

融合无模型强化学习的永磁同步电机混沌抗扰控制

谭浚楷^{1,2}, 薛霜思^{1,2}, 郭子航^{1,2}, 曹晖^{1,2}

(1. 西安交通大学电气工程学院, 陕西 西安 710049;
2. 西安交通大学电力设备电气绝缘国家重点实验室, 陕西 西安 710049)

摘要: 针对永磁同步电机 (permanent magnet synchronous motor, PMSM) 控制中出现的由非线性震荡引发的混沌行为, 提出了一种利用电机历史运行数据迭代获取最优控制策略的无模型强化学习方法, 该方法解决了不同工作环境下因负载扰动不确定、电机模型参数不确定带来的问题。首先针对外部负载扭矩的不确定性, 建立电机最优控制与外部最坏扰动之间的零和博弈, 设计迭代形式的黎卡提方程, 基于该方程进一步构建了一种基于模型的抗干扰最优控制器。在该控制器基础上, 引入无模型强化学习思想, 设计了 PMSM 混沌现象的数据驱动稳定方法, 利用电机历史运行数据实现优化求解无模型抗扰动稳定控制器, 实现对电机混沌行为的抗扰稳定控制。最后通过多个数值仿真实验对比验证了所设计方法的性能。实验结果显示, 在外界负载扰动不确定的情况下, 提出的方法相较传统有限时间控制方法在累计资源消耗损失性能方面提升了 39.04%, 在电机模型参数不确定的情况下, 提出的方法相比线性二次型调节器的控制成功率提高 10.71%。

关键词: 永磁同步电机; 混沌现象; 强化学习; 零和博弈; 前馈补偿控制器

中图分类号: TP39

文献标志码: A

doi: 10.11959/j.issn.2096-6652.202523

Chaotic disturbance rejection control of permanent magnet synchronous motor integrating model-free reinforcement learning

TAN Junkai^{1,2}, XUE Shuangxi^{1,2}, GUO Zihang^{1,2}, CAO Hui^{1,2}

1. School of Electrical Engineering, Xi'an Jiaotong University, Xi'an 710049, China
2. State Key Laboratory of Electrical Insulation and Power Equipment, Xi'an Jiaotong University, Xi'an 710049, China

Abstract: For the chaotic behavior of nonlinear oscillations occurring in the control of permanent magnet synchronous motor (PMSM), a data-driven model-free reinforcement learning method was investigated, which only utilized the historical state data of the motor to obtain the optimal controller. This method solved the problem of uncertain external load and uncertain motor dynamic under varying and unknown working conditions. Firstly, to solve the uncertainty of external torque load, a zero-sum game between the controller and external disturbance was constructed. By redesigning the Riccati equation in iteration form, a model-based robust optimal controller was obtained. Based on this controller, by introducing the model-free reinforcement learning method, a data-driven stabilization method for chaotic phenomena in PMSM was proposed, which obtained a model-free robust optimal controller by learning from the historical operation data. Finally, the performance of the designed method was verified by comparing several simulations, and the results showed that the energy effort saving performance of our proposed method was improved by 39.04% compared with the traditional finite-time control method under the uncertainty of external load perturbation, and the success rate was improved by 10.71% compared with the linear quadratic regulator under the uncertainty of the motor model parameters.

收稿日期: 2025-03-07; 修回日期: 2025-05-26

通信作者: 薛霜思, xssxjt@xjtu.edu.cn

基金项目: 中国博士后科学基金 (No.2024M762602)

Foundation Item: The China Postdoctoral Science Foundation (No.2024M762602)

Key words: PMSM, chaos, reinforcement learning, zero-sum game, feedforward compensation controller

0 引言

永磁同步电机 (permanent magnet synchronous motor, PMSM) 因具有高效率、高性能等特点, 被广泛应用于工业生产、交通传动等领域^[1-5]。然而, 电机的电磁特性导致其有可能进入运行不稳定的混沌状态, 对电机本身的电控系统与机械结构造成不可逆的伤害, 其不可预测性甚至可能对操作者的人身安全造成威胁。

在无控制输入的情况下, PMSM 的运行状态因混沌现象而不断抖动, 这给电机系统的安全运行带来巨大挑战。针对 PMSM 中的混沌现象, 学术界进行了大量的研究^[6-8]。文献[9]对 PMSM 中混沌动力学进行建模, 并设计了一种带非线性项补偿的控制器。文献[10-11]分别引入了汉密尔顿-雅可比-贝尔曼方程和代数黎卡提方程求解永磁同步电机混沌现象抑制的最优控制器。文献[12]设计了抑制带负载 PMSM 混沌的有限时间控制器。在 PMSM 的混沌现象抑制研究中, 常用的方法还有自适应控制^[13-17]。文献[18]研究了一种基于神经网络的自适应反步法控制策略, 并设计了障碍李雅普诺夫函数来确保系统状态满足约束条件。文献[19]利用两个等效输入扰动估计器对可能产生扰动的非线性项进行补偿, 以实现混沌抑制控制。然而, 上述方法都是基于模型的。PMSM 容易受到外部环境因素如温度、电磁等的影响, 从而造成模型参数偏移, 使得原本的控制不再准确, 从而显著降低混沌抑制的性能。因此, 研究无模型的混沌现象抑制器具有显著的现实意义。

