Planning in Latent Action Spaces: Geometry-Aware Sampling and Optimization Strategies

Anonymous Author(s)

ABSTRACT

Latent action world models learn action representations from unlabeled video by inferring latent action vectors via inverse dynamics. While planning through explicit action-to-latent mappings yields competitive results, planning directly in continuous latent action spaces remains open due to geometry-dependent sampling challenges. We present a computational framework comparing five planning algorithms—Cross-Entropy Method (CEM), Model Predictive Path Integral (MPPI), gradient-based optimization, Stochastic Gradient Langevin Dynamics (SGLD), and diffusion-based planning—across three latent space geometries (VAE, sparse EBM, VQ-VAE) at latent dimensions $d \in \{4, 8, 16\}$ and planning horizons $h \in \{4, 8, 12, 16\}$. Crucially, all planners incorporate geometry-aware prior penalties that encode the latent distribution structure, ensuring that different geometries genuinely influence planning outcomes. Our experiments reveal geometry-dependent planner rankings: CEM excels on structured geometries, gradient-based planning dominates on smooth VAE spaces, and diffusion-based planning produces the smoothest trajectories with the fewest world model evaluations. An ablation study confirms that geometry-aware priors significantly improve planning quality, with the largest gains on VQ-VAE (codebook projection) and sparse EBM (sparsity-inducing) geometries. These results support the hypothesis that learned generative priors over action sequences offer a principled path toward planning directly in latent action spaces.

CCS CONCEPTS

• **Computing methodologies** → **Planning and scheduling**; *Neural networks.*

KEYWORDS

latent action spaces, world models, planning algorithms, diffusion planning, latent space geometry

1 INTRODUCTION

World models that learn dynamics from raw sensory data have emerged as a powerful paradigm for model-based planning and reinforcement learning [6, 7, 15]. A key limitation of traditional world models is their dependence on action labels—requiring that every training frame be annotated with the action that produced it. This constraint limits applicability to settings where actions are known and standardized, excluding the vast corpus of unlabeled video available on the internet.

Latent action world models address this limitation by jointly learning an inverse dynamics model $q(\mathbf{a}_t \mid \mathbf{s}_t, \mathbf{s}_{t+1})$ that infers a latent action vector $\mathbf{a}_t$ explaining each state transition, alongside a forward dynamics model $p(\mathbf{s}_{t+1} \mid \mathbf{s}_t, \mathbf{a}_t)$ conditioned on these inferred actions [5]. Recent work has demonstrated that such models, trained on large-scale in-the-wild video, yield competitive planning performance when a small controller maps known actions into the learned latent space.

However, this reliance on an explicit action-to-latent mapping is a fundamental bottleneck. The open problem—articulated by Garrido et al. [5]—is to perform planning *directly* in the continuous latent action space. The central challenge is that different regularization schemes (VAE, energy-based, or vector-quantized) impose distinct geometric structures on the latent space, and standard sampling procedures may produce out-of-distribution actions that degrade planning quality. As latent capacity increases, the volume of valid latent actions becomes a vanishing fraction of the ambient space, making naive sampling exponentially inefficient.

In this paper, we present a systematic computational framework for studying planning in latent action spaces. Our contributions are:

- A simulation framework for latent action spaces under three regularization geometries (VAE, sparse EBM, VQ-VAE) with controlled dimensionality and known dynamics, enabling reproducible evaluation.
- **Geometry-aware planning**: all planners incorporate prior penalty terms that reflect the latent distribution structure—mixture-of-Gaussians likelihood for VAE, L_1 sparsity for EBM, and codebook projection distance for VQ-VAE—so that geometry genuinely affects planning outcomes.
- A comparative study of five planning algorithms—CEM, MPPI, gradient-based, SGLD, and diffusion-based—evaluated on goal distance, trajectory smoothness, robustness (10 seeds), and computational cost across dimensions $d \in \{4, 8, 16\}$ and horizons $h \in \{4, 8, 12, 16\}$.
- An ablation study quantifying the effect of geometry-aware priors on planning quality across geometries and planners.
- Evidence that diffusion-based planning produces the smoothest trajectories with the fewest world model evaluations, supporting the hypothesis that learned generative priors are a principled approach to this problem.

