Q-Learning-Based Model Predictive Variable Impedance Control for Physical Human-Robot Collaboration (Extended Abstract)*

Loris Roveda[1] , **Andrea Testa**[2] , **Asad Ali Shahid**[1] , **Francesco Braghin**[2] and **Dario Piga**[1]
[1]Istituto Dalle Molle di studi sull'Intelligenza Artificiale (IDSIA), Scuola Universitaria Professionale della Svizzera Italiana (SUPSI), Università della Svizzera italiana (USI), via la Santa 1, 6962, Lugano, Switzerland
[2]Politecnico di Milano, Department of Mechanical Engineering, via La Masa 1, 20156, Milano, Italy
loris.roveda@idsia.ch, andrea8.testa@mail.polimi.it, asadali.shahid@idsia.ch, francesco.braghin@polimi.it, dario.piga@supsi.ch

Abstract

Physical human-robot collaboration is increasingly required in many contexts. To implement an effective collaboration, the robot should be able to recognize the human's intentions and guarantee safe and adaptive behavior along the desired directions of motion. The robot-control strategies with such attributes are particularly demanded in the industrial field. Indeed, with this aim, this work proposes a Q-Learning-based Model Predictive Variable Impedance Control (Q-LMPVIC) to assist the operators in physical human-robot collaboration (pHRC) tasks. A Cartesian impedance control loop is designed to implement the decoupled compliant robot dynamics. The impedance control parameters (*i.e.*, setpoint and damping parameters) are then optimized in an online manner to maximize the performance of the pHRC. First, an ensemble of neural networks is designed to learn the model of human-robot interaction dynamics while capturing the associated uncertainties. The derived model is then used by the model predictive controller (MPC), enhanced with stability guarantees through Lyapunov constraints. The MPC is solved by making use of a Q-Learning method that, in its online implementation, uses an actor-critic algorithm to approximate the exact solution. The Q-learning method provides an accurate and highly efficient solution (in terms of computational time and resources). The proposed approach has been validated through experimental tests on a Franka EMIKA panda robot.

1 Introduction

To meet customers' needs, which are becoming more and more oriented on tailor-made products, companies are updating their production processes by means of new flexible and agile tools [Fragapane *et al.*, 2020]. In this context, collaborative robotics plays a key role [Makris, 2021], providing powerful solutions to assist the operators in the execution of different activities, such as co-manipulation [Roveda *et al.*, 2018a; Roveda *et al.*, 2019], task's knowledge transfer to the robotic system [Roveda *et al.*, 2021b; Roveda *et al.*, 2021a], easily programmable and deployable applications [Vicentini *et al.*, 2020], etc. Physical human-robot collaboration (pHRC) is currently one of the most investigated topics [Galin and Meshcheryakov, 2020]. In fact, pHRC is nowadays demanded in many fields of applications, both for collaborative robots [Roveda *et al.*, 2018b] and exoskeletons [Mauri *et al.*, 2019]. However, many open issues in the state of the art are still to be overcome, in particular considering safety/stability guarantees in the human-robot interaction, human-robot dynamics modeling, human intention recognition (for active assistance/empowering purposes), and computation efficiency (for real-time control adaptation and optimization).

To tackle the above-mentioned issues within the pHRC scenario, this paper proposes a Q-Learning-based Model Predictive Variable Impedance Control (Q-LMPVIC) to assist the operator while physically interacting with a collaborative robot. Based on Cartesian impedance control (providing the controlled manipulator a compliant and decoupled behavior in the Cartesian space), an MPC is designed in order to online optimize its parameters (*i.e.*, setpoint and damping parameters) to assist the user along the detected intended motion direction(s), maximizing the collaboration performance. The MPC exploits a learned human-robot interaction dynamics model, obtained by means of an ensemble of neural networks. Therefore, the lack of sophisticated analytical models for the human-robot interaction dynamics is overcome, by employing a method that is capable to capture the complexity and uncertainties of such dynamics. An MPC objective function is designed in order to minimize the user's effort during the collaboration with the robot. Indeed, the user's intention of motion can be detected, making it possible to assist him/her along the intended direction(s) of motion. The designed MPC is also enhanced with stability guarantees by means of Lyapunov constraints. In such a way, safety/stability issues are tackled by the proposed methodology. The MPC is then (online) solved by making use of a Q-Learning method, exploit-

