

# 基于强化学习自适应鲁棒控制的异构双阀协调控制策略研究

苏世杰<sup>1\*</sup> 程泳钦<sup>1</sup> 胡毅<sup>1,2</sup> 何建辉<sup>1</sup> 杨书吉<sup>1</sup>

1. 江苏科技大学机械工程学院, 镇江, 212003

2. 航天晨光股份有限公司, 南京, 211100

**摘要:** 采用小流量伺服阀和大流量比例阀并联驱动同一个执行元件的异构双阀电液伺服系统具有成本低、流量大、精度高的优势, 但当系统参数发生变化或受到内外扰动时, 其控制性能和稳定性会急剧下降。为此提出一种融合强化学习 SAC (soft actor-critic) 算法与自适应鲁棒控制算法 (ARC) 的异构双阀协调控制策略。该控制策略一方面通过设计的流量分配策略减小比例阀与伺服阀在工作切换中产生的瞬态误差, 另一方面通过上层 SAC 算法学习目标电液伺服系统的动态非线性特性, 进而实现对下层 ARC 算法控制参数的动态调节, 以增强系统的控制性能与鲁棒性。该研究为后续的仿真和实验验证提供了坚实的理论基础。

**关键词:** 电液伺服系统; 双阀并联控制; 流量分配; 伺服控制

**中图分类号:** TP273

DOI:10.3969/j.issn.1004-132X.2025.10.022

开放科学(资源服务)标识码(OSID):



## Coordination Control of Dual-valve Electrohydraulic Servo Systems Based on Integration of Reinforcement Learning and Adaptive Robust Control

SU Shijie<sup>1\*</sup> CHENG Yongqin<sup>1</sup> HU Yi<sup>1,2</sup> HE Jianhui<sup>1</sup> YANG Shuji<sup>1</sup>

1. School of Mechanical Engineering, Jiangsu University of Science and Technology, Zhenjiang, Jiangsu, 212003

2. Aerosun Co., Ltd., Nanjing, 211100

**Abstract:** The dual-valve electrohydraulic servo systems, which employed a small-flow servo valve and a large-flow proportional valve to drive the same actuator in parallel, offered advantages such as low cost, high flow rate and high accuracy. However, the control performance of the systems was compromised by parameter uncertainties, system nonlinearities and disturbances. To address these issues, a dual-valve coordinated control strategy was proposed which integrates the SAC reinforcement learning algorithm with the ARC algorithm. This control strategy aimed to reduce transient errors generated by the proportional valves and the servo valves during work switching through a specifically designed flow allocation strategy. Additionally, the upper SAC algorithm learned the dynamic nonlinearities of the target electrohydraulic servo systems. Consequently, the control parameters of the lower ARC algorithm were dynamically adjusted, thereby enhancing the system's control performance and robustness. The findings of this study establish a solid theoretical foundation for subsequent simulation and experimental validation.

**Key words:** electrohydraulic servo system; dual-valve parallel control; flow allocation; servo control

### 0 引言

随着技术的进步, 航空航天、船舶与海洋工程装备逐步向着重型化、高精度的方向发展, 作为相关机械装备动力核心的电液伺服系统也不断向着大流量、高精度及低成本等方向演进<sup>[1-3]</sup>。目前,

大流量电液伺服系统的核心控制元件为电液比例阀或大流量电液伺服阀。但电液比例阀存在死区、滞环等非线性因素, 使得系统的控制性能有限; 而大流量电液伺服阀不仅成本高、维护困难, 且系统的动态响应性能会随着伺服阀额定流量的增大而逐步降低。

实际上, 许多用于重型装备的大流量电液伺服系统的工作过程可以看作一系列快速运动阶段和精确定位阶段的组合。在快速运动阶段, 系统需要快速大量的输出流量, 而在精确定位阶段, 系

收稿日期: 2024-09-10

基金项目: 江苏省高等学校基础学科研究重大资助项目(23KJA40005); 江苏省镇江市产业前瞻与共性关键技术项目(GY2024003)

统需要精确地控制输出流量。因此,有学者提出通过采用两个电液比例/伺服阀并联驱动一个执行元件的方案,并设计了相应的系统结构与控制算法<sup>[4]</sup>。YU等<sup>[5]</sup>提出采取迭代学习控制与自适应控制相结合的控制策略控制两个伺服阀同步驱动一个作动器。SU等<sup>[6]</sup>提出基于自适应鲁棒控制策略的双阀协调控制策略,该策略采用自适应控制机制对系统的未知参数进行预测,并通过鲁棒控制策略提高系统的鲁棒性。

随着人工智能技术的快速发展<sup>[7-9]</sup>,为进一步提高电液伺服系统在复杂多变工况下的性能与鲁棒性,国内外学者提出了融合机器/深度学习与经典/现代控制理论的智能控制算法<sup>[10]</sup>,其目的是通过智能体动态地调整控制算法的相关参数,从而增强被控系统的控制性能。CHEN等<sup>[11]</sup>提出将强化学习 DDPG(deep deterministic policy gradient)算法与 PID(proportional integral derivative)算法相结合并应用于电液伺服控制系统中,结果表明其控制性能优于PID及改进的PID控制算法。HE等<sup>[12]</sup>通过强化学习与PID控制算法相结合来控制比例阀实现电液伺服系统位置控制,结果表明该控制策略在面对内外干扰时能有效地提高位置跟踪精度。YUAN等<sup>[13]</sup>通过强化学习 TD3(twin delayed deep deterministic policy gradient)算法来实现液压控制系统的位置控制,其控制性能优于PID算法。YU等<sup>[14]</sup>将强化学习与PID控制算法相结合并应用于机器人的轨迹规划中,当系统不能精确建模以及运行环境发生相应的变化时,所提控制算法能根据环境的变化调节PID控制参数,实时地补偿路径与移动机器人的误差。

