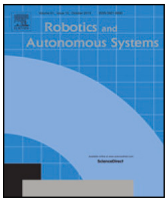
Robotics and Autonomous Systems 170 (2023) 104531

Contents lists available at ScienceDirect

Robotics and Autonomous Systems

journal homepage: www.elsevier.com/locate/robot



Model-based variable impedance learning control for robotic manipulation

Akhil S. Anand [a,*], Jan Tommy Gravdahl [a], Fares J. Abu-Dakka [b]

[a] Department of Engineering Cybernetics at Norwegian University of Science and Technology (NTNU), Trondheim, Norway
[b] Munich Institute of Robotics and Machine Intelligence, Technical University of Munich, Munich, Germany

ARTICLE INFO

Keywords:
Variable impedance control
Model predictive control
Robot learning

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 for enabling real-world force interaction tasks with human-like dexterity. In this study, we introduce a novel approach, we call "deep Model Predictive Variable Impedance Controller (MPVIC)" for compliant robotic manipulation, which combines Variable Impedance Control with Model Predictive Control (MPC). The method involves learning a generalized Cartesian impedance model of a robot manipulator through an exploration strategy to maximize information gain. Within the MPC framework, this learned model is utilized to adapt the impedance parameters of a low-level variable impedance controller, thereby achieving the desired compliance behavior for various manipulation tasks without requiring retraining or finetuning. We assess the efficacy of the proposed deep MPVIC approach using a Franka Emika Panda robotic manipulator in simulations and real-world experiments involving diverse manipulation tasks. Comparative evaluations against model-free and model-based reinforcement learning approaches in variable impedance control are conducted, considering aspects such as transferability between tasks and performance. The results demonstrate the effectiveness and potential of the presented approach for advancing robotic manipulation capabilities.

1. Introduction

Human interaction with the real world heavily relies on the ability to manipulate objects with remarkable dexterity; despite the limitation of low-frequency biological feedback loops. The precise motor control mechanisms responsible for such adept manipulation skills remain largely elusive. Nevertheless, research has suggested that the modulation of arm impedance plays a pivotal role in achieving these capabilities [1–3]. In contrast, robotic manipulators, benefiting from higher-frequency feedback control loops, have struggled to achieve comparable levels of dexterity in real-world applications. Traditionally, these applications have predominantly relied on trajectory planning and position control, which prove to be inadequate in terms of dexterity, safety, energy efficiency, and constrained interactions. Notably, human muscle actuators possess impedance properties, such as stiffness and damping [4], which can be adapted by the neural control to achieve various manipulation behaviors.

Drawing inspiration from the adaptability of human manipulation, Impedance Control (IC) for robot control, as introduced by Hogan in [5], seeks to establish a strong coupling between the manipulator's dynamics with its environment instead of treating it as an isolated system when designing control strategies. In contrast to conventional control approaches, IC aims to establish a dynamic relation between manipulator variables such as end-point positions and forces rather than controlling these variables independently. By adopting IC, it becomes possible to address position uncertainties effectively, thus mitigating the risk of encountering significant impact forces. This is achieved by allowing robots to adjust their motion or compliance based on force feedback [6]. In doing so, IC offers a viable and effective solution for managing positional uncertainties and promoting safer interactions with the environment.

IC naturally extendes to VIC, where the impedance parameters are varied during the task [7–9]. VIC has gained popularity in robotic research due to its ability to provide scalability for IC in handling complex robotic manipulation tasks. However, formulating variable impedance laws for complex tasks is nontrivial, and hand-designing them often proves impossible. To address this challenge, VILC emerges as an alternative approach that combines learning algorithms with VIC. In VILC, a learned policy is used to adapt the impedance gains in the VIC framework. Readers can refer to [10] for an in-depth review of diverse learning approaches applied to VIC.

RL is the most prominent approach in recent VILC research owing to its inherent flexibility and scalability. However, when applying RL to VILC, or robotics in general, significant drawbacks are evident, particularly in terms of date-efficiency and constraint guarantees. RL

[*] Corresponding author.
E-mail addresses: akhil.s.anand@ntnu.no (A.S. Anand), jan.tommy.gravdahl@ntnu.no (J.T. Gravdahl), fares.abu-dakka@tum.de (F.J. Abu-Dakka).

https://doi.org/10.1016/j.robot.2023.104531
Received 17 September 2022; Received in revised form 14 September 2023; Accepted 18 September 2023
Available online 25 September 2023

typically requires a substantial amount of data to learn effective control policies, which can be impractical in real-world robotic systems. Moreover, the control policies obtained through RL tend to be task or scenario-specific, making them challenging to transfer seamlessly to new tasks or scenarios. This lack of generalizability can hinder the broader application of RL-based VILC methods in diverse robotic manipulation contexts. As a result, researchers continue to explore ways to address these limitations and enhance the practicality and robustness of RL-based approaches in the field of VILC.

MPC provides a systematic framework for designing control systems by formulating them as optimization problems, utilizing a system model and an optimization objective [11]. MPC approaches are widely used in robotic control, especially when a reliable model of the system dynamics is available. By employing MPC, robotic control systems can predict future system behavior based on the model and optimize control inputs over a finite time horizon to achieve desired objectives. The MPC methodology facilitates the incorporation of constraints, adaptability, and precise tracking of desired trajectories, making it a valuable choice for a variety of robotic manipulation tasks where an accurate model is at hand.

When controlling a robot controlled with a Cartesian space VIC approach, it becomes feasible to learn a Cartesian impedance model of the robot and integrate it into an MPC framework for optimizing impedance profiles. By employing MPC-based VIC, a potential alternative to RL-based VILC methods arises, offering advantages in terms of data efficiency, transferability, and constraint satisfaction. However, the effectiveness of an MPC scheme heavily relies on the quality of the model employed. While various modeling approaches exist, deterministic Neural Network (NN) are well-suited for learning complex dynamics but may suffer from over-fitting and do not provide a quantification of uncertainties. In contrast, Gaussian Processes (GP) models can account for uncertainty and are used from impedance learning [12,13]. However, they face challenges in scaling with high-dimensional data, limiting their practicality in some cases, especially for robotic systems. Addressing these limitations, the PENN models introduced in [14] offer a promising solution. These models combine the strengths of NN and GP models, enabling the quantification of both aleatoric and epistemic uncertainties while being scalable. By incorporating PENN models into the MPC-based VIC framework, it becomes possible to enhance the robustness and reliability of impedance control, making it more suitable for complex robotic manipulation tasks with uncertain environments.

In this paper, we propose a novel approach called deep Model Predictive Variable Impedance Control (MPVIC) framework. Within this framework, we leverage a PENN based Cartesian impedance model of the robotic manipulator and combine it with a CEM-based MPC strategy. The objective is to achieve online adaptation of the impedance parameters while executing tasks that require VIC skills. With the deep MPVIC framework, we aim to facilitate the learning of effective impedance adaptation strategies for a wide range of robotic manipulation tasks by defining a suitable cost function. Primary contributions of our paper include:

- a novel VIC framework, we call it deep MPVIC that seamlessly integrates a CEM-based MPC with PENN dynamical model, allowing for real-time adaptation of impedance parameters, offering the following properties.
  - transferability: the key property of the deep MPVIC framework. It allows for the seamless transfer of the impedance adaptation strategy between various manipulation tasks without the need for relearning or fine-tuning.
  - data efficiency and scalability: The proposed framework is effective in learning VIC with high efficiency, requiring fewer data samples while it is scalable to complex manipulation tasks.
- an uncertainty-based exploration scheme is integrated into the proposed framework to facilitate learning a generalized Cartesian impedance model of the robot in a data-efficient manner.
- 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 introduces the necessary background knowledge, 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

In this section, we present a concise review of relevant related works that are pertinent to this paper, covering the area of VILC, MPC for VIC, and finally, on uncertainty targeted exploration.

