Dynamically Constraining Diffusion Policies During Inference

Niklas Fiedler Florian Vahl Jianwei Zhang

I. INTRODUCTION

In recent years, the capabilities of mobile robots with manipulation skills have made drastic improvements in diverse areas. This was accelerated further by large investments especially in generalizable humanoid robots. Robots have become more efficient in learning tasks and generalizing across more diverse situations. An efficient approach to robot learning is learning from demonstration. The knowledge on solving tasks is often available in form of other systems or humans capable of it. A number of demonstrations collected from such an expert is both the definition of the goal to accomplish as well as of a strategy on how to accomplish it. Prominent examples of learning from demonstration are behavioral cloning [1] and inverse reinforcement learning [2]. Since the publication of the original paper on diffusion policies by Chi et al. [3], they emerged as one of the most efficient approaches for learning from demonstration.

When creating demonstrations for diffusion policies, the robots running them need to be considered from early on. This includes for example the type of control and the action space. Especially when recording demonstrations in joint space, the policies are very specific to the used robot. When collecting demonstrations by teleoperating diverse robots or directly recording human motions, the movement range of a specific system might me surpassed.

In case of the Universal Manipulation Interface (UMI) [4], the human demonstrations need to be executed by a robot. Its action space does not match the demonstrator. The issue is approached by removing demonstrations which do not fit the robot's action space from the demonstration dataset. This is not transferable to other robots and environments with smaller action spaces without filtering the dataset for the new parameters and retraining the policy. The authors of DexGraspAnything [5] improve the performance of their grasping policy by introducing physics constraints during training and inference. Therefore, similar to the approach of UMI, temporary or dynamic restrictions in the action space cannot be accounted for. Retraining is necessary to accommodate for changes in the constraints.

The goal for this work is to develop a strategy capable of guiding pretrained diffusion policies to avoid sampling from restricted areas in the action space without retraining the policy. We aim for an approach which sacrifices success rates for the ability to spontaneously adapt to dynamic constraints in the action space of diffusion policies. This approach allows to collect demonstrations in diverse setups to train a diffusion policy which can then be applied to a specific system with its own constrains without retraining. Especially when dynamically avoiding human collaborators in the workspace, the weakness of the approach, the inability to plan around constraints becomes less relevant as a collaborator can be expected to move out of the way of the robot at some point.

The authors are with TAMS, Faculty of Mathematics, Informatics and Natural Sciences, University of Hamburg, Germany niklas.fiedler@uni-hamburg.de

II. APPROACH

To prevent a diffusion policy from producing actions within restricted areas of the action space, we propose an approach which generally aims to avoid sampling from restricted areas as well as directing the denoising process away from them. This process is only affecting the backward-process of removing noise from the initial sample. Further, it is only active during inference and is not used during training. The network is trained with a diverse set of demonstrations, the actions in some of which are in the restricted area. This is achieved by extending the DDPM algorithm [6] in two ways. One modification targets the noise sampling component and the other the application of the predicted noise.

We utilize an external function to determine whether a certain action is within the restricted area or not. It takes the action as input and returns a boolean indicating whether the action is allowed or not. This allows for flexible definitions of restricted areas. The first strategy aims to avoid sampling from restricted areas in the first place. As soon as a sample is detected within the restricted area at any step of the inference process, it is resampled. The new sample is generated by combining the original with a value randomly sampled in the action space from a normal distribution around its center. The ratio of the diffused and the random sample depends on the step in the denoising process. At the beginning, the random sample is used exclusively. Its weight is reduced linearly down to zero at the end of the process.. Whether a sample is located within the restricted area is determined by an external function.

Second, inspired by the classifier guided diffusion approach presented by Dhariwal et al. [7], a gradient-based approach for guiding samples out of and away from the restricted zones is proposed. If a sample is located within a restricted zone, a vector pointing out of it is cast and set to a unit length. The vector is generated by an external function depending on the shape of the restricted area. Before adding it to the current sample, the vector is multiplied by a parameter which scales the correction for a single diffusion step. This allows to control how firmly sampling in the restricted area is avoided.

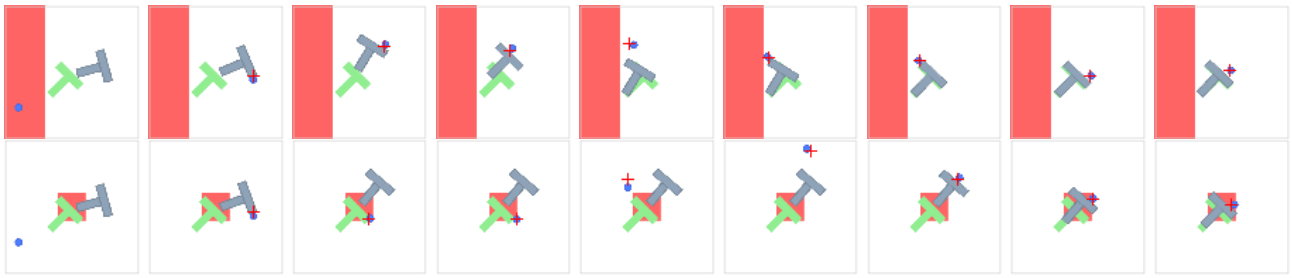

Fig. 1. Exemplary execution of the T-pushing task in Resampling + Gradient 10 configuration in a 0.3 left (upper row) and 0.2 center (lower row) scenario. The manipulator controlled by the diffusion policy is shown in **blue**, the manipulated T-object in **gray**, the goal area for the T-object in **green**, and the restricted area in **red**.

