第 52 卷 第 3 期：1146-1158 高电压技术 Vol.52, No.3: 1146-1158
2026 年 3 月 31 日 High Voltage Engineering March 31, 2026
DOI: 10.13336/j.1003-6520.hve.20250420

基于交叉熵的 AC 算法动态调参微网变换器自抗扰稳压控制

周雪松[1]，马立聪[1]，赵　明[2]，陈韵霏[1]，耿盛意[1]，李　双[1]
（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. 天津河西光启能源科技有限公司，天津 300202）

摘　要：DC-DC 接口变换器对直流微网强不确定性抑制具有重要作用，但其母线电压波动、负载变化和建模误差等不确定性因素，对系统稳压控制带来了严峻挑战。为抑制直流微网输出电压波动，考虑自抗扰不依赖模型信息和抗扰能力强的优点，融合演员-评论家(actor-critic, AC)算法自学习的特征，提出一种基于 AC 算法的 DC-DC 接口变换器强化学习自抗扰控制策略，将自抗扰控制（active disturbance rejection control, ADRC）的控制器参数调节问题表征为马尔科夫决策过程，根据智能体与系统交互数据持续学习微网模型复杂规律，训练生成误差驱动的鲁棒最优参数动态调节曲线。数值仿真和物理实验结果表明，对比传统线性自抗扰控制(linear active disturbance rejection control, LADRC)，所提策略在母线电压波动、负载变化时的输出电压最大偏移量分别减小了 60.8%和 58.6%，对应的恢复时间缩短了 22 ms 和 23 ms，表现出更优越的抗扰性能，同时对电容电感参数变化具有较强的鲁棒性。
关键词：直流微网；DC-DC 变换器；强化学习；AC 算法；自抗扰控制

AC Algorithm Based on Cross Entropy for Dynamic Parameter Tuning of Microgrid Converter with Self Disturbance Rejection and Voltage Stabilization Control

ZHOU Xuesong[1], MA Licong[1], ZHAO Ming[2], CHEN Yunfei[1], GENG Shengyi[1], LI Shuang[1]
(1. Tianjin Key Laboratory of New Energy Power Conversion Transmission and Intelligent Control (Tianjin University of Technology),Tianjin 300384, China; 2.Tianjin Hexi Guangqi Energy Technology Co., Ltd., Tianjin 300202, China)

Abstract：The load-side interface converter plays an important role in suppressing strong uncertainty in DC microgrids, but its input voltage fluctuations, modeling errors, and load changes pose severe challenges to system voltage regulation control. To suppress voltage fluctuations in DC microgrids, the advantages of independent model information and strong anti-interference ability are taken into account, and the self-learning features of actor-critic (AC) algorithm are integrated; thus a DC-DC interface converter reinforcement learning self disturbance control strategy based on AC algorithm is proposed. The parameter tuning problem of active disturbance rejection control (ADRC) is characterized as a Markov decision process, and the complex rules of the microgrid model are continuously learned based on the interaction data between the intelligent agent and the system. A robust optimal parameter dynamic adjustment curve driven by errors is trained and generated. Numerical simulation and physical experiment results show that, compared with traditional linear active disturbance rejection control (LADRC), the proposed strategy reduces the maximum output voltage offset by 60.8% and 58.6%, respectively, during bus voltage fluctuations and load changes, and shortens the corresponding recovery time by 22 ms and 23 ms, demonstrating superior disturbance rejection performance and strong robustness to changes in capacitance and inductance parameters.
Key words：DC microgrid; DC-DC converter; reinforcement learning; actor-critic algorithm; active disturbance rejection control

基金资助项目：国家自然科学基金重大项目(U24B6011)；国家自然科学基金重点项目(U23B20142)。
Project supported by Major Project of National Natural Science Foundation of China (U24B6011), Key Project of National Natural Science Foundation of China (U23B20142).

0　引言

直流微网具有灵活高效、绿色清洁等优点，是新型电力系统的重要组成部分，然而风光出力的随机性和负载不确定性对直流微网的正常工作造成严峻挑战[1-3]。电压稳定性是衡量直流微网正常运行的关键因素[4-5]。作为负载端与直流微网的接口设备，DC-DC 变换器的快速稳定控制是实现能量有序流动和维持电压稳定的关键[6-7]。专家学者们对 DC-DC 变换器的稳压控制策略展开了深入研究，如 PI 控制[8]、滑膜控制[9]、模糊控制[10]、自抗扰控制[11]等。PI 控制通过比例-积分调节实现稳态精度，但其依赖精确模型参数，对非线性扰动适应性较差；滑模控制基于滑模面设计实现强鲁棒性，但存在高频抖振问题，且参数整定依赖经验；模糊控制通过模糊规则库处理不确定性，但规则设计复杂、计算量大，难以适应高频开关场景的实时性需求。相较之下，自抗扰控制抗扰性能出色且不依赖模型信息，因此在该领域具有显著优势。

传统 ADRC[12]具有较强的非线性时变特性且控制参数较多，调参过程和整定效果往往依赖经验和试凑。为此，文献[13]提出线性自抗扰控制器(linear active disturbance rejection control，LADRC)，运用带宽法的思想将极点配置到实轴某一位置，大大简化了控制器参数整定，但其结果自由度不足。文献[14]和文献[15]分别运用 D 分割法和定量反馈理论两种图解方式整定 LADRC 参数，虽然细化了颗粒度但调参过程较为繁琐。文献[16-18]根据稳定性理论计算参数稳定域，在此基础上运用智能优化算法的全局搜索能力寻找最优控制参数。虽然改善了系统的控制效果，但离线参数整定方式仍难以适应直流微网的复杂扰动工况。文献[19]利用 RBF 神经网络方法在线整定 ADRC 参数，虽然改善了传统离线整定方法的局限性，但仍然存在规则设计复杂、计算量大的问题。

近年来，人工智能凭借其自适应能力强的优点，在电力系统领域得到广泛应用[20-21]。目前有学者将强化学习算法应用到整定 ADRC 参数上，文献[22]利用 Q-learning 算法通过训练 Q 表实现参数整定，但其依赖离散动作空间的设计导致控制精度受限。文献[23]提出基于深度 Q 网络(Deep Q-Network)的在线调参策略，虽通过神经网络缓解维度灾难问题，却因需离散化连续参数空间而面临策略震荡与计算效率低下的挑战。演员-评论家(actor-critic, AC)算法通过策略梯度与值函数学习的协同机制，支持连续动作空间下的参数动态优化，既避免了离散化导致的精度损失，又显著提升了策略稳定性与计算效率。为此，本文主要开展了如下工作：

1）首次将 AC 强化学习算法与 LADRC 融合用于 DC-DC 变换器稳压控制，通过数据与模型联合驱动的控制方式有效改善强不确定性直流微网的输出电压质量。

2）结合 actor-critic 网络捕捉时序状态变化特征，挖掘连续动作域下控制参数与动态特性的相互关联，构建场景工况自适应的 LADRC 控制参数动态调节机制。

1　系统结构与 LADRC 策略

1.1　直流微网结构及数学建模

直流微网结构示意图如图 1 所示，接口变换器在微网中扮演着至关重要的角色，它是各单元与直流母线间能量流动的关键中枢，本文研究了 Buck 变换器在直流母线和低压负载侧之间的应用，其拓扑结构如图 2 所示。

图 2 拓扑结构图中，u_g 和 u_o 分别为输入和输出电压；L、C、R_l 分别为电感、电容和负载；S、Q 分别为开关管和二极管。i_c、i_L、i_o 分别为电容电

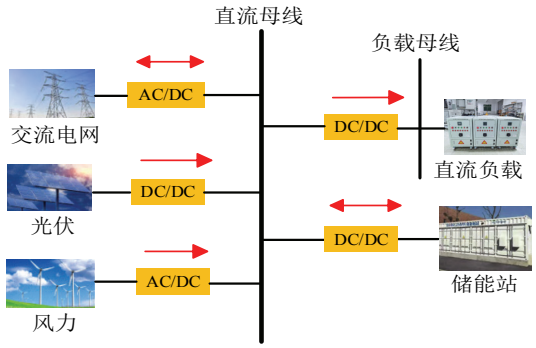


图 1　直流微网结构框图

Fig.1　DC microgrid structure diagram

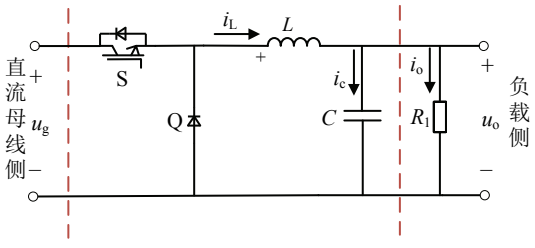


图 2　Buck 电路拓扑结构图

Fig.2　Buck circuit topology diagram

流、电感电流、电阻电流。

在忽略电感和电容等效内阻的前提下，考虑一个周期内开关管的导通与关断，将电路的工作状态在每个周期 T_s 分为[0,DT_s)和[DT_s,T_s]两部分，其中 D 为开关管的占空比，通过对两个工作状态下的电容电流 $i_c(t)$ 和电感电压 $u_L(t)$ 进行建模，并采用状态空间平均法[24]可得：

$$\begin{bmatrix} \dfrac{\mathrm{d}i_L(t)}{\mathrm{d}t} \\ \dfrac{\mathrm{d}u_o(t)}{\mathrm{d}t} \end{bmatrix} = \begin{bmatrix} 0 & -\dfrac{1}{L} \\ \dfrac{1}{C} & -\dfrac{1}{RC} \end{bmatrix} \begin{bmatrix} i_L(t) \\ u_o(t) \end{bmatrix} + \begin{bmatrix} \dfrac{D}{L} \\ 0 \end{bmatrix} u_g(t) \tag{1}$$

