Q \in \mathcal{S}_{D,\kappa}$ and that predictive training select this geometry. For the Euclidean task metric $Q = I_D$, L_Q is the trace-normalized state covariance, showing how the learnable target can compensate for the inverse-covariance weighting induced by isotropic regularization without modifying the planner.

The finite-horizon construction of Theorem 1 realizes this compensation explicitly. For the trace-normalized two-level covariance family in Eq. (79), the construction uses $W = I_D$, so $R = \Sigma^{-1}$. With $\kappa = \operatorname{cond}(\Sigma)$, Eq. (13) uniquely selects $L \to \Sigma$, and hence

$$A^\top A \longrightarrow qI_D, \qquad \operatorname{Regret}_{\eta,H}(U_{T,\eta}) \longrightarrow 0. \tag{15}$$

The isotropic-target planner on the same family retains the positive limiting regret in Eq. (6), while both objectives have vanishing prediction loss and fixed-horizon encoded rollout error.

Appendix C gives closed-form regret analysis, while Appendix D analyzes higher-dimensional, distribution-dependent spectrum selection.

5 Experiments

5.1 Experimental setup

We evaluate on TwoRoom, Reacher, PushT, and Cube using the datasets, model architecture, and visual goal-planning protocol of LeWM (Maes et al., 2026). AnisoWM uses a 192-dimensional

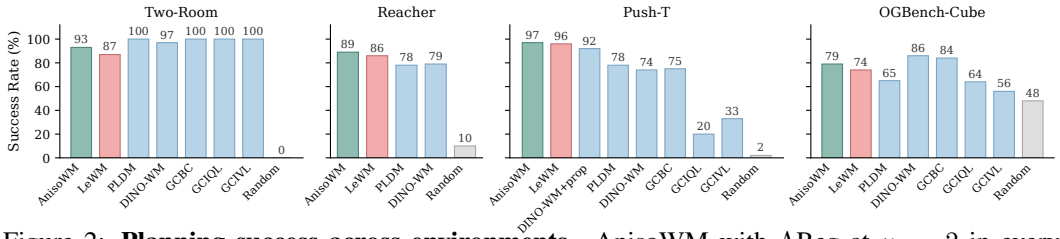

Figure 2: **Planning success across environments.** AnisoWM with ΛReg at $\kappa = 2$ in every environment, compared with LeWM and the reference baselines. Reported AnisoWM and LeWM values are means over three training seeds.

Table 1: **Ordering recorded action sequences by the outcome they reached.** For each initial–goal pair, both models score the same sequences, and the entry is the fraction of sequence pairs whose cost ordering agrees with their outcome ordering. J_{enc} encodes the observation a sequence actually reached; J_{pred} is the cost used for planning and additionally includes the predictor rollout.

	J_{enc} (representation only)			J_{pred} (planning cost)		
Task	LeWM	AnisoWM	Gap	LeWM	AnisoWM	Gap
TwoRoom	0.564	0.661	+0.097	0.575	0.646	+0.071
Reacher	0.790	0.893	+0.103	0.676	0.723	+0.048
PushT	0.647	0.641	−0.005	0.594	0.621	+0.028
Cube	0.555	0.576	+0.021	0.537	0.553	+0.016

representation and regularization weight $\lambda = 0.09$. Implementation and regularizer hyperparameters are provided in Appendix F.1. Each environment in the primary comparison is trained with three seeds, and the reported value is the mean over the three. For the primary comparison, we use a single shared anisotropy bound $\kappa = 2$ across all environments, which permits at most a twofold ratio between target variances while avoiding environment-specific tuning. Prediction and CEM planning operate in the original latent coordinates. Success rates for PLDM (Sobal et al., 2025), DINO-WM (Zhou et al., 2025), GCBC (Ghosh et al., 2021), GCIQL (Kostrikov et al., 2022; Park et al., 2025), GCIVL (Park et al., 2025), and Random are taken from Maes et al. (2026). The LeWM baseline is trained in our pipeline with the released code and identical settings ($\lambda = 0.09$, training seeds, and evaluation pairs); it corresponds to $\kappa = 1$ in our parameterization.

5.2 Results and Analysis

Planning performance. AnisoWM plans more successfully than LeWM in all four environments (Figure 2): 93% against 87% in TwoRoom, 89% against 86% in Reacher, 97% against 96% in PushT, and 79% against 74% in Cube. Because AnisoWM changes only the representation regularization, these gains preserve LeWM's predictor and Euclidean planner. Although some reference baselines achieve higher success in TwoRoom and Cube, AnisoWM retains LeWM's efficient latent-space planning while achieving competitive performance.

Action ranking. Motivated by Theorem 1, we compare how LeWM and AnisoWM order the same recorded action sequences against their task outcomes; Appendix F.2 gives the full protocol. We report $J_{\text{pred}}(u) = \|\hat{z}_H(u) - z_g\|^2$, the cost used by CEM, and $J_{\text{enc}}(u) = \|f_\theta(o_H(u)) - f_\theta(o_g)\|^2$, which evaluates the realized terminal observation without predictor rollout. As shown in Table 1, AnisoWM improves J_{pred} ordering in all four environments. Under J_{enc}, the gain is positive in TwoRoom, Reacher, and Cube, while PushT is nearly unchanged, indicating that the contribution of representation geometry and predictor rollout differs across environments.

Local cost geometry. Figure 3 visualizes the latent planning cost around goals in Cube and PushT. AnisoWM produces low-cost neighborhoods that more closely follow the corresponding task cost in physical distance, whereas LeWM assigns low cost to positions that can lie farther from the goal. Spearman's rank correlation (ρ) between latent cost and task cost increases from 0.15 to 0.91 in Cube and from 0.65 to 0.92 in PushT. Additional examples are provided in Appendix F.3.

Learned target spectra. Despite the shared bound $\kappa = 2$, the learned spectra differ across environments: the number of coordinates above the mean target variance ranges from 95 in PushT to 124 in Reacher (Figure 4). In TwoRoom, the allocation continues to evolve after reaching the condition-number bound, showing that κ constrains but does not determine the learned spectrum.

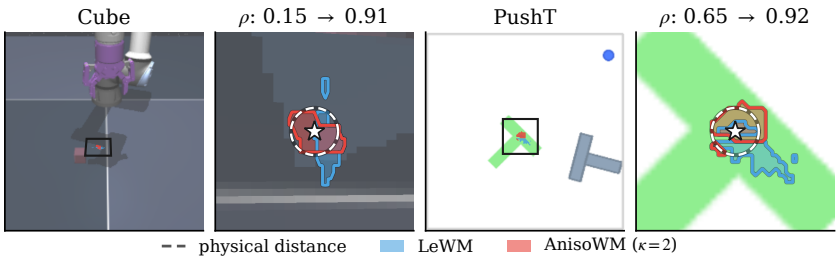

Figure 3: **Latent cost neighborhoods around the goal.** For Cube and PushT, the left panels show the rendered arena and evaluated region, and the right panels show positions in the lowest 10% of latent planning cost around the goal ($\star$) for LeWM (blue) and AnisoWM (κ=2, red). The dashed circle shows the corresponding neighborhood under physical Euclidean distance.

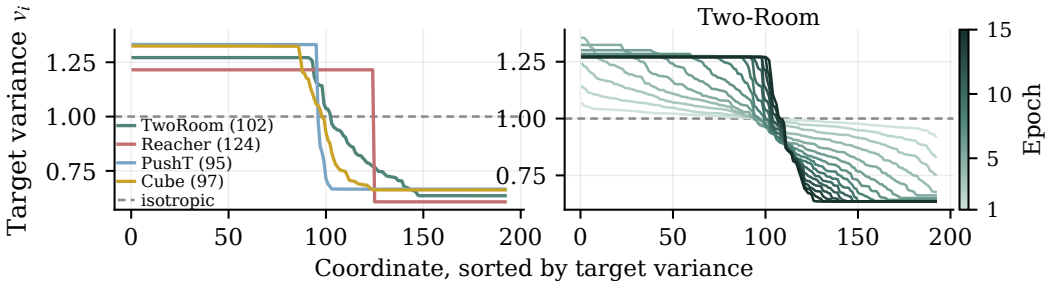

Figure 4: **Learned target spectra.** Left: learned targets at $\kappa = 2$, sorted by variance, with the isotropic target marked. The trace and maximum variance ratio are fixed, while the allocation is learned; the legend shows the number of coordinates above the mean variance. Right: target spectrum during one TwoRoom run, showing continued evolution after the condition-number bound is reached.

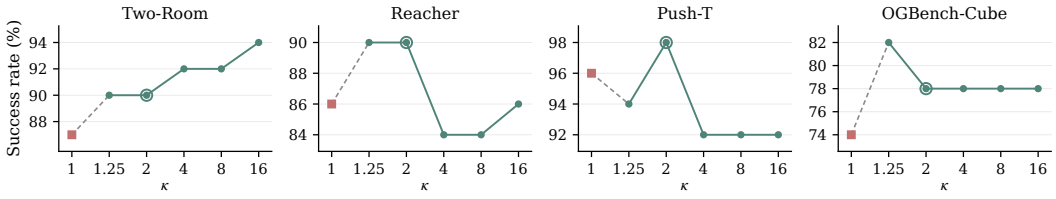

Figure 5: **Planning success across anisotropy bounds.** Each point with $\kappa > 1$ reports one training run evaluated on the same set of initial–goal pairs. The $\kappa = 1$ marker shows the isotropic LeWM baseline from the primary comparison.

Sensitivity to the anisotropy bound. Figure 5 shows non-monotone sensitivity to κ, with the highest observed anisotropic setting differing across environments. Because each $\kappa > 1$ point is a single training run, we treat the sweep as a sensitivity diagnostic rather than a tuned comparison. Appendix F.5 reports the numerical values.

Additional experimental details are provided in Appendix F.

6 Conclusion

We studied how isotropic Gaussian regularization shapes the geometry used for latent planning. Our analysis shows that accurate prediction and noncollapsed representations do not guarantee a task-aligned planning cost. We introduced AnisoWM with ΛReg, which learns a constrained anisotropic Gaussian target while leaving the prediction objective and Euclidean planner unchanged. Our analysis characterizes the prediction-driven selection of representation geometry and establishes conditions under which it reduces planning regret. Empirically, AnisoWM achieves higher planning success than LeWM across four visual control environments, and its latent costs show higher agreement with recorded task outcomes.

AI USE STATEMENT

AI tools assisted with writing and editing the manuscript and with selected mathematical checks using Lean. The authors are responsible for the final manuscript and all reported results.

REFERENCES

Robert Jenkinson Alvarez. Beyond isotropy in JEPAs: Hamiltonian geometry and symplectic prediction. *arXiv:2605.20107*, 2026.

Mahmoud Assran, Adrien Bardes, David Fan, Quentin Garrido, Russell Howes, Mojtaba Komeili, Matthew Muckley, Ammar Rizvi, Claire Roberts, Koustuv Sinha, Artem Zholus, Sergio Arnaud, Abha Gejji, Ada Martin, Francois Robert Hogan, Daniel Dugas, Piotr Bojanowski, Vasil Khalidov, Patrick Labatut, Francisco Massa, Marc Szafraniec, Kapil Krishnakumar, Yong Li, Xiaodong Ma, Sarath Chandar, Franziska Meier, Yann LeCun, Michael Rabbat, and Nicolas Ballas. V-JEPA 2: Self-supervised video models enable understanding, prediction and planning. *arXiv:2506.09985*, 2025.

Randall Balestriero and Yann LeCun. LeJEPA: Provable and scalable self-supervised learning without the heuristics. *arXiv:2511.08544*, 2025.

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

Aleksandr Ermolov, Aliaksandr Siarohin, Enver Sangineto, and Nicu Sebe. Whitening for self-supervised representation learning. In *Proceedings of the International Conference on Machine Learning (ICML)*, 2021.

Norm Ferns, Prakash Panangaden, and Doina Precup. Metrics for finite markov decision processes. In *Proceedings of the Conference on Uncertainty in Artificial Intelligence (UAI)*, 2004.

Yuntian Gao and Xiangyu Xu. Fast LeWorldModel. *arXiv:2606.26217*, 2026.

Carles Gelada, Saurabh Kumar, Jacob Buckman, Ofir Nachum, and Marc G. Bellemare. DeepMDP: Learning continuous latent space models for representation learning. In *Proceedings of the International Conference on Machine Learning (ICML)*, 2019.

Dibya Ghosh, Abhishek Gupta, Ashwin Reddy, Justin Fu, Coline Devin, Benjamin Eysenbach, and Sergey Levine. Learning to reach goals via iterated supervised learning. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2021.

Danijar Hafner, Timothy Lillicrap, Ian Fischer, Ruben Villegas, David Ha, Honglak Lee, and James Davidson. Learning latent dynamics for planning from pixels. In *Proceedings of the International Conference on Machine Learning (ICML)*, 2019.

Nicklas A. Hansen, Hao Su, and Xiaolong Wang. Temporal difference learning for model predictive control. In *Proceedings of the International Conference on Machine Learning (ICML)*, 2022.

Jiaming Hu, Yan Zheng, and Tian Wang. SCALE: State-calibrated latent embeddings for JEPA planning in the right geometry. *arXiv:2608.16287*, 2026.

David Klindt, Yann LeCun, and Randall Balestriero. When does LeJEPA learn a world model? *arXiv:2605.26379*, 2026.

Ilya Kostrikov, Ashvin Nair, and Sergey Levine. Offline reinforcement learning with implicit Q-learning. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2022.

Yilun Kuang, Yash Dagade, Deep Chakraborty, Erik Learned-Miller, Randall Balestriero, Tim G. J. Rudner, and Yann LeCun. Radial-VCReg: More informative representation learning through radial gaussianization. *arXiv:2602.14272*, 2026a.

Yilun Kuang, Yash Dagade, Quentin Le Lidec, Lucas Maes, Randall Balestriero, and Yann LeCun. LpWM: A case for sparse representations in world models. *arXiv:2608.22764*, 2026b.

Yilun Kuang, Yash Dagade, Tim G. J. Rudner, Randall Balestriero, and Yann LeCun. Rectified Lp-JEPA: Joint-embedding predictive architectures with sparse and maximum-entropy representations. *arXiv:2602.01456*, 2026c.

Liangyu Li, Shengzhi Wang, Libin Qiu, Mingliang Xiong, and Qingwen Liu. World model control by trajectory reachability metrics. *arXiv:2605.22164*, 2026a.

Wenyuan Li, Guang Li, Keisuke Maeda, Takahiro Ogawa, and Miki Haseyama. Predictive but not plannable: Rc-aux for latent world models. *arXiv:2605.07278*, 2026b.

Chang Liu, Fei Suo, Yanzhou Jin, Zeyu Ping, Yusuke Iwasawa, Yutaka Matsuo, and Yaonan Zhu. Temporally centered SIGReg improves LeWorldModel representations for robot policy learning. *arXiv:2607.26924*, 2026.

Lucas Maes, Quentin Le Lidec, Damien Scieur, Yann LeCun, and Randall Balestriero. LeWorldModel: Stable end-to-end joint-embedding predictive architecture from pixels. *arXiv:2603.19312*, 2026.

Seohong Park, Kevin Frans, Benjamin Eysenbach, and Sergey Levine. Ogbench: Benchmarking offline goal-conditioned rl. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2025.

Uladzislau Sobal, Wancong Zhang, Kyunghyun Cho, Randall Balestriero, Tim G. J. Rudner, and Yann LeCun. Learning from reward-free offline data: A case for planning with latent dynamics models. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2025.

Jiawei Wang, Ke Rui, Yushen Zuo, Yichun Feng, and Minglei Li. Decision-metric alignment in latent world models: Diagnostics and action-conditioned objectives for mpc planning. *arXiv:2608.18746*, 2026a.

Ying Wang, Oumayma Bounou, Gaoyue Zhou, Randall Balestriero, Tim G. J. Rudner, Yann LeCun, and Mengye Ren. Temporal straightening for latent planning. In *Proceedings of the International Conference on Machine Learning (ICML)*, 2026b.

Jure Zbontar, Li Jing, Ishan Misra, Yann LeCun, and Stephane Deny. Barlow Twins: Self-supervised learning via redundancy reduction. In *Proceedings of the International Conference on Machine Learning (ICML)*, 2021.

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

Gaoyue Zhou, Hengkai Pan, Yann LeCun, and Lerrel Pinto. DINO-WM: World models on pre-trained visual features enable zero-shot planning. In *Proceedings of the International Conference on Machine Learning (ICML)*, 2025.

Eric Zimmermann, Harley Wiltzer, Justin Szeto, David Alvarez-Melis, and Lester Mackey. KerJEPA: Kernel discrepancies for euclidean self-supervised learning. *arXiv:2512.19605*, 2025.

APPENDIX

A Notation **13**

B Technical Preliminaries and a Two-Dimensional Closed-Form Specialization **14**

B.1 Latent metric and isotropic covariance 14

B.2 Expected finite-batch SIGReg 15

B.3 Two-dimensional closed-form specialization 16

B.4 Transition construction and prediction reduction for the specialization 17

B.5 Proof of Theorem 2 18

B.6 General residuals and a fixed evaluation state 19

C Closed-Form Analysis of the Learnable Gaussian Target **20**

C.1 Proof of Theorem 3 21

C.2 A target-coupled rescaling path 22

C.3 Spectral bounds for admissible targets 24

C.4 Scope of the analysis 25

D Prediction-Driven Target Spectrum Selection **25**

D.1 Spectrum selection in arbitrary dimension 25

D.2 Two training distributions with different selected spectra 27

E General-Dimensional Geometry, Finite-Horizon Planning, and Nonlinear Stability **29**

E.1 Standing regularizer assumptions 29

E.2 General Gaussian training model and isotropic joint optimum 29

E.3 Proof of Proposition 2 30

E.4 Exact rollouts and finite-horizon regret 31

E.5 Separation in every dimension and at every finite horizon 32

E.6 Stability of the planning separation 34

E.7 Smooth nonlinear perturbations 35

F Experimental Details **37**

F.1 Experimental protocol and per-seed results 37

F.2 Latent-cost ordering diagnostic 39

F.3 Additional local cost geometry 40

F.4 Learned target allocation 40

F.5 Sensitivity to the anisotropy bound 42

F.6 Realized representation geometry 42

F.7 Additional prediction diagnostics 43

F.8 Qualitative rollouts 43

A NOTATION

Table 2: Notation used throughout the paper.