*This work is described in the paper [Roveda et al., 2022] Loris Roveda, Andrea Testa, Asad Ali Shahid, Francesco Braghin, and Dario Piga. Q-learning-based model predictive variable impedance control for physical human-robot collaboration. Artificial Intelligence, 312:103771, 2022.

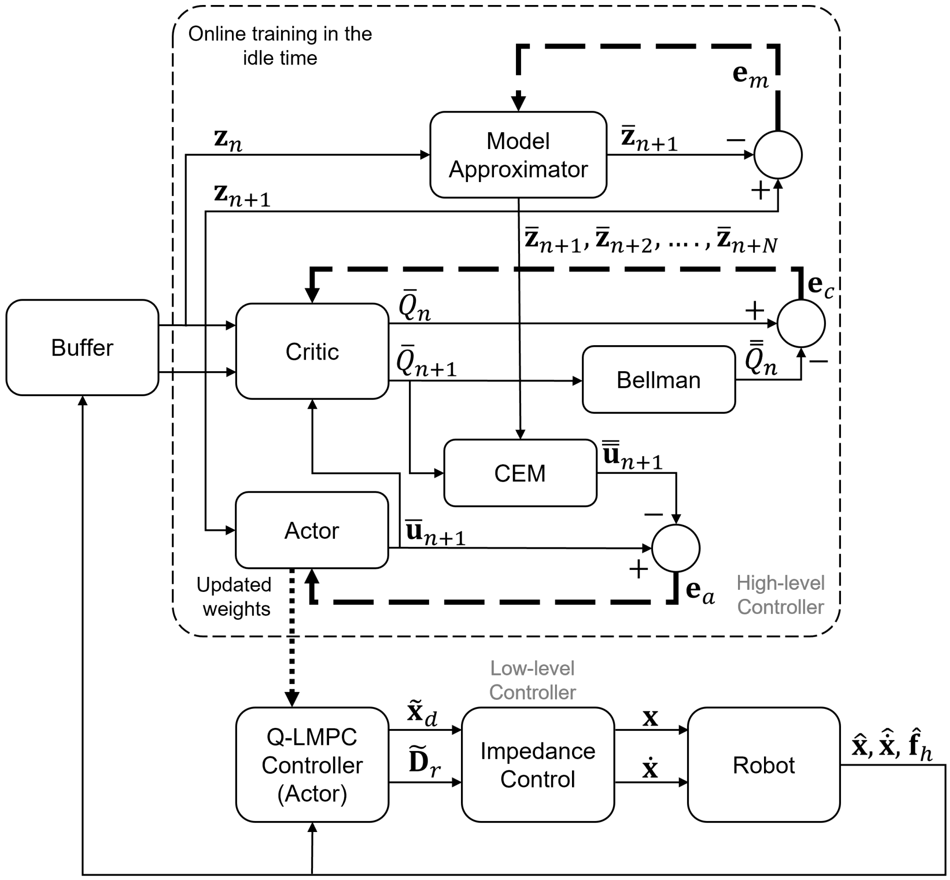

Figure 1: Overall Q-LMPVIC scheme. The training operations inside the dashed square are performed once the buffer is full, leading to the update of the Q-LMPC controller, which regularly sends the optimal setpoint and damping parameters to the impedance controller. The optimization variables (*i.e.*, the Cartesian impedance control setpoint $\tilde{\mathbf{x}}_d$ and damping $\tilde{\mathbf{D}}_r$) are used as the control input $\bar{\mathbf{u}}$, to be computed on the basis of the robot state $\mathbf{z}$

.

ing an actor-critic algorithm to approximate its exact solution. The obtained solution is accurate and highly efficient, being able to tackle the issue related to computation efficiency that might compromise the implementation of the controller for real applications. The proposed approach has been applied to two complex use cases (a collaborative assembly task and a collaborative deposition task) in order to show its applicability to real industrial tasks.

