自动化学报
ACTA AUTOMATICA SINICA

不可靠通信下矿物磨选过程多速率分层学习控制

任鹏旭 代伟 张淇瑞 杨春雨

Multi-rate Layered Learning Control of Mineral Grinding Process Subject to Unreliable Communication

REN Peng-Xu, DAI Wei, ZHANG Qi-Rui, YANG Chun-Yu

在线阅读 View online:
http://www.aas.net.cn/article/current/ebcf6710f5acc76ce6cc55f5dacb144cf078d15e3ac6e562ba116f29b1da59f5

您可能感兴趣的其他文章

不确定工业过程运行指标异步更新强化学习决策算法

Asynchronous Updating Reinforcement Learning Algorithm for Decision-making Operational Indices of Uncertain Industrial Processes

自动化学报. 2023, 49(2): 461-472 https://doi.org/10.16383/j.aas.c210983

具有指数图信息通信的大规模情境多智能体强化学习

Large-scale Episodic Multi-agent Reinforcement Learning With Exponential Graph Information Communication

自动化学报. 2026, 52(6): 1304-1318 https://doi.org/10.16383/j.aas.c250633

基于IT2FBLS强化学习PID的MSWI过程炉膛温度控制

Furnace Temperature Control Using IT2FBLS-based Reinforcement Learning PID for MSWI Process

自动化学报. 2025, 51(7): 1626-1641 https://doi.org/10.16383/j.aas.c240074

复杂工业过程非串级双速率组合分散运行优化控制

Non-cascade Dual-rate Composite Decentralized Operational Optimal Control for Complex Industrial Processes

自动化学报. 2023, 49(1): 172-184 https://doi.org/10.16383/j.aas.c210897

多智能体强化学习控制与决策研究综述

Survey on Multi-agent Reinforcement Learning for Control and Decision-making

自动化学报. 2025, 51(3): 510-539 https://doi.org/10.16383/j.aas.c240392

基于无模型策略梯度强化学习的未知随机系统最优控制

Model-free Policy Gradient-based Reinforcement Learning Algorithms for Optimal Control of Unknown Stochastic Systems

自动化学报. 2025, 51(10): 2245-2255 https://doi.org/10.16383/j.aas.c250156

不可靠通信下矿物磨选过程多速率分层学习控制

任鹏旭 [1] 代 伟 [1] 张淇瑞 [1] 杨春雨 [1]

摘 要 矿物磨选过程运行优化控制通常采用基础回路层和运行层双层结构, 涉及不同时间尺度的被控对象, 其运行层动态机理复杂难以建模, 且层级间存在不同采样速率与通信丢包问题, 进一步增加了控制设计难度. 针对矿物磨选过程运行优化控制中存在的多速率、不可靠通信问题, 提出一种带有通信补偿的多速率分层学习控制方法. 该方法在基础回路层采用提升技术和模型预测控制实现多速率下的设定值跟踪; 在此基础上, 通过递归提升将回路动态引入运行层, 采用强化学习技术, 结合史密斯预估器的思想, 设计带有通信补偿的快慢耦合逆学习控制算法, 以解决性能指标权重参数依赖人工经验设定、调参困难的问题, 利用演示运行数据逆向学习性能指标权重参数的同时在线更新回路设定值, 进而实现运行指标的优化控制. 理论分析和工业应用验证了所提方法的有效性.

关键词 矿物磨选过程; 运行优化控制; 多速率; 不可靠通信; 强化学习

引用格式 任鹏旭, 代伟, 张淇瑞, 杨春雨. 不可靠通信下矿物磨选过程多速率分层学习控制. 自动化学报, xxxx, **xx**(x): x−xx

DOI 10.16383/j.aas.c250584 **CSTR** 32138.14.j.aas.c250584

Multi-rate Layered Learning Control of Mineral Grinding Process Subject to Unreliable Communication

REN Peng-Xu[1] DAI Wei[1] ZHANG Qi-Rui[1] YANG Chun-Yu[1]

Abstract The operational optimal control of mineral grinding process typically adopt a two-layer structure consisting of the basic loop layer and the operational layer, involving controlled objects operating at different time scales. The operational layer is characterized by complex dynamic mechanisms that are difficult to model. Moreover, the problem of different sampling rates and communication packet loss between layers further increases the difficulty of control design. To address the problems of multi-rate and unreliable communication in the operational optimal control of mineral grinding process, a multi-rate layered learning control method with communication compensation is proposed. At the basic loop layer, lifting technology and model predictive control are used to achieve setpoint tracking under multi-rate conditions; Building on this, recursive lifting is employed to introduce loop dynamics into the operational layer, where reinforcement learning technology combined with the Smith predictor is used to design a fast-slow coupled inverse learning control algorithm with communication compensation. The proposed method addresses the reliance of performance index weight parameters on empirical manual setting and tuning difficulty by inversely learning them from demonstration operation data while updating loop setpoints online, thereby achieving optimal control of operational indices. The effectiveness of the proposed method is verified by theoretical analysis and industrial applications.

Keywords mineral grinding process; operational optimal control; multi-rate; unreliable communication; reinforcement learning

Citation Ren Peng-Xu, Dai Wei, Zhang Qi-Rui, Yang Chun-Yu. Multi-rate layered learning control of mineral grinding process subject to unreliable communication. *Acta Automatica Sinica*, xxxx, **xx**(x): x−xx

矿物磨选运行优化控制通常采用分层控制架构, 包括运行层与基础回路层. 运行层主要面向矿物磨选运行过程, 通过以太网获取现场反馈信息, 并基于运行控制系统动态调整磨机给矿量、泵池补水量等基础回路设定值, 使矿物磨选粒度、循环负荷等关键运行指标控制在目标范围内; 基础回路层则直接面向底层设备, 通过设备网传输控制输入和输出数据, 其主要目标是实现对设定值的快速跟踪, 确保各执行机构的稳定运行[1−3].

由于国外矿石品位较高且成分比较稳定, 通常以配矿的方式实现矿石的“均一化”处理, 在生产运行过程中给矿性质比较稳定, 生产过程平稳, 可以

收稿日期 2025-10-30 录用日期 2026-04-12
Manuscript received October 30, 2025; accepted April 12, 2026
江苏省自然科学基金 (BK20240102, BK20231062), 国家自然科学基金 (62373361, 62403469), 江苏省研究生科研与实践创新计划 (KYCX25_2846) 资助
Supported by Natural Science Foundation of Jiangsu Province (BK20240102, BK20231062), National Natural Science Foundation of China (62373361, 62403469), and Postgraduate Research & Practice Innovation Program of Jiangsu Province (KYCX25_2846)
本文责任编委 段海滨
Recommended by Associate Editor DUAN Hai-Bin
1. 中国矿业大学信息与控制工程学院 徐州 221116
1. School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221116

建立近似动态模型[4]. 因此, 可以通过多变量解耦控制、实时优化 (real time optimization, RTO) 和模型预测控制 (model predictive control, MPC) 等基于模型的方法实现运行优化控制目标. 文献 [5] 基于动态线性模型, 采用 RTO 方法选择可以使产量最大化的回路设定值分配给下层基础控制回路, 同时将粒度指标保持在目标范围内. 文献 [6–8] 利用 MPC 调整回路设定值, 使磨矿运行过程具有良好的经济效益. 为解决随机干扰对矿物磨选粒度的影响, 文献 [9] 提出一种基于干扰观测器的磨矿过程多变量解耦控制方法. 文献 [10] 将多变量干扰观测器与 MPC 相结合, 用以改善传统反馈控制的干扰抑制性能, 并在磨矿过程中进行有效性验证. 然而, RTO 需要稳态模型[11], 而多变量解耦控制和 MPC 依赖于精确的工艺动态模型. 由于国内矿石品位低, 矿石性质和成分波动较大, 运行指标与回路被控变量之间的动态机理不清, 很难建立所需的动态模型. 因此, 上述模型驱动的控制方法无法应用于实际工业矿物磨选过程.

智能技术凭借自适应、自学习及无需显式模型的优势, 在矿物磨选运行优化控制中受到广泛关注, 常用方法包括专家系统[12]、模糊推理[13]、案例推理[14–15] 等. 文献 [13] 针对闭路再磨过程的运行控制问题, 提出一种包含模糊切换控制器、多 PI 控制器的混合智能控制方法, 并应用于某大型赤铁矿选矿厂的再磨流程, 有效提高了粒度指标合格率. 文献 [15] 针对二段磨矿过程, 将案例推理、模糊控制与神经网络相结合, 提出带有回路设定值优化、神经网络软测量、模糊逻辑动态调节器等多模块的混合智能控制策略, 工业应用表明了其有效性. 然而, 上述基于智能技术的控制器设计主要是依靠模拟领域专家的经验与知识, 与环境交互能力有限, 当新工况出现时, 依然需要领域专家对控制器参数进行调整, 其参数整定过程具有主观性, 难以保证系统的优化运行.

在实际的矿物磨选过程中存在大量的生产过程数据, 这些数据不仅能够揭示矿物磨选的运行机理, 还能反映其运行状态[16]. 因此, 数据驱动的矿物磨选运行优化控制方法相继被提出. 文献 [17] 针对赤铁矿磨矿工艺, 提出一种基于双层分级结构的数据驱动安全运行优化控制方法. 文献 [18] 针对磨矿特性, 提出 PI 与强化学习集成的数据驱动分层运行优化控制方法. 文献 [19] 基于强化学习技术, 提出一种针对磨矿过程的数据驱动输入约束运行控制方法. 文献 [20] 基于奇异摄动技术将大规模多时间尺度系统分解为快慢变量分离的降阶子系统, 提出一种具有信号补偿的复合控制器设计方法. 上述分层运行优化控制方法仅考虑层级之间的快慢时间尺度特性, 没有考虑由于检测装置采样周期不同导致的控制周期和采样周期难以统一的问题. 针对具有多种速率的分层结构工业过程, 文献 [21] 提出一种基于 MPC、提升技术和 Q 学习的集成运行优化控制方法. 与此同时, 随着工业以太网与现场总线的广泛应用, 矿物磨选分层控制过程演变为典型的网络化控制系统 (networked control systems, NCSs)[22–23], 实现层级间以及反馈回路的实时信息交互. 然而, NCSs 中存在的随机丢包等不可靠通信问题[24], 使未显式考虑网络通信问题的多速率分层运行优化控制方法[17–21] 面临新的实现难题.

近年来, 针对 NCSs 中的数据丢包与传输时延等问题, 学术界已取得丰硕成果[25–29]. 文献 [26] 针对网络环境下浓密过程的控制问题, 设计一种基于 Q 学习的流量控制器和一种抗丢包的流量设定值控制器, 在无精确模型下实现对浓度设定值的有效跟踪和流量过程设定值的优化. 文献 [28–29] 研究具有网络诱导丢包和时延的自适应离散时间线性 H_∞ 控制问题, 虽然未考虑多速率采样特性, 但为网络环境下运行优化控制方法的设计提供了重要指导. 尽管上述控制方法在解决网络环境下运行闭环反馈控制方面取得一定的研究进展, 但在实际矿物磨选运行优化控制中, 仍面临一个更为本质的难题: 运行层性能指标难设定.

在实际矿物磨选运行优化控制中, 运行层性能指标权重参数的取值将直接影响优化控制的效果, 而选择合适的性能指标权重参数存在困难: 现场操作人员熟悉工艺流程但难以显式表达控制目标, 而算法开发人员虽具备控制设计能力, 但缺乏对工艺细节的深入理解, 导致权重设定常依赖经验或试错, 缺乏理论支持与工艺适应性. 虽然现有逆学习控制技术[30–31] 为从演示数据中重构性能指标提供理论手段, 但直接将其应用于矿物磨选 NCSs 将面临严峻的快慢时间耦合挑战: 运行层操作策略 (慢时间尺度) 本质上是基于对底层设备响应能力 (快时间尺度) 和网络通信质量的综合判断. 现有的逆学习控制方法多假设理想通信与单一速率采样, 忽略底层动态与网络丢包的耦合影响, 这极大限制了现有逆学习控制方法的应用.

矿物磨选作为选矿工业的核心环节, 面临多时间尺度耦合及通信不可靠等复杂约束, 传统控制方法难以实现理想的运行优化控制效果. 针对上述问题, 本文提出一种融合 MPC、提升技术、史密斯预估器与强化学习的集成分层运行优化控制方法. 通

过两次提升技术将多速率分层控制问题统一到同一时间尺度, 并基于 MPC 设计基础回路控制器实现设定值跟踪, 构建能够表征整体系统动态特性的广义被控对象. 在此基础上, 引入强化学习技术, 首次将史密斯预估器应用于矿物磨选运行优化控制中, 利用历史有效数据对通信丢包引起的状态信息缺失进行预测与补偿, 提升系统在不可靠通信条件下的鲁棒性. 针对运行层性能指标权重参数依赖人工经验设定、调节过程复杂的问题, 本文进一步提出快慢耦合逆学习控制方法; 在上述广义被控模型基础上进行逆学习, 使运行层性能指标权重参数的学习显式考虑底层设备动态特性及通信丢包对设定值优化的影响, 通过演示运行数据逆向学习权重参数并直接用于回路设定值动态优化, 实现矿物磨选运行优化控制目标. 最后, 通过闭路矿物磨选过程仿真验证了所提方法的有效性.

1　控制问题描述

1.1　矿物磨选过程描述

矿物磨选过程是对大颗粒的原矿石加以粉碎, 并将金属含量高的矿物或其他有价值矿物与大量无用的脉石分离, 富集精矿、抛弃无用的脉石及尾矿. 本文选取如图 1 所示的典型矿物磨选过程, 其由电振给矿机、球磨机和水力旋流器等设备组成. 在矿物磨选生产过程中, 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.

通过上述分析, 矿物磨选过程运行指标粒度 $r_1(t)$ 和循环负荷 $r_2(t)$ 受给矿量 $y_1(t)$、泵池补水量 $y_2(t)$ 的影响. 给矿量和泵池补水量这两个变量可以通过给矿机电振频率 $u_1(t)$ 和泵池补加水阀门开度 $u_2(t)$ 来控制.

矿物磨选过程可以被描述为基础回路层和运行层双层结构. 基础回路过程通常围绕某一稳态工作点运行, 从而使其在该点附近可以用线性模型进行近似描述[21]:

$$\begin{cases} \dot{x}_f(t) = A_{c1}x_f(t) + B_{c1}u(t) \\ y(t) = C_{c1}x_f(t) + D_{c1}u(t) \end{cases} \tag{1}$$

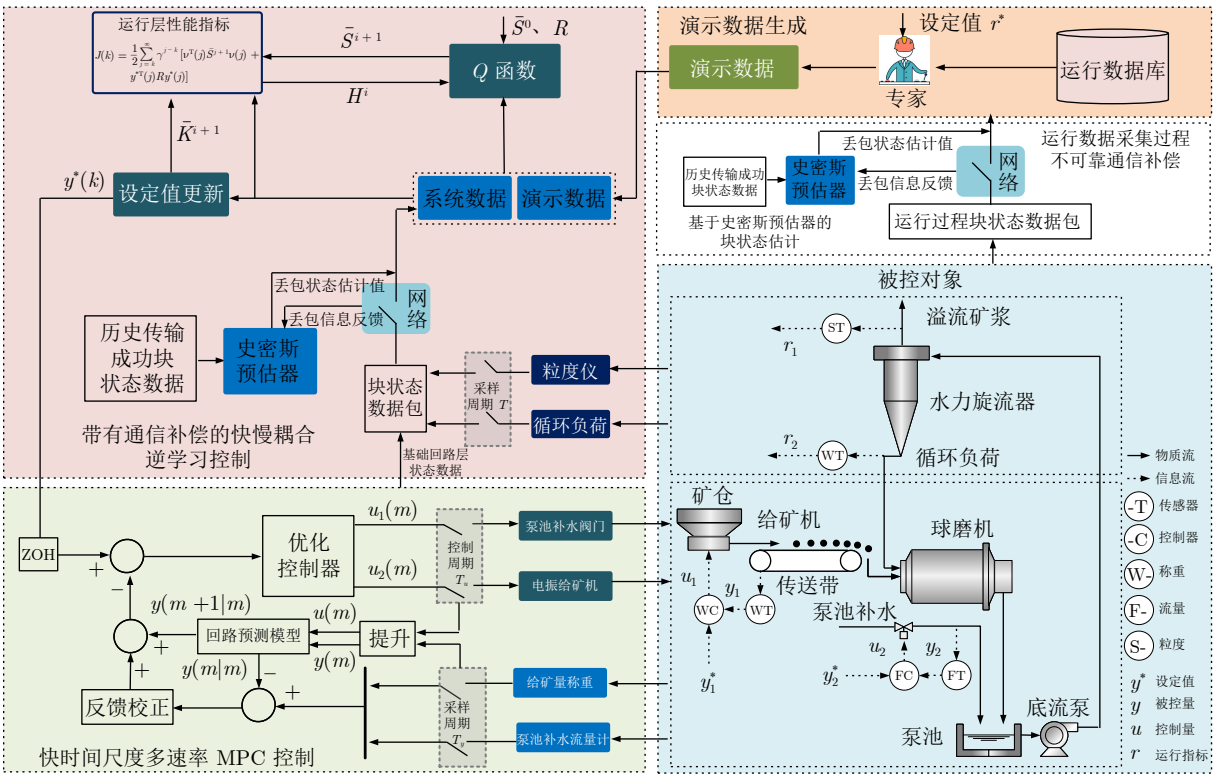


图 1　不可靠通信下矿物磨选过程多速率分层学习控制框图

Fig. 1　Block diagram of multi-rate layered learning control of mineral grinding process subject to unreliable communication

其中, $x_f(t)$ 为 2×1 维状态向量; $u(t)$ 为 2×1 维输入向量; $y(t)$ 为 2×1 维输出向量; A_{c1}、B_{c1}、C_{c1} 和 $D_{c1}=\mathbf{0}$ 为适当维度的系统矩阵. 然后, 选择 $y(t)=[y_1(t)\ \ y_2(t)]^{\mathrm{T}}$ 作为运行层的输入向量, 输出向量可以表示为 $r(t)=[r_1(t)\ \ r_2(t)]^{\mathrm{T}}$. 运行层模型可以表示为:

$$\begin{cases}\dot{x}_s(t)=A_{c2}x_s(t)+B_{c2}y(t)\\ r(t)=C_{c2}x_s(t)+D_{c2}y(t)\end{cases}\tag{2}$$

其中, $x_s(t)$ 为 5×1 维状态向量; A_{c2}、B_{c2}、C_{c2} 和 $D_{c2}=\mathbf{0}$ 为适当维度的系统矩阵.

