INVERSE-CONDITIONED WORLD MODELS: SHAPING ACTION RESPONSES FOR VISUAL CONTROL

Anonymous authors
Paper under double-blind review

ABSTRACT

Visual world models support goal-directed control by predicting the consequences of candidate actions in a learned state space. However, regularizing the latent marginal alone does not bound the sensitivity of local action recovery to latent errors. We show that even exact marginal isotropy permits arbitrarily ill-conditioned local action responses and introduce Inverse-Conditioned LeWorldModel (IC-LeWM), a joint-embedding predictive model offering two independently usable approaches to action-response learning beyond marginal regularization. Inverse co-training (ICT) supervises the encoder by recovering actions from observed trajectory pairs; action-Jacobian regularization (AJR) shapes the predictor's response to nearby actions. A linear-Gaussian analysis motivates AJR's scaled-isometry target: equal nonzero singular values minimize action-recovery risk at fixed rank and total gain. Across four visual-control domains and five control interfaces, IC-LeWM configurations using ICT or AJR achieve a mean absolute success-rate gain of 5.12% over LeWM, with equal weight per domain–interface pair. The gains averaged across domains are 4.75% for CEM planning and 3.83% for direct control with freshly trained inverse heads. Geometric diagnostics further show that improved predictor conditioning alone does not ensure better control, so response shaping must also preserve predictive utility.

1 INTRODUCTION

Visual world models support goal-directed control by predicting the consequences of candidate actions before those actions are executed (Hafner et al., 2019b; Hansen et al., 2024). Joint-embedding predictive architectures (JEPAs) provide a reconstruction-free framework for learning predictive representations: encoders map observations to latent states, and predictors estimate target representations rather than pixels (LeCun et al., 2022; Assran et al., 2023; Bardes et al., 2024). LeWorldModel (LeWM) instantiates this framework for control with an action-conditioned predictor, trains the encoder and predictor end-to-end using latent prediction and a sketched isotropic-Gaussian regularizer (SIGReg), and plans in the learned space (Maes et al., 2026; Balestriero & LeCun, 2025). SIGReg prevents collapse by constraining the distribution of latent states, but control also requires that the representation capture how actions change those states. This raises a control-specific question: what does a well-regularized representation tell us about the local effects of alternative actions?

Answering this question requires distinguishing marginal regularity from local action-response geometry. A marginal regularizer describes how states populate the latent space, whereas local control depends on how changing an action moves the next state within that space. At a fixed state, one action direction may have a much weaker latent response than another, so small latent errors can obscure the weaker effect. To formalize why marginal isotropy does not resolve this imbalance, we construct a smooth, invertible reparameterization that preserves the entire standard-Gaussian marginal while making the true action-to-latent Jacobian arbitrarily ill-conditioned on its identifiable directions (Proposition 1). The condition number on these directions is also the tight worst-case factor for relative error amplification in linearized action recovery (Proposition 3). Together, these results show that preventing marginal collapse alone cannot bound local inverse sensitivity.

To identify a geometric target for action recovery, we analyze a linear-Gaussian inverse problem in which equal nonzero singular values minimize action-recovery risk at fixed rank and total gain (Proposition 2). This motivates a scaled-isometry preference for local action responses.

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.

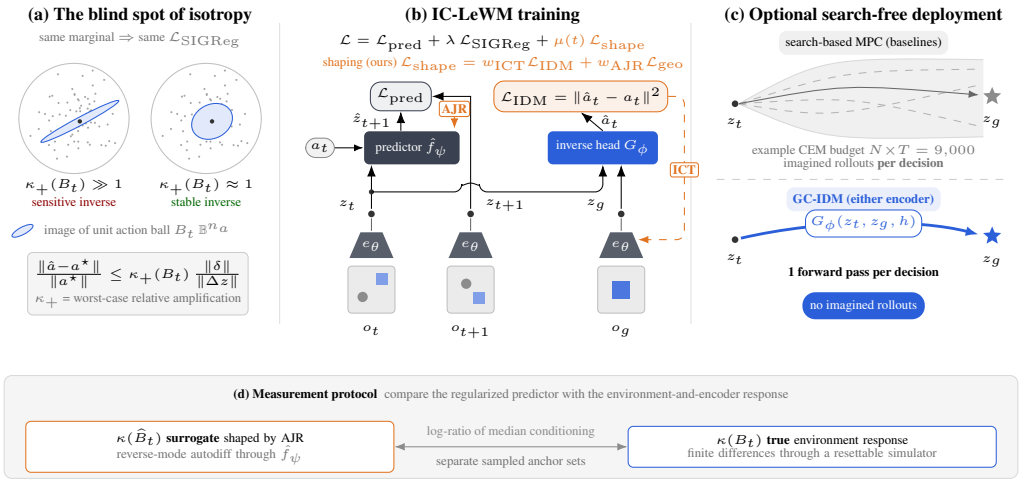

Figure 1: **IC-LeWM offers two independently usable action-response learning approaches.** (a) Identical isotropic latent marginals can coexist with different action-to-latent conditioning. Blue ellipses depict the image of a unit action ball. (b) ICT updates the encoder through inverse regression, while AJR regularizes the predictor's action Jacobian; either signal can be enabled independently. Solid arrows denote forward computation and dashed orange arrows denote the added gradient paths. (c) A separately trained goal-conditioned head provides an optional direct-control interface without imagined rollouts. The search-based example uses 300 candidates and 30 iterations; the direct controller uses one head evaluation. (d) Separate surrogate and simulator-based measurements assess how the learned geometry relates to the environment response.

Building on this analysis, we introduce **Inverse-Conditioned LeWorldModel (IC-LeWM)** within the LeWM backbone. Retaining latent prediction and SIGReg, IC-LeWM offers two independently usable training signals motivated by the same gap: marginal regularity alone does not ensure useful action-response geometry. **Inverse co-training (ICT)** regresses actions from real trajectory pairs, including future goals beyond the prediction window, and propagates this supervision into the encoder. **Action-Jacobian regularization (AJR)** regularizes the predictor's local responses to action perturbations at a fixed latent context. We provide a finite-difference implementation of the scaled-isometry penalty that avoids higher-order differentiation through attention.

Search-based controllers compare the predictor's outputs for candidate action sequences, whereas a goal-conditioned inverse head maps the current representation, goal representation and horizon directly to an action (Ghosh et al., 2019; Brandfonbrener et al., 2023). This optional search-free interface applies to both baseline and IC-LeWM representations and requires one head evaluation per decision, without imagined rollouts (Figure 1c). Figure 2 previews the cross-environment planning results. We assess closed-loop performance across four domains and five control interfaces. The main comparison yields a mean absolute success-rate gain of 5.12% over LeWM, with equal weight per domain–interface pair (Table 2). Averaged across domains, CEM success rises from 85.75% to 90.50%, while direct GC-IDM success rises from 80.50% to 84.33%. For direct control, we freeze each encoder and train a fresh head under matched settings. Our main results use three training seeds; CEM baselines are published LeWM results, and the other baselines are our unshaped controls. Fixed-controller ablations and a separate held-out-episode study examine the training signals in more detail. Geometric diagnostics complement these control results by examining the relationship between response shaping and predictive utility.

Our contributions are:

- **Separation of marginal regularity and local inverse sensitivity.** We show that exact marginal isotropy permits arbitrarily ill-conditioned local action responses, connect conditioning to relative-error amplification, and identify the minimum-risk spectrum for action recovery at fixed rank and total gain in a linear-Gaussian model.

- **IC-LeWM with independently usable ICT and AJR objectives.** We augment LeWM with ICT and AJR objectives: ICT supervises the encoder through action recovery, while AJR shapes the predictor's local response. We provide both Gram-matrix and finite-difference implementations of the geometric penalty.
- **Evaluation across control interfaces.** Across four domains and five interfaces, we compare search-based and direct control, using freshly trained inverse heads and fixed-controller ablations to examine the effects of the training signals. Separate diagnostics compare predictor and simulator-mediated response geometry.

2 Related Work

Inverse dynamics and amortized policies. Inverse dynamics has long provided a representation-learning signal (Agrawal et al., 2016; Pathak et al., 2017) and a pretraining objective for imitation (Brandfonbrener et al., 2023). Goal-conditioned supervised learning and offline reinforcement learning also train direct goal-reaching policies (Ghosh et al., 2019; Park et al., 2023; Eysenbach et al., 2022; Wang et al., 2023; Levine et al., 2020), with benchmarks such as OGBench supporting their evaluation (Park et al., 2025). These approaches instantiate the broader idea of amortizing repeated inference (Gershman & Goodman, 2014). IC-LeWM uses inverse regression as auxiliary supervision for a JEPA encoder, independently of whether deployment uses a search planner or an inverse head. The separate action-Jacobian regularizer acts on the predictive branch, while the fixed-gain analysis supplies a local spectral preference under its stated assumptions.

Action-supervised JEPA world models. PLDM includes inverse dynamics within a composite latent-learning objective (Sobal et al., 2025). More recently, SMWM uses inverse dynamics for end-to-end predictive learning and collapse prevention (Ivashkov et al., 2026), while Delta-JEPA reconstructs actions from latent differences and extends this objective to multistep action sequences (Zhang et al., 2026). Concurrent work combines inverse dynamics with physical-state alignment (Liu et al., 2026). These methods make action-dependent supervision a direct point of comparison. IC-LeWM retains marginal regularization and offers independently usable horizon-conditioned trajectory-pair supervision and predictor action-Jacobian regularization. Its geometric analysis further distinguishes marginal regularity from local inverse sensitivity, and its diagnostics separate predictor geometry from environment responses. The shared-backbone SMWM objective comparison in Appendix E.2 tests one close alternative; it is not a reproduction of the full SMWM system or a matched comparison with Delta-JEPA.

3 Preliminaries

Goal-reaching from pixels. We consider a controlled system with states s, primitive actions $u \in \mathbb{R}^{n_u}$, deterministic transition $s' = \mathrm{env}(s, u)$ and rendering $o = \mathrm{obs}(s)$. The agent acts at a coarser rate than the simulator. A *world-model action* $a_t \in \mathbb{R}^{n_a}$, $n_a = k\, n_u$, concatenates k primitive actions (frameskip k), and we write $s_{t+1} = \mathrm{Env}(s_t, a_t)$. A task is a pair (o_0, o_g) of initial and goal observations, and a controller is scored by whether it reaches the goal within a fixed interaction budget.

JEPA world models. An encoder $e_\theta : \mathbb{R}^{C\times H\times W} \to \mathbb{R}^d$, where C, H, and W denote the image channels, height, and width, produces d-dimensional latents $z_t = e_\theta(o_t)$. An action encoder embeds a_t, and a causal transformer predictor $\hat{f}_\psi$ maps a length-m latent/action context to the next latent, $\hat{z}_{t+1} = \hat{f}_\psi(z_{t-m+1:t}, a_{t-m+1:t})$. Training minimizes latent prediction error plus an anti-collapse term,

$$\mathcal{L}_{\mathrm{pred}} = \mathbb{E}\big\|\hat{f}_\psi(z_{t-m+1:t}, a_{t-m+1:t}) - e_\theta(o_{t+1})\big\|_2^2, \qquad \mathcal{L}_{\mathrm{LeWM}} = \mathcal{L}_{\mathrm{pred}} + \lambda\, \mathcal{L}_{\mathrm{SIGReg}}(\{z\}), \quad (1)$$

where $\mathcal{L}_{\mathrm{SIGReg}}$ is the sketched isotropic-Gaussian regularizer of Balestriero & LeCun (2025). At the population level and in expectation over sketch directions, the regularizer vanishes if and only if the latent marginal is $\mathcal{N}(0, I_d)$. Our analysis uses the fact that this regularizer depends only on the *marginal law* of z. The background discussion in the appendix describes its underlying statistic.

Search and amortization. In the control formulation, H denotes the planning horizon. Given a goal latent $z_g = e_\theta(o_g)$, model predictive control (MPC) minimizes the imagined terminal cost $\min_{a_{t:t+H-1}} \|\hat{z}_{t+H} - z_g\|_2^2$, executes a prefix and replans. Sampling-based solvers such as the cross-entropy method (CEM), its improved variant iCEM, and model predictive path integral (MPPI) evaluate $N \times T$ latent rollouts for N candidates and T refinement iterations (Rubinstein, 1999; Pinneri et al., 2021; Williams et al., 2017). The amortized alternative trains a goal-conditioned inverse-dynamics model (GC-IDM) head $G_\phi(z_t, z_g, h) \mapsto a_t$ by regression on offline trajectories,

$$\mathcal{L}_{\mathrm{IDM}}(\phi) = \mathbb{E}_{t,\ h\sim\mathrm{Unif}\{1,\dots,H_{\max}\}} \big\| G_\phi(z_t, z_{t+h}, h) - a_t \big\|_2^2, \tag{2}$$

where $H_{\max}$ is the inverse-head horizon cap, z_{t+h} is a future latent in the same episode, and a_t is the first action toward it. At test time, the head predicts an action directly from the current state, goal and horizon.

Action-response geometry. We distinguish the environment response from its learned approximation. Local inversion then provides a way to analyze the sensitivity of both maps.

Definition 1 (Action-to-latent maps)**.** Fix a state s_t with observation o_t and a nominal action a_t. The *true* action-to-latent Jacobian is $B_t = \frac{\partial}{\partial a}\, e_\theta\big(\mathrm{obs}(\mathrm{Env}(s_t, a))\big)\big|_{a=a_t} \in \mathbb{R}^{d\times n_a}$. The *surrogate* Jacobian is the corresponding derivative of the learned predictor, $\widehat{B}_t = \frac{\partial}{\partial a} \hat{f}_\psi\big(z_{t-m+1:t}, (a_{t-m+1:t-1}, a)\big)\big|_{a=a_t}$. The *encoder* Jacobian is $J_e = \partial e_\theta(o)/\partial o|_{o_t}$. For a matrix $M \in \mathbb{R}^{d\times n}$ with $d \geq n$ and singular values $\sigma_1 \geq \cdots \geq \sigma_n$ we write $\kappa(M) = \sigma_1/\sigma_n$, with the convention $\kappa(M) = \infty$ when $\sigma_n = 0$. For a nonzero map of rank r, define $\kappa_+(M) = \sigma_1/\sigma_r$ on its *identifiable action subspace* $\mathcal{A}_M = (\ker M)^\perp$. It equals $\kappa(M)$ when M has full column rank.

The predictor provides a differentiable surrogate for the environment response. In our offline training setting, $\widehat{B}_t$ is available for direct regularization, whereas estimating B_t requires interventions in a resettable simulator. These local derivatives describe smooth response regions; near contact events or pixel-rendering discontinuities, finite differences instead characterize responses at the chosen perturbation scale.

Different primitive-action sequences can reach the same endpoint, so local action recovery is defined on the identifiable subspace $\mathcal{A}_{B_t}$. Conditioning describes sensitivity within this subspace, while the remaining directions contribute ambiguity to the regression target. A goal-reaching controller can resolve this ambiguity by selecting any successful action sequence.

