Model Based Diffusion for Trajectory Optimization – NeurIPS-2024

Chaoyi Pan, Zeji Yi, Guanya Shi, Guannan Qu

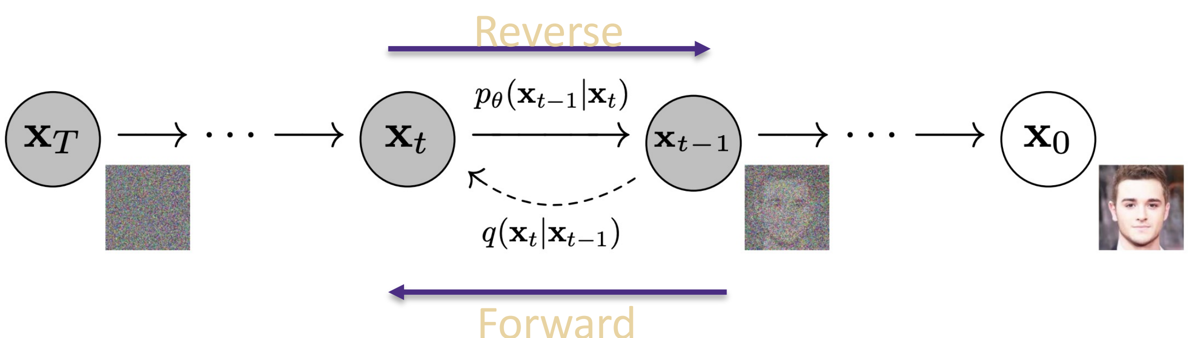
UNIVERSITY *of* WASHINGTON

Preliminaries

- Trajectory Optimization:

$$
\begin{aligned}
\min_{x_{1:T}, u_{1:T}} \; & J(x_{1:T}; u_{1:T}) = l_T(x_T) + \sum_{t=0}^{T-1} l_t(x_t, u_t) \\
\text{s.t.} \quad & x_0 = x_{\text{init}} \\
& x_{t+1} = f_t(x_t, u_t), \quad \forall t = 0, 1, \ldots, T-1, \\
& g_t(x_t, u_t) \le 0, \quad \forall t = 0, 1, \ldots, T-1.3
\end{aligned}
$$

- Diffusion Models:



$$L_{\text{simple}}(\theta) := \mathbb{E}_{t, \mathbf{x}_0, \boldsymbol{\epsilon}}\left[\left\| \boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}, t)\right\|^2\right]$$

PROBLEM

Trajectory Optimization Challenges:

- **High Dimensional State and Control action space:** Example: Humanoid and dexterous hand gripper.
- **Non-linear discontinuous dynamics and Non-convex Objective / Constraints.**

Challenges with Diffusion Models:

- **Model Free:** This nature limits generalization across morphologies, Not a problem with vision and language tasks.
- Needs high quality / dynamically feasible data to learn high dimensional and complex distribution through iterative sampling.

How could diffusion models be directly used as trajectory optimization solver?

APPROACH

- Optimization Objective:

$$\begin{aligned}\min_{x_{1:T}, u_{1:T}} \quad & J(x_{1:T}; u_{1:T}) = l_T(x_T) + \sum_{t=0}^{T-1} l_t(x_t, u_t) \\ \text{s.t.} \quad & x_0 = x_{\text{init}} \\ & x_{t+1} = f_t(x_t, u_t), \quad \forall t = 0, 1, \ldots, T-1, \\ & g_t(x_t, u_t) \leq 0, \quad \forall t = 0, 1, \ldots, T-1.3\end{aligned}$$

- MFD requires data to learn high dimensional distribution (HDD) but in TO we already know optimum distribution or HDD given by:

$$\boxed{p_0(Y) \propto p_d(Y) p_J(Y) p_g(Y)}$$

$$p_d(Y) \propto \prod_{t=1}^{T} \mathbf{1}(x_t = f_{t-1}(x_{t-1}, u_{t-1}))$$

$$p_g(Y) \propto \prod_{t=1}^{T} \mathbf{1}(g_t(x_t, u_t) \leq 0)$$

$$p_J(Y) \propto \exp\left(-\frac{J(Y)}{\lambda}\right)$$

Sampling will lead to Optimum Trajectory (Y^*) as $\lambda \to 0$ during convergence.

