This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2019.2912018, IEEE Transactions on Industrial Informatics

New Methods for Optimal Operational Control of Industrial Processes using Reinforcement Learning on Two Time-Scales

Wenqian Xue, Jialu Fan, *Member, IEEE*, Victor G. Lopez, *Student Member, IEEE*, Jinna Li, *Member, IEEE*, Yi Jiang, *Student Member, IEEE*, Tianyou Chai, *Fellow, IEEE*, and Frank L. Lewis, *Fellow, IEEE*

***Abstract*—Current challenges in industrial processes control include achieving optimum operation for systems with dual-rate dynamics and unknown models. This paper presents for the first time the integration of singular perturbation theory and reinforcement learning (RL) to solve this problem. To this end, an optimal operational control (OOC) problem with two-time-scale is formulated to reach the desired operational indices. Then, a singularly perturbed dynamics for two-time-scale industrial operational processes is developed by introducing a perturbed scale, resulting in the separation of the original system dynamics. Thus the original optimization problem is decomposed into a reduced slow subproblem and boundary fast subproblem. The fact that the sum of the separate solutions of these subproblems is approximately equal to the solution of the OOC problem is proven. Then, two Q-learning algorithms are proposed to obtain a composite feedback control. Finally, an industrial thickener example is employed to show the effectiveness of the proposed method.**

***Index Terms*—Optimal operational control (OOC); two-time-scales industrial processes; singular perturbation; Q-learning; reinforcement learning (RL).**

I. INTRODUCTION

Manuscript received August 1, 2018; revised January 21, 2019; accepted March 31, 2019. This work was supported in part by the National Natural Science Foundation of China under Grant 61533015, Grant 61304028, and Grant 61673280, in part by 111 Project B08015, the Fundamental Research Funds for the Central Universities N180804001. Paper no. TII-19-0201. *(Corresponding author: Jialu Fan and Jinna Li.)*

W. Xue, J. Fan, Y. Jiang and T. Chai are with the State Key Laboratory of Synthetical Automation for Process Industries and International Joint Research Laboratory of Integrated Automation, Northeastern University, Shenyang 110819, China. (e-mail: xuewenqian23@163.com, jlfan@mail.neu.edu.cn, JY369356904@163.com, tychai@mail.neu.edu.cn.)

Victor G. Lopez and F.L. Lewis are with the UTA Research Institute, the University of Texas at Arlington, Texas 76118, USA and F.L. Lewis is also a Qian Ren Consulting Professor, the State Key Laboratory of Synthetical Automation for Process Industries and International Joint Research Laboratory of Integrated Automation, Northeastern University, Shenyang 110819, China (email: victor.lopezmejia@mavs.uta.edu; lewis@uta.edu).

J. Li is with the School of Information and Control Engineering, Liaoning Shihua University, Fushun 113001, P.R. China (e-mail: lijinna_721@126.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier

HOW to reach the optimum operational status for industrial processes, e.g., ovens, grinders, thickeners, etc, is a central problem duo to optimizing the so-called operational indices is strongly desired by owners of plants for economic benefit and production safety [1]–[7]. The operational indices are performance measures that usually include economic benefit, product quality, efficiency and energy consumptions. Facing the fierce competition of global market economics, the optimal operational control (OOC) problem for industrial processes has gained critical importance and has attracted the increasing attention from many scholars. Moreover, the fast development of artificial intelligent techniques has contributed to a great progress in the OOC of industrial processes [1]–[3], [8]–[14].

The main objective in OOC is to determine the optimal controller for the unit device to ensure that the operational indices stay within their target ranges. The existing methods to solve the OOC problem include designing optimal setpoints in two-layered optimization strategies [1]–[3], [5], [8]–[12], [14]–[16] and designing optimal controllers in one-layer optimization schemes [13], [17].

Real-time optimization (RTO) was proposed to achieve optimal operation of the systems, but it requires that the lower-layer control loop stays in its steady-state value while solving the upper-layer optimization problem [17], [18]. Existing dynamic characteristics and exogenous disturbances make this approach impractical in most industrial operational processes. In this context, some approaches such as dynamic RTO, RTO integrated with model predictive control (MPC), nonlinear MPC methods and economic MPC (EMPC) have emerged to solve the OOC problem for industrial processes [3], [8]–[10], [13], [17]. These methods, however, are based on the complete knowledge of the mathematical models of the industrial processes. It is difficult to establish accurate models for industrial operational processes due to high complexity of variables, large scales and uncertain mechanism. Therefore, data-driven optimal control has become an important method preferred by many researchers and has provided increasingly promising results in applications like industrial processes, smart grids and unmanned vehicles [11], [12], [18], [19].

In recent years, case-based-reasoning intelligent control methods and RL methods have been reported to design or correct prescribed optimal controllers for large-scale complex industrial processes. RL techniques were first employed to find

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2019.2912018, IEEE Transactions on Industrial Informatics

the optimal control policy to regulation problem [20]–[24], and it is currently one of the widely used machine learning methods to seek the optimal policy in uncertain systems [25]. Neural network and other data-driven algorithms are used to calculate the control policy without requiring complete knowledge of the system dynamics [11], [12], [26]. More recent applications on neural network and machine learning are shown in [27]–[33]. RL and other intelligent control methods are applied to achieve data-driven OOC of industrial processes [1], [2], [18], [19], [34]. In [1], [2], the increment of correction depends on the operator's experience. [18] presented the neural-network based set-points design on the premise that the optimal performance indices are known a priori. The OOC algorithm in [34] does not need completely dynamics. The goal of our research is to solve the OOC problem using only data measured from the industrial processes.

It is well known that control loops for industrial devices run generally on a fast time scale, while the operational indices change on a slow time scale. The two time-scales poses a great challenge of optimizing operational indices [5], [10]–[12], [14], [16]. In [5], [10]–[12], [14], [16], [34], the lifting technique [35] is utilized for dealing with two time-scales, but this technique regards merely difference between two time-scale as sampling period difference, and is only useful for the case when the period of sampling of operational process is an integer multiple of that of unit device process. This case is impractical and is seldom true in the industrial field. Also the computation is increases as the multiple increases.

Singular perturbation approach is an alternative method to handle multi-time-scales. It is first used as a method to obtain approximate solution to differential equation in [36], and then it is applied to solve the multiple time-scale problem in control field from 1960s [37]–[40]. Compared to the lifting technique, decomposing the original OOC problem into reduced subproblems by using the singular perturbation approach considerably reduces the computational burden. Besides, using singular perturbation does not need the assumption that integer multiple relationship between fast and slow time scales and handle the time-scale problem in an improved way. To the best of our knowledge, integrating RL and singular perturbation theory to learn the solution to OOC problem for industrial operational processes with dual-rate problem using only measurable data has not been reported in the literature.

In this paper, a novel method that combines singular perturbation theory and RL techniques is designed to handle OOC problems with two-time-scales and unknown models. The proposed method can be implemented without information from either the device layer dynamics or the operational layer dynamics. The contributions of this paper are as follows:

1) In contrast to the lifting technique universally used in OOC of industrial processes with two-time-scale, this paper uses singular perturbation theory to solve the challenges induced by the time-scale problem, greatly reducing the computation burden and allowing the use of arbitrarily different time-scales.
2) Different from the model-based controllers existing in the literature, this paper develops a data-driven application of singular perturbation techniques. The resulting control design does not require knowledge of the system dynamics.
3) Singular perturbation theory and RL techniques are combined for the first time to learn in real time the approximately optimal controller for OOC problem of industrial processes in a data-driven way. This procedure results in an universally applicable method for OOC of industrial processes with two arbitrarily different time scales. The decomposition of the original OOC problem into two subproblems increases the efficiency and reduces the complexity of the control design procedure. The stabilization of the proposed composite controller derived by integrating singular perturbation theory into RL method is rigorously proven.

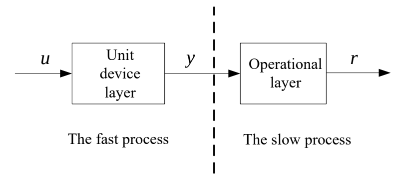


Fig. 1. The industrial process with two-time-scales.

The remainder of this paper is arranged as follows. Section II formulates singular perturbation separation of industrial processes with dual-rate sampling. In Section III, the OOC problem for the discrete-time (DT) singular perturbed system is presented. Moreover, the relationship between the solutions of the original OOC problem and the slow and the fast subproblems is given. Section IV presents slow and fast RL algorithms for finding the composite controller that drives the operational indices to the desired value at an approximate approach. Section V verifies the effectiveness of the proposed method applied to a thickener process. Conclusions are stated in Section VI.

II. Singular Perturbation Formulation for Industrial Processes with Two-Time-Scales

In this section, singular perturbation theory is applied to decompose the mathematical model of industrial processes into a fast subsystem and a slow subsystem. A detailed description of dual-rate sampling is also provided.

A. Model of an Industrial Process with Two-Time-Scales

Operation in multi-time scales is a general property of practical industrial processes. The unit devices may be ovens, grinders, thickeners, etc., and operate on a fast time scale of seconds. Their performance is measured in terms of operational indices, such as product quality, energy efficiency and resource usage, that change slowly and can only be measured on time scales corresponding to hours [9]–[12]. The structure of an industrial process with two time scales is shown in Fig. 1. The industrial operational processes are composed of unit device processes running on a fast time scale and operational indices measured on a slow time scale.

Since the unit devices in practical industrial operations are usually running at some steady states, then their nonlinear

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2019.2912018, IEEE Transactions on Industrial Informatics

dynamics can be linearized near the steady operating points [5], [9]–[12], [15], [16], [41]. Thus, the following linear dynamics is considered for the device processes

$$\dot{y}(t) = A_1 y(t) + B_1 u(t) \tag{1}$$

where $y \in \mathbb{R}^{n_y}$ is the system state vector of the process and $u \in \mathbb{R}^{n_u}$ is the control input and $\mathbb{R}$ denotes the set of the real numbers. A_1 and B_1 are matrices with appropriate dimensions.

Assumption 1. Matrix A_1 is nonsingular.

The dynamics of operational indices is also assumed to be given by linear dynamics as

$$\begin{cases} \dot{x}(t) = A_2 x(t) + B_2 y(t) \\ r(t) = C_2 x(t) \end{cases} \tag{2}$$

where $x \in \mathbb{R}^{n_x}$ is the state of the operational process and $r \in \mathbb{R}^{n_r}$ represents the operational indices. A_2, B_2 and C_2 are matrices with appropriate dimensions.

B. Singular Perturbation Approach for the Separation of Two-Time-Scales Industrial Operational Processes

Since the variables $y(t)$ in the unit device control layer change faster than the variables $x(t)$ in the operational layer, there exists the so-called gap in changing rate between the two layers. The difference in running time-scales means that the whole system has high order and strong stiffness, which leads to huge difficulties in calculation, modeling and control [11], [12], [37], [42].

The basic principle in singular perturbation theory is to use the gap between the fast and slow variables to decompose a high-order system into two or more reduced order systems [37], [42]–[44]. In this subsection, the singular perturbation method is used to decompose the global system (1)-(2) with different time-scales, resulting in two subsystems with reduced dimensions.

Introduce a small scaling factor ε, as the least speed ratio from the fast variables $y(t)$ to the slow variables $x(t)$. Define $y(t) = \varepsilon\xi(t)$, $A_1 = \frac{1}{\varepsilon}A_f$, $B_2 = \frac{1}{\varepsilon}B_s$. Substituting these new variables into (1) and (2) yields the following singularly perturbed system

$$\varepsilon\dot{\xi}(t) = A_f \xi(t) + B_1 u(t), \tag{3}$$
$$\dot{x}(t) = A_2 x(t) + B_s \xi(t), \tag{4}$$
$$r(t) = C_2 x(t). \tag{5}$$

Here, we use the same procedure as in [37] to decompose the singularly perturbed system (3)-(5). According to the singular perturbation principle, $u(t)$ and $\xi(t)$ are respectively replaced by $u(t) = \bar{u}(t) + \tilde{u}(t)$ and $\xi(t) = \bar{\xi}(t) + \tilde{\xi}(t)$, where $\bar{u}$ and $\bar{\xi}$ are the slow components of the system variables, while $\tilde{u}$ and $\tilde{\xi}$ are the fast components of the system variables.