1.2 运行优化控制结构与问题分析

矿物磨选运行优化控制目标是使用反馈运行数据选取合适的基础回路设定值 y_1^*、y_2^*, 并通过调整基础回路给矿机电振频率 u_1 和泵池补加水阀门开度 u_2 来调节回路给矿量 y_1 和泵池补水量 y_2 跟踪其设定值, 从而控制粒度 r_1 和循环负荷 r_2 保持在期望值 r_1^*、r_2^* 附近.

矿物磨选运行优化控制问题可以表述为:

$$\begin{aligned}&\min_{y^*}J(k)=\frac{1}{2}\sum_{j=k}^{\infty}\gamma^{j-k}\Big[\big(r(j)-r^*(j)\big)^{\mathrm{T}}Q\big(r(j)-\\&\quad r^*(j)\big)+y^{*\mathrm{T}}(j)Ry^*(j)\Big]\\&\text{s.t.}\begin{cases}x_s(k+1)=A_{d2}x_s(k)+B_{d2}y(k)\\ r(k)=C_{d2}x_s(k)+D_{d2}y(k)\end{cases}\\&\quad\begin{cases}x_f(n+1)=A_{d1}x_f(n)+B_{d1}u(n)\\ y(n)=C_{d1}x_f(n)+D_{d1}u(n)\end{cases}\\&\quad u(n)=p(x_f(n),\ y^*(n))\end{aligned}\tag{3}$$

其中, $0<\gamma<1$ 为折扣因子; Q 和 R 为适当维度的正定加权矩阵; k、n 分别为运行层和基础回路层对应周期下的采样点; $p(\cdot)$ 为基础回路层控制律; A_{d1}、B_{d1}、C_{d1}、D_{d1} 和 A_{d2}、B_{d2}、C_{d2}、D_{d2} 分别为基础回路层模型 (1) 和运行层模型 (2) 离散化后的系统矩阵, 具体细节将在第 2 节和第 3 节中详细阐述.

矿物磨选运行优化控制包含两层动态: 基础回路层和运行层, 并且运行过程机理复杂难以建模, 即模型参数 A_{d2}、B_{d2}、C_{d2}、D_{d2} 不容易获取. 其中基础回路过程直接与执行机构以及检测装置相接触, 受仪表物理特性所限, 其控制更新周期 T_u 与采样周期 T_y 并不一致, 且通常 $T_u<T_y$. 在实际应用中, 常见做法是将控制频率降低, 使控制周期与采样周期保持一致, 以便简化控制器设计, 但这种处理方式往往导致控制性能下降. 鉴于此, 本文在矿物磨选过程的背景下考虑 $T_u<T_y$ 的情形, 使得基础回路能够更好地发挥快速调节的作用, 从而进一步提升运行优化控制性能. 此外, 表征矿物磨选品质的粒度和循环负荷等运行指标通常按固定周期上传至控制系统, 且对设定值的刷新频率要求较低, 通常将运行层控制与运行指标采样统一为同一周期 T. 相比之下, 运行过程具有慢时间尺度特性, 并且 T_u、$T_y\ll T$. 因此, 矿物磨选过程运行优化控制是具有三种速率的分层结构, 增加了控制方法的设计难度.

现有的矿物磨选运行优化控制问题研究如何根据给定的运行层性能指标求解最优设定值策略, 即式 (3) 中权重参数矩阵 Q、R 预先给定, 而性能指标权重参数会直接影响最终的控制效果. 例如, 在采用强化学习解决矿物磨选运行优化控制问题 (3) 时, 当性能指标中跟踪误差项的权重参数矩阵 Q 较小时, 会导致跟踪误差较大; 当权重参数矩阵 Q 太大时, 学习过程可能不稳定甚至直接导致学习无法进行. 在实际应用中, 性能指标权重参数通常依赖控制算法工程师的经验或反复试错来调整, 不仅效率低、代价高, 选择不当还会导致控制效果不理想.

此外, 随着互联网技术的工业应用, 矿物磨选过程基础回路层控制输入和输出信息通过设备网进行传输, 而运行过程中回路设定控制的状态反馈信息传输采用工业以太网 (如图 1 所示), 网络传输过程中产生的丢包等问题会导致反馈控制性能下降.

综上所述, 矿物磨选运行优化控制是一个多层次、部分模型未知且受网络不可靠通信影响的复杂多速率控制问题, 对现有的优化控制方法提出了严峻的挑战. 如何在性能指标权重参数缺乏合理设置方法、多时间尺度、不可靠通信以及运行层动态特性难以精确建模等综合复杂性的背景下, 实现矿物磨选运行优化控制目标是本文研究的重点.

2 多速率基础回路模型预测控制器设计

以基础周期 h 离散化系统 (1), 可以得到:

$$\begin{cases}x_f(n+1)=A_{d1}x_f(n)+B_{d1}u(n)\\ y(n)=C_{d1}x_f(n)+D_{d1}u(n)\end{cases}\tag{4}$$

其中, $A_{d1}=\mathrm{e}^{A_{c1}h}$; $B_{d1}=(\int_0^h\mathrm{e}^{A_{c1}h}\mathrm{d}t)B_{c1}$; $C_{d1}=C_{c1}$; $D_{d1}=D_{c1}$ 为回路过程离散化后的系统矩阵.

注 1. 对于基础回路层动态系统 (4), 本文假设其系统矩阵 A_{d1}、B_{d1}、C_{d1}、D_{d1} 已知. 在通常情况下, C_{d1} 为单位矩阵, 此时 $x_f(n)$ 与 $y(n)$ 具有相同的物理含义, 即 $x_f(n)=[x_{f1}(n)\ \ x_{f2}(n)]^{\mathrm{T}}$ 表示给

矿量和泵池补水量.

控制周期与采样周期存在如下关系: $T_u = ah$, $T_y = bh$, $a < b$ 且互为质数, 框架周期可表示为 $T_0 = abh$. 为解决基础回路输入输出多速率问题, 对系统 (4) 采用提升技术, 将输入向量维度提升 b 倍, 输出向量维度提升 a 倍, 即将系统 (4) 的控制周期和输出采样周期提升至框架周期 T_0, m 为周期 T_0 下的采样点. 提升过程定义如下:

$$\bar{u}(m) = L_b u(mT_0) = \begin{bmatrix} u(mT_0) \\ u(mT_0 + T_u) \\ \vdots \\ u(mT_0 + (b-1)T_u) \end{bmatrix}$$

$$\bar{y}(m) = L_a y(mT_0) = \begin{bmatrix} y(mT_0) \\ y(mT_0 + T_y) \\ \vdots \\ y(mT_0 + (a-1)T_y) \end{bmatrix}$$

其中, L_a、L_b 为提升算子; $\bar{u}(m) \in \mathbf{R}^{N_u}$, $N_u = 2b$; $\bar{y}(m) \in \mathbf{R}^{N_y}$, $N_y = 2a$.

提升后的基础回路模型为:

$$\begin{cases} x_f(m+1) = \tilde{A}_{d1} x_f(m) + \tilde{B}_{d1} \bar{u}(m) \\ \bar{y}(m) = \tilde{C}_{d1} x_f(m) + \tilde{D}_{d1} \bar{u}(m) \end{cases} \tag{5}$$

其中, $\tilde{A}_{d1}$、$\tilde{B}_{d1}$、$\tilde{C}_{d1}$ 和 $\tilde{D}_{d1}$ 的提升推导过程详见文献 [21, 32], 此处不再赘述.

对于提升后的系统 (5), 可以采用 MPC 对其进行控制, 预测时域和控制时域均可设置为 1. 详细推导过程如下:

$$x_f(m+1|m) = \tilde{A}_{d1} x_f(m) + \tilde{B}_{d1} \bar{u}(m) \tag{6}$$

$$\begin{aligned} \bar{y}(m+1|m) = \tilde{C}_{d1} x_f(m+1|m) + \tilde{D}_{d1} \bar{u}(m+1) \ = \\ \tilde{C}_{d1} \tilde{A}_{d1} x_f(m) + \tilde{D}_{d1} \tilde{B}_{d1} \bar{u}(m) + \\ \tilde{D}_{d1} \bar{u}(m+1) \end{aligned} \tag{7}$$

由于基础回路控制周期和采样周期通常为秒级或毫秒级, 前后两个时刻系统动态变化不大, 可将 $\bar{u}(m+1)$ 用 $\bar{u}(m)$ 近似, 得到预测方程:

$$\bar{y}(m+1|m) = \psi_x x_f(m) + \psi_u \bar{u}(m) \tag{8}$$

其中, $\psi_x = \tilde{C}_{d1} \tilde{A}_{d1}$; $\psi_u = \tilde{D}_{d1} \tilde{B}_{d1} + \tilde{D}_{d1}$.

针对提升后的基础回路预测方程 (8), 目标函数可以定义为如下形式:

$$\begin{aligned} J_1(m) = \|\Gamma_{\bar{y}}(\bar{y}(m+1|m) - \alpha y^*(m+1))\|^2 + \\ \|\Gamma_{\bar{u}} \bar{u}(m)\|^2 \end{aligned} \tag{9}$$

其中, $y^*(m+1)$ 为基础回路设定值; α 为与 $\bar{y}(m+1)$ 相同维度的单位矩阵; $\Gamma_{\bar{y}}$ 和 $\Gamma_{\bar{u}}$ 分别为提升后系统控制输出和控制输入的加权因子且

$$\begin{cases} \Gamma_{\bar{y}} := \mathrm{diag}\{\Gamma_{\bar{y}_1}, \Gamma_{\bar{y}_2}, \cdots, \Gamma_{\bar{y}_{N_y}}\} \\ \Gamma_{\bar{u}} := \mathrm{diag}\{\Gamma_{\bar{u}_1}, \Gamma_{\bar{u}_2}, \cdots, \Gamma_{\bar{u}_{N_u}}\} \end{cases}$$

为求解基础回路最优控制输入, 可将目标函数 (9) 写成如下紧凑形式:

$$J_1(m) = \varrho^{\mathrm{T}} \varrho \tag{10}$$

其中

$$\varrho := \begin{bmatrix} \Gamma_{\bar{y}}(\bar{y}(m+1|m) - \alpha y^*(m+1)) \\ \Gamma_{\bar{u}} \bar{u}(m) \end{bmatrix}$$

根据极值必要条件, 将预测方程 (8) 代入式 (10) 中, 并令 $\frac{\mathrm{d}J_1(m)}{\mathrm{d}\bar{u}(m)} = 0$, 基础回路层最优控制输入为:

$$\bar{u}(m) = L_{\mathrm{mpc}}(\alpha y^*(m+1) - \psi_x x_f(m)) \tag{11}$$

其中

$$L_{\mathrm{mpc}} = (\psi_u^{\mathrm{T}} \Gamma_{\bar{y}}^{\mathrm{T}} \Gamma_{\bar{y}} \psi_u + \Gamma_{\bar{u}}^{\mathrm{T}} \Gamma_{\bar{u}})^{-1} \psi_u^{\mathrm{T}} \Gamma_{\bar{y}}^{\mathrm{T}} \Gamma_{\bar{y}} \tag{12}$$

由于运行层是在慢时间尺度下给出设定值, 在运行控制周期 T ($\zeta = T/T_0$) 内保持不变, 需要零阶保持器 (zero-order holder, ZOH) 将慢信号维持到快信号, 所以有 $y^*(m+j) = y^*(m)$, $j = 1, 2, \cdots, \zeta - 1$. 最优控制输入 (11) 可以写成如下形式:

$$\bar{u}(m) = L_{\mathrm{mpc}}(\alpha y^*(m) - \psi_x x_f(m)) \tag{13}$$

将控制输入 (13) 代入提升后的基础回路模型 (5), 可以得到基础回路层闭环动态模型:

$$\begin{cases} x_f(m+1) = \tilde{A} x_f(m) + \tilde{B} y^*(m) \\ \bar{y}(m) = \tilde{C} x_f(m) + \tilde{D} y^*(m) \end{cases} \tag{14}$$

其中, $\tilde{A} = \tilde{A}_{d1} - \tilde{B}_{d1} L_{\mathrm{mpc}} \psi_x$; $\tilde{B} = \tilde{B}_{d1} L_{\mathrm{mpc}} \alpha$; $\tilde{C} = \tilde{C}_{d1} - \tilde{D}_{d1} L_{\mathrm{mpc}} \psi_x$; $\tilde{D} = \tilde{D}_{d1} L_{\mathrm{mpc}} \alpha$.

3　快慢耦合逆学习控制

3.1　快慢耦合动态处理方法

由于基础回路层闭环动态模型 (14) 周期为 T_0, 而运行层控制周期为 T, 且 $T \gg T_0$. 为解决层级之间的快慢时间耦合问题, 通过递推提升技术, 将基础回路动态提升到 T 周期, k 为运行层采样周期 T 下的运行步数, 详细过程如下:

$$
\begin{aligned}
&x_f((k+1)T) = x_f(kT+\zeta T_0) = \\
&\tilde{A}x_f(kT+(\zeta-1)T_0)+\tilde{B}y^*(kT) = \\
&\tilde{A}(\tilde{A}x_f(kT+(\zeta-2)T_0)+\tilde{B}y^*(kT))+ \\
&\tilde{B}y^*(kT) = \tilde{A}^{\zeta}x_f(kT)+\sum_{i=0}^{\zeta-1}\tilde{A}^i\tilde{B}y^*(kT)
\end{aligned} \tag{15}
$$

递归提升后的基础回路层模型可以表示为:

$$
\begin{cases}
x_f(k+1) = \check{A}x_f(k)+\check{B}y^*(k) \\
\bar{y}(k) = \check{C}x_f(k)+\check{D}y^*(k)
\end{cases} \tag{16}
$$

其中, $\check{A}=\tilde{A}^{\zeta}$; $\check{B}=\sum_{i=0}^{\zeta-1}\tilde{A}^i\tilde{B}$; $\check{C}=\tilde{C}$; $\check{D}=\tilde{D}$.

对式 (2) 以周期 T 进行离散化可以得到:

$$
\begin{cases}
x_s(k+1) = A_{d2}x_s(k)+B_{d2}y(k) \\
r(k) = C_{d2}x_s(k)+D_{d2}y(k)
\end{cases} \tag{17}
$$

其中, $A_{d2}=\mathrm{e}^{A_{c2}T}$; $B_{d2}=(\int_0^T \mathrm{e}^{A_{c2}t}\mathrm{d}t)B_{c2}$; $C_{d2}=C_{c2}$; $D_{d2}=D_{c2}$ 为运行过程离散化后系统矩阵.

注 2. 矿物磨选过程运行层动态机理复杂且难以精确建模, 通常无法仅通过物理机理直接构造出与实际工艺变量一一对应的状态变量 $x_s(k)$. 虽然这些状态缺乏明确的物理解释, 但它们能够有效刻画运行层的动态特性. 在实际应用中, 运行层的状态信息可以通过软测量方法由可测数据间接获得.

在 T 周期下的基础回路模型 (16) 可与运行层结合构成运行层的广义被控模型.

易知 $\bar{y}(k)$ 与 $y(k)$ 存在如下关系:

$$
y(k) = \theta\bar{y}(k) \tag{18}
$$

其中

$$
\theta = \begin{bmatrix} I_{2\times 2} & \underbrace{\mathbf{0}_{2\times 2}\cdots\mathbf{0}_{2\times 2}}_{a-1} \end{bmatrix}
$$

结合基础回路层模型 (16) 和运行层模型 (17), 定义状态向量 $x(k+1)=[x_f^{\mathrm{T}}(k+1)\ \ x_s^{\mathrm{T}}(k+1)]^{\mathrm{T}}$, 可以得到运行层广义被控对象模型:

$$
\begin{cases}
x(k+1) = Ax(k)+By^*(k) \\
r(k) = Cx(k)+Dy^*(k)
\end{cases} \tag{19}
$$

其中

$$
A = \begin{bmatrix} \check{A} & 0 \\ B_{d2}\theta\check{C} & A_{d2} \end{bmatrix},\ B = \begin{bmatrix} \check{B} \\ B_{d2}\theta\check{D} \end{bmatrix}
$$

$$
C = \begin{bmatrix} D_{d2}\theta\check{C} & C_{d2} \end{bmatrix},\ D = D_{d2}\theta\check{D}
$$

状态 $x(k)$ 通过通信网络传递到设定值控制器, 并计算设定值 $y^*(k)$(如图 1 所示). 假设状态信息采用单包传输工作方式, 并且状态数据包在传输中会随机丢失, 即网络引起的丢包, 从而引发不可靠通信的问题. 因此, 当丢包发生时, 设定值控制器可用的状态 $x_\sigma(k)$ 为:

$$
x_\sigma(k) = \sigma(k)x(k)+(1-\sigma(k))x_\sigma(k-1) \tag{20}
$$

其中, $\sigma(k)\in(0,\ 1)$ 表示第 k 个运行层采样周期反馈数据的接收状态, $\sigma(k)=0$ 表示丢包发生, $\sigma(k)=1$ 表示正常传输. 考虑到工业现场通信丢包的发生通常具有突发性与短时随机性, 在运行层分钟级采样尺度下, 相邻采样周期之间的丢包事件相关性较弱, 可近似视为独立随机事件. 因此, 可采用伯努利随机变量对丢包过程进行建模, 即 $P(\sigma(k)=0)=p$, $P(\sigma(k)=1)=1-p$, 其中 p 为丢包概率.

为保证系统始终工作在闭环控制框架下, 避免因长时间连续丢包而退化为开环运行, 有必要对连续丢包行为作出合理约束. 因此, 作如下假设:

假设 1. 连续反馈丢包数是有界的, 且满足

$$
\sum_{i=0}^{\sigma_{f_{\max}}}\sigma(k-i) > 0 \tag{21}
$$

其中, $\sigma_{f_{\max}}$ 为最大连续丢包数.

