

# 工业过程多速率分层运行优化控制

代伟<sup>1</sup> 陆文捷<sup>1</sup> 付俊<sup>2</sup> 马小平<sup>1</sup>

**摘要** 工业过程运行优化控制通常采用基础回路层和运行层两层结构, 涉及不同时间尺度特性的被控对象, 且由于检测装置采样周期不同难以统一控制与采样周期; 此外, 运行层动态往往机理复杂难以建模. 因此针对这一多层次、多时间尺度且部分模型未知的复杂多速率控制问题, 本文提出一种工业过程多速率分层运行优化控制方法. 该方法在使用提升技术解决分层多速率问题的基础上, 采用一种基于 Q-学习的数据驱动运行层设定值优化方法, 更新基础回路层的设定值; 并针对提升后的系统采用模型预测控制 (Model predictive control, MPC) 方法设计基础回路层控制器以跟踪设定值, 从而实现运行指标的优化控制. 对典型工业闭路磨矿过程进行了仿真实验, 验证了本文所提方法的有效性.

**关键词** 多速率, 多时间尺度, 分层运行优化控制, Q-学习, 模型预测控制

**引用格式** 代伟, 陆文捷, 付俊, 马小平. 工业过程多速率分层运行优化控制. 自动化学报, 2019, 45(10): 1946–1959

**DOI** 10.16383/j.aas.2018.c180300

## Multi-rate Layered Optimal Operational Control of Industrial Processes

DAI Wei<sup>1</sup> LU Wen-Jie<sup>1</sup> FU Jun<sup>2</sup> MA Xiao-Ping<sup>1</sup>

**Abstract** The optimal operational control of industrial processes usually adopts a two-layer structure (i.e., the operation layer and basic loop layer), which involves the controlled objects with different-time-scale characteristics, and has unsynchronized control and the sampling periods because of the difference of the sensor sampling periods. In addition, the dynamics of the operation layer is usually too complex to be modeled. Therefore, in order to solve this complex multi-rate control problem with characteristics of multi-layer, multi-time scale and unknown partial dynamics, this paper proposes a multi-rate layered optimal operational control method for industrial processes. On the basis of using the lifting technology to solve the multi-rate layered problem, a data-driven set point optimization method based on Q-learning is first employed to update the basic loop setpoints; and then a model predictive control is used to design the base loop controller according the lifted system to track the updated setpoints, thereby realizing the optimization control of operation indices. Experiments have been carried out on a closed-loop grinding process, which shows the effectiveness of the proposed method.

**Key words** Multi-rate, multi-time scale, layered optimal operational control, Q-learning, model predictive control (MPC)

**Citation** Dai Wei, Lu Wen-Jie, Fu Jun, Ma Xiao-Ping. Multi-rate layered optimal operational control of industrial processes. *Acta Automatica Sinica*, 2019, 45(10): 1946–1959

先进控制与优化技术一直被认为是复杂工业过程提质增效、实现经济利润提高的关键. 如今, 其前

沿核心技术之一是工业过程运行优化控制<sup>[1-2]</sup>, 内涵是采用信息技术, 围绕生产过程的知识与数据信息进行集成, 通过过程运行优化与控制的智能化和集成化, 在保证过程安全运行的条件下, 不仅使基础回路输出很好地跟踪设定值, 而且控制整个运行过程, 使其在生产条件约束下实现质量、效率和能耗等运行指标的最优化<sup>[3]</sup>.

当前工业过程运行优化控制通常采用分层结构, 其优点在于它提供了优化和控制层任务的清晰界限, 在被控对象和时间尺度上均有不同. 运行层面向工业运行过程, 具有慢时间尺度特性, 其目标是优化设定值; 基础回路层面向基础设备/回路过程, 具有快时间尺度特性, 其目标是实现设定值的跟踪.

在工业过程运行优化控制的分层结构下, 已取得了大量基于模型的研究成果, 如 SOC (Self-optimizing control) 以及 SOC 与 NCO tracking

收稿日期 2018-05-12 录用日期 2018-10-06  
Manuscript received May 12, 2018; accepted October 6, 2018  
国家自然科学基金 (61603393, 61503384, 61741318), 江苏省自然科学基金 (BK20160275), 中国博士后科学基金 (2015M581885, 2018T110571), 流程工业综合自动化国家重点实验室开放基金 (PAL-N201706) 资助  
Supported by National Natural Science Foundation of China (61603393, 61503384, 61741318), Natural Science Foundation of Jiangsu Provinces (BK20160275), the Postdoctoral Science Foundation of China (2015M581885, 2018T110571), State Key Laboratory of Synthetical Automation for Process Industries (PAL-N201706)  
本文责任编辑 王鼎  
Recommended by Associate Editor WANG Ding  
1. 中国矿业大学信息与控制工程学院 徐州 221116 2. 东北大学流程工业综合自动化国家重点实验室 沈阳 110004  
1. School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221116 2. State Key Laboratory of Synthetical Automation for Process Industries, Northeastern University, Shenyang 110004

(Necessary conditions of optimality tracking) 集成、RTO (Real-time optimization) 以及 RTO 与 MPC (Model predictive control) 集成的优化控制方法。

SOC 以静态经济效益模型为目标函数, 离线求解一组满足生产约束的被控变量及其设定值, 使工业过程受扰后, 不改变设定值仍可处于近似经济最优状态<sup>[4-5]</sup>。SOC 与 NCO tracking 的集成方法是在 SOC 选择基础回路层的被控变量后, 采用 NCO tracking 更新设定值, 以提高动态性能<sup>[6]</sup>。但对于干扰源众多或者干扰变化幅度较大的工业过程来说, 难以利用 SOC 确定被控变量, 或者满足 SOC 基本条件的被控变量根本不存在。基于 RTO 的运行优化控制方法, 是基于基础回路层稳态的过程模型求解运行层的最优设定值<sup>[7]</sup>, 但其只能在系统达到稳定时才进行优化, 优化周期较长, 难以处理动态扰动。基于拟稳态概念的小周期采样的 RTO 方法<sup>[8]</sup>由于时间尺度难以把握, 有时会导致生产不可控制的问题, 难以工业应用。RTO 与 MPC 的集成方法<sup>[9-11]</sup>在基础回路层采用 MPC 方法, 运行层采用 RTO 求解期望的稳态最优值, 并设计运行层 MPC 计算基础回路层的被控变量和 MPC 输入量的设定值, 上层 MPC 优化与底层 MPC 控制采用相同的周期, 实现扰动产生后设定值的快速调整。但 RTO 与 MPC 的集成方法存在运行层稳态模型和基础回路层动态模型失配以及网络问题, 影响优化效果, 为此, 运行层采用过程动态模型的 DRTO (Dynamic RTO) 与 MPC 的集成方法<sup>[12]</sup>、双层 MPC 方法<sup>[13]</sup>以及 RTO 与 Network-based MPC 的集成方法<sup>[14]</sup>被相继提出。但实际工业过程, 如冶金、选矿, 运行过程机理复杂, 加之设备状况及生产环境的变化影响, 导致数学模型难以精确建立<sup>[1-2]</sup>, 从而制约了上述基于模型的分层运行优化控制方法的实际工业应用。

知识驱动的分层运行优化控制方法在假设基础回路控制器可保证设定值快速跟踪的条件下, 集中研究基于专家系统<sup>[15]</sup>、案例推理<sup>[16]</sup>、模糊规则推理技术<sup>[17]</sup>及各种技术相集成<sup>[18]</sup>的运行层智能设定值优化方法。知识驱动的优化控制方法主要是依靠模拟领域专家的经验与知识, 与环境的交互能力差, 当工况变化时, 常常依然需要领域专家对控制器进行调整, 而人的主观性和随意性导致系统难以优化运行。实际工业生产中存在大量能够反映过程运行机理和运行状态的实时与历史运行数据, 因此采用数据驱动技术来研究分层运行优化控制方法已成为当前的热点。

自适应评价设计 (Adaptive critic design, ACD) 是强化学习在控制理论领域中的一种近似

方法, 被广泛用于解决模型未知的最优控制问题<sup>[19]</sup>。与 ACD 相类似的名称还包括自适应动态规划<sup>[20]</sup>、神经动态规划<sup>[21]</sup>等。文献 [22] 借鉴 ACD 思想, 在假设基础回路层稳定跟踪的基础上, 在运行层提出了由评价网络和执行网络串联组成的数据驱动控制回路设定值优化方法。文献 [23] 针对磨矿特性, 提出强化学习与传统 PI 集成的分层运行优化控制方法。文献 [24] 在基础回路层采用多回路 PI 控制的基础上, 提出一种基于 Q-学习的次优设定值设计方法。文献 [25] 提出一种零和博弈脱策强化学习与传统 PI 相集成的分层运行优化控制。上述分层优化控制方法考虑了两层快慢时间尺度特性, 实现了设定值的动态调整。然而, 实际工业过程控制系统中, 因各检测装置的采样周期不同, 使得基础回路层的采样和控制周期可能不同, 即基础回路层也存在多速率问题, 因此, 工业过程的运行优化控制是具有三种及以上速率的多速率分层结构, 导致上述算法难以直接使用。

本文结合工业过程分层控制结构下的多速率特点, 针对一类基础回路层模型已知, 运行层模型未知的工业过程, 提出一种基于 Q-学习、MPC 与提升技术的集成运行优化控制方法。其首先在基础回路层采用提升技术对采样与控制周期不一致的动态模型提升至框架周期, 对增维后的动态模型采用 MPC 进行控制器设计; 其次, 运行层采用一种数据驱动的 Q-学习方法, 借助工业过程中采集的数据优化基础回路层的设定值, 从而实现运行优化控制的目标; 最后以闭路磨矿为背景进行实验研究, 验证了方法的有效性。

## 1 分层运行优化控制问题描述

本部分首先简要介绍工业过程运行优化控制的层级架构, 然后对多速率分层运行优化控制问题及难点进行分析, 并给出本文的控制策略。

### 1.1 工业过程分层运行优化控制架构分析

工业过程分层运行优化控制结构如图 1 所示。其中  $r$  表示工业过程运行指标, 分层控制的目标即是通过选取一个合适的基础回路设定值  $w$ , 并通过调节基础回路控制量  $u$  使回路输出  $y$  跟踪  $w$ , 从而控制实际运行指标  $r$  在期望运行指标  $r^*$  附近。

工业过程包括回路过程和运行过程两层动态, 其运行过程往往机理复杂难以建模, 且整个系统因层级功能以及所涉及的被控对象特性的差异导致层级间时间尺度各异, 以及由于信号变化速率相差较大, 各检测装置的采样周期不同, 难以统一基础回路的控制与采样周期。由此可以看出, 工业过程运行优化控制是一个多层次、多时间尺度且部分模型未知

的多速率控制问题,对现有控制理论提出了挑战.如何针对多时间尺度被控对象特性,在分层的多速率控制结构下,将运行层优化与基础回路层控制相集成,模型与数据相结合,实现运行优化控制目标是本文研究的重点.

