

DOI: 10.19783/j.cnki.pspc.240480

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

周雪松¹, 韩静¹, 马幼捷¹, 陶珑¹, 问虎龙², 赵明³

(1. 天津理工大学, 天津 300384; 2. 天津瑞能电气有限公司, 天津 300381;
3. 承德电智尚节能科技有限公司, 河北 承德 067000)

摘要: 在直流微电网中, 为了保证直流母线与负载之间能量流动的稳定性, 解决在能量流动中不确定因素产生的扰动问题。在建立 DC-DC 变换器数学模型的基础上, 设计了一种基于深度强化学习的 DC-DC 变换器自抗扰控制策略。利用线性扩张观测器对总扰动的估计补偿和线性误差反馈控制特性对自抗扰控制器结构进行简化设计, 并结合深度强化学习对其控制器参数进行在线优化。根据不同工况下的负载侧电压波形, 分析了 DC-DC 变换器在该控制策略、线性自抗扰控制与比例积分控制下的稳定性、抗扰性和鲁棒性, 验证了该控制策略的正确性和有效性。最后, 在参数扰动下进行了蒙特卡洛实验, 仿真结果表明该控制策略具有较好的鲁棒性。

关键词: 直流微电网; 深度强化学习; DQN 算法; DC-DC 变换器; 线性自抗扰控制

Active disturbance rejection control strategy of a DC microgrid load interface converter based on a DQN algorithm

ZHOU Xuesong¹, HAN Jing¹, MA Youjie¹, TAO Long¹, WEN Hulong², ZHAO Ming³

(1. Tianjin University of Technology, Tianjin 300384, China; 2. Tianjin Ruineng Electric Co., Ltd., Tianjin 300381, China;
3. Chengde Dianzhishang Energy Saving Technology Co., Ltd., Chengde 067000, China)

Abstract: In the DC microgrid, to ensure the stability of the energy flow between the DC bus and the load, the disturbance problem caused by the uncertain factors in the energy flow is solved. Based on the mathematical model of DC-DC converter, an active disturbance rejection control strategy for a DC-DC converter based on deep reinforcement learning is designed. The active disturbance rejection control structure is simplified using the estimation compensation of the total disturbance and the linear error feedback control characteristics of the linear expansion observer, and the controller parameters are optimized online by deep reinforcement learning. From the load-side voltage waveform in different working conditions, the stability, immunity and robustness of the DC-DC converter using the control strategy, linear active disturbance rejection control and proportional integral control are analyzed, and the correctness and effectiveness of the control strategy are verified. Finally, Monte Carlo experiments are carried out under parameter perturbation, and the simulation results show that the control strategy has good robustness.

This work is supported by the Key Project of National Natural Science Foundation of China (No. U23B20142).

Key words: DC microgrid; deep reinforcement learning; deep-Q-network (DQN) algorithm; DC-DC converters; linear active disturbance rejection control

0 引言

我国“十四五”规划提出的“双碳”目标, 推进了能源结构体制的改革。由于分布式能源发电技术具有清洁、安全、高效、资源丰富以及可省去远

距离输电和调控环节等优点, 被广泛应用于微电网中^[1-2]。直流微电网作为一种微电网类型, 其电力系统主要采用直流供电方式, 并作为未来智能电力配供系统的关键组成部分, 对于促进节能减排和实现可持续能源发展具有重要意义^[3-4]。相比交流微电网, 直流微电网更高效可靠地接纳风、光等分布式可再生能源、储能单元、电动汽车及其他直流用电负荷。直流微电网需要 DC-DC 变换器将直流母线

基金项目: 国家自然科学基金重点项目资助 (U23B20142);
天津市研究生科研创新项目资助 (2022BKYZ036)

电压进行变换, 以实现与各动态元件的连接^[5]。

DC-DC 变换器可以实现直流母线与负载之间能量流动, 其控制策略决定了能量流动的效率与稳定。由于微电网容量较小且抗扰能力差, 运维方式改变时容易引起直流母线电压剧烈波动, 使得电能质量大大降低^[6]。针对以上问题, 文献[7-8]将模糊控制与比例积分控制器(proportion integration controller, PI)相结合, 实现了控制器参数在线自调整, 提高了系统的控制精度和鲁棒性; 文献[9]利用采样降维观测器设计了采样输出反馈控制器, 跟踪速度虽然比PI快, 但速度仍有提升的空间; 文献[10]将自抗扰控制(active disturbance rejection control, ADRC)用于双向DC-DC变换器控制, 提高了系统的抗扰性和鲁棒性; 文献[11]从频域分析入手, 分析了线性自抗扰控制器对扰动的抑制能力, 并提出了控制参数的工程配置方法, 交替调节控制器参数, 以此获得满意的动态性能。文献[12]将自抗扰控制与模糊神经网络相结合实现了控制器参数的在线优化, 显著地提高了系统的稳定性和抗扰性。根据以上文献分析可知, 将智能算法和控制理论与自抗扰相结合对系统参数进行整定是一种行之有效的控制策略, 有利于提高系统性能。自抗扰控制相比于传统的PI控制, 将串联积分作为系统的规范模型, 并将规范模型以外的信息作为系统总扰动, 通过扩张状态观测器(extended state observer, ESO)估计补偿系统总扰动。因此, 自抗扰控制有着较强的鲁棒性和抗扰性, 并且不依赖系统的精确模型。

针对以上问题, 本文以传统自抗扰控制策略为基础, 提出一种基于深度Q网络的线性自抗扰控制(deep-Q-network-LADRC, DQN-LADRC)策略。由于自抗扰控制策略的控制过程可以归为具有马尔可夫性的一类决策问题, 因此可以引入深度强化学习解决该类决策问题, 寻找系统运行下的优化参数。该策略以线性扩张状态观测器为核心, 利用带宽法配置自抗扰控制器中的观测器带宽和控制器带宽, 进而简化自抗扰控制系统整体结构。在此基础上, 利用DQN算法对自抗扰控制器的观测器带宽参数进行整定, 从而实现参数优化。通过仿真分析验证, 该控制策略能够有效解决直流微电网不同工况下, 由于不确定性干扰导致的控制效果恶化问题。