Setting $\varepsilon = 0$ in (3) yields

$$0 = A_f \bar{\xi}(t) + B_1 \bar{u}(t). \tag{6}$$

Now, the quasi-steady state value of the slow system variables can be obtained from (6) as

$$\bar{\xi}(t) = -A_f^{-1} B_1 \bar{u}(t) \tag{7}$$

where matrix A_f is invertible because matrix A_1 in (1) was assumed to be nonsingular. Substituting (7) into (5) yields the reduced order slow subsystem dynamics

$$\dot{\bar{x}}(t) = A_2 \bar{x}(t) - B_s A_f^{-1} B_1 \bar{u}(t). \tag{8}$$

The slow component $\bar{\xi}(t)$ is keeping constant during every fast time interval. By (3) and (7), the fast subsystem dynamics is given by

$$\begin{aligned} \varepsilon\dot{\xi}(t) =& \varepsilon(\dot{\bar{\xi}}(t) + \dot{\tilde{\xi}}(t)) \\ =& A_f(\bar{\xi}(t) + \tilde{\xi}(t)) + B_1(\bar{u}(t) + \tilde{u}(t)) \\ =& -A_f A_f^{-1} B_1 \bar{u}(t) + A_f \tilde{\xi}(t) + B_1(\bar{u}(t) + \tilde{u}(t)) \\ =& A_f \tilde{\xi}(t) + B_1 \tilde{u}(t). \end{aligned} \tag{9}$$

Now, (9) and (8) can be expressed as the fast and slow subsystems respectively by

$$\varepsilon\dot{\tilde{\xi}}(t) = A_f \tilde{\xi}(t) + B_1 \tilde{u}(t) \tag{10}$$
$$\dot{\bar{x}}(t) = A_2 \bar{x}(t) - B_s A_f^{-1} B_1 \bar{u}(t). \tag{11}$$

Defining $\tau = \frac{t}{\varepsilon}$, (10) can be rewritten as

$$\frac{d\tilde{\xi}}{d\tau} = \dot{\tilde{\xi}}(\tau) = A_f \tilde{\xi}(\tau) + B_1 \tilde{u}(\tau) \tag{12}$$

which is also known as a boundary layer subsystem.

According to the Tikhonov Theorem [38], [45], there exists an $\varepsilon^* > 0$ such that if $\varepsilon \in (0, \varepsilon^*)$, then

$$\xi(t) = \tilde{\xi}(\tau) + \bar{\xi}(t) + o(\varepsilon) \tag{13}$$
$$x(t) = \bar{x}(t) + o(\varepsilon) \tag{14}$$

where $o(\varepsilon)$ is an error term of order ε.

After the singular perturbation transformation, the variables in the unit device process and the operational indices are separated as the slow and the fast components in the singularly perturbed system composed of (11) and (12) respectively. The task now is to design $\bar{u}$ and $\tilde{u}$ to obtain optimal operation in the whole industrial process.

III. DT Formulation of Two-Time-Scales Industrial Processes for Operational Control

In this section, the DT formulation of the decomposed industrial subsystems defined above is obtained. An optimal control formulation is given to find the optimal operational indices for the original system (1)-(2). Then two optimisation subproblems are defined corresponding to the fast subsystem (12) and the slow subsystem (11). The main result in theorem 1 shows that the optimal operational control problem is solved by solving the fast and slow subproblems.

A. Discretization of the Singularly Perturbed Systems

Since a digital controller is generally employed by the device control process, the reduced order slow subsystem (11) and fast subsystem (12) are now discretized by using a sampling period T such that $t = kT$. This procedure yields

$$\bar{x}(k+1) = M_s \bar{x}(k) + N_s \bar{u}(k) \tag{15}$$
$$\tilde{\xi}(k+1) = M_f \tilde{\xi}(k) + N_f \tilde{u}(k) \tag{16}$$

where $M_f = e^{A_f T}$, $N_f = \int_0^T e^{A_f t} B_1 dt$, $M_s = e^{A_2 T}$, $N_s = -\int_0^T e^{A_2 t} B_s A_f^{-1} B_1 dt$.

Remark 1. Notice that for the slow process (11), the sampling period is $\Delta t_s = T$ based on t-scale, while for the fast process (12) the sampling period is $\Delta\tau = T = \frac{\Delta t_f}{\varepsilon}$, such that $\Delta t_f = \varepsilon T$. This means that the sampling period of (12) is εT based on t-scale.

For the DT subsystems (15)-(16), according to the Tikhonov Theorem [38], [45], there exists $\varepsilon^* > 0$ such that if $\varepsilon \in (0, \varepsilon^*)$, one has

$$\xi(k) = \tilde{\xi}(k) + \bar{\xi}(k) + o(\varepsilon) \tag{17}$$

$$x(k) = \bar{x}(k) + o(\varepsilon). \tag{18}$$

Moreover, from (5), the following form of the operational index holds

$$r(k) = C_2\bar{x}(k) + o(\varepsilon) = \bar{r}(k) + o(\varepsilon). \tag{19}$$

It is desired to make the operational indices follow prescribed values or trajectories. Define the desired trajectories of the operational indices as

$$r^*(k+1) = Fr^*(k) \tag{20}$$

where F is a square matrix, $k \in \mathbb{N}$ denotes the measurement time instant and $\mathbb{N}$ denotes the set of the positive integers.

The reference trajectory for the operational process is produced by the command generator model (20). Set $\eta(k) = [\bar{x}^T(k), (r^*(k))^T]^T$, by (15), (19) and (20), one has the augmented dynamics of the reduced slow subsystem (11) given by

$$\begin{cases} \eta(k+1) = \bar{M}_s\eta(k) + \bar{N}_s\bar{u}(k) \\ \bar{r}(k) = C_2\bar{x}(k) \end{cases} \tag{21}$$

where

$$\bar{M}_s = \begin{bmatrix} M_s & 0 \\ 0 & F \end{bmatrix},$$

and

$$\bar{N}_s = \begin{bmatrix} N_s \\ 0 \end{bmatrix}.$$

B. Formulation of the OOC Problem for the Singular Perturbed System

The objective of the OOC problem is to design the optimal controller u for the original system (1)-(2) to make the operational indices r follow the prescribed values r^*. To formulate the DT OOC problem, define a quadratic cost function J as

$$\begin{aligned} J = \min_{u(i)} \sum_{i=k}^{\infty} \gamma^{i-k} [&(r(i)-r^*(i))^T Q_1 (r(i)-r^*(i)) \\ &+ (y(i)-\bar{y}(i))^T Q_2 (y(i)-\bar{y}(i)) + w(i)^T R w(i)] \end{aligned} \tag{22}$$

where $0 < \gamma \leq 1$ is a discount factor, $\gamma = 1$ when $r^* = 0$, $w(k) = [\bar{u}^T(k), \tilde{u}^T(k)]^T$, $u(k) = \bar{u}(k) + \tilde{u}(k)$, Q_1 and Q_2 are positive definite matrices, $R = \text{diag}\{R_s, R_f\}$, R_s and R_f are positive definite matrices. $k \in \mathbb{N}$ denotes the fast sampling time instant. $\bar{y}(k)$ is the quasi-steady state of the output of system (1). The term $y(k) - \bar{y}(k)$ is added to reduce high frequency transients of the device outputs.

Remark 2. In this performance index, the tracking errors $e_1(k) = r(k) - r^*(k)$ and $e_2(k) = y(k) - \bar{y}(k)$ are both taken into account to guarantee an optimal tracking performance. Furthermore, J also considers the minimization of the energy consumption of the control input, represented by $w(k)$.

Considering the original global system (1)-(2), the OOC problem for industrial processes can now be defined as

Problem 1: For the original globle system (1)-(2), the performance index is defined as

$$\begin{aligned} J = \min_{u(i)} \sum_{i=k}^{\infty} \gamma^{i-k} [&(r(i)-r^*(i))^T Q_1 (r(i)-r^*(i)) \\ &+ (y(i)-\bar{y}(i))^T Q_2 (y(i)-\bar{y}(i)) + w(i)^T R w(i)] \\ \text{s.t.} \quad & \dot{y}(k+1) = \bar{A}_1 y(k) + \bar{B}_1 u(k) \\ & \dot{x}(k+1) = \bar{A}_2 x(k) + \bar{B}_2 y(k) \\ & r(k) = C_2 x(k) \\ & r^*(k+1) = Fr^*(k). \end{aligned} \tag{23}$$

where $\bar{A}_1 = e^{A_1 T}$, $\bar{B}_1 = \int_0^T e^{A_1 t} B_1 dt$, $\bar{A}_2 = e^{A_2 T}$, $\bar{B}_2 = -\int_0^T e^{A_2 t} B_2 dt$.

Remark 3. The OOC problem with two time scales defined in Problem 1 is viewed as an optimization problem subject to a high-order system. By singular perturbation theory, the steady solution in the slow phenomena that has dominant effect on the system and the correction term of boundary layer calculated in “stretched” time-scale can both be obtained.

Thus, for OOC Problem 1, two optimization subproblems are here presented for the two singularly perturbed subsystems (16) and (21).

Problem 2: For the DT fast subsystem (16), the performance index is defined as

$$\begin{aligned} J_1 = \min_{\tilde{u}} \sum_{i=k}^{\infty} \gamma^{i-k} [&\tilde{\xi}(i)^T Q_3 \tilde{\xi}(i) + \tilde{u}(i)^T R_f \tilde{u}(i)] \\ \text{s.t.} \quad & \tilde{\xi}(k+1) = M_f \tilde{\xi}(k) + N_f \tilde{u}(k) \end{aligned} \tag{24}$$

where $Q_3 = \varepsilon^2 Q_2 \in \mathbb{R}^{n_y \times n_y}$ is a positive definite matrix.

Remark 4. The OOC problem for the fast subsystem consists in determining the control input $\tilde{u}$ that stabilizes $\tilde{\xi}$ in (16) and minimizes J_1 in (24).

Problem 3: For the DT augmented slow subsystem (21), the performance index is defined as

$$\begin{aligned} J_2 = \min_{\bar{u}} \sum_{i=k}^{\infty} \gamma^{i-k} [&(\bar{r}(i)-r^*(i))^T Q_1 (\bar{r}(i)-r^*(i)) + \bar{u}(i)^T R_s \bar{u}(i)] \\ \text{s.t.} \quad & \eta(k+1) = \bar{M}_s \eta(k) + \bar{N}_s \bar{u}(k) \\ & \bar{r}(k) = C_2 \bar{x}(k). \end{aligned} \tag{25}$$

Remark 5. The OOC problem for the slow subsystem is to realize the tracking performance of $\bar{r}$ to the reference trajectory r^* with minimum control input $\bar{u}$. The reference trajectory r^* can be achieved by operational index r if $\bar{r}$ tends to r^* due to the relationship (19) [37].

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2019.2912018, IEEE Transactions on Industrial Informatics

The following theorem shows the equivalence of the solutions of Problem 2 and Problem 3, with respect to the solution of Problem 1.