Notation	Description
$o_t,\ o_g$	Observation at time t and goal observation.
$f_\theta,\ g_\phi$	Neural encoder and action-conditioned latent predictor.
$z_t,\ z_g$	Latent representation at time t and goal representation.
$a_t,\ U$	Per-step action and finite-horizon action sequence $U = (a_0, \dots, a_{H-1})$.
H	Planning horizon.
$\hat{z}_H(U)$	Predicted terminal representation after rolling out U.
$J_z(U)$	Euclidean latent terminal cost $\|\hat{z}_H(U) - z_g\|^2$.
$x,\ x'$	Current and successor physical states in the theoretical analysis.
x_g	Generic physical goal state.
$D,\ d_a$	State/representation dimension and action dimension in the general theory.
A	Square linear encoder in the theoretical analysis.
p	Linear predictor in the theoretical analysis.
$M = A^\top A$	State-space metric induced by Euclidean latent distance.
Q	Positive-definite task metric.
Σ	Training-state covariance.
$F_\eta,\ G,\ \Gamma$	Transition matrix, control matrix, and training-action covariance.
$\eta,\ W$	Process-noise scale and noise-shape covariance.
$R = \Sigma^{-1/2} W \Sigma^{-1/2}$	State-whitened process-noise covariance.
$K = (A\Sigma^{1/2})^\top (A\Sigma^{1/2})$	Encoder Gram matrix in state-whitened coordinates.
O	Orthogonal factor in $A\Sigma^{1/2} = OK^{1/2}$.
$N,\ n_{\mathrm{proj}}$	Batch size per regularized time slice and number of projection directions.
$\boldsymbol{\omega}$	Collection of sampled SIGReg projection directions.
$\widehat{\mathcal{R}}_N,\ \mathcal{R}_N$	Finite-batch and expected finite-batch SIGReg statistics.
q	Scalar such that qI_D minimizes the expected finite-batch statistic.
$\Lambda = \mathrm{diag}(v_1, \dots, v_D)$	Learnable diagonal Gaussian target covariance.
κ	Condition-number bound on Λ; $\kappa = 1$ gives the isotropic target.
$\mathcal{T}_{D,\kappa}$	Feasible diagonal target family.
$L = O^\top \Lambda O$	Target geometry in state-whitened coordinates.
$\mathcal{S}_{D,\kappa}$	Feasible oriented target family.
$\lambda, \lambda_B, \lambda_T$	Regularization weights; subscripts distinguish the isotropic- and learned-target theory.
$\mathcal{U},\ y$	Feasible action-sequence set and goal used in the explicit construction.
$\mu_{\eta,H}(U)$	Conditional mean of the terminal physical state.
$\Omega_{\eta,H}$	Terminal process-noise covariance.
$J_{*,\eta,H}(U)$	Expected terminal task cost.
$\mathrm{Regret}_{\eta,H}(U)$	Excess task cost over the best feasible sequence.
Δ_H	Positive limiting regret gap in the separation construction.
$\bar{u},\ r_H$	Per-step action bound and limiting reachable-ball radius.
$J_{\mathrm{pred}}(U)$	Predicted terminal latent cost available to the planner.
$J_{\mathrm{enc}}(U)$	Encoded cost of the realized reached observation.
$\ell_j,\ \pi_j$	Feature-covariance eigenvalue and its normalized spectral weight.
r_{ent}	Entropy effective rank of the measured feature covariance.

B Technical Preliminaries and a Two-Dimensional Closed-Form Specialization

This appendix collects the finite-batch SIGReg facts used throughout the analysis and gives a two-dimensional specialization in which the selected action and planning regret are available in closed form. The general joint-optimum and finite-horizon results used in the main text are proved in Appendix E. All matrix square roots are symmetric positive-semidefinite roots. Matrix derivatives use the Frobenius inner product on symmetric matrices. Every limit as $\eta \downarrow 0$ fixes the batch size, quadrature, noise-shape parameters, action variance, and strictly positive regularization weights. Scaling a regularization weight with η defines a different asymptotic regime.

B.1 Latent metric and isotropic covariance

Proof of Lemma 1. The covariance identity

$$A\Sigma A^\top = sI$$

gives

$$\Sigma = sA^{-1}A^{-\top},$$

and hence

$$A^\top A = s\Sigma^{-1}.$$

Therefore

$$\|A(x - x_g)\|^2 = (x - x_g)^\top A^\top A(x - x_g) = s(x - x_g)^\top \Sigma^{-1}(x - x_g).$$

Finally, two positive-definite quadratic forms agree up to a common positive scale for every residual if and only if their symmetric matrices are proportional. Thus the latent Euclidean cost agrees up to positive scale with $\Delta x^\top Q \Delta x$ for all residuals if and only if Q is proportional to Σ^{-1}. □

Ranking preservation under linear reparameterization. For completeness, consider an invertible linear transform T applied to arbitrary residuals. It preserves all pairwise Euclidean cost rankings, including ties, if and only if

$$T^\top T = cI$$

for some $c > 0$.

To see this, let $M = T^\top T \succ 0$. If $M = cI$, then

$$\|Te\|^2 = c\,\|e\|^2$$

for every residual e, so every ranking and tie is preserved. Conversely, preservation of all ties requires $v^\top M v$ to be constant over the Euclidean unit sphere. Evaluating the quadratic form at the eigenvectors of M shows that all eigenvalues of M must coincide, and hence $M = cI$.

Ranking reversals between nonproportional metrics. More generally, let $M, Q \succ 0$ be nonproportional. Then there exist residuals e_1, e_2 whose ordering is reversed by the two quadratic costs. Indeed, let v_1, v_2 be orthonormal eigenvectors of $M^{-1/2}QM^{-1/2}$ with eigenvalues $\mu_1 < \mu_2$. Choose $1 < s < \mu_2/\mu_1$ and set

$$e_1 = \sqrt{s}\, M^{-1/2}v_1, \qquad e_2 = M^{-1/2}v_2.$$

Then

$$e_1^\top M e_1 = s > 1 = e_2^\top M e_2,$$

whereas

$$e_1^\top Q e_1 = s\mu_1 < \mu_2 = e_2^\top Q e_2.$$

Thus nonproportional metrics can induce different rankings of feasible outcomes. The finite-horizon construction in Appendix E.5 realizes this geometric ranking mismatch through feasible action sequences.

The affine-encoder assumption in Lemma 1 is essential. In Cartesian coordinates, let $\mathrm{Rot}(\psi)$ be the planar rotation through angle ψ. The smooth map $x \mapsto \mathrm{Rot}(\|x\|^2)x$, or $(r, \theta) \mapsto (r, \theta + r^2)$ in polar coordinates, preserves standard two-dimensional Gaussian measure: the radius is unchanged, and the conditional angle remains uniform. Its Cartesian Jacobian is nonorthogonal away from the origin. A Gaussian marginal therefore places weaker restrictions on the local metric of a nonlinear encoder.

B.2 EXPECTED FINITE-BATCH SIGREG

Let $Z = (z_1, \dots, z_N)$ contain $N > 2$ independent samples from $\mathcal{N}(0, C)$. Let $\omega_1, \dots, \omega_{n_{\text{proj}}}$, with $n_{\text{proj}} \geq 1$, be independent uniform directions on $\mathbb{S}^{D-1}$, independent of the batch. For finitely many nonnegative knots t_k and weights w_k, define

$$\widehat{\mathcal{R}}_N(Z; \boldsymbol{\omega}) = \frac{1}{n_{\text{proj}}} \sum_{\ell=1}^{n_{\text{proj}}} \sum_k w_k N \left| \frac{1}{N} \sum_{b=1}^{N} e^{i t_k \omega_\ell^\top z_b} - e^{-t_k^2/2} \right|^2, \tag{16}$$

where $\boldsymbol{\omega} = (\omega_1, \dots, \omega_{n_{\text{proj}}})$. Assume that at least one positive-weight knot is strictly positive. Averaging over samples and directions gives

$$\begin{aligned} \mathcal{R}_N(C) &= \mathbb{E}_v r_N(v^\top C v), \\ r_N(s) &= \sum_k w_k \big[1 + (N-1)e^{-t_k^2 s} - 2N e^{-t_k^2 (s+1)/2} + N e^{-t_k^2}\big], \end{aligned} \tag{17}$$

where v is uniform on $\mathbb{S}^{D-1}$. We assume $r_N'(0) < 0$, a condition determined by the batch size and quadrature.

Lemma 2 (Covariance minimum of the expected statistic)**.** *There is a unique $q \in (0, 1)$ satisfying $r_N'(q) = 0$, and $\mathcal{R}_N$ is uniquely minimized over $C \succeq 0$ at qI_D. With $c_N = r_N''(q) > 0$, its Hessian satisfies*

$$D^2 \mathcal{R}_N(qI_D)[E, E] = \frac{c_N}{D(D+2)} \left[(\operatorname{tr} E)^2 + 2 \operatorname{tr}(E^2) \right] \tag{18}$$

for every symmetric matrix E.

The contraction from I_D to qI_D reflects finite-batch bias in the squared discrepancy between the empirical and reference characteristic functions (Balestriero & LeCun, 2025). The distinction between population discrepancies and their finite-sample approximations also appears in kernel formulations of SIGReg (Zimmermann et al., 2025). The lemma characterizes Gaussian covariances after averaging over samples and directions.

Proof of Lemma 2. For a fixed direction with projected variance $s = v^\top C v \geq 0$, let $Y_b = e^{i t v^\top z_b}$ and $\widehat{\phi} = N^{-1} \sum_b Y_b$. Independence gives

$$\mathbb{E}|\widehat{\phi}|^2 = \frac{1}{N} + \frac{N-1}{N} e^{-t^2 s}, \qquad \mathbb{E}\widehat{\phi} = e^{-t^2 s/2}. \tag{19}$$

Expanding $N\mathbb{E}|\widehat{\phi} - e^{-t^2/2}|^2$ yields (17), including at $s = 0$. This is the finite-sample expectation of the original biased empirical statistic used by SIGReg (Balestriero & LeCun, 2025).

Writing $a_k = t_k^2$, differentiation gives

$$r_N'(s) = \sum_k w_k a_k \left[-(N-1)e^{-a_k s} + N e^{-a_k (s+1)/2} \right], \tag{20}$$

$$r_N''(s) = \sum_k w_k a_k^2 \left[(N-1)e^{-a_k s} - \tfrac{N}{2} e^{-a_k (s+1)/2} \right]. \tag{21}$$

For $s \in [0, 1]$, the bracket in the second line equals

$$e^{-a_k (s+1)/2} \left[(N-1) e^{a_k (1-s)/2} - \tfrac{N}{2} \right].$$

It is strictly positive when $a_k > 0$, since $N > 2$. Thus r_N' is strictly increasing on $[0, 1]$. The assumption $r_N'(0) < 0$ and the identity

$$r_N'(1) = \sum_k w_k a_k e^{-a_k} > 0$$

give a unique zero $q \in (0,1)$. For $s \geq 1$, the bracket in the first derivative is at least $e^{-a_k(s+1)/2}$, so $r'_N(s) > 0$. Consequently, q is the unique global minimizer of r_N on $[0,\infty)$, and $c_N = r''_N(q) > 0$.

Pointwise minimization gives $\mathcal{R}_N(C) \geq r_N(q)$. Equality requires $v^\top C v = q$ for almost every unit vector v. Continuity extends this identity to every unit vector, which implies $C = qI_D$. Conversely, qI_D attains equality. This proves uniqueness over the entire positive-semidefinite cone, including singular covariances.

The quadrature is finite and the sphere is compact, so the integrand's matrix derivatives are uniformly bounded on each compact covariance set. Differentiation under the expectation is therefore justified by dominated convergence. At qI_D this gives

$$D^2\mathcal{R}_N(qI_D)[E,E] = c_N \mathbb{E}(v^\top E v)^2.$$

The uniform-sphere fourth moments are

$$\mathbb{E}v_i^4 = \frac{3}{D(D+2)}, \qquad \mathbb{E}v_i^2 v_j^2 = \frac{1}{D(D+2)} \quad (i \neq j).$$

Diagonalizing E and expanding the square proves (18). □

LeWM's SIGReg statistic (Maes et al., 2026) uses knots $t_k = 3k/16$, $k = 0, \ldots, 16$, and the weights are $w_k = c_k(3/16)e^{-t_k^2/2}$, with $c_0 = c_{16} = 1$ and $c_k = 2$ otherwise. These are symmetry-doubled trapezoidal weights on $[0,3]$, multiplied by the Gaussian window. A sufficient condition for $r'_N(0) < 0$ is $N \geq 58$. For every positive knot,

$$1 - e^{-t_k^2/2} \geq \frac{t_k^2/2}{1 + t_k^2/2} \geq \frac{9}{521}, \qquad 1 - N(1 - e^{-t_k^2/2}) < 0. \tag{22}$$

The first inequality follows from $e^x \geq 1 + x$, and the final strict inequality follows from $58 \cdot 9 > 521$. Every nonzero summand of $r'_N(0)$ is therefore negative. The sufficient condition includes $N = 128$, the batch size at each time point in our training setup. A smaller batch threshold may also suffice.

Independent normalized Gaussian directions are uniform on the sphere, as required by the calculation. We compute the empirical characteristic function across examples at each time point and average the resulting statistics over time. Each training batch contains four time points with $N = 128$ examples each. If these slices have the same Gaussian marginal covariance C and independent examples within each slice, the expected average remains $\mathcal{R}_N(C)$ under dependence across slices. With different covariances C_t, the expectation is the average of $\mathcal{R}_N(C_t)$; dependence among examples within a slice requires a different finite-sample calculation.

B.3 Two-dimensional closed-form specialization

The general results in Section 3 do not require a two-dimensional state space. The following specialization is useful because both the isotropic and learned-target planners admit closed-form limiting actions and regrets.

Consider independent reset transitions

$$\begin{aligned} x' &= F_\eta x + gu + \xi, \\ x &\sim \mathcal{N}(0,\Sigma), \qquad u \sim \mathcal{N}(0,\tau^2), \qquad \xi \sim \mathcal{N}(0,W_\eta), \end{aligned} \tag{23}$$

where x, u, ξ are mutually independent. Set

$$\Sigma = \mathrm{diag}(1+\delta, 1-\delta), \qquad R_0 = \mathrm{diag}(\rho_1,\rho_2), \qquad g = (1,1)^\top,$$

with $0 < \delta < 1$, $\rho_1, \rho_2 > 0$, and $0 < \tau^2 < (1-\delta^2)/2$, and let

$$W_\eta = \eta \Sigma^{1/2} R_0 \Sigma^{1/2}.$$

The covariance-preserving transition matrix F_η is defined in Eq. (28) below. For sufficiently small $\eta > 0$, it is invertible and gives x' the same covariance Σ as x.

For evaluation, let $d = (1,-1)^\top$, take $x_0 = F_\eta^{-1} d$, set $x_g = 0$, and restrict $u \in [-1,1]$. The terminal conditional mean is

$$d + gu = (1+u, u-1)^\top.$$

We evaluate actions using squared Euclidean distance in state space:

$$J_*(u) = \mathbb{E}\,\|d + gu + \xi\|^2 = 2 + 2u^2 + \operatorname{tr} W_\eta,$$
$$u_* = 0, \qquad \operatorname{Regret}(u) = J_*(u) - J_*(0) = 2u^2. \tag{24}$$

The noise contributes the same additive term to every action and therefore does not affect their ranking.

We train linear encoders $A \in \mathbb{R}^{2\times 2}$, including singular encoders, together with predictors linear in (Ax, u) under

$$\mathcal{L}_B(A,p) = \underbrace{\tfrac{1}{2}\mathbb{E}\,\|p(Ax,u) - Ax'\|^2}_{\mathcal{L}_{\mathrm{pred}}(A,p)} + \lambda_B \mathcal{R}_N(A\Sigma A^\top), \qquad \lambda_B > 0. \tag{25}$$

The latent planner minimizes

$$J_{z,\eta}(u) = \|p(Ax_0, u) - Ax_g\|^2$$

over the same feasible interval.

Theorem 2 (Closed-form isotropic-target specialization)**.** *Fix the preceding control parameters and the statistic assumptions of Lemma 2, and let $\lambda_B > 0$. For all sufficiently small $\eta > 0$, Eq. (25) attains a global minimum over all linear encoders and predictors. Every minimizing encoder is invertible, and every minimizing predictor recovers the exact encoded conditional mean,*

$$p(Ax, u) = A(F_\eta x + gu).$$

All global minimizers induce the same

$$K_\eta = (A\Sigma^{1/2})^\top (A\Sigma^{1/2}) \longrightarrow qI,$$

and therefore

$$M_\eta = A^\top A = \Sigma^{-1/2} K_\eta \Sigma^{-1/2} \longrightarrow q\Sigma^{-1}. \tag{26}$$

If u_B minimizes $J_{z,\eta}$ on $[-1,1]$, then

$$\mathcal{L}_{\mathrm{pred}}(A,p) = \tfrac{\eta}{2}\operatorname{tr}(K_\eta R_0) \longrightarrow 0, \qquad u_B \longrightarrow \delta, \qquad \operatorname{Regret}(u_B) \longrightarrow 2\delta^2 > 0. \tag{27}$$

For all sufficiently small positive η, the latent cost strictly prefers u_B to the task-optimal action 0, whereas the task cost strictly prefers 0 to u_B.

Equation (26) makes the mechanism explicit: the lower-variance second state coordinate receives greater weight under Σ^{-1}, so the latent planner trades the two terminal errors differently from the equal-weight task cost. This specialization is not needed for the general separation in Section 3.3; it is used in Appendix C to obtain a closed-form dependence on the target anisotropy bound.

B.4 Transition construction and prediction reduction for the specialization

For the independent reset model in (23), fix $0 < \delta < 1$, $\rho_1, \rho_2 > 0$, and $0 < \tau^2 < (1-\delta^2)/2$. With the training-state covariance $\Sigma = \operatorname{diag}(1+\delta, 1-\delta)$, noise shape $R_0 = \operatorname{diag}(\rho_1, \rho_2)$, and $g = (1,1)^\top$, set

$$W_\eta = \eta \Sigma^{1/2} R_0 \Sigma^{1/2}, \qquad F_\eta = (\Sigma - \tau^2 gg^\top - W_\eta)^{1/2}\Sigma^{-1/2}. \tag{28}$$

The model is well defined for sufficiently small η. The rank-one positive-definiteness criterion gives

$$\begin{aligned} \Sigma - \tau^2 gg^\top \succ 0 \quad &\Longleftrightarrow \quad \tau^2 g^\top \Sigma^{-1} g < 1 \\ &\Longleftrightarrow \quad \tau^2 < \frac{1-\delta^2}{2}. \end{aligned} \tag{29}$$

Positive definiteness persists after subtracting W_η for small $\eta > 0$. The resulting F_η is invertible and satisfies

$$F_\eta \Sigma F_\eta^\top + \tau^2 gg^\top + W_\eta = \Sigma.$$

Thus x and x' have the same Gaussian marginal.

The same covariance identity gives a stationary trajectory construction. Initialize $x_0 \sim \mathcal{N}(0, \Sigma)$ and draw iid innovation pairs $(u_t, \xi_t) \sim \mathcal{N}(0, \tau^2) \otimes \mathcal{N}(0, W_\eta)$, with the pair sequence independent of x_0. Induction on the transition gives $x_t \sim \mathcal{N}(0, \Sigma)$ at every time. Each transition has the same one-step tuple law as (23), so N independent trajectories provide iid examples within each time slice.

For fixed encoder, predictor, and target parameters, the expected prediction loss averaged over any fixed finite set of times equals that of the reset construction. The same equality holds for the average of SIGReg statistics computed separately at each time slice.

For any encoder A, including singular A, independence and the zero mean of ξ yield

$$\begin{aligned}\mathcal{L}_{\mathrm{pred}}(A, p) &= \tfrac{1}{2}\mathbb{E}\,\|p(Ax, u) - AF_\eta x - Agu\|^2 + \tfrac{1}{2}\operatorname{tr}(AW_\eta A^\top) \\ &\geq \tfrac{\eta}{2}\operatorname{tr}(KR_0), \qquad B = A\Sigma^{1/2}, \quad K = B^\top B \succeq 0. \end{aligned} \tag{30}$$

For invertible A, equality is attained by the linear encoded conditional-mean predictor

$$p_A(z, u) = AF_\eta A^{-1} z + Agu. \tag{31}$$

Equality uniquely determines the predictor. The covariance of (Ax, u) is positive definite, so two linear predictions that agree almost surely have identical coefficient matrices.