4 Inverse-Conditioned World Models

4.1 Objective

IC-LeWM keeps the predictive world model and augments its encoder and predictor with separate training signals. The encoder receives inverse supervision from observed trajectory pairs, and the predictor receives a local action-response penalty:

$$\mathcal{L} \;=\; \mathcal{L}_{\mathrm{pred}} \;+\; \lambda\, \mathcal{L}_{\mathrm{SIGReg}} \;+\; \mu(t)\, \mathcal{L}_{\mathrm{shape}}, \qquad \mathcal{L}_{\mathrm{shape}} \;=\; w_{\mathrm{ICT}}\, \mathcal{L}_{\mathrm{IDM}} \;+\; w_{\mathrm{AJR}}\, \mathcal{L}_{\mathrm{geo}}, \tag{3}$$

where λ weights SIGReg and w_{ICT} and w_{AJR} control the two components. Nonzero weights enable either signal alone or their optional combination. During an initial fraction of training, $\mu(t)$ ramps linearly from 0 to $\mu^\star$, so prediction learning precedes the full shaping signal. Setting $w_{\mathrm{ICT}} = w_{\mathrm{AJR}} = 0$ recovers the unmodified world-model objective and defines the matched baseline.

4.2 Inverse co-training (ICT)

ICT uses an auxiliary inverse head to supervise the representation, independently of the controller used after training. Given an observation and a later frame from the same episode, the head predicts the first action taken between them. We train G_ϕ with (2) and let its gradients pass through both the current and goal latents to the encoder. The prediction objective continues to train future-state prediction, while inverse regression supplies an additional criterion for the encoded observations.

The head is a residual multilayer perceptron (MLP) on $[z_t \,\|\, z_g]$, conditioned on the remaining horizon through AdaLN-Zero (Peebles & Xie, 2023) (Appendix B). Goal frames are sampled from the

same episode independently of the prediction window and encoded in one extra batched pass. This allows inverse training to use goals beyond the prediction window (Appendix B). Two ablations examine the gradient path and the goal source by detaching both latents or substituting one-step predictor outputs for real future latents, respectively.

4.3 Action-Jacobian regularization (AJR)

ICT supervises sampled trajectory pairs; AJR directly constrains the predictor's local response to actions at a fixed context. It regularizes the Jacobian $\widehat{B}_t$ of the map,

$$a \mapsto \hat{f}_\psi(z_{t-m+1:t}, (a_{t-m+1:t-1}, a)).$$

The latent context is detached, so AJR directly updates the predictor and action encoder. Any effect on the observation encoder is indirect through subsequent joint predictive training. We consider two implementations of the local geometric preference.

Finite-difference variant (AJR-FD). We perturb the current action twice and compare the predicted next latents. The loss encourages the distance between these predictions to be proportional to the distance between the actions. For Gaussian perturbations $\delta, \delta' \sim \mathcal{N}(0, \epsilon^2 I)$ around a_t, where ϵ is the perturbation standard deviation and I is the $n_a \times n_a$ identity, we suppress the fixed predictor context in $\hat{f}$ and write

$$\mathcal{L}^{\mathrm{FD}}_{\mathrm{geo}} = \mathbb{E}\Big(\big\| \hat{f}(a_t + \delta) - \hat{f}(a_t + \delta') \big\|_2 \; - \; c \, \|\delta - \delta'\|_2 \Big)^2, \tag{4}$$

where $\Delta\hat{z}$ and Δa are the paired predicted-latent and action displacements, and c tracks the batch median ratio $\|\Delta\hat{z}\|/\|\Delta a\|$ through an exponential moving average (EMA). The tracker uses the current perturbation batch and is detached from the gradient. AJR-FD requires ordinary predictor evaluations and backpropagation, and Lemma 2 connects its distance loss to local Jacobian geometry.

Gram-matrix variant (AJR-Gram). AJR-Gram penalizes $\|\widehat{B}_t^\top \widehat{B}_t - \alpha I\|_F^2$ relative to a detached EMA scale α, with an optional hinge on batch-average gain. The default zero gain floor leaves the hinge inactive. Appendix B.1 gives the full objective and exact and matrix-free estimators. We use AJR-FD by default to avoid higher-order differentiation through attention; AJR-Gram supplies a direct spectral ablation.

4.4 Action responses and local inverse sensitivity

Local inversion provides a tractable way to study why action-response geometry matters. We first examine what marginal isotropy tells us about this inverse problem, then identify the spectra that minimize action-recovery risk under fixed rank and gain. Using local linearized maps and a linear-Gaussian observation model, the analysis provides a geometric preference for the training objective. Proofs are in Appendix A.

Proposition 1 (Distributional isotropy does not constrain conditioning)**.** *Let e_θ have latent marginal $\mathcal{N}(0, I_d)$ with $d \geq 3$. Fix a transition with successor latent $z_{t+1} \neq 0$ and true Jacobian B_t of rank $r \geq 2$. For every $M > 0$ there is a C^∞ diffeomorphism $T : \mathbb{R}^d \to \mathbb{R}^d$ with $T_\# \mathcal{N}(0, I_d) = \mathcal{N}(0, I_d)$ such that the encoder $e' = T \circ e_\theta$ satisfies $\kappa_+(B'_t) \geq M$, where $T_\#$ denotes pushforward, DT is the Jacobian of T, and $B'_t = DT(z_{t+1})B_t$ has the same rank as B_t. Any population regularizer that depends only on the latent marginal, including SIGReg and the variance and covariance terms of Variance-Invariance-Covariance Regularization (VICReg), takes the same value on e_θ and e'.*

The proof constructs a radial twist within a two-dimensional plane. The transformation preserves both Gaussian density and volume, while its differential contains a shear whose magnitude can grow without bound. Thus, identical latent marginals can conceal very different local responses to actions, as illustrated in Figure 1(a).

The construction isolates the role of marginal regularization: invertible reparameterization preserves information and unrestricted Bayes regression risk while allowing Euclidean sensitivity, finite-head approximation, and prediction loss to change (Remark 2).

For linearized recovery on the identifiable action subspace, $\kappa_+(B_t)$ is the tight worst-case factor amplifying relative latent error into relative action error (Proposition 3, Appendix A.3). Null action components cannot be recovered by the minimum-norm inverse.

Inverse regression combines conditional action ambiguity with head approximation error (Appendix A.4). A Gaussian observation model makes the ambiguity term explicit through the response spectrum.

Proposition 2 (Inverse risk at fixed rank and gain)**.** *Let $z_{t+h} = \Lambda z_t + Ba + \eta$, where $\Lambda \in \mathbb{R}^{d\times d}$, $B \in \mathbb{R}^{d\times n_a}$ has rank $r \geq 1$, and $d \geq n_a$. Let $a \sim \mathcal{N}(0, I_{n_a})$ and $\eta \sim \mathcal{N}(0, \tau^2 I_d)$ for $\tau > 0$, and assume that a, η, and z_t are mutually independent. Then*

$$R(B) = \min_G \mathbb{E}\|G(z_t, z_{t+h}) - a\|^2 = \mathrm{tr}[(I + B^\top B/\tau^2)^{-1}] = n_a - r + \sum_{i=1}^{r}(1 + \sigma_i^2/\tau^2)^{-1}. \quad (5)$$

At fixed rank r and gain $\|B\|_F^2 = E > 0$, (i) $R(B) \geq n_a - r + r(1 + E/(r\tau^2))^{-1}$, with equality if and only if $\sigma_1 = \cdots = \sigma_r$; and (ii) $R(B) \geq n_a - r + (1 + E/(\kappa_+(B)^2\tau^2))^{-1}$. If the model additionally has stationary covariance $\mathrm{Cov}(z) = I_d$, then $BB^\top \preceq (1-\tau^2)I_d$ and $E \leq r(1-\tau^2)$, so positive gain requires $\tau^2 < 1$.

The term $n_a - r$ is uncertainty in unobservable action directions. Among the remaining directions, equal singular values minimize risk at fixed gain; for $r \geq 2$ and fixed gain, a diverging κ_+ makes at least one more direction approach its unit prior uncertainty. Here B describes the effective response over the chosen horizon.

AJR applies this spectral preference to the available one-step predictor Jacobian. The theoretical optimum concerns the effective response over a fixed horizon, so its connection to per-step regularization and closed-loop success is evaluated empirically.

Proposition 4 shows that the Gram penalty controls spectral spread and deviation from its target scale. A sufficiently small residual relative to a positive target bounds conditioning, but the self-scaled penalty alone also vanishes under collapse. Lemma 2 connects AJR-FD to the same scaled-isometric maps: with bounded second derivatives and a fixed positive scale, the finite-distance loss differs from its linearized counterpart by $O(\epsilon^3)$. Appendix A gives both statements, proofs, and the finite-batch qualification for the implemented median/EMA tracker. At the default $\epsilon = 0.1$, the nonlinear remainder determines the accuracy of this local approximation.

Transfer requires derivative fidelity. AJR shapes $\widehat{B}_t$, not B_t. Transferring conditioning requires paired derivative error smaller than the surrogate's least singular value (Remark 3, Appendix A.7). A full-rank target can conflict with fidelity when the environment response is rank deficient; AJR's weight balances this preference against prediction fidelity. Separately sampled condition-number medians do not test this transfer premise.

4.5 Control after training

The trained encoder and predictor support the search procedures described in the preliminaries. Direct control instead uses the encoder and a goal-conditioned inverse head without predictor rollouts. In the matched-head evaluations, we freeze each trained encoder and fit a fresh head with the same architecture and training settings within each domain. This separates the auxiliary ICT training head from the head used to compare representations.

The blocked interface executes k primitive actions per head evaluation; the main Push-T and Reacher evaluations use the official base-action interface and observe after each primitive action. Both condition on the remaining budget and permit caching the goal encoding. Appendix B.3 specifies both interfaces.

4.6 Measuring conditioning

We measure $\widehat{B}_t$ with exact reverse-mode Jacobians, B_t with central finite differences from reset simulator states, and J_e with random pixel-space tangent sketches. The stabilized ratio is $\kappa_{\mathrm{num}}(M) = \sigma_{\max}(M)/\max\{\sigma_{\min}(M), 10^{-12}\}$. We report medians and interquartile ranges, interpreting the ratio alongside singular values because near-null directions can dominate it. The

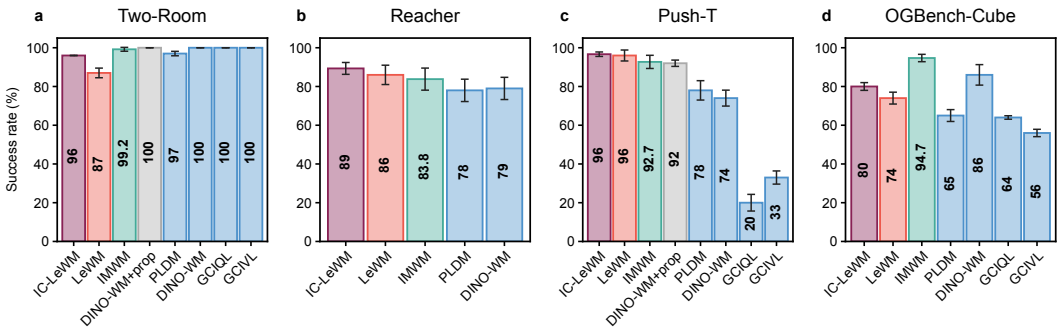

Figure 2: **Goal-reaching performance across four environments.** IC-LeWM uses ICT on Push-T/Reacher and AJR-FD on Two-Room/Cube. IC-LeWM bars show mean CEM success over three independent training seeds, with SD error bars; integer labels truncate the means. LeWM and the other baseline scores and error bars are from Maes et al. (2026); IMWM uses mean and SD over 12 evaluation cells (Gao et al., 2026). IMWM uses 300×30 CEM queries on all tasks; ours uses 300×30 on Push-T and 300×10 elsewhere. This is a descriptive cross-system comparison.

conditioning gap is the natural log-ratio of separately sampled true and surrogate medians, not a paired derivative error. Appendix B.4 specifies these separately sampled diagnostics; Appendix E.4 adds matched recorded-state anchors, numerical rank thresholds, gain, and paired derivative errors at three perturbation scales.

5 Experiments

We evaluate whether action supervision and response shaping improve goal reaching across controllers, then use component ablations and geometric diagnostics to examine their contributions.

5.1 Domains and evaluation protocol

Tasks and evaluation protocol. We consider Push-T (Chi et al., 2025), Two-Room, OGBench-Cube (Park et al., 2025), and Reacher (Tassa et al., 2018). Following the three-training-seed evaluation setting of LeWorldModel (Maes et al., 2026, Appendix G, Table 5), we report main results as mean and standard deviation across three independent training seeds. Each evaluation uses 50 starts, goals 25 primitive steps ahead, and an interaction budget of 50. Reacher uses a per-joint error threshold of 0.1 radians. The main evaluation follows the in-dataset protocol; a separate episode-level $80/10/10$ split tests transfer to held-out episodes (Appendix E.2).

Models and controllers. All models retain LeWM's ViT-Tiny encoder, 192-dimensional latents, six-layer predictor, and optimization settings (Dosovitskiy et al., 2020; Maes et al., 2026; Loshchilov & Hutter, 2017). Frozen encoders receive fresh GC-IDM heads trained for 50 epochs under matched within-domain settings. We also evaluate CEM, iCEM, MPPI, and gradient planning. S1 denotes unshaped LeWM; S2 uses ICT alone on Push-T/Reacher and AJR-FD alone on Two-Room/Cube. CEM S1 values are published references, while the remaining S1 results use our unshaped controls. Appendix C.1 gives action interfaces and budgets.

5.2 Control gains depend on the deployment interface

Direct control with freshly trained heads. ICT improves mean GC-IDM success from 82.67% to 87.33% on Push-T and from 42.00% to 51.33% on Reacher (Table 2); Two-Room and Cube are at or near saturation. The heads share the same deployment architecture and training budget, so the comparison evaluates the learned representations through a common readout. In the separate held-out Push-T study, ICT improves success from 70% to 78% (Appendix E.2).

Search benefits depend on the controller. The selected IC-LeWM CEM means exceed the published LeWM scores in all four domains (Figure 2). Within our other planner comparisons, ICT

Table 1: **Goal-reaching mean success rates (%).** CEM S1 uses official LeWorldModel scores (Maes et al., 2026); other S1 entries are local unshaped controls. S2* uses ICT and S2† uses AJR-FD. Means are truncated to integers for display. Bold compares the untruncated means, including ties; equal displayed integers need not indicate equal means. Full means and uncertainties appear in Table 2. Sampling budgets count rollouts; gradient budgets count updates. The slash separates Push-T/other domains for CEM, and non-Reacher/Reacher for the other search planners.