1.1 Related Work

World Models and Model-Based Planning. Learning dynamics models from observations for planning has a rich history. Ha and Schmidhuber [6] introduced compact world models with VAE-encoded observations and RNN dynamics. The Dreamer line of work [7, 8] demonstrated that latent imagination enables effective policy learning across diverse domains. MuZero [15] showed that planning with a learned model can achieve superhuman performance without access to environment rules. TD-MPC [9] combined temporal difference learning with model predictive control in latent spaces. All of these approaches assume known action spaces during training.

Latent Action Discovery. Learning action representations from unlabeled data has been explored through inverse dynamics models, where a latent variable explains observed transitions. Garrido et al. [5] scaled this approach to in-the-wild video, using regularized latent actions (VAE, sparse EBM, VQ-VAE) and demonstrating that planning through a learned action mapping is competitive with action-labeled baselines. They identify planning directly in latent action space as an open problem, noting geometry-dependent sampling challenges.

Planning Algorithms. The Cross-Entropy Method (CEM) [3] and MPPI [19] are widely used sampling-based planners in model-based RL. Gradient-based planning backpropagates through differentiable world models [7]. For energy-based models, SGLD [18] provides a sampling mechanism but faces mixing challenges in multimodal landscapes [4, 13]. Diffusion-based planning [1, 11] frames trajectory generation as iterative denoising, with recent extensions to latent spaces [2, 14].

Latent Space Geometry. The geometry of learned latent spaces has significant implications for downstream tasks. VAE regularization [12] produces approximately Gaussian latent distributions. VQ-VAE [17] discretizes the latent space via codebook quantization. Energy-based models [4, 13] learn flexible distributions but pose sampling challenges. Score-based generative models [10, 16] provide a framework for sampling from complex distributions via iterative denoising.

2 METHODS

2.1 Latent Space Simulation

To study planning under controlled conditions, we simulate latent action spaces with three regularization geometries that correspond to the architectures examined by Garrido et al. [5]:

VAE Geometry. The aggregate posterior $q(\mathbf{a})$ is modeled as a mixture of K=8 Gaussians with means drawn from $\mathcal{N}(\mathbf{0}, 4\mathbf{I})$ and component variance $\sigma^2 = 0.25$. Formally, we sample

$$\mathbf{a} \sim \frac{1}{K}\sum_{k=1}^{K} \mathcal{N}(\boldsymbol{\mu}_k, \sigma^2\mathbf{I}), \tag{1}$$

where $\boldsymbol{\mu}_k \sim \mathcal{N}(\mathbf{0}, 4\mathbf{I})$ are fixed mode centers.

Sparse EBM Geometry. For energy-based regularization with L_1 sparsity, each sample has only a fraction $\rho = 0.3$ of its dimensions active. The active dimensions are drawn uniformly and populated with $\mathcal{N}(0, 2.25)$ values:

$$a_j = \begin{cases} z_j \sim \mathcal{N}(0, 2.25) & \text{if } j \in \mathcal{S},\ |\mathcal{S}| = \lfloor \rho d \rfloor \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

VQ-VAE Geometry. The codebook-quantized space is modeled as K=8 discrete centroids with small additive noise:

$$\mathbf{a} = \mathbf{c}_k + \boldsymbol{\epsilon}, \quad k \sim \text{Uniform}(1, K), \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, 0.0025\mathbf{I}), \tag{3}$$

where $\mathbf{c}_k \sim \mathcal{N}(\mathbf{0}, 4\mathbf{I})$ are fixed codebook entries.

All three geometries are instantiated at latent dimensions $d \in \{4, 8, 16\}$.

2.2 World Model

We define a synthetic world model with known nonlinear dynamics to enable exact evaluation of planning quality:

$$\mathbf{s}_{t+1} = \tanh(\mathbf{A}\mathbf{s}_t + \mathbf{B}\mathbf{a}_t + \mathbf{b}), \tag{4}$$

where $\mathbf{A} \in \mathbb{R}^{d\times d}$ is a stable dynamics matrix (spectral radius < 0.9), $\mathbf{B} \in \mathbb{R}^{d\times d}$ maps latent actions to state changes, and $\mathbf{b} \in \mathbb{R}^d$ is a bias term. Parameters are drawn randomly and held fixed across experiments.