由式(1)得：

$$\begin{cases} L\dfrac{\mathrm{d}i_L(t)}{\mathrm{d}t} = Du_g(t) - u_o(t) \\ C\dfrac{\mathrm{d}u_o(t)}{\mathrm{d}t} = i_L(t) - \dfrac{u_o(t)}{R_l} \end{cases} \tag{2}$$

通过拉氏变换和小信号分析法，得到变换器控制量与输出电压之间的传递函数：

$$G_{ud}(s) = \frac{\mathrm{d}u_o(s)}{\mathrm{d}s} = \frac{u_g}{CLs^2 + \dfrac{L}{R_l}s + 1} \tag{3}$$

1.2 二阶线性自抗扰控制器

传统 LADRC 主要组成部分有：线性扩张状态观测器(linear extended state error observer, LESO)、线性误差反馈控制率(linear state error feedback control, LSEFC)。线性扩张状态观测器用于实时估计系统的状态和总扰动，以提高控制系统的稳定性和抗扰动能力。线性误差反馈控制律用于根据系统误差生成控制信号，以实现期望的系统响应和性能目标[25-26]。二阶 LADRC 控制框图如图 3 所示。

含扰动的二阶系统被控对象可以表示为：

$$\begin{aligned} \ddot{y} &= f(y, \dot{y}.\omega, t) + bu = \\ &\quad -a_1\dot{y} - a_2 y + \omega + bu \end{aligned} \tag{4}$$

式中：u 为系统控制量输入；y 为系统输出量；ω 为扰动。a_1、a_2、ω 都是未知的，b 为控制器增益，通常情况下 b 部分已知(已知部分记为 b_0)[27]。则式(4)可写成：

$$\ddot{y} = -a_1\dot{y} - a_2 y + \omega + (b - b_0)u + b_0 u = f + b_0 u \tag{5}$$

其中 $f = -a_1\dot{y} - a_2 y + \omega + (b - b_0)u + b_0 u$ 为包含了内扰与外扰的总扰动。

对式(3)进行拉普拉斯反变换可得：

$$\begin{aligned} \ddot{y}(t) &= -a_1\dot{y}(t) - a_2 y(t) + \\ &\quad b_0 u(t) + (b - b_0)u(t) + \omega(t) \end{aligned} \tag{6}$$

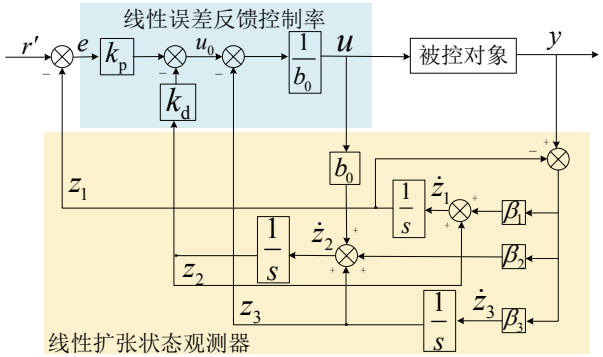


图 3 二阶 LADRC 控制框图

Fig.3 Block diagram of the second-order LADRC control

其中，$a_1 = \dfrac{1}{CR_l}$，$a_2 = \dfrac{1}{CL}$，$b_0 = \dfrac{u_g}{CL}$。

选取状态变量 $x_1 = y$，$x_2 = \dot{y}$，将总扰动 f 视为系统的扩展状态，并且设 $x_3 = f$，则 $\boldsymbol{x} = [y \quad \dot{y} \quad f]^T$ 为包含扰动的扩张状态，将式(5)转化成连续状态空间表达式：

$$\begin{cases} \dot{x} = \boldsymbol{A}x + \boldsymbol{B}u + \boldsymbol{E}\dot{f} \\ y = \boldsymbol{C}x \end{cases} \tag{7}$$

其中：$\boldsymbol{A} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$，$\boldsymbol{B} = \begin{bmatrix} 0 \\ b_0 \\ 0 \end{bmatrix}$，$\boldsymbol{E} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$，$\boldsymbol{C} = [1 \quad 0 \quad 0]$。

则连续 LESO 为：

$$\begin{cases} \dot{z} = \boldsymbol{A}z + \boldsymbol{B}u + \boldsymbol{L}(y - \hat{y}) \\ \hat{y} = \boldsymbol{C}z \end{cases} \tag{8}$$

式中：z 为状态变量的估计量；$\boldsymbol{L} = (\beta_1 \quad \beta_2 \quad \beta_3)^T$ 为观测器误差反馈增益矩阵。选取合适的观测器增益后，LESO 可实时跟踪各状态变量。同时，扰动补偿模块能对总扰动进行实时修正。

设计控制律为：

$$u = \frac{u_0 - z_3}{b_0} \tag{9}$$

设计线性状态误差反馈控制率(LSEFC)为：

$$u_0 = k_p(r' - z_1) - k_d z_2 \tag{10}$$

式中：u_0 为控制器输出值；k_p、k_d 为控制器增益；r' 为图 3 的输入电压。

扰动误差 $z_3 - f$ 收敛于 0 时，将式(9)代入式(5)可得：

$$\ddot{y} = u_0 \tag{11}$$

通过 LESO 的实时估计和控制补偿，原系统可近似为双积分串联结构。采用带宽法[13]的思想将参

数配置于极点位置。

可得观测器参数：

$$\begin{cases}\beta_1=3\omega_o\\ \beta_2=3\omega_o^{\ 2}\\ \beta_3=\omega_o^{\ 3}\end{cases} \tag{12}$$

式中：ω_o 为观测器带宽。

控制器参数：

$$\begin{cases}k_p=\omega_c^2\\ k_d=2\omega_c\end{cases} \tag{13}$$

式中：ω_c 为控制器带宽。

1.3　收敛性分析

跟踪误差可表示为 $e_1(t)=z_1(t)-y(t)$，$e_2(t)=z_2(t)-\dot{y}(t)$，$e_3(t)=z_3(t)-f$。

根据式(8)可得到 $z_1(t)$、$z_2(t)$、$z_3(t)$ 的传递函数：

$$\begin{cases}Z_1(s)=\dfrac{3\omega_o s^2+3\omega_o^2 s+\omega_o^3}{(s+\omega_o)^3}Y(s)+\dfrac{b_0 s}{(s+\omega_o)^3}U(s)\\ Z_2(s)=\dfrac{3\omega_o^2 s^2+\omega_o^3 s}{(s+\omega_o)^3}Y(s)+\dfrac{b_0 s^2+b_0 3\omega_o s}{(s+\omega_o)^3}U(s)\\ Z_3(s)=\dfrac{\omega_o^3 s^2}{(s+\omega_o)^3}Y(s)-\dfrac{\omega_o^3 b_0}{(s+\omega_o)^3}U(s)\end{cases} \tag{14}$$

式(5)结合式(14)通过 Laplace 变换可得到：

$$\begin{cases}E_1(s)=-\dfrac{s^3}{(s+\omega_o)^3}Y(s)+\dfrac{s}{(s+\omega_o)^3}b_0U(s)\\ E_2(s)=-\dfrac{s^4+3\omega_o s^3}{(s+\omega_o)^3}Y(s)+\dfrac{s^2+3\omega_o s}{(s+\omega_o)^3}b_0U(s)\\ E_3(s)=\dfrac{s^2(\omega_o^3-(s+\omega_o)^3)}{(s+\omega_o)^3}Y(s)+\dfrac{s^3+3\omega_o s^2+3\omega_o^2 s}{(s+\omega_o)^3}b_0U(s)\end{cases} \tag{15}$$

稳态误差可表示为：

$$\begin{cases}E_{1s}=\lim\limits_{s\to0}(sE_1(s))=0\\ E_{2s}=\lim\limits_{s\to0}(sE_2(s))=0\\ E_{3s}=\lim\limits_{s\to0}(sE_3(s))=0\end{cases} \tag{16}$$

由式(16)可知：LESO 具有较好的收敛性，通过线性观测器结构，在保证误差快速收敛的同时，可对低频扰动实现零稳态偏差的精确观测。

1.4　抗扰性分析

根据式(9)、式(10)和式(13)可得：

$$u=\frac{1}{b_0}(\omega_c^2(r'-z_1)-2\omega_c z_2-z_3) \tag{17}$$

将(17)代入(14)得：

$$u=((s+\omega_o)^3/b_0((s+\omega_o)^3+2\omega_c s^2+(\omega_c^2+6\omega_c\omega_o)s-\omega_o^3))\cdot(\omega_c^2 r'-\begin{pmatrix}(3\omega_c^2\omega_o+6\omega_c\omega_o^2+\omega_o^3)s^2+\\ (3\omega_c^2\omega_o^2+2\omega_c\omega_o^3)s+\omega_c^2\omega_o^2\end{pmatrix}\cdot y/(s+\omega_o)^3) \tag{18}$$