		Push-T		Two-Room		Cube		Reacher	
Planner	**Search budget**	S1	S2*	S1	S2†	S1	S2†	S1	S2*
CEM	9000/3000	96	**96**	87	**96**	74	**80**	86	**89**
iCEM	9000/3000	90	**90**	88	**96**	74	**77**	**88**	85
MPPI	9000/3000	58	**78**	80	**89**	52	**58**	**67**	65
Gradient	30/10	6	**10**	40	**61**	47	**57**	**83**	74
GC-IDM	0	82	**87**	**100**	**100**	97	**98**	42	**51**

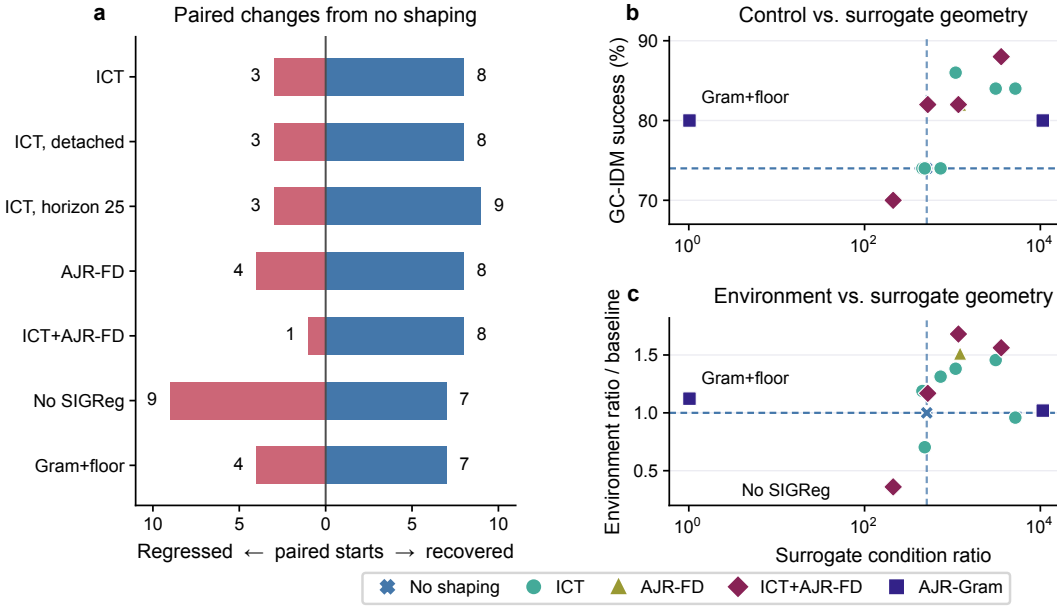

Figure 3: **Paired outcomes and response geometry separate the effects of shaping.** All panels use the Push-T parameter study in Table 5. **(a)** Selected controls report paired changes relative to the unshaped model: recovered starts are baseline failures that become successes, whereas regressed starts are baseline successes that become failures. The combined model uses $\mu^\star = 0.1$. **(b)** Across all fourteen configurations, lower surrogate condition ratios do not order GC-IDM success. **(c)** Environment ratios, normalized by the unshaped median, do not mirror the sharp surrogate change induced by the Gram gain floor. Dashed lines mark the corresponding unshaped references. Geometry points are medians over separate sets of 256 predictor contexts and 64 simulator anchors. The stabilized ratios include near-null directions and are neither paired derivative errors nor estimates of variation across training seeds.

raises Push-T MPPI success from 58.00% to 78.00%, and AJR-FD raises Two-Room gradient planning from 40.00% to 61.33%. Reacher iCEM instead decreases from 88.67% to 85.33%. The results support evaluating each training signal through its intended control interface rather than assuming that one representation ranks best for every solver.

5.3 Fixed-controller ablations: gains, supervision and training-path controls

The Push-T parameter study fixes the GC-IDM interface, head width, head-training schedule, goal offset and interaction budget. All fourteen world models use epoch-10 checkpoints with seed 42, receive fresh heads with $H_{\max} = 50$, and are evaluated on exactly the same 50 episode/start pairs. The 74% unshaped reference is a separate checkpoint from those in Table 1. Tables 5 and 13 compare ICT and AJR individually and in combination, alongside training choices and response geometry under this protocol. Combining the signals need not improve over either alone, as illustrated by the Cube comparison in Appendix D.2.

The largest gain comes with fewer regressions on solved starts. ICT, AJR-FD, and their combination attain 84%, 82%, and 88% success, respectively, versus 74% without shaping (Table 5). On the same 50 starts, the combined model recovers eight baseline failures while losing one baseline success (Figure 3a). ICT also recovers eight failures but loses three successes. These paired outcomes show that the aggregate gain is accompanied by fewer regressions relative to the unshaped model. These fixed-controller, single-seed results complement the three-seed main comparison.

Supporting controls. Removing SIGReg reduces combined-model success from 88% to 70%. Real future goals outperform predictor-generated goals (84% versus 74%), but this comparison also changes temporal coverage. Training-horizon controls favor longer trajectory pairs within these runs without a monotonic trend. Combined weights $\mu^\star \in \{0.05, 0.1, 0.2\}$ yield 82%, 88%, and 82%; the strongest action fit occurs at 0.2, not at the best control score. Appendix D.3 provides the paired-outcome, horizon, and action-fit analyses. The detached-ICT control does not isolate the effect of encoder gradients; Appendix E.1 examines its update path.

5.4 Predictor conditioning is not a standalone explanation of success

Near-isometry can change the diagnostic without changing control. Adding a gain floor to AJR-Gram reduces the median surrogate ratio from 10654.98 to 1.02, yet GC-IDM success remains 80% (Figure 3b). The best parameter-study result, 88%, instead occurs at a ratio of 3585.54, compared with 509.98 for the unshaped baseline. Surrogate conditioning therefore neither ranks these controllers nor identifies the best observed configuration. The fixed-rank, fixed-gain inverse-risk result isolates spectral effects, whereas rank, gain, observation noise, and inverse-head approximation error are not held fixed across these learned checkpoints. Moreover, GC-IDM does not query the predictor at deployment, so any AJR effect must be mediated by joint representation training.

Prediction fidelity limits the usefulness of spectral shaping. A separate shape–gain study explains why conditioning alone is an insufficient criterion (Appendix E.3). Adding a normalized shape penalty to gain control reduces the predictor condition number from 2350.8 to 1.06, but increases validation prediction loss from 0.0057 to 0.2324 and reduces CEM success from 96% to 12%. These exact-Jacobian controls use different objectives from default AJR-FD and expose a tradeoff between aggressive spectral equalization and predictive fidelity. Paired simulator measurements likewise show that a near-isometric predictor need not approximate the environment response accurately (Appendix E.4). Together, the diagnostics motivate assessing response geometry alongside prediction and closed-loop performance.

6 Conclusion

IC-LeWM offers independently usable encoder-side action recovery and predictor-side response shaping. Its analysis separates marginal isotropy from local inverse sensitivity and identifies balanced nonzero spectra as optimal at fixed rank and gain in a Gaussian inverse model. The main control comparison shows average gains across domains and interfaces, with further support from fixed-controller and held-out-episode studies. The benefits depend on the deployment interface. Paired response diagnostics and shape–gain controls show that near-isometric predictor responses can coexist with poor derivative fidelity and predictive utility. Action-response shaping should therefore be assessed jointly with prediction fidelity and performance under the intended controller.

AI USE STATEMENT

Generative AI tools assisted with language polishing and editing of the manuscript and with figure preparation. The authors are responsible for the final manuscript, proofs, citations, code, and experimental reporting.

REFERENCES

Pulkit Agrawal, Ashvin V Nair, Pieter Abbeel, Jitendra Malik, and Sergey Levine. Learning to poke by poking: Experiential learning of intuitive physics. *Advances in neural information processing systems*, 29, 2016.

Mahmoud Assran, Quentin Duval, Ishan Misra, Piotr Bojanowski, Pascal Vincent, Michael Rabbat, Yann LeCun, and Nicolas Ballas. Self-supervised learning from images with a joint-embedding predictive architecture. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15619–15629. IEEE, 2023.

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

Nitin Bansal, Xiaohan Chen, and Zhangyang Wang. Can we gain more from orthogonality regularizations in training deep networks? *Advances in neural information processing systems*, 31, 2018.

Adrien Bardes, Jean Ponce, and Yann LeCun. Vicreg: Variance-invariance-covariance regularization for self-supervised learning. *arXiv preprint arXiv:2105.04906*, 2021.

Adrien Bardes, Quentin Garrido, Jean Ponce, Xinlei Chen, Michael Rabbat, Yann LeCun, Mahmoud Assran, and Nicolas Ballas. Revisiting feature prediction for learning visual representations from video. *arXiv preprint arXiv:2404.08471*, 2024.

David Brandfonbrener, Ofir Nachum, and Joan Bruna. Inverse dynamics pretraining learns good representations for multitask imitation. *Advances in Neural Information Processing Systems*, 36: 66953–66978, 2023.

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pp. 1597–1607. PmLR, 2020.

Cheng Chi, Zhenjia Xu, Siyuan Feng, Eric Cousineau, Yilun Du, Benjamin Burchfiel, Russ Tedrake, and Shuran Song. Diffusion policy: Visuomotor policy learning via action diffusion. *The International Journal of Robotics Research*, 44(10-11):1684–1704, 2025.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

Thomas W Epps and Lawrence B Pulley. A test for normality based on the empirical characteristic function. *Biometrika*, 70(3):723–726, 1983.

Benjamin Eysenbach, Tianjun Zhang, Sergey Levine, and Russ R Salakhutdinov. Contrastive learning as goal-conditioned reinforcement learning. *Advances in Neural Information Processing Systems*, 35:35603–35620, 2022.

Baoqi Gao, Ruize Han, Miao Wang, and Song Wang. IMWM: Intuition models complement world models for latent planning. *arXiv preprint arXiv:2606.01626*, 2026. URL https://arxiv.org/abs/2606.01626.

Quentin Garrido, Randall Balestriero, Laurent Najman, and Yann Lecun. Rankme: Assessing the downstream performance of pretrained self-supervised representations by their rank. In *International conference on machine learning*, pp. 10929–10974. PMLR, 2023.

Samuel Gershman and Noah Goodman. Amortized inference in probabilistic reasoning. In *Proceedings of the annual meeting of the cognitive science society*, volume 36, 2014.

Dibya Ghosh, Abhishek Gupta, Ashwin Reddy, Justin Fu, Coline Devin, Benjamin Eysenbach, and Sergey Levine. Learning to reach goals via iterated supervised learning. *arXiv preprint arXiv:1912.06088*, 2019.

Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pires, Zhaohan Guo, Mohammad Gheshlaghi Azar, et al. Bootstrap your own latent-a new approach to self-supervised learning. *Advances in neural information processing systems*, 33:21271–21284, 2020.

Amos Gropp, Matan Atzmon, and Yaron Lipman. Isometric autoencoders. *arXiv preprint arXiv:2006.09289*, 2020.

David Ha and Jürgen Schmidhuber. Recurrent world models facilitate policy evolution. *Advances in neural information processing systems*, 31, 2018.

Danijar Hafner, Timothy Lillicrap, Jimmy Ba, and Mohammad Norouzi. Dream to control: Learning behaviors by latent imagination. *arXiv preprint arXiv:1912.01603*, 2019a.

Danijar Hafner, Timothy Lillicrap, Ian Fischer, Ruben Villegas, David Ha, Honglak Lee, and James Davidson. Learning latent dynamics for planning from pixels. In *International conference on machine learning*, pp. 2555–2565. PMLR, 2019b.

Danijar Hafner, Jurgis Pasukonis, Jimmy Ba, and Timothy Lillicrap. Mastering diverse control tasks through world models. *Nature*, 640(8059):647–653, 2025.

Nick Hansen, Hao Su, and Xiaolong Wang. Td-mpc2: Scalable, robust world models for continuous control. In B. Kim, Y. Yue, S. Chaudhuri, K. Fragkiadaki, M. Khan, and Y. Sun (eds.), *International Conference on Learning Representations*, volume 2024, pp. 47376–47405, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/file/cf73d57b6dcda32b293df7c2d5341f49-Paper-Conference.pdf.

Olivier J Hénaff, Robbe LT Goris, and Eero P Simoncelli. Perceptual straightening of natural videos. *Nature neuroscience*, 22(6):984–991, 2019.

Judy Hoffman, Daniel A Roberts, and Sho Yaida. Robust learning with jacobian regularization. *arXiv preprint arXiv:1908.02729*, 2019.

Roger A Horn and Charles R Johnson. *Matrix analysis*. Cambridge university press, 2012.

Michael F Hutchinson. A stochastic estimator of the trace of the influence matrix for laplacian smoothing splines. *Communications in Statistics-Simulation and Computation*, 18(3):1059–1076, 1989.

Petr Ivashkov, Randall Balestriero, and Bernhard Schölkopf. Sensorimotor world models: Perception for action via inverse dynamics, 2026. URL https://arxiv.org/abs/2606.20104.

Yann LeCun et al. A path towards autonomous machine intelligence version 0.9. 2, 2022-06-27. *Open Review*, 62(1):1–62, 2022.

Sergey Levine, Aviral Kumar, George Tucker, and Justin Fu. Offline reinforcement learning: Tutorial, review, and perspectives on open problems. *arXiv preprint arXiv:2005.01643*, 2020.

Muyuan Liu, Yue Huang, Zheng Liang, and Xiang Gao. Toward physically grounded JEPA world models for goal-conditioned robotic planning, 2026. URL https://arxiv.org/abs/2609.03565.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017.

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

Takeru Miyato, Toshiki Kataoka, Masanori Koyama, and Yuichi Yoshida. Spectral normalization for generative adversarial networks. *arXiv preprint arXiv:1802.05957*, 2018.

Seohong Park, Dibya Ghosh, Benjamin Eysenbach, and Sergey Levine. Hiql: Offline goal-conditioned rl with latent states as actions. *Advances in Neural Information Processing Systems*, 36:34866–34891, 2023.

Seohong Park, Kevin Frans, Benjamin Eysenbach, and Sergey Levine. Ogbench: Benchmarking offline goal-conditioned rl. In Y. Yue, A. Garg, N. Peng, F. Sha, and R. Yu (eds.), *International Conference on Learning Representations*, volume 2025, pp. 94937–94982, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/file/ecd92623ac899357312aaa8915853699-Paper-Conference.pdf.

Deepak Pathak, Pulkit Agrawal, Alexei A Efros, and Trevor Darrell. Curiosity-driven exploration by self-supervised prediction. In *International conference on machine learning*, pp. 2778–2787. PMLR, 2017.

William Peebles and Saining Xie. Scalable diffusion models with transformers. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 4172–4182. IEEE, 2023.

Cristina Pinneri, Shambhuraj Sawant, Sebastian Blaes, Jan Achterhold, Joerg Stueckler, Michal Rolinek, and Georg Martius. Sample-efficient cross-entropy method for real-time planning. In *Conference on Robot Learning*, pp. 1049–1065. PMLR, 2021.

Salah Rifai. Contractive auto-encoders: Explicit invariance during feature extraction. In *International Conference on Machine Learning (ICML)*, pp. 833–840, 2011.

Olivier Roy and Martin Vetterli. The effective rank: A measure of effective dimensionality. In *2007 15th European signal processing conference*, pp. 606–610. IEEE, 2007.

Reuven Rubinstein. The cross-entropy method for combinatorial and continuous optimization. *Methodology and computing in applied probability*, 1(2):127–190, 1999.

Julian Schrittwieser, Ioannis Antonoglou, Thomas Hubert, Karen Simonyan, Laurent Sifre, Simon Schmitt, Arthur Guez, Edward Lockhart, Demis Hassabis, Thore Graepel, et al. Mastering atari, go, chess and shogi by planning with a learned model. *Nature*, 588(7839):604–609, 2020.

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 D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen (eds.), *Advances in Neural Information Processing Systems*, volume 38, Main Conference, pp. 43905–43941. Curran Associates, Inc., 2025. doi: 10.52202/085713-1465. URL https://proceedings.neurips.cc/paper_files/paper/2025/file/3e7cf447f21cd11c846463affefce665-Paper-Conference.pdf.

Yuval Tassa, Yotam Doron, Alistair Muldal, Tom Erez, Yazhe Li, Diego de Las Casas, David Budden, Abbas Abdolmaleki, Josh Merel, Andrew Lefrancq, et al. Deepmind control suite. *arXiv preprint arXiv:1801.00690*, 2018.

Tongzhou Wang, Antonio Torralba, Phillip Isola, and Amy Zhang. Optimal goal-reaching reinforcement learning via quasimetric learning. In *International Conference on Machine Learning*, pp. 36411–36430. PMLR, 2023.