强化学习是一种强大的机器学习范式, 它通过智能体与环境的交互来学习最优策略, 该领域包含基于模型和无模型两类方法。其中, 无模型强化学习由于不依赖精确的系统动力学模型, 在处理具有不确定性或未知动态的复杂系统方面显示出巨大潜力, 它能直接从数据中学习控制策略, 不需要精确模型。无模型强化学习的关键技术包括: 利用数据辨识系统模型再设计控制器的方法 (如结合神经网络与神经动态规划的方法^[20-21])、提高数据效率的异策略强化学习方法^[22], 以及通过虚拟系统指导实际控制的平行动态优化技术^[23-24]。具体到 PMSM 的转速控制, 文献[2,25]研究了无模型异策略强化

学习, 通过收集非最优策略的运行数据来训练完全未知无模型信息下的最优速度跟踪控制器。文献[26]设计了演员-评论家-系统辨识的控制框架, 实现了伺服速度控制器和系统模型的在线学习。文献[27-28]考虑了存在动态扰动下的 PMSM 强化学习控制策略。文献[29]则研究了无具体混沌系统的自适应动态规划控制方法。以上强化学习方法可以解决 PMSM 因温度、湿度等环境条件变化而导致的模型动力学参数变化的问题。尽管上述这些技术取得了显著进展, 但将它们应用于具有快速动态和非线性的 PMSM 混沌控制仍面临挑战。

综上, 针对 PMSM 控制中出现的混沌现象的稳定控制问题, 本文提出了一种负载和模型未知条件下的融合无模型强化学习的最优控制方法, 主要内容如下。

(1) 针对 PMSM 的电磁特性, 建立了电机的非线性连续时间混沌动力学模型, 提出了一种非线性项的补偿控制项, 消去运行模型状态中的强耦合项, 简化后续策略迭代控制的被控对象。

(2) 电机运行过程中可能出现部分不稳定的、未知的外部扰动, 针对可能出现的最坏外部扰动建立了零和博弈, 根据 H 无穷理论, 求解迭代黎卡提方程, 获得在敌对性扰动情况下具有鲁棒性的基于模型的控制策略。

(3) 针对 PMSM 运行模型参数易受环境影响的特性, 设计了一种融合强化学习的策略迭代方法, 通过对电机的历史运行状态数据进行训练, 从而得到最优控制策略, 实现了无模型下的 PMSM 的混沌现象控制, 解决了电机模型参数漂移、不易控制的问题。

综上所述, 本文的主要贡献在于提出了一种融合非线性前馈补偿与无模型强化学习的新型控制框架, 专门用于解决负载和模型参数未知条件下的 PMSM 混沌稳定控制问题。该框架通过模型补偿简化控制任务, 并通过数据驱动的强化学习解决参数漂移和未知扰动难题, 旨在实现比单一方法更优的鲁棒性和自适应性能。

1 PMSM 的混沌现象建模

在 $d-q$ 坐标系下, PMSM 的矢量控制有如下所

示的非线性动力学数学模型^[2]:

$$\begin{cases} \frac{d\omega}{dt} = \frac{n_p \psi_f}{J} i_q - \frac{B}{J} \omega - \frac{1}{J} T_L \\ \frac{di_q}{dt} = -\frac{R_s}{L} i_q - n_p \omega i_d - \frac{n_p \psi_f}{L} \omega + \frac{1}{L} u_q \\ \frac{di_d}{dt} = -\frac{R_s}{L} i_d + n_p \omega i_q + \frac{1}{L} u_d \end{cases} \quad (1)$$

其中, ω 、 n_p 、 T_L 和 J 分别表示角速度、磁极对数、负载转矩和转子惯量, i_q 、 i_d 、 u_q 和 u_d 为定子 in q 轴和 d 轴上的电流和电压, L 为定子电感, R_s 为定子电阻, B_f 和 ψ_f 为粘性摩擦系数和转子力矩。本文假设 PMSM 的状态变量 ω 、 i_q 和 i_d 是可测的。

为方便后续混沌现象控制研究, 将 PMSM 的电磁参数进行合并, 并在时间尺度上进行变换, 其混沌动力学模型可建立为如下形式:

$$\begin{cases} \frac{dx_1(\tau)}{d\tau} = \sigma - d_L \\ \frac{dx_2(\tau)}{d\tau} = -x_2(\tau) - x_1(\tau)(x_3(\tau) - \gamma) + u_1 \\ \frac{dx_3(\tau)}{d\tau} = -x_3(\tau) + x_1(\tau)x_2(\tau) + u_2 \end{cases} \quad (2)$$

其中, 时间变换尺度 $\tau = R_s t / L$, 参数 $\sigma = B_f L / J R_s$ 、 $\gamma = n_p \psi_f / B_f R_s$, 控制输入量 $u_1 = n_p L \psi_f u_d / R_s^2 B_f$, $u_2 = n_p L \psi_f u_q / R_s^2 B_f$, 负载 $d_L = L^2 T_L / R_s^2 J$, 变换状态 $x_1 = R_s \omega / L$ 、 $x_2 = n_p L \psi_f i_q / R_s B_f$, $x_3 = n_p L \psi_f i_d / R_s B_f$ 。