Theorem 1. *For Problem 1, Problem 2 and Problem 3, the relationship $J = J_1 + J_2 + o(\varepsilon)$ holds.*

Proof: By, (24) and (25), one has

$$\begin{aligned} J =& \min_{u(i)} \sum_{i=k}^{\infty} \gamma^{i-k}[(r(i)-r^*(i))^T Q_1(r(i)-r^*(i)) \\ &+ (y(i)-\bar{y}(i))^T Q_2(y(i)-\bar{y}(i)) + w(i)^T Rw(i)] \\ =& \min_{u(i)} \sum_{i=k}^{\infty} \gamma^{i-k}[(\bar{r}(i)+o(\varepsilon)-r^*(i))^T Q_1(\bar{r}(i)+o(\varepsilon)-r^*(i)) \\ &+ (\tilde{y}(i)+o(\varepsilon))^T Q_2(\tilde{y}(i)+o(\varepsilon)) + w(i)^T Rw(i)] \\ =& \min_{u(i)} \sum_{i=k}^{\infty} \gamma^{i-k}[(\bar{r}(i)-r^*(i))^T Q_1(\bar{r}(i)-r^*(i)) + \tilde{\xi}(i)^T Q_3 \tilde{\xi}(i) \\ &+ \bar{u}(i)^T R_s \bar{u}(i) + \tilde{u}(i)^T R_f \tilde{u}(i)] + o(\varepsilon). \end{aligned} \tag{26}$$

Note that $o(\varepsilon)$ goes to zero when ε goes to zero, which is proven in [40], [45]. Then we have

$$\begin{aligned} J =& \min_{\bar{u}(i)} \sum_{i=k}^{\infty} \gamma^{i-k}[(\bar{r}(i)-r^*(i))^T Q_1(\bar{r}(i)-r^*(i)) + \bar{u}(i)^T R_s \bar{u}(i)] \\ &+ \min_{\tilde{u}(i)} \sum_{i=k}^{\infty} \gamma^{i-k}[\tilde{\xi}(i)^T Q_3 \tilde{\xi}(i) + \tilde{u}(i)^T R_f \tilde{u}(i)] + o(\varepsilon) \\ =& J_1 + J_2 + o(\varepsilon). \end{aligned} \tag{27}$$

The proof is completed. □

Remark 6. Theorem 1 implies that separately designing the optimal controllers for the two reduced-order subsystems (16) and (21) can approximately reach the optimal operation of the entire industrial process since ε is small enough.

Remark 7. The original OOC Problem 1 can be decomposed into two optimization subproblems subject to reduced subsystems by using singular perturbation approach. In contrast to lifting technique-based set-point design using two-layer hierarchical structure [5], [9]–[12], [14], [16] and one-layer controller design with the high-order augment systems [13], [17], the singular perturbation based decomposition of fast and slow processes implies less computational complexity and has no integer-multiple limitation of difference between two time scales, making it easier and more efficient to solve OOC Problem 1.

IV. RL-based Solution to the OOC Problem

In this section two approximate optimal controllers for Problem 2 and 3 are determined by using RL techniques to achieve optimality of the performance functions given in (24) and (25). This method requires only data measured from the industrial operational processes, without the need to know the dynamic models of either the unit device process or the operational process. This data-driven method to deal with OOC of industrial processes with two time-scales is developed for the first time in this paper. To this end, two RL algorithms for two optimization subproblems are developed to separately design fast control $\tilde{u}^*$ and slow control $\bar{u}^*$ based on Q-learning [26].

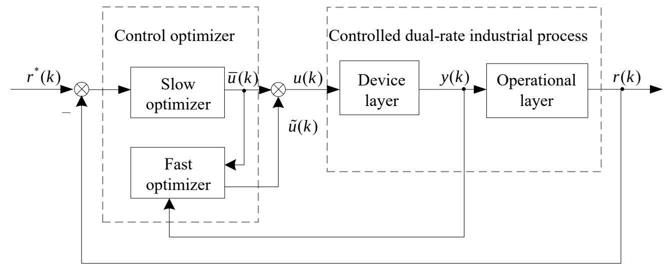


Fig. 2. The control block diagram of DT singularly perturbed system.

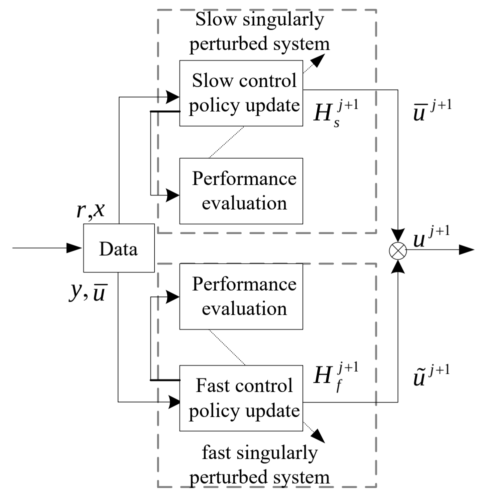


Fig. 3. The relationship inside the control optimizer.

The structure of the two time-scales controller of Theorem 1 is given in Fig. 2, and the detailed structure of the industrial singular perturbation RL optimizer is shown in Fig. 3. The overall convergence of the algorithm covering the slow RL Algorithm and the fast RL Algorithm is shown in section C.

A. Model-Free Optimal Controller Design for the Slow Operational Process

A RL technique based approach to solve the slow Problem 3 is developed in this subsection. The result is Algorithm 1.

Referring to the cost function (25) and the system (21), a value function is defined as

$$\begin{aligned} V_s(\eta(k)) =& \sum_{i=k}^{\infty} \gamma^{i-k}[(\bar{r}(i)-r^*(i))^T Q_1(\bar{r}(i)-r^*(i)) \\ &+ \bar{u}(i)^T R_s \bar{u}(i)] \\ =& \sum_{i=k}^{\infty} \gamma^{i-k}[\eta(i)^T Q_s \eta(i) + \bar{u}(i)^T R_s \bar{u}(i)] \\ =& \rho_s(k) + \gamma \sum_{i=k+1}^{\infty} \rho_s(i) \end{aligned} \tag{28}$$

where $\eta(k) = [\bar{x}^T(k), (r^*(k))^T]^T$, $Q_s = [C_2, -I]^T Q_1 [C_2, -I]$ and $\rho_s(k) = \eta(k)^T Q_s \eta(k) + \bar{u}(k)^T R_s \bar{u}(k)$.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2019.2912018, IEEE Transactions on Industrial Informatics

The Bellman equation of the slow operational process can now be obtained as

$$V_s(\eta(k)) = \rho_s(k) + \gamma V_s(\eta(k+1)). \tag{29}$$

Define the corresponding Hamiltonian function as

$$H(\eta, \bar{u}, V_s) = \rho_s(k) + \gamma V_s(\eta(k+1)) - V_s(\eta(k)). \tag{30}$$

It is well known that the optimal value function satisfies the Hamiltonian-Jacobi-Bellman (HJB) equation [26], [46]–[49], that is,

$$V_s{}^*(\eta(k)) = \min_{\bar{u}_k}(\rho_s(k) + \gamma V_s{}^*(\eta(k+1))). \tag{31}$$

According to the HJB equation (31), the optimal Q-function-based HJB equation is defined as

$$Q^*(\eta(k), \bar{u}^*(k)) = \min_{\bar{u}}[\rho_s(k) + \gamma Q^*(\eta(k+1), \bar{u}^*(k+1)] \tag{32}$$

where

$$V_s{}^*(\eta(k)) = \min_{\bar{u}} Q^*(\eta(k), \bar{u}) = Q^*(\eta(k), \bar{u}^*). \tag{33}$$

The value function is known to have quadratic form in terms of the states for linear systems, thus we have

$$V_s(\eta(k)) = \eta^T(k) P_s \eta(k). \tag{34}$$

Following the idea in [26], the Q-function-based HJB equation can be written as

$$\begin{aligned}
&Q^*(\eta(k), \bar{u}^*(k)) = \eta^T(k) Q_s \eta(k) + (\bar{u}^*(k))^T R_s \bar{u}^*(k) \\
&+ \gamma \eta^T(k+1) P_s \eta(k+1) \\
&= \eta^T(k) Q_s \eta(k) + (\bar{u}^*(k))^T R_s \bar{u}^*(k) \\
&+ \gamma(\bar{M}_s \eta(k) + \bar{N}_s \bar{u}^*(k))^T P_s (\bar{M}_s \eta(k) + \bar{N}_s \bar{u}^*(k)) \\
&= \begin{bmatrix} \eta(k) \\ \bar{u}^*(k) \end{bmatrix}^T \begin{bmatrix} Q_s + \gamma \bar{M}_s^T P_s \bar{M}_s & \gamma \bar{M}_s^T P_s \bar{N}_s \\ \gamma \bar{N}_s^T P_s \bar{M}_s & R_s + \gamma \bar{N}_s^T P_s \bar{N}_s \end{bmatrix} \begin{bmatrix} \eta(k) \\ \bar{u}^*(k) \end{bmatrix} \\
&= \begin{bmatrix} \eta(k) \\ \bar{u}^*(k) \end{bmatrix}^T H_s \begin{bmatrix} \eta(k) \\ \bar{u}^*(k) \end{bmatrix}
\end{aligned} \tag{35}$$

where

$$H_s = \begin{bmatrix} H_{s\eta\eta} & H_{s\eta\bar{u}} \\ H_{s\bar{u}\eta} & H_{s\bar{u}\bar{u}} \end{bmatrix} \tag{36}$$

and $H_{s\eta\eta} = Q_s + \gamma \bar{M}_s^T P_s \bar{M}_s$, $H_{s\eta\bar{u}} = \gamma \bar{M}_s^T P_s \bar{N}_s$, $H_{s\bar{u}\eta} = \gamma \bar{N}_s^T P_s \bar{M}_s$, $H_{s\bar{u}\bar{u}} = R_s + \gamma \bar{N}_s^T P_s \bar{N}_s$.

According to the necessary condition of optimality, setting $\frac{\partial Q(\eta_k, \bar{u}_k)}{\partial \bar{u}_k} = 0$ yields the optimal control law for the slow operational process as follows

$$\begin{aligned}
\bar{u}^*(k) &= - H_{s\bar{u}\bar{u}}^{-1} H_{s\bar{u}\eta} \eta(k) \\
&= - K_s^* \eta(k)
\end{aligned} \tag{37}$$

and using the definitions in (36), yields

$$\bar{u}^*(k) = -(R_s + \gamma \bar{N}_s^T P_s \bar{N}_s)^{-1} \gamma \bar{N}_s^T P_s \bar{M}_s \eta(k). \tag{38}$$

Notice that $\bar{x}(k)$ cannot be measured, since the Tiknohov Theorem (18) holds, to achieve data-driven control, set

$$\bar{Z}(k) = \begin{bmatrix} x(k) \\ r^*(k) \\ \bar{u}(k) \end{bmatrix}$$

and replace $\eta(k)$ by $[x^T(k), (r^*(k))^T]^T$ to approximately calculate H_s. According to (35), the Q-function based on the Bellman equation (32) can be written as

$$\bar{Z}^T(k) H_s \bar{Z}(k) = \rho_s(k) + \gamma \bar{Z}^T(k+1) H_s \bar{Z}(k+1). \tag{39}$$

Algorithm 1 is given bellow for learning the optimal control policy $\bar{u}^*(k)$ by value iteration approach.

Algorithm 1 Q-learning for the slow subsystem

Step 1: **Initialization.** Select an admissible policy $\bar{u}^o(k) = -K_s^{j_s} \eta(k) + e_s(k)$ as the input where $e_s(k)$ is the probing noise, and choose a discount factor $0 < \gamma \leq 1$. Let $j_s = 0$, where j_s denotes iteration step;

Step 2: **Policy Evaluation.** Solve Q-function matrix $H_s{}^{j_s+1}$ satisfying

$$\begin{aligned}
\bar{Z}^T(k) H_s^{j_s+1} \bar{Z}(k) =& \rho_s^{j_s}(k) \\
&+ \gamma \bar{Z}^T(k+1) H_s^{j_s} \bar{Z}(k+1);
\end{aligned} \tag{40}$$

Step 3: **Policy Improvement.**

$$\begin{aligned}
\bar{u}^{j_s+1}(k) =& - (H_{s\bar{u}\bar{u}}^{j_s+1})^{-1} H_{s\bar{u}\eta}^{j_s+1} \\
&\times [x^T(k), (r^*(k))^T]^T;
\end{aligned} \tag{41}$$

Step 4: Set $j_s = j_s + 1$ and go to step 2. Stop when $\|H_s{}^{j_s} - H_s{}^{j_s+1}\| \leq \sigma_1$ for an arbitrary small positive constant σ_1.