根据式(18)可以将系统结构表示为图 4。

$$G_a(s)=\frac{(s+\omega_o)^3}{(s+\omega_o)^3+2\omega_c s^2+(\omega_c^2+6\omega_c\omega_o)s-\omega_o^3} \tag{19}$$

$$H(s)=\begin{pmatrix}(3\omega_c^2\omega_o+6\omega_c\omega_o^2+\omega_o^3)s^2+\\ (3\omega_c^2\omega_o^2+2\omega_c\omega_o^3)s+\omega_c^2\omega_o^2\end{pmatrix}\Big/(s+\omega_o)^3 \tag{20}$$

求得系统的闭环传递函数为：

$$G(s)=\frac{\omega_c^2 G_1(s)G_a(s)/b_0}{1+G_1(s)G_a(s)H(s)/b_0} \tag{21}$$

式中：$G_1(s)$ 为 1.1 节所求变换器的传递函数，对式(21)化简可得：

$$G(s)=\frac{(s+\omega_c)^2+3\omega_o(s+2\omega_c+\omega_o)}{(s+\omega_o)^3(s+\omega_c)^2}sf+\frac{\omega_c^2}{(s+\omega_c)^2}r' \tag{22}$$

由式(22)可知，系统的扰动项受 ω_o、ω_c 的影响。选取 $\omega_c=6\ 000$，$\omega_o=10\ 000$、12 000、14 000、16 000、18 000，其频域特性曲线如图 5 所示。同理选择 $\omega_o=18\ 000$，$\omega_c=4\ 000$、6 000、8 000、10 000、12 000，其频域特性曲线如图 6 所示。

由图 5、图 6 可以看出，随着 ω_o 的增大，幅频响应对应的信号频率也向高频方向移动，系统的低频扰动增益减小，系统的抗扰性增强。在一定范围

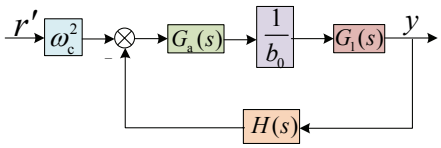


图 4　系统结构图

Fig.4　System architecture diagram

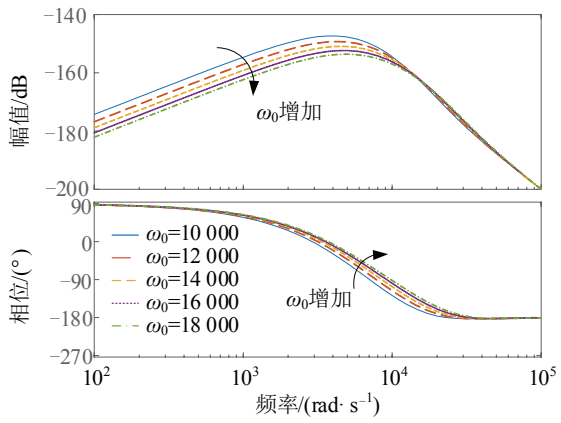


图 5　频域特性曲线(ω_o 变化)

Fig.5　Frequency domain characteristic curve (ω_o variation)

内，增大 ω_c 可加快系统响应速度。但过高的 ω_o (ω_c) 会放大高频噪声，系统可能会变得不稳定，出现超调甚至振荡现象。由于两者之间存在耦合会相互影响，致使对系统性能的调控需同时兼顾二者，增加了参数优化难度。

带宽法给出了 LADRC 参数的解析设计表达式，简化参数整定流程的同时也使颗粒度变得粗糙，接下来本文将融合 AC 强化学习算法实现参数动态调节。强化学习的复杂度随参数维度增加呈指数级上升，若同时调整 ω_c 和 ω_o，动作空间变为二维，可能导致策略网络难以收敛。由于 ω_o 和 ω_c 之间存在倍数关系：$\omega_o = (3 \sim 5)\omega_c$ [13]。可以将双参数优化问题转化为单参数优化，通过动态整定控制器带宽 ω_c，进而实现了对观测器带宽 ω_o 的整定，以满足复杂多变工况的控制需求。

2 AC-LADRC 设计

2.1 交叉熵

交叉熵(cross-entropy)是信息论中的一个重要概念，用于衡量两个概率分布之间的差异程度。对于两个概率分布 $p(x)$和 $q(x)$，其交叉熵的公式为：

$$H(p,q) = -\sum_x p(x)\ln q(x) \tag{23}$$

式中：$p(x)$表示真实分布；$q(x)$表示预测分布。

交叉熵能够量化以概率分布 $q(x)$来表征概率分布$p(x)$时所面临的复杂程度,其本质是对两个概率分布之间差异程度的一种精准刻画。当交叉熵的数值越小时，这表明概率分布 $q(x)$与概率分布 $p(x)$在形态和特征上的相似度越高，二者之间的距离也就越近，即 $q(x)$能够以更高的精准度来逼近 $p(x)$。

引入交叉熵的目的是为了更有效地优化控制策略，提高系统在复杂工况下的性能表现。通过计算策略网络输出的动作概率分布与目标动作概率分布之间的交叉熵，能够衡量当前策略与理想策略的偏离程度，从而为策略的改进提供方向。

2.2 融合强化学习的 LADRC 控制框架

强化学习(reinforcement learing, RL)是通过智能体(Agent)与环境的不断交互，旨在找到一种最优策略，以最大化期望回报值[28]。其示意图见图 7。

图 7 中，a_t 表示 t 时刻智能体的动作；r_t 表示 t 时刻环境反馈的奖励；q_t 表示 t 时刻动作的长期价值期望，衡量该动作未来能获得的累计奖励 3 者构成“选动作-获奖励-评价值”的强化学习交互闭环。

强化学习不需要标记样本，而是通过考虑动作信息(action)和环境信息(environment)，并采用连续马尔可夫决策过程(Markov decision process, MDP)[29]来进行学习和控制。马尔可夫决策过程主要由 5 个元素构成，分别为 s 为状态空间、a 为动作空间、p 为状态转移概率、r 为奖励函数，γ 为折扣因子，用来权衡未来奖励对累计奖励的影响。马尔科夫决策过程如图 8 所示。

本文所提出的 AC-LADRC 控制策略中，整个 buck 变换器 LADRC 系统被视为环境，控制效果用作奖励标准。智能体从环境中获取当前状态，基于策略选择并执行动作，随后转移到新的状态并获取

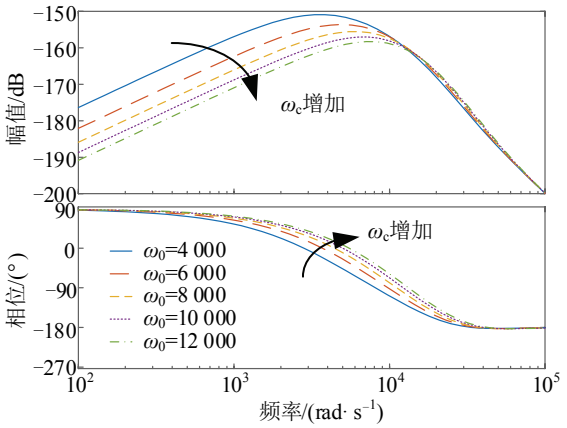


图 6 频域特性曲线(ω_c 变化)

Fig.6 Frequency domain characteristic curve (ω_c variation)

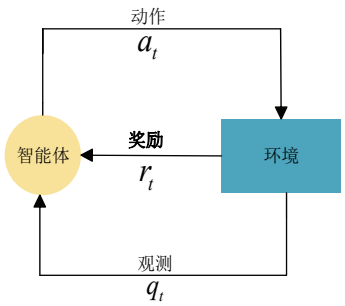


图 7 强化学习示意图

Fig.7 Schematic diagram of reinforcement learning

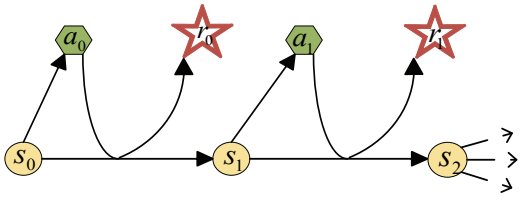


图 8 马尔可夫决策过程

Fig.8 Markov decision process

奖励。通过奖励反馈和状态变化，智能体不断更新策略与价值函数，直到奖励趋于稳定。AC-LADRC 结构框架如图 9 所示。

根据被控对象和 ADRC 模型，确定系统的状态空间 s、动作空间 a 以及奖励函数 r。

状态空间包括 buck 变换器的输出电压 y 与 r' 之间的误差 e_1、观测值 z_1 与 buck 变换器的输出电压 y 之间的误差 e 和变换器的输出电压 y。策略网络输出动作概率分布的熵值 $H(P(a|s))(a|s$ 是在当前状态 s 给定的前提下，选择动作 a 的条件概率)和策略网络输出与目标动作概率分布之间的交叉熵值 $H_{ce}(P,Q)$；动作空间选取控制器带宽 ω_c。

为了减少最终误差，利用变换器的输出电压 y 与参考值 r' 之间的误差 e_1 来设计分段奖励函数，以便训练智能体学习有效的控制策略。奖励函数可表示为：

$$r=\begin{cases}100+5\times H_{ce}(P,Q), & |e_1|<0.5\\ 50+3\times H_{ce}(P,Q), & 0.5\le|e_1|<2\\ 20+1\times H_{ce}(P,Q), & 2\le|e_1|<7\\ -20-1\times H_{ce}(P,Q), & 7\le|e_1|<15\\ -50-3\times H_{ce}(P,Q), & 15\le|e_1|<30\\ -70-5\times H_{ce}(P,Q), & 30\le|e_1|<50\\ -100-7\times H_{ce}(P,Q), & |e_1|\ge 50\end{cases} \tag{24}$$