While other approaches have been developed to deal with the proposed topic (*i.e.*, pHRC) [Roveda *et al.*, 2020; Cremer *et al.*, 2019; Roveda *et al.*, 2019; Dimeas and Aspragathos, 2015; Medina *et al.*, 2019; Peternel *et al.*, 2017; Zhang *et al.*, 2020], they are indeed characterized by some difficulties in the simultaneous optimization of the impedance parameters. The online optimization of the setpoint and the other impedance parameters is, in fact, important to obtain an active target-oriented and compliant behavior of the manipulator during pHRC task. In addition, it is commonly difficult to ensure stability in such an optimization problem. Moreover, a reliable model of the target dynamics is not always available, making it difficult to optimize the collaboration. Besides, most of the strategies are based on the internal simulation of the state evolution, exploiting the prediction capabilities of dynamic models or Q-functions. However, the former strategies are often not enough accurate or efficient to be implemented for real applications execution, and the latter ones are usually approximated with fuzzy logic, requiring a precise setting of the fuzzy rules and membership functions to solve the ”curse of dimensionality” problems.

2 Methodology

The proposed Q-Learning-based Model Predictive Variable Impedance Control (Q-LMPVIC) for pHRC tasks is made up of two main levels. In the first one (*i.e.*, the low-level control loop), a variable Cartesian impedance controller is realized, such that the outer high-level controller could work considering the manipulator as a decoupled mass-spring-damper system in the Cartesian space. The outer high-level controller is then used to update the setpoint and the damping parameters of the inner controller in order to optimize the pHRC performance (*i.e.*, minimize the interaction force between the human and the robot, and, therefore, the operator’s effort). The outer high-level controller is composed of an actor and

(a) Assembly task (b) Deposition task

Figure 2: The tested (a) assembly task (where the gear has to be assembled into its shaft) and (b) deposition task (where the material has to be deposited along the highlighted path) are shown.

a critic ANN, which implements a Q-Learning algorithm for the resolution of a nonlinear optimal control problem. An ensemble of ANNs is exploited to estimate the model of the system. An MPC enhanced with stability guarantees (by means of Lyapunov constraints) is implemented to online compute the low-level impedance control parameters, maximizing the pHRC performance.

Figure 1 shows the proposed Q-LMPVIC schema, highlighting each element composing the proposed methodology. In the following, all the elements composing the Q-LMPVIC (*i.e.*, the low-level Cartesian impedance control, the human-robot interaction dynamics, the modeling estimation methodology based on the ensemble of ANNs, the L-MPC, the Q-learning methodology, its actor-critic ANNs, and the CEM algorithm) are described.

Remark 1. All the theoretical derivation of the approach is available in the authors' work [Roveda *et al.*, 2022].

3 Experimental Validation

The developed controller has been employed in two industrial tasks to provide insights into its usage in real production scenarios. In particular, (i) a collaborative assembly task of a gear into its shaft [Roveda *et al.*, 2021b] (Figure 2 (a)) and (ii) a collaborative deposition task [Roveda *et al.*, 2021a] (Figure 2 (b)) have been considered as target use-cases. (i) considers the robot equipped with a gripper manipulating a gear to be collaboratively assembled. (ii) considers the robot equipped with a sealing gun operated by the human along the deposition path, strictly related to the activities developed in the H2020 CS2 ASSASSINN project. Both tasks can be executed to demonstrate the application to the robot (*e.g.*, in the context of programming-by-demonstration) or to collaboratively perform it.