But, Difficult to sample from.

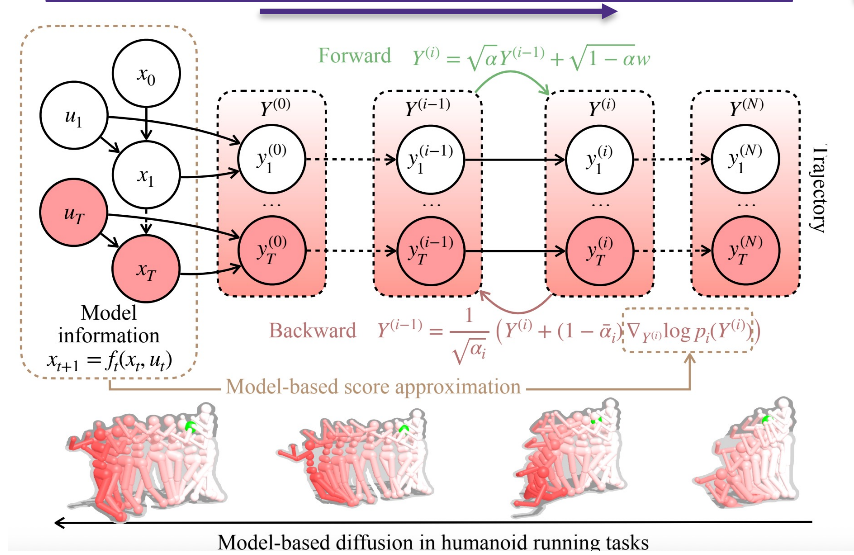
APPROACH: Model-Based Diffusion

Forward Trajectory corruption to an isotropic distribution for ease in sampling

$$p_{i|i-1}(\cdot|Y^{(i-1)}) \sim \mathcal{N}(\sqrt{\alpha_i}Y^{(i-1)}, (1-\alpha_i)I)$$

$$p_{i|0}(\cdot|Y^{(0)}) \sim \mathcal{N}(\sqrt{\bar{\alpha}_i}Y^{(0)}, (1-\bar{\alpha}_i)I), \quad \bar{\alpha}_i = \prod_{k=1}^{i} \alpha_k.$$

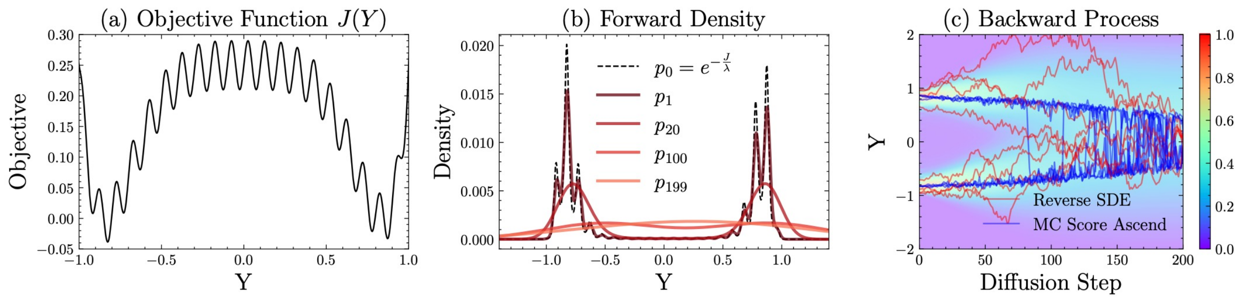
Has closed form solution due to independent noise at each time step



$$p_{i-1}(Y^{(i-1)}) = \int p_{i-1|i}(Y^{(i-1)}|Y^{(i)}) p_i(Y^{(i)})\, dY^{(i)},$$

$$p_0(Y^{(0)}) = \int p_N(Y^{(N)}) \prod_{i=N}^{1} p_{i-1|i}(Y^{(i-1)}|Y^{(i)}) dY^{(1:N)}$$