上述模型存在非线性项, 为优化后续无模型强化学习控制的设计与训练流程, 本文提出如下形式的前馈补偿控制器, 以补偿混沌动力学中的非线性部分:

$$\begin{cases} u_{1c} = x_1(\tau)x_3(\tau) \\ u_{2c} = -x_1(\tau)x_2(\tau) \end{cases} \quad (3)$$

定义向量形式的状态变量 $\mathbf{x} = [x_1, x_2, x_3]^T \in \mathbb{R}^3$, 控制输入组合 $\mathbf{u}_f = [u_{1f}, u_{2f}]^T \in \mathbb{R}^2$, 输出 $y = x_3 = R_s \omega / L$, 则上述系统可改写为以下简化线性系统:

$$\begin{cases} \dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u}_f + \mathbf{D}d_L \\ y = \mathbf{C}\mathbf{x} = \frac{R_s}{L} \omega \end{cases} \quad (4)$$

其中, $\mathbf{A} = \begin{bmatrix} -\sigma & \sigma & 0 \\ \gamma & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$, $\mathbf{B} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$, $\mathbf{D} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ 。

故非线性系统的控制器 u_1 和 u_2 由前馈补偿控制器和反馈控制器两部分组成:

$$\begin{cases} u_1 = u_{1f} + u_{1c} \\ u_2 = u_{2f} + u_{2c} \end{cases} \quad (5)$$

至此, 本节将 PMSM 的动力学模型通过时间尺度变换转换为混沌动力学中的非线性模型, 利用前馈补偿控制器将模型中的非线性部分抵偿, 以便于后续数据驱动方法的数据收集与训练。

本文提出的控制策略采用混合方法。该方法首先利用已知的 PMSM 的基本非线性动力学结构 (见式 (1)) 及其变换形式 (见式 (2)) 来设计前馈补偿器 (见式 (3)), 以处理系统主要的、已知的非线性混沌动力学。随后, 结合了一个强化学习控制器。在此框架下, “无模型” 特指该强化学习控制器在处理系统不确定性方面的能力, 即它能够通过与环境直接交互进行学习, 从而补偿那些未被前馈控制器完全覆盖或模型本身未精确描述的部分。这些部分包括关键参数如电阻、电感、转子惯量的不确定性或时变性、未建模动态如摩擦的精确特性, 以及未知外部扰动如负载转矩等。强化学习控制器在应对这些影响时, 不需要建立这些未知或不确定因素的精确数学模型。因此, 该方法融合了模型知识处理已知动力学与无模型学习自适应处理不确定性和扰动, 旨在增强整体控制系统的鲁棒性和自适应性。

2 抗干扰控制器设计

为处理 PMSM 的参数漂移和未建模动态, 本节提出融合无模型强化学习的控制策略。该策略首先利用已知模型设计前馈补偿以简化控制; 然后, 对补偿后的系统采用基于策略迭代的强化学习, 利用运行数据学习最优策略, 以自适应补偿参数不确定性、残余动态和未知扰动。该策略结合了模型补偿与无模型学习, 利用已知信息提升性能, 并通过数据驱动学习确保对未知因素的鲁棒性, 尤其适用于参数多变、难建模的 PMSM 混沌控制。

2.1 面向零和博弈的纳什均衡

本节针对 PMSM 的混沌现象设计抑制混沌状态控制器。考虑可能出现的最坏扰动, 本节基于博弈理论设置零和博弈的奖励函数, 利用最优控制中基于模型的黎卡提方程迭代方法, 实现抗扰最优控制器的求解。零和博弈指的是控制者争夺有限的资源, 在零和博弈背景下, 由于总资源有限, 每个控制者的利益是互斥的, 又根据最优理论中每个控制者的目标是最大化自身利益, 故不同控制者之间是

非合作的。具有敌对性的干扰与抗扰控制之间的冲突属于利益冲突型的博弈，因此本文通过设置特定的奖励函数以代表零和博弈。基于最优控制理论和博弈论，设计如下形式的线性二次型性能指标：

$$J(\mathbf{x}, \mathbf{u}, \mathbf{d}) = \int_{t_0}^{\infty} r(\mathbf{x}, \mathbf{u}, \mathbf{d}, \tau) d\tau \quad (6)$$

其中， $r(\mathbf{x}, \mathbf{u}, \mathbf{d}, \tau) = \mathbf{x}^T \mathbf{Q} \mathbf{x} + \mathbf{u}^T \mathbf{R} \mathbf{u} - \gamma^2 \mathbf{d}^T \mathbf{d}$ 为最优控制器 \mathbf{u} 和扰动 \mathbf{d} 的零和博弈损失函数， $\mathbf{Q} \in \mathbb{R}^{3 \times 3} \geq 0$ 和 $\mathbf{R} \in \mathbb{R}^{2 \times 2} > 0$ 为对称矩阵， $\gamma \in \mathbb{R}^+$ 为衰减系数。

