Sequential Learning Unification Controller from Human Demonstrations for Robotic Compliant Manipulation Communicated by Dr Chenguang Yang

Journal Pre-proof

Sequential Learning Unification Controller from Human Demonstrations for Robotic Compliant Manipulation

Jianghua Duan, Yongsheng Ou, Sheng Xu, Ming Liu

 PII:
 S0925-2312(19)31088-4

 DOI:
 https://doi.org/10.1016/j.neucom.2019.07.081

 Reference:
 NEUCOM 21144

To appear in: *Neurocomputing*

Received date:28 January 2019Revised date:18 June 2019Accepted date:26 July 2019

Please cite this article as: Jianghua Duan, Yongsheng Ou, Sheng Xu, Ming Liu, Sequential Learning Unification Controller from Human Demonstrations for Robotic Compliant Manipulation, *Neurocomputing* (2019), doi: https://doi.org/10.1016/j.neucom.2019.07.081

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2019 Published by Elsevier B.V.



Sequential Learning Unification Controller from Human Demonstrations for Robotic Compliant Manipulation

Jianghua Duan^{a,c}, Yongsheng Ou^{b,c,*}, Sheng Xu^{a,b}, Ming Liu^d

 ^a Guangdong Provincial Key Laboratory of Robotics and Intelligent System, Shenzhen Institutes of Advanced Technology (SIAT), Chinese Academy of Sciences (CAS), Shenzhen 518055, China
 ^b CAS Key Laboratory of Human-Machine-Intelligence Synergic Systems, Shenzhen Institutes of Advanced Technology

(SIAT), Shenzhen 518055, China

^c University of Chinese Academy of Sciences, Beijing 100049, China ^dDepartment of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Hong Kong

Abstract

Robotic compliant manipulation not only contains robot motion but also embodies interaction with the environment. Frequently endowing the compliant manipulation skills to the robot by manual programming or off-line training is complicated and time-consuming. In this paper, we propose a sequential learning framework to take both kinematic profile and variable impedance parameter profile into consideration to model a unified control strategy with "motion generation" and "compliant control". In order to acquire this unification controller efficiently, we use a sequential learning neural network to encode robot motion and a new force-based variable impedance learning algorithm to estimate varying damping and stiffness profiles in three directions. Furthermore, the state-independent stability constraints for variable impedance control are presented. The effectiveness of the proposed learning framework is validated by a set of experiments using the 4-DoF Barrett WAM.

Key words: Robot learning from demonstration (RLfD), variable impedance control, sequential learning, physical Human-Robot interaction, stability analysis.

1. Introduction

When a robot will be used in a new task, a very time-consuming programming work by the professional technicians is needed to endow the robot with a new skill [1]. This traditional way for preprogramming robots is not suitable for the multi-product and small batch manufacturing lines. To endow robots with the learning ability instead of being explicitly programmed in every possible situation is a most effective way to reduce the threshold to start using robot technology and increase implementation efficiency. Robot learning from demonstration (RLfD) [2] is a well-known methodology to extract the task-relevant information and the new sensorimotor knowledge by observing the correct demonstrations. Then a learned model acquired from the data of demonstration can be used for autonomous task execution by the robot [3].

Traditionally, RLfD is used to model the kinematic profiles of the tasks then to produce robot motion.

 $^{^{*}}$ Corresponding author

JOURNAL PRE-PROOF



Figure 1: Illustration of the necessity of controlling both the robot motion and its physical interaction with the environment.

Many excellent methods have been proposed, such as the dynamic movement primitives (DMPs)[4, 5], the stable estimator of dynamical systems (SEDS) approach [6, 7], the fast and stable modeling for dynamical systems [8]. However, two problems have emerged in practical application: Firstly, these methods collected and modeled the robot motion (kinematic profile) in an off-line way. These processes are time-consuming. And even longer than the manual programming, so the advantages of the RLfD with off-line way is not ideal in the multipurpose production line (new manipulation skills are needed frequently). Secondly, these methods are usually used to learn the robot motion. But a stiff motion controller is not enough to implement compliant behaviors [9]. The successful executions of compliant tasks not only require the motion generation (the kinematic profiles of the tasks) but also need the interaction with the environment, see Fig. 1. In this graph, the red line represents a portion of the path and the belted-ellipsoid illustrates the robot stiffness at the contact point in Cartesian space between the end-effector and the computer screen. Motion generation creates a path in task implementation (from starting point to reach the final state). This path is named kinematic profile of the task in this paper. The term "interaction with the environment" refers to regulation of the robot's dynamic behavior at its contact points/surfaces (for instance making the robot compliant or stiff when necessary). Both of these skills are essential for safe and successful execution of many robotic tasks. Cleaning the computer screen with a pure motion generation is challenging because slight uncertainty could result in a excessive or inadequate force, and eventually lead to breakage or unresolved dirt on the screen. Besides, the typical tasks always involve contact or require an appropriate response to unforeseen physical perturbations. The question of how to regulate the compliance to accommodate the requirements of the task have not been fully tackled in RLfD [10]. There are several ways to make the robots compliant: contact detection using sensors (artificial tactile skin or force/ torque sensor) [11, 12], passively compliant by design (reducing the weight and hardness of the robot structure or implementing elastic elements) [13, 14] and active torque control strategies [15, 16, 17]. The last way is seen as the good alternative to the formers, because of its initiative and precision [18]. Therefore, we apply the suitable impedance strategies to active control the torque of the robot in this paper. Impedance control can regulate the dynamic responses of the robot to interaction forces by establishing an appropriate virtual mass-spring-damper system on the

end-effector [15]. Similar to the human limb impedance adjusting, parameters of the impedance control model are regulated over time through learning based on certain criterias [19]. This learning policy has been explored in many previous studies [20, 21, 22, 23].

In [24, 25], learning variable impedance control strategy was formulated as an optimal control problem. In these works, the optimal control formulations were proposed to compute both the variable stiffness profiles and motion reference adaptation. However, in these studies, the impedance profile was tailored to each specific robotic platform, which lost the generality.

Reinforcement learning (RL) has also been applied for learning variable impedance policies [21, 26, 27], which is closely related to the optimal control, and successfully applied in via-point trajectory following and pancakes flipping. Even though RL can be applied to compute the impedance of the robot to improve the task performance, it usually needs a large number of iterations to find an optimal policy, which is not preferred in the multi-product and small batch manufacturing lines.

In recent years, the research about seeking suitable impedance control policies for compliant behaviors is going toward learning and imitation of impedance strategies of humans [9, 28, 10]. The learning of the teacher's arm impedance is not easy. Depending on the differences of task and the teaching interface, the corresponding learning methods are also different.

In [29], the concept of teleimpedance was proposed to make the human impedance regulation skills applied for robot interacting with uncertain environments. In this case, a suitable human-machine interface was used to provide the trajectory and stiffness references. Stiffness reference was computed using the electromyogram signals (EMG) recorded from the arm of the operator. However, as it is a real-time teleoperate robot by a user, it did not learn the task for autonomous execution. Recently, [30, 31] not only used the surface EMG to estimate demonstrator limb stiffness but also used the DMP method to encode both movement trajectories and 1-DoFs stiffness profile (estimated from EMG) to achieve variable impedance skill transfer and generalization.

Calinon et al. [28] derived variable stiffness for an active compliant controller from the demonstrations of the kinematic profiles. The stiffness profile was shaped inversely proportional to the variance along the trajectory [9]. This approach depends upon the assumption that a large position variety in demonstration will not have a negative impact on the task execution which may not always be reasonable [10], especially in the narrow task space (such as Minimally Invasive Surgery). In [32], the trajectory of the task was taught using kinesthetic teaching firstly. When the kinesthetic profile of the task was learned, the teacher demonstrated the forces using a haptic device while the robot was executing the learned motion. Then, the model describing the desired contact forces was built, and used by the robot to determine the desired force during the task execution. The teaching process was divided into two successive steps (kinesthetic teaching first and then haptic teaching), and also used the off-line learning method to encode position and force profiles.

In addition, none of these works addressed the issue of execution stability of the learned variable impedance model [33]. For the variable impedance control, stability must be guaranteed for all the pos-

sible range of variation of the impedance parameters. This issue has been recently addressed in [34], wherein a variable impedance control policy was proposed, by designing a tank-based admittance control strategy and the impedance parameters all can be passively changed. The main idea of this method is to evaluate the power balance of the robot and it ensures that the amount of energy pumped into the system is always less than the dissipated energy, and thus the system remains passive. This approach analyzed the stability for the cases that the reference trajectory is fixed throughout the motion [35].