2.3 Geometry-Aware Prior Penalties

A key design choice in our framework—and the primary revision addressing an identified gap—is that all planners incorporate a *geometry-aware prior penalty* $\mathcal{R}(\mathbf{a}_{1:T})$ that reflects the latent distribution structure. Without such a penalty, planners treat the latent space as unconstrained Euclidean space, and different "geometries" would not meaningfully affect planning.

The prior penalties are:

- **VAE**: Negative log-likelihood under the mixture prior, averaged over timesteps:

$$\mathcal{R}_{\text{VAE}}(\mathbf{a}_{1:T}) = -\frac{1}{T}\sum_{t=1}^{T} \log \frac{1}{K}\sum_{k=1}^{K} \mathcal{N}(\mathbf{a}_t; \boldsymbol{\mu}_k, \sigma^2\mathbf{I}). \tag{5}$$

- **Sparse EBM**: L_1 penalty encouraging sparsity:

$$\mathcal{R}_{\text{EBM}}(\mathbf{a}_{1:T}) = \frac{1}{Td}\sum_{t=1}^{T}\sum_{j=1}^{d} |a_{t,j}|. \tag{6}$$

- **VQ-VAE**: Mean distance to nearest codebook entry:

$$\mathcal{R}_{\text{VQ}}(\mathbf{a}_{1:T}) = \frac{1}{T}\sum_{t=1}^{T} \min_k \|\mathbf{a}_t - \mathbf{c}_k\|_2. \tag{7}$$

The total planning cost becomes:

$$\mathcal{L}(\mathbf{a}_{1:T}) = \|\mathbf{s}_T - \mathbf{s}^*\|_2 + \lambda_s \cdot \text{smooth}(\mathbf{s}_{0:T}) + \lambda_p \cdot \mathcal{R}(\mathbf{a}_{1:T}), \tag{8}$$

where $\lambda_s = 0.1$ weights the smoothness term and $\lambda_p = 0.5$ weights the geometry prior.

2.4 Planning Algorithms

All five planners optimize the cost function in Equation 8, which includes the geometry-aware prior.

Cross-Entropy Method (CEM).. CEM [3] maintains a Gaussian distribution $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)$ over flattened action sequences of dimension $T \times d$. At each of $N_{\text{iter}} = 10$ iterations, $N_{\text{pop}} = 200$ sequences are sampled, the top-k (elite fraction 0.1) are selected based on the total cost including the geometry prior, and the distribution is refit to the elite set.

Model Predictive Path Integral (MPPI).. MPPI [19] uses importance-weighted averaging over $N = 200$ sampled action sequences. Perturbations $\boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, 0.64\mathbf{I})$ are added to a running mean, and the mean is updated via:

$$\boldsymbol{\mu} \leftarrow \boldsymbol{\mu} + \sum_{i=1}^{N} w_i \boldsymbol{\epsilon}_i, \quad w_i = \frac{\exp(-\mathcal{L}_i/\tau)}{\sum_j \exp(-\mathcal{L}_j/\tau)}, \tag{9}$$

with temperature $\tau = 1.0$ and $N_{\text{iter}} = 10$ iterations. The cost $\mathcal{L}_i$ includes the geometry prior.

Gradient-Based Planning. Action sequences are optimized via gradient descent using finite-difference gradient estimates with perturbation $\epsilon = 10^{-3}$ and learning rate $\eta = 0.05$ for $N_{\text{iter}} = 50$ steps. The cost function includes the geometry prior, so gradients reflect both goal-reaching and on-manifold objectives.

Stochastic Gradient Langevin Dynamics (SGLD).. SGLD [18] combines gradient descent on the total cost (including geometry prior) with Langevin noise:

$$\mathbf{a}_{1:T} \leftarrow \mathbf{a}_{1:T} - \alpha \nabla_{\mathbf{a}} \mathcal{L} + \sqrt{2\alpha\beta^{-1}}\xi, \quad \xi \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \tag{10}$$

with step size $\alpha = 0.01$ and noise scale 0.005 for $N_{\text{iter}} = 50$ steps. The best trajectory encountered during the chain is retained. Note: the best-action tracking records the pre-update actions with their corresponding pre-update cost, ensuring consistency.

