Robot Skill Adaptation via Soft Actor-Critic Gaussian Mixture Models

Iman Nematollahi*, Erick Rosete-Beas*, Adrian Röfer, Tim Welschehold, Abhinav Valada, Wolfram Burgard

Abstract— **A core challenge for an autonomous agent acting in the real world is to adapt its repertoire of skills to cope with its noisy perception and dynamics. To scale learning of skills to long-horizon tasks, robots should be able to learn and later refine their skills in a structured manner through trajectories rather than making instantaneous decisions individually at each time step. To this end, we propose the Soft Actor-Critic Gaussian Mixture Model (SAC-GMM), a novel hybrid approach that learns robot skills through a dynamical system and adapts the learned skills in their own trajectory distribution space through interactions with the environment. Our approach combines classical robotics techniques of learning from demonstration with the deep reinforcement learning framework and exploits their complementary nature. We show that our method utilizes sensors solely available during the execution of preliminarily learned skills to extract relevant features that lead to faster skill refinement. Extensive evaluations in both simulation and real-world environments demonstrate the effectiveness of our method in refining robot skills by leveraging physical interactions, high-dimensional sensory data, and sparse task completion rewards. Videos, code, and pre-trained models are available at `http://sac-gmm.cs.uni-freiburg.de`.**

I. Introduction

Thinking ahead is a hallmark of human intelligence. From early infancy, we form rich primitive object concepts through our physical interactions with the real world and apply this knowledge as an intuitive model of physics for reasoning about physically plausible trajectories and adapting them to suit our purposes [1]. This is at odds with most current deep imitation and reinforcement learning paradigms for robot sensorimotor control, which, despite recent progress [2]–[5], are typically trained to make isolated decisions at each time step of the trajectory. In fact, most existing methods for learning manipulation skills are end-to-end high-capacity models that map directly from pixels to actions [6]–[8]. However, although these approaches can capture complex relationships and are flexible to adapt in face of noisy perception, they require extensive amounts of data, and the trained agent is typically bound to take a distinct decision at every time step.

Learning from demonstration [9] is the classical paradigm to tackle the problem of representing skills with a trajectory-space policy. In this context, dynamical systems have shown to be a physically plausible motion generation mechanism that provides a high level of reactivity and robustness against perturbations in the environment [10]–[14]. Despite the great success of dynamical systems in affording flexible robotic systems for industry, where a high-precision state of the environment is available, they are still of limited use in more complex real-world robotics scenarios. The main limitations of current dynamical systems in contrast to deep sensorimotor learning methods are their incompetence in handling raw high-dimensional sensory data such as images, and their susceptibility to noise in the perception pipeline.

In this paper, we advocate for hybrid models in learning robot skills: "Soft Actor-Critic Gaussian Mixture Models" (SAC-GMMs). SAC-GMMs learn and refine robot skills in the real-world and present a hybrid model that combines dynamical systems and deep reinforcement learning in order to leverage their complementary nature. More precisely, SAC-GMMs learn a trajectory-based Gaussian mixture policy of skills from demonstrations and refine it by physical interactions of a soft actor-critic agent with the world. Our hybrid formulation allows the dynamical system to utilize high-dimensional observation spaces and cope with noise in demonstrations and sensory observations while maintaining a reactive and robust trajectory-based policy when interacting with dynamic environments. We argue that maintaining this physically meaningful structure within the reinforcement learning refinement will yield enhanced performance and stability compared to residual corrections or direct learning of desired end-effector velocities. The method is simple, sample efficient and readily applicable in a variety of robotics scenarios. We exemplify this, by using our hybrid model for simulated peg insertion and power lever sliding skills, and a real-world door-opening skill. We demonstrate that SAC-GMM is able to successfully open a door in the real world after half an hour of physical interaction.

The main contributions of this paper are: 1) a hybrid model for learning and refining skills in trajectory distribution space, 2) exploiting high-dimensional sensory inputs obtainable solely during skill adaptation through physical interaction, such as tactile images, gripper camera images, and static camera depth maps to refine parameters of a dynamical system, 3) mitigating amount of robot exploration efforts for learning skills in sparse reward settings through a dynamical system model learned from few demonstrations, and 4) learning to refine two simulated and one real-world robot manipulation skills.

