PRESERVING UNSTABLE MODES THROUGH INVERSE DYNAMICS IN JEPA WORLD MODELS

Anonymous authors
Paper under double-blind review

ABSTRACT

Robotic systems often exhibit unstable modes, along which small perturbations and disturbances can cause unbounded growth unless corrected through feedback. Controlling such systems from high-dimensional visual observations requires representations that preserve these modes. Joint-embedding predictive architectures (JEPAs) provide a natural framework for learning representations and their dynamics from visual data. However, we demonstrate that next step prediction combined with anti-collapse regularization provides no such guarantee: the training loss can be minimized while controllable unstable modes are collapsed, making stabilization from the learned representation impossible. To circumvent this, we augment predictive world-model training with an action reconstruction objective (i.e., an inverse dynamics loss) that encourages control-aware representations, namely, visual representations that preserve crucial features for control. We prove that exact action reconstruction makes the encoder injective on the finite-horizon reachable subspace; hence, no state direction produced by an action sequence within H steps can be discarded. Moreover, as H grows, the dominant eigenspace of the finite-horizon controllability Gramian reveals the controllable unstable subspace. Although our theory considers linear systems, experiments on nonlinear visual control tasks (CartPole, Walker2D, and PointMaze) validate our findings and show the benefits of control-aware representation learning.

1 INTRODUCTION

Quadrotors and legged robots operate around unstable equilibria where small perturbations can grow catastrophically without corrective feedback (Mellinger & Kumar, 2011; Ames et al., 2014). Controlling such systems requires actions that not only minimize a cost (e.g., goal reaching), but also *stabilize* the physical dynamics; otherwise, incorrect feedback can rapidly drive the system from its intended operating regime and compromise safe deployment (Ames et al., 2016; Marco et al., 2021).

Latent world models have found their way into robotics as a computationally efficient solution for learning dynamics and designing control actions from high-dimensional sensory observations. They compress pixel observations into a lower-dimensional representation, learn action-conditioned latent dynamics, and optimize actions by rolling out these dynamics, with limited interaction with the physical system (Ha & Schmidhuber, 2018; Hafner et al., 2019; 2020; 2021; 2023). Joint-embedding predictive architectures (JEPAs) follow this recipe without pixel-level reconstruction (LeCun, 2022; Sobal et al., 2025; Balestriero & LeCun, 2025): an encoder extracts a latent representation, a predictor forecasts its evolution, and anti-collapse mechanisms (e.g., SIGReg (Balestriero & LeCun, 2025), EMA (Assran et al., 2023), and others (Bardes et al., 2021))

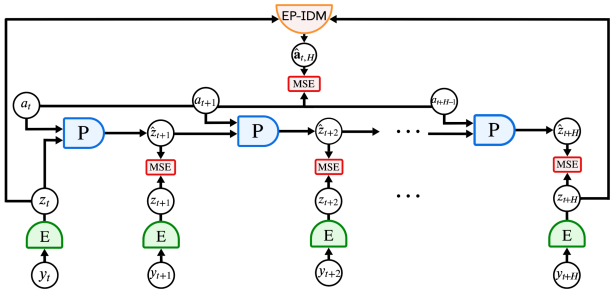

Figure 1: Architecture of our proposed world model. At each step, an encoder (E) maps observation y_t to a latent state z_t, and a predictor (P) rolls out latent predictions $\hat{z}_{t+1}, \ldots, \hat{z}_{t+H}$ conditioned on actions. Training losses are applied at each predicted latent: 1SP/MSP (single- or multi-step prediction) aligns $\hat{z}_{t+k}$ with the encoded ground-truth z_{t+k}, and SIG (SIGReg) prevents representation collapse. The EP-IDM estimates the intermediate actions $\hat{\mathbf{a}}_{t,H} = (\hat{a}_t, \hat{a}_{t+1}, \ldots, \hat{a}_{t+H-1})$ from z_t and z_{t+H}.

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.

prevent representation collapse, enabling efficient planning directly in latent space (Zhou et al., 2024; Toso et al., 2026; Wang et al., 2026a). However, deploying JEPAs on unstable systems remains largely unexplored.

When actions are designed from the learned latent representation and dynamics, the representation must preserve the physical information required for feedback. For open-loop unstable systems, this requires preserving the unstable modes that feedback must observe and correct (Lemma 1) (Hu et al., 2022; Werner & Peherstorfer, 2024; Toso et al., 2025; Lutkus et al., 2025). That is, if a controller cannot "see" an unstable mode, it cannot stabilize it.

However, latent predictive accuracy and anti-collapse regularization alone do not guarantee that this control-theoretic information is preserved. In fact, next-step prediction combined with anti-collapse regularization, the standard recipe for JEPAs, can be minimized while the encoder discards every unstable mode (Lemma 2). To preserve these modes, we augment the JEPA training with an endpoint inverse dynamics (EP-IDM) loss, which reconstructs the underlying action sequence from the initial and final latent states (see Figure 1), so that the encoder cannot discard directions the actions can reach (Mhammedi et al., 2020; Ivashkov et al., 2026). Our contributions are summarized below.

- **Prediction does not guarantee stabilization.** We show that a controller acting on the latent states can stabilize the system if and only if the encoder keeps every unstable mode (Lemma 1). We show that next-step prediction with SIGReg can reach its minimum while the encoder discards all unstable modes (Lemma 2)
- **Inverse dynamics preserves unstable modes.** We introduce EP-IDM, which reconstructs the action sequence from the initial and final latent states. We prove that, with sufficient action excitation, exact reconstruction keeps every direction reachable within H steps, and hence every unstable mode reachable within H steps (Theorem 1).
- **Unstable subspace recovery.** We prove that, as H grows, the dominant eigenspace of the finite-horizon controllability Gramian converges to the unstable subspace (Theorem 2).
- **Numerical validation.** We show on synthetic linear systems and a linearized CartPole that EP-IDM preserves the unstable direction and allows latent stabilization, while SIGReg does not, and we validate the benefit of EP-IDM on nonlinear visual control tasks (CartPole, Walker2D, and PointMaze) learned from images and proprioception.

