

DOI: 10.19783/j.cnki.pspc.241244

融入 SAC 算法的光储微网混合储能自驱优级联自抗扰控制

周雪松¹, 张宇轩¹, 马幼捷¹, 王馨悦¹, 陶珑¹, 问虎龙^{2,3}

(1. 天津市新能源电力变换传输与智能控制重点实验室(天津理工大学), 天津 300384; 2. 天津瑞能电气有限公司, 天津 300385; 3. 天津瑞源电气有限公司, 天津 300308)

摘要: 母线电压稳定是实现新能源高水平消纳的重要前提。针对光储直流微电网混合储能系统因源荷不确定性扰动导致的母线电压波动问题, 提出一种融入深度强化学习柔性动作评价(soft actor-critic, SAC)算法的自驱优级联自抗扰控制策略。首先, 设计了级联扩张状态观测器来实时估计和补偿系统中的不确定性扰动, 以提升系统的扰动估计精度。其次, 针对系统建立了马尔可夫决策模型, 并设计了状态奖励与信息熵综合评估的 SAC 智能体, 融入控制器参数优化中。通过其在线学习和经验回放实现了控制参数的自驱优整定, 进一步提高了系统的抗扰性和鲁棒性。最后, 通过仿真实验对比了 3 种控制策略在典型工况下的控制性能, 验证了所提策略的有效性和优越性。

关键词: 光储直流微电网; 混合储能; 自抗扰控制; 深度强化学习; SAC 算法

Self-driven optimal cascade active disturbance rejection control for PV-storage microgrid with hybrid energy storage integrated with the SAC algorithm

ZHOU Xuesong¹, ZHANG Yuxuan¹, MA Youjie¹, WANG Xinyue¹, TAO Long¹, WEN Hulong^{2,3}

(1. Tianjin Key Laboratory of New Energy Power Conversion, Transmission and Intelligent Control (Tianjin University of Technology), Tianjin 300384, China; 2. Tianjin Ruineng Electric Co., Ltd., Tianjin 300385, China; 3. Tianjin Ruiyuan Electric Co., Ltd., Tianjin 300308, China)

Abstract: Bus voltage stability is a critical prerequisite for achieving high-level integration of new energy. To address the problem of bus voltage fluctuations in PV-storage DC microgrid with hybrid energy storage systems caused by source-load uncertainties, a self-driven optimal cascade active disturbance rejection control strategy integrated with the soft actor-critic (SAC) deep reinforcement learning algorithm is proposed. First, a cascaded extended state observer is designed to estimate and compensate for system uncertainties in real time, improving the accuracy of disturbance estimation. Then, a Markov decision model is established for the system, and a SAC agent, designed with a comprehensive evaluation of state rewards and information entropy, is integrated into the controller parameter optimization. By leveraging online learning and experience replay, the control parameters are autonomously and optimally tuned, further enhancing the system's disturbance rejection capability and robustness. Finally, the performance of three control strategies under typical working conditions is compared by simulation experiment, validating the effectiveness and superiority of the proposed approach.

This work is supported by the Major Program of National Natural Science Foundation of China (No. U24B6011).

Key words: PV-storage DC microgrid; hybrid energy storage; active disturbance rejection control; deep reinforcement learning; SAC algorithm

0 引言

随着全球能源危机日益严峻, 以光伏为主的分布式电源的开发利用已成为应对未来能源需求、实

现“双碳”目标的重要途径^[1-2]。直流微电网通过电力电子变换器将分布式电源与各类负荷高效互联, 显著提升了新能源的消纳水平^[3-4]。然而, 由于分布式电源具有间歇性和波动性的特点, 严重影响微电网的稳定运行, 因此需要配置储能设备来实现削峰填谷, 维持系统供需功率平衡^[5-6]。目前, 单一储能难以同时兼具高能量密度和高功率密度的特性, 因

基金项目: 国家自然科学基金重大项目资助(U24B6011);
国家自然科学基金重点项目资助(U23B20142)

此由蓄电池和超级电容组成的混合储能单元已被广泛应用于光储微网，针对储能单元控制技术的研究成为了焦点^[7-8]。

由于光储微网具有多变量、强耦合、非线性等特征^[9]，传统 PI 控制在维持母线电压稳定时面临响应速度慢、抗扰性能差等问题。为满足系统高性能控制需求，诸如滑模控制^[10]、模型预测控制^[11]、自抗扰控制^[12]等多种先进控制策略应运而生。其中，自抗扰控制(active disturbance rejection control, ADRC)因其对模型依赖性低、鲁棒性强的优点被广泛应用于微电网领域，其核心在于通过扩张状态观测器(extended state observer, ESO)实时观测总扰动，并通过反馈即时补偿，因此具备较强的抗扰性。相比于非线性 ADRC 由于参数众多难以实现工程应用，线性自抗扰控制(linear active disturbance rejection control, LADRC)提高了控制算法的实用性，但其鲁棒性有所降低，因此，有许多学者对 LADRC 进行改进以提高其控制性能。文献[13]在 ESO 的设计中引入了总扰动微分量，增强了光储系统对低频扰动的抑制能力，但也恶化了系统对高频噪声的滤波性能。文献[14]设计了一种改进型滑模自抗扰控制策略，提高了系统的响应速度和抗扰能力，但该策略涉及的控制参数较多，难以根据实际工况整定至理想效果。因此为提升系统在复杂工况下的适应能力，一些研究引入了模糊控制算法^[15]、启发式算法^[16]、神经网络^[17]等人工智能手段对 ADRC 参数进行实时整定。文献[18]针对储能侧直流变换器，利用模糊神经网络在线优化 LADRC 参数，以解决工况不确定性导致的系统性能降低的问题。然而，以上智能算法对经验初值较为敏感，且通常仅针对特定运行状态输出恒定控制参数，形成静态控制律，难以适应微电网动态变化的环境，对系统鲁棒性和适应性的提升仍有局限性。

基于前述分析及问题剖析，本文针对光储微网混合储能设计了一种融合深度强化学习柔性动作评价(soft actor-critic, SAC)算法的级联自抗扰控制策略(cascaded LADRC with SAC, SAC-CLADRC)。首先，在前级 LESO 的基础上增加二级观测器，用于估计前级观测器的扰动估计残差，从而提高系统对复杂扰动的估计精度。同时，基于系统状态构建马尔可夫决策模型，并设计状态奖励与信息熵综合评估的 SAC 智能体，将其融入控制器参数优化过程，形成自驱优级联自抗扰控制策略。该策略与微电网环境交互学习，并将数据存入经验回放池中，无需依赖预定数据集，具有内在的自适应性，能够更有效地应对实际工况的不确定性。最后，通过 3 种控

制策略在光储微网典型工况下的仿真实验对比，验证了所提策略的有效性和优越性。

1 光储直流微电网系统结构

直流微电网整体结构如图 1 所示，光伏阵列、风机等分布式电源分别通过 Boost 变换器、AC-DC 整流器接入直流母线。由储能电池和超级电容组成的混合储能单元，分别通过双向 DC-DC 变换器与直流母线相连，实现能量的双向流动，从而平抑系统功率波动。负荷类型包括阻性负载和恒功率负载，其中恒功率负载又分为直流负载与交流负载。图 2 为本文所研究的光储直流微网简化模型。



图 1 直流微电网结构图

Fig. 1 Structure diagram of DC microgrid

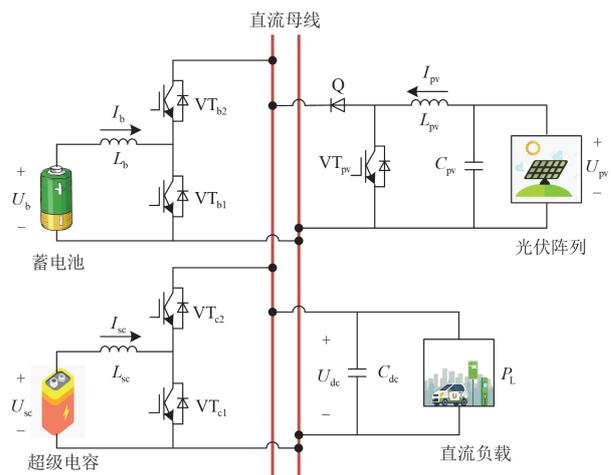


图 2 光储直流微网简化模型

Fig. 2 Simplified model of PV storage DC microgrid

图 2 中： U_{pv} 、 I_{pv} 、 C_{pv} 、 L_{pv} 、 VT_{pv} 、 Q 分别为光伏侧的输出电压、电感电流、滤波电容、电感、

开关管和二极管; U_b 、 I_b 、 L_b 、 VT_{b1} 和 VT_{b2} 分别为蓄电池侧的端电压、电感电流、电感、互补导通的开关管; U_{sc} 、 I_{sc} 、 L_{sc} 、 VT_{c1} 和 VT_{c2} 分别为超级电容侧的端电压、电感电流、电感、互补导通的开关管; P_L 为等效直流负载; C_{dc} 、 U_{dc} 分别为母线侧稳压电容、母线电压。