The selected tasks allow the assessment of the adoption of the proposed controller in real human-robot collaborative industrial tasks. In fact, such tasks require a smooth human-robot interaction, having the robot capable of quickly reacting to the intention of motion of the human in order to properly execute the target task, guaranteeing the stability of the interaction. In fact, instabilities, vibrations, or delays in the controller reaction might cause task failures or unacceptable output quality. Indeed, the proposed use-cases allow the evaluation of the effectiveness of the developed controller in complex human-robot collaborative applications.

A video showing the performed assembly task is available at the link, while a video showing the performed deposition task is available at this link. As can be highlighted, the proposed controller allows the implementation of a smooth and reactive robot behavior (reflecting the performance evaluation provided in the previous Sections), making it possible to perform complex human-robot collaborative tasks. Thus, the proposed controller is proven to be applicable to real industrial tasks.

Remark 2. The complete experimental analysis of the developed approach is available in the authors' work [Roveda *et al.*, 2022], where the performance of the proposed controller are compared with the ones achieved by [Roveda *et al.*, 2020].

4 Conclusions

The presented work in this paper proposes a Q-Learning-based Model Predictive Variable Impedance Control Q-LMPVIC to assist the operators in physical human-robot collaboration (pHRC) tasks. The proposed methodology is composed of two control loops, a low-level Cartesian impedance controller, and a high-level methodology for the online optimization of impedance control parameters (*i.e.*, the setpoint and damping parameters), allowing to minimize the human-robot interaction force during the collaboration. The outer high-level controller is composed of an actor and a critic ANN, which implement a Q-Learning algorithm for the resolution of a nonlinear optimal control problem. An ensemble of ANNs is exploited to estimate the model of the system. The MPC enhanced with stability guarantees (by means of Lyapunov constraints) is, indeed, implemented to compute the low-level impedance control parameters online. The proposed Q-LMPVIC has been tested to assess its performance. Two complex use-cases (a collaborative assembly task and a collaborative deposition task) have been setup to show the applicability of the proposed approach to real industrial tasks.

Future work is devoted to increasing the high-level control loop frequency (imposed equal to 6 Hz in this paper) in order to further improve the pHRC performance. To solve this issue, parallelization with modern GPUs can be exploited. In addition, external sensors (such as EMGs) can be exploited to better capture the human-robot interaction modeling and collaboration performance. The proposed strategy will also be adapted (in terms of modeling and control objectives) and applied to an exoskeleton device.

Ethical Statement

There are no ethical issues.

Acknowledgments

The work has been developed within the project ASSASSINN, funded from H2020 CleanSky 2 under grant agreement n. 886977.

The work has been developed within the project ExoRescue, funded from Eureka Eurostars under grant agreement n. E!115182.

References

[Cremer *et al.*, 2019] Sven Cremer, Sumit Kumar Das, Indika B Wijayasinghe, Dan O Popa, and Frank L Lewis. Model-free online neuroadaptive controller with intent estimation for physical human–robot interaction. *IEEE Transactions on Robotics*, 36(1):240–253, 2019.

[Dimeas and Aspragathos, 2015] Fotios Dimeas and Nikos Aspragathos. Reinforcement learning of variable admittance control for human-robot co-manipulation. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1011–1016. IEEE, 2015.

[Fragapane *et al.*, 2020] Giuseppe Fragapane, Dmitry Ivanov, Mirco Peron, Fabio Sgarbossa, and Jan Ola Strandhagen. Increasing flexibility and productivity in industry 4.0 production networks with autonomous mobile robots and smart intralogistics. *Annals of operations research*, pages 1–19, 2020.

[Galin and Meshcheryakov, 2020] Rinat R Galin and Roman V Meshcheryakov. Human-robot interaction efficiency and human-robot collaboration. In *Robotics: Industry 4.0 Issues & New Intelligent Control Paradigms*, pages 55–63. Springer, 2020.

[Makris, 2021] Sotiris Makris. *Cooperating Robots for Flexible Manufacturing*. Springer, 2021.