2 Related Work

Joint embedding predictive world models. JEPAs learn representations by predicting future latents rather than reconstructing pixels (LeCun, 2022; Assran et al., 2023). Recent work applies this recipe to action-conditioned prediction and planning using pre-trained visual encoders (Zhou et al., 2024; Toso et al., 2026), end-to-end anti-collapse mechanisms (Balestriero & LeCun, 2025; Maes et al., 2026; Kuang et al., 2026), or objectives that shape latent dynamics and geometry (Sobal et al., 2025; Parthasarathy et al., 2025; Wang et al., 2026a; Zhang et al., 2026; Nath et al., 2026). While these works primarily address total representation collapse and planning, we demonstrate that standard JEPA objectives can selectively collapse directions required for feedback stabilization.

Of particular relevance is Ivashkov et al. (2026), who use one-step inverse dynamics to prevent collapse. We share the principle that actions provide a control-relevant learning signal but we study feedback stabilization of open-loop unstable systems. Our EP-IDM reconstructs the entire action sequence from the endpoint latents. We prove that exact reconstruction preserves the finite-horizon reachable subspace and characterize its convergence to the controllable unstable subspace, connecting action reconstruction to detectability and latent stabilizability.

Latent stabilization. Control from visual observations requires retaining the state information needed for stabilization. Prior work studies stabilization of linear systems under unknown nonlinear observation maps (Mhammedi et al., 2020), characterizes the sample complexity of learning to stabilize (Hu et al., 2022; Werner & Peherstorfer, 2024), and learns low-dimensional unstable subspace representations (Toso et al., 2025; Lutkus et al., 2025). In particular, preserving only the unstable subspace can suffice as stable modes already decay without correction (Toso et al., 2025). More than just constructing those sufficient representations, we also ask whether standard JEPA training can learn them in the first place. In particular, we connect latent world model training objectives to the control-theoretic information that their representations must preserved.

3 SETUP

Let us introduce the dynamical system, latent world model, and latent world model training objectives considered throughout the paper.

Learning an encoder and a predictor. We consider the discrete-time dynamical system

$$x_{t+1} = f(x_t, a_t), \qquad y_t = g(x_t) \in \mathcal{Y} \subset \mathbb{R}^p. \tag{1}$$

for all $t \geq 0$, where $x_t \in \mathcal{X} \subset \mathbb{R}^n$ is the physical state and $a_t \in \mathcal{A} \subset \mathbb{R}^m$ is the applied action at time t. The transition map $f : \mathcal{X} \times \mathcal{A} \to \mathcal{X}$ describes the physical dynamics, while the observation map $g : \mathcal{X} \to \mathcal{Y}$ maps the state to the available sensory observation y_t. This observation may contain images, proprioceptive information (i.e., partial or full physical state information), both, or other modalities (Wang et al., 2026b; Huang et al., 2026).

Our goal is to learn an observation encoder $E_\theta : \mathcal{Y} \to \mathcal{Z}$, an action encoder $G_\theta : \mathcal{A} \to \mathcal{U}$, and a latent dynamics model (predictor) $P_\theta : \mathcal{Z} \times \mathcal{U} \to \mathcal{Z}$ such that

$$z_{t+1} = P_\theta\big(z_t, G_\theta(a_t)\big) \text{ with latent state } z_t = E_\theta(y_t), \tag{2}$$

where $\mathcal{Z} \subset \mathbb{R}^d$ is the latent space, with $d \ll p$. For notation simplicity, we denote by $\theta \in \mathbb{R}^{d_\theta}$ the trainable parameters of the observation encoder, action encoder, and predictor.

Standard training objective. Let $\mathcal{D}$ denote the distribution of finite-horizon trajectory windows collected from the system in Eq. (1), where each window is composed of $H+1$ consecutive observations and H actions. A sampled window $\tau \sim \mathcal{D}$ starts at a trajectory-dependent time $t_\tau \geq 0$. In our experiments, the encoder uses two consecutive image observations, the difference between them (to further capture velocities), and the current proprioceptive state, so that

$$z_{t_\tau+k} = E_\theta\big(y_{t_\tau+k-1}, y_{t_\tau+k}, y_{t_\tau+k} - y_{t_\tau+k-1}, x_{t_\tau+k}\big), \qquad k = 0, \ldots, H. \tag{3}$$

We train the model using the next-step prediction objective

$$\mathcal{L}_{\text{pred}}(\theta) = \mathbb{E}_{\tau\sim\mathcal{D}}\left[\frac{1}{H}\sum_{k=0}^{H-1} \big\| P_\theta\big(z_{t_\tau+k}, G_\theta(a_{t_\tau+k})\big) - z_{t_\tau+k+1}\big\|_2^2\right]. \tag{4}$$

We combine this loss with SIGReg (Balestriero & LeCun, 2025) to prevent representation collapse:

$$\min_\theta \ \mathcal{L}_{\text{pred}}(\theta) + \lambda_{\text{reg}}\mathcal{L}_{\text{reg}}(\theta), \tag{5}$$

where λ_{reg} controls the weight of the regularizer (i.e., later we consider SIGReg and EP-IDM loss).

Planning in the latent space. Given an initial observation y_0 and a goal observation y_{g}, we encode $z_0 = E_\theta(y_0)$ and $z_{\text{g}} = E_\theta(y_{\text{g}})$ and optimize the action sequence through the learned latent dynamics. For a planning horizon T, we solve

$$\min_{\boldsymbol{a}} \quad \|z_T - z_{\text{g}}\|_2^2 \qquad \text{subject to} \qquad z_{t+1} = P_\theta\big(z_t, G_\theta(a_t)\big), \qquad t = 0, \ldots, T-1, \tag{6}$$

where $\boldsymbol{a} = (a_0, \ldots, a_{T-1})$ is the designed action sequence. In practice, we use a receding-horizon implementation. More details in Appendix A.3.

Open-loop stability. Let x_{e} be an equilibrium under zero input, i.e., $x_{\text{e}} = f(x_{\text{e}}, 0)$. The system is *locally open-loop stable* around x_{e} if trajectories initialized sufficiently close to x_{e} remain close to it when $a_t = 0$ for all t. It is *locally asymptotically stable* if these trajectories additionally converge to x_{e}, and it is *open-loop unstable* if the trajectories diverge from x_{e}.

