Planning and Control in Uncertain and Dynamic Environments using Generative Models

Ralf Römer

ralf.roemer@tum.de

Supervisors: Angela Schoellig

Collaborators: Lukas Brunke (TUM), Martin Schuck (TUM), Oliver Hausdörder (TUM) and SiQi Zhou (TUM)

CONVEY

Motivation

Artificial intelligence (AI) has seen tremendous success in the last decade and has made a large impact in many scientific domains, such as healthcare, autonomous driving, and natural language processing (NLP). Yet, the creation of **embodied AI**, i.e., the integration of AI techniques into physical bodies that autonomously operate in and interact with the real world, remains a largely unsolved problem. We explore the use of generative models, such as diffusion models and transformers, for decision-making, as they can deal with high-dimensional and multimodal data distributions, which are common in real-world robotics scenarios.

Background

In classical control, the policy $\boldsymbol{\pi}$ maps the current state $\boldsymbol{s}$ to an action $\boldsymbol{a}$, i.e., $\boldsymbol{a} = \boldsymbol{\pi}(\boldsymbol{s})$. Reinforcement learning (RL) can use (visual) observations $\boldsymbol{o}$ as input, but RL policies are either also deterministic or Gaussian. Compared to these paradigms, generative modeling allows for a more flexible policy parameterization.

Definition 1: Generative Models for Decision-Making

Given the current (and potentially past) observations $\boldsymbol{o}$ and potential additional conditioning variables $\boldsymbol{c}$, the next action (or sequence of actions) $\boldsymbol{a}$ is obtained by sampling from a learned conditional distribution $p_{\boldsymbol{\theta}}$, i.e.,

$$\boldsymbol{a} \sim p_{\boldsymbol{\theta}}(\cdot|\boldsymbol{o}, \boldsymbol{c}). \tag{1}$$

This formulation offers several advantages:

- **Scalability**: We can deal with high-dimensional inputs, such as images, and generate high-dimensional outputs, such as desired future trajectories.
- **Multimodality**: We can capture multimodal distributions that are common in imitation learning and offline RL.
- **Conditioning**: The action generation can be flexibly conditioned on additional variables, such as constraints to satisfy or skills to execute.

Safe Offline RL using Diffusion Models

Consider a discrete-time dynamical system with state $\boldsymbol{s}_k$ and action $\boldsymbol{a}_k$ at timestep k. We represent a trajectory as $\boldsymbol{\tau} = (\boldsymbol{s}_0, \boldsymbol{a}_0, \dots, \boldsymbol{s}_T, \boldsymbol{a}_T)$ and denote its cumulative reward by $R(\boldsymbol{\tau})$. Given a static offline dataset $\mathcal{D} = \{\boldsymbol{\tau}_i\}_{i=1}^{N}$, we can learn the distribution

$$p_{\boldsymbol{\theta}}(\boldsymbol{\tau}|\mathcal{O}) \propto p(\boldsymbol{\tau})p(\mathcal{O}|\boldsymbol{\tau}), \tag{2}$$

where $\mathcal{O}$ is a binary variable denoting the optimality of $\boldsymbol{\tau}$ with respect to $R(\boldsymbol{\tau})$, by training a conditional diffusion model [1, 2]. Online, we can generate desired future trajectories by sampling from this distribution, execute the first action in the selected trajectory and repeat this procedure in a receding horizon manner. This formulation can be extended to include additional conditioning variables, such as constraints or visual observations, and we can reduce the complexity by sampling only actions, cf. (1).

Incorporating Constraints

In robotics, we often need to satisfy hard safety constraints $\boldsymbol{s}_k \in \mathcal{S}_k$ and/or $\boldsymbol{a}_k \in \mathcal{A}_k$, $k = 0, 1, \dots$, that may not be known prior to deployment, for example, when operating in dynamic environments. To deal with such unseen test-time constraints, we propose to incorporate a projection method into the backward diffusion process. After each backward diffusion step, we project the noisy trajectory into the safe set. In this way, we obtain a trajectory that is both dynamically feasible and guaranteed to satisfy the constraints.

Evaluation

We evaluate our approach via simulation of two robotic systems: a mobile robot with acceleration commands and a 2D quadrotor with thrust commands. The task is to reach a desired goal while avoiding static and dynamic obstacles.

