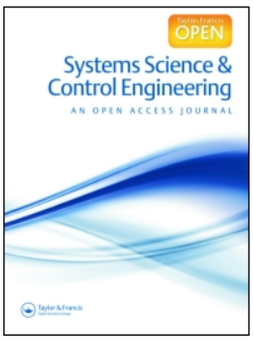

Systems Science & Control Engineering

An Open Access Journal

ISSN: (Print) 2164-2583 (Online) Journal homepage: http://www.tandfonline.com/loi/tssc20

Adaptive impedance control of robot manipulators based on Q-learning and disturbance observer

Runxian Yang, Chenguang Yang, Mou Chen & Jing Na

To cite this article: Runxian Yang, Chenguang Yang, Mou Chen & Jing Na (2017) Adaptive impedance control of robot manipulators based on Q-learning and disturbance observer, Systems Science & Control Engineering, 5:1, 287-300, DOI: 10.1080/21642583.2017.1347532

To link to this article: http://dx.doi.org/10.1080/21642583.2017.1347532

Published online: 10 Jul 2017.

SYSTEMS SCIENCE & CONTROL ENGINEERING: AN OPEN ACCESS JOURNAL, 2017
VOL. 5, 287–300
https://doi.org/10.1080/21642583.2017.1347532

Taylor & Francis
Taylor & Francis Group

Adaptive impedance control of robot manipulators based on Q-learning and disturbance observer

Runxian Yang[a,b,d], Chenguang Yang[b], Mou Chen[a] and Jing Na[c]

[a]College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, China; [b]Zienkiewicz Centre for Computational Engineering, Swansea University, Swansea, UK; [c]Faculty of Mechanical and Electrical Engineering, Kunming University of Science and Technolog, China; [d]College of Electric and IT, Yangzhou Polytechnic Institute, Yangzhou, China

ABSTRACT

In this paper, an adaptive impedance control combined with disturbance observer (DOB) is developed for a general class of uncertain robot manipulators in discrete time. The impedance control is applied to realize the interaction force control of robot manipulators in unknown, time-varying environments. The optimal reference trajectory is produced by impedance control, and the impedance parameters are achieved using Q-learning technique, which is implemented based on trajectory tracking errors. The position control with DOB of robot manipulators is implemented to track the virtual desired trajectory, and the DOB is designed to compensate for unknown compounded disturbance function by bounding both tracking error inputs and compounded disturbance inputs in a permitted control region, of which the compounded disturbance function is taken into account of all uncertain terms and external disturbances. The appropriate DOB parameters are selected applying linear matrix inequalities (LMIs) method. Both the impedance control and the bounded DOB control can well guarantee semiglobal uniform boundness of the closed-loop robot systems based on Lyapunov analysis and Schur complement theory. Simulation results are performed to test and verify effectiveness of the investigated combining adaptive impedance control with DOB.

ARTICLE HISTORY
Received 27 February 2017
Accepted 23 June 2017

KEYWORDS
Discrete-time system; time-varying environment; robot manipulator; disturbance observer; Trajectory tracking; interaction control

1. Introduction

Applications of robot manipulators have been extended to many fields, such as domestic service, medical care, industrial production and so on, and robot manipulators are anticipated to work by interacting with fragile object, other machines and even humans (Peshkin et al., 2001; Lambercy et al., 2007). On the one hand, the interaction is in unknown, time-varying, complex environment, which makes the trajectory tracking problem of nonlinear multiple-input multiple-output (MIMO) robot manipulators becomes more difficult, and on the other hand, most robot manipulators in practical application have unmodeled dynamics and uncertainties (Lewis, Dawson, & Abdallah 2004; Lewis, Jagannathan, & Yesildirak, 1998; Yang, Yang, Chen, & Na, 2016).

The problem of interaction control between robot manipulators and working environment has became increasingly important and popular. Studies of interaction control mainly involve force control and impedance control (Hogan, 1985). The impedance control focuses on selecting appropriate impedance parameters compared with force control method. The impedance control is preferred to force control in interaction, because it does not rely on a direction decomposition. Many research findings of impedance control have been applied to robot manipulators in recent two decades. The impedance control approach was firstly proposed in Hogan (1985) to introduce an ideal dynamic behaviour to the interaction control between robot manipulator and environment. In Johansson and Spong (1994) and Matinfar and Hashtrudi-Zaad (2005), the impedance control is investigated, and the impedance parameters is properly selected by applying an optimal control method as the linear quadratic regulator(LQR). The system control obtained satisfying trajectory tracking performance and force regulation, but the environment dynamics are completely known. In Jung and Hsia (2010), Hosseinzadeh, Aghabalaie, Talebi, and Shafie (2010) and Li, Sam Ge, and Yang (2012), desirable impedance parameters are chosen as constant values, while in many tasks, interaction environment is time-varying, uncertain and unstructured, the conventional impedance control methods are incapable of incorporating environment properties.

Preliminary work on estimation of impedance parameters for a robot manipulator working in an unknown environment has been studied in Diolaiti, Melchiorri, and

CONTACT Chenguang Yang cyang@theiet.org

Stramigioli (2005), a desired impedance model is constructed by precisely estimating the stiffness and damping parameters of the interaction environment. Furthermore, research work of time-varying force control for a robot manipulator is investigated in Xie, Sun, Liu, Cheng and Liu (2009), where a cosine wave reference force is tracked. However, impedance control is referred, which is only briefly mentioned in simulation section, and no theoretical analysis is provided. An automatic cell injection system is proposed in Xie, Sun, Liu, Tse and Cheng (2010), and the research focuses on time-varying force trajectory tracking. However, the method is only studied for one-link manipulator.

In our previous work Ge, Li, and Wang (2014), the developed method is verified for the time-invariant environment dynamics, such that the method is inapplicable to the time-varying environment interacted with the end-effector of robot manipulator. In Wang, Li, Ge, and Lee (2015), the optimal critic learning is proposed for unknown and time-varying environment, however, the uncertain effect of robot manipulator for trajectory tracking is not considered.

To compensate for uncertainties, many research works focus on disturbance observer (DOB) of states and external disturbances. In Wen, Zhou, Liu, and Su (2011), a robust adaptive control with DOB is designed for a class of nonlinear systems with uncertainty. The adaptive parameters are properly selected by saturating input states and compensating for external disturbances. In Xu, Lu, Zhou, and Yang$ (2004), a DOB control based on saturation of inputs and compensation for external disturbances is designed, and the state feedback theory is added to DOB control. In Yang, Fukushima, and Qin (2012), an adaptive robust control method is proposed for robot manipulators, the decentralized controller is designed by introducing a DOB and an adaptive sliding mode term to compensate for uncertainties of robot manipulators.

Most DOB methods are usually subject to compensate for external disturbances, which have be widely used in the field of trajectory tracking control for robot manipulators. However, most research studies are concentrated in continuous time. In Zeinali and Notash (2010), the dynamic model of robot manipulator is divided into two terms, the known-structure dynamics and the unknown-structure dynamics. Corresponding with the known term, a known system controller term is designed, and a feedback control and an adaptive control terms are proposed to correspond with the unmodeled dynamics. In Chen (2011), neural network (NN) control is proposed, a satisfying control performance is achieved by introducing the neural fuzzy network method, observer and sliding-mode method. In these studies, the stability of closed-loop robot control systems are reliably guaranteed, and the trajectory tracking control has obtained satisfying performance. Moreover, the digital controller of robot manipulator is applied more and more extensively at present, and the quick run speed of the digital implementation is more important in practical industry application. Recent relevant research works for nolinear uncertain robot manipulators focus on trajectory tracking control in discrete time.

In Li, Ma, Yang, and Fu (2015a), a adaptive controller is designed for a class of robot manipulators in discrete time, which have unknown fixed terms or time-varying payload uncertain terms. A satisfying control performance is obtained based on estimation for the external payload terms. However, it assumes that the uncertain terms of robot manipulators are bounded in a fixed range, and the structure of controller is complex, such that their applications in practice are limited.

Based on the above discussion, we will extend our previous works to propose an adaptive impedance control based on Q-learning and disturbance observer for an unknown, time-varying environment and an uncertain time-varying robot system. The objective of this paper is to achieve the optimal control performance of trajectory tracking requiring little knowledge of the environment and the robot dynamics. As discussed above, impedance iterative learning method, adaptive impedance control and DOB method have been developed and applied, but very few control methods have been proposed both for environments with unknown time-varying parameters and robot manipulator with nonliear uncertainties. This is the motivation to develop novel trajectory tracking control using optimal impedance control with DOB in the rest of this paper.

We highlight the contributions of this paper as follows:

- The uncertain time-varying damping-stiffness environment is described as linear stiffness system with unknown dynamic parameters.
- The optimal virtual desired reference trajectory is derived subject to unknown environment dynamics in Cartesian space by applying the impedance control with Q-learning, and the online adaptation of impedance parameters are achieved.
- The optimal position trajectory to track the virtual desired reference trajectory is obtained subject to uncertain robot system in joint space, and the compounded effect of uncertainties and disturbances is compensated by DOB with saturation.

Throughout this paper, the notations used are detailed in Table 1..

Table 1. NOMENCLATURE

Notation	Description
$\|\cdot\|$	the Euclidean norm of vectors and induced norm of matrices
$[\ ]^T$	the transpose of a vector or a matrix
$[\ ]^{-1}$	the inverse of a n-order reversible matrix
$\mathbf{0}_n$	n-dimensional zero vector
$\mathbf{0}_{a\times b}$	$a\times b$-dimension zero matrix
$\mathbf{I}_m$	m-dimensional identity matrix
x	n-dimensional position vector
f_e	n-dimensional impedance force vector
x_d	n-dimensional desired trajectory in Cartesian space
x_r	n-dimensional virtual desired reference trajectory
q	n-dimensional joint position
q_r	n-dimensional virtual desired reference joint position
τ	n-dimensional vector of control input torque
τ_e	n-dimensional external force torque

2. Preliminaries

2.1. System structure

In this paper, Study of the whole system includes a class of rigid robot manipulators and an unknown time-varying environment.