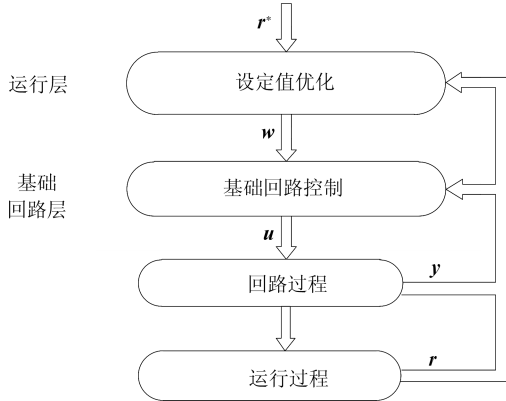


图 1 多速率工业过程的双层层级架构

Fig.1 Two-layer structure of multi-rate industrial processes

### 1.2 运行优化控制策略

由上述分析可知,回路过程直接与检测装置以及执行机构相接触,而各种仪表因其自身的特殊性,获取信息和处理信息的速度不同,加之控制节点分散,因此在很多复杂的实际工业过程控制中,系统的控制更新速率与采样速率并不一致,且通常对控制输入的在线更新速度要求较高.因此本文考虑的基础回路层多速率采样控制策略是针对基础回路过程以  $T_2$  为周期进行等周期采样,而基础回路控制器的输出由周期为  $T_1$  的保持器转换得到的情况,  $T_1$  与  $T_2$  之间满足关系式  $T_1 < T_2$ . 此外,由于工业过程的运行指标与质量、效率和能耗相关,往往以固定的周期统一上报至控制系统,加之系统对设定值的在线更新速度要求不高,因此通常将运行层控制周期与运行指标采样周期一致设置为  $T$ ,且相对于基础回路过程,运行过程具有慢时间尺度特性,因此  $T$  远大于  $T_1$  和  $T_2$ .

本文解决上述多速率分层运行优化控制系统设计问题的思路是:首先,通过提升技术将基础回路层的控制与采样周期均提升至框架周期  $T_0$ ,使该层速率一致,并采用 MPC 设计基础回路控制器;然后,针对框架周期  $T_0$  普遍小于运行层控制周期  $T$  ( $T = \xi T_0$ ,  $\xi$  为正整数) 的问题,进一步将基础回路层的周期  $T_0$  提升至运行层控制周期  $T$ ,使基础回路层的快时间尺度动态特性与运行层的慢时间尺度一致,从而构建增广状态模型用于描述由基础回路控制器、回路过程和运行过程组成的广义被控对象;继而针对这一部分模型未知的广义被控对象,提出

一种基于 Q-学习的数据驱动运行层设定值优化方法,以根据系统运行状态在线更新设定值.

## 2 基于提升技术和 MPC 的基础回路层控制器设计

### 2.1 基于提升技术的基础回路层多速率处理方法

工业实际中的回路过程往往为非线性动态模型,然而它们通常在运行点附近稳态运行,因此能在运行点附近线性化.

考虑如下连续时间基础回路过程模型:

$$\begin{cases} \dot{\mathbf{x}}_f(t) = A_c \mathbf{x}_f(t) + B_c \mathbf{u}(t) \\ \mathbf{y}(t) = C_c \mathbf{x}_f(t) + D_c \mathbf{u}(t) \end{cases}, D_c = 0 \quad (1)$$

其中,  $\mathbf{x}_f \in \mathbf{R}^{n_x}$ ,  $\mathbf{u} \in \mathbf{R}^{n_u}$ ,  $\mathbf{y} \in \mathbf{R}^{n_y}$  分别为基础回路层被控对象即回路过程的状态、控制输入和控制输出变量,  $n_x$ 、 $n_u$  和  $n_y$  表示相应变量的维数;  $A_c$ ,  $B_c$ ,  $C_c$  和  $D_c$  为回路过程模型的系统矩阵.

控制周期  $T_1$  与采样周期  $T_2$  有如下关系:  $T_1 = ph$ ,  $T_2 = qh$ ,  $p$  小于  $q$  且互为质数,  $h$  为基周期,由此可得框架周期  $T_0 = pqh$ ,即  $T_1$  和  $T_2$  的最小公倍数.采用提升技术使基础回路层的控制与采样周期均提升至框架周期  $T_0$ ,以解决基础回路层中的输入输出多速率控制问题,具体方法如下.

首先以基周期  $h$  将基础回路层模型离散化为

$$\begin{cases} \mathbf{x}_f(n+1) = A_f \mathbf{x}_f(n) + B_f \mathbf{u}(n) \\ \mathbf{y}(n) = C_f \mathbf{x}_f(n) + D_f \mathbf{u}(n) \end{cases} \quad (2)$$

其中,  $A_f = e^{A_c h}$ ,  $B_f = \int_{kh}^{(k+1)h} e^{[(k+1)h-\tau]} d\tau B_c = \int_0^h e^{A_c t} dt B_c$ ,  $C_f = C_c$ ,  $D_f = D_c$ .

对离散化后的系统进行提升,将输入向量维度提升  $q$  倍,输出向量维度提升  $p$  倍,即将  $\mathbf{u}$  和  $\mathbf{y}$  的周期均提升至  $T_0$ ,  $m$  表示采样周期  $T_0$  下的运行步数.提升过程定义如下:

$$\bar{\mathbf{u}}(m) = L_q \mathbf{u}(mT_0) = \begin{bmatrix} \mathbf{u}(mT_0) \\ \mathbf{u}(mT_0 + T_1) \\ \vdots \\ \mathbf{u}(mT_0 + (q-1)T_1) \end{bmatrix}$$

$$\bar{\mathbf{y}}(m) = L_p \mathbf{y}(mT_0) = \begin{bmatrix} \mathbf{y}(mT_0) \\ \mathbf{y}(mT_0 + T_2) \\ \vdots \\ \mathbf{y}(mT_0 + (p-1)T_2) \end{bmatrix}$$

其中,  $\bar{\mathbf{u}}(m) \in \mathbf{R}^{N_u}$ ,  $\bar{\mathbf{y}}(m) \in \mathbf{R}^{N_y}$ ,  $N_u = qn_u$ ,  $N_y = pn_y$ .

提升后的基础回路层状态空间模型如下:

$$\begin{cases} \mathbf{x}_f(m+1) = \bar{\mathbf{A}}\mathbf{x}_f(m) + \bar{\mathbf{B}}\bar{\mathbf{u}}(m) \\ \bar{\mathbf{y}}(m) = \bar{\mathbf{C}}\mathbf{x}_f(m) + \bar{\mathbf{D}}\bar{\mathbf{u}}(m) \end{cases} \quad (3)$$

为了便于表示各系统矩阵先做如下定义:

$$F(i, p) := \sum_{j=0}^{p-1} A_f^{i+j} = A_f^i + A_f^{i+1} + \cdots + A_f^{i+p-1}$$

$$F(0, 0) = I, p > 0; i \geq 0$$

$$A_f^j F(i, p) = F(i+j, p), j \geq 0$$

$$A_{fp} := e^{A_c T_1} = A_f^p, A_{fq} := e^{A_c T_2} = A_f^q$$

$$B_{fp} := \int_0^{T_1=ph} e^{A_c t} dt B_c =$$

$$(A_f^{p-1} + A_f^{p-2} + \cdots + A_f + I)B_f = F(0, p)B_f$$

则

$$\bar{\mathbf{A}} = e^{A_c T} = A_f^{pq} = A_{fp}^q = A_{fq}^p \in \mathbf{R}^{n_x \times n_x}$$

$$\bar{\mathbf{B}} = [A_{fp}^{q-1} B_{fp}, A_{fp}^{q-2} B_{fp}, \cdots, A_{fp} B_{fp}, B_{fp}] =$$

$$[F(pq-p, p)B_f, F(pq-2p, p)B_f, \cdots,$$

$$F(0, p)B_f] \in \mathbf{R}^{n_x \times N_u}$$

$$\bar{\mathbf{C}} = \begin{bmatrix} C_f \\ C_f A_{fq} \\ \vdots \\ C_f A_{fq}^{(p-1)} \end{bmatrix} \in \mathbf{R}^{N_y \times n_x}$$

$$\bar{\mathbf{D}} = \begin{bmatrix} D_f & 0 & \cdots & \cdots & \cdots \\ D_{f10} & D_{f11} & \cdots & D_{f1c_1} & 0 \\ D_{f20} & D_{f21} & \cdots & \cdots & D_{f2c_2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ D_{fp-1,0} & D_{fp-1,1} & \cdots & \cdots & \cdots \\ \cdots & \cdots & 0 & \vdots & \vdots \\ \cdots & \cdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \vdots & \vdots & \vdots \\ \ddots & \ddots & \vdots & \vdots & \vdots \\ \cdots & D_{fp-1, c_{p-1}} & 0 & \vdots & \vdots \end{bmatrix} \in \mathbf{R}^{N_y \times N_u}$$

$$D_{fij} = C_f F(iq - jp - p)B_f,$$

$$j = 0, 1, 2, \cdots, c_i - 1$$

$$D_{fic_i} = C_f F(0, d_i)B_f + D_f$$

**注 1.** 对多时间尺度系统, 当  $p$  和  $q$  为互质整数时, 对每一个  $i$ ,  $0 \leq i \leq p-1$ , 存在整数  $c_i \geq 0$  和  $0 \leq d_i \leq p$  满足  $iq = c_i p + d_i$ ,  $i = 1, 2, \cdots, p-1$ .

**注 2.** 经过提升后的系统 (3), 本质上是系统将多速率系统以“块”的思想对输入输出信号进行“堆叠”, 以一个大于系统所有周期的时间为大周期即框架周期  $T_0$ , 对系统信息进行采集, 使得输入和输出在框架周期  $T_0$  内以其各自的实际周期进行增维. 从而所组成的新系统方程既不损失原系统信息, 又将多速率系统化为单速率, 便于系统设计与分析.

## 2.2 基础回路层控制器设计

提升至框架周期  $T_0$  之后的基础回路层控制系统  $(\bar{\mathbf{A}}, \bar{\mathbf{B}}, \bar{\mathbf{C}}, \bar{\mathbf{D}})$  为一个多输入多输出的非方系统, 且相比于原系统, 不仅输入输出的维度大大增加, 而且变量之间的耦合性也随之加强, 难以实施常规的 PID 控制方法. 因此, 本文采用 MPC 对其进行控制. 为简单起见, 推导过程中的预测时域和控制时域均设置为 1, 具体如下.

根据式 (3) 可得到预测模型为

$$\mathbf{x}_f(m+1 | m) = \bar{\mathbf{A}}\mathbf{x}_f(m) + \bar{\mathbf{B}}\bar{\mathbf{u}}(m) \quad (4)$$

$$\begin{aligned} \bar{\mathbf{y}}(m+1 | m) &= \bar{\mathbf{C}}\mathbf{x}_f(m+1 | m) + \bar{\mathbf{D}}\bar{\mathbf{u}}(m+1) = \\ &= \bar{\mathbf{C}}\bar{\mathbf{A}}\mathbf{x}_f(m) + \bar{\mathbf{C}}\bar{\mathbf{B}}\bar{\mathbf{u}}(m) + \bar{\mathbf{D}}\bar{\mathbf{u}}(m+1) \end{aligned} \quad (5)$$

由于工业过程均为连续变化的系统, 且底层基础回路控制采样与控制周期较小, 通常为秒级或毫秒级, 前后连续两个时刻系统变化不大, 因此, 为简单起见, 将  $\bar{\mathbf{u}}(m+1)$  用  $\bar{\mathbf{u}}(m)$  来代替, 则上式改写为

$$\bar{\mathbf{y}}(m+1 | m) = S_x \mathbf{x}_f(m) + S_u \bar{\mathbf{u}}(m) \quad (6)$$

其中,  $S_x = \bar{\mathbf{C}}\bar{\mathbf{A}}$ ,  $S_u = \bar{\mathbf{C}}\bar{\mathbf{B}} + \bar{\mathbf{D}}$ .