通过在奖励函数中引入交叉熵，为策略网络的参数更新提供明确的梯度导向，助力算法加速收敛至全局最优策略，同时增强智能体在不同工况下对最优控制策略的泛化能力，提升系统应对复杂运行场景的稳健性。

2.3　AC 算法训练模型

AC 算法是结合了策略梯度和值函数学习的强化学习算法，包括策略网络(Actor)和价值网络(Critic)两部分[30]。Actor 网络对输入样本予以分析和计算，进而输出与环境交互的动作。Critic 网络依据交互的先验知识，计算并输出在当前动作策略下观测环境状态可获得的期望回报价值。Critic 将提升该期望值的梯度信息反馈至 Actor 网络，从而不断提升 Agent 的动作策略性能。环境在交互过程中，把折扣化的奖励信息反馈给 Critic 网络，使其不断提升对期望回报值的评估能力。与其他强化学习方法相比，AC 算法在平衡探索与利用方面表现更佳，提高了学习效率和稳定性，并且适用于处理连续动作空间的问题[31]。AC 算法的结构框架如图 10 所示。

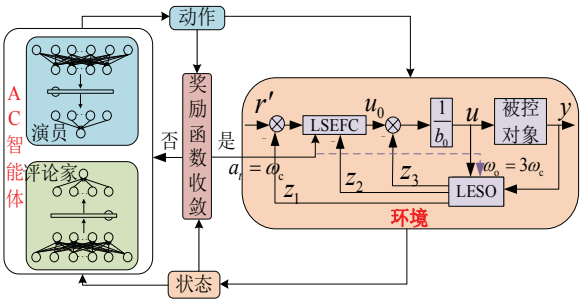


图 9　AC-LADRC 结构框架

Fig.9　AC-LADRC structural framework

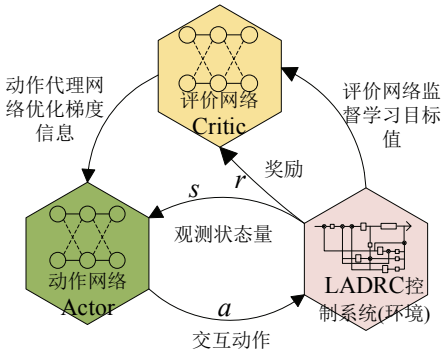


图 10　AC 算法结构框架

Fig.10　AC algorithm architecture framework

对该 Agent 进行训练的最终目标是找到一个基于 Actor 网络的最优动作策略 π^*，该策略可以使得 Critic 评价网络的输出最大化，即优化目标函数为：

$$\pi^*=\arg\max J(\pi)\approx\arg\max v_\pi(s_0) \tag{25}$$

式中：$J(\pi)$ 为期望回报；$v_\pi(s_0)$ 为初始状态的状态价值函数。

在交互学习中，需要不断提升 Actor 网络的性能以寻找最优策略。为增强强化学习过程的稳定性，可以在 Actor 网络参数的梯度优化函数的价值函数中加入基准值信息 $v_\pi(s)$，采用 TD 网络优化方法[32]，由此得到的 Actor 网络的优化梯度为：

$$\begin{aligned}&\theta\leftarrow\theta+\alpha\nabla\ln\pi_\theta(s_t,a_t)r(s_t,a_t)+\\&\alpha\nabla\ln\pi_\theta(s_t,a_t)(\gamma v_\pi(s_{t+1})-v_\pi(s_t))+\\&\alpha\nabla(-\sum_a Q(a|s)\ln P(a|s))\end{aligned} \tag{26}$$

式中：$r(s_t,a_t)$ 为 t 时刻智能体在系统状态 s_t 下，执行动作 a_t 后，环境(即 DC-DC 变换器 LADRC 系统)反馈的即时奖励值；α 为 Actor 网络学习率；θ 为 Actor 深度学习网络参数；s_t、a_t 表示 t 时刻的状态空间与动作空间；$Q(a|s)$ 为目标动作概率分布 $P(a|s)$ 为策略网络输出的动作概率分布。

Critic 网络的优化目标是最小化其输出期望值与实际最大折扣期望回报值之间的误差。由于无法预先确认最优策略 π^*，希望 Critic 网络在已有的先验知识下，能够输出基于当前策略 π 的最大折扣期望回报值，当搜寻到的策略 $\pi \approx \pi^*$ 时，即可得到 Critic 网络的输出期望值为实际最大折扣期望回报值。

Critic 网络的优化目标函数为：

$$\omega_{\mathrm{c}}^{*}=\arg\min(r(s_t,a_t)+\gamma v_{\pi}(s_{t+1})-v_{\pi}(s_t))^2 \quad (27)$$

式中：ω_{c}^{*} 为 Critic 深度学习网络参数。

根据式(26)、(27)可得到训练目标的损失函数：

$$L=(r(s_t,a_t)+\gamma v_{\pi}(s_{t+1})-v_{\pi}(s_t))^2+\lambda(-\sum_{a}Q(a\mid s)\ln P(a\mid s))-\ln\pi_{\theta}(s_t,a_t)\delta(t) \quad (28)$$

式中：λ 为交叉熵项系数；$\delta(t)$ 为时间差分误差；$\pi_{\theta}(s_t,a_t)$ 由参数 θ 确定的策略函数在状态 s_t 下输出动作 a_t 的概率密度。

2.4 AC 算法收敛性证明

由于演员-评论家算法是基于梯度的，只能期望证明收敛到一个局部最优策略。策略梯度算法通过计算性能指标 $J(\theta)$ 相对于策略参数 θ 的梯度 $\nabla J(\theta)$，并沿着梯度上升方向更新参数：$\theta_{t+1}=\theta_t+\alpha\nabla J(\theta_t)$。在实际应用中，这通常会转化为收敛到 $J(\theta)$ 的局部最小值。$\delta(t)$ 为当前状态价值估计与理想价值估计的偏差。$\delta(t)$ 评论家通常会收敛到值函数的期望投影的近似，因此相应的收敛结果必然较弱，只保证 $\nabla J(\theta_k)$ 变得很小。下面做进一步的假设。

对于每个 $\boldsymbol{\theta}\in\mathbf{R}^n$ 定义 $m\times m$ 矩阵 $\boldsymbol{G}(\boldsymbol{\theta})$：

$$\boldsymbol{G}(\boldsymbol{\theta})=\sum_{\boldsymbol{s},\boldsymbol{a}}\boldsymbol{\eta}_{\boldsymbol{\theta}}(\boldsymbol{s},\boldsymbol{a})\boldsymbol{\phi}_{\boldsymbol{\theta}}(\boldsymbol{s},\boldsymbol{a})\boldsymbol{\phi}_{\boldsymbol{\theta}}(\boldsymbol{s},\boldsymbol{a})^{\mathrm{T}} \quad (29)$$

式中：$\boldsymbol{\eta}_{\boldsymbol{\theta}}(\boldsymbol{s},\boldsymbol{a})$ 为状态-动作对的稳态分布概率；$\boldsymbol{\phi}_{\boldsymbol{\theta}}(\boldsymbol{s},\boldsymbol{a})$ 为策略对数概率关于参数 $\boldsymbol{\theta}$ 的梯度。

假设 $\boldsymbol{G}(\boldsymbol{\theta})$ 是一致正定的，即存在一个 $\varepsilon>0$ 对于所有 $\boldsymbol{\theta}\in\mathbf{R}^n$ 和所有非零向量 $\boldsymbol{d}\in\mathbf{R}^m$ 有：

$$\boldsymbol{d}^{\mathrm{T}}\boldsymbol{G}(\boldsymbol{\theta})\boldsymbol{d}\geq\varepsilon\|\boldsymbol{d}\|^2 \quad (30)$$

假设步长序列$\{\alpha_{\mathrm{c}k}\}$，$\{\beta_{\mathrm{c}k}\}$是正的，不递增的，并且满足：

$$\forall k,\alpha_{\mathrm{c}k}>0,\beta_{\mathrm{c}k}>0$$

$$\sum_{k=1}^{\infty}\alpha_{\mathrm{c}k}=\infty,\sum_{k=1}^{\infty}\beta_{\mathrm{c}k}=\infty$$

$$\sum_{k=1}^{\infty}\alpha_{\mathrm{c}k}^2<\infty,\sum_{k=1}^{\infty}\beta_{\mathrm{c}k}^2<\infty$$

$$\frac{\alpha_{\mathrm{c}k}}{\beta_{\mathrm{c}k}}\to 0$$

当 $\alpha_{\mathrm{c}k}/\beta_{\mathrm{c}k}\to 0$ 时，与评论家更新的大小相比，演员更新的大小可以忽略不计。就评论家而言，演员看起来是静止的。因此，通过对 $\delta(t)$ 评论家的分析，认为第 k 次迭代时评论家参数 $\boldsymbol{\omega}_k$ 与理想目标评论家参数 $\boldsymbol{\omega}^*$ 之间的距离随着迭代次数的增加逐渐趋近于 0。

评论家更新公式：

$$\boldsymbol{\omega}_{k+1}=\boldsymbol{\omega}_k+\beta_{\mathrm{c}k}\delta_k\nabla V_1(s;\boldsymbol{\omega}_k) \quad (31)$$