此时, 定义 k 时刻前连续丢包数为 $\sigma_{fn}(k)$. 那么, 在 k 时刻设定值反馈控制器获得的最新状态数据为:

$$
x_\sigma(k) = x(k-\sigma_{fn}(k)) \tag{22}
$$

当 $x(k)$ 正常传输时, $\sigma_{fn}(k)=0$; 当 $x(k)$ 出现丢包时, $\sigma_{fn}(k)=\sigma_i$, 且 $1\le\sigma_i\le\sigma_{f_{\max}}$.

注 3. 网络控制系统通信协议一般是 TCP 或 UDP. 在 TCP 情况下, 存在对接收数据包的确认, 丢包可由协议层直接感知并记录; 在 UDP 情况下, 通信链路不提供反馈, 可通过添加序号或者时间戳对丢包进行检测. 因此, 无论使用 TCP 还是 UDP, 丢包情况 $\sigma_{fn}(k)$ 是可观测的, 可以用来改进控制器设计, 用于工业实际.

3.2 不可靠通信下运行层设定值优化策略

运行指标矿物磨选粒度和循环负荷的期望值一般是恒定的, 可以表示为如下形式:

$$
r^*(k+1) = Fr^*(k) \tag{23}
$$

其中, F 为单位矩阵. 根据式 (19) 和式 (23), 引入增广向量 $X(k)=\left[x^{\mathrm{T}}(k)\ \ r^{*\mathrm{T}}(k)\right]^{\mathrm{T}}$, 得到增广系统模型:

$$
X(k+1) = \bar{A}X(k)+\bar{B}y^*(k) \tag{24}
$$

其中

$$\bar{A}=\begin{bmatrix} A & 0 \\ 0 & F \end{bmatrix},\ \bar{B}=\begin{bmatrix} B \\ 0 \end{bmatrix}$$

选取设定值形式为:

$$y^*(k)=-KX(k)=-K_x x(k)-K_r r^*(k) \tag{25}$$

其中, K 为设定值增益; K_x、K_r 分别为前馈增益和反馈增益. 根据增广向量 $X(k)$ 的定义, 矿物磨选运行优化控制问题的性能指标 (3) 可以重新表示为:

$$V(X(k),\ y(k))=\frac{1}{2}\sum_{j=k}^{\infty}\gamma^{j-k}\Big[X^{\mathrm{T}}(j)SX(j)+ y^{*\mathrm{T}}(j)Ry^*(j)\Big] \tag{26}$$

其中, $S=C_1^{\mathrm{T}}QC_1$, $C_1=[C-DK_x\ -DK_r-I]$.

根据最优性的必要条件[33], 最优设定值 $y^*(k)=-KX(k)$ 最小化 $V(X(k),\ y^*(k))$, 如下等式成立.

$$V^*(X(k),\ y^*(k))=\min_{y^*}\frac{1}{2}\sum_{j=k}^{\infty}\gamma^{j-k}\Big[X^{\mathrm{T}}(j)SX(j)+ y^{*\mathrm{T}}(j)Ry^*(j)\Big]=X^{\mathrm{T}}(k)PX(k) \tag{27}$$

其中, $P>0$ 为正定对称矩阵.

基于式 (24)、(26) 和 (27), 有如下贝尔曼方程:

$$X^{\mathrm{T}}(k)PX(k)=X^{\mathrm{T}}(k)SX(k)+y^{*\mathrm{T}}(k)Ry^*(k)+ \gamma X^{\mathrm{T}}(k+1)PX(k+1) \tag{28}$$

根据贝尔曼方程 (28), 可以得到哈密尔顿函数:

$$H(X(k),\ y^*(k))=X^{\mathrm{T}}(k)SX(k)+y^{*\mathrm{T}}(k)Ry^*(k)+ \gamma X^{\mathrm{T}}(k+1)PX(k+1)\ -X^{\mathrm{T}}(k)PX(k) \tag{29}$$

进一步, 对应的最优设定值有如下形式:

$$y^*(k)=-KX(k)= -\left(R+\gamma\bar{B}^{\mathrm{T}}P\bar{B}\right)^{-1}\gamma\bar{B}^{\mathrm{T}}P\bar{A}X(k) \tag{30}$$

且 P 满足如下代数黎卡提方程

$$S-P+\gamma\bar{A}^{\mathrm{T}}P\bar{A}- \gamma^2\bar{A}^{\mathrm{T}}P\bar{B}(R+\gamma\bar{B}^{\mathrm{T}}P\bar{B})^{-1}\bar{B}^{\mathrm{T}}P\bar{A}=0 \tag{31}$$

当通信丢包发生时, 根据连续反馈丢包数的定义, 网络引起的状态数据丢包可视为随机的有界时延. 利用史密斯预估器的思想, 根据状态和控制输入历史信息预测当前丢失状态, 即

$$x(k)=A^{\sigma_{fn}(k)}x(k-\sigma_{fn}(k))+ \sum_{i=1}^{\sigma_{fn}(k)}A^{i-1}By^*(k-i) \tag{32}$$

结合式 (32), 增广向量 $X(k)$ 可以写为如下形式:

$$X(k)=\Phi\nu(k) \tag{33}$$

其中

$$\Phi=\begin{bmatrix} I & A\cdots A^{\sigma_{f\max}} & B & AB\cdots A^{\sigma_{f\max}-1}B & 0 \\ 0 & 0\cdots\ \ 0 & 0 & 0\ \ \cdots\ \ 0 & I \end{bmatrix} \tag{34}$$

且对于 $\sigma_{fn}(k)=0,\ 1,\ \cdots,\ \sigma_{f_{\max}}$, $\nu(k)\in\mathbf{R}^{n_\nu}$ $(n_\nu=7(\sigma_{f_{\max}}+1)+4\sigma_{f_{\max}}+2)$ 可表示为:

$$\begin{aligned}
&\sigma_{fn}(k)=0,\\
&\quad\nu(k)=\Big[\underbrace{x_\sigma^{\mathrm{T}}(k)\ 0\ \cdots\ 0}_{\sigma_{f_{\max}+1}}\ \underbrace{0\ \cdots\ 0}_{\sigma_{f_{\max}}}\ r^*(k)\Big]^{\mathrm{T}}\\
&\sigma_{fn}(k)=1,\\
&\quad\nu(k)=\Big[\underbrace{0\ x_\sigma^{\mathrm{T}}(k)\ \cdots\ 0}_{\sigma_{f_{\max}+1}}\\
&\qquad\underbrace{y^{*\mathrm{T}}(k-1)\ \cdots\ 0}_{\sigma_{f_{\max}}}\ r^*(k)\Big]^{\mathrm{T}}\\
&\qquad\vdots\\
&\sigma_{fn}(k)=\sigma_{f_{\max}},\\
&\quad\nu(k)=\Big[\underbrace{0\ 0\ \cdots\ x_\sigma^{\mathrm{T}}(k)}_{\sigma_{f_{\max}+1}}\\
&\qquad\underbrace{y^{*\mathrm{T}}(k-1)\ \cdots\ y^{*\mathrm{T}}(k-\sigma_{f_{\max}})}_{\sigma_{f_{\max}}}\ r^*(k)\Big]^{\mathrm{T}}
\end{aligned} \tag{35}$$

将式 (33) 代入式 (24), 可以得到

$$X(k+1)=\Phi\nu(k+1)=\bar{A}\Phi\nu(k)+\bar{B}y^*(k) \tag{36}$$

由式 (36) 可得

$$\nu(k+1)=\hat{A}\nu(k)+\hat{B}y^*(k) \tag{37}$$

其中, $\hat{A}=\Phi^*\bar{A}\Phi$, $\Phi^*=\Phi^{\mathrm{T}}(\Phi\Phi^{\mathrm{T}})^{-1}$ 是 Φ 的右逆; $\hat{B}=\Phi^*\bar{B}$.

当状态丢包发生时, 根据式 (36), 最优值函数 (27) 对于状态 $\nu(k)$ 具有如下二次型形式:

$$V^*(\nu(k),\ y^*(k)) = \min_{y^*}\frac{1}{2}\sum_{j=k}^{\infty}\gamma^{j-k}\Big[\nu^{\mathrm{T}}(j)\bar{S}\nu(j) + y^{*\mathrm{T}}(j)Ry^*(j)\Big] = \nu^{\mathrm{T}}(k)\bar{P}\nu(k) \tag{38}$$

其中, $\bar{S} = \Phi^{\mathrm{T}}S\Phi$; $\bar{P} = \Phi^{\mathrm{T}}P\Phi$.

对于式 (28), 用 $\nu(k)$ 代替 $X(k)$, 贝尔曼方程 (28) 可以表示为:

$$\nu^{\mathrm{T}}(k)\bar{P}\nu(k) = \nu^{\mathrm{T}}(k)\bar{S}\nu(k) + y^{*\mathrm{T}}(k)Ry^*(k) + \gamma\nu^{\mathrm{T}}(k+1)\bar{P}\nu(k+1) \tag{39}$$

对应的哈密尔顿函数可以表示为:

$$H(\nu(k),\ y^*(k)) = \nu^{\mathrm{T}}(k)\bar{S}\nu(k) + y^{*\mathrm{T}}(k)Ry^*(k) + \gamma\nu^{\mathrm{T}}(k+1)\bar{P}\nu(k+1)\ -\nu^{\mathrm{T}}(k)\bar{P}\nu(k) \tag{40}$$

根据极值必要条件 $\frac{\partial H(\nu(k),\ y^*(k))}{\partial y^*(k)} = 0$, 结合式 (37), 最优设定值可以表示为:

$$y^*(k) = -\bar{K}\nu(k) = -\left(R + \gamma\hat{B}^{\mathrm{T}}\bar{P}\hat{B}\right)^{-1}\gamma\hat{B}^{\mathrm{T}}\bar{P}\hat{A}\nu(k) \tag{41}$$

其中, $\bar{K} = K\Phi$; $\bar{P}$ 满足如下黎卡提方程:

$$\bar{S} - \bar{P} + \gamma\hat{A}^{\mathrm{T}}\bar{P}\hat{A} - \gamma^2\hat{A}^{\mathrm{T}}\bar{P}\hat{B}(R + \gamma\hat{B}^{\mathrm{T}}\bar{P}\hat{B})^{-1}\hat{B}^{\mathrm{T}}\bar{P}\hat{A} = 0 \tag{42}$$

注 4. 当状态丢包发生时, 史密斯预估补偿误差来源于通信丢包期间状态信号的重构误差, 由于连续丢包长度有界, 该误差在丢包区间内保持有界, 并等效为作用于反馈通道中的加性扰动, 而不改变闭环矩阵 $\hat{A} - \hat{B}\bar{K}$ 的结构. 由下文定理 2 和定理 3 可知闭环系统稳定, 因此该有界误差仅引起有限性能退化, 而不会破坏系统稳定性与算法收敛性, 具有一定的鲁棒性. 类似结论已在 Astrom 等[34] 提出的史密斯预估结构中得到系统性的分析与论证.

针对上述黎卡提方程 (42) 的求解, 可通过文献 [35] 等基于强化学习的方法实现. 然而, 对于控制算法工程师而言, 如何选择一组合适的性能指标权重参数 ($\bar{S}$, R) 仍然是一个具有挑战性的问题, 通常依赖人工试错设定. 在实际的矿物磨选运行优化控制中, ($\bar{S}$, R) 对控制性能具有决定性影响. 当 $\bar{S}$ 取值过大时, 控制器更倾向于保证粒度精度, 能够有效提升产品合格率, 但由于粒度与循环负荷之间存在强耦合关系, 为保持粒度稳定, 系统往往需要大幅度调整给矿量和补水量, 从而引发循环负荷波动并导致能耗增加; 当 $\bar{S}$ 取值过小时, 控制器对粒度偏差的抑制作用减弱, 执行动作相对温和, 虽然循环负荷较为平稳且能耗较低, 但粒度跟踪精度下降, 合格率降低, 产品质量难以保证.

为解决这一问题, 本文设计一种带有通信补偿的快慢耦合逆学习控制算法 (如图 1 所示), 利用演示运行数据逆向学习性能指标权重参数, 并将学习到的性能指标直接应用于回路设定值的动态优化, 从而实现矿物磨选运行优化控制目标.

3.3 带有通信补偿的快慢耦合逆学习控制

3.3.1 演示数据生成

在实际矿物磨选生产过程中会产生大量运行数据, 这些数据能够反映系统的动态特性和操作行为. 经验丰富的专家结合工艺知识对历史运行数据进行筛选和整理, 主要包括运行指标 (如粒度、循环负荷) 和关键控制变量 (如给矿量、补水量), 从而形成能够代表期望控制行为 $y_d^*(k)$ 的演示数据. 为便于算法推导, 本文将演示运行数据视为由如下演示系统产生, 其形式如下所示:

$$X_d(k+1) = \bar{A}X_d(k) + \bar{B}y_d^*(k) \tag{43}$$

其中, $X_d(k) = \left[x_d^{\mathrm{T}}(k)\ \ r_d^{*\mathrm{T}}(k)\right]^{\mathrm{T}}$ 为演示系统增广状态, $x_d(k)$ 为广义被控对象演示状态, $r_d^*(k)$ 为演示系统期望运行指标; $y_d^*(k)$ 为设定值输入, 演示轨迹可以表示为 (X_d, y_d^*).

对应地, 假设演示系统 (43) 隐含如下二次型性能指标函数:

$$V_d(X_d(k),\ y_d^*(k)) = \frac{1}{2}\sum_{j=k}^{\infty}\gamma^{j-k}\Big[X_d^{\mathrm{T}}(j)S_dX_d(j) + y_d^{*\mathrm{T}}(j)R_dy_d^*(j)\Big] \tag{44}$$

其中, $S_d = C_1^{\mathrm{T}}Q_dC_1$, Q_d 和 $R_d > 0$ 分别为演示系统运行指标跟踪误差和设定值输入的权重矩阵.

注 5. 与增广系统 (24) 类似, 演示系统 (43) 可根据式 (4) ~ (24) 的过程整理得到. 两者区别在于演示系统 (43) 是已经实现较为理想控制效果的闭环系统, 其中 $y_d^*(k)$ 是由经验丰富的专家结合工艺知识对历史运行数据进行筛选与整理后确定的设定值输入, 能够满足实际生产需求, 而被控系统 (24) 需要进行控制方法的设计来达到理想控制效果. 性能指标 (44) 中权重参数 S_d、$R_d > 0$ 为演示系统隐含的目标信息, 其对于被控系统是未知的, 该演示系统仅作为演示数据的发生器. 与传统依赖人工经验反复调节性能指标权重的方法不同, 本文方法中工程经验仅用于演示数据的筛选阶段, 而控制器设计过程中性能指标权重通过本文方法学习获得, 从

而显著降低了对人工经验的整体依赖程度.

下面将给出上述演示系统所满足的最优性条件. 根据最优控制理论, 最优设定值 $y_d^*(k) = -K_d \times X_d(k)$ 最小化 $V_d(X_d(k),\, y_d^*(k))$, 即

$$V_d^*(X_d(k),\, y_d^*(k)) = \min_{y_d^*} \frac{1}{2}\sum_{j=k}^{\infty}\gamma^{j-k}\big[X_d^{\mathrm{T}}(j)S_dX_d(j) + y_d^{*\mathrm{T}}(j)R_dy_d^*(j)\big] = X_d^{\mathrm{T}}(k)P_dX_d(k) \tag{45}$$

其中, K_d 为演示系统设定值增益; $P_d > 0$ 为对称矩阵.

基于式 (43) ~ (45), 可以给出如下贝尔曼方程:

$$X_d^{\mathrm{T}}(k)P_dX_d(k) = X_d^{\mathrm{T}}(k)S_dX_d(k) + y_d^{*\mathrm{T}}(k)R_dy_d^*(k) + \gamma X_d^{\mathrm{T}}(k+1)P_dX_d(k+1) \tag{46}$$

根据贝尔曼方程 (46), 可以得到哈密尔顿函数:

$$H_d(X_d(k),\, y_d^*(k)) = X_d^{\mathrm{T}}(k)S_dX_d(k) + y_d^{*\mathrm{T}}(k)R_dy_d^*(k) + \gamma X_d^{\mathrm{T}}(k+1)P_dX_d(k+1) - X_d^{\mathrm{T}}(k)P_dX_d(k) \tag{47}$$

进一步, 根据极值必要条件, 对应的最优设定值增益 K_d 为:

$$K_d = \left(R_d + \gamma\bar{B}^{\mathrm{T}}P_d\bar{B}\right)^{-1}\gamma\bar{B}^{\mathrm{T}}P_d\bar{A} \tag{48}$$

且 P_d 满足如下代数黎卡提方程

$$S_d - P_d + \gamma\bar{A}^{\mathrm{T}}P_d\bar{A} - \gamma^2\bar{A}^{\mathrm{T}}P_d\bar{B}(R_d + \gamma\bar{B}^{\mathrm{T}}P_d\bar{B})^{-1}\bar{B}^{\mathrm{T}}P_d\bar{A} = 0 \tag{49}$$

上述式 (43) ~ (49) 是最优控制问题标准推导过程, 旨在说明演示系统 (43) 满足的隐含条件, 接下来将在此基础上对发生丢包时演示系统的隐含性质进行推导.

演示状态 $x_d(k)$ 在收集过程中同样存在随机丢失的情况. 当演示系统 (43) 的状态 $x_d(k)$ 发生丢失时, 最近可达状态 $x_{d\delta}(k)$ 为:

$$x_{d\delta}(k) = x_d(k - \delta_{fn}(k)) \tag{50}$$

其中, $\delta_{fn}(k)$ 为演示系统 k 时刻丢包数, 同样满足假设 1, 最大连续丢包数为 $\delta_{f_{\max}}$.

与式 (32) ~ (35) 类似, 演示状态 $x_d(k)$ 发生丢包时, $X_d(k)$ 可以通过史密斯预估器进行补偿预测:

$$X_d(k+1) = \Phi\nu_d(k+1) = \bar{A}\Phi\nu_d(k) + \bar{B}y_d^*(k) \tag{51}$$

由式 (51) 可得

$$\nu_d(k+1) = \hat{A}\nu_d(k) + \hat{B}y_d^*(k) \tag{52}$$

其中, Φ 和 $\nu_d(k)$ 具有与 (34) 和 (35) 相似的形式.