The matrices $BB^\top$ and $B^\top B$ are orthogonally conjugate, including when B is singular. Orthogonal invariance of $\mathcal{R}_N$ therefore gives $\mathcal{R}_N(A\Sigma A^\top) = \mathcal{R}_N(K)$. The loss of every encoder and predictor pair is bounded below by the reduced objective

$$\mathcal{F}_\eta(K) = \tfrac{\eta}{2}\operatorname{tr}(KR_0) + \lambda_B \mathcal{R}_N(K), \qquad K \succeq 0. \tag{32}$$

Every positive-definite K attains this bound with $A = K^{1/2}\Sigma^{-1/2}$ and the predictor in (31). It remains to show that all reduced global minimizers are positive definite.

B.5 Proof of Theorem 2

Existence and uniform localization. For every fixed $\eta > 0$, $\mathcal{F}_\eta$ is continuous and coercive on the positive-semidefinite cone, since $R_0 \succ 0$ and $\mathcal{R}_N \geq 0$. It therefore attains its minimum. Comparing any minimizer K_η with qI gives

$$\frac{\eta}{2}\operatorname{tr}(K_\eta R_0) + \lambda_B\big[\mathcal{R}_N(K_\eta) - r_N(q)\big] \leq \frac{\eta q}{2}\operatorname{tr}(R_0). \tag{33}$$

Both terms on the left are nonnegative. In particular,

$$\operatorname{tr}(K_\eta R_0) \leq q\operatorname{tr}(R_0), \qquad 0 \leq \mathcal{R}_N(K_\eta) - r_N(q) \leq \frac{\eta q}{2\lambda_B}\operatorname{tr}(R_0).$$

The trace bound is uniform over $\eta > 0$ and all reduced global minimizers. Every cluster point as $\eta \downarrow 0$ minimizes $\mathcal{R}_N$, so Lemma 2 gives $K_\eta \to qI$. To verify uniform convergence over the minimizing sets, suppose a sequence of minimizers lies outside a fixed neighborhood of qI. The trace bound gives a convergent subsequence whose limit must be qI, a contradiction.

Uniqueness and the first-order perturbation. The Hessian in (18) is positive definite at qI and, by continuity, throughout a sufficiently small convex neighborhood. All global minimizers eventually lie in this neighborhood. Strict convexity therefore gives a unique positive-definite reduced minimizer K_η for small $\eta > 0$. Conjugation by $J = \operatorname{diag}(1, -1)$ preserves both terms of (32). Uniqueness implies $JK_\eta J = K_\eta$, so K_η is diagonal.

The stationarity equation is

$$\frac{\eta}{2}R_0 + \lambda_B \nabla\mathcal{R}_N(K_\eta) = 0.$$

In two dimensions, the Hessian operator at qI is

$$E \longmapsto \frac{c_N}{8}\big[\operatorname{tr}(E)I + 2E\big]. \tag{34}$$

It is invertible on the symmetric matrices. The implicit function theorem gives a smooth local solution, which coincides with the unique global minimizer for small positive η. Differentiating at zero yields

$$K_0' = -\frac{1}{\lambda_B c_N}\left(2R_0 - \bar{\rho} I\right), \qquad \bar{\rho} = \frac{\rho_1 + \rho_2}{2}. \tag{35}$$

Consequently,

$$K_\eta = qI - \frac{\eta}{\lambda_B c_N}\left(2R_0 - \tfrac{1}{2}\operatorname{tr}(R_0) I\right) + O(\eta^2). \tag{36}$$

Write $K_\eta = \operatorname{diag}(c_1, c_2)$ and $\beta_\eta = (c_1 - c_2)/(c_1 + c_2)$. Then

$$\beta_\eta = \frac{\eta(\rho_2 - \rho_1)}{\lambda_B q c_N} + O(\eta^2). \tag{37}$$

For sufficiently small positive η, convergence gives $|\beta_\eta| < \delta$ and $c_i > q/2$. Under the additional ordering $\rho_1 < \rho_2$, the expansion also gives $\beta_\eta > 0$ and $c_1\rho_1 - c_2\rho_2 < 0$. If $\rho_1 = \rho_2$, the objective is invariant under every orthogonal conjugation, so uniqueness forces K_η to be a scalar matrix and $\beta_\eta = 0$.

The positive-definite reduced minimizer is attained by an invertible encoder and (31), so the full training objective has a global minimum. Every full minimizer must attain both the reduced minimum and the prediction bound. Failure to attain either would give a strictly larger objective value. All full global minimizers therefore induce the same K_η and recover the exact encoded conditional mean. The feature covariance $BB^\top$ has eigenvalues $c_1, c_2 > q/2$, proving the noncollapse bound.

Planning and feasible regret. At x_0, the encoded conditional mean is $A(d + gu)$. The latent planner therefore minimizes

$$\begin{aligned} J_{z,\eta}(u) &= (d + gu)^\top M_\eta (d + gu), \\ M_\eta &= A^\top A = \Sigma^{-1/2} K_\eta \Sigma^{-1/2} = \operatorname{diag}\left(\frac{c_1}{1+\delta}, \frac{c_2}{1-\delta}\right). \end{aligned} \tag{38}$$

Writing the diagonal entries as $m_1, m_2 > 0$, the objective is $m_1(1+u)^2 + m_2(u-1)^2$. Its unique unconstrained minimizer is

$$u_B = \frac{m_2 - m_1}{m_1 + m_2} = \frac{\delta - \beta_\eta}{1 - \delta\beta_\eta} \in (0, 1),$$

which is feasible for sufficiently small η. When $\rho_1 < \rho_2$, it lies in $(0, \delta)$. Equation (24) gives $\operatorname{Regret}(u_B) = 2u_B^2 \to 2\delta^2$. The attained prediction loss is $\eta \operatorname{tr}(K_\eta R_0)/2$ and tends to zero. Strict convexity and $u_B \neq 0$ imply $J_{z,\eta}(u_B) < J_{z,\eta}(0)$ and $J_*(0) < J_*(u_B)$. This proves the ranking reversal. □

The strict inequalities between u_B and 0 extend by continuity to sufficiently small feasible neighborhoods of those actions. A candidate pool containing actions from both neighborhoods therefore reverses their ordering under the latent and physical task costs. Finite-budget CEM performance also depends on candidate sampling and the search procedure.

B.6 General residuals and a fixed evaluation state

Let the terminal conditional-mean residual be $\tilde{d} = (d_1, d_2)^\top$, with goal zero, $g = (1, 1)^\top$, and task metric I. For a diagonal latent metric $M = \operatorname{diag}(m_1, m_2) \succ 0$, the unconstrained task and latent optima are

$$u_*^0 = -\frac{d_1 + d_2}{2}, \qquad u_z^0 = -\frac{m_1 d_1 + m_2 d_2}{m_1 + m_2}.$$

Subtracting and completing the square in the task cost gives

$$u_z^0 - u_*^0 = \frac{(m_2 - m_1)(d_1 - d_2)}{2(m_1 + m_2)}, \qquad J_*(u) - J_*(u_*^0) = 2(u - u_*^0)^2. \tag{39}$$

When both optima are interior to $[-1, 1]$, the resulting regret is

$$\operatorname{Regret}(u_z^0) = \frac{(d_1 - d_2)^2}{2}\left(\frac{m_2 - m_1}{m_1 + m_2}\right)^2.$$

For the baseline metric, this converges to $\delta^2(d_1 - d_2)^2/2$. For the favorable learned-target limit in Theorem 3, the corresponding limit is

$$\frac{(d_1 - d_2)^2}{2}\left(\frac{c-\kappa}{c+\kappa}\right)^2. \tag{40}$$

Assume that the task, baseline, and target limiting optima are interior and that $d_1 \neq d_2$. For each fixed residual satisfying these conditions, the target limiting regret is strictly lower than the baseline limit exactly when $1 < \kappa < c^2$. A uniform positive improvement margin over a set of residuals further requires a positive lower bound on $|d_1 - d_2|$.

For arbitrary residuals, let $\Pi(t) = \min\{1, \max\{-1, t\}\}$. The constrained optima are $u_* = \Pi(u_*^0)$ and $u_z = \Pi(u_z^0)$, and the exact regret is

$$\mathrm{Regret}(u_z) = 2\big[(u_z - u_*^0)^2 - (u_* - u_*^0)^2\big].$$

Distinct unconstrained optima can project to the same endpoint. For example, $(m_1, m_2) = (1, 2)$ and $(d_1, d_2) = (2, 3)$ give $u_*^0 = -5/2$ and $u_z^0 = -8/3$, both projecting to -1. The strictness statement therefore uses the interior assumption.

The original evaluation state $x_0 = F_\eta^{-1} d$ fixes the residual $d = (1, -1)^\top$ at every noise level. Alternatively, take the fixed state $\bar{x}_0 = F_0^{-1} d$, where $F_0 = (\Sigma - \tau^2 g g^\top)^{1/2}\Sigma^{-1/2}$. Continuity of the positive-definite square root gives $F_\eta \bar{x}_0 \to d$. The projected quadratic argmin is continuous in the residual and the positive-definite metric, so the actions and regrets have the limits in Theorems 2 and 3. At positive noise, regret in this variant is measured relative to its task-optimal action $\Pi(-g^\top F_\eta \bar{x}_0/2)$, which tends to zero.

C Closed-Form Analysis of the Learnable Gaussian Target

This appendix analyzes ΛReg in the two-dimensional specialization of Appendix B.3. The purpose of the specialization is to expose the dependence of the selected action and regret on the anisotropy bound κ in closed form. The general learned-target metric is treated in Appendix E.3, and arbitrary-dimensional spectrum selection is treated in Appendix D.1.

Assume the two-dimensional model and evaluation setup of Theorem 2. For a diagonal target $\Lambda \in \mathcal{T}_{2,\kappa}$, the expected learned-target objective is

$$\mathcal{L}_T(A, p, \Lambda) = \mathcal{L}_{\mathrm{pred}}(A, p) + \lambda \mathcal{R}_N\left(\Lambda^{-1/2} A \Sigma A^\top \Lambda^{-1/2}\right), \qquad \Lambda \in \mathcal{T}_{2,\kappa}. \tag{41}$$

We focus first on the ordering $\rho_1 < \rho_2$, under which predictive training favors a variance reallocation that counteracts the inverse-covariance weighting of the isotropic solution. The reverse ordering is treated in Appendix C.4.

Theorem 3 (Selected geometry and planning regret in the two-dimensional specialization)**.** *Under the assumptions of Theorem 2, suppose $\rho_1 < \rho_2$, fix $\kappa > 1$ and $\lambda > 0$, and define*

$$c = \mathrm{cond}(\Sigma) = \frac{1+\delta}{1-\delta}, \qquad b = \frac{\kappa - 1}{\kappa + 1}.$$

For all sufficiently small $\eta > 0$, Eq. (41) attains a global minimum. Every minimizing encoder is invertible, and every minimizing predictor recovers the exact encoded conditional mean. For each global minimizer, let

$$K_{\eta,T} = (A\Sigma^{1/2})^\top (A\Sigma^{1/2}),$$

and let u_T minimize the learned latent cost over $[-1, 1]$. Then, as $\eta \downarrow 0$,

$$K_{\eta,T} \longrightarrow q\,\mathrm{diag}(1+b, 1-b), \qquad u_T \longrightarrow \frac{c-\kappa}{c+\kappa}, \qquad \mathrm{Regret}(u_T) \longrightarrow 2\left(\frac{c-\kappa}{c+\kappa}\right)^2. \tag{42}$$

The limiting regret is zero at $\kappa = c$. Relative to the isotropic baseline at any fixed $\lambda_B > 0$, it is strictly lower if and only if $1 < \kappa < c^2$, equal at $\kappa = c^2$, and strictly higher for $\kappa > c^2$.

At $\kappa = c$, the limiting state-space metric is qI, so the learned target exactly cancels the inverse-covariance weighting of the isotropic solution in this specialization. The improvement is not monotone in κ: excessive anisotropy eventually increases regret. This specialization makes the effect of the anisotropy bound explicit; Appendix D characterizes the target spectra selected in higher dimensions.

C.1 Proof of Theorem 3

The prediction-loss normalization is the same as in Appendix B.3, so the reduced prediction bound in Eq. (30) applies directly.

Reduction with a free encoder orientation. For independent samples $x_b \sim \mathcal{N}(0, \Sigma)$ and $Z = (Ax_1, \dots, Ax_N)$, the ΛReg statistic satisfies

$$\mathbb{E}_{Z,\boldsymbol{\omega}} \widehat{\mathcal{R}}_{N,\Lambda}(Z; \boldsymbol{\omega}) = \mathcal{R}_N \left(\Lambda^{-1/2} A \Sigma A^\top \Lambda^{-1/2} \right).$$

This gives the expected objective in (41). To separate the target spectrum from the encoder orientation, let

$$\mathcal{S}_\kappa = \{ L \succ 0 : \operatorname{tr} L = 2,\ \operatorname{cond}(L) \le \kappa \}$$

be the family of orthogonal rotations of the diagonal target covariances. Write $B = A\Sigma^{1/2} = OK^{1/2}$ by polar decomposition, with $K = B^\top B$ and an orthogonal extension when B is singular. Set $L = O^\top \Lambda O \in \mathcal{S}_\kappa$. Since $\Lambda^{-1/2} = OL^{-1/2}O^\top$,

$$\begin{aligned} \Lambda^{-1/2} A \Sigma A^\top \Lambda^{-1/2} &= \Lambda^{-1/2} B B^\top \Lambda^{-1/2} \\ &= O\big(L^{-1/2} K L^{-1/2}\big) O^\top. \end{aligned}$$

Orthogonal invariance of $\mathcal{R}_N$ and the prediction lower bound in (30) give the reduced lower bound

$$\begin{gathered} \mathcal{G}_\eta(K, L) = \frac{\eta}{2} \operatorname{tr}(K R_0) + \lambda \mathcal{R}_N\big(L^{-1/2} K L^{-1/2}\big), \\ K \succeq 0, \qquad L \in \mathcal{S}_\kappa. \end{gathered} \tag{43}$$

Conversely, every admissible pair (K, L) with $K \succ 0$ and $L \in \mathcal{S}_\kappa$ is realizable. Choose an orthogonal O such that $OLO^\top = \Lambda$ is diagonal, set $A = OK^{1/2}\Sigma^{-1/2}$, and use the predictor in (31). This pair attains the reduced bound. The encoder's orthogonal freedom realizes every orientation in $\mathcal{S}_\kappa$ with a diagonal output target.

The unique limiting allocation. For $b = (\kappa - 1)/(\kappa + 1)$, every $L \in \mathcal{S}_\kappa$ has the form

$$L = I + \begin{pmatrix} s & t \\ t & -s \end{pmatrix}, \qquad s^2 + t^2 \le b^2. \tag{44}$$

Its eigenvalues are $1 \pm \sqrt{s^2 + t^2}$, whose ratio is at most κ precisely on this disk. Moreover,

$$\operatorname{tr}(L R_0) = \rho_1 + \rho_2 - (\rho_2 - \rho_1) s.$$

Since $\rho_2 > \rho_1$, the unique minimum over the disk occurs at $s = b, t = 0$, giving $L_* = \operatorname{diag}(1 + b, 1 - b)$.

For fixed $\eta > 0$, $\mathcal{G}_\eta$ attains its minimum. The set $\mathcal{S}_\kappa$ is compact with a positive lower eigenvalue bound, and the nonnegative prediction term bounds $\operatorname{tr}(KR_0)$ on every sublevel set. Comparing a minimizing pair $(K_{\eta,T}, L_\eta)$ with (qL_*, L_*) gives

$$\begin{gathered} \frac{\eta}{2} \operatorname{tr}(K_{\eta,T} R_0) + \lambda \big[\mathcal{R}_N(C_\eta) - r_N(q) \big] \le \frac{\eta q}{2} \operatorname{tr}(L_* R_0), \\ C_\eta = L_\eta^{-1/2} K_{\eta,T} L_\eta^{-1/2}. \end{gathered} \tag{45}$$

Both terms on the left are nonnegative, so $\operatorname{tr}(K_{\eta,T} R_0) \le q \operatorname{tr}(L_* R_0)$ uniformly over all minimizers. The lower eigenvalue bound on L_η also bounds C_η. The regularizer gap tends to zero, and every cluster point of C_η equals qI by Lemma 2. Hence $C_\eta \to qI$ uniformly over all minimizers.

For any subsequence with $L_\eta \to \bar{L}$,

$$K_{\eta,T} = L_\eta^{1/2} C_\eta L_\eta^{1/2} \longrightarrow q\bar{L}.$$

The trace bound from (45) implies $\mathrm{tr}(\bar{L}R_0) \le \mathrm{tr}(L_* R_0)$. Uniqueness of the minimizer over the disk in (44) forces $\bar{L} = L_*$. Thus

$$L_\eta \longrightarrow L_*, \qquad K_{\eta,T} \longrightarrow qL_*$$

uniformly over the minimizing sets. Every minimizing $K_{\eta,T}$ is therefore positive definite for sufficiently small η, with

$$\lambda_{\min}(K_{\eta,T}) > \frac{q(1-b)}{2}.$$

The positive-definite reduced minimizers can all be realized by an encoder, predictor, and diagonal target, so the full objective attains the reduced minimum. As in Appendix B.5, every full global minimizer must attain both the reduced minimum and the prediction lower bound. Its encoder is invertible and its predictor is p_A from (31). The preceding uniform limits therefore hold for the Gram matrices of all full global minimizers.

The limiting action and the improvement range. For a positive-definite state metric M, the unconstrained minimizing action is

$$u(M) = -\frac{g^\top M d}{g^\top M g},$$

which is continuous in M. The limiting state metric is

$$q\,\mathrm{diag}\left(\frac{1+b}{1+\delta}, \frac{1-b}{1-\delta}\right),$$

so

$$u_T \longrightarrow \frac{\delta - b}{1 - \delta b} = \frac{c-\kappa}{c+\kappa}, \qquad c = \frac{1+\delta}{1-\delta}.$$

This limit lies strictly in $(-1, 1)$ for $\delta, b \in (0, 1)$. The unconstrained action is therefore eventually feasible, uniformly over the minimizing sets. Positive definiteness makes the quadratic cost strictly convex, so the minimizing action is unique. The task cost in (24) gives the regret limit in (42).

Relative to any baseline with fixed $\lambda_B > 0$, the limiting regret improvement is

$$\begin{aligned} \Delta_{\mathrm{imp}}(c, \kappa) &= 2\left(\frac{c-1}{c+1}\right)^2 - 2\left(\frac{c-\kappa}{c+\kappa}\right)^2 \\ &= \frac{8c(c^2-\kappa)(\kappa-1)}{(c+1)^2(c+\kappa)^2}. \end{aligned} \tag{46}$$

Since $c > 1$ and $\kappa > 1$, the sign of Δ_{imp} is the sign of $c^2 - \kappa$. Limiting regret is therefore lower than the baseline for $1 < \kappa < c^2$, equal at $\kappa = c^2$, and higher for $\kappa > c^2$. It vanishes at $\kappa = c$. Uniform convergence over the minimizing sets of both objectives gives uniform convergence of their regret difference to Δ_{imp}. When $\Delta_{\mathrm{imp}} > 0$, choose η sufficiently small that the deviation is less than $\Delta_{\mathrm{imp}}/2$. The finite-noise improvement then exceeds $\Delta_{\mathrm{imp}}/2$. At $\kappa = c^2$, higher-order terms determine the finite-noise sign. □

The limiting regret depends on $|\log \kappa - \log c|$, so it is symmetric about $\kappa = c$ on the logarithmic scale.