A novel trajectory tracking control method, integrating an adaptive impedance control and a DOB controlling both trajectory tracking errors and all uncertain terms, is proposed to achieve a satisfying interaction performance and a satisfying trajectory tracking performance.

In particular, the system control framework is shown in Figure 1. The framework consists of two parts: an optimal impedance control and a bounded DOB control.

In the first part, a certain optimal interaction performance between the environment and the end-effector of robot system is achieved by founding a proper impedance model, and an optimal reference trajectory is provided to the second part as the virtual desired reference. However, it is extremely difficult to identity the time-varying parameters of working environment. In this regard, the research of this paper focuses on adopting ideal Q-learning to derive a desired optimal impedance function.

In the second part, joint position control of robot manipulators is implemented to track the virtual desired trajectory produced by the impedance control in the first part. Furthermore, the DOB is designed to approximate and compensate for all uncertainties and external disturbances of robot manipulators.

2.2. System model

In this paper, we consider a system in which a class of rigid robot manipulators is physically interacting with an unknown time-varying environment.

2.2.1. Environment model

The second part of the control system in (1) is considered using a typical damping-stiffness environment, the interaction of environment and robot is described in Figure 2.

In the model, the contact parameters relate the end-effector position x to the interaction force f_e at each contact effector, C_e and K_e are unknown time-varying damping and stiffness matrices of the dynamics, respectively. Introducing an environment model proposed (Wang et al., 2015), we define that k describes the time-step index, the unknown time-varying environment

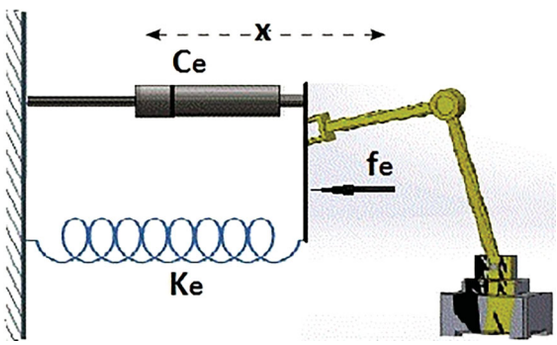

Figure 2. Interaction environment.

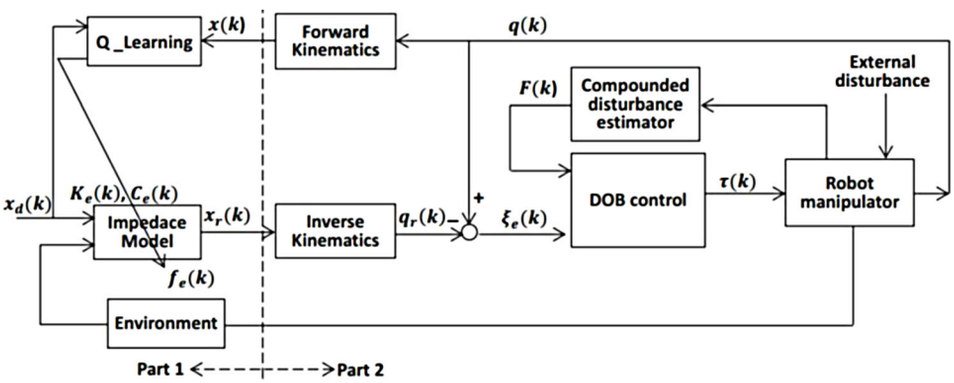

Figure 1. Control framework.

dynamics in discrete time is given as follows:

$$x(k+1) = A_e(k)x(k) + B_e(k)f_e(k) \tag{1}$$

where $x(k) \in \Re^n$ is the position state vector of end-effector, $f_e(k) \in \Re^n$ is the interaction force imposed by the environment, and $A_e(k)$ and $B_e(k)$ are parameter matrices of the environment, and they are also unknown time-varying functions of the damping matrix C_e and the stiffness matrix K_e.

This kind of damping-stiffness environment model stands for a large range of connection environment with robot system, it may represent a class of viscoelastic objects in robot works.

Assumption 2.1: The environment parameters $A_e(k)$ and $B_e(k)$ are assumed to be unknown time-varying matrices, and they are stabilizable.

Compare with the previous studies in Matinfar and Hashtrudi-Zaad (2005) and Ge et al. (2014), in this paper, research of the interaction force control and position control based on the Assumption 2.1 are more practical and more complicated.

A class of robot manipulators are required for the damping-stiffness environment model in (1) to achieve a satisfying interaction performance.

2.2.2. Impedance control

The impedance control method is introduced to obtain an optimal control performance in (1) by using a Q-learning to approximate impedance parameters.

In this paper, we adopt the desired target impedance model to implement impedance control in Cartesian space as follows Li et al. (2012):

$$\begin{aligned} -f_e(k) &= \Psi(x_d(k), x_r(k)) \\ &= C_e(k)\dot{e}_{rd}(k) + K_e(k)e_{rd}(k) \end{aligned} \tag{2}$$

where Ψ is the target impedance function, $x_d(k) \in \Re^n$ is the desired trajectory and $x_r(k) \in \Re^n$ is the virtual reference trajectory of the robot end-effector in Cartesian space, and $e_{rd}(k) = x_r(k) - x_d(k)$ is the corresponding desired tracking error.

Obviously, the end-effector of robot manipulator is described in Cartesian space, and intermediate links of the kinematic chain are to be represented in this space. However, joints of robot manipulator are in joint space. We need proceed the map between Cartesian space and joint space by inverse kinematics and forward kinematics. Furthermore, we can obtain virtual desired joint angles and virtual reference angles according to (2).

Let T represents the sampling time interval and the robot joint angels be $q \in \Re^n$ in continuous time, and the sampled joint angles $q(k) = q(t_k)$ at time $t_k = kT$. The relationship between the position in Cartesian space and the joint angles in joint space can be obtained by

$$\begin{aligned} q_r(k) &= \varphi(x_r(k)) \\ x(k) &= \psi(q(k)) \end{aligned} \tag{3}$$

where $q_r(k) \in \Re^n$ is the virtual desired joint angles in joint space, $\varphi(\cdot)$ and $\psi(\cdot)$ are the backward kinematics function and forward kinematics function of robot manipulators, respectively. The position control target is designed to make $\lim_{k\to\infty} x(k) = x_r(k)$.

2.2.3. Robot manipulator model

In this paper, the end-effector of robot manipulator physically interacts with the environment, of which the model is defined in (1), and the trajectory tracking control will be considered for n-degrees of freedom (DOF) rigid robot manipulators. The robot dynamic model is described in continuous time as follows:

$$M(q)\ddot{q} + C(q,\dot{q})\dot{q} + G(q) = \tau - \tau_e \tag{4}$$

where $q \in \Re^n$, $\dot{q} \in \Re^n$ and $\ddot{q} \in \Re^n$ are the joint angle position, velocity and acceleration, and $M(q) \in \Re^{n\times n}$, $C(q,\dot{q}) \in \Re^{n\times n}$ and $G(q) \in \Re^n$ are the symmetric positive definite inertia matrix, the Coriolis-Centrifugal torque matrix and the gravity torque vector, and $\tau \in \Re^n$ and $\tau_e = J_\tau^{\mathrm{T}}(q)f_e(k) \in \Re^n$ are the control input torque vector and the external force vector mapped to the generalized coordinates with $J_\tau(q)$ as the Jacobian matrix, respectively.

The dynamic model of robot manipulator (4) has the following properties (Lewis et al., 2004):

Property 2.1: The inertia matrix $M(q)$ is uniformly bounded, $g_1 > 0$ and $g_2 > 0$ are constants, and thus, $M(q)$ satisfies the following inequality

$$g_1 \le \|M(q)\| \le g_2 \tag{5}$$

Property 2.2: The Coriolis-Centrifugal torque matrix $C(q,\dot{q})$ and the gravity vector $G(q)$ are bounded by $\|C(q,\dot{q})\| \le \rho_c\|\dot{q}\|^2$, $\|G(q)\| \le \rho_g$, respectively, where ρ_c, ρ_g, are positive constants.

Property 2.3: The matrix $[\dot{M}(q) - C(q,\dot{q})]$ is skew symmetric, i.e.,

$$y^{\mathrm{T}}[\tfrac{1}{2}\dot{M}(q) - C(q,\dot{q})]y = 0, \quad \forall y \neq 0 \tag{6}$$

3. Impedance adaptation learning

As discussed in Section 2, an impedance control is proposed based on Q-learning method to obtain the optimal virtual desired reference trajectory $x_r(k)$.

3.1. Q-function construction

In the following section, the desired trajectory in Cartesian space is generated by adopting an exogenous system, and the *Q*-learning method is introduced to derive the optimal control, and in which we does not rely on prior information of environment and robot system.

In fact, the traditional optimal problem can be regard as the robot desired trajectory is zero, which is a special case. Further, relative robot studies are needed to make the problems identical.

In particular, the following assumption is considered:

Assumption 3.1: Assume that the desired trajectory $x_d(k)$ is generated by the following exogenous system:

$$\begin{aligned} \sigma(k+1) &= W_d\sigma(k) \\ x_d(k) &= U_d\sigma(k) \end{aligned} \tag{7}$$

where $\sigma(k) \in \Re^n$ is an observable auxiliary vector, $W_d \in \Re^{n\times n}$ and $U_d \in \Re^{n\times n}$ are known matrices, and (W_d, U_d) is also observable.

It is noted that a wide class of desired trajectory $x_d(k)$ can be determined by the exogenous system.

Assumption 3.2: The desired position trajectory $x_d(k)$ is bounded in Cartesian space.