工业过程控制系统通常希望过程变量与设定值尽可能接近外, 还希望生产过程中的各类消耗尽可能低, 故采用如下目标函数:

$$J_{\text{mpc}} = \|\Gamma_{\bar{\mathbf{y}}}(\bar{\mathbf{y}}(m+1 | m) - \alpha \mathbf{w}(m+1))\|^2 + \|\Gamma_{\bar{\mathbf{u}}}\bar{\mathbf{u}}(m)\|^2 \quad (7)$$

其中,  $\alpha = [I_{n_y \times n_y} \cdots I_{n_y \times n_y}]^T \in \mathbf{R}^{N_y \times n_y}$  表示将设定值  $\mathbf{w}$  的维度提升至与  $\bar{\mathbf{y}}$  相同的系数矩阵,  $\Gamma_{\bar{\mathbf{y}}}$ ,  $\Gamma_{\bar{\mathbf{u}}}$  分别为预测控制输出和控制量的加权因子, 其中,

$$\Gamma_{\bar{\mathbf{u}}} := \text{diag}\{\Gamma_{\bar{u}_1}, \Gamma_{\bar{u}_2}, \cdots, \Gamma_{\bar{u}_{N_u}}\}$$

$$\Gamma_{\bar{\mathbf{y}}} := \text{diag}\{\Gamma_{\bar{y}_1}, \Gamma_{\bar{y}_2}, \cdots, \Gamma_{\bar{y}_{N_y}}\}$$

为求解  $\min_{\bar{\mathbf{u}}(m)} J_{\text{mpc}}(\mathbf{x}_f(m), \bar{\mathbf{u}}(m))$  的最优控制问题, 首先定义辅助变量:

$$\rho := \begin{bmatrix} \Gamma_{\bar{\mathbf{y}}}(\bar{\mathbf{y}}(m+1 | m) - \alpha \mathbf{w}(m+1)) \\ \Gamma_{\bar{\mathbf{u}}}\bar{\mathbf{u}}(m) \end{bmatrix} \quad (8)$$

则目标函数 (7) 变为

$$J_{\text{mpc}} = \rho^T \rho \quad (9)$$

将预测方程 (6) 代入式 (8), 可得:

$$\rho = \begin{bmatrix} \Gamma_{\bar{y}} S_u \\ \Gamma_{\bar{u}} \end{bmatrix} \bar{\mathbf{u}}(m) - \begin{bmatrix} \Gamma_{\bar{y}}(\alpha \mathbf{w}(m+1) - S_x \mathbf{x}_f(m)) \\ 0 \end{bmatrix} \quad (10)$$

求  $J_{\text{mpc}}$  的极值, 即令  $\frac{d\rho^T \rho}{d\bar{\mathbf{u}}(m)} = 0$ , 不难得到最优控制律为

$$\bar{\mathbf{u}}(m) = K_{\text{mpc}}(\alpha \mathbf{w}(m+1) - S_x \mathbf{x}_f(m)) \quad (11)$$

其中,  $K_{\text{mpc}} = (S_u^T \Gamma_{\bar{y}}^T \Gamma_{\bar{y}} S_u + \Gamma_{\bar{u}}^T \Gamma_{\bar{u}})^{-1} S_u^T \Gamma_{\bar{y}}^T \Gamma_{\bar{y}}$ .

由于运行层具有慢时间尺度特性, 其优化出的设定值  $\mathbf{w}$  在运行控制周期  $T$  内保持不变, 故  $\mathbf{w}(m+i) = \mathbf{w}(m)$ ,  $i = 1, 2, \dots, \xi - 1$  ( $\xi = T/T_0$ ), 由此可得:

$$\bar{\mathbf{u}}(m) = K_{\text{mpc}}(\alpha \mathbf{w}(m) - S_x \mathbf{x}_f(m)) \quad (12)$$

将式 (12) 代入式 (3), 可得系统闭环方程:

$$\begin{cases} \mathbf{x}_f(m+1) = \hat{A} \mathbf{x}_f(m) + \hat{B} \mathbf{w}(m) \\ \bar{\mathbf{y}}(m) = \hat{C} \mathbf{x}_f(m) + \hat{D} \mathbf{w}(m) \end{cases} \quad (13)$$

其中,  $\hat{A} = \bar{A} - \bar{B} K_{\text{mpc}} S_x$ ,  $\hat{B} = \bar{B} K_{\text{mpc}} \alpha$ ,  $\hat{C} = \bar{C} - \bar{D} K_{\text{mpc}} S_x$ ,  $\hat{D} = \bar{D} K_{\text{mpc}} \alpha$ .

**注 3.** 将  $\bar{\mathbf{u}}(m+1)$  用  $\bar{\mathbf{u}}(m)$  来代替只是工程上的一种近似处理方法, 对于变化较快的工业过程, 可计算求解两个控制时间序列即  $\bar{\mathbf{u}}(m)$  和  $\bar{\mathbf{u}}(m+1)$ , 但每一时刻只实施当前的控制量.

### 3 基于递归提升和 Q-学习方法的运行层控制器设计

#### 3.1 基于递归提升的运行层多速率处理方法

如式 (13) 所示, 基础回路层闭环控制周期为  $T_0$ , 而运行层的控制周期为  $T$ , 通常情况下  $T > T_0$ , 此时运行层控制器的设计仍然是一个多速率控制问题, 因此首先需要将系统  $(\hat{A}, \hat{B}, \hat{C}, \hat{D})$  提升到慢时间尺度  $T$ , 通过递推容易得到:

$$\begin{aligned} \mathbf{x}_f((k+1)T) &= \mathbf{x}_f(kT + \xi T_0) = \\ &\hat{A} \mathbf{x}_f(kT + (\xi - 1)T_0) + \hat{B} \mathbf{w}(kT) = \\ &\hat{A}(\hat{A} \mathbf{x}_f(kT + (\xi - 2)T_0) + \hat{B} \mathbf{w}(kT)) + \\ &\hat{B} \mathbf{w}(kT) = \hat{A}^\xi \mathbf{x}_f(kT) + \sum_{i=0}^{\xi-1} \hat{A}^i \hat{B} \mathbf{w}(kT) \end{aligned} \quad (14)$$

则通过递归提升后的系统可以表示为

$$\begin{cases} \mathbf{x}_f(k+1) = \tilde{A} \mathbf{x}_f(k) + \tilde{B} \mathbf{w}(k) \\ \bar{\mathbf{y}}(k) = \tilde{C} \mathbf{x}_f(k) + \tilde{D} \mathbf{w}(k) \end{cases} \quad (15)$$

其中,  $\tilde{A} = \hat{A}^\xi$ ,  $\tilde{B} = \sum_{i=0}^{\xi-1} \hat{A}^i \hat{B}$ ,  $\tilde{C} = \hat{C}$ ,  $\tilde{D} = \hat{D}$ .

式 (15) 即是基础回路层在慢时间尺度  $T$  下的闭环模型, 可与运行过程模型相结合构成运行层的广义被控对象.

考虑如下连续时间运行过程模型:

$$\begin{cases} \dot{\mathbf{x}}_s(t) = A_{c'} \mathbf{x}_s(t) + B_{c'} \mathbf{y}(t) \\ \mathbf{r}(t) = C_{c'} \mathbf{x}_s(t) + D_{c'} \mathbf{y}(t) \end{cases}, \quad D_{c'} = 0 \quad (16)$$

其中,  $\mathbf{x}_s \in \mathbf{R}^{n_{x'}}$ ,  $\mathbf{r} \in \mathbf{R}^{n_r}$  分别为运行过程的状态和运行指标变量,  $n_{x'}$  和  $n_r$  分别为状态和运行指标变量的维数;  $A_{c'}$ ,  $B_{c'}$  和  $C_{c'}$  为运行过程模型的系统矩阵.

对式 (16) 按周期  $T$  进行离散化可得:

$$\begin{cases} \mathbf{x}_s(k+1) = A_s \mathbf{x}_s(k) + B_s \mathbf{y}(k) \\ \mathbf{r}(k) = C_s \mathbf{x}_s(k) + D_s \mathbf{y}(k) \end{cases} \quad (17)$$

其中,  $A_s = e^{A_{c'} T}$ ,  $B_s = \int_{kT}^{(k+1)T} e^{[(k+1)T-\tau]} d\tau B_{c'}$  =  $\int_0^T e^{A_{c'} t} dt B_{c'}$ ,  $C_s = C_{c'}$ ,  $D_s = D_{c'}$ .

易知  $\bar{\mathbf{y}}(k)$  和  $\mathbf{y}(k)$  存在如下关系:

$$\mathbf{y}(k) = \beta \bar{\mathbf{y}}(k) \quad (18)$$

$$\text{其中, } \beta = \begin{bmatrix} I_{n_y \times n_y} & \underbrace{0_{n_y \times n_y} \cdots 0_{n_y \times n_y}}_{p-1} \end{bmatrix}$$

结合式 (15)、式 (17) 两式联立得到如下增广矩阵:

$$\mathbf{x}(k+1) = \begin{bmatrix} \mathbf{x}_f(k+1) \\ \mathbf{x}_s(k+1) \end{bmatrix} := A \mathbf{x}(k) + B \mathbf{w}(k) \quad (19)$$

$$\text{其中, } A = \begin{bmatrix} \tilde{A} & 0 \\ B_s \beta \tilde{C} & A_s \end{bmatrix}, \quad B = \begin{bmatrix} \tilde{B} \\ B_s \beta \tilde{D} \end{bmatrix}$$

由此可得, 统一时间尺度后的运行层广义被控对象为

$$\begin{cases} \mathbf{x}(k+1) = A \mathbf{x}(k) + B \mathbf{w}(k) \\ \mathbf{r}(k) = C \mathbf{x}(k) + D \mathbf{w}(k) \end{cases} \quad (20)$$

其中,  $C = \begin{bmatrix} D_s \beta \tilde{C} & C_s \end{bmatrix}$ ,  $D = D_s \beta \tilde{D}$ .

### 3.2 运行层设定值优化问题

为了公式的清晰表达, 在下文中多处将  $\mathbf{r}(k)$ ,  $\mathbf{r}^*(k)$ ,  $\mathbf{w}(k)$ ,  $\mathbf{w}^*(k)$ ,  $\mathbf{x}(k)$  和  $\mathbf{X}(k)$  表示为  $\mathbf{r}_k$ ,  $\mathbf{r}_k^*$ ,  $\mathbf{w}_k$ ,  $\mathbf{w}_k^*$ ,  $\mathbf{x}_k$  和  $\mathbf{X}_k$ , 两种表示方法同义.