For a linear system $x_{t+1} = Ax_t + Ba_t$, open-loop asymptotic stability is equivalent to A being Schur stable, i.e., $\rho(A) < 1$. Eigenvalues satisfying $|\lambda| > 1$ correspond to unstable modes, while those with $|\lambda| = 1$ are marginally stable. We emphasize that Schur stability prevents $\|x_t\| \to \infty$ as $t \to \infty$. Likewise, stabilization allows us to design the feedback that sends $\|x_t\|$ to zero.

Stabilization and unstable modes. For an open-loop unstable system, the goal is to design a feedback controller that selects actions from the available observations and makes the desired equilibrium asymptotically stable. Under a linear feedback $a_t = Kx_t$, this amounts to finding K such that the closed-loop matrix $A + BK$ is Schur stable. When the controller acts only on the learned latent state, this is possible only if the encoder preserves every unstable or marginally stable direction that must be corrected through feedback (Toso et al., 2025). In this work, we demonstrate that this requirement is not implied by an accurate latent next-step predictor alone.

Next, we formalize necessary and sufficient conditions for latent stabilization and demonstrate that the objective above in Eq. (5) does not necessarily satisfy them.

4 PREDICTION DOES NOT GUARANTEE STABILIZATION

An encoder can discard unstable modes while still learning predictable, non-constant representations. In the linear setting, we characterize the conditions required for stabilization and show that prediction loss and SIGReg do not necessarily satisfy them. Numerical examples are also provided in the end of this section and in Section 6, with additional details in Appendix A.5.

Linear setting. We consider a linear dynamical system (LDS), obtained by linearizing Eq.(1) around an equilibrium point

$$x_{t+1} = Ax_t + Ba_t, \qquad y_t = Cx_t, \tag{7}$$

for all $t \in \mathbb{Z}_+$, where $x_t \in \mathbb{R}^n$ and $a_t \in \mathbb{R}^m$ and A, B, and C are the transition, control, and observation matrices, respectively. Here, we consider an open-loop unstable system (i.e., $\rho(A) > 1$).

Definition 1 (Stabilizability, observability, and detectability)**.** *The pair (A, B) is* stabilizable *if there exists a gain K such that $A + BK$ is Schur stable under $a_t = Kx_t$. Given $q_t = Mx_t$, the pair (A, M) is* observable *if x_0 can be recovered from finitely many observations, and* detectable *if every unobservable mode is Schur stable. By the Popov-Belevitch-Hautus (PBH) test (Hautus, 1969), detectability is also equivalent to*

$$\operatorname{rank} \begin{bmatrix} \lambda I - A \\ M \end{bmatrix} = n \quad \textit{for every } \lambda \in \operatorname{spec}(A) \textit{ such that } |\lambda| \geq 1, \tag{8}$$

where $\operatorname{spec}(A)$ denotes the set of eigenvalues of A.

Assumption 1. *For the system in Eq. (7), the pair (A, B) is stabilizable and (A, C) is observable.*

We consider an affine observation encoder and a linear predictor:

$$E_\theta(y) = Wy + b = z \in \mathbb{R}^d \qquad P_\theta(z, a) = A_z z + B_z a \in \mathbb{R}^d \tag{9}$$

Here, $\theta = (W, b, A_z, B_z)$ collects all trainable parameters. We write the linear predictor directly in terms of the actions, as a linear action encoder can be absorbed into B_z without loss of expressivity.

4.1 NECESSARY AND SUFFICIENT CONDITIONS FOR STABILIZATION OF LDS

Let us first characterize the conditions under which a controller acting only on the learned latent representation can stabilize the physical system. We defer the proof of Lemma 1 to Appendix A.6.2.

Lemma 1. *Let $F = WC$ denote the state-to-latent map and consider the centered latent observation $\tilde{z}_t = z_t - b = Fx_t$. Suppose that Assumption 1 holds. Then there exists a causal dynamic controller that uses only the latent observations $\{\tilde{z}_\tau\}_{\tau=0}^t$ and asymptotically stabilizes Eq. (7) if and only if (A, F) is detectable. Equivalently, every unstable or marginally stable mode of A is observable through F, or, by the PBH test, there does not exist a nonzero vector $v \in \mathbb{C}^n$ such that*

$$Av = \lambda v, \qquad Fv = 0, \qquad |\lambda| \geq 1. \tag{10}$$

Lemma 1 demonstrates that unstable modes and marginally stable modes need to remain observable through the learned representation. In the next subsection, we demonstrate that minimizing prediction loss together with SIGReg does not guarantee this property.

4.2 THE COLLAPSE OF UNSTABLE MODES

Let $\Phi_{\geq 1}$ denote the unstable subspace of A associated with eigenvalues satisfying $|\lambda| \geq 1$. We next demonstrate that standard predictive training objectives (e.g., Eq. 5) can discard the entire subspace $\Phi_{\geq 1}$ while still reaching their minimum.

Lemma 2. *There exist open-loop unstable linear systems satisfying Assumption 1 and data distributions for which the prediction loss and SIGReg objective are simultaneously minimized by an encoder satisfying $\Phi_{\geq 1} \subset \ker(F)$.*

Proof. We begin our proof by setting the encoder bias to zero and consider coordinates in which

$$A = \begin{bmatrix} A_u & 0 \\ \Delta & A_s \end{bmatrix}, \; B = \begin{bmatrix} B_u \\ B_s \end{bmatrix}, \text{ and } x_t = \begin{bmatrix} x_{u,t} \\ x_{s,t} \end{bmatrix}, \tag{11}$$