式中：δ_k 为 TD 误差；$\beta_{\mathrm{c}k}$ 为步长。

TD 误差的公式为：

$$\delta_k=r+\gamma v_{\pi}(s_{t+1})-v_{\pi}(s_t) \quad (32)$$

演员更新公式：

$$\boldsymbol{\theta}_{k+1}=\boldsymbol{\theta}_k+\alpha_{\mathrm{c}k}\nabla J(\boldsymbol{\theta}_k) \quad (33)$$

式中：$\boldsymbol{\theta}_k$ 是演员的参数；$\alpha_{\mathrm{c}k}$ 为步长。

如果 $r_{\mathrm{a}}(\boldsymbol{\theta})$ 表示当演员参数固定为 $\boldsymbol{\theta}$ 时，评论家收敛到的值，则对该演员的更新可以重写为：

$$\boldsymbol{\theta}_{k+1}=\boldsymbol{\theta}_k-\alpha_{\mathrm{c}k}(r_{\mathrm{a}}(\boldsymbol{\theta}_k))\nabla V_1(s;\boldsymbol{\omega}_k)+\alpha_{\mathrm{c}k}\boldsymbol{e}_k \quad (34)$$

式中：$\boldsymbol{e}_k$ 是一个渐近可以忽略的误差；$\nabla V_1(s;\omega_k)$ 是当前评论家值函数的梯度；$r_{\mathrm{a}}(\boldsymbol{\theta}_k)$ 是评论家在固定在 $\boldsymbol{\theta}_k$ 时的值函数收敛值。

构造李雅普诺夫(Lyapunov)函数：

$$V(\boldsymbol{\theta},\boldsymbol{\omega})=\left\|\boldsymbol{\theta}-\boldsymbol{\theta}^*\right\|^2+\left\|\boldsymbol{\omega}-\boldsymbol{\omega}^*\right\|^2 \quad (35)$$

$\{\boldsymbol{\theta}_k\}$ 会收敛到一个局部最优点，即 $J(\boldsymbol{\theta})$ 收敛到局部最小值。

3 系统模型训练

3.1 AC-LADRC 控制策略框架

采用强化学习实现 LADRC 控制器带宽的自动调节，利用 AC 算法加速网络的收敛。通过训练控制器带宽 ω_{c}，可得 LSEFC 的参数,并通过 LESO 与 LSEFC 参数的关系，进而到 LESO 的参数。利用 AC 算法整定 LADRC 参数的流程图如图 11 所示。AC-LADRC 的控制逻辑如图 12 所示。

3.2 算法训练过程

在 AC 算法中，学习率和折扣因子是两个重要的超参数。学习率控制参数更新的步长，影响收敛速度和稳定性。折扣因子控制未来回报的重要性，影响策略的长期性和短期性，取值范围为[0,1][33]。通过多次调试，最终确定 AC 参数，如表 1 所示。

在强化学习中，平均奖励(Average Reward)是评

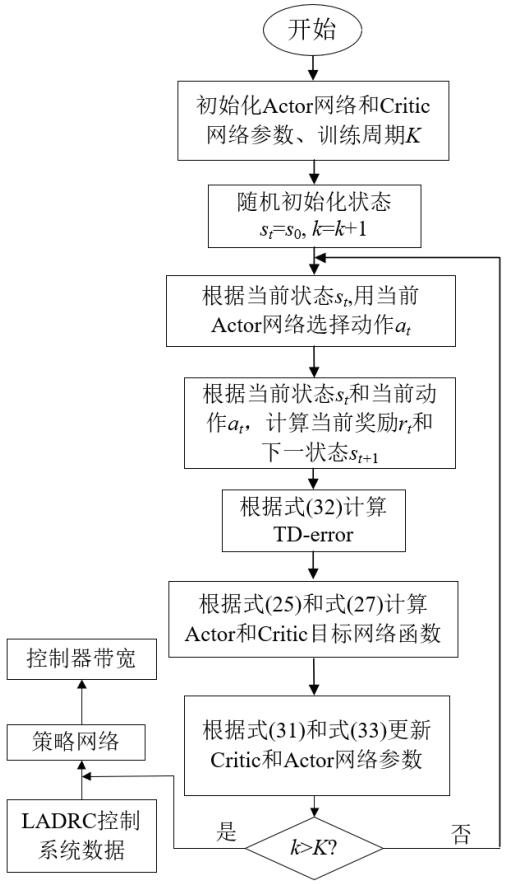


图 11　AC 算法整定 LADRC 参数流程图

Fig.11　Flowchart of AC algorithm tuning LADRC parameters

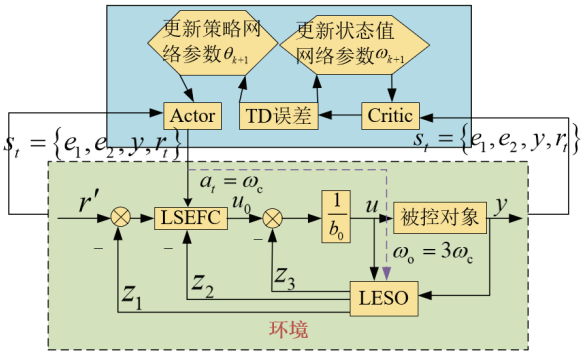


图 12　AC-LADRC 控制策略

Fig.12　AC-LADRC control strategy

估智能体性能的一种重要指标。它通常表示智能体在一段时间或若干回合内的平均奖励，用于判断训练是否收敛。图 11 中训练周期 K 值的选取基于收敛性的分析，本文分别训练了 K 为 400 和 1 200 步长进行分析对比,其 AC 算法的平均奖励曲线分别如图 13、图 14 所示。

由图 13、图 14 分析可得：过低的 K 会使策略未充分学习，进而降低系统对电压波动的抑制能力和稳定性。本文选取 K 为 1 200，为了覆盖收敛后

表 1　AC 算法的参数配置

Table 1　Parameter configuration of AC algorithm

超参数	数值
演员网络学习率	0.000 1
评论家网络学习率	0.001
小批量采样样本大小	128
折扣因子	0.97
采样时间	0.000 1
每个回合持续的最大时间步长	100
训练步长	1 200

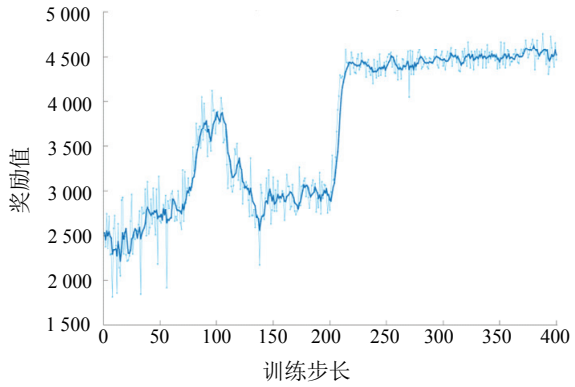


图 13　AC 算法平均奖励曲线(K=400)

Fig.13　AC algorithm average reward curve(K=400)

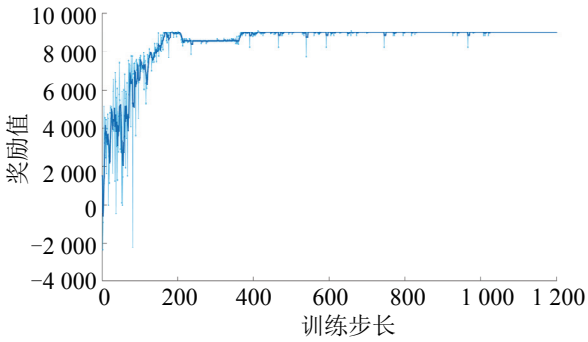


图 14　AC 算法平均奖励曲线(K=1 200)

Fig.14　AC algorithm average reward curve(K=1 200)

的稳定期，确保策略网络充分探索参数空间，避免了因过早停止训练导致的局部最优问题。该智能体大约训练 420 步平均奖励达到收敛的状态，收敛后得平均奖励值为 9 030，训练效果显著。

4　仿真分析

在 Matlab/Simulink 平台搭建仿真模型，对比分析双闭环 PI、LADRC 和 AC-LADRC 3 种控制策略在电压突变、指令阶跃和负载突变工况下的性能，验证所提策略的有效性。系统模型参数如表 2 所示。

表 2 系统模型参数

Table 2 System model parameters

参数	数值
直流母线电压/V	550
开关频率/kHz	100
电阻/Ω	5
电容/μF	300
参考电压/V	450
电感/μH	120

其中，双闭环 PI 控制器电压外环的比例系数 k_{Pv} 和积分系数 k_{Iv} 的参数分别为 0.704 和 527.753；电流内环的比例系数 k_{Pi} 和积分系数 k_{Ii} 的参数分别为 0.003 和 8.756。

经多次仿真调试，平衡响应速度与稳定性，利用带宽法设计 LADRC 的观测器参数 β_1、β_2、β_3 分别为 54 000、9.72×10^8、5.832×10^{12}，其中 ω_o 为 18 000；控制器参数 k_p 和 k_d 分别为 3.6×10^7 和 12 000，其中 ω_c 为 6 000。

4.1 ω_c 动作曲线分析

系统中加入 AC 算法实时整定控制器参数 ω_c 和输出电压仿真结果如图 15 和图 16 所示。

由图 15 和图 16 分析可得：在 0.000 3 s 处输出电压和控制器参数 ω_c 同时达到稳定状态，通过 AC 算法实时整定控制器参数，使输出电压和控制器参数同步快速达到稳定状态，确保系统在动态响应过程中实时调整优化参数以维持输出电压的稳定性。

