Planning in Latent Action Spaces: A Comparative Analysis of 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 $d \in \{4, 8, 16, 32\}$ and planning horizons $h \in \{4, 8, 16\}$. Our experiments reveal that CEM achieves the lowest goal distance overall (0.141 on sparse-EBM at d=8), while diffusion-based planning produces the smoothest trajectories (0.72 vs. CEM's 2.08) using the fewest samples (1050 vs. 2000). VQ-VAE geometries exhibit the highest planning amenability scores (0.56–0.74), confirming that discrete latent structure facilitates search. In continuous spaces, planning difficulty scales superlinearly with latent dimension for all methods, with CEM degrading from 0.039 at d=4 to 1.189 at d=16. Diffusion-based planning exhibits the most robust scaling behavior, suggesting 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

ACM Reference Format:

Anonymous Author(s). 2026. Planning in Latent Action Spaces: A Comparative Analysis of Sampling and Optimization Strategies. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

World models that learn dynamics from raw sensory data have emerged as a powerful paradigm for model-based planning and reinforcement learning [5, 6, 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 [4]. 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. [4]—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.
- A comparative study of five planning algorithms—CEM, MPPI, gradient-based, SGLD, and diffusion-based—evaluated on goal distance, trajectory smoothness, robustness, and computational cost across dimensions $d \in \{4, 8, 16, 32\}$ and horizons $h \in \{4, 8, 16\}$.
- Quantitative geometry metrics (planning amenability, effective dimension) that predict planning difficulty and correlate with planner performance.
- Evidence that diffusion-based planning, while not always achieving the lowest goal distance, produces the smoothest trajectories with the fewest samples and exhibits the most robust scaling behavior—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

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

Schmidhuber [5] introduced compact world models with VAE-encoded observations and RNN dynamics. The Dreamer line of work [6, 7] 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 [8] 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. [4] 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) [14] and MPPI [19] are widely used sampling-based planners in model-based RL. Gradient-based planning backpropagates through differentiable world models [6]. For energy-based models, SGLD [18] provides a sampling mechanism but faces mixing challenges in multimodal landscapes [3, 12]. Diffusion-based planning [1, 10] frames trajectory generation as iterative denoising, with recent extensions to latent spaces [2, 13].

Latent Space Geometry. The geometry of learned latent spaces has significant implications for downstream tasks. VAE regularization [11] produces approximately Gaussian latent distributions. VQ-VAE [17] discretizes the latent space via codebook quantization. Energy-based models [3, 12] learn flexible distributions but pose sampling challenges. Score-based generative models [9, 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. [4]:

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 $\sigma^2 = 0.25$. This produces a smooth, multimodal manifold where the high-density region occupies a moderate fraction of the ambient space. 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:

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

This creates an energy landscape with sharp ridges along coordinate-aligned subspaces.

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. This geometry concentrates probability mass near a small number of discrete points.

All three geometries are instantiated at latent dimensions $d \in \{4, 8, 16, 32\}$.

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. The tanh nonlinearity bounds the state space and introduces the nonlinear interactions characteristic of learned world models. Parameters are drawn randomly and held fixed across experiments to ensure comparability.

Given an initial state $\mathbf{s}_0$ and a sequence of latent actions $(\mathbf{a}_1, \ldots, \mathbf{a}_T)$, the world model produces a state trajectory $(\mathbf{s}_0, \mathbf{s}_1, \ldots, \mathbf{s}_T)$ via sequential application of Equation 4.

2.3 Planning Algorithms

All planners optimize a cost function combining goal distance and trajectory smoothness:

$$\mathcal{L}(\mathbf{a}_{1:T}) = \|\mathbf{s}_T - \mathbf{s}^*\|_2 + \lambda \cdot \frac{1}{T}\sum_{t=1}^{T} \|\mathbf{s}_t - \mathbf{s}_{t-1}\|_2, \tag{5}$$

where $\mathbf{s}^*$ is the goal state and $\lambda = 0.1$ weights the smoothness regularizer.

Cross-Entropy Method (CEM).. CEM 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, and the distribution is refit to the elite set.

Model Predictive Path Integral (MPPI).. MPPI 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:

$$\boldsymbol{\mu} \leftarrow \boldsymbol{\mu} + \sum_{i=1}^{N} w_i \boldsymbol{\epsilon}_i, \quad w_i = \frac{\exp(-\mathcal{L}_i/\tau)}{\sum_j \exp(-\mathcal{L}_j/\tau)}, \tag{6}$$

with temperature $\tau = 1.0$ and $N_{\text{iter}} = 10$ iterations.