零和博弈在式 (6) 中主要体现在奖励函数 r 的项 $\mathbf{u}^T \mathbf{R} \mathbf{u}$ 和 $\gamma^2 \mathbf{d}^T \mathbf{d}$ 的系数符号相反，故控制器 \mathbf{u} 和扰动 \mathbf{d} 之间的目的相冲突。

假设 1^[22,30]：(\mathbf{A}, \mathbf{B}) 可控，($\mathbf{A}, \sqrt{\mathbf{Q}}$) 可观测。

定义 1：若 $\forall \mathbf{d} \in L_2[0, \infty)$ ，满足 $\left[\int_{t_0}^{\infty} (\mathbf{x}^T \mathbf{Q} \mathbf{x} + \mathbf{u}^T \mathbf{R} \mathbf{u}) d\tau \right] / \left[\int_{t_0}^{\infty} (\mathbf{d}^T \mathbf{d}) d\tau \right] \leq \gamma^2$ ，则上述动力学系统具有 L_2 有界的衰减增益 γ 。

定义 2：当输入-扰动组合 $(\mathbf{u}^*, \mathbf{d}^*)$ 满足 $J(\mathbf{x}(t_0), \mathbf{u}^*, \mathbf{d}^*) \leq J(\mathbf{x}(t_0), \mathbf{u}^*, \mathbf{d}^*) \leq J(\mathbf{x}(t_0), \mathbf{u}, \mathbf{d}^*)$ 时，该组合为零和博弈下的纳什均衡。

在假设 1 的条件下，同时满足定义 1 的条件时，上述混沌动力学系统是可稳定的。根据最优控制理论，最优控制输入 \mathbf{u}^* 和最坏扰动 \mathbf{d}^* 可以表示为：

$$\begin{cases} \mathbf{u}^* = -\mathbf{R}^{-1} \mathbf{B}^T \mathbf{P} \mathbf{x} \\ \mathbf{d}^* = \frac{1}{\gamma^2} \mathbf{D} \mathbf{P} \mathbf{x} \end{cases} \quad (7)$$

其中， \mathbf{P} 为强化学习中价值函数的系数矩阵。

根据定义 2， $(\mathbf{u}^*, \mathbf{d}^*)$ 为零和博弈下的纳什均衡，即如下的 min-max 问题的解：

$$J(\mathbf{x}_0, \mathbf{u}^*, \mathbf{d}^*) = \min_{\mathbf{u}} \max_{\mathbf{d}} \int_{t_0}^{\infty} r(\cdot, \tau) d\tau = \mathbf{x}^T \mathbf{P}^* \mathbf{x} \quad (8)$$

其中， \mathbf{x}_0 为系统的初始状态， \mathbf{P}^* 为价值函数的最优系数矩阵， $\mathbf{P} = \mathbf{P}^T > 0$ 满足博弈代数黎卡提方程：

$$\mathbf{A}^T \mathbf{P}^* + \mathbf{P}^* \mathbf{A} + \mathbf{Q} - \mathbf{P}^* \mathbf{B} \mathbf{R}^{-1} \mathbf{B}^T \mathbf{P}^* + \frac{1}{\gamma^2} \mathbf{P}^* \mathbf{D} \mathbf{D}^T \mathbf{P}^* = 0 \quad (9)$$

2.2 融合系统模型的黎卡提方程迭代求解

为求解最优控制策略与负载扭矩扰动之间零和

博弈的纳什均衡，本文设计融合模型的黎卡提方程迭代求解方法，后续将其改进为数据驱动无模型强化学习方法。首先给出零和博弈存在最优解的充分条件。

定理 1：给定 $\mathbf{R} > 0$ 和 $\gamma > 0$ ，如果存在 \mathbf{Q} 、 \mathbf{R} 、 γ 和解 \mathbf{P} 满足黎卡提方程 (9) 和以下等式：

$$(\mathbf{A} - \mathbf{B} \mathbf{K}^*)^T \mathbf{P} + \mathbf{P} (\mathbf{A} - \mathbf{B} \mathbf{K}^*) + \mathbf{Q} + \frac{1}{\gamma^2} \mathbf{P} \mathbf{D} \mathbf{D}^T \mathbf{P} + \mathbf{K}^{*T} \mathbf{R} \mathbf{K}^* = 0 \quad (10)$$

则对应的控制增益 \mathbf{K} 等于最优控制增益 \mathbf{K}^* 。

证明：黎卡提方程重写为以下形式：

$$(\mathbf{A} - \mathbf{B} \mathbf{K}^*)^T \mathbf{P} + \mathbf{P} (\mathbf{A} - \mathbf{B} \mathbf{K}^*) + \frac{1}{\gamma^2} \mathbf{P} \mathbf{D} \mathbf{D}^T \mathbf{P} + \mathbf{Q} + \mathbf{K}^T \mathbf{R} \mathbf{K}^* + \mathbf{K}^{*T} \mathbf{R} \mathbf{K} - \mathbf{K}^T \mathbf{R} \mathbf{K} = 0 \quad (11)$$