To ensure the parameter convergence of the linear time-varying environment model (1) (Zhang, Ge, Hang, & Chai, 2000), we design a a control input to formulated the optimal control problem as follows:

$$f_e(k) = -L(k)x_r(k) \tag{8}$$

where $L(k) \in \Re^n$ is the control gain vector, which minimizes the system cost function defined in quadratic form:

$$J(k) = \sum_{k=1}^{\infty}[e_{rd}(k)^{\mathrm{T}}Se_{rd}(k) + f_e^{\mathrm{T}}(k)Rf_e(k)] \tag{9}$$

where $S \in \Re^{n\times n}$ and $R \in \Re^{n\times n}$ are weights of the end-effector position tracking error and interaction force, respectively, which satisfy $S = S^{\mathrm{T}} \geq 0$ and $R = R^{\mathrm{T}} \geq 0$.

The stabilizing feedback gain vector $L(k)$ can be calculated by using solution sequence of algebraic Riccati equation (DARE) in discrete time. According to the heuristic dynamic programming in Landelius (1997), the solution sequence $P(k+1)$ is derived as

$$\begin{aligned} P(k+1) &= A_e^{\mathrm{T}}(k)P(k)A_e(k) + S - A^{\mathrm{T}}(k)P(k)B_e(k) \\ &\quad \times [R + B_e^{\mathrm{T}}(k)P(k)B_e(k)]^{-1}B_e^{\mathrm{T}}(k)P(k)A_e(k) \\ P(0) &= n \times n \end{aligned} \tag{10}$$

Then, the control gain $L(k)$ is obtained that

$$\begin{aligned} L(k) = &-[R + B_e^{\mathrm{T}}(k)P(k+1)B_e(k)]^{-1} \\ &\times B_e^{\mathrm{T}}(k)P(k+1)A_e(k) \end{aligned} \tag{11}$$

After enough iterations, $P(k+1)$ can converges to the solution of the DARE.

Introducing the auxiliary state $\sigma(k)$ in (7) into environment model in (1), then, an extended state vector $\eta(k) \in \Re^{2n}$ is defined as follows:

$$\eta(k) = [x_r^{\mathrm{T}}(k), \quad \sigma^{\mathrm{T}}(k)]^{\mathrm{T}} \tag{12}$$

The augmented matrices of system (1) can be defined as follows:

$$\begin{aligned} \bar{A}_e(k) &= \begin{bmatrix} A_e(k) & 0 \\ 0 & W_d \end{bmatrix}, \quad \bar{B}_e(k) = \begin{bmatrix} B_e(k) \\ 0 \end{bmatrix} \\ \bar{S} &= \begin{bmatrix} S & -SU_d \\ -U_d^{\mathrm{T}}S & U_d^{\mathrm{T}}SU_d \end{bmatrix}, \quad \bar{R} = R \end{aligned} \tag{13}$$

Then, the environment model (1) can be renewed as:

$$\eta(k+1) = \bar{A}_e(k)\eta(k) + \bar{B}_e(k)f_e(k) \tag{14}$$

The corresponding function with the system cost function in (9) can be rewritten as

$$\bar{J}(k) = \sum_{k=1}^{\infty}[\eta^{\mathrm{T}}\bar{S}\eta + f_e^{\mathrm{T}}(k)\bar{R}f_e(k)] \tag{15}$$

It is noted that the cost function $\bar{J}(k)$ correlates with extended system state $\eta(k)$ and impedance force $f_e(k)$.

Similarly, the control input law (8) can be renewed as

$$f_e(k) = -\bar{L}(k)\eta(k) \tag{16}$$

where $\bar{L} = [\bar{L}_1^{\mathrm{T}}(k), \bar{L}_2^{\mathrm{T}}(k)]^{\mathrm{T}}$, $\bar{L}_1(k) \in \Re^{n\times n}$ and $\bar{L}_2(k) \in \Re^{n\times n}$ are control gains for the state system $x(k)$ and the auxiliary state $\sigma(k)$, respectively.

According to Remark 2.1, we know that the matrix $A_e(k) + B_e(k)\bar{L}(k)$ has all its eigenvalues in the open unit disc, which also applies to the work environment (14).

We can define a cost-to-go function $V(x(k))$ with a quadratic form:

$$V(k) = \sum_{i=k}^{\infty}[\eta(i)^{\mathrm{T}}\bar{S}\eta(i) + f_e^{\mathrm{T}}(i)\bar{R}f_e(i)] = r(\eta(k), f_e(k)) + V(k+1) \tag{17}$$

where $r(\eta(k), f_e(k)) = \eta(k)^{\mathrm{T}}\bar{S}\eta(k) + f_e(k)^{\mathrm{T}}\bar{R}f_e(k)$.

The cost function (17) is minimized by finding the appropriate $f_e(k)$ in (16).

Assume the optimal impedance force $f_e^* = \arg\lim_{f_e(k)} V(k)$ exists, corresponding with cost-to-go function in (17), $V^*(k)$ is quadratic, and it can be described as follows:

$$V^*(k) = \min_{f_e(k)} V(k) = \sum_{i=k}^{\infty}[\eta(k)^{\mathrm{T}}\bar{S}\eta(k) + f_e^{*^{\mathrm{T}}}(k)\bar{R}f_e^*(k)] = \eta(k)^{\mathrm{T}}\mathbb{P}(k)\eta(k) \tag{18}$$

where $\mathbb{P}(k)$ is the solution sequence of the DARE, which is derived as

$$\begin{aligned}\mathbb{P}(k+1) = {} & \bar{A}_e^{\mathrm{T}}(k)\mathbb{P}(k)\bar{A}_e(k) + \bar{S} - \bar{A}^{\mathrm{T}}(k)\mathbb{P}(k)\bar{B}_e(k) \\ & \times [\bar{R} + \bar{B}_e^{\mathrm{T}}(k)\mathbb{P}(k)\bar{B}_e(k)]^{-1}\bar{B}_e^{\mathrm{T}}(k)\mathbb{P}(k)\bar{A}_e(k) \\ \mathbb{P}(0) = {} & \mathbf{0}_{2n\times 2n}\end{aligned} \tag{19}$$

Further more, $\bar{L}(k)$ in (16) can be calculated by using solution sequence of DARE, such that we have

$$\bar{L}(k) = -[\bar{R} + \bar{B}_e^{\mathrm{T}}(k)\mathbb{P}(k+1)\bar{B}_e(k)]^{-1} \times \bar{B}_e^{\mathrm{T}}(k)\mathbb{P}(k+1)\bar{A}_e(k) \tag{20}$$

Consider our previous results in Wang et al. (2015) and the cost-to-go function in (17), a Q-function with quadratic form is introduced as follows:

$$\begin{aligned}Q(\eta(k), f_e(k)) &= \sum_{i=k}^{\infty}[\eta(i)^{\mathrm{T}}\bar{S}\eta(i) + f_e^{\mathrm{T}}(i)\bar{R}f_e(i)] \\ &= r(\eta(k), f_e(k)) + Q(\eta(k+1), f_e(k+1)) \\ &= \begin{bmatrix}\eta(k) \\ f_e(k)\end{bmatrix}^{\mathrm{T}} H(k) \begin{bmatrix}\eta(k) \\ f_e(k)\end{bmatrix}\end{aligned} \tag{21}$$

where $H(k)$ is a parameter matrix, and it is written as follows

$$H(k) = \begin{bmatrix}H_{11} & H_{12} \\ H_{21} & H_{22}\end{bmatrix} \tag{22}$$

where

$$\begin{aligned}H_{11} &= \bar{A}_e^{\mathrm{T}}(k)\mathbb{P}(k+1)\bar{A}_e(k) + \bar{S} \\ H_{12} &= \bar{A}_e^{\mathrm{T}}(k)\mathbb{P}(k+1)\bar{B}_e(k) \\ H_{21} &= H_{12}^{\mathrm{T}} \\ H_{22} &= \bar{B}_e^{\mathrm{T}}(k)\mathbb{P}(k+1)\bar{B}_e(k) + \bar{R}\end{aligned} \tag{23}$$

It is easy to prove that the matrix H describing the Q-function is positive semi-definite.

The goal of $f_e(k)$ is to determine the optimal control law:

$$f_e^*(k) = \arg\lim_{f_e(k)} Q(\eta(k), f_e(k)) \tag{24}$$

Note the corresponding Q-function $Q^*(\eta(k), f_e(k)) = \lim_{f_e(k)} Q(\eta(k), f_e(k))$ is also quadratic, when $Q^*(\eta(k), f_e^*(k))$ exists:

$$Q^*(\eta(k), f_e^*(k)) = \eta(k)^{\mathrm{T}}\mathbb{P}(k)\eta(k) \tag{25}$$

Employ the optimization algorithm based on the gradient as below:

$$\frac{\partial Q(\eta(k), f_e(k))}{\partial f_e(k)} = (\bar{B}_e^{\mathrm{T}}(k)\mathbb{P}(k+1)\bar{B}_e(k) + \bar{R})^{-1}\bar{B}_e^{\mathrm{T}}(k)\mathbb{P}(k+1)\bar{A}_e(k)\eta(k) \tag{26}$$

Compare (23) with (26), the optimal control policy is acquired as:

$$f_e(k) = -\bar{L}(k)\eta(k) = -H_{22}^{-1}H_{21}\eta(k) \tag{27}$$

Noting (21) and (22), we know if the parameter $H(k)$ can be obtained by an identification method, then, the system dynamic parameters will no longer be needed. In particular, (21) equals to (25) when $f_e^*(k)$ exists, and the optimal performance will is achieved.

By introducing $f_e(k)$ into (21) and (25) without external disturbance, we have

$$\mathbb{P}(k) = [\mathbf{I}_n, \bar{L}^{\mathrm{T}}(k)]H(k)[\mathbf{I}_n, \bar{L}^{\mathrm{T}}(k)]^{\mathrm{T}} \tag{28}$$

Based on the above discussion, $Q(\eta(k), f_e(k))$ will converge to $Q^*(\eta(k), f_e^*(k))$ with the optimal control input $f_e^*(k)$.