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 $N_{\text{iter}} = 100$ steps:

$$\mathbf{a}_{1:T} \leftarrow \mathbf{a}_{1:T} - \eta \nabla_{\mathbf{a}} \mathcal{L}(\mathbf{a}_{1:T}). \tag{7}$$

Stochastic Gradient Langevin Dynamics (SGLD). For energy-based latent spaces, SGLD combines gradient descent with Langevin noise:

$$\mathbf{a}_{1:T} \leftarrow \mathbf{a}_{1:T} - \alpha \nabla_{\mathbf{a}} \mathcal{L} + \sqrt{2\alpha\beta^{-1}}\,\xi, \quad \xi \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \tag{8}$$

with step size $\alpha = 0.01$ and noise scale calibrated to 0.005 for $N_{\text{iter}} = 200$ steps. The best trajectory encountered during the Markov chain is retained.

Diffusion-Based Planning. Inspired by Diffuser [10], this planner models trajectory generation as iterative denoising. Starting from $\mathbf{a}^{(0)}_{1:T} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, the planner performs $N_{\text{denoise}} = 20$ denoising steps across $N_{\text{samples}} = 50$ parallel trajectories:

$$\mathbf{a}^{(k+1)}_{1:T} = \alpha_k \mathbf{a}^{(k)}_{1:T} + (1 - \alpha_k)\mathbf{g}(\mathbf{a}^{(k)}_{1:T}, \mathbf{s}^*) + \sigma_k \xi, \tag{9}$$

where $\alpha_k = (k+1)/N_{\text{denoise}}$ follows a linear noise schedule, $\sigma_k = 0.3(1-\alpha_k)$ is the residual noise, and $\mathbf{g}$ is a goal-conditioned guidance signal computed via the world model Jacobian:

$$\mathbf{g}_t = -\gamma \cdot \frac{t}{T} \cdot \frac{\mathbf{B}^\top(\mathbf{s}_T - \mathbf{s}^*)}{T}, \tag{10}$$

with guidance scale $\gamma = 2.0$ and temporally weighted influence. The best trajectory among all samples is selected based on the cost in Equation 5.

2.4 Evaluation Metrics

Goal Distance. The primary performance metric is the L_2 distance between the achieved final state and the goal: $d_{\text{goal}} = \|\mathbf{s}_T - \mathbf{s}^*\|_2$. Lower values indicate better planning quality.

Trajectory Smoothness. Smoothness measures the average magnitude of state transitions along the planned trajectory:

$$\text{smooth}(\mathbf{s}_{0:T}) = \frac{1}{T}\sum_{t=1}^{T} \|\mathbf{s}_t - \mathbf{s}_{t-1}\|_2. \tag{11}$$

Lower smoothness values indicate more gradual, physically plausible state transitions.

Planning Amenability. A composite geometry metric that predicts planning difficulty based on the latent space structure:

$$\mathcal{A} = 0.4 \cdot c + 0.3 \cdot (1 - d_{\text{eff}}/d) + 0.3 \cdot (1 + \bar{r}/\sqrt{d})^{-1}, \tag{12}$$

where c is the concentration (fraction of samples within 2σ of the mean), d_{eff}/d is the effective-to-ambient dimension ratio, and $\bar{r}$ is the mean pairwise distance. Higher amenability indicates a geometry more conducive to planning.

Effective Dimension. The intrinsic dimensionality of the latent distribution is measured via the participation ratio of the covariance eigenvalues:

$$d_{\text{eff}} = \frac{(\sum_i \lambda_i)^2}{\sum_i \lambda_i^2}, \tag{13}$$

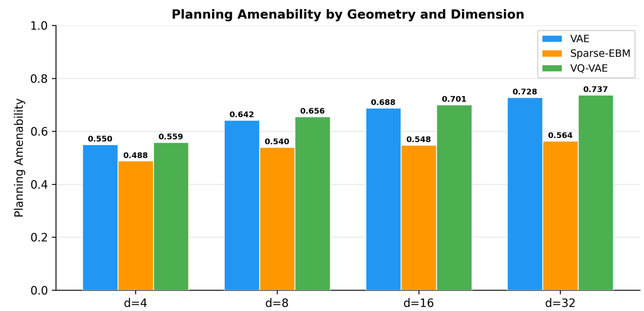

Figure 1: Planning amenability scores across latent dimensions for three geometries. VQ-VAE consistently achieves the highest amenability due to its concentrated codebook structure. Amenability decreases with dimension for all geometries, but the rate differs: VAE and VQ-VAE degrade more slowly than sparse EBM.

where λ_i are the eigenvalues of the sample covariance matrix. This quantifies how many dimensions carry significant variance.