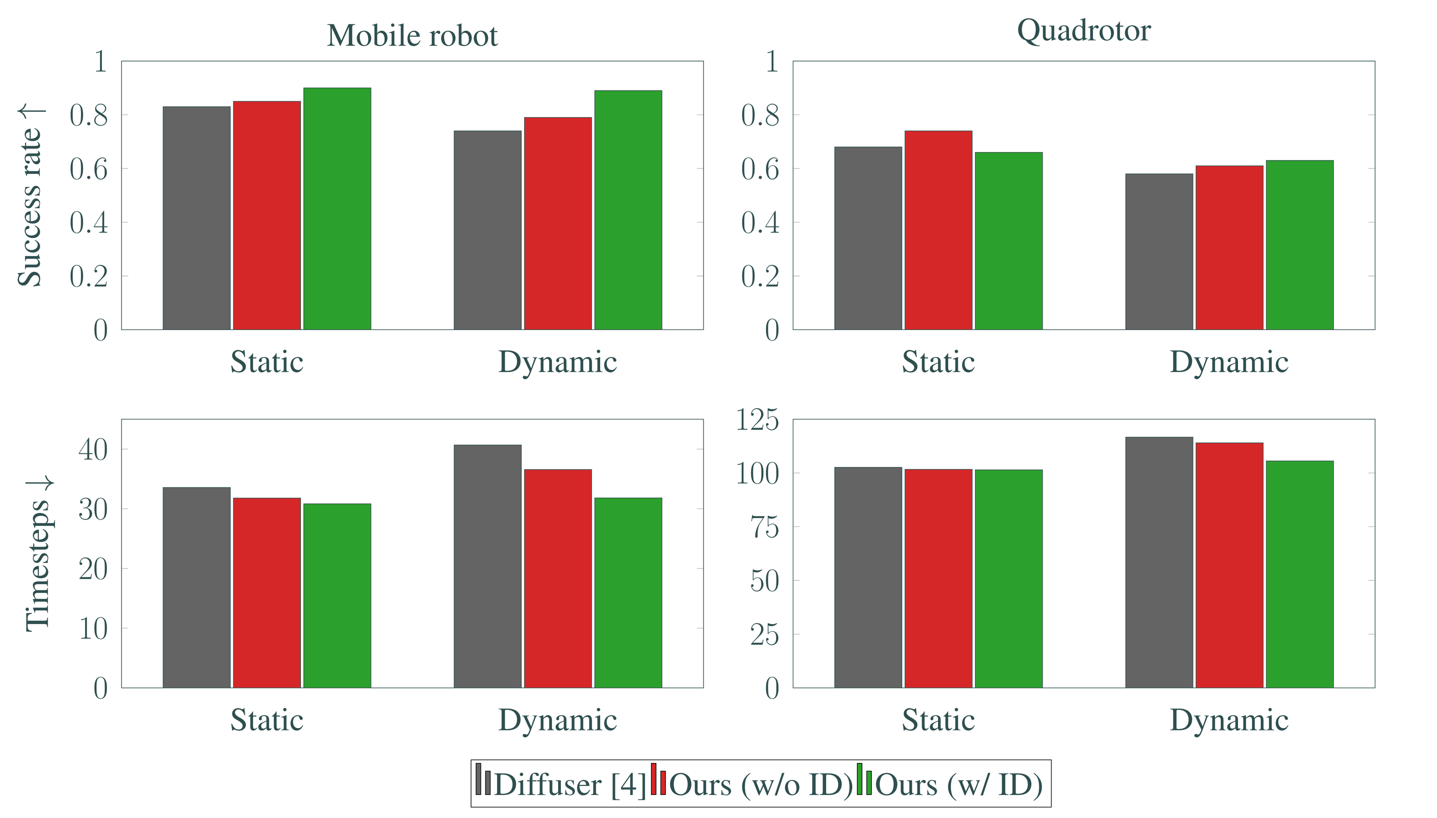

Figure 1: Our proposed projection method reduces the number of constraint violations, resulting in a higher success rate of reaching the goal.

References

[1] P. Dhariwal et al. "Diffusion models beat gans on image synthesis". In: *Advances in neural information processing systems* 34 (2021), pp. 8780–8794.

[2] J. Ho et al. "Classifier-free diffusion guidance". In: *arXiv preprint arXiv:2207.12598* (2022).

[3] J. Ho et al. "Denoising diffusion probabilistic models". In: *Advances in Neural Information Processing Systems* 33 (2020), pp. 6840–6851.

[4] M. Janner et al. "Planning with Diffusion for Flexible Behavior Synthesis". In: *Proceedings of the 39th International Conference on Machine Learning*. Vol. 162. Proceedings of Machine Learning Research. 2022, pp. 9902–9915.