4.2 电压跟踪性能对比分析

对起始阶段的 450 V 启动信号的跟踪性仿真结果如图 17 所示，在 0.03 s 时刻电压指令从 450 V 阶跃到 460 V 跟踪性仿真如图 18 所示，其超调量 σ 与调节时间如表 3 和表 4 所示。

由图 17、图 18、表 3 和表 4 分析得：在初始阶段，双闭环 PI 表现出显著的超调，且调节时间明显长于传统 LADRC 和 AC-LADRC，而 AC-LADRC 在跟踪输入信号时没有出现超调，且能够实现更快的调节时间。在跟踪稳态下 0.03 s 处进行+10 V 电压调整时，AC-LADRC 在抑制超调和改善跟踪性方面显著优于双闭环 PI 和传统 LADRC。

4.3 母线电压波动抑制性能对比分析

考虑到微网中直流母线电压波动的影响，在 0.03 s 时加入直流母线侧电压的突增和突减(分别为 10%)。其输出电压变化曲线见图 19，不同扰动条件下的最大电压偏移量与恢复时间如表 5 和表 6 所示。

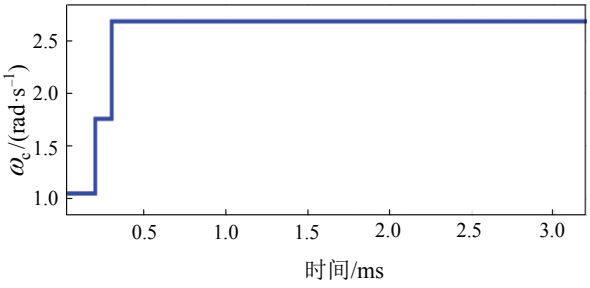


图 15 AC-LADRC 的控制器参数 ω_c 的仿真结果

Fig.15 Simulation results of controller parameter ω_c for AC-LADRC

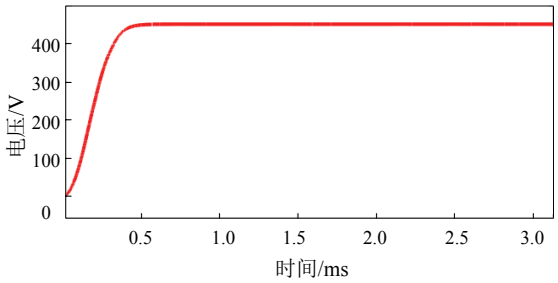


图 16 AC-LADRC 的输出电压仿真结果

Fig.16 Comparison of inhibitory effects of increasing light intensity

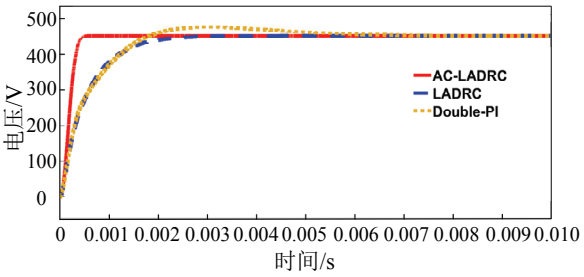


图 17 对起始阶段的 450 V 启动信号的跟踪

Fig.17 Tracking of 450 V startup signal in the initial stage

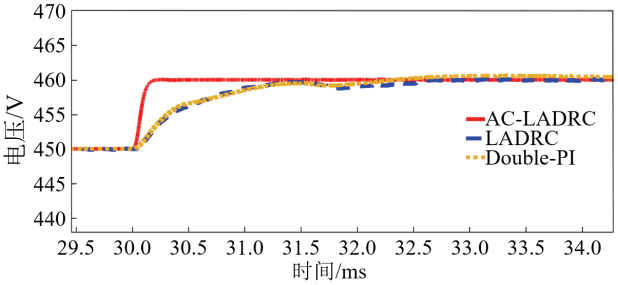


图 18 电压指令阶跃为 460 V 的跟踪

Fig.18 Tracking with voltage command step of 460 V

由图 19、表 5 和表 6 可得；在直流母线电压突增的情况下，与双闭环 PI 与传统 LADRC 相比，AC-LADRC 的最大电压偏移量减少了 90.1%和 93.3%，在直流母线电压突减的情况下，AC-LADRC

表 3　跟踪 450 V 目标函数效果数据

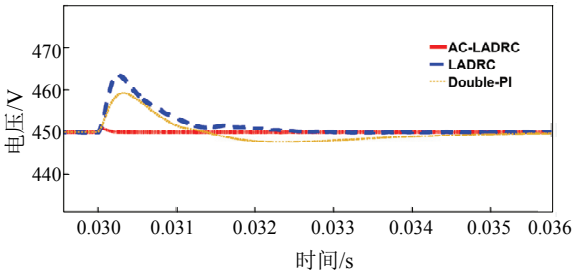
Table 3　Tracking 450 V objective function performance data

性能指标	Double-PI	LADRC	AC-LADRC
σ/%	5.5	0.07	0
调节时间/ms	7.942	4.912	0.687

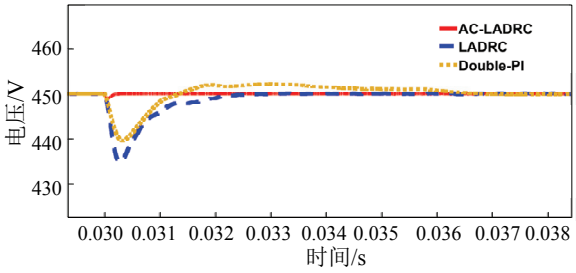
表 4　电压指令阶跃为 460 V 跟踪效果数据

Table 4　Tracking effect data with voltage command step of 460 V

性能指标	Double-PI	LADRC	AC-LADRC
σ/%	0.1	0.02	0
调节时间/ms	5.551	3.091	0.206



(a) 母线电压突增 10%



(b) 母线电压突减 10%

图 19　母线电压波动抑制对比

Fig.19　Comparison of bus voltage fluctuation suppression

表 5　电压突增时的性能指标

Table 5　Performance indicators during sudden voltage increase

性能指标	Double-PI	LADRC	AC-LADRC
最大电压偏移量/V	9.09	13.41	0.9
恢复时间/ms	6.5	3	0.262

的最大电压偏移量较双闭环 PI 和线性 LADRC 减少了 91.3%和 94%，且能更快的恢复稳定。

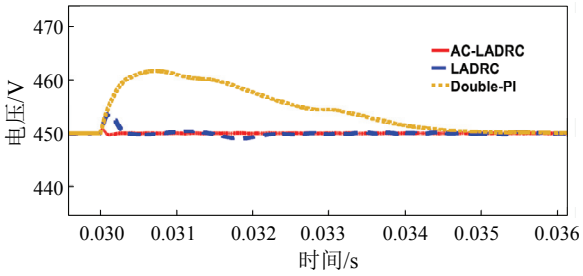
4.4　负载波动抑制性能对比分析

考虑负载突变对系统稳定性的影响，在 0.03 s 处分别施加 20%负载突增和突减的扰动。得到其输出电压曲线如图 20 所示，这两种工况下电压最大偏移量与恢复时间如表 7 和表 8 所示。

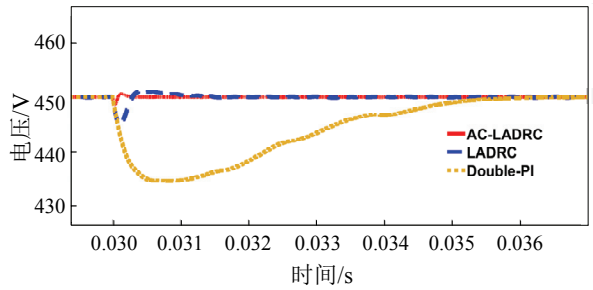
表 6　电压突减时的性能指标

Table 6　Performance indicators during sudden voltage drop

性能指标	Double-PI	LADRC	AC-LADRC
最大电压偏移量/V	10.305	15.12	0.9
恢复时间/ms	6.83	2.8	0.263



(a) 负载突增 20%



(b) 负载突减 20%

图 20　负载波动抑制对比

Fig.20　Comparison of load fluctuation suppression

表 7　负载突增时的性能指标

Table 7　Performance indicators during sudden load increase

性能指标	Double-PI	LADRC	AC-LADRC
最大电压偏移量/V	11.7	3.51	0.9
调节时间/ms	6.1	0.342	0.265

表 8　负载突减时的性能指标

Table 8　Performance indicators during sudden load reduction

性能指标	Double-PI	LADRC	AC-LADRC
最大电压偏移量/V	15.39	4.95	1.395
调节时间/ms	6.48	1.5	0.287

由图 20、表 7 和表 8 可得：在受到负载突增 20%的扰动时，与双闭环 PI 和传统 LADRC 相比，AC-LADRC 的最大电压偏移量减少 92.3%和 74.4%；在受到负载突减 20%的扰动时，AC-LADRC 的电压最大偏移量较双闭环 PI 和线性 LADRC 减少 90.9%和 71.8%，且在两种工况下 AC-LADRC 的恢复时间均优于其它两种控制方式。