现有的双阀协调控制系统主要依靠现代控制算法来改善双阀电液伺服系统的位置跟踪精度,但与常规的单阀及双阀并联电液伺服系统相比,由于异构双阀电液伺服系统存在更强的阀间耦合干扰,当系统参数发生变化或受到内外扰动时,其控制性能和稳定性会急剧下降。本文提出一种融合深度强化学习和自适应鲁棒控制的异构双阀协调控制策略,通过设计的智能体与环境的持续交互自主学习系统在各种不确定工况下的最优控制参数调节策略,进而有效提高异构双阀电液伺服系统在参数发生变化和受到内外扰动时的稳定性和控制精度。其主要工作与贡献如下:

1)建立了异构双阀协调控制电液伺服系统数学模型,并提出了一种由上层强化学习 SAC(soft

actor critic)算法、下层自适应鲁棒控制(adaptive robust control, ARC)算法及异构双阀流量分配策略组成的SAC-ARC异构双阀协调控制策略。

2)提出了一种基于SAC算法的ARC控制参数自适应自调整机制。该机制能够根据电液伺服系统当前的状态动态调节ARC算法的控制超参数,使得系统在复杂多变的环境下也能精确跟踪不断变化的目标轨迹。

3)提出了一种基于比例阀输入信号补偿的伺服阀-比例阀流量分配策略,有效地减小了伺服阀与比例阀工作切换时产生的瞬态误差。

### 1 异构双阀电液伺服系统数学建模

图1所示的异构双阀电液伺服系统主要由液压泵、电机、比例阀、伺服阀、液压缸、位移传感器以及控制器等组成。控制器周期性地向比例阀和伺服阀发出控制信号,并通过位移传感器获取液压缸活塞杆的当前位置,从而实现对液压缸活塞位置的闭环控制。

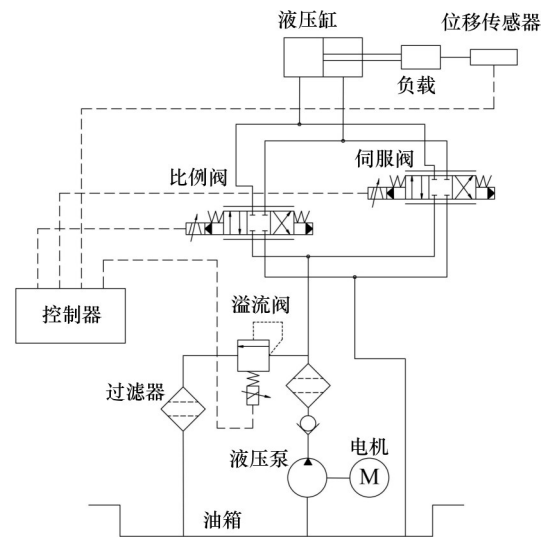


图1 异构双阀协调控制系统的工作原理

Fig.1 The design of heterogeneous dual-valve control system

根据比例阀与伺服阀的工作特性,其线性流量方程可表示为<sup>[15]</sup>

$$\left. \begin{aligned} q_{PL} &= K_{Pq} x_{Pv} - K_{Pc} p_L \\ q_{SL} &= K_{Sq} x_{Sv} - K_{Sc} p_L \end{aligned} \right\} \quad (1)$$

式中: $q_{PL}$ 、 $q_{SL}$ 分别为比例阀、伺服阀的输出流量; $K_{Pq}$ 、 $K_{Sq}$ 分别为比例阀、伺服阀的流量增益系数; $K_{Pc}$ 、 $K_{Sc}$ 分别为比例阀、伺服阀的流量-压力系数; $x_{Pv}$ 、 $x_{Sv}$ 分别为比例阀阀芯、伺服阀阀芯的输出位移; $p_L$ 为负载压力。

比例阀和伺服阀阀芯的位移与电信号之间的关系可表示为<sup>[16]</sup>

$$\begin{cases} x_{p_v} = K_{p_a} K_{p_s} u_p \\ x_{s_v} = K_{s_a} K_{s_s} u_s \end{cases} \quad (2)$$

式中： $K_{p_a}$ 为比例阀放大器增益； $K_{p_s}$ 为比例阀增益； $u_p$ 为比例阀输入信号； $K_{s_a}$ 为伺服阀放大器增益； $K_{s_s}$ 为伺服阀增益； $u_s$ 为伺服阀输入信号。

结合式(1)、式(2)，系统实际输出的总流量为

$$Q_L = q_{pL} + q_{sL} = K_{p_q} K_{p_a} K_{p_s} u_p + K_{s_q} K_{s_a} K_{s_s} u_s - (K_{p_c} + K_{s_c}) p_L \quad (3)$$

由流量连续性定理可知，非对称液压缸的系统流量连续方程为<sup>[17]</sup>

$$q_L = A \frac{dx_p}{dt} + \frac{V_t}{2(1+n^2)\beta_e} \frac{dp_L}{dt} + C_t p_L \quad (4)$$

式中： $A$ 为液压缸活塞的有效面积； $x_p$ 为液压缸活塞位移； $C_t$ 为液压缸外泄漏系数； $V_t$ 为管道及液压缸油腔总容积； $\beta_e$ 为液压油体积弹性模量； $n$ 为液压缸有杆腔与无杆腔有效作用面积之比。

非对称液压缸的力平衡方程为

$$A p_L = m_t \frac{d^2 x_p}{dt^2} + B_p \frac{dx_p}{dt} + K_p x_p + F \quad (5)$$