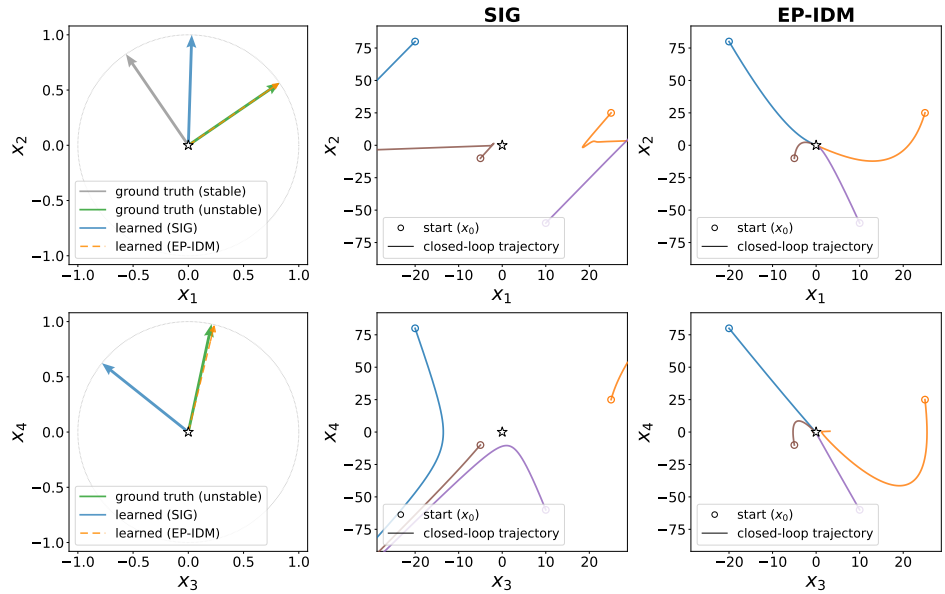

Figure 2: **SIG vs. EP-IDM across two examples.** Each row compares a representation trained with SIGReg (SIG) against one trained with the endpoint inverse-dynamics (action-reconstruction) loss (EP-IDM). **Top row: synthetic open-loop unstable LDS. Bottom row: linearized CartPole. Left:** true unstable (v_u, green) and stable (v_s, grey) eigenvectors versus the learned predictor's dominant eigenvector, for SIG (blue) and EP-IDM (dashed orange). **Center:** closed-loop trajectories under the latent LQR controller from four initial states (circles) for the model trained with SIGReg. **Right:** closed-loop trajectories under the latent LQR controller for the model trained with the inverse dynamics loss (EP-IDM).

and, for simplicity, we assume that $\Delta \equiv 0$. Here, we note that A_u is open-loop unstable and A_s is stable. We choose $F = [0\ F_s]$, where F_s is invertible on the stable coordinates. Then

$$z_{t+1} = F_s A_s F_s^{-1} z_t + F_s B_s u_t. \tag{12}$$

Therefore, choosing $A_z = F_s A_s F_s^{-1}$ and $B_z = F_s B_s$ implies zero one-step error, while every unstable direction lies in $\ker(F)$.

Note that SIGReg (Balestriero & LeCun, 2025) does not rule out this degenerate solution. Let $x_s \sim \mathcal{N}(0, \Sigma_s)$, with $\Sigma_s \succ 0$, and take $F_s = Q\Sigma_s^{-1/2}$ for any orthogonal matrix Q. Then z_t is distributed according to a standard Gaussian. The SIGReg objective in Eq. (5) is minimized despite the unstable directions being absent in the latent space. □

The constructed encoder therefore minimizes both objectives while discarding the entire unstable subspace. By Lemma 1, the physical system cannot be stabilized from this resulting latent representation. Our linear analysis removes representation capacity as a confounding factor: even with a linear observation-to-latent map and little compression, standard predictive objectives can discard precisely the state directions required for stabilization.

We illustrate Lemma 2 with a synthetic open-loop unstable system (top row of Figure 2) and a linearized CartPole system (bottom row). Appendix A.5 provides details and an additional open-loop stable example. The left panels compare the learned unstable eigenvector, mapped back to the physical state space, with the ground-truth unstable eigenvector. The center panels show closed-loop trajectories for the physical systems under latent LQR controllers trained with prediction and SIGReg through Eq. (5). In both systems, SIG learns a substantially misaligned unstable direction, and the resulting controller fails to stabilize the physical dynamics.

5 Preserving Unstable Directions Through Inverse Dynamics

To encourage the learned representation to preserve unstable modes, we augment the world-model training objective with an inverse dynamics loss (IDM). In particular, our endpoint inverse dynamics (EP-IDM) loss reconstructs the entire action sequence over the prediction horizon from the initial and final latent representations (see Fig. 1 for the complete architecture).

5.1 ENDPOINT INVERSE DYNAMICS

We train an inverse dynamics model $D_\theta : \mathcal{Z} \times \mathcal{Z} \to \mathcal{A}^H$ to reconstruct the action sequence from the latent representations of the initial and final observations[1]:

$$\mathcal{L}_{\text{EP-IDM}}(\theta) = \mathbb{E}_{\tau\sim\mathcal{D}}\left[\frac{1}{H}\sum_{t=0}^{H-1}\left\|\left[D_\theta(z_{t_\tau},\, z_{t_\tau+H})\right]_t - a_{t_\tau+t}\right\|_2^2\right]. \tag{13}$$

where $\left[D_\theta(z_{t_\tau},\, z_{t_\tau+H})\right]_t$ denotes the predicted action at timestep t. We jointly optimize the encoders, predictor, and inverse dynamics model solving Eq. (5) with $\mathcal{L}_{\text{EP-IDM}}(\theta)$ in lieu of $\mathcal{L}_{\text{reg}}(\theta)$.

5.2 PRESERVATION OF REACHABLE DIRECTIONS

We show that exact endpoint action reconstruction requires the encoder to preserve every direction reachable within H steps. For the linear analysis, we consider an affine inverse dynamics model

$$D_\theta(z_0, z_H) = W_H[z_0\ z_H] + b_H \text{ with } W_H \in \mathbb{R}^{mH\times 2d} \text{ and } b_H \in \mathbb{R}^{mH}. \tag{14}$$

Let $\mathcal{C}_H = \begin{bmatrix}A^{H-1}B & A^{H-2}B & \cdots & B\end{bmatrix}$ denote the finite-horizon controllability matrix and let $\mathcal{R}_H = \text{range}(\mathcal{C}_H)$ be the H-step reachable subspace.