II. SAC-GMM

We propose the hybrid approach Soft Actor-Critic Gaussian Mixture Models (SAC-GMM) consisting of two phases. In the first, we learn a dynamical system parameterization in

*These authors contributed equally. All authors are with University of Freiburg, Germany. This work has been supported by German Federal Ministry of Education and Research under contract number 01IS18040B-OML and the BrainLinks-BrainTools center of the University of Freiburg.

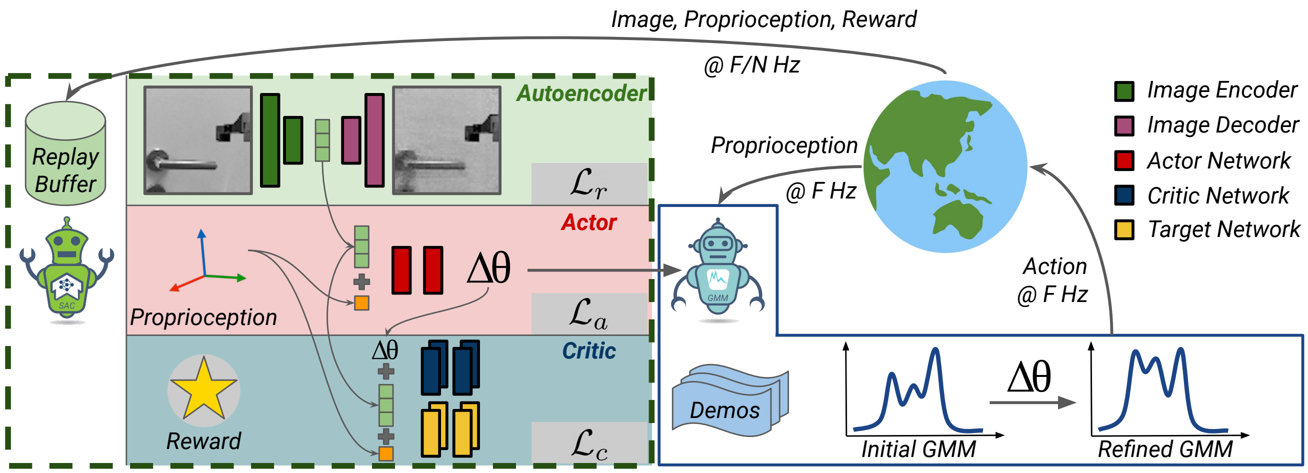

Fig. 1: Structure of SAC-GMM: Given a set of human demonstrations, the GMM agent learns an initial dynamical system parameterization θ that provides an analytical description of the robot's skill trajectory. After N interactions by the GMM agent with the environment (working with frequency F), the SAC agent receives a high-dimensional observation, robot state, and a sparse success reward from the environment. It then refines the initial GMM agent's trajectory parameters by $\Delta\theta$ for the next N interactions to optimize for skill success.

form of a Gaussian mixture model from a few demonstrations. In the second, we refine this dynamical system with the soft actor-critic algorithm through physical interactions with the world. The architecture of our hybrid model is shown in Figure 1.

A. Dynamical System: Gaussian Mixture Model Agent

Dynamical systems afford an analytical representation of a motion's progression over time, and accordingly, they enable the robot to generate trajectories while being robust in the face of perturbations.We formulate a robot skill as a control law driven by an autonomous dynamical system, defined by the robot pose ξ:

$$\dot{\xi} = f_\theta(\xi) + \boldsymbol{\epsilon}, \tag{1}$$

where f_θ is the robot skill model, a parametric, non-linear, steady, and continuously differentiable function, and $\boldsymbol{\epsilon}$ is a zero-mean additive Gaussian noise. From a machine learning perspective, learning the noise-free estimate of f from data is a regression problem and can be addressed by a mixture of Gaussians. Given a set of reference demonstrations $\mathbf{X}$ for a robot skill, we parametrize Eq. (1) through Gaussian Mixture Regression (GMR) [15]. We first estimate the joint probability density $\mathcal{P}(\dot{\xi}, \xi)$ of the robot pose and its corresponding first-order derivative by a Gaussian mixture. Thereby, we parametrize the robot skill model f by $\theta = \{\pi_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\}_{k=1}^K$, where π_k are the priors (or mixing weights), $\boldsymbol{\mu}_k$ the means and $\boldsymbol{\Sigma}_k$ the covariances of the k Gaussian functions.