[Mauri *et al.*, 2019] Alessandro Mauri, Jacopo Lettori, Giovanni Fusi, Davide Fausti, Maurizio Mor, Francesco Braghin, Giovanni Legnani, and Loris Roveda. Mechanical and control design of an industrial exoskeleton for advanced human empowering in heavy parts manipulation tasks. *Robotics*, 8(3):65, 2019.

[Medina *et al.*, 2019] José R Medina, Hendrik Börner, Satoshi Endo, and Sandra Hirche. Impedance-based gaussian processes for modeling human motor behavior in physical and non-physical interaction. *IEEE Transactions on Biomedical Engineering*, 66(9):2499–2511, 2019.

[Peternel *et al.*, 2017] Luka Peternel, Nikos Tsagarakis, and Arash Ajoudani. A human–robot co-manipulation approach based on human sensorimotor information. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25(7):811–822, 2017.

[Roveda *et al.*, 2018a] Loris Roveda, Nicola Castaman, Stefano Ghidoni, Paolo Franceschi, Nicolò Boscolo, Enrico Pagello, and Nicola Pedrocchi. Human-robot cooperative interaction control for the installation of heavy and bulky components. In *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 339–344. IEEE, 2018.

[Roveda *et al.*, 2018b] Loris Roveda, Shaghavezh Haghshenas, Alessio Prini, Tito Dinon, Nicola Pedrocchi, Francesco Braghin, and Lorenzo Molinari Tosatti. Fuzzy impedance control for enhancing capabilities of humans in onerous tasks execution. In *2018 15th international conference on ubiquitous robots (UR)*, pages 406–411. IEEE, 2018.

[Roveda *et al.*, 2019] Loris Roveda, Shaghayegh Haghshenas, Marco Caimmi, Nicola Pedrocchi, and Lorenzo Molinari Tosatti. Assisting operators in heavy industrial tasks: on the design of an optimized cooperative impedance fuzzy-controller with embedded safety rules. *Frontiers in Robotics and AI*, 6:75, 2019.

[Roveda *et al.*, 2020] Loris Roveda, Jeyhoon Maskani, Paolo Franceschi, Arash Abdi, Francesco Braghin, Lorenzo Molinari Tosatti, and Nicola Pedrocchi. Model-based reinforcement learning variable impedance control for human-robot collaboration. *Journal of Intelligent & Robotic Systems*, 100(2):417–433, 2020.

[Roveda *et al.*, 2021a] Loris Roveda, Beatrice Maggioni, Elia Marescotti, Asad Shahid, Andrea Maria Zanchettin, Alberto Bemporad, and Dario Piga. Pairwise preferences-based optimization of a path-based velocity planner in robotic sealing tasks. *IEEE Robotics and Automation Letters*, 2021.

[Roveda *et al.*, 2021b] Loris Roveda, Mauro Magni, Martina Cantoni, Dario Piga, and Giuseppe Bucca. Human–robot collaboration in sensorless assembly task learning enhanced by uncertainties adaptation via bayesian optimization. *Robotics and Autonomous Systems*, 136:103711, 2021.

[Roveda *et al.*, 2022] Loris Roveda, Andrea Testa, Asad Ali Shahid, Francesco Braghin, and Dario Piga. Q-learning-based model predictive variable impedance control for physical human-robot collaboration. *Artificial Intelligence*, 312:103771, 2022.

[Vicentini *et al.*, 2020] Federico Vicentini, Nicola Pedrocchi, Manuel Beschi, Matteo Giussani, Niccolò Iannacci, Paolo Magnoni, Stefania Pellegrinelli, Loris Roveda, Enrico Villagrossi, Mehrnoosh Askarpour, et al. Piros: Cooperative, safe and reconfigurable robotic companion for cnc pallets load/unload stations. In *Bringing Innovative Robotic Technologies from Research Labs to Industrial End-users*, pages 57–96. Springer, 2020.

[Zhang *et al.*, 2020] Hao Zhang, Shaoyuan Li, and Yi Zheng. Q-learning-based model predictive control for nonlinear continuous-time systems. *Industrial & Engineering Chemistry Research*, 59(40):17987–17999, 2020.