We require the training data to contain sufficiently rich action excitation. For a fixed initial state x_{t_τ}, let $\text{supp}(\mathbf{a}_{\tau,H} \mid x_{t_\tau})$ denote the support of the conditional distribution of the stacked action sequence $\mathbf{a}_{\tau,H} := \begin{bmatrix}a_{t_\tau}^\top & \cdots & a_{t_\tau+H-1}^\top\end{bmatrix}^\top \in \mathbb{R}^{mH}$. The proof of Theorem 1 is provided in Appendix A.6.3

Theorem 1. *Consider the linear system and suppose that, for almost every initial state x_{t_τ}, the conditional support $\text{supp}(\mathbf{a}_{\tau,H} \mid x_{t_\tau})$ contains a nonempty open subset of $\mathbb{R}^{mH}$. If an affine inverse dynamics model achieves $\mathcal{L}_{\text{EP-IDM}} = 0$, then*

$$\ker(F) \cap \mathcal{R}_H = \{0\}. \tag{15}$$

Equivalently, the encoder is injective on the H-step reachable subspace: if $v \in \mathcal{R}_H$ and $Fv = 0$, then $v = 0$. Thus, no state direction that can be generated by an action sequence within H steps is discarded by the encoder. In particular, if $\Phi_{\geq 1} \subseteq \mathcal{R}_H$, then $\ker(F) \cap \Phi_{\geq 1} = \{0\}$. Exact reconstruction $\mathcal{L}_{\text{EP-IDM}} = 0$ holds only if $\text{rank}(F\mathcal{C}_H) = mH \leq \min\{d, n\}$.

We note that under Assumption 1 and a sufficiently large horizon H, we have that $\Phi_{\geq 1} \subseteq \mathcal{R}_H$. Therefore, the theorem above ensures that all unstable and marginally unstable modes remain observable through the representation, satisfying the detectability condition in Lemma 1.

Theorem 1 describes the idealized exact-reconstruction regime. We note that empirically, we use $\mathcal{L}_{\text{EP-IDM}}$ as a soft inductive bias and do not require it to be zero. Indeed, when $mH > d$, exact recovery of an arbitrary action sequence is not feasible. Nevertheless, minimizing the loss can still encourage the representation to keep the action-induced directions required for control (Section 6).

Returning to the examples in Fig. 2, training the latent world model with endpoint inverse dynamics loss (EP-IDM) preserves the unstable direction of the physical system in the latent space. The learned latent dynamics captures the unstable mode that must be corrected through feedback, allowing for latent stabilization, as shown in the right panels of Fig. 2.

In addition, our result in Theorem 1 motivates the question we address next: *Which controllable directions preserved by the action reconstruction are the most identifiable ones?*

5.3 REACHABILITY AND UNSTABLE DYNAMICS

We consider the finite-horizon controllability Gramian given by $\Pi_H = \sum_{k=0}^{H-1} A^k BB^\top (A^k)^\top$. For simplicity, we exclude eigenvalues on the unit circle, so $\Phi_{\geq 1}$ coincides with the strictly unstable subspace Φ. Let $\|\sin\Theta(\hat{\Phi}, \Phi)\|_2$ denote their subspace distance, as defined in Definition 2.

Theorem 2. *Under the mild assumptions given in Appendix A.6.4, let $\hat{\Phi}$ be the span of the top r eigenvectors of Π_H. Then, we have*

$$\|\sin\Theta(\hat{\Phi}, \Phi)\|_2 \to 0 \qquad \text{as } H \to \infty. \tag{16}$$

[1] We trained two other variants of this inverse dynamics loss that are detailed in Appendix A.3.4.

The convergence rates and proof are provided in Appendix A.6.4.

Taken together, Theorems 1 and 2 demonstrate that action reconstruction, in particular, EP-IDM, prevents the encoder from discarding finite-horizon controllable directions, while the controllability Gramian explains why the dominant action-induced controllable directions increasingly align with the unstable subspace as H grows, i.e., why the unstable directions are the most identifiable ones.

Note that for simplicity, Theorem 2 does not consider eigenvalues on the unit circle, therefore the instantiation of this result to the linearized CartPole example in Figure 2 is only characterized by its strictly unstable mode which is depicted in the left panel of Figure 2.

6 Experiments

We now validate our approach on nonlinear visual-control tasks in MuJoCo (Todorov et al., 2012). Our experiments address three questions: (i) whether standard predictive JEPA objectives (i.e., solving Eq. 5 with SIGReg as regularizer) preserve the information required for feedback stabilization, (ii) whether EP-IDM preserves that information, and (iii) whether the resulting representations remain useful for stable systems. We study CartPole, Walker2D, and PointMaze systems. Additional experiments and ablations are provided in Appendix A.1.

6.1 Setup

We consider both encoders trained from scratch and frozen pretrained encoders (DINOv2 (Oquab et al., 2023) and iBOT (Zhou et al., 2021), with results reported in Appendix A.2), for which we train a projector on top. Our training objectives combine either one-step prediction (1SP) or autoregressive multi-step prediction (MSP) with either SIGReg (SIG) or an inverse-dynamics objective (IDM).

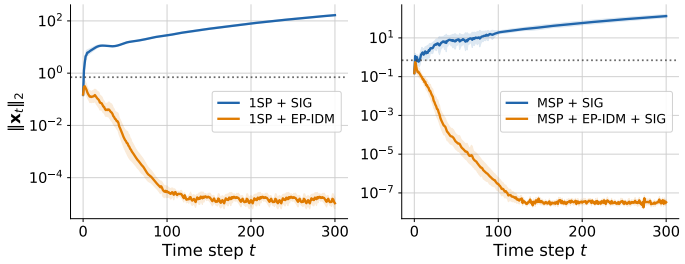

Figure 3: **CartPole** state norm $\|x_t\|_2$ under the designed latent LQR controller over 300 steps.