According to the current research progress and present problems, the main contents in this paper are as follows:

1) Firstly, we emphasize that the kinematic profile (the trajectory of task) is very important for completing tasks in some applications. Especially teaching large position deviation [9, 10] in a narrow task space is not feasible, such as Minimally Invasive Surgery. Therefore, we propose a learning framework based on a simple teaching interface, so as the stiffness and damping profile can be learned from demonstrations without destroying the desired kinematic profiles (It is unnecessary to intentionally make a deviation from the teaching trajectory). That also means there is no need to provide kinematic and compliant teaching separately, the kinematic and 3-DoFs impedance profiles can be learned simultaneously from demonstrations.

2) Owing to an increasing number of robots are working on multiproduct and small-batch production lines or providing a wide range of services to meet people's daily needs, robots need to master new skills frequently and rapidly, the learning speed of algorithms is an important evaluation metric for learning skills from demonstrations. Therefore an effective sequential learning method is developed in this work, which sequentially exploits the new input data of kinematic profiles and interactive forces to learn the control model in demonstration phase. Then the learned controller is used to reproducing the robotic compliant manipulation. After the demonstration phase (learning phase), the learned controller is fixed until the next demonstration start. The proposed learning method does not require the time-consuming off-line training process, and thus the robot learning system becomes high efficiency and easy using. In addition, the learning process does not need a large number of samples, a control policy can be learned from one or more demonstrations.

3) A force-based variable impedance learning process along the x, y and z axes is developed. More specifically, the compliances along the three directions change independently which makes robot more flexible than the previous works using uniform compliance changing in three directions.

4) Furthermore, the stability analysis and the corresponding constraints for the learned impedance parameters are presented. Therefore, the global asymptotic stability of variable impedance control can be guaranteed through the stability considerations in the learning process, which allows the robot to safely operate in the environments where human beings and robot share the workspace.

To validate the proposed learning framework, we test in three different experiments. The first experiment illustrates the capability of the sequential learning algorithm in modeling kinematic profiles while the learned model being able to generalize to unseen initial points and reach the desired position precisely. The second experiment shows the proposed method can learn variable impedance parameters (stiffness and damping value) in selective directions based on the teaching forces. Meanwhile, the stability constraints of variable impedance control are confirmed by the test. The third experiment shows how a unification controller can be learned in a realistic task teaching.

The rest of the paper is organized as follows. Section 2 presents the control framework of this work. The proposed interface for demonstrations and the learning framework of the compliance behaviors are described in Section 3. In Section 4, we develop the global asymptotic stability conditions of the proposed methods. Section 5 presents the experimental evaluations. Finally, the conclusion and future work are presented in Section 6.

2. Controller and Problem Formulation

Impedance control is a very effective way of implementing the robotic compliant behavior, which makes a manipulator equivalent to a multidimensional mass-spring-damper system with the desired impedance parameters (inertia, stiffness and damping matrix). The goal of impedance control in the task space is to maintain the relationship between the position error $\tilde{x} = x_d - x$ and the exerted force F_e , the controller is given by:

$$M\ddot{\tilde{x}} = F_e - B_q \dot{\tilde{x}} - K_q \dot{\tilde{x}}.$$
(1)

where x and $x_d \in \mathbb{R}^3$ represent the current and desired position of the robot end-effector, $M, B_q, K_q \in \mathbb{R}^{3\times 3}$ denote the desired inertia matrix, time varying damping matrix and time varying stiffness matrix respectively [36]. M, B_q and K_q are symmetric and positive definite. In this work we assign a diagonal structure to the stiffness matrix $K_q = diag([k^x_q, k^y_q, k^z_q]^T)$ by decoupling it. This is under an assumption that the positional errors only result in force variations and won't influence the orientation of the robot [10]. $F_e \in \mathbb{R}^3$ is called external fore, disturbing force or teaching force, which is obtained by a six-axis force/torque sensor on the robot end-effector in the teaching phase. But in the reproduction stage, the force F_e is real-time calculated through formula (1) according to the learned impedance parameters, desired position and the current state.

Robot motion generation can be modeled by dynamic system (DS) technology [8] as

$$\dot{x} = f(x) \tag{2}$$

where x and x are recorded in every sample interval Δt , f(.) is a nonlinearly differentiable function and is estimated from neural network modeling in motion learning stage (Section 3.2). The input to the learning process is a dataset $D = \{x^n, \dot{x}^n\}$, $n \in [1, N]$, where x^n and \dot{x}^n indicate the *n*th collected position and velocity data. Then adopting algorithm to the DS is to learn the mapping $f(\cdot) : x \to \dot{x}$. Accordingly, the equation (2) can be rewrited to discrete representation as $\dot{x}^{n-1} = \frac{x^n - x^{n-1}}{\Delta t} = f(x^{n-1})$ or $x^n = f(x^{n-1}).\Delta t + x^{n-1}$. Also the desired points can be calculated by $x_d^n = f(x_d^{n-1}).\Delta t + x_d^{n-1}$ in the reproduction stage, where $x_d^0 = x^0$ is the initial position of the robot end-effector. And the *n*th position/velocity/acceleration errors between desired point and current position are computed as follows

$$\begin{cases} \tilde{x}^n = x_d^n - x^n = G(x_d^{n-1}) - x^n \leftarrow Set : G(x_d^{n-1}) = f(x_d^{n-1}) \cdot \Delta t + x_d^{n-1} \\ \dot{\tilde{x}}^n = \dot{x}_d^n - \dot{x}^n = f(x_d^n) - \dot{x}^n = f[G(x_d^{n-1})] - \dot{x}^n \\ \ddot{\tilde{x}}^n = \ddot{x}_d^n - \ddot{x}^n = \dot{f}[G(x_d^{n-1})] - \ddot{x}^n. \end{cases}$$
(3)

Substituting the equation (3) into (1), we can obtained a unified control strategy which is combines kinematic (x_d) stiffness and damping profile as

$$F_{e}^{n} = M[\dot{f}[G(x_{d}^{n-1})] - \ddot{x}^{n}] + B_{n,\Delta t}[f[G(x_{d}^{n-1})] - \dot{x}^{n}] + K_{n,\Delta t}[G(x_{d}^{n-1}) - x^{n}]$$

$$(4)$$

where M is known in this paper, f(.), $B_{n,\Delta t}$ and $K_{n,\Delta t}$ are unknown and needed to be estimated in the learning phase. And the *n*th robot end-effector position x^n is acquired by the joint angles and the forward kinematics. x_d^n can be calculated by the learned f(.) and the previous n - 1th position x_d^{n-1} . The serial desired positions $\{x_d^n, n \in [1, N]\}$ of the robot end-effector determine the robot motion. The impedance parameters $B_{n,\Delta t}$ and $K_{n,\Delta t}$ determine the compliant behaviors of the robot when subjected to external disturbance. In this paper, we formulate the problem to learning stability unification controller with "motion generation f(.) (produce kinematic profiles x_d^n)" and "variable impedance skills $(B_{n,\Delta t} \text{ and } K_{n,\Delta t})$ " from human demonstrations.

3. Learning the Unified Control Strategy from Demonstrations

We start by giving an overview of our robotic compliant manipulation learning system. As illustrated in the schematic of Fig. 2, the system is divided into two phase (learning and reproduction). In the first phase (learning), demonstrations of the task are online collected to sequentially learn the motor skills and impedance parameters (variable stiffness and damping). In the reproduction phase, the learned motion generation and compliant skills (variable impedance parameters) are used for reproducing the robotic compliant behavior. This section shows how we address these challenging aspects by proposing a novel RLfD structure that combines a simple demonstration interface, neural network learning, and a force-based variable impedance learning process for reproducing compliant manipulation.