将式 (11) 与式 (9) 相减，并根据奖励函数的定义中 $\mathbf{R} > 0$ ，得 $\mathbf{K} = \mathbf{K}^*$ 。证明完毕。

定理 1 给出了零和博弈下最优控制问题存在解满足的充分条件。现设计黎卡提方程迭代求解形式为：

$$\begin{aligned} & (\mathbf{A} - \mathbf{B} \mathbf{K}^i)^T \mathbf{P}^i + \mathbf{P}^i (\mathbf{A} - \mathbf{B} \mathbf{K}^i) + \\ & \frac{1}{\gamma^2} \mathbf{P}^i \mathbf{D} \mathbf{D}^T \mathbf{P}^i + \mathbf{Q} + (\mathbf{K}^i)^T \mathbf{R} \mathbf{K}^i = 0 \end{aligned} \quad (12)$$

利用求解得到的 \mathbf{P}^i ，结合式 (7) 的最优控制器形式，得到迭代形式的控制器，如下所示：

$$\begin{cases} \mathbf{u}^{i+1} = -\mathbf{K}^{i+1} \mathbf{x} = -\mathbf{R}^{-1} \mathbf{B}^T \mathbf{P}^i \mathbf{x} \\ \mathbf{d}^{i+1} = \mathbf{L}^{i+1} \mathbf{x} = \frac{1}{\gamma^2} \mathbf{D} \mathbf{P}^i \mathbf{x} \end{cases} \quad (13)$$

重复利用式 (12) 求解 \mathbf{P}^i 和利用式 (13) 更新控制律 \mathbf{u}^{i+1} 和 \mathbf{d}^{i+1} ，直到 $|\mathbf{P}^{i+1} - \mathbf{P}^i| < \varepsilon$ ， ε 为阈值，得到的就是融合系统模型的黎卡提方程迭代求解算法，或称为融合模型的策略迭代强化学习方法，这种方法迭代求解速度较快，但前提是需要 PMSM 准确的系统动态信息，然而，控制器性能可能会因环境变化和系统参数变化而大幅降低，因此，基于上述算法的局限性进行改进，本文提出了不需要系统模型参数的无模型强化学习方法。

3 融合无模型强化学习的数据驱动控制

由于 PMSM 的运行状况易受环境的影响，本节设计了一种数据驱动的无模型策略强化学习 (model-free off-policy reinforcement learning, MFOPRL)

方法，其中异策略指的是预先定义好的非最优策略，用于强化学习的探索阶段执行。MFOPRL方法只利用执行非最优策略时收集得到的数据 $(\mathbf{x}, \mathbf{u}, \mathbf{d})$ 驱动迭代得到第 i 次迭代的价值系数 \mathbf{P}^i 、第 $i+1$ 次迭代的控制器增益 \mathbf{K}^{i+1} 和第 $i+1$ 次迭代的扰动增益 \mathbf{L}^{i+1} 。首先利用预先定义的非最优策略 \mathbf{u}_E 和干扰 \mathbf{d}_E 进行探索，收集状态数据 \mathbf{x}_E ：

$$\dot{\mathbf{x}}_E = \mathbf{A}\mathbf{x}_E + \mathbf{B}\mathbf{u}_E + \mathbf{D}\mathbf{d}_E \quad (14)$$

其中，非最优预定义异策略 $\mathbf{u}_E = -\mathbf{K}_E\mathbf{x}_E + \mathbf{e}$ 为一个预设的固定负反馈控制策略加一个具有随机性的探索噪音 \mathbf{e} 。受文献[31]的启发，在探索阶段，执行异策略 \mathbf{u}_E 可以使收集的数据满足暂态激励条件。在收集到足够的 $(\mathbf{x}_E, \mathbf{u}_E, \mathbf{d}_E)$ 数据后，进行控制器的训练。设 \mathbf{K}^{i+1} 为第 i 次迭代的控制器增益，上述系统可改写为：

$$\dot{\mathbf{x}}_E = \mathbf{A}\mathbf{x}_E - \mathbf{B}\mathbf{K}^i\mathbf{x}_E + \mathbf{D}\mathbf{d}_E + \mathbf{B}(\mathbf{u}_E + \mathbf{K}^i\mathbf{x}_E) \quad (15)$$

将该系统动态代入代数黎卡提方程(11)并积分，每段积分从 t 到 $t+T$ ， $T>0$ 为每段积分的长度：

$$\begin{aligned} & \mathbf{x}_E(t+T)^T \mathbf{P}^i \mathbf{x}_E(t+T) - \mathbf{x}_E(t)^T \mathbf{P}^i \mathbf{x}_E(t) - \\ & 2 \int_t^{t+T} \left((\mathbf{u}_E + \mathbf{K}_E \mathbf{x}_E)^T \mathbf{R} \mathbf{K}^{i+1} \mathbf{x} + \gamma^2 \mathbf{d}^T \mathbf{L}^{i+1} \mathbf{x}_E \right) d\tau \quad (16) \\ & = - \int_t^{t+T} \mathbf{x}_E^T \left(\mathbf{Q}^i + \mathbf{K}_E^T \mathbf{R} \mathbf{K}_E + \gamma^2 (\mathbf{L}^i)^T \mathbf{L}^i \right) \mathbf{x}_E d\tau \end{aligned}$$