VILC approaches: A wide variety of learning-based approaches have been integrated with VIC to develop diverse VILC methods [10]. Prominent examples of such learning-based approaches include Imitation Learning (IL), Iterative learning control (ILC), and RL. IL has been used in many recent VILC works [15–18]. IL-based VILC methods are generally some form of Learning from Demonstration (LfD) methods as they often rely on demonstrations to learn from [19]. IL can be useful in developing highly sample efficient VILC [10]. However such learning strategies can be biased to the demonstration which is often suboptimal and potentially limits the performance and generalization of the learned policies. IL is useful for tasks that are easy to demonstrate and which do not have a clear optimal way of execution, whereas RL is well suited for highly dynamic tasks, where there is a clear measure of the success of the task [20]. Optimizing variable impedance gains/parameters can be done using ILC where the robot improves its performance iteratively. ILC based methods have been used for VILC in a range of works [21–24]. The key difference between ILC and RL is that, in RL, the control law is derived by maximizing a reward function defined by the task requirements. One advantage of ILC compared to RL is its sample efficiency. But even when a model of the dynamics is not available, RL offers better performance and can be applied to a broader range of problems [25].

RL-based VILC: Recently, RL has been explored largely for VILC research. However, RL demands a large amount of data samples/interactions to obtain high performance. Refs. [26–30] 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. [31] 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 cannot be estimated directly from kinesthetic demonstrations used in [31]. Ref. [32] demonstrated model-free RL based VILC using Dynamic Movement Primitive (DMP) policy and Policy Improvement with Path Integrals (PI^2), which is sample efficient but fails to scale to complex policies. In comparison, 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 easily transfer 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 uses a generalized Cartesian impedance model of the robot with an MPC policy that can be used for multiple tasks by designing suitable cost functions.

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
[26,27,29,30]	Low	–	Model-free	Low	Force
[28,31]	Low	–	Model-free	Low	Position
[32]	High	–	Model-free	Low	Force
[33,35]	High	–	Model-based	High	Position
[34]	High	–	Model-based	High	force
Our MPVIC	High	✓	Model-based	High	force

Model-based Reinforcement Learning (MBRL) for VILC: Alternatively, MBRL approaches offer a sample efficient framework leveraging on the model. In [13,33] MBRL is used for learning position-based VIC on industrial robots using GP models. Ref. [34] used a similar approach for force-based VIC for contact-sensitive tasks. All of these approaches utilize GP models and the PILCO algorithm limiting its use to less complex tasks with smooth dynamics and relatively simple policies and reward structure. Considering such limitations, in this paper, we aim to go beyond such classical approaches to develop a scalable VILC approach. In [35], a PETS approach is utilized for learning position-based VIC strategy for Human-Robot Collaboration (HRC) tasks. This was extended in [36] combining CEM with Q-learning and enhanced with the stability guarantees by means of Lyapunov constraints. Similar to RL approaches referenced earlier, all of these MBRL-based approaches are task-specific and generally lack the performance of model-free RL approaches [37]. 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 generalized Cartesian impedance model with an MPC scheme. A comparison between existing RL-based VILC approaches is summarized in Table 1.

MPC for VIC: In literature, MPC is used in robotic interaction control for manipulations tasks [38,39], 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 [38] where it can be used as a low-level optimizer to solve additional constraints. Haningeret al. [40] 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 [40]. Unlike [40], we optimize the impedance parameters for a force-based VIC in our deep MPVIC framework using PENN to model the Cartesian impedance behavior of the robot manipulator.

Uncertainty-based exploration: For efficient model learning in terms of sample efficiency, uncertainty-based exploration with ensembles of NNs has been proposed in prior works [41–44]. The basis for uncertainty-based exploration for model learning is derived from the expected information gain formulation in [45]. In [46] 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

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

where $\dot{\mathbf{x}}, \ddot{\mathbf{x}}$ are velocity and acceleration of the robot end-effector in task space, $\mathbf{f}_{\mathbf{c}}$ is the task space control force, $\mathbf{f}_{\mathbf{ext}}$ is the external force, $\boldsymbol{\Gamma}(\mathbf{q},\dot{\mathbf{q}}) \in \mathbb{R}^{6\times6}$ is a matrix representing the centrifugal and Coriolis effects, and $\boldsymbol{\eta}(\mathbf{q}) = \mathbf{J}^{-\mathbf{T}}\mathbf{g}(\mathbf{q}) \in \mathbb{R}^{6\times1}$ is the gravitational force, where $\mathbf{g}(\mathbf{q})$ is the joint space forces and torques. The Cartesian inertia matrix is denoted as $\boldsymbol{\Lambda}(\mathbf{q}) = (\mathbf{J}\mathbf{H}(\mathbf{q})^{-\mathbf{1}}\mathbf{J}^{\mathbf{T}})^{-\mathbf{1}} \in \mathbb{R}^{6\times6}$, 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 $\boldsymbol{\Gamma}(\mathbf{q},\dot{\mathbf{q}}) = \mathbf{J}^{-T}\mathbf{V}(\mathbf{q},\dot{\mathbf{q}})\mathbf{J}^{-1} - \boldsymbol{\Lambda}(\mathbf{q})\dot{\mathbf{J}}\mathbf{J}^{-1}$.

3.2. Variable impedance control

VIC is designed to achieve force regulation by adjusting the system impedance [47], via the adaptation of the inertia, damping, and stiffness components. In the presence of a force and torque sensor measuring $\mathbf{f}_{\mathbf{ext}}$, impedance control can be implemented by enabling inertia shaping [48]. Casting the control law

$$\mathbf{f}_{\mathbf{c}} = \boldsymbol{\Lambda}(\mathbf{q})\boldsymbol{\alpha} + \boldsymbol{\Gamma}(\mathbf{q},\dot{\mathbf{q}})\dot{\mathbf{x}} + \boldsymbol{\eta}(\mathbf{q}) + \mathbf{f}_{\mathbf{ext}}\,, \tag{2}$$

into the dynamic model in (1) results in $\ddot{\mathbf{x}} = \boldsymbol{\alpha}$, $\boldsymbol{\alpha}$ being the control input denoting acceleration with respect to the base frame. In task space IC, the objective is to maintain a dynamics relationship (3) between the external force, $\mathbf{f}_{\mathbf{ext}}$, and the error in position $\boldsymbol{\delta}\mathbf{x} = \mathbf{x}^{\mathbf{r}} - \mathbf{x}$, 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

$$\mathbf{M}\delta\ddot{\mathbf{x}} + \mathbf{D}\delta\dot{\mathbf{x}} + \mathbf{K}\delta\mathbf{x} = \mathbf{f}_{\mathbf{ext}}\,, \tag{3}$$

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 behavior (3) can be achieved using the following control law,

$$\boldsymbol{\alpha} = \ddot{\mathbf{x}}^{r} + \mathbf{M}^{-1}(\mathbf{D}\delta\dot{\mathbf{x}} + \mathbf{K}\delta\mathbf{x} - \mathbf{f}_{\mathbf{ext}})\,. \tag{4}$$

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 (3).