3.1. Interface for Demonstrations

In order to teach the robot with the desired kinematics and compliance, the teacher should communicate with the robot such that the robot knows what compliant behavior should be executed. We use a simple and low-cost interface for the demonstration collection, which only depends on the robot arm equipped with a six-axis force and torque (F/T) sensor (Barrett Technology, Inc) at the endpoint (considered as the wrist center), see Fig. 3. In the teaching stage, the teacher shows the desired kinematic profile with one hand meanwhile teacher's another hand physically demonstrates the appropriate force to the F/T sensor at corresponding direction. Meaning that while the teaching is start, the robot is set into gravity compensation



Figure 2: Block diagram of the proposed learning framework.

mode, and the links are manually moved (the forces are manually exerted) by demonstrator. Note that the teacher backs up the robot arm to eliminate the spatial displacements caused by the teaching forces along the movement. The newly received kinematic profile and force are preserved and shown to the teacher himself on a computer screen. Both force and position measurements are acquired at a sampling frequency of 500 Hz and force data are filtered by a 3-order low-pass Butterworth filter with cutoff frequency 10 Hz to eliminate the noise.



Figure 3: The figure shows the equipment of the proposed interface. The forces are measured by the F/T sensor and shown on a computer screen. The positions of robot end-effector are real-time calculated by the joint angles and Forward Kinematics.

3.2. Sequential Learning of the Kinematic Profile

Various techniques can be applied to capture the kinematics of the task, such as DMP [4], Dynamical System (DS) [6, 7] and so on. Owing to the benefits of the DS, we model the kinematic profile of the task

as an autonomous DS. In our previous work [8], the DS is modeled by the ELM:

$$\dot{\boldsymbol{x}} = O = f(\boldsymbol{x}) = \sum_{i=1}^{\tilde{N}} \beta_i g(\boldsymbol{w}_i^T \boldsymbol{x} + b_i)$$
(5)

where \tilde{N} denotes the number of hidden nodes and $b = (b_1, \dots, b_{\tilde{N}})$ expresses the hidden layer biases; in addition, we refer to the input weights w^{iutput} as $W = (w_1, w_2, ..., w_{\tilde{N}})$ with dimension $d \times \tilde{N}$ (*d* denotes the number of nodes of input and output layers), and the output weights w^{output} as $\beta = (\beta_1, \beta_2, ..., \beta_{\tilde{N}})$ with the same dimension.

Eq.(5) can be written compactly as

$$O = H\beta^T.$$
(6)

In the training phase, W and b are randomly initialized and the output weights are calculated from the following norm least-squares problem

$$\min_{\beta} ||H\beta^T - O|| \tag{7}$$

where

$$H(W,b) = \begin{bmatrix} g(\boldsymbol{w}_1^T \boldsymbol{x}_1 + b_1) & \dots & g(\boldsymbol{w}_{\tilde{N}}^T \boldsymbol{x}_1 + b_{\tilde{N}}) \\ \vdots & \dots & \vdots \\ g(\boldsymbol{w}_1^T \boldsymbol{x}_N + b_1) & \dots & g(\boldsymbol{w}_{\tilde{N}}^T \boldsymbol{x}_N + b_{\tilde{N}}) \end{bmatrix}$$

is the hidden layer output matrix, and $O = (o_1, ..., o_N)^T$ are the target values from the demonstrations. N is the total number of training samples. Owing to the fixed input weights and hidden layer biases, H(W, b)is also fixed, and thus the solution of the minimum norm least-squares is $\hat{\beta}^T = H^{\dagger}O$, where H^{\dagger} is the *Moore-Penrose generalized inverse* of the matrix H [38].

Through above analysis, our previous work [8] assumes that the training data collection is completed and the data are all available for the ELM training. Actually, the training data arrives chunk-by-chunk (in some case is one-by-one) in the teaching phase. In order to further improve the efficiency of RLfD, an effective training algorithm should be developed. Due to the input weights and the bias of the ELM are randomly generated, only the output weights are calculated in the training process. We take inspiration from online sequential extreme learning machine (OS-ELM) [39], and further investigate the online updating of output weights based on the sequentially arriving data in the DS.

By the definition of the ELM, the output weight matrix $\hat{\beta}^T = H^{\dagger}O$ is a least-squares solution of (6). Under the condition $rank(H) = \tilde{N}$, where \tilde{N} is the number of hidden nodes, H^{\dagger} (the Moore-Penrose generalized inverse of the matrix H) is given by

$$H^{\dagger} = (H^T H)^{-1} H^T.$$
(8)

Substituting (8) into $\hat{\beta}^T = H^{\dagger}O, \ \hat{\beta}^T$ becomes

$$\hat{\beta}^T = (H^T H)^{-1} H^T O. \tag{9}$$

Eq.(9) is called the least-squares solution to Eq.(6).

Step 1 Initialization phase: Given an initial data set of training $D_0 = \{\boldsymbol{x}(n), \dot{\boldsymbol{x}}(n)\}_{n=1}^{N_0}$, and $N_0 \geq \tilde{N}$.

(a) Calculate hidden layer output matrix

$$H_0(W,b) = \begin{bmatrix} g(\boldsymbol{w}_1^T \boldsymbol{x}_1 + b_1) & \dots & g(\boldsymbol{w}_{\tilde{N}}^T \boldsymbol{x}_1 + b_{\tilde{N}}) \\ \vdots & \dots & \vdots \\ g(\boldsymbol{w}_1^T \boldsymbol{x}_{N_0} + b_1) & \dots & g(\boldsymbol{w}_{\tilde{N}}^T \boldsymbol{x}_{N_0} + b_{\tilde{N}}) \end{bmatrix}$$

and $O_0 = (o_1, ..., o_{N_0})^T = (\dot{x}_1, ..., \dot{x}_{N_0})^T.$

(b) Calculate initial output weight matrix $\hat{\beta}_0^T = \phi_0^{-1} H_0^T O_0$ where $\phi_0 = H_0^T H_0$.

(c) Set q = 0.

Step 2 Sequence learning phase: When the (q+1) th chunk of data $D_{q+1} = \{ \boldsymbol{x}(n), \dot{\boldsymbol{x}}(n) \}_{n=(\sum_{j=0}^{q} N_j)+1}^{\sum_{j=0}^{q+1} N_j}$ arrives

(a) Calculate hidden layer output matrix of new added data

$$H_{q+1} = \begin{cases} g(w_1^T x_{(\sum_{j=0}^{q} N_j)+1} + b_1) & \cdots & g(w_{\tilde{N}}^T x_{(\sum_{j=0}^{q} N_j)+1} + b_{\tilde{N}}) \\ \vdots & \ddots & \vdots \\ g(w_1^T x_{\sum_{j=0}^{q+1} N_j} + b_1) & \cdots & g(w_{\tilde{N}}^T x_{\sum_{j=0}^{q+1} N_j} + b_{\tilde{N}}) \end{cases} \end{bmatrix},$$

$$O_{q+1} = (o_{(\sum_{j=0}^{q} N_j)+1}, \dots, o_{\sum_{j=0}^{q+1} N_j})^T$$
(10)

(b) Let $\delta_{q+1} = \phi_{q+1}^{-1}$.

(c) Calculate hidden layer output weight matrix of new added data

$$\delta_{q+1} = \delta_q - \delta_q H_{q+1}^T (I + H_{q+1} \delta_q H_{q+1}^T)^{-1} H_{q+1} \delta_q$$

$$\beta_{q+1}^T = \beta_q^T + \delta_{q+1} H_{q+1}^T (O_{q+1} - H_{q+1} \beta_q^T).$$
(11)

(d) Set q = q + 1, and go back to step 2.

Eq. (11) provides the recursive formulation for updating the output weight matrix, which is the sequential calculation of the least-squares solution (9). When the data update is finished, the last output weight matrix is obtained. Substituting the last output weight matrix into Eq.(5), and this learned dynamic system can be applied to reproduce the kinematic profile.

3.3. Sequential Learning of the Varying Impedance Parameters

According to the proposed interface for demonstration, we can obtain varying interaction forces. In this subsection, we will explore the variability of collected force to adjust the impedance parameters (stiffness and damping matrix). The kernel thought of the stiffness adjustment is: the larger disturbing force the robot can resist, the greater stiffness the robot will become. The stiffness matrix is symmetric and positive definite, which is built around the eigenvectors of the covariance matrix of the force data. The stiffness value along each eigenvector set to be proportional to the corresponding eigenvalue of the covariance of the

collected force perturbations.