Grady Williams, Andrew Aldrich, and Evangelos A Theodorou. Model predictive path integral control: From theory to parallel computation. *Journal of Guidance, Control, and Dynamics*, 40 (2):344–357, 2017.

Jure Zbontar, Li Jing, Ishan Misra, Yann LeCun, and Stéphane Deny. Barlow twins: Self-supervised learning via redundancy reduction. In *International conference on machine learning*, pp. 12310–12320. PMLR, 2021.

Zhenghao Zhang, Yuanxiang Wang, Zhenyu Guan, Yujia Yang, Bingkang Shi, Tianyu Zong, Hongzhu Yi, Guoqing Chao, Xingchen Chen, Tiankun Yang, Chenxi Bao, Tao Yu, Jingjing Zhou, and Jungang Xu. Delta-JEPA: Learning action-sensitive world models via latent difference decoding, 2026. URL https://arxiv.org/abs/2606.31232.

Gaoyue Zhou, Hengkai Pan, Yann LeCun, and Lerrel Pinto. Dino-wm: World models on pre-trained visual features enable zero-shot planning. *arXiv preprint arXiv:2411.04983*, 2024.

The appendix contains proofs (Appendix A), implementation details (Appendix B), experimental protocols (Appendix C), supplementary results and ablations (Appendix D), additional controlled studies (Appendix E), and extended related work (Appendix F).

A PROOFS

A.1 A SHEAR LEMMA

Lemma 1. *Let $A \in \mathbb{R}^{d\times n}$ have rank $n \geq 2$, let $e_1, e_2 \in \mathbb{R}^d$ be orthonormal with $\gamma := \|A^\top e_1\| > 0$, and set $S_s = I_d + s\, e_2 e_1^\top$ for $s > 0$. Then*

$$\kappa(S_s A) \;\geq\; \frac{s\gamma - \sigma_{\max}(A)}{\sigma_{\max}(A)} \;\xrightarrow[s\to\infty]{}\; \infty. \tag{6}$$

Proof. A shear can stretch one action direction without stretching all of them. To exhibit these two directions, let $c_1 = A^\top e_1/\gamma$ and choose a unit $c_0 \perp c_1$, which exists because $n \geq 2$. Then

$$S_s A c_1 = A c_1 + s\gamma e_2, \qquad S_s A c_0 = A c_0.$$

The variational characterizations of the extreme singular values give

$$\sigma_{\max}(S_s A) \geq s\gamma - \sigma_{\max}(A), \qquad \sigma_{\min}(S_s A) \leq \|A c_0\| \leq \sigma_{\max}(A).$$

Since $e_1^\top e_2 = 0$, the inverse shear is $S_s^{-1} = I_d - s e_2 e_1^\top$; thus $S_s A$ retains full column rank. Dividing the bounds proves the claim whenever the numerator is nonnegative. Otherwise the displayed lower bound is negative and is automatic. Its divergence follows from $\gamma > 0$. □

A.2 PROOF OF PROPOSITION 1

Proof. We construct a Gaussian-preserving twist whose derivative is an arbitrarily strong shear at the chosen successor. First remove the null action directions: let $V_r \in \mathbb{R}^{n_a \times r}$ have orthonormal columns spanning $\mathcal{A}_{B_t}$ and put $C_t = B_t V_r$. Then C_t has full column rank and the same positive singular values as B_t.

Choose a plane intersecting the response. Choose a unit v such that $v^\top C_t \neq 0$ and $v^\top z_{t+1} \neq 0$. Such a choice exists because the excluded sets are two proper linear subspaces. As $d \geq 3$, choose a unit $y \perp \operatorname{span}\{v, z_{t+1}\}$. In the plane $P = \operatorname{span}\{v, y\}$, write $\Pi_P z = \rho u(\varphi)$ with $u(\varphi) = \cos\varphi\, v + \sin\varphi\, y$. At the successor, $\rho_t = |v^\top z_{t+1}| > 0$ and $u(\varphi_t) = \pm v$.

Twist the plane without changing the Gaussian law. Take $0 < \rho_1 < \rho_t < \rho_2$ and a smooth angle ϑ supported inside (ρ_1, ρ_2). Define

$$T(z) = \rho u\big(\varphi + \vartheta(\rho)\big) + (I - \Pi_P) z. \tag{7}$$

The map is smooth at $\rho = 0$ because the angle vanishes nearby, and its inverse replaces ϑ by $-\vartheta$. It preserves the radius and the perpendicular component. In polar coordinates its Jacobian is triangular with determinant one; the polar volume factor ρ is unchanged. Hence T preserves both volume and $\|z\|$, and therefore the standard-Gaussian density.

Make the local response arbitrarily ill-conditioned. In orthonormal radial/tangential frames, the differential on P is

$$\begin{pmatrix} 1 & 0 \\ \rho\vartheta'(\rho) & 1 \end{pmatrix},$$

and it is the identity on $P^\perp$. Thus $DT(z_{t+1}) = Q_2 S_s Q_1^\top$ for orthogonal Q_1, Q_2, where $s = \rho_t \vartheta'(\rho_t)$ and $Q_1 e_1 = \pm v$. Any $s > 0$ can be realized by $\vartheta(\rho) = (s/\rho_t)(\rho - \rho_t)\chi(\rho)$, where χ is a smooth bump equal to one near ρ_t. Because DT is invertible, the chain rule preserves the action kernel and rank: $B_t' = DT(z_{t+1}) B_t$ and $\ker B_t' = \ker B_t$. Applying Lemma 1 to $A = Q_1^\top C_t$ gives

$$\kappa_+(B_t') = \kappa(S_s A) \geq \frac{s\|C_t^\top v\| - \sigma_{\max}(C_t)}{\sigma_{\max}(C_t)}.$$

The right-hand side exceeds any prescribed M for sufficiently large s. Meanwhile $T_\# \mathcal{N}(0, I_d) = \mathcal{N}(0, I_d)$, so every population functional of the latent marginal retains exactly its original value. □

Remark 1. The twist angle is supported on the annulus $\{\rho_1 \le \|\Pi_P z\| \le \rho_2\}$, whose Gaussian measure can be made arbitrarily small around ρ_t. Thus the construction can localize the coordinate change to a set of arbitrarily small measure.

Remark 2 (Geometry is distinct from information). Proposition 1 concerns population marginal regularization, not the entire training objective. An invertible reparameterization preserves the information in the latent pair and hence unrestricted Bayes action-regression risk. It can change Euclidean error sensitivity, finite-head approximation and the coordinate-dependent prediction loss. Proposition 2 compares response maps under a fixed isotropic noise model rather than asserting that a change of coordinates destroys information.

A.3 Error amplification: statement and proof

We now consider how a small error in the goal latent affects the recovered action. The following result gives a tight worst-case bound for the linearized inverse.

Proposition 3 (Error amplification on identifiable directions)**.** *Let B_t have rank $r \ge 1$, let $a^\star \in \mathcal{A}_{B_t}$ be a nonzero local action perturbation, and set $\Delta z = B_t a^\star$. The minimum-norm linearized inverse $\hat{a}(\cdot) = B_t^+(\cdot)$, where B_t^+ is the Moore–Penrose pseudoinverse, obeys*

$$\frac{\|\hat{a}(\Delta z + \delta) - a^\star\|}{\|a^\star\|} \;\le\; \kappa_+(B_t)\,\frac{\|\delta\|}{\|\Delta z\|}. \tag{8}$$

The constant is tight. Equality is attained when $a^\star$ is a top right singular vector and δ is parallel to the left singular vector for σ_r. For an unrestricted action a, the noiseless inverse recovers its orthogonal projection onto $\mathcal{A}_{B_t}$ rather than its component in $\ker B_t$.

Proof. The pseudoinverse satisfies $B_t^+ B_t = P_{\mathcal{A}_{B_t}}$, the orthogonal projection onto $\mathcal{A}_{B_t}$. Hence $B_t^+ B_t a^\star = a^\star$ and the error is $B_t^+\delta$. Using $\|B_t^+\|_2 = 1/\sigma_r$ and $\|\Delta z\| \le \sigma_1\|a^\star\|$ gives (8). To attain the bound, we place the signal in the strongest direction and the error in the weakest identifiable direction. For a compact SVD $B_t = U_r\Sigma_r V_r^\top$, take $a^\star = v_1$ and $\delta = \varepsilon u_r$ with $\varepsilon > 0$. Then $\Delta z = \sigma_1 u_1$ and $B_t^+\delta = (\varepsilon/\sigma_r)v_r$, so both sides equal ε/σ_r. For arbitrary a, the projection identity gives the final claim. □

A.4 Proof of Proposition 2

The inverse-regression loss also has a distribution-free interpretation. Write $X = (z_t, z_{t+h}, h)$ and $m(X) = \mathbb{E}[a_t \mid X]$. For any offline distribution with $\mathbb{E}\|a_t\|^2 < \infty$ and any measurable head with $\mathbb{E}\|G(X)\|^2 < \infty$,

$$\mathbb{E}\|G(X) - a_t\|^2 = \mathbb{E}\,\mathrm{tr}\,\mathrm{Cov}(a_t \mid X) + \mathbb{E}\|G(X) - m(X)\|^2. \tag{9}$$

The first term measures action ambiguity given the features, and the second measures the head's approximation error. ICT optimizes their sum within the chosen encoder and head classes; a reduction need not arise from both terms. To express the ambiguity term through a singular spectrum, we next consider a Gaussian observation model at a fixed horizon.

Proof. The conditional mean separates unavoidable ambiguity from approximation error. Writing $G(X) - a_t = (G(X) - m(X)) + (m(X) - a_t)$, the cross term has zero expectation conditional on X. This proves (9).

Compute the optimal inverse risk. Conditioning on z_t, subtract the known term Λz_t to obtain $y = z_{t+h} - \Lambda z_t = Ba + \eta$. Since (a, η) is independent of z_t, conditioning additionally on z_t does not change $p(a \mid y)$. Completing the square yields the posterior covariance and mean

$$\Sigma_{a|y} = (I + B^\top B/\tau^2)^{-1}, \qquad m(y) = \Sigma_{a|y} B^\top y/\tau^2.$$

The conditional mean minimizes squared error, so the Bayes risk is $\mathrm{tr}\,\Sigma_{a|y}$. The r positive eigenvalues of $B^\top B$ contribute $(1 + \sigma_i^2/\tau^2)^{-1}$ each; its $n_a - r$ zero eigenvalues contribute one each. This gives (5).

Distribute a fixed gain across observable directions. Let $g(x) = (1 + x/\tau^2)^{-1}$. Since $g''(x) = 2\tau^{-4}(1 + x/\tau^2)^{-3} > 0$, Jensen's inequality and $\sum_{i=1}^{r} \sigma_i^2 = E$ imply

$$\sum_{i=1}^{r} g(\sigma_i^2) \geq r g(E/r).$$

Equality holds exactly when all positive singular values coincide, proving (i). For (ii), retain only the weakest positive-direction contribution and use $\sigma_r^2 = \sigma_1^2/\kappa_+(B)^2 \leq E/\kappa_+(B)^2$. As g is decreasing, $R(B) \geq n_a - r + g(E/\kappa_+(B)^2)$.

Account for the optional stationary covariance budget. Independence and $\mathrm{Cov}(z_t) = \mathrm{Cov}(z_{t+h}) = I_d$ give

$$I_d = \Lambda\Lambda^\top + BB^\top + \tau^2 I_d.$$

Therefore $BB^\top \preceq (1 - \tau^2)I_d$ and $E \leq r(1 - \tau^2)$, with positive gain requiring $\tau^2 < 1$. For $0 < E \leq r(1 - \tau^2)$ the equal-spectrum solution is compatible with this budget: choose $\Lambda\Lambda^\top = (1 - \tau^2)I_d - BB^\top \succeq 0$. □

A.5 Geometric penalty: statement and proof

Proposition 4 (What the geometric penalty optimizes)**.** *Let $B \in \mathbb{R}^{d\times n}$ with $d \geq n$, put $\bar{\alpha} = \|B\|_F^2/n$, and let $\alpha \geq 0$ be a target. (i) $\|B^\top B - \alpha I\|_F^2 = \sum_i (\sigma_i^2 - \alpha)^2 = n\,\mathrm{Var}(\sigma^2) + n(\bar{\alpha} - \alpha)^2$, so at $\alpha = \bar{\alpha}$ the penalty is n times the variance of $\{\sigma_i^2\}$, and for B of rank n it vanishes if and only if $\kappa(B) = 1$. (ii) If $\|B^\top B - \alpha I\|_F \leq \epsilon < \alpha$ then $\kappa(B) \leq \sqrt{(\alpha + \epsilon)/(\alpha - \epsilon)} \leq 1 + \epsilon/(\alpha - \epsilon)$. (iii) At $\alpha = \bar{\alpha}$ the penalty is scale-covariant, $\|(cB)^\top(cB) - c^2\bar{\alpha}I\|_F^2 = c^4\|B^\top B - \bar{\alpha}I\|_F^2$, so on its own it is driven to zero by $B \to 0$. A per-matrix hinge $\xi[c_0 - \bar{\alpha}]_+^2$ with $c_0 > 0$ and $\xi > 0$ makes $B = 0$ incur positive loss, whereas a scaled isometry with $\bar{\alpha} \geq c_0$ has zero loss at its self-consistent target.*

Part (ii) shows that a sufficiently small Gram residual guarantees good conditioning relative to the positive target scale. In the implementation, α follows a moving average rather than the current $\bar{\alpha}$. Part (i) then shows that the loss also penalizes changes in scale through $n(\bar{\alpha} - \alpha)^2$, while the bound in part (ii) still applies. The next lemma connects the finite-difference objective to the same scaled-isometric maps.

Proof. (i) Let $B = U\Sigma V^\top$ be a thin SVD. Then $B^\top B - \alpha I = V\mathrm{diag}(\sigma_i^2 - \alpha)V^\top$, and unitary invariance of the Frobenius norm gives $\|B^\top B - \alpha I\|_F^2 = \sum_i (\sigma_i^2 - \alpha)^2$. Expanding around the mean $\bar{\alpha} = \frac{1}{n}\sum_i \sigma_i^2$ yields $\sum_i (\sigma_i^2 - \alpha)^2 = \sum_i (\sigma_i^2 - \bar{\alpha})^2 + n(\bar{\alpha} - \alpha)^2$ because the cross term vanishes by definition of the mean. The first sum is $n\,\mathrm{Var}(\sigma^2)$. At $\alpha = \bar{\alpha}$, the expression vanishes if and only if all σ_i^2 are equal; because B has rank n, this condition is equivalent to $\kappa(B) = 1$.

(ii) A bound on the total Gram residual bounds each spectral deviation separately. From (i), $|\sigma_i^2 - \alpha| \leq \left(\sum_j (\sigma_j^2 - \alpha)^2\right)^{1/2} \leq \epsilon$ for every i. Hence $\sigma_i^2 \in [\alpha - \epsilon, \alpha + \epsilon]$ and $\kappa(B)^2 = \sigma_1^2/\sigma_n^2 \leq (\alpha + \epsilon)/(\alpha - \epsilon)$. If $x = 2\epsilon/(\alpha - \epsilon) \geq 0$, the last ratio equals $1 + x$, and $\sqrt{1 + x} \leq 1 + x/2 = 1 + \epsilon/(\alpha - \epsilon)$.