当演示系统状态发生丢包时, 根据式 (51), 最优值函数 (45) 对于状态 $\nu_d(k)$ 具有二次型形式:

$$V_d^*(\nu_d(k),\, y_d^*(k)) = \min_{y_d^*}\frac{1}{2}\sum_{j=k}^{\infty}\gamma^{j-k}\big[\nu_d^{\mathrm{T}}(j)\bar{S}_d\nu_d(j) + y_d^{*\mathrm{T}}(j)R_dy_d^*(j)\big] = \nu_d^{\mathrm{T}}(k)\bar{P}_d\nu_d(k) \tag{53}$$

其中, $\bar{P}_d = \Phi^{\mathrm{T}}P_d\Phi$; $\bar{S}_d = \Phi^{\mathrm{T}}S_d\Phi$.

对于式 (46), 用 $\nu_d(k)$ 代替 $X_d(k)$, 贝尔曼方程可以表示为:

$$\nu_d^{\mathrm{T}}(k)\bar{P}_d\nu_d(k) = \nu_d^{\mathrm{T}}(k)\bar{S}_d\nu(k) + y_d^{*\mathrm{T}}(k)R_dy_d^*(k) + \gamma\nu_d^{\mathrm{T}}(k+1)\bar{P}_d\nu_d(k+1) \tag{54}$$

对应的哈密尔顿函数可以表示为:

$$H_d(\nu_d(k),\, y_d^*(k)) = \nu_d^{\mathrm{T}}(k)\bar{S}_d\nu_d(k) + y_d^{*\mathrm{T}}(k)R_dy_d^*(k) + \gamma\nu_d^{\mathrm{T}}(k+1)\bar{P}_d\nu_d(k+1) - \nu_d^{\mathrm{T}}(k)\bar{P}_d\nu_d(k) \tag{55}$$

此时, 根据极值必要条件 $\frac{\partial H_d}{\partial y_d^*} = 0$, 结合式 (52), 不可靠通信下最优设定值为:

$$y_d^*(k) = -\bar{K}_d\nu_d(k) = -\left(R_d + \gamma\hat{B}^{\mathrm{T}}\bar{P}_d\hat{B}\right)^{-1}\gamma\hat{B}^{\mathrm{T}}\bar{P}_d\hat{A}\nu_d(k) \tag{56}$$

其中, $\bar{K}_d = K_d\Phi$; $\bar{P}_d$ 满足如下黎卡提方程:

$$\bar{S}_d - \bar{P}_d + \gamma\hat{A}^{\mathrm{T}}\bar{P}_d\hat{A} - \gamma^2\hat{A}^{\mathrm{T}}\bar{P}_d\hat{B}(R_d + \gamma\hat{B}^{\mathrm{T}}\bar{P}_d\hat{B})^{-1}\hat{B}^{\mathrm{T}}\bar{P}_d\hat{A} = 0 \tag{57}$$

假设 2. 演示系统性能指标 (53) 中权重参数 $(\bar{S}_d,\, R_d)$, 式 (56) 和 (57) 中的控制参数 $(\bar{K}_d,\, \bar{P}_d)$ 对于被控系统 (24) 是未知的, 仅演示轨迹数据 $(\nu_d(k),\, y_d^*(k))$ 可知.

3.3.2　带有通信补偿的模型驱动逆学习控制算法

在定理 1 中, 本文给出了逆学习问题的解需要满足的条件, 这为开发带有通信补偿的模型驱动逆学习控制算法提供了理论基础.

定理 1. 如果核矩阵 $\bar{P}$ 同时满足式 (42) 和

$$\bar{S} + \gamma(\hat{A} - \hat{B}\bar{K}_d)^{\mathrm{T}}P(\hat{A} - \hat{B}\bar{K}_d) + \bar{K}_d^{\mathrm{T}}R\bar{K}_d - \bar{P} = 0 \tag{58}$$

那么, 式 (41) 中的 $\bar{K}$ 和式 (56) 中的 $\bar{K}_d$ 是相等的.

证明. 式 (42) 可以重写为如下形式:

$$\bar{S} + \gamma(\hat{A} - \hat{B}\bar{K})^{\mathrm{T}}\bar{P}(\hat{A} - \hat{B}\bar{K}) + \bar{K}^{\mathrm{T}}R\bar{K} - \bar{P} = 0 \tag{59}$$

并减去式 (58), 可以得到:

$$\gamma(\hat{A}-\hat{B}\bar{K})^{\mathrm{T}}\bar{P}(\hat{A}-\hat{B}\bar{K})+\bar{K}^{\mathrm{T}}R\bar{K}-(\hat{A}-\hat{B}\bar{K}_d)^{\mathrm{T}}\bar{P}(\hat{A}-\hat{B}\bar{K}_d)-\bar{K}_d^{\mathrm{T}}R\bar{K}_d=0 \tag{60}$$

展开后整理可得:

$$\gamma\hat{A}^{\mathrm{T}}\bar{P}\hat{B}(\bar{K}_d-\bar{K})+\gamma(\bar{K}_d-\bar{K})^{\mathrm{T}}\hat{B}^{\mathrm{T}}\bar{P}\hat{A}+\gamma\bar{K}^{\mathrm{T}}\hat{B}^{\mathrm{T}}\bar{P}\hat{B}\bar{K}-\gamma\bar{K}_d^{\mathrm{T}}\hat{B}^{\mathrm{T}}\bar{P}\hat{B}\bar{K}_d+\bar{K}^{\mathrm{T}}R\bar{K}-K_d^{\mathrm{T}}RK_d=0 \tag{61}$$

根据式 (41), 下面的等式成立

$$\gamma\hat{B}^{\mathrm{T}}\bar{P}\hat{A}=(R+\gamma\hat{B}^{\mathrm{T}}\bar{P}\hat{B})\bar{K} \tag{62}$$

将式 (62) 代入式 (61) 中, 得到:

$$\bar{K}^{\mathrm{T}}(R+\gamma\hat{B}^{\mathrm{T}}\bar{P}\hat{B})(\bar{K}_d-\bar{K})+\bar{K}^{\mathrm{T}}(R+\gamma\hat{B}^{\mathrm{T}}\bar{P}\hat{B})\bar{K}+(\bar{K}_d-\bar{K})^{\mathrm{T}}(R+\gamma\hat{B}^{\mathrm{T}}\bar{P}\hat{B})\bar{K}-\bar{K}_d^{\mathrm{T}}(R+\gamma\hat{B}^{\mathrm{T}}\bar{P}\hat{B})\bar{K}_d=0 \tag{63}$$

将式 (63) 展开后整理可得:

$$(\bar{K}_d-\bar{K})^{\mathrm{T}}(R+\gamma\hat{B}^{\mathrm{T}}\bar{P}\hat{B})(\bar{K}-\bar{K}_d)=0 \tag{64}$$

由于 $(R+\gamma\hat{B}^{\mathrm{T}}\bar{P}\hat{B})>0$, 从式 (64) 可以看出 $\bar{K}=\bar{K}_d$. □

定理 1 表明, 如果核矩阵 $\bar{P}$ 同时满足式 (42) 和式 (58), 则由逆学习得到的设定值增益 $\bar{K}$ 与演示系统隐含的最优设定值增益 $\bar{K}_d$ 是相等的, 这一结果为后续逆学习算法的设计提供了理论依据. 第 3.3.3 节提出一种带有通信补偿的数据驱动逆学习控制算法, 通过逆学习恢复一组与演示系统等价的性能指标权重参数, 使被控系统能够实现与演示系统一致的设定值策略.

接下来, 任意选择 $R>0$ 并固定, 修正更新状态惩罚权重系数 $\bar{S}$, 使得 $\bar{K}=\bar{K}_d$. 为此, 基于定理 1 提出一种带有通信补偿的模型驱动快慢耦合逆学习控制算法, 所提出的算法分为 3 个步骤:

步骤 1. 根据式 (58) 更新核矩阵 $\bar{P}$, 使得式 (41) 中的 $\bar{K}$ 接近式 (56) 中的 $\bar{K}_d$;

步骤 2. 利用式 (30) 与修正的核矩阵 $\bar{P}$ 更新设定值增益 $\bar{K}$;

步骤 3. 结合逆最优控制理论[35] 和式 (59) 修正更新满足条件的状态惩罚权重 $\bar{S}$.

为了实现这 3 个步骤, 算法 1 给出了带有通信补偿的模型驱动快慢耦合逆学习控制算法, 详细实现过程如 1) 中所述, 式 (65) 为式 (58) 的迭代形式, 用于更新核矩阵 $\bar{P}^{i+1}$. 然后, 如 3) 中所述, 结合逆最优控制的概念[35], 用式 (42) 的迭代形式更新 $\bar{S}^{i+1}$, 并引入标量参数 $\lambda\in(0,\,1]$, 得到式 (67) 的更新形式, 以保证参数更新的平滑性以及稳定性.

注 6. 在算法 1 中, 利用式 (65) 更新 $\bar{P}$, 目的是使式 (41) 计算出的 $\bar{K}$ 接近式 (56) 中的目标策略 $\bar{K}_d$; 式 (66) 基于最优控制理论, 利用更新后的 $\bar{P}$ 来求解最优设定值增益 $\bar{K}$; 式 (67) 是一个逆最优控制过程, 确保更新后的 $\bar{S}$ 满足逆最优控制条件. 从理论上看, 本文所采用的性能指标权重更新机制可视为逆最优控制在二次型性能指标情形下的一种实现形式.

算法 1. 带有通信补偿的模型驱动逆学习控制算法

初始化: 给定 $R>0$, 初始状态惩罚权重系数为 $\bar{S}^i=\Phi^{\mathrm{T}}S^i\Phi>0$, 选择小正常数 ϵ, 令 $i=0$.

1) 策略评估 (由式 (58) 计算核矩阵 $\bar{P}^{i+1}$):

$$\bar{P}^{i+1}=\bar{S}^i+\gamma(\hat{A}-\hat{B}\bar{K}_d)^{\mathrm{T}}\bar{P}^{i+1}(\hat{A}-\hat{B}\bar{K}_d)+\bar{K}_d^{\mathrm{T}}R\bar{K}_d \tag{65}$$

2) 策略提升 (由式 (30) 更新控制增益 $\bar{K}^{i+1}$):

$$\bar{K}^{i+1}=(R+\gamma\hat{B}^{\mathrm{T}}\bar{P}^{i+1}\hat{B})^{-1}\gamma\hat{B}^{\mathrm{T}}\bar{P}^{i+1}\hat{A} \tag{66}$$

3) 更新状态惩罚权重系数 $\bar{S}^{i+1}$:

$$\bar{S}^{i+1}=(1-\lambda)\bar{S}^i+\lambda\hat{S}^i \tag{67}$$

$$\hat{S}^i=\bar{P}^{i+1}-\gamma\hat{A}^{\mathrm{T}}\bar{P}^{i+1}\hat{A}+\gamma^2\hat{A}^{\mathrm{T}}\bar{P}^{i+1}\hat{B}(R+\gamma\hat{B}^{\mathrm{T}}\bar{P}^{i+1}\hat{B})^{-1}\hat{B}^{\mathrm{T}}\bar{P}^{i+1}\hat{A} \tag{68}$$

其中, $0<\lambda\le 1$ 为学习步长.

4) 如果 $\|\bar{K}^{i+1}-\bar{K}^i\|<\epsilon$, 结束; 否则, $i=i+1$, 返回 1).

3.3.3 带有通信补偿的数据驱动逆学习控制算法

第 3.3.2 节基于模型的逆学习控制算法需要了解系统动力学 $\hat{A}$ 和 $\hat{B}$ 以及设定值增益 $\bar{K}_d$, 这在实际的矿物磨选过程中是难以实现的. 因此, 在上述基于模型的算法基础上提出一种数据驱动的逆学习控制算法, 推导过程如下:

将式 (65) 两边左乘 $\nu_d^{\mathrm{T}}(k)$, 右乘 $\nu_d(k)$, 可以得到:

$$\nu_d^{\mathrm{T}}(k)\bar{S}^i\nu_d(k)+\nu_d^{\mathrm{T}}(k)\bar{K}_d^{\mathrm{T}}R\bar{K}_d\nu_d(k)-\nu_d^{\mathrm{T}}(k)\bar{P}^{i+1}\nu_d(k)+\gamma\nu_d^{\mathrm{T}}(k)(\hat{A}-\hat{B}\bar{K}_d)^{\mathrm{T}}\bar{P}^{i+1}(\hat{A}-\hat{B}\bar{K}_d)\nu_d(k)=0 \tag{69}$$

根据式 (52), 式 (69) 可以写成如下形式:

$$\nu_d^{\mathrm{T}}(k)\bar{S}^i\nu_d(k)+y_d^{*\mathrm{T}}(k)Ry_d^*(k)-\nu_d^{\mathrm{T}}(k)\bar{P}^{i+1}\nu_d(k)+\gamma\nu_d^{\mathrm{T}}(k+1)\bar{P}^{i+1}\nu_d(k+1)=0 \tag{70}$$

结合式 (70), Q 函数可定义为如下形式:

$$
\begin{aligned}
&Q(\nu_d,\ y_d^*,\ \bar{P}^{i+1}) = \nu_d^{\mathrm{T}}(k)\bar{P}^{i+1}\nu_d(k) = \\
&\quad \nu_d^{\mathrm{T}}(k)\bar{S}^i\nu_d(k) + y_d^{*\mathrm{T}}(k)Ry_d^*(k) - \nu_d^{\mathrm{T}}(k)\bar{P}^{i+1}\nu_d(k) + \\
&\quad \gamma\nu_d^{\mathrm{T}}(k+1)\bar{P}^{i+1}\nu_d(k+1) = \\
&\quad \nu_d^{\mathrm{T}}(k)\bar{S}^i\nu_d(k) + y_d^{*\mathrm{T}}(k)Ry_d^*(k) - \nu_d^{\mathrm{T}}(k)\bar{P}^{i+1}\nu_d(k) + \\
&\quad \gamma(\hat{A}\nu_d(k) + \hat{B}y_d^*(k))^{\mathrm{T}}\bar{P}^{i+1}(\hat{A}\nu_d(k) + \hat{B}y_d^*(k)) = \\
&\quad \begin{bmatrix}\nu_d(k)\\ y_d^*(k)\end{bmatrix}^{\mathrm{T}} H^{i+1}\begin{bmatrix}\nu_d(k)\\ y_d^*(k)\end{bmatrix} = \\
&\quad \begin{bmatrix}\nu_d(k)\\ y_d^*(k)\end{bmatrix}^{\mathrm{T}}\begin{bmatrix}H_{\nu\nu}^{i+1} & H_{\nu y^*}^{i+1}\\ H_{y^*\nu}^{i+1} & H_{y^*y^*}^{i+1}\end{bmatrix}\begin{bmatrix}\nu_d(k)\\ y_d^*(k)\end{bmatrix}
\end{aligned}
\tag{71}
$$

其中

$$
\begin{aligned}
H_{\nu\nu}^{i+1} &= \bar{S}^i + \gamma\hat{A}^{\mathrm{T}}\bar{P}^{i+1}\hat{A}\\
H_{\nu y^*}^{i+1} &= (H_{y^*\nu}^{i+1})^{\mathrm{T}} = \gamma\hat{A}^{\mathrm{T}}\bar{P}^{i+1}\hat{B}\\
H_{y^*y^*}^{i+1} &= R + \gamma\hat{B}^{\mathrm{T}}\bar{P}^{i+1}\hat{B}
\end{aligned}
$$

定义 $Z_d(k) = \begin{bmatrix}\nu_d(k)\\ y_d^*(k)\end{bmatrix}$, 式 (71) 可以写为:

$$
Q(\nu_d,\ y_d^*,\ \bar{P}^{i+1}) = Z_d^{\mathrm{T}}(k)H^{i+1}Z_d(k) \tag{72}
$$

根据极值必要条件, 令 $\frac{\partial Q}{\partial y_d^*} = 0$, 式 (66) 中的 $\bar{K}^{i+1}$ 可以表示为:

$$
\bar{K}^{i+1} = (H_{y^*y^*}^{i+1})^{-1}H_{y^*\nu}^{i+1} \tag{73}
$$

将式 (71) 代入 (70) 中, H^{i+1} 可由下式计算得到:

$$
\begin{aligned}
&Z_d^{\mathrm{T}}(k)H^{i+1}Z_d(k) = \nu_d^{\mathrm{T}}(k)\bar{S}^i\nu_d(k) + y_d^{*\mathrm{T}}(k)Ry_d^*(k) + \\
&\quad \gamma Z_d^{\mathrm{T}}(k+1)H^{i+1}Z_d(k+1)
\end{aligned}
\tag{74}
$$

为得到 $\bar{S}^{i+1}$ 的数据驱动计算方法, 式 (67) 可以重新写为如下形式:

$$
\begin{aligned}
\bar{S}^{i+1} = (1-\lambda)\bar{S}^i + \lambda(&\bar{P}^{i+1} - (\bar{K}^{i+1})^{\mathrm{T}}R\bar{K}^{i+1} - \\
&\gamma(\hat{A} - \hat{B}\bar{K}^{i+1})^{\mathrm{T}}\bar{P}(\hat{A} - \hat{B}\bar{K}^{i+1}))
\end{aligned}
\tag{75}
$$

式 (75) 两边同时左乘 $\nu^{\mathrm{T}}(k)$, 右乘 $\nu(k)$, 结合式 (37), 可以得到:

$$
\begin{aligned}
&\nu^{\mathrm{T}}(k)\bar{S}^{i+1}\nu(k) = \\
&\quad (1-\lambda)\nu^{\mathrm{T}}(k)\bar{S}^i\nu(k) + \lambda(\nu^{\mathrm{T}}(k)\bar{P}^{i+1}\nu(k) - \\
&\quad y^{*\mathrm{T}}(k)Ry^*(k) - \gamma\nu^{\mathrm{T}}(k+1)\bar{P}^{i+1}\nu(k+1))
\end{aligned}
\tag{76}
$$

其中, $y^*(k) = -\bar{K}^{i+1}\nu(k)$. 由于 H^{i+1} 已知, 而 $\bar{P}^{i+1}$ 未知, 要使用式 (76) 更新 $\bar{S}^{i+1}$, 需要先用 H^{i+1} 评估式 (76) 中含有 $\bar{P}^{i+1}$ 的项.