C.2 A TARGET-COUPLED RESCALING PATH

To examine how target anisotropy trades prediction loss against planning regret, we construct a coupled rescaling of an isotropic-target optimum. The path preserves the standardized features seen by ΛReg while changing the encoder's representation geometry. Let c_1, c_2 be the baseline Gram eigenvalues from Appendix B.5. We use the canonical baseline encoder

$$A_B = \mathrm{diag}(\sqrt{c_1}, \sqrt{c_2})\Sigma^{-1/2}$$

and predictor $p_B = p_{A_B}$ from (31) as the starting pair. Define the encoder, predictor, and target along the path by

$$\begin{aligned} &\Lambda_h = \mathrm{diag}(1+h, 1-h), \qquad S_h = \Lambda_h^{1/2}, \qquad 0 \le h < 1, \\ &A_h = S_h A_B, \qquad p_h(z, u) = S_h p_B(S_h^{-1} z, u). \end{aligned} \tag{47}$$

Each target has trace two. Under a fixed condition-number bound κ, this path belongs to $\mathcal{T}_{2,\kappa}$ precisely for $0 \le h \le b = (\kappa-1)/(\kappa+1)$. The full interval $0 \le h < 1$ describes the effect of varying the allowed anisotropy.

Since $\Lambda_h^{-1/2} A_h = A_B$, ΛReg is constant along this path: for every batch X and common set of projection directions $\boldsymbol{\omega}$,

$$\widehat{\mathcal{R}}_{N,\Lambda_h}(A_h X; \boldsymbol{\omega}) = \widehat{\mathcal{R}}_N(A_B X; \boldsymbol{\omega}).$$

For the evaluation state and goal in Appendix B.3, let

$$u_h = \operatorname*{arg\,min}_{u \in [-1,1]} \left\| p_h(A_h x_0, u) - A_h x_g \right\|^2 .$$

Proposition 3 (A target-coupled rescaling path)**.** *Under the baseline setup of Theorem 2, fix $\lambda_B > 0$ and assume $\rho_1 < \rho_2$. For all sufficiently small $\eta > 0$ and every $0 \le h < 1$,*

$$\mathcal{L}_{\mathrm{pred}}(A_h, p_h) - \mathcal{L}_{\mathrm{pred}}(A_B, p_B) = \frac{\eta h}{2}(c_1\rho_1 - c_2\rho_2) < 0 \qquad (h > 0), \tag{48}$$

$$\mathrm{Regret}(u_h) = 2\left(\frac{u_B - h}{1 - u_B h}\right)^2 . \tag{49}$$

Along the full path, regret strictly decreases on $[0, u_B]$, reaches zero at $h = u_B$, and exceeds its baseline value when $h > 2u_B/(1+u_B^2)$. At every $h > 0$, the expected original isotropic-target regularizer is strictly larger than its baseline value. Its increase, weighted by λ_B, exceeds the decrease in expected prediction loss.

Increasing h continues to reduce latent prediction loss after regret reaches zero. Each predictor along the path recovers the encoded conditional mean, so the change in (48) comes from the encoder's weighting of process noise. Proposition 3 compares objectives and planning costs along this prescribed path in parameter space. Theorem 3 identifies the limiting geometry selected by the constrained joint objective.

Proof of Proposition 3. The baseline predictor p_B recovers the encoded conditional mean. The definitions of A_h and p_h apply S_h to its prediction and target, giving

$$\begin{aligned}\mathcal{L}_{\mathrm{pred}}(A_h, p_h) &= \frac{1}{2}\operatorname{tr}(A_h W_\eta A_h^\top) \\ &= \frac{\eta}{2}\big[(1+h)c_1\rho_1 + (1-h)c_2\rho_2\big].\end{aligned}$$

Subtracting the baseline loss proves (48). Appendix B.5 establishes $c_1\rho_1 - c_2\rho_2 < 0$ for sufficiently small positive η.

Write the baseline state-metric entries as $m_1 = c_1/(1+\delta)$ and $m_2 = c_2/(1-\delta)$, so that $u_B = (m_2 - m_1)/(m_1 + m_2)$. Along the path, these entries become $(1+h)m_1$ and $(1-h)m_2$. The minimizing action is therefore

$$u_h = \frac{(1-h)m_2 - (1+h)m_1}{(1+h)m_1 + (1-h)m_2} = \frac{u_B - h}{1 - u_B h}.$$

For $0 \le h < 1$, it lies strictly in $(-1, 1)$ and satisfies

$$\frac{du_h}{dh} = -\frac{1 - u_B^2}{(1 - u_B h)^2} < 0.$$

Thus u_h^2 strictly decreases until $h = u_B$ and strictly increases thereafter. Since $\mathrm{Regret}(u) = 2u^2$, this gives (49). Solving $u_h^2 < u_B^2$ yields

$$0 < h < \frac{2u_B}{1 + u_B^2},$$

with equality at $h = 0$ and at the upper threshold.

The unstandardized Gram matrix along the path is

$$K_h = \mathrm{diag}\big((1+h)c_1, (1-h)c_2\big) \neq K_\eta \qquad (h > 0).$$

Uniqueness of the minimizer of (32) implies

$$\begin{aligned} 0 &< \mathcal{F}_\eta(K_h) - \mathcal{F}_\eta(K_\eta) \\ &= \Delta\mathcal{L}_{\mathrm{pred}} + \lambda_B\big[\mathcal{R}_N(K_h) - \mathcal{R}_N(K_\eta)\big], \end{aligned}$$

where $\Delta\mathcal{L}_{\mathrm{pred}} = \mathcal{L}_{\mathrm{pred}}(A_h, p_h) - \mathcal{L}_{\mathrm{pred}}(A_B, p_B) < 0$. Consequently, the original isotropic-target regularizer is larger than its baseline value, and its weighted increase exceeds $-\Delta\mathcal{L}_{\mathrm{pred}}$. □

C.3 Spectral bounds for admissible targets

Corollary 1 (Spectral bounds for admissible targets)**.** *Let $D \geq 2$ and $\kappa \geq 1$. Every target $\Lambda = \mathrm{diag}(v_1, \ldots, v_D) \in \mathcal{T}_{D,\kappa}$ satisfies*

$$\begin{gathered} \frac{D}{1 + (D-1)\kappa} \leq v_i \leq \frac{D\kappa}{\kappa + D - 1}, \\ d_{\mathrm{eff}}(\Lambda) := \frac{(\mathrm{tr}\,\Lambda)^2}{\mathrm{tr}(\Lambda^2)} \geq \frac{4\kappa}{(\kappa+1)^2} D. \end{gathered} \tag{50}$$

A fixed condition-number bound gives a lower bound on target effective dimension; at $\kappa = 2$, the bound is $8D/9$. This bound quantifies variance concentration in the target covariance; the covariance of a trained neural representation also depends on the encoder. In two dimensions, trace two and the variance floor $v_i \geq 2/(\kappa+1)$ describe the same feasible spectra as the condition-number constraint.

Proof. Write $v_{\min} = \min_i v_i$ and $v_{\max} = \max_i v_i$. The trace constraint and $v_{\max} \leq \kappa v_{\min}$ imply

$$\begin{aligned} D = \sum_i v_i &\leq v_{\min} + (D-1)\kappa v_{\min}, \\ D = \sum_i v_i &\geq v_{\max} + (D-1) v_{\max}/\kappa. \end{aligned}$$

Thus $v_{\min} \geq D/[1 + (D-1)\kappa]$ and $v_{\max} \leq D\kappa/(\kappa + D - 1)$, giving the eigenvalue bounds. Also, $v_i \in [v_{\min}, \kappa v_{\min}]$ implies

$$(v_i - v_{\min})(\kappa v_{\min} - v_i) \geq 0, \qquad v_i^2 \leq (\kappa+1) v_{\min} v_i - \kappa v_{\min}^2.$$

Summing and completing the square yields

$$\sum_i v_i^2 \leq D\big[(\kappa+1)v_{\min} - \kappa v_{\min}^2\big] \leq \frac{D(\kappa+1)^2}{4\kappa}.$$

Since $\sum_i v_i = D$, the effective-dimension bound follows. For $\kappa > 1$, equality requires all eigenvalues to be in $\{v_{\min}, \kappa v_{\min}\}$ and $v_{\min} = (\kappa+1)/(2\kappa)$. The trace then requires exactly $D/(\kappa+1)$ upper eigenvalues. Thus equality holds precisely when this number is an integer and the spectrum has that multiplicity. For $D = 2$ and trace two, the eigenvalues are $v_{\min}$ and $2 - v_{\min}$, so

$$\mathrm{cond}(\Lambda) \leq \kappa \quad \Longleftrightarrow \quad \frac{2 - v_{\min}}{v_{\min}} \leq \kappa \quad \Longleftrightarrow \quad v_{\min} \geq \frac{2}{\kappa+1}.$$

□

The logit parameterization in Section 4.2 covers every admissible target. For any $\Lambda = \mathrm{diag}(v_1, \ldots, v_D) \in \mathcal{T}_{D,\kappa}$, let $v_{\min} = \min_i v_i$ and $v_{\max} = \max_i v_i$. Choose $\alpha_i = \log v_i -$

$(\log v_{\max} + \log v_{\min})/2$. Since $\log v_{\max} - \log v_{\min} \le \log \kappa$, these logits lie in $[-\frac{1}{2}\log\kappa, \frac{1}{2}\log\kappa]$. The common shift cancels in the normalized exponential, and $\sum_i v_i = D$ gives $De^{\alpha_i}/\sum_j e^{\alpha_j} = v_i$.

C.4 Scope of the analysis

Theorem 2 holds for any $\rho_1, \rho_2 > 0$. Theorem 3 assumes $\rho_1 < \rho_2$: minimizing $\mathrm{tr}(LR_0)$ then assigns the larger target variance to the direction with larger state variance. For $\rho_1 > \rho_2$, the trace over the disk in (44) is minimized at $s = -b, t = 0$. The same compactness and localization argument gives

$$K_{\eta,T} \longrightarrow q\,\mathrm{diag}(1-b, 1+b), \qquad u_T \longrightarrow \frac{\delta+b}{1+\delta b} = \frac{c\kappa - 1}{c\kappa + 1}.$$

Since

$$\frac{\delta+b}{1+\delta b} - \delta = \frac{b(1-\delta^2)}{1+\delta b} > 0,$$

every $\kappa > 1$ gives strictly higher limiting regret than the baseline. If $\rho_1 = \rho_2$, all target orientations tie in the leading trace term.

Two-dimensional target spectrum. In two dimensions, the target spectrum is fully determined by κ up to permutation. In higher dimensions, Proposition 4 additionally selects the multiplicity of the larger target variance as a function of the predictive structure.

For the same control family with diagonal task metric $Q = \mathrm{diag}(\gamma_1, \gamma_2) \succ 0$, set $\kappa_Q = c\gamma_1/\gamma_2$. Under $\rho_1 < \rho_2$, if $\kappa_Q > 1$, choosing $\kappa = \kappa_Q$ in (42) gives

$$M_{\eta,T} \longrightarrow \frac{2q}{\mathrm{tr}(\Sigma Q)} Q.$$

Indeed, the limiting target allocation is $2\Sigma Q/\,\mathrm{tr}(\Sigma Q)$. The planner therefore attains zero limiting regret for the Q-weighted task cost. The case $Q = I$ recovers $\kappa_Q = c$. This calibration uses the state covariance and task metric. Under a physical-state coordinate change $\bar{x} = Tx$, the same physical cost has metric $T^{-\top} Q T^{-1}$.

The asymptotic comparisons above hold the regularization weights fixed and strictly positive as $\eta \downarrow 0$. Allowing the regularization weights to scale with η defines a different asymptotic regime.

D Prediction-Driven Target Spectrum Selection

This appendix characterizes the target spectrum selected by the joint prediction–regularization objective. We first establish its structure in arbitrary dimension and identify the role of the state-whitened noise covariance. We then give two three-dimensional training distributions that select different variance allocations under the same trace and condition-number constraints. As in Appendix C, A and p denote the encoder and predictor in the original latent coordinates. We use the prediction-loss normalization $\frac{1}{2}\mathbb{E}\,\|p(Ax,u) - Ax'\|^2$ throughout. For dimension D, multiplying the coordinate-averaged implementation objective by $D/2$ gives this convention with a correspondingly rescaled positive regularization weight; this does not change the global minimizers or planning costs.

D.1 Spectrum selection in arbitrary dimension

Target learning selects both the variance values and the number of directions assigned to each value. We characterize this selection in arbitrary dimension before applying it to the two control models. For $D \ge 2$, $\kappa > 1$, and a positive-definite state-whitened noise covariance R, define

$$\begin{aligned} \mathcal{S}_{D,\kappa} &= \{L \succ 0 : \mathrm{tr}\, L = D,\ \mathrm{cond}(L) \le \kappa\}, \\ \mathcal{G}_\eta(K, L) &= \tfrac{\eta}{2}\,\mathrm{tr}(KR) + \lambda \mathcal{R}_N(L^{-1/2} K L^{-1/2}), \qquad K \succeq 0, \quad L \in \mathcal{S}_{D,\kappa}. \end{aligned} \tag{51}$$

Let $0 < r_1 \le \cdots \le r_D$ be the eigenvalues of R, and set

$$a_m = \frac{D}{D + m(\kappa - 1)}, \qquad \phi_m = a_m \left[\mathrm{tr}\, R + (\kappa - 1)\sum_{i=1}^{m} r_i\right], \quad 1 \le m < D. \tag{52}$$

Here a_m and κa_m are the lower and upper target variances, and ϕ_m is the trace cost $\mathrm{tr}(LR)$ when the upper variance is assigned to the m lowest-noise directions.

Proposition 4 (Structure of optimal target spectra)**.** *Assume the statistic conditions of Lemma 2, and fix $D \geq 2$, $R \succ 0$, $\kappa > 1$, and $\lambda > 0$. For all sufficiently small $\eta > 0$, every global minimizer of* (51) *has a target geometry of the form*

$$L_\eta = a_m\big[I + (\kappa - 1)P\big], \qquad P = P^\top = P^2, \quad \operatorname{rank} P = m, \quad 1 \leq m < D.$$

Its spectrum has two values, κa_m and a_m, and its condition number equals κ. Moreover, $L_\eta^{-1/2} K_\eta L_\eta^{-1/2} \to qI$ uniformly over the global minimizers as $\eta \downarrow 0$.
For every sequence $\eta_n \downarrow 0$ and choice of global minimizers (K_n, L_n), each accumulation point of L_n minimizes $\operatorname{tr}(LR)$ on $\mathcal{S}_{D,\kappa}$. Its multiplicity m minimizes (52)*, and its larger-variance eigenspace is a span of m lowest-noise eigendirections of R. If ϕ_m has a unique minimizing index m_* and $r_{m_*} < r_{m_*+1}$, the target geometries converge to the unique limit in any matrix norm, uniformly over the global minimizers.*

Proof. The set $\mathcal{S}_{D,\kappa}$ is compact by the eigenvalue bounds in (50). It is convex because $\lambda_{\max}(L) - \kappa\lambda_{\min}(L)$ is a convex function of the symmetric matrix L. For each fixed $\eta > 0$, the joint objective is coercive in K uniformly over L. Compactness of the target set therefore gives a global minimum. For an effective noise covariance $Y \succ 0$, define the value obtained after optimizing the standardized feature covariance:

$$\Phi_\eta(Y) = \min_{C \succeq 0} \left\{ \lambda \mathcal{R}_N(C) + \tfrac{\eta}{2} \operatorname{tr}(CY) \right\}.$$

The minimum exists by continuity and coercivity. Substituting $K = L^{1/2} C L^{1/2}$ in (51) gives the coefficient $L^{1/2} R L^{1/2}$ in the trace. This matrix and $R^{1/2} L R^{1/2}$ have the same eigenvalues. Orthogonal invariance of $\mathcal{R}_N$ therefore gives

$$\min_{K \succeq 0} \mathcal{G}_\eta(K, L) = \Phi_\eta(R^{1/2} L R^{1/2}). \tag{53}$$

The matrices $Y = R^{1/2} L R^{1/2}$ range over a compact convex set with a uniform positive lower eigenvalue bound. Comparing any minimizer C in the definition of $\Phi_\eta(Y)$ with qI gives

$$\tfrac{\eta}{2} \operatorname{tr}(CY) + \lambda[\mathcal{R}_N(C) - \mathcal{R}_N(qI)] \leq \tfrac{\eta q}{2} \operatorname{tr} Y.$$

Both left-hand terms are nonnegative. This bounds C uniformly and makes its regularizer gap tend uniformly to zero. By compactness and Lemma 2, all such minimizers approach qI uniformly and are positive definite for sufficiently small $\eta > 0$. Orthogonal conjugacy transfers this convergence to the standardized covariance $L^{-1/2} K L^{-1/2}$ in the original reduced objective.

As an infimum of affine functions of Y, Φ_η is concave. It is strictly concave on the displayed set for small positive η. Indeed, equality in the concavity inequality for distinct Y_0, Y_1 would force a minimizer at their proper convex combination to minimize both endpoint objectives. This minimizer is interior, so its two stationarity equations give

$$\lambda \nabla_C \mathcal{R}_N(C) + \tfrac{\eta}{2} Y_i = 0, \qquad i = 0, 1,$$

contradicting $Y_0 \neq Y_1$. The map $L \mapsto R^{1/2} L R^{1/2}$ is linear and injective. Hence (53) is strictly concave in L, and every minimizing L is an extreme point of $\mathcal{S}_{D,\kappa}$.

We now characterize those extreme points. If $\operatorname{cond}(L) < \kappa$, small traceless symmetric perturbations of both signs remain feasible, so L is not extreme. Suppose $\operatorname{cond}(L) = \kappa$, with minimum eigenvalue a. If any eigenvalue lies strictly between a and κa, perturb all upper endpoint eigenvalues by $\kappa\epsilon$ and all lower endpoint eigenvalues by ϵ. Distribute the opposite total perturbation among the intermediate eigenvalues to preserve the trace. Both signs of sufficiently small ϵ remain feasible. Such an L is again nonextreme.

The remaining matrices have the stated form $L = a_m[I + (\kappa - 1)P]$. To show that each is extreme, suppose $L = tX + (1 - t)Y$ with $0 < t < 1$ and $X, Y \in \mathcal{S}_{D,\kappa}$. For unit vectors $e_+ \in \operatorname{ran} P$ and $e_- \in \ker P$,

$$e_+^\top X e_+ \leq \lambda_{\max}(X) \leq \kappa \lambda_{\min}(X) \leq \kappa e_-^\top X e_-,$$

and the same inequalities hold for Y. Since $e_+^\top L e_+ = \kappa e_-^\top L e_-$, equality holds throughout both chains. Thus the two subspaces are respectively maximal and minimal eigenspaces of X and Y. The trace condition forces $X = Y = L$.