由于本文主要针对混合储能单元的控制策略展开研究, 光伏侧 Boost 变换器采用最大功率点跟踪算法控制, 根据图 2 可建立混合储能单元数学模型, 如式(1)所示。

$$\begin{cases} \frac{dI_b}{dt} = \frac{U_b}{L_b} - \frac{U_{dc}(1-D_b)}{L_b} \\ \frac{dI_{sc}}{dt} = \frac{U_{sc}}{L_{sc}} - \frac{U_{dc}(1-D_c)}{L_{sc}} \\ \frac{dU_{dc}}{dt} = \frac{I_b(1-D_b)}{C_{dc}} + \frac{I_{sc}(1-D_c)}{C_{dc}} - \frac{P_L}{C_{dc}U_{dc}} \end{cases} \quad (1)$$

式中: D_b 为开关管 VT_{b1} 的导通占空比; D_c 为开关管 VT_{c1} 的导通占空比。

2 混合储能 SAC-CLADRC 策略

混合储能系统整体控制策略如图 3 所示, 其中: I_{ref} 为总输出电流参考值; I_b^* 、 I_{sc}^* 分别为蓄电池、超级电容参考电流。本文所提的 SAC-CLADRC 策略主要针对电压环设计, 以提升直流母线电压的抗干扰能力和系统的鲁棒性能, 电流环采用传统 PI 控制以实现系统动态响应。当忽略线路阻抗时, 功率分配可等效为电流分配, 通过低通滤波环节对总输出电流参考值进行解耦^[19], 由蓄电池响应低频分量, 同时将剩余分量交由超级电容来补偿, 包含高频功率分量以及蓄电池未能及时响应的功率分量。

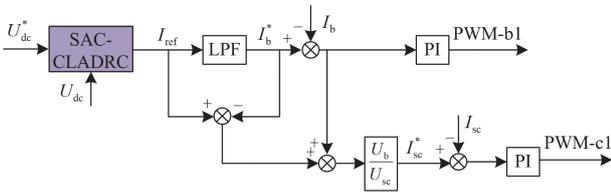


图 3 混合储能整体控制策略

Fig. 3 Overall control strategy for hybrid energy storage

2.1 级联型自抗扰控制器设计

通常对于一个含扰的 n 阶系统可描述为

$$y^{(n)} = g(y, y^{(1)}, y^{(2)}, \dots, y^{(n-1)}, w, t) + bu \quad (2)$$

式中: y 为系统输出; b 为控制量增益, 通常用近似估计值 b_0 来表示; u 为控制量; $g(\cdot)$ 为反映系统各状态变量间耦合关系的函数; w 为外部扰动。

根据一阶自抗扰范式, 将系统内部不确定的高阶分量、外部扰动以及控制量增益不确定部分定义为广义总扰动 f , 可表示为

$$f = g(y, y^{(1)}, y^{(2)}, \dots, y^{(n-1)}, w, t) + (b - b_0)u - y^{(n)} + y^{(1)} \quad (3)$$

结合式(2)、式(3), 并定义状态变量为 $x_1 = y$ 、 $x_2 = f$, 则式(2)的高阶系统可描述为

$$\begin{cases} \dot{x}_1 = x_2 + b_0 u \\ \dot{x}_2 = \hat{f} \\ y = x_1 \end{cases} \quad (4)$$

根据线性扩张状态观测器(linear extended state observer, LESO)设计准则, 可设计 LESO1 为

$$\begin{cases} \dot{z}_1 = z_2 + b_0 u - l_1 e_z \\ \dot{z}_2 = -l_2 e_z \\ e_z = z_1 - x_1 \end{cases} \quad (5)$$

式中: e_z 为 LESO1 的系统输出估计误差; z_1 、 z_2 分别为 x_1 、 x_2 的估计值; l_1 、 l_2 为 LESO1 的观测增益, 可通过带宽法^[20]设计, 即 $l_1 = 2\omega_0$, $l_2 = \omega_0^2$, ω_0 为观测器带宽。

结合式(4)、式(5)可得, LESO1 中由总扰动 f 到扰动观测误差 e_{z2} 的传递函数 $E_L(s)$ 为

$$E_L(s) = \frac{z_2(s) - f(s)}{f(s)} = -\frac{s^2 + 2\omega_0 s}{(s + \omega_0)^2} \quad (6)$$

若 f 取典型斜坡函数, 则 e_{z2} 的时域形式如式(7)所示。

$$e_{z2} = 2ke^{-\omega_0 t} / \omega_0 + kte^{-\omega_0 t} - 2k / \omega_0 \quad (7)$$

式中: k 为斜坡函数斜率。

由式(7)可知, LESO1 在斜坡扰动下仅依靠调节 ω_0 并不能实现零误差收敛, 仍然存在稳态观测偏差 $-2k / \omega_0$, 为提高观测器对于斜坡函数等复杂扰动的观测性能与收敛精度, 引入二级观测器 LESO2 与 LESO1 形成级联型观测器(Cascaded LESO, CLESO), LESO2 在前级扰动估计的基础上观测扰动估计残差, 数学模型如式(8)所示, 将 LESO1 的扰动估计值 z_2 作为已知扰动输入。

$$\begin{cases} \dot{n}_1 = n_2 + b_0 u + z_2 - \beta_1 e_n \\ \dot{n}_2 = -\beta_2 e_n \\ e_n = n_1 - y \end{cases} \quad (8)$$

式中: e_n 为 LESO2 的系统输出估计误差; n_1 为 y 的估计值; n_2 为 LESO1 的扰动估计残差的重构值; β_1 、 β_2 为 LESO2 的观测增益, 为简化参数配置, 与 LESO1 的增益设计一致。

CLESO 中由 f 到扰动估计误差 e_f 的传递函数 $E_{CL}(s)$ 为

$$E_{CL}(s) = \frac{z_2(s) + n_2(s) - f(s)}{f(s)} = -\frac{s^2(s + 2\omega_0)^2}{(s + \omega_0)^4} \quad (9)$$

同理取 f 为典型斜坡函数, e_f 的时域表达式如式(10)所示, 当 $t \rightarrow \infty$, CLESO 在斜坡扰动下可实现零误差收敛, 因此其具备更强的扰动估计性能。

$$e_f = -(kte^{-\omega_0 t} + k\omega_0 t^2 e^{-\omega_0 t} + k\omega_0^2 t^3 e^{-\omega_0 t}/6) \quad (10)$$

基于 CLESO 重构的系统状态变量, 可通过线性状态误差反馈率(linear state error feedback, LSEF)实时补偿, 使复杂系统在稳态时转化为积分串联标准型结构, LSEF 环节设计为

$$u = \frac{k_p(U_{dc}^* - n_1) - (z_2 + n_2)}{b_0} \quad (11)$$

式中: U_{dc}^* 为母线电压给定值; k_p 为控制器增益, 通过带宽法表示为 $k_p = \omega_c$, ω_c 为控制器带宽。

因此, 级联型线性自抗扰控制器的系统控制框图如图 4 所示。

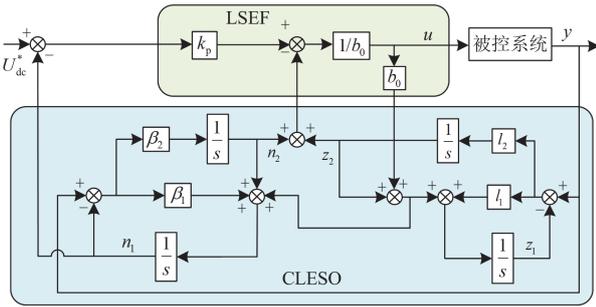


图 4 CLADRC 的结构框图

Fig. 4 Structure diagram of the CLADRC

2.2 混合储能系统马尔可夫决策模型

马尔可夫决策过程^[21](Markov decision process, MDP)是解决控制器增益参数优化问题的有效方式, 一个 MDP 由状态空间 S 、动作空间 A 、状态转移概率函数 P 、折扣因子 γ 以及奖励函数 R 组成, 可记为 $\{S, A, P, \gamma, R\}$ 。所提 SAC-CLADRC 算法将整个光储微网混合储能系统视作环境, 以母线电压跟踪给定值的效果为奖励, 评估 SAC 智能体调参的优劣, 通过智能体与环境的持续交互优化策略, 直至奖励函数收敛, 最终实现最优控制策略的自适应学习, 即最优 CLADRC 参数。根据母线电压稳定的控制目标, 可定义状态空间、动作空间、奖励函数、状态转移过程如下。