式中： $m_t$ 为活塞和负载的总质量； $B_p$ 为活塞和负载黏性阻尼系数； $K_p$ 为弹性刚度系数； $F$ 为作用在活塞上的外负载力。

联合式(3)~式(5)可得

$$K_1 u_p + K_2 u_s = M_y \ddot{x}_p + B \dot{x}_p + C x_p + D x_p + G \quad (6)$$

$$M_y = \frac{V_t m_t}{2(1+n^2)\beta_e A}$$

$$B = \frac{V_t B_p + 2(1+n^2)\beta_e K_{CE} m_t}{2(1+n^2)\beta_e A}$$

$$C = \frac{2(1+n^2)\beta_e A^2 + K_p V_t + 4\beta_e K_{CE} B_p}{2(1+n^2)\beta_e A}$$

$$D = \frac{K_{CE} K_p}{A}$$

$$G = \frac{V_t}{2(1+n^2)A} \dot{F} + \frac{K_{CE} F}{A} \quad K_1 = K_{p_q} K_{p_a} K_{p_s}$$

$$K_2 = K_{s_q} K_{s_a} K_{s_s} \quad K_{CE} = K_{p_c} + K_{s_c} + C_t$$

当液压缸处于正向运动时，有

$$K_q = C_w \sqrt{\frac{2(p_s - p_L)}{\rho(1+n^3)}}$$

$$K_{CE} = C_t - C_d w x \sqrt{\frac{1}{2\rho(p_s - p_L)(1+n^3)}}$$

反之，则有

$$K_q = C_d w \sqrt{\frac{2(np_s + p_L)}{\rho(1+n^3)}}$$

$$K_{CE} = C_t - C_d w x \sqrt{\frac{1}{2\rho(np_s + p_L)(1+n^3)}}$$

## 2 异构双阀协调控制策略设计

如图2所示，本文提出的异构双阀协调控制策略主要由SAC算法、ARC算法及流量分配策略

组成。上层SAC算法根据被控系统的状态动态调整下层ARC算法的控制超参数，从而精确地计算被控系统所需要的实时流量，然后通过流量分配策略计算出伺服阀与比例阀各自的输出信号，最终实现对液压缸运动的精确控制。

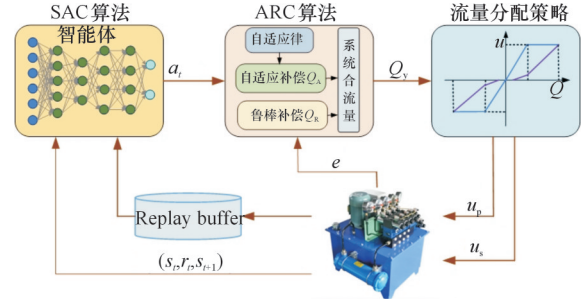


图2 异构双阀协调控制SAC-ARC概念模型

Fig.2 Structure of the proposed SAC-ARC scheme

### 2.1 ARC算法设计

根据式(6)建立液压缸活塞杆流量平衡方程，总流量 $Q_y$ 为

$$Q_y = K_1 u_p + K_2 u_s = M_y \ddot{x}_p + B \dot{x}_p + C x_p + D x_p + d \quad (7)$$

$$d = \bar{d} + \Delta$$

式中： $d$ 为系统干扰； $\bar{d}$ 为 $d$ 的常值分量； $\Delta$ 为 $d$ 的时变分量。

根据式(7)定义未知参数：

$$\eta_d = [\eta_{d1} \quad \eta_{d2} \quad \dots \quad \eta_{d5}]^T = [M_y \quad B \quad C \quad D \quad \bar{d}]^T \quad (8)$$

定义液压缸活塞杆运动轨迹的跟踪误差 $e$ ：

$$e = x_p - x_d \quad (9)$$

式中： $x_p$ 为期望的运动轨迹； $x_d$ 为实际的运动轨迹。

定义系统的切换函数为

$$z = \dot{e} + K_v e \quad (10)$$

式中： $K_v$ 为正增益系数。

设计关于 $z$ 和 $e$ 的Lyapunov函数：

$$V = \frac{1}{2} B (z^2 + e^2) \quad (11)$$

分别对式(10)和式(11)求导，联立式(9)可得

$$\dot{V} = zB\dot{z} + eB\dot{e} = z(Q_y - \varphi\eta_d - \Delta) + eB\dot{e} \quad (12)$$

其中， $\varphi$ 为回归量矩阵，其定义如下：

$$\varphi = [\dot{x}_d + \ddot{e} \quad \ddot{x}_d - K_v \dot{e} \quad \dot{x}_d + \dot{e} \quad x_d + e \quad 1] \quad (13)$$

带有自适应补偿项 $Q_A$ 和鲁棒补偿项 $Q_R$ 的ARC控制策略设计为

$$Q_y = Q_A + Q_R \quad (14)$$

为补偿系统中未建模的误差，自适应补偿项 $Q_A$ 设计为

$$Q_A = \varphi \hat{\eta}_d \quad (15)$$

$$\hat{\eta}_d = \hat{\eta}_d - \eta_d$$

其中， $\hat{\eta}_d$ 为系统不确定参数估计值与实际值的误差。 $Q_A$ 中的 $\hat{\eta}_d$ 为 $\eta_d$ 的估计，其更新方式为<sup>[18]</sup>

$$\hat{\eta}_d = \eta_{d0} + \int_0^T Proj(\Gamma\tau)dt \quad (16)$$

其中,  $\eta_{d0}$  为  $\eta_d$  的初值;  $\Gamma$  为正对角矩阵;  $\tau$  为自适应函数, 映射函数  $Proj_{\hat{\eta}_d}(\cdot)$  表示为

$$Proj_{\hat{\eta}_d}(\cdot) = \begin{cases} 0 & \hat{\eta}_{di} = \eta_{di\max}, \cdot i > 0 \\ 0 & \hat{\eta}_{di} = \eta_{di\min}, \cdot i < 0 \\ \cdot i & \text{其他} \end{cases} \quad (17)$$

为了消除式(7)中时变分量  $\Delta$  和式(15)中不确定参数估计误差  $\tilde{\eta}_d$ , 设计鲁棒补偿项  $Q_R$  如下:

$$Q_R = Q_{R1} + Q_{R2} \quad Q_{R1} = -K_{RS}z \quad Q_{R2} = -\frac{1}{4\xi_1}\lambda^2z \quad (18)$$

其中,  $Q_{R1}$  为比例补偿量;  $K_{RS}$  为常数且  $K_{RS} > \eta_{2\max}/(2K_V)$ ;  $Q_{R2}$  为鲁棒补偿量;  $\xi_1$  为很小的正常数;  $\lambda$  为平滑函数, 且满足<sup>[19]</sup>

$$\lambda \geq \|\eta_{di\max} - \eta_{di\min}\| \|\varphi^T\| + \Delta_{\max} \quad (19)$$

