

Deep Model Predictive Variable Impedance Control

Akhil S Anand^{a,*}, Fares J. Abu-Dakka^b, Jan Tommy Gravdahl^a

^a*Dept. of Engineering Cybernetics at Norwegian University of Science and Technology (NTNU), Trondheim, Norway.*
^b*Intelligent Robotics Group, Department of Electrical Engineering and Automation (EEA) at Aalto University, Aalto, Finland.*

Abstract

The capability to adapt compliance by varying muscle stiffness is crucial for dexterous manipulation skills in humans. Incorporating compliance in robot motor control is crucial to performing real-world force interaction tasks with human-level dexterity. This work presents a Deep Model Predictive Variable Impedance Controller for compliant robotic manipulation which combines Variable Impedance Control with Model Predictive Control (MPC). A generalized Cartesian impedance model of a robot manipulator is learned using an exploration strategy maximizing the information gain. This model is used within an MPC framework to adapt the impedance parameters of a low-level variable impedance controller to achieve the desired compliance behavior for different manipulation tasks without any retraining or finetuning. The deep Model Predictive Variable Impedance Control approach is evaluated using a Franka Emika Panda robotic manipulator operating on different manipulation tasks in simulations and real experiments. The proposed approach was compared with model-free and model-based reinforcement approaches in variable impedance control for transferability between tasks and performance.

Keywords: Variable impedance control, Model predictive control, Robot learning .

1. Introduction

Manipulating objects is central to how humans interact with the real-world and even with a limitation of low-frequency biological feedback loops, we possess dexterous manipulation skills. Although the exact motor control mechanisms responsible for such skills remain unknown, the impedance modulation of the arm has been proposed as key mechanism [1, 2, 3]. While in robotics the feedback control loops can be operated at much higher frequencies, human-level dexterity is seldom achieved in real-world applications. Most real-world applications using robotic manipulators traditionally relied on trajectory planning and position control which is undesirable for dexterity, safety, energy efficiency, and constrained interactions. Human muscle actuators posses the impedance properties (stiffness and damping) [4] which can be adapted by the neural control to achieve various manipulation behaviours. Motivated by human manipulation, Impedance Control (IC) for

robot control, introduced by Hogan in [5], aims to couple the manipulator dynamics with its environment instead of treating it as an isolated system while designing control strategies. Unlike the more conventional control approaches, IC attempts to implement a dynamic relation between manipulator variables such as end-point positions and forces rather than just control these variables independently. The use of IC provides a feasible solution to overcome position uncertainties in order to avoid large impact forces since robots are controlled to modulate their motion or compliance according to force feedback [6].

IC is naturally extended to Variable Impedance Control (VIC) where the impedance parameters are varied during the task [7, 8]. VIC gained popularity in robotic research due the adaptability and safety properties. The emergence of learning algorithms has generated great interest in learning based VIC, where a learned policy is used to adapt the impedance gains in Variable Impedance Learning Control (VILC) framework. In [9], a detailed review of various learning approaches applied to VIC is presented. While using Reinforcement Learning (RL) for VILC, or in general to robotics, the control policies obtained are task or scenario-specific, based on what they encountered during the learning. Model

*Corresponding author
 Email addresses: akhil.s.anand@ntnu.no (Akhil S Anand), fares.abu-dakka@aalto.fi (Fares J. Abu-Dakka), jan.tommy.gravdahl@ntnu.no (Jan Tommy Gravdahl)

Table 1: Comparison among state-of-the-art of VILC approaches

	Data-efficiency	Task transferability	Model-based/ Model-free	Computation time	force-/position-based VIC
[10, 11, 12, 13]	low	-	model-free	low	force
[14, 15]	low	-	model-free	low	position
[16]	high	-	model-free	low	force
[17, 18]	high	-	model-based	high	position
[19]	high	-	model-based	high	force
Our MPVIC	high	✓	model-based	high	force

Predictive Control (MPC) offers a framework to formulate the control systems as an optimization problem based on a system model and an optimization objective [20]. MPC approaches are widely used in robotic control when a model of the system dynamics is available. In the case of a robot controlled using a Cartesian space VIC, it is possible to learn a Cartesian impedance model of the robot and utilize it in an MPC framework to optimize control policies over various tasks by providing suitable optimization objectives. While complex dynamics can be estimated from data using Neural Network (NN)s, they suffer from over-fitting and can not quantify uncertainties. Gaussian Processes (GP) models are capable of modelling the uncertainty, but they don't scale well with high-dimensional data. Probabilistic Ensemble NN (PE) models introduced in [21] overcome these limitations of NNs, offering a way to quantify both aleatoric and epistemic uncertainties.

In this paper, we propose a deep Model Predictive Variable Impedance Control (MPVIC) framework, where a NN based Cartesian impedance model of the robotic manipulator is used in a Cross Entropy Method (CEM)-based MPC for online adaptation of the impedance parameters of a VIC. This deep MPVIC framework is utilized to learn impedance adaptation strategy for various robotic manipulation tasks by specifying a suitable cost function. The main contributions of this paper are:

- a novel VIC framework, we call it deep MPVIC, which combines a CEM-based MPC with PE dynamical model for compliant robotic manipulation.
- the deep MPVIC framework learns a generalized Cartesian impedance model of the robot to facilitate the transferability between completely different manipulation tasks without any need of relearning the model.
- an uncertainty-based exploration scheme is integrated into the proposed framework to facilitate

learning a generalized model efficiently from fewer samples.

- an extensive evaluation in simulation and real setups, in addition to a comparison between our approach and the state-of-the art model-free and model-based RL approaches on transferability and performance.

The rest of the paper is organized as follows. Section 2 describes the existing references relevant to our work. Section 3 briefly introduces the necessary background knowledge, in Section 4 presents the details of the deep MPVIC framework proposed. Section 5 presents the evaluation of our approach on simulation and experimental setups using Franka Panda robotic manipulator. Detailed discussion on the results and the limitations of our approach is presented in Section 6 and conclusion in Section 7.

2. Related Work

Recently, RL has been explored largely for VILC research. However, RL demands large amount of data samples/interactions to obtain high performance. References [10, 11, 14, 12, 13] are some examples of using deep RL for VILC applied to different robotic manipulation tasks. All these approaches could learn complex VIC policies for specific tasks, however, at the expense of sample efficiency. Ref. [15] combines human demonstrations with RL, providing improved sample efficiency for learning stiffness control policies. But it is not suitable for force-based VIC, as unlike stiffness values the impedance parameters can not be estimated directly from kinesthetic demonstrations used in [15]. Ref. [16] demonstrated model-free RL based VILC using Dynamic Movement Primitive (DMP) policy and Policy Improvement with Path Integrals (PI²), which is sample efficient but fails to scale to complex policies. Whereas our MPC based approach is scalable to complex problems with a NN dynamics model.