We let $\Xi = \{F_n, q_n\}_{n=0}^N$ represent the set of observed teaching forces (exerting on the endpoint of the robot) with their corresponding time stamps, where N is the number of the acquired perturbation data. These force perturbations are counteracted by the teacher through maintaining the position on the desired trajectory. We assumed that these teaching forces don't cause the motion deviations. At time q, a stiffness matrix is estimated based on the data in Ξ with the time stamps [q - (w - 1), q]. Thus, the stiffness value is acquired based on a sliding temporal window-view of the length w over the observed interaction force data. Let U_q and L_q indicate the upper and lower bounds for the indices of data points inside the sliding window [q - (w - 1), q] which has the

$$\begin{cases}
L_q = \max\{n \in [1, 2, ..., N] : q_n \le q - (w - 1)\} \\
U_q = \min\{n \in [1, 2, ..., N] : q \le q_n\}.
\end{cases}$$
(12)

Let $W_q = U_q - L_q + 1$ denote the number of data points in the temporal window at time q. The corresponding covariance matrix of the force data in the window is represented by Σ_q that:

$$\Sigma_q = E_q^2 - E_q^1 (E_q^1)^T$$
(13)

where

$$E_q^1 = \frac{1}{W_q} \sum_{\substack{n=L_q}}^{U_q} F_n$$

$$E_q^2 = \frac{1}{W_q} \sum_{\substack{n=L_q}}^{U_q} F_n F_n^T.$$
(14)

At time q, Σ_q determines the stiffness matrix and this desired stiffness will be sent to the robot. Due to the covariance matrix Σ_q is symmetric and positive definite, it can be decomposed as the following form $\Sigma_q = P\Lambda P^{-1}$ where Λ is a diagonal matrix consisting of the eigenvalues $\lambda_q^j > 0, j = 1, 2, 3$. The columns of the matrix P are the orthonormal eigenvectors. We constructed a stiffness matrix K_q using the same eigenvectors as Σ_q :

$$K_{q} = P \begin{bmatrix} g(\lambda_{q}^{1}) & 0 & 0 \\ 0 & g(\lambda_{q}^{2}) & 0 \\ 0 & 0 & g(\lambda_{q}^{3}) \end{bmatrix} P^{-1}$$
(15)

where the eigenvalues λ_q^j are proportional to the corresponding eigenvalue of the covariance matrix Σ_q :

$$g(\lambda_q^j) = \begin{cases} k_{\max}^j, & \lambda_{\max}^j < \lambda_q^j \\ (k_{\max}^j - k_{\min}^j) \frac{\lambda_q^j - \lambda_{\min}^j}{\lambda_{\max}^j - \lambda_{\min}^j} + k_{\min}^j, & \lambda_{\min}^j < \lambda_q^j < \lambda_{\max}^j \\ k_{\min}^j, & \lambda_q^j < \lambda_{\min}^j \end{cases}$$
(16)

where the allowable values of the stiffness are bounded by $K_{max} = \text{diag}([k_{max}^1, k_{max}^2, k_{max}^3]^T)$ and $K_{min} = \text{diag}([k_{min}^1, k_{min}^2, k_{min}^3]^T)$. Besides, $\lambda_{max} = \text{diag}([\lambda_{max}^1, \lambda_{max}^2, \lambda_{max}^3]^T)$ and $\lambda_{min} = \text{diag}([\lambda_{min}^1, \lambda_{min}^2, \lambda_{min}^3]^T)$ are tunable parameters of the

 $\lambda_{max} = \text{diag}([\lambda_{max}^1, \lambda_{max}^2, \lambda_{max}^3]^T)$ and $\lambda_{min} = \text{diag}([\lambda_{min}^1, \lambda_{min}^2, \lambda_{min}^3]^T)$ are tunable parameters of the system.



Figure 4: The figure shows the sliding windows from time q to (q + 1).

As the force teaching continues, the (q + 1) th chunk of force data at time (q + 1) is added to Ξ . Let U_{q+1} and L_{q+1} indicate the upper and lower bounds for the indices of data points inside the sliding window [q - w, q + 1]. The number of data points in the temporal window at time (q + 1) is $W_{q+1} = U_{q+1} - L_{q+1} + 1$. The new data set is sorted in chronological order, and the newly arrived perturbation data is simply tucked in the correct place according to the time stamp. The sliding windows from time q to (q + 1) are shown in Fig. 4. According to the Eq. (14), the E_{q+1}^1 and E_{q+1}^2 can be calculated as:

$$E_{q+1}^{1} = \frac{1}{W_{q+1}} \sum_{i=L_{q+1}}^{U_{q+1}} F_{i} = \frac{1}{W_{q+1}} \left[\sum_{i=L_{q+1}}^{U_{q}} F_{i} + \sum_{i=U_{q}+1}^{U_{q+1}} F_{i} \right]$$

$$= \frac{1}{W_{q+1}} \left[\left(\sum_{i=L_{q}}^{U_{q}} F_{i} - \sum_{i=L_{q}}^{L_{q+1}-1} F_{i} \right) + \sum_{i=U_{q}+1}^{U_{q+1}} F_{i} \right]$$

$$= \frac{1}{W_{q+1}} \left[\left(W_{q} E_{q}^{1} - \sum_{i=L_{q}}^{L_{q}} F_{i} \right) + \sum_{i=U_{q}+1}^{U_{q+1}} F_{i} \right]$$

$$= \frac{W_{q}}{W_{q+1}} E_{q}^{1} + \frac{1}{W_{q+1}} \left(\sum_{i=U_{q}+1}^{U_{q}+1} F_{i} - \sum_{i=L_{q}}^{L_{q}+1-1} F_{i} \right)$$

$$E_{q+1}^{2} = \frac{1}{W_{q+1}} \sum_{i=L_{q+1}}^{U_{q+1}} F_{i} F_{i}^{T} = \frac{1}{W_{q+1}} \left[\sum_{i=L_{q+1}}^{U_{q}} F_{i} F_{i}^{T} + \sum_{i=U_{q}+1}^{U_{q+1}} F_{i} F_{i}^{T} \right]$$

$$= \frac{1}{W_{q+1}} \left[\left(\sum_{i=L_{q}}^{U_{q}} F_{i} F_{i}^{T} - \sum_{i=L_{q}}^{L_{q}+1-1} F_{i} F_{i}^{T} \right) + \sum_{i=U_{q}+1}^{U_{q+1}} F_{i} F_{i}^{T} \right]$$

$$= \frac{1}{W_{q+1}} \left[\left(W_{q} E_{q}^{2} - \sum_{i=L_{q}}^{L_{q+1}-1} F_{i} F_{i}^{T} \right) + \sum_{i=U_{q}+1}^{U_{q+1}} F_{i} F_{i}^{T} \right]$$

$$= \frac{W_{q}}{W_{q+1}} E_{q}^{2} + \frac{1}{W_{q+1}} \left(\sum_{i=U_{q}+1}^{U_{q+1}} F_{i} F_{i}^{T} - \sum_{i=L_{q}}^{U_{q+1}-1} F_{i} F_{i} F_{i}^{T} \right).$$
(18)

The empirical covariance of a new data set is recalculated by this algorithm. The sequential calculation of the covariance matrix from time q to q + 1 is shown below:

$$\Sigma_{q+1} = E_{q+1}^2 - E_{q+1}^1 (E_{q+1}^1)^T$$
(19)

where E_{q+1}^1 and E_{q+1}^2 are calculated from Eq. (17), (18) based on the previous E_q^1 , E_q^2 and newly arrived

data.

According to Eq. (15) and (16), we can calculate the new stiffness values K_{q+1} based on the empirical covariance of the new data set Σ_{q+1} .

Some researchers investigated the relationship between the stiffness matrix K and the damping matrix B [41, 42, 37]. Studies have shown that the damping ratio is a constant, which indicates a linear relationship between square root stiffness and the damping matrix. Accordingly, the damping matrix B of the control system can be presented as:

$$B = \gamma \sqrt{K} \tag{20}$$

where the constant $\gamma > 0$ can be tuned to adjust different tasks.

4. Stability Analysis

Section 3 proposed the sequential learning method to acquire the robot motion dynamic system $\dot{x} = f(x)$ and variable impedance parameters from demonstrations. Some researchers found that the variability of the impedance parameters produce the potential energy [34]. The potential energy is injected into the system, and it may destroy the passivity of the system. The loss of passivity severely affects the impedance control: the stable interaction is no more guaranteed, the tracking error \tilde{x} can hardly converge to zero. It is crucial to guarantee the stability of the variable impedance controller. This is especially important in the applications that the robots work close to humans [33].