By using this estimated joint probability density function, we employ Gaussian mixture regression (GMR) to retrieve $\dot{\xi}$ given ξ as the conditional distribution $\mathcal{P}(\dot{\xi} \mid \xi)$. This way our skill model can reproduce the demonstrated skill by estimating the next velocity at the current robot pose and thus generate a trajectory by updating the pose ξ with the generated velocity $\dot{\xi}$ scaled by a time step and proceeding iteratively. For detailed insights on using GMM encoded dynamic system for imitation learning we refer the reader to the extensively available literature [11], [12], [15]–[17].

B. Dynamical System Adaptation: Soft Actor-Critic Agent

Having learned the skill model f_θ, we can now leverage robot interactions with the world to explore and refine the initial model. We formulate this refinement as a reinforcement learning problem in which the agent has to modify the learned skill in the trajectory space and has only access to sparse rewards. In our skill refinement scenario, the agent receives high-dimensional sensory measurements such as RGB images, tactile measurements, or depth maps which are encoded to a latent representation $\mathbf{z}$ by an autoencoder. Together with the robot pose ξ, these form our continuous state space. The action space is also continuous and consists of the desired adaptation in the skill trajectory parameters $\Delta\theta$. Moreover, the environment emits a sparse reward only if the robot executes the skill effectively. Namely, if $\mathbf{s}_t$, $\mathbf{a}_t$ and $\mathbf{z}_t$ define the robot's state, action, and latent representation of observation respectively at time step t, then

$$\mathbf{s}_t := \{\xi_t, \mathbf{z}_t\}, \quad \mathbf{a}_t := \{\Delta\pi_k, \Delta\boldsymbol{\mu}_k, \Delta\boldsymbol{\Sigma}_k\}_{k=1}^K, \tag{2}$$

and consequently

$$\Delta\theta = \pi_\phi(\mathbf{a}_t \mid \mathbf{s}_t), \tag{3}$$

where π_ϕ is the robot skill refinement policy. We use $\rho_{\pi_\phi}(\mathbf{s}_t, \mathbf{a}_t)$ to denote the state-action marginal of the trajectory distribution induced by the policy π_ϕ. The robot has to learn this policy from its interactions with the world, such that it maximizes the expected total reward of the refined skill trajectory. Our particular choice for the reinforcement learning framework to learn the skill refinement policy is the soft actor-critic (SAC) algorithm [5].

Our SAC agent stores a collection of $\{\mathbf{s}_t, \mathbf{o}_t, \mathbf{a}_t, r_t, \mathbf{s}_{t+1}, \mathbf{o}_{t+1}\}_{i=1}^T$ transition tuples in a replay buffer $\mathcal{D}$, and concurrently learns an autoencoder AE_ω, a policy π_ϕ and two Q-functions Q_{ψ_1} and Q_{ψ_2} (to prevent overly optimistic value estimates) and their target networks. More concretely, we use the autoencoder AE_ω to learn a low-dimensional latent representation of the robot's high-dimensional observations.

Model \ Task	Peg Insertion				Lever Sliding			
	No Noise		With Noise		No Noise		With Noise	
	No Tactile	With Tactile	No Tactile	With Tactile	No Depth	With Depth	No Depth	With Depth
GMM	20%	x	10%	x	54%	x	39%	x
SAC	0%	0%	0%	0%	0%	0%	0%	0%
Res-GMM	24%	30%	26%	23%	49%	62%	29%	38%
SAC-GMM	**44%**	**86%**	**33%**	**56%**	**68%**	**81%**	**42%**	**52%**

TABLE I: The average success rate of skill models over 100 trials per five different random seeds, under various noise and sensors settings for the simulated robot skills.

C. Full Model

Figure 1 shows how our hybrid model learns and refines a robot skill. The GMM agent is fitted on the provided demonstrations and represents a dynamical system, controlling the motion in the trajectory space. After each N interaction steps with the world driven by the GMM encoded dynamics, the SAC agent receives the current state $\mathbf{s}_t$ consisting of the latent observation $\mathbf{z}_t$ and the robot state ξ_t, and additionally a reward r_t for the previous step. It then generates an action $\mathbf{a}_t := \{\Delta\pi_k, \Delta\boldsymbol{\mu}_k, \Delta\boldsymbol{\Sigma}_k\}_{k=1}^{K}$ according to the current $\mathbf{s}_t$ to adapt the original GMM for the next N interactions with the environment.