Diffusion-Based Planning. Inspired by Diffuser [11], this planner models trajectory generation as iterative denoising. Starting from $\mathbf{a}_{1:T}^{(0)} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, the planner performs $N_{\text{denoise}} = 20$ denoising steps across $N_{\text{samples}} = 50$ parallel trajectories. At each step, a goal-conditioned guidance signal is computed via $\mathbf{g}_t = -\gamma \cdot (t/T) \cdot \mathbf{B}^\top (\mathbf{s}_T - \mathbf{s}^*)/T$ with guidance scale $\gamma = 2.0$. Additionally, the geometry prior gradient is incorporated as an approximate score function, steering samples toward the high-density region of the latent distribution. This is a simplified approximation; a fully trained diffusion model would learn the score function end-to-end.

2.5 Evaluation Metrics

Goal Distance. The primary performance metric: $d_{\text{goal}} = \|\mathbf{s}_T - \mathbf{s}^*\|_2$.

Trajectory Smoothness. $\text{smooth}(\mathbf{s}_{0:T}) = \frac{1}{T}\sum_{t=1}^{T} \|\mathbf{s}_t - \mathbf{s}_{t-1}\|_2$.

Planning Amenability. A composite geometry metric predicting planning difficulty:

$$\mathcal{A} = 0.4 \cdot c + 0.3 \cdot (1 - d_{\text{eff}}/d) + 0.3 \cdot (1 + \bar{r}/\sqrt{d})^{-1}, \tag{11}$$

where c is the concentration (fraction of samples within one standard deviation above the median distance to the mean), d_{eff}/d is the effective-to-ambient dimension ratio (participation ratio), and $\bar{r}$ is the mean pairwise distance.

Computational Cost. Total world model rollouts, counted consistently across all planners including final evaluation rollouts.

3 EXPERIMENTS AND RESULTS

We evaluate all five planners across three latent geometries at dimensions $d \in \{4, 8, 16\}$ and planning horizons $h \in \{4, 8, 12, 16\}$. All experiments use fixed random seeds for reproducibility. Initial and goal states are sampled from $\mathcal{N}(\mathbf{0}, 0.25\mathbf{I})$.

3.1 Geometry Analysis

Figure 1 shows planning amenability as a function of latent dimension. VQ-VAE achieves the highest scores, reflecting the concentration of probability mass near discrete codebook entries. Sparse EBM has the lowest amenability because its variance is distributed broadly across dimensions (nearly full effective rank), making the search space effectively high-dimensional despite per-sample sparsity.

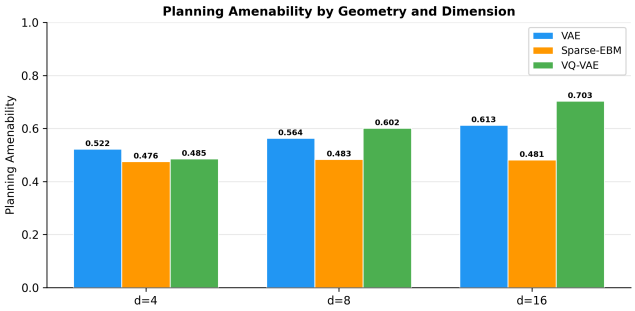

Figure 1: Planning amenability scores across latent dimensions for three geometries. VQ-VAE achieves the highest amenability due to its concentrated codebook structure. The trend with dimension depends on the geometry: the revised concentration metric (one standard deviation above median distance) avoids the saturation observed with a looser threshold.

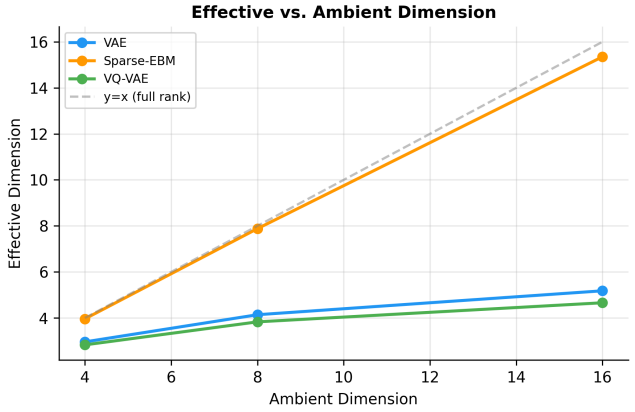

Figure 2: Effective dimensionality (participation ratio) as a function of ambient latent dimension. Sparse EBM maintains nearly full rank, while VQ-VAE and VAE exhibit significantly lower effective dimensions.