3.3. Probabilistic Ensemble NN (PENN)

PENN [14] is a NN based model approach capable of learning uncertainty-aware NN dynamics models including both aleatoric and epistemic uncertainties. Aleatoric uncertainty refers to the inherent stochasticity of the system. Whereas epistemic uncertainty is a systematic uncertainty arising from issues one could in principle avoid but does not in practice, such as inaccurate measurement, lack of data, modeling errors, etc. 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 epistemic uncertainty. In contrary, an ensemble of deterministic NN can only quantify epistemic uncertainty. In [14] a thorough comparison of PENN with an ensemble of deterministic NN is provided, demonstrating the advantages of PENN for modeling dynamics. The predictive PENN model is trained with negative log prediction probability as a loss function, $\mathrm{loss}_{\mathrm{P}}(\boldsymbol{\theta}) = -\sum_{t=1}^{N}\log\widetilde{f}_{\theta}\left(s_{t+1} \mid s_t, u_t\right)$. Where s_t is the state of the system at time step t, u_t is the applied

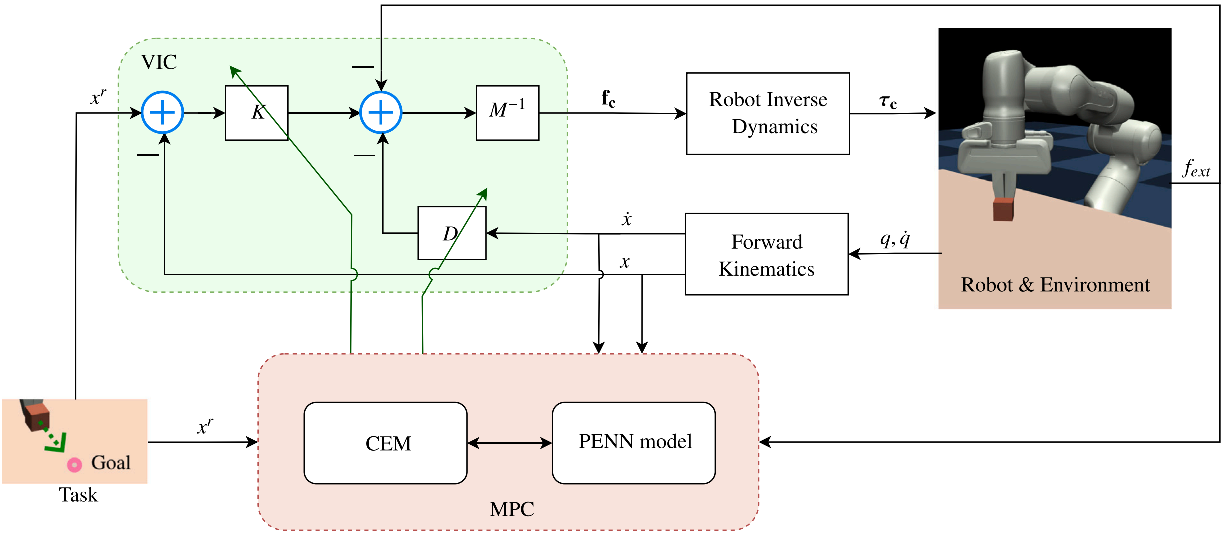


Fig. 1. Block diagram of the deep MPVIC with PENN Cartesian impedance model and the proposed CEM-based MPC scheme for impedance adaptation. This impedance adaptation scheme along with the VIC forms the deep MPVIC framework. The task objective is represented by (6).

action, and s_{t+1} is the next state. The PENN model is defined to output a Gaussian distribution with mean μ and diagonal covariances Σ parameterized by θ s.t, $\tilde{f} = \Pr\left(s_{t+1} \mid s_t, u_t\right) = \mathcal{N}\left(\mu_\theta\left(s_t, u_t\right), \Sigma_\theta\left(s_t, u_t\right)\right)$ [14]. The network output in this fashion parameterizes a Gaussian distribution allowing for modeling the aleatoric uncertainty. In this work, we use the PENN model to learn a Cartesian impedance model of the robot (see Section 4.1).

3.4. CEM-based MPC

The CEM [49] offers a gradient-free optimization scheme, and 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 $u_{t:t+H} \sim \mathcal{N}\left(\mu_{t:t+H}, \operatorname{diag}\left(\sigma^2_{t:t+H}\right)\right)$, where these action sequences are evaluated on the learned dynamical model with respect to a cost function. The sampling distribution, $\mu_{t:t+H}, \sigma^2_{t:t+H}$ is then updated based on best $\mathcal{N}$ trajectories. Safety can be directly incorporated into CEM-based optimization by sorting the samples based on constraint satisfaction values [50], but we are not considering constraints in the MPC scheme for this work.

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

The deep MPVIC framework is formulated to optimize a VIC utilizing a learned PENN 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 PENN model in an MBRL setting alternating between model learning and CEM based exploration strategy. The Cartesian impedance model represents the environment-robot dynamic relationship in (3). We define the state s_t as $[\mathbf{x}_t, \dot{\mathbf{x}}_t]$, and action u_t as $[\mathbf{f}^t{}_{\mathbf{ext}}, \mathbf{K}_t]$. $\mathbf{K}_t$ is given by the CEM-based MPC scheme. $\mathbf{f}^t{}_{\mathbf{ext}}$ is the sensed external force acting on the robot at time instant t, this is an uncertain external factor the VIC needs to compensate for. The damping parameters are chosen according to the critical damping condition, $\mathbf{D} = 2\sqrt{\mathbf{K}}$.

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 that maximize the epistemic uncertainty estimate from PENN. Given a PENN model $\tilde{f}$ of B bootstrap models $\tilde{f}_b$, the uncertainty of the model prediction at the current state can be estimated by calculating the model variance [42], $\rho = \sigma^2$, given by

$$\rho(s,u) = \frac{1}{B-1}\sum_{b=1}^{B}\left(\tilde{f}_b(s,u) - \overline{\tilde{f}(s,u)}\right)^2 . \qquad (5)$$

The designed exploration scheme will excite the system in areas in its state space where the model is more uncertain, thereby maximizing the information gained during exploration. The exploration scheme relies on a control strategy that chooses the actions that provide the highest uncertainty estimate from any given state according to (5). We employ a CEM-based MPC strategy to optimize for the actions that will excite the system to the most uncertain areas. In order to achieve this we define the MPC cost to maximize the variance of the outputs from all the individual NN models in the PENN, $C_\rho = \rho(s_t, u_t)$. At any given state s_t, CEM-based MPC scheme works by (i) sampling a set of actions from the defined time-evolving distribution (we use Gaussian distribution), (ii) sorting the actions according to the uncertainty estimate in (5), (iii) apply the action u^*_t with the highest value of ρ, and (iv) update the Gaussian distribution. This exploration strategy enables learning a generalized model in a sample-efficient way. The model learning approach is summarized in Algorithm 1. Learning a model with low 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^r_t and a control input u_t transitions to the next state s_{t+1}. The dynamics model shown in Fig. 1 represents a generalized Cartesian behavior of an 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 $\tilde{f}$ can be utilized in a MPC