To calculate the Q-function matrix $H_s{}^{j_s+1}$, let (40) be written as follows

$$(h_s^{j_s+1})^T \bar{z}(k) = \rho_s(k) + \gamma (h_s^{j_s})^T \bar{z}(k+1) \tag{42}$$

where $h_s^{j_s} = \mathrm{vec}(H_s^{j_s})$, $\bar{z}(k) = \bar{Z}(k) \otimes \bar{Z}(k)$. $\otimes$ denotes the Kronecker product and $\mathrm{vec}(H_s^{j_s})$ is the vector formed by stacking the columns of matrix $H_s^{j_s}$.

Now, online least-squares (LS) method can be employed to calculate $h_s^{j_s}$ satisfying (42) by using only data generated by the trajectory of the slow operational process under the control action $\bar{u}^{j_s}(k)$.

B. Model-Free Optimal Controller Design for the Fast Device Control System

The same procedure as in Section IV(A) is now performed to design a RL optimal controller for the fast problem 2 of the unit devices, the result is Algorithm 2.

A new value function is defined by referring to the cost function (24) as

$$\begin{aligned}
V_f(\tilde{\xi}(k)) =& \sum_{i=k}^{\infty} \gamma^{i-k} [\tilde{\xi}(i)^T Q_3 \tilde{\xi}(i) + \tilde{u}(i)^T R_f \tilde{u}(i)] \\
=& \sum_{i=k}^{\infty} \rho_f(i) \\
=& \rho_f(k) + \gamma \sum_{i=k+1}^{\infty} \rho_f(i)
\end{aligned} \tag{43}$$

where $\rho_f(k) = \tilde{\xi}(k)^T Q_3 \tilde{\xi}(k) + \tilde{u}(k)^T R_f \tilde{u}(k)$. Again, we obtain the Bellman equation of the fast device control process as

$$V_f(\tilde{\xi}(k)) = \rho_f(k) + \gamma V_f(\tilde{\xi}(k+1)) \tag{44}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2019.2912018, IEEE Transactions on Industrial Informatics

and Hamiltonian function

$$H(\tilde{\xi},\tilde{u},V_f)=\rho_f(k)+\gamma V_f(\tilde{\xi}(k+1))-V_f(\tilde{\xi}(k)). \quad (45)$$

The optimal Q-function for the fast device control process can be defined as

$$\begin{aligned}V_f^*(\tilde{\xi}(k))&=\min_{\tilde{u}}Q^*(\tilde{\xi}(k),\tilde{u})\\&=Q^*(\tilde{\xi}(k),\tilde{u}^*)\end{aligned} \quad (46)$$

and the Q-function based Bellman equation of the fast device control process is

$$Q(\tilde{\xi}(k),\tilde{u})=\rho_f(k)+\gamma Q(\tilde{\xi}(k+1),\tilde{u}). \quad (47)$$

Since $V_f(\tilde{\xi}(k))=\tilde{\xi}_k^T P_f\tilde{\xi}_k$ with $P_f>0$ and the relationship between $V_f^*(\tilde{\xi}(k))$ and $Q^*(\tilde{\xi}(k),\tilde{u}^*)$ is given by (46), then (47) becomes

$$\begin{aligned}Q(\tilde{\xi}(k),\tilde{u}^*)=&\rho_f(k)+\gamma V_f^*(\tilde{\xi}(k+1))\\=&\tilde{\xi}^T(k)Q_3\tilde{\xi}(k)+\tilde{u}^{*T}R_f\tilde{u}^*\\&+\gamma\tilde{\xi}^T(k+1)P_f\tilde{\xi}(k+1)\\=&\tilde{\xi}^T(k)Q_3\tilde{\xi}(k)+\tilde{u}^{*T}R_f\tilde{u}^*\\&+\gamma(M_f\tilde{\xi}(k)+N_f\tilde{u}^*)^TP_f(M_f\tilde{\xi}(k)+N_f\tilde{u}^*)\\=&\begin{bmatrix}\tilde{\xi}(k)\\\tilde{u}^*\end{bmatrix}^T H_f\begin{bmatrix}\tilde{\xi}(k)\\\tilde{u}^*\end{bmatrix}\end{aligned} \quad (48)$$

where

$$\begin{aligned}H_f&=\begin{bmatrix}H_{f\tilde{\xi}\tilde{\xi}} & H_{f\tilde{\xi}\tilde{u}}\\H_{f\tilde{u}\tilde{\xi}} & H_{f\tilde{u}\tilde{u}}\end{bmatrix}\\&=\begin{bmatrix}Q_3+\gamma M_f{}^TP_fM_f & \gamma M_f{}^TP_fN_f\\\gamma N_f{}^TP_fM_f & R_f+\gamma N_f{}^TP_fN_f\end{bmatrix}.\end{aligned}$$

The Q-function for the fast device control process is therefore quadratic with the form

$$Q(\tilde{\xi}(k),\tilde{u})=\begin{bmatrix}\tilde{\xi}(k)\\\tilde{u}\end{bmatrix}^T H_f\begin{bmatrix}\tilde{\xi}(k)\\\tilde{u}\end{bmatrix}. \quad (49)$$

To achieve the optimal value $V_f^*(\tilde{\xi}(k))$, setting $\frac{\partial Q(\tilde{\xi}_k,\tilde{u}_k)}{\partial\tilde{u}_k}=0$ yields

$$\tilde{u}^*(k)=-H_{f\tilde{u}\tilde{u}}^{-1}H_{f\tilde{u}\tilde{\xi}}\tilde{\xi}(k) \quad (50)$$

which can also be expressed as

$$\tilde{u}^*(k)=-(R_f+\gamma N_f{}^TP_fN_f)^{-1}\gamma N_f{}^TP_fM_f\tilde{\xi}(k). \quad (51)$$

Notice from (51) that $\tilde{\xi}(k)$ needs to be available to compute $\tilde{u}^*(k)$; however, $\tilde{\xi}(k)$ cannot be measured directly. It is an objective of this paper to find the optimal controllers using only measured data. According to the Tikhonov Theorem [45], $\xi(k)=\bar{\xi}(k)+\tilde{\xi}(k)+o(\varepsilon)$ holds; since the parameter ε is small enough, $\tilde{\xi}(k)$ can be approximated as $\tilde{\xi}(k)\approx\xi(k)-\bar{\xi}(k)$. Since $y(k)=\varepsilon\xi(k)$ and $\bar{\xi}(k)=-A_f^{-1}B_1\bar{u}(k)$, the Q-function (49) can be written by replacing $\tilde{\xi}(k)$ by $\xi(k)-\bar{\xi}(k)=\frac{1}{\varepsilon}y(k)+A_f^{-1}B_1\bar{u}(k)$ as

$$\begin{aligned}Q(\delta(k),\tilde{u}(k))&=\begin{bmatrix}\tilde{\xi}(k)\\\tilde{u}(k)\end{bmatrix}^T\begin{bmatrix}H_{f\tilde{\xi}\tilde{\xi}} & H_{f\tilde{\xi}\tilde{u}}\\H_{f\tilde{u}\tilde{\xi}} & H_{f\tilde{u}\tilde{u}}\end{bmatrix}\begin{bmatrix}\tilde{\xi}(k)\\\tilde{u}(k)\end{bmatrix}\\&=\begin{bmatrix}\xi(k)-\bar{\xi}(k)\\\tilde{u}(k)\end{bmatrix}^T\begin{bmatrix}H_{f\tilde{\xi}\tilde{\xi}} & H_{f\tilde{\xi}\tilde{u}}\\H_{f\tilde{u}\tilde{\xi}} & H_{f\tilde{u}\tilde{u}}\end{bmatrix}\begin{bmatrix}\xi(k)-\bar{\xi}(k)\\\tilde{u}(k)\end{bmatrix}\\&=\begin{bmatrix}y(k)\\\bar{u}(k)\\\tilde{u}(k)\end{bmatrix}^T\begin{bmatrix}\frac{1}{\varepsilon^2}H_{f\tilde{\xi}\tilde{\xi}} & \frac{1}{\varepsilon}H_{f\tilde{\xi}\tilde{\xi}}A_f^{-1}B_1 & \frac{1}{\varepsilon}H_{f\tilde{\xi}\tilde{u}}\\ * & \Pi_f & \Lambda_f\\ * & * & H_{f\tilde{u}\tilde{u}}\end{bmatrix}\\&\times\begin{bmatrix}y(k)\\\bar{u}(k)\\\tilde{u}(k)\end{bmatrix}\\&=\begin{bmatrix}\delta(k)\\\tilde{u}(k)\end{bmatrix}^T\begin{bmatrix}Hf_{\delta\delta} & Hf_{\delta\tilde{u}}\\Hf_{\tilde{u}\delta} & Hf_{\tilde{u}\tilde{u}}\end{bmatrix}\begin{bmatrix}\delta(k)\\\tilde{u}(k)\end{bmatrix}\end{aligned} \quad (52)$$

where

$$\delta(k)=\begin{bmatrix}y(k)\\\bar{u}(k)\end{bmatrix},$$

$$\Pi_f=(A_f^{-1}B_1)^TH_{f\tilde{\xi}\tilde{\xi}}A_f^{-1}B_1,$$

$$\Lambda_f=(A_f^{-1}B_1)^TH_{f\tilde{\xi}\tilde{u}},$$

$$Hf_{\tilde{u}\tilde{u}}=H_{f\tilde{u}\tilde{u}},$$

$$Hf_{\delta\tilde{u}}=\begin{bmatrix}\frac{1}{\varepsilon}H_{f\tilde{\xi}\tilde{u}}\\\Lambda_f\end{bmatrix}^T,$$

$$Hf_{\delta\delta}=\begin{bmatrix}\frac{1}{\varepsilon^2}H_{f\tilde{\xi}\tilde{\xi}} & \frac{1}{\varepsilon}H_{f\tilde{\xi}\tilde{\xi}}A_f^{-1}B_1\\ * & \Pi_f\end{bmatrix}.$$

Now, the Q-function (47) based on the Bellman equation of the fast device control process and the optimal controller (50) are respectively written as

$$\tilde{Z}^T(k)Hf\tilde{Z}(k)=\rho_f(k)+\gamma\tilde{Z}^T(k+1)Hf\tilde{Z}(k+1) \quad (53)$$

and

$$\begin{aligned}\tilde{u}^*(k)&=-H_{f\tilde{u}\tilde{u}}^{-1}H_{f\tilde{u}\tilde{y}}\tilde{\xi}(k)\\&=-H_{f\tilde{u}\tilde{u}}^{-1}H_{f\tilde{u}\tilde{y}}(\xi(k)-\bar{\xi}(k))\\&=-H_{f\tilde{u}\tilde{u}}^{-1}H_{f\tilde{u}\tilde{y}}\frac{1}{\varepsilon}y(k)+H_{f\tilde{u}\tilde{u}}^{-1}H_{f\tilde{u}\tilde{y}}(A_f^{-1}B_1)\bar{u}(k)\\&=-Hf_{\tilde{u}\tilde{u}}^{-1}Hf_{\tilde{u}\delta}\delta(k)\\&=-K_f^*\delta(k)\end{aligned} \quad (54)$$

where

$$\tilde{Z}(k)=\begin{bmatrix}\delta(k)\\\tilde{u}(k)\end{bmatrix},$$

$$Hf=\begin{bmatrix}Hf_{\delta\delta} & Hf_{\delta\tilde{u}}\\Hf_{\tilde{u}\delta} & Hf_{\tilde{u}\tilde{u}}\end{bmatrix}.$$

To solve Hf^{j_f+1} in Algorithm 2, (55) is rewritten as

$$\begin{aligned}(h_f^{j_f+1})^T\tilde{z}(k)=&\delta^T(k)Q_f\delta(k)+(\tilde{u}^{j_f}(k))^TR_f\tilde{u}^{j_f}(k)\\&+\gamma(h_f^{j_f})^T\tilde{z}(k+1)\end{aligned} \quad (57)$$

where $h_f^{j_f+1}=\mathrm{vec}(Hf^{j_f+1})$, $\tilde{z}(k)=\tilde{Z}(k)\otimes\tilde{Z}(k)$.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2019.2912018, IEEE Transactions on Industrial Informatics

Algorithm 2 Q-learning for the fast subsystem

Step 1: **Initialization.** Select an admissible policy $\tilde{u}^o(k) = -K_f^{j_f}\delta(k) + e_f(k)$ as the input where $e_f(k)$ is the probing noise, and choose a discount factor $0 < \gamma \leq 1$. Let $j_f = 0$, where j_f denotes iteration step;

Step 2: **Policy Evaluation.** Solve Q-function matrix Hf^{j_f+1} satisfying

$$\begin{aligned}\tilde{Z}^T(k)Hf^{j_f+1}\tilde{Z}(k) =& \gamma\tilde{Z}^T(k+1)Hf^{j_f}\tilde{Z}(k+1)\\ &+\rho_f^{j_f}(k);\end{aligned} \tag{55}$$

Step 3: **Policy Improvement.**

$$\tilde{u}^{j_f+1}(k) = -(Hf_{\tilde{u}\tilde{u}}^{j_f+1})^{-1}Hf_{\tilde{u}\delta}^{j_f+1}\delta(k); \tag{56}$$

Step 4: Set $j_f = j_f + 1$ and go to step 2. Stop when $\|Hf^{j_f} - Hf^{j_f+1}\| \leq \sigma_2$ for an arbitrary small positive constant σ_2.