1 系统结构

在直流微电网中, 系统内包含光伏发电单元、风力发电单元等波动性分布式电源、微型燃气轮机和燃料电池等可控型分布式电源、电池等储能装置以及本地交直流负荷、负载、电力电子变换器。直

流微电网系统结构如图1所示。

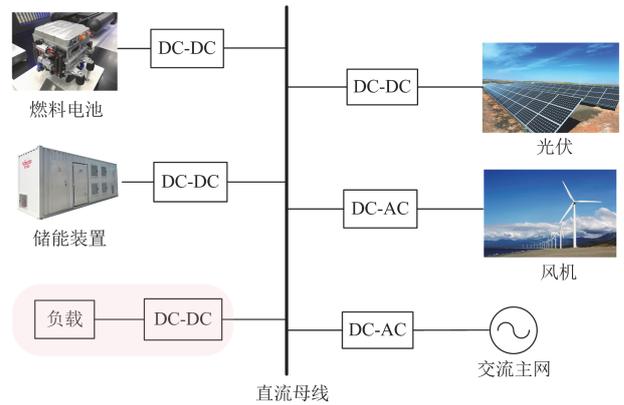


图1 直流微电网系统结构

Fig. 1 DC microgrid system structure

变换器是连接直流母线与各类电源和负载的核心“通道”, DC-DC变换器可以对直流母线电压进行降压处理, 从而对低压负载进行供能, 实现直流母线与负载之间能量流动^[13]。

2 DC-DC 电路建模

本文所研究的直流微电网 Buck 变换器拓扑结构如图2所示。考虑到DC-DC变换器的开关频率远大于微电网状态的变化速率, 所以在短时间内认为直流母线侧功率恒定, 可以等效为恒功率负载。

以电感电流 i_L 和电容电压 u_C 为状态变量, 输出电压为 U_0 , 输入电压为 U_g , 开关管和二极管分别为S和Q。

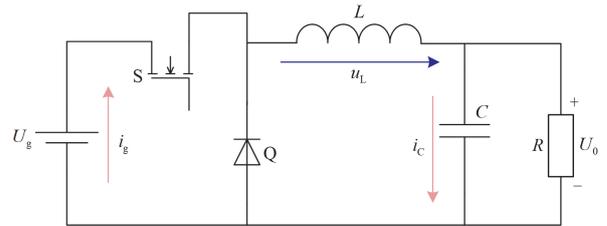


图2 Buck变换器拓扑结构

Fig. 2 Buck converter topology

电感元件 L 和电容元件 C 的电压电流关系为

$$\begin{cases} u_L(t) = L \frac{di_L(t)}{dt} \\ i_C(t) = C \frac{du_C(t)}{dt} \end{cases} \quad (1)$$

针对一个周期 T_s 内开关管的导通与关断, 开关管被分为两个工作状态, 开关管在 $[t, t+dT_s]$ 内导通, 在 $[t+dT_s, t+T_s]$ 内关断, 其中 d 为占空比。列出整个开关周期内Buck电路的微分方程:

$$\begin{cases} v_L(t) = L \frac{di_L(t)}{dt} = \begin{cases} v_g(t) - v(t), [t, t + dT_s] \\ -v(t), (t + dT_s, t + T_s] \end{cases} \\ i_C(t) = C \frac{du_C(t)}{dt} = i_L(t) - \frac{v(t)}{R}, [t, t + T_s] \\ i_g = \begin{cases} i_L(t), [t, t + dT_s] \\ 0, (t + dT_s, t + T_s] \end{cases} \end{cases} \quad (2)$$

在小信号扰动条件下, 输入电压 $v_g(t)$ 、输出电压 $v(t)$ 、电感电流 $i_L(t)$ 在一个开关周期中变化很小, 可以近似于开关周期内平均值 $\langle v_g(t) \rangle_{T_s}$ 、 $\langle v(t) \rangle_{T_s}$ 、 $\langle i_L(t) \rangle_{T_s}$, 即

$$\begin{cases} v_g(t) \approx \langle v_g(t) \rangle_{T_s} \\ v(t) \approx \langle v(t) \rangle_{T_s} \\ i_L(t) \approx \langle i_L(t) \rangle_{T_s} \end{cases} \quad (3)$$

根据式(2)、式(3)可得, 电感电压、电容电流、输入电流的开关周期平均值为

$$\begin{cases} \langle v_L(t) \rangle_{T_s} = L \frac{d\langle i_L(t) \rangle_{T_s}}{dt} = \langle v_g(t) \rangle_{T_s} d(t) - \langle v(t) \rangle_{T_s} \\ \langle i_C(t) \rangle_{T_s} = C \frac{d\langle v(t) \rangle_{T_s}}{dt} = \langle i_L(t) \rangle_{T_s} - \frac{\langle v(t) \rangle_{T_s}}{R} \\ \langle i_g(t) \rangle_{T_s} = d(t) \langle i_L(t) \rangle_{T_s} \end{cases} \quad (4)$$

状态空间平均方程和输出方程中含有控制变量 $d(t)$ 和状态变量 $v_g(t)$ 、 $i_L(t)$ 的乘积项, 使系统具有非线性特性, 需要引入小信号扰动, 其中 D 为稳态下的占空比, $\hat{d}(t)$ 为占空比的小信号扰动。

$$d(t) = D + \hat{d}(t) \quad (5)$$

电感电流、输出电压和输入电压的小信号扰动分别为 $\hat{i}_L(t)$ 、 $\hat{v}(t)$ 和 $\hat{v}_g(t)$ 。联立式(4)和式(5)可得:

$$\begin{cases} \langle i_L(t) \rangle_{T_s} = I_L + \hat{i}_L(t) \\ \langle v(t) \rangle_{T_s} = U + \hat{v}(t) \end{cases} \quad (6)$$