为此, 将式 (37) 代入 $\gamma\nu^{\mathrm{T}}(k+1)\bar{P}^{i+1}\nu(k+1)$, 可以得到:

$$
\begin{aligned}
&\gamma\nu^{\mathrm{T}}(k+1)\bar{P}^{i+1}\nu(k+1) = \\
&\quad \gamma(\hat{A}\nu(k) + \hat{B}y^*(k))^{\mathrm{T}}\bar{P}^{i+1}(\hat{A}\nu(k) + \hat{B}y^*(k)) = \\
&\quad \begin{bmatrix}\nu(k)\\ y^*(k)\end{bmatrix}^{\mathrm{T}} H_L^{i+1}\begin{bmatrix}\nu(k)\\ y^*(k)\end{bmatrix} = \\
&\quad \begin{bmatrix}\nu(k)\\ y^*(k)\end{bmatrix}^{\mathrm{T}}\begin{bmatrix}\gamma\hat{A}^{\mathrm{T}}\bar{P}^{i+1}\hat{A} & \gamma\hat{A}^{\mathrm{T}}\bar{P}^{i+1}\hat{B}\\ \gamma\hat{B}^{\mathrm{T}}\bar{P}^{i+1}\hat{A} & \gamma\hat{B}^{\mathrm{T}}\bar{P}^{i+1}\hat{B}\end{bmatrix}\begin{bmatrix}\nu(k)\\ y^*(k)\end{bmatrix}
\end{aligned}
\tag{77}
$$

定义 $Z(k) = \begin{bmatrix}\nu(k)\\ y^*(k)\end{bmatrix}$, 式 (77) 可以写成:

$$
\gamma\nu^{\mathrm{T}}(k+1)\bar{P}^{i+1}\nu(k+1) = Z^{\mathrm{T}}(k)H_L^{i+1}Z(k) \tag{78}
$$

同样地, 将式 (37) 代入 $\nu^{\mathrm{T}}(k)\bar{P}^{i+1}\nu(k)$ 中, 可得:

$$
\begin{aligned}
&\nu^{\mathrm{T}}(k)\bar{P}^{i+1}\nu(k) = \\
&\quad (\hat{A}\nu(k-1) + \hat{B}y^*(k-1))^{\mathrm{T}}\bar{P}^{i+1}(\hat{A}\nu(k-1) + \\
&\quad \hat{B}y^*(k-1)) = \\
&\quad \frac{1}{\gamma}\begin{bmatrix}\nu(k-1)\\ y^*(k-1)\end{bmatrix}^{\mathrm{T}} H_L^{i+1}\begin{bmatrix}\nu(k-1)\\ y^*(k-1)\end{bmatrix} = \\
&\quad \frac{1}{\gamma}Z^{\mathrm{T}}(k-1)H_L^{i+1}Z(k-1)
\end{aligned}
\tag{79}
$$

结合式 (71), 下面的等式成立:

$$
H_L^{i+1} = H^{i+1} - \begin{bmatrix}\bar{S}^i & 0\\ 0 & R\end{bmatrix} \tag{80}
$$

因此, 根据式 (77) ~ (80), 式 (76) 可以表示为:

$$
\begin{aligned}
&\nu(k)\bar{S}^{i+1}\nu(k) = \\
&\quad (1-\lambda)\nu^{\mathrm{T}}(k)\bar{S}^i\nu(k) - \lambda y^{*\mathrm{T}}(k)Ry^*(k) + \\
&\quad \lambda Z^{\mathrm{T}}(k-1)H_L^{i+1}Z(k-1) - \\
&\quad \frac{\lambda}{\gamma}Z^{\mathrm{T}}(k)H_L^{i+1}Z(k)
\end{aligned}
\tag{81}
$$

注 7. 由式 (34) 和 (39) 可知, $\bar{S}^{i+1}$ 为丢包发生时性能指标权重参数, 其中包含 $\hat{A}$、$\hat{B}$ 的信息. 虽然 $\hat{A}$、$\hat{B}$ 在实际应用中无法获取, 但是其维度已知. 只需根据其维度信息即可给出 $\bar{S}^{i+1}$ 初始值, 进而实现所提出的设定值逆学习控制算法. 因此, 该问题不影响所提出算法的实际运行.

式 (73)、(74) 和 (81) 构成了数据驱动快慢耦合

逆学习控制算法, 无需了解矿物磨选过程系统动态信息. 基于上述推导, 带有通信补偿的数据驱动逆学习控制算法如算法 2 所示.

算法 2. 带有通信补偿的数据驱动逆学习控制算法

初始化: 给定 $R>0$, 初始状态惩罚权重系数为 $\bar{S}^i=\Phi^{\mathrm{T}}S^i\Phi>0$, 选择小正常数$\epsilon$, 令$i=0$.

1) 策略评估: 利用式 (74) 计算 H^{i+1};

2) 策略提升: 利用式 (73) 更新设定值增益 $\bar{K}^{i+1}$;

3) 性能指标权重参数更新校正: 利用式 (80) 计算 H_L^{i+1}, 进而利用式 (81) 更新校正 $\bar{S}^{i+1}$;

4) 如果 $\|\bar{K}^{i+1}-\bar{K}^i\|<\epsilon$, 结束; 否则, $i=i+1$, 返回 1).

3.3.4 算法复杂度分析

本文所提出控制方法的在线计算复杂度主要来源于基础回路控制计算以及运行层逆学习算法 (算法 2). 运行层控制周期为 $T=\zeta T_0$, 其中基础回路在一个运行层周期内执行 ζ 次. 在一个运行层周期内, 基础回路解析 MPC 控制律的在线计算复杂度为 $\mathrm{O}(\zeta N_u N_y)$. 对于算法 2, 单次迭代的计算量主要由策略评估、策略提升以及性能指标权重更新构成, 其计算复杂度分别为 $\mathrm{O}(Nd^2+d^3)$、$\mathrm{O}(n_y^3+n_y^2n_\nu)$、$\mathrm{O}(Nd^2)$, 其中 N 表示求解 H^i 所需的样本数, $d=n_\nu+n_y$, n_y 为回路设定值向量维度. 因此, 忽略低阶项后整体计算复杂度可近似表示为 $\mathrm{O}(\zeta N_u N_y+Nd^2+d^3)$. 基础回路控制律为解析形式, 仅涉及低维矩阵运算, 其计算复杂度较低, 能够在 PLC 等工业控制器的快速采样周期内完成. 算法 2 所涉及的矩阵运算复杂度均为多项式量级, 结合工业控制器或上位机的计算能力, 所提出方法能够满足在线实时运行要求.

3.4 控制性能分析

本节将对上述提出方法进行收敛性、稳定性以及逆最优性的理论分析.

定理 2. 给定初始状态惩罚权重 $\bar{S}^0>0$, 算法 1 中 $\bar{S}^i$、$\bar{P}^i$、$\bar{K}^i$ 收敛至 $\bar{S}^\infty$、$\bar{P}^\infty$、$\bar{K}^\infty$ 且满足

$$\begin{aligned}&\bar{S}^\infty-\bar{P}^\infty+\gamma\hat{A}^{\mathrm{T}}\bar{P}^\infty\hat{A}-\gamma^2\hat{A}^{\mathrm{T}}\bar{P}^\infty\hat{B}(R+\\&\gamma\hat{B}^{\mathrm{T}}\bar{P}^\infty\hat{B})^{-1}\hat{B}^{\mathrm{T}}\bar{P}^\infty\hat{A}=0\end{aligned}\tag{82}$$

$$\bar{K}^\infty=\left(R+\gamma\hat{B}^{\mathrm{T}}\bar{P}^\infty\hat{B}\right)^{-1}\gamma\hat{B}^{\mathrm{T}}\bar{P}^\infty\hat{A}=\bar{K}_d\tag{83}$$

证明. 首先证明算法 1 的收敛性. 式 (65) 和式 (68) 可写为如下形式:

$$\bar{S}^i=\bar{P}^{i+1}-\gamma(\hat{A}-\hat{B}\bar{K}_d)^{\mathrm{T}}\bar{P}^{i+1}(\hat{A}-\hat{B}\bar{K}_d)-\bar{K}_d^{\mathrm{T}}R\bar{K}_d\tag{84}$$

$$\begin{aligned}\hat{S}^i=&\bar{P}^{i+1}-\gamma\hat{A}^{\mathrm{T}}\bar{P}^{i+1}\hat{A}+\\&\gamma^2\hat{A}^{\mathrm{T}}\bar{P}^{i+1}\hat{B}(R+\gamma\hat{B}^{\mathrm{T}}\bar{P}^{i+1}\hat{B})^{-1}\hat{B}^{\mathrm{T}}\bar{P}^{i+1}\hat{A}\end{aligned}\tag{85}$$

将式 (85) 减去式 (84) 可以得到:

$$\begin{aligned}\hat{S}^i-\bar{S}^i=&\gamma^2\hat{A}^{\mathrm{T}}\bar{P}^{i+1}\hat{B}(R+\\&\gamma\hat{B}^{\mathrm{T}}\bar{P}^{i+1}\hat{B})^{-1}\hat{B}^{\mathrm{T}}\bar{P}^{i+1}\hat{A}-\\&\gamma\hat{A}^{\mathrm{T}}\bar{P}^{i+1}\hat{B}\bar{K}_d-\gamma\bar{K}_d^{\mathrm{T}}\hat{B}^{\mathrm{T}}\bar{P}^{i+1}\hat{A}+\\&\gamma\bar{K}_d^{\mathrm{T}}\hat{B}^{\mathrm{T}}\bar{P}^{i+1}\hat{B}\bar{K}_d+\bar{K}_d^{\mathrm{T}}R\bar{K}_d=\\&(\bar{K}_d-(R+\gamma\hat{B}^{\mathrm{T}}\bar{P}^{i+1}\hat{B})^{-1}\hat{B}^{\mathrm{T}}\bar{P}^{i+1}\hat{A})^{\mathrm{T}}\times\\&(R+\gamma\hat{B}^{\mathrm{T}}\bar{P}^{i+1}\hat{B})\times\\&(\bar{K}_d-(R+\gamma\hat{B}^{\mathrm{T}}\bar{P}^{i+1}\hat{B})^{-1}\hat{B}^{\mathrm{T}}\bar{P}^{i+1}\hat{A})=\\&(\bar{K}_d-\bar{K}^{i+1})^{\mathrm{T}}(R+\gamma\hat{B}^{\mathrm{T}}\bar{P}^{i+1}\hat{B})(\bar{K}_d-\bar{K}^{i+1})\end{aligned}$$

然后, 将式 (86) 代入式 (67), 得到:

$$\bar{S}^{i+1}=\bar{S}^i+\lambda\Lambda^i\tag{86}$$

其中

$$\begin{aligned}\Lambda^i=&(\bar{K}_d-\bar{K}^{i+1})^{\mathrm{T}}(R+\gamma\hat{B}^{\mathrm{T}}\bar{P}^{i+1}\hat{B})\times\\&(\bar{K}_d-\bar{K}^{i+1})\geq0\end{aligned}\tag{87}$$

由式 (86) 可知 $\bar{S}^i$ 是单调递增序列. 接下来, 通过选择合适的 $\lambda\in(0,\,1]$ 说明增量 Λ^i 随着 $\bar{S}^i$ 的增大而减小. 根据文献 [35], 基于逆最优控制的特性, 存在无穷组不同于 $\bar{S}_d$、R_d 的 $\bar{S}\geq0$、R 使得 $\bar{K}_d$ 为最优. 给定 $R>0$ 并固定, 如果初始权重 $\bar{S}^0=\bar{S}$, 那么由式 (66) 可以得到 $\bar{K}^1=\bar{K}_d$, 此时迭代停止. 如果 $\bar{S}^0\leq\bar{S}$, 那么 $\bar{K}^1\neq\bar{K}_d$, 通过选择适当的 λ, 可以得到 $\bar{S}^0\leq\bar{S}^1\leq\bar{S}$, 此时 $\bar{S}^1$ 相比 $\bar{S}^0$ 更接近 $\bar{S}$, 这也意味着 $\bar{K}^1$ 更接近 $\bar{K}_d$, 即 $0\leq\Lambda^2\leq\Lambda^1$. 以此类推, 随着 $i\to\infty$, 由式 (86) 得到 $\bar{S}^{i+1}\to\bar{S}^i\to\bar{S}$, 同时意味着 $\bar{P}^{i+1}\to\bar{P}^i$ 和 $\bar{K}^{i+1}\to\bar{K}^i$.

接下来对收敛解进行分析. 式 (67) 和 (68) 可以写为如下形式:

$$\begin{aligned}\bar{S}^i=&(1-\lambda)\bar{S}^{i-1}+\lambda(\bar{P}^i-\gamma\hat{A}^{\mathrm{T}}\bar{P}^i\hat{A}+\\&\gamma^2\hat{A}^{\mathrm{T}}\bar{P}^i\hat{B}(R+\gamma\hat{B}^{\mathrm{T}}\bar{P}^i\hat{B})^{-1}\hat{B}^{\mathrm{T}}\bar{P}^i\hat{A})\end{aligned}\tag{88}$$

将式 (88) 代入式 (65), 可以得到:

$$\begin{aligned}&(1-\lambda)\bar{S}^{i-1}+\lambda(\bar{P}^i-\gamma\hat{A}^{\mathrm{T}}\bar{P}^i\hat{A}+\\&\gamma^2\hat{A}^{\mathrm{T}}\bar{P}^i\hat{B}(R+\gamma\hat{B}^{\mathrm{T}}\bar{P}^i\hat{B})^{-1}\hat{B}^{\mathrm{T}}\bar{P}^i\hat{A})=\\&\bar{P}^{i+1}-\gamma(\hat{A}-\hat{B}\bar{K}_d)^{\mathrm{T}}\bar{P}^{i+1}(\hat{A}-\hat{B}\bar{K}_d)-\bar{K}_d^{\mathrm{T}}R\bar{K}_d\end{aligned}\tag{89}$$

定义 $\bar{P}^{i+1}=\bar{P}^{i}=\bar{P}^{\infty}$, $\bar{S}^{i+1}=\bar{S}^{i}=\bar{S}^{\infty}$, 式 (89) 可以写为如下形式:

$$\begin{aligned}&(1-\lambda)(\bar{S}^{\infty}+\bar{P}^{\infty}-\gamma\hat{A}^{\mathrm{T}}\bar{P}^{\infty}\hat{A}+\\&\gamma^{2}\hat{A}^{\mathrm{T}}\bar{P}^{\infty}\hat{B}(R+\gamma\hat{B}^{\mathrm{T}}\bar{P}^{\infty}\hat{B})^{-1}\hat{B}^{\mathrm{T}}\bar{P}^{\infty}\hat{A})=\\&\gamma\hat{A}^{\mathrm{T}}\bar{P}^{\infty}\hat{B}\bar{K}_{d}+\gamma\bar{K}_{d}^{\mathrm{T}}\hat{B}^{\mathrm{T}}\bar{P}^{\infty}\hat{A}-\\&\gamma\bar{K}_{d}^{\mathrm{T}}\hat{B}^{\mathrm{T}}\bar{P}^{\infty}\hat{B}\bar{K}_{d}-\bar{K}_{d}^{\mathrm{T}}R\bar{K}_{d}-\\&\gamma^{2}\hat{A}^{\mathrm{T}}\bar{P}^{\infty}\hat{B}(R+\gamma\hat{B}^{\mathrm{T}}\bar{P}^{\infty}\hat{B})^{-1}\hat{B}^{\mathrm{T}}\bar{P}^{\infty}\hat{A}=\\&-(\bar{K}_{d}-\bar{K}^{\infty})^{\mathrm{T}}(R+\gamma\hat{B}^{\mathrm{T}}\bar{P}^{\infty}\hat{B})(\bar{K}_{d}-\bar{K}^{\infty})\end{aligned}\tag{90}$$

其中

$$\bar{K}^{\infty}=\left(R+\gamma\hat{B}^{\mathrm{T}}\bar{P}^{\infty}\hat{B}\right)^{-1}\gamma\hat{B}^{\mathrm{T}}\bar{P}^{\infty}\hat{A}\tag{91}$$

由于收敛解满足式 (67) 和 (68), 因此可以得到:

$$\begin{aligned}&\bar{S}^{\infty}-\bar{P}^{\infty}+\gamma\hat{A}^{\mathrm{T}}\bar{P}^{\infty}\hat{A}-\\&\gamma^{2}\hat{A}^{\mathrm{T}}\bar{P}^{\infty}\hat{B}(R+\gamma\hat{B}^{\mathrm{T}}\bar{P}^{\infty}\hat{B})^{-1}\hat{B}^{\mathrm{T}}\bar{P}^{\infty}\hat{A}=0\end{aligned}\tag{92}$$

因为 $R+\gamma\hat{B}^{\mathrm{T}}\bar{P}^{\infty}\hat{B}>0$, 根据式 (91) 推断出 $\bar{K}_{d}=\bar{K}^{\infty}$. □

定理 3. 给定 $R>0$ 和初始状态惩罚权重 $\bar{S}^{0}>0$, 算法 1 中的每一次迭代 $i=0,\ 1,\ \cdots$ 得到的 $\bar{P}^{i}$ 和 $\bar{S}^{i}$ 都能在控制输入 (66) 的作用下使系统 (24) 稳定.

证明. 给定初始 $\bar{S}^{0}>0$, 由式 (86) 可知, 对于 $i=0,\ 1,\ \cdots,\ \bar{S}^{i+1}\geq 0$ 始终成立. 因此可以得到对于任意的 $\lambda\in(0,\ 1]$, $\bar{S}^{i+1}-(1-\lambda)\bar{S}^{i}\geq 0$ 成立.

由于 $\bar{P}^{i+1}$ 满足式 (65), 并且 $\bar{S}^{i}\geq 0$, 根据式 (67)、(68) 和文献 [33] 可以得到:

$$\begin{aligned}&\bar{P}^{i+1}-\gamma\hat{A}^{\mathrm{T}}\bar{P}^{i+1}\hat{A}+\gamma^{2}\hat{A}^{\mathrm{T}}\bar{P}^{i+1}\hat{B}(R+\\&\gamma\hat{B}^{\mathrm{T}}\bar{P}^{i+1}\hat{B})^{-1}\hat{B}^{\mathrm{T}}\bar{P}^{i+1}\hat{A}\geq 0\end{aligned}$$

其中, $(\bar{P}^{i+1})^{\mathrm{T}}=\bar{P}^{i+1}>0$.