1) 状态空间 S : 为更全面地表征系统状态, 令母线电压 $U_{dc}(t)$ 、电压跟踪误差 $e_{dc}(t)$ 为确定系统状

态的底层信号, 将其延迟信号 $U_{dc}(t-1)$ 、 $e_{dc}(t-1)$ 视为补充状态, 因此状态 $s_t \in S$ 可设计为

$$s_t = \{U_{dc}(t), e_{dc}(t), U_{dc}(t-1), e_{dc}(t-1)\} \quad (12)$$

2) 动作空间 A : 动作 $a_t \in A$ 根据 CLADRC 参数设计为二维变量, 如式(13)所示。

$$a_t = \{\omega_0, \omega_c\} \quad (13)$$

3) 奖励函数 R : r_t 为智能体的即时奖励, 对母线电压波动量、调节时间等性能指标进行定量分析, 当 r_t 达到最大时, 优化结果最佳, 可设计为

$$r_t = \begin{cases} -15|e_{dc}(t)| & |e_{dc}(t)| \geq 10 \\ -10|e_{dc}(t)| & 4 \leq |e_{dc}(t)| < 10 \\ -5|e_{dc}(t)| & 0.5 \leq |e_{dc}(t)| < 4 \\ -|e_{dc}(t)| & \text{other} \end{cases} \quad (14)$$

4) 状态转移函数 P : 系统状态变化过程可以描述为

$$\begin{cases} a_t = \pi(s_t) \\ s_{t+1} = s_t + a_t \end{cases} \quad (15)$$

式中: $\pi(s_t)$ 为策略函数, 表示从状态 s_t 到动作 a_t 的映射关系。

2.3 SAC-CLADRC 算法

与传统深度强化学习算法仅专注于最大化累积奖励期望的策略不同, SAC 算法还在学习目标中引入信息熵指标, 从而量化策略的不确定性, 使其在面对复杂环境时, 具备更强的探索能力和鲁棒性。因此, 将该算法引入光储微网自抗扰控制参数的自适应优化过程中, 能够有效应对工况变化、系统参数波动等不确定性因素的影响。SAC 算法的最优策略 π^* 定义如式(16)所示。

$$\begin{cases} \pi^* = \arg \max_{\pi} E_{(s_t, a_t) \sim \rho_{\pi}} \left[\sum_{t=0}^{\infty} \gamma^t [r(s_t, a_t) + \alpha H(\pi(\cdot | s_t))] \right] \\ H(\pi(\cdot | s_t)) = E[-\log \pi(\cdot | s_t)] \end{cases} \quad (16)$$

式中: E 为数学期望; ρ_{π} 表示由 π 产生的所有状态-动作分布, $(s_t, a_t) \sim \rho_{\pi}$ 表示 s_t 和 a_t 服从分布 ρ_{π} ; $r(s_t, a_t)$ 表示在 s_t 和 a_t 下的即时奖励; $H(\pi(\cdot | s_t))$ 表示策略 π 的熵值, 熵值越高表明 π 的不确定性越大; α 为温度系数, 用于平衡熵对奖励的重要性。

SAC 的价值函数通过引入熵正则项的柔性贝尔曼方程进行迭代求解, 如式(17)所示。

$$\begin{cases} Q(s_t, a_t) = E_{s_{t+1} \sim P} [r(s_t, a_t) + \gamma V(s_{t+1})] \\ V(s_t) = E_{a_t \sim \pi} [Q(s_t, a_t) + \alpha H(\pi(\cdot | s_t))] \end{cases} \quad (17)$$

式中: $Q(s_t, a_t)$ 为柔性动作价值函数, 表示在 (s_t, a_t) 下对累积奖励和策略熵的期望; $V(s_t)$ 为柔性状态价值函数, 表示在 s_t 下对累积奖励和策略熵的期望。

SAC 智能体中含 1 个策略网络(Actor network)、2 个 Q 价值网络(Critic network)以及经验回放池, 其中每个价值网络又包含 1 个目标网络, 用于更新价值网络参数。策略网络负责与环境进行交互, 价值网络用于辅助策略网络训练, 指导其执行更优的策略。SAC-CLADRC 算法框架如图 5 所示。

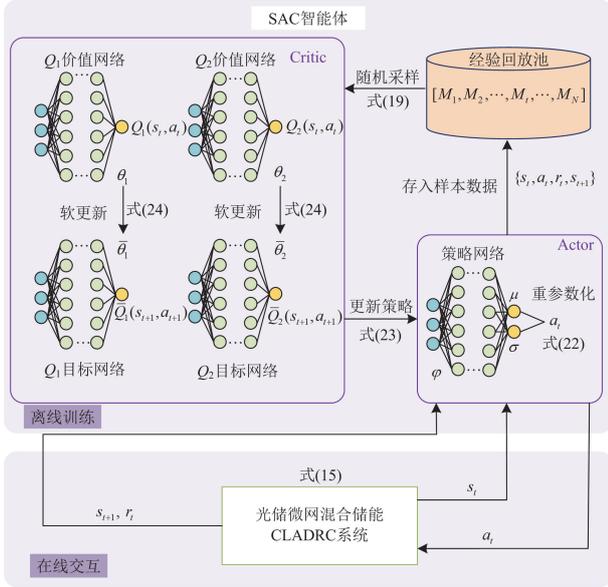


图 5 SAC-CLADRC 算法框架

Fig. 5 Framework of the SAC-CLADRC algorithm

图 5 中, SAC 智能体与系统在线交互获得一组样本数据存入经验池中, 当经验池中的样本数据足够时, 随机采样小批量样本离线更新 $Q_i (i=1,2)$ 价值网络和策略网络参数, 同时根据策略优化程度自适应优化 α 。目标网络参数的优化采用定期软更新策略实现渐进式调整, 更新过程如下。

价值网络参数 θ 可通过最小化柔性贝尔曼误差来更新, 损失函数 $J_Q(\theta)$ 为

$$\begin{cases} J_Q(\theta) = E_{(s_t, a_t) \sim M} \left[\frac{1}{2} (Q_\theta(s_t, a_t) - (r(s_t, a_t) + \gamma E_{s_{t+1} \sim P} [V_{\bar{\theta}}(s_{t+1})]))^2 \right] \\ V_{\bar{\theta}}(s_t) = E_{a_t \sim \pi} [Q_{\bar{\theta}}(s_t, a_t) - \alpha \log \pi(a_t | s_t)] \end{cases} \quad (18)$$

式中: M 表示经验元组; $\bar{\theta}$ 表示目标网络的参数。

参数 θ_i 通过式(19)所示的梯度下降算法进行迭代更新。

$$\theta_i \leftarrow \theta_i - \lambda_Q \hat{\nabla}_{\theta_i} J_Q(\theta_i), i \in \{1, 2\} \quad (19)$$

式中: λ_Q 为价值网络的学习率; $\hat{\nabla}_{\theta_i}$ 表示对 θ_i 求导。

策略网络参数的更新中, 通过最小化预期的 KL 散度^[22]来优化策略函数, 如式(20)所示。

$$\pi_{\text{new}} = \arg \min_{\pi' \in \Pi} D_{\text{KL}} \left(\pi'(\cdot | s_t) \left\| \frac{\exp(Q^{\pi_{\text{old}}}(s_t, \cdot) / \alpha)}{Z^{\pi_{\text{old}}}(s_t)} \right\| \right) \quad (20)$$

式中: π_{new} 表示新策略; D_{KL} 为 KL 散度; Π 为可行的策略分布集合; $Q^{\pi_{\text{old}}}(s_t, \cdot)$ 表示旧策略的价值函数; $Z^{\pi_{\text{old}}}(s_t)$ 为标准归一化系数。

根据式(20)可得更新策略网络的损失函数为

$$J_\pi(\varphi) = E_{s_t \sim M, a_t \sim \pi_\varphi} [\alpha \log \pi_\varphi(a_t | s_t) - Q_\theta(s_t, a_t)] \quad (21)$$

式中: φ 为策略网络参数。

由于 a_t 是从策略网络输出的概率分布中采样获取的随机变量, 无法对其求导, 因此采用重参数化技巧^[23]重构为式(22)。

$$a_t = f_\varphi(\varepsilon_t; s_t) = f_\varphi^\mu(s_t) + \varepsilon_t \odot f_\varphi^\sigma(s_t) \quad (22)$$