The $f_e^*(k)$ satisfies a time-varying temporal difference equation as follows:

$$Q^*(\eta(k), f_e^*(k)) = \eta(k)^{\mathrm{T}}\bar{S}\eta(k) + f_e^{*^{\mathrm{T}}}(k)\bar{R}f_e^*(k) + Q^*(\eta(k+1), f_e^*(k+1)) \tag{29}$$

It is obvious that the unknown and time-varying environment (1) has damping-stiffness dynamics, and it is more complex by using the traditional impedance for the systems. Considering the structure of Q-function in (29) or (21), the optimal impedance control is proposed by using Q-learning in discrete time.

3.2. Impedance adaptation control with Q-learning

In this subsection, we will employ a successive Q-learning method to solve (10) to obtain the sequence matrix

$\mathbb{P}(k)$ (Wang et al., 2015), and impedance parameters are obtained by applying the Q-learning method.

The algorithm is summarized as follows:

(a) Choose a stable control vector $u_0(k)$ when the iteration index $j=0$.
(b) The evaluation solver of $Q(\eta(k), f_e(k))$ at the $(j+1)$th iteration is calculated as follows:

$$\begin{aligned} Q_{j+1}(\eta(k), f_e(k)) &= z^{\mathrm{T}}(k)\bar{H}_{j+1}(k)z(k) \\ &= z^{\mathrm{T}}(k)Dz(k) + Q_j(\eta(k+1), u_j(k)) \\ &= z^{\mathrm{T}}(k)Dz(k) + z^{\mathrm{T}}(k+1)\bar{H}_j(k+1)z^{\mathrm{T}}(k+1) \end{aligned} \tag{30}$$

where $\bar{H}(k)_{j+1}$ is the approximation of $H(k)$ at the $(j+1)$th iteration, $D = [\bar{S}\ 0; 0\ \bar{R}]$, and $z(k) = [\eta^{\mathrm{T}}(k)\ f^{\mathrm{T}}(k)]^{\mathrm{T}}$, and $z(k+1) = [\eta^{\mathrm{T}}(k+1)\ \bar{L}_j(k+1)\eta(k+1)]^{\mathrm{T}}$.
(c) $\mathbb{P}_j(k+1)$ is obtained by solving DARE (19).
(d) $\bar{L}_j$ can is obtained by solving (20).
(e) The control vector can be updated by

$$u_{j+1}(\eta(k)) = \arg\min_{f_e(k)} Q_{j+1}(\eta(k), f_e(k)) \tag{31}$$

(f) Update $j \longleftarrow j+1$, and go back to (30).

To obtain the approximative solver in (30), and achieve the optimal control performance, the recursive time-varying least square method is applied.

The $(j+1)$th step of Q-function is introduced as follows:

$$\begin{aligned} Q_{j+1}(\eta(k), f_e(k)) &= z^{\mathrm{T}}(k)Dz(k) + \bar{h}_j(k+1)^{\mathrm{T}}z(k+1) \\ &\quad \otimes z(k+1) \\ &= \bar{h}_{j+1}(k)^{\mathrm{T}}(z(k) \otimes z(k)) \end{aligned} \tag{32}$$

where $\bar{h}_{j+1}(k) = \mathrm{vec}(\bar{H}_{j+1}(k))$, of which vec() represents a linear transformation to convert a matrix into a column vector, and $\bar{h}_j(k+1) = \mathrm{vec}(\bar{H}_j(k+1))$.

Then, we can rewrite (32) by the following linear-in-parameters form as:

$$Q_{j+1}(\eta(k), f_e(k)) = \bar{h}_{j+1}(k)^{\mathrm{T}}(z(k) \otimes z(k)) = \theta^{\mathrm{T}}(k)\phi(k) \tag{33}$$

where $\theta(k)$ is system parameter vector and $\phi(k)$ is regressive vector.

The time-varying parameter $\theta(k)$ is able to be identify, the exponentially weighted RLSs (EWRLS) method (Astrom and Wittenmark, 1989) is introduced to minimize the following blockwise mean squared error (MSE):

$$D(\theta, k) = \frac{1}{2}\sum_{i=1}^{k} \alpha^{k-i}(\rho(i) - \theta^{\mathrm{T}}(i)\phi(i))^2 \tag{34}$$

where α is a design parameter with $0 < \alpha < 1$, $\rho(i) = z^{\mathrm{T}}(k)Dz(k) + \bar{h}_j(k+1)^{\mathrm{T}}z(k+1) \otimes z(k+1)$.

The parameter $\theta(k)$, minimizing (34), is given by

$$\hat{\theta}(k+1) = \hat{\theta}(k) + g(k+1)(\rho(k+1) - \hat{\theta}^{\mathrm{T}}(k)\phi(k+1)) \tag{35}$$

where the estimation gain matrix $g(x)$ is designed as follows:

$$g(k+1) = N(k)\phi(k+1)(\alpha I_n + \phi^{\mathrm{T}}(k+1)N(k)\phi(k+1))^{-1} \tag{36}$$

where the covariance matrix $N(k)$ at the kth step with

$$N(k+1) = \frac{1}{\alpha}(I_n - g(k+1)\phi^{\mathrm{T}}(k+1)N(k)) \tag{37}$$

To avoid $N(k)$ becoming too close to singularity, we define ϱ_0 and ϱ_1 are the positive scalars, and assume $\lambda_{\min}(N(k)) \le \varrho_1$. Then the covariance matrix is designed as follows:

$$N(k) = \varrho_0 I_n, \tag{38}$$

where $\lambda(\cdot)$ denotes the eigenvalue of a matrix.

Based on the above discuss about impedance control policy design and Q-learning, consider the exogenous system of the desired trajectory $x_d(k)$ in (7) and the impedance control in (27), we rewrite (27) as follows:

$$\begin{aligned} -f_e(k) &= \bar{L}(k)\eta(k) = H_{22}^{-1}H_{21}\eta(k) \\ &= \bar{L}_1(k)x_r(k) + \bar{L}_2\sigma(k) \\ &= \bar{L}_1(k)x_r(k) + \bar{L}_2(U_d^{\mathrm{T}}U_d)^{-1}U_d^{\mathrm{T}})x_d(k) \end{aligned} \tag{39}$$

Compare the optimal impedance control (39) with the desired target impedance model(2), it is obvious that the considered damping-stiffness environment has been changed to the stiffness environment.

The proposed adaptive impedance control by using Q-learning is investigated to simplify the structure of the interaction environment model, and only the stiffness term exists to achieve optimal interaction performance between the damping-stiffness environment and the robot manipulators.

4. Discrete-time trajectory tracking controller of robot manipulator

Consider the robot dynamic model (4) and the actual desired reference trajectory $x_r(k)$ in Section 2, the $q_r(k)$ is able to derive according to (2).

Define $\bar{q} = [q^{\mathrm{T}}, \dot{q}^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{2n}$, the dynamics corresponding with the model (4) is written as Li, Ma, Yang, and

Fu (2015b)

$$\dot{\bar{q}} = \Phi(q,\dot{q})\bar{q} + \Gamma(q)(\tau - \tau_e - G(q)) \tag{40}$$

with

$$\Phi(q,\dot{q}) = \begin{bmatrix} \mathbf{0}_{n\times n} & \mathbf{I}_{n\times n} \\ \mathbf{0}_{n\times n} & -M^{-1}(q)C(q,\dot{q}) \end{bmatrix}, \quad \Gamma(q) = \begin{bmatrix} \mathbf{0}_{n\times n} \\ M^{-1}(q) \end{bmatrix}$$

T is sampling time and $q(k)$ is the sampled joint angle, which are defined in (3), $v(k) = \dot{q}(t_k)$ is the sampled joint angle velocity, $\tau(k) = \tau(t_k)$ is the control torque and $\tau_e(k) = \tau_e(t_k)$ is the external torque at the sampling time instant $t_k = kT$, respectively. The equivalent robot dynamics can be derived as

$$\bar{q}(k+1) = \Theta(k)\bar{q}(k) + \Pi(k)(\tau(k) - \tau_e(k) - G(k)) \tag{41}$$

where $\bar{q}(k) = [q^{\mathrm{T}}(k), v^{\mathrm{T}}(k)]^{\mathrm{T}}$, and $G(k) = G(q(k))$ is the gravity torque vector in discrete time. $\Theta(k) \in \mathbb{R}^{2n\times 2n}$ and $\Pi(k) \in \mathbb{R}^{2n\times n}$ are counter-part matrices in discrete time corresponding with the matrices $\Phi(q,\dot{q})$ and $\Gamma(q)$ in (40) in continuous time.

$\Theta(k), \Pi(k)$ are analyzed as following:

$$\Theta(k) = \mathrm{e}^{\Phi(q(k),v(k))T}, \quad \Pi(k) = \int_{(k-1)T}^{kT} \mathrm{e}^{\Phi(q,\dot{q})t}\Gamma(q)\,\mathrm{d}t \tag{42}$$

We only can only obtain joint angle $q(k)$ and joint velocity $v(k)$ in practice, and the estimation values of $\hat{\Theta}(k) \in \mathbb{R}^{2n\times 2n}$ and $\hat{\Pi}(k) \in \mathbb{R}^{2n\times n}$ are obtained as following:

$$\hat{\Theta}(k) = \mathrm{e}^{\Phi(k)T}, \quad \hat{\Pi}(k) = \int_{(k-1)T}^{kT} \mathrm{e}^{\Phi(k)t}\Gamma(k)\,\mathrm{d}t \tag{43}$$

with $\Phi(k) = \Phi(q(k),\dot{q}(k))$ and $\Gamma(k) = \Gamma(q(k))$.

At each sampling time, the matrix $\Phi(k)$ is determined, and $\hat{\Theta}(k)$, $\hat{\Pi}(k)$ can be obtained via a numerical method at sampling time $t_k = kT$.