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 $u_{t:t+T}$ ~CEM($\cdot$), 1 to CEM Iterations do
            Evaluate and sort the actions by based on the uncertainty estimate in (5).
        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 PENN 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 $C$ (6) for all $K_{t:t+T}$ on $\tilde{f}$ with actions $[K_{t:t+T}, f_t, s_t^r]$ using trajectory sampling (Section 4.2).
        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
```

framework to adapt the impedance parameters of the VIC by designing a suitable optimization objective as shown in Fig. 1. The MPC scheme uses the prediction of the PENN model $\tilde{f}$ to plan action trajectories yielding the highest reward. At every time-step, an MPC with a horizon length of n, samples the current state and optimizes a control trajectory $u_{t:t+n}$ for n future time-steps and applies the first control input, u_t, to the system. The action optimal action sequence is chosen by: $\arg\min_{u_{t:t+n}} \sum_{i=t}^{t+n} \mathbb{E}_{\tilde{f}} \left[C\left(s_i, u_i\right) \right]$, where C is the cost function. A gradient-free optimization method, CEM is used in an MPC setting to optimize the controller over the PENN model. CEM samples actions from a distribution closer to previous action samples achieved the minimum cost.

In order to calculate the cumulative cost of the action trajectories, we use particle-based propagation as they are specifically suited for PENN dynamics models, [14]. P particles are created from the current state, $s_{t=0}^p = s_0 \forall p$ in order to predict the state trajectories using particle-based propagation. Each of these particles are propagated along the PENN model as, $s_{t+1}^p \sim \tilde{f}_{b(p,t)}\left(s_t^p, u_t\right)$ based on a bootstrap $b(p,t)$ in $\{1, \ldots, B\}$. We keep the particle bootstrap index constant during a trial as it allows us to separate between aleatoric and epistemic uncertainties [51]. The aleatoric uncertainty can be quantified using the average variance of particles of the same bootstrap whereas epistemic uncertainty can be quantified using the variance of the average of particles of the same bootstrap indexes.

The proposed deep MPVIC approach utilizing PENN 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,

$$C\left(s_t, u_t\right) = \delta s_t^T \mathbf{Q}_t \delta s_t + \lambda(K_t)^T \mathbf{R}_t \lambda(K_t), \tag{6}$$

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 robot's states. The MPC output behavior 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 x, y, z directions 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 the PENN model with ensembles of 5 NN with 3 hidden layers, each with 256 neurons. The network structure is chosen based on one-step prediction accuracy empirically over a pre-collected dataset. 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 $\mathbf{D} = 2\sqrt{\mathbf{K}}$. The mass matrix $\mathbf{M}$ is kept constant to avoid stability issues during the experiment. CEM is used to optimize the exploration strategy based on uncertainty maximization. The control frequency for low-level VIC is set at 100 Hz.

For learning the model, the robot manipulator is excited at every time-step with $f_{ext} \sim \mathrm{U}(-20, 20)$ N and s_t^r, where $x_t^r, y_t^r, z_t^r \sim \mathrm{U}(-10, 10)$ 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 scalar values α_Q and α_R as $\mathbf{Q_{new}} = \mathbf{Q} * \alpha_Q$ and $\mathbf{R_{new}} = \mathbf{R} * \alpha_R$ 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 10 Hz which is equivalent to 2.77 h of real-world training. The quality of the model was evaluated using a randomly sampled evaluation dataset as shown in Fig. 2. For experiments, a prior model trained in simulations over 50 000 time-steps are fine-tuned offline in the experimental scenario instead of training 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 using ropes attached to the gripper.

After learning the Cartesian impedance model of the manipulator, 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. Tasks requiring real-time stiffness adaptation are suitable for evaluating the stiffness profile generated by the deep MPVIC controller. For all chosen tasks, the requirements can be defined as achieving a desired goal pose for the robot end-effector. However, the robot is also required to be highly compliant whenever it is possible or be stiff only when it is necessary. This is achieved by using a weighted reward in (6) for the task requirement (first term) and maximizing compliance (second term). In the considered tasks, we are, essentially, trying for a trade-off between position control and compliance.

The three different simulation tasks are modeled in the MuJoCo physics simulation framework [52], see Fig. 3(a), (b), and (c). The two real experimental scenarios are shown in Fig. 3(d) and (e). The aleatoric uncertainties in these robotic tasks is majorly due to measurement noise, whereas the epistemic uncertainty we target during exploration arises from not having enough data to model the Cartesian impedance dynamics in the system state-space we are interested in. *In simulations*,

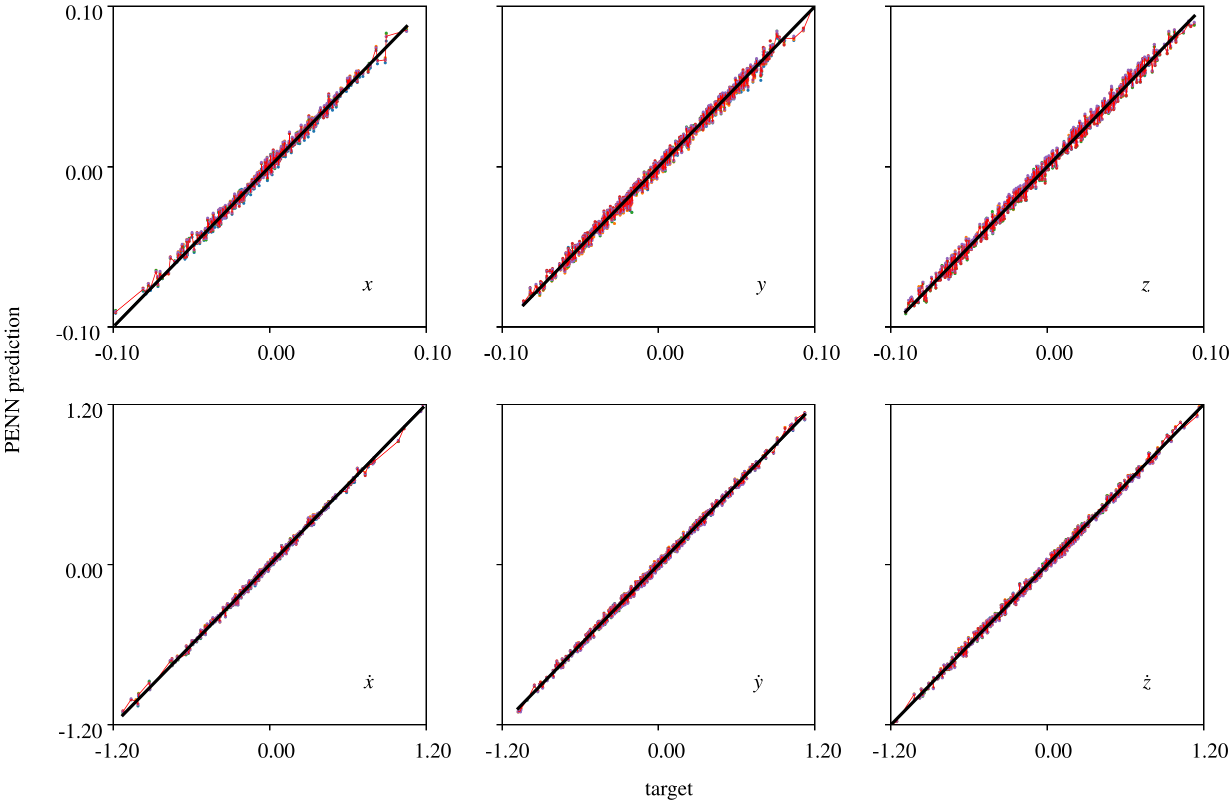


Fig. 2. Scatter plot demonstrating the prediction quality of the PENN model along all the six state dimensions. The plot compares the ground truth (target) with the model's prediction. This model evaluation is estimated over a randomly sampled evaluation dataset from the same training set-up on simulation. The plot shows the mean prediction as well as the individual predictions of each ensemble member of the PENN model. The mean prediction from the PENN model is shown in red, and other colors in the scatter plot represent the predictions from the individual networks in the PENN model. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

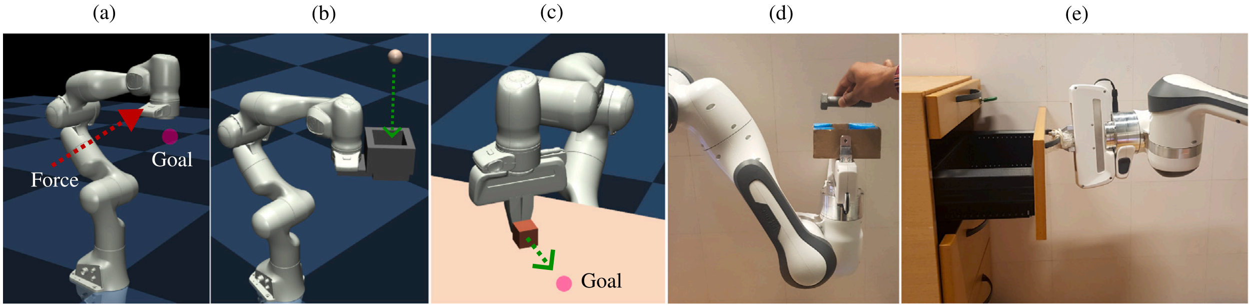


Fig. 3. Three simulation tasks, (a) Cartesian compliance task: the robot manipulator end-effector should hold its pose in the Cartesian space compliantly while reacting to the external forces acting on it. (b) Reacting to falling object: The robot manipulator with cup end-effector should hold a Cartesian position while smoothly catching a ball of weight 0.5 g 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: the 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.

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 to 5. *While for the real experiments*, the control frequency is set as 5 Hz. 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. More conservative control frequency and CEM parameters are chosen for the experiment due to the computational time limitations of the proposed method as discussed in Section 6.3. 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. 3(a)), the robot is expected to behave highly compliant to hold its pose allowing only small deviations. Upon applying an external force to the robot's 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 for testing 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

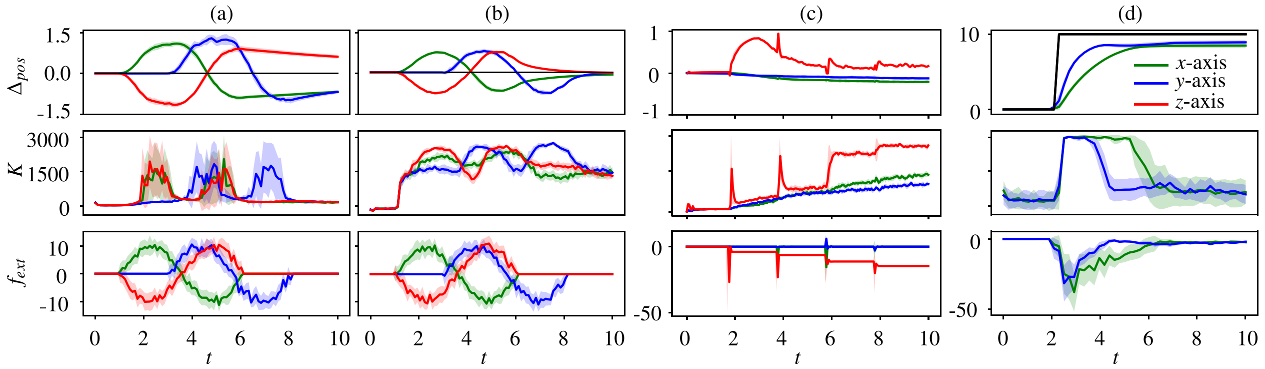