4.5　噪声干扰抑制性能对比分析

考虑到噪声干扰对系统稳定性的影响，在 0.03 s

处对输出电压施加噪声干扰。3 种策略的输出电压曲线如图 21 所示，其超调量如表 9 所示。

由图 21 和表 9 可得：在受到噪声干扰时，双闭环 PI 和 LADRC 表现出明显的超调，AC-LADRC 相较于双闭环 PI 和 LADRC，超调量分别减少 88% 和 85.6%，表现出优越的抑制噪声干扰能力。

4.6 鲁棒性性能对比

控制器的初始参数保持不变，使 buck 变换器的电容和电感值在±20%内随机波动，在 3 种控制策略下分别进行 20 次重复实验。选取平方误差积分(integration of square error，ISE)以及 ISE 分布(distribution of ISE，dISE)作为性能指标。ISE 是衡量系统输出与目标值之间误差的平方和的指标，dISE 表示每个 ISE 指标与该组 ISE 指标平均值之间的误差值。3 种控制策略作用下所对应的散点图对比分析如图 22 所示。

性能指标越接近原点且分布越集中，表示系统具有更好的抗扰性和鲁棒性。由图 22 可知，在系统内部参数偏离原始数据的情况下，AC-LADRC 相比于双闭环 PI 和传统 LADRC 的性能指标的值更接近原点，同时分布更加聚集，说明 AC-LADRC 在鲁棒性优于其它两种控制器。

5 物理实验验证

为进一步验证微网负载侧接口变换器在 AC-LADRC 策略下的有效性，搭建实验平台进行物理实验验证，实验平台设备如图 23 所示，实验参数如表 10 所示。

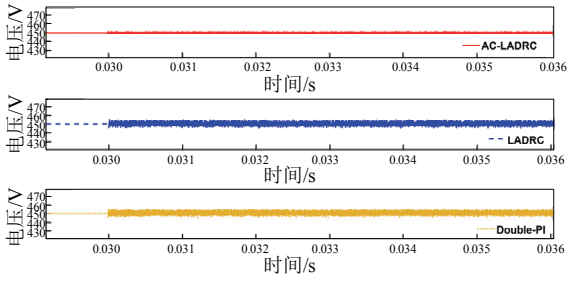


图 21 噪声干扰抑制对比

Fig.21 Comparison of noise interference suppression

表 9 受到噪声干扰时的性能指标

Table 9 Performance indicators under noise interference

性能指标	Double-PI	LADRC	AC-LADRC
σ/%	1.33	1.11	0.16

由数值仿真结果可知 PI 控制效果相对较弱，因此实验环节仅对比 LADRC 和 AC-LADRC 两种策略。通过可编程电源模拟微电网母线电压跳变，使用负载跳变开关模拟负载突变，这两种工况对应的输出波形如图 24 所示，同时基于电压最大变化量 $\triangle u$ 和调节时间 t_s 指标考察系统抗扰性能。

由图 24 分析可得：在 4 种工况下，AC-LADRC 的电压最大变化量和恢复时间均优于传统控制策略，与 LADRC 相比，该策略的电压最大变化量平均减小了 60%，恢复时间平均缩短 22 ms，展现出更强的抗干扰能力。综上所述，该策略能够有效减小外部干扰对直流微电网稳定性的影响，从而确保系统的稳定运行。

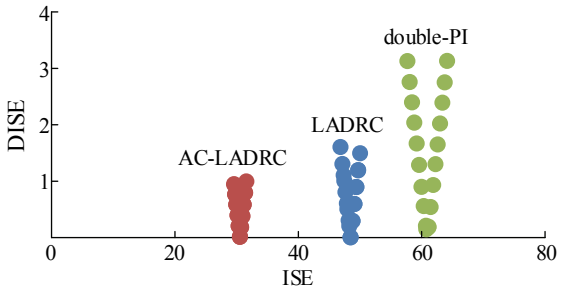


图 22 系统鲁棒性分析散点图

Fig.22 Scatter plot of system robustness analysis

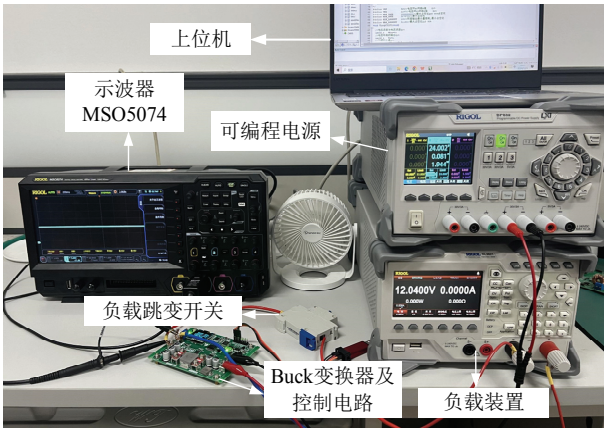


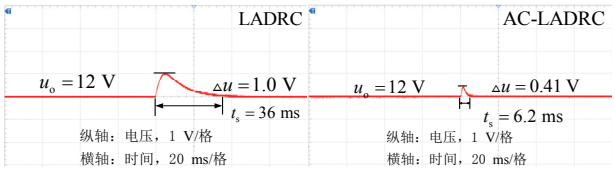
图 23 微网模拟实验平台

Fig.23 Experiment platform for microgrid simulation

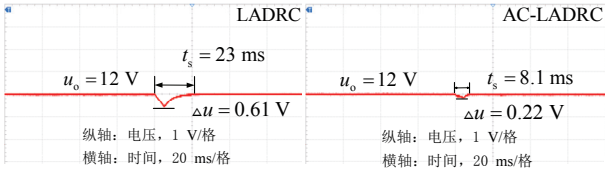
表 10 实验设备参数

Table 10 Experimental equipment parameters

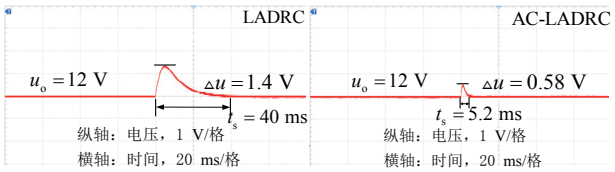
参数	数值
输入电压/V	24
输出电压/V	12
输出功率/W	24
开关频率/kHz	200
主芯片型号	STM32F334
开关管型号	BSC070N10NS



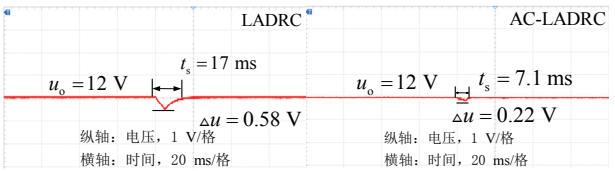
(a) 母线电压突增扰动



(b) 母线电压突减扰动



(c) 负载突增扰动



(d) 负载突减扰动

图 24　不同扰动工况下的输出波形

Fig.24　Output waveforms under different disturbance conditions

6　结论

本文以直流微网输出电压控制场景为切口，尝试探索运用人工智能技术和先进控制理论处理新型电力系统复杂问题，提出一种强化学习动态调参的 DC-DC 变换器自抗扰稳压控制策略，数值仿真和物理实验结果表明：

1）在母线电压波动和负载变化工况下，传统 LADRC 控制，采用本文所提策略时，电压偏移量平均减小了 60%，恢复时间平均缩短了 77%，显著提升了微网系统的输出电压质量。

2）相较于传统控制器人为设置、固定参数的整定和控制方式，AC-LADRC 的带宽参数可根据场景工况变化动态调整，无需人工干预和反复调参，有效提升了系统对复杂工况的适应能力，具有一定的工程应用价值。

下一步将尝试将该策略拓展至多机协同、交直流混合微网等场景。

参考文献　References

[1] 高红均，郭明浩，刘挺坚，等. 新型电力系统电力电量平衡分析研究综述[J]. 高电压技术，2023，49(7)：2683-2696.
GAO Hongjun, GUO Minghao, LIU Tingjian, et al. Review on electric power and energy balance analysis of new-generation power system[J]. High Voltage Engineering, 2023, 49(7): 2683-2696.

[2] 李国庆，刘先超，辛业春，等. 含高比例新能源的电力系统频率稳定研究综述[J]. 高电压技术，2024，50(3)：1165-1181.
LI Guoqing, LIU Xianchao, XIN Yechun, et al. Research on frequency stability of power system with high penetration renewable energy: a review[J]. High Voltage Engineering, 2024, 50(3): 1165-1181.

[3] 侯　恺，刘泽宇，贾宏杰，等. 含高比例可再生能源的电力系统运行可靠性解析评估方法综述[J]. 高电压技术，2023，49(7)：2697-2710.
HOU Kai, LIU Zeyu, JIA Hongjie, et al. Review of analytical methods for operation reliability assessment of power systems with high-penetration renewable energy[J]. High Voltage Engineering, 2023, 49(7): 2697-2710.

[4] 谢小荣，贺静波，毛航银，等. “双高”电力系统稳定性的新问题及分类探讨[J]. 中国电机工程学报，2021，41(2)：461-474.
XIE Xiaorong, HE Jingbo, MAO Hangyin, et al. New issues and classification of power system stability with high shares of renewables and power electronics[J]. Proceedings of the CSEE, 2021, 41(2): 461-474.

[5] 徐贵清，郑开玮，张　崇，等. 高风电渗透率下电网暂态电压的稳定性评估[J]. 电工技术，2023(22)：30-34.
XU Guiqing, ZHENG Kaiwei, ZHANG Chong, et al. Evaluation method of transient voltage stability for grids with high proportions of wind power sources[J]. Electrical Technology, 2023(22): 30-34.