In the adaptive control, it is common to construct the energy functions based on the weighted sums of the velocity and position errors. The same approach can be used for the varying impedance control to establish state-independent stability conditions relating to the stiffness and damping profiles. Consider the following Lyapunov candidate function:

$$V = \frac{1}{2}Q^T M Q + \frac{1}{2}\tilde{x}^T A_q \tilde{x}$$
⁽²¹⁾

where

$$\begin{cases} Q = \dot{\tilde{x}} + \alpha \tilde{x} \\ A_q = -\alpha^2 M + \alpha B_q + K_q. \end{cases}$$
(22)

We can guarantee that A_q is positive semidefinite through selecting an appropriate positive α . This candidate function is a generalized version of a Lyapunov function which is used for the analysis of time-varying scalar systems in [43, 44].

Stability conditions: Let M be a constant, symmetric and positive definite matrix. Let K_q and B_q be symmetric, positive definite and continuously differentiable varying stiffness and damping profiles. The equilibrium point of the system (Eq. (1)) is $\{\tilde{x} = 0, \dot{\tilde{x}} = 0\}$ ($\{x = x_d, \dot{x} = \dot{x}_d\}$ and F = 0), that is stable if there exists an $\alpha > 0$ such that $\forall q > 0$:

$$1) - \alpha^2 M + \alpha B_q + K_q \succeq 0 \tag{23a}$$

$$2)\alpha M - B_q \preceq 0 \tag{23b}$$

$$3)\dot{K}_q - 2\alpha K_q + \alpha \dot{B}_q \preceq 0. \tag{23c}$$

If the semi-definiteness in condition 3) is replaced with definiteness, the system becomes asymptotic stability.

Proof. Please see Appendix A.

Inspection of the condition 2) in Eq.(23), we can use the least conservative constrain to chose α . Due to M and B are symmetric, the condition 2) can be rewritten as:

$$\sup_{\|v\|=1} v^{T}(\alpha M)v - \inf_{\|v\|=1} v^{T} B_{q} v \le 0.$$
(24)

Because $\alpha > 0$, M and B are positive, the condition 2) can be rewritten as:

$$0 < \alpha \le \min_{q} \frac{\inf_{\|v\|=1} v^T B_q v}{\sup_{\|v\|=1} v^T M v}.$$
(25)

If condition 2) is satisfied, condition 2) can be rewritten as $-(\alpha)(\alpha M - B_q) \succeq 0$. And we can learn that the condition 1) is fulfilled as well by substituting the rewritten condition 2) $(-\alpha^2 M + \alpha B_q \succeq 0)$ into condition 1). Therefore, with the choice of Eq. (25), condition 1) and 2) in Eq.(23) are satisfied for all q.

Substituting Eq. (20) into Eq. (1), a scalar system is given by $m\ddot{\tilde{x}} + b(q)\dot{\tilde{x}} + k(q)\tilde{x} = 0$ with the constant damping ratio $b(q) = \gamma \sqrt{k(q)}$. Now consider condition 3) in Eq.(23). The damping ratio is constant and $\gamma > 0$. We can further obtain the formula $\dot{b}(q) = \frac{\gamma \dot{k}(q)}{2\sqrt{k(q)}}$. Substituting $\dot{b}(q) = \frac{\gamma \dot{k}(q)}{2\sqrt{k(q)}}$ into the condition 3) of Eq.(23), we have:

$$\dot{k}(q) - 2\alpha k(q) + \alpha \frac{\gamma \dot{k}(q)}{2\sqrt{k(q)}} \le 0.$$
(26)

Further more, the upper bound of the stiffness time-derivative in Eq.(23) is given by:

$$\dot{k}(q) \le \frac{4\alpha k^{\frac{3}{2}}(q)}{2\sqrt{k(q)} + \alpha\gamma}$$
(27)

5. Experimental Evaluations

We evaluate the performance of the proposed learning framework via three experiments. The first experiment is performed the kinematic profile of the task on the 4-DoF Baxter WAM. From this experiment, we illustrate the capability of the sequential learning algorithm in modeling kinematic profiles. Furthermore, this experiment also shows that the learned model is able to generalize to unseen initial points and reach the desired position precisely. In the second experiment, we show that the proposed method can learn variable impedance parameters (stiffness and damping value) in three directions based on the teaching forces. Meanwhile, the stability analysis of variable impedance control is confirmed by the test. The third experiment reports on the performance quantification of the sequential learning unification controller in encoding kinematic profiles and the impedance profiles of the task.

During these experiments, the motions and interaction forces are measurement through the teaching

interface presented in Section 3.1. In the teaching phase of the second and third experiments, the teacher's one hand demonstrates the interaction force and another hand backs up the robot arm to counteract the disturbances and strictly guarantees the smooth accomplishment of the demonstration task. The kinematic profiles of the task and the forces applied for demonstrations are acquired at 500 Hz. In the first experiment, we only use the kinematic profile (trajectory) data to evaluate the generalization of the learned motion generator. The second experiment only uses the force data to verify the availability of the 3-DoF stiffness and damping learning and the stability of the leaned variable compliant controller. In the third experiment, we use both the kinematic profile and force data to show the feasibility of the learned unification controller. Therefore, the purpose of three experiments is to detailed evaluate the generalization of the learned motion generator, stability of the leaned variable compliant controller and the feasibility of the learned unification controller.

5.1. Evaluation of Kinematic Profile Learning in the Operational Space

To evaluate the capability of our approach in modeling motions, we teach a 4-DoF Barrett WAM how to place a tennis ball to the desired location in front of it. We use the Cartesian coordinate system to represent the robot motions and consider the positions of the end-effector as state variables. This task is demonstrated by a human teacher four times (four start positions are different, target positions are the same), see Fig. 5. Meanwhile, the learning method as described in section 3.2 keeps receiving new data and updating the parameter β of the model until the demonstration ends. We set the number of hidden layer neurons as 150, and use the sigmoid activation function in the kinematic profile learning model. When the demonstrations ended, a robot motion generator (Eq.(2)) is acquired immediately based on the last updated parameters. In the reproduction phase, the initial input is the current robot end-effector position (initial point). The subsequent inputs are the previous calculated desired points. To better illustrate the extensiveness feature of the learned robot motion generator, we let the robot start from unseen locations (initial points different from demonstrations) in the operational space to reproduce the learned motion (place a tennis ball to the desired location). 12 trajectories generated by learned robot motion generator are displayed in Fig. 6. Figure 6 shows the executions of the task from 12 different initial points. We can see that the reproductions have the similar shapes with demonstrations and the robot can reach the target location in all the cases precisely. The mean distance between the 12 final points (generated by the learned robot motion generator) and the desired point (average points of four demonstrated target) is only 1.8565mm. This precision is more than the requirement for the accurate execution of the task.

5.2. Stability Evaluation of Variable Impedance Learning

The goal of this experiment is to verify the idea that the force-based method can learn the stiffness and damping variations in three directions independently. Besides, the stability conditions are validated in this experiment. According to Eq.(23) and Eq. (25), the conditions 1) and 2) can be guaranteed by selecting an



Figure 5: Collection of four demonstrations through kinesthetic teaching. A teacher demonstrates the robot how to places a ball in the desired position from four different starting points.



Figure 6: Generalization to unseen situations. Execution of the task from 12 arbitrary initial points after learning from 4 demonstrations.



Figure 7: Collection of interaction forces through physical teaching. The teacher applies forces in x, y, z directions respectively.

appropriate α , the stability of the varying impedance control is confirmed if all eigenvalues of Eq. (28) are negative

$$\Gamma = \dot{K}_q - 2\alpha K_q + \alpha \dot{\beta}_q.$$
⁽²⁸⁾

In the teaching phase, the teacher applies force in the z, y, x directions at 7.5s 18s, 20.5s 30s, 31s 43.5s respectively. Fig. 7 shows the recorded interaction forces through the teaching interface. In the learning process, the width of the sliding window w is set to 1 second. γ is manually set as 2. α is chosen as Eq. (25). The minimum and maximum values for the stiffness are set to $K_{min} = 150N/m$, $K_{max} = 800N/m$. The parameters λ_{min} and λ_{max} are set to 0.05 and 3, respectively.