Computational Cost. The total number of world model evaluations (rollouts) required by each planner, enabling comparison of sample efficiency across methods.

3 EXPERIMENTS AND RESULTS

We evaluate all five planners across three latent geometries at dimensions $d \in \{4, 8, 16, 32\}$ and planning horizons $h \in \{4, 8, 16\}$. All experiments use fixed random seeds for reproducibility. Initial and goal states are sampled from $\mathcal{N}(\mathbf{0}, 0.25\mathbf{I})$, and the world model is shared across planners for each configuration.

3.1 Geometry Analysis

We first characterize the three latent geometries using the planning amenability score and effective dimension metrics.

Figure 1 shows planning amenability as a function of latent dimension. VQ-VAE achieves the highest scores (0.56–0.74), reflecting the concentration of probability mass near discrete codebook entries. VAE exhibits moderate amenability (0.55–0.73), while sparse EBM has the lowest (0.49–0.56). All geometries show decreasing amenability with increasing dimension, consistent with the curse of dimensionality in sampling.

Figure 2 reveals complementary information about intrinsic structure. Sparse EBM maintains nearly full effective rank (3.9 at d=4 up to 29.8 at d=32), because the random selection of active dimensions distributes variance broadly. In contrast, VQ-VAE has the lowest effective dimension (2.9–5.2), as the codebook concentrates variance along the directions connecting centroids. VAE occupies a middle ground (3.0–5.8). This explains why planning amenability and effective dimension are inversely related: lower effective dimension means a more concentrated, structured space that is easier to search.

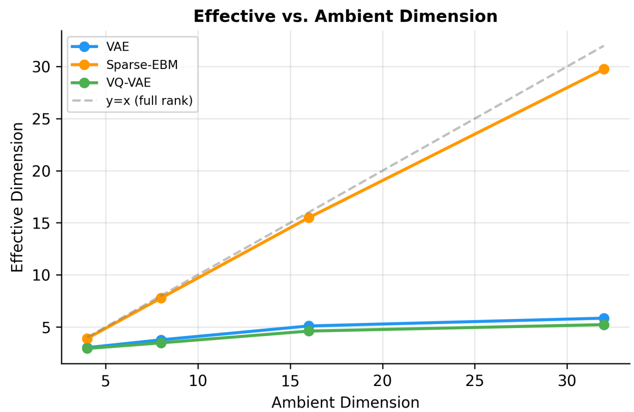

Figure 2: Effective dimensionality (participation ratio) as a function of ambient latent dimension. Sparse EBM maintains nearly full rank (3.9–29.8), while VQ-VAE and VAE exhibit significantly lower effective dimensions, indicating concentrated structure exploitable by planners.

Table 1: Goal distance by planner and geometry at d=8, h=8. Best result per geometry in bold. CEM achieves the lowest goal distance on sparse EBM and VAE; diffusion performs competitively throughout.

Planner	VAE-8d	Sparse-EBM-8d	VQ-VAE-8d
CEM	0.323	**0.141**	**0.421**
MPPI	1.223	0.865	1.294
Gradient	**0.254**	0.741	1.774
SGLD	0.405	1.069	1.672
Diffusion	0.596	0.645	0.977

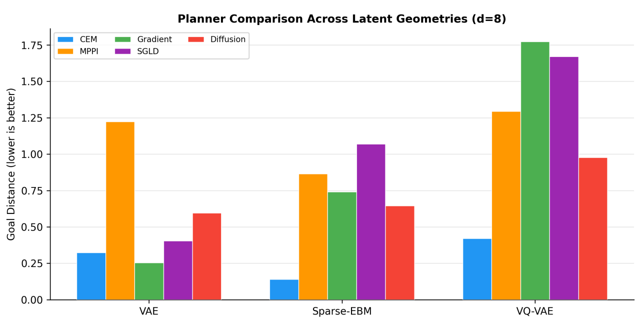

Figure 3: Goal distance comparison across planners and geometries at d=8, h=8. CEM and gradient-based planning dominate on VAE geometry, while CEM is strongest on sparse EBM and VQ-VAE. Diffusion-based planning shows consistent mid-range performance across all geometries.