为实现运行指标跟踪理想运行指标  $\mathbf{r}^*$ , 定义如下性能指标:

$$J = \frac{1}{2} \sum_{i=k}^{\infty} ((\mathbf{r}_i - \mathbf{r}_i^*)^T Q (\mathbf{r}_i - \mathbf{r}_i^*) + \mathbf{w}_i^T R \mathbf{w}_i) \quad (21)$$

其中,  $Q$  和  $R$  分别为半正定矩阵和正定矩阵; 运行指标的期望轨迹采用类似文献 [26] 的处理方法, 即

$$\mathbf{r}_{k+1}^* = F \mathbf{r}_k^* \quad (22)$$

**注 4.** 对于跟踪问题, 性能指标的前一项为了使跟踪误差足够小, 保证实际运行指标能够跟踪期望值; 后一项的引入是为了限定控制输入的变化程度, 避免设定值过大.

**注 5.** 为保证优化控制的可实现性, 还需修定性能指标, 这是因为性能指标函数中包含回路设定值部分, 如果当  $k \rightarrow \infty$ ,  $\mathbf{r}_k^*$  不收敛于 0, 而设定值  $\mathbf{w}(k)$  由于依赖系统状态  $\mathbf{x}(k)$  和理想运行指标  $\mathbf{r}^*(k)$ , 也不收敛于 0, 性能指标 (21) 可能无界. 实际工业过程中, 运行指标表征工业生产中间过程的产品质量、能耗与物耗等, 无法满足  $k \rightarrow \infty$  时  $\mathbf{r}_k^*$  收敛到 0 的条件, 为此, 需要在性能指标函数中引入折扣因子  $\gamma$  ( $0 < \gamma < 1$ ) 避免对  $\mathbf{r}_k^*$  收敛到 0 的限制, 使得只要系统输入输出有界, 则  $J$  有界.

结合式 (20) 和 (21), 运行层的设定值优化转化为求解如下带有折扣因子的线性二次跟踪 (Linear quadratic tracking, LQT) 问题.

$$\begin{cases} \min_{\mathbf{w}_k} \frac{1}{2} \sum_{i=k}^{\infty} \gamma^{i-k} ((\mathbf{r}_i - \mathbf{r}_i^*)^T Q (\mathbf{r}_i - \mathbf{r}_i^*) + \mathbf{w}_i^T R \mathbf{w}_i) \\ \text{s. t. 式 (20) 和 (22)} \end{cases} \quad (23)$$

当广义被控对象模型即式 (20) 已知时, 可通过引入贝尔曼方程及哈密尔顿函数求解上述带有折扣因子的 LQT 问题. 但对于工业过程, 其运行过程机理复杂难以获得模型 ( $A_s, B_s, C_s, D_s$ ), 因此, 式 (20) 含有部分未知模型, 导致难以采用传统 LQT 求解方法. 下文首先将求解 LQT 问题转化为线性二次调节 (Linear quadratic regular, LQR) 问题. 进而引入贝尔曼方程及哈密尔顿函数进行求解, 阐明该方法对系统全动态模型的依赖, 继而采用增强学习, 设计一种基于 Q-学习的数据驱动方法.

### 3.3 基于模型的运行层设定值优化

首先引入增广向量  $\mathbf{X}(k) = \begin{bmatrix} \mathbf{x}(k) \\ \mathbf{r}^*(k) \end{bmatrix}$ , 增广后的广义被控对象的状态方程为:

$$\mathbf{X}(k+1) = \begin{bmatrix} \mathbf{x}(k+1) \\ \mathbf{r}^*(k+1) \end{bmatrix} = \begin{bmatrix} A & 0 \\ 0 & F \end{bmatrix} \begin{bmatrix} \mathbf{x}(k) \\ \mathbf{r}^*(k) \end{bmatrix} + \begin{bmatrix} B \\ 0 \end{bmatrix} \mathbf{w}(k) \equiv T \mathbf{X}(k) + B_1 \mathbf{w}(k) \quad (24)$$

此时, 设计系统的设定值为  $\mathbf{w}_k = K_x \mathbf{x}_k + K_r \mathbf{r}_k^*$ , 将其代入到  $\mathbf{r}_k = C \mathbf{x}_k + D \mathbf{w}_k$  中, 则性能指标可以表示为如下值函数:

$$V(\mathbf{x}_k, \mathbf{r}_k^*) = \frac{1}{2} \sum_{i=k}^{\infty} \gamma^{i-k} (\mathbf{X}_i^T Q_1 \mathbf{X}_i + \mathbf{w}_i^T R \mathbf{w}_i) \quad (25)$$

其中,  $Q_1 = C_1^T Q C_1$ ,  $C_1 = \begin{bmatrix} C + D K_x & D K_r - I \end{bmatrix}$ .

则有折扣因子的 LQT 问题 (23) 可重写为如下 LQR:

$$\begin{cases} \min_{\mathbf{w}_k} \frac{1}{2} \sum_{i=k}^{\infty} \gamma^{i-k} (\mathbf{X}_i^T Q_1 \mathbf{X}_i + \mathbf{w}_i^T R \mathbf{w}_i) \\ \text{s. t. (24)} \end{cases} \quad (26)$$

进一步将  $\mathbf{w}_k = K_x \mathbf{x}_k + K_r \mathbf{r}_k^*$  代入式 (25) 中, 可得:

$$\begin{aligned} V(\mathbf{x}_k, \mathbf{r}_k^*) &= \frac{1}{2} \sum_{i=k}^{\infty} \gamma^{i-k} ((\varepsilon \mathbf{x}_i + \delta \mathbf{r}_i^*)^T Q (\varepsilon \mathbf{x}_i + \delta \mathbf{r}_i^*) + \\ & (K_x \mathbf{x}_i + K_r \mathbf{r}_i^*)^T R (K_x \mathbf{x}_i + K_r \mathbf{r}_i^*)) = \\ & \frac{1}{2} \sum_{i=0}^{\infty} \gamma^i (\mathbf{x}_{i+k}^T (\varepsilon^T Q \varepsilon + K_x^T R K_x) \mathbf{x}_{i+k} + \\ & \mathbf{x}_{i+k}^T (\varepsilon^T Q \delta + K_x^T R K_r) \mathbf{r}_{i+k}^* + \\ & (\mathbf{r}_{i+k}^*)^T (\delta^T Q \varepsilon + K_r^T R K_x) \mathbf{x}_{i+k} + \\ & (\mathbf{r}_{i+k}^*)^T (\delta^T Q \delta + K_r^T R K_r) \mathbf{r}_{i+k}^*) \end{aligned} \quad (27)$$

其中,  $\varepsilon = C + D K_x$ ,  $\delta = D K_r - I$ , 根据式 (20) 和 (22) 可得:

$$\mathbf{r}_{i+k}^* = F^i \mathbf{r}_k^* \quad (28)$$

$$\mathbf{x}_{i+k} = G^i \mathbf{x}_k + M \mathbf{r}_k^* \quad (29)$$

其中,  $G = A + B K_x$ ,  $M = \sum_{n=0}^{i-1} G^{i-n-1} B K_r F^n$ , 将以上两式代入式 (27) 中可得:

$$\begin{aligned} V(\mathbf{x}_k, \mathbf{r}_k^*) &= \frac{1}{2} \mathbf{x}_k^T P_{11} \mathbf{x}_k + \frac{1}{2} \mathbf{x}_k^T P_{12} \mathbf{r}_k^* + \\ & \frac{1}{2} (\mathbf{r}_k^*)^T P_{21} \mathbf{x}_k + \frac{1}{2} (\mathbf{r}_k^*)^T P_{22} \mathbf{r}_k^* \end{aligned} \quad (30)$$

其中

$$\begin{aligned}
 P_{11} &= \sum_{i=0}^{\infty} \gamma^i ((G^i)^T (\varepsilon^T Q \varepsilon + K_x^T R K_x) G^i) \\
 P_{12} &= \sum_{i=0}^{\infty} \gamma^i ((G^i)^T (\varepsilon^T Q \delta + K_x^T R K_r) F^i) + \\
 &\quad (G^i)^T (\varepsilon^T Q \varepsilon + K_x^T R K_x) M \\
 P_{21} &= \sum_{i=0}^{\infty} \gamma^i ((F^i)^T (\delta^T Q \varepsilon + K_r^T R K_x) G^i) + \\
 &\quad M^T (\varepsilon^T Q \varepsilon + K_x^T R K_x) G^i \\
 P_{22} &= \sum_{i=0}^{\infty} \gamma^i (M^T (\varepsilon^T Q \delta + K_x^T R K_r) F^i + M^T \\
 &\quad (\varepsilon^T Q \varepsilon + K_x^T R K_x) M + (F^i)^T (\delta^T Q \varepsilon + \\
 &\quad K_r^T R K_x) M + (F^i)^T (Q + K_r^T R K_r) F^i)
 \end{aligned}$$

则值函数表示为

$$V(\mathbf{x}_k, \mathbf{r}_k^*) = V(\mathbf{X}_k) = \frac{1}{2} \mathbf{X}_k^T P \mathbf{X}_k \quad (31)$$

$$\text{其中, } P = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix}.$$

由式 (25) 和 (31), 可得如下贝尔曼方程:

$$V(\mathbf{X}_k) = \frac{1}{2} (\mathbf{X}_k^T Q_1 \mathbf{X}_k + \mathbf{w}_k^T R \mathbf{w}_k) + \gamma V(\mathbf{X}_{k+1}, \mathbf{w}_{k+1}) \quad (32)$$

将式 (31) 代入式 (32), 有

$$\mathbf{X}_k^T P \mathbf{X}_k = \mathbf{X}_k^T Q_1 \mathbf{X}_k + \mathbf{w}_k^T R \mathbf{w}_k + \gamma \mathbf{X}_{k+1}^T P \mathbf{X}_{k+1} \quad (33)$$

由此, 得到如下 LQR 问题的哈密顿函数:

$$H(\mathbf{X}_k, \mathbf{w}_k) = \mathbf{X}_k^T Q_1 \mathbf{X}_k + \mathbf{w}_k^T R \mathbf{w}_k + \gamma \mathbf{X}_{k+1}^T P \mathbf{X}_{k+1} - \mathbf{X}_k^T P \mathbf{X}_k \quad (34)$$

根据文献 [27-28], 求解 LQR 问题需满足最优性的必要条件:

$$\begin{aligned}
 \frac{\partial H(\mathbf{X}_k, \mathbf{w}_k)}{\partial \mathbf{w}_k} &= 2R\mathbf{w}_k + \\
 \gamma \left( \frac{\partial \mathbf{X}_{k+1}}{\partial \mathbf{w}_k} \right)^T \frac{\partial V(\mathbf{X}_{k+1})}{\partial \mathbf{X}_{k+1}} &= \\
 2R\mathbf{w}_k + 2\gamma B_1^T P \mathbf{X}_{k+1} &= 0 \quad (35)
 \end{aligned}$$

将式 (24) 代入上式, 得到:

$$\mathbf{w}_k^* = K^* \mathbf{X}_k \quad (36)$$

其中,  $K^* = -(R + \gamma B_1^T P B_1)^{-1} \gamma B_1^T P T$ , 矩阵  $P$  可由式 (24) 和 (35) 代入贝尔曼方程 (33) 中离线计算

获得, 即求解如下的黎卡提方程 (Algebraic riccati equation, ARE).

$$\begin{aligned}
 Q_1 - P + \gamma T^T P T - \gamma^2 T^T P B_1 \times \\
 (R + \gamma B_1^T P B_1)^{-1} B_1^T P T = 0 \quad (37)
 \end{aligned}$$

也可采用在线策略迭代算法逐步求解矩阵  $P$  和最优控制率  $K^*$ , 具体算法如下:

**算法 1.** 基于贝尔曼方程的在线策略迭代算法

**步骤 1.** 策略评估 (用贝尔曼方程计算  $P^{j+1}$ ,  $j = 1, 2, \dots$ );

$$\begin{aligned}
 \mathbf{X}_k^T P^{j+1} \mathbf{X}_k = \mathbf{X}_k^T (Q_1 + (K^j)^T R K^j) \mathbf{X}_k + \\
 \gamma \mathbf{X}_{k+1}^T P^{j+1} \mathbf{X}_{k+1} \quad (38)
 \end{aligned}$$

**步骤 2.** 策略提升

$$K^{j+1} = -(R + \gamma B_1^T P^{j+1} B_1)^{-1} \gamma B_1^T P^{j+1} T \quad (39)$$

在线策略迭代算法在初始值  $K^1$  能镇定系统的情况下, 可通过迭代 LQR 贝尔曼方程 (38), 采用最小二乘法 (Least squares, LS) 或其递推算法 (Recursive least squares, RLS) 计算正定矩阵  $P^{j+1}$ , 并改进设定值优化策略  $\mathbf{w}_k^{j+1} = K^{j+1} \mathbf{X}_k$ , 直到收敛, 从而获得最优的设定值.