式中: ε_t 表示输入噪声量, 从标准正态分布中采样; f_φ 相当于重构的 π_φ ; μ 和 σ 分别表示动作分布的均值和方差。

为减小策略优化过程中价值的过估计, 选择较小的 Q 值评估策略。 φ 的更新过程如式(23)所示。

$$\begin{cases} \hat{\nabla}_\varphi J_\pi(\varphi) = \hat{\nabla}_\varphi [\alpha \log \pi_\varphi(f_\varphi(\varepsilon_t; s_t) | s_t) - \min_{i=1,2} Q_{\theta_i}(s_t, f_\varphi(\varepsilon_t; s_t))] \\ \varphi \leftarrow \varphi - \lambda_\pi \hat{\nabla}_\varphi J_\pi(\varphi) \end{cases} \quad (23)$$

式中: λ_π 为策略网络的学习率。

目标网络参数 $\bar{\theta}$ 通过式(24)更新, 而温度系数 α 的损失函数 $J(\alpha)$ 及更新过程如式(25)所示。

$$\bar{\theta}_i \leftarrow \tau \theta_i + (1 - \tau) \bar{\theta}_i, i \in \{1, 2\} \quad (24)$$

$$\begin{cases} J(\alpha) = E_{a_t \sim \pi_t} [-\alpha \log(\pi_t(a_t | s_t) - \alpha \bar{H})] \\ \alpha \leftarrow \alpha - \lambda_\alpha \hat{\nabla}_\alpha J(\alpha) \end{cases} \quad (25)$$

式中: τ 为软更新因子; \bar{H} 为目标熵值, 设计为动作维数的负数; λ_α 为策略熵的学习率。

2.4 SAC 智能体设计

SAC 智能体的设计主要包括 Actor、Critic 神经网络结构以及超参数, 由于本文是通过神经网络拟合非线性的策略函数和价值函数, 无需输入图像, 因此 Actor、Critic 网络均设计为全连接层神经网络, 为增强神经网络对于数据的泛化能力, Actor 与 Critic 网络均设计为两层隐藏层, 每层含有 128 个神经元, 同时为防止网络更新过程中梯度消失和爆炸, 选择 ReLU 函数作为激活函数。神经网络结构如图 6 所示, Actor 网络根据输入的四维状态 s_t 输出随机性动作参数 a_t , Critic 网络根据输入的 s_t 和 a_t 输出价值, 以评估状态 s_t 下执行 a_t 的优劣。

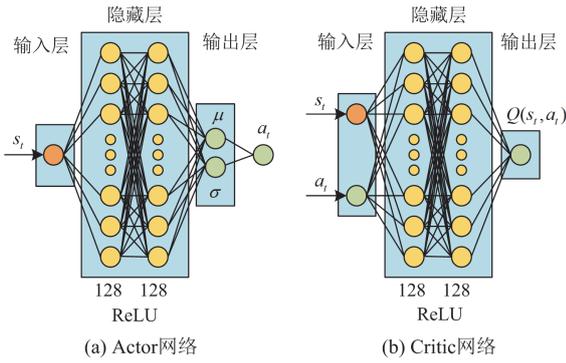


图6 Actor-Critic 神经网络结构

Fig. 6 Actor-Critic neural network structure

在 SAC 智能体参数设计中,主要包括神经网络学习率、折扣因子以及经验池大小等,增大学习率可加快智能体寻优速度,但过大的学习率会导致网络参数难以收敛至最优值,表 1 为本文经过多次调试最终设计的 SAC 智能体参数。

表 1 SAC 智能体参数

Table 1 SAC agent parameters

参数	值
Actor 网络学习率 λ_π	0.0001
Critic 网络学习率 λ_Q	0.001
温度系数 α 学习率 λ_α	0.0001
折扣因子 γ	0.99
经验池大小	1 000 000
小批量采样样本	128
软更新因子 τ	0.001
优化器	Adam
训练最大时间步长	100

3 SAC-CLADRC 性能分析

3.1 CLADRC 稳定性分析

根据式(4)、式(5)和式(8),可得 CLADRC 的估计误差传递函数为

$$e_z(s) = \frac{-s}{G}(sy(s) - b_0u(s)) \quad (26)$$

$$e_n(s) = \frac{-(s^3 + l_1s^2)}{GH}(sy(s) - b_0u(s)) \quad (27)$$

$$e_f(s) = \frac{-(s^3 + l_1s^2)(s + \beta_1)}{GH}(sy(s) - b_0u(s)) \quad (28)$$

式中: $G = s^2 + l_1s + l_2$; $H = s^2 + \beta_1s + \beta_2$; e_f 为扰动估计误差, $e_f = z_2 + n_2 - f$ 。

对于式(26),由带宽法配置 LESO1 的增益 l_1 、 l_2 ,将极点配置于 s 域左半平面,即可保证系统稳定。由式(27)、式(28)可得 LESO2 和 e_f 的特征方程为

$$GH = s^4 + c_1s^3 + c_2s^2 + c_3s + c_4 \quad (29)$$

式中: $c_1 = l_1 + \beta_1$; $c_2 = l_2 + \beta_2 + l_1\beta_1$; $c_3 = l_2\beta_1 + l_1\beta_2$; $c_4 = l_2\beta_2$ 。

根据 Routh 判据,系统稳定的充要条件为

$$\begin{cases} c_1c_2 - c_3 > 0 \\ c_1(c_2c_3 - c_1c_4) - c_3^2 > 0 \end{cases} \quad (30)$$

根据 2.1 节的带宽法配置增益,可得式(30)成立,即特征方程所有极点均位于 s 域的左半平面,因此所设计的级联扩张状态观测器可保持稳定。

3.2 SAC 收敛性分析

SAC 算法迭代过程主要分为两个阶段,即策略评估和策略提升,二者交替优化从而学习一个最大化奖励期望和熵期望的策略。

引理 1: 对于一个固定的策略 π 和任意函数 $Q_0: S \times A \rightarrow R$ 且 $|A| < \infty$,存在贝尔曼备份算子 T^π ,满足 $Q_{k+1} = T^\pi Q_k$,则当 $k \rightarrow \infty$ 时, Q_k 将收敛到 π 的柔性 Q 值。

证明: 首先将式(17)改写为

$$T^\pi Q(s_t, a_t) = r(s_t, a_t) + \gamma E_{s_{t+1} \sim P} [V(s_{t+1})] \quad (31)$$

定义熵增奖励为

$$r_\pi(s_t, a_t) = r(s_t, a_t) + E_{s_{t+1} \sim P} [H(\pi(\cdot | s_{t+1}))] \quad (32)$$

则价值函数更新可表示为

$$Q(s_t, a_t) \leftarrow r_\pi(s_t, a_t) + \gamma E_{s_{t+1} \sim P, a_{t+1} \sim \pi} [Q(s_{t+1}, a_{t+1})] \quad (33)$$

当 $|A| < \infty$ 时可保证熵增奖励有界。由 Banach 不动点定理可知,迭代 $Q_{k+1} = T^\pi Q_k$,将收敛到 π 的唯一 Q 值^[24]。

引理 2: 对所有 $(s_t, a_t) \in S \times A$ 且 $|A| < \infty$,均有 $Q^{\pi_{old}}(s_t, a_t) \leq Q^{\pi_{new}}(s_t, a_t)$ 成立。

证明: 令 $\pi_{old} \in \Pi$,定义 $Q^{\pi_{old}}$ 和 $V^{\pi_{old}}$ 分别为对应柔性动作价值和状态价值,将式(20)改写为

$$\pi_{new}(\cdot | s_t) = \arg \min_{\pi' \in \Pi} J_{\pi_{old}}(\pi'(\cdot | s_t)) \quad (34)$$

由于 $\pi_{new} = \pi_{old} \in \Pi$ 成立,因此有

$$J_{\pi_{old}}(\pi_{new}(\cdot | s_t)) \leq J_{\pi_{old}}(\pi_{old}(\cdot | s_t)) \quad (35)$$

因归一化系数 $Z^{\pi_{old}}(s_t)$ 仅依赖于状态,式(35)可简化为

$$E_{a_t \sim \pi_{new}} [Q^{\pi_{old}}(s_t, a_t) - \alpha \log \pi_{new}(a_t | s_t)] \geq V^{\pi_{old}}(s_t) \quad (36)$$

因此,结合式(36)和引理 1,柔性贝尔曼方程为

$$\begin{aligned} Q^{\pi_{old}}(s_t, a_t) &= r(s_t, a_t) + \gamma E_{s_{t+1} \sim P} [V^{\pi_{old}}(s_{t+1})] \leq \\ &= r(s_t, a_t) + \gamma E_{s_{t+1} \sim P} [E_{a_{t+1} \sim \pi_{new}} [Q^{\pi_{old}}(s_{t+1}, a_{t+1}) - \\ &\quad \alpha \log \pi_{new}(a_{t+1} | s_{t+1})]] \leq \dots \leq Q^{\pi_{new}}(s_t, a_t) \end{aligned} \quad (37)$$