Considering uncertain terms and estimation errors of robot manipulators in (43), we develop the following structures for $\Theta(k)$, $\Pi(k)$ and $G(k)$ in (41):

$$\begin{aligned} \Theta(k) &= \hat{\Theta} + \Delta\Theta(k) \\ \Pi(k) &= \hat{\Pi} + \Delta\Pi(k) \\ G(k) &= \hat{G} + \Delta G(k) \end{aligned} \tag{44}$$

where $\hat{\Theta} \in \mathbb{R}^{2n\times 2n}, \hat{\Pi} \in \mathbb{R}^{2n\times n}, \hat{G} \in \mathbb{R}^n$ are the known parts, and $\Delta\Theta(k), \Delta\Pi(k), \Delta G(k)$ are the unknown parts of $\Theta(k), \Pi(k), G(k)$, respectively.

Further, we define $u(k) = \tau(k) - \hat{G}(k) \in \Re^n$ is a corresponding control input in the presence of uncertainties and disturbances, such that a standard dynamics corresponding with system model (41) is derived as follows:

$$\begin{aligned} \bar{q}(k+1) = {} & \hat{\Theta}\bar{q}(k) + \hat{\Pi}u(k) - \hat{\Pi}\tau_e(k) + d(k) \\ & + \Delta\Theta(k)\bar{q}(k) + \Delta\Pi(k)u(k) \\ & - \Delta\Pi(k)\Delta G(k) - \hat{\Pi}\Delta G(k) \\ & - \Delta\Pi(k)\tau_e(k) \end{aligned} \tag{45}$$

where $d(k) \in \mathbb{R}^{2n}$ represents a external disturbance vector which is bounded.

Assume the matching conditions, for example, system's structure property, are satisfied, then, all uncertain terms are guaranteed in a range space. Then, an unknown function vector $F(k) \in \mathbb{R}^{2n}$, consisting of uncertainties including external disturbances in uncertain elements in (45), is defined as follows:

$$\begin{aligned} F(k) &= F(\bar{q}(k), \tau(k)) \\ &= \Delta\Theta(k)\bar{q}(k) + \Delta\Pi(k)u(k) \\ &\quad - \Delta\Pi(k)\Delta G(k) - \hat{\Pi}\Delta G(k) \\ &\quad - \Delta\Pi(k)\tau_e(k) \end{aligned} \tag{46}$$

Substituting (46) into (45) yields

$$\bar{q}(k+1) = \hat{\Theta}\bar{q}(k) + \hat{\Pi}u(k) - \hat{\Pi}\tau_e(k) + F(k) \tag{47}$$

To track the reference trajectory $q_r(k)$, a new error vector is defined as $\xi_e(k) = \bar{q}(k) - \bar{q}_r(k) \in \Re^{2n}$ with $\bar{q}_r(k) = [q_r(k), \dot{q}_r(k)]$. Therefore, an error dynamics model can be described as follows:

$$\xi_e(k+1) = \hat{\Theta}\xi_e(k) + \hat{\Pi}u(k) - \hat{\Pi}\tau_e(k) + F(k) + \Lambda(k) \tag{48}$$

where $\Lambda(k) = \hat{\Theta}\bar{q}_r(k) - \bar{q}_r(k+1) \in \Re^{2n}$, and $F(k)$ can be formulated under the following assumption.

Assumption 4.1: The unknown complicated function $F(k)$ in (48) can be formulated as the following exogenous system:

$$\begin{aligned} w(k+1) &= W_f w(k) \\ F(k) &= U_f w(k) \end{aligned} \tag{49}$$

where $w(k) \in \mathbb{R}^{2n}$ is the observer parameter, and $W_f \in \mathbb{R}^{2n\times 2n}$ and $U_f \in \mathbb{R}^{2n\times 2n}$ are auxiliary matrices.

With respect to the error dynamic represented in (48), we provided the following assumptions.

Assumption 4.2: The function $F(k)$ and its partial derivatives both are continuous, and they locally uniformly are bounded in Euclidian norm as follows:

$$\|F(k)\| \le F^* \tag{50}$$

with $F^* > 0$ as a constant.

Assumption 4.3: Considering Property 2.2 and the bounded reference trajectory, we assume that the vector $\Lambda(k)$ is bounded as follows:

$$\|\Lambda(k)\| \leq \Lambda^* \tag{51}$$

with $\Lambda^* > 0$ as a constant.

To compensate the effect of robot uncertainties, we introduce the saturation method to design the position controller. The saturation function is considered as follows:

Assumption 4.4: We assume $\text{sat}(\phi(k))$ is a saturated nonlinear function, and we define $\text{sat}(\phi(k))$ as

$$\text{sat}(\phi(k)) = [\text{sat}(\phi_1(k)), \ldots, \text{sat}(\phi_n(k))]^{\mathsf{T}}, \quad i = 1, \ldots, n \tag{52}$$

Define an auxiliary control vector $\tau_u(k) = \text{sat}(K_1\xi_e(k) + K_2\hat{F}(k))$, of which $K_1 \in \mathbb{R}^{n*2n}, K_2 \in \mathbb{R}^{n*2n}$ are feedback gain matrices, and $\hat{F}(k)$ is estimation of the unknown complicated function $F(k)$, and a bounded controller for system (48) is designed as follows:

$$\begin{aligned}\hat{u}(k) &= \tau_u(k) + \tau_e - \hat{\Pi}^+ Q_r(k) \\ &= \text{sat}(K_1\xi_e(k) + K_2\hat{F}(k)) + \tau_e - \hat{\Pi}^+ Q_r(k)\end{aligned} \tag{53}$$

where $\hat{\Pi}^+$ represents pseudo inverse matrix of $\hat{\Pi}$.

For design the bounded, saturated disturbance observer, we apply the following Lemmas and Definition as:

Lemma 4.1 (Song and Wang, 2013): *Assume that $\mathcal{D} = \{D_1, D_2, \ldots, D_{2^n}\}$ is the set of $n \times n$ diagonal matrices, of which diagonal elements are either 1 or 0, if $D_l \in \mathcal{D}$, we have that $D_l^- = \boldsymbol{I}_n - D_l$ with $l = 1, 2, \ldots, 2^n$.*

Lemma 4.2 (Zheng and Wu, 2008): *we assume that $v(k) = [v_1, \ldots, v_n]^{\mathsf{T}} \in \mathbb{R}^n$ is existent auxiliary vector, if $|v_i| \leq \tau_{u_{i\max}}$, the saturated input $\text{sat}(\tau_u(k))$ is denoted as*

$$\text{sat}(\tau_u(k)) = \sum_{l=1}^{2^n} \eta_l (D_l \tau_u(k) + D_l^- v(k))$$

where $i = 1, \ldots, n$, and η_l is limited as $0 < \eta_l < 1$ and $\sum_{l=1}^{2^n} \eta_i = 1$.

Definition 4.1 (Wu, Chen, & Chen, 2015): The robot control input $\tau_u(k)$ are saturated in $\tau_{u_{i\max}}$ in a linear region, which is defined as

$$\wp(V_1, V_2) = (\xi_e(k), \hat{F}(k)) : \|V_{i,1}\xi_e(k) + V_{i,2}\hat{F}(k)\| \leq \tau_{u_{i\max}} \tag{54}$$

where $\wp(V_1, V_2) \in \mathbb{R}^{4n}$, $V_1 = [V_{1,1}, \ldots, V_{2n,1}]^{\mathsf{T}} \in \mathbb{R}^{n\times 2n}$ with $V_{i,1} \in \mathbb{R}^{1\times 2n}$, $V_2 = [V_{1,2}, \ldots, V_{2n,2}]^{\mathsf{T}} \in \mathbb{R}^{n\times 2n}$ with $V_{i,2} \in \mathbb{R}^{1\times 2n}$, and $i = 1, 2, \ldots n$.

We assume $v_i = V_{i,1}\xi_e(k) + V_{i,2}\hat{F}(k)$ satisfies $|v_i| \leq \tau_{u_{i\max}}$, such that the control input $\tau_u(k)$ in (53) can be saturated in $\tau_{u_{i\max}}$. We further define the following saturated control input $\tau_u(k)$ as

$$\begin{aligned}\tau_u(k) &= \text{sat}(K_1\xi_e(k) + K_2\hat{F}(k)) \\ &= \sum_{l=1}^{2^n} \eta_l D_l (K_1\xi_e(k) + K_2\hat{F}(k)) \\ &\quad + \sum_{l=1}^{2^n} \eta_l D_l^- (V_1\xi_e(k) + V_2\hat{F}(k))\end{aligned} \tag{55}$$

$\hat{F}(k)$, the estimation value of $F(k)$, is achieved designing the following observer as

$$\begin{aligned}\hat{w}(k) &= b(k) - K_3\xi_e(k) \\ b(k+1) &= (W + K_3U_f)\hat{w}(k) + K_3(\hat{\Theta}\xi_e(k) + \hat{\Pi}u(k) \\ &\quad - \hat{\Pi}\tau_e(k) + Q_r(k))\end{aligned} \tag{56}$$

where $w(k)$ defined in (49), $b(k) \in \mathbb{R}^{2n}$ is an auxiliary vector as the observer, $K_3 \in \mathbb{R}^{2n\times 2n}$ is design as feedback gain matrix.

Note equations (49), (48) and (56), the estimation error of uncertain terms $\tilde{F}(k) = \hat{F}(k) - F(k)$ is derived as

$$\begin{aligned}\tilde{w}(k+1) &= \hat{w}(k+1) - w(k+1) \\ &= (W + K_3U_f)\tilde{w}(k)\end{aligned} \tag{57}$$

Substituting (53) into (48), the closed loop system formulated by

$$\begin{aligned}\xi_e(k+1) &= \sum_{l=1}^{2^n} \eta_l \{(\hat{\Theta} + \Upsilon_1)\xi_e(k) + \Upsilon_2\tilde{w}(k)\} \\ &\quad + \sum_{l=1}^{2^n} \eta_l \{\Upsilon_3 w(k)\}\end{aligned} \tag{58}$$

where $\Upsilon_1 = \hat{\Pi}(D_lK_1 + D_l^-V_1)$, $\Upsilon_2 = \hat{\Pi}(D_lK_2 + D_l^-V_2)U_f$ and $\Upsilon_3 = \hat{\Pi}(D_lK_2 + D_l^-V_2)U_f + U_f$.