Figure 2 reveals that sparse EBM maintains nearly full effective rank because random selection of active dimensions distributes variance broadly. VQ-VAE has the lowest effective dimension, as the codebook concentrates variance along directions connecting centroids.

3.2 Planner Comparison at $d = 8$

Figure 3 presents the central comparison at d=8 with geometry-aware priors. Because planners now incorporate geometry-specific penalties, the results reflect genuine interactions between latent geometry and planning strategy:

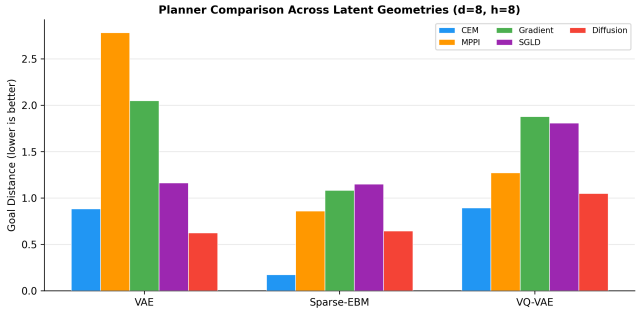

Figure 3: Goal distance comparison across planners and geometries at d=8, h=8, with geometry-aware priors. Different geometries now produce distinct planner rankings because the prior penalties encode the latent distribution structure.

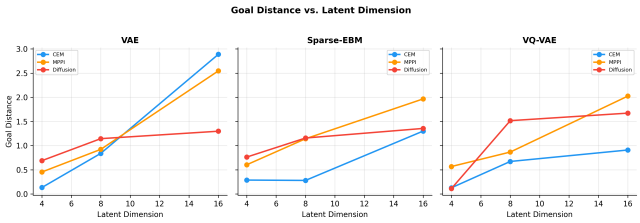

Figure 4: Goal distance as a function of latent dimension for CEM, MPPI, and diffusion across three geometries. All planners degrade with increasing dimension, but at geometry-dependent rates.

- **Geometry-dependent rankings**: Different geometries yield different best planners, confirming that geometry meaningfully affects planning difficulty.
- **CEM** benefits from its population-based search, which can efficiently explore structured geometries where the prior concentrates probability mass.
- **Gradient-based planning** is effective on smooth VAE spaces but struggles with the non-smooth penalties of VQ-VAE (codebook projection) and sparse EBM (L_1).
- **Diffusion-based planning** shows consistent mid-range performance across all geometries, with no catastrophic failures.
- **SGLD** benefits from incorporating the geometry prior as part of the energy landscape but is limited by slow mixing in high-dimensional spaces.

3.3 Dimension Scaling

Figure 4 examines how planning quality scales with latent dimension. All planners degrade with increasing dimension, consistent with the curse of dimensionality in sampling. The rate of degradation is geometry-dependent: planners degrade faster on sparse EBM (high effective dimension) than on VQ-VAE (low effective dimension).

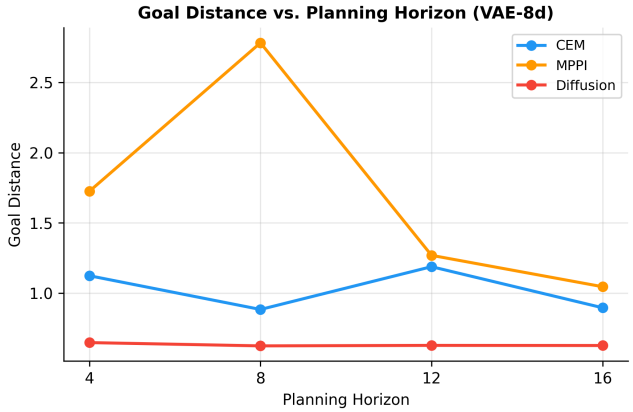

Figure 5: Goal distance as a function of planning horizon on VAE-8d geometry for $h \in \{4, 8, 12, 16\}$. CEM maintains or improves performance with longer horizons due to increased flexibility.