TABLE I

EXPERIMENT RESULTS OF FIVE CONDITIONING CONFIGURATIONS IN FOUR SCENARIOS. AVERAGE SUCCESS RATE (**s**) AND NUMBER OF SAMPLES WITHIN THE RESTRICTED AREA (**r**).

	Scenario		0.1 left	0.5 left	0.1 center	0.2 center
Configuration	None	s	0.92	0.96	0.944	0.91
		r	2.36	104.56	6.32	32.28
	Resampling	s	0.95	0.528	0.914	0.953
		r	0.08	103.48	5.8	26.8
	Gradient 1	s	0.927	0.781	0.953	0.959
		r	0.6	107.6	5.68	20.28
	Gradient 10	s	0.913	0.426	0.893	0.877
		r	0.16	38.6	3.24	6.56
	Resampling + Gradient 1	s	0.858	0.515	0.948	0.915
		r	0.24	97.56	2.36	26.6
	Resampling + Gradient 10	s	0.916	0.475	0.88	0.761
		r	0.0	3.8	2.36	9.48

III. EXPERIMENTS

We selected the pushT task from the original paper on diffusion policies [3] as evaluation scenario. The goal is for the policy to push a two-dimensional T-piece into a predefined goal area. The action space is defined by the pushers position on an X-Y-plane in a discrete 512×512 grid. For our experiments, we restrict moving into a rectangular area. Figure 1 shows two manipulation sequences in restricted action spaces generated for exemplary scenarios. As evaluation metrics, the number of action steps sampled within the constrained area as well as the achieved score are used. The score is defined as the relation of the area of overlap between the pushed T and the goal shape divided by 0.95 clipped between 0 and 1. This results in a perfect score of 1 for an overlap of 0.95 or higher between goal and achieved pose. Every run performs 200 action steps. For each experiment run, the same 25 randomly generated scenarios are used. Each scene is performed once. For reproducability and to demonstrate that no specific training is required, the pretrained weights provided with the code of the original diffusion policy paper are used.

We perform a series of experiments to demonstrate and investigate the overall functionality and behavior of the systems parameters in various scenarios. Four scenarios are investigated. Two restrict movements into an area spread over the full height of the workspace starting from the left side. The first covers a tenth of the overall workspace and the second half of it. This simulates a situation in which the robot workspace is smaller than that of the demonstrator. In the two other scenarios, constrained areas with a width and height of 0.1 and 0.2 of the workspace size are spawned right in its center, mimicking an obstacle. The *None* configuration without any constraint handling is used as a baseline. We test both components of the approach, resampling points from the restricted area and pushing them out following a gradient separately and together. Values of 1 and 10 for the scaling of the gradient are evaluated. Table I lists the results of the experiments.

The experiments show, that depending on the configuration areas of the action space can be reliably avoided while the policy is still able to achieve the desired goal in many cases. Note that small numbers of actions within the restricted areas are to be expected as the starting position of the actor might be randomly sampled within the restricted area. Of course, as the restricted area grows and is moved towards the goal position (center), the success score is lowered. Large scaling parameters for the gradient based approach in combination with resampling are especially effective in avoiding the restricted areas.

Further, this is not just applicable directly to the action space but also indirectly via more complex functions which decide whether a sample point is within a feasible area. This allows for example to avoid areas in Cartesian space while the actions are performed in joint space. Such types of restrictions might however produce complex shapes of restricted areas leading to a high failure rate of the policy. Also, when predicting the next steps relative to an intermediate pose, the actions can be transformed into a global frame and restricted accordingly.

Similar to DexGraspAnything, static constraints such as physics can be introduced in this manner, but these can and should be introduced during training and not just at the time of interference. This would allow the policy to actively handle the constrains which is especially relevant for concave shapes of restricted areas.

IV. CONCLUSION AND FUTURE WORK

The approach is successful in adapting a diffusion policy to slight restrictions of the action space without retraining and dynamic avoidance of moving obstacles. Adapting the diffusion process is far more efficient than fully resampling the action from scratch and hoping for a result which does not interfere with the restricted area. In future work, the approach should be evaluated with dynamically changing restrictions for example in a human collaboration scenario.

REFERENCES

[1] F. Torabi, G. Warnell, and P. Stone, "Behavioral cloning from observation," *arXiv preprint arXiv:1805.01954*, 2018.

[2] S. Arora and P. Doshi, "A survey of inverse reinforcement learning: Challenges, methods and progress," *Artificial Intelligence*, vol. 297, p. 103500, Aug. 2021.

[3] C. Chi, Z. Xu, S. Feng, E. Cousineau, Y. Du, B. Burchfiel, R. Tedrake, and S. Song, "Diffusion Policy: Visuomotor Policy Learning via Action Diffusion," https://arxiv.org/abs/2303.04137v5, Mar. 2023.

[4] C. Chi, Z. Xu, C. Pan, E. Cousineau, B. Burchfiel, S. Feng, R. Tedrake, and S. Song, "Universal Manipulation Interface: In-The-Wild Robot Teaching Without In-The-Wild Robots," in *Robotics: Science and Systems XX*. Robotics: Science and Systems Foundation, Jul. 2024.

[5] Y. Zhong, Q. Jiang, J. Yu, and Y. Ma, "DexGrasp Anything: Towards Universal Robotic Dexterous Grasping with Physics Awareness," Mar. 2025.

[6] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *Advances in neural information processing systems*, vol. 33, pp. 6840–6851, 2020.

[7] P. Dhariwal and A. Nichol, "Diffusion models beat gans on image synthesis," *Advances in neural information processing systems*, vol. 34, pp. 8780–8794, 2021.