进而得到 Buck 变换器的近似小信号模型:

$$\begin{cases} L \frac{d\hat{i}_L(t)}{dt} = D\hat{v}_g(t) + U_g \hat{d}(t) - \hat{v}(t) \\ C \frac{d\hat{v}(t)}{dt} = \hat{i}_L(t) - \frac{\hat{v}(t)}{R} \\ \hat{i}_g(t) = I_L \hat{d}(t) + D\hat{i}_L(t) \end{cases} \quad (7)$$

最后, 进行拉式变换得到 Buck 变换器输入电压到占空比的传递函数 $G_{vd}(s)$, 如式(8)所示。

$$G_{vd}(s) = \frac{U_g}{LCs^2 + \frac{L}{R}s + 1} \quad (8)$$

3 DQN-LADRC 的设计

3.1 线性自抗扰控制器

1) 线性扩张观测器

在考虑外扰与内扰等不确定因素下, 二阶被控对象^[14]可表示为

$$\ddot{y} = f(y, \dot{y}, \omega, t) + bu \quad (9)$$

式中: y 为系统输出量; u 为系统输入控制量; ω 为内外扰动等不确定量; b 为控制量增益; $f(\cdot)$ 为被控系统模型变量之间的耦合关系。在实际系统中, 系统的控制量增益 b 无法确定, 其中已知部分记为 b_0 , 则式(9)可写为

$$\ddot{y} = f(y, \dot{y}, \omega, t) + (b - b_0)u + b_0u \quad (10)$$

将未知扰动 $f(y, \dot{y}, \omega, t) + (b - b_0)u$ 当作总扰动 $h(t)$, 则式(10)可写为

$$\ddot{y} = h(t) + b_0u \quad (11)$$

假设总扰动可导, 选取状态变量: $x_1 = y$ 、

$x_2 = \dot{y}$ 、 $x_3 = h$, 则 $\mathbf{x} = [y \ \dot{y} \ h]^T$ 为包括了扩张扰动的三维状态变量, 将式(11)转化为状态空间可以表示为

$$\begin{cases} \dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}u + \mathbf{E}\dot{h} \\ \mathbf{y} = \mathbf{C}\mathbf{x} \end{cases} \quad (12)$$

式中:

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}; \quad \mathbf{B} = \begin{bmatrix} 0 \\ b_0 \\ 0 \end{bmatrix}; \quad \mathbf{E} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}; \quad \mathbf{C} = [1 \ 0 \ 0].$$

其对应的线性扩张状态观测器为

$$\begin{cases} \dot{\mathbf{z}} = \mathbf{A}\mathbf{z} + \mathbf{B}u + \mathbf{L}(y - \hat{y}) \\ \hat{\mathbf{y}} = \mathbf{C}\mathbf{x} \end{cases} \quad (13)$$

式中: \mathbf{z} 为观测器的状态变量; \mathbf{L} 为观测器的增益矩阵。

利用带宽法^[15]进行参数化, 把特征方程的极点配置同一个位置 $-\omega_0$ 上, ω_0 为观测器带宽, 则观测器的增益矩阵为

$$\mathbf{L} = [3\omega_0 \ 3\omega_0^2 \ \omega_0^3]^T \quad (14)$$

2) 线性状态误差反馈控制

在 ESO 估计出扰动并补偿后, 线性状态误差反馈(linear states error feedback, LSEF)控制本质上是一个比例微分控制器, LSEF 数学模型即为

$$\begin{cases} u_0 = k_p(v - z_1) - k_d z_2 \\ u = (u_0 - z_3)/b_0 \end{cases} \quad (15)$$

式中: v 为给定输入电压值; $z_1 - z_3$ 为 ESO 观测变量; k_p 、 k_d 为控制器增益系数; u_0 为控制器输出;

u 为系统最终控制量。

控制器增益矩阵 \mathbf{K} 为

$$\mathbf{K} = [k_p \quad k_d]^T \quad (16)$$

联立式(15)、式(16)并进行频域变换, 可得到一个没有零点的二阶系统, 其闭环传递函数 $G(s)$ 为

$$G(s) = \frac{k_p}{s^2 + k_d s + k_p} \quad (17)$$

通过参数化, 将特征方程的极点配置在同一个位置 $-\omega_c$ 处, ω_c 为控制器带宽。则控制器增益矩阵 \mathbf{K} 为

$$\mathbf{K} = [\omega_c^2 \quad 2\omega_c]^T \quad (18)$$

3) 控制量增益

在被控对象已知的情况下, 控制量增益可根据被控对象与系统模型求出, 如式(19)所示。

$$b_0 = \frac{U_g}{LC} \quad (19)$$

线性自抗扰控制器整体结构如图 3 所示。

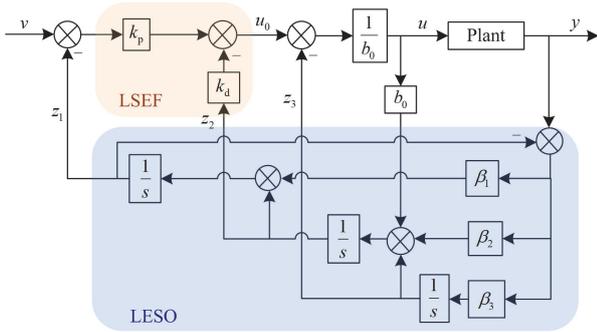


图 3 线性自抗扰控制器

Fig. 3 Linear active disturbance rejection controller

3.2 深度强化学习原理

强化学习是用于解决在智能体与环境交互过程中通过学习策略以达到回报最大化或者实现特定目标的问题。强化学习算法根据已采取的动作、环境状态观测以及获得的奖励值来优化策略^[16-17]。

深度强化学习通过深度神经网络来表示和优化智能体的行为策略, 通过试错和延时回报机制, 不断优化控制策略, 使得智能体可以在复杂环境中进行学习和决策, 并生成优化动作以适应复杂环境。深度强化学习原理如图 4 所示。