Fig. 4. Simulations: (a) and (b), (Cartesian compliance behavior), results from 20 trials where a sinusoidal force profile with amplitude of 10 N with random noise of (±5) N is applied to the robot end-effector. (a) High compliant behavior optimized using a cost function with larger compliance factor $\alpha_R = 0.1$, (b) Low compliant behavior 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 $K \to 0$. Objects of different weights are dropped at regular intervals of 2 s, 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 $K \to 0$. 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 directions (shown in solid black line) on a surface. The results shown here are over 10 trials with objects of random weights between $(0.5 - 3.0)$ kg and $\alpha_R = 0.1$. A common legend for all four figures is provided in (d). The black line in (a), (b) and (d) represents the goal position. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

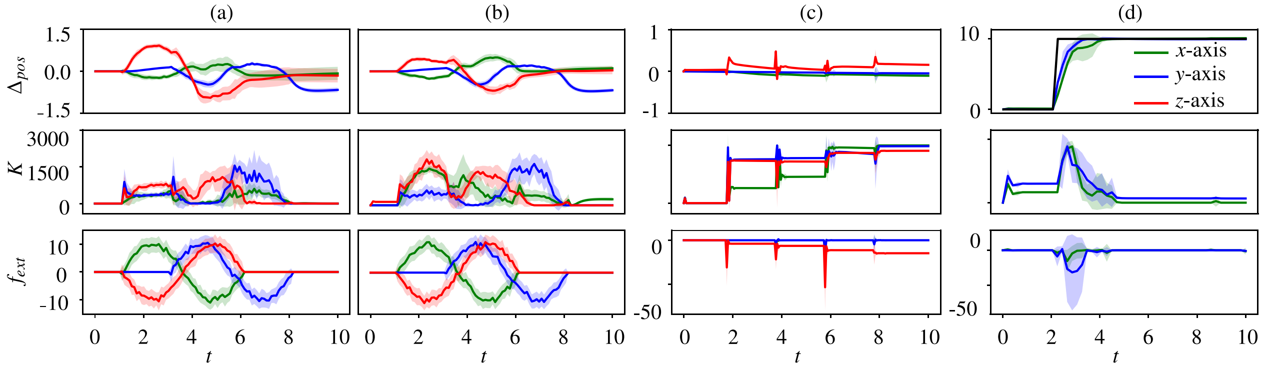


Fig. 5. Corresponding results from Model-free RL policy for the simulation tasks shown in Fig. 4. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

maximization component in the cost function. The results in Fig. 4(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 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. 3(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 falling objects. The robot is additionally expected to be as compliant as possible and be stiff only when necessary as in (6). The resulting robot behavior is shown in Fig. 4(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. 3(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. 4(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 PENN 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 [26,29,30]. 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. 3 are trained using SAC implementation from *stable-baselines* [53] for 500 000 time-steps.

In addition, we compare our approach with the MBRL approach PETS [14]. In the case of PETS, the simulation tasks are trained for 100 000 time-steps. The PETS policies were trained with the same CEM parameters and cost functions used for the corresponding tasks in our deep MPVIC. The performance and the transferability of the learned

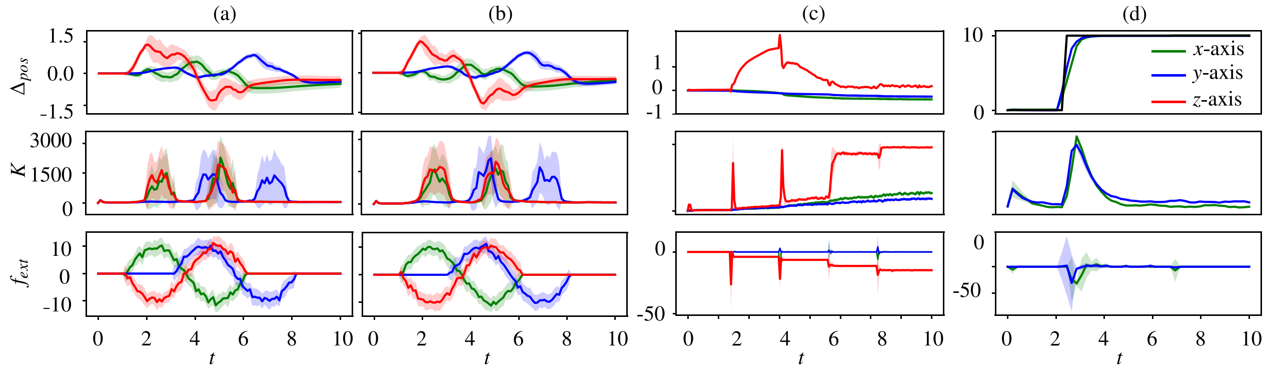


Fig. 6. Corresponding results from PETS policy for the simulation tasks shown in Fig. 4. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

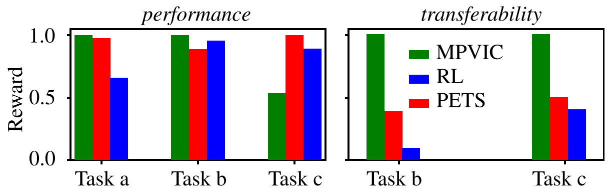


Fig. 7. (left) Comparing the normalized value of the reward (mean value over 20 trials) obtained using Model-free RL, PETS, and our MPVIC framework on all the three simulation tasks. (right) Comparing the transferability of the Model-free RL and PETS based policy with our MPVIC framework based on the normalized value of the mean reward over 20 trials. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

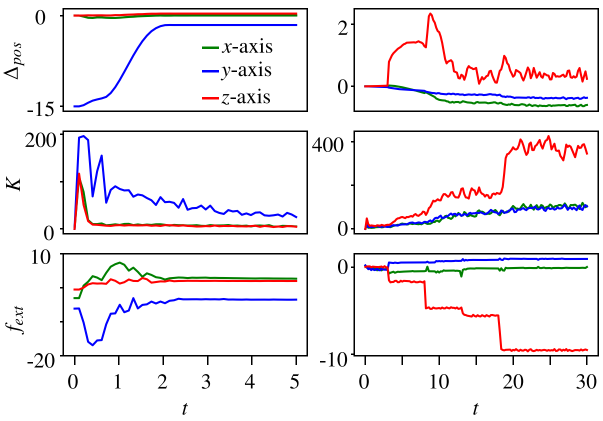


Fig. 8. Experiments: (left) task (e), Robot manipulator opening a table drawer. (right) task (d) The robot manipulator with a tray holding its pose while objects are dropped to the tray. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
Comparison on transferability between tasks.

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

policies in both of these approaches were compared with our MPVIC approach in Fig. 7.

Performance: The resulting robot behavior on applying the learned model-free RL and PETS policies on the three simulation tasks are shown in Figs. 5 and 6 respectively. We compare the performance in terms of the reward obtained by the final policies from each of the approaches on the three simulation tasks. We do not compare the reward during the learning process as the MPC scheme does not have any policy learning process. The rewards obtained while applying the learned policies are shown in Fig. 7. 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. The performance of MPVIC is lower in *task c* as the model is learned on task a which has different dynamics. The key difference between the dynamics of *task a* and *task c* is the robot movement is unconstrained in *task a*, whereas in *task c* it is constrained by the object. That means the model cannot accurately predict the dynamics as in other unconstrained tasks.

Task transferability: In order to evaluate how efficiently 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 PENN model trained on *task a*. Fig. 7-*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 of the additional *data samples/time steps* required for retraining the models/policies for the tasks are shown in Table 2. For example, the model-free RL policy trained on *task a* needed additional training on *task b* with 38.6×10^5 data samples to learn the task b. Whereas MPVIC did not need any training at all (shown as 0 training samples in the table). The number of additional training samples required is correlated with the computational time. While RL approaches demand additional computational/training time to perform a new task, the proposed deep MPVIC can be deployed without any additional computational effort.

5.3. Real-world experiments

Reacting to falling objects: The experimental setup is shown in Fig. 3(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), 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. 8 (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 is similar but in the opposite direction of the pushing task. The experimental setup is shown in Fig. 3(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. 8 (left-column), show the impedance adaptation behavior similar to the pushing task in the simulation where the robot increases its stiffness initially to overcome the inertia of the drawer and then decreases once the drawer starts to move closer to the desired position.

6. Discussion and limitations

6.1. Variable impedance learning control

The deep MPVIC-based approach presented in the work is evaluated over different tasks in Section 5 for optimizing impedance adaptation strategies. 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 behaviors. We considered three simulations and two experimental tasks for evaluating the proposed method. In all the tasks, the task requirement is defined by achieving a desired goal pose for the robot end-effector. The performance of the impedance adaptation strategy is evaluated based on how well it is able to achieve this requirement while being maximally compliant. In all the evaluation scenarios, both in simulation and experiments, the stiffness adaptation guarantees a high level of compliance unless there is a large deviation from the target position or an external force is applied to it. The deep MPVIC scheme is able to adapt the impedance profiles to counteract the external forces and also to trade-off effectively between position accuracy and compliance during the task.

The modeling approach using PENN 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 generalizing a single model to solve multiple manipulation tasks. The versatility 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 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 [26,29,30,35] to solve complex manipulation tasks. The results show that the 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 demands 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. It is important to note that the transferability of deep MPVIC is dependent on the quality of the model. This is evident in its lower performance in the *task c* where the model cannot accurately predict the dynamics of a constrained task. 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.