从上述计算过程可以看出, 无论是采用 ARE 离线求解还是采用在线策略迭代算法, 均依赖系统的全部动态信息 ( $T, B_1$ ), 对于无法建立运行过程的工业系统, 两种算法均难以使用. 因此, 在下节中将利用 Q-学习来求解带有折扣因子的 LQR 问题, 实现设定值优化.

### 3.4 基于 Q-学习的数据驱动运行层设定值优化

基于 Q-学习的设定值优化方法是一种在线迭代算法, 其不需要系统动态信息 ( $T, B_1$ ), 具体算法如下:

根据贝尔曼方程 (33), 定义 Q-函数为

$$\begin{aligned}
 Q(\mathbf{X}_k, \mathbf{w}_k) = \frac{1}{2} \mathbf{X}_k^T Q_1 \mathbf{X}_k + \frac{1}{2} \mathbf{w}_k^T R \mathbf{w}_k + \\
 \frac{1}{2} \gamma \mathbf{X}_{k+1}^T P \mathbf{X}_{k+1} \quad (40)
 \end{aligned}$$

由增广系统 (24) 可改写式 (40) 为

$$\begin{aligned}
 Q(\mathbf{X}_k, \mathbf{w}_k) = \frac{1}{2} \mathbf{X}_k^T Q_1 \mathbf{X}_k + \\
 \frac{1}{2} \mathbf{w}_k^T R \mathbf{w}_k + \frac{1}{2} \gamma (T \mathbf{X}_k + B_1 \mathbf{w}_k)^T \times \\
 P (T \mathbf{X}_k + B_1 \mathbf{w}_k) = \frac{1}{2} \begin{bmatrix} \mathbf{X}_k \\ \mathbf{w}_k \end{bmatrix}^T \times \\
 \begin{bmatrix} Q_1 + \gamma T^T P T & \gamma T^T P B_1 \\ \gamma B_1^T P T & R + \gamma B_1^T P B_1 \end{bmatrix} \begin{bmatrix} \mathbf{X}_k \\ \mathbf{w}_k \end{bmatrix} \quad (41)
 \end{aligned}$$

因此, 定义:

$$Q(\mathbf{X}_k, \mathbf{w}_k) = \frac{1}{2} \begin{bmatrix} \mathbf{X}_k \\ \mathbf{w}_k \end{bmatrix}^T \begin{bmatrix} H_{XX} & H_{Xw} \\ H_{wX} & H_{ww} \end{bmatrix} \begin{bmatrix} \mathbf{X}_k \\ \mathbf{w}_k \end{bmatrix} = \frac{1}{2} \begin{bmatrix} \mathbf{X}_k \\ \mathbf{w}_k \end{bmatrix}^T H \begin{bmatrix} \mathbf{X}_k \\ \mathbf{w}_k \end{bmatrix} \quad (42)$$

求解最优设定值, 需对式 (42) 求极值, 即使  $\frac{\partial Q(\mathbf{X}_k, \mathbf{w}_k)}{\partial \mathbf{w}_k} = 0$ , 有:

$$\mathbf{w}_k^* = -H_{ww}^{-1} H_{wX} \mathbf{X}_k \quad (43)$$

由此可以看出最优设定值的求解只与矩阵  $H$  相关, 类似第 3.3 节中算法 1 的在线策略迭代算法, 本文通过在线采集的数据  $\mathbf{X}_k, \mathbf{Z}_k, \mathbf{Z}_{k+1}$  对矩阵  $H$  进行逐步估计, 从而求得最优设定值  $\mathbf{w}^*$ . 具体算法如下.

首先定义:  $\mathbf{Z}_k = \begin{bmatrix} \mathbf{X}_k^T & \mathbf{w}_k^T \end{bmatrix}^T$ , 代入式 (42) 可将其改写为

$$Q(\mathbf{X}_k, \mathbf{w}_k) = \frac{1}{2} \mathbf{Z}_k^T H \mathbf{Z}_k \quad (44)$$

由于 Q-函数满足如下贝尔曼方程:

$$Q(\mathbf{X}_k, \mathbf{w}_k) = \frac{1}{2} \mathbf{X}_k^T Q_1 \mathbf{X}_k + \frac{1}{2} \mathbf{w}_k^T R \mathbf{w}_k + \gamma Q(\mathbf{X}_{k+1}, \mathbf{w}_{k+1}) \quad (45)$$

将上式 (44) 代入式 (45) 可得:

$$\mathbf{Z}_k^T H \mathbf{Z}_k = \mathbf{X}_k^T Q_1 \mathbf{X}_k + \mathbf{w}_k^T R \mathbf{w}_k + \gamma \mathbf{Z}_{k+1}^T H \mathbf{Z}_{k+1} \quad (46)$$

根据式 (43) 和 (46), 可采用算法 2 所示的在线策略迭代算法逐步求解矩阵  $H$  和最优设定值  $\mathbf{w}^*$ .

**算法 2. 基于 Q-函数的策略迭代算法**

**步骤 1. 策略评估**

$$\mathbf{Z}_k^T H^{j+1} \mathbf{Z}_k = \mathbf{X}_k^T Q_1 \mathbf{X}_k + (\mathbf{w}_k^j)^T R \mathbf{w}_k^j + \gamma \mathbf{Z}_{k+1}^T H^{j+1} \mathbf{Z}_{k+1} \quad (47)$$

**步骤 2. 策略提升**

$$\mathbf{w}_k^{j+1} = -(H_{ww}^{j+1})^{-1} H_{wX}^{j+1} \mathbf{X}_k \quad (48)$$

在每一次策略评估时, 矩阵  $H^{j+1}$  可利用数据  $\mathbf{X}_k, \mathbf{Z}_k, \mathbf{Z}_{k+1}$ , 采用 LS 或 RLS 获得; 此后将  $H^{j+1}$  用于策略提升, 更新后的设定值  $\mathbf{w}_k^{j+1}$  在基础回路层 MPC 控制器的作用下, 产生新的工业过程数据, 再次用于策略评估, 如此循环, 当算法收敛时, 通过式 (48) 即可求得最优设定值  $\mathbf{w}^*$ . 从上述算法过程可以

看出, 基于 Q-学习的设定值优化方法无需任何系统动态先验知识, 是一种无模型的数据驱动方法.

**注 6.** 为保证式 (47) 中  $H^{j+1}$  能够在 LS 或 RLS 算法下被准确估计, 要求数据充分, 因此在执行策略迭代时需要加入持续激励条件. 实际上, 工业过程在运行时不可避免会受到测量扰动的影响, 因此, 一定程度上保证算法所需的充分持续激励条件, 从而使得  $H^{j+1}$  能够被准确估计. 从强化学习的角度看, 干扰可以帮助算法更有效地探索周围的环境, 增加选择最优值的几率.

**注 7.** 由式 (47) 可以看出,  $H^{j+1}$  是控制参数, 本文直接采用 LS 或 RLS 算法对  $H^{j+1}$  进行估计, 从而计算控制输出, 不需要辨识系统模型. 因此, 从自适应控制的角度看, 是一种直接自适应控制方法, 对于具有不确定性和时变的被控对象具有一定的自学习和自适应能力.

**3.5 数据驱动运行层设定值优化算法的收敛性分析**

由于  $V(\mathbf{X}_k)$  和  $Q(\mathbf{X}_k, \mathbf{w}_k)$  在数学公式上等价, 结合式 (31) 和 (41) 可以得到:

$$Q(\mathbf{X}_k, \mathbf{w}_k) = \frac{1}{2} \mathbf{X}_k^T P \mathbf{X}_k = \frac{1}{2} \mathbf{Z}_k^T H \mathbf{Z}_k \quad (49)$$

由此可以看出 Q-函数和  $H$  矩阵的同趋性, 因此, 当 Q-函数随着  $j \rightarrow \infty$  而趋于最优值时,  $H^j$  和  $K^j$  都会趋于最优值, 本节即通过这个思想来证明算法的收敛性.

在 LS 算法充分求解, 保证式 (47) 中  $H^{j+1}$  能被准确估计, 且在初始控制策略保证运行指标稳定的前提下, 可得引理 1.

**引理 1.**

$$H^{j+1} = \begin{bmatrix} Q_1 & 0 \\ 0 & R \end{bmatrix} + \gamma \begin{bmatrix} T & B_1 \\ K^j T & K^j B_1 \end{bmatrix}^T \times H^{j+1} \begin{bmatrix} T & B_1 \\ K^j T & K^j B_1 \end{bmatrix} \quad (50)$$

**证明.** 由  $\mathbf{w}_k^j = K^j \mathbf{X}_k$  和式 (47) 可得:

$$\begin{aligned} \mathbf{Z}_k^T H^{j+1} \mathbf{Z}_k &= \begin{bmatrix} \mathbf{X}_k \\ \mathbf{w}_k \end{bmatrix}^T \begin{bmatrix} Q_1 & 0 \\ 0 & R \end{bmatrix} \begin{bmatrix} \mathbf{X}_k \\ \mathbf{w}_k \end{bmatrix} + \gamma \begin{bmatrix} \mathbf{X}_k \\ \mathbf{w}_k \end{bmatrix}^T \times \\ &\begin{bmatrix} T & B_1 \\ K^j T & K^j B_1 \end{bmatrix} H^{j+1} \begin{bmatrix} T & B_1 \\ K^j T & K^j B_1 \end{bmatrix} \times \\ &\begin{bmatrix} \mathbf{X}_k \\ \mathbf{w}_k \end{bmatrix} = \mathbf{Z}_k^T \begin{bmatrix} Q_1 & 0 \\ 0 & R \end{bmatrix} \mathbf{Z}_k + \gamma \mathbf{Z}_k^T \times \end{aligned}$$



$$\begin{bmatrix} T & B_1 \\ K^j T & K^j B_1 \end{bmatrix} H^{j+1} \begin{bmatrix} T & B_1 \\ K^j T & K^j B_1 \end{bmatrix} \mathbf{Z}_k \quad (51)$$

由此易得式 (50).  $\square$

引理 2.

$$H^{j+1} = \begin{bmatrix} Q_1 + \gamma T^T P^{j+1} T & \gamma T^T P^{j+1} B_1 \\ \gamma B_1^T P^{j+1} T & R + \gamma B_1^T P^{j+1} B_1 \end{bmatrix} \quad (52)$$

$K^{j+1}$  和  $P^{j+1}$  的关系如式 (39) 所示, 其中,

$$P^{j+1} = \frac{1}{\gamma} \begin{bmatrix} I & (K^j)^T \end{bmatrix} H^{j+1} \begin{bmatrix} I & (K^j)^T \end{bmatrix}^T \quad (53)$$

证明. 式 (52) 可直接由式 (41) 得到, 根据式 (49) 可得:

$$\gamma \mathbf{X}_k^T P^{j+1} \mathbf{X}_k = \begin{bmatrix} \mathbf{X}_k \\ \mathbf{w}_k \end{bmatrix}^T H^{j+1} \begin{bmatrix} \mathbf{X}_k \\ \mathbf{w}_k \end{bmatrix} \quad (54)$$

结合  $\mathbf{w}_k^j = K^j \mathbf{X}_k$ , 进一步可得:

$$\gamma \mathbf{X}_k^T P^{j+1} \mathbf{X}_k = \mathbf{X}_k^T \begin{bmatrix} I & (K^j)^T \\ I & (K^j)^T \end{bmatrix} H^{j+1} \times \begin{bmatrix} I & (K^j)^T \end{bmatrix}^T \mathbf{X}_k \quad (55)$$

由此易得式 (53).  $\square$

引理 3.  $P^{j+1}$  满足如下迭代过程:

$$\begin{aligned} \gamma P^{j+1} &= Q_1 + (K^j)^T R K^{j+1} + \\ &\gamma (T^T + (K^j)^T B_1^T) P^{j+1} (T + B_1 K^j) \end{aligned} \quad (56)$$

$p^{j+1}$  和  $H^{j+1}$  是同步且等价的.