The closed system (58) and the uncertain error (57) are combined and formulated as:

$$\bar{\xi}_e(k+1) = \sum_{l=1}^{2^n} \eta_l \{A_s(k)\bar{\xi}_e(k) + B_s w(k)\} \tag{59}$$

with

$$\begin{aligned}\bar{\xi}_e(k) &= \begin{bmatrix}\xi_e(k) \\ \tilde{w}(k)\end{bmatrix} \quad \bar{B}_s = \begin{bmatrix}\Upsilon_3 \\ \mathbf{0}_{2n}\end{bmatrix} \\ A_s(k) &= \begin{bmatrix}\hat{\Theta} + \Upsilon_1 & \Upsilon_2 \\ \mathbf{0}_{2n} & W_f + K_3U_f\end{bmatrix}\end{aligned} \tag{60}$$

Stability of the controller and control performances can be achieved by the proof in next section.

5. Controller realization and stable analysing

Further, to guarantee that the closed control system (59) is asymptotically stable, the design parameter matrices K_1, K_2, K_3, V_1, V_2 of the bounded observer can be achieved applying Schur complement Lemma and stability method as follows:

Lemma 5.1 (Ouellette, 1981): *Give constant matrices S_{11}, S_{22} and S_{12}, which are the symmetric constant matrices, then, $S_{22} < 0$ and $S_{11} - S_{12}S_{22}^{-1}S_{12}^{\mathrm{T}} < 0$ hold if and only if*

$$\begin{bmatrix} S_{11} & S_{12} \\ S_{12}^{\mathrm{T}} & S_{22} \end{bmatrix} < 0 \tag{61}$$

In this section, the feedback gain matrix K_1 and observer gain matrix K_2 can be derived by applying the LMIs theory, and the stable of closed-loop system and the robust control performance for uncertainty, nonliear, and vary-time can be given.

We define the Lyapunov function as follows:

$$V(k) = \bar{\xi}_e^{\mathrm{T}}(k)\bar{P}\bar{\xi}_e(k) \tag{62}$$

where symmetric positive matrix $\bar{P} \in \mathbb{R}^{4n\times 4n}$ can guarantee the closed system is stable.

Further, assume the matrix $\bar{P}(k)$ exists, and we define it as

$$\bar{P}(k) = \begin{bmatrix} \bar{Q}_1^{-1} & \mathbf{0}_{2n} \\ \mathbf{0}_{2n} & \bar{P}_2 \end{bmatrix} > 0 \tag{63}$$

with $\bar{Q}_1^{-1} = \bar{P}_1 \in \mathbb{R}^{2n\times 2n} > 0$ and $\bar{P}_2 \in \mathbb{R}^{2n\times 2n} > 0$.

Then, we have $\Delta V(k) = V(k+1) - V(k)$, which can be further analyzed that

$$\Delta V(k) \le \max_{l\in[1,2^n]} \begin{bmatrix} \bar{\xi}_e(k) \\ w(k) \end{bmatrix}^{\mathrm{T}} S_1 \begin{bmatrix} \bar{\xi}_e(k) \\ w(k) \end{bmatrix} \tag{64}$$

where S_1 is a matrix, which represents as:

$$S_1 = \begin{bmatrix} A_s^{\mathrm{T}}\bar{P}A_s - \bar{P} & A_s^{\mathrm{T}}\bar{P}B_s \\ B_s^{\mathrm{T}}\bar{P}A_s & B_s^{\mathrm{T}}\bar{P}B_s \end{bmatrix} \tag{65}$$

It is obvious that $\Delta V < 0$ in (64) holds if $S_1 < 0$.

Applying the Schur complement Lemma 5.1, a new matrix $S_2 < 0$ can be obtained from matrix $S_1 < 0$, and there $S_2 < 0 \Leftrightarrow S_1 < 0$, such that the matrix S_2 can be derived as

$$S_2 = \begin{bmatrix} -\bar{P} & * & * \\ \mathbf{0}_{2n} & \mathbf{0}_{2n} & * \\ A_s & B_s & -\bar{P}^{-1} \end{bmatrix} < 0 \tag{66}$$

where '$*$' in $S_2(i,j)$ represents the transpose matrix of $S_2(j,i)$ with $i =$ and j as the index of row and column, respectively, and the following expressions are similar.

Thus, it is shown that $\Delta V < 0$ holds if and only if $S_2 < 0$ under existing positive symmetric defined matrix $\bar{P}$. Moreover, under ensuring the system is stable, the design parameters of bounded observer are achieved using the following computing and analyzing.

Substituting (63) and (59) into (67), we have

$$S_3 = \begin{bmatrix} -\bar{Q}_1^{-1} & * & * & * & * \\ \mathbf{0}_{2n} & -\bar{P}_2 & * & * & * \\ \mathbf{0}_{2n} & \mathbf{0}_{2n} & \mathbf{0}_{2n} & * & * \\ \hat{\Theta} + \Upsilon_1 & \Upsilon_2 & \Upsilon_3 & -\bar{Q}_1 & * \\ \mathbf{0}_{2n} & W_f + K_3U_f & \mathbf{0}_{2n} & \mathbf{0}_{2n} & -\bar{P}_2^{-1} \end{bmatrix} < 0 \tag{67}$$

Furthermore, we define auxiliary matrices as follows:

$$\Omega_1 = \mathrm{diag}\{\bar{Q}_1, \mathbf{I}_{2n}, \mathbf{I}_{2n}, \mathbf{I}_{2n}, \mathbf{I}_{2n}\}$$
$$\Omega_2 = \mathrm{diag}\{\mathbf{I}_{2n}, \mathbf{I}_{2n}, \mathbf{I}_{2n}, \mathbf{I}_{2n}, \bar{P}_2\}$$

Thus, a new matrix $S_4 = \Omega_2^{\mathrm{T}}(\Omega_1^{\mathrm{T}}S_3\Omega_1)\Omega_2$ can be obtained as

$$S_4 = \begin{bmatrix} -\bar{Q}_1 & * & * & * & * \\ \mathbf{0}_{2n} & -\bar{P}_2 & * & * & * \\ \mathbf{0}_{2n} & \mathbf{0}_{2n} & \mathbf{0}_{2n} & * & * \\ T_1 & \Upsilon_2 & \Upsilon_3 & -\bar{Q}_1 & * \\ \mathbf{0}_{2n} & P_2W_f + X_3U_f & \mathbf{0}_{2n} & \mathbf{0}_{2n} & -\bar{P}_2 \end{bmatrix} \tag{68}$$

where $T_1 = \hat{\Theta}\bar{Q}_1 + \hat{\Theta}D_lX_1 + \hat{\Theta}D_l^-X_2$.

It is shown that $\Delta V(k) < 0$ if and only if $S_4 < 0$, which implies that $q(k) \to q_r(k)$ and $\tilde{w}(k) \to 0$ as $k \to \infty$. Thus, the following Theorem is derived and is described as:

Theorem 5.1: *Giving auxiliary matrices U_f, W_f, if existing symmetric positive-defined matrices $\bar{P}_1 = Q_1^{-1} > 0$, $\bar{P}_2 > 0$, if existing matrices X_1, X_2, X_3 satisfy $S_4 < 0$, then, the closed-loop robot system in (59) is asymptotically stable based on the impedance control and the bounded observer, and the system control has satisfying robustness for robot manipulators with uncertainty under designing the parameters as follows:*

$$K_1 = X_1\bar{Q}_1^{-1}$$
$$V_1 = X_2\bar{Q}_1^{-1}$$
$$K_3 = \bar{P}_2^{-1}X_3$$

6. Simulation studies

To verify the validity of the proposed control method, a 2-DOF rigid robot manipulator is considered, of which the end-effector has a physical interact with the damping-stiffness environment.

The parameters of 2-DOF rigid robot manipulator are given in Table 2.

Table 2. Structure parameters of 2-DOF robot manipulator.

Parameter	Description	Value
m_1	Mass of link 1	1.0 kg
m_2	Mass of link 2	1.0 kg
l_1	Length of link 1	0.2 m
l_2	Length of link 2	0.2 m
I_1	Inertia of z-axis of link 1	0.003 kgm^2
I_2	Inertia of z-axis of link 2	0.003 kg m^2
l_{c1}	Mass centre distance of link 1	0.1 m
l_{c2}	Mass centre distance of link 2	0.1 m

The damping-stiffness environment model is described in (69):

$$A_e(k) = 1 - \frac{0.004}{0.1(\sin(5 \times 10^{-4}k) + 1.1)}$$
$$B_e(k) = -\frac{0.01}{0.1(\sin(5 \times 10^{-4}k) + 1.1)} \quad (69)$$

In joint space, the robot initial coordinates are given as $q_r(0) = q(0) = [\pi/3, 2\pi/3]^{\mathrm{T}}$. It is noted that the initial position in Cartesian space is $x_r(0) = x(0) = [0.4, 0]^{\mathrm{T}}$, the control input torque $\tau(0) = [0, 0]^{\mathrm{T}}$, the auxiliary $b(0) = [0, 0, 0, 0]^{\mathrm{T}}$, the observer design vector $w(0) = [0, 0]^{\mathrm{T}}$, the uncertain function in joint space is determined with $U_f = \mathbf{I}_4$ and $W_f = [1, 0.1, -0.1, 1; -0.2, 1, 0.3, 1; 0, -1, 0.1, 3; 0.2, 0, -1, 0.1]$.

The known components of the robot manipulator are assumed as, $\hat{\Theta} = 0.1 \times \mathbf{I}_4$, $\hat{\Pi} = [10; 01; 10; 01]$, and $\hat{G} = [0.001; 0.001]$.