APPROACH: Model-Based Diffusion



$$Y^{(i-1)} = \frac{1}{\sqrt{\alpha_i}}\left(Y^{(i)} + (1-\bar{\alpha}_i)\nabla_{Y^{(i)}} \log p_i(Y^{(i)})\right)$$

$$\nabla_{Y^{(i)}} \log p_i(Y^{(i)}) = \frac{\nabla_{Y^{(i)}} \int p_{i|0}(Y^{(i)} \mid Y^{(0)}) p_0(Y^{(0)})\, dY^{(0)}}{\int p_{i|0}(Y^{(i)} \mid Y^{(0)}) p_0(Y^{(0)})\, dY^{(0)}} = \frac{\int \nabla_{Y^{(i)}} p_{i|0}(Y^{(i)} \mid Y^{(0)}) p_0(Y^{(0)})\, dY^{(0)}}{\int p_{i|0}(Y^{(i)} \mid Y^{(0)}) p_0(Y^{(0)})\, dY^{(0)}}$$

$$\boxed{\approx -\frac{Y^{(i)}}{1-\bar{\alpha}_i} + \frac{\sqrt{\bar{\alpha}_i}}{1-\bar{\alpha}_i} \underbrace{\frac{\sum_{Y^{(0)} \in \mathcal{Y}^{(i)}} Y^{(0)} p_0(Y^{(0)})}{\sum_{Y^{(0)} \in \mathcal{Y}^{(i)}} p_0(Y^{(0)})}}_{\text{Monte Carlo Approximation}} := -\frac{Y^{(i)}}{1-\bar{\alpha}_i} + \frac{\sqrt{\bar{\alpha}_i}}{1-\bar{\alpha}_i} \bar{Y}^{(0)}(\mathcal{Y}^{(i)})}$$

$$\phi_i(Y^{(0)}) \propto p_{i|0}(Y^{(i)} \mid Y^{(0)}) \propto \mathcal{N}\left(\frac{Y^{(i)}}{\sqrt{\bar{\alpha}_i}}, \frac{I}{\bar{\alpha}_i} - I\right)$$

APPROACH: MBD Algorithm

$$p_0(Y) \propto \exp\left(-\frac{J(Y)}{\lambda}\right)$$

Algorithm 1 Model-based Diffusion for Generic Optimization

1: **Input:** $Y^{(N)} \sim \mathcal{N}(\mathbf{0}, I)$
2: **for** $i = N$ to 1 **do**
3: Sample $\mathcal{Y}^{(i)} \sim \mathcal{N}(\frac{Y^{(i)}}{\sqrt{\bar{\alpha}_{i-1}}}, (\frac{1}{\bar{\alpha}_{i-1}} - 1)I)$
4: Calculate Eq. (9b) $\bar{Y}^{(0)}(\mathcal{Y}^{(i)}) = \frac{\sum_{Y^{(0)} \in \mathcal{Y}^{(i)}} Y^{(0)} p_0(Y^{(0)})}{\sum_{Y^{(0)} \in \mathcal{Y}^{(i)}} p_0(Y^{(0)})}$
5: Estimate the score Eq. (9a): $\nabla_{Y^{(i)}} \log p_i(Y^{(i)}) \approx -\frac{Y^{(i)}}{1-\bar{\alpha}_i} + \frac{\sqrt{\bar{\alpha}_i}}{1-\bar{\alpha}_i} \bar{Y}^{(0)}(\mathcal{Y}^{(i)})$
6: Monte Carlo score ascent Eq. (6): $Y^{(i-1)} \leftarrow \frac{1}{\sqrt{\alpha_i}} \left(Y^{(i)} + (1 - \bar{\alpha}_i) \nabla_{Y^{(i)}} \log p_i(Y^{(i)})\right)$
7: **end for**

Model Based Diffusion for Trajectory Optimization: $p_0(Y) \propto p_d(Y) p_J(Y) p_g(Y)$