Our primary objective is endpoint action reconstruction (EP-IDM), for which the decoder reconstructs the complete action sequence from the first and last latent representations. We additionally evaluate one-step action reconstruction (IDM), which reconstructs each action from consecutive latent representations, and multi-step action reconstruction (MS-IDM), which reconstructs the complete action sequence from the corresponding latent trajectory. Further details are provided in Appendix A.3.4. Architecture, optimization, and dataset details are provided in Appendix A.3.

Systems. CartPole requires stabilizing an open-loop unstable upright equilibrium, where small angular perturbations can make the pole fall. Walker2D is also open-loop unstable because maintaining a walking gait requires continuous feedback. PointMaze instead evaluates goal reaching in a U-shaped maze with open-loop stable point-mass dynamics.

Controllers and planners. We evaluate three uses of the learned world model. Latent LQR linearizes the predictor at the equilibrium and provides our most direct diagnostic: a collapsed unstable mode cannot be identified or corrected by the resulting feedback controller. CEM searches over sampled open-loop rollouts with repeated replanning, whereas GBP differentiates the terminal latent cost through the predicted rollout. Unlike LQR, neither requires the local linearization to capture the unstable modes. We report success over ten trials and provide all hyperparameters in Appendix A.3.

Metrics. We report success rate (SR). For CartPole, success requires $\|x_T\|_2 \leq 0.7$ after the 300-step rollout, and the mean fraction of steps (MFS) satisfying $\|x_t\|_2 \leq 0.7$ measures how consistently the controller remains near the upright equilibrium. For PointMaze, success requires $\|p_T - p_g\|_2 \leq 0.5$, where p_T and p_g are the final and goal positions. For Walker2D, we evaluate whether the learned latent dynamics preserve the limit cycle induced by the locomotion policy used to collect the training data. Additional details are provided in Appendix A.3.

6.2 CartPole: Stabilization Around an Unstable Equilibrium

Table 1 shows the benefit of action reconstruction when the encoder and predictor are trained from scratch. Prediction with SIGReg yields zero latent-LQR success and low MFS, whereas adding EP-

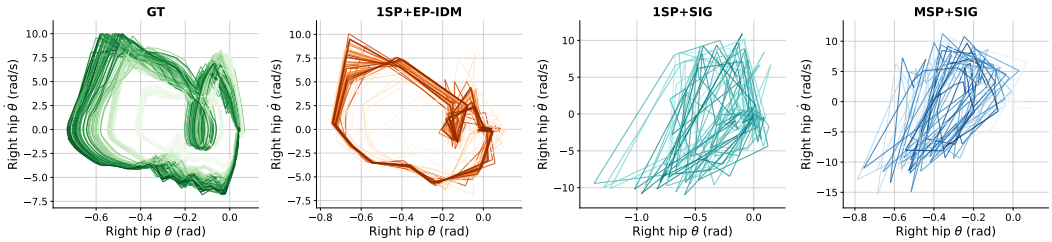

Figure 4: **Walker2D.** Phase portraits of right-hip angle vs. angular velocity for GT dynamics, 1SP+EP-IDM, 1SP+SIG, and MSP+SIG. Each trajectory is a 600-step closed-loop rollout: GT uses real MuJoCo physics, model panels roll out in latent space under an MLP-decoded state, with a SAC (Huang et al., 2022; Haarnoja et al., 2018) policy acting on the decoded observation.

Table 1: **CartPole** stabilization success rate (%).

Model	LQR (SR)	LQR (MFS)	CEM (SR)	GBP (SR)
Ground truth (GT)	**100%**	**1.000**	**100%**	**100%**
1SP + SIG	0%	0.202	**100%**	90%
MSP + SIG	0%	0.247	**100%**	**100%**
1SP + EP-IDM	**100%**	0.999	**100%**	**100%**
MSP + EP-IDM + SIG	**100%**	0.996	**100%**	**100%**

IDM achieves 100% success. This confirms that standard prediction and anti-collapse regularization can discard controllable unstable directions, while action reconstruction encourages the representation to preserve the finite-horizon reachable directions containing them.

CEM and GBP succeed even with SIGReg models (Table 1). While LQR relies on a local linearization and directly verifies whether the learned model supports local feedback, CEM and GBP optimize finite action sequences through nonlinear predictor rollouts. Their success indicates that SIGReg models retain enough information for nonlinear replanning, even when their local latent linearizations do not yield stabilizing LQR controllers. Figure 3 makes the failure of the standard JEPA recipe more explicit: under 1SP+SIG and MSP+SIG, the state norm grows unbounded, whereas EP-IDM enables the latent LQR controller to drive the state toward the equilibrium and keep it there.

6.3 Walker2D: Preserving a Walking Gait

For Walker2D, we evaluate whether action reconstruction preserves control-relevant dynamics beyond a single unstable equilibrium. A walking gait forms a limit cycle whose deviations must be continuously corrected. Figure 4 compares the right-hip phase portrait from the ground-truth MuJoCo simulator with latent rollouts from 1SP+EP-IDM, MSP+SIG, and 1SP+SIG. EP-IDM preserves the cyclic structure and range of joint angles and angular velocities, whereas 1SP+SIG and MSP+SIG fail to capture the limit cycle induced by the locomotion policy. Thus, action reconstruction also preserves control-relevant information associated with a locomotion orbit.

6.4 Diagnostics

Figure 5 compares the local dynamics learned by 1SP+EP-IDM and MSP+SIG. 1SP+EP-IDM reproduces the ground-truth vector field near the upright equilibrium with low three-step prediction error, whereas MSP+SIG captures only parts of the field and exhibits larger, anisotropic error. Both models produce a latent goal cost with a minimum near the equilibrium, which CEM and GBP can exploit through finite-horizon optimization, while LQR requires an accurate local linearization.

Table 2: **PointMaze** navigation success rate (%).

Encoder	Objective	CEM (SR)	LQR (SR)	GBP (SR)
GT		**100%**	**80%**	90.0%*
From scratch	1SP + IDM	**100%**	70%	**100%**
	MSP + MS-IDM	**100%**	**80%**	90.0%
	MSP + EP-IDM	90.0%	**80%**	90.0%
	1SP + EP-IDM	**100%**	**80%**	80.0%
	MSP + SIG	90.0%	60%	**100%**
	1SP + IDM + MS-IDM	90.0%	**80%**	90.0%
DINOv2	MSP	50.0%	20%	80%
iBOT	MSP	**100%**	**80%**	80%
DINO-WM		**80%**	**80%**	60%

* Because MuJoCo is non-differentiable, we estimate gradients with SPSA using two rollouts per optimization step (Spall, 1998).