[6] 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.

[7] 张继红，赵　锐，高　雷，等. 直流微网母线电压稳定控制策略[J]. 电网技术，2021，45(12)：4922-4929.
ZHANG Jihong, ZHAO Rui, GAO Lei, et al. DC bus voltage stability control strategy for DC microgrid[J]. Power System Technology, 2021, 45(12): 4922-4929.

[8] 张同国，高文进，张　冀，等. 基于 PI 控制的 DC/DC 变换器峰值电流控制策略[J]. 机电工程技术，2022，51(5)：98-101.
ZHANG Tongguo, GAO Wenjin, ZHANG Ji, et al. Peak current control strategy of DC/DC converter based on PI control[J]. Mechanical & Electrical Engineering Technology, 2022, 51(5): 98-101.

[9] 王丙元，姜　建. 基于滑模预测控制的光伏储能双向 DC/DC 变换器[J]. 电气自动化，2022，44(4)：29-33.
WANG Bingyuan, JIANG Jian. Bidirectional DC/DC converter for photovoltaic energy storage based on sliding mode predictive control[J]. Electrical Automation, 2022, 44(4): 29-33.

[10] 陈　桃. 基于改进模糊 PI 控制的双有源桥 DC-DC 变换器[J]. 计算技术与自动化，2023，42(4)：1-8.
CHEN Tao. Dual active bridge DC-DC converter based on improved fuzzy PI control[J]. Computing Technology and Automation, 2023, 42(4): 1-8

[11] 李　乔，吴　捷. 自抗扰控制及其在 DC-DC 变换器中的应用[J]. 电工技术学报，2005，20(1)：83-88.
LI Qiao, WU Jie. Auto-disturbance rejection control and its application to DC-DC converter[J]. Transactions of China Electrotechnical Society, 2005, 20(1): 83-88.

[12] 徐晓宁，周雪松，马幼捷，等. 基于自抗扰控制技术的微网运行控

制器[J]. 高电压技术，2016，42(10)：3336-3346.
XU Xiaoning, ZHOU Xuesong, MA Youjie, et al. Micro grid operation controller based on ADRC[J]. High Voltage Engineering, 2016, 42(10): 3336-3346

[13] GAO Z Q. Scaling and bandwidth-parameterization based controller tuning[C]//Proceedings of the 2003 American Control Conference. Denver, USA: IEEE, 2003: 4989-4996.

[14] 顾浦阳，张　兴，陈少龙，等. 基于 D 分割法的燃料电池 DC/DC 变换器控制设计[J]. 电力电子技术，2022，56(5)：130-133.
GU Puyang, ZHANG Xing, CHEN Shaolong, et al. Control design of fuel cell DC/DC converter based on D partition method[J]. Power Electronics, 2022, 56(5): 130-133.

[15] 赵天烽，程　赟，华　亮，等. 基于定量反馈理论的时滞系统自抗扰控制参数整定[J]. 控制理论与应用，2021，38(5)：578-586.
ZHAO Tianfeng, CHENG Yun, HUA Liang, et al. Tuning of active disturbance rejection control for time-delay systems via quantitative feedback theory[J]. Control Theory & Applications, 2021, 38(5): 578-586.

[16] 孙雨萌，张旭秀. 改进遗传算法整定自抗扰控制器参数及应用[J]. 自动化与仪表，2020，35(3)：13-17，45.
SUN Yumeng, ZHANG Xuxiu. Parameters optimization and application of auto-disturbance-rejection-controller based on improved genetic algorithm[J]. Automation & Instrumentation, 2020, 35(3): 13-17, 45.

[17] 侯永强，王增宝. 基于粒子群算法的自抗扰控制器参数优化[J]. 计算机与数字工程，2012，40(2)：26-28，49.
HOU Yongqiang, WANG Zengbao. Active disturbance rejection controller parameters optimitation based on particle swarm optimized[J]. Computer and Digital Engineering, 2012, 40(2): 26-28, 49.

[18] 陈路伟. 萤火虫算法在自抗扰控制器参数整定中的方法研究[J]. 舰船电子工程，2019，39(2)：32-34，37.
CHEN Luwei. Research on parameters optimization of ADRC controller based on firefly algorithm[J]. Ship Electronic Engineering, 2019, 39(2): 32-34, 37.

[19] 唐　冲，童仲志，侯远龙. 改进型 LM-RBF 神经网络在自抗扰控制器上的应用[J]. 兵工自动化，2020，39(7)：11-15，28.
TANG Chong, TONG Zhongzhi, HOU Yuanlong. Application of improved LM-RBF neural network in ADRC[J]. Ordnance Industry Automation, 2020, 39(7): 11-15, 28.

[20] 孙永辉，孟雲帆，葛磊蛟，等. 人工智能赋能微电网运行优化的应用及展望[J]. 高电压技术，2023，49(6)：2239-2252.
SUN Yonghui, MENG Yunfan, GE Leijiao, et al. Application and prospect of microgrid operation optimization enabled by artificial intelligence[J]. High Voltage Engineering, 2023, 49(6): 2239-2252.

[21] 葛磊蛟，范延赫，来金钢，等. 面向低碳经济的人工智能赋能微电网优化运行技术[J]. 高电压技术，2023，49(6)：2219-2238.
GE Leijiao, FAN Yanhe, LAI Jin'gang, et al. Artificial intelligence enabled microgrid optimization technology for low carbon economy[J]. High Voltage Engineering, 2023, 49(6): 2219-2238.

[22] 马幼捷，刘熠铭，周雪松，等. 微网储能侧 DC-DC 变换器的强化学习自抗扰控制策略[J]. 太阳能学报，2025，46(3)：63-72.
MA Youjie, LIU Yiming, ZHOU Xuesong, et al. Reinforcement learning active disturbance rejection control strategy for microgrid energy storage side DC-DC converter[J]. Journal of Solar Energy, 2025, 46(3): 63-72

[23] 周雪松，韩　静，马幼捷，等. 基于 DQN 算法的直流微电网负载接口变换器自抗扰控制策略[J]. 电力系统保护与控制，2025，53(1)：95-103.
ZHOU Xuesong, HAN Jing, MA Youjie, et al. Active disturbance rejection control strategy of a DC microgrid load interface converter based on a DQN algorithm[J]. Power System Protection and Control, 2025, 53(1): 95-103.

[24] 张玮麟，刘东立，张耀昌. 非理想 Buck 变换器建模与控制研究[J]. 电器与能效管理技术，2020(7)：25-28.
ZHANG Weilin, LIU Dongli, ZHANG Yaochang. Research of modeling and control of non-ideal Buck converter[J]. Electrical & Energy Management Technology, 2020(7): 25-28.

[25] 韩京清. 自抗扰控制技术[J]. 前沿科学，2007(1)：24-31.
HAN Jingqing. Auto disturbance rejection control technique[J]. Frontier Science, 2007(1): 24-31.

[26] 高志强. 自抗扰控制思想探究[J]. 控制理论与应用，2013，30(12)：1498-1510.
GAO Zhiqiang. On the foundation of active disturbance rejection control[J]. Control Theory & Applications, 2013, 30(12): 1498-1510.

[27] 韩京清. 自抗扰控制器及其应用[J]. 控制与决策，1998，13(1)：19-23.
HAN Jingqing. Auto-disturbances-rejection controller and its applications[J]. Control and Decision, 1998, 13(1): 19-23.

[28] SUTTON R S, BARTO A G. Reinforcement learning: an introduction[M]. 2nd ed. Cambridge, UK: MIT Press, 2018.

[29] JIA S D, SHEN L C, XUE H T. Continuous-time Markov decision process with average reward: using reinforcement learning method[C]//Proceedings of the 34th Chinese Control Conference (CCC). Hangzhou, China: IEEE, 2015: 3097-3100.

[30] 唐　伦，贺小雨，王　晓，等. 基于异步优势演员-评论家学习的服务功能链资源分配算法[J]. 电子与信息学报，2021，43(6)：1733-1741.
TANG Lun, HE Xiaoyu, WANG Xiao, et al. Resource allocation algorithm of service function chain based on asynchronous advantage actor-critic learning[J]. Journal of Electronics & Information Technology, 2021, 43(6): 1733-1741.

[31] DEGRIS T, PILARSKI P M, SUTTON R S. Model-free reinforcement learning with continuous action in practice[C]//Proceedings of the 2012 American Control Conference. Montreal, Canada: IEEE, 2012: 2177-2182.

[32] DENG B, DING H, GONG C. Energy management strategy for two-mode hybrid electric vehicles using double actor-critic with TD error-driven regularization algorithm[J]. International Journal of Automotive Technology, 2025(prepublish): 1-14.

[33] 刘　全，翟建伟，章宗长，等. 深度强化学习综述[J]. 计算机学报，2018，41(1)：1-27.
LIU Quan, ZHAI Jianwei, ZHANG Zongchang, et al. A survey on deep reinforcement learning[J]. Chinese Journal of Computers, 2018, 41(1): 1-27.



周雪松(通信作者)
1964—，男，博士，教授，博导
主要研究方向为新能源发电、电力系统控制
E-mail: sjteam2023@126.com

ZHOU Xuesong
Ph.D., Professor
Corresponding author

收稿日期　2025-03-31　在线出版日期　2025-07-18　　编辑　陈　蔓