(iii) Replacing B by cB multiplies every σ_i^2 and $\bar{\alpha}$ by c^2, so by (i) the penalty at $\alpha = \bar{\alpha}$ scales by c^4; it therefore tends to 0 as $c \to 0$ for any fixed spectrum shape. The zero set of the self-scaled term includes both $B = 0$ and all nonzero scaled isometries. If $\xi > 0$ and $c_0 > 0$, adding $\xi[c_0 - \bar{\alpha}]_+^2$ gives loss $\xi c_0^2 > 0$ at $B = 0$, whereas $B^\top B = \beta I_n$ with $\beta \geq c_0$ has zero total loss at $\alpha = \bar{\alpha} = \beta$. Thus a scaled isometry attains a strictly lower loss than the collapsed matrix. In (11), the hinge applies this gain constraint to the batch average. □

A.6 Finite-difference control: statement and proof

Lemma 2 (Finite-difference control of the local distance loss)**.** *Fix a measurable predictor map $\hat{f} : \mathbb{R}^n \to \mathbb{R}^d$ that is differentiable at a_t with Jacobian B, and a target scale $c > 0$. For independent $\delta, \delta' \sim \mathcal{N}(0, \epsilon^2 I_n)$, let $w = \delta - \delta'$,*

$$Q_\epsilon(B, c) = \mathbb{E}(\|Bw\| - c\|w\|)^2, \qquad \varrho = \hat{f}(a_t + \delta) - \hat{f}(a_t + \delta') - Bw, \qquad D_\epsilon^2 = \mathbb{E}\|\varrho\|^2 < \infty.$$

Then, without a rank assumption,

$$\left|\mathcal{L}_{\text{geo}}^{\text{FD}} - Q_\epsilon(B,c)\right| \le 2D_\epsilon\sqrt{Q_\epsilon(B,c)} + D_\epsilon^2. \tag{10}$$

If $\hat{f}$ is C^2 with globally bounded bilinear operator norm $\sup_a \|D^2\hat{f}(a)\| \le L$, then $D_\epsilon \le L\sqrt{n(n+1)}\,\epsilon^2$, so $\mathcal{L}_{\text{geo}}^{\text{FD}} = Q_\epsilon(B,c) + O(\epsilon^3)$ as $\epsilon \to 0$ for fixed B, c, n, L. For an affine map the equality is exact at every perturbation scale. For $\epsilon > 0$, $Q_\epsilon(B,c) = 0$ if and only if $B^\top B = c^2 I_n$, and its minimizer over $c \ge 0$ is $c^\star = \mathbb{E}[\|Bw\|\|w\|]/\mathbb{E}\|w\|^2$.

The comparison bound (10) also holds for each realized batch if expectations are replaced by averages and both losses use the same batch-dependent scale and floored distance $\max\{\|w\|, 10^{-8}\}$. This statement concerns the comparison bound, not the zero set or the fixed-c asymptotic expansion, and covers the implemented median/EMA tracker. At the default perturbation scale $\epsilon = 0.1$, the nonlinear remainder determines how closely the finite-distance objective approximates the local Jacobian loss. The tracker supplies the target scale, while $c^\star$ characterizes the fixed-matrix least-squares optimum.

Proof. We separate the linearized distance error from the change introduced by the nonlinear remainder. Put $v = \|Bw\| - c\|w\|$ and $e = \|Bw + \varrho\| - \|Bw\|$. The reverse triangle inequality gives $|e| \le \|\varrho\|$ even when $Bw = 0$. Consequently,

$$|\mathbb{E}(v+e)^2 - \mathbb{E}v^2| \le 2\sqrt{\mathbb{E}v^2\,\mathbb{E}e^2} + \mathbb{E}e^2 \le 2\sqrt{Q_\epsilon(B,c)}D_\epsilon + D_\epsilon^2.$$

With the second-derivative bound, Taylor's theorem gives $\|\varrho\| \le (L/2)(\|\delta\|^2 + \|\delta'\|^2)$. The Gaussian moments satisfy $\mathbb{E}(\|\delta\|^2+\|\delta'\|^2)^2 = 4n(n+1)\epsilon^4$, which bounds D_ϵ as stated. Gaussian scaling gives $Q_\epsilon(B,c) = \epsilon^2 Q_1(B,c)$. With $C = L\sqrt{n(n+1)}$, the comparison bound becomes

$$|\mathcal{L}_{\text{geo}}^{\text{FD}} - Q_\epsilon(B,c)| \le 2C\sqrt{Q_1(B,c)}\,\epsilon^3 + C^2\epsilon^4.$$

This proves the expansion as $\epsilon \to 0$ for fixed B, c, n, L, and makes the dependence of the remainder constant explicit. The first inequality uses only Cauchy–Schwarz and the reverse triangle inequality, so it also holds for finite averages with any realized adaptive scales and the same floored action distances in the nonlinear and linearized losses. If $Q_1 = 0$, the bound improves to $O(\epsilon^4)$. For an affine map $\varrho = 0$.

Since w has full support and $w \mapsto \|Bw\| - c\|w\|$ is continuous, $Q_\epsilon(B,c) = 0$ if and only if $\|Bw\| = c\|w\|$ for every w. Squaring both sides gives $B^\top B = c^2 I_n$. The expression is quadratic in c with positive leading coefficient $\mathbb{E}\|w\|^2$, so differentiation gives $c^\star$. For $B = 0$, this optimum is 0; otherwise it is positive.

The cubic remainder can occur even for a smooth predictor. Take $\hat{f}(a) = a^2$, $a_t = 0$ and $c = 1$. Here $B = 0$ and direct Gaussian integration gives $\mathcal{L}_{\text{geo}}^{\text{FD}} = 2\epsilon^2 - 8\epsilon^3/\sqrt{\pi} + 4\epsilon^4$. The nonzero cubic term arises from the norm at zero, despite the symmetric perturbations. □

A.7 Transfer to the environment response

Remark 3 (Transfer requires derivative fidelity). The penalty acts on $\widehat{B}_t$, not B_t. At the same state, action and latent context, Weyl's inequality (Horn & Johnson, 2012) gives $|\sigma_i(\widehat{B}_t) - \sigma_i(B_t)| \le \|\widehat{B}_t - B_t\|_2 =: \epsilon_{\text{pred}}$. Consequently,

$$\kappa(B_t) \le \frac{\sigma_{\max}(\widehat{B}_t) + \epsilon_{\text{pred}}}{\sigma_{\min}(\widehat{B}_t) - \epsilon_{\text{pred}}} \quad \text{when} \quad \sigma_{\min}(\widehat{B}_t) > \epsilon_{\text{pred}}.$$

The rank limitation is substantive: if B_t is rank deficient, any unit $v \in \ker B_t$ yields $\|\widehat{B}_t - B_t\|_2 \ge \|\widehat{B}_t v\| \ge \sigma_{\min}(\widehat{B}_t)$. A positive full-rank target can therefore compete with derivative fidelity in null action directions. A ratio of separately sampled condition-number medians does not test the paired-error premise of this bound.

A.8 THE MATRIX-FREE ESTIMATOR

Lemma 3. *Fix $\widehat{B}$ and α, let $M = \widehat{B}^\top \widehat{B} - \alpha I_{n_a}$, and let v have zero mean and $\mathbb{E}[vv^\top] = I_{n_a}$. Then $\mathbb{E}_v\|Mv\|^2 = \|M\|_F^2$ and $\mathbb{E}_v\|\widehat{B}v\|^2 = \mathrm{tr}(\widehat{B}^\top\widehat{B})$, and $Mv = \widehat{B}^\top(\widehat{B}v) - \alpha v$ is computable with one forward-mode Jacobian-vector product and one reverse-mode vector-Jacobian product, without forming $\widehat{B}$.*

Proof. Isotropic probes recover the sum over coordinate directions through an expectation. $\mathbb{E}_v\|Mv\|^2 = \mathbb{E}_v\,\mathrm{tr}(v^\top M^\top M v) = \mathrm{tr}\left(M^\top M\,\mathbb{E}_v[vv^\top]\right) = \mathrm{tr}(M^\top M) = \|M\|_F^2$; this is Hutchinson's trace identity (Hutchinson, 1989) applied to $M^\top M$, and the same computation with $\widehat{B}$ in place of M gives the trace. The decomposition of Mv is immediate, and $\widehat{B}v$ is by definition the directional derivative of the action-to-latent closure at a_t along v, while $\widehat{B}^\top y$ is its pullback of y. □

Remark 4. Probes with independent standard Gaussian entries or independent Rademacher entries both satisfy the hypothesis. The identities condition on $\widehat{B}$ and a target α fixed independently of the current probes; estimating that target from the same probes need not preserve unbiasedness. For $H = M^\top M$, their single-probe variances are $2\|H\|_F^2$ and $2\sum_{i\neq j} H_{ij}^2$, respectively. Rademacher probes therefore have no greater variance, with strict inequality when $M \neq 0$, for these scalar estimators at a fixed matrix and target.

B IMPLEMENTATION DETAILS

B.1 GRAM-MATRIX REGULARIZATION

Gram-matrix variant (AJR-Gram). The second variant penalizes deviations of the action Gram matrix from a scaled identity,

$$\mathcal{L}_{\mathrm{geo}}^{\mathrm{Gram}} = \frac{1}{q}\sum_{j=1}^{q}\left\|\widehat{B}_j^\top\widehat{B}_j - \alpha I_{n_a}\right\|_F^2 + \xi\left[c_0 - \frac{1}{q}\sum_{j=1}^{q}\frac{\mathrm{tr}(\widehat{B}_j^\top\widehat{B}_j)}{n_a}\right]_+^2, \tag{11}$$

where j indexes the q samples in the geometric sub-batch and α tracks the batch mean of $\mathrm{tr}(\widehat{B}_j^\top\widehat{B}_j)/n_a$ through a detached EMA. The first term regularizes each sample's spectrum relative to a shared scale. In the second term, c_0 is the gain floor and ξ is its weight, and $[x]_+ = \max\{x, 0\}$. With $c_0 > 0$ and $\xi > 0$, the hinge penalizes collapse of the batch-average gain; it does not impose a gain floor on every sample. The default $c_0 = 0$ leaves the floor inactive, and a positive floor is an ablation setting.

The exact estimator forms each sample's $\widehat{B}_t$ with a vectorized reverse-mode Jacobian at a cost of d pullbacks. A matrix-free alternative combines forward-mode products $\widehat{B}_t v$ with reverse-mode products $\widehat{B}_t^\top(\cdot)$. For a fixed target α, it uses Hutchinson's identity $\|M\|_F^2 = \mathbb{E}_v\|Mv\|^2$ for $M = \widehat{B}_t^\top\widehat{B}_t - \alpha I$ (Hutchinson, 1989), where the probes v have zero mean and $\mathbb{E}[vv^\top] = I_{n_a}$. It requires K probe pairs without materializing $\widehat{B}_t$. We use AJR-FD as the default because it avoids higher-order differentiation through attention. AJR-Gram provides a direct spectral penalty for ablation, and its exact Jacobian computation also supports the measurement protocol in Section 4.6.

B.2 TRAINING AND CONTROLLER IMPLEMENTATION

Inverse-head initialization. The residual MLP uses AdaLN-Zero modulation of the remaining horizon. Its scale-and-shift projection is initialized to zero, so the initial head is horizon-agnostic; training learns how the remaining horizon should affect the action.

Gradient paths. The shaping terms reuse the prediction-loss forward pass. For ICT, the anchor latent comes from the final context frame, the goal latents come from a separate encoder pass, and the target action is detached. The detached-ICT variant detaches both latent inputs. AJR instead uses a detached latent context and updates only the predictor and action encoder. This division

assigns encoder-side shaping to inverse regression and predictor-side shaping to the action-response objective.

Goal-frame sampling for ICT. The prediction loss uses short windows of $m{+}1$ frames, so sampling $(t, t+h)$ only within each window would unnecessarily restrict the inverse-training horizon. We instead sample $P{=}4$ goals independently from the same episode, with horizons drawn uniformly from $\{1, \ldots, \min(H_{\max}, \text{frames left in the episode})\}$. One extra batched encoder pass produces all goal latents without changing the prediction-loss dataloader.

Jacobian computation. The exact surrogate estimator computes vectorized, per-sample reverse-mode Jacobians. For each batch element, the predictor is differentiated with respect to its current action at a fixed context. Recovering all output coordinates requires d pullbacks. Constructing the same matrix in forward mode would require n_a passes and forward-mode support in the attention kernel. Because AJR-Gram differentiates through this Jacobian during training, we use the math attention backend for compatibility with vectorized mapping and higher-order derivatives; the remaining forward pass is unchanged. The matrix-free variant avoids materializing $\widehat{B}_t$ when either the latent or action dimension is large.

Cost of the shaping term. ICT adds one batched encoder pass over P goal frames and one MLP evaluation. AJR uses a sub-batch of 16 windows: AJR-FD adds two predictor evaluations per window, whereas AJR-Gram uses either d reverse-mode pullbacks or K matrix-free probe pairs before differentiating the geometric loss. These operations involve the six-layer predictor over m tokens and do not require additional vision-encoder passes.

Scale trackers. The isometry scale c in AJR-FD and the isotropy target α in AJR-Gram use exponential moving averages initialized with their first observation. Both targets are detached from the gradient, so the loss updates the predictor relative to a slowly varying scale estimate.

Blocked-action controller. The head predicts blocked world-model actions, whereas the evaluation harness accepts one primitive action per step. At each buffer refill, the controller encodes the current observation and goal, computes the remaining horizon, evaluates the head, and splits its output into k primitive actions. It then executes one buffered action at each environment step and clears the buffer at episode boundaries or termination.

Collapse diagnostics. We monitor latent standard deviation at every step and periodically measure effective rank and surrogate Jacobian scale to detect representation collapse and diminishing action sensitivity during shaping.

B.3 Control interfaces

The blocked-action interface follows (2): the controller sets $h = \min(\text{blocked actions left}, H_{\max})$, predicts one blocked action and executes its k primitive actions before observing again. The official base-action interface used in the main Push-T and Reacher direct evaluations, and in the supplementary head study, instead extracts latents at frameskip 1 and trains the head to predict one primitive action. It conditions on the remaining budget and observes after each action. Appendix C specifies the evaluation interfaces. Both require one head evaluation per decision; the goal encoding can be cached within an episode.

B.4 Conditioning measurement protocol

We distinguish three Jacobians because they characterize different parts of the control pipeline. For the **surrogate Jacobian** $\widehat{B}_t$, we compute exact per-sample reverse-mode Jacobians over sampled context windows. For the **true Jacobian** B_t, we use central finite differences in a resettable simulator. Each perturbation starts from the same recorded state, applies a coordinate-wise perturbed blocked action, and renders and encodes the successor. This estimator measures environment transitions without a learned dynamics model. For the **encoder Jacobian** J_e, we use forward-mode

Jacobian-vector products along random unit pixel-space tangents, since the full pixel-space Jacobian is too large to form. The resulting sketch measures encoder sensitivity along the sampled tangents.

We report the stabilized condition ratio $\kappa_{\mathrm{num}}(M) = \sigma_{\max}(M)/\max\{\sigma_{\min}(M), 10^{-12}\}$, alongside the singular values, and summarize it by the median and interquartile range. The denominator floor yields a finite statistic, including zero for a zero matrix, so we interpret this ratio alongside the singular values to identify small or null directions. The conditioning gap is the natural log-ratio of true and surrogate medians over separately sampled anchors. It summarizes their spectral difference, whereas the transfer bound in Remark 3 uses paired Jacobian error. Rank correlations are descriptive summaries of the association between conditioning and control performance across the measured checkpoints.