For any trace minimizer L_*, comparison of a global minimizer (K_η, L_η) with (qL_*, L_*) gives $\operatorname{tr}(K_\eta R) \le q \operatorname{tr}(L_* R)$. Along any convergent subsequence $L_\eta \to \bar{L}$, the uniform covariance limit gives $K_\eta \to q\bar{L}$. Hence $\bar{L}$ minimizes $\operatorname{tr}(LR)$. The extreme-point set is a finite union of compact projection orbits, so $\bar{L}$ has the same two-value spectrum. The rearrangement inequality places the larger eigenvalues along the m smallest eigenvalues of R, giving ϕ_m. A unique minimizing index and the corresponding eigengap make the spectral projector, and therefore $\bar{L}$, unique. □

For a square linear encoder with training-state covariance $\Sigma \succ 0$, write $B_\eta = A_\eta \Sigma^{1/2} = O_\eta K_\eta^{1/2}$ and $L_\eta = O_\eta^\top \Lambda_\eta O_\eta$, where O_η is orthogonal. This rotation transfers the spectrum of L_η to the diagonal output target Λ_η. At the global optima in the proposition, $K_\eta - qL_\eta \to 0$, so the feature covariance satisfies $A_\eta \Sigma A_\eta^\top - q\Lambda_\eta \to 0$. Its finite-noise spectrum may have more than two distinct eigenvalues. The selector ϕ_m depends on the relative magnitudes of the noise eigenvalues.

Selection among tied spectra. For fixed D, $R \succ 0$, $\lambda > 0$, and $\kappa > 1$, uniform localization and the positive-definite Hessian give a common smooth minimizer branch in ηY by the implicit-function theorem. Expanding the optimum value gives, uniformly for $L \in \mathcal{S}_{D,\kappa}$,

$$\begin{aligned}\Phi_\eta(Y) = \lambda \mathcal{R}_N(qI) + \frac{\eta q}{2} \operatorname{tr} Y \\ - \frac{\eta^2 D}{16 \lambda c_N} \left[(D+2) \operatorname{tr}(Y^2) - (\operatorname{tr} Y)^2 \right] + O(\eta^3),\end{aligned} \tag{54}$$

where $Y = R^{1/2} L R^{1/2}$ and $c_N = r_N''(q) > 0$.

Write $\phi(L) = \operatorname{tr}(LR)$ and $\psi(L) = \operatorname{tr}((LR)^2) = \operatorname{tr}(Y^2)$. Every target limit $\bar{L}$ minimizes ϕ. Compare a convergent sequence of global minimizers with any minimizer L_* of ϕ and divide the value inequality by η^2. The nonnegative first-order gap and uniform remainder give $\psi(\bar{L}) \ge \psi(L_*)$. Thus its two-level rank maximizes

$$\psi_m = a_m^2 \left[\kappa^2 \sum_{i=1}^{m} r_i^2 + \sum_{i=m+1}^{D} r_i^2 \right] \tag{55}$$

among indices tied for the smallest ϕ_m. A unique maximizing index m and $r_m < r_{m+1}$ give a unique target limit. Remaining ties in ψ_m depend on higher-order terms. For $D = 3$, $\kappa = 2$, and $R = \operatorname{diag}(1, 2, 4)$, $\phi_1 = \phi_2 = 6$ and $\psi_1 = 27/2 > 324/25 = \psi_2$, so every optimum at sufficiently small positive noise uses the $m = 1$ branch.

D.2 Two training distributions with different selected spectra

Let $j \in \{1, 2\}$ index the models. Define

$$\begin{aligned} R_1 &= \operatorname{diag}(1, 10, 11), & \Sigma_1 &= \operatorname{diag}(3/2, 3/4, 3/4), \\ R_2 &= \operatorname{diag}(1, 2, 10), & \Sigma_2 &= \operatorname{diag}(6/5, 6/5, 3/5), \\ g_1 &= (1, 1, 0)^\top, & d_1 &= (1, -1, 0)^\top, \\ g_2 &= (0, 1, 1)^\top, & d_2 &= (0, 1, -1)^\top. \end{aligned} \tag{56}$$

Fix $\tau^2 = 1/10$ and set

$$W_{j,\eta} = \eta \Sigma_j^{1/2} R_j \Sigma_j^{1/2}, \qquad F_{j,\eta} = (\Sigma_j - \tau^2 g_j g_j^\top - W_{j,\eta})^{1/2} \Sigma_j^{-1/2}.$$

Since $g_1^\top \Sigma_1^{-1} g_1 = 2$ and $g_2^\top \Sigma_2^{-1} g_2 = 5/2$, the matrix inside the square root is positive definite for sufficiently small $\eta > 0$. Thus $F_{j,\eta}$ is invertible. Draw independent reset tuples

$$x \sim \mathcal{N}(0, \Sigma_j), \quad u \sim \mathcal{N}(0, \tau^2), \quad \xi \sim \mathcal{N}(0, W_{j,\eta}), \qquad x' = F_{j,\eta} x + g_j u + \xi.$$

The marginal covariance of x' is also Σ_j. We use independent tuples within each batch and the same finite-statistic assumptions as Lemma 2, which applies in dimension three.

We optimize an unrestricted encoder $A \in \mathbb{R}^{3\times 3}$ and a predictor p that is linear in (Ax, u). For any fixed $\lambda > 0$, their joint training objective with the target is

$$\mathcal{L}_{j,\eta}(A, p, \Lambda) = \tfrac{1}{2} \mathbb{E} \left\| p(Ax, u) - Ax' \right\|^2 + \lambda \mathcal{R}_N\big(\Lambda^{-1/2} A \Sigma_j A^\top \Lambda^{-1/2}\big), \quad \Lambda \in \mathcal{T}_{3,2}. \tag{57}$$

For evaluation, use $x_0 = F_{j,\eta}^{-1} d_j$, goal $x_g = 0$, and $u \in [-1, 1]$. Both physical costs satisfy

$$J_{*,j}(u) = \mathbb{E}\,\|d_j + g_j u + \xi\|^2 = 2 + 2u^2 + \operatorname{tr} W_{j,\eta}, \qquad \mathrm{Regret}_j(u) = 2u^2. \tag{58}$$

The task-optimal action is zero; the task cost is used only for evaluation.

Proposition 5 (Distribution-dependent selected spectra)**.** *For the two three-dimensional models above, fix $\kappa = 2$ and strictly positive regularization weights, held constant as $\eta \downarrow 0$. At global training optima, the target geometries in state-whitened coordinates converge to*

$$L_1^* = \Sigma_1 = \mathrm{diag}(3/2, 3/4, 3/4), \qquad L_2^* = \Sigma_2 = \mathrm{diag}(6/5, 6/5, 3/5).$$

The limiting target spectrum therefore assigns the larger variance to one direction in the first model and two directions in the second, despite the identical ordering of their noise eigenvalues. Every global optimum is noncollapsed for sufficiently small positive noise, and its state-space metric converges to qI. The exact Euclidean planner consequently has zero limiting task regret in each model. These limits hold uniformly over global minimizers as $\eta \downarrow 0$.

Proof. **Selected target geometries.**

Write $B = A\Sigma_j^{1/2} = OK^{1/2}$ and $L = O^\top \Lambda O$. Appendix C.1 gives

$$\begin{aligned} \mathcal{G}_{j,\eta}(K, L) &= \tfrac{\eta}{2} \operatorname{tr}(KR_j) + \lambda \mathcal{R}_N(L^{-1/2} K L^{-1/2}), \\ K \succeq 0, &\qquad L \in \mathcal{S} := \{L \succ 0 : \operatorname{tr} L = 3,\ \mathrm{cond}(L) \le 2\}. \end{aligned} \tag{59}$$

For $K \succ 0$, the polar construction realizes each pair with a diagonal output target, encoder orientation, and $p_A(z, u) = AF_{j,\eta} A^{-1} z + A g_j u$. The target domain $\mathcal{S}$ is compact, with eigenvalues uniformly bounded away from zero by (50).

Let L_* be the unique minimizer of $\operatorname{tr}(LR_j)$ over $\mathcal{S}$, identified below. For each $\eta > 0$, continuity and coercivity give a reduced global minimum. The comparison argument in the proof of Proposition 4 gives

$$\tfrac{\eta}{2} \operatorname{tr}(KR_j) + \lambda[\mathcal{R}_N(C) - \mathcal{R}_N(qI)] \le \tfrac{\eta q}{2} \operatorname{tr}(L_* R_j), \qquad C = L^{-1/2} K L^{-1/2}. \tag{60}$$

Since $R_j \succeq I$, the trace bound controls K. The uniform lower eigenvalue bound on L also bounds C. The regularizer gap is $O(\eta)$, so compactness and Lemma 2 give $C \to qI$ uniformly. Hence $K = L^{1/2} C L^{1/2} \to q\bar{L}$ along every subsequence $L \to \bar{L}$. Equation (60) and uniqueness force $\bar{L} = L_*$. Compactness then gives convergence uniformly over all reduced minimizers:

$$L \longrightarrow L_*, \qquad K \longrightarrow qL_*. \tag{61}$$

Thus $K \succ 0$ for every reduced minimizer at sufficiently small positive noise. Each positive-definite reduced minimizer is realizable by an encoder, predictor, and admissible target. Thus the full and reduced objectives have the same minimum value. Every full minimizer induces a reduced minimizer and attains the conditional-mean prediction bound. Its feature covariance shares K's eigenvalues and is noncollapsed.

We minimize $\operatorname{tr}(LR_j)$ over both the target spectrum and its orientation. Each R_j has strictly increasing diagonal entries. For fixed eigenvalues of L, the trace is minimized by aligning its largest eigenvalue with the smallest noise entry. To see this, expand $\operatorname{tr}(\mathrm{diag}(r) V \,\mathrm{diag}(v) V^\top)$ as $\sum_{i,k} r_i v_k V_{ik}^2$. The squared entries form a doubly stochastic matrix. Minimizing the linear expression over such matrices admits a permutation minimizer, and the rearrangement inequality orders the eigenvalues oppositely. Strict ordering of the noise entries makes the minimizing matrix unique. Rotations within a repeated target eigenspace leave that matrix unchanged.

It remains to optimize $v_1 \ge v_2 \ge v_3$ subject to $\sum_i v_i = 3$ and $v_1 \le 2v_3$. This feasible region is the triangle with vertices $(1, 1, 1)$, $(3/2, 3/4, 3/4)$, and $(6/5, 6/5, 3/5)$.

The corresponding trace values are

Spectrum	$\operatorname{tr}(LR_1)$	$\operatorname{tr}(LR_2)$
$(1, 1, 1)$	22	13
$(3/2, 3/4, 3/4)$	$69/4$	$21/2$
$(6/5, 6/5, 3/5)$	$99/5$	$48/5$

Each column has a unique minimizing vertex. Thus $L^*_{1,\mathrm{learn}} = \Sigma_1$ and $L^*_{2,\mathrm{learn}} = \Sigma_2$. Although the two noise spectra have the same ordering, their relative magnitudes select different numbers of larger target eigenvalues.

Planning regret. Since $L^*_{j,\mathrm{learn}} = \Sigma_j$,

$$A^\top A = \Sigma_j^{-1/2} K \Sigma_j^{-1/2} \longrightarrow qI.$$

The Euclidean planner minimizes $(d_j + g_j u)^\top A^\top A(d_j + g_j u)$. Its unconstrained action converges to zero and hence is feasible for sufficiently small noise. The limiting task regret is zero in both models.

□

The construction sets $\Sigma_j = L^*_{j,\mathrm{learn}}$, yielding the full limiting state metric qI. In general, agreement with Euclidean state costs over all residuals requires L_* proportional to Σ; under Proposition 4, the state covariance must therefore have exactly two distinct eigenvalues with ratio κ, up to an overall scale. Agreement on a restricted feasible residual set can hold under weaker conditions.

E General-Dimensional Geometry, Finite-Horizon Planning, and Nonlinear Stability

This appendix proves the arbitrary-dimensional joint-optimum result used in Section 3.2, gives the explicit finite-horizon construction underlying Theorem 1, and records the corresponding learned-target and nonlinear extensions. Throughout, $\|\cdot\|$ denotes the Euclidean norm for vectors and the operator norm for matrices. Matrix square roots are symmetric positive-semidefinite roots. All limits fix the dimension, horizon, batch size, quadrature, action statistics, and positive regularization weights.

E.1 Standing regularizer assumptions

We use the expected finite-batch statistic $\mathcal{R}_N$ of Appendix B.2 in dimension D. Specifically, $N > 2$, the quadrature has finitely many nonnegative knots t_k and weights w_k, at least one $w_k t_k$ is positive, projection directions are uniform on $\mathbb{S}^{D-1}$, and $r'_N(0) < 0$ in Eq. (17). Lemma 2 and its proof show that $\mathcal{R}_N$ is continuous, nonnegative, and invariant under orthogonal conjugation on the positive-semidefinite cone. Its unique minimum is qI_D, where $q \in (0, 1)$, and it is smooth near this minimum. With $c_N = r''_N(q) > 0$, Eq. (18) gives

$$D^2\mathcal{R}_N(qI_D)[E, E] = \frac{c_N}{D(D+2)}\left[(\operatorname{tr} E)^2 + 2\operatorname{tr}(E^2)\right]$$

for every symmetric E. In particular, the Hessian is positive definite on symmetric matrices. These are the regularizer properties used below; the finite-statistic proof need not be repeated.

E.2 General Gaussian training model and isotropic joint optimum

For the proof of Proposition 1, consider the Gaussian training model

$$\begin{aligned} x' &= F_\eta x + Ga + \xi, \\ x &\sim \mathcal{N}(0, \Sigma), \qquad a \sim \mathcal{N}(0, \Gamma), \qquad \xi \sim \mathcal{N}(0, \eta W), \end{aligned} \tag{62}$$

where x, a, and ξ are mutually independent, $\Sigma, W \in \mathbb{R}^{D\times D}$ and $\Gamma \in \mathbb{R}^{d_a \times d_a}$ are positive definite, and $G \in \mathbb{R}^{D\times d_a}$. Encoders $A \in \mathbb{R}^{D\times D}$ may be singular; predictors are linear in (Ax, a). Write

$$\mathcal{L}_{\mathrm{pred}}(A, p) = \tfrac{1}{2}\mathbb{E}\left\|p(Ax, a) - Ax'\right\|^2, \qquad R = \Sigma^{-1/2} W \Sigma^{-1/2}.$$

Multiplying a coordinate-averaged prediction objective by $D/2$ gives this normalization, with its regularization weight multiplied by the same factor. If the regularizer is averaged over current and successor features, the stationary-marginal condition

$$F_\eta \Sigma F_\eta^\top + G\Gamma G^\top + \eta W = \Sigma$$

makes the two expected statistics identical. Independent trajectories also suffice when examples are independent within each regularized time slice.

For reference, write the isotropic objective as

$$\mathcal{L}_B(A,p) = \mathcal{L}_{\text{pred}}(A,p) + \lambda_B \mathcal{R}_N(A\Sigma A^\top). \tag{63}$$

Proof of Proposition 1. Prediction reduction. Let $B = A\Sigma^{1/2}$ and $K = B^\top B \succeq 0$. Independence and zero-mean noise give, including for singular A,

$$\mathcal{L}_{\text{pred}}(A,p) = \tfrac{1}{2}\mathbb{E}\,\|p(Ax,a) - A(F_\eta x + Ga)\|^2 + \tfrac{\eta}{2}\operatorname{tr}(KR) \geq \tfrac{\eta}{2}\operatorname{tr}(KR). \tag{64}$$

Every positive-definite K attains this lower bound with an invertible encoder and

$$p_A(z,a) = AF_\eta A^{-1} z + AGa. \tag{65}$$

Isotropic target. The matrices $BB^\top$ and $B^\top B$ are orthogonally similar. The reduced objective is therefore

$$f_\eta(K) = \tfrac{\eta}{2}\operatorname{tr}(KR) + \lambda_B \mathcal{R}_N(K), \qquad K \succeq 0.$$

It is continuous and coercive for every $\eta > 0$ since $R \succ 0$ and $\mathcal{R}_N \geq 0$. Comparing any minimizer with qI_D yields

$$\tfrac{\eta}{2}\operatorname{tr}(KR) + \lambda_B[\mathcal{R}_N(K) - \mathcal{R}_N(qI_D)] \leq \tfrac{\eta q}{2}\operatorname{tr} R. \tag{66}$$

Both left-hand terms are nonnegative. Thus all minimizing K lie in a common compact set and their regularizer gaps tend to zero. Every accumulation point is qI_D. This localization is uniform: otherwise a sequence of minimizers staying a fixed distance from qI_D would have a subsequence converging to it. By the Hessian formula and continuity, $\mathcal{R}_N$ is strictly convex on a small convex neighborhood of qI_D. All global minimizers eventually lie there, so the reduced minimizer is unique and positive definite.

It is attained by $A = K^{1/2}\Sigma^{-1/2}$ and Eq. (65). A global minimizer of the full objective must attain both the reduced minimum and equality in Eq. (64). Its encoder is invertible; since $\operatorname{Cov}(Ax,a) = \operatorname{diag}(A\Sigma A^\top, \Gamma)$ is positive definite, equality in prediction identifies the linear predictor coefficients. Finally $A^\top A = \Sigma^{-1/2} K \Sigma^{-1/2}$ gives Eq. (5), and the prediction-loss identity follows from Eq. (64). □

E.3 Proof of Proposition 2

In addition to the diagonal target family $\mathcal{T}_{D,\kappa}$ in Eq. (7), define

$$\mathcal{S}_{D,\kappa} = \{L \succ 0 : \operatorname{tr} L = D,\ \operatorname{cond}(L) \leq \kappa\}.$$

The set $\mathcal{S}_{D,\kappa}$ incorporates the encoder's free orientation; it does not introduce a full-covariance target parameterization.

Proof. Let $B = A\Sigma^{1/2}$ and write the polar decomposition as $B = OK^{1/2}$, using an orthogonal extension if B is singular. Set $L = O^\top \Lambda O$. Then

$$\Lambda^{-1/2} A\Sigma A^\top \Lambda^{-1/2} = O(L^{-1/2} K L^{-1/2}) O^\top.$$

Together with the prediction reduction in Eq. (64) and orthogonal invariance of $\mathcal{R}_N$, the learned-target objective reduces to

$$g_\eta(K,L) = \tfrac{\eta}{2}\operatorname{tr}(KR) + \lambda_T \mathcal{R}_N(L^{-1/2} K L^{-1/2}), \qquad K \succeq 0, \quad L \in \mathcal{S}_{D,\kappa}.$$

Every admissible pair (K,L) with $K \succ 0$ is realizable by the full model: choose O such that $OLO^\top$ is diagonal, set $\Lambda = OLO^\top$ and $A = OK^{1/2}\Sigma^{-1/2}$, and use the predictor in Eq. (65). For $L \in \mathcal{S}_{D,\kappa}$,

$$\lambda_{\min}(L) \geq \frac{D}{1 + (D-1)\kappa}, \qquad \lambda_{\max}(L) \leq D,$$

so $\mathcal{S}_{D,\kappa}$ is compact. Since the reduced objective is continuous and coercive in K, it attains a minimum for every $\eta > 0$.