定义新的丢包状态向量 $\bar{\nu}(k)=\gamma^{\frac{k}{2}}\nu(k)$. 根据式 (38), 有如下值函数:

$$V^{i+1}(\bar{\nu}(k))=\bar{\nu}^{\mathrm{T}}(k)\bar{P}^{i+1}\bar{\nu}(k)\tag{93}$$

根据式 (66)、(93) 可以得到:

$$\begin{aligned}&V^{i+1}(\bar{\nu}(k+1))-V^{i+1}(\bar{\nu}(k))=\\&\bar{\nu}^{\mathrm{T}}(k+1)\bar{P}^{i+1}\bar{\nu}(k+1)-\bar{\nu}^{\mathrm{T}}(k)\bar{P}^{i+1}\bar{\nu}(k)\end{aligned}\tag{94}$$

利用 $\nu(k)=\gamma^{-\frac{k}{2}}\bar{\nu}(k)$ 和式 (24) 中的设定值 (41), 可得:

$$\begin{aligned}&\bar{\nu}(k+1)=\gamma^{\frac{1}{2}}\hat{A}\bar{\nu}(k)+\gamma^{\frac{1}{2}}\hat{B}\bar{y}^{*}(k)=\\&\gamma^{\frac{1}{2}}(\hat{A}-\hat{B}(R+\gamma\hat{B}^{\mathrm{T}}\bar{P}^{i+1}\hat{B})^{-1}\hat{B}^{\mathrm{T}}\bar{P}^{i+1}\hat{A})\bar{\nu}(k)\end{aligned}\tag{95}$$

其中

$$\bar{y}^{*}(k)=-\bar{K}^{i+1}\bar{\nu}(k)\tag{96}$$

将式 (95) 代入式 (94), 整理可得:

$$\begin{aligned}&V^{i+1}(\bar{\nu}(k+1))-V^{i+1}(\bar{\nu}(k))=\\&\bar{\nu}^{\mathrm{T}}(k)\big[-\bar{P}^{i+1}+\gamma\hat{A}^{\mathrm{T}}\bar{P}^{i+1}\hat{A}-(\bar{K}^{i+1})^{\mathrm{T}}R\bar{K}^{i+1}-\\&\gamma^{2}\hat{A}^{\mathrm{T}}\bar{P}^{i+1}\hat{B}(R+\gamma\hat{B}^{\mathrm{T}}\bar{P}^{i+1}\hat{B})^{-1}\hat{B}^{\mathrm{T}}\bar{P}^{i+1}\hat{A}\big]\bar{\nu}(k)\end{aligned}\tag{97}$$

根据式 (67) 和 (68) 可以得到:

$$\begin{aligned}&-\bar{P}^{i+1}+\gamma\hat{A}^{\mathrm{T}}\bar{P}^{i+1}\hat{A}-\gamma^{2}\hat{A}^{\mathrm{T}}\bar{P}^{i+1}\hat{B}(R+\\&\gamma\hat{B}^{\mathrm{T}}\bar{P}^{i+1}\hat{B})^{-1}\hat{B}^{\mathrm{T}}\bar{P}^{i+1}\hat{A}=\\&-\frac{1}{\lambda}(\bar{S}^{i+1}-(1-\lambda)\bar{S}^{i})\end{aligned}\tag{98}$$

那么, 式 (97) 可以写成如下形式:

$$\begin{aligned}&V^{i+1}(\bar{\nu}(k+1))-V^{i+1}(\bar{\nu}(k))=\\&\bar{\nu}^{\mathrm{T}}(k)\Big[-\frac{1}{\lambda}(\bar{S}^{i+1}-(1-\lambda)\bar{S}^{i})-\\&(\bar{K}^{i+1})^{\mathrm{T}}R\bar{K}^{i+1}\Big]\bar{\nu}(k)\end{aligned}\tag{99}$$

由于 $\bar{S}^{i+1}-(1-\lambda)\bar{S}^{i}\geq 0$ 且 $R>0$, 可以得到 $V^{i+1}(\bar{\nu}(k+1))-V^{i+1}(\bar{\nu}(k))<0$. □

定理 4. 对于由式 (67) 更新得到的 $\bar{S}^{i+1}$, 考虑对应的性能指标函数:

$$\begin{aligned}&V^{i+1}(\bar{\nu}(k))=\gamma^{-k}\frac{1}{2}\sum_{j=k}^{\infty}\big[\bar{\nu}^{\mathrm{T}}(j)\bar{S}^{i+1}\bar{\nu}(j)+\\&\bar{y}^{*\mathrm{T}}(j)R\bar{y}^{*}(j)\big]\end{aligned}\tag{100}$$

则控制输入 (96) 是使该性能指标函数取得最小值的最优控制输入.

证明. 根据 $\bar{\nu}(k)$ 和 $\bar{y}^{*}(k)$, 式 (38) 可以写成:

$$\begin{aligned}&V^{i+1}(\bar{\nu}(k))=\gamma^{-k}\frac{1}{2}\sum_{j=k}^{\infty}\big[\bar{\nu}^{\mathrm{T}}(j)\bar{S}^{i+1}\bar{\nu}(j)+\\&\bar{y}^{*\mathrm{T}}(j)R\bar{y}^{*}(j)\big]\end{aligned}\tag{101}$$

事实上, 最小化系统 (36) 的性能指标函数 (38) 等价于最小化系统 (95) 的性能指标函数 (101). 考虑如下等式:

$$\begin{aligned}&\gamma^{-k}\sum_{j=k}^{\infty}\big[V^{i+1}(\bar{\nu}(j+1))-V^{i+1}(\bar{\nu}(j))\big]=\\&\gamma^{-k}\sum_{j=k}^{\infty}\big[(\hat{A}\bar{\nu}(j)+\hat{B}\bar{y}^{*}(j))^{\mathrm{T}}\bar{P}^{i+1}(\hat{A}\bar{\nu}(j)+\end{aligned}$$

$$\hat{B}\bar{y}^*(j)) - \bar{\nu}(j)^{\mathrm{T}}\bar{P}^{i+1}\bar{\nu}(j)\big] =$$
$$\gamma^{-k}\sum_{j=k}^{\infty}\big[\bar{\nu}^{\mathrm{T}}(j)(\hat{A}^{\mathrm{T}}\bar{P}^{i+1}\hat{A} - \bar{P}^{i+1})\bar{\nu}(j) +$$
$$\bar{y}^{*\mathrm{T}}(j)\hat{B}^{\mathrm{T}}\bar{P}^{i+1}\hat{A}\bar{\nu}(j) + \bar{\nu}^{\mathrm{T}}(j)\hat{A}^{\mathrm{T}}\bar{P}^{i+1}\hat{B}\bar{y}^*(j) +$$
$$\bar{y}^{*\mathrm{T}}(j)\hat{B}^{\mathrm{T}}\bar{P}^{i+1}\hat{B}\bar{y}^*(j)\big] \tag{102}$$

注意到

$$\gamma^{-k}\sum_{j=k}^{\infty}\big[V^{i+1}(\bar{\nu}(j+1)) - V^{i+1}(\bar{\nu}(j))\big] =$$
$$\bar{\nu}^{\mathrm{T}}(\infty)\bar{P}^{i+1}\bar{\nu}(\infty) - \bar{\nu}^{\mathrm{T}}(k)\bar{P}^{i+1}\bar{\nu}(k) \tag{103}$$

在式 (102) 和式 (103) 中使用 $\bar{\nu}(\infty) = 0$ 和式 (67) 得出:

$$\gamma^{-k}\sum_{j=k}^{\infty}\big[\bar{\nu}^{\mathrm{T}}(j)(-\bar{S}^{i+1} + \hat{A}^{\mathrm{T}}\bar{P}^{i+1}\hat{B}(R +$$
$$\hat{B}^{\mathrm{T}}\bar{P}^{i+1}\hat{B})^{-1}\hat{B}^{\mathrm{T}}\bar{P}^{i+1}\hat{A})\bar{\nu}(j) +$$
$$\gamma\bar{y}^{*\mathrm{T}}(j)\hat{B}^{\mathrm{T}}\bar{P}^{i+1}\hat{A}\bar{\nu}(j) + \gamma\bar{\nu}^{\mathrm{T}}(j)\hat{A}^{\mathrm{T}}\bar{P}^{i+1}\hat{B}\bar{y}^*(j) +$$
$$\gamma\bar{y}^{*\mathrm{T}}(j)\hat{B}^{\mathrm{T}}\bar{P}^{i+1}\hat{B}\bar{y}^*(j)\big] + \bar{\nu}^{\mathrm{T}}(k)\bar{P}^{i+1}\bar{\nu}(k) = 0 \tag{104}$$

其中, $\bar{S}^{i+1}$ 如式 (68) 所示. 将式 (104) 的两边乘以 λ, 并将结果加到式 (101) 的右边, 然后进行完全平方运算并结合式 (100), 整理可以得到:

$$V^{i+1}(\bar{\nu}(k)) = \gamma^{-k}\sum_{j=k}^{\infty}\big[\bar{\nu}^{\mathrm{T}}(j)\bar{S}^{i+1}\bar{\nu}(j) +$$
$$\bar{y}^{*\mathrm{T}}(j)R\bar{y}^*(j)\big] + \lambda\gamma^{-k}\sum_{j=k}^{\infty}\big[(\bar{y}^*(j) + (R +$$
$$\gamma\hat{B}^{\mathrm{T}}\bar{P}^{i+1}\hat{B})^{-1}\gamma\hat{B}^{\mathrm{T}}\bar{P}^{i+1}\hat{A}\bar{\nu}(j))^{\mathrm{T}}(R +$$
$$\gamma B^{\mathrm{T}}P^{i+1}B)(\bar{y}^*(j) + (R +$$
$$\gamma\hat{B}^{\mathrm{T}}\bar{P}^{i+1}\hat{B})^{-1}\gamma\hat{B}^{\mathrm{T}}\bar{P}^{i+1}\hat{A}\bar{\nu}(j))\big] \tag{105}$$

由于 $\lambda \in (0,\ 1]$, $\bar{S}^{i+1} > 0$, 且 $R > 0$, 当 $\bar{y}^*(k) = -\bar{K}^{i+1}\bar{\nu}(k)$ 时, 式 (101) 达到最小值, 其中 $\bar{K}^{i+1}$ 由式 (66) 给出. □

4 仿真研究

针对矿物磨选过程 (1) 和 (2), 对本文提出的带有通信补偿的多速率分层学习控制方法进行仿真研究. 本节首先给出算法 2 的仿真研究验证其控制效果, 然后进行 2 组对比仿真来进一步说明算法 2 的优越性.

4.1 仿真研究参数选择

根据文献 [21], 结合实际的矿物磨选过程, 基础回路层和运行层动态可由状态空间模型 (106) 和 (107) 近似模拟.

对两层连续时间状态空间模型 (106) 和 (107) 分别以基周期 h 和运行周期 T 离散化后得到离散化形式. 折扣因子设为 $\gamma = 0.9$. 演示系统的隐含性能指标权重为 $Q_d = \begin{bmatrix} 2\,000 & 0 \\ 0 & 2\,000 \end{bmatrix}$、$R_d = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, 性能指标参数 $\bar{S}_d$ 可以根据式 (54) 计算得到. 上述参数仅用于获取演示运行数据, 其在控制器设计时并未使用.

被控系统性能指标 (3) 的初始权重参数选择为 $Q^0 = \begin{bmatrix} 750 & 0 \\ 0 & 100 \end{bmatrix}$、$R = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}$, 初始 $\bar{S}^0$ 可通过计算获得.

$$\begin{cases} \dot{x}_f(t) = \begin{bmatrix} -0.065\,4 & 0 \\ 0 & -0.082\,3 \end{bmatrix} x_f(t) + \begin{bmatrix} 0.145\,7 & 0 \\ 0 & 0.180\,7 \end{bmatrix} u(t) \\ y(t) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} x_f(t) \end{cases} \tag{106}$$

$$\begin{cases} \dot{x}_s(t) = \begin{bmatrix} -0.286 & 0 & 0 & 0 & 0 \\ 0 & -0.313 & 0 & 0 & 0 \\ 0 & 0 & -1.500 & -0.500 & 0 \\ 0 & 0 & 1.000 & 0 & 0 \\ 0 & 0 & 0 & 0 & -0.182 \end{bmatrix} x_s(t) + \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} y(t) \\ r(t) = \begin{bmatrix} -0.114 & 0.250 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.010 & 0.455 \end{bmatrix} x_s(t) \end{cases} \tag{107}$$

本文仿真研究的矿物磨选运行优化控制目标是控制矿物磨选粒度 r_1 尽可能接近 70%, 循环负荷 r_2 尽可能接近 150 t/h. 此外, 在本文方法下无需依赖人工经验试错调整运行层性能指标 (53) 权重参数 $\bar{S}$ 和 R.

仿真阶段, 将基周期 h 设置为 1 s, 电振给矿机频率 u_1 和泵池补水阀门开度 u_2 两个基础控制回路的控制周期 T_u 设置为 2 s, 磨机给矿量 y_1 和泵池补水量 y_2 的采样周期 T_y 设置为 3 s, 框架周期即基础回路层控制周期 T_0 设置为 6 s、运行层周期 T 设置为 60 s.

4.2　本文方法的仿真结果及分析

首先基于演示系统连续采集 30 个运行控制周期即 1 800 s 的运行数据构建演示数据集. 该数据集完整记录了初始运行指标与期望值存在较大偏差时, 通过调整设定值使运行指标逐步逼近期望目标 $r^* = [70,\ 150]^{\mathrm{T}}$, 实现优化控制的过程, 包括设定值变化、基础回路控制过程数据等. 在实际矿物磨选过程中, 经验丰富的专家结合工艺知识、不同工况和不同运行指标对历史运行数据进行筛选和整理, 形成演示数据库. 演示系统采集数据集的过程可用于模拟控制工程师根据工况和所要实现的运行指标从演示数据库选择演示数据集的过程.

4.2.1　不同丢包率下算法仿真研究

本节展示本文方法的有效性以及不同丢包率下的控制性能. 图 2 ~ 图 7 为一组使用本文方法在最大连续反馈丢包数设置为 $\sigma_{f_{\max}} = \delta_{f_{\max}} = 1$, 丢包率设置为 30% 所得到的仿真结果曲线. 图 2 为演示数据和被控系统状态到控制器的反馈丢包频率. 图 3 给出了运行指标矿物磨选粒度和循环负荷期望值与实际值控制曲线; 从图 3 可以看出, 在运行指标初始状态与期望值有较大偏差的情况下, 控制系统在 6 个运行控制周期即 360 s 内实现了运行指标的优化控制. 图 4 给出了磨机给矿量和泵池补水量两个控制回路设定值的调整曲线以及实际跟踪情况; 在运行之初, 由于矿浆粒度较高、循环负荷较低, 通过降低给矿量和少量降低泵池补水量以使矿浆粒度逐渐降低至期望值, 循环负荷逐渐提高至期望值; 从图 4 中可以看出, 在每一次回路设定值优化设定后, 过程变量能够在 10 s 左右即小于运行控制周期的时间内快速跟踪设定值. 图 6 给出了参数的收敛曲线结果, 可以看出在本文方法下, 性能指标权重参数 $\bar{S}^i$ 和设定值控制增益 $\bar{K}^i$ 能够在 6 次迭代收敛至稳定值.

为了更好地说明本文所提出方法的控制效果, 在第 1 800 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、韧性和粒度分布等) 的变化对系统的扰动. 扰动出现后, 从图 3 和图 4 的仿真结果可以看出, 本文方法凭借其良好的自适应能力, 控制系统能够在 7 个运行控