Apart from sample efficiency a major drawback of the referenced RL based approaches is their inability to easy transfer of a learned policy to a different task. In practice, retraining the policy is necessary, which is difficult in real-world robotic tasks. In contrast, our deep-MPVIC framework use a generalized Cartesian impedance model of the robot with an MPC policy that can be used for multiple tasks by designing suitable cost functions.

Alternatively, Model-based Reinforcement Learning (MBRL) approaches offer a sample efficient framework leveraging on the model. In [17] MBRL is used for learning position-based VIC on industrial robots using GP models. Ref. [19] used a similar approach for force-based VIC for contact sensitive tasks. Both of these approaches utilizes GP models and PILCO algorithm limiting its use to less complex tasks with smooth dynamics and relatively simple policies and reward structure. In [18], a Probabilistic Ensembles with Trajectory Sampling (PETS) approach is utilized for learning position-based VIC strategy for Human-Robot Collaboration (HRC) tasks. Similar to RL approaches referenced earlier, all of these MBRL-based approaches are task specific and generally lacks the performance of model-free RL approaches [22]. Unlike aforementioned VILC approaches, our deep MPVIC is not only able to adapt to new situations of the same task, but also it is transferable to new tasks using the same trained model without any need to re-train or train a new model. Transferability between tasks is achieved by combining a generalised Cartesian impedance model with an MPC scheme. A comparison between existing RL-based VILC approaches is summarised in Table 1.

In literature, MPC is used in robotic interaction control for manipulations tasks [23, 24], where MPC optimizes the robot control input but not the stiffness itself, while in our approach the MPC adapt the stiffness values directly. It is possible to couple our deep MPVIC with the approach in [23] where it can be used as a low level optimizer to solve additional constraints. Haninger *et al.* [25] used an MPC scheme with GP models for human-robot interaction tasks. The MPC scheme used could optimize the impedance parameters for an admittance controller, but it is task specific as the human force model is estimated from demonstrations as a function of robot states. Using GP models limits the complexity and generalizability of the model as pointed out by the authors in [25]. Unlike [25], we optimize the impedance parameters for a force-based VIC in our deep MPVIC framework using PE to model the Cartesian impedance behaviour of the robot manipulator.

For efficient model learning in terms of sample effi-

ciency, uncertainty-based exploration with ensembles of NNs has been proposed in prior works [26, 27, 28, 29]. The basis for uncertainty-based exploration for model learning is derived from the expected information gain formulation in [30]. In [31] this approach is termed *curiosity-driven exploration*. The model uncertainty is evaluated based on the variance of the model in predicting the next state. We incorporated *curiosity-driven exploration* to our deep MPVIC framework to learn a generalized Cartesian impedance model sample efficiently.

3. Background

3.1. Robot Manipulator Dynamics

For a rigid n -DOF robotic arm, the task space formulation of the robot dynamics is given by

$$\Lambda(\mathbf{q})\ddot{\mathbf{x}} + \Gamma(\mathbf{q}, \dot{\mathbf{q}})\dot{\mathbf{x}} + \boldsymbol{\eta}(\mathbf{q}) = \mathbf{f}_c - \mathbf{f}_{ext}, \quad (1)$$

where $\dot{\mathbf{x}}, \ddot{\mathbf{x}}$ are velocity and acceleration of the robot end-effector in task space, \mathbf{f}_c is the task space control force, \mathbf{f}_{ext} is the external force, $\Gamma(\mathbf{q}, \dot{\mathbf{q}}) \in \mathbb{R}^{6 \times 6}$ is a matrix representing the centrifugal and Coriolis effects, and $\boldsymbol{\eta}(\mathbf{q}) = \mathbf{J}^{-T}\mathbf{g}(\mathbf{q}) \in \mathbb{R}^{6 \times 1}$ is the gravitational force, where $\mathbf{g}(\mathbf{q})$ is the joint space forces and torques. The Cartesian inertia matrix is denoted as $\Lambda(\mathbf{q}) = (\mathbf{J}\mathbf{H}(\mathbf{q})^{-1}\mathbf{J}^T)^{-1} \in \mathbb{R}^{6 \times 6}$, where $\mathbf{H}(\mathbf{q}) \in \mathbb{R}^{n \times n}$ is the joint space inertia matrix and \mathbf{J} is the end-effector geometric Jacobian. By additionally knowing the joint space centrifugal and Coriolis matrix, $\mathbf{V}(\mathbf{q}, \dot{\mathbf{q}})$, the corresponding task space matrix is given by,

$$\Gamma(\mathbf{q}, \dot{\mathbf{q}}) = \mathbf{J}^{-T}\mathbf{V}(\mathbf{q}, \dot{\mathbf{q}})\mathbf{J}^{-1} - \Lambda(\mathbf{q})\mathbf{J}\mathbf{J}^{-1}. \quad (2)$$

3.2. Variable Impedance Control

VIC is designed to achieve force regulation by adjusting the system impedance [32], via the adaptation of the inertia, damping, and stiffness components. In the presence of a force and torque sensor measuring \mathbf{f}_{ext} , impedance control can be implemented by enabling inertia shaping [33]. Casting the control law

$$\mathbf{f}_c = \Lambda(\mathbf{q})\alpha + \Gamma(\mathbf{q}, \dot{\mathbf{q}})\dot{\mathbf{x}} + \boldsymbol{\eta}(\mathbf{q}) + \mathbf{f}_{ext}, \quad (3)$$

into the dynamic model in (1) results in $\ddot{\mathbf{x}} = \alpha$, α being the control input denoting acceleration with respect to the base frame. In task space IC, the objective is to maintain a dynamics relationship (4) between the external force, \mathbf{f}_{ext} , and the error in position $\delta\mathbf{x} = \mathbf{x}^r - \mathbf{x}$, velocity $\delta\dot{\mathbf{x}} = \dot{\mathbf{x}}^r - \dot{\mathbf{x}}$ and acceleration $\delta\ddot{\mathbf{x}} = \ddot{\mathbf{x}}^r - \ddot{\mathbf{x}}$. This dynamic relationship that governs the interaction is modeled as a mass-spring-damper system as follows

$$\mathbf{M}\delta\ddot{\mathbf{x}} + \mathbf{D}\delta\dot{\mathbf{x}} + \mathbf{K}\delta\mathbf{x} = \mathbf{f}_{ext}, \quad (4)$$

where \mathbf{M} , \mathbf{D} and \mathbf{K} are Symmetric Positive Definite (SPD) matrices, adjustable impedance parameters, representing inertia, damping and stiffness terms, respectively. This desired dynamic behaviour (4) can be achieved using the following control law,

$$\alpha = \ddot{\mathbf{x}} + \mathbf{M}^{-1}(\mathbf{D}\delta\dot{\mathbf{x}} + \mathbf{K}\delta\mathbf{x} - \mathbf{f}_{\text{ext}}). \quad (5)$$

With no external force acting on the manipulator, under this control scheme, the end-effector will asymptotically follow the desired trajectory. In the presence of external forces, the compliant behavior of the end-effector is described by (4).