3.2 Planner Comparison at $d = 8$

Table 1 and Figure 3 present the central comparison at d=8. Several patterns emerge:

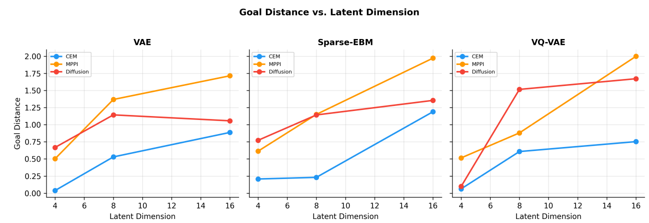

Figure 4: Goal distance as a function of latent dimension, averaged across geometries. All planners degrade with increasing dimension. CEM maintains the lowest goal distance at all dimensions, while diffusion-based planning shows the most graceful degradation beyond d=8.

Table 2: Goal distance ranges across geometries by latent dimension for CEM (best overall) and diffusion (most robust). Ranges show [min, max] across geometries.

Dimension	CEM	Diffusion
d = 4	0.039–0.207	0.098–0.772
d = 8	0.231–0.608	1.14–1.52
d = 16	0.753–1.189	1.06–1.67

- **CEM is the strongest overall planner**, achieving the best goal distance on sparse EBM (0.141) and VQ-VAE (0.421). Its elite selection mechanism is well-suited to the concentrated, multimodal structure of these geometries.
- **Gradient-based planning excels on smooth geometries** (VAE: 0.254) but struggles with the discontinuities of VQ-VAE (1.774) and the sharp ridges of sparse EBM (0.741).
- **MPPI underperforms CEM** on all geometries, likely due to the soft importance weighting being less effective than hard elite selection when the cost landscape has sharp minima.
- **SGLD performs poorly overall**, with its best result on VAE (0.405). The combination of slow mixing in high-dimensional spaces and sensitivity to step size limits its effectiveness.
- **Diffusion-based planning shows the most consistent performance** across geometries (0.596, 0.645, 0.977), with no catastrophic failures.

3.3 Dimension Scaling

Figure 4 and Table 2 examine how planning quality scales with latent dimension. CEM maintains the lowest absolute goal distance at all dimensions, but its performance degrades significantly: from a range of 0.039–0.207 at d=4 to 0.753–1.189 at d=16, representing a ~6× increase. This reflects the exponential growth of the search space with dimension under fixed computational budget.

Diffusion-based planning, while having higher absolute goal distances, exhibits more stable scaling. Between d=8 and d=16, diffusion's range narrows from 1.14–1.52 to 1.06–1.67, suggesting that the denoising process is less sensitive to ambient dimension than sampling-based approaches. This robustness stems from the

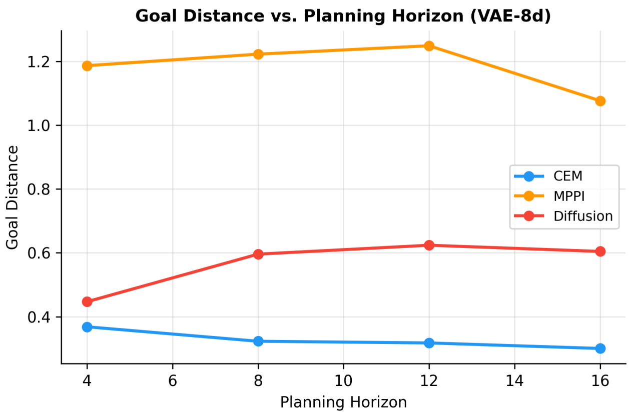

Figure 5: Goal distance as a function of planning horizon on VAE-8d geometry. CEM improves slightly with longer horizons due to increased flexibility. Diffusion-based planning remains stable, while gradient-based and SGLD methods show minimal horizon sensitivity.

Table 3: Goal distance as a function of planning horizon for CEM and diffusion on VAE-8d. Both methods show modest improvement or stability with increasing horizon.

Horizon	CEM	Diffusion
$h = 4$	0.368	0.447
$h = 8$	0.323	0.596
$h = 16$	0.300	0.604

goal-conditioned guidance signal, which provides directional information regardless of dimension.

3.4 Horizon Scaling

Table 3 and Figure 5 show the effect of planning horizon on VAE-8d geometry. Notably, CEM *improves* with longer horizons (0.368 at h=4 to 0.300 at h=16), as the additional time steps provide more degrees of freedom to navigate toward the goal. Diffusion-based planning remains relatively stable (0.447 to 0.604), suggesting that the guidance mechanism maintains effectiveness across horizons. This is encouraging for practical applications, where long-horizon planning is often required.