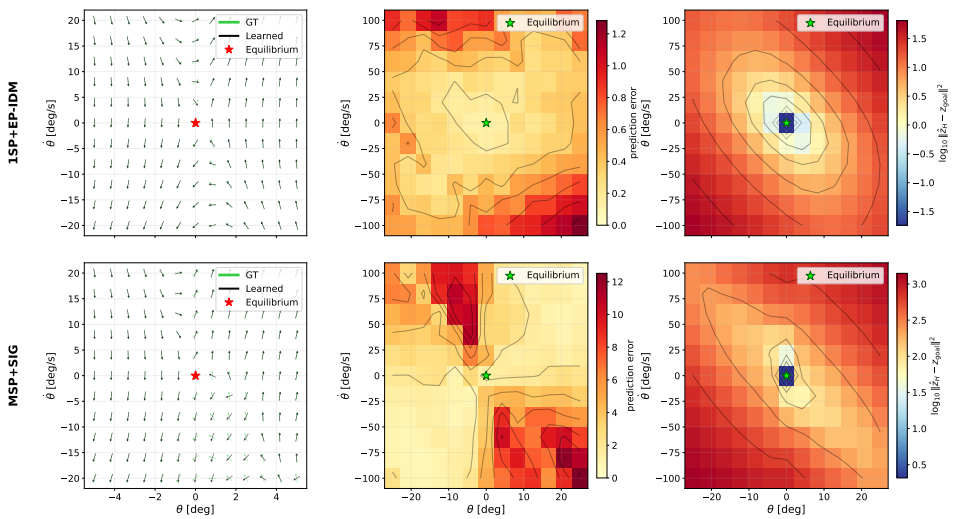

Figure 5: **CartPole** diagnostics. **Left:** Phase portraits of ground-truth and learned vector fields. **Center:** $H = 3$ open-loop prediction error across $(\theta, \dot{\theta})$. **Right:** Zero-action planning cost.

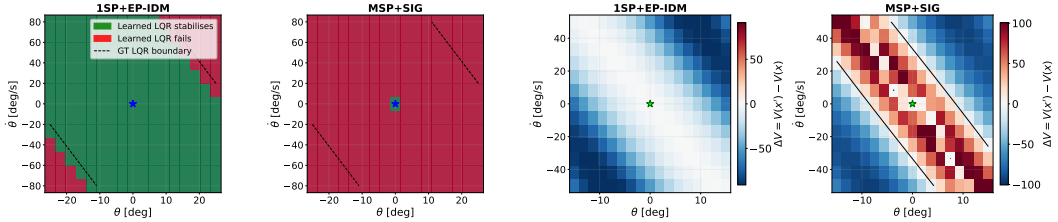

Figure 6: Local stability on **CartPole**. **First two panels:** Empirical region of attraction of the learned LQR controller when deployed in ground-truth system. The dashed line indicates the ground-truth LQR boundary. **Last two panels:** Lyapunov decrease $\Delta V = V(x') - V(x)$ under the latent LQR controller.

Figure 6 shows the empirical region of attraction (ROA) and Lyapunov difference under the latent LQR controllers. The controller obtained from 1SP+EP-IDM stabilizes almost every evaluated initial condition and yields $\Delta V = V(x_{t+1}) - V(x_t) < 0$ around the equilibrium, providing an empirical certificate of local closed-loop stability. In contrast, the MSP+SIG controller stabilizes only a small neighborhood around the equilibrium.

Although MSP+SIG provides a terminal cost suitable for nonlinear planning (Figure 5), its local representation and predictor do not yield a latent stabilizing controller. Action reconstruction instead learns local dynamics from which a stabilizing controller can be designed, consistent with Section 5.

We evaluate whether action reconstruction remains effective for systems with no unstable modes using PointMaze (Table 2). We note that all action-reconstruction variants trained from scratch reach at least 90% CEM success and at least 70% LQR success.

7 Conclusions and Future Work

In this work, we have shown a fundamental limitation of next-step prediction combined with standard anti-collapse regularization (e.g., SIGReg) in JEPAs: these objectives do not necessarily preserve the control-theoretic information required to stabilize an open-loop unstable system. They can be minimized while controllable unstable modes are discarded from the latent representation. Therefore, the latent representation may not collapse on the training distribution, yet be insufficient for designing a stabilizing latent feedback controller. To overcome this limitation, we proposed endpoint inverse dynamics (EP-IDM), which reconstructs the action sequence from the initial and final latent representations. For linear systems, we proved that exact reconstruction preserves the finite-horizon reachable subspace and prevents the collapse of reachable unstable modes, while the dominant eigenspace of the controllability Gramian converges to the controllable unstable subspace as the prediction horizon grows. Empirically, EP-IDM increases latent-LQR success from 0% to 100% on nonlinear CartPole, preserves the Walker2D limit cycle, and achieves strong planning performance on PointMaze. Future work could combine control-aware representations with latent-geometry objectives for efficient and robust planning, including bisimulation (Toso et al., 2026) and temporal straightening (Wang et al., 2026a).

AI USE STATEMENT

We used generative AI tools only to edit and polish the manuscript for clarity, grammar, and readability. We did not use generative AI tools to conceive the project idea or develop the theoretical results, formulate mathematical claims, construct proofs, process data, or interpret the results. All AI-assisted edits were reviewed and verified by the authors, who take full responsibility for the final content of this work.

REFERENCES

Aaron D Ames, Kevin Galloway, Koushil Sreenath, and Jessy W Grizzle. Rapidly exponentially stabilizing control lyapunov functions and hybrid zero dynamics. *IEEE Transactions on Automatic Control*, 59(4):876–891, 2014.

Aaron D Ames, Xiangru Xu, Jessy W Grizzle, and Paulo Tabuada. Control barrier function based quadratic programs for safety critical systems. *IEEE transactions on automatic control*, 62(8): 3861–3876, 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 15619–15629, 2023.