3.3. Probabilistic Ensemble NN (PE)

PE [21] is a NN based model approach capable of learning uncertainty-aware NN dynamics models including both aleatoric and epistemic uncertainties. The output neurons of the probabilistic NN parameterize a probability distribution function, which can capture the aleatoric uncertainty of the model. Using multiple such networks in an ensemble can capture the epistemic uncertainty. The predictive PE model trained with negative log prediction probability as a loss function can be defined to output a Gaussian distribution with diagonal covariance [21].

3.4. CEM based MPC

The CEM [34] offers a gradient free optimization scheme, coupling it with an MPC allows us to optimize an action sequence using the learned model. CEM samples multiple action sequences from a time-evolving distribution which is usually modeled as a Gaussian distribution $u_{t:t+H} \sim \mathcal{N}(\mu_{t:t+H}, \text{diag}(\sigma_{t:t+H}^2))$, where these action sequences are evaluated on the learned dynamical model with respect to a cost function. The sampling distribution, $\mu_{t:t+H}, \sigma_{t:t+H}^2$ is then updated based on best N trajectories. Safety can be directly incorporated into CEM-based optimization by sorting the samples based on constraint satisfaction values [35].

4. Deep Model Predictive Variable Impedance Control (MPVIC) Framework

The deep MPVIC framework is formulated to optimize a VIC utilizing a learned PE based Cartesian impedance model of the robot manipulator within a CEM based MPC.

4.1. Learning Cartesian Impedance Model

A Cartesian impedance model of the robot manipulator system controlled using a VIC is learned as a PE model in an MBRL setting alternating between model learning and CEM based exploration strategy. To learn a generalized model, an exploration strategy is designed to minimize the epistemic uncertainty of the model across the entire state space. The exploration strategy chooses the actions which maximized the epistemic uncertainty estimate from PE. Given a PE model \tilde{f} of B bootstrap models \tilde{f}_b , the uncertainty of the model prediction at current state can be estimated by calculating the model variance, $\rho = \sigma^2$, given by

$$\rho(s, u) = \frac{1}{B-1} \sum_{b=1}^B \left(\tilde{f}_b(s, u) - \overline{\tilde{f}(s, u)} \right)^2. \quad (6)$$

The designed exploration strategy will excite the system in areas in its state space where the model is more uncertain, thereby maximizing the information gain during exploration. This exploration strategy enables learning a generalized model in a sample-efficient way. A CEM-based MPC is used for exploration while the cost is defined to maximize the variance of the outputs from all the individual NN models in the PE, $C = \sigma^2(s_t, u_t)$. The model learning approach is summarized in Algorithm 1. Learning a model with low epistemic uncertainty over the entire state-space facilitates reusing the model for different tasks.

A free-space unconstrained manipulation task where the robot has to interact with its external environment can be described by a scenario where a robot in its current state s_t under the influence of an external force or sensed force f_t provided with a goal state s_t^* and a control input u_t transitions to the next state s_{t+1} . For a robot manipulator controlled by a VIC the stiffness matrix \mathbf{K} can be considered as the control input u_t , where the damping matrix $\mathbf{D} = 2\sqrt{\mathbf{K}}$. The dynamics model shown in Fig. 1 represents a generalized Cartesian behaviour of a unconstrained end-effector of a robot manipulator controlled by a VIC.

4.2. Impedance Adaptation

The compliant behavior of the robot end-effector can be optimized by designing a suitable impedance adaptation strategy. The Cartesian impedance model of the robotic system can be utilized in a MPC framework to adapt the impedance parameters of the VIC by designing a suitable optimization objective as shown in Fig. 1. At every time-step, an MPC with a horizon length of n , samples the current state and optimizes a control

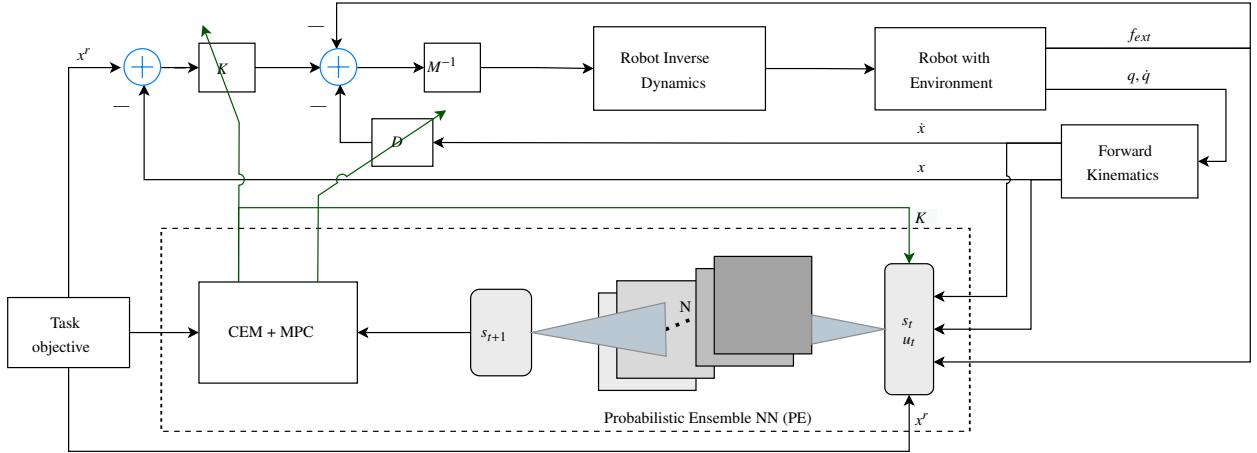


Figure 1: Block diagram of the deep MPVIC with PE Cartesian impedance model.