传统的强化学习方法通常适用于具有较小的动作和状态空间问题, 而深度学习可以从大规模数据中学习复杂的模式和结构。通过在强化学习中嵌入深度学习技术, 可有效处理更大、更复杂的状态和动作空间, 从而提高系统性能。深度强化学习结合了强化学习的优化目标和深度学习的表征学习能

力, 为处理复杂的环境提供了更有效的方法^[18]。

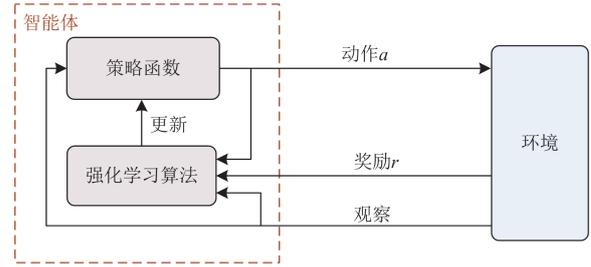


图 4 强化学习原理图

Fig. 4 Reinforcement learning schematic

3.3 马尔可夫决策过程模型

由于整个强化学习是随着动作、回报和状态而更新并得到新的信息, 因此可用马尔可夫决策过程 (Markov decision process, MDP) 对该过程进行表示^[19]。

1) 马尔可夫过程

马尔可夫过程是一个具有马尔可夫性质的离散随机过程。马尔可夫性质是指下一个状态只取决于当前状态, 与过去状态无关。用有限状态集合 S 和状态转移矩阵 \mathbf{P} 可以将马尔可夫过程表示为 $\langle S, \mathbf{P} \rangle$ 。

$$\mathbf{P}(S_{t+1} | S_t) = \mathbf{P}(S_{t+1} | S_0, S_1, \dots, S_t) \quad (20)$$

2) 马尔可夫奖励过程

为了能够进行对智能体的奖励反馈, 马尔可夫过程增加了一个奖励机制。该机制将马尔可夫过程扩展成了 $\langle S, \mathbf{P}, R, \gamma \rangle$, 其中 R 是奖励函数, γ 是奖励折扣因子。

奖励与状态的关系为

$$R_t = R(S_t) \quad (21)$$

价值函数 $V(s)$ 是智能体在状态 s 的期望回报, 可以表示为

$$V^\pi(s) = E[R(\tau) | S_0 = s] \quad (22)$$

3) 马尔可夫决策过程

给定一个动作, 就有动作价值函数在给定基于状态和动作下的期望回报, 可以表示为

$$Q^\pi(s, a) = E \left[\sum_{t=0}^{\infty} \gamma^t R(S_t, A_t) | S_0 = s, A_0 = a \right] \quad (23)$$

3.4 DQN 算法

DQN 算法结合了神经网络的强大表征能力和 Q 学习的基本思想, 将深度学习的函数近似能力和表征能力融入到强化学习中。为了提高算法的稳定性和效率, 使其能够更好地处理复杂的环境和大规模数据, 在 Q 学习的基础上, 引入了目标 Q 值网络, 并采用了经验回放机制, 为强化学习在复杂环境中的应用提供了一种有效的方法^[20-23]。

DQN 算法是基于价值函数的学习, 对于基于价值函数的智能体, 函数的输入为状态和该状态下的可能动作, 并输出采取该动作的价值。智能体在该过程中将奖励作为新信息, 并根据贝尔曼方程更新前一状态和刚采取的动作对应的价值, 贝尔曼方程如式(24)所示。

$$Q_{\text{new}}(s, a) = Q(s, a) + \alpha[(R(s, a) + \gamma \max_{a'} Q'(s', a') - Q(s, a)] \quad (24)$$

1) 观察值

智能体通过与环境交互学习策略, 为学习适应该环境下的优化策略, 该过程中需要不断获取环境信息, 以更好地适应复杂环境。状态是对智能体当前所处环境的完整描述, 对于状态来说环境的所有信息都是可知的。而观察值则是一个状态的部分描述, 智能体通过对观察值的观测, 获取该环境信息学习策略, 优化动作^[16,22]。

本文的观察值为 3 个, 分别为输出电压 y 与给定输入电压 v 之间的误差、输出电压观测值 z_1 与 v 之间的误差以及 z_2 与输出微分值 \dot{y} 之间的误差。

2) 动作空间

DQN 智能体通过观察值与环境进行交互学习, 输出其动作, 根据奖励函数累积奖励值, 优化其动作输出, 为使训练快速达到收敛, 将动作空间根据系统环境进行范围确定。

根据式(10)一式(16)可得:

$$\begin{cases} z_1 = \frac{3\omega_0^2 s^2 + 3\omega_0^2 s + \omega_0^3}{(s + \omega_0)^3} y + \frac{b_0 s}{(s + \omega_0)^3} u \\ z_2 = \frac{(3\omega_0 s + \omega_0^3)s}{(s + \omega_0)^3} y + \frac{b_0(s + 3\omega_0)s}{(s + \omega_0)^3} u \\ z_3 = \frac{\omega_0^3 s^2}{(s + \omega_0)^3} y - \frac{b_0 \omega_0^3}{(s + \omega_0)^3} u \end{cases} \quad (25)$$

根据式(11)、式(12)和式(15)可得:

$$u = \frac{1}{b_0}(\omega_c^2(v - z_1) - 2\omega_c z_2 - z_3) \quad (26)$$

将式(25)代入式(26)可得:

$$u = \frac{1}{b_0} \cdot \frac{(s + \omega_0)^3}{(s + \omega_0)^3 + 2\omega_c s^2 + (\omega_c^2 + 6\omega_c \omega_0)s - \omega_0^3} = \omega_c^2 v - \frac{(3\omega_c^2 \omega_0 + 6\omega_c \omega_0^2 + \omega_0^3)s^2 y}{(s + \omega_0)^3} + \frac{(3\omega_c^2 \omega_0^2 + 2\omega_c \omega_0^3)s y}{(s + \omega_0)^3} + \frac{\omega_c^2 \omega_0^2 y}{(s + \omega_0)^3} \quad (27)$$