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

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

Chandler Davis and William Morton Kahan. The rotation of eigenvectors by a perturbation. iii. *SIAM Journal on Numerical Analysis*, 7(1):1–46, 1970.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255. Ieee, 2009.

David Ha and Jürgen Schmidhuber. World models. *arXiv preprint arXiv:1803.10122*, 2(3), 2018.

Tuomas Haarnoja, Aurick Zhou, Pieter Abbeel, and Sergey Levine. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In *International conference on machine learning*, pp. 1861–1870. Pmlr, 2018.

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, 2019.

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

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

Danijar Hafner, Jurgis Pasukonis, Jimmy Ba, and Timothy Lillicrap. Mastering diverse domains through world models. *arXiv preprint arXiv:2301.04104*, 2023.

M.L.J. Hautus. Controllability and observability conditions of linear autonomous systems. *Indagationes Mathematicae (Proceedings)*, 72(5):443–448, 1969. ISSN 1385-7258.

Yang Hu, Adam Wierman, and Guannan Qu. On the sample complexity of stabilizing lti systems on a single trajectory. *Advances in Neural Information Processing Systems*, 35:16989–17002, 2022.

Hai Huang, Yann LeCun, and Randall Balestriero. Llm-jepa: Large language models meet joint embedding predictive architectures. In *International Conference on Learning Representations*, volume 2026, pp. 105717–105737, 2026.

Shengyi Huang, Rousslan Fernand Julien Dossa, Chang Ye, Jeff Braga, Dipam Chakraborty, Kinal Mehta, and João GM Araújo. Cleanrl: High-quality single-file implementations of deep reinforcement learning algorithms. *Journal of Machine Learning Research*, 23(274):1–18, 2022.

Petr Ivashkov, Randall Balestriero, and Bernhard Schölkopf. Sensorimotor world models: Perception for action via inverse dynamics. *arXiv preprint arXiv:2606.20104*, 2026.

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

Yann LeCun. 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.

Paul Lutkus, Kaiyuan Wang, Lars Lindemann, and Stephen Tu. Latent representations for control design with provable stability and safety guarantees. In *2025 IEEE 64th Conference on Decision and Control (CDC)*, pp. 2937–2944. IEEE, 2025.

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.

Alonso Marco, Dominik Baumann, Majid Khadiv, Philipp Hennig, Ludovic Righetti, and Sebastian Trimpe. Robot learning with crash constraints. *IEEE Robotics and Automation Letters*, 6(2): 1439–1446, 2021.

Daniel Mellinger and Vijay Kumar. Minimum snap trajectory generation and control for quadrotors. In *2011 IEEE international conference on robotics and automation*, pp. 2520–2525. Ieee, 2011.

Zakaria Mhammedi, Dylan J Foster, Max Simchowitz, Dipendra Misra, Wen Sun, Akshay Krishnamurthy, Alexander Rakhlin, and John Langford. Learning the linear quadratic regulator from nonlinear observations. *Advances in Neural Information Processing Systems*, 33:14532–14543, 2020.

Devesh Nath, Anutam Srinivasan, Haoran Yin, Ruitong Jiang, Jeffrey Fang, and Glen Chou. Pixels to proofs: Probabilistically-safe latent world model control via parallel conformal robust mpc. *arXiv preprint arXiv:2606.15594*, 2026.

Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, et al. Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193*, 2023.

Arjun Parthasarathy, Nimit Kalra, Rohun Agrawal, Yann LeCun, Oumayma Bounou, Pavel Izmailov, and Micah Goldblum. Closing the train-test gap in world models for gradient-based planning. *arXiv preprint arXiv:2512.09929*, 2025.

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.

Vlad Sobal, Wancong Zhang, Kyunghyun Cho, Randall Balestriero, Tim GJ Rudner, and Yann LeCun. Learning from reward-free offline data: A case for planning with latent dynamics models. *arXiv preprint arXiv:2502.14819*, 2025.

James C Spall. An overview of the simultaneous perturbation method for efficient optimization. *Johns Hopkins apl technical digest*, 19(4):482–492, 1998.

Emanuel Todorov, Tom Erez, and Yuval Tassa. Mujoco: A physics engine for model-based control. In *2012 IEEE/RSJ international conference on intelligent robots and systems*, pp. 5026–5033. IEEE, 2012.

Leonardo F Toso, Lintao Ye, and James Anderson. Learning stabilizing policies via an unstable subspace representation. In *2025 IEEE 64th Conference on Decision and Control (CDC)*, pp. 7543–7550. IEEE, 2025.

Leonardo F Toso, Davit Shadunts, Yunyang Lu, Nihal Sharma, Donglin Zhan, Nam H Nguyen, and James Anderson. Learning invariant visual representations for planning with joint-embedding predictive world models. *arXiv preprint arXiv:2602.18639*, 2026.

Ying Wang, Oumayma Bounou, Gaoyue Zhou, Randall Balestriero, Tim GJ Rudner, Yann LeCun, and Mengye Ren. Temporal straightening for latent planning. *arXiv preprint arXiv:2603.12231*, 2026a.

Ziyu Wang, Kun Fang, and Yann LeCun. Music-jepa: Learning a world model of sound from action. *arXiv preprint arXiv:2607.22000*, 2026b.

Steffen WR Werner and Benjamin Peherstorfer. On the sample complexity of stabilizing linear dynamical systems from data. *Foundations of Computational Mathematics*, 24(3):955–987, 2024.

Wancong Zhang, Basile Terver, Artem Zholus, Soham Chitnis, Harsh Sutaria, Mido Assran, Randall Balestriero, Amir Bar, Adrien Bardes, Yann LeCun, et al. Hierarchical planning with latent world models. *arXiv preprint arXiv:2604.03208*, 2026.

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.

Jinghao Zhou, Chen Wei, Huiyu Wang, Wei Shen, Cihang Xie, Alan Yuille, and Tao Kong. ibot: Image bert pre-training with online tokenizer. *arXiv preprint arXiv:2111.07832*, 2021.