These collected forces are shown in Fig. 7. Fig. 8 shows the resulting variable stiffness in x, y and z directions. The varying damping profiles are shown in Fig. 9, which are similar to the stiffness profiles. This is because that the relationship between the root square stiffness and the damping matrix is linear, see Eq.(20). From Fig. 8 and Fig. 9, the robot can learn the 3-DoF stiffness and damping profiles based on the proposed method. So that the robot can execute more dexterous tasks than 1-DoF stiffness and damping profiles. The force based 3D variable impedance learning method is one of our improvement than the previous 1D variable impedance learning method [10, 30]. Fig.10 shows the evolution of the eigenvalues in Eq. (28) for the variable stiffness and damping profiles. The eigenvalues remain negative definite over the time, and thus the impedance profile is validated to yield stable control. The stability of variable impedance for the stability and security of the learned controller.

5.3. Learning Compliant Task —Putting a Doll into a Box

In this section, we show that the learning framework can learn a unified control strategy as instructed by the teacher, and the controller is capable of adjusting both the robot motion and its physical interaction with the environment. The interface for compliant behavior learning has described in section 3.1. In order



Figure 8: Varying stiffness profiles resulting from the interaction forces showed in Fig. 7



Figure 9: Varying damping profiles resulting from the interaction forces showed in Fig. 7.



Figure 10: The evolution of the eigenvalues in Eq.(28) for the variable stiffness and damping profiles.

to implement the control strategy in Eq.(4), the control law for the actuator torques τ_c takes the form below:

$$\tau_c = \tau_i(\vartheta) + J(\vartheta)^T F \tag{29}$$

where $\tau_i(\vartheta)$ is a term of the feedforward gravity compensation, ϑ indicates the joint angles and $J(\vartheta)$ is the Jacobian matrix of the manipulator. F is calculated by Eq. (4) in the reproduction phase. The minimum and maximum values for stiffness are set to $K_{min} = 150N/m$, $K_{max} = 800N/m$. Empirically, the minimum stiffness is used to overcome static friction and the gravity of a handheld object. The maximum value of the stiffness is set based on the safety grounds. The parameters λ_{max} and λ_{min} are set to 3 and 0.05.

5.3.1. Description

We chose a realistic task which gets the benefits from a varying impedance profiles. The task consists of transporting a doll over an obstacle (the blue case) and reaching the target point upon a purple box, then the robot drops the doll into it (the purple box). The snapshot of the task execution is shown in the Fig. 11. The desired qualitative characteristics for this task are depicted as follows:

a). From the starting point to the right top side of the obstacle, the robot should be compliant in all directions. Because the position errors are not crucial to the completion of the task at this stage. If the robot corrects the position errors with high stiffness, the end-effector will have high accelerations which are dangerous to the nearby workers.

b). Overflying obstacle (from the right top of the obstacle to the left top of the obstacle). Once the robot reaches the right top of the obstacle, the robot should stiffen up in the z-direction until arriving at the left top of the obstacle. The goal is to avoid the robot hit the lower obstacle (the blue case) in the z-direction when a disturbance of the z-direction occurs.

c), d). Upon the purple box and put the doll in. The robot should stiffen up in the x-y directions to hold its x-y positions so that the robot can drop the doll into the purple box accurately. The reasonable position changes in z-direction do not prevent the doll falling into the purple box.



Figure 11: (a): From the starting point to the right top side of the obstacle. (b): Overflying obstacle (from the right top of the obstacle). (c, d): Upon the purple box and put the doll in.

JOURNAL PRE-PROOF

5.3.2. Teaching

According to the description above, the robot complies in all directions from the starting point to the right top side of the obstacle. Then, the robot stiffens up in the z-direction upon the obstacle and move over it. Finally, the robot stiffens up in x-y directions when it reaches the target point upon the purple box. These three steps guarantee the successful completion of the task.

In the teaching process, the teacher moves the robot to perform the desired kinematic profile with one hand meanwhile teacher's another hand exerts different disturbing forces on the end-effector according to the needs of the task at different stages. Teacher backs up the robot arm to eliminate the spatial displacements caused by the teaching forces along the movement. The robot motion and force data are acquired with 500 Hz measurement frequency. The x,y,z components of the collected force information are shown in Fig. (12). From the starting point to the right top side of the obstacle (0 - 4.1s), the teacher doesn't exert any force. At 4.1 - 9.1s, the robot is moving over the obstacle while the teacher imposes force perturbations along the z-direction. Between 17.1s and 27.2s, the robot reaches the target point over the purple box and drops the doll in. The teacher applies force in the x, y directions mainly. Note that the force applied in one direction also causes slight changes in other directions. This slight changes are inevitable and the effects on task performance can be ignored.

5.3.3. Results

The stiffness profiles have acquired from the teaching round are shown in Fig.??. Note that the stiffness in the z-direction reaches the maximum during the robot moving over the obstacle, the stiffness in x, y directions reach their maximum during the robot reaching the target and putting the doll in. We attribute this to the fact that, the obstacle (blue case) and the box (purple box) are placed parallel to the x-y plane. And when the robot moving over the obstacle, position deviation in the z-direction would result in false collisions. When the robot reaches the target point upon the purple box, the robot would fail to drop the





Figure 12: Figures show the force changes in the x, y and z directions during teaching.



Figure 14: Results of producing compliant manipulation with the learned unification controller.

doll into the purple box because of the position deviations along the x and y-directions.

In the reproduction phase, we adopt a torque control model and set the calculated desired position as the input of the neural network. Illustration of reproducing demonstration task with the learned unification controller is shown in Fig. 14. The red solid line indicates the reproduction trajectory of the task. The stiffness matrices are plotted as the ellipsoids at a sequence of points along the trajectory of the task. The ellipsoids are plot with semi-axis, the lengths of semi-axis (xr, yr, zr) are inversely proportional to the corresponding stiffness values. For example, a low stiffness along x-direction is depicted by an ellipsoid with a long semi-axis in the x-direction. The ellipsoid semi-axes in three directions from the initial point to the first blue dotted line and from the second blue dotted line to the target point are longer than the ellipsoid semi-axis between the two blue dashed lines. Longer semi-axis corresponding to the lower stiffness at those points. Between the two blue dashed lines (crossing obstacle phase), we can see that the semi-axis lengths of the ellipses in the z-direction are very small, and it shows that the stiffness of the z-direction is large (to prevent robot collides obstacle below when disturbance arise). At the target point, the semi-axis lengths of the ellipse in the x, y directions are small which denotes the stiffness in the x, y directions are large. It is determined by the requirements of the successful completion of the task. The kinematic, stiffness and damping profiles are modeled simultaneously without teaching kinematic profile and force separately in this experiment, which is a very natural way. Our learning framework exploits the data of kinematic profiles and teaching forces to sequentially learn compliant manipulation, which is a more efficient way than the training after data collection completed [8, 35] and manual programming.

6. Conclusion and future work

In this paper, we presented a learning framework to model a unification controller for robotic compliant manipulation. This framework employs a sequential learning neural network to encode robot motion, and uses a new force-based variable impedance learning method to estimate varying damping and stiffness matrix in three directions. When the demonstration is over, a unified control strategy has automatically acquired without hand-coding. Furthermore, we derived the stability constraints of the variable impedance control to guarantee the safety of the learned unification controller when it interacts with environments. Three experiments have been carried out to verify the performance of the proposed method.

In this work, we assumed that the locations of the obstacle and target point are known exactly. In the further studies, we would like to take the visual information into the learning framework such that it can timely regulate control policy once the environment (the target point and obstacle positions) changes. Additionally, we will investigate the automatic detection of failed controller. When the learned controller is no longer adaptive to the new changing environment (detected through a visual system), the system can automatically alert the users to supply new demonstrations. This will not only keep the safety but also provide a better interaction between the robot and human.