Let L_0 minimize $\operatorname{tr}(LR)$ on $\mathcal{S}_{D,\kappa}$. For any reduced global minimizer (K, L), comparison with (qL_0, L_0) gives, with $C = L^{-1/2}KL^{-1/2}$,

$$\tfrac{\eta}{2}\operatorname{tr}(KR) + \lambda_T[\mathcal{R}_N(C) - \mathcal{R}_N(qI_D)] \le \tfrac{\eta q}{2}\operatorname{tr}(L_0 R). \tag{67}$$

The two terms on the left are nonnegative. The comparison therefore bounds K uniformly over minimizing pairs and forces the regularizer gap to vanish as $\eta \downarrow 0$. Compactness of the minimizing sets and the uniqueness of the covariance minimizer of $\mathcal{R}_N$ then give, uniformly over reduced global minimizers,

$$L^{-1/2}KL^{-1/2} \longrightarrow qI_D. \tag{68}$$

Consequently,

$$K - qL = L^{1/2}\Big(L^{-1/2}KL^{-1/2} - qI_D\Big)L^{1/2} \longrightarrow 0$$

uniformly. In particular, every minimizing K is positive definite for all sufficiently small η. The reduced minimum is therefore attained by the full objective, proving existence of a global minimum in Eq. (11). Moreover, every full global minimizer must attain both the reduced minimum and equality in the prediction bound, so its encoder is invertible and its predictor is the exact encoded conditional mean in Eq. (65).

Along any sequence of global minimizers with $L \to \bar{L}$, Eq. (67) and $K - qL \to 0$ imply

$$\operatorname{tr}(\bar{L}R) \le \operatorname{tr}(L_0 R).$$

Hence every accumulation point $\bar{L}$ minimizes $\operatorname{tr}(LR)$ on $\mathcal{S}_{D,\kappa}$. If this trace minimizer is unique, compactness gives uniform convergence $L \to L_0$ over global minimizers. Finally,

$$A^\top A = \Sigma^{-1/2}K\Sigma^{-1/2},$$

so conjugating $K - qL \to 0$ by $\Sigma^{-1/2}$ yields the metric limit in Eq. (12). The trace bounds in Eq. (67) also imply $\mathcal{L}_{\text{pred}} = \eta\operatorname{tr}(KR)/2 \to 0$. The same bounds give uniform upper and lower bounds on the encoder singular values for sufficiently small noise. This proves Proposition 2. □

The task-alignment condition used in Section 4.3 follows directly from the limiting metric in Eq. (12). For a task metric $Q \succ 0$, proportionality to Q together with $\operatorname{tr} L = D$ gives

$$L_* = \frac{D\Sigma^{1/2}Q\Sigma^{1/2}}{\operatorname{tr}(\Sigma Q)},$$

which is Eq. (14). Here $L = O^\top \Lambda O$ incorporates the encoder orientation. The diagonal target alone therefore does not determine the state metric: the aligned geometry must be both feasible and selected by predictive training.

E.4 Exact rollouts and finite-horizon regret

For a deterministic planned sequence $U = (a_0, \dots, a_{H-1})$, use independent innovations in Eq. (62) and a fixed initial state x_0. Set

$$\mu_{\eta,H}(U) = F_\eta^H x_0 + \sum_{t=0}^{H-1} F_\eta^{H-1-t} G a_t, \tag{69}$$

$$\Omega_{\eta,H} = \eta \sum_{j=0}^{H-1} F_\eta^j W (F_\eta^j)^\top. \tag{70}$$

For a fixed goal y and $Q \succ 0$, let $J_{*,\eta,H}(U) = \mathbb{E}\,\|x_H - y\|_Q^2$, where $\|v\|_Q^2 = v^\top Q v$. All planning comparisons below use the same feasible sequence set.

Lemma 3 (Exact finite-horizon encoded means). *At every global optimum covered by Proposition 1 or Proposition 2, autoregressive prediction initialized at Ax_0 satisfies*

$$\hat{z}_H(U) = A\mu_{\eta,H}(U) = \mathbb{E}[Ax_H \mid x_0, U]. \tag{71}$$

Moreover,

$$\mathbb{E}\,\|\hat{z}_H(U) - Ax_H\|^2 = \operatorname{tr}(A\Omega_{\eta,H}A^\top), \tag{72}$$

$$J_{*,\eta,H}(U) = \|\mu_{\eta,H}(U) - y\|_Q^2 + \operatorname{tr}(Q\Omega_{\eta,H}). \tag{73}$$

If $F_\eta \to F_0$ and H is fixed, Eq. (72) is $O(\eta)$ uniformly over all training optima and deterministic planned sequences.

Proof. Induction with Eq. (65) gives Eq. (71). The independent, zero-mean innovations give Eq. (70); expanding squared error gives the two remaining identities. A finite sum of bounded powers of F_η, together with the uniform bound on $\|A\|$, gives the $O(\eta)$ estimate. The task-noise term is independent of the planned sequence and cancels in open-loop regret. □

Lemma 4 (Convergence of constrained planning regret). *Let $\mathcal{U} \subset \mathbb{R}^{d_a H}$ be nonempty, compact, and convex, let $F_\eta \to F_0$, and suppose the learned metric $M_\eta = A^\top A$ converges to $M_0 \succ 0$ uniformly over the training optima under consideration. Define*

$$\mathcal{Y}_0 = \{\mu_{0,H}(U) : U \in \mathcal{U}\}, \qquad v_M = \operatorname*{arg\,min}_{v\in\mathcal{Y}_0} \|v - y\|_{M_0}^2, \qquad v_Q = \operatorname*{arg\,min}_{v\in\mathcal{Y}_0} \|v - y\|_Q^2.$$

The two terminal minimizers are unique. Every exact latent-planning minimizer U_η satisfies, uniformly over training and planning minimizers,

$$J_{*,\eta,H}(U_\eta) - \min_{U\in\mathcal{U}} J_{*,\eta,H}(U) \longrightarrow \|v_M - y\|_Q^2 - \|v_Q - y\|_Q^2. \tag{74}$$

The limit is positive if and only if $v_M \neq v_Q$.

Proof. The terminal set is compact and convex, so strict convexity of each quadratic gives a unique minimizing terminal state. Optimal action sequences themselves need not be unique. Since H is fixed and $\mathcal{U}$ is compact, the mean maps and the latent costs converge uniformly. For any $\eta_n \downarrow 0$ and choices of training and planning minimizers, compactness gives a subsequence $U_{\eta_n} \to \bar{U}$. Passing the minimizing inequality to the limit shows that $\mu_{0,H}(\bar{U}) = v_M$. Uniform convergence of the mean task costs gives convergence of their minima and proves Eq. (74), since the common noise term cancels. If convergence were not uniform, a sequence with a fixed nonzero deviation would have the same convergent-subsequence property, a contradiction. Strict convexity of the task cost on $\mathcal{Y}_0$ proves the last claim. □

E.5 Separation in every dimension and at every finite horizon

Theorem 4 (Explicit finite-horizon construction and target compensation). *Fix $D \geq 2$, a finite $H \geq 1$, $\alpha \in (0,1)$, $\bar{u} > 0$, and any nonscalar covariance $\Sigma \succ 0$. Set*

$$F_\eta = (\alpha^2 I_D - \eta\Sigma^{-1})^{1/2}, \qquad \Gamma = (1-\alpha^2)\Sigma, \tag{75}$$

and train on independent Gaussian tuples

$$x' = F_\eta x + a + \xi, \qquad x \sim \mathcal{N}(0,\Sigma), \quad a \sim \mathcal{N}(0,\Gamma), \quad \xi \sim \mathcal{N}(0,\eta I_D). \tag{76}$$

Assume the statistic conditions in Appendix E.1. Evaluate from $x_0 = 0$, with $\|a_t\| \leq \bar{u}$ at every step, using expected squared Euclidean terminal error to a fixed goal y. Put

$$\mathcal{U} = \{(a_0,\dots,a_{H-1}) : \|a_t\| \leq \bar{u}\}, \qquad r_H = \bar{u}\sum_{j=0}^{H-1}\alpha^j,$$

and choose $\|y\| > r_H$ with $\Sigma^{-1}y$ not parallel to y.

For any fixed $\lambda_B > 0$ and all sufficiently small $\eta > 0$, every isotropic-training global optimum is invertible and has exact encoded conditional-mean rollouts. For every exact H-step latent-planning minimizer $U_{B,\eta}$,

$$\begin{aligned} &\mathrm{Regret}_{\eta,H}(U_{B,\eta}) \longrightarrow \Delta_H > 0, \\ &\mathrm{Regret}_{\eta,H}(U) := J_{*,\eta,H}(U) - \min_{V \in \mathcal{U}} J_{*,\eta,H}(V), \end{aligned} \tag{77}$$

where

$$\begin{aligned} &v_B = (I_D + \nu\Sigma)^{-1} y, \qquad \|v_B\| = r_H, \qquad \nu > 0, \\ &\Delta_H = \|v_B - y\|^2 - (\|y\| - r_H)^2. \end{aligned} \tag{78}$$

The scalar ν is unique.
Suppose additionally that

$$\begin{aligned} &\Sigma = s[I_D + (c-1)P], \qquad s = \frac{D}{D + r(c-1)}, \qquad c > 1, \\ &P = P^\top = P^2, \qquad \operatorname{rank} P = r \in \{1, \dots, D-1\}. \end{aligned} \tag{79}$$

For any fixed $\lambda_T > 0$ and $\kappa = c$, every learned-target global optimum then has

$$L \longrightarrow \Sigma, \qquad A^\top A \longrightarrow q I_D, \qquad \mathrm{Regret}_{\eta,H}(U_{T,\eta}) \longrightarrow 0. \tag{80}$$

All limits are uniform over global training optima and exact planning minimizers. Both methods have vanishing training prediction loss and fixed-H encoded rollout MSE. The learned-target regret is strictly smaller for all sufficiently small positive noise.

Proof. Valid training distribution. For $0 < \eta < \alpha^2 \lambda_{\min}(\Sigma)$, F_η is positive definite, commutes with Σ, and satisfies

$$F_\eta \Sigma F_\eta^\top = \alpha^2 \Sigma - \eta I_D.$$

Adding Γ and ηI_D proves that x' has covariance Σ. Initializing independent training trajectories with this Gaussian marginal and using independent Gaussian actions preserves it at every time slice. The control matrix is I_D, so every state coordinate is actuated. The Gaussian training policy and the bounded evaluation action set are distinct; the evaluation bound applies equally to both planners.

Isotropic metric and reachable means. Proposition 1 applies with $R = \Sigma^{-1}$ and gives $M_{B,\eta} \to q\Sigma^{-1}$. Since $F_\eta \to \alpha I_D$,

$$\mu_{0,H}(U) = \sum_{t=0}^{H-1} \alpha^{H-1-t} a_t.$$

Its image over $\mathcal{U}$ is exactly the ball $\{v : \|v\| \le r_H\}$. The triangle inequality proves one inclusion. For the reverse inclusion, choose $a_t = v/(\sum_{j=0}^{H-1} \alpha^j)$ for every t. All these actions satisfy the bound. Every action time has a nonzero coefficient.

Different optimal terminal states. The Euclidean projection of y onto this ball is $v_* = r_H y/\|y\|$. The baseline instead minimizes $(v-y)^\top \Sigma^{-1}(v-y)$ on the ball. Since y is outside it, the optimum is on the boundary and satisfies

$$\Sigma^{-1}(v_B - y) + \nu v_B = 0, \qquad \nu > 0.$$

This gives Eq. (78). The norm of $(I_D + \nu\Sigma)^{-1} y$ decreases continuously and strictly from $\|y\|$ to zero with ν, proving uniqueness. If $v_B = v_*$, the stationarity equation requires $\Sigma^{-1} y$ to be parallel to y, contrary to the assumption. The task projection is unique, so $\Delta_H > 0$. Such goals exist for every nonscalar Σ: take a vector with nonzero components in two distinct eigenspaces and scale it beyond radius r_H. Lemma 4 proves Eq. (77) for the positive-noise systems.

Unique learned selector for Eq. (79). Here $\operatorname{tr}\Sigma = D$ and $\mathrm{cond}(\Sigma) = c$. For $L \in \mathcal{S}_{D,c}$ let $T = \operatorname{tr}(PL)$. Then

$$\frac{T}{r} \le \lambda_{\max}(L) \le c\lambda_{\min}(L) \le c\frac{D-T}{D-r}, \qquad T \le crs. \tag{81}$$

Since $\Sigma^{-1} = s^{-1}I_D - (c-1)(sc)^{-1}P$,

$$\operatorname{tr}(L\Sigma^{-1}) = \frac{D}{s} - \frac{c-1}{sc}T \geq D,$$

and $L = \Sigma$ attains equality. Equality requires equality throughout Eq. (81). In particular, $\operatorname{tr}(P[\lambda_{\max}(L)I_D - L]) = 0$. The bracket is positive semidefinite, so each vector in an orthonormal basis of $\operatorname{range}(P)$ has zero quadratic form under it and is annihilated by it. Thus that range is a maximal-eigenvalue subspace of L. The same argument makes its orthogonal complement a minimal-eigenvalue subspace. Equality in the middle inequality and the trace constraint fix their eigenvalues to sc and s, respectively. Consequently $L = \Sigma$ is the unique trace minimizer.

Proposition 2 gives $L \to \Sigma$ and $M_{T,\eta} \to qI_D$. Lemma 4 now gives Eq. (80), uniformly over all minimizer choices. Lemma 3 gives exact encoded terminal means and vanishing rollout MSE; the training losses vanish by Proposition 2. Uniform convergence and $\Delta_H > 0$ give strict improvement at small positive noise. □

The theorem concerns regret relative to the best feasible sequence, not zero terminal error: the selected goal is outside the limiting reachable ball. It establishes a family in every dimension and at every fixed finite horizon. The admissible noise threshold can depend on these parameters. The learned-target conclusion uses Eq. (79) and its matching bound; the isotropic failure does not require the two-level restriction. The theorem analyzes open-loop sequence selection over the stated feasible set.

The isotropic part of Theorem 4 proves Theorem 1 in the main text; the additional two-level condition gives the learned-target compensation used in the method analysis.

E.6 Stability of the planning separation

The next lemma controls both changes to the planning cost and changes to the physical task cost. It also allows approximate minimization of the perturbed latent cost.

Lemma 5 (A quantitative regret margin)**.** *Let $\mathcal{U}$ be a nonempty compact set, and let b_0, j_0 be continuous functions on it. Put $j_0^* = \min_{\mathcal{U}} j_0$ and suppose*

$$\Delta = \min_{U \in \arg\min_{\mathcal{U}} b_0} [j_0(U) - j_0^*] > 0.$$

For $q > 0$, define $t_0 = qj_0$ and

$$\mathcal{E} = \{U \in \mathcal{U} : j_0(U) - j_0^* \leq 3\Delta/4\}, \qquad \gamma = \min_{U \in \mathcal{E}} b_0(U) - \min_{U \in \mathcal{U}} b_0(U) > 0. \tag{82}$$

Let continuous costs b, t, j satisfy

$$\|b - b_0\|_\infty \leq \epsilon_B, \qquad \|t - t_0\|_\infty \leq \epsilon_T, \qquad \|j - j_0\|_\infty \leq \epsilon_J.$$

Let U_B, U_T have respective optimization errors $\zeta_B, \zeta_T \geq 0$, i.e., $b(U_B) \leq \min_{\mathcal{U}} b + \zeta_B$ and $t(U_T) \leq \min_{\mathcal{U}} t + \zeta_T$. If $2\epsilon_B + \zeta_B < \gamma$, then

$$j(U_B) - \min_{\mathcal{U}} j > 3\Delta/4 - 2\epsilon_J, \tag{83}$$

$$j(U_T) - \min_{\mathcal{U}} j \leq (2\epsilon_T + \zeta_T)/q + 2\epsilon_J. \tag{84}$$

In particular, if also $\epsilon_J < \Delta/16$ and $2\epsilon_T + \zeta_T < q\Delta/8$, then the first regret exceeds $\Delta/2$ and the second is less than $\Delta/4$.

Proof. The set $\mathcal{E}$ is nonempty and compact, and is disjoint from $\arg\min_{\mathcal{U}} b_0$. Thus $\gamma > 0$. The approximate minimizing inequality gives

$$b_0(U_B) \leq \min_{\mathcal{U}} b_0 + 2\epsilon_B + \zeta_B < \min_{\mathcal{U}} b_0 + \gamma,$$

so $U_B \notin \mathcal{E}$. Also $qj_0(U_T) \leq qj_0^* + 2\epsilon_T + \zeta_T$. Finally $|\min_{\mathcal{U}} j - \min_{\mathcal{U}} j_0| \leq \epsilon_J$; applying this bound and the pointwise task-cost bound proves Eqs. (83)–(84). The stated thresholds give respectively a lower bound greater than $5\Delta/8$ and an upper bound less than $\Delta/4$. □

E.7 SMOOTH NONLINEAR PERTURBATIONS

We now realize the cost perturbations in Lemma 5 through nonlinear dynamics and encoders. The reference models are the joint optima of Theorem 4, including the two-level condition for its learned-target conclusion. The nonlinear predictors may themselves be exact one-step conditional means.

Corollary 2 (Nonlinear robustness of finite-horizon separation)**.** *Fix the parameters, action set, and goal of Theorem 4 with Eq.* (79)*. For each sufficiently small $\eta > 0$, choose any global linear-training optima with encoders $A_{B,\eta}$ and $A_{T,\eta}$. Perturb the physical dynamics to*

$$X_{t+1} = F_\eta X_t + a_t + d(X_t, a_t) + \sqrt{\eta} Z_t, \quad X_0 = 0, \quad Z_t \overset{\text{iid}}{\sim} \mathcal{N}(0, I_D), \tag{85}$$

where d is continuously differentiable and $\sup_{x,\|a\|\le\bar{u}} \|d(x,a)\| \le \varepsilon_d$. For $\chi \in \{\mathrm{B}, \mathrm{T}\}$, let

$$\begin{aligned} f_\chi(x) &= A_{\chi,\eta} x + h_\chi(x), \qquad \sup_x \|h_\chi(x)\| \le \varepsilon_e, \\ \sup_x \|Dh_\chi(x)\| &\le \varepsilon_e < \sigma_{\min}(A_{\chi,\eta}), \end{aligned} \tag{86}$$

with h_χ continuously differentiable. These bounds are global. Define the exact encoded one-step conditional mean by

$$p^{\mathrm{cm}}_\chi(z,a) = \mathbb{E}_Z f_\chi\big(F_\eta f_\chi^{-1}(z) + a + d(f_\chi^{-1}(z), a) + \sqrt{\eta} Z\big)\,. \tag{87}$$

Use any continuous predictor $\widetilde{p}_\chi$ satisfying

$$\sup_{z,\|a\|\le\bar{u}} \big\|\widetilde{p}_\chi(z,a) - p^{\mathrm{cm}}_\chi(z,a)\big\| \le \varepsilon_p. \tag{88}$$

Roll it out from $f_\chi(0)$ and use the cost $c_\chi(U) = \|\hat{z}_{\chi,H}(U) - f_\chi(y)\|^2$. The physical cost is $j(U) = \mathbb{E}\,\|X_H(U) - y\|^2$. Suppose each returned sequence $\widetilde{U}_\chi$ has latent-cost optimization error at most $\zeta \ge 0$.
There exist $\eta_0, \varepsilon_0, \zeta_0 > 0$, depending only on the fixed reference parameters, such that for $0 < \eta < \eta_0$, $\max\{\varepsilon_d, \varepsilon_e, \varepsilon_p\} < \varepsilon_0$, and $0 \le \zeta < \zeta_0$, every such choice satisfies

$$\begin{aligned} \mathrm{Regret}_j(\widetilde{U}_B) > \Delta_H/2, \qquad \mathrm{Regret}_j(\widetilde{U}_T) < \Delta_H/4, \\ \mathrm{Regret}_j(U) = j(U) - \min_{V\in\mathcal{U}} j(V). \end{aligned} \tag{89}$$

Both encoders are globally invertible and have uniformly nonsingular Jacobians. For fixed D, H, their encoded rollout mean squared errors satisfy

$$\sup_{U\in\mathcal{U}} \mathbb{E}\,\|\hat{z}_{\chi,H}(U) - f_\chi(X_H(U))\|^2 = O\big(\eta + (\varepsilon_d + \varepsilon_e + \varepsilon_p)^2\big)\,. \tag{90}$$

The constants and conclusions are uniform over the reference global training optima. In particular, Eq. (89) *allows $\varepsilon_p = \zeta = 0$: exact one-step conditional-mean prediction and exact sequence optimization do not remove the separation in this nonlinear neighborhood.*

Proof. For this proof write $A_\chi = A_{\chi,\eta}$, $F = F_\eta$, and $S_H(\beta) = \sum_{j=0}^{H-1} \beta^j$ for $\beta \ge 0$.