根据式(27)可简化结构, 得到

$$P(s) = \frac{(s + \omega_0)^3}{(s + \omega_0)^3 + 2\omega_c s^2 + (\omega_c^2 + 6\omega_c \omega_0)s - \omega_0^3}$$

$$H(s) = \frac{(3\omega_c^2 \omega_0 + 6\omega_c \omega_0^2 + \omega_0^3)s^2 + (3\omega_c^2 \omega_0^2 + 2\omega_c \omega_0^3)s + \omega_c^2 \omega_0^2}{(s + \omega_0)^3}$$

系统简化结构如图 5 所示。

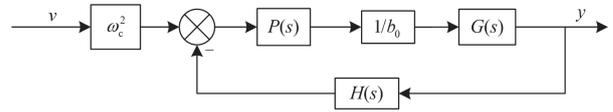


图 5 系统简化结构

Fig. 5 Simplified system structure

求得系统的闭环传递函数 $G_c(s)$ 为

$$G_c(s) = \frac{\omega_c^2 P(s)G(s)/b_0}{1 + P(s)G(s)H(s)/b_0} \quad (28)$$

式中, $G(s)$ 为被控对象。根据劳斯-赫尔维兹稳定性判据以及文献[21]可得, 动作输出范围为 $10^2 \sim 10^7$ 。

3) 奖励函数

DQN 算法旨在学习状态到动作的映射, 以实现累积奖励最大化, 奖励函数的选取与算法的收敛情况息息相关。在训练过程中, 智能体根据策略在当前所处状态选择一个动作, 环境对这些动作做出相应的响应, 转移到新状态中的同时产生一个奖励信号。

奖励的折扣累加和为回报, 是智能体在动作选择过程中想要最大化的目标。奖励是智能体目标的一种形式化、数值化的表征, 其本质是向智能体指明任务目标^[24-25]。

本文选取的奖励函数如式(29)所示, 其中, e 为输出电压 y 与给定输入电压 v 之间的误差。

$$R = \begin{cases} (300 - e) \times 2000, & |e| < 0.2 \\ (300 - e) \times 100, & 0.2 \leq |e| < 1 \\ -(300 - e) \times 10, & 1 \leq |e| < 10 \\ -100, & |e| \geq 10 \end{cases} \quad (29)$$

3.5 控制策略的设计

ADRC 控制器下一时刻的控制量只与当前控制误差有关。因此 ADRC 具有马尔可夫性, 所以控制过程可以看作一类决策问题。在不改变传统 LADRC 结构的基础上, 将输出电压 y 与给定输入电压 v 之间的误差、输出电压观测值 z_1 与 v 之间的误差以及 z_2 与输出微分值 \dot{y} 之间的误差作为 DQN 算法的输入观察值, DQN 算法的输出动作值为 ω_0 , 从而对 LADRC 中的观测器带宽进行在线优化。DQN-LADRC 控制策略框图如图 6 所示。

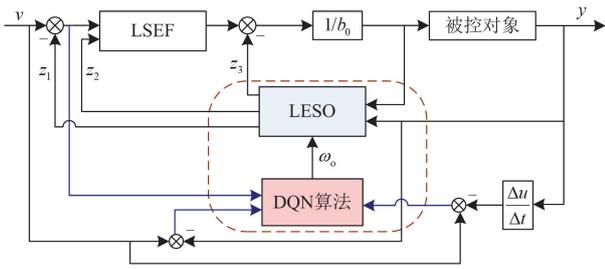


图 6 DQN-LADRC 控制策略框图

Fig. 6 Block diagram of the DQN-LADRC control strategy

4 仿真分析

4.1 仿真参数

为验证所提 DQN-LADRC 的正确性与有效性，在 Matlab/Simulink 中搭建数字仿真模型，并进行仿真分析。在不同工况下，分别对比分析双闭环 PI 控制、LADRC 与 DQN-LADRC 下的负载侧输出电压波形。

Buck 变换器的仿真参数如表 1 所示，控制器参数如表 2 所示。

表 1 Buck 变换器参数

Table 1 Buck converter parameters

参数	数值	参数	数值
直流母线电压/V	500	电感/ μH	300
开关频率/kHz	100	储能侧电压/V	300
电容/ μF	500	负载/ Ω	10

表 2 控制器参数

Table 2 Controller parameters

控制方式	参数	数值
PI 电流内环	k_{p1}	0.0113
	k_{i1}	60.65
PI 电压外环	k_{p2}	0.8526
	k_{i2}	563.5
LADRC	ω_c	3750
	ω_o	20000

在对比分析前，需将 DQN 智能体进行多次离线优化，训练控制器参数，再将离线训练好的智能体用于在线控制，以期得到更好的控制效果，DQN 超参数如表 3 所示。图 7 为平均奖励值收敛曲线，由该曲线图可以看出，奖励均值在 250 回合前快速变化，在 270 回合后趋于平稳，训练平稳后智能体即可实现平稳输出，其输出值则为动作优化值。

4.2 仿真结果与分析

1) 工况一：直流母线侧电压突变

考虑能量的流动性，在 0.04 s 处分别施加电压

突增 10%、电压突降 10% 的扰动，对比分析不同控制策略在电压突变扰动下的抗扰性。直流母线负载侧电压波形变化如图 8 所示，不同状况下的超调量和调节时间分别如表 4 和表 5 所示。