APPENDIX A PROOF OF STABILITY CONDITIONS

Proof. We used Eq. (21) as a Lyapunov function candidate for the system (1). Condition 1) is intended to ensure the positive definite of the Lyapunov candidate function V. Substituting Eq. (22) into Eq. (21), yields the Lyapunov function candidate as shown below:

$$V = \frac{1}{2} (\dot{\tilde{x}} + \alpha \tilde{x})^T M (\dot{\tilde{x}} + \alpha \tilde{x}) + \frac{1}{2} \tilde{x}^T [-\alpha^2 M + \alpha B_q + K_q] \tilde{x}.$$
(30)

The derivative of V is given by

$$\dot{V} = (\dot{\tilde{x}} + \alpha \tilde{x})^T M (\ddot{\tilde{x}} + \alpha \dot{\tilde{x}}) + \frac{1}{2} \tilde{x}^T (\alpha \dot{B}_q + \dot{K}_q) \tilde{x} + \tilde{x}^T (-\alpha^2 M + \alpha B_q + K_q) \dot{\tilde{x}} = \dot{\tilde{x}}^T M \ddot{\tilde{x}} + \alpha \tilde{x}^T M \ddot{\tilde{x}} + \dot{\tilde{x}}^T \alpha M \dot{\tilde{x}} + \tilde{x}^T \alpha^2 M \dot{\tilde{x}} + \tilde{x}^T [-\alpha^2 M + \alpha B_q + K_q] \dot{\tilde{x}} + \frac{1}{2} \tilde{x}^T [\alpha \dot{B}_q + \dot{K}_q] \tilde{x}.$$

$$(31)$$

Substituting F = 0 into Eq. (1) yields $M\ddot{\tilde{x}} = -B_q\dot{\tilde{x}} - K_q\tilde{x}$ and with some rearrangement we obtain:

$$\begin{split} \dot{V} &= \dot{\tilde{x}}^{T} \left[-B_{q} \dot{\tilde{x}} - K_{q} \tilde{x} \right] + \alpha \tilde{x}^{T} \left[-B_{q} \dot{\tilde{x}} - K_{q} \tilde{x} \right] + \dot{\tilde{x}}^{T} \alpha M \dot{\tilde{x}} \\ &+ \tilde{x}^{T} \alpha^{2} M \dot{\tilde{x}} + \tilde{x}^{T} \left[-\alpha^{2} M + \alpha B_{q} + K_{q} \right] \dot{\tilde{x}} \\ &+ \frac{1}{2} \tilde{x}^{T} \left[\alpha \dot{B}_{q} + \dot{K}_{q} \right] \tilde{x} \\ &= \dot{\tilde{x}}^{T} \left[\alpha M - B_{q} \right] \dot{\tilde{x}} \\ &+ \dot{\tilde{x}}^{T} \left[-K_{q} - \alpha B_{q} + \alpha^{2} M - \alpha^{2} M + \alpha B_{q} + K_{q} \right] \tilde{x} \\ &+ \tilde{x}^{T} \left[-\alpha K_{q} + \frac{1}{2} \alpha \dot{B}_{q} + \frac{1}{2} \dot{K}_{q} \right] \tilde{x} \\ &= \dot{\tilde{x}}^{T} \underbrace{\left[\alpha M - B_{q} \right]}_{\leq 0(see \ Theorem \ 1)} \dot{\tilde{x}} \\ &+ \tilde{x}^{T} \underbrace{\left[-\alpha K_{q} + \frac{1}{2} \alpha \dot{B}_{q} + \frac{1}{2} \dot{K}_{q} \right]}_{\leq 0(see \ Theorem \ 1)} \end{split}$$
(32)

Note that V is positive definite, decrescent and radially unbounded. We can also substitute the conditions from Theorem 1 in to Eq. (32) to confirm $\dot{V} \leq 0$ for all $\tilde{x}, \dot{\tilde{x}} \neq 0$, which concludes the proof. Replacing semi-definiteness in condition 2) or 3) of Theorem 1 with negative definiteness analogously yields a V < 0for all $\tilde{x}, \dot{\tilde{x}} \neq 0$. Therefore, the asymptotic stability is proven.

Conflict of Interest Statement

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

We confirm that the manuscript has been read and approved by all named authors and that there are no other persons who satisfied the criteria for authorship but are not listed. We further confirm that the order of authors listed in the manuscript has been approved by all of us.

We confirm that we have given due consideration to the protection of intellectual property associated with this work and that there are no impediments to publication, including the timing of publication, with respect to intellectual property. In so doing we confirm that we have followed the regulations of our institutions concerning intellectual property.

We understand that the Corresponding Author is the sole contact for the Editorial process (including Editorial Manager and direct communications with the office). He is responsible for communicating with the other authors about progress, submissions of revisions and final approval of proofs. We confirm that we have provided a correct email address which is accessible by the Corresponding Author (Yongsheng Ou: ys.ou@siat.ac.cn). Signed by all authors as follows:

Jianghua Duan jh.duan@siat.ac.cn Yongsheng Ou ys.ou@siat.ac.cn Sheng Xu sheng.xu@siat.ac.cn Ming Liu eelium@ust.hk

Acknowledgement

This research was supported in part by the National Natural Science Foundation of China under Grant U1613210, in part by the Guangdong Special Support Program (2017TX04X265), in part by the Primary Research & Development Plan of Guangdong Province (2019B090915002), in part by the Shenzhen Fundamental Research Programs (JCYJ20170413165528221).

References

- B. D. Argall, S. Chernova, M. Veloso, and B. Browning, "A survey of robot learning from demonstration," *Robot. Auto. Syst.*, vol. 57, no. 5, pp. 469–483, 2009.
- [2] A. Billard, S. Calinon, and R. Dillmann, Springer Handbook of Robotics. Cham, Switzerland: Springer International Publishing, 2016, pp. 1995–2014.
- [3] S. M. Khansari-Zadeh, K. Kronander, and A. Billard, "Learning to play minigolf: A dynamical systembased approach," Advanc. Robotic., vol. 26, no. 17, pp. 1967–1993, 2012.
- [4] A. J. Ijspeert and S. Schaal, "Learning attractor landscapes for learning motor primitives," in Proc. Adv. in Neur. Inform. Proc. Syst., Jan. 2002, pp. 1523–1530.
- [5] R. Wang, Y. Wu, W. L. Chan, and K. P. Tee, "Dynamic movement primitives plus: For enhanced reproduction quality and efficient trajectory modification using truncated kernels and local biases," in Proc. IEEE Int. Conf. Intell. Robot. and Syst. (IROS), Oct. 2016, pp. 3765–3771.
- [6] S. M. Khansari-Zadeh and A. Billard, "Learning stable nonlinear dynamical systems with gaussian mixture models," *IEEE Trans. Robot.*, vol. 27, no. 5, pp. 943–957, 2011.
- [7] K. Neumann and J. J. Steil, "Learning robot motions with stable dynamical systems under diffeomorphic transformations," *Robot. Auto. Syst.*, vol. 70, pp. 1–15, 2015.
- [8] J. Duan, Y. Ou, J. Hu, Z. Wang, S. Jin, and C. Xu, "Fast and stable learning of dynamical systems based on extreme learning machine," *IEEE Trans. Syst. Man. and Cyber.: Syst.*, vol. 49, no. 6, pp. 1175–1185, 2019.
- [9] K. Kronander and A. Billard, "Online learning of varying stiffness through physical human-robot interaction," in Proc. IEEE Int. Conf. Robot. Autom. (ICRA), May. 2012, pp. 1842–1849.

- [10] K. Kronander and A. Billard, "Learning compliant manipulation through kinesthetic and tactile humanrobot interaction," *IEEE Trans. Hapt.*, vol. 7, no. 3, pp. 367–380, Jul. 2014.
- [11] A. Schmitz, P. Maiolino, M. Maggiali, L. Natale, G. Cannata, and G. Metta, "Methods and technologies for the implementation of large-scale robot tactile sensors," *IEEE Trans. Robot.*, vol. 27, no. 3, pp. 389–400, Jun. 2011.
- [12] P. Mittendorfer and G. Cheng, "Humanoid multimodal tactile-sensing modules," *IEEE Trans. Robot.* , vol. 27, no. 3, pp. 401–410, June 2011.
- [13] R. V. Ham, T. G. Sugar, B. Vanderborght, K. W. Hollander, and D. Lefeber, "Compliant actuator designs," *IEEE Robot. Autom. Magaz.*, vol. 16, no. 3, pp. 81–94, Sep. 2009.
- [14] C. Laschi, M. Cianchetti, B. Mazzolai, L. Margheri, M. Follador, and P. Dario, "Soft robot arm inspired by the octopus," *Advanc. Robotic.*, vol. 26, no. 7, pp. 709–727, 2012.
- [15] F. Ficuciello, L. Villani, and B. Siciliano, "Variable impedance control of redundant manipulators for intuitive human-robot physical interaction," *IEEE Trans. Robot.*, vol. 31, no. 4, pp. 850–863, Aug. 2015.
- [16] J. K. Salisbury, "Active stiffness control of a manipulator in cartesian coordinates," in Proc. IEEE Int. Conf. Decis. and Contr., 1980, pp. 95–100.
- [17] W. He and Y. Dong, "Adaptive fuzzy neural network control for a constrained robot using impedance learning," *IEEE Trans. Neur. Netw. Learn. Syst.*, vol. PP, no. 99, pp. 1–13, 2017.
- [18] M. Deniša, A. Gams, A. Ude, and T. Petrič, "Learning compliant movement primitives through demonstration and statistical generalization," *IEEE Trans. Mechatr.*, vol. 21, no. 5, pp. 2581–2594, Oct. 2016.
- [19] Y. Li and S. S. Ge, "Impedance learning for robots interacting with unknown environments," IEEE Trans. Contr. Syst. Techn., vol. 22, no. 4, pp. 1422–1432, Jul. 2014.
- [20] H. Sadeghian, L. Villani, M. Keshmiri, and B. Siciliano, "Task-space control of robot manipulators with null-space compliance," *IEEE Trans. Robot.*, vol. 30, no. 2, pp. 493–506, Apr. 2014.
- [21] J. Buchli, F. Stulp, E. Theodorou, and S. Schaal, "Learning variable impedance control," Int. J. Rob. Res., vol. 30, no. 7, pp. 820–833, Jun. 2011.
- [22] L. Peternel, L. Rozo, D. Caldwell, and A. Ajoudani, "A method for derivation of robot task-frame control authority from repeated sensory observations," *IEEE Robot. Autom. Lett.*, vol. 2, no. 2, pp. 719–726, Apr. 2017.