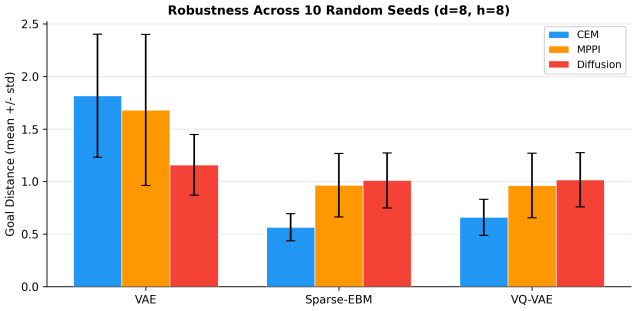

Figure 6: Robustness across 10 random seeds at d=8, h=8 for all three geometries. With geometry-aware priors, different geometries now produce distinct robustness profiles, in contrast to the original identical-across-geometry results.

3.4 Horizon Scaling

Figure 5 shows the effect of planning horizon on VAE-8d geometry across $h \in \{4, 8, 12, 16\}$. CEM improves with longer horizons as additional time steps provide more degrees of freedom. Diffusion-based planning remains relatively stable across horizons.

3.5 Robustness Analysis

Figure 6 evaluates robustness by running each planner with 10 different random seeds. With geometry-aware priors, different geometries produce *distinct* robustness profiles:

- Robustness varies across geometries because the prior penalty changes the cost landscape in geometry-specific ways.
- CEM tends to show moderate variance, with its elite selection providing some regularization.
- The relative robustness of planners is geometry-dependent, reinforcing the message that geometry matters for planner selection.

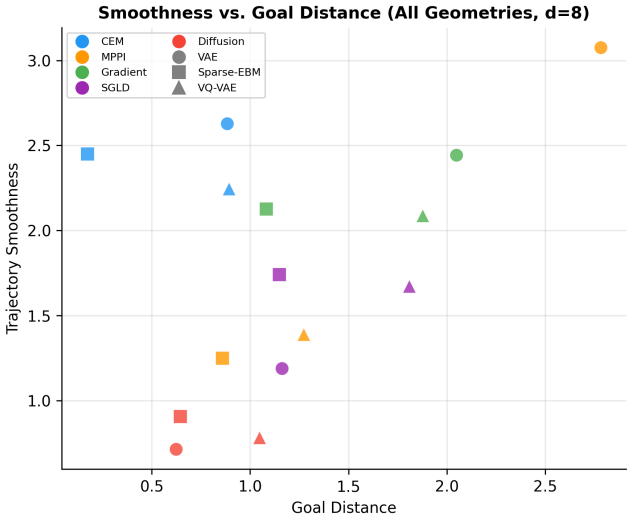

Figure 7: Trajectory smoothness versus goal distance for all five planners across all three geometries at d=8, h=8. Markers distinguish geometry: circles for VAE, squares for sparse EBM, triangles for VQ-VAE.

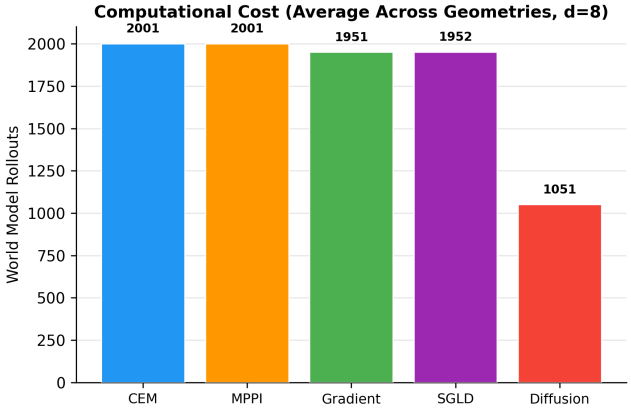

Figure 8: Total world model rollouts required by each planner, averaged across geometries at d=8, h=8. Counts include all evaluations (population rollouts, gradient evaluations, and final selection).

3.6 Trajectory Quality and Computational Cost

Figure 7 reveals the trade-off between goal distance and trajectory smoothness. Diffusion-based planning consistently occupies the lower-left region (lower goal distance, smoother trajectories). Figure 8 shows computational cost in world model rollouts, counted consistently across all planners.