LS method is employed to approximate $h_f^{j_f+1}$ satisfying (57) resulting in getting the optimal control policy $\tilde{u}^*(k)$.

Notice that there exists a coupling relationship when implementing Algorithm 1 and Algorithm 2 since $\bar{u}^j(k)$ needs to be used in (55) due to $\delta(k) = [y(k), \bar{u}(k)]^T$, and that data used in Algorithm 1 and Algorithm 2 are generated online under the composite controller $u^j(k) = \tilde{u}^{j_f}(k) + \bar{u}^{j_s}(k)$. The following algorithm shows how to learn the approximate optimal controller for OOC problem by combining Algorithm 1 and Algorithm 2.

Algorithm 3 Q-learning for the system with two time-scales

Initialization:

Step 1: Select an admissible policy $\tilde{u}^o(k) = -K_f^{j_f}\delta(k) + e_f(k)$ and $\bar{u}^o = -K_s^{j_s}\eta(k) + e_s(k)$ to get initial input $u^0(k) = \tilde{u}^0(k) + \bar{u}^0(k)$ where $e_f(k)$ and $e_s(k)$ are the probing noises, and choose a discount factor $0 < \gamma \leq 1$. Let $j_f = 0$ and $j_s = 0$ where j_s and j_f denote iteration steps;

Slow Process Learning (Algorithm 1):

Step 2: **Policy Evaluation.** Solve Q-function matrix $H_s{}^{j_s+1}$ satisfying (40);

Step 3: **Policy Improvement.** Get $\bar{u}^{j_s+1}(k)$ by implementing (41);

Fast Process Learning (Algorithm 2):

Step 4: **Policy Evaluation.** Solve Q-function matrix Hf^{j_f+1} satisfying (55);

Step 5: **Policy Improvement.** Get $\tilde{u}^{j_f+1}(k)$ by implementing (56);

Step 6: Apply $u^{j+1}(k) = \bar{u}^{j_s+1}(k) + \tilde{u}^{j_f+1}(k)$ to the original system (1);

Step 7: Set $j_s = j_s + 1$, $j_f = j_f + 1$ and go to step 2. Stop when $\|H_s^{j_s} - H_s{}^{j_s+1}\| \leq \sigma_1$ and $\|Hf^{j_f} - Hf^{j_f+1}\| \leq \sigma_2$ for arbitrary small positive constants σ_1 and σ_2.

Finally, a composite controller for the whole system is obtained as

$$u^*(k) = \bar{u}^*(k) + \tilde{u}^*(k). \tag{58}$$

Remark 8. If proper probing noises $e_f(k)$ and $e_s(k)$ are added into the control inputs $\tilde{u}^{j_f}(k)$ and $\bar{u}^{j_s}(k)$, the persistent excitation condition can be guaranteed to accurately calculate $h_f{}^{j_f+1}$ and $h_s^{j_s+1}$ [12], [26], [50].

Remark 9. As proven in Theorem 1 and the analysis about singular perturbation in [37], [42], [43], the composite controller is the approximately optimal controller of OOC Problem 1. Thus, by Algorithm 3, an approximately optimal controller can be found for OOC Problem 1 using only data generated by the systems without knowing any information about the device control process and the operation process, which is different from the traditional model-based OOC methods [9]–[12].

Remark 10. Notice that the reported OOC methods for industrial processes in [1], [2] are based on operator's experience when selecting properly the increment of correction of set-points. The optimality of industrial operation is then hard to guarantee. Unlike [18] where the neural-network based set-points design requires the known optimal performance indices as a priori, the proposed Algorithm 3 needs only data from the system output and the real operational indices to be measured.

C. Stability Analysis for Combination of Singular Perturbation and RL

In this subsection, the convergence of Algorithm 3, as well as the stability of the original system under the controller given by Algorithm 3 are proven.

Lemma 1. Define

$$P_s = \begin{bmatrix} P_{s11} & P_{s12} \\ P_{s21} & P_{s22} \end{bmatrix}$$

in the value function for slow subsystem (34), and the definitions of the four subparts of P_s can be found in [47]. The control policy $\bar{u}^{j_s}$ derived by Algorithm 1 converges to the optimal control policy $\bar{u}^*$ as $j_s \to \infty$, i.e. $\lim_{j_s\to\infty}\bar{u}^{j_s} = \bar{u}^*$. Moreover, the slow operational process (21) is stable and can reach the optimum of performance under $\bar{u}^*$ if $F\gamma^{0.5}$ is stable and

$$0 < (P_{s11} - Q_1)(P_{s11} + G_s)^{-1} < \gamma^2 I \tag{59}$$

where $G_s = P_{s11}N_s(R_s + N_s{}^TP_{s11}N_s)^{-1}R_s(R_s + N_s{}^TP_{s11}N_s)^{-1}N_s{}^TP_{s11}$.

Proof: The detailed proof can be found in [47] so that is omitted here. □

Lemma 2. The control policy u^j derived by Algorithm 3 converges to the optimal control policy u^* as $j \to \infty$, i.e. $\lim_{j\to\infty}u^j = u^*$. Moreover, the original globle system (1)-(2) is asymptotically stable and can reach the optimum of performance in (22) under u^*.

Proof: The LS problem (42) and (57) can be solved if the persistent excitation condition is satisfied. Thus $\lim_{j_s\to\infty}\bar{u}^{j_s} = \bar{u}^*$ and $\lim_{j_f\to\infty}\tilde{u}^{j_f} = \tilde{u}^*$ holds, which has been proved in [26]. Moreover, it is true that $\bar{u}^*$ and $\tilde{u}^*$ can guarantee the stability and optimality of the decomposed slow subsystem (15) and fast subsystem (16) with the goals (25) and (24) as proven in

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2019.2912018, IEEE Transactions on Industrial Informatics

Fig. 4. Thickening industrial process.

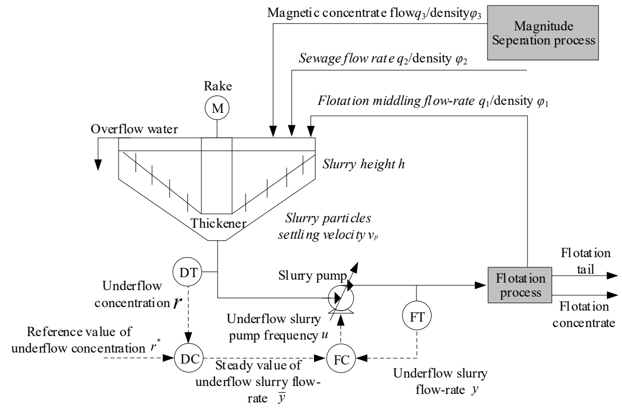


Fig. 5. Schematic illustration of MSTP.

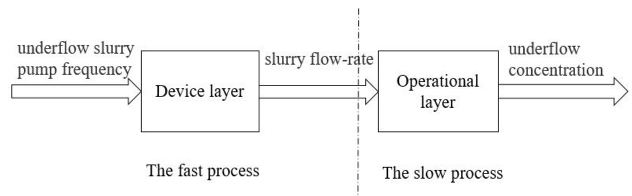


Fig. 6. The two-layer structure of MSTP.

[26], [47]. Then, according to the Tikhonov Theorem (18) and singular perturbation decomposition principle $u = \bar{u} + \tilde{u}$, the control input u is convergent and goes to the optimal $u^* = \bar{u}^* + \tilde{u}^*$ which stabilize the globle system (1)-(2). This completes the proof. □

Lemma 3. The tracking errors $e_1(k) = r(k) - r^*(k)$ and $e_2(k) = y(k) - \bar{y}(k)$ are stable.

Proof: With convergence of performance indices (28) and (24), which is proven in [26], [51], the term $\eta_k{}^T Q_s \eta_k$ in (28) and the term $\tilde{\xi}_k^T Q_3 \tilde{\xi}_k$ in (24) are also convergent. According to the Tikhonov Theorem (18) and (28), the tracking error of operational process

$$e_1(k) = r(k) - r^*(k) = \bar{r}(k) - r^*(k) + o(\varepsilon) \tag{60}$$

is convergent. According to (18) and the transformation $y = \varepsilon\xi$, the tracking error of device process

$$e_2(k) = y(k) - \bar{y}(k) = \varepsilon\tilde{\xi}(k) + o(\varepsilon) \tag{61}$$

is also convergent. This completes the proof. □

Theorem 2. *The complete operational process in Algorithm 3 is stable under the composite controller* (58) *if* $\varepsilon \in (0, \varepsilon^*)$.

Proof: By Lemma 1 and Lemma 2, the two singularly perturbed system (16) and (21) can be stabilized under respectively the control policies $\tilde{u}^*$ and $\bar{u}^*$. Reference [38] has pointed out that if the singularly perturbed system (16) and (21) is stable and $\varepsilon \in (0, \varepsilon^*)$, then the original operational process (1)-(2) is also stable. The proof is completed. □

V. SMULATION RESULTS

In this section, the mixed separation thickening process (MSTP) of hematite beneficiation with two time-scales [52], [53] is taken as an experimental example to verify the effectiveness of the proposed method, combining the singular perturbation technique and RL. We verify that the control policy calculated by Q-learning based on the singularly perturbed system is also the optimal control of original system and that it performs satisfactorily.

Fig. 4 shows a production site of the industrial project where the proposed control scheme can be applied. The thickening process plays an important role in mineral separation industrial processes. It is the process after the grinding procedure [34] and before the flotation process, and is used mainly to increase the concentration of slurry from grinding process and provide the qualified slurry concentration and the pulp flow for the flotation process. This allows for the flotation process to get satisfactory concentrate grade and tail grade, and guarantees the production safety and efficiency [19], [54], [55]. A schematic illustration of MSTP is shown in Fig. 5. Underflow concentration is the control output as well as an operational index of the MSTP.

A. Control Goal of MSTP

MSTP has a two-layer structure. The underflow slurry pump frequency is the input of the device layer and the slurry flow-rate is its output, while the underflow concentration is the output of the operational layer as shown in Fig. 6. This is a system with multiple time scales with the device layer as the fast process and the operational layer as the slow process. In this example, the control goal is described by

$$31 \le r(k) \le 35, \tag{62}$$

$$10 \le y(k) \le 60. \tag{63}$$

These quantities are selected with respect to the industrial physical limitations. Considering (62) and (63), set the reference value as $r^* = 33$.