式中:  $\Delta_{\max}$  为时变分量最大值。

$Q_{R2}$  满足以下性质<sup>[20]</sup>:

$$\left. \begin{aligned} z(Q_{R2} - \tilde{\eta}_d\varphi - \Delta) &\leq \xi_1 \\ zQ_{R2} &\leq 0 \end{aligned} \right\} \quad (20)$$

由  $Q_{R2}$  的性质可知, 可将不确定参数估计误差  $\tilde{\eta}_d$  或时变分量  $\Delta$  等不确定性对系统的影响减小到一个可忽略的范围。

根据式(18), 鲁棒补偿项整理为

$$Q_R = -K_S z \quad (21)$$

式中:  $K_S$  为控制参数, 且  $K_S \geq K_{RS} + \lambda^2/(4\xi_1)$ 。

结合式(10)、式(14)、式(15)和式(21), 总流量  $Q_y$  可表示为

$$Q_y = \varphi\hat{\eta}_d - K_S(\dot{e} + K_V e) \quad (22)$$

选取控制超参数满足  $K_S \geq K_{RS} + \lambda^2/(4\xi_1)$ ,  $K_V > B/(4K_{RS})$ , 则设计的自适应鲁棒控制器(式(14))具有以下性质:

**性质 1:** 系统中所有信号均有界。

**性质 2:** 系统在有限时间  $t_0$  后, 如果只存在参数不确定性 ( $\Delta = 0, \forall t \geq t_0$ ), 则除了满足性质 1 外, 系统还可以实现渐进跟踪, 即  $t \rightarrow \infty$  时,  $e \rightarrow 0, z \rightarrow 0$ 。

**证明:** 将式(10)和式(14)代入式(12), 得

$$\dot{V} = z(Q_y - \varphi\eta_d - \Delta) + eB\dot{e} \leq -(K_{RS}z^2 - eBz + BK_V e^2) + \xi_1 = -X^T Q_1 X + \xi_1 \quad (23)$$

故式(23)可以等效为矩阵形式:

$$\dot{V} \leq -X^T Q_1 X + \xi_1 \leq -\sigma_{\min}(Q_1) \|X\|_2^2 + \xi_1 = -2\sigma_{\min}(Q_1)/B + \xi_1 \quad (24)$$

$$X = [e \quad z]^T \quad Q_1 = \begin{bmatrix} K_{RS} & -B/2 \\ -B/2 & BK_V \end{bmatrix}$$

其中,  $\sigma_{\min}(Q_1)$  为矩阵  $Q_1$  的最小特征值, 当  $K_V >$

$B/(4K_{RS})$  时,  $Q_1$  为正定对角矩阵。

对式(24)两边积分变换后可得

$$V(t) \exp\left(\frac{2\sigma_{\min}(Q_1)}{B}t\right) \leq V(0) + \xi_1 \int_0^t \exp\left(\frac{2\sigma_{\min}(Q_1)}{B}s\right) ds \quad (25)$$

整理式(25), 得

$$V(t) \leq V(0) \exp\left(-\frac{2\sigma_{\min}(Q_1)}{B}t\right) + \frac{\xi_1 B}{2\sigma_{\min}(Q_1)} (1 - \exp\left(-\frac{2\sigma_{\min}(Q_1)}{B}t\right)) \quad (26)$$

由于 Lyapunov 函数  $V$  有上界, 由式(11)知, 切换函数  $z$  和跟踪误差  $e$  均有界, 故可推出系统中所有信号均有界, 性质 1 得证。

为证明性质 2, 建立一个新的 Lyapunov 函数  $V_\theta$ :

$$V_\theta = V + \frac{1}{2} \tilde{\eta}_d^T \Gamma^{-1} \tilde{\eta}_d \quad (27)$$

将  $\Delta = 0$  和式(20)中第二式代入式(27), 可得

$$\begin{aligned} \dot{V}_\theta &\leq -X^T Q_1 X + (z\varphi\tilde{\eta}_d + zQ_{R2}) + \tilde{\eta}_d^T \Gamma^{-1} \dot{\tilde{\eta}}_d \leq \\ &\quad -X^T Q_1 X - \tau^T \tilde{\eta}_d + \tilde{\eta}_d^T \Gamma^{-1} \dot{\tilde{\eta}}_d = \\ &\quad -X^T Q_1 X + \tilde{\eta}_d^T [\Gamma^{-1} Proj(\Gamma\tau) - \tau] \leq -X^T Q_1 X \end{aligned} \quad (28)$$

由式(28)得,  $\dot{V}_\theta$  为负定, 由式(11)和式(16)中的  $\Gamma$  得  $V_\theta$  是正定的, 则性质 2 得证, 当  $t \rightarrow \infty$  时,  $X \rightarrow 0$ , 得  $e \rightarrow 0, z \rightarrow 0$ 。

根据性质 1 和性质 2, 对强化学习 SAC 输出动作  $a_t = \{K_V, K_S\}$  数值范围进行了限制, 将  $K_S$  的调节范围限制为  $[K_{RS} + \lambda^2/(4\xi_1), +\infty)$ , 将  $K_V$  的调节范围限制为  $(B/(4K_{RS}), +\infty)$ , 这样既可以让上层 SAC 控制算法根据系统的当前状态调整下层控制参数  $K_S, K_V$ , 又能保证系统的渐进稳定。

## 2.2 异构双阀流量分配策略

当 ARC 控制算法计算出系统的合流量后, 进一步通过流量分配策略分解伺服阀与比例阀各自的输出流量。

伺服阀的流量与输入信号之间的关系可以近似为以下线性方程:

$$Q_s(t) = K_S u_s(t) \quad (29)$$

式中:  $Q_s$  为伺服阀的输出流量;  $K_S$  为伺服阀的流量-信号比例系数;  $u_s$  为伺服阀的输入信号。

比例阀由于存在死区等非线性因素, 会在实际的工作过程中存在响应滞后的问题。比例阀的流量与输入信号之间的关系可近似为以下线性分段函数:

$$Q_p(t) = \begin{cases} K_P(u_p(t) - Zu_{P\max}) & u_p(t) \geq Zu_{P\max} \\ 0 & |u_p(t)| < Zu_P \\ K_P(u_p(t) + Zu_{P\max}) & u_p(t) \leq -Zu_{P\max} \end{cases} \quad (30)$$