6.2. Stability analysis

MBRL approaches could improve sample efficiency and can be useful in providing stability and safety guarantees, but there is a need for further research in this direction facilitating complex model structures such as Deep Neural Networks (DNN) to build scalable and sample efficient VILC approaches with theoretical guarantees. In general, the stability of a dynamic system is not necessarily guaranteed when it is coupled to a stable dynamic environment. However, [54] showed stability of the manipulator is preserved when it is coupled to a large class of stable environments if the manipulator has the behavior of a simple impedance. An impedance controller with constant gains makes the closed-loop robot-environment system passive and hence stable in interaction with passive environments [54]. However, this passivity property is lost if the impedance parameters are varied. If the learning-based controller could identify the optimal impedance parameters, one could achieve complex compliant manipulation skills with safety and stability guarantees. But this is not obvious while using RL or CEM-based MPC. One alternative in the case of RL is to use structured policies as done by [55] where the authors use Integrated MOtion Generator and Impedance Controller (iMOGIC) framework to guarantee stable VILC with model-free RL. Even though safety can be achieved using constrained-CEM method [50] in the proposed MPVIC framework, stability guarantees are difficult due to the PENN dynamical models.

Guaranteeing stability and robustness for controllers in a complex robotic manipulator operating in uncertain environments is challenging. In the case of VIC, often passivity theory is used to provide theoretical guarantees under relatively general working assumptions. However, this approach is model-based and the passivity property is lost if arbitrary variations of the impedance parameters are allowed. Passivity-based approaches are often concerned with the analysis of variable impedance profiles that already exist prior to task execution [56]. This is not suitable for guaranteeing the stability of state-dependent real-time impedance variations. In another recent approach, a modified impedance control strategy allows the reproduction of a variable stiffness while preserving the passivity, and therefore a stable behavior both in free motion and in interaction with partially known environments, of the robot [57]. In [57], the goal is to modify the impedance control in order to allow stiffness variations while preserving passivity and, consequently, stable interactive behavior and asymptotic tracking in free motion. This tank-based strategy has been shown very well suited for VIC, in spite of some difficulties in tuning its parameters. Nevertheless, it is dependent on the states of the system, measured during task execution and so can only be applied online. An approach based on the combination of passivity conditions with an adaptation law on the impedance profile was proposed in [58]. This method allows for verifying whether a given profile is passive and if it is not, it provides a method to modify it in a way to guarantee passivity. But none of these approaches are directly extendable to the proposed MPVIC framework to guarantee stability. Ref. [59] proposed an approach using a designed Lyapunov candidate function to stabilize the learned impedance system with an optimal input law in analytical form. But in the case of our MPVIC, this requires solving an additional convex optimization problem at every MPC solution, which could be computationally very expensive and not feasible practically.

The safe learning approaches described in [60] are interesting to explore for model-based VILC. A feasible approach in this direction could be to provide probabilistic safety and stability guarantees using Control Barrier Functions (CBF) and Control Lyapunov Functions (CLF) and solving constrained optimization problems over the GP model [61]. Guaranteeing stability properties to the resulting VILC is challenging as guarantees have to be provided in real-time in an online fashion as the stiffness values predicted by the policy are state-dependent. The approach proposed in [59] by designing a quadratic Lyapunov candidate function could be coupled with GP models to provide probabilistic stability guarantees similar to safety guarantees in [61]. But in the proposed MPVIC framework with PENN dynamics model, such methods are not straightforward to apply. This problem in CEM-based MPC is partly addressed in [62] using available prior models and an auxiliary controller based on CLF and CBF to provide guarantees.

6.3. Limitations

While guaranteeing stability is one major challenge, there are other identified limitations to the proposed approach. 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 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. Even though VIC can be operated generally at lower control frequencies, in tasks with complex contact dynamics this might not be sufficient. This drawback of CEM can affect the model-learning aspect in high-dimensional tasks in terms of sample efficiency. Applying a sample efficient CEM approach [63] and efficient parallelization could help in addressing this issue partly.

The performance of the MPVIC framework can be affected by the quality of the model as the single-step prediction errors get compounded during the multi-step MPC optimization. This can be addressed by training the model over multi-step predictions. Additionally, the MPC horizon can be made adaptive based on the uncertainty estimate from the PENN model. The level of impedance adaptation or the compliance behavior can be adjusted by tuning the Q and R parameters in the cost function (6). However, it is not obvious how to find optimal values for these parameters as the cost function needs to additionally account for model inaccuracies to guarantee optimal performance. Therefore, instead of focusing on the true cost, finding a surrogate cost using Bayesian optimization or evolutionary methods as proposed in [64] offers a way to design the cost. Another interesting approach for cost designing is to iteratively update a baseline cost using RL for closed-loop performance [65]. But here we need to rely on a good enough baseline cost and model so that the RL agent can fine-tune the cost for optimal performance.

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 PENN, a Cartesian impedance model of the robot is learned using an exploration strategy that maximizes the information gain. The PENN 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 future work, we aim to extend this approach to scenarios with constraints, such as in-contact interaction tasks.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This work was partially supported by the project “Dynamic Robot Interaction and Motion Compensation” funded by the Research Council of Norway under contract number 270941, and by euROBIN funded by EU-Horizon-Framework under grant agreement No 101070596