表 3 DQN 超参数

Table 3 Hyperparameters of the DQN

DQN 超参数	数值
折扣因子 γ	0.99
学习率 α	1×10^{-4}
经验回放池 D	1×10^6
最小取样批次 M	128

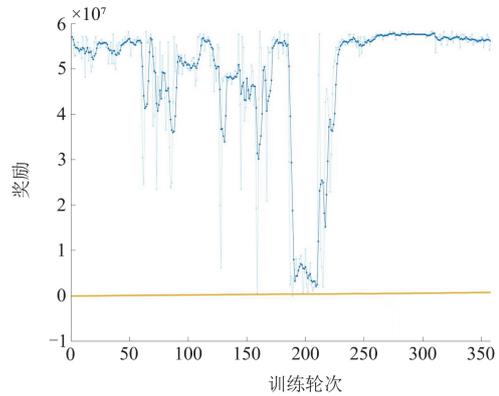
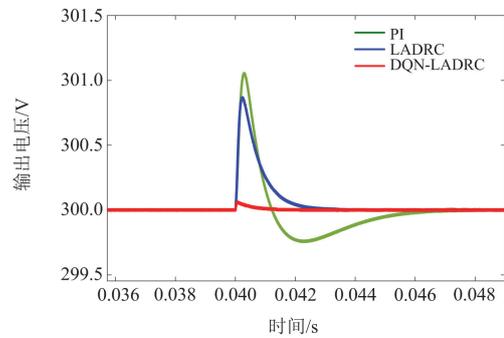
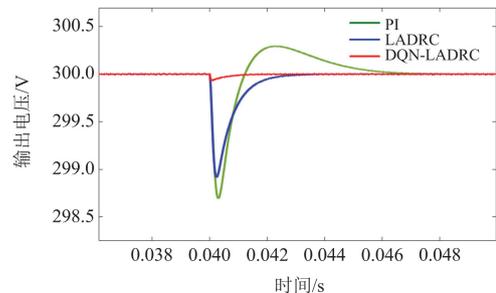


图 7 平均奖励值收敛曲线

Fig. 7 Convergence curve of average reward value



(a) 直流母线侧电压突增 10%



(b) 直流母线侧电压突降 10%

图 8 直流母线侧电压突变时负载侧电压

Fig. 8 Sudden change in voltage on the DC bus side

表 4 直流母线侧电压突变下的超调量

Table 4 Overshoot of DC bus side voltage change

扰动类型	PI	LADRC	DQN-LADRC
电压突增 10%	0.37%	0.29%	0.021%
电压突降 10%	0.43%	0.36%	0.024%

表 5 直流母线侧电压突变下的调节时间

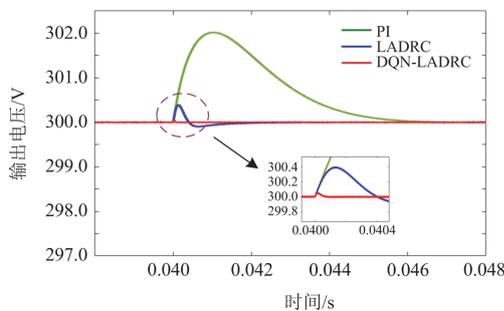
Table 5 Adjustment time of DC bus side voltage change

扰动类型	PI	LADRC	DQN-LADRC
电压突增 10%	7.276 ms	4.033 ms	1.731 ms
电压突降 10%	7.160 ms	4.112 ms	1.827 ms

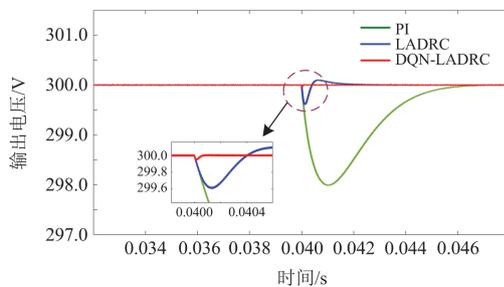
通过仿真分析以及表 4、表 5 可知, 在电压突增扰动下, DQN-LADRC 相比于 LADRC 和 PI 控制, 其超调量分别降低了 0.269%、0.349%, 调节时间分别降低了 2.302 ms、5.545 ms; 在电压突降扰动下, DQN-LADRC 相比于 LADRC 和 PI 控制, 其超调量分别降低了 0.336%、0.406%, 调节时间分别降低了 2.285 ms、5.333 ms。因此, DQN-LADRC 的控制效果优于 LADRC、PI 控制。

2) 工况二: 直流母线负载突变

考虑直流母线侧负载突变对系统的干扰, 在 0.04 s 时分别施加负载突增 10%、负载突降 10% 的扰动。直流母线负载侧电压波形如图 9 所示, 不同状态下的超调量和调节时间分别如表 6 和表 7 所示。



(a) 负载突增10%



(b) 负载突降10%

图 9 负载突变时直流母线负载侧电压

Fig. 9 Voltage of DC bus load side when load mutations

表 6 直流母线侧负载突变下的超调量

Table 6 Overshoot of DC bus side load change

扰动类型	PI	LADRC	DQN-LADRC
负载突增 10%	0.68%	0.14%	0.02%
负载突降 10%	0.67%	0.15%	0.019%

表 7 直流母线侧负载突变下的调节时间

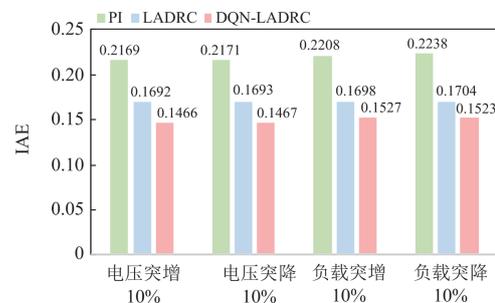
Table 7 Adjustment time of DC bus side load change

扰动类型	PI	LADRC	DQN-LADRC
负载突增 10%	6.697 ms	3.251 ms	0.512 ms
负载突降 10%	7.022 ms	3.49 ms	0.602 ms