Algorithm 1 Learning a generalized Cartesian impedance model

Initialize dynamics model \tilde{f} .
 Populate dataset \mathcal{D} using random controller for n initial trials.
for $k \leftarrow 1$ to K Trials **do**
 Train dynamics model \tilde{f} on \mathcal{D} .
for $t \leftarrow 1$ to $TaskHorizon$ **do**
for Actions $u_{t:t+T} \sim CEM(\cdot)$, 1 to CEM Iterations **do**
 Evaluate and sort the actions by maximizing the uncertainty estimate in (6).
end
 Execute first action u_t^* from optimal action sequence $u_{t:t+T}^*$.
 Record outcome: $\mathcal{D} \leftarrow \mathcal{D} \cup (s_t, u_t, s_{t+1})$.
end
end

Algorithm 2 deep MPVIC

Given a cost function C and a PE dynamics model \tilde{f} .
MPC based optimization
for $t \leftarrow 1$ to $TaskHorizon$ **do**
CEM-based optimization
for $i \leftarrow 1$ to CEM Iterations **do**
 Generate N samples .
 Sample N stiffness profiles $K_{t:t+T} \sim CEM(\cdot)$.
 Evaluate samples .
 Calculate C for all $K_{t:t+T}$ on \tilde{f} with actions $[K_{t:t+T}, f_t, s_t]$ using trajectory sampling [21].
 Sort stiffness profiles K based on C .
 Update $CEM(\cdot)$ distribution .
 Choose optimal K^* where C is minimum .
end
 Adapt the impedance parameters of VIC .
 Execute first action K_t^* from optimal action sequence $K_{t:t+T}^*$.
end

trajectory $u_{t:t+n}$ for n future time-steps and applies the first control input, u_t , to the system. A gradient-free optimization method, CEM is used in an MPC setting to optimize the controller over the PE model. The proposed deep MPVIC approach utilizing PE models is described in Algorithm 2. The objective of the impedance adaptation strategy is to achieve the manipulation task requirement while executing a desired level of compliance. A cost function describing the task objective and the compliance objective is designed for the CEM-based MPC as,

$$C(s_t, u_t) = \delta s_t^T \mathbf{Q}_t \delta s_t + \lambda(K_t)^T \mathbf{R}_t \lambda(K_t), \quad (7)$$

where $\lambda(K_t)$ are the eigenvalues of the stiffness matrix represented in a vector form, $\delta s_t = s_t^r - s_t$ and \mathbf{Q}_t and \mathbf{R}_t are diagonal gain matrices for task and compliance components respectively. These gain matrices can be either constant or can be a function of the robots states. The MPC output behaviour will be tightly coupled with the gain matrices. In case of reference tracking tasks we chose \mathbf{Q}_t to be a linear function of $\|\delta s_t\|$ so that MPC will penalize larger deviations from target more than small deviations.

5. Experiments and Evaluation

For evaluation we consider only the stiffness adaptation along the position of the robot manipulator while keeping the stiffness values along orientations constant. However, before evaluation, we first need to learn the Cartesian impedance model of the robot manipulator. To do so, a free-space goal reaching task with random external force is used to train PE model with ensembles of 5 NNs with 3 hidden layers, each with 256 neurons. Its state space is chosen as $s = [x, y, z, \dot{x}, \dot{y}, \dot{z}]$, while the sensed external forces are denoted as $f = [f_{ext}^x, f_{ext}^y, f_{ext}^z]$. $s^r = [x^r, y^r, z^r]$ represents the target positions in x, y and z directions, \mathbf{K} denotes the Cartesian stiffness matrix. The damping matrix is chosen as $\mathbf{D} = 2\sqrt{\mathbf{K}}$. CEM is used to optimize the exploration strategy based on uncertainty maximization. The control frequency for low-level VIC is set at 100Hz. For learning the model, the robot manipulator is excited at every time-step with random $f_{ext} \in [-20, 20]$ N and s_t^r , where $x_t^r, y_t^r, z_t^r \in [-10, 10]$ cm. The gain matrices \mathbf{Q} and \mathbf{R} are kept constants for a specific task. However, while transferring to a new task, they can be scaled using a scalar values α_Q and α_R as $\mathbf{Q}_{\text{new}} = \mathbf{Q} * \alpha_Q$ and $\mathbf{R}_{\text{new}} = \mathbf{R} * \alpha_R$ respectively to trade off between compliance and accuracy depending on the task requirement. The model was trained for 100 000 time-steps

with a control-frequency of 10Hz which is equivalent to 2.77 h of real-world training. For experiments, a prior model estimated in simulations over 50 000 time-steps is fine-tuned in the experimental scenario instead of learning from scratch. The model was fine-tuned for 10 000 time-steps which is equivalent to 33.33 min of real-world training. Similar to in simulations random external forces were manually applied to the robot end-effector.

After learning the Cartesian impedance model of the manipulator, and to evaluate the effectiveness of the proposed deep MPVIC, three different simulation tasks and two experimental tasks using a Franka Emika Panda manipulator are designed. The tasks demanding real time stiffness adaptation are chosen for evaluating the stiffness profile generated by the deep MPVIC controller. The three different simulation tasks are modeled in the MuJoCo physics simulation framework [36], see Fig. 2 (a), (b) and (c). The two real experimental scenarios are shown in Fig. 2 (d) and (e). *In simulations*, the population size for CEM is chosen as 200 and elite size of 40 and learning rate of 0.1 and number of CEM iterations as 10. The MPC planning horizon is set as 5. *While for the real experiments*, the control frequency is set as 5Hz. The CEM is chosen as 64 and elite size of 32 and learning rate of 0.5, number of CEM iterations as 5 and MPC planning horizon is set as 5. In all the simulations and experiments the model described here is used without any further fine-tuning. Here we consider only fixed goal states, therefore s_t^r is a constant value, s^r for all timesteps.