Algorithm 2 Model-based Diffusion for Trajectory Optimization

1: **Input:** $Y^{(N)} \sim \mathcal{N}(\mathbf{0}, I)$
2: **for** $i = N$ to 1 **do**
3: Sample $\mathcal{Y}^{(i)} \sim \mathcal{N}(\frac{Y^{(i)}}{\sqrt{\bar{\alpha}_{i-1}}}, (\frac{1}{\bar{\alpha}_{i-1}} - 1)I)$
4: Get dynamically feasible samples: $\mathcal{Y}_d^{(i)} \leftarrow \text{rollout}(\mathcal{Y}^{(i)})$
5: Calculate $\bar{Y}^{(0)}$ with Eq. (10d) (model only) or Eq. (13) (model + demonstration)
6: Estimate the score Eq. (10c): $\nabla_{Y^{(i)}} \log p_i(Y^{(i)}) \approx -\frac{Y^{(i)}}{1-\bar{\alpha}_i} + \frac{\sqrt{\bar{\alpha}_i}}{1-\bar{\alpha}_i} \bar{Y}^{(0)}$
7: Monte Carlo score ascent Eq. (6): $Y^{(i-1)} \leftarrow \frac{1}{\sqrt{\alpha_i}} \left(Y^{(i)} + (1 - \bar{\alpha}_i) \nabla_{Y^{(i)}} \log p_i(Y^{(i)})\right)$
8: **end for**

$$\bar{Y}^{(0)} = \frac{\sum_{Y^{(0)} \in \mathcal{Y}_d^{(i)}} Y^{(0)} w(Y^{(0)})}{\sum_{Y^{(0)} \in \mathcal{Y}_d^{(i)}} w(Y^{(0)})}, \quad w(Y^{(0)}) = p_J(Y^{(0)}) p_g(Y^{(0)})$$

APPROACH: MBD Algorithm

Model Based Diffusion with Demonstration:

Assumption: Imperfect or partial-state demonstration are noisy samples from desired trajectory. $p(Y_{\text{demo}} \mid Y^{(0)}) \sim \mathcal{N}(Y^{(0)}, \sigma^2 I)$.

Given suboptimal demonstrations sampling from posterior $p(Y^{(0)} \mid Y_{\text{demo}})$ could lead to poor solutions.

$$p(Y^{(0)} \mid Y_{\text{demo}}) \propto \underline{p_0(Y^{(0)})}\,\underline{p(Y_{\text{demo}} \mid Y^{(0)})}$$

Because $p(Y^{(0)} \mid Y_{\text{demo}})$ could dominate model-based distribution.

$$p_0'(Y^{(0)}) \propto (1-\eta)p_d(Y^{(0)})p_J(Y^{(0)})p_g(Y^{(0)}) + \eta p_{\text{demo}}(Y^{(0)})p_J(Y_{\text{demo}})p_g(Y_{\text{demo}})$$

$$\eta = \begin{cases} 1 & p_d(Y^{(0)})p_J(Y^{(0)})p_g(Y^{(0)}) < p_{\text{demo}}(Y^{(0)})p_J(Y_{\text{demo}})p_g(Y_{\text{demo}}) \\ 0 & p_d(Y^{(0)})p_J(Y^{(0)})p_g(Y^{(0)}) \geq p_{\text{demo}}(Y^{(0)})p_J(Y_{\text{demo}})p_g(Y_{\text{demo}}). \end{cases}$$

$$\bar{Y}^{(0)} = \frac{\sum_{Y^{(0)} \in \mathcal{Y}_d^{(i)}} Y^{(0)} w(Y^{(0)})}{\sum_{Y^{(0)} \in \mathcal{Y}_d^{(i)}} w(Y^{(0)})}, \quad w(Y^{(0)}) = \max\left\{ \begin{array}{c} p_d(Y^{(0)})p_J(Y^{(0)})p_g(Y^{(0)}), \\ p_{\text{demo}}(Y^{(0)})p_J(Y_{\text{demo}})p_g(Y_{\text{demo}}) \end{array} \right\}$$

EXPERIMENTATIONS:

- Model Based Diffusion in contact rich task without data augmentation:

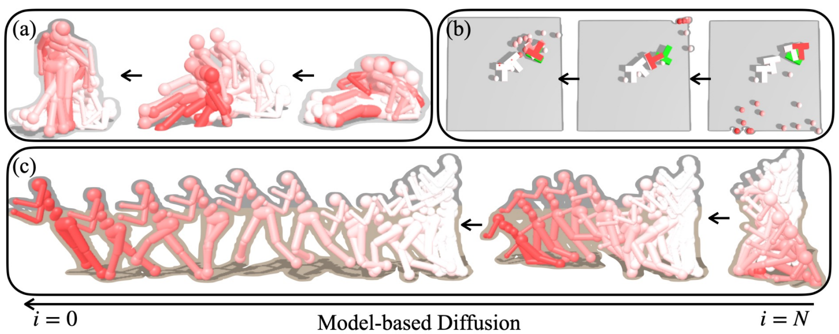


Figure 3: Optimization process of MBD on the (a) Humanoid Standup, (b) Push T, and (c) Humanoid Running tasks. The trajectory is iteratively refined to achieve the desired objective in the high dimensional space with model information.

Task	CMA-ES	CEM	MPPI	RL*	MBD
Hopper	1.12 ± 0.10	0.65 ± 0.12	0.91 ± 0.15	1.40 ± 0.04	$\mathbf{1.53 \pm 0.03}$
Half Cheetah	0.44 ± 0.10	0.22 ± 0.15	0.20 ± 0.14	1.59 ± 0.05	$\mathbf{2.31 \pm 0.19}$
Ant	1.18 ± 0.52	0.85 ± 0.17	0.33 ± 0.45	3.26 ± 1.61	$\mathbf{3.80 \pm 0.35}$
Walker2D	0.83 ± 0.04	1.06 ± 0.04	0.90 ± 0.05	1.09 ± 0.28	$\mathbf{2.63 \pm 0.23}$
Humanoid Standup	0.58 ± 0.01	0.47 ± 0.01	0.53 ± 0.05	0.83 ± 0.02	$\mathbf{0.99 \pm 0.07}$
Humanoid Running	0.60 ± 0.11	0.41 ± 0.16	0.59 ± 0.14	1.80 ± 0.03	$\mathbf{2.92 \pm 0.26}$
Push T	0.39 ± 0.07	0.25 ± 0.09	-0.13 ± 0.09	-0.63 ± 0.16	$\mathbf{0.67 \pm 0.10}$

Table 2: Reward of different methods on non-continuous tasks. *RL requires offline training and generate a closed-loop policy so it is not an apple-to-apple baseline.

Task	CMA-ES	CEM	MPPI	RL	MBD
Hopper	29.3 s	26.5 s	26.4 s	17 m 45.63 s	26.5 s
Half Cheetah	29.5 s	26.4 s	26.7 s	4 m 18.8 s	26.8 s
Ant	18.4 s	16.1 s	16.0 s	2 m 46.8 s	16.2 s
Walker2D	37.5 s	34.5 s	34.7 s	5 m 1.5 s	34.6 s
Humanoid Standup	20.8 s	17.6 s	17.7 s	4 m 29 s	17.7 s
Humanoid Running	30.8 s	29.7 s	29.6 s	3 m 34.7 s	30.0 s
Push T	10 m 40.0 s	10 m 32.0 s	10 m 32.3 s	67 m 25.6 s	10 m 32.8 s

Table 3: Computational time of different methods on non-continuous tasks.

EXPERIMENTATIONS:

- Model Based Diffusion with data augmentation:

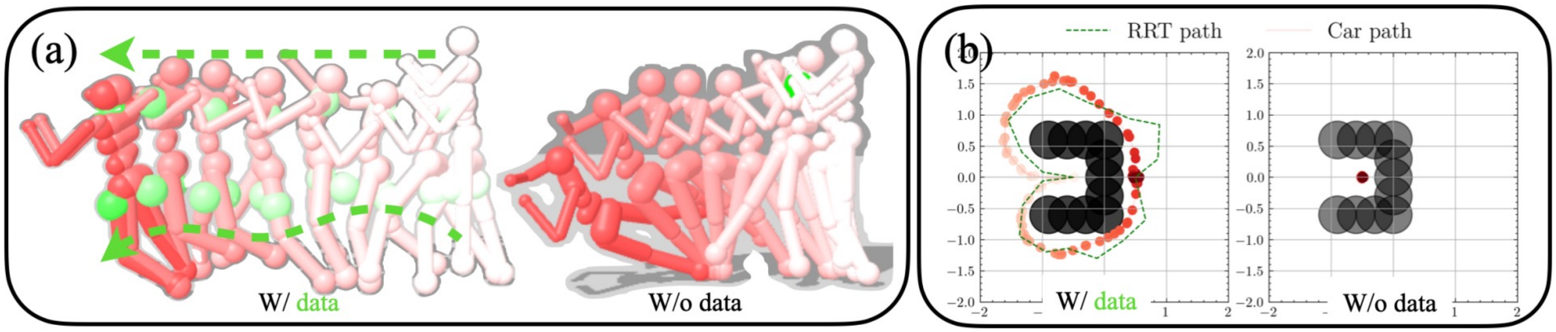


Figure 4: MBD optimized trajectory with data augmentation on the (a) Humanoid Jogging and (b) Car UMaze Navigation tasks. With data augmentation, the trajectory is regularized and refined to achieve the desired objective.

EXPERIMENTATION:

- Black Box Optimization with MBD:

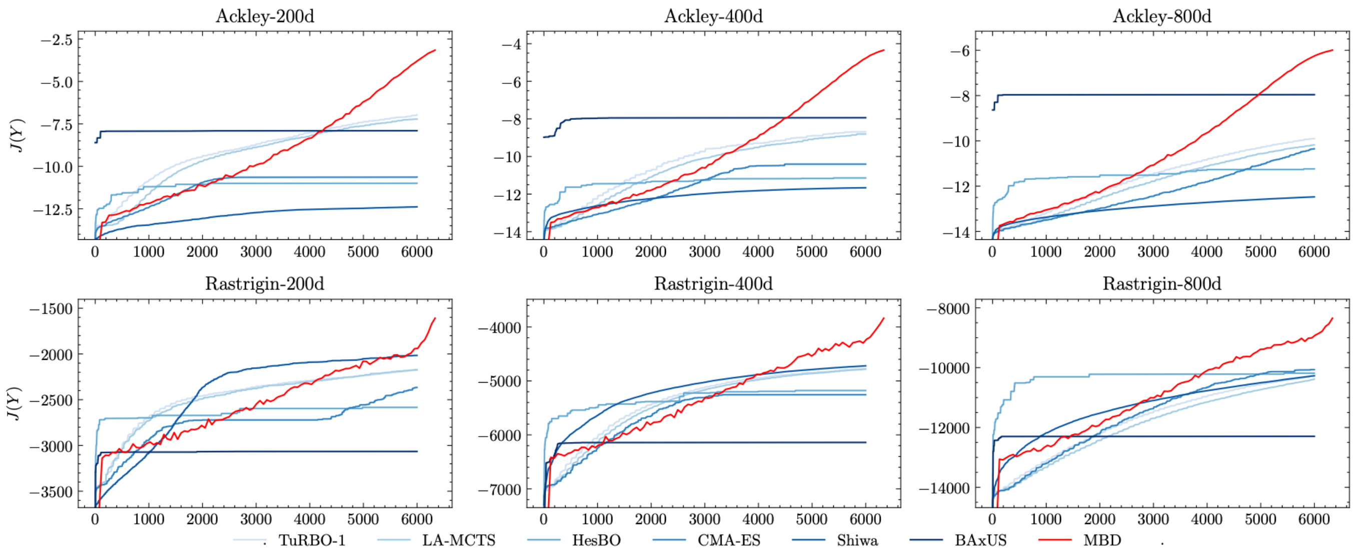


Figure 5: Performance of MBD on high-dimensional black-box optimization benchmarks. MBD outperforms other Gaussian Process-based Bayesian Optimization methods by a clear margin.

Conclusion:

- Model Based Diffusion proposes a trajectory optimization solution without learning the target high dimensional distribution by leveraging the dynamics model to estimate the score function.
- Outperforms other sampling-based optimization techniques, and model free RL.
- If integrated with noisy demonstration/ Imperfect data, leads to improvement in performance of trajectory optimization.