III. Experimental Evaluation

We evaluate SAC-GMM for learning robot skills in both simulated and real-world environments. The goals of these experiments are to investigate: (i) whether our hybrid model is effective in performing skills in realistic noisy environments, (ii) if exploiting high-dimensional data boosts the dynamical system adaptation, and (iii) how refining robot skills in trajectory space compares with alternative exploration policies in terms of accuracy and exploration budget.

A. Evaluation Protocol

For skills in simulation, we collect 20 demonstrations by teleoperating the robot. For the door opening skill in the real-world, we use OpenPose [18] to track the human hand and collect 5 human demonstrations. We compare our skill model against the following models:

GMM: This baseline corresponds to the same dynamical system that we learn with the provided demonstrations.

SAC: We employ the soft actor-critic agent [5] to explore and learn the skills. We initialize the replay buffer of the SAC agent with the demonstrations of skills.

Res-GMM: Analogous to our approach, this baseline first learns a GMM agent using the demonstrations and then employs a SAC agent for skill refinement. In contrast to our approach,this SAC agent acts at each time step (instead of each N step), and instead of predicting change in trajectory parameters, it predicts a residual velocity which is summed up with the GMM agent's predicted velocity.

B. Experiments in Simulation

We start by evaluating our method on the peg insertion and power lever sliding skills in simulation. Quantitative results of the average success rate of each skill model over 100 trials per five different random seeds are reported in Table I. Our SAC-GMM successfully performs the peg insertion and lever sliding skills with an average final success rate of 86% and 81% respectively. It achieves significantly higher success rates than the GMM baseline, proving its effectiveness in refining the robot skills through physical interaction.

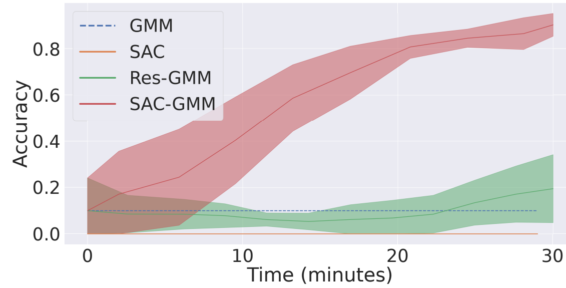

Fig. 2: Real-world door opening skill: SAC-GMM learns to open a real door after half an hour of physical interactions.

To analyze the influence of noise in the perception pipeline and the efficacy of SAC-GMM in exploiting high-dimensional sensory data, we conduct several experiments on the simulated environments (see Table I). We observe that SAC-GMM can fully leverage high-dimensional data such as tactile measurements and depth maps in all scenarios to achieve a superior skill success rate. Table I also shows the success rate of SAC-GMM with and without having access to the tactile sensors [19] during refinement of the peg insertion skill and to the depth camera during refinement of the lever sliding skill in the noisy setup. We find that SAC-GMM utilizes the high-dimensional observation to deal better with noise and learn the skill faster.

C. Real-World Door Opening

Fig. 2 reports the results for the door opening skill in the real world. We find that, although the initial dynamical system (the GMM agent fitted on human demonstrations) enables the robot to reach the door handle, the robot misses the proper position to apply its force and can only open the door with a 10% success rate. This failure is due to the robot's noisy perception and dynamics. Our SAC-GMM exploits the wrist-mounted camera RGB images and sparse door opening rewards and achieves a 90% success rate after only half an hour of physical interactions ($\sim$100 episodes) with the door. The SAC baseline fails to learn the skill and the Res-GMM model performs poorly, as adding residual velocities at each time step results in non-smooth trajectories.

IV. Conclusions

In this work, we present "Soft Actor-Critic Gaussian Mixture Models" as a new framework for learning robot skills. This hybrid model leverages reinforcement learning to refine robot skills represented via dynamical systems in their trajectory distribution space and exploits the natural synergy between data-driven and analytical frameworks. Extensive experiments carried out in both simulation and real-world settings, demonstrate that our proposed skill model: 1) learns to refine robot skills through physical interactions in realistic noisy environments, 2) exploits high-dimensional sensory inputs available during skill refinement to cope better with noise, and 3) performs robot skills significantly better than comparable alternatives considering the performance accuracy and exploration costs.