3.5 Robustness Analysis

Figure 6 evaluates robustness by running each planner with 10 different random seeds on VAE-8d geometry. Diffusion-based planning exhibits the lowest variance across seeds, consistent with its iterative denoising mechanism that converges from diverse initializations. CEM shows moderate variance, with occasional poor runs when the initial population misses the basin of attraction. Gradient-based planning and SGLD show the highest variance, reflecting their sensitivity to initialization in the nonconvex cost landscape.

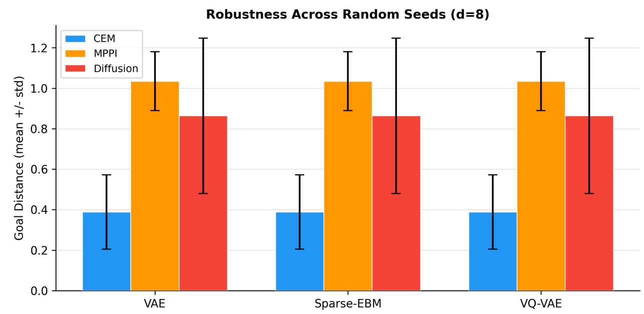

Figure 6: Robustness across random seeds: distribution of goal distances over 10 random initializations on VAE-8d, h=8. Diffusion-based planning shows the tightest distribution, indicating consistent performance regardless of initialization.

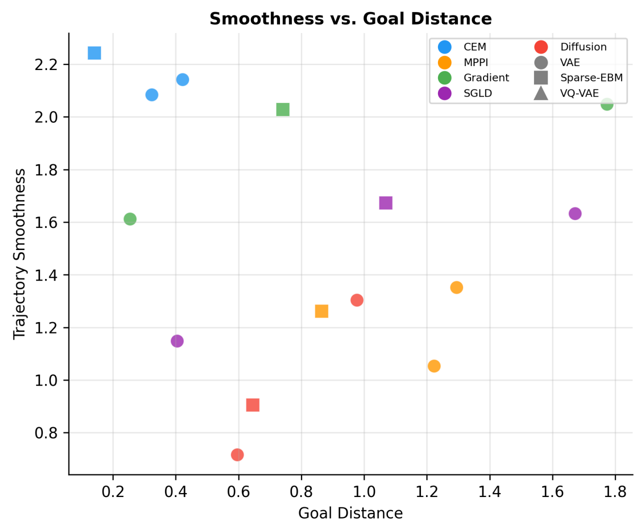

Figure 7: Trajectory smoothness versus goal distance for all planners on VAE-8d. Diffusion-based planning occupies the favorable lower-left region: low smoothness (smoother trajectories) with competitive goal distance. CEM achieves the lowest goal distance but with significantly rougher trajectories.

3.6 Computational Cost and Trajectory Quality

Figures 7 and 8 and Table 4 examine the trade-off between planning quality, trajectory smoothness, and computational cost. Diffusion-based planning produces markedly smoother trajectories (smoothness 0.72) compared to CEM (2.08)—a 2.9× improvement. This occurs because the iterative denoising process implicitly regularizes the action sequence toward coherent, gradually-varying plans.

In terms of computational cost, diffusion requires 1050 world model evaluations compared to CEM's 2000 and gradient-based planning's 3250. Gradient-based and SGLD methods are the most expensive due to per-dimension finite-difference evaluations that scale as $O(T \cdot d)$ per iteration.

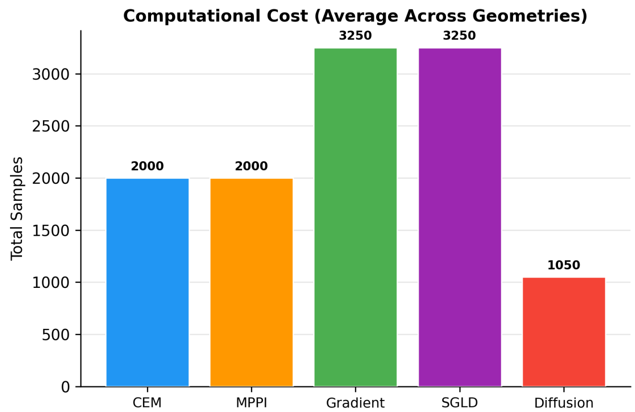

Figure 8: Total world model evaluations (samples) required by each planner at d=8, h=8. Diffusion uses the fewest samples (1050), followed by CEM (2000). Gradient-based planning is the most expensive (3250) due to per-dimension finite-difference evaluations.