证明. 将引理 2 中的式 (52) 代入式 (53) 可得:

$$P^{j+1} = \frac{1}{\gamma} \begin{bmatrix} I \\ K^j \end{bmatrix}^T \begin{bmatrix} Q_1 + \gamma T^T P^{j+1} T \\ \gamma B_1^T P^{j+1} T \\ \gamma T^T P^{j+1} B_1 \\ R + \gamma B_1^T P^{j+1} B_1 \end{bmatrix} \begin{bmatrix} I \\ K^j \end{bmatrix} \quad (57)$$

展开后经过整理即可得到式 (56).  $\square$

定理 1. 假设 LQR 问题在状态反馈信息下是可解的且具有一个值结构. 那么, 初始值  $H^0$ 、 $K^0$  在稳定范围内时, 通过定理 1 中式 (50) 的迭代,  $H^j$  会最终趋近于  $H^*$ . 其中  $H^*$  为 Q-函数迭代至最优值  $Q^*(\mathbf{X}_k, \mathbf{w}_k)$  时的取值, 此时  $P^*$  的值对应式 (37) 所能求出的理想解.

证明. 已有文献已证明广义 ARE 通过迭代可以保证初始值  $P^0 = 0$  时,  $P$  收敛于  $P^*$ [29], 从而说明本文 ARE 可通过式 (55) 的迭代使得  $P^j$  收敛于  $P^*$ . 又引理 2 和 3 说明了  $P^j$  和  $H^j$  是同步变化的, 所以  $H^j$  最终也会趋近于期望值  $H^*$ , 即

$$H^j \rightarrow \begin{bmatrix} Q_1 + \gamma T^T P^* T & \gamma T^T P^* B_1 \\ \gamma B_1^T P^* T & R + \gamma B_1^T P^* B_1 \end{bmatrix} \quad (58)$$

上述算法的收敛性分析是在 LS 可充分求解的前提下讨论的, 对于采用 RLS 算法下的控制器收敛性分析, 详见文献 [30].  $\square$

## 4 仿真实验

为了验证本文方法的有效性, 本文以典型工业闭路磨矿过程为对象进行仿真实验研究.

### 4.1 闭路磨矿工艺流程

磨矿过程是对矿石经过物理的研磨、分级处理, 将颗粒由大变小, 从而将有用矿物从脉石中分离出来. 本文选取如图 2 所示的典型磨机过程, 其由电振给矿机、球磨机和水力旋流器等设备组成. 磨矿生产过程中, 原矿首先和一定比例的水被输送至球磨机, 球磨机通过自身旋转带动机内钢球对矿石进行研磨. 研磨后的矿石随矿浆流动从球磨机出口处排出流入泵池, 矿浆经稀释后由底流泵打入水力旋流器进行粒度分级, 形成含有细颗粒物的溢流矿浆和含有粗颗粒物的底流矿浆. 底流矿浆返回至球磨机再进行研磨, 形成循环负荷. 由于粒度过粗或过细都不利于有用矿粒的选别, 因此溢流矿浆的产品粒度  $\mathbf{r}_1$  是关键的运行指标之一; 此外, 循环负荷  $\mathbf{r}_2$  反映了磨矿能耗水平, 是体现磨矿运行效率的重要运行指标. 磨矿过程控制的目标即是实现产品粒度  $\mathbf{r}_1$  和循环负荷  $\mathbf{r}_2$  的优化控制.

由于产品粒度  $\mathbf{r}_1$  和循环负荷  $\mathbf{r}_2$  与磨机给矿量和泵池补加水密切相关, 因此, 磨矿系统通常设置磨机给矿和泵池补加水两个基础控制回路, 运行过程以磨机给矿量和泵池补加水量为输入, 以运行指标为输出. 由于给矿和给水的动态过程较快, 而矿石研磨需要较长的时间, 因此, 磨矿过程的回路过程和运行过程具有不同时间尺度特性, 是一个典型的多层次、多时间尺度的控制问题. 为了验证本文所提方法, 将电振给矿机频率  $\mathbf{u}_1$  和泵池补水阀门开度  $\mathbf{u}_2$  两个基础控制回路的控制周期  $T_1$  设置为 2s, 磨机给矿量  $\mathbf{y}_1$ 、和泵池补水流量  $\mathbf{y}_2$  的采样频率  $T_2$  设置为 3s, 运行层周期  $T$  为 60s, 即  $\xi = 10$ . 根据本文所提的多速率分层运行优化控制方法, 可以得到基周期  $h$  为 1s, 框架周期即基础回路层控制周期  $T_0$  为 6s.

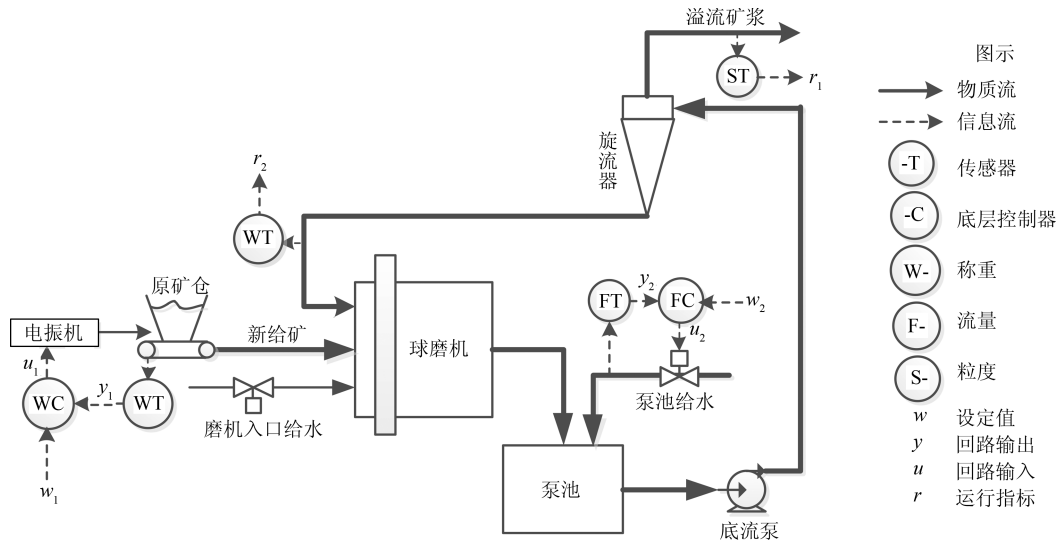


图 2 闭路磨矿过程工艺流程图

Fig. 2 Flow chart of closed-circuit mineral grinding process

根据文献 [31], 结合实际磨矿过程, 两层动态模型分别采用式 (49) 和式 (50) 来近似模拟.

$$\begin{bmatrix} y_1(s) \\ y_2(s) \end{bmatrix} = \begin{bmatrix} \frac{0.1457}{s + 0.0654} & 0 \\ 0 & \frac{0.1807}{s + 0.0823} \end{bmatrix} \begin{bmatrix} u_1(s) \\ u_2(s) \end{bmatrix} \quad (59)$$

$$\begin{bmatrix} r_1(s) \\ r_2(s) \end{bmatrix} = \begin{bmatrix} \frac{-0.4}{3.5s + 1} & \frac{3.42(1 - 0.994)}{(s + 1)(2s + 1)} \\ \frac{0.8}{3.2s + 1} & \frac{2.5}{5.5s + 1} \end{bmatrix} \begin{bmatrix} y_1(s) \\ y_2(s) \end{bmatrix} \quad (60)$$

$$\begin{cases} \dot{x}_s(t) = \begin{bmatrix} -0.286 & 0 & 0 & 0 \\ 0 & -0.313 & 0 & 0 \\ 0 & 0 & -1.5 & -0.5 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} x_s(t) + \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} y(t) \\ r(t) = \begin{bmatrix} -0.114 & 0.25 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.01 & 0.455 \end{bmatrix} x_s(t) \end{cases} \quad (62)$$

由于本文方法不需要运行过程动态信息, 因此运行过程模型 (52) 只用于被控对象仿真, 其在控制器设计时并未使用.

将上述模型转换为控制器设计所需的状态空间模型, 即

$$\begin{cases} \dot{x}_f(t) = \begin{bmatrix} -0.0654 & 0 \\ 0 & -0.0823 \end{bmatrix} x_f(t) + \begin{bmatrix} 0.1457 & 0 \\ 0 & 0.1807 \end{bmatrix} u(t) \\ y(t) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} x_f(t) \end{cases} \quad (61)$$

### 4.2 仿真实验研究

实验设置期望的磨矿粒度为 70%, 期望的循环负荷为 150 t/h, 即  $r^* = [70, 150]^T$ . 本文方法首先依靠系统运行产生的新数据, 通过 Q-学习给出新的决策量  $w$  作为给矿量和泵池补水量的设定值. 然后, 基础回路层通过 MPC 控制器调整给矿量控制量和泵池补水量控制量, 使给矿量、补水量跟踪运行层所给的设定值  $w$ .

采用本文所提方法的控制效果如图 3~5 所示, 图 3 描绘的是运行层磨矿粒度和循环负荷跟踪其期望值的变化曲线; 图 4 给出了基础回路层给矿量和

泵池补水量跟踪其设定值的变化曲线; 图 5 是基础回路层给矿机电振频率和泵池补加水阀门开度的变化曲线。

从图 3 可以看出, 在运行指标初始状态与期望值较大偏离的情况下, 控制系统在 6 个运行控制周期即 360 s 内快速实现了运行指标  $r$  的优化控制, 并在每一次磨机给矿量和泵池补水流量回路的设定值  $w_1$  和  $w_2$  优化设定后, 其过程变量  $y_1$  和  $y_2$  均能够在远小于运行控制周期的时间内跟踪设定值. 由此可以看出, 本文所提出的工业过程多速率分层运行优化控制方法在基础回路层和运行层均能够获得良好的控制效果. 此外, 在第 900 s 仿真时间处, 使模型参数发生突变, 以模拟实际生产过程中矿石性质(如矿石硬度和矿石粒度)的变化对系统的扰动. 从仿真结果可以看出, 在出现扰动后, 本文方法凭借其自身的自适应能力, 可以使运行指标快速回到其期望值, 具有良好的稳定性.