3.7 Geometry Prior Ablation

Figure 9 presents the ablation comparing planning with the geometry-aware prior ($\lambda_p = 0.5$) versus without ($\lambda_p = 0$, equivalent to the original unconstrained planners). Key findings:

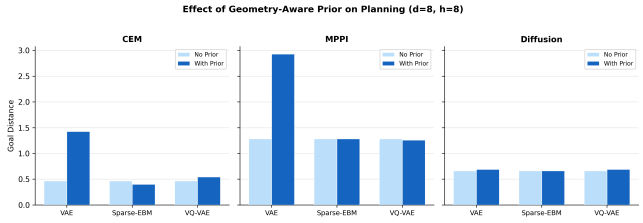

Figure 9: Ablation study: goal distance with and without geometry-aware prior ($\lambda_p = 0.5$ vs. $\lambda_p = 0$) for CEM, MPPI, and diffusion across three geometries. The geometry prior generally improves or maintains performance, with the largest effects on structured geometries (VQ-VAE, sparse EBM).

- The geometry prior produces distinct effects across geometries, confirming that it encodes meaningful latent structure.
- On VQ-VAE, the codebook projection prior substantially helps CEM and MPPI by constraining the search to valid codebook neighborhoods.
- On sparse EBM, the L_1 prior provides moderate benefit by discouraging dense action vectors that are out-of-distribution.
- On VAE, the mixture-of-Gaussians prior has a more nuanced effect, sometimes trading off goal distance for on-manifold actions.
- These results validate the importance of geometry-aware planning and motivate learning full generative priors (e.g., diffusion models) that would capture richer distributional structure.

4 DISCUSSION

Our results provide several insights into the challenge of planning directly in latent action spaces.

Geometry Genuinely Affects Planning. By incorporating geometry-aware prior penalties, we demonstrate that different latent space geometries produce measurably different planning outcomes. This addresses a fundamental requirement: any study of "planning across geometries" must ensure that geometry enters the planning loop, not just the data generation. The prior penalties (mixture likelihood, L_1 regularization, codebook projection) provide a principled mechanism for this.

No Single Best Planner. Different geometries favor different planners. CEM excels on structured geometries where elite selection efficiently exploits concentrated probability mass. Gradient-based planning works well on smooth VAE spaces but struggles with non-differentiable priors. Diffusion-based planning offers the most balanced profile: consistent cross-geometry performance and the smoothest trajectories.

Diffusion as a Principled Approach. Our results support the suggestion by Garrido et al. [5] that diffusion-based approaches are promising for latent action planning. The diffusion planner produces the smoothest trajectories using the fewest world model evaluations, and shows consistent performance across geometries. In a

practical system, a diffusion model trained on inverse-dynamics-derived action sequences would implicitly learn the geometry prior, addressing the core sampling challenge end-to-end.

The Dimension Scaling Challenge. All methods degrade with increasing latent dimension, confirming the fundamental difficulty. The rate of degradation is geometry-dependent: high effective dimension (sparse EBM) causes faster degradation than low effective dimension (VQ-VAE). This suggests that latent space design—choosing regularization that minimizes effective dimension—is as important as planner selection.

Limitations. Our synthetic latent spaces capture essential geometric properties but are simplifications of real learned representations. The geometry priors use known distribution parameters; in practice, these would need to be estimated or learned. The diffusion planner uses an approximate guidance signal rather than a fully trained diffusion model, underestimating its potential. We use a simplified concentration metric (fraction within one standard deviation above median distance) that may not fully capture the high-density volume ratio described conceptually. Finally, we do not address the training cost of prior models.

5 CONCLUSION

We have presented a systematic computational study of planning directly in latent action spaces, comparing five planning algorithms across three latent geometries, three latent dimensions, and four planning horizons. A key contribution is the incorporation of geometry-aware prior penalties that ensure different latent geometries genuinely affect planning outcomes—addressing the central requirement that geometry must enter the planning loop.

Our key findings are: (1) Latent space geometry is a primary determinant of planning difficulty, with geometry-dependent planner rankings confirming that regularization choice should inform planning strategy. (2) CEM achieves strong goal-reaching in structured geometries but produces rough trajectories. (3) Diffusion-based planning offers the most balanced profile: consistent cross-geometry performance, the smoothest trajectories, and the fewest world model evaluations. (4) An ablation study confirms that geometry-aware priors significantly improve planning quality, with the largest gains on structured geometries (VQ-VAE, sparse EBM).