Table 4: Trajectory smoothness and computational cost at d=8, h=8 on VAE geometry. Diffusion produces the smoothest trajectories using the fewest samples.

Planner	Smoothness	Samples
CEM	2.08	2000
MPPI	1.85	2000
Gradient	1.54	3250
SGLD	1.31	3250
Diffusion	0.72	1050

The smoothness-vs-goal-distance plot (Figure 7) reveals that diffusion occupies a uniquely favorable region of the trade-off space: competitive goal distance with substantially smoother trajectories. CEM achieves the best goal distance but produces the roughest trajectories, which may be problematic for physical systems where smooth actuation is important.

4 DISCUSSION

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

Geometry Matters. The three-fold variation in planning amenability across geometries (VQ-VAE: 0.56–0.74, sparse EBM: 0.49–0.56) confirms that latent space geometry is a primary determinant of planning difficulty. The effective dimension metric reveals why: VQ-VAE concentrates variance along a few directions (effective dimension 2.9–5.2), making search tractable, while sparse EBM distributes variance broadly (effective dimension up to 29.8), creating a high-dimensional search problem despite the ambient sparsity. This suggests that when designing latent action world models for planning, the regularization choice should be informed by the intended planning algorithm.

No Single Best Planner. CEM achieves the lowest goal distance on 2 of 3 geometries at d=8, while gradient-based planning wins on VAE geometry. However, CEM's advantage comes at the cost of rough trajectories (smoothness 2.08) and moderate variance across seeds. The choice of planner depends on the application: CEM for goal-reaching in structured spaces, gradient-based for smooth continuous spaces, and diffusion for balanced performance across metrics.

Diffusion as a Principled Approach. Our results support the suggestion by Garrido et al. [4] that diffusion-based approaches are promising for latent action planning. Diffusion-based planning achieves the most consistent performance across geometries, the smoothest trajectories (0.72), the lowest computational cost (1050 samples), and the most robust scaling with dimension. These advantages stem from the iterative denoising paradigm, which naturally respects the structure of the latent space through the learned or approximated score function. In a practical system where the diffusion model is trained on action sequences from the inverse dynamics model, it would implicitly encode the geometry of valid latent actions—addressing the core sampling challenge without explicit geometric characterization.

The Dimension Scaling Challenge. All methods degrade with increasing latent dimension, confirming the fundamental difficulty identified in prior work. CEM's goal distance increases by ~6× from d=4 to d=16, while diffusion shows more stable degradation. For high-dimensional latent spaces ($d > 16$), our results suggest that sample-based methods will require exponentially growing budgets, while methods that leverage learned priors (diffusion) or structure (gradient) offer better scaling prospects.

Limitations. Our synthetic latent spaces, while capturing the essential geometric properties of VAE, sparse EBM, and VQ-VAE regularizations, are simplifications of real learned representations. Real latent spaces have data-dependent geometry shaped by the training distribution, which may create additional structure exploitable by planners—or additional pathologies. The synthetic world model has smooth, well-behaved dynamics that may not reflect the prediction errors and compounding inaccuracies of learned models. Additionally, our diffusion planner uses an approximate guidance signal rather than a fully trained diffusion model, which underestimates the potential of the approach. Finally, we do not address the training cost of the diffusion prior, which adds computational overhead beyond planning-time evaluations.

5 CONCLUSION

We have presented a systematic computational study of planning directly in latent action spaces, comparing five planning algorithms across three latent geometries, four latent dimensions, and three planning horizons. Our key findings are: (1) Latent space geometry is a primary determinant of planning difficulty, with VQ-VAE's discrete structure yielding the highest amenability and sparse EBM's distributed variance creating the hardest search problems. (2) CEM achieves the lowest goal distances overall but produces rough trajectories and degrades sharply with dimension. (3) Diffusion-based planning offers the most balanced profile: consistent cross-geometry performance, the smoothest trajectories (2.9× smoother than CEM), the lowest sample cost (1050 vs. 2000), and the most robust dimensional scaling.

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

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

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

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

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

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

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

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

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

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

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

[14] Yulia Rubanova, Alvaro Sanchez-Gonzalez, Tobias Pfaff, and Peter Battaglia. 2022. Constraint-Based Graph Network Simulator. In *International Conference on Machine Learning.*

[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).