- [23] M. Suomalainen and V. Kyrki, "Learning compliant assembly motions from demonstration," in Proc. IEEE Int. Conf. Intell. Robot. and Syst. (IROS), Oct. 2016, pp. 871–876.
- [24] D. Braun, M. Howard, and S. Vijayakumar, "Optimal variable stiffness control: formulation and application to explosive movement tasks," *Auton. Robots*, vol. 33, no. 3, pp. 237–253, Oct. 2012.
- [25] J. R. Medina, D. Sieber, and S. Hirche, "Risk-sensitive interaction control in uncertain manipulation tasks," in Proc. IEEE Int. Conf. Robot. Autom. (ICRA), May. 2013, pp. 502–507.
- [26] P. Kormushev, S. Calinon, and D. G. Caldwell, "Robot motor skill coordination with em-based reinforcement learning," in Proc. IEEE Int. Conf. Intell. Robot. and Syst. (IROS), Oct. 2010, pp. 3232–3237.
- [27] N. Roy, P. Newman, and S. Srinivasa, Tendon-Driven Variable Impedance Control Using Reinforcement Learning. MA, USA: MIT Press, 2013, pp. 504–512.
- [28] S. Calinon, I. Sardellitti, and D. G. Caldwell, "Learning-based control strategy for safe human-robot interaction exploiting task and robot redundancies," in Proc. IEEE Int. Conf. Intell. Robot. and Syst. (IROS), Oct. 2010, pp. 249–254.
- [29] A. Ajoudani, N. Tsagarakis, and A. Bicchi, "Tele-impedance: Teleoperation with Impedance Regulation Using a Body-Machine Interface," Int. J. Rob. Res., vol. 31, no. 13, pp. 1642–1656, Oct. 2012.
- [30] C. Yang, C. Zeng, Y. Cong, N. Wang and M. Wang, "A Learning Framework of Adaptive Manipulative Skills From Human to Robot," *IEEE Trans. Industr. Inform.*, vol. 15, no. 2, pp. 1153–1161, Feb. 2019.
- [31] C. Yang, C. Zeng, C. Fang, W. He and Z. Li, "A DMPs-Based Framework for Robot Learning and Generalization of Humanlike Variable Impedance Skills," *IEEE Trans. Mechatr.*, vol. 23, no. 3, pp. 1193–1203, Jun. 2018.
- [32] P. Kormushev, S. Calinon, and D. G. Caldwell, "Imitation learning of positional and force skills demonstrated via kinesthetic teaching and haptic input," Advanc. Robotic., vol. 25, no. 5, pp. 581–603, 2011.
- [33] K. Kronander and A. Billard, "Stability considerations for variable impedance control," IEEE Trans. Robot., vol. 32, no. 5, pp. 1298–1305, Oct. 2016.
- [34] F. Ferraguti, C. Secchi, and C. Fantuzzi, "A tank-based approach to impedance control with variable stiffness," in Proc. IEEE Int. Conf. Robot. Autom. (ICRA), May. 2013, pp. 4948–4953.
- [35] S. M. Khansari-Zadeh and O. Khatib, "Learning potential functions from human demonstrations with encapsulated dynamic and compliant behaviors," *Auton. Robots*, vol. 41, no. 1, pp. 45–69, 2017.
- [36] N. Hogan, "Impedance control: An approach to manipulation," in Proc. American Contr. Conf., Jun. 1984, pp. 304–313.

- [37] Z. Li, Y. Kang, Z. Xiao, and W. Song, "Human-robot coordination control of robotic exoskeletons by skill transfers," *IEEE Trans. Industr. Electr.*, vol. 64, no. 6, pp. 5171–5181, Jun. 2017.
- [38] D. Serre, Matrices: Theory and applications. New York, USA: Springer 2002.
- [39] N. Liang, G. Huang, P. Saratchandran, and N. Sundararajan, "A fast and accurate online sequential learning algorithm for feedforward networks," *IEEE Trans. Neur. Netw.*, vol. 17, no. 6, pp.1411–1423, Nov. 2006.
- [40] G. H. Golub and C. F. Van Loan, Matrix computations (3rd ed.). Baltimore, MD, USA: Johns Hopkins University Press, 1996.
- [41] H. Gomi and R. Osu, "Task-dependent viscoelasticity of human multijoint arm and its spatial characteristics for interaction with environments," *The J. Neu.*, vol. 18, no. 21, pp. 8965–8978, Nov. 1998.
- [42] R. Kearney and I. Hunter, "System identification of human joint dynamics," Crit. Rev. Biomed. Eng., vol. 18, no. 1, pp. 55–87, 1990.
- [43] J.-J. E. Slotine and W. Li, Applied nonlinear control. Englewood Cliffs, NJ, USA: Prentice Hall, 1991.
- [44] H. K. Khalil, Nonlinear systems. Englewood Cliffs, NJ, USA: Prentice Hall, 2002.



Jianghua Duan is currently pursuing the Ph.D. degree in pattern recognition and intelligent system at University of Chinese Academy of Sciences. His research interest includes robot learning from human demonstrations in an unstructured environment and compliant control for safe human robot interaction.



Yongsheng Ou received the Ph.D. degree in Department of Mechanical and Automation from the Chinese University of Hong Kong, Hong Kong, China, in 2004. He was a Postdoctoral Researcher in Department of Mechanical Engineering, Lehigh University, USA, for Five years. He is currently a Professor at Shenzhen Institutes of Advanced Technology, Chinese Academy

of Sciences. He is the author of more than 150 papers in major journals and international conferences, as well as a coauthor of the monograph on Control of Single Wheel Robots (Springer, 2005). He also serves as Associate Editor of IEEE Robotics & Automation Letters. His research interests include control of complex systems, learning human control from demonstrations, and robot navigation.



Sheng Xu received B.S. degree from Shandong University, Shandong, China, in 2011, the M.S. degree from Beihang University, Beijing, China, in 2014, both in Electrical Engineering, and the Ph.D. degree in Telecommunications Engineering from ITR, University of South Australia, Australia, in 2017.

He is currently a postdoctoral fellow with the Center for Intelligent and Biomimetic Systems, Shenzhen Institutes of Advanced Technology (SIAT), Chinese Academy of Sciences (CAS), Shenzhen, China. His main research interests include target tracking, nonlinear estimation, statistical signal processing, servo system control and robot control.



Ming Liu (S'12M'12SM'18) received the B.A. degree in automation from Tongji University, Shanghai, China, in 2005, and the Ph.D. degree from the Department of Mechanical Engineering and Process Engineering, ETH Zurich, Zurich, Switzerland, in 2013. He was a Visiting Scholar with Erlangen Nurnberg University, Erlangen, Germany, and the Fraunhofer Institute IISB, Erlangen. He is currently an Assistant Professor with the Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Hong Kong. His current research interests include autonomous mapping, visual navigation, topological mapping, and environment modeling.