式中： $K_P$ 为比例阀的流量信号比例系数； $Z$ 为比例阀的死区系数； $Q_y$ 为比例阀的输出流量； $u_p$ 、 $u_{pmax}$ 分别为比例阀的实际输入信号与最大输入信号。

比例阀、伺服阀流量分配的基本原则如下：当系统需要的总流量  $Q_y < Q_{Smax}$  时，伺服阀的输出流量  $Q_S = Q_y$ 、比例阀的输出流量  $Q_P = 0$ ；反之， $Q_S = Q_{Smax}$ 、 $Q_P = Q_y - Q_{Smax}$ 。其中， $Q_{Smax}$ 为伺服阀的最大输出流量。

由于死区的存在，当比例阀的输入信号在  $\pm Zu_{pmax}$  内时，实际的输出流量均为0。为减少死区对比例阀输出流量连续性的影响，在伺服阀单独工作 ( $Q_y < Q_{Smax}$ ) 时，给予比例阀相应的控制信号补偿 (图3)。

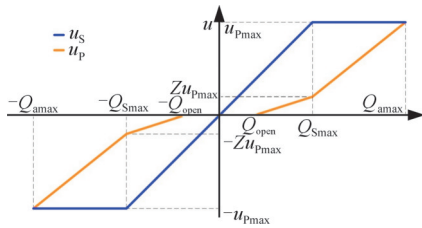


图3 伺服阀与比例阀的输入信号与系统总流量的映射关系

Fig.3 The relationship of the servo valve and proportional directional valve between the input signal and system the total flow

图3中， $Q_{amax}$ 为伺服阀和比例阀的最大合输出流量， $Q_{open}$ 为比例阀控制信号补偿起始流量，并且满足  $Q_{open} \geq 0$ 。定义异构双阀协调系统中伺服阀和比例阀与系统总流量的映射关系如下：

$$u_s = \begin{cases} Q_y/K_S & |Q_y| \leq Q_{Smax} \\ u_{Smax} & Q_y > Q_{Smax} \\ -u_{Smax} & Q_y < -Q_{Smax} \end{cases} \quad (31)$$

$$u_p = \begin{cases} 0 & |Q_y| \leq Q_{open} \\ [Zu_{pmax}/(Q_{Smax} - Q_{open})](Q_y - Q_{open}) & Q_{open} < |Q_y| \leq Q_{Smax} \\ Zu_{pmax} + [Q_y - Q_{Smax}]/K_P & Q_y > Q_{Smax} \\ -Zu_{pmax} + [Q_y + Q_{Smax}]/K_P & Q_y < -Q_{Smax} \end{cases} \quad (32)$$

### 2.3 SAC算法结构

SAC算法主要由 Actor 网络和 Critic 网络组成。Actor 网络为策略网络，作用是与环境进行交互并学习最优策略  $\pi_\phi$ ，进而根据当前的状态  $s_t$  产生一组优化控制参数  $a_t = \{K_S, K_V\}$ ，这组控制参数将传递到下层自适应鲁棒控制器中。

Critic 网络为价值网络，其作用是评价给定策略下“行为-状态”的优劣，根据当前的状态  $s_t$ 、控制参数  $a_t$  以及奖励值  $r_t$  评估 Actor 网络所生成的控制参数  $a_t$  的“行为-状态”价值，并为 Actor 网络

提供优化方向，指导其更新参数  $\phi$ 。

SAC 算法以熵最大化为目标来完成复杂任务的策略学习，其计算公式如下<sup>[21]</sup>：

$$J(\pi) = \sum_{t=0}^T E(s_t, a_t) \sim \rho_\pi [r(s_t, a_t) + \alpha H(\pi(\cdot|s_t))] \quad (33)$$

其中， $\alpha$  为温度系数。行为-状态价值函数  $Q(s_t, a_t)$  为

$$Q(s_t, a_t) = r(s_t, a_t) + \gamma E_{s_{t+1} \sim p} [V(s_{t+1})] \quad (34)$$

$$V(s_t) = E_{a_t \sim \pi} [Q(s_t, a_t) - \alpha \log \pi(a_t|s_t)] \quad (35)$$

通过最小损失函数来更新的 Critic 网络的计算公式如下<sup>[22]</sup>：

$$J_Q(\theta) = E_{(s_t, a_t) \sim D} \left[ \frac{1}{2} (Q_\theta(s_t, a_t) - \hat{Q}(s_t, a_t))^2 \right] \quad (36)$$

$$\hat{Q}(s_t, a_t) = r(s_t, a_t) + \gamma E_{s_{t+1} \sim p} [V_\theta(s_{t+1})] \quad (37)$$

由 Actor 网络更新的  $\pi_\phi$  策略为<sup>[23]</sup>

$$J_\pi(\phi) = E_{s_t \sim D, \epsilon_t \sim \pi_\phi} [\log \pi_\phi(f(\epsilon_t; s_t)|s_t) - Q_\theta(s_t, f(\epsilon_t; s_t))] \quad (38)$$

式中： $\epsilon_t$  为输入的噪声向量。

温度系数  $\alpha$  的更新方式如下：

$$J(\alpha) = E_D [-\alpha \log \pi(a_t|s_t) - \alpha H] \quad (39)$$

### 2.4 SAC-ARC 控制策略设计

图4所示的 SAC-ARC 控制策略以 SAC 算法作为上层控制策略<sup>[24-25]</sup>、ARC 控制作为下层控制策略。将上层 SAC 控制策略的输出动作  $a_t = \{K_V, K_S\}$  数值范围分别设置为  $(B/(4K_{RS}), +\infty)$  和  $[K_{RS} + \lambda^2/(4\hat{\epsilon}_1), +\infty)$ ，通过上层控制策略不断地与系统环境进行信息交互，修正下层控制策略中的控制参数  $K_V$  和  $K_S$ ，从而提升系统的控制性能。