1. Invertibility and conditional means. By the derivative bound, h_χ is globally ε_e-Lipschitz. For any x, x',

$$(\sigma_{\min}(A_\chi) - \varepsilon_e)\,\|x - x'\| \le \|f_\chi(x) - f_\chi(x')\| \le (\|A_\chi\| + \varepsilon_e)\,\|x - x'\|\,. \tag{91}$$

For every z, the map $x \mapsto A_\chi^{-1}(z - h_\chi(x))$ is a contraction on $\mathbb{R}^D$, so it has a unique fixed point. This proves surjectivity as well as injectivity. The Jacobian is nonsingular by the same lower bound, so f_χ is a global C^1 diffeomorphism. Thus Eq. (87) is well defined and is the exact conditional mean of the next encoded state. Its expectation is finite; bounded h_χ and continuity, together with the Gaussian first moment, also give continuity.

2. Uniform perturbation of autoregressive prediction. Let $p^0_\chi(z,a) = A_\chi F A_\chi^{-1} z + A_\chi a$ and $\ell_\chi = \|A_\chi F A_\chi^{-1}\|$. At $z = f_\chi(x)$, zero-mean noise gives

$$p^{\mathrm{cm}}_\chi(f_\chi(x),a) - p^0_\chi(f_\chi(x),a) = A_\chi d(x,a) - A_\chi F A_\chi^{-1} h_\chi(x) + \mathbb{E}_Z h_\chi(Fx + a + d(x,a) + \sqrt{\eta} Z).$$

Since f_χ is onto, for every z and feasible a,

$$\|\widetilde{p}_\chi(z,a) - p^0_\chi(z,a)\| \le b_\chi := \|A_\chi\|\,\varepsilon_d + (1+\ell_\chi)\varepsilon_e + \varepsilon_p. \tag{92}$$

Comparison with the reference rollout $A_\chi \mu_{\eta,t}(U)$ and induction yield

$$\|\hat{z}_{\chi,H}(U) - A_\chi \mu_{\eta,H}(U)\| \le \ell_\chi^H \varepsilon_e + S_H(\ell_\chi) b_\chi. \tag{93}$$

Including the goal-encoding perturbation, put

$$e_\chi = (\ell_\chi^H + 1)\varepsilon_e + S_H(\ell_\chi) b_\chi, \qquad B_\chi = \|A_\chi\|\,(r_H + \|y\|). \tag{94}$$

The reference latent cost is $c^0_{\chi,\eta}(U) = \|A_\chi(\mu_{\eta,H}(U) - y)\|^2$. Using $\|F_\eta\| \le \alpha$ and $\|\mu_{\eta,H}(U)\| \le r_H$, we obtain

$$\|c_\chi - c^0_{\chi,\eta}\|_\infty \le 2B_\chi e_\chi + e_\chi^2. \tag{95}$$

This argument does not commute expectation with a nonlinear rollout.

3. Uniform perturbation of the physical cost. Couple Eq. (85) to the unperturbed linear process X^0_t using the same innovations and planned actions. Pathwise,

$$\|X_{t+1} - X^0_{t+1}\| \le \alpha \|X_t - X^0_t\| + \varepsilon_d, \qquad \|X_H - X^0_H\| \le d_H := \varepsilon_d S_H(\alpha).$$

Let $j^0_\eta(U) = \mathbb{E}\,\|X^0_H(U) - y\|^2$ and define

$$B_* := \left[(r_H + \|y\|)^2 + \eta D S_H(\alpha^2)\right]^{1/2}.$$

Cauchy–Schwarz and the pathwise bound give

$$\|j - j^0_\eta\|_\infty \le 2B_* d_H + d_H^2. \tag{96}$$

The global bound on d provides a uniform square-integrable envelope, so j is continuous on $\mathcal{U}$ even though the innovations are unbounded.

4. Applying the limiting regret margin. Define on the same compact action set

$$\mu_0(U) = \sum_{t=0}^{H-1} \alpha^{H-1-t} a_t, \qquad j_0(U) = \|\mu_0(U) - y\|^2,$$

$$b_0(U) = q\,\|\mu_0(U) - y\|^2_{\Sigma^{-1}}, \quad t_0(U) = q j_0(U).$$

By Theorem 4, all minimizers of b_0 have j_0 regret Δ_H, while t_0 has the same minimizers as j_0. Propositions 1 and 2, together with $F_\eta \to \alpha I_D$, give, uniformly over reference training optima,

$$c^0_{B,\eta} \to b_0, \qquad c^0_{T,\eta} \to t_0, \qquad j^0_\eta \to j_0$$

in the uniform norm on $\mathcal{U}$. The encoder norms and inverse norms are uniformly bounded, so ℓ_χ, B_χ, and all finite-horizon constants above are uniformly bounded as well. Equations (95) and (96) therefore make the total cost perturbations arbitrarily small by choosing η and $\varepsilon_d, \varepsilon_e, \varepsilon_p$ sufficiently small. There is also a common positive lower bound on $\sigma_{\min}(A_\chi)$, so Eqs. (86) and (91) hold uniformly after reducing ε_0 if necessary. Choose the thresholds to satisfy Lemma 5 with $\Delta = \Delta_H$, and then take a common sufficiently small ζ_0. This proves Eq. (89) for every allowed choice, including nonunique or approximate planning minimizers.

5. Encoded rollout error. Using the same coupling and the L^2 triangle inequality,

$$\left(\mathbb{E}\,\|\hat{z}_{\chi,H} - f_\chi(X_H)\|^2\right)^{1/2} \le \ell_\chi^H \varepsilon_e + S_H(\ell_\chi) b_\chi + \|A_\chi\|\,d_H + \varepsilon_e + \|A_\chi\|\sqrt{\eta D S_H(\alpha^2)}.$$

All bounds are uniform in U and the reference optima. Squaring gives Eq. (90). □

Corollary 2 concerns a neighborhood of the jointly selected linear models. It proves stability of their planning behavior, not selection of nearby encoders by unrestricted nonlinear joint optimization. For nonlinear encoders, an exact one-step encoded conditional mean need not compose into an exact terminal conditional mean; the explicit rollout bound above handles this distinction. The global perturbation bounds accommodate the unbounded support of Gaussian innovations; compact action constraints alone do not bound noisy state trajectories.

F EXPERIMENTAL DETAILS

This appendix provides the implementation and evaluation details underlying the experiments in Section 5. We first report the evaluation protocol and per-seed planning results, then describe the latent-cost ordering diagnostic used in Table 1. We next examine the target allocation learned by ΛReg, sensitivity to the anisotropy bound, and additional representation and prediction diagnostics.

F.1 EXPERIMENTAL PROTOCOL AND PER-SEED RESULTS

Data and models. We use the TwoRoom, Reacher, PushT, and Cube datasets and preprocessing protocol of LeWM (Maes et al., 2026). Training clips contain four observations with the temporal spacing used in that protocol. A randomly initialized ViT-Tiny followed by a projection MLP produces 192-dimensional representations, and the causal predictor uses a three-observation context. Prediction and planning operate in the original latent coordinates, whereas ΛReg is evaluated on the standardized representation $\Lambda^{-1/2}z$.

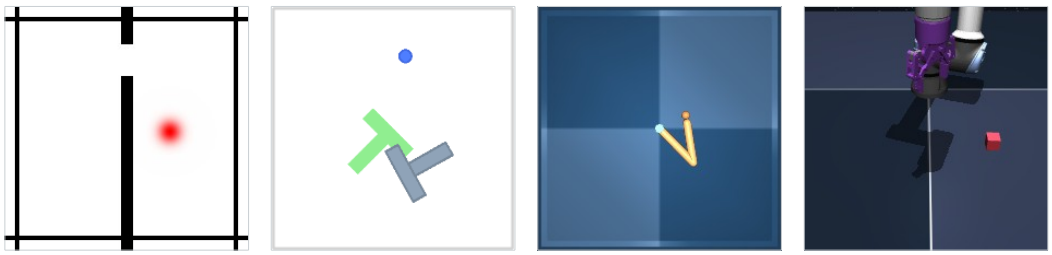

Figure 6: **Evaluation environments.** From left to right: TwoRoom, PushT, Reacher, and Cube. All four environments have continuous action spaces, and visual goal planning is performed directly from image observations.

Training and planning settings. Table 3 summarizes the settings shared across the experiments. Unless otherwise stated, AnisoWM uses $\kappa = 2$ and $\lambda = 0.09$. The LeWM baseline corresponds to $\kappa = 1$ and is trained with the same model architecture, optimizer settings, training seeds, and evaluation pairs.

Table 3: **Implementation settings.** Target logits use zero weight decay and are clipped after each update as described in Section 4.2.

Setting	Value
Epochs	15
Training seeds	3072, 1234, 42
Optimizer / batch size	AdamW / 128
Peak learning rate	5×10^{-5}
Encoder and predictor weight decay	10^{-3}
Target-logit learning-rate schedule	Same as the model
Representation dimension	192
Primary anisotropy bound / regularization weight	2 / 0.09
Projection directions / quadrature knots	1,024 / 17
CEM iterations / candidates / elites	30 / 300 / 30
Initial–goal pairs per environment	50
Environment-step budget per pair	50

Implementation pseudocode. Algorithm 1 gives the implementation of ΛReg. Relative to SIGReg, the only change to the statistic is to standardize each latent coordinate by the learned diagonal Gaussian target before computing the projected empirical characteristic functions. This standardization is used only inside the regularizer; prediction and planning operate on the original latent representation.

Algorithm 1 ΛReg with the Epps–Pulley statistic and DDP support. The target variances are parameterized by logits α and normalized to have trace D. After standardizing the features by the resulting diagonal target covariance, the remainder is the original SIGReg computation.

```
def LambdaReg(x, alpha, global_step, num_slices=1024):
    """x is (N, D); alpha is a learnable vector of size D."""

    # --- learned diagonal Gaussian target ---
    D = x.size(1)
    v = D * alpha.softmax(dim=0)    # sum(v) = D

    # standardize only inside LambdaReg
    x = x * v.rsqrt()

    # --- original SIGReg statistic ---
    # slice sampling -- synchronized across devices
    dev = dict(device=x.device)
    g = torch.Generator(**dev)
    g.manual_seed(global_step)

    proj_shape = (x.size(1), num_slices)
    A = torch.randn(proj_shape, generator=g, **dev)
    A /= A.norm(p=2, dim=0)

    # Epps--Pulley integration points and Gaussian window
    t = torch.linspace(0, 3, 17, **dev)
    exp_f = torch.exp(-0.5 * t**2)

    # empirical characteristic function -- gathered across devices
    x_t = (x @ A).unsqueeze(2) * t       # (N, M, T)
    ecf = (1j * x_t).exp().mean(0)
    ecf = all_reduce(ecf, op="AVG")

    # weighted squared distance to the Gaussian characteristic function
    err = (ecf - exp_f).abs().square().mul(exp_f)

    N = x.size(0) * world_size
    T = torch.trapz(err, t, dim=1) * N
    return T.mean()
```

The target variances are

$$v_i = D\frac{\exp(\alpha_i)}{\sum_{j=1}^{D}\exp(\alpha_j)},$$

which guarantees $\mathrm{tr}(\Lambda) = D$ for $\Lambda = \mathrm{diag}(v_1, \ldots, v_D)$. The logits are initialized at zero and, after each optimizer update, clipped as

$$\alpha_i \leftarrow \mathrm{clip}\left(\alpha_i, -\frac{1}{2}\log\kappa, \frac{1}{2}\log\kappa\right),$$

which ensures $\mathrm{cond}(\Lambda) \le \kappa$. The target parameters receive gradients only through ΛReg.

Evaluation protocol. Evaluation follows the protocol of LeWM (Maes et al., 2026). For each environment, planning is evaluated on 50 initial–goal pairs, with goals sampled 25 steps ahead within the same dataset trajectory. CEM uses 300 candidate action sequences, 30 elites, and 30 optimization iterations, and each evaluation has a 50-step environment budget.

Regularizer hyperparameters. We use the SIGReg settings reported by Maes et al. (2026), including $\lambda = 0.09$, 1,024 projection directions, and 17 integration knots, for both LeWM and AnisoWM. Maes et al. (2026) select $\lambda = 0.09$ based on their reported SIGReg hyperparameter study, so we retain this value rather than retuning the baseline for AnisoWM. ΛReg additionally introduces the anisotropy bound κ; the primary comparison uses $\kappa = 2$, and Appendix F.5 reports its sensitivity.

Per-seed planning results. The main-paper planning results are means over three training seeds. Table 4 reports the corresponding per-seed AnisoWM results.

Table 4: **Per-seed AnisoWM planning success (%).** Each entry reports AnisoWM planning success under the same evaluation protocol used for Figure 2. The final column reports the mean across the three training seeds, which the main text rounds to an integer.

Task	Seed 3072	Seed 1234	Seed 42	Mean
TwoRoom	90	94	94	92.7
Reacher	90	88	88	88.7
PushT	98	96	96	96.7
Cube	78	80	80	79.3

F.2 Latent-cost ordering diagnostic

The finite-horizon analysis in Section 3 concerns the ordering of feasible outcomes under the latent and task costs: a planner selects by that ordering, and the analysis shows it can disagree with the task cost even when prediction is exact. We therefore measure the ordering directly, in a diagnostic where LeWM and AnisoWM score the same recorded action sequences and are compared against the outcomes those sequences reached.

Candidate sequences and outcomes. For each initial–goal pair, stored evaluations provide executed action sequences together with the state each of them reached. We label a sequence by the scalar of Table 5, evaluated at the state reached after H environment steps and normalized by the environment's success threshold, so that a value of one is the success boundary and smaller values are better. The label is continuous rather than binary, and is read at the horizon rather than minimized along the trajectory. The set excludes trajectories generated by either of the two checkpoints being compared, and both models score the same remaining sequences. A case is discarded when fewer than eight candidates remain, or when all candidates reach the same outcome and the case induces no ordering. The horizon H is fixed per environment before the compared models are loaded, and is independent of the CEM planning horizon.

Table 5: **Outcome scalar and diagnostic horizon.** Each scalar is normalized by its threshold, so one is the success boundary and lower is better. Cases are counted out of the fifty initial–goal pairs of the evaluation protocol.

Task	Outcome scalar	Threshold	H	Cases
TwoRoom	agent–goal distance	16 px	20	45
Reacher	largest joint-position error	0.05 rad	25	50
PushT	$\max\big(\text{position}/20\,\text{px},\ \text{angle}/(\pi/9)\big)$	1	25	48
Cube	block–target distance	0.04 m	15	33

Encoded and predicted latent costs. The cost a planner minimizes combines the learned representation with a predictor rollout, while ΛReg changes only the target the encoder is trained against. To separate the two we score each sequence twice. For a candidate sequence U with realized terminal observation $o_H(U)$,

$$J_{\mathrm{enc}}(U) = \|f_\theta(o_H(U)) - f_\theta(o_g)\|^2, \qquad J_{\mathrm{pred}}(U) = \|\hat{z}_H(U) - f_\theta(o_g)\|^2. \tag{97}$$

J_{enc} applies the learned representation directly to the realized outcome and therefore removes rollout prediction from the diagnostic. J_{pred} is the cost available to the planner and additionally includes the predictor rollout.

Pairwise ordering accuracy. Within a case we consider every pair of candidates whose outcomes differ; pairs with identical outcomes induce no ordering and are excluded. The accuracy is the fraction of these pairs on which the latent-cost ordering agrees with the outcome ordering, counting a pair with exactly equal costs as one half. We evaluate this within each case and average over cases, so that

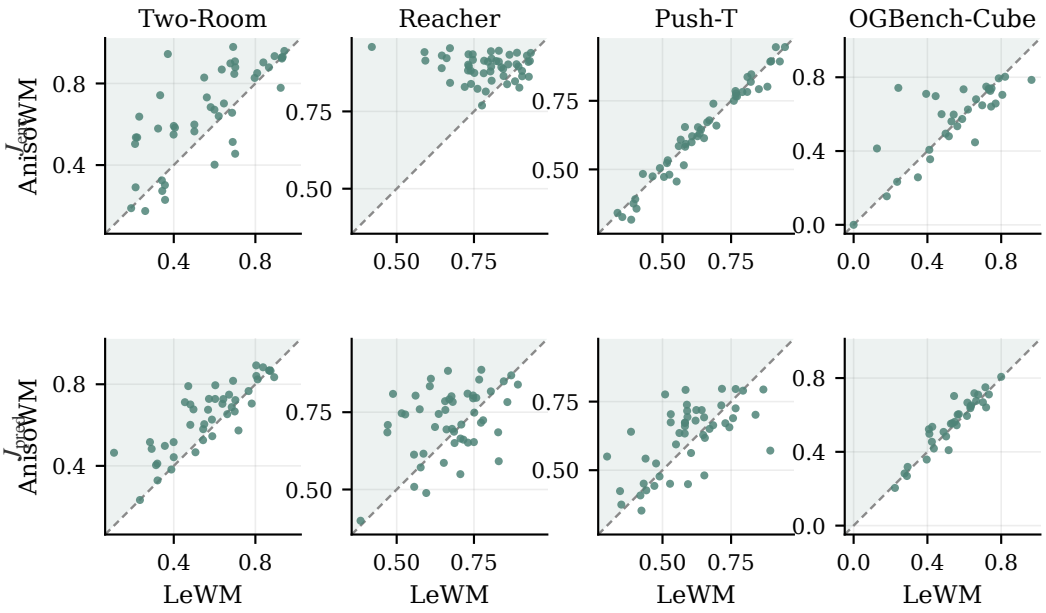

Figure 7: **Ranking accuracy, pair by pair.** Each point is one initial–goal pair: how well LeWM's cost orders its sequences on the horizontal axis against how well AnisoWM's does on the vertical one, both scoring the same sequences. The shaded half-plane above the diagonal is where AnisoWM orders that pair better. Top row: J_{enc}, which leaves the representation on its own. Bottom row: J_{pred}, the cost the planner minimizes. Rows of a column share their axis limits; each panel shows the seed Table 1 reports.

each case contributes equally regardless of its number of candidates. Table 1 reports this quantity for LeWM and AnisoWM on identical candidate sets.

Comparing the two costs localizes where a difference between the models arises. A gap under J_{enc} is present in the representation itself, before any rollout; a gap that appears only under J_{pred} arises once the predictor is involved. TwoRoom and Reacher show the former, with AnisoWM ahead under both costs, while in PushT the representation-only ordering is nearly unchanged and only the predicted cost improves (Figure 7).