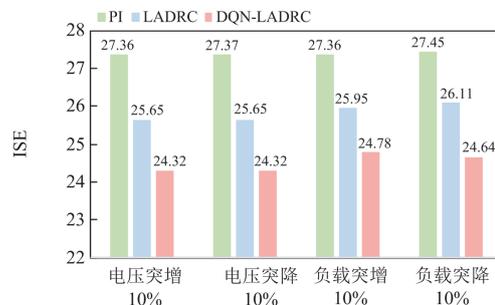
通过仿真分析以及表 6、表 7 可知, 在负载突增扰动下, DQN-LADRC 相比于 LADRC 和 PI 控制, 其超调量分别降低了 0.12%、0.66%, 调节时间分别降低了 2.739 ms、6.482 s; 在负载突降扰动下, DQN-LADRC 相比于 LADRC 和 PI 控制, 其超调量分别降低了 0.131%、0.651%, 调节时间分别降低了 2.888 ms、6.42 ms。因此, DQN-LADRC 的控制效果优于 LADRC 和 PI 控制。

3) 工况三: 不同工况下整体性能分析

为了分析系统中控制器在不同工况下控制效果的稳态性能, 选用偏差绝对值积分值(integral absolute error, IAE)、偏差平方值积分(integral square error, ISE)作为性能指标, 不同工况下的 IAE 和 ISE 如图 10 所示。



(a) 不同工况下 IAE



(b) 不同工况下 ISE

图 10 不同工况下的 IAE 和 ISE

Fig. 10 IAE and ISE under different operating conditions

通过对比不同控制器在不同工况下的 IAE 和 ISE 可知, 在不同工况下, DQN-LADRC 的 IAE 和 ISE 均比 LADRC 和 PI 控制低。因此, DQN-LADRC 的整体性能优于 LADRC 和 PI 控制。

4.3 蒙特卡洛实验仿真结果

为了分析不同控制器的鲁棒性, 在控制器参数不变的情况下, 令被控系统内部参数 C 、 L 产生 $\pm 15\%$ 的随机波动, 对 3 种控制器各进行重复实验, 对比分析控制器的控制效果。为更好地对比分析不同控制策略的差异性, 选取 IAE、ISE 作为性能指标并绘制散点图, 如图 11 所示。

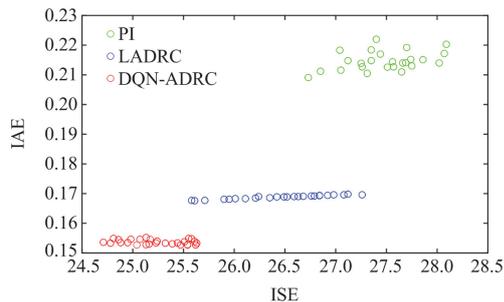


图 11 蒙特卡洛实验结果散点图

Fig. 11 Scatter plot of Monte Carlo experiment results

由图 11 可以看出, 在系统参数扰动下 DQN-LADRC 和 LADRC 的鲁棒性优于 PI 控制。本文所提出的 DQN-LADRC 在 LADRC 的基础上, 使用 DQN 算法在线优化控制器的参数, 从而优化了控制效果。根据蒙特卡洛实验的散点分布表明, 本文所提出的 DQN-LADRC 具有更好的鲁棒性。

5 结论

为了保证负载与直流母线间能量流动的效率和稳定性, 优化其控制, 解决不同工况下的扰动问题, 本文结合自抗扰鲁棒性强和深度强化学习自适应性强的特点, 设计了一种基于 DQN 算法的自抗扰控制策略, 具有以下特点:

1) 通过求取线性扩张状态观测器和线性误差反馈控制下控制系统的精确结构模型, 对其进行结构简化设计, 并利用简化的结构求取观测器带宽参数稳定优化范围, 进一步提高控制精度和控制效果;

2) 结合深度强化学习算法, 在所求取的自抗扰控制器参数稳定优化范围中进一步利用 DQN 算法实现参数整定和优化调整, 最终实现控制器控制精度的有效提高以及系统鲁棒性能、抗扰性能的增强。

通过对控制策略的算法设计和仿真验证, 说明了 DQN-LADRC 策略在多种工况下均具有良好的抗扰性能和鲁棒性能。深度强化学习与自抗扰控制

器结合, 对于未来探索研究控制方面问题具有重要的意义。

参考文献

- [1] 程林, 万宇翔, 齐宁, 等. 含多种分布式资源的配用电系统运行可靠性研究评述及展望[J]. 电力系统自动化, 2021, 45(22): 191-207.
CHENG Lin, WAN Yuxiang, QI Ning, et al. Review and prospect of research on operation reliability of power distribution and consumption system considering various distributed energy resources[J]. Automation of Electric Power Systems, 2021, 45(22): 191-207.
- [2] 梅文龙, 施佳余, 张旭东, 等. 微电网发展研究[J]. 电源技术, 2017, 41(2): 334-336.
MEI Wenlong, SHI Jiayu, ZHANG Xudong, et al. Research on microgrid development[J]. Chinese Journal of Power Sources, 2017, 41(2): 334-336.
- [3] 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 microgrid hybrid energy storage system[J]. Power System Protection and Control, 2021, 49(3): 177-187.
- [4] 王成山, 李微, 王议锋, 等. 直流微电网母线电压波动分类及抑制方法综述[J]. 中国电机工程学报, 2017, 37(1): 84-98.
WANG Chengshan, LI Wei, WANG Yifeng, et al. DC bus voltage fluctuation classification and restraint methods review for DC microgrid[J]. Proceedings of the CSEE, 2017, 37(1): 84-98.
- [5] 赵忠斌, 张靖, 马蕊, 等. 互联直流微电网多模式协调控制策略[J]. 智慧电力, 2020, 48(4): 28-35.
ZHAO Zhongbin, ZHANG Jing, MA Rui, et al. Multi-mode coordinated control strategy of interconnected DC microgrid[J]. Smart Power, 2020, 48(4): 28-35.
- [6] 陈景文, 周媛, 李晓飞, 等. 光储直流微网混合储能控制策略研究[J]. 智慧电力, 2022, 50(1): 14-20, 87.
CHEN Jingwen, ZHOU Yuan, LI Xiaofei, et al. Hybrid energy storage control strategy of optical storage DC microgrid[J]. Smart Power, 2022, 50(1): 14-20, 87.
- [7] 张化光, 路文贇, 刘秀翀. 基于电流模式的 Buck 变换器的模糊控制[J]. 控制工程, 2009, 16(4): 472-474.
ZHANG Huaguang, LU Wenzhe, LIU Xiuchong. Fuzzy control for buck converter based on current mode control[J]. Control Engineering of China, 2009, 16(4): 472-474.
- [8] 孙永军, 周茜, 赵国强, 等. 基于自适应 PID 控制的 Buck 变换器设计与实现[J]. 煤矿机电, 2012, 33(5): 1-4, 8.
SUN Yongjun, ZHOU Qian, ZHAO Guoqiang, et al. Design and implementation of Buck converter based on self-adaptive PID control[J]. Colliery Mechanical & Electrical Technology, 2012, 33(5): 1-4, 8.
- [9] WANG Y, FANG S, HU J. Active disturbance rejection