为方便后续对式(16)的迭代求解，将异策略收集的状态 \mathbf{x}_E 和迭代第 i 轮的解 \mathbf{P}^i 转为向量形式 $\bar{\mathbf{x}}_E \in \mathbb{R}^{\bar{n}}$ 和 $\bar{\mathbf{P}}^i \in \mathbb{R}^{\bar{n}}$ ， $\bar{\mathbf{x}}_E$ 由 \mathbf{x}_E 克罗内克积得到： $\bar{\mathbf{x}}_E = [x_{E1}^2, 2x_{E1}x_{E2}, \dots, x_{E2}^2, 2x_{E2}x_{E3}, \dots, x_{En}^2]^T$ ， $\bar{\mathbf{P}}^i$ 为 \mathbf{P}^i 矩阵的拉直运算，由式(9)的条件可知， \mathbf{P}^i 为对称矩阵，故 $\bar{\mathbf{P}}^i = [\mathbf{P}_{11}^i, \mathbf{P}_{12}^i, \dots, \mathbf{P}_{22}^i, \mathbf{P}_{23}^i, \dots, \mathbf{P}_{nn}^i]^T$ 。

根据克罗内克积的表示形式，可知式(16)中的二次型项可以转化为如下形式（具体推导过程如附录所示）：

$$\mathbf{x}_E^T \mathbf{P}^i \mathbf{x}_E = (\mathbf{x}_E^T \otimes \mathbf{x}_E^T) \text{vec}(\mathbf{P}^i) = \bar{\mathbf{x}}_E^T \bar{\mathbf{P}}^i \quad (17)$$

$$\begin{aligned} & (\mathbf{u}_E + \mathbf{K}_E \mathbf{x}_E)^T \mathbf{R} \mathbf{K}^{i+1} \mathbf{x}_E = \mathbf{e}^T \mathbf{R} \mathbf{K}^{i+1} \mathbf{x}_E \\ & = (\mathbf{x}_E^T \otimes \mathbf{e}^T) (\mathbf{I}_n \otimes \mathbf{R}) \text{vec}(\mathbf{K}^{i+1}) \quad (18) \end{aligned}$$

$$\gamma^2 \mathbf{d}^T \mathbf{L}^{i+1} \mathbf{x}_E = \gamma^2 (\mathbf{x}_E^T \otimes \mathbf{d}^T) \text{vec}(\mathbf{L}^{i+1}) \quad (19)$$

$$\text{设 } \mathbf{I}_{\bar{\mathbf{x}}} = \left[\int_t^{t+T} \bar{\mathbf{x}}_E d\tau, \dots, \int_{t+(l-1)T}^{t+lT} \bar{\mathbf{x}}_E d\tau \right]^T, \quad \delta_{\mathbf{x}} = [\bar{\mathbf{x}}_E$$

$(t+T) - \bar{\mathbf{x}}_E(t), \dots, \bar{\mathbf{x}}_E(t+lT) - \bar{\mathbf{x}}_E(t+(l-1)T)]^T$ 和 $\mathbf{I}_{\mathbf{x}} = \left[\int_t^{t+T} \mathbf{x}_E \otimes d\tau, \dots, \int_{t+(l-1)T}^{t+lT} \mathbf{x}_E \otimes d\tau \right]^T$ ，则式(16)左侧变为 $\Psi \Theta^i$ ，右侧变为 Ξ^i ：

$$\Psi = [\delta_{\mathbf{x}}, -2\mathbf{I}_{\mathbf{x}e}(\mathbf{I}_n \otimes \mathbf{R}), -2\gamma^2 \mathbf{I}_{\mathbf{x}d}]^T \quad (20)$$

$$\Theta^i = \left[(\bar{\mathbf{P}}^i)^T, \text{vec}(\mathbf{K}^{i+1})^T, \text{vec}(\mathbf{L}^{i+1})^T \right]^T \quad (21)$$

$$\Xi^i = - \left[\int_t^{t+T} \mathbf{r}^i d\tau, \dots, \int_{t+(l-1)T}^{t+lT} \mathbf{r}^i d\tau \right]^T \quad (22)$$

其中， $l \geq n(n+1)/2 + nm + nz$ 。利用批量最小二乘法， $\bar{\mathbf{P}}^i$ 、 \mathbf{K}^{i+1} 和 \mathbf{L}^{i+1} 可以计算得到：

$$\Theta^i = (\Psi^T \Psi)^{-1} \Psi^T \Xi^i \quad (23)$$

根据上述提出的数据驱动方法，在无系统模型参数的条件下，通过收集执行异策略下的电机运行状态 $(\mathbf{x}, \mathbf{u}, \mathbf{d})$ 的数据，利用策略迭代获取最优控制增益，同时对可能出现的最坏扰动 \mathbf{d}^* 具有抵抗能力。无模型的异策略数据驱动强化学习算法如下。