References

[1] B. M. Lake, T. D. Ullman, J. B. Tenenbaum, and S. J. Gershman, “Building machines that learn and think like people,” *Behavioral and brain sciences*, vol. 40, 2017.

[2] J. Ho and S. Ermon, “Generative adversarial imitation learning,” *Advances in Neural Information Processing Systems*, vol. 29, pp. 4565–4573, 2016.

[3] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, “Proximal policy optimization algorithms,” *arXiv preprint arXiv:1707.06347*, 2017.

[4] S. Fujimoto, H. Hoof, and D. Meger, “Addressing function approximation error in actor-critic methods,” in *Int. Conf. on Machine Learning*, 2018, pp. 1587–1596.

[5] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, “Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor,” in *Int. Conf. on Machine Learning*, 2018, pp. 1861–1870.

[6] C. Finn, T. Yu, T. Zhang, P. Abbeel, and S. Levine, “One-shot visual imitation learning via meta-learning,” in *Conf. on Robot Learning*, 2017, pp. 357–368.

[7] A. Zeng, S. Song, K.-T. Yu, E. Donlon, F. R. Hogan, M. Bauza, D. Ma, O. Taylor, M. Liu, E. Romo, *et al.*, “Robotic pick-and-place of novel objects in clutter with multi-affordance grasping and cross-domain image matching,” in *Int. Conf. on Robotics & Automation*, 2018.

[8] L. Berscheid, P. Meißner, and T. Kröger, “Self-supervised learning for precise pick-and-place without object model,” *IEEE Robotics and Automation Letters*, vol. 5, no. 3, pp. 4828–4835, 2020.

[9] A. Billard, S. Calinon, R. Dillmann, and S. Schaal, “Survey: Robot programming by demonstration,” *Springer Handbook of Robotics*, pp. 1371–1394, 2008.

[10] S. Schaal, S. Kotosaka, and D. Sternad, “Nonlinear dynamical systems as movement primitives,” *International Journal of Humanoid Robotics*.

[11] S. M. Khansari-Zadeh and A. Billard, “Learning stable nonlinear dynamical systems with gaussian mixture models,” *IEEE Transactions on Robotics*, vol. 27, no. 5, pp. 943–957, 2011.

[12] A. J. Ijspeert, J. Nakanishi, H. Hoffmann, P. Pastor, and S. Schaal, “Dynamical movement primitives: learning attractor models for motor behaviors,” *Neural computation*, vol. 25, no. 2, pp. 328–373, 2013.

[13] J. Silvério, L. Rozo, S. Calinon, and D. G. Caldwell, “Learning bimanual end-effector poses from demonstrations using task-parameterized dynamical systems,” in *Int. Conf. on Intelligent Robots and Systems*, 2015, pp. 464–470.

[14] S. Manschitz, M. Gienger, J. Kober, and J. Peters, “Mixture of attractors: A novel movement primitive representation for learning motor skills from demonstrations,” *IEEE Robotics and Automation Letters*, vol. 3, no. 2, pp. 926–933, 2018.

[15] S. Calinon, F. D’halluin, E. L. Sauser, D. G. Caldwell, and A. G. Billard, “Learning and reproduction of gestures by imitation,” *IEEE Robotics & Automation Magazine*, vol. 17, no. 2, pp. 44–54, 2010.

[16] P. Pastor, H. Hoffmann, T. Asfour, and S. Schaal, “Learning and generalization of motor skills by learning from demonstration,” in *Int. Conf. on Robotics & Automation*, 2009, pp. 763–768.

[17] N. Figueroa and A. Billard, “A physically-consistent bayesian non-parametric mixture model for dynamical system learning.” in *Conf. on Robot Learning*, 2018, pp. 927–946.

[18] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, “Openpose: realtime multi-person 2d pose estimation using part affinity fields,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 43, no. 1, pp. 172–186, 2019.

[19] S. Wang, M. Lambeta, P.-W. Chou, and R. Calandra, “Tacto: A fast, flexible and open-source simulator for high-resolution vision-based tactile sensors,” *arXiv preprint arXiv:2012.08456*, 2020.