- control based on deep reinforcement learning of PMSM for more electric aircraft[J]. IEEE Transactions on Power Electronics, 2022, 38(1): 406-416.
- [10] 杨惠, 骆姗, 孙向东, 等. 光伏储能双向 DC-DC 变换器的自抗扰控制方法研究[J]. 太阳能学报, 2018, 39(5): 1342-1350.
YANG Hui, LUO Shan, SUN Xiangdong, et al. Research on ADRC method for bidirectional DC-DC converter of solar energy storage system[J]. Acta Energetica Sinica, 2018, 39(5): 1342-1350.
- [11] 袁东, 马晓军, 曾庆含, 等. 二阶系统线性自抗扰控制器频带特性与参数配置研究[J]. 控制理论与应用, 2013, 30(12): 1630-1640.
YUAN Dong, MA Xiaojun, ZENG Qinghan, et al. Research on frequency band characteristics and parameters configuration of linear active disturbance rejection control for second-order systems[J]. Control Theory & Applications, 2013, 30(12): 1630-1640.
- [12] 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.
- [13] 蒋谦, 黄志豪, 程启明, 等. 基于 MMC 的 PET 中间隔离级 DC-DC 变换器的新型模型预测控制策略[J]. 电力建设, 2022, 43(4): 49-57.
JIANG Qian, HUANG Zhihao, CHENG Qiming, et al. Novel model predictive control strategy for MMC-based DC-DC converter in the intermediate isolation stage of PET[J]. Electric Power Construction, 2022, 43(4): 49-57.
- [14] 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.
- [15] 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.
- [16] 张延宇, 饶新朋, 周书奎, 等. 基于深度强化学习的电动汽车充电调度算法研究进展[J]. 电力系统保护与控制, 2022, 50(16): 179-187.
ZHANG Yanyu, RAO Xinpeng, ZHOU Shukui, et al. Research progress of electric vehicle charging scheduling algorithms based on deep reinforcement learning[J]. Power System Protection and Control, 2022, 50(16): 179-187.
- [17] 胡正伟, 夏思懿, 王文彬, 等. 基于深度强化学习的 II 型阻抗匹配网络多参数最优求解方法[J]. 电力系统保护与控制, 2024, 52(6): 152-163.
HU Zhengwei, XIA Siyi, WANG Wenbin, et al. Multi-parameter optimal solution method for II-type impedance matching networks based on deep reinforcement learning[J]. Power System Protection and Control, 2024, 52(6): 152-163.
- [18] SU J, MA C, LI S, et al. An AGC dynamic control method based on DQN algorithm[J]. IOP Conference Series: Materials Science and Engineering, 2020, 729(1): 012009.
- [19] 周雪松, 张心茹, 赵浩宇, 等. 基于 DDPG 算法的微电网负载端接口变换器自抗扰控制[J]. 电力系统保护与控制, 2023, 51(21): 66-75.
ZHOU Xuesong, ZHANG Xinru, ZHAO Hanyu, et al. Active disturbance rejection control of a microgrid load-side interface converter based on a DDPG algorithm[J]. Power System Protection and Control, 2023, 51(21): 66-75.
- [20] 裴培, 何绍溟, 王江, 等. 一种深度强化学习制导控制一体化算法[J]. 宇航学报, 2021, 42(10): 1293-1304.
PEI Pei, HE Shaoming, WANG Jiang, et al. An integrated algorithm of deep reinforcement learning guidance and control[J]. Journal of Astronautics, 2021, 42(10): 1293-1304.
- [21] GUO J, WANG Y, AN H, et al. IIDQN: an incentive improved DQN algorithm in EBSN recommender system[J]. Security & Communication Networks, 2022, 12: 1-12.
- [22] ZHENG Y, SUN Q, CHEN Z, et al. Deep Q-network based real-time active disturbance rejection controller parameter tuning for multi-area interconnected power systems[J]. Neurocomputing, 2021: 360-373.
- [23] 赵恬恬, 孔建国, 梁海军, 等. 未知环境下基于 Dueling DQN 的无人机路径规划研究[J]. 现代计算机, 2024, 30(5): 37-43.
ZHAO Tiantian, KONG Jianguo, LIANG Haijun, et al. Research on UAV path planning based on Dueling DQN in unknown environment[J]. Modern Computer, 2024, 30(5): 37-43.
- [24] LI Y, QIU X, LIU X, et al. Deep reinforcement learning and its application in autonomous fitting optimization for attack areas of UCAVs[J]. Journal of Systems Engineering and Electronics, 2020, 31(4): 734-742.
- [25] LIANG Y, GUO C, DING Z, et al. Agent-based modeling in electricity market using deep deterministic policy gradient algorithm[J]. IEEE Transactions on Power Systems, 2020, 35(6): 4180-4192.

收稿日期: 2024-04-21; 修回日期: 2024-06-03

作者简介:

周雪松(1964—), 男, 通信作者, 博士, 教授, 研究方向为新能源与智能电网; E-mail: zxsmyj@126.com

韩静(1999—), 女, 硕士研究生, 研究方向为智能电网、电力系统保护; E-mail: 1776556885@qq.com

马幼捷(1964—), 女, 博士, 教授, 研究方向为微电网运行与控制。

(编辑 姜新丽)