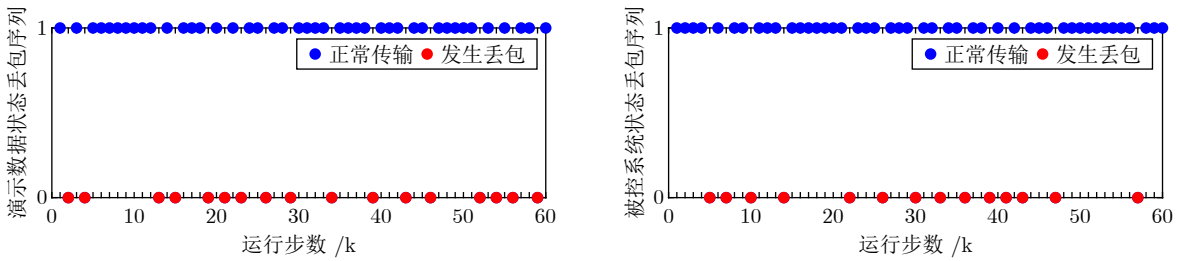


图 2　演示数据与被控系统状态丢包频率

Fig. 2　State packet loss frequency of demonstration data and controlled system

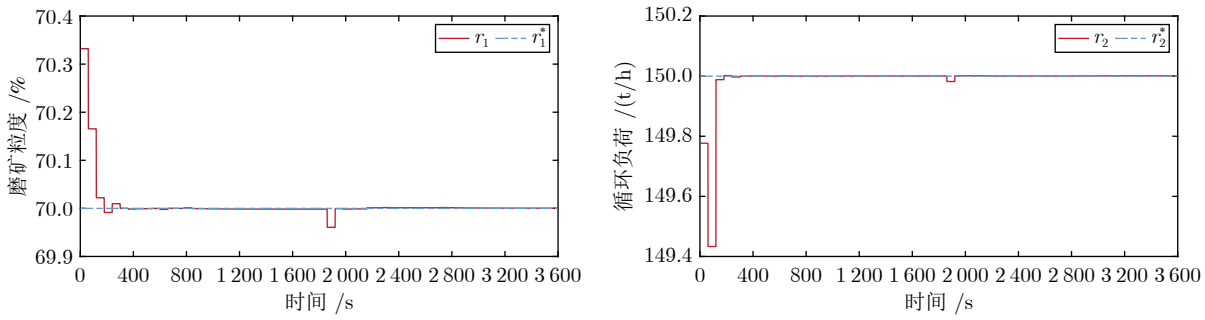


图 3　运行指标控制曲线

Fig. 3　The control curve of the operational index

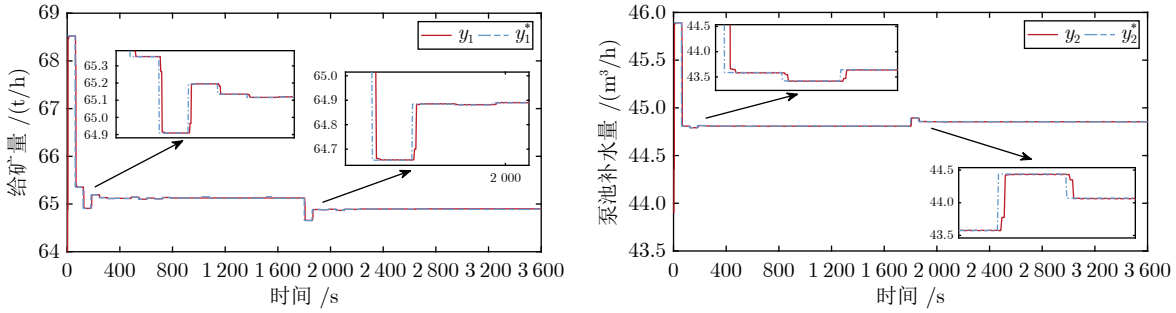


图 4 基础回路层设定值跟踪曲线

Fig. 4 The tracking curves for the set points of the basic loop layer

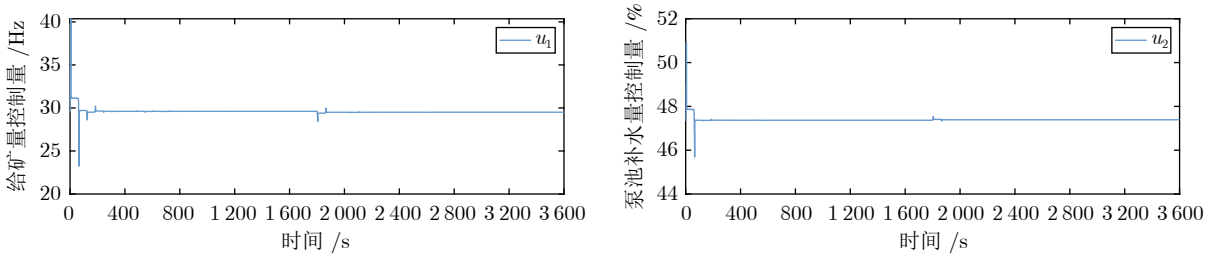


图 5 基础回路层控制输入曲线

Fig. 5 The control inputs curve of the basic loop layer

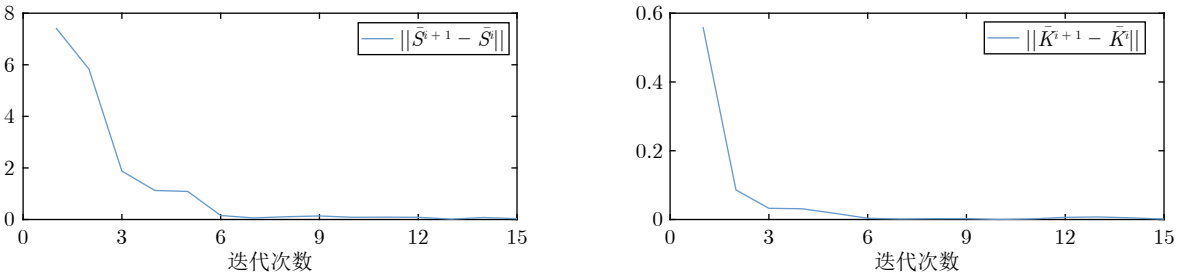


图 6 参数收敛曲线

Fig. 6 The convergence curve of the parameters

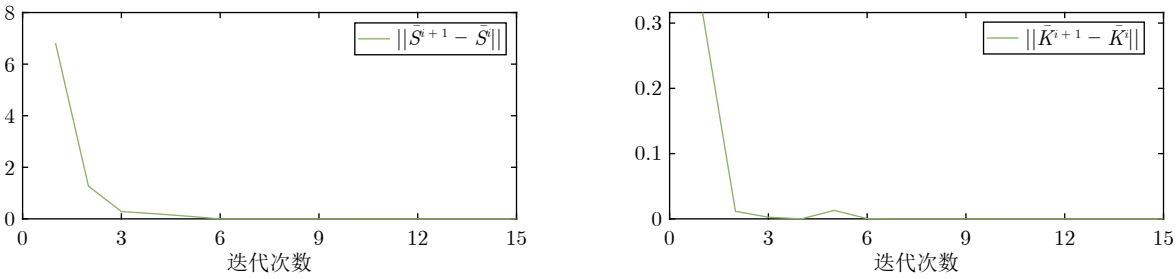


图 7 模型变化后参数收敛曲线

Fig. 7 The convergence curve of the parameters after the model change

制周期即 420 s 内使运行指标跟踪至期望值, 体现出较好的稳定性. 图 7 给出了模型变化后参数的收敛曲线结果, 可以观察到在本文方法下, 性能指标权重参数 $\bar{S}^i$ 和设定值控制增益 $\bar{K}^i$ 仍能在 7 次迭代收敛至稳定值.

为了量化运行指标的跟踪效果差异, 引入如下积分绝对误差 (integral absolute error, IAE) 和均方误差 (mean square error, MSE) 来评价算法

性能:

$$\mathrm{IAE}=\sum_{k=1}^{k^*}|r^*(k)-r(k)| \tag{108}$$

$$\mathrm{MSE}=\sqrt{\frac{1}{k^*}\sum_{k=1}^{k^*}|r^*(k)-r(k)|^2} \tag{109}$$

其中, k^* 为总仿真步数.

为进一步验证本文方法在不同丢包率下的鲁棒性, 将丢包率分别设置为 0%、10%、30%、50%、70%, 得到表 1 中所示的对比数据. 从表 1 中数据可以看出, 随着丢包率的增加, 各项误差指标呈小幅度上升趋势, 但整体仍保持在较低量级. 例如, 在 70% 丢包率下, r_1 和 r_2 的 MSE 指标维持在 10^{-4} 量级内, 表明均方误差仍然较小, 未出现发散现象; 同时, IAE 指标仍控制在 10^{-2} 量级内. 上述结果表明, 本文方法在不同丢包率下仍然能够保持较好的控制性能.

表 1　本文方法在不同丢包率下的评价指标

Table 1　Evaluation indices of the proposed method under different packet loss rates

丢包率 (%)	r_1 (IAE)	r_2 (IAE)	r_1 (MSE)	r_2 (MSE)	代码运行时间 (s)
0	0.0002	0.0011	2.8763e−08	1.2812e−06	0.1123
10	0.0023	0.0030	5.3894e−06	8.7299e−06	0.1124
30	0.0036	0.0024	1.3208e−05	5.8225e−06	0.1245
50	0.0034	0.0050	1.1482e−05	2.2387e−05	0.1305
70	0.0053	0.0152	2.7620e−05	2.3004e−04	0.1527

4.2.2　连续丢包场景下算法仿真研究

本节展示本文方法在连续丢包工况下的控制性能. 将最大连续反馈丢包数设置为 $\sigma_{f_{\max}}=\delta_{f_{\max}}=3$, 丢包率设置为 30%, 其他条件与第 4.2.1 节相同.

图 8 ~ 图 11 为本文方法在连续丢包工况下的仿真结果. 如图 8 所示, 给出了演示数据和被控系统状态到控制器的连续丢包频率. 图 9 给出了连续丢包工况下运行指标的控制结果. 可以观察到, 在初始阶段, 受初始值偏离以及丢包的影响, 运行指标出现一定幅度的波动; 随后, 随着通信补偿与逆学习算法的协同作用, 系统运行指标能够在 6 个运行周期即 360 s 内实现运行指标的跟踪. 图 10 展示了连续丢包工况下性能指标权重参数 $\bar{S}^i$ 与设定值控制增益 $\bar{K}^i$ 的迭代收敛过程. 可以看出, 在算法初始迭代阶段, 参数更新幅度较大, 表明学习算法能够根据演示数据和运行反馈迅速调整参数; 随着迭代次数增加, 参数变化幅度逐步减小, 并在第 7 次迭代收敛至零附近. 在 1800 s 仿真时间处模型参数发生突变, 从图 9 可以观察到, 模型参数突变后本文方法能够在 5 个运行周期内实现运行指标的跟踪; 图 11 给出了模型变化后参数的收敛情况, 可以观察到, 性能指标权重参数 $\bar{S}^i$ 和设定值控制增益 $\bar{K}^i$

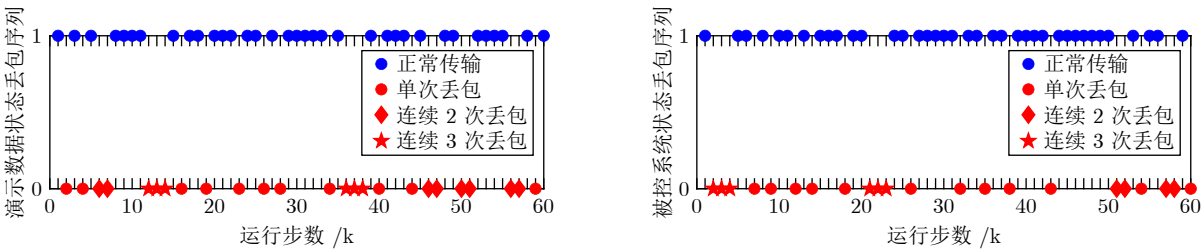


图 8　演示数据与被控系统状态连续丢包频率

Fig. 8　State continuous packet loss frequency of demonstration data and controlled system

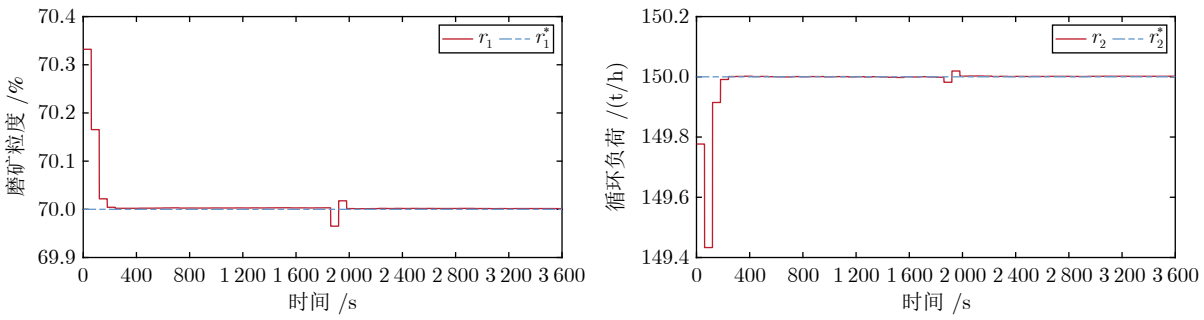


图 9　连续丢包下运行指标控制曲线

Fig. 9　The control curve of the operational index subject to continuous packet loss

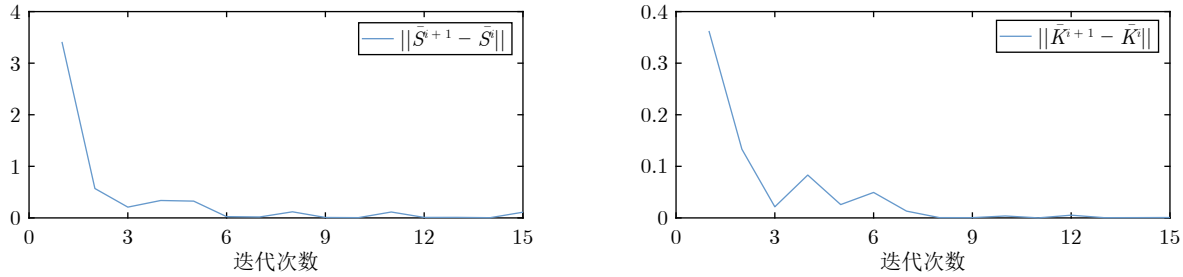


图 10 连续丢包下参数收敛曲线

Fig. 10 The convergence curve of the parameters subject to continuous packet loss

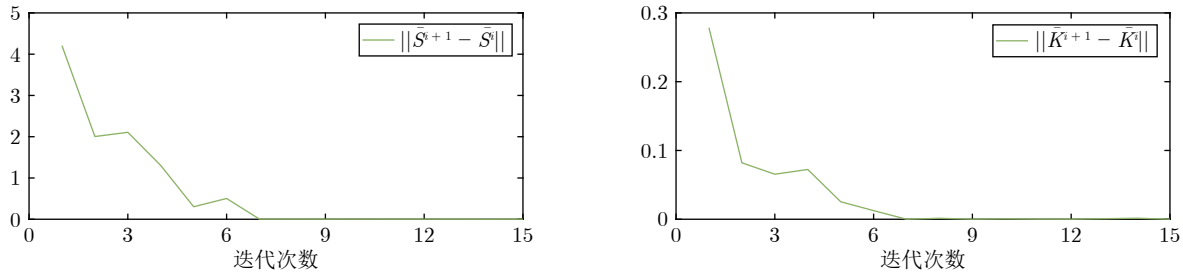


图 11 连续丢包下模型变化后参数收敛曲线

Fig. 11 The convergence curve of the parameters after the model change subject to continuous packet loss

仍能在 7 次迭代收敛至稳定值. 在本节的仿真条件下, r_1 和 r_2 的 IAE 指标分别为 0.004 2、0.005 1, 控制在 10^{-2} 量级内; MSE 指标分别为 2.929 0e−05、3.031 0e−05, 维持在 10^{-5} 量级内. 从评价指标数据可以看出本文方法在连续丢包工况下依然能保持较好的控制性能.

综合图 8 ~ 图 11 的仿真结果以及评价指标数据可以得出, 在连续丢包工况下, 本文方法仍能够实现运行指标的稳定控制, 并保证参数学习过程的收敛性. 仿真结果进一步验证了本文控制方法在不可靠通信环境下的有效性.

本文仿真研究在 CPU 为 i7-9700、主频为 3 GHz、RAM 为 8 GB 的平台上采用 MATLAB R2019b 进行. 根据表 1 给出的代码运行时间, 在本文的仿真条件下代码能够在 1 s 内完成参数学习与设定值的优化更新, 远小于运行层控制周期 60 s; 进一步结合第 3.3.4 节算法复杂度的分析可知, 本文方法整体计算负担较轻, 能够在工业控制器的计算能力约束下实时运行, 满足矿物磨选运行优化控制对在线计算效率的要求.

4.3 通信处理方法对比仿真研究

为验证本文所提出通信补偿方案的优越性, 本节将其与文献 [24] 中针对不可靠网络通信的处理方法进行对比. 文献 [24] 采用了状态保持策略, 没有对丢包状态数据进行补偿, 在状态信息能够正常传输时, 直接利用该状态进行设定值的计算; 当发生丢包时, 则采用最近一次成功接收的状态信息来代替当前丢失的状态数据进行设定值计算. 对比实验阶段, 最大连续反馈丢包数设置为 $\sigma_{f_{\max}}=\delta_{f_{\max}}=1$, 丢包率设置为 30%, 其他初始条件与第 4.2 节相同. 除此之外, 本部分采用的算法仅与第 4.2 节中通信补偿方案不同, 其他算法部分完全一致.

使用状态保持策略所得到的控制结果如图 12 ~ 图 13 所示. 图 12 给出了运行指标矿物磨选粒度和循环负荷的期望值与实际控制结果的对比曲线. 从图 12 可以看出, 虽然该方法可以实现粒度和循环负荷两个运行指标的跟踪, 但由于在状态信息丢失时设定值控制器无法及时获得最新状态信息, 系统需要较长的调节时间 (约 1 000 s), 明显长于本文提出的通信补偿方案下的调节时间. 图 13 展示了性能指标权重参数 $\bar{S}^i$ 和设定值控制增益 $\bar{K}^i$ 的迭代收敛过程, 可以观察到在该方法下, 参数收敛速度较慢, 需要 17 次迭代才能达到稳定状态. 在本节的仿真条件下, 评价指标数据如表 2 所示. 从表 2 中数据可知, 本文通信补偿方法与文献 [24] 通信丢包处理方法相比, r_1 和 r_2 的 IAE 指标从 38.298 1、49.868 3 降低至 0.003 6、0.002 4, MSE 指标从 0.072 3、0.193 2 降低至 1.320 8e−05、5.822 5e−06. 综合对比结果, 本文方法不仅显著缩短了调节时间, 还在不可靠通信条件下表现出更优的控制性能.