5.1. Simulations

Cartesian compliant behavior : In this task (Fig. 2 (a)), the robot is expected to behave highly compliant to hold its pose thereby spending minimum energy at rest. Upon applying an external force to the robot-end-effector it is expected to counter the force by adapting its stiffness such that it achieves a new rest position close to the initial position. This task is ideal to test the impedance adaptation strategy as it needs to increase the stiffness in case of large external forces and larger deviation from its initial position. Two scenarios with different compliance behavior are evaluated here by changing the compliance maximization component in the cost function. The results in Fig. 3 (a) and (b) show that the robot which is highly compliant at rest adapts the stiffness in response to the external forces and deviation from the rest position. Having a higher value of compliance factor α_R allows for larger deviations from the initial position when applied with an external force while having a lower α_R limits this deviation. It is also

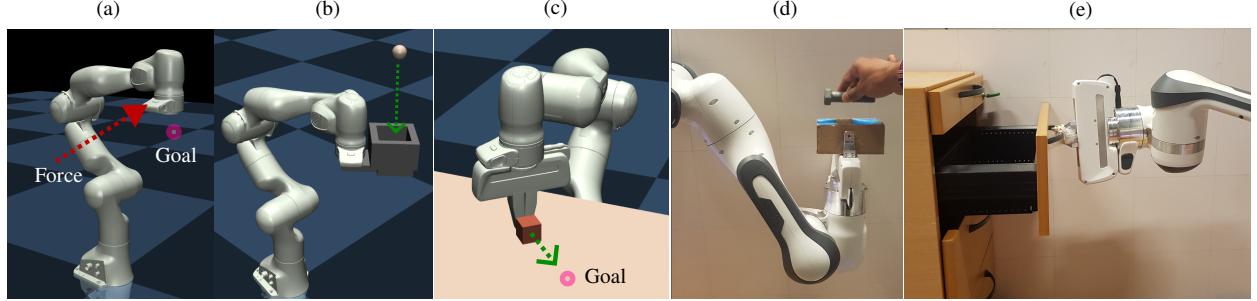


Figure 2: Three simulation tasks, (a) Cartesian compliance task: the robot manipulator end-effector should hold its pose in the Cartesian space compliantly while reacting to the external forces acting on it. (b) Reacting to falling object: The robot manipulator with cup end-effector should hold a Cartesian position while smoothly catching a ball of weight 0.5g falling into the cup. (c) Pushing task: A robot manipulator with a gripper end-effector should push an object over a rigid surface with friction to a target position. Two experimental tasks, (d) Reacting to falling objects: robot end-effector is fitted with a tray, where objects of different weights are dropped into the tray at regular intervals. (e) Drawer opening task: Robot manipulator opening a table drawer.

noted that higher α_R results in noisy stiffness adaption behavior as larger Δ_{pos} (the deviation from the desired pose) creates larger gradients in the cost function.

Reacting to falling object: In this task (Fig. 2 (b)), a robot with a cup end-effector that is highly compliant at rest position is expected to react optimally to objects falling into the cup end-effector. Four different objects are dropped from different heights to the cup in different trials resulting in large variations in the impact force. The desired behavior of the robot is not to deviate largely from the rest position while reacting to the falling objects while not being very stiff. The resulting robot behavior is shown in Fig. 3 (c), which shows a sudden increase in K_z upon a spike in f_{ext} in z direction induced by the impact of the falling object. The robot increases its stiffness every time a new object is falling to the cup and maintains a higher level of stiffness during the later phases to hold the robot back to a new rest position.

Pushing task: In this task (Fig. 2 (c)), the robot is expected to push a cube-shaped object to a target position on a surface with friction. Here, K_z is set constant as 1000 as the robot is not expected to move in z direction. Stiffness in x and y directions are optimized to push the object to the target while being compliant and stiff only when necessary. The results in Fig. 3 (d) show that the stiffness is increased to its upper limit in the pushing directions initially to overcome the static friction. Upon reaching close to the target position the stiffness is decreased to be more compliant.

5.2. Comparison with Model-free/based RL:

The deep MPVIC is compared with RL based VILC approaches for their transferability between tasks which

is the main contribution of this work while also comparing their performance. Specifically, in these comparisons, we utilize the PE model trained with curiosity driven exploration with our deep MPVIC for different tasks without retraining or fine tuning the model. This enables the deep MPVIC to generalize over multiple tasks where the RL approaches are task specific.

Model-free RL approaches have been successfully used in VILC for robotic manipulation tasks in multiple previous works [10, 12, 13]. Out of which we have chosen the off-policy RL algorithm Soft Actor Critic (SAC) because of its high sample efficiency. All the three simulation tasks shown in Fig. 2 are trained using SAC implementation from *stable-baselines* [37] for 500 000 time-steps.

In addition, we compare our approach with the MBRL approach PETS [21]. In case of PETS, the simulation tasks are trained for 100 000 time-steps. The PETS policies were trained with same CEM parameters and cost functions used for the corresponding tasks in our deep MPVIC. The performance and the transferability of the learned policies in both of these approaches were compared with our MPVIC approach in Fig. 6.

Performance: The resulting robot behaviour on applying the learned model-free RL and PETS policies on the three simulation tasks are shown in Fig. 4 and Fig. 5 respectively. The reward obtained while applying the learned policies are shown in Fig. 6. Deep MPVIC performed better on *task a*, the performance was similar on *task b* and model-free RL and PETS policies performed better on the *task c* by minimizing the stiffness more effectively.

Task transferability: In order to evaluate how effi-

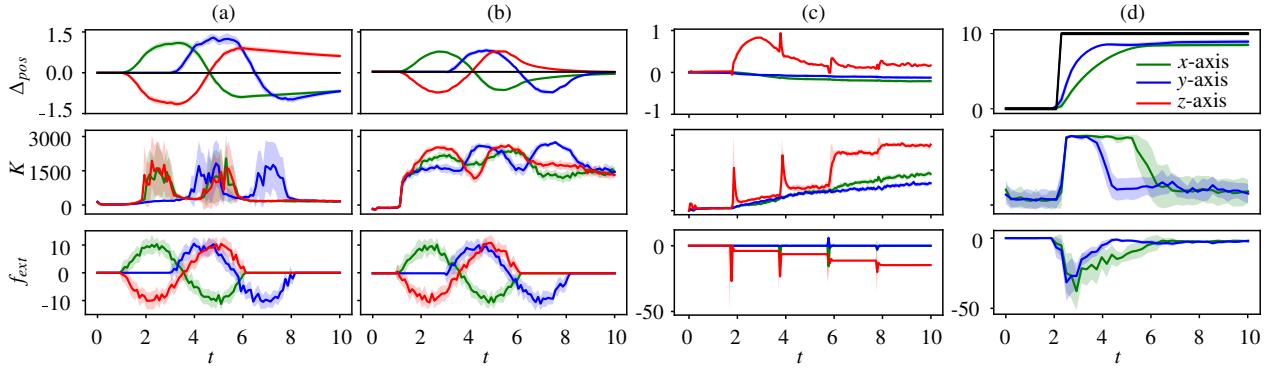


Figure 3: Simulations: (a) and (b), (Cartesian compliance behaviour), results from 20 trials where a sinusoidal force profile with amplitude of 10 N with a random noise of (± 5) N is applied to the robot end-effector. (a) High compliant behaviour optimized using a cost function with larger compliance factor $\alpha_R = 0.1$, (b) Low compliant behaviour optimized using a cost function with $\alpha_R = 0.01$. (c), (Reacting to falling objects) The robot is initialized at a rest position being very compliant with $K \rightarrow 0$. Objects of different weights are dropped at regular intervals of 2s, from random heights between $(0.5 - 1.0)$ m. Results shown here are over 10 such random trials with $\alpha_R = 0.1$. (d), (Pushing task) Robot with a gripper end-effector is at rest with $K \rightarrow 0$. At $t = 1$ s, it is commanded to push an object to a target position given by Δ_{pos} of 10 cm in x and y direction on a surface. The results shown here are over 10 trials with objects of random weights between $(0.5 - 3.0)$ kg and $\alpha_R = 0.1$.