Algorithm 1 伪代码如下所示：

**Algorithm 1.** SAC-ARC 控制策略训练过程算法伪代码  
初始化网络参数、全局记忆缓冲区以及训练回合数  $E$

**for** episode=1:  $E$  **do**

$\tau = 0$

**for**  $t=1, 2, \dots, T_t$  **do**

$$e(t) = x_p(t) - x_d(t)$$

$$Q_y(t) = -\varphi \hat{\eta}_d - K_S (\dot{e}(t) + K_V e(t))$$

$$u_s = \begin{cases} Q_y/K_S & |Q_y| \leq Q_{Smax} \\ u_{Smax} & Q_y > Q_{Smax} \\ -u_{Smax} & Q_y < -Q_{Smax} \end{cases}$$

$$u_p =$$

$$\begin{cases} 0 & Q_y < Q_{open} \\ [Zu_{pmax}/(Q_{Smax} - Q_{open})](Q_y - Q_{open}) & Q_{open} \leq |Q_y| < Q_{Smax} \\ Zu_{pmax} + [Q_y - Q_{Smax}]/K_P & Q_y > Q_{Smax} \\ -Zu_{pmax} + [Q_y + Q_{Smax}]/K_P & Q_y < -Q_{Smax} \end{cases}$$

**if**  $t=10, 20, \dots$  **do**

$$\tau = \tau + 1$$

获取当前环境状态  $s_t$

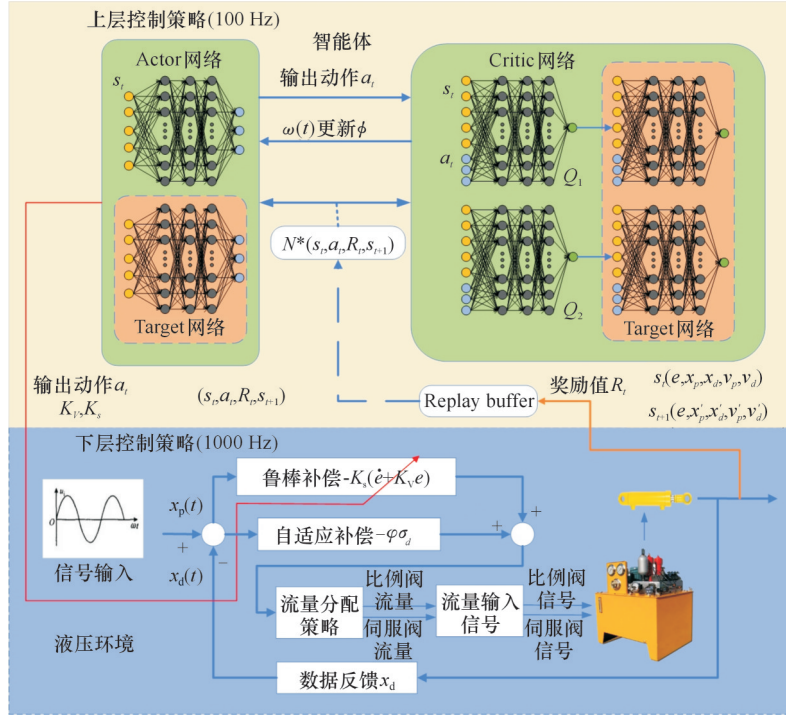


图4 基于SAC-ARC控制的异构双阀协调控制系统模型

Fig.4 The main framework of the dual-valve control system based on the SAC-ARC control strategy

$$a_\tau \leftarrow \pi_\phi(a_\tau | s_\tau)$$

$$s_{\tau+1} \leftarrow p(s_{\tau+1} | a_\tau, s_\tau)$$

$$r_\tau := r(s_\tau, a_\tau)$$

$$D \leftarrow \{s_\tau, a_\tau, r_\tau, s_{\tau+1}\}$$

**For** 每一更新步长 **then**

$$\theta_i \leftarrow \theta_i - \lambda_Q \hat{\nabla}_{\theta_i} J(\theta_i) \text{ for } i \in \{1, 2\}$$

$$\phi \leftarrow \phi - \lambda_\pi \hat{\nabla}_\phi J(\phi)$$

$$\alpha \leftarrow \alpha - \lambda_V \hat{\nabla}_\alpha J(\alpha)$$

更新目标网络参数

**End for**

**End if**

**End for**

**End for**

在模型训练阶段，SAC算法输出动作  $a_\tau = \{K_V, K_S\}$ ，调节ARC算法对应的控制参数。被控系统通过式(22)计算系统的总流量；根据式(31)和式(32)计算流量与伺服阀和比例阀的输入信号关系，从而控制液压缸的精确运动。当采样时间为10的倍数时，上层控制策略在观测当前状态  $s_\tau$ ，执行动作  $a_\tau$  并修正ARC控制算法的控制参数  $K_V$  和  $K_S$  后，计算出当前的奖励值  $r_\tau$  预测状态  $s_{\tau+1}$ ，然后把状态  $s_\tau, s_{\tau+1}$ 、奖励值  $r_\tau$  及动作  $a_\tau$  存入经验池中，最后通过式(36)、式(38)、式(39)更新参数  $\theta, \phi$  和  $\alpha$ 。

Algorithm 2 伪代码如下所示：

**Algorithm 2.** SAC-ARC控制策略运行过程算法伪代码  
 加载完成训练后的SAC算法网络模型  
 $\tau = 1; t = 0$

**Repeat**

$$e(t) = x_p(t) - x_d(t)$$

$$Q_y(t) = -\varphi \hat{\eta}_d - K_S(\dot{e}(t) + K_V e(t))$$

$$u_S = \begin{cases} Q_y/K_S & |Q_y| \leq Q_{Smax} \\ u_{Smax} & Q_y > Q_{Smax} \\ -u_{Smax} & Q_y < -Q_{Smax} \end{cases}$$

$$u_P = \begin{cases} 0 & Q_y < Q_{open} \\ [Zu_{Pmax}/(Q_{Smax} - Q_{open})](Q_y - Q_{open}) & Q_{open} \leq |Q_y| < Q_{Smax} \\ Zu_{Pmax} + [Q_y - Q_{Smax}]/K_P & Q_y > Q_{Smax} \\ -Zu_{Pmax} + [Q_y + Q_{Smax}]/K_P & Q_y < -Q_{Smax} \end{cases}$$