C Experimental protocol

C.1 Main-table training and readout controls

Training and readout controls. All models use LeWM's ViT-Tiny encoder, 192-dimensional latents, and six-layer predictor (Dosovitskiy et al., 2020; Maes et al., 2026), together with its optimizer and learning-rate schedule (Loshchilov & Hutter, 2017). We augment this objective with ICT, AJR-FD, or both. For direct control, each encoder is frozen and paired with a fresh head trained for 50 epochs with 1,000,000 sampled pairs per epoch and 10,000 validation pairs. Architecture and training settings are matched within each domain. Fresh heads provide a common readout for the independently trained representations.

Controllers and comparison settings. We evaluate CEM, iCEM, MPPI, gradient planning, and GC-IDM. CEM uses 300 candidates, 30 elites and 30 iterations on Push-T or 10 elsewhere. Non-Reacher iCEM and MPPI use 300 candidates and 30 iterations, and gradient planning uses 30 updates. Reacher's release-aligned matrix uses 10 updates, with 300 candidates for sampling solvers and 100 for gradient planning. World-model actions concatenate five primitive actions; the main Push-T and Reacher GC-IDM heads instead predict one primitive action and observe at every step. Figure 2 and Table 1 compare LeWM (S1) with IC-LeWM (S2). S2 uses ICT on Push-T and Reacher and AJR-FD on Two-Room and Cube across all five controllers.

C.2 Detailed training and evaluation settings

World-model settings for the parameter study. The Push-T parameter study uses 10 training epochs, seed 42, context length 3, batch size 128, and AdamW with learning rate 5×10^{-5} and weight decay 10^{-3}. The shaping weight defaults to $\mu^\star = 0.1$ and warms up over the first 20% of training. ICT uses four sampled goals with horizon cap 50; active component weights are $w_{\mathrm{ICT}} = w_{\mathrm{AJR}} = 1$ unless explicitly varied. AJR uses a geometric sub-batch of 16, a perturbation standard deviation of 0.1, and EMA momentum 0.99. AJR-Gram uses the exact Jacobian, unit hinge weight, and zero gain floor except in the $c_0 = 0.5$ ablation. Table 5 lists the changes to these defaults.

Head training. Frozen-encoder embeddings are extracted once per checkpoint for the complete dataset. Pairs (z_t, z_{t+h}, a_t, h) are then sampled online within episodes with $h \sim \mathrm{Unif}\{1, \ldots, H_{\max}\}$, which decouples the reachable horizon from the world-model context window. The matched-head study trains each head for 50 epochs with AdamW, a learning rate of 10^{-3}, weight decay of 10^{-4}, batch size 1024, gradient clipping at norm 1, and a cosine schedule. Each epoch uses 1,000,000 training pairs, and validation uses 10,000 validation pairs. The capacity control doubles the hidden width, whereas the horizon controls vary $H_{\max}$ independently.

Evaluation and pairing. The CEM comparison in Table 1 uses 30 iterations on Push-T and 10 on Two-Room, Cube, and Reacher, with 300 candidates, 30 elites, horizon 5, action block 5 and receding execution. Goals lie 25 primitive steps ahead in the recorded trajectory, with an interaction budget of 50 primitive steps. The main Reacher matrix uses the release evaluator, qpos threshold 0.1 radians and 10 updates for all four search solvers. In the separate local comparison, ICT reaches

92% CEM success versus 82% for the unshaped control. The 86% CEM baseline in Table 1 is the published LeWM result, rather than this paired control.

The main Push-T and Reacher GC-IDM evaluations use the official base-action interface with frameskip 1, horizon cap 50, and remaining-budget conditioning. The additional paired comparison of unshaped and combined models shares episode/start indices within each domain and uses 30 iterations for sampling planners and 30 gradient updates. Because this planning budget differs from the main Reacher protocol, the two sets of results are analyzed separately. The alternative blocked controller in Section 4 executes five primitive actions before re-encoding.

Supplementary ablation protocol. The supplementary study trains 30 heads with seed 42 and hidden width 512. Each head is trained for 50 epochs with 1,000,000 training pairs per epoch and 10,000 validation pairs. Each world-model encoder is frozen and paired with a fresh head to standardize the downstream readout. These evaluations use the official base-action GC-IDM interface: latent extraction uses frameskip 1, the head predicts one primitive action, and the controller conditions on the remaining interaction budget. Their action interface therefore differs from the blocked-action controller described in Section 4. The default head horizon cap is 50; the stress-test control uses 100.

Push-T, Two-Room, and Cube use the evaluation protocol of Section 5 with seed and split seed 42, goal offset 25, budget 50 and `episode_split=all`. The fourteen Push-T parameter configurations share identical episode/start pairs. Their Δ values are relative to the parameter-study baseline of 74%. The component and weight studies use separate checkpoints and are compared within their respective groups. CEM and GC-IDM share the recorded model paths and episode/start lists for the unshaped, ICT, AJR-FD, and $\mu^\star = 0.1$ combined configurations. Pairing is not asserted for the remaining rows of Table 13.

The supplementary Reacher ablations use the isolated release evaluator for both controllers, with qpos threshold 0.1, seed 42, offset 25, budget 50 and the same 50 starts. CEM uses 300 samples, 30 elites, variance 1, and 10 iterations; its planning horizon, receding horizon, and action block are all set to 5. GC-IDM uses the release FeedForwardPolicy, preprocessing and success evaluator, with no iterative solver.

D SUPPLEMENTARY RESULTS AND ABLATIONS

D.1 FULL CONTROL RESULTS

Table 2 reports the three-training-seed main results for ICT on Push-T/Reacher and AJR-FD on Two-Room/Cube, following the LeWorldModel evaluation setting. CEM S1 values are published LeWorldModel references.

Table 2: **Control success rates and uncertainty (%).** S1/S2 configurations follow Table 1. Following the LeWM Maes et al. (2026) configuration, results are mean ± SD over three independent training seeds. The last column identifies the IC-LeWM configuration used for S2. Bold compares means within each dataset–controller pair.

Domain	Controller	S1	S2	S2 configuration
Push-T	CEM	96.00 ± 2.83	$\mathbf{96.67 \pm 1.15}$	ICT
	iCEM	90.00 ± 2.00	$\mathbf{90.67 \pm 1.15}$	ICT
	MPPI	58.00 ± 2.00	$\mathbf{78.00 \pm 8.72}$	ICT
	Gradient	6.00 ± 2.00	$\mathbf{10.00 \pm 2.00}$	ICT
	GC-IDM	82.67 ± 1.15	$\mathbf{87.33 \pm 1.15}$	ICT
Two-Room	CEM	87.00 ± 2.49	$\mathbf{96.00 \pm 0.00}$	AJR-FD
	iCEM	88.00 ± 2.00	$\mathbf{96.00 \pm 0.00}$	AJR-FD
	MPPI	80.67 ± 2.31	$\mathbf{89.33 \pm 2.31}$	AJR-FD
	Gradient	40.00 ± 3.46	$\mathbf{61.33 \pm 3.06}$	AJR-FD
	GC-IDM	$\mathbf{100.00 \pm 0.00}$	$\mathbf{100.00 \pm 0.00}$	AJR-FD
Cube	CEM	74.00 ± 3.06	$\mathbf{80.00 \pm 2.00}$	AJR-FD
	iCEM	74.00 ± 2.00	$\mathbf{77.33 \pm 3.06}$	AJR-FD
	MPPI	52.00 ± 2.00	$\mathbf{58.00 \pm 0.00}$	AJR-FD
	Gradient	47.33 ± 2.31	$\mathbf{57.33 \pm 1.15}$	AJR-FD
	GC-IDM	97.33 ± 2.31	$\mathbf{98.67 \pm 1.15}$	AJR-FD
Reacher	CEM	86.00 ± 4.99	$\mathbf{89.33 \pm 3.06}$	ICT
	iCEM	$\mathbf{88.67 \pm 4.62}$	85.33 ± 1.15	ICT
	MPPI	$\mathbf{67.33 \pm 1.15}$	65.33 ± 2.31	ICT
	Gradient	$\mathbf{83.33 \pm 4.62}$	74.67 ± 4.16	ICT
	GC-IDM	42.00 ± 6.00	$\mathbf{51.33 \pm 4.16}$	ICT

D.2 Interaction between ICT and AJR-FD on Cube

Table 3 compares the combined objective with AJR-FD alone under the same Cube planning settings as the main evaluation. Each combined result summarizes three training seeds. CEM retains the AJR-FD mean of 80.00%, whereas iCEM, MPPI, and gradient planning show absolute decreases of 2.00%, 0.67%, and 4.00%, respectively. The combination therefore provides no mean improvement over AJR-FD alone in these four comparisons; the small sampling-planner differences do not by themselves establish a statistically reliable disadvantage.

Table 3: **Combining ICT and AJR-FD on Cube.** Success rates (%) under the main planning protocol, reported as mean ± SD over three training seeds. AJR-FD values are from Table 2; comparisons concern aggregate means.

Controller	AJR-FD Mean ± SD	ICT+AJR-FD Mean ± SD
CEM	80.00 ± 2.00	80.00 ± 2.00
iCEM	77.33 ± 3.06	75.33 ± 1.15
MPPI	58.00 ± 0.00	57.33 ± 4.16
Gradient	57.33 ± 1.15	53.33 ± 3.06

This result contrasts with the Push-T GC-IDM ablation, where the combination attains 88% versus 84% for ICT and 82% for AJR-FD. The objectives address distinct properties: ICT trains representations for action recovery between observed endpoints, whereas AJR constrains the predictor's local action response. Search uses the predicted trajectory, so better inverse readout does not necessarily improve its cost landscape. The Cube results motivate evaluating their combination for the intended task and controller because the benefits of the two components need not add. The shape–gain study in Appendix E.3 provides a separate example of tension with prediction fidelity; it does not diagnose the Cube models.

D.3 Supporting fixed-controller analyses

Marginal regularization remains useful in this configuration. Removing SIGReg reduces the combined model from 88% to 70%. Although the no-SIGReg model recovers seven baseline failures, it loses nine baseline successes. The resulting decline reflects regressions on previously solved starts rather than a complete absence of newly solved cases. Within this setting, marginal regularization therefore remains complementary to the two action-dependent signals.

Temporal coverage is more informative than the real/predicted-goal label. Real future goals yield 84% success, compared with 74% for predictor-generated goals (Table 5). Because the predictor-goal variant also narrows the temporal span of supervision, it does not isolate goal accuracy. The horizon controls provide a more focused comparison: ICT training caps of 5 and 10 yield 74%, whereas a cap of 25 reaches 86%; all evaluation heads retain cap 50. Longer-range trajectory pairs accompany the gain within these runs, but the default training cap of 50 attains 84%, so the evidence does not imply monotonic improvement with horizon.

Better action fit is not a sufficient selection criterion. Combined shaping weights $\mu^\star \in \{0.05, 0.1, 0.2\}$ yield 82%, 88%, and 82% success, respectively (Table 5). All three exceed the unshaped reference, yet the largest weight produces the highest action fit ($R^2 = 0.834$) rather than the highest success rate. Validation action fit and closed-loop control therefore provide distinct selection criteria.

D.4 Predictive utility across comparison groups

Search-based control supplies a separate test of predictive utility. The release-aligned Reacher AJR-Gram run attains 42% CEM success and 28% GC-IDM success despite a lower surrogate ratio than several other Reacher ablations (Table 10). The additional paired Push-T comparison also reports that combining ICT and AJR-FD raises GC-IDM success from 76% to 84% while reducing

Table 4: **Published IMWM reference (success rate, %).** Values reproduce Table 2 of Gao et al. (2026): mean ± SD across 12 data/solver-seed evaluation cells per task, with frozen world models and 50 episodes per cell. All tasks use 300×30 CEM queries per replan. These are evaluation-cell SDs, not world-model training-seed SDs; local non-Push-T CEM uses 300×10. Bold compares only the two methods within this published protocol.

Domain	World-model only	IMWM
Two-Room	87.7 ± 4.7	$\mathbf{99.2 \pm 1.0}$
Reacher	83.2 ± 4.1	$\mathbf{83.8 \pm 5.7}$
Push-T	89.8 ± 4.8	$\mathbf{92.7 \pm 3.4}$
OGBench-Cube	66.2 ± 5.7	$\mathbf{94.7 \pm 1.9}$

Table 5: **Push-T ablations under one fixed GC-IDM protocol.** Success rates (%) use the same 50 starts; R^2 measures validation action fit. Each row uses one world-model/head seed. The detached-ICT row is a training-path diagnostic (Appendix E.1). The combined reference uses $\mu^\star = 0.1$, and horizon caps in the last block concern ICT training only.

Configuration / controlled change	**Success (%)**	R^2
Components		
No shaping	74	0.819
ICT	84	0.826
AJR-FD	82	0.821
ICT+AJR-FD	88	0.827
Gradient path, goal source and geometric objective		
ICT, detached encoder inputs	84	0.818
ICT, predictor-generated goal	74	0.820
ICT+AJR-FD, no SIGReg	70	0.790
AJR-Gram	80	0.816
AJR-Gram, gain floor $c_0 = 0.5$	80	0.816
Hyperparameter sensitivity		
ICT+AJR-FD, $\mu^\star = 0.05$	82	0.825
ICT+AJR-FD, $\mu^\star = 0.2$	82	0.834
ICT, training horizon cap 5	74	0.820
ICT, training horizon cap 10	74	0.819
ICT, training horizon cap 25	86	0.826

CEM success from 92% to 60% (Appendix D.5). These groups are not pooled with the main parameter study. Together they show that local spectral diagnostics and performance under the intended planner assess different aspects of predictive utility. The shape–gain and paired-response studies in Appendices E.3–E.4 examine how spectral targets interact with prediction and derivative fidelity.

D.5 Supplementary control comparisons

Paired evaluation of the combined objective. An additional study compares the unmodified objective with ICT+AJR-FD rather than with the controller-specific configurations selected for the main table. On Push-T, inverse-head success rises from 76% to 84%, whereas CEM success falls from 92% to 60%. On Reacher, the combination reduces inverse-head success from 50% to 24% under the supplementary protocol, which differs from the main-table ICT comparison. The combination therefore has different effects across deployment settings.

Offline action fit is not closed-loop success. The heads trained on the combined encoders yield higher held-out action R^2 in all four domains of the additional paired study (Table 8). On Reacher, however, closed-loop success declines even as R^2 increases from 0.0648 to 0.0753. Higher offline action fit can therefore coexist with worse goal reaching.

Table 6: Supplementary component comparisons under the CEM budgets of Table 1. Success rates (%) use 50 episodes. S2 denotes the configuration selected within this study. Dashes denote unreported entries.

Domain	No shaping	ICT	AJR-FD	S2
Push-T	94	96	70	94
Two-Room	90	92	96	96
Cube	76	76	78	78
Reacher	84	92	90	90

Table 7: Additional component comparisons. The scores of Push-T and ICT appear in Table 6. The Two-Room and Cube ICT+AJR-FD scores use separate training runs. The epoch-10 Reacher ICT result is 92%.

	Push-T	Two-Room	Cube	Reacher
Success (%)	96	94	80	90
Component	ICT	ICT+AJR-FD	ICT+AJR-FD	AJR-FD

Component comparisons. Table 6 compares ICT and AJR-FD at epoch 10. Table 7 gives additional configurations evaluated with different checkpoints, as specified in its caption.