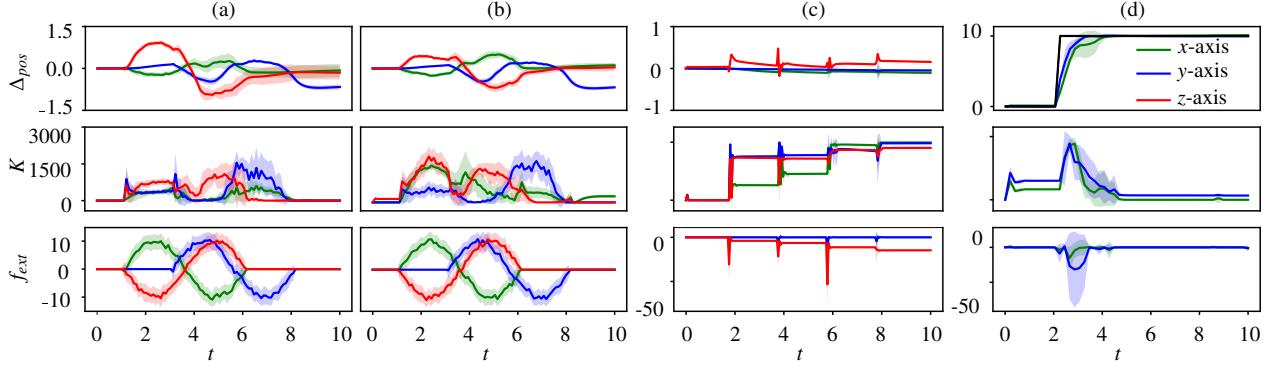


Figure 4: Corresponding results from Model-free RL policy for the simulation tasks shown in Fig. 3

ciently the policy learned on a task can be transferred to another task, the model-free RL and PETS policies learned on the simulation *task a* was tested on *task b* and *task c* without retraining the policy/model. The performance of the transferred model-free RL and PETS policies on *task b* and *c* were compared with the corresponding performance of deep MPVIC using the PE model trained on *task a*. Figure 6-right illustrates the transferability of our deep MPVIC in comparison with RL-based approaches, where deep MPVIC demonstrates the major advantage (green bars). Further, the model-free RL and PETS policies have been retrained to achieve similar performance as our deep MPVIC. A comparison on the additional *data samples/time steps* required for retraining the models/policies for the tasks is shown in Table 2. The number of additional training samples required is correlated with the computational time. While RL approaches demanding additional computational/training time to perform a new task, the proposed

deep MPVIC can be deployed without any additional computational effort.

Table 2: Comparison on transferability between tasks

	Training samples ($\times 10^5$)		
	Transferability to		
	<i>Task a</i>	<i>Task b</i>	<i>Task c</i>
Model-free RL	50	38.6	27.95
PETS	10	3.2	3.9
Our MPVIC	10	0	0

5.3. Real-World Experiments

Reacting to falling objects: The experimental setup is shown in Fig. 2 (d) where the robot end-effector is fitted with a tray and four objects of different weights are added to the tray at regular intervals. The optimization objective here is similar to the simulation task (c),

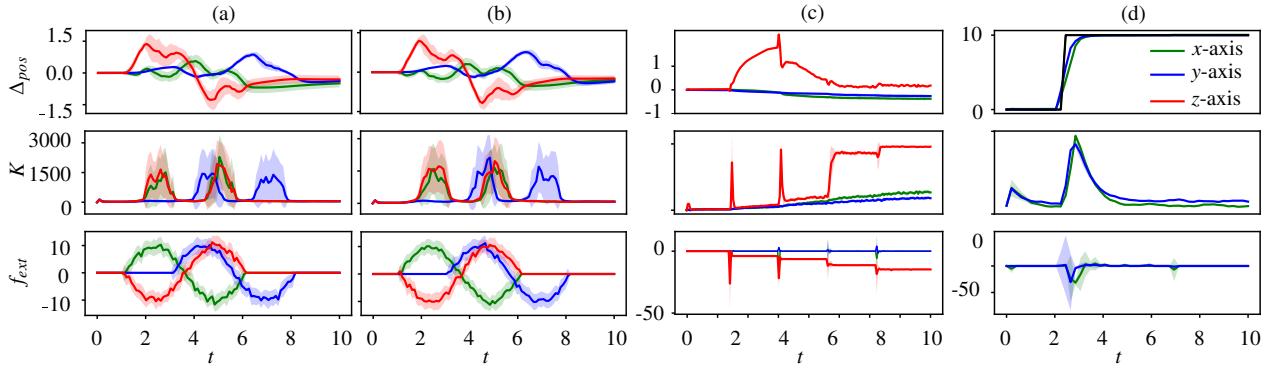


Figure 5: Corresponding results from PETS policy for the simulation tasks shown in Fig. 3

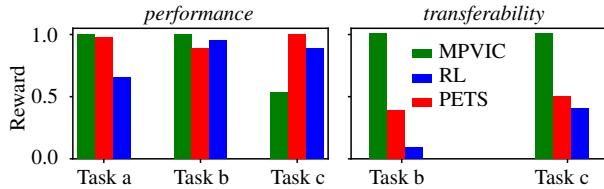


Figure 6: (left) Comparing the normalized value of reward (mean value over 20 trials) obtained using Model-free RL, PETS, and our MPVIC framework on all the three simulation tasks. (right) Comparing the transferability of the Model-free RL and PETS based policy with our MPVIC framework based on normalized value of the mean reward over 20 trials.