B. The Dynamic Formulation of MSTP

The dynamics of the device layer is given by

$$\frac{\mathrm{d}y(t)}{\mathrm{d}t} = -\frac{y(t)}{\sigma} + \frac{1}{\sigma}\sqrt{\frac{k_0 u^2(t) - \frac{\Delta p(h,y)\lambda y(t)}{g} + C}{K}} \tag{64}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2019.2912018, IEEE Transactions on Industrial Informatics

TABLE I
PARAMETERS IN MSTP AND THEIR PHYSICAL MEANINGS

parameter	description
$q_1(t)\varphi_1(t)$	Medium Mineral Slurry Concentration and Flow
$h(\cdot)$	Slurry height
$q_2(t)\varphi_2(t)$	Sewage concentration and flow
$v_p(\cdot)$	Slurry particle sedimentation velocity
$q_3(t)\varphi_3(t)$	Magnetic slurry concentration and flow
σ	Time constant
A	Thickener cross-sectional area
K, k_0	Constants related to pulp pump structure
$\upsilon(t)$	Total solute of thickener
k_1, k_2, k_3, k_4	Constants related to thickener structure
$\Delta p(\cdot)$	Differential pressure at both ends of slurry pump
$\lambda(t)$	Underflow slurry particle concentration coefficient
μ, p, ρ_s, ρ_l	Constants related to the nature of the pulp
C	Loss of resistance
g	Gravity acceleration

where the underflow slurry pump frequency $u(t)$ is the input of the device layer, and the underflow slurry flow-rate $y(t)$ is the output of the device layer.

The dynamic of the operational layer is

$$\frac{\mathrm{d}r(t)}{\mathrm{d}t} = \frac{1}{k_2 h(\upsilon, y, r, A)}\Big[\frac{-r^2(t)y(t)}{r(t)+k_3\upsilon(t)} + k_1 v_p(\upsilon, y, r)\upsilon(t) + \frac{k_1(k_i - k_3)v_p(\upsilon, y, r)\upsilon(t)}{r(t)+k_3\upsilon(t)}\Big] \tag{65}$$

where the underflow concentration $r(t)$ is the output of the operational layer, $k_1 = Ak_i$, $k_2 = Ap_i$, $k_3 = k_i - \mu(\rho_s - \rho_l)/Ap$ and $\upsilon(t) = q_1(t)\varphi_1(t) + q_2(t)\varphi_2(t) + q_3(t)\varphi_3(t)$. The parameter definitions involved are given in Table I.

The balanced point values of the parameters are taken as $k_0 = 47.97$, $k_i = 0.001$, $k_1 = 1.9625$, $K_2 = 98.13$, $k_3 = 0.0049$, $v_p = 1.825$, $h = 6$m, $\sigma = 1.47$, $C = 100000$, $\Delta p/g\lambda = 151.0748$, $\upsilon_1 = 340$, $K = 1.12$. Linearizing (64) and (65) on its balance state yields

$$\begin{cases} \dot{y}(t) = -0.68y(t) + 2.6u(t), \\ \dot{r}(t) = -0.057r(t) + 0.055y(t). \end{cases} \tag{66}$$

Select the parameter value of the optimal controller as $\gamma = 0.9$, $Q_1 = 10$, $Q_3 = 200$, $R_s = 1$, $R_f = 1$. Set the initial value of the approximate matrices as

$$Hf = \begin{bmatrix} 2583 & -9877 & 260 \\ -9877 & 37766 & -994 \\ 260 & -994 & 59 \end{bmatrix}$$

and

$$H_s = \begin{bmatrix} 1056.8 & -301.6 & 1004.5 \\ -301.6 & 425.5 & -1597.2 \\ 1004.5 & -1597.2 & 2157.2 \end{bmatrix}.$$

Algorithm 3 is employed to determine the optimal controller for the linearized thickening system (66). The simulation results are shown from Fig. 7 to Fig. 10.

C. Simulation Results

Fig. 7 shows that the operational index r of thickening process tracks its setpoint r^* satisfactorily. From Fig. 8, the transient value of the output of the fast device layer converges

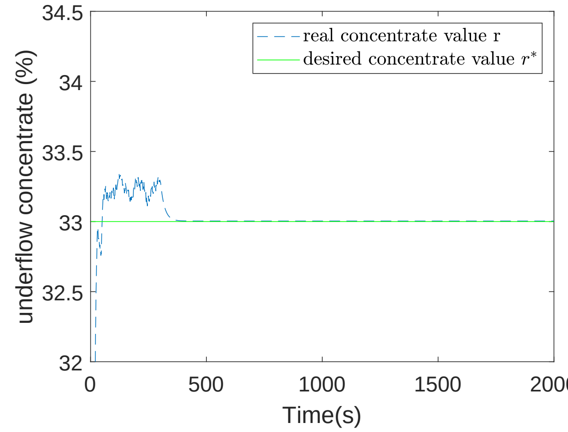


Fig. 7. The tracking performance of the underflow concentration to its setpoint.

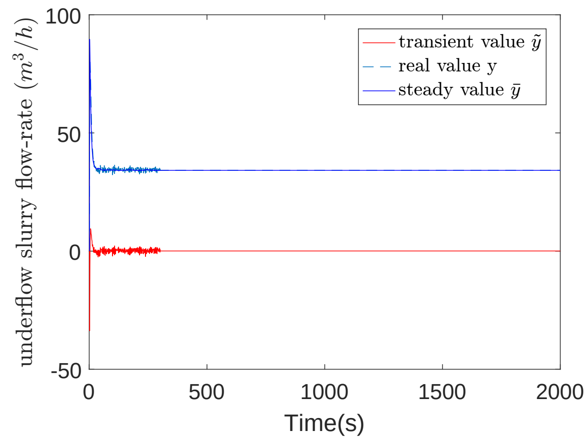


Fig. 8. Convergence of the underflow slurry flow-rate to its quasi-steady-state.

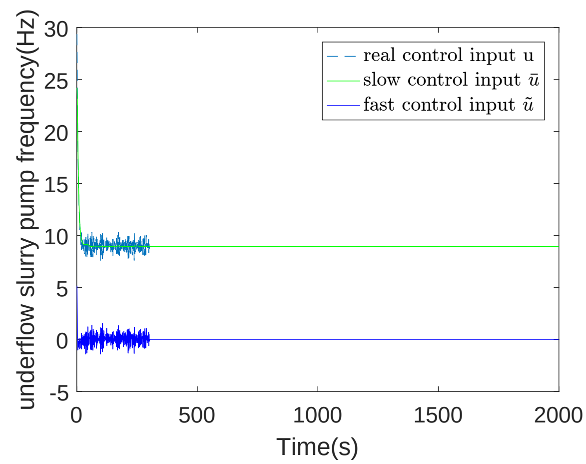


Fig. 9. The real control input and the singularly perturbed fast and slow inputs of MSTP.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2019.2912018, IEEE Transactions on Industrial Informatics

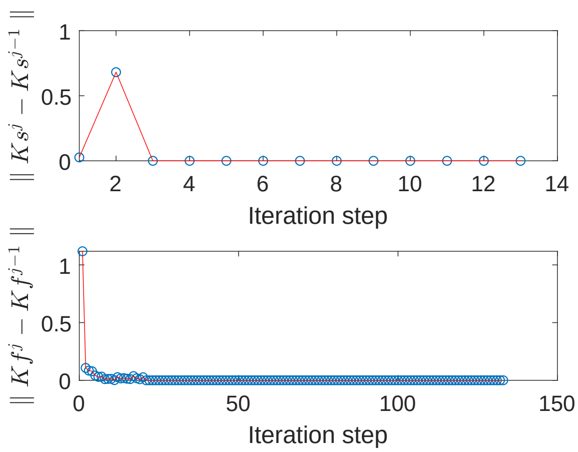


Fig. 10. Convergence of the slow and the fast control policies.

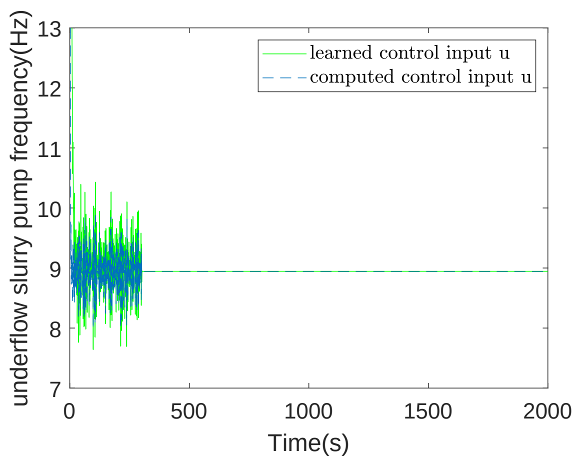


Fig. 11. Trajectories of the control input learned by Algorithm 3 and optimal computed control input.

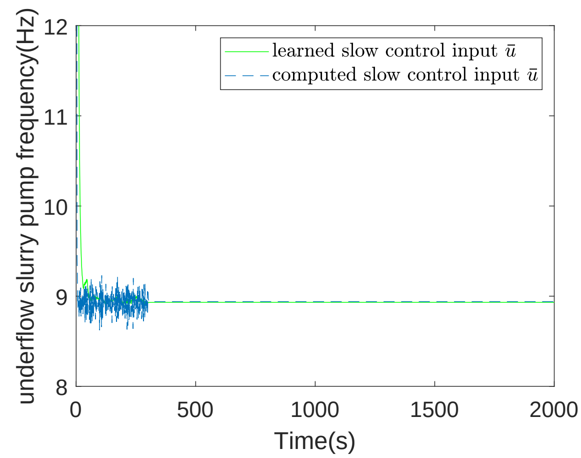


Fig. 12. Trajectories of the slow control input learned by Algorithm 3 and slow optimal computed control input.

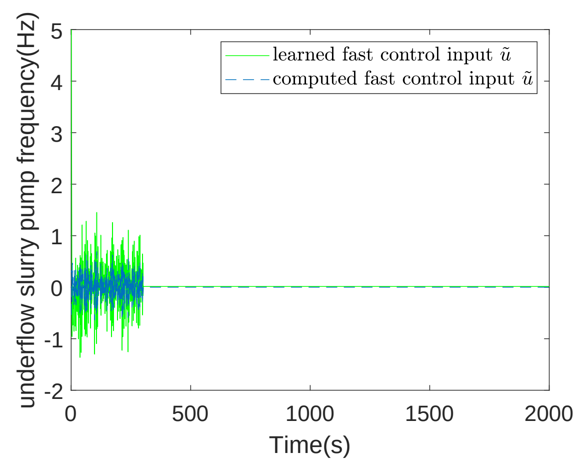


Fig. 13. Trajectories of the fast control input learned by Algorithm 3 and fast optimal computed control input.

to zero and finally the output y can reach its quasi-steady state $\bar{y}$. Fig. 9 shows that $u = \bar{u} + \tilde{u}$, and the fast control input $\tilde{u}$ for the boundary layer system goes zero. Fig. 10 shows convergence of the the control policies. Thus, the method proposed in this paper guarantees the tracking performance of the system with two time scales, making it reach a steady and optimal running state.

D. Comparison with optimal computed control

In this subsection, the comparison simulation results of the optimal control input learned by Algorithm 3 and the optimal control policy computed by solving the ARE with models of the fast and the slow subsystems are given. Fig. 11 shows the learned and computed composite control inputs of the original system, Fig. 12 shows the learned and computed control input of the slow decomposed subsystem and Fig. 13 shows that of the fast decomposed subsystem.

We observe that the solutions learned by Algorithm 3 match very well to the optimal control input obtained by solving two AREs of the decomposed subsystems with known models.

E. Comparison Simulation Experiment Using Q-learning

In this subsection, a comparison simulation experiment using only Q-learning without solving two time-scales problem is presented.

The simulation results are given from Fig. 14 to Fig. 16. From Subsection C and this Subsection, it is seen that the tracking performance of our two time-scales Q-learning approach are better than that of the comparison approach, also the operational indices of our approach converge faster than that of comparison approach.