D.6 Response geometry

Figure 4 compares predictor and environment response geometry with GC-IDM success for the same models. Table 11 gives the medians, interquartile ranges and correlations, together with the encoder-sketch measurements.

Conditioning measurement. Surrogate Jacobians are measured on context windows drawn from the training distribution. Measuring the environment response requires a resettable simulator. Before each of the $2n_a$ perturbed rollouts, the environment is reset to a recorded state; the resulting frame is re-rendered and re-encoded, and the Jacobian columns are assembled by central differences with a fixed step size. Random unit pixel-space tangents provide a sketch of the encoder spectrum. We report medians and interquartile ranges over anchors. The five-checkpoint correlations pair each success rate with the median measured on the same checkpoint. The separate parameter-study batch is reported in Table 14.

D.7 Deployment stress tests and complete ablations

Increasing the evaluation-head horizon does not consistently improve long-budget control. For ICT, the default $H_{\max} = 50$ head attains 80% and 76% at budgets 75 and 100, compared with 72% and 66% for a separately trained $H_{\max} = 100$ head. The offset sweep varies the goal offset and interaction budget together, so it measures their joint effect rather than goal difficulty at a fixed budget. These deployment tests are distinct from the ICT training-horizon ablations.

Statistics and scope. The fixed-controller ablations use one model per configuration and 50 paired starts. Configuration choices in Table 6 use these evaluation outcomes. The held-out study instead fixes the final training epoch before testing (Appendix E.2).

Table 8: Held-out action-regression R^2 of the matched GC-IDM heads. These scores measure offline action fit. Cube† uses a combined model trained with a different schedule from its baseline in the supplementary paired study.

Domain	No shaping	ICT+AJR-FD
Push-T	0.8190	0.8297
Two-Room	0.1101	0.1296
Cube†	0.6146	0.6787
Reacher	0.0648	0.0753

Table 9: GC-IDM component and regularization-weight comparisons. Success rates (%) use 50 starts per row, and all heads share the supplementary training settings. These studies use different checkpoints from the Push-T parameter study. The final two columns report median numerical condition ratios; the encoder column is a tangent sketch. No simulator Jacobians were measured for Two-Room or Cube.

Domain	Configuration	GC-IDM	R^2	$\kappa_{\mathrm{num}}(\widehat{B})$	$\kappa_{\mathrm{num}}(J_e)$
Push-T	AJR-FD (component)	88	0.818	393.14	116.14
Two-Room	No shaping (component)	100	0.111	3133.45	93.19
Two-Room	ICT (component)	98	0.113	2763.52	188.42
Two-Room	AJR-FD (component)	100	0.133	2126.42	152.11
Cube	ICT (component)	94	0.655	2.02×10^8	130.14
Cube	AJR-FD (component)	98	0.641	1.40×10^8	176.74
Push-T	AJR-FD, $w_{\mathrm{AJR}}=0.25$	78	0.819	2205.33	133.22
Push-T	AJR-FD, $w_{\mathrm{AJR}}=0.5$	88	0.822	31269.32	127.75
Two-Room	AJR-FD, $w_{\mathrm{AJR}}=0.25$	100	0.126	2102.63	129.80
Two-Room	AJR-FD, $w_{\mathrm{AJR}}=0.5$	100	0.129	3184.93	124.45
Two-Room	AJR-Gram (screen)	100	0.120	1.06	123.57

Table 10: Reacher ablations under the isolated release evaluator. Both success columns use the same 50 episode/start pairs and qpos threshold 0.1. CEM uses 10 iterations; GC-IDM uses the base-action feed-forward controller. Geometry columns report surrogate and encoder-sketch medians; true simulator Jacobians are unavailable.

Configuration	CEM-10	GC-IDM	R^2	$\kappa_{\mathrm{num}}(\widehat{B})$	$\kappa_{\mathrm{num}}(J_e)$
ICT, detached encoder inputs	90	28	0.065	2596.17	403.88
AJR-FD, $w_{\mathrm{AJR}}=0.25$	90	22	0.075	1909.93	271.81
AJR-FD, $w_{\mathrm{AJR}}=0.5$	80	24	0.072	3813.39	242.92
AJR-Gram	42	28	0.065	695.78	356.44

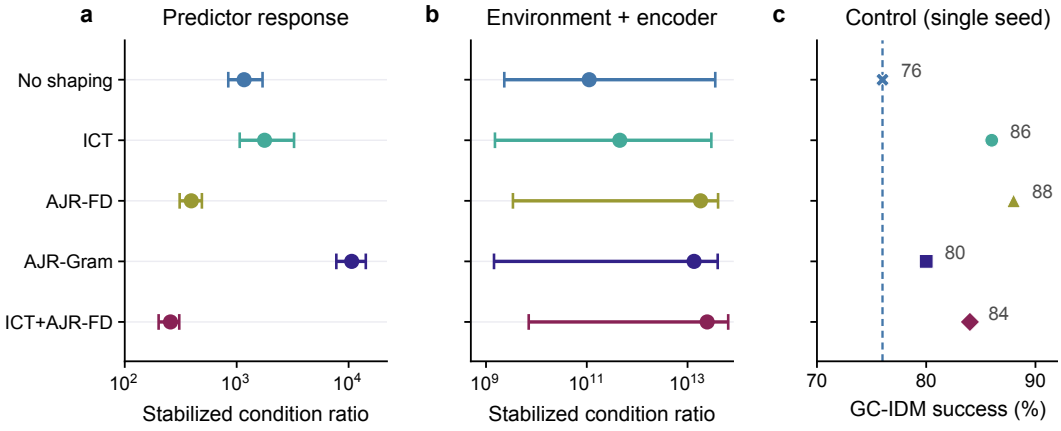

Figure 4: **Response geometry does not determine the control ranking. (a)** Median surrogate condition ratios with interquartile ranges over 256 predictor contexts. **(b)** Median simulator-mediated condition ratios with interquartile ranges over 64 reset anchors. **(c)** GC-IDM success from the same world-model checkpoints, each evaluated with a fresh head. Rows and colors identify the same five configurations across panels, and the dashed line in (c) marks the unshaped score. Intervals in (a,b) describe variation across measurement contexts or anchors, not variation across training seeds.

Table 11: Push-T response geometry and GC-IDM success for five checkpoints from the paired, component and parameter studies. Geometry entries are medians with anchor interquartile ranges; the stabilized numerical ratio is defined in Section 4.6. The surrogate, simulator and encoder measurements use 256 contexts, 64 reset anchors and 16 frames with 64 tangents, respectively. Gap is the natural log-ratio of environment and surrogate medians. Geometry and control are evaluated on the same checkpoints. Correlations are descriptive across the five checkpoints; the two-sided p-values use an asymptotic approximation.

Shaping	$\kappa_{\mathrm{num}}(\widehat{B}_t)$ **(surrogate)**	$\kappa_{\mathrm{num}}(B_t)$ **(true)**	**Gap**	$\kappa_{\mathrm{num}}(J_e)$	**GC-IDM SR (%)**
No shaping	1165 [841, 1703]	1.12×10^{11} $[2.30\times10^{9}, 3.55\times10^{13}]$	18.38	142 [130, 162]	76
ICT	1775 [1064, 3255]	4.52×10^{11} $[1.50\times10^{9}, 2.97\times10^{13}]$	19.36	106 [97, 117]	86
AJR-FD	393 [309, 489]	1.81×10^{13} $[3.42\times10^{9}, 4.04\times10^{13}]$	24.55	116 [103, 132]	88
AJR-Gram	10655 [7767, 14241]	1.35×10^{13} $[1.44\times10^{9}, 3.97\times10^{13}]$	20.96	153 [130, 175]	80
ICT + AJR	256 [201, 307]	2.45×10^{13} $[7.04\times10^{9}, 6.41\times10^{13}]$	25.28	91 [80, 102]	84
Spearman ρ: surrogate κ vs. success		$\rho = -0.30$, $p = 0.624$		$n = 5$ pairs	
Spearman ρ: true κ vs. success		$\rho = 0.50$, $p = 0.391$		$n = 5$ pairs	

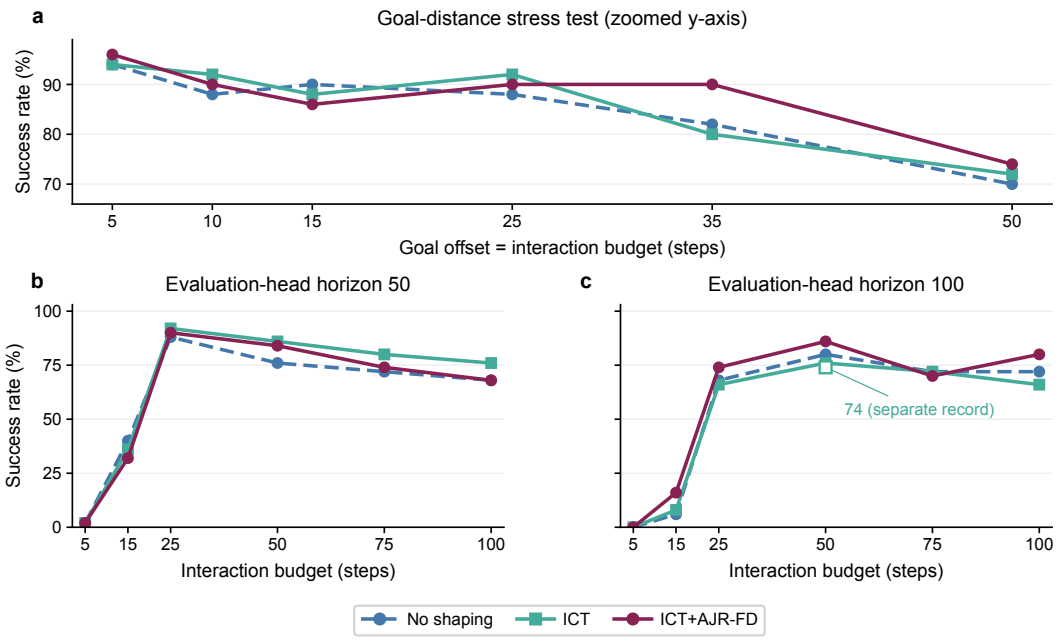

Figure 5: **Deployment performance depends on goal distance and head horizon.** Push-T GC-IDM success under **(a)** a joint sweep of goal offset and interaction budget with head cap 50, and **(b,c)** interaction-budget sweeps at fixed offset 25 with head caps 50 and 100. Panel (a) uses a zoomed vertical axis, as indicated in its title. Each point reports one 50-start evaluation for a single checkpoint/head seed; lines connect only evaluated settings. The unshaped and combined models are from the additional paired study, whereas ICT uses the selected epoch-10 backbone. The open square in (c) marks a 74% ICT evaluation at the same setting and starts as the 76% sweep result (Table 12).

Table 12: Push-T stress sweeps (success rate, %). ICT uses the selected epoch-10 backbone trained with seed 3072. The unshaped and ICT+AJR-FD models are from the additional paired study; each cell is a separate 50-start evaluation. Panel (a) sets budget equal to goal offset. Panel (b) fixes offset 25 and varies interaction budget and evaluation-head horizon cap. The $H_{\max} = 100$ ICT head is trained separately. These head-horizon controls differ from the ICT world-model training-horizon controls in Table 13. Both values are shown because the one-episode discrepancy remains unresolved.

(a) goal offset (primitive steps)

Backbone	5	10	15	25	35	50
LeWM (no shaping)	94	88	90	88	82	70
ICT+AJR-FD	96	90	86	90	90	74
ICT	94	92	88	92	80	72

(b) interaction budget (primitive steps)

Backbone	$H_{\max}$	5	15	25	50	75	100
LeWM (no shaping)	50	2	40	88	76	72	68
LeWM (no shaping)	100	0	6	68	80	72	72
ICT+AJR-FD	50	2	32	90	84	74	68
ICT+AJR-FD	100	0	16	74	86	70	80
ICT	50	2	36	92	86	80	76
ICT	100	0	8	66	76	72	66

Table 13: Push-T parameter ablations with one world-model/head seed per configuration. GC-IDM and CEM-30 report success (%); Δ subtracts the 74% parameter baseline from the displayed GC-IDM success rate. R^2 measures offline validation action fit. CEM and GC-IDM are task-paired for the unshaped, ICT, AJR-FD, and $\mu^\star = 0.1$ combined rows; other rows are configuration-level comparisons. ICT alone corresponds to the encoder-gradient row. The ICT training-horizon controls change world-model training, while all standard evaluation heads use $H_{\max} = 50$. † The wider-head control is descriptive because task-level pairing with the standard-head evaluation is unverified.

Configuration	$\kappa_{\mathrm{num}}(\widehat{B})$	**GC-IDM**	Δ	R^2	**CEM**
No shaping	509.98	74	0	0.819	98
ICT+AJR-FD, $\mu^\star = 0.05$	1168.85	82	+8	0.825	74
ICT+AJR-FD, $\mu^\star = 0.1$	3585.54	88	+14	0.827	68
ICT+AJR-FD, $\mu^\star = 0.2$	523.81	82	+8	0.834	70
ICT, encoder gradients	3105.00	84	+10	0.826	96
ICT, detached encoder inputs	5193.50	84	+10	0.818	86
AJR-FD	1221.39	82	+8	0.821	70
AJR-Gram	10654.98	80	+6	0.816	12
AJR-Gram, $c_0 = 0.5$	1.02	80	+6	0.816	16
ICT, predictor goal	454.87	74	0	0.820	96
ICT+AJR-FD, no SIGReg	211.46	70	-4	0.790	14
ICT, training horizon cap 5	483.90	74	0	0.820	96
ICT, training horizon cap 10	732.91	74	0	0.819	96
ICT, training horizon cap 25	1086.64	86	+12	0.826	92
No shaping, 2× head width†	–	84	–	0.824	98

Table 14: Response geometry of the Push-T ablation models. The columns report median stabilized condition ratios over 256 predictor contexts, 64 simulator-reset anchors, and 16 encoder frames, respectively. The simulator ratios include near-null directions under the stabilization convention in Section 4.6. The measurement contexts and anchors are sampled separately from those in Table 11.

Configuration	$\kappa_{\mathrm{num}}(\widehat{B})$	$\kappa_{\mathrm{num}}(B)$	$\kappa_{\mathrm{num}}(J_e)$
No shaping	509.98	1.45×10^{13}	126.42
ICT+AJR-FD, $\mu^\star = 0.05$	1168.85	2.44×10^{13}	122.55
ICT+AJR-FD, $\mu^\star = 0.1$	3585.54	2.27×10^{13}	104.25
ICT+AJR-FD, $\mu^\star = 0.2$	523.81	1.70×10^{13}	75.77
ICT, encoder gradients	3105.00	2.12×10^{13}	111.25
ICT, detached encoder inputs	5193.50	1.39×10^{13}	122.34
AJR-FD	1221.39	2.20×10^{13}	124.20
AJR-Gram	10654.98	1.48×10^{13}	153.11
AJR-Gram, $c_0 = 0.5$	1.02	1.63×10^{13}	142.37
ICT, predictor goal	454.87	1.73×10^{13}	147.76
ICT+AJR-FD, no SIGReg	211.46	5.23×10^{12}	233.45
ICT, training horizon cap 5	483.90	1.02×10^{13}	121.21
ICT, training horizon cap 10	732.91	1.91×10^{13}	120.64
ICT, training horizon cap 25	1086.64	2.01×10^{13}	119.58
AJR-FD (component)	393.14	2.29×10^{13}	116.14
AJR-FD, $w_{\mathrm{AJR}} = 0.25$	2205.33	1.94×10^{12}	133.22
AJR-FD, $w_{\mathrm{AJR}} = 0.5$	31269.32	1.65×10^{13}	127.75

E ADDITIONAL CONTROLLED STUDIES

We examine the ICT gradient path, generalization to held-out episodes, and the relationship between response geometry and predictive utility. Each study uses its own matched reference model.