By applying LMIs theory, the following parameters is obtained as:

$$K_{11} = [-50, -6; -5, -20]$$
$$K_{12} = [-5.5, -0.5; -2, -1]$$
$$K_{21} = 0.1 * [0.15, 0.25; 0.35, 0.5]$$
$$K_{22} = 0.1 * [0.15, 0.25; 0.35, 0.5]$$
$$K_{31} = [0.1, 0.2; -0.1, 0.2; 0.1, -0.1; -0.001, 0.1]$$
$$K_{32} = [0.3, 0.4; 0.1, 0.3; 0.6, 0.4; -0.001, 0.1]$$
$$V_{11} = [-0.4967, 0.0014; 0.0037, -0.5073]$$
$$V_{12} = [-0.4981, -0.0082; 0.0037, -0.4898]$$
$$V_2 = [0.5, 0, 0.5, 0; 0.5, 0.5, 0.5, 0]$$

The interaction force between environment and end-effector of robot manipulator is regulated to imposed along with the x-axis and the y-axis. The desired trajectory in Cartesian space is determined with $U_d = 1$ and $W_d = 1$.

To verify effectiveness of the investigated combining adaptive impedance control with DOB, LQR method is applied to obtained the desired impedance control based on the DARE, and the environment parameters $A_e(k)$ and $B_e(k)$ are known in simulation. The LQR method is compared with the desired impedance obtained by the proposed Q-learning method, which does not rely on the environment knowledge.

We design the saturated observer based on system state $\bar{\xi}_e(k)$ and unknown function $\hat{F}(k)$. The following simulation process are showed under the system sampling interval $T = 0.01$ s.

To show the effectiveness of the proposed method, using above design parameters K_1, K_2, K_3, V_1, V_2, the interaction performance and trajectory tracking control are shown in Figures 3–7.

In the Cartesian space, the simulation results of impedance control based Q-learning are shown in Figures 3–5, the weights $\bar{S}$ and $\bar{R}$ are given by $\bar{S} = 1$ and $\bar{R} = 0.2$. Figure 3 shows that the convergence of control gain $\bar{L}$ is demonstrated and compared. Figure 4 shows

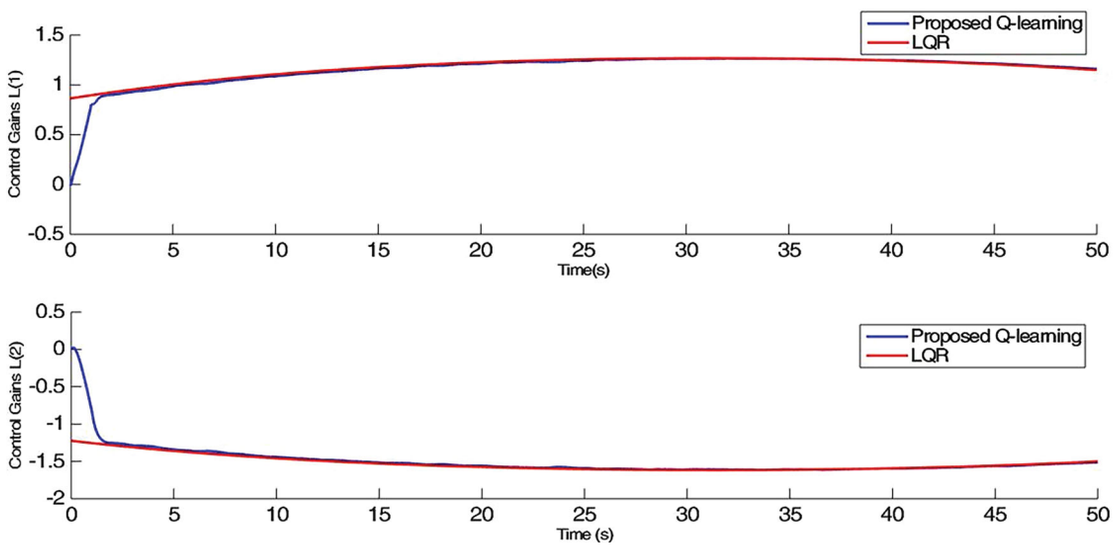

Figure 3. Impedance control gain.

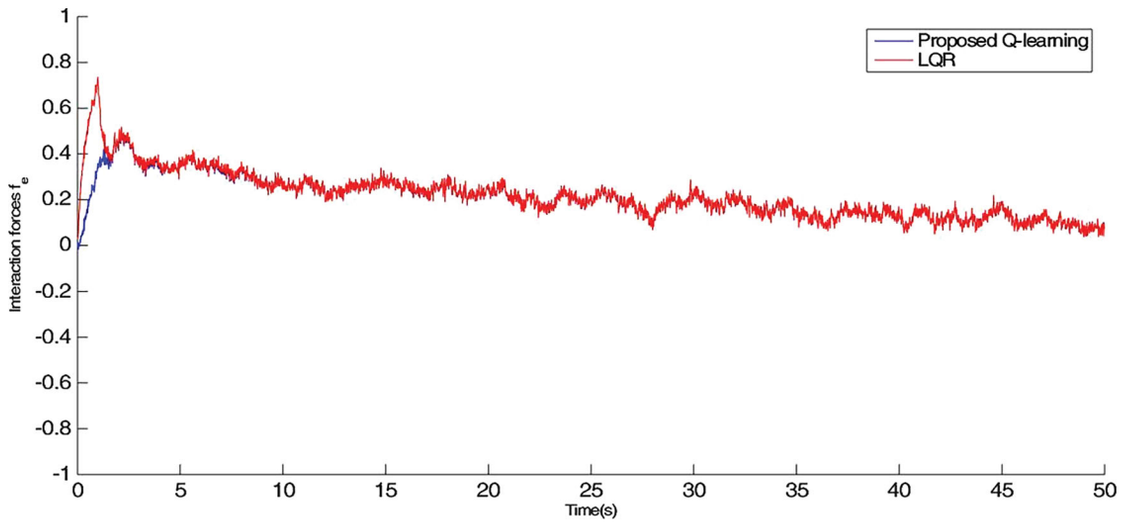

Figure 4. Interaction force between robot and environment.

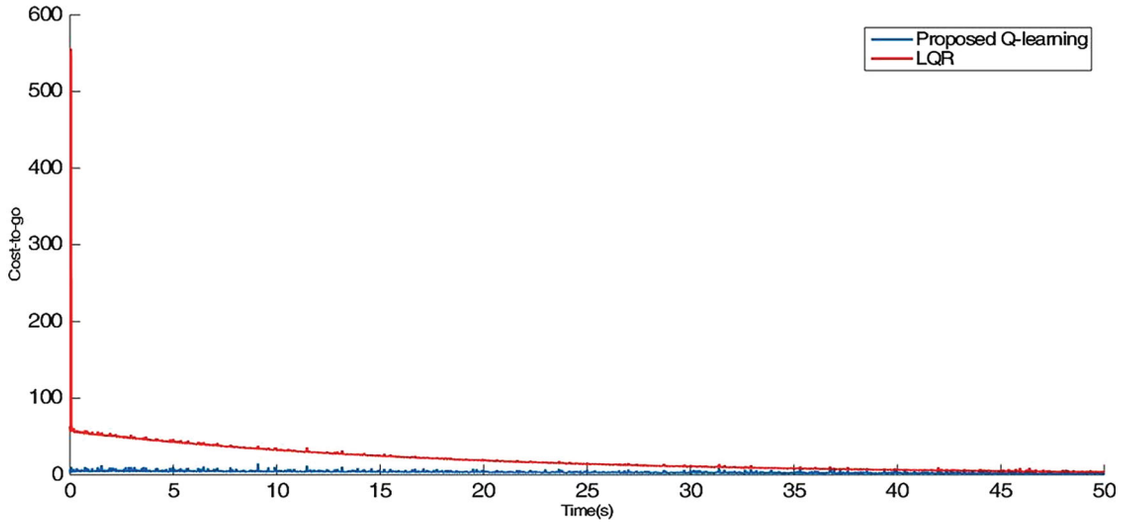

Figure 5. Cost-to-go function.

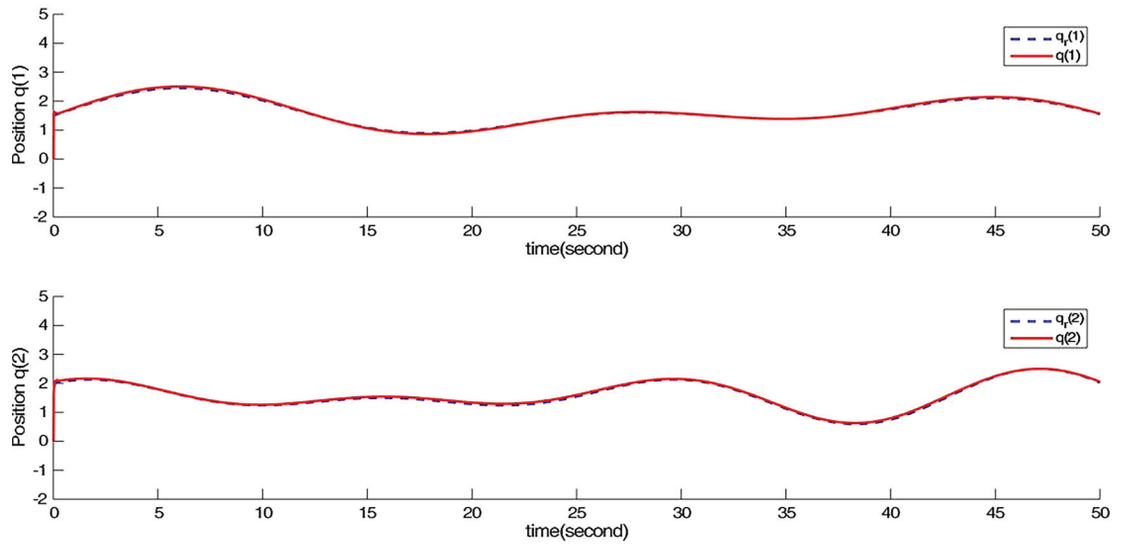

Figure 6. Joint position trajectory.