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.robot.2023.104531.

References

[1] E. Bizzi, N. Accornero, W. Chapple, N. Hogan, Posture control and trajectory formation during arm movement, J. Neurosci. 4 (11) (1984) 2738–2744.

[2] N. Hogan, An organizing principle for a class of voluntary movements, J. Neurosci. 4 (11) (1984) 2745–2754.

[3] S.D. Kennedy, A.B. Schwartz, Stiffness as a control factor for object manipulation, J. Neurophysiol. 122 (2) (2019) 707–720.

[4] A.V. Hill, The series elastic component of muscle, Proc. R. Soc. Lond. Ser. B (1950) 273–280.

[5] N. Hogan, Impedance control: An approach to manipulation, in: 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 J. Robot. Autom. 3 (1) (1987) 43–53.

[7] R. Ikeura, H. Inooka, Variable impedance control of a robot for cooperation with a human, in: IEEE International Conference on Robotics and Automation, IEEE, Nagoya, Japan, 1995, pp. 3097–3102.

[8] F.J. Abu-Dakka, L. Rozo, D.G. Caldwell, Force-based variable impedance learning for robotic manipulation, Robot. Auton. Syst. 109 (2018) 156–167.

[9] E. Caldarelli, A. Colomé, C. Torras, Perturbation-based stiffness inference in variable impedance control, IEEE Robot. Autom. Lett. 7 (4) (2022) 8823–8830.

[10] F.J. Abu-Dakka, M. Saveriano, Variable impedance control and learning—A review, Frontiers Robotics and AI 7 (2020).

[11] E.F. Camacho, C.B. Alba, Model Predictive Control, Springer science & business media, 2013.

[12] L. Deng, Z. Li, Y. Pan, Sparse online Gaussian process impedance learning for multi-DoF robotic arms, in: IEEE International Conference on Advanced Robotics and Mechatronics, IEEE, Chongqing, China, 2021, pp. 199–206.

[13] V. van Spaandonk, Learning variable impedance control: A model-based approach using Gaussian processes, 2016.

[14] K. Chua, R. Calandra, R. McAllister, S. Levine, Deep reinforcement learning in a handful of trials using probabilistic dynamics models, in: S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, R. Garnett (Eds.), Advances in Neural Information Processing Systems, 31, Curran Associates, Inc., 2018.

[15] S. Calinon, I. Sardellitti, D.G. Caldwell, Learning-based control strategy for safe human-robot interaction exploiting task and robot redundancies, in: IEEE/RSJ International Conference on Intelligent Robots and Systems, IEEE, Taipei, Taiwan, 2010, pp. 249–254.

[16] S.M. Khansari-Zadeh, K. Kronander, A. Billard, Modeling robot discrete movements with state-varying stiffness and damping: A framework for integrated motion generation and impedance control, in: D. Fox, L.E. Kavraki, H. Kurniawati (Eds.), Robotics: Science and systems X, Berkeley, USA, 2014, p. 2014.

[17] D. Lee, C. Ott, Incremental kinesthetic teaching of motion primitives using the motion refinement tube, Auton. Robots 31 (2) (2011) 115–131.

[18] M. Saveriano, S.-i. An, D. Lee, Incremental kinesthetic teaching of end-effector and null-space motion primitives, in: 2015 IEEE International Conference on Robotics and Automation, IEEE, Seattle, WA, USA, 2015, pp. 3570–3575.

[19] A. Hussein, M.M. Gaber, E. Elyan, C. Jayne, Imitation learning: A survey of learning methods, ACM Comput. Surv. 50 (2) (2017) 1–35.

[20] P. Kormushev, S. Calinon, D.G. Caldwell, Reinforcement learning in robotics: Applications and real-world challenges, Robotics 2 (3) (2013) 122–148.

[21] C.-C. Cheah, D. Wang, Learning impedance control for robotic manipulators, IEEE Trans. Robot. Autom. 14 (3) (1998) 452–465.

[22] A. Gams, B. Nemec, A.J. Ijspeert, A. Ude, Coupling movement primitives: Interaction with the environment and bimanual tasks, IEEE Trans. Robot. 30 (4) (2014) 816–830.

[23] F.J. Abu-Dakka, B. Nemec, J.A. Jørgensen, T.R. Savarimuthu, N. Krüger, A. Ude, Adaptation of manipulation skills in physical contact with the environment to reference force profiles, Auton. Robots 39 (2) (2015) 199–217.

[24] A. Kramberger, E. Shahriari, A. Gams, B. Nemec, A. Ude, S. Haddadin, Passivity based iterative learning of admittance-coupled dynamic movement primitives for interaction with changing environments, in: IEEE/RSJ International Conference on Intelligent Robots and Systems, IEEE, Madrid, Spain, 2018, pp. 6023–6028.

[25] Y. Zhang, B. Chu, Z. Shu, A preliminary study on the relationship between iterative learning control and reinforcement learning, IFAC-PapersOnLine 52 (29) (2019) 314–319.

[26] 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: IEEE/RSJ International Conference on Intelligent Robots and Systems, IEEE, Macau, China, 2019, pp. 1010–1017.

[27] 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, Appl. Sci. 10 (19) (2020) 6923.

[28] 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 Robot. Autom. Lett. 5 (4) (2020) 5709–5716.

[29] M. Bogdanovic, M. Khadiv, L. Righetti, Learning variable impedance control for contact sensitive tasks, IEEE Robot. Autom. Lett. 5 (4) (2020) 6129–6136.

[30] P. Varin, L. Grossman, S. Kuindersma, A comparison of action spaces for learning manipulation tasks, in: IEEE/RSJ International Conference on Intelligent Robots and Systems, IEEE, Macau, China, 2019, pp. 6015–6021.

[31] M. Kim, S. Niekum, A.D. Deshpande, SCAPE: Learning stiffness control from augmented position control experiences, in: Conference on Robot Learning, PMLR, London, UK, 2022, pp. 1512–1521.

[32] J. Buchli, F. Stulp, E. Theodorou, S. Schaal, Learning variable impedance control, Int. J. Robot. Res. 30 (7) (2011) 820–833.

[33] 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 (8) (2018) 2539.

[34] A.S. Anand, M.H. Myrestrand, J.T. Gravdahl, Evaluation of variable impedance- and hybrid force/MotionControllers for learning force tracking skills, in: IEEE/SICE International Symposium on System Integration, IEEE, Narvik, Norway, 2022, pp. 83–89.

[35] 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, J. Intell. Robot. Syst. 100 (2) (2020) 417–433.

[36] L. Roveda, A. Testa, A.A. Shahid, F. Braghin, D. Piga, Q-Learning-based model predictive variable impedance control for physical human-robot collaboration, Artificial Intelligence 312 (2022) 103771.

[37] 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, 2019, arXiv preprint arXiv:1907.02057.

[38] M.V. Minniti, R. Grandia, K. Fäh, F. Farshidian, M. Hutter, Model predictive robot-environment interaction control for mobile manipulation tasks, in: IEEE International Conference on Robotics and Automation, IEEE, Xi'an, China, 2021, pp. 1651–1657.

[39] T. Gold, A. Völz, K. Graichen, Model predictive interaction control for robotic manipulation tasks, IEEE Trans. Robot. 39 (2023) 76–89.

[40] K. Haninger, C. Hegeler, L. Peternel, Model predictive control with Gaussian processes for flexible multi-modal physical human robot interaction, in: IEEE International Conference on Robotics and Automation, IEEE, Philadelphia, PA, USA, 2022, pp. 6948–6955.

[41] P. Shyam, W. Jaśkowski, F. Gomez, Model-based active exploration, in: International Conference on Machine Learning, PMLR, Long Beach, California, USA, 2019, pp. 5779–5788.

[42] 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, Online, 2020, pp. 8583–8592.