E.1 ICT GRADIENT-PATH CONTROL

A detached ICT loss has zero world-model gradient, but auxiliary computation can still change random-number consumption, gradient clipping, or running statistics. We compare unshaped and detached-ICT updates with identical model and optimizer states, prediction batches, and a common initial AdamW step. The test uses the Push-T architecture in deterministic fp32 with batch size 4, isolates auxiliary random draws, separates model/head gradient clipping, and prevents auxiliary passes from updating BatchNorm running statistics.

Across three subsequent updates, detached ICT exactly matches the unshaped model and optimizer states, whereas ordinary ICT produces nonzero auxiliary encoder gradients (Table 15). This verifies the intended gradient paths over the audited updates. The 74% versus 84% detached comparison remains a training-path diagnostic: its initial random-number and optimizer states are unavailable for a matched full-training comparison.

Table 15: **Deterministic update-path audit.** Columns 2–3 compare the unshaped and detached-ICT branches after each update. Gradient columns give auxiliary encoder gradient L_1 norms, not downstream performance. All three checks use fp32 and batch 4.

Step	Model-state max. difference	Optimizer max. difference	Detached	ICT
1	0	0	0	2.2225×10^4
2	0	0	0	2.4226×10^4
3	0	0	0	2.8629×10^4

E.2 HELD-OUT EPISODES AND RELATED INVERSE OBJECTIVES

Protocol. We fix an episode-level $80/10/10$ train/validation/test split before training: $14{,}948/1{,}868/1{,}869$ episodes on Push-T and $8{,}000/1{,}000/1{,}000$ on Reacher. World-model training, action normalization, and head training use only training episodes. Models share seed 42, ten epochs, batch size 128, history length 3, five primitive actions per transition, and AdamW with learning rate 5×10^{-5} and weight decay 10^{-3}. We use bf16 and the auxiliary-state handling above. The final epoch is fixed before testing.

Inverse objectives. The unshaped reference uses prediction loss and SIGReg. ICT retains SIGReg and uses a width-512 auxiliary head conditioned on horizons up to 50 primitive steps, with four goals per context. One-step IDM+SIGReg uses a width-256 two-layer ReLU head to predict the action block between adjacent frames, inverse weight 0.1, 20% warmup, and SIGReg weight 0.09. The adapted SMWM objective (Ivashkov et al., 2026) uses a width-256 inverse head without SIGReg and a constant inverse weight of 30 on Push-T or 5 on Reacher. This shared-backbone comparison evaluates complete training objectives; auxiliary capacity, temporal supervision, and regularization are not individually controlled. It is not a full-system SMWM reproduction.

Readout and planning. All frozen encoders receive the same width-512 GC-IDM readout, trained for 50 epochs with seed 42, 10^6 pairs per epoch, 10^4 validation pairs, and maximum horizon 50 primitive steps. Evaluation uses the same 50 test tasks within each domain, goal offset 25, and interaction budget 50. GC-IDM predicts one primitive action per decision. Search uses 300 candidates and 30 elites, with CEM-30 on Push-T and release-aligned iCEM-10 on Reacher (qpos threshold 0.1 radians).

Results. ICT raises held-out Push-T GC-IDM success from 70% to 78%. One-step IDM+SIGReg and adapted SMWM attain 76% and 78%, respectively. Under CEM, ICT scores 90% versus

Table 16: **Control on held-out episodes.** Each cell uses training seed 42, the final epoch-10 world model, and 50 fixed test tasks from held-out episodes. Search denotes CEM-30 on Push-T and iCEM-10 on Reacher. Success rates are in %. Unlike Table 1, this study adopts an episode-level held-out evaluation protocol and retains SMWM's inverse-head architecture and loss weights within our shared training setup.

	Push-T		Reacher	
Training objective	GC-IDM	Search	GC-IDM	Search
No shaping	70	94	32	94
ICT	78	90	44	92
One-step IDM+SIGReg	76	96	26	94
SMWM objective	78	72	92	98

SMWM's 72%, while the baseline and one-step IDM reach 94% and 96%. Thus ICT improves direct control on this split, while the unshaped and one-step IDM objectives retain higher CEM success. ICT also supports stronger planning than the adapted SMWM objective on Push-T. On Reacher, SMWM reaches 92% GC-IDM and 98% iCEM success. The comparison evaluates complete training objectives with matched readout heads and one training run per objective.

E.3 Separating spectral shape from response gain

Controlled objectives. The self-scaled Gram loss couples spectral spread to response magnitude. To examine these contributions separately, let $J = \widehat{B}_t$, $G = J^\top J$, $n = n_a$, and $g = \mathrm{tr}(G)/n$. We compare the normalized penalties

$$\ell_{\mathrm{gain}} = (g/g_{\mathrm{ref}} - 1)^2, \qquad \ell_{\mathrm{shape}} = \frac{1}{n}\left\| \frac{G}{\max\{g, 10^{-6} g_{\mathrm{ref}}\}} - I_n \right\|_F^2 . \tag{12}$$

The three shaped configurations retain ICT and substitute gain, shape, or both terms for the default AJR-FD term. Exact action Jacobians are used in these controls. The fixed reference $g_{\mathrm{ref}} = 1.314854$ is the median gain on 256 windows from the training subset of an unshaped calibration model. On eight batches, the median ratios of the default FD gradient norm to each penalty's gradient norm set coefficients 0.070624 for gain and 0.144615 for shape. This calibration uses only initial gradient norms to set comparable starting scales for the three penalties.

All four models, including a separately trained unshaped reference, use seed 42, ten epochs, context length 3, batch size 128, and the parameter-study optimizer. The shaped models use $\mu^\star = 0.1$, a 20% warmup, and geometric sub-batches of 16. Initialization is explicitly seeded and auxiliary random-number draws are isolated. This group precedes the complete update-path corrections in Appendix E.1; it is not a full-training placebo test. The group uses a random-window $90/10$ training/validation split and in-dataset control evaluation. Its epoch-10 baseline is therefore analyzed separately from both the historical 74% parameter baseline and the held-out 70% baseline.

Prediction and control outcomes. Table 17 reports final-epoch validation prediction MSE, averaged over latent coordinates and validation samples, together with latent standard deviation, predictor geometry, and closed-loop success on the same 50 tasks. All four encoders receive fresh matched GC-IDM heads. Shape regularization yields near-isometric predictor responses but much larger validation prediction error and severe CEM degradation. Adding gain control keeps the response magnitude near its reference without resolving either deterioration. Latent standard deviations remain near one, so the prediction-error difference is not explained by a gross change in overall latent scale. Nevertheless, each model defines its own latent prediction target; these losses are not errors in a common physical coordinate system.

The gain-only and joint-penalty rows differ by the shape term and exhibit a pronounced prediction/control tradeoff. This observation is consistent with competition between spectral equalization and predictive fidelity in these exact-Jacobian controls. GC-IDM does not query the predictor online and is less severely affected than CEM in the joint-penalty row, but none of the shaped models exceeds this group's unshaped GC-IDM reference. These negative results limit the use of near-isometry as a model-selection criterion.

Table 17: **Spectral equalization and predictive utility.** Single-seed, epoch-10 shape–gain study. MSE and latent SD are final-epoch validation averages. $\kappa(J)$ and g are medians over 64 predictor contexts. Control columns give success (%) on 50 matched starts. Every shaped row retains ICT; these are separate objectives from the historical AJR-Gram and AJR-FD rows.

Configuration	Pred. MSE	Latent SD	$\kappa(J)$	g	GC-IDM	CEM
No shaping	0.00333	0.9960	1295.18	1.139	90	92
ICT + gain	0.00569	1.0078	2350.79	1.296	84	96
ICT + shape	0.24232	1.0036	1.059	3.393	78	24
ICT + shape + gain	0.23240	1.0039	1.063	1.313	88	12

E.4 Paired finite-response measurements

Anchors, coordinates and numerical rank. For each model in Appendix E.3, we use the same 64 recorded episode/window anchors and nominal actions. The predictor Jacobian uses the recorded three-frame context. The simulator is reset to the corresponding recorded state, executes the five-action block, and returns a rendered successor to the same encoder. Both responses are differentiated with respect to z-score-normalized primitive-action coordinates, so the action block has dimension 10. The predictor derivative is computed by automatic differentiation; the simulator estimate B_δ uses central differences with coordinate step $\delta \in \{0.003, 0.01, 0.03\}$.

For singular values $\sigma_1 \geq \cdots \geq \sigma_{10}$, we record the full spectrum and define numerical rank and a truncated condition ratio by

$$r_\eta(M) = \#\{i : \sigma_i(M) > \eta\sigma_1(M)\}, \qquad \kappa_\eta(M) = \sigma_1(M)/\sigma_{r_\eta(M)}(M).$$

We use $\eta \in \{10^{-2}, 10^{-3}, 10^{-4}\}$. For a zero response, $r_\eta = 0$ and κ_η is undefined; such anchors are included in rank and gain summaries and excluded only from condition-ratio medians. We also measure $g(M) = \|M\|_F^2/10$, the paired spectral error $\|J - B_\delta\|_2$, and the relative Frobenius error $e_\delta = \|J - B_\delta\|_F / \max\{\|B_\delta\|_F, 10^{-12}\}$. These are finite-scale diagnostics; numerical rank is not a certified mathematical rank, and κ_η is not an exact estimate of the theoretical κ_+.

Table 18: **Paired simulator-response diagnostics at $\delta = 0.03$.** Entries are medians over 64 corresponding recorded-state anchors. Each model has eight zero-response anchors; condition-ratio medians use the remaining 56. The triples in the rank and ratio columns use thresholds $10^{-2}/10^{-3}/10^{-4}$. Gain and error medians include all anchors.

Configuration	$r_\eta(B_\delta)$	$\kappa_\eta(B_\delta)$	$g(B_\delta)$	$\|J - B_\delta\|_2$	e_δ
No shaping	3/3/3	9.00/9.65/9.65	8.464	7.611	0.949
ICT + gain	3/3/3	6.25/8.45/8.45	9.264	7.581	0.963
ICT + shape	3/3/3	7.00/7.90/7.90	9.536	9.306	1.125
ICT + shape + gain	3/3/3	7.43/8.28/8.28	11.749	9.238	1.042

Table 19: **Normalized singular spectra.** Coordinate-wise medians of σ_i/σ_1 for the same geometry models. J uses all 64 predictor contexts; $B_{0.03}$ uses the 56 nonzero simulator-response anchors. Values rounded to 0.000 need not be exactly zero. Normalization exposes spectral shape; absolute gain is reported in Tables 17 and 18.

Configuration	Map	1	2	3	4	5	6	7	8	9	10
No shaping	J	1.000	0.716	0.119	0.060	0.034	0.023	0.016	0.011	0.006	0.001
	$B_{0.03}$	1.000	0.521	0.115	0.002	0.000	0.000	0.000	0.000	0.000	0.000
ICT + gain	J	1.000	0.647	0.131	0.071	0.042	0.028	0.021	0.013	0.007	0.000
	$B_{0.03}$	1.000	0.469	0.131	0.003	0.000	0.000	0.000	0.000	0.000	0.000
ICT + shape	J	1.000	0.987	0.981	0.976	0.972	0.969	0.965	0.961	0.956	0.944
	$B_{0.03}$	1.000	0.495	0.130	0.003	0.000	0.000	0.000	0.000	0.000	0.000
ICT + shape + gain	J	1.000	0.989	0.981	0.975	0.971	0.967	0.964	0.958	0.952	0.940
	$B_{0.03}$	1.000	0.472	0.126	0.003	0.000	0.000	0.000	0.000	0.000	0.000

Scale dependence and transfer. At $\delta = 0.03$, all four models have median resolved simulator rank 3, whereas the shape-regularized predictors are full-rank and nearly isometric. Their paired relative errors remain near or above one. Thus successful optimization of the predictor's spectral target does not establish close approximation of the measured environment response. The conditional median truncated ratios vary moderately across models, but changes in rank, gain, and coordinate-dependent encoder responses prevent interpreting these ratios as an isolated inverse-risk comparison.

At steps 0.003, 0.01, and 0.03, respectively, 40, 22, and 8 of the 64 anchors have zero measured simulator response for every model. This strong scale dependence is compatible with contact and rendering resolution effects, and makes relative-error summaries with floored zero denominators especially uninformative at the smallest scale. The tables report zero-response counts alongside the spectra and errors rather than treating zero-response anchors as well-conditioned observations.

Reset fidelity. Recorded-state pairing does not guarantee identical latent contexts after simulator reset. The median norm between the re-rendered anchor latent and the recorded final context latent is 0.595, 0.708, 0.641, and 0.603 for the unshaped, gain, shape, and joint-penalty models, respectively. The measured derivative discrepancies therefore include reset/rendering differences as well as predictive mismatch. The comparison characterizes surrogate fidelity at corresponding recorded states; applying the transfer bound in Remark 3 would additionally require identical latent contexts.

F Extended Related Work

Latent world models and planning. World models support control through several computational strategies, including recurrent latent prediction (Ha & Schmidhuber, 2018; Hafner et al., 2019b), actor–critic learning from imagined trajectories (Hafner et al., 2019a; 2025), search with learned models (Schrittwieser et al., 2020), and latent MPC (Hansen et al., 2024). Reconstruction-free approaches learn features through joint-embedding objectives (LeCun et al., 2022; Assran et al., 2023; Bardes et al., 2024) and use them for action-conditioned prediction and planning (Zhou et al., 2024; Sobal et al., 2025; Maes et al., 2026). Our focus is the representation of action effects during predictive world-model training. We study independently usable inverse supervision and local action-response regularization during training, then evaluate the resulting models under both search-based and direct control.

Distributional regularization and collapse prevention. Self-supervised representation learning prevents collapse through several complementary mechanisms. VICReg and Barlow Twins use variance or cross-correlation constraints (Bardes et al., 2021; Zbontar et al., 2021), while BYOL uses asymmetric prediction with a moving-average target (Grill et al., 2020) and SimCLR uses a contrastive objective (Chen et al., 2020). LeJEPA explicitly regularizes the latent marginal with a sketched normality test (Balestriero & LeCun, 2025; Epps & Pulley, 1983). Rank-based measures summarize embedding diversity (Garrido et al., 2023; Roy & Vetterli, 2007). SIGReg compares random one-dimensional projections of an embedding batch with the standard normal through an Epps–Pulley characteristic-function statistic (Epps & Pulley, 1983). IC-LeWM complements this marginal regularization with training signals for action recovery and action-dependent latent geometry.

Geometry-aware representation learning. Contractive autoencoders (Rifai, 2011), Jacobian regularization (Hoffman et al., 2019), isometric autoencoders (Gropp et al., 2020), orthogonality penalties (Bansal et al., 2018) and spectral normalization (Miyato et al., 2018) constrain different aspects of a learned map's sensitivity. IC-LeWM applies geometric regularization to the predictor's action-to-latent map and connects its spectrum to inverse risk. Latent path straightness (Hénaff et al., 2019) provides a complementary view of trajectory geometry.