因此对于所有 $(s_i, a_i) \in S \times A$ 且 $|A| < \infty$, 均有 $Q^{\pi^{\text{old}}}(s_i, a_i) \leq Q^{\pi^{\text{new}}}(s_i, a_i)$ 成立。

策略迭代: 令 π_i 为迭代 i 次的策略, 根据引理 2 可知, Q^{π_i} 单调递增, 由于 Q^{π} 在 $\pi \in \Pi$ 上有界(奖励和熵均有界), 因此 Q^{π_i} 必然收敛到一个 π^* 。同时, 根据式(35)可得: 对于所有 $\pi \in \Pi$ 且 $\pi \neq \pi^*$, 必然存在 $J_{\pi^*}(\pi^*(\cdot|s_i)) < J_{\pi}(\pi(\cdot|s_i))$, 因此对于所有 $(s_i, a_i) \in S \times A$, 均可得到 $Q^{\pi^*}(s_i, a_i) > Q^{\pi}(s_i, a_i)$, 即 Π 上所有策略的柔性 Q 值均小于收敛策略的柔性价值, 所以 π^* 是最优的。

4 仿真实验验证

为验证本文所提 SAC-CLADRC 策略的可行性和优越性, 基于 Matlab/Simulink 平台搭建图 2 所示光储直流微电网混合储能系统仿真模型, 并使用 PI 控制、LADRC 进行性能对比。系统初始条件设定为: 光照强度为 1000 W/m^2 、环境温度为 $25 \text{ }^\circ\text{C}$, 光伏侧通过电导增量法控制, 实现最大功率点跟踪。光储系统参数如表 2 所示, 控制器参数如表 3 所示。

表 2 光储系统参数

Table 2 Parameters of the photovoltaic-storage system

参数	值	参数	值
母线参考电压 U_{dc}^*/V	750	蓄电池侧电感 L_b/mH	10
母线电容 C_{dc}/mF	5	蓄电池初始 SOC	50%
开关频率 f_s/kHz	10	超级电容端电压 U_{sc}/V	240
等效负载 P_L/kW	5	超级电容侧电感 L_{sc}/mH	10
蓄电池端电压 U_b/V	240	超级电容初始 SOC	70%

表 3 控制器参数

Table 3 Controller parameters

控制策略	电压环	电流环
PI	$K_{Pv}=1, K_{Iv}=20$	
LADRC	$\omega_o=500, \omega_c=80, b_0=200$	$K_{Pi}=200, K_{Ii}=0.5$
SAC-CLADRC	ω_o, ω_c (图 8), $b_0=200$	
LPF 时间常数	0.015	

4.1 SAC-CLADRC 预训练过程

在 SAC 智能体预训练之前, 为模拟实际微电网复杂多变的工况, 在环境中设置多组源侧不确定性因素(以光照强度变化为例)、荷侧不确定性因素(以负载投切为例)。为提高训练速度, 结合实际仿真设置动作临界值: $\omega_{o-\min}=100$ 、 $\omega_{c-\min}=60$, 训练过程中奖励变化曲线如图 7 所示。图 8 为平均奖励收敛后, SAC 智能体在训练 130 次时输出的自驱优带宽。

由图 7 可知, 整个预训练过程可分为两个阶段,

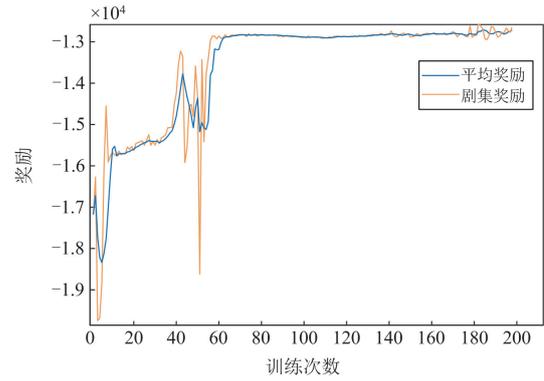


图 7 奖励变化曲线

Fig. 7 Reward variation curve

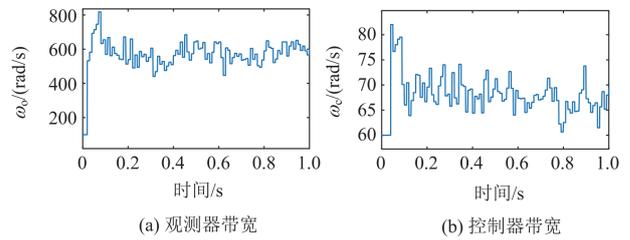


图 8 SAC-CLADRC 自驱优带宽

Fig. 8 Self-driven optimal bandwidth of SAC-CLADRC

分别为: SAC 智能体探索学习阶段(0~65 次)以及收敛阶段(65~195 次)。由图 8 可知, SAC 智能体根据已学习的当前最优策略输出适应于系统不同状态的动态带宽 (ω_o, ω_c) , 以增强系统在不确定工况下的适应性和鲁棒性。

4.2 光伏输出功率波动工况

在实际微网工况中, 光伏输出功率会随着光照强度、温度的改变而变化, 从而引起直流母线电压的波动。本节保持负载以 5 kW 恒功率运行, 改变光照强度使光伏阵列输出功率波动, 为模拟复杂型工况, 光照强度在 0.4 s 时由 1000 W/m^2 经斜坡变化为 500 W/m^2 , 在 0.8 s 和 1 s 时阶跃为 900 W/m^2 和 600 W/m^2 , 光伏输出功率如图 9 所示。

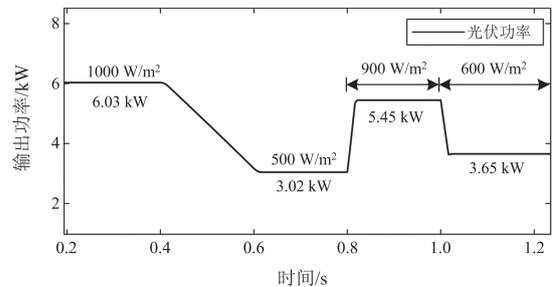
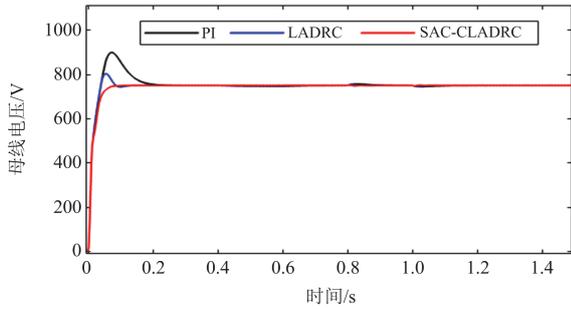


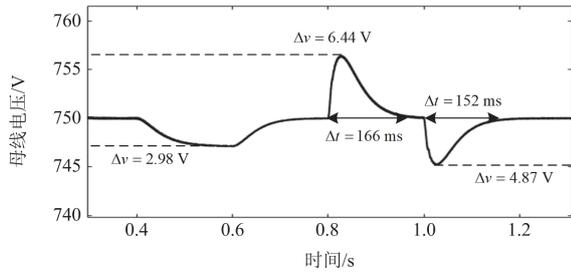
图 9 光伏阵列输出功率

Fig. 9 Output power of PV arrays

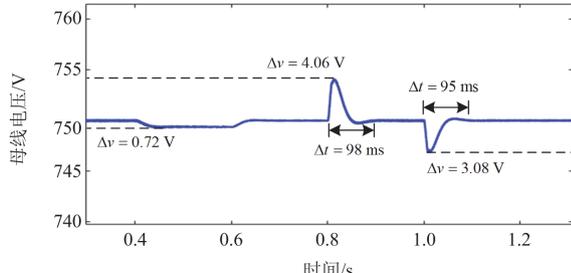
由图 9 可知,在 0.4~0.6 s 时段内,光伏输出功率由大于负载需求逐渐转变为小于负载需求,母线电压出现跌落现象,此时混合储能由吸收母线多余功率转变为补偿差额功率。在 0.8 s 和 1.0 s 时,根据供需功率原则,母线电压出现短暂的突增和跌落,混合储能吸收多余功率和补偿差额功率后恢复稳定。图 10 为母线电压在 3 种控制策略下的电压波动情况。