[43] Y. Yao, L. Xiao, Z. An, W. Zhang, D. Luo, Sample efficient reinforcement learning via model-ensemble exploration and exploitation, in: IEEE International Conference on Robotics and Automation, IEEE, Xi'an, China, 2021, pp. 4202–4208.

[44] D. Pathak, D. Gandhi, A. Gupta, Self-supervised exploration via disagreement, in: International Conference on Machine Learning, PMLR, Long Beach, California, USA, 2019, pp. 5062–5071.

[45] D.V. Lindley, On a measure of the information provided by an experiment, Ann. Math. Stat. 27 (4) (1956) 986–1005.

[46] D. Pathak, P. Agrawal, A.A. Efros, T. Darrell, Curiosity-driven exploration by self-supervised prediction, in: International Conference on Machine Learning, PMLR, Sydney, Australia, 2017, pp. 2778–2787.

[47] H.-P. Huang, S.-S. Chen, Compliant motion control of robots by using variable impedance, Int. J. Adv. Manuf. Technol. 7 (6) (1992) 322–332.

[48] L. Villani, J. De Schutter, Force control, in: B. Siciliano, O. Khatib (Eds.), Springer Handbook of Robotics, Springer International Publishing, 2016, pp. 195–220.

[49] Z.I. Botev, D.P. Kroese, R.Y. Rubinstein, P. L'Ecuyer, The cross-entropy method for optimization, in: C. Rao, V. Govindaraju (Eds.), Handbook of Statistics, 31, Elsevier, 2013, pp. 35–59.

[50] M. Wen, U. Topcu, Constrained cross-entropy method for safe reinforcement learning, Adv. Neural Inf. Process. Syst. 31 (2018).

[51] S. Depeweg, J.-M. Hernandez-Lobato, F. Doshi-Velez, S. Udluft, Decomposition of uncertainty in Bayesian deep learning for efficient and risk-sensitive learning, in: International Conference on Machine Learning, PMLR, Stockholm, Sweden, 2018, pp. 1184–1193.

[52] E. Todorov, T. Erez, Y. Tassa, Mujoco: A physics engine for model-based control, in: IEEE/RSJ International Conference on Intelligent Robots and Systems, IEEE, Vilamoura-Algarve, Portugal, 2012, pp. 5026–5033.

[53] 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, 2018, https://github.com/hill-a/stable-baselines.

[54] N. Hogan, On the stability of manipulators performing contact tasks, IEEE J. Robot. Autom. 4 (6) (1988) 677–686.

[55] S.A. Khader, H. Yin, P. Falco, D. Kragic, Stability-guaranteed reinforcement learning for contact-rich manipulation, IEEE Robot. Autom. Lett. 6 (1) (2020) 1–8.

[56] K. Kronander, A. Billard, Stability considerations for variable impedance control, IEEE Trans. Robot. 32 (5) (2016) 1298–1305.

[57] F. Ferraguti, C. Secchi, C. Fantuzzi, A tank-based approach to impedance control with variable stiffness, in: IEEE International Conference on Robotics and Automation, IEEE, Karlsruhe, Germany, 2013, pp. 4948–4953.

[58] M. Bednarczyk, H. Omran, B. Bayle, Passivity filter for variable impedance control, in: IEEE/RSJ International Conference on Intelligent Robots and Systems, IEEE, Las Vegas, NV, USA, 2020, pp. 7159–7164.

[59] Z. Jin, A. Liu, W.-a. Zhang, L. Yu, An optimal variable impedance control with consideration of the stability, IEEE Robot. Autom. Lett. 7 (2) (2022) 1737–1744.

[60] A. Anand, K. Seel, V. Gjærum, A. Håkansson, H. Robinson, A. Saad, Safe learning for control using control Lyapunov functions and control barrier functions: A review, Procedia Comput. Sci. 192 (2021) 3987–3997.

[61] M.J. Khojasteh, V. Dhiman, M. Franceschetti, N. Atanasov, Probabilistic safety constraints for learned high relative degree system dynamics, in: A.M. Bayen, A. Jadbabaie, G. Pappas, P.A. Parrilo, B. Recht, C. Tomlin, M. Zeilinger (Eds.), Learning for Dynamics and Control, 120, PMLR, 2020, pp. 781–792.

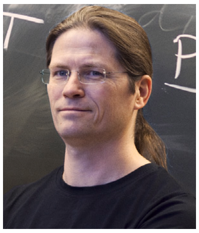
[62] L. Zheng, R. Yang, Z. Wu, J. Pan, H. Cheng, Safe learning-based gradient-free model predictive control based on cross-entropy method, Eng. Appl. Artif. Intell. 110 (2022) 104731.

[63] C. Pinneri, S. Sawant, S. Blaes, J. Achterhold, J. Stueckler, M. Rolinek, G. Martius, Sample-efficient cross-entropy method for real-time planning, in: Conference on Robot Learning, PMLR, Cambridge MA, USA., 2021, pp. 1049–1065.

[64] A. Jain, L. Chan, D.S. Brown, A.D. Dragan, Optimal cost design for model predictive control, in: Learning for Dynamics and Control, PMLR, 2021, pp. 1205–1217.

[65] S. Gros, M. Zanon, Data-driven economic nmpc using reinforcement learning, IEEE Trans. Automat. Control 65 (2) (2019) 636–648.

Akhil S. Anand is a Postdoc at the Norwegian University of Science and Technology (NTNU). He received his B.Tech degree in mechanical engineering from Indian Institute of Technology (IIT), Patna, India in 2013 and his M.Sc in Mechatronics from Universität Siegen, Germany in 2019, and Ph.D. in Robot Learning from NTNU in 2023. His research areas are on robot control and learning with a specific focus on learning-based complainant control and reinforcement learning.



Jan Tommy Gravdahl is a professor at the Norwegian University of Science and Technology (NTNU). He was born in 1969 and graduated siv.ing (1994) and dr.ing (1998) in Engineering Cybernetics, NTNU. From 1998 to 2001 he was a postdoctoral researcher with the Department of Engineering Cybernetics. He was appointed Associate Professor (2001) and Professor (2005) at the same department, where he served as deputy department head 2006-07 and department head in 2008/09. In 2007/08 he was with The Centre for

Complex Dynamic Systems and Control (CDSC), The University of Newcastle, Australia. His current research interests include mathematical modeling and nonlinear control in general and with application to turbomachinery, spacecraft, robots, ships and nanopositioning devices. He has supervised the graduation of 100 M.Sc. and 10 PhD. He has published more than 200 papers at conferences and in international journals. He received the IEEE Transactions on Control Systems Technology Outstanding Paper Award in 2000 and in 2017. He is author of Compressor Surge and Rotating Stall: Modeling and Control (Springer 1999), co-author of Modeling and Simulation for Automatic Control (Marine Cybernetics 2002), Snake Robots: Modeling, Mechatronics, and Control (Springer 2013), Vehicle-Manipulator Systems: Modeling for Simulation, Analysis and Control (Springer 2014) and co-editor of Group Coordination and Cooperative Control (Springer 2006).

Fares J. Abu-Dakka received his B.Sc. degree in mechanical engineering from Birzeit University, Palestine, in 2003, and his DEA and Ph.D. degrees in robotics motion planning from the Polytechnic University of Valencia, Spain, in 2006 and 2011, respectively. In 2012 he started his first postdoc at Jozef Stefan Institute ´, Slovenia. Between 2013 and 2016 he held a visiting professorship at ISA within the Carlos III University of Madrid, Spain. From 2016 to 2019, he was a Postdoc at Istituto Italiano di Tecnologia (IIT). From 2019 to 2022, he was a Research Fellow at Aalto University. Then in 2022, he moved to MIRMI, Technical University of Munich, Germany, to serve as a Senior Scientist and leader of the Robot Learning group. Currently, he is a lecturer and researcher at the Faculty of Engineering, Mondragon University, Bilbao, Spain. His research lies in the intersection of control theory, differential geometry, and machine learning, in order to enhance robot manipulation performance and safety. He is an Associate Editor for ICRA, IROS, and RA-L. Webpage: https://sites.google.com/view/abudakka/.