To evaluate the control performance, the integral absolute error (IAE) and the mean square error (MSE) [55] are used and the evaluation equations are given by

$$\mathrm{IAE} = \sum_{i=k^*}^{k^*+n} |r(i) - r^*(i)| \tag{67}$$

$$\mathrm{MSE} = \sqrt{\frac{1}{n} \sum_{i=k^*}^{k^*+n} |r(i) - r^*(i)|^2} \tag{68}$$

and the comparison data is given in Table II.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2019.2912018, IEEE Transactions on Industrial Informatics

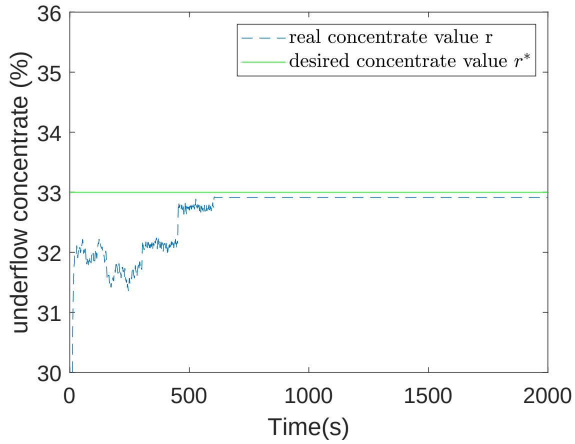


Fig. 14. The tracking performance of the underflow concentration to its setpoint.

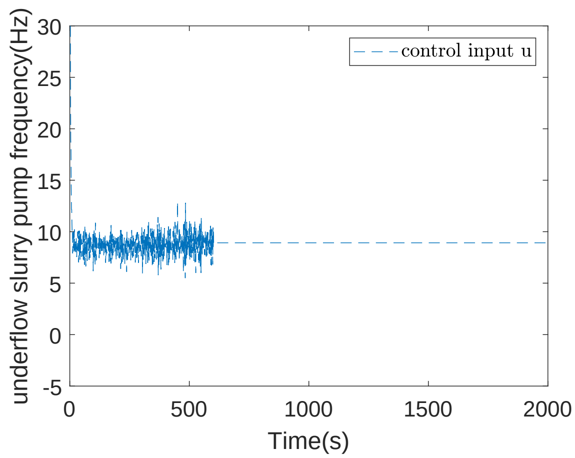


Fig. 15. The control input of the thickening process.

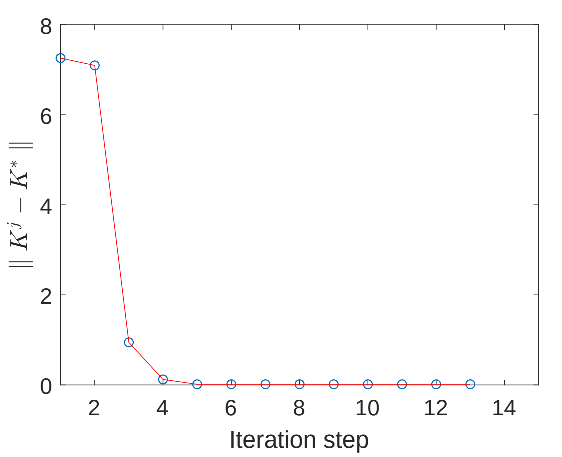


Fig. 16. The convergence of control policy to the optimal control policy.

TABLE II
DATA COMPARISON OF SIMULATION RESULTS

$800 < k^*$, $n \geq 1$	IAE	MSE
Algorithm 3	0.0522	0.0037
Comparison method	1.6408	0.1319

From Table II, the IAE and MSE of Algorithm 3 are both improved with respect to Q-learning without singular perturbation, implying that the tracking performance of the proposed controller is also improved.

VI. CONCLUSION

The data-driven optimal operation problem is addressed for two-time-scales industrial processes by integrating Q-learning algorithm and singular perturbation technique without requiring the completely knowledge of dynamics of device-layer control systems and operational indices. Using singular perturbation method decomposes the OOC problem of industrial processes into two reduced optimality problems, so that the fast and the slow controllers can be separately designed by applying two Q-learning algorithms. The final composite controller, obtained using only measured data from the plant, is employed to follow the desired operational indices with an approximately optimal approach. Furthermore, the convergence of the proposed algorithms, the stability of the system and the optimality of the operational process are guaranteed. A MSTP example shows the effectiveness and advantages of the proposed method. This approach saves computational cost compared with the classical design of the controller for systems with operational processes and device control processes, and improves control efficiency and tracking performance compared with optimal control for this kind of systems without solving its two-time-scales problem as shown in comparison experiment results.

Complex industrial processes present additional challenges that are considered as future work for this research. Input constraints considerations are required to address production safety concerns, data dropouts in data transmission, etc. Furthermore, applicability of our control procedure can be improved using nonlinear systems design.

REFERENCES

[1] T. Chai, S. J. Qin, and H. Wang, "Optimal operational control for complex industrial processes," *Annual Reviews in Contr.*, vol. 38, no. 1, pp. 81–92, Apr. 2014.

[2] T. Chai, J. Ding, and F. Wu, "Hybrid intelligent control for optimal operation of shaft furnace roasting process," *Contr. Eng. Practice*, vol. 19, no. 3, pp. 264–275, Mar. 2011.

[3] S. Engell, "Feedback control for optimal process operation," *J. Process Contr.*, vol. 17, no. 3, pp. 203–219, Mar. 2007.

[4] Y. Jiang, "Operational feedback multi-rate interval switch control of flotation processes," Master's thesis, College of Inf. Sci. Eng., Noreastern University, Shenyang, China, 2016.

[5] J. Fan, Y. Jiang, and T. Chai, "Operational feedback control of industrial processes in a wireless network environment," *Acta Automatica Sinica*, vol. 42, no. 8, pp. 1166–1174, Aug. 2016.

[6] W. Yu and C. Zhao, "Sparse exponential discriminant analysis and its application to fault diagnosis," *IEEE Trans. Ind. Electron.*, vol. 65, no. 7, pp. 5931–5940, Jul. 2018.

[7] ——, "Recursive exponential slow feature analysis for fine-scale adaptive processes monitoring with comprehensive operation status identification," *IEEE Trans. Ind. Informa.*, to be accepted, DOI: 10.1109/TII.2018.2878405.

[8] M. Ellis and P. D. Christofides, "Integrating dynamic economic optimization and model predictive control for optimal operation of nonlinear process systems," *Contr. Eng. Practice*, vol. 22, pp. 242–251, Jan. 2014.

[9] T. Chai, L. Zhao, J. Qiu, F. Liu, and J. Fan, "Integrated network-based model predictive control for setpoints compensation in industrial processes," *IEEE Trans. Ind. Inf.*, vol. 9, no. 1, pp. 417–426, Feb. 2013.

[10] F. Liu, H. Gao, J. Qiu, S. Yin, J. Fan, and T. Chai, "Networked multirate output feedback control for setpoints compensation and its application to rougher flotation process," *IEEE Trans. Ind. Electron.*, vol. 61, no. 1, pp. 460–468, Jan. 2014.

[11] J. Li, B. Kiumarsi, T. Chai, F. L. Lewis, and J. Fan, "Off-policy reinforcement learning: Optimal operational control for two-time-scale industrial processes," *IEEE Trans. Cybern.*, vol. 47, no. 12, pp. 4547–4558, Dec. 2017.

[12] J. Li, T. Chai, F. L. Lewis, J. Fan, Z. Ding, and J. Ding, "Off-policy q-learning: Set-point design for optimizing dual-rate rougher flotation operational processes," *IEEE Trans. Ind. Electron.*, vol. 65, no. 5, pp. 4092–4102, May 2018.

[13] A. Zanin, M. T. de Gouvea, and D. Odloak, "Industrial implementation of a real-time optimization strategy for maximizing production of lpg in a fcc unit," *Comput. Chem. Eng.*, vol. 24, no. 2-7, pp. 525–531, Jul. 2000.

[14] Y. Jiang, J. Fan, T. Chai, and F. L. Lewis, "Dual-rate operational optimal control for flotation industrial process with unknown operational model," *IEEE Trans. Ind. Electron.*, vol. 66, no. 6, pp. 4587–4599, Jun. 2019.

[15] J. Fan, Y. Jiang, and T. Chai, "Mpc-based setpoint compensation with unreliable wireless communications and constrained operational conditions," *Neurocomputing*, vol. 270, pp. 110–121, Dec. 2017.

[16] Y. Jiang, J. Fan, T. Chai, and T. Chen, "Setpoint dynamic compensation via output feedback control with network induced time delays," in *Proc. Amer. Contr. Conf.* Chicago, IL, USA, Jul. 2015, pp. 5384–5389.

[17] A. Zanin, M. T. De Gouvea, and D. Odloak, "Integrating real-time optimization into the model predictive controller of the fcc system," *Contr. Eng. Pract.*, vol. 10, no. 8, pp. 819–831, Aug. 2002.

[18] W. Dai, T. Chai, and S. X. Yang, "Data-driven optimization control for safety operation of hematite grinding process," *IEEE Trans. Ind. Electron.*, vol. 62, no. 5, pp. 2930–2941, May 2015.

[19] Y. Jiang, J. Fan, T. Chai, J. Li, and F. L. Lewis, "Data-driven flotation industrial process operational optimal control based on reinforcement learning," *IEEE Trans. Ind. Informa.*, vol. 14, no. 5, pp. 1974–1989, May 2018.

[20] P. J. Werbos, *A menu of designs for reinforcement learning over time*. MIT Press, Cambridge, MA, 1990.

[21] ——, "Neural networks for control and system identification," in *Proc. IEEE 28th CDC*. IEEE, Dec. 1989, pp. 260–265.

[22] D. Wang, D. Liu, Q. Wei, D. Zhao, and N. Jin, "Optimal control of unknown nonaffine nonlinear discrete-time systems based on adaptive dynamic programming," *Automatica*, vol. 48, no. 8, pp. 1825–1832, Aug. 2012.

[23] W. Gao and Z. P. Jiang, "Adaptive dynamic programming and adaptive optimal output regulation of linear systems," *IEEE Trans. Autom. Control.*, vol. 61, no. 12, pp. 4164–4169, Dec. 2016.

[24] C. Mu, D. Wang, and H. He, "Novel iterative neural dynamic programming for data-based approximate optimal control design," *Automatica*, vol. 81, pp. 240–252, Jul. 2017.

[25] B. A. G. Sutton R S, *Reinforcement learning: An introduction*. Cambridge, MT: MIT Press, 1998.

[26] B. Kiumarsi, F. L. Lewis, H. Modares, A. Karimpour, and M.-B. Naghibi-Sistani, "Reinforcement q-learning for optimal tracking control of linear discrete-time systems with unknown dynamics," *Automatica*, vol. 50, no. 4, pp. 1167–1175, Apr. 2014.

[27] W. Gao and Z.-P. Jiang, "Learning-based adaptive optimal tracking control of strict-feedback nonlinear systems," *IEEE Trans. Neural Netw. Learning Syst.*, vol. 29, no. 6, pp. 2614–2624, Jun. 2018.

[28] Z. Liu, Z. Wu, T. Li, J. Li, and C. Shen, "Gmm and cnn hybrid method for short utterance speaker recognition," *IEEE Trans. Ind. Informa.*, vol. 14, no. 7, pp. 3244–3252, Jul. 2018.

[29] W. He, T. Meng, X. He, and S. S. Ge, "Unified iterative learning control for flexible structures with input constraints," *Automatica*, vol. 96, pp. 326–336, Oct. 2018.

[30] W. He, Z. Yan, C. Sun, and Y. Chen, "Adaptive neural network control of a flapping wing micro aerial vehicle with disturbance observer," *IEEE Trans. Cybern.*, vol. 47, no. 10, pp. 3452–3465, Sep. 2017.

[31] B. Gupta, D. P. Agrawal, and S. Yamaguchi, *Handbook of research on modern cryptographic solutions for computer and cyber security*. IGI Global, 2016.

[32] X. Chen, J. Li, X. Huang, J. Ma, and W. Lou, "New publicly verifiable databases with efficient updates," *IEEE Trans. Dependable Secur. Comput.*, vol. 12, no. 5, pp. 546–556, Oct. 2015.