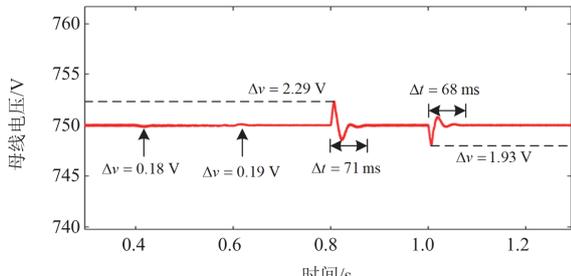
(a) 母线电压起始动态响应



(b) PI



(c) LADRC



(d) SAC-CLADRC

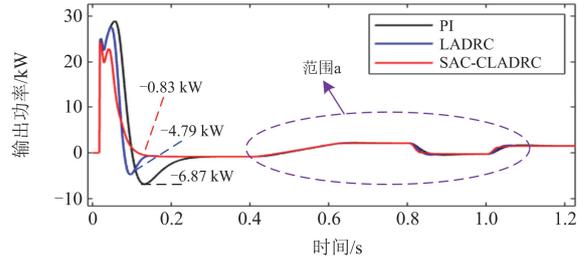
图 10 光伏功率波动时母线电压对比

Fig. 10 Comparison of bus voltage during PV power fluctuations

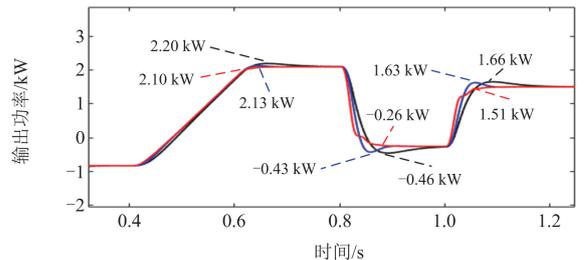
由图 10(a)可知,在 PI、LADRC 以及 SAC-CLADRC 策略下,母线电压起始动态响应的最大电压波动量分别为 148 V、54.1 V 和 0.2 V,调节时间分别为 0.22 s、0.15 s 和 0.1 s,由此可见,所提 SAC-CLADRC 策略在母线电压跟踪性能上表现出更加优异的控制效果。

通过对比分析图 10(b)—图 10(d)可知,在光照强度斜坡变化过程中,PI 控制与 LADRC 均存在着一定的母线电压偏差,直至光伏输出功率稳定时,母线电压才得以恢复至期望值,反观 SAC-CLADRC 策略仅在扰动出现和结束时产生 0.18 V 和 0.19 V 的电压波动量,母线电压随即迅速恢复稳定,实现对期望值的零误差跟踪。此外,在光照强度阶跃变化时,以 0.8 s 为例,相比于 PI、LADRC, SAC-CLADRC 策略的最大电压波动量分别减小了 4.15 V、1.77 V,调节时间分别缩短了 95 ms 和 27 ms,由以上可知,所提 SAC-CLADRC 策略在光伏功率波动时能够有效提高母线电压的抗扰性和动态性能。

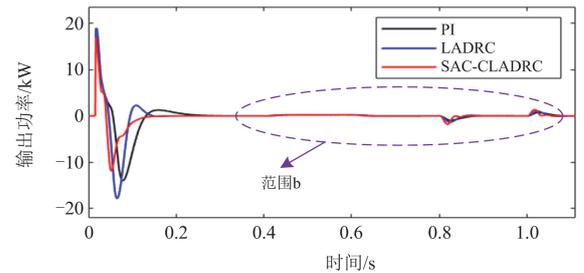
图 11 为光伏输出功率波动时蓄电池和超级电容的出力情况。通过对比分析图 11(a)、图 11(b)可



(a) 蓄电池输出功率



(b) 范围 a 放大图



(c) 超级电容输出功率

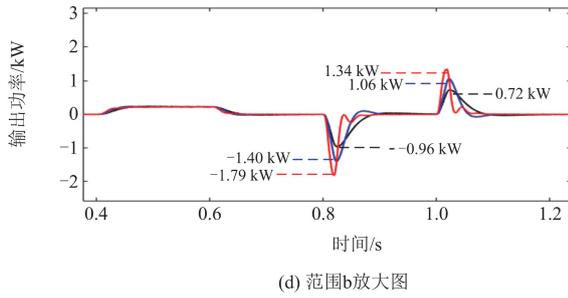


图 11 光伏功率波动时混合储能出力对比

Fig. 11 Comparison of hybrid energy storage output during PV power fluctuations

知, 在母线电压起始响应阶段以及稳态受扰波动阶段, 在本文所提控制策略下蓄电池的输出功率过冲现象明显减小, 以 1.0 s 时为例, 电池在 SAC-CLADRC 策略下的出力峰值为 1.51 kW, 相比在 PI、LADRC 策略下分别减少了 0.15 kW、0.12 kW, 充放电转换更加平稳, 有效减轻了其受到的短时功率冲击。因此, 所提策略对于延长电池寿命、提高经济性具有显著效果。此外, 通过对比分析图 11(c)、图 11(d) 可知, 在母线电压起始响应阶段, 基于 SAC-CLADRC 的协同控制策略, 超级电容在有效抑制过度出力的同时, 显著提升了光伏波动工况下的短时功率支撑能力。基于以上分析可知, 本文所提 SAC-CLADRC 策略在光伏输出功率波动时, 可以较好地协调混合储能各单元出力。

4.3 负载功率波动工况

本节保持光照强度恒为 1000 W/m^2 , 光伏输出功率为 6.03 kW, 通过负载投切改变负荷功率。图 12 为负载功率波动情况, 在 0.4 s 和 0.8 s 时, 负载功率由 5 kW 突增为 8 kW 和 7 kW, 此时光伏输出功率小于负载需求, 由混合储能补偿差额功率; 在 0.6 s 时, 负载功率由 8 kW 功率突减 5 kW, 此时光伏输出功率大于负载需求, 由混合储能吸收多余功率。图 13 为负载功率波动时, 母线电压在 3 种控制策略下的波动情况。

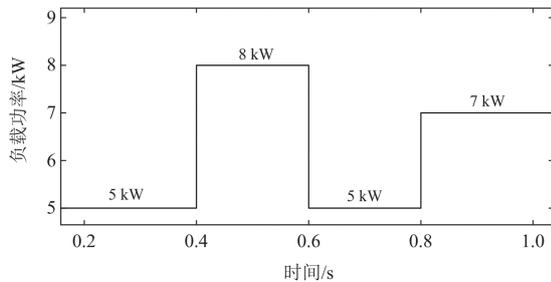
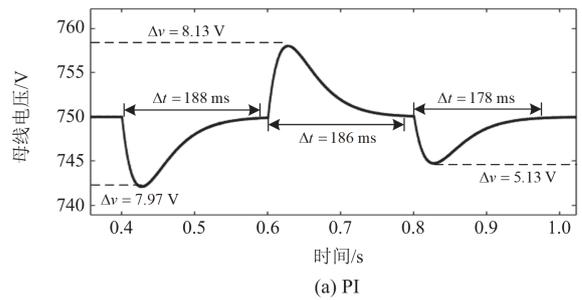
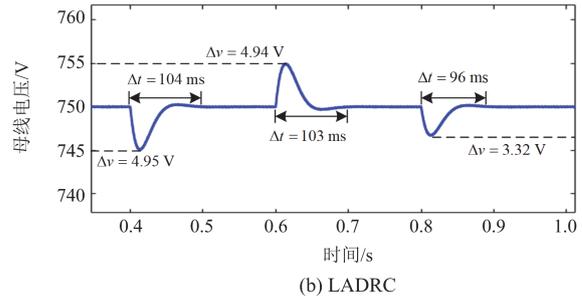


图 12 负载功率波动

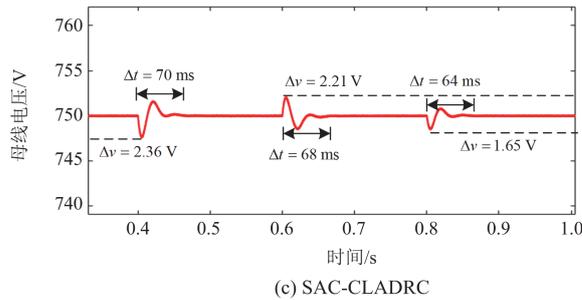
Fig. 12 Power fluctuation of loads



(a) PI



(b) LADRC



(c) SAC-CLADRC

图 13 负载功率波动时母线电压对比

Fig. 13 Comparison of bus voltage during load power fluctuations

通过对比分析图 13 可知, 在负载功率波动工况下, 本文所提策略能够显著减小母线电压的最大波动量, 并缩短调节时间。以 0.4 s 为例, 当负载功率突增为 8 kW 时, 母线电压在 SAC-CLADRC 策略下的最大跌落量为 2.36 V, 调节时间为 70 ms, 而在 PI、LADRC 策略下的最大跌落量分别为 7.97 V、4.95 V, 调节时间分别为 188 ms、104 ms。相比之下, SAC-CLADRC 策略的最大电压跌落量分别减小了 5.61 V、2.59 V, 调节时间分别缩短了 118 ms、34 ms, 由此可见, 本文所提策略在负载功率波动工况下能够有效提升母线电压的抗扰性能和动态响应能力。