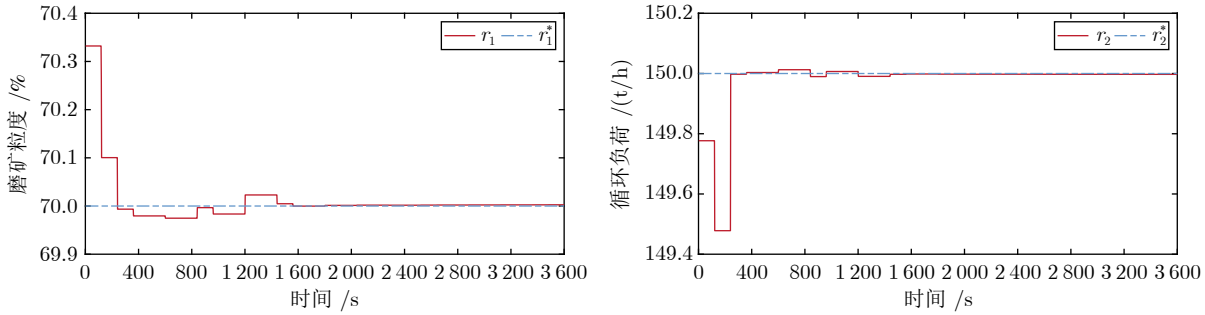


图 12　状态保持策略下运行指标控制曲线

Fig. 12　The control curve of the operational index under state-holding strategy

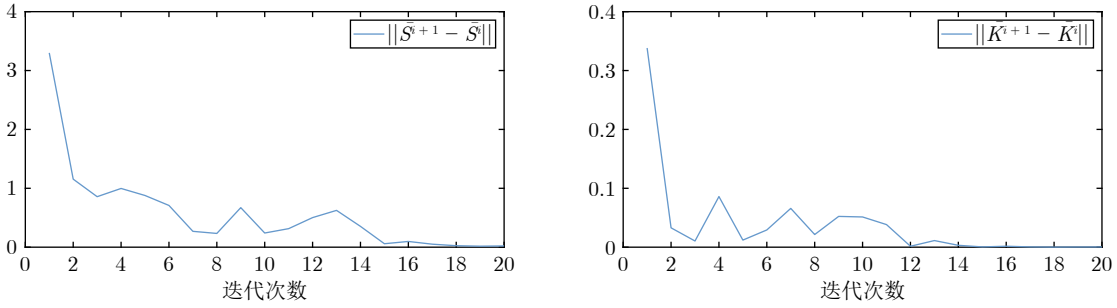


图 13　状态保持策略下参数收敛曲线

Fig. 13　The convergence curve of the parameters under state-holding strategy

表 2　不同通信补偿方案的评价指标

Table 2　Evaluation indices of different communication compensation schemes

丢包处理方法	r_1 (IAE)	r_2 (IAE)	r_1 (MSE)	r_2 (MSE)
本文方法	**0.0036**	**0.0024**	**1.3208e−05**	**5.8225e−06**
文献 [24]	38.2981	49.8683	0.0723	0.1932

4.4　多速率分层方法对比仿真研究

本节展示了本文方法与多速率分层 MPC、多速率分层 Q 学习算法[21] 的对比结果, 以体现逆学习性能指标权重参数的优越性. 对于该部分仿真研究, 考虑如图 2 所示的相同的状态丢失序列. 在进行多速率分层 MPC 仿真研究时, 由于运行过程模型参数未知, 需采用系统辨识方法得到近似模型, 可参考文献 [36], 在此基础上定义如下性能指标:

$$J_{MPC}(k) = \|\Gamma_r(r(k+1) - r^*(k+1))\|^2 + \|\Gamma_y y(k)\|^2 \tag{110}$$

其中, Γ_r 和 Γ_y 分别为运行指标跟踪误差和设定值输入的加权因子, 并设置为:

$$\Gamma_r = \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix},\ \Gamma_y = \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix}$$

在进行多速率分层 Q 学习算法[21] 的仿真研究时, 运行层性能指标权重参数设置为:

$$Q = \begin{bmatrix} 750 & 0 \\ 0 & 100 \end{bmatrix},\ R = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}$$

$\bar{S}$ 可通过计算获得. 上述两组性能指标权重参数均由人工经验设定, 旨在模拟工程实践中算法工程师通过反复试错调节权重参数的过程.

图 14 为多速率分层 MPC 方法下运行指标跟踪曲线, 可以看出, 在运行指标初始状态与期望值有较大偏差的情况下, 控制系统在 10 个运行控制周期即 600 s 内实现了运行指标的稳态, 并趋近于运行指标期望值. 图 15 为多速率分层 Q 学习方法下运行指标跟踪曲线, 可以看出, 控制系统在 11 个运行控制周期即 660 s 内实现了运行指标的稳态, 并趋近于运行指标期望值. 表 3 进一步给出了不同方法下运行指标跟踪效果的评价指标数据, 可以看出, 与多速率分层 MPC 方法和多速率分层 Q 学习方法相比, 本文方法 r_1 和 r_2 的 IAE 指标从 438.7964、629.7690 以及 47.2822、107.7708 降低至 0.0036、0.0024; 同时, MSE 指标从 0.6073、0.8281 以及 0.0047、0.0189 降低至 1.3208e−05、5.8225e−06, 进一步验证了本文方法的优越性.

综合图 14、图 15 的运行指标动态响应以及表 3 的定量评价结果可以看出, 当性能指标权重由

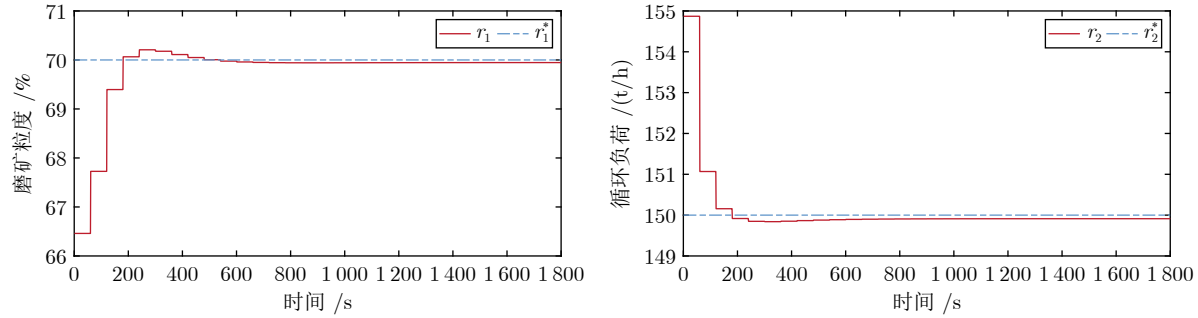


图 14 多速率分层 MPC 方法下运行指标控制曲线

Fig. 14 The control curve of the operational index under multi-rate layered MPC method

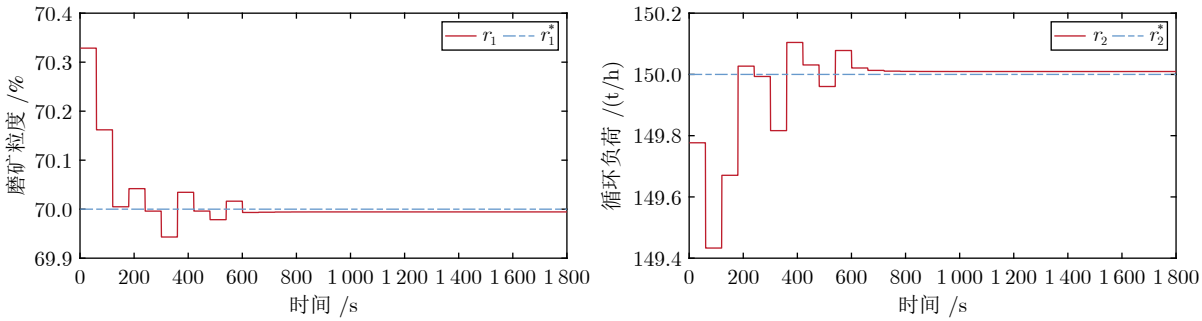


图 15 多速率分层 Q 学习方法下运行指标控制曲线

Fig. 15 The control curve of the operational index under multi-rate layered Q-learning method

表 3 不同方法的评价指标

Table 3 Evaluation indices of different methods

对比方法	r_1 (IAE)	r_2 (IAE)	r_1 (MSE)	r_2 (MSE)
本文方法	**0.0036**	**0.0024**	**1.3208e−05**	**5.8225e−06**
多速率分层 MPC	438.7964	629.7690	0.6073	0.8281
多速率分层 Q 学习	47.2822	107.7708	0.0047	0.0189

人工经验设定时, 若权重参数选择不合理, 不仅会影响系统的调节过程, 还可能导致稳态阶段出现较大误差, 从而限制运行指标的最终优化效果. 此外, 权重参数往往需要算法工程师经过多次试错调试才能取得较为理想的控制性能, 该过程耗时耗力且难以适应不同工况变化.

由第 4 节仿真研究的结果分析可以看出, 本文所提出的不可靠通信下矿物磨选过程多速率分层学习控制方法, 具有稳定性好、响应快的特点, 显著降低了运行层设定值优化对人工经验调节性能指标权重的依赖, 有效实现了在不可靠通信、多时间尺度等综合问题下对矿物磨选过程运行指标的稳定控制, 提质增效效果显著.

4.5 方法适应性分析

在本文方法下, 运行层性能指标权重参数是逆学习过程中的待恢复变量, 其初始值决定学习初期各性能指标之间的相对主导关系. 当磨选工艺结构较为简单、耦合程度较低但单位能耗敏感时, 如开路磨矿过程, 若运行指标跟踪误差项权重初值过大, 逆学习过程将强化快速调节行为, 导致控制输入频繁波动并增加能耗. 在此情况下, 权重初值宜适当提高控制输入项权重、相对降低跟踪误差项权重, 使学习过程在满足基本精度约束的前提下抑制过度调节. 当磨选系统流程较长、设备数量多、耦合关系复杂且对连续稳定运行要求较高时, 如多段磨矿过程, 若控制输入项权重过大则可能削弱设定值的调节强度, 影响系统稳定区间. 此时权重初值应提高跟踪误差项权重, 使关键指标在逆学习过程中优先保持在合理范围内, 在此基础上逐步实现权重参数的学习优化. 由此可见, 不同磨选工艺条件下权重初值的调整, 体现为误差项与控制输入项权重比例的结构性配置. 初值决定逆学习的演化方向及性能权衡路径, 是本文方法实现工程适应性的关键环节.

5 结束语

本文针对多时间尺度、运行层模型未知、网络传输存在不可靠通信的复杂矿物磨选过程, 提出一种融合 MPC、提升技术、史密斯预估器和强化学习的集成运行优化控制方法. 该方法使用两次提升技

术将多速率问题统一到同一时间尺度, 在基础回路层设计多速率 MPC 控制器实现设定值的跟踪, 进而将回路动态引入运行层, 采用强化学习技术, 设计带有通信补偿的快慢耦合逆学习控制算法, 利用演示数据逆向学习性能指标并优化求解设定值, 进而实现运行指标的优化控制. 针对一段闭环矿物磨选过程进行了仿真实验, 仿真结果验证了提出方法的有效性. 未来工作中, 将进一步研究同时存在随机丢包、传输时延或数据乱序情况下的多速率分层运行优化控制问题, 从而进一步提升所提出方法在复杂工业网络环境中的适用性与鲁棒性.

参考文献

1 Chai Tian-You, Zheng Rui, Xing Fang-Xin, Jia Yao, Zheng Xiu-Ping. Intelligence for industrial process control: Development and prospects. *Science China Information Sciences*, 2025, **55**: 1555−1570
(柴天佑, 郑锐, 邢方新, 贾瑶, 郑秀萍. 工业过程控制智能化及未来发展展望. 中国科学: 信息科学, 2025, **55**: 1555−1570)

2 Fan Jia-Lu, Zhang Ye-Wei, Chai Tian-You. Optimal operational feedback control for a class of industrial processes. *Acta Automatica Sinica*, 2015, **41**(10): 1754−1761
(范家璐, 张也维, 柴天佑. 一类工业过程运行反馈优化控制方法. 自动化学报, 2015, **41**(10): 1754−1761)

3 Zhou P, Chai T Y, Wang H. Intelligent optimal-setting control for grinding circuits of mineral processing process. *IEEE Transactions on Automation Science and Engineering*, 2009, **6**(4): 730−743

4 Aguila-Camacho N, le Roux J D, Duarte-Mermoud M A, Orchard M E. Control of a grinding mill circuit using fractional order controllers. *Journal of Process Control*, 2017, **53**: 80−94

5 Lestage R, Pomerleau A, Hodouin D. Conbstrained real-time optimization of a grinding circuit using steady-state linear programming supervisory control. *Powder Technology*, 2002, **124**(3): 254−263

6 le Roux J D, Padhi R, Craig I K. Optimal control of grinding mill circuit using model predictive static programming: A new nonlinear MPC paradigm. *Journal of Process Control*, 2014, **24**(12): 29−40

7 Minchala L I, Zhang Y, Garza-Castanón L. Predictive control of a closed grinding circuit system in cement industry. *IEEE Transactions on Industrial Electronics*, 2017, **65**(5): 4070−4079

8 Yamashita A S, Martins W T, Pinto T V B, Raffo G V, Euzébio T A M. Multiobjective tuning technique for MPC in grinding circuits. *IEEE Access*, 2023, **11**: 43041−43054

9 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

10 Zhou P, Dai W, Chai T Y. Multivariable disturbance observer based advanced feedback control design and its application to a grinding circuit. *IEEE Transactions on Control Systems Technology*, 2013, **22**(4): 1474−1485

11 Engell S. Feedback control for optimal process operation. *Journal of Process Control*, 2007, **17**(3): 203−219

12 Baraheni M, Azarhoushang B, Daneshi A, Kadivar M, Amini S. Development of an expert system for optimal design of the grinding process. *The International Journal of Advanced Manufacturing Technology*, 2021, **116**(9): 2823−2833

13 Zhao D, Chai T Y, Wang H, Fu J. Hybrid intelligent control for regrinding process in hematite beneficiation. *Control Engineering Practice*, 2014, **22**: 217−230

14 Dai Wei, Wang Xian-Wei, Lu Xing-Long, Chai Tian-You. Case-based reasoning and reinforcement learning integrated set-point optimization method for grinding process. *Control Theory & Applications*, 2019, **36**(1): 53−64
(代伟, 王献伟, 路兴龙, 柴天佑. 基于案例推理增强学习的磨矿过程设定值优化. 控制理论与应用, 2019, **36**(1): 53−64)

15 Zhou P, Chai T Y, Sun J. Intelligence-based supervisory control for optimal operation of a DCS-controlled grinding system. *IEEE Transactions on Control Systems Technology*, 2012, **21**(1): 162−175

16 Li Jin-Na, Gao Xi-Ze, Chai Tian-You, Fan Jia-Lu. Data-driven operational optimization control of industrial processes. *Control Theory & 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)

17 Dai W, Chai T Y, Yang S X. Data-driven optimization control for safety operation of hematite grinding process. *IEEE Transactions on Industrial Electronics*, 2014, **62**(5): 2930−2941

18 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 & Applications*, 2016, **10**(12): 1348−1356

19 Lu X L, Kiumarsi B, Chai T Y, Jiang Y, Lewis F L. Operational control of mineral grinding processes using adaptive dynamic programming and reference governor. *IEEE Transactions on Industrial Informatics*, 2018, **15**(4): 2210−2221

20 Li J, Yang M, Lewis F L. Optimal operational self-learning control for multi-time scale industrial processes with signal compensations. *Engineering Applications of Artificial Intelligence*, 2023, **126**: Article No. 107065

21 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
(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)

22 Liu J X, Shen H, Wang J, Cao J D, Rutkowski L. H_∞ control for interconnected systems with unknown system dynamics: A two-stage reinforcement learning method. *IEEE Transactions on Automation Science and Engineering*, 2024, **22**: 6388−6397

23 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)

24 Chai T Y. Industrial process control systems: Research status and development direction. *Scientia Sinica Informationis*, 2016, **46**(8): 1003−1015

25 Chai T Y, Zhao L, Qiu J, Fan J. Integrated network-based model predictive control for setpoints compensation in industrial processes. *IEEE Transactions on Industrial Informatics*, 2012, **9**(1): 417−426

26 Fan Jia-Lu, Jiang Yi, Chai Tian-You. Operational feedback control of industrial processes in a wireless network environment. *Acta Automatica Sinica*, 2016, **42**(8): 1166−1174
(范家璐, 姜艺, 柴天佑. 无线网络环境下工业过程运行反馈控制方法. 自动化学报, 2016, **42**(8): 1166−1174)

27 Jiang Y, Liu L, Feng G. Adaptive optimal control of networked nonlinear systems with stochastic sensor and actuator dropouts based on reinforcement learning. *IEEE Transactions on Neural Networks and Learning Systems*, 2022, **35**(3): 3107−3120

28 Zhang X M, Han Q L, Ge X. A novel approach to H_∞ performance analysis of discrete-time networked systems subject to network-induced delays and malicious packet dropouts. *Automatica*, 2022, **136**: Article No. 110010

29 Jiang Y, Yang T, Gao W, Wu J, Chai T Y, Lewis F L. Off-policy reinforcement learning for H_∞ control of linear discrete-time systems with network induced dropouts. *IEEE Transactions on Automatic Control*, 2025, **70**(12): 8000−8015

30 Lian B, Xue W, Lewis F L, Davoudi A. Inverse value iteration and Q-learning: Algorithms, stability, and robustness. *IEEE Transactions on Neural Networks and Learning Systems*, 2024, **36**(4): 6970−6980

31 Wu H, Hu Q, Zheng J, Dong F, Ouyang Z, Li D. Discounted inverse reinforcement learning for linear quadratic control. *IEEE*

Transactions on Cybernetics, 2025, **55**(4): 1995−2007

32 Chen T, Francis B A. *Optimal Sampled-data Control Systems*. London: Springer Science & Business Media, 2012.

33 Lewis F L, Vrabie D, Syrmos V L. *Optimal Control*. Hoboken: John Wiley & Sons, 2012.

34 Astrom K J, Hang C C, Lim B C. A new Smith predictor for controlling a process with an integrator and long dead-time. *IEEE Transactions on Automatic Control*, 1994, **39**(2): 343−345

35 Xue W Q, Lian B S, Fan J L, Kolaric P, Chai T Y, Lewis F L. Inverse reinforcement Q-learning through expert imitation for discrete-time systems. *IEEE Transactions on Neural Networks and Learning Systems*, 2021, **34**(5): 2386−2399

36 Jiang Y, Li Y N, Fan J L. Model predictive control-based setpoint regulation in industrial processes. *Control Engineering of China*, 2018, **25**(6): 980−984



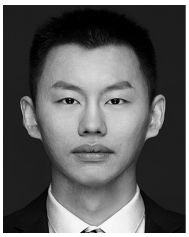
任鹏旭 中国矿业大学信息与控制工程学院博士研究生. 主要研究方向为工业过程运行优化与控制, 强化学习. E-mail: px.ren@cumt.edu.cn

(**REN Peng-Xu** Ph.D. candidate at the School of Information and Control Engineering, China University of Mining and Technology. His research interests include operational optimization and control for industrial process, and reinforcement learning.)



代 伟 中国矿业大学信息与控制工程学院教授. 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** Professor at the School of Information and Control Engineering, China University of Mining and Technology. His research interests include modeling, operational optimization and control for complex industrial process. Corresponding author of this paper.)



张淇瑞 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的安全控制和脆弱性分析. E-mail: qiruizhang@cumt.edu.cn

(**ZHANG Qi-Rui** Associate professor at the School of Information and Control Engineering, China University of Mining and Technology. His research interests include secure control and vulnerability analysis for complex industrial process.)



杨春雨 中国矿业大学信息与控制工程学院教授. 主要研究方向为奇异摄动系统, 工业过程运行控制, 网络物理系统和鲁棒控制.
E-mail: chunyuyang@cumt.edu.cn

(**YANG Chun-Yu** Professor at the School of Information and Control Engineering, China University of Mining and Technology. His research interests include singularly perturbed systems, industrial process operational control, cyber-physical systems, and robust control.)