[33] J. Li, T. Chai, F. L. Lewis, Z. Ding, and Y. Jiang, "Off-policy interleaved q-learning: optimal control for affine nonlinear discrete-time systems," *IEEE Trans. Neural Netw. Learning Syst.*, to be published, DOI: 10.1109/TNNLS.2018.2861945.

[34] X. Lu, B. Kiumarsi, T. Chai, Y. Jiang, and F. L. Lewis, "Operational control of mineral grinding processes using adaptive dynamic programming and reference governor," *IEEE Trans. Ind. Informa.*, vol. 15, no. 4, pp. 2210–2221, Apr. 2019.

[35] T. Chen and B. A. Francis, *Optimal sampled-data control systems*. Springer Science & Business Media, 2012.

[36] L. Prandtl, "Über flussigkeitsbewegung bei sehr kleiner reibung," *Verhandl. III, Internat. Math.-Kong., Heidelberg, Teubner, Leipzig, 1904*, pp. 484–491, 1904.

[37] F. L. Lewis, S. Jagannathan, and A. Yesildirak, *Neural network control of robot manipulators and non-linear systems*. CRC Press, Taylor and Francls, UK, 1998.

[38] A. Klimushchev and N. Krasovskii, "Uniform asymptotic stability of systems of differential equations with a small parameter in the derivative terms," *J. Appl. Math. Mech.*, vol. 25, no. 9, pp. 1011–1025, Sep. 1962.

[39] X. Chen, M. Heidarinejad, J. Liu, and P. D. Christofides, "Composite fast-slow mpc design for nonlinear singularly perturbed systems," *AICHE J.*, vol. 58, no. 6, pp. 1802–1811, Mar. 2012.

[40] J. Chow and P. Kokotovic, "A decomposition of near-optimum regulators for systems with slow and fast modes," *IEEE Trans. Autom. Control*, vol. 21, no. 5, pp. 701–705, Oct. 1976.

[41] Rojas, D. Cipriano, and Aldo, "Model based predictive control of a rougher flotation circuit considering grade estimation in intermediate cells," *Dyna*, vol. 78, no. 166, pp. 29–37, Apr. 2011.

[42] K. Xu, *Singular Perturbation Control Theory*. Science Press, Beijing, CN, 1986.

[43] P. V. Kokotović, "Applications of singular perturbation techniques to control problems," *SIAM Rev.*, vol. 26, no. 4, pp. 501–550, Oct. 1984.

[44] V. R. Saksena, J. O'reilly, and P. V. Kokotovic, "Singular perturbations and time-scale methods in control theory: survey 1976–1983," *Automatica*, vol. 20, no. 3, pp. 273–293, May. 1984.

[45] H. K. Khalil, *Noninear systems*. Prentice-Hall, New Jersey, USA, 1996.

[46] D. Wang, D. Liu, Q. Zhang, and D. Zhao, "Data-based adaptive critic designs for nonlinear robust optimal control with uncertain dynamics," *IEEE Trans. Syst. Man, Cybern. A: Syst.*, vol. 46, no. 11, pp. 1544–1555, Nov. 2016.

[47] Y. Jiang, J. Fan, T. Chai, F. L. Lewis, and J. Li, "Tracking control for linear discrete-time networked control systems with unknown dynamics and dropout," *IEEE Trans. Neural Netw. Learning Syst.*, vol. 29, no. 10, pp. 4607–4620, Oct. 2018.

[48] B. Luo, D. Liu, T. Huang, and D. Wang, "Model-free optimal tracking control via critic-only q-learning," *IEEE Trans. Neural Netw. Learning Syst.*, vol. 27, no. 10, pp. 2134–2144, Oct. 2016.

[49] B. Luo, D. Liu, H. Wu, D. Wang, and F. L. Lewis, "Policy gradient adaptive dynamic programming for data-based optimal control," *IEEE Trans. Cybern.*, vol. 47, no. 10, pp. 3341–3354, Oct. 2017.

[50] A. Al-Tamimi, F. L. Lewis, and M. Abu-Khalaf, "Model-free q-learning designs for linear discrete-time zero-sum games with application to h-infinity control," *Automatica*, vol. 43, no. 3, pp. 473–481, Mar. 2007.

[51] ——, "Discrete-time nonlinear hjb solution using approximate dynamic programming: Convergence proof," *IEEE Trans. Syst. Man Cybern. B*, vol. 38, no. 4, pp. 943–949, Apr. 2008.

[52] Q. Wu, J. Fan, Y. Jiang, and T. Chai, "Data-driven dual-rate control for mixed separation thickening process in a wireless network environment," *Acta Automatica Sinica*, to be published, DOI: 1016383/j.aas.c180202.

[53] L. Wang, Y. Jia, T. Chai, and W. Xie, "Dual-rate adaptive control for mixed separation thickening process using compensation signal based approach," *IEEE Trans. Ind. Electron.*, vol. 65, no. 4, pp. 3621–3632, Apr. 2018.

[54] W. Xue, J. Fan, and Y. Jiang, "Flotation process with model free adaptive control," in *IEEE Inter. Conf. Inf. Automa.* Macao, CN, Jul. 2017, pp. 442–447.

[55] Y. Jiang, J. Fan, Y. Jia, and T. Chai, "Data-driven flotation process operational feedback decoupling control," *Acta Automatica Sinica*, vol. 45, no. 4, pp. 759–770, Apr. 2019.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2019.2912018, IEEE Transactions on Industrial Informatics

Wenqian Xue received the B.E. degree in automation from Qingdao University, Qingdao, China, in 2015, and the M.S. degree in control theory and control engineering from State Key Laboratory of Synthetical Automation for Process Industries in Northeastern University, Shenyang, Liaoning, China, in 2018, respectively, where she is currently working toward the Ph.D. degree. Her research interests include industrial process operational control and reinforcement learning.

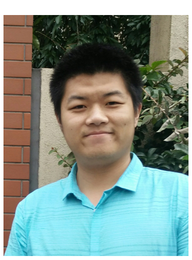
Jialu Fan (M'12) received the B.E. degree in automation from Northeastern University, Shenyang, China, in 2006, and the Ph.D. degree in control science and engineering from Zhejiang University, Hangzhou, China, in 2011. She was a Visiting Scholar with the Pennsylvania State University during 2009-2010. She was a recipient of the ICAMechS 2018 Best Paper Award and the Best Student Paper Award of the 29th Chinese Process Control Conference.

Currently, she is a Tenured Associate Professor with the State Key Laboratory of Synthetical Automation for Process Industries, Northeastern University, Shenyang, China. Her research interests include networked control systems, industrial process operational control, reinforcement learning, and mobile social networks.

Victor G. Lopez (S'17) received the B.S. degree from the Universidad Autonoma de Campeche, Mexico, in 2010 and the M.S. degree from the Research and Advanced Studies Center (CINVESTAV), Mexico, in 2013. He is currently a Ph.D. student at the University of Texas at Arlington, TX, USA. Victor was a Lecturer at the Western Technologic Institute of Superior Studies (ITESO) in Guadalajara, Mexico, in 2015. His research interests include cyber-physical systems, game theory, distributed control, reinforcement learning and robust control.

Jinna Li (M'12) received the M.S. degree and the Ph.D. degree from Northeastern University, Shenyang, China, 2006 and 2009, respectively. She is a professor at School of Information and Control Engineering, Liaoning Shihua University, Fushun, P.R. China.

She was a Visiting Scholar granted by China Scholarship Council with Energy Research Institute, Nanyang Technological University, Singapore from June to June, 2015, a Domestic Young Core Visiting Scholar granted by Ministry of Education of China with State Key Lab of Synthetical Automation for Process Industries, Northeastern University from September 2015 to June 2016, a visiting scholar with the University of Manchester, the United Kingdom From January to July, 2017. Her current research interests include operational optimal control, reinforcement learning, distributed optimization control.



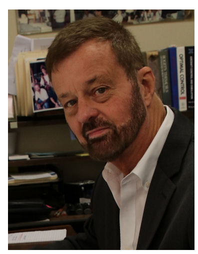
Yi Jiang (S'14) received the B.E. degree in automation and M.S. degree in control theory and control engineering from Information Science and Engineering College and State Key Laboratory of Synthetical Automation for Process Industries in Northeastern University, Shenyang, China in 2014 and 2016, respectively, where he is currently working toward the Ph.D. degree.

From January to July, 2017, he was a Visiting Scholar with the UTA Research Institute, University of Texas at Arlington, TX, USA, and from March 2018 to March 2019, he was a Research Assistant with the University of Alberta, Edmonton, Canada.His research interests include networked control systems, industrial process operational control, reinforcement learning, event-triggered estimation and control.

Tianyou Chai (M'90-SM'97-F'08) received the Ph.D. degree in control theory and engineering from Northeastern University, Shenyang, China, in 1985.

He has been with the Research Center of Automation, Northeastern University, Shenyang, China, since 1985, where he became a Professor in 1988 and a Chair Professor in 2004. He is the founder and Director of the Center of Automation, which became a National Engineering and Technology Research Center in 1997. He has made a number of important contributions in control technologies and applications. He has authored and coauthored two monographs, 84 peer reviewed international journal papers and around 219 international conference papers. He has been invited to deliver more than 20 plenary speeches in international conferences of IFAC and IEEE. His current research interests include adaptive control, intelligent decoupling control, integrated plant control and systems, and the development of control technologies with applications to various industrial processes.

Prof. Chai is a member of the Chinese Academy of Engineering, an academician of International Eurasian Academy of Sciences, and an IFAC Fellow. He is a Distinguished Visiting Fellow of The Royal Academy of Engineering (UK) and an Invitational Fellow of Japan Society for the Promotion of Science (JSPS). For his contributions, he has won three prestigious awards of National Science and Technology Progress, the 2002 Technological Science Progress Award from the Ho Leung Ho Lee Foundation, the 2007 Industry Award for Excellence in Transitional Control Research from the IEEE Control Systems Society, and the 2010 Yang Jia-Chi Science and Technology Award from the Chinese Association of Automation.



Frank L. Lewis (S'70-M'81-SM'86-F'94) received the bachelors degree in physics/electrical engineering and the M.S.E.E. degree from Rice University, Houston, TX, USA, the M.S. degree in aeronautical engineering from the University of West Florida, Pensacola, FL, USA, and the Ph.D. degree from the Georgia Institute of Technology, Atlanta, GA, USA.

He is currently a Distinguished Scholar Professor and Distinguished Teaching Professor with the University of Texas at Arlington, Fort Worth, TX, USA, the Moncrief-ODonnell Chair of the University of Texas at Arlington Research Institute, Fort Worth, the Qian Ren Thousand Talents Professor and Project 111 Professor with Northeastern University, Shenyang, China, and a Distinguished Visiting Professor with the Nanjing University of Science and Technology, Nanjing, China. He is involved in feedback control, intelligent systems, cooperative control systems, and nonlinear systems. He has authored six U.S. patents, numerous journal special issues, journal papers, and 20 books, including Optimal Control, Aircraft Control, Optimal Estimation, and Robot Manipulator Control, which are used as university textbooks worldwide.

Prof. Lewis is a member of the National Academy of Inventors, a fellow of the International Federation of Automatic Control and the Institute of Measurement and Control, U.K., a Texas Board of Professional Engineer, a U.K. Chartered Engineer, and a founding member of the Board of Governors of the Mediterranean Control Association. He was a recipient of the Fulbright Research Award, the NSF Research Initiation Grant, the Terman Award from the American Society for Engineering Education, the Gabor Award from the International Neural Network Society, the Honeywell Field Engineering Medal from the Institute of Measurement and Control, U.K., the Neural Networks Pioneer Award from IEEE Computational Intelligence Society, the Outstanding Service Award from the Dallas IEEE Section, and the Texas Regents Outstanding Teaching Award in 2013. He was elected as an Engineer of the year by the Fort Worth IEEE Section. He was listed in the Fort Worth Business Press Top 200 Leaders in Manufacturing.