图 14 为负载功率波动下混合储能出力情况。通过对比分析图 14(b) 可知, 在母线电压受负载功率波动影响阶段, 采用 SAC-CLADRC 策略后, 蓄电池充放电转换更加平稳, 这极大地减轻了电池因短时

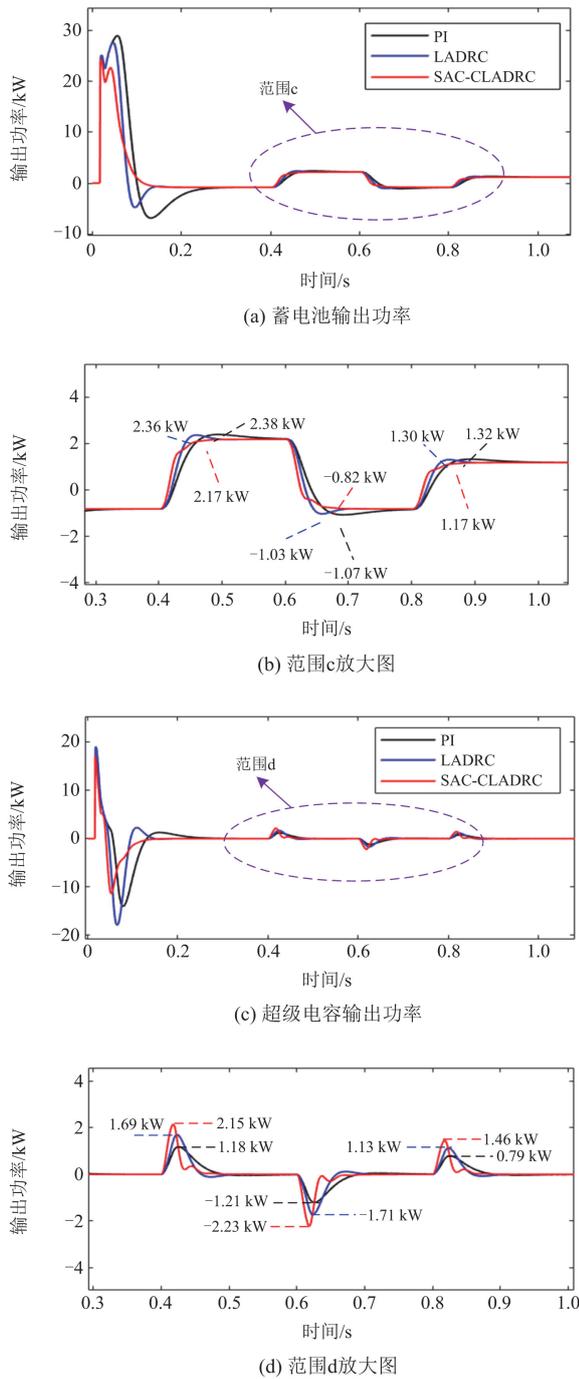


图 14 负载功率波动时混合储能出力对比

Fig. 14 Comparison of hybrid energy storage output during load power fluctuations

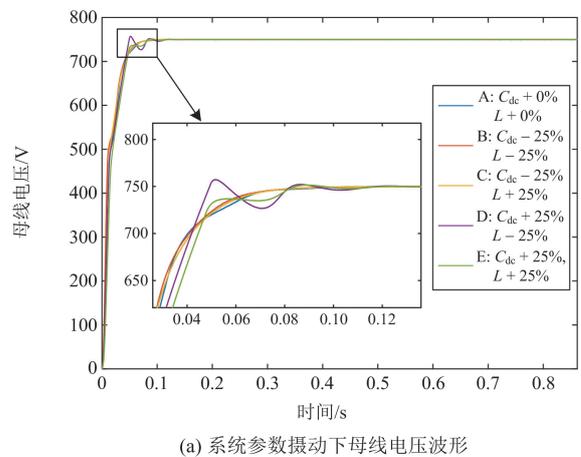
功率波动所受到的冲击，对于延长其使用寿命以及提升整个系统的经济性具有显著优势。同时，分析图 14(d)可知，相比 PI 和 LADRC 策略，SAC-CLADRC 策略能够有效提高超级电容对系统短时供需功率调节的出力，使其承担更大的短时功率冲击，以 0.8 s

为例，超级电容在 SAC-CLADRC 策略下的出力峰值为 1.46 kW，相比 PI 和 LADRC 策略，提高了 0.33 kW 和 0.67 kW。基于以上分析，本文所提策略在负载功率波动时，能够较好地协调混合储能各单元出力。

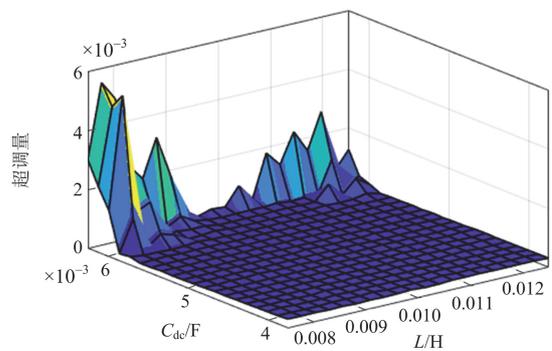
4.4 系统内部参数摄动下性能分析

考虑到混合储能单元 BDC 参数通常会受到出厂设计误差、使用时限以及外部环境等因素的影响，导致 BDC 的电容和电感参数与标称值之间产生一定偏差。为验证本文所提策略的鲁棒性以及系统在内部参数摄动下的动态性能，保持光伏输出功率恒为 6.03 kW、负载功率恒为 5 kW，令超级电容和蓄电池侧电感 $L(L_b, L_{sc})$ 以及母线电容 C_{dc} 参数在标称值 $[-25\%, +25\%]$ 内变化，图 15(a)展示了母线电压在不同参数摄动下的动态响应过程，同时，经过 400 次均匀分布的参数摄动实验，母线电压的动态性能指标与电容和电感参数摄动量之间的数值分布关系如图 15(b)、图 15(c)所示。

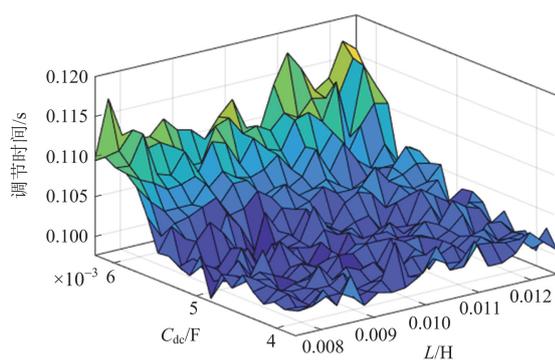
分析图 15(a)可知，在考虑参数摄动量最大化的工况下(A,B,C,D,E)，母线电压在经历短暂的动态过程后均能稳定跟踪至电压给定值。此外，根据图 15(b)、图 15(c)中电压超调量和调节时间随电容、电感变化的数值分布曲面，可以看出在参数摄动极大



(a) 系统参数摄动下母线电压波形



(b) 系统参数摄动下的超调量分布



(c) 系统参数扰动下的调节时间分布

图 15 系统参数扰动下的动态性能分析

Fig. 15 Dynamic performance analysis under system parameter perturbations

范围内, 电压波形的动态超调量可稳定趋近于零, 且在大扰动量工况下, 最大超调量仅为 0.56%。同时, 调节时间随电容 C_{dc} 、电感 L 的变化呈一定的相关特性, 但总体来看, 调节时间在参数扰动下可保持在 0.11 s 左右。因此, 综上所述可知, 在系统内部参数扰动下, 本文所提控制策略表现出较强的抗扰性和鲁棒性, 能够有效保证系统的稳定性和动态性能。

5 结论

为实现以光伏为主的新能源高水平消纳, 针对光储直流微电网混合储能系统因光伏出力波动、负荷功率波动等不确定性因素引起的母线电压波动和系统鲁棒性差的问题, 提出一种融合深度强化学习 SAC 算法的自驱优级联自抗扰控制策略, 通过理论分析与仿真实验验证了本文所提策略的有效性和优越性, 结论如下。

1) 针对传统 LADRC 中 LESO 扰动估计精度有限的问题, 设计了 CLESO, 从而提高了系统对于复杂扰动的观测性能与收敛精度。

2) 基于系统状态构建了混合储能马尔可夫决策模型, 并设计状态奖励与信息熵综合评估的 SAC 智能体融入 CLADRC 参数优化过程中, 通过其在线学习和经验回放能力实现了控制参数的自驱优整定, 从而提高了系统对于不确定工况的抗扰性与鲁棒性。

通过仿真实验可知, 本文所提混合储能控制策略在源荷功率波动以及系统内部参数扰动等内外扰动的抑制方面, 均有较好的抗扰性和鲁棒性, 可以有效提高直流母线电压的稳定性。此外, 该策略还可较好地协调混合储能各单元出力。

参考文献

[1] 王大兴, 宁妍, 汪敬培, 等. 构建新型电力系统背景下

的微电网鲁棒简化建模[J]. 中国电力, 2024, 57(1): 148-157.