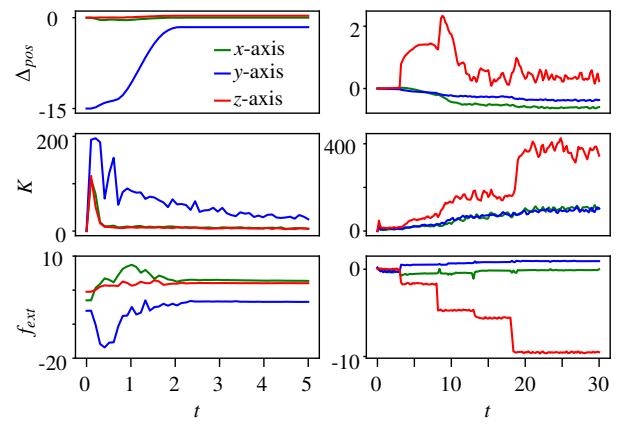


Figure 7: Experiments: (left) task (e), Robot manipulator opening a table drawer. (right) task (d) The robot manipulator with a tray holding its pose while objects are dropped to the tray.

the robot is expected to hold a pose while being highly compliant and becoming stiffer with extra weights being introduced to the tray. In Fig. 7 (right-column) the robot with a very low initial stiffness increases the stiffness every instant a new object is introduced to the tray in order to maintain it at the desired pose.

Opening a drawer: The pulling task, which is similar but in the opposite direction of the pushing task. The experimental setup is shown in Fig. 2 (e) where the robot is opening a table drawer to a desired position (15 cm in x direction) in the Cartesian space. The results shown in Fig. 7 (left-column), shows the impedance adaptation behaviour similar to the pushing task in simulation where the robot increases its stiffness initially to overcome the inertia of the drawer and then decreased once the drawer starts to move closer to the desired position.

6. Discussion and Limitations

Discussion: The deep MPVIC-based approach presented in the work is evaluated over different tasks in Section 5 for optimizing impedance adaptation strate-

gies. The objective in all experiments has been consistent in having high stiffness values for the VIC only when the task objective demands that. This objective is motivated by human manipulation behavior and can increase the dexterity of the robot while encouraging energy efficient and safe behaviours. In all the evaluation scenarios, both in simulation and experiments, the stiffness adaptation guarantee a high level of compliance unless there is a large deviation from the target position or an external force is applied to it. The modeling approach using PE combined with uncertainty targeted exploration has been found to be very useful in learning a generalized unconstrained Cartesian impedance model of the robot. In addition, combining it with MPC based optimization has enabled to solve different manipulation tasks demanding stiffness adaptation. The proposed deep MPVIC approach succeeds in generalising a single model to solve multiple manipulation tasks. The ver-

satility of the impedance adaptation strategy is evident in the scenarios of impact force from falling objects, overcoming the inertia of the objects in the pushing and drawer opening tasks respectively. While a majority of robot manipulation tasks rely on trajectory planning and tracking, our approach is not straightforward in solving complex manipulation problems. Nevertheless, it can be combined with a high-level planning approach where the low-level VIC will modify the given trajectory to ensure compliant behavior. Incorporating such compliant behaviors could improve the manipulation skills, especially in tasks involving contacts.

The deep MPVIC framework was compared with model-free and model-based RL approaches utilized successfully in various previous works [10, 12, 13, 18] to solve complex manipulation tasks. The results show that deep MPVIC framework is able to achieve similar performance to model-free and model-based RL approaches while being highly sample efficient and able to seamlessly transfer the controller between different tasks without any further training of the model. Whereas in model-free and model-based RL, transferring policy between different tasks demand relearning the policy on the new task or extensive fine tuning of the existing policy. PETS shows better task transferability compared to model-free RL, this can be justified by the use of a model in PETS for impedance optimization even though it is not a generalized model as in deep MPVIC. RL has the potential to solve very complex tasks at the expense of high sample complexity. It would be ideal to combine this aspect of RL with sample efficiency and easy transferability of the learned controller between tasks as in our deep MPVIC framework. Further extending the model-based RL approaches for VILC could be a promising approach in this direction.

Limitations: Applying our approach to tasks with non-continuous contacts is not possible as the model is not aware of the contact dynamics, which could lead to unstable behavior. Detecting contact discontinuities and switching to a different contact re-establish policy could be a solution to this issue. Whereas a more general approach could be to learn a model aware of contact constraints, incorporating such constraints into the model state-space is challenging. In the future work we will explore ways to sufficiently incorporate contact constraints to the model to aid faster fine-tuning of the VILC for different manipulation tasks. In addition, there are limitations inherited from applying CEM to a real robotic system because of the high computation time, where the trade-off is between optimization performance and the control frequency. Eventhough VIC can be operated generally at lower control frequencies,

in tasks with complex contact dynamics this might not be sufficient. The level of impedance adaptation or the compliance behaviour can be adjusted by tuning the Q and R parameters in the cost function (7). However, it is not obvious how to find optimal values for these parameters.

7. Conclusion

In this work, we presented a deep MPVIC approach for compliant manipulation skills for a robotic manipulator by optimizing the impedance parameters. By utilizing PE, a Cartesian impedance model of the robot is learned using an exploration strategy maximizing the information gain. The PE dynamic model is coupled with a CEM-based MPC to optimize impedance parameters of a low-level VIC. We identified an impedance optimization objective-based human manipulation skill and replicated it on a robot manipulator for simplified scenarios in simulations and experiments. The deep MPVIC was compared with model-free and model-based RL approaches in VILC. The approach proved experimentally to be beneficial for solving multiple tasks without any need to relearn the model or policy as opposed to other VILC approaches. In the future work, we aim to extend this approach to scenarios with constraints, such as in-contact interaction tasks.

Acknowledgments

This work was part of the project “Dynamic Robot Interaction and Motion Compensation” funded by the Research Council of Norway under contract number 270941.

References

- [1] E. Bizzi, N. Accornero, W. Chapple, N. Hogan, Posture control and trajectory formation during arm movement, *Journal of Neuroscience* 4 (1984) 2738–2744.
- [2] N. Hogan, An organizing principle for a class of voluntary movements, *Journal of neuroscience* 4 (1984) 2745–2754.
- [3] S. D. Kennedy, A. B. Schwartz, Stiffness as a control factor for object manipulation, *Journal of Neurophysiology* 122 (2019) 707–720.
- [4] A. V. Hill, The series elastic component of muscle, *Proceedings of the Royal Society of London. Series B, Biological Sciences* (1950) 273–280.
- [5] N. Hogan, Impedance control: An approach to manipulation, in: 1984 American control conference, IEEE, 1984, pp. 304–313.
- [6] O. Khatib, A unified approach for motion and force control of robot manipulators: The operational space formulation, *IEEE Journal on Robotics and Automation* 3 (1987) 43–53.