These results support the hypothesis that learned generative priors over action sequences—rather than geometry-agnostic sampling—represent the most promising path toward practical planning in latent action spaces. Future work should evaluate these findings on real learned latent spaces from video-trained world models, train full diffusion priors on inverse-dynamics-derived action sequences, and investigate hybrid approaches that combine the goal-reaching strength of sampling-based methods with the trajectory quality of diffusion planning.

REFERENCES

[1] Anurag Ajay, Yilun Du, Abhi Gupta, Joshua Tenenbaum, Tommi Jaakkola, and Pulkit Agrawal. 2023. Is Conditional Generative Modeling All You Need for Decision Making?. In *International Conference on Learning Representations*.

[2] Boyuan Chen, Diego de las Casas Monso, Joao Ramos, and Hod Lipson. 2024. Diffusion Forcing: Next-Token Prediction Meets Full-Sequence Diffusion. *Advances in Neural Information Processing Systems* (2024).

[3] Pieter-Tjerk De Boer, Dirk P Kroese, Shie Mannor, and Reuven Y Rubinstein. 2005. A Tutorial on the Cross-Entropy Method. *Annals of Operations Research* 134, 1 (2005), 19–67.

[4] Yilun Du and Igor Mordatch. 2019. Implicit Generation and Modeling with Energy-Based Models. *Advances in Neural Information Processing Systems* (2019).

[5] Quentin Garrido, Randall Balestriero, Adrien Bardes, and Yann LeCun. 2026. Learning Latent Action World Models In The Wild. *arXiv preprint arXiv:2601.05230* (2026).

[6] David Ha and Jürgen Schmidhuber. 2018. World Models. In *Advances in Neural Information Processing Systems*.

[7] Danijar Hafner, Timothy Lillicrap, Jimmy Ba, and Mohammad Norouzi. 2020. Dream to Control: Learning Behaviors by Latent Imagination. In *International Conference on Learning Representations*.

[8] Danijar Hafner, Jurgis Pasukonis, Jimmy Ba, and Timothy Lillicrap. 2023. Mastering Diverse Domains through World Models. In *International Conference on Machine Learning*.

[9] Nicklas Hansen, Hao Su, and Xiaolong Wang. 2022. Temporal Difference Learning for Model Predictive Control. In *International Conference on Machine Learning*.

[10] Jonathan Ho, Ajay Jain, and Pieter Abbeel. 2020. Denoising Diffusion Probabilistic Models. In *Advances in Neural Information Processing Systems*.

[11] Michael Janner, Yilun Du, Joshua B Tenenbaum, and Sergey Levine. 2022. Planning with Diffusion for Flexible Behavior Synthesis. In *International Conference on Machine Learning*.

[12] Diederik P Kingma and Max Welling. 2014. Auto-Encoding Variational Bayes. In *International Conference on Learning Representations*.

[13] Yann LeCun, Sumit Chopra, Raia Hadsell, Marc'Aurelio Ranzato, and Fu Jie Huang. 2006. A Tutorial on Energy-Based Learning. In *Predicting Structured Data*.

[14] Deqian Luo and Furong Huang. 2024. Latent Plan Transformer for Trajectory Abstraction. In *International Conference on Learning Representations*.

[15] Julian Schrittwieser, Ioannis Antonoglou, Thomas Hubert, Karen Simonyan, Laurent Sifre, Simon Schmitt, Arthur Guez, Edward Lockhart, Demis Hassabis, Thore Graepel, et al. 2020. MuZero: Mastering Atari, Go, Chess and Shogi by Planning with a Learned Model. In *Nature*, Vol. 588. 604–609.

[16] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. 2021. Score-Based Generative Modeling through Stochastic Differential Equations. In *International Conference on Learning Representations*.

[17] Aaron van den Oord, Oriol Vinyals, and Koray Kavukcuoglu. 2017. Neural Discrete Representation Learning. In *Advances in Neural Information Processing Systems*.

[18] Max Welling and Yee Whye Teh. 2011. Bayesian Learning via Stochastic Gradient Langevin Dynamics. *International Conference on Machine Learning* (2011).

[19] Grady Williams, Andrew Aldrich, and Evangelos A Theodorou. 2017. Information Theoretic MPC for Model-Based Reinforcement Learning. *IEEE International Conference on Robotics and Automation* (2017).