将所提方法与 PI+MPC 和 PI+PI 两种控制方法, 在相同实验条件下进行对比. 其中 PI+MPC 控制方法在运行层采用 PI 控制器, 控制器参数取  $K_P^H = \begin{bmatrix} -5 & 0 \\ 0 & 1 \end{bmatrix}$ ,  $K_I^H = \begin{bmatrix} -1 & 0 \\ 0 & 0.25 \end{bmatrix}$ , 基础回路层采用本文所提出的基于提升技术和 MPC 的

控制器, 其控制结果如图 6 和图 7 所示. PI+PI 控制方法在运行层和基础回路层均采用 PI 控制方法以单速率设计控制器, 运行层的控制器参数取  $K_P^H = \begin{bmatrix} -5 & 0 \\ 0 & 1 \end{bmatrix}$ ,  $K_I^H = \begin{bmatrix} -1 & 0 \\ 0 & 0.25 \end{bmatrix}$ ; 基础回路层的参数取  $K_P^L = \begin{bmatrix} 5 & 0 \\ 0 & 6 \end{bmatrix}$ ,  $K_I^L = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$ , 其控制结果如图 8 和图 9 所示.

从图 7 和图 9 可以看出, PI+MPC 控制方法由于基础回路层采用了提升技术和 MPC, 基础回路层的跟踪效果好于 PI+PI 控制方法, 这是因为采用提升技术和 MPC 设计的控制器一方面可以很好地解决多速率导致的时变特性, 另一方面能够对系统的耦合进行有效的处理. 由于 PI+MPC 与 PI+PI 两种控制方法在运行层均采用 PI 控制器, 因此对运行指标的控制效果相似, 这可从图 6 和图 8 看出. 虽然两种控制方法均能实现磨矿粒度和循环负荷两个运行指标的跟踪, 但调节时间约 1000 s, 均远远大于本文所提方法下的 360 s. 由此可以看出, 本文所提出的工业过程多速率分层运行优化控制方法具有响应速度快, 且稳定好的特点, 对稳定运行指标, 实现工业过程的提质增效效果显著.

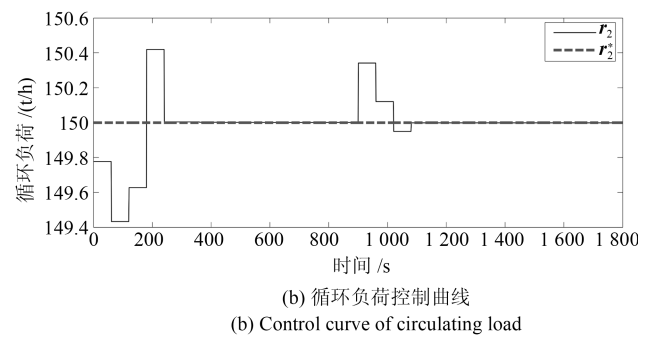
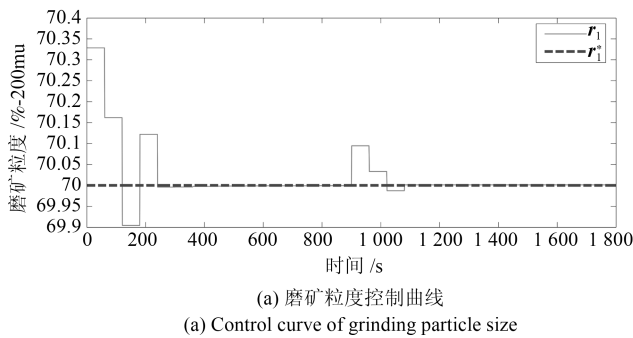


图 3 本文方法下的运行指标控制曲线

Fig. 3 Control curve of operational indices using the proposed method

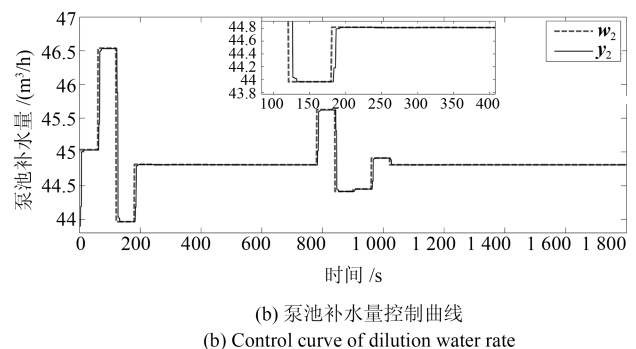
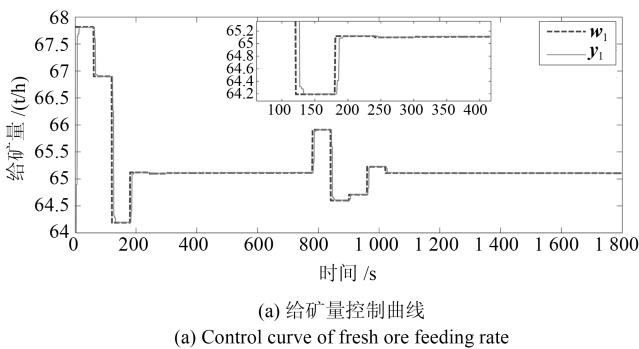