算法1 无模型的异策略数据驱动强化学习算法

步骤1, 初始化损失的权重 \mathbf{R} 、 \mathbf{Q} 和 γ ，选择 \mathbf{P}^0 、控制增益 \mathbf{K}^0 、干扰增益 \mathbf{L}^0 以及阈值 ε

步骤2, 收集数据：执行策略 \mathbf{u}_E ，收集式(20)和式(21)的数据

步骤3, 策略改进和评估：利用式(23)更新解 $\bar{\mathbf{P}}^i$ 、控制增益 \mathbf{K}^{i+1} 和最坏扰动增益 \mathbf{L}^{i+1}

步骤4, 判定结束：重复步骤3，直到 $|\bar{\mathbf{P}}^{i+1} - \bar{\mathbf{P}}^i| < \varepsilon$

定理2：如果对所有 $L \geq l_o$ 、 $K \geq k_o$ ，存在 $l_o > 0$ 、 $k_o > 0$ ，使得：

$$\begin{cases} \text{rank}([I_{\mathbf{x}\mathbf{x}}, I_{\mathbf{x}\mathbf{u}}, I_{\mathbf{x}\mathbf{d}}]) = n(n+1)/2 + nm + nz \\ \text{rank}(I_{\mathbf{x}\mathbf{x}}) = n(n+1)/2 \end{cases} \quad (24)$$

则式(23)中批量最小二乘法求得的解具有唯一性。

证明：为了证明式(23)的解唯一，相当于要证明：

$$\Psi \Omega = 0 \quad (25)$$

有唯一的平凡解 $\Omega = 0$ 。

下面利用反证法进行证明。假设 $\Omega = [\bar{\mathbf{X}}^T, \bar{\mathbf{Y}}^T, \bar{\mathbf{Z}}^T]^T \in \mathbb{R}^{n(n+1)/2 + nm + nz}$ 是式(25)的非零解，其中，

$\bar{X} \in \mathbb{R}^{(n+1)/2}$ 、 $\bar{Y} \in \mathbb{R}^m$ 、 $\bar{Z} \in \mathbb{R}^n$ 。根据矩阵的拉直运算法则, 存在对称矩阵 X 、 Y 、 Z 与 Ω 唯一对应: $\bar{X} = \text{vec}(X)$, $\bar{Y} = \text{vec}(Y)$, $\bar{Z} = \text{vec}(Z)$ 。对式 (16) 拉直运算得:

$$\Psi\Omega = I_{xx} \text{vec}(G) + 2I_{xu} \text{vec}(M) + 2I_{xd} \text{vec}(N) \quad (26)$$

其中, $M = B^T X - RY$, $N = D^T X - \gamma^2 Z$, $G = A^T X + XA - K_E^T RY - Y^T R K_E$ 。

由于 L 是对称矩阵, 故有 $I_{xx} \text{vec}(G) = I_{\bar{x}} \bar{G}$, $\bar{G} = [G_{11}, 2G_{12}, \dots, G_{22}, 2G_{23}, \dots, G_{nn}]^T$, 代入式 (26), 可得:

$$\begin{bmatrix} I_{\bar{x}}, 2I_{xu}, 2I_{xd} \end{bmatrix} \begin{bmatrix} \bar{G}, \text{vec}(M), \text{vec}(N) \end{bmatrix}^T = 0 \quad (27)$$

根据式 (24) 定理2的条件可知, $\begin{bmatrix} I_{\bar{x}}, 2I_{xu}, 2I_{xd} \end{bmatrix}$ 满秩, 故式 (27) 存在唯一解 $\bar{G} = 0$ 、 $\text{vec}(M) = 0$ 、 $\text{vec}(N) = 0$:

$$\begin{cases} K_E^T RY + Y^T R K_E = A^T X + XA \\ RY = B^T X \\ \gamma^2 Z = D^T X \end{cases} \quad (28)$$

由于 $A - BK_E$ 具有赫尔维茨稳定性, 将式 (28) 进行合并, 可得 $X = 0$ 。由于 $R > 0$ 、 $\gamma^2 > 0$, 故有 $Y = 0$ 、 $Z = 0$, 即 $\Omega = 0$ 。因此在式 (26) 定理2的条件下, 式 (25) 的解具有唯一性。证毕。

本节提出了应用于 PMSM 混沌现象的数据驱动 MFOPRL 方法, 其可以在收集满足条件的数据后, 训练得到最优控制器。该方法不需要电机的具体模型信息, 并借助 H 无穷理论加入对最坏扰动的考虑, 具备对可能存在最坏扰动的抗干扰性。下节将设置实验验证本节提出的 MFOPRL 方法, 并通过对比实验测试其鲁棒性。

本研究采用前馈补偿与强化学习相结合的混合控制策略。首先利用前馈补偿控制器补偿大部分已

知的 PMSM 模型系统动力学, 将系统简化为近似线性系统, 这使得后续施加的强化学习控制器不必学习整个复杂的非线性动力学, 而是可以专注于处理剩余的模型不确定性、参数变化和外部扰动等未建模或未知的部分。相较于直接将标准强化学习算法应用于整个非线性系统, 这种设计通过简化学习任务, 显著提高了强化学习的学习效率和控制精度。因此, 该策略融合了前馈模型知识和数据驱动学习的优势, 并通过改进强化学习算法本身来优化整体控制性能、学习效率和鲁棒性, 未来可以通过引入深度强化学习方法来进一步改进 MFOPRL 在不同场景下的泛化能力^[32]。