F.3 Additional local cost geometry

Figure 8 shows four additional Cube/PushT example pairs using the visualization of Figure 3. The values above the zoomed panels report Spearman's rank correlation ρ between latent planning cost and task cost within the evaluated region. Across all additional examples, AnisoWM gives higher rank correlation than LeWM.

F.4 Learned target allocation

The trace and condition-number constraints determine the feasible family of target covariances but do not specify how variance is allocated across latent coordinates. This section examines how that allocation evolves during neural world-model training.

Environment-dependent target spectra. Figure 4 in the main paper shows the learned target spectra at the common bound $\kappa = 2$. The number of coordinates above the mean target variance differs across environments: 102 in TwoRoom, 124 in Reacher, 95 in PushT, and 97 in Cube. Since the spectra are sorted independently, these numbers describe differences in the learned variance distribution rather than aligned semantic coordinates.

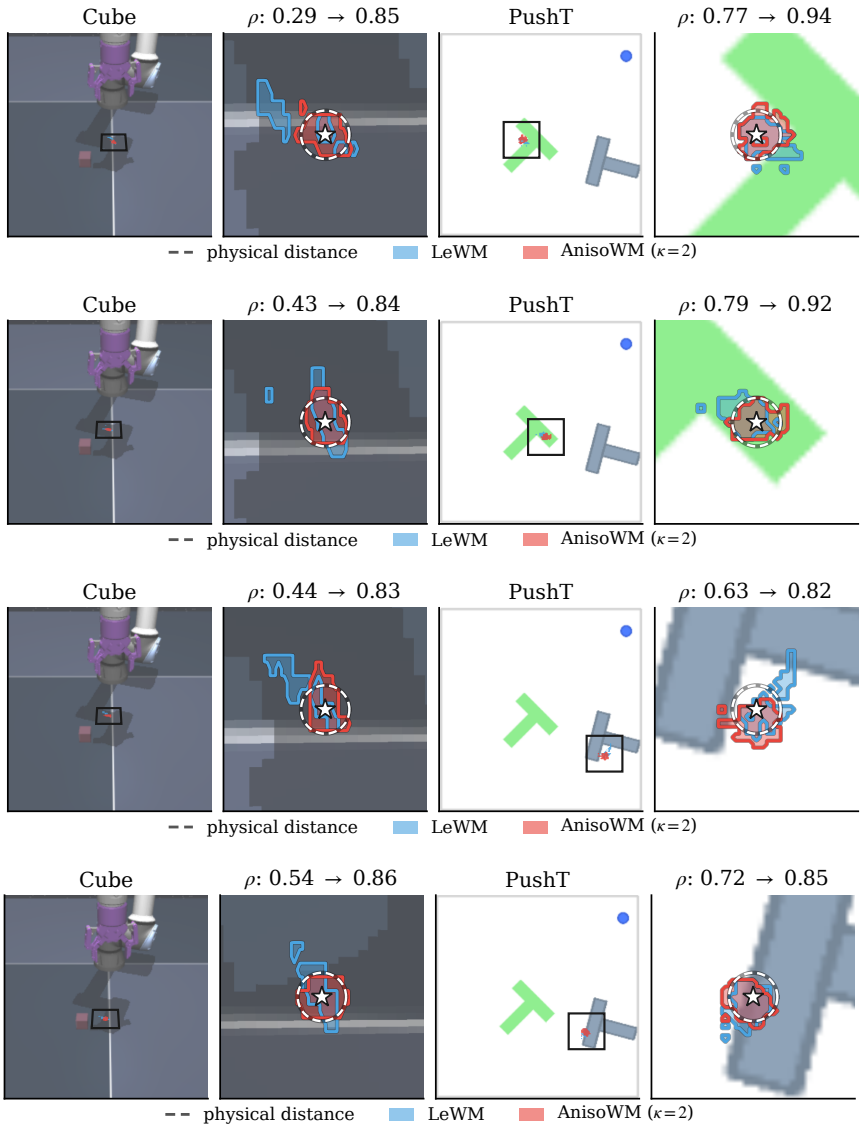

Figure 8: **Additional latent-cost neighborhoods around the goal.** Four additional Cube/PushT example pairs, shown using the same visualization as Figure 3. The dashed circle denotes the task-cost neighborhood in physical space; blue and red contours denote the lowest-cost regions under LeWM and AnisoWM ($\kappa = 2$), respectively.

Evolution during training. All target logits are initialized at zero, so training begins from the isotropic target. The right panel of Figure 4 shows the target spectrum throughout one TwoRoom run. The spectrum continues to evolve even after the learned condition number reaches the permitted bound, indicating that reaching the boundary does not uniquely determine the variance allocation.

Use of the anisotropy budget. The learned target need not always saturate the condition-number constraint. Figure 9 reports the target condition number during training for the anisotropy bounds used in the TwoRoom sweep. The bound is reached for $\kappa \leq 8$, whereas the $\kappa = 16$ run ends the fifteen-epoch training budget at a condition number of 11.3. Thus the nominal bound need not equal the anisotropy realized by training.

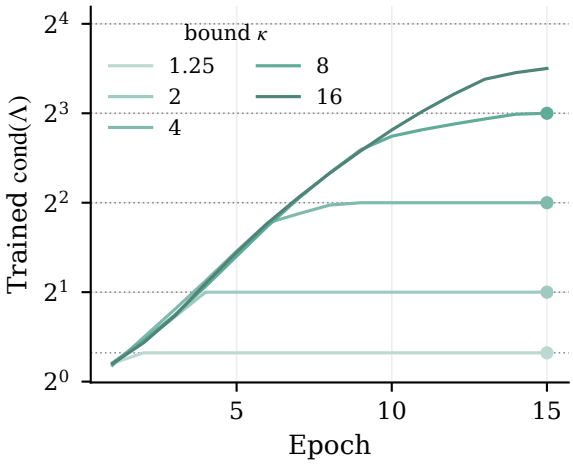

Figure 9: **Evolution of the learned target condition number.** Each curve shows one TwoRoom training run with the indicated anisotropy bound, and the dotted line at each bound is the ratio that bound permits. Filled markers indicate that the learned target reaches its bound.

F.5 Sensitivity to the anisotropy bound

The primary experiments use the same $\kappa = 2$ in all four environments. To examine sensitivity to this choice, we additionally train models at several anisotropic bounds. The numerical values underlying Figure 5 are reported in Table 6.

Table 6: **Planning success across anisotropic target bounds (%).** Each entry corresponds to the training run used in the sweep of Figure 5.

κ	TwoRoom	Reacher	PushT	Cube
1.25	90	90	94	82
2	90	90	98	78
4	92	84	92	78
8	92	84	92	78
16	94	86	92	78

The preferred value differs across environments. In particular, increasing the permitted variance contrast is not monotonically beneficial: TwoRoom continues to improve across the displayed range, whereas the other environments attain their best anisotropic result at smaller bounds. These sweep results are single training runs and should therefore be interpreted as sensitivity diagnostics rather than as replacements for the three-seed primary comparison.

F.6 Realized representation geometry

The target covariance Λ is used inside ΛReg, but it does not directly constrain the empirical covariance of the learned neural features. We therefore measure the centered feature covariance on a fixed validation probe.

Feature covariance spectrum. For each run, let

$$\ell_1 \geq \ell_2 \geq \cdots \geq \ell_D$$

be the eigenvalues of the measured feature covariance. To compare spectra across runs, we normalize each eigenvalue by the mean eigenvalue.

Changing κ changes the measured feature spectrum as well as the target spectrum. For example, in TwoRoom the eigenvalue at index 100, normalized by the mean eigenvalue, increases from 0.009 at $\kappa = 1.25$ to 0.054 at $\kappa = 16$.

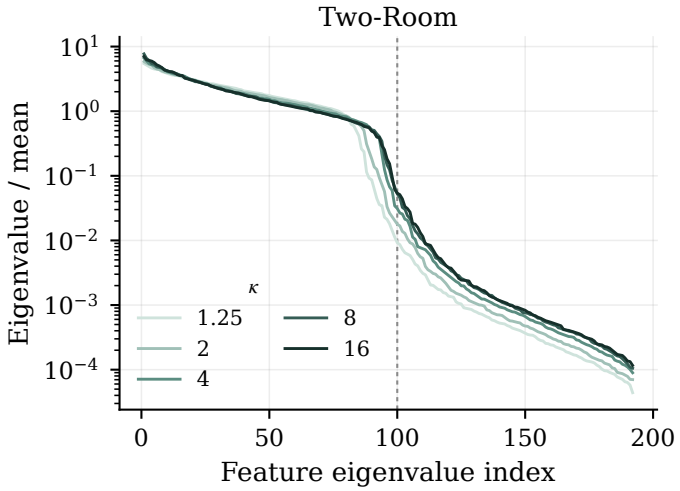

Figure 10: **Measured feature-covariance spectra in TwoRoom.** Eigenvalues of the centered feature covariance are normalized by their mean and sorted independently for each displayed anisotropy bound; the dashed line marks index 100, the value quoted in the text. These empirical feature covariances are distinct from the learned target covariance Λ.

Representation variation. With

$$\pi_j = \frac{\ell_j}{\sum_k \ell_k},$$

we compute the entropy effective rank

$$r_{\mathrm{ent}} = \exp\left(- \sum_{j:\pi_j>0} \pi_j \log \pi_j \right). \tag{98}$$

Across the reported runs, the entropy effective rank ranges from 74 to 111 out of 192 dimensions. This empirical statistic is distinct from the participation-ratio lower bound on the target covariance in Corollary 1.

F.7 Additional prediction diagnostics

Latent prediction loss and planning. Figure 11 compares latent prediction loss with planning success across the anisotropy sweep. The run with the smallest latent MSE is not consistently the run with the highest planning success. Because each model learns its own representation, however, latent MSE is measured in a different coordinate system for each run and should not be interpreted as a common physical prediction error. The figure is therefore included only as a diagnostic of the relationship between the reported latent training loss and planning performance.

F.8 Qualitative rollouts

Figures 12 and 13 show recorded rollout pairs for the four environments. These examples illustrate behavior under the same evaluation planner and are not used as an aggregate measure of performance.

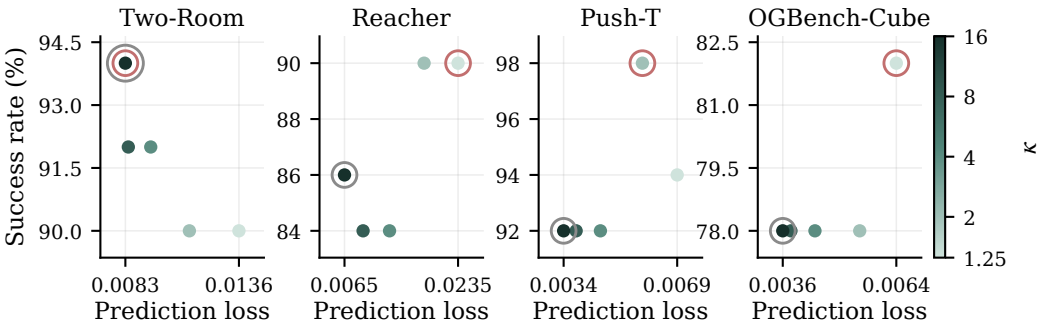

Figure 11: **Latent prediction loss and planning success.** One marker per trained run, coloured by its anisotropy bound. A run further left predicts better and a run higher plans better. The red ring marks the best-planning run and the grey ring the lowest-loss one; where a single run is both, the two rings are concentric. The MSE is measured in each model's own learned representation space.

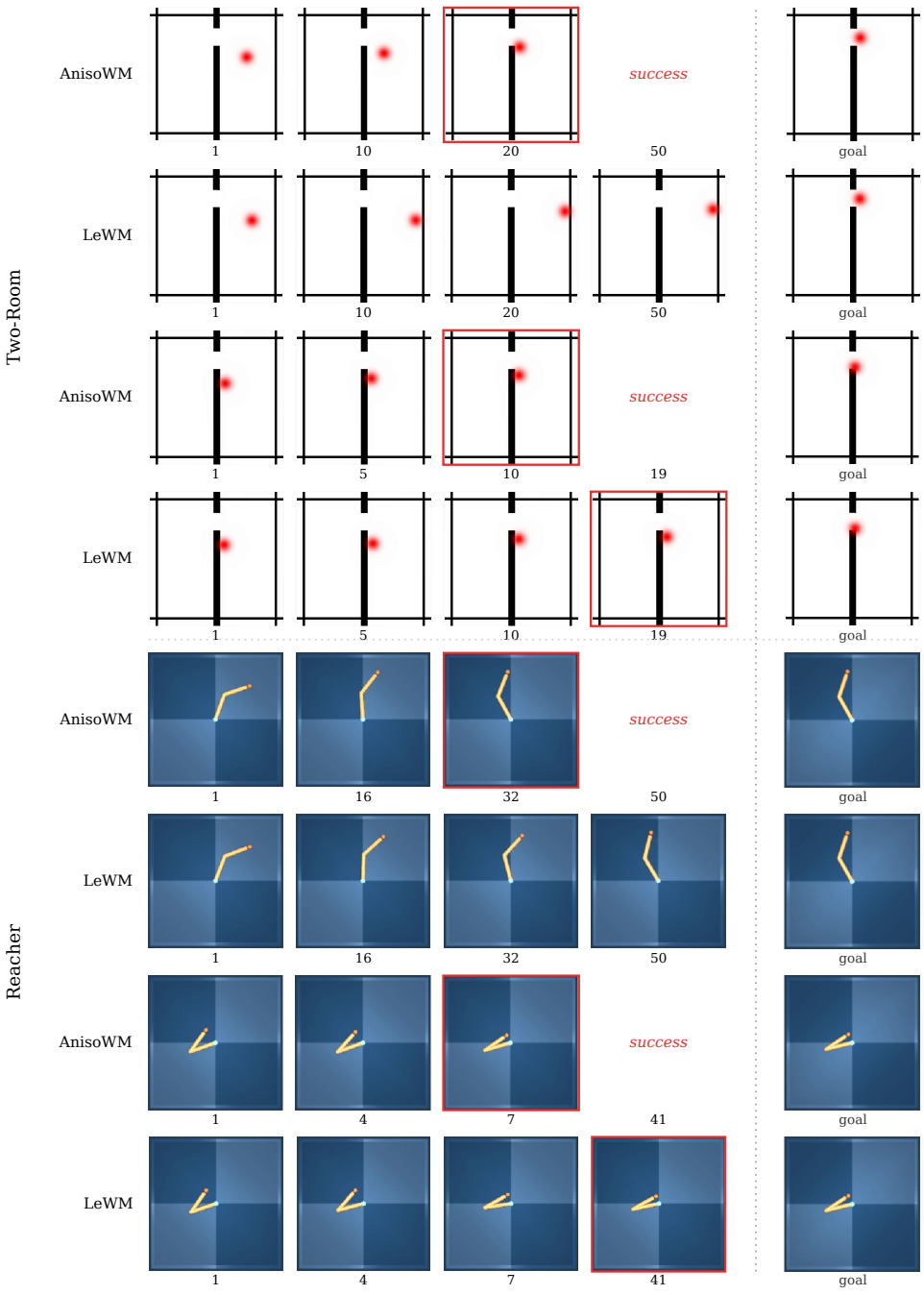

Figure 12: **Rollouts in TwoRoom and Reacher.** Two rollout pairs per environment compare AnisoWM at $\kappa = 2$ with LeWM under the same evaluation planner and seed. Columns show the indicated environment steps and the goal observation; red outlines indicate recorded goal attainment.

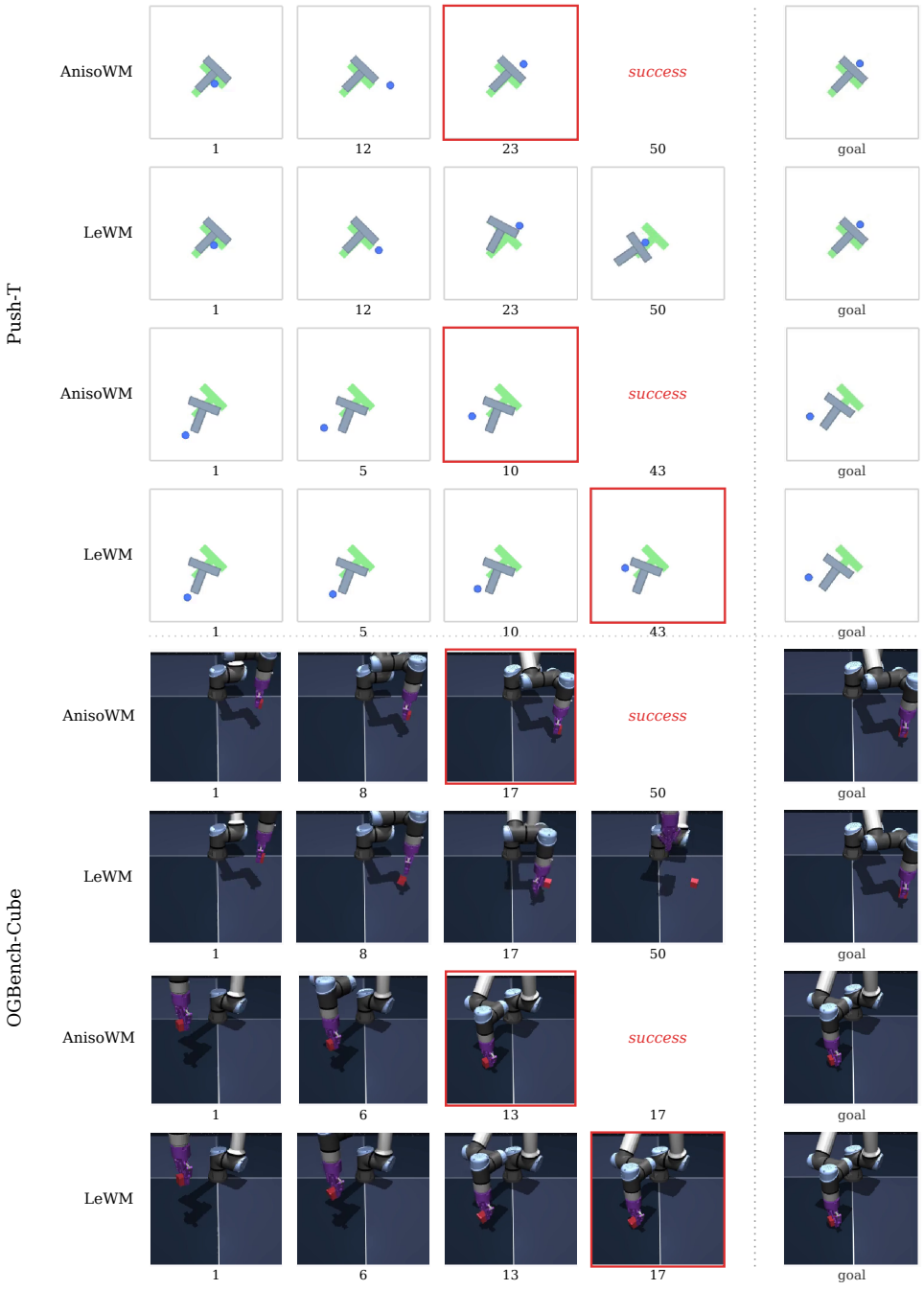

Figure 13: **Rollouts in PushT and Cube.** Two rollout pairs per environment compare AnisoWM at $\kappa = 2$ with LeWM under the same evaluation planner and seed. Columns show the indicated environment steps and the goal observation; red outlines indicate recorded goal attainment.