图 4 本文方法下的基础回路层输出曲线

Fig. 4 Output curve of basic loop layer using the proposed method

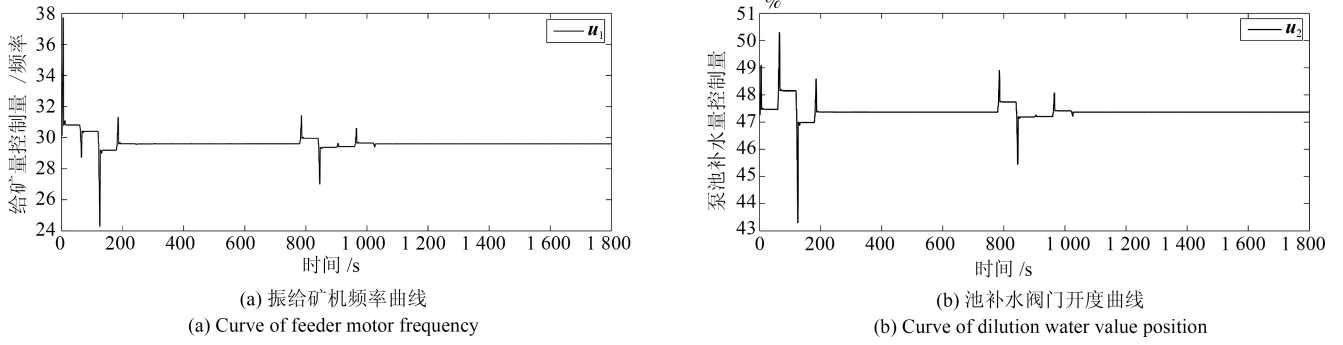


图 5 本文方法下的基础回路层输入曲线

Fig. 5 Input curve of basic loop layer using the proposed method

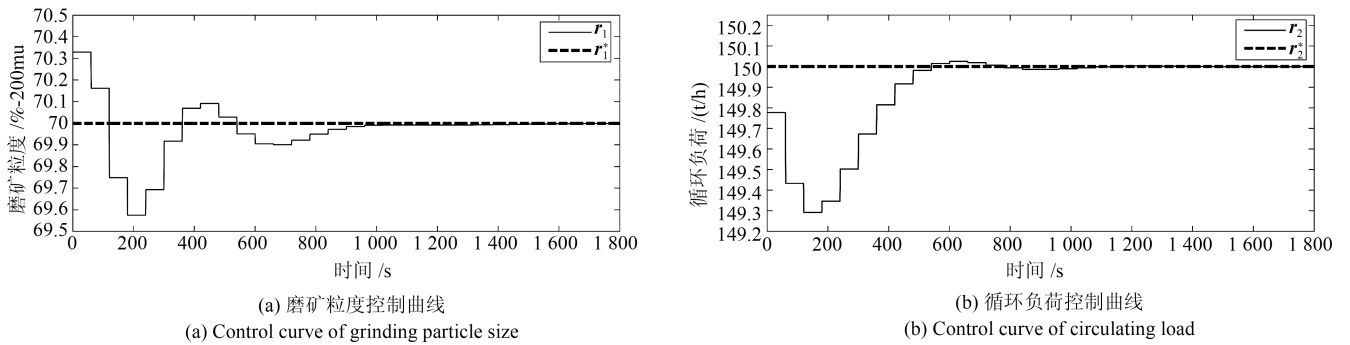


图 6 PI+MPC 方法下的运行指标控制曲线

Fig. 6 Control curve of operational indices using the PI+MPC method

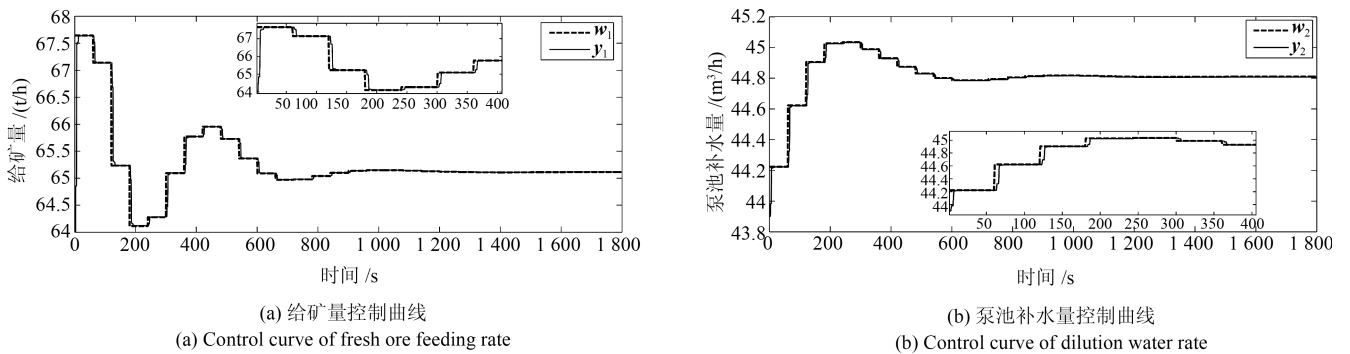


图 7 PI+MPC 方法下的基础回路层输出曲线

Fig. 7 Output curve of basic loop layer using the PI+MPC method

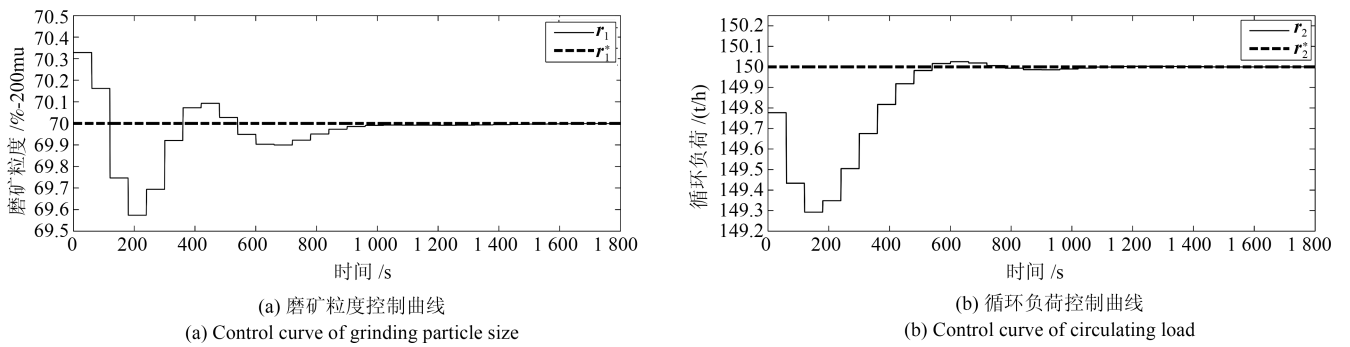


图 8 PI+PI 方法下的运行指标控制曲线

Fig. 8 Control curve of operational indices using the PI+PI method

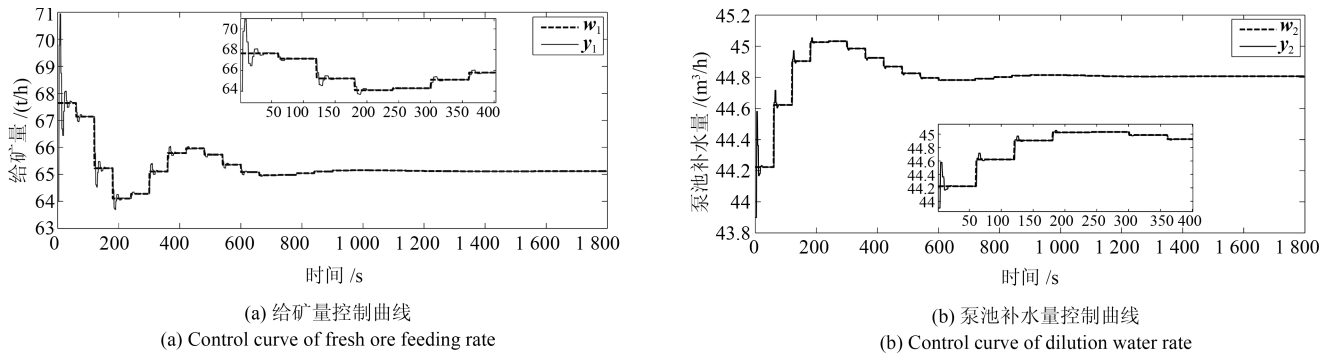


图9 PI+PI方法下的基础回路层输出曲线

Fig. 9 Output curve of basic loop layer using the PI+PI method

## 5 结论

本文针对多层次、多时间尺度、且运行层模型未知的复杂工业过程,创新性的考虑了多速率分层问题,将提升技术、模型预测与增强学习方法相结合,提出一种工业过程多速率分层运行优化控制方法.其通过两种提升方法将多速率分层问题统一到一个时间尺度,采用一种数据驱动的Q-学习算法,求解基础回路最优设定值,同时设计基础回路层的MPC控制器实现设定值的快速跟踪,从而最终实现运行指标对其期望值的跟踪.将本文方法针对一段闭环磨矿过程进行了仿真实验研究,表明了其有效性.本文方法无需采用运行层模型来设计运行层控制器,仅利用数据通过自学习实现设定值的在线优化,对运行层机理复杂难以建立模型的复杂工业过程控制器的设计具有参考价值.

## References

- Chai Tian-You. Operational optimization and feedback control for complex industrial processes. *Acta Automatica Sinica*, 2013, **39**(11): 1744–1757  
(柴天佑. 复杂工业过程运行优化与反馈控制. 自动化学报, 2013, **39**(11): 1744–1757)
- Gui Wei-Hua, Yang Chun-Hua. *Intelligent Modeling, Control And Optimization of Complex Nonferrous Metallurgical Process*. Beijing: Science Press, 2010  
(桂卫华, 阳春华. 复杂有色冶金生产过程智能建模、控制与优化. 北京: 科学出版社, 2010)
- Chai Tian-You. Industrial process control systems: research status and development direction. *Scientia Sinica Informationis*, 2016, **46**: 1003–1015  
(柴天佑. 工业过程控制系统研究现状与发展方向. 中国科学: 信息科学, 2016, **46**: 1003–1015)
- Skogestad S. Plantwide control: the search for the self-optimizing control structure. *Journal of Process Control*, 2000, **10**(5): 487–507
- Ye L J, Cao Y, Yuan X F, Song Z H. Retrofit self-optimizing control: a step forward toward real implementation. *IEEE Transactions on Industrial Electronics*, 2017, **64**(6): 4662–4670
- Jaschke J, Skogestad S. NCO tracking and self-optimizing control in the context of real-time optimization. *Journal of Process Control*, 2011, **21**: 1047–1416
- Darbya M L, Nikolaou M, Jones J, Nicholson D. RTO: an overview and assessment of current practice. *Journal of Process Control*, 2011, **21**(6): 874–884
- Sequeira S E, Graells M, Puigjaner L. Real-time evolution for online optimization of continuous processes. *Industrial and Engineering Chemistry Research*, 2002, **41**(7): 1815–1825
- Sun Z J, Qin S J, Singhal A, Megan L. Control performance monitoring of LP-MPC cascade systems. In: *Proceedings of the 2011 American Control Conference*. San Francisco, USA: IEEE, 2011. 4422–4427
- Souza G D, Odloaka D, Zanin A C. Real time optimization (RTO) with model predictive control (MPC). *Computers and Chemical Engineering*, 2010, **34**(12): 1999–2006
- Al-Shammari A A, Forbes J F. Post-optimality approach to prevent cycling in linear MPC target calculation. *European Journal of Control*, 2012, **18**(6): 558–569
- Pontes K V, Wolf I J, Embirucy M, Marquardt W. Dynamic real-time optimization of industrial polymerization processes with fast dynamics. *Industrial and Engineering Chemistry Research*, 2015, **54**(47): 11881–11893
- Ellis M, Christofides P D. Integrating dynamic economic optimization and model predictive control for optimal operation of nonlinear process systems. *Control Engineering Practice*, 2014, **22**(22): 242–251
- Chai T Y, Zhao L, Qiu J B, Liu F Z, Fan J L. Integrated network-based model predictive control for setpoints compensation in industrial processes. *IEEE Transactions on Industrial Informatics*, 2012, **9**(1): 417–426
- Zyl V F, Paquot F, Fouche F, Gomez A. Implementation of a SAG grinding expert system at Kansanshi Mine-Zambia. *IFAC Proceedings Volumes*, 2013, **46**(16): 178–181
- Qiao Jun-Fei, Han Gai-Tang, Zhou Hong-Biao. Knowledge-based intelligent optimal control for wastewater biochemical treatment process. *Acta Automatica Sinica*, 2017, **43**(6): 1038–1046  
(乔俊飞, 韩改堂, 周红标. 基于知识的污水生化处理过程智能优化方法. 自动化学报, 2017, **43**(6): 1038–1046)

- 17 Li H X, Guan S P. Hybrid intelligent control strategy. Supervising a DCS-controlled batch process. *IEEE Control Systems Magazine*, 2001, **21**(3): 36–48
- 18 Xie S W, Xie Y F, Li F B, Yang C H, Gui W H. Optimal setting and control for iron removal process based on adaptive neural network soft-sensor. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2018, **PP**(99): 1–13
- 19 Wang D, He H B, Liu D R. Adaptive critic nonlinear robust control: a survey. *IEEE Transactions on Cybernetics*, 2017, **47**(10): 3429–3451
- 20 Lewis F L, Vrabie D. Reinforcement learning and adaptive dynamic programming for feedback control. *IEEE circuits and systems magazine*, 2009, **9**(3): 32–50
- 21 Wang Ding, Mu Chao-Xu, Liu De-Rong. Data-driven nonlinear near-optimal regulation based on iterative neural dynamic programming. *Acta Automatica Sinica*, 2017, **43**(3): 366–375  
(王鼎, 穆朝絮, 刘德荣. 基于迭代神经动态规划的数据驱动非线性近似最优调节. *自动化学报*, 2017, **43**(3): 366–375)
- 22 Dai Wei, Chai Tian-You. Data-driven optimal operational control of complex grinding processes. *Acta Automatica Sinica*, 2014, **40**(9): 2005–2014  
(代伟, 柴天佑. 数据驱动的复杂磨矿过程运行优化控制方法. *自动化学报*, 2014, **40**(9): 2005–2014)
- 23 Lu X L, Kiumarsi B, Chai T Y, Lewis F L. Data-driven optimal control of operational indices for a class of industrial processes. *IET Control Theory and Applications*, 2016, **10**(12): 1348–1356
- 24 Li Jin-Na, Gao Xi-Ze, Chai Tian-You, Fan Jia-Lu. Data-driven operational optimization control of industrial processes. *Control Theory and Applications*, 2016, **33**(12): 1584–1592  
(李金娜, 高溪泽, 柴天佑, 范家璐. 数据驱动的工业过程运行优化控制. *控制理论与应用*, 2016, **33**(12): 1584–1592)
- 25 Li J N, Kiumarsi B, Chai T Y, Fan J L. Off-policy reinforcement learning: optimal operational control for two-time-scale industrial processes. *IEEE Transactions on Cybernetics*, 2017, **47**(12): 4547–4558
- 26 Kiumarsi B, Lewis F L, Modares H, Karimpour A, Naghibi M. Reinforcement Q-learning for optimal tracking control of linear discrete-time systems with unknown dynamics. *Automatica*, 2014, **50**(4): 1167–1175
- 27 Lewis F L, Vrabie D, Vamvoudakis K G. Reinforcement learning and feedback control: using natural decision methods to design optimal adaptive controllers. *IEEE Control Systems*, 2012, **32**(6): 76–105
- 28 Wei Q L, Song R B, Sun Q F. Nonlinear neuro-optimal tracking control via stable iterative Q-learning algorithm. *Neurocomputing*, 2015, **168**: 520–528
- 29 Hewer G A. An iterative technique for the computation of steady state gains for the discrete optimal regulator. *IEEE Transactions on Automation Control*, 1971, **16**(4): 382–384
- 30 Bradtke S J, Ydstie B E, Barto A G. Adaptive linear quadratic control using policy iteration. In: Proceedings of the 1994 American Control Conference. Baltimore, USA: IEEE, 1994. 3475–3475

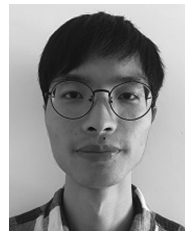
- 31 Chen X S, Yang J, Li S H, Li Q. Disturbance observer based multi-variable control of ball mill grinding circuits. *Journal of Process Control*, 2009, **19**(7): 1205–1213



代伟 中国矿业大学信息与控制工程学院副教授. 主要研究方向为复杂工业过程建模、运行优化与控制. 本文通信作者. E-mail: weidai@cumt.edu.cn

(DAI Wei Associate professor at the School of Information and Control Engineering, China University of Mining and Technology. His research interest

covers modeling, operational optimization and control for complex industrial process. Corresponding author of this paper.)



陆文捷 中国矿业大学信息与控制学院硕士研究生. 主要研究方向为复杂工业过程运行优化与控制.

E-mail: luwenjiecumt@163.com

(LU Wen-Jie Master student at the School of Information and Control Engineering, China University of Mining and Technology. His research interest

covers operational optimization and control for complex industrial process.)



付俊 东北大学流程工业综合自动化国家重点实验室教授. 主要研究方向为动态优化, 切换系统, 非线性控制.

E-mail: junfu@mail.neu.edu.cn

(FU Jun Professor at the State Key Laboratory of Synthetical Automation for Process Industries, Northeastern University. His research interest

covers dynamic optimization, switching system, and nonlinear control.)



马小平 中国矿业大学信息与控制工程学院教授. 主要研究方向为过程控制, 网络控制, 故障诊断.

E-mail: xpma@cumt.edu.cn

(MA Xiao-Ping Professor at the School of Information and Control Engineering, China University of Mining and Technology. His research interest

covers process control, networked control, and fault detection.)