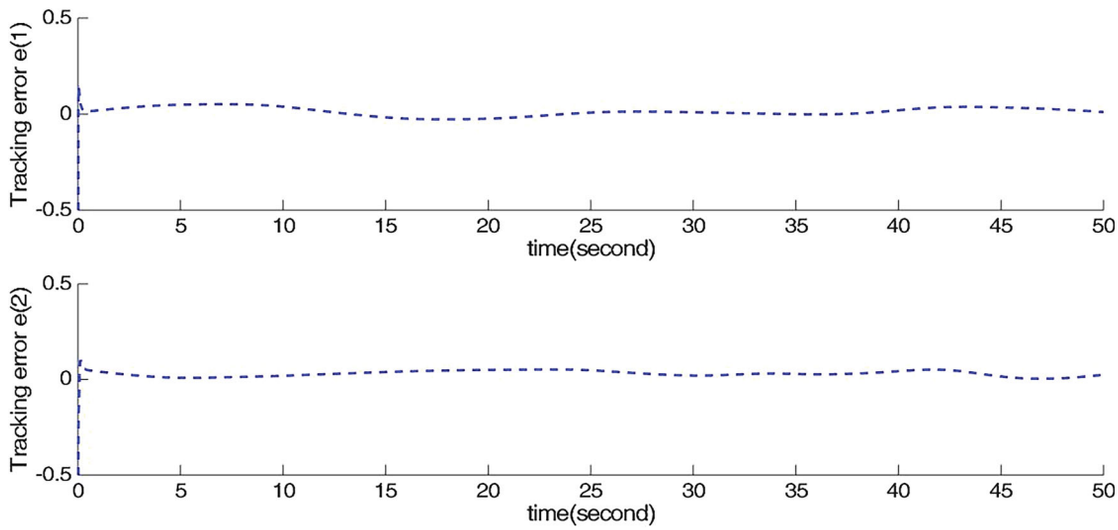

Figure 7. Joint position trajectory error.

that the interaction force f_e obtained by applying the proposed impedance control method has high tracking performance. Figure 5 shows the cost-to-go performance using proposed method, and the convergence to zero is satisfying.

In the joint space, the simulation results of impedance control based on DOB are shown in Figures 6 and 7. Figure 6 shows actual Joint position trajectories $q(1)$ and $q(2)$ compared with the desired virtual reference trajectories $q_r(1)$ and $q_r(2)$, and Figure 7 shows position tracking errors of q_1 and q_2 for the desired virtual reference trajectories $q_r(1)$ and $q_r(2)$.

Analyze the simulation results, the adaptive adjustment takes time lead to the initial errors, which are small away from the desired reference trajectories for less than 5s. We can improve control performance at the initial stage if some prior knowledge of the environment and uncertain robot have been given, and initial parameters can be properly selected.

7. Conclusion

In this paper, a new method is proposed to realize the interaction force control of uncertain robot manipulators and unknown environments. The adaptive impedance control is introduced to obtain optimal virtual reference trajectory, and the impedance parameters are adjusted by the Q-learning method in Cartesian space. The position control with bounded DOB is investigated to obtain optimal virtual trajectory for tracking the virtual reference trajectory in the joint space, and the effect of uncertainties and disturbances is compensated by bounding them in a permitted control region. The method combined Q-learning and DOB is proposed to realize the impedance adaptation, such that we obtained the optimal trajectory tracking performance in both Cartesian space and joint space, where the optimal impedance parameters of system are properly selected online without any prior knowledge both of the environment dynamics and robot dynamics. Simulation results are performed to test and verify effectiveness of the proposed adaptive impedance control method.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Astrom, K. J., & Wittenmark, B. (1989). *Adaptive control*. Reading, MA: Addison-Wesley.

Chen, C.-S. (2011). Robust self-organizing neural-fuzzy control with uncertainty observer for MIMO nonlinear systems. *IEEE Transactions on Fuzzy Systems*, *19*(4), 694–706.

Diolaiti, N., Melchiorri, C., & Stramigioli, S. (2005). Contact impedance estimation for robotic systems. *IEEE Transactions on Robotics*, *21*(5), 925–935.

Ge, S. S., Li, Y., & Wang, C. (2014). Impedance adaptation for optimal robot–environment interaction. *International Journal of Control*, *87*(2), 249–263.

Hogan, N. (1985). Impedance control: An approach to manipulation: Part II–implementation. *Journal of Dynamic Systems, Measurement, and Control*, *107*(1), 8–16.

Hosseinzadeh, M., Aghabalaie, P., Talebi, H., & Shafie, M. (2010). Adaptive hybrid impedance control of robotic manipulators. In *IECON 2010–36th Annual Conference on IEEE Industrial Electronics Society* (pp. 1442–1446). IEEE.

Johansson, R., & Spong, M. W. (1994). Quadratic optimization of impedance control. In *1994 IEEE International Conference on Robotics and Automation* (pp. 616–621). IEEE.

Jung, S., & Hsia, T. (2010). Reference compensation technique of neural force tracking impedance control for robot manipulators. In *2010 8th World Congress on Intelligent Control and Automation (WCICA)* (pp. 650–655). IEEE.

Lambercy, O., Dovat, L., Gassert, R., Burdet, E., Teo, C. L., & Milner, T. (2007). A haptic knob for rehabilitation of hand function. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *15*(3), 356–366.

Landelius, T. (1997). *Reinforcement learning and distributed local model synthesis* (Dissertations No. 469). Linköping Studies in Science and Technology, Linköping, Sweden.

Lewis, L., Dawson, D. M., & Abdallah, C. T. (2004). *Robot manipulator control theory and practice* (2nd ed.). New York: Marcel Dekker, Inc.

Lewis, F., Jagannathan, S., & Yesildirak, A. (1998). *Neural network control of robot manipulators and non-linear systems*. Bristol, PA: Taylor & Francis, Inc.

Li, J., Ma, H., Yang, C., & Fu, M. (2015a). Discrete-time adaptive control of robot manipulator with payload uncertainties. In *2015 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER)* (pp. 1971–1976). IEEE.

Li, J., Ma, H., Yang, C., & Fu, M. (2015b). Discrete-time adaptive control of robot manipulator with payload uncertainties. In *The 5th Annual IEEE International Conference on Cyber Technology in Automation, Control and Intelligent Systems* (pp. 8–12). IEEE.

Li, Y., Sam Ge, S., & Yang, C. (2012). Learning impedance control for physical robot–environment interaction. *International Journal of Control*, *85*(2), 182–193.

Matinfar, M., & Hashtrudi-Zaad, K. (2005). Optimization-based robot compliance control: Geometric and linear quadratic approaches. *The International Journal of Robotics Research*, *24*(8), 645–656.

Ouellette, D. V. (1981). Schur complements and statistics. *Linear Algebra and its Applications*, *36*, 187–295.

Peshkin, M. A., Colgate, J. E., Wannasuphoprasit, W., Moore, C. A., Gillespie, R. B., & Akella, P. (2001). Cobot architecture. *IEEE Transactions on Robotics and Automation*, *17*(4), 377–390.

Song, G., & Wang, Z. (2013). A delay partitioning approach to output feedback control for uncertain discrete time-delay systems with actuator saturation. *Nonlinear Dynamics*, *74*(1-2), 189–202.

Wang, C., Li, Y., Ge, S. S., & Lee, T. H. (2015). Optimal critic learning for robot control in time-varying environments. *IEEE Transactions on Neural Networks and Learning Systems*, *26*(10), 2301–2310.

Wen, C., Zhou, J., Liu, Z., & Su, H. (2011). Robust adaptive control of uncertain nonlinear systems in the presence of input saturation and external disturbance. *IEEE Transactions on Automatic Control*, *56*(7), 1672–1678.

Wu, B., Chen, M., & Chen, X. (2015). Observer-based bounded control for discrete time-delay uncertain nonlinear systems. *Discrete Dynamics in Nature and Society*, *2015* (2015), 1–16.

Xie, Y., Sun, D., Liu, C., Cheng, S. H., & Liu, Y. H. (2009). A force control based cell injection approach in a bio-robotics system. In *IEEE International Conference on Robotics and Automation, 2009. ICRA'09* (pp. 3443–3448). IEEE.

Xie, Y., Sun, D., Liu, C., Tse, H. Y., & Cheng, S. H. (2010). A force control approach to a robot-assisted cell microinjection system. *The International Journal of Robotics Research*, *29*(9), 1222–1232.

Xu, S., Lu, J., Zhou, S., & Yang, C. (2004). Design of observers for a class of discrete-time uncertain nonlinear systems with time delay. *Journal of the Franklin Institute*, *341*(3), 295–308.

Yang, R., Yang, C., Chen, M., & Na, J. (2016). Robust control for robot manipulators with time-varying uncertainty based on bounded observer in discrete time. In *2016 22nd International Conference on Automation and Computing (ICAC)* (pp. 383–388). IEEE.

Yang, Z.-J., Fukushima, Y., & Qin, P. (2012). Decentralized adaptive robust control of robot manipulators using disturbance observers. *IEEE Transactions on Control Systems Technology*, *20*(5), 1357–1365.

Zeinali, M., & Notash, L. (2010). Adaptive sliding mode control with uncertainty estimator for robot manipulators. *Mechanism and Machine Theory*, *45*(1), 80–90.

Zhang, T., Ge, S. S., Hang, C. C., & Chai, T. (2000). Adaptive control of first-order systems with nonlinear parameterization. *IEEE Transactions on Automatic Control*, *45*(8), 1512–1516.

Zheng, Q., & Wu, F. (2008). Output feedback control of saturated discrete-time linear systems using parameter-dependent lyapunov functions. *Systems & Control Letters*, *57*(11), 896–903.