**if**  $t = 10, 20, \dots$  **do**

$$\tau = \tau + 1$$

加载当前环境状态  $s_\tau$

$$a_\tau \leftarrow \pi_\phi(a_\tau | s_\tau)$$

$$s_{\tau+1} \leftarrow p(s_{\tau+1} | a_\tau, s_\tau)$$

$$r_\tau := r(s_\tau, a_\tau)$$

$$D \leftarrow \{s_\tau, a_\tau, r_\tau, s_{\tau+1}\}$$

**End if**

$$t = t + 1$$

**Until**  $t \geq T_f$

在系统运行阶段时，上层SAC算法加载训练好的网络模型参数，在每个更新步长  $\tau$  根据实际的环境状态  $s_\tau$  的变化，输出动作  $a_\tau$  不断地调整ARC控制的参数，以增强系统的控制性能与鲁棒性。

### 2.5 奖励函数设计

奖励函数用于评价智能体在每一回合的训练效果，并引导智能体往奖励值大的方向去更新网

络模型参数,以不断优化被控系统的控制性能。本文的奖励函数围绕以下目标进行设计:①跟踪误差最小化。减小系统的跟踪误差从而提高系统的跟踪精度。②防止系统振荡。当液压系统出现内外干扰时,通过系统的加速度的负奖励来限制系统的振荡,从而使系统快速恢复稳定。综上,本文设计的奖励函数如下:

$$\left. \begin{aligned} r(s) &= r_1 + r_2 \\ r_1 &= k_1 |e_t| \\ r_2 &= k_2 |a| \end{aligned} \right\} \quad (40)$$

式中: $e_t$ 为当前时刻的液压缸活塞杆跟踪误差; $t$ 为采样时间; $a$ 为当前时刻的液压缸活塞杆加速度; $k_1$ 、 $k_2$ 为负增益系数。

### 3 结论

1)本文建立了异构双阀电液伺服系统的数学模型,在此基础上构建了异构双阀协调控制策略,并进一步从理论上证明了在该控制策略作用下的异构双阀电液伺服系统的稳定性。

2)本文提出了一种异构双阀协调控制策略SAC-ARC。一方面通过设计的流量分配策略减少比例阀与伺服阀在切换时产生的瞬态误差;另一方面通过SAC算法学习电液伺服系统的动态非线性特性,从而实现了ARC算法控制参数的高效动态调节。

3)本文所提出的SAC-ARC控制策略以强化学习SAC算法作为上层控制策略、ARC控制算法作为下层控制策略。上层控制策略周期性地与液压系统环境进行交互,并不断地修正下层ARC控制策略中相关的控制参数。

由于篇幅所限,本文未涉及仿真与实验验证部分,后续将验证所提SAC-ARC控制策略在不同工况下的有效性与鲁棒性。

#### 参考文献:

[1] YANG Huayong, SHI Hu, GONG Guofang, et al. Electro-hydraulic Proportional Control of Thrust System for Shield Tunneling Machine[J]. Automation in Construction, 2009, 18(7): 950-956.

[2] NGUYEN M T, DANG T D, AHN K K. Application of Electro-hydraulic Actuator System to Control Continuously Variable Transmission in Wind Energy Converter[J]. Energies, 2019, 12(13): 2499.

[3] QianLYU, YU Xiaoling, MA Haihui, et al. Applications of Machine Learning to Reciprocating Compressor Fault Diagnosis: a Review[J]. Processes, 2021, 9(6): 909.

[4] BAI Yanhong, QUAN Long. Improving Electro-hydraulic System Performance by Double-valve Actuation[J]. Transactions of the Canadian Society for Mechanical Engineering, 2016, 40(3): 289-301.

[5] YU Shaojuan, SONG Junjun. Iterative Learning Control of Double Servo Valve Controlled Electro Hydraulic Servo System[C]//2011 Seventh International Conference on Computational Intelligence and Security. IEEE, 2011: 278-282.

[6] SU Shijie, XUE Ting, CHEN Yun, et al. Harmonic Control of a Dual-valve Hydraulic Servo System with Dynamically Allocated Flows[J]. Asian Journal of Control, 2023, 25(3): 1939-1956.

[7] 焦宗夏, 吴帅, 李洋, 等. 液压元件及系统智能化发展现状及趋势思考[J]. 机械工程学报, 2023, 59(20): 357-384.

JIAO Zongxia, WU Shuai, LI Yang, et al. Development Status and Trends of the Intelligence of Hydraulic Components and Systems[J]. Journal of Mechanical Engineering, 2023, 59(20): 357-384.

[8] 郭具涛, 吕佑龙, 戴铮, 等. 基于复合规则和强化学习的混流装配线调度方法[J]. 中国机械工程, 2023, 34(21): 2600-2606.

GUO Jutao, LYU Youlong, DAI Zheng, et al. Compound Rules and Reinforcement Learning Based Scheduling Method for Mixed Model Assembly Lines [J]. China Mechanical Engineering, 2023, 34(21): 2600-2606.

[9] CORONATO A, NAEEM M, de PIETRO G, et al. Reinforcement Learning for Intelligent Healthcare Applications: a Survey [J]. Artificial Intelligence in Medicine, 2020, 109: 101964.

[10] 石晴晴, 张润锋, 张连洪, 等. 基于强化学习算法的水下滑翔机路径跟踪研究[J]. 中国机械工程, 2023, 34(9): 1100-1110.

SHI Qingqing, ZHANG Runfeng, ZHANG Lianhong, et al. Research on Underwater Gliders Path Tracking Based on Reinforcement Learning Algorithm[J]. China Mechanical Engineering, 2023, 34(9): 1100-1110.

[11] CHEN Pengzhan, HE Zhiqiang, CHEN Chuanxi, et al. Control Strategy of Speed Servo Systems Based on Deep Reinforcement Learning[J]. Algorithms, 2018, 11(5): 65.