WANG Daxing, NING Yan, WANG Jingpei, et al. Robust simplified modeling of microgrid in the context of constructing new power systems[J]. Electric Power, 2024, 57(1): 148-157.

[2] 严小珊, 唐惠玲, 吴杰康, 等. 基于 MPC 的光-储协同调频优化策略[J]. 南方能源建设, 2024, 11(2): 125-138. YAN Xiaoshan, TANG Huiling, WU Jiekang, et al. Optimization strategy for collaborative frequency modulation of PVs-ESs based on MPC[J]. Southern Energy Construction, 2024, 11(2): 125-138.

[3] 张琳, 谢洪途, 赵路路, 等. 混合能源直流微电网能源优化管控策略研究[J]. 电力系统保护与控制, 2024, 52(3): 141-151.

ZHANG Lin, XIE Hongtu, ZHAO Lulu, et al. Energy optimization and control strategy for a hybrid energy DC microgrid[J]. Power System Protection and Control, 2024, 52(3): 141-151.

[4] 杨新法, 苏剑, 吕志鹏, 等. 微电网技术综述[J]. 中国电机工程学报, 2014, 34(1): 57-70.

YANG Xinfu, SU Jian, LÜ Zhipeng, et al. Overview of microgrid technology[J]. Proceedings of the CSEE, 2014, 34(1): 57-70.

[5] 李嘉伟, 巨云涛, 张璐, 等. 基于分布鲁棒模型预测控制的微电网多时间尺度优化调度[J]. 电力工程技术, 2024, 43(4): 45-55.

LI Jiawei, JU Yuntao, ZHANG Lu, et al. Multi-time scale distributed robust optimal scheduling of microgrid based on model predictive control[J]. Electric Power Engineering Technology, 2024, 43(4): 45-55.

[6] 张智刚, 康重庆. 碳中和目标下构建新型电力系统的挑战与展望[J]. 中国电机工程学报, 2022, 42(8): 2806-2819. ZHANG Zhigang, KANG Chongqing. Challenges and prospects for constructing the new-type power system towards a carbon neutrality future[J]. Proceedings of the CSEE, 2022, 42(8): 2806-2819.

[7] 刘畅, 卓建坤, 赵东明, 等. 利用储能系统实现可再生能源微电网灵活安全运行的研究综述[J]. 中国电机工程学报, 2020, 40(1): 1-18, 369.

LIU Chang, ZHUO Jiankun, ZHAO Dongming, et al. A review on the utilization of energy storage system for the flexible and safe operation of renewable energy microgrids[J]. Proceedings of the CSEE, 2020, 40(1): 1-18, 369.

[8] 张宇涵, 杜贵平, 雷雁雄, 等. 直流微网混合储能系统控制策略现状及展望[J]. 电力系统保护与控制, 2021, 49(3): 177-187.

ZHANG Yuhan, DU Guiping, LEI Yanxiong, et al.

- Current status and prospects of control strategy for a DC micro grid hybrid energy storage system[J]. *Power System Protection and Control*, 2021, 49(3): 177-187.
- [9] 郭立东, 雷鸣宇, 杨子龙, 等. 光储微网系统多目标协调控制策略[J]. *电工技术学报*, 2021, 36(19): 4121-4131.
GUO Lidong, LEI Mingyu, YANG Zilong, et al. Multi-objective coordinated control strategy for photovoltaic and energy-storage microgrid system[J]. *Transactions of China Electrotechnical Society*, 2021, 36(19): 4121-4131.
- [10] 张世欣, 皇金锋, 杨艺. 基于平坦理论的直流微电网双向 DC-DC 变换器改进滑模自抗扰控制[J]. *电力系统保护与控制*, 2023, 51(5): 107-116.
ZHANG Shixin, HUANG Jinfeng, YANG Yi. Improved sliding mode and active disturbance rejection control based on flatness theory for a bi-directional DC-DC converter in a DC microgrid[J]. *Power System Protection and Control*, 2023, 51(5): 107-116.
- [11] GORBACHEV S, GUO J, MANI A, et al. MPC-based LFC for interconnected power systems with PVA and ESS under model uncertainty and communication delay[J]. *Protection and Control of Modern Power Systems*, 2023, 8(4): 1-17.
- [12] LI Z, WANG B, XIAN L, et al. Decentralized active disturbance rejection control for hybrid energy storage system in DC microgrid[J]. *IEEE Transactions on Industrial Electronics*, 2024, 71(11): 14232-14243.
- [13] 王雨祺, 赵兴勇, 赖建中, 等. 基于自抗扰控制的光伏微电网储能控制策略[J]. *电气工程学报*, 2024, 19(3): 423-431.
WANG Yuqi, ZHAO Xingyong, LAI Jianzhong, et al. Energy storage control strategy for photovoltaic microgrid based on active disturbance rejection control[J]. *Journal of Electrical Engineering*, 2024, 19(3): 423-431.
- [14] 皇金锋, 杨振宇, 李帅杰. 基于混合储能的光储直流微网改进型滑模自抗扰控制[J]. *电力系统保护与控制*, 2023, 51(23): 151-159.
HUANG Jinfeng, YANG Zhenyu, LI Shuaijie. Improved sliding mode self-disturbance resistance control of an optical storage DC micro-grid based on hybrid energy storage[J]. *Power System Protection and Control*, 2023, 51(23): 151-159.
- [15] SUN C, LIU M, LIU C, et al. An industrial quadrotor UAV control method based on fuzzy adaptive linear active disturbance rejection control[J]. *Electronics*, 2021, 10(4): 376.
- [16] 赵希梅, 陈广国, 金鸿雁. 基于改进灰狼优化算法的 PMSM 滑模自抗扰控制[J]. *电机与控制学报*, 2022, 26(11): 132-140.
ZHAO Ximei, CHEN Guangguo, JIN Hongyan. Sliding mode active disturbance rejection control for PMSM based on improved grey wolf optimization algorithm[J]. *Electric Machines and Control*, 2022, 26(11): 132-140.
- [17] LIU W, ZHAO T, WU Z, et al. Linear active disturbance rejection control for hysteresis compensation based on backpropagation neural networks adaptive control[J]. *Transactions of the Institute of Measurement and Control*, 2021, 43(4): 915-924.
- [18] 马幼捷, 杨清, 周雪松, 等. 基于模糊神经网络的储能侧双向 DC-DC 变换器自抗扰控制策略[J]. *太阳能学报*, 2023, 44(10): 488-495.
MA Youjie, YANG Qing, ZHOU Xuesong, et al. Active disturbance rejection control strategy of bidirectional DC-DC converter on energy storage side based on fuzzy neural network[J]. *Acta Energetica Sinica*, 2023, 44(10): 488-495.
- [19] MANANDHAR U, TUMMURU N R, KOLLIMALLA S K, et al. Validation of faster joint control strategy for battery-and supercapacitor-based energy storage system[J]. *IEEE Transactions on Industrial Electronics*, 2017, 65(4): 3286-3295.
- [20] GAO Z. Scaling and bandwidth-parameterization based controller tuning[C] // *Proceedings of the 2003 American Control Conference*, June 4-6, 2003, Denver, CO, USA: 4989-4996.
- [21] ZHENG Y, TAO J, HARTIKAINEN J, et al. DDPG based LADRC trajectory tracking control for underactuated unmanned ship under environmental disturbances[J]. *Ocean Engineering*, 2023, 271.
- [22] HAARNOJA T, ZHOU A, ABBEEL P, et al. Soft actor-critic: off-policy maximum entropy deep reinforcement learning with a stochastic actor[C] // *International Conference on Machine Learning*, PMLR, 2018: 1861-1870.
- [23] LIANG Z, RUAN J, WANG Z, et al. Soft actor-critic-based EMS design for dual motor battery electric bus[J]. *Energy*, 2024, 288.
- [24] SUTTON R S, BARTO A G. Reinforcement learning: an introduction[J]. *Robotica*, 1999, 17(2): 229-235.

收稿日期: 2024-09-13; 修回日期: 2025-01-13

作者简介:

周雪松(1964—), 男, 博士, 教授, 研究方向为新能源发电、电力系统控制; E-mail: sjteam2023@163.com

张宇轩(2001—), 男, 硕士研究生, 研究方向为电力电子变换器复杂控制与储能技术; E-mail: 1187392255@qq.com

马幼捷(1964—), 女, 通信作者, 博士, 教授, 研究方向为微电网运行与控制。

(编辑 许威)