- [7] R. Ikeura, H. Inooka, Variable impedance control of a robot for cooperation with a human, in: Proceedings of 1995 IEEE International Conference on Robotics and Automation, volume 3, IEEE, 1995, pp. 3097–3102.
- [8] E. Caldarelli, A. Colomé, C. Torras, Perturbation-based stiffness inference in variable impedance control, *IEEE Robotics and Automation Letters* 7 (2022) 8823–8830.
- [9] F. J. Abu-Dakka, M. Saveriano, Variable impedance control and learning—a review, *Frontiers in Robotics and AI* 7 (2020).
- [10] R. Martín-Martín, M. A. Lee, R. Gardner, S. Savarese, J. Bohg, A. Garg, Variable impedance control in end-effector space: An action space for reinforcement learning in contact-rich tasks, in: 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, 2019, pp. 1010–1017.
- [11] C. C. Beltran-Hernandez, D. Petit, I. G. Ramirez-Alpizar, K. Harada, Variable compliance control for robotic peg-in-hole assembly: A deep-reinforcement-learning approach, *Applied Sciences* 10 (2020) 6923.
- [12] M. Bogdanovic, M. Khadiv, L. Righetti, Learning variable impedance control for contact sensitive tasks, *IEEE Robotics and Automation Letters* 5 (2020) 6129–6136.
- [13] P. Varin, L. Grossman, S. Kuindersma, A comparison of action spaces for learning manipulation tasks, in: 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, 2019, pp. 6015–6021.
- [14] C. C. Beltran-Hernandez, D. Petit, I. G. Ramirez-Alpizar, T. Nishi, S. Kikuchi, T. Matsubara, K. Harada, Learning force control for contact-rich manipulation tasks with rigid position-controlled robots, *IEEE Robotics and Automation Letters* 5 (2020) 5709–5716.
- [15] M. Kim, S. Niekum, A. D. Deshpande, Scape: Learning stiffness control from augmented position control experiences, in: Conference on Robot Learning, PMLR, 2022, pp. 1512–1521.
- [16] J. Buchli, F. Stulp, E. Theodorou, S. Schaal, Learning variable impedance control, *The International Journal of Robotics Research* 30 (2011) 820–833.
- [17] C. Li, Z. Zhang, G. Xia, X. Xie, Q. Zhu, Efficient force control learning system for industrial robots based on variable impedance control, *Sensors* 18 (2018) 2539.
- [18] L. Roveda, J. Maskani, P. Franceschi, A. Abdi, F. Braghin, L. M. Tosatti, N. Pedrocchi, Model-based reinforcement learning variable impedance control for human-robot collaboration, *Journal of Intelligent & Robotic Systems* 100 (2020) 417–433.
- [19] A. S. Anand, M. H. Myrestrand, J. T. Gravdahl, Evaluation of variable impedance-and hybrid force/motioncontrollers for learning force tracking skills, in: 2022 IEEE/SICE International Symposium on System Integration (SII), IEEE, 2022, pp. 83–89.
- [20] E. F. Camacho, C. B. Alba, *Model predictive control*, Springer science & business media, 2013.
- [21] K. Chua, R. Calandra, R. McAllister, S. Levine, Deep reinforcement learning in a handful of trials using probabilistic dynamics models, *arXiv preprint arXiv:1805.12114* (2018).
- [22] T. Wang, X. Bao, I. Clavera, J. Hoang, Y. Wen, E. Langlois, S. Zhang, G. Zhang, P. Abbeel, J. Ba, Benchmarking model-based reinforcement learning, *arXiv preprint arXiv:1907.02057* (2019).
- [23] M. V. Minniti, R. Grandia, K. Fäh, F. Farshidian, M. Hutter, Model predictive robot-environment interaction control for mobile manipulation tasks, in: 2021 IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2021, pp. 1651–1657.
- [24] T. Gold, A. Völz, K. Graichen, Model predictive interaction control for robotic manipulation tasks, *IEEE Transactions on Robotics* (2022).
- [25] K. Haninger, C. Hegeler, L. Peternel, Model predictive control with gaussian processes for flexible multi-modal physical human robot interaction, in: 2022 International Conference on Robotics and Automation (ICRA), IEEE, 2022, pp. 6948–6955.
- [26] P. Shyam, W. Jaśkowski, F. Gomez, Model-based active exploration, in: International conference on machine learning, PMLR, 2019, pp. 5779–5788.
- [27] R. Sekar, O. Rybkin, K. Daniilidis, P. Abbeel, D. Hafner, D. Pathak, Planning to explore via self-supervised world models, in: International Conference on Machine Learning, PMLR, 2020, pp. 8583–8592.
- [28] Y. Yao, L. Xiao, Z. An, W. Zhang, D. Luo, Sample efficient reinforcement learning via model-ensemble exploration and exploitation, in: 2021 IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2021, pp. 4202–4208.
- [29] D. Pathak, D. Gandhi, A. Gupta, Self-supervised exploration via disagreement, in: International conference on machine learning, PMLR, 2019, pp. 5062–5071.
- [30] D. V. Lindley, On a measure of the information provided by an experiment, *The Annals of Mathematical Statistics* 27 (1956) 986–1005.
- [31] D. Pathak, P. Agrawal, A. A. Efros, T. Darrell, Curiosity-driven exploration by self-supervised prediction, in: International conference on machine learning, PMLR, 2017, pp. 2778–2787.
- [32] H.-P. Huang, S.-S. Chen, Compliant motion control of robots by using variable impedance, *The International Journal of Advanced Manufacturing Technology* 7 (1992) 322–332.
- [33] L. Villani, J. De Schutter, Force control, in: *Springer handbook of robotics*, Springer, 2016, pp. 195–220.
- [34] Z. I. Botev, D. P. Kroese, R. Y. Rubinstein, P. L’Ecuyer, The cross-entropy method for optimization, in: *Handbook of statistics*, volume 31, Elsevier, 2013, pp. 35–59.
- [35] M. Wen, U. Topcu, Constrained cross-entropy method for safe reinforcement learning, *Advances in Neural Information Processing Systems* 31 (2018).
- [36] E. Todorov, T. Erez, Y. Tassa, Mujoco: A physics engine for model-based control, in: 2012 IEEE/RSJ international conference on intelligent robots and systems, IEEE, 2012, pp. 5026–5033.
- [37] A. Hill, A. Raffin, M. Ernestus, A. Gleave, A. Kanervisto, R. Traore, P. Dhariwal, C. Hesse, O. Klimov, A. Nichol, M. Plappert, A. Radford, J. Schulman, S. Sidor, Y. Wu, Stable baselines, <https://github.com/hill-a/stable-baselines>, 2018.