[12] HE Jianhui, SU Shijie, WANG Hairong, et al. Online PID Tuning Strategy for Hydraulic Servo Control Systems via SAC-based Deep Reinforcement Learning[J]. Machines, 2023, 11(6): 593.

[13] YUAN Xiaoming, WANG Yu, ZHANG Ruicong, et al. Reinforcement Learning Control of Hydraulic

- Servo System Based on TD3 Algorithm[J]. *Machines*, 2022, 10(12): 1244.
- [14] YU Xinyi, FAN Yuehai, XU Siyu, et al. A Self-adaptive SAC-PID Control Approach Based on Reinforcement Learning for Mobile Robots[J]. *International Journal of Robust and Nonlinear Control*, 2022, 32(18): 9625-9643.
- [15] ZHUANG Huixuan, SUN Qinglin, CHEN Zengqiang. Sliding Mode Control for Electro-hydraulic Proportional Directional Valve-controlled Position Tracking System Based on an Extended State Observer[J]. *Asian Journal of Control*, 2021, 23(4): 1855-1869.
- [16] HE Jianhui, ZHOU Lijun, LI Cunjun, et al. Control Strategy of Hydraulic Servo Control Systems Based on the Integration of Soft Actor-Critic and Adaptive Robust Control[J]. *IEEE Access*, 2024, 12: 63629-63643.
- [17] 苏世杰, 游有鹏, 齐继阳, 等. 电液伺服试验机力控系统负载刚度自适应控制[J]. *控制理论与应用*, 2018, 35(4): 429-437.
- SU Shijie, YOU Youpeng, QI Jiyang, et al. Load Rigidity Adaptive Control of Electro-hydraulic Servo Universal Testing Machine Force Control System [J]. *Control Theory & Applications*, 2018, 35(4): 429-437.
- [18] CHEN Zheng, YAO Bin, WANG Qingfeng.  $\mu$ -synthesis-based Adaptive Robust Control of Linear Motor Driven Stages with High-frequency Dynamics; a Case Study[J]. *IEEE/ASME Transactions on Mechatronics*, 2015, 20(3): 1482-1490.
- [19] YAO Jianyong, JIAO Zongxia, YAO Bin, et al. Nonlinear Adaptive Robust Force Control of Hydraulic Load Simulator[J]. *Chinese Journal of Aeronautics*, 2012, 25(5): 766-775.
- [20] LI Chao, CHEN Zheng, YAO Bin. Adaptive Robust Synchronization Control of a Dual-linear-motor-driven Gantry with Rotational Dynamics and Accurate Online Parameter Estimation[J]. *IEEE Transactions on Industrial Informatics*, 2018, 14(7): 3013-3022.
- [21] TANG Hengliang, WANG Anqi, XUE Fei, et al. A Novel Hierarchical Soft Actor-Critic Algorithm for Multi-logistics Robots Task Allocation [J]. *IEEE Access*, 2021, 9: 42568-42582.
- [22] LEE M H, MOON J. Deep Reinforcement Learning-based Model-free Path Planning and Collision Avoidance for UAVs: a Soft Actor-Critic with Hindsight Experience Replay Approach[J]. *ICT Express*, 2023, 9(3): 403-408.
- [23] DING Feng, MA Guanfeng, CHEN Zhikui, et al. Averaged Soft Actor-Critic for Deep Reinforcement Learning [J]. *Complexity*, 2021, 2021(1): 6658724.
- [24] WONG C C, CHIEN S Y, FENG H M, et al. Motion Planning for Dual-arm Robot Based on Soft Actor-critic [J]. *IEEE Access*, 2021, 9: 26871-26885.
- [25] CHU Zhenzhong, WANG Fulun, LEI Tingjun, et al. Path Planning Based on Deep Reinforcement Learning for Autonomous Underwater Vehicles under Ocean Current Disturbance[J]. *IEEE Transactions on Intelligent Vehicles*, 2023, 8(1): 108-120.

(编辑 陈勇)

作者简介: 苏世杰\*(通信作者), 男, 1981年生, 教授、博士。研究方向为电液伺服控制、智能控制以及仿生机械设计等。E-mail: sushijie@just.edu.cn.

#### 本文引用格式:

苏世杰, 程泳钦, 胡毅, 等. 基于强化学习自适应鲁棒控制的异构双阀协调控制策略研究[J]. *China Mechanical Engineering*, 2025, 36(10): 2335-2342.

SU Shijie, CHENG Yongqin, HU Yi, et al. Coordination Control of Dual-valve Electrohydraulic Servo Systems Based on Integration of Reinforcement Learning and Adaptive Robust Control[J]. *China Mechanical Engineering*, 2025, 36(10): 2335-2342.

(上接第 2334 页)

- [19] GUO M, XIA W, WU C, et al. A Surface Quality Prediction Model Considering the Machine-tool-material Interactions [J]. *The International Journal of Advanced Manufacturing Technology*, 2024, 131(7): 3937-3955.
- [20] CAI Y, YANG Y, WANG Y, et al. Model for Surface Topography Prediction in the Ultra-precision Grinding of Silicon Wafers Considering Volumetric Errors[J]. *Measurement*, 2024, 234: 114825.

(编辑 袁兴玲)

作者简介: 闫祖龙, 男, 1999年生, 硕士研究生。研究方向为基于深度学习的精密加工光学元件的表面形貌预测和性能分析。E-mail: yanzulong@njfu.edu.cn. 庞启龙\*(通信作者), 男, 1979年生, 副教授。研究方向为表面微观结构的分析与重构、parylene镀膜工艺与设备、分子动力学仿真。E-mail: qlpang@njfu.edu.cn.

#### 本文引用格式:

闫祖龙, 庞启龙, 熊建龙. 基于深度学习的KDP晶体三维已加工表面形貌预测[J]. *中国机械工程*, 2025, 36(10): 2329-2334.

YAN Zulong, PANG Qilong, XIONG Jianlong. Prediction of the Three-dimensional Machined Surface Topography of KDP Crystals Based on Deep Learning[J]. *China Mechanical Engineering*, 2025, 36(10): 2329-2334.