4 仿真结果与分析

本节从以下3个方面对本文提出算法的有效性进行验证。

(1) 异策略收集混沌现象的状态变量, 并离线训练稳定控制策略。

(2) 在负载扰动下, MFOPRL 方法对比其他控制方法的抗扰性。

(3) 在模型不确定下, MFOPRL 方法对比线性二次型调节器的性能。

本实验采用 MATLAB/Simulink 进行数值仿真。基于 MFOPRL 的 PMSM 抗扰控制的控制系统如图1所示, 用于训练 MFOPRL 控制器的参数见表1。

4.1 PMSM 中混沌现象的稳定

为验证本文提出算法的有效性, 对 PMSM 混沌现象稳定过程进行仿真, 仿真主要展示 MFOPRL 的训练收敛过程以及该方法对混沌现象的稳定作

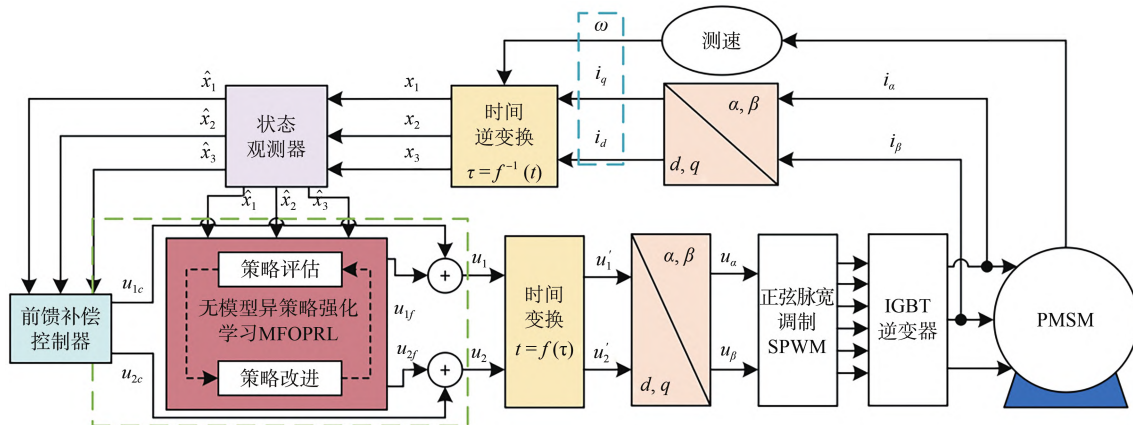
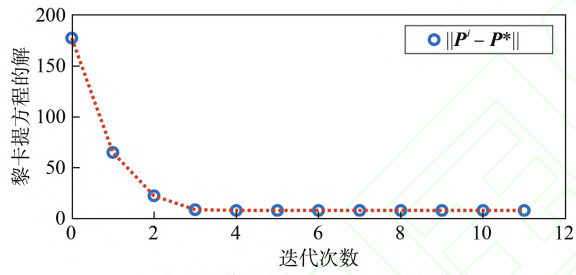


图1 基于MFOPRL的PMSM抗扰控制的控制系统

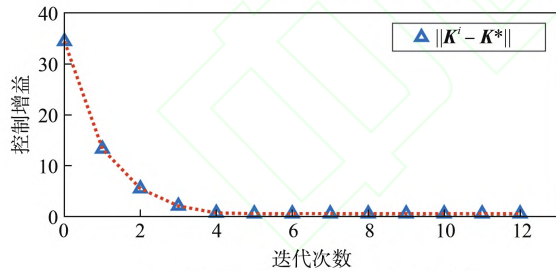
用。MFOPRL 训练过程的迭代收敛曲线如图 2 所示。本文提出的 MFOPRL 方法对混沌现象的抑制仿真结果如图 3~图 5 所示。

表 1 用于训练 MFOPRL 控制器的参数

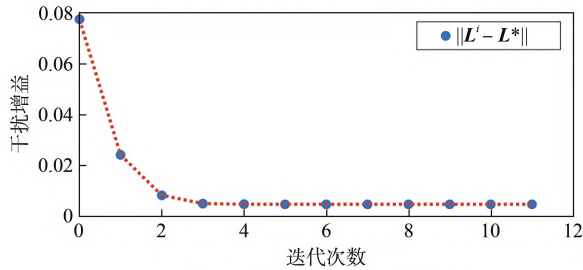
参数	数值
Q	[10,10;10,10]
R	10
T	0.02
l	2 000
α	0.1
γ	40
P^0	[358,231,0;231,150,0;0,0,4]
K^0	[11.3,9.6,1.5;1.2,-1,2,3]
L^0	[0.76,-0.24,1.73]
ε	10^{-10}
x_0	[-1,10,-5]
E	[10,0,0;0,10,0;0,0,10]



(a) 迭代黎卡提方程解 P 的收敛曲线



(b) 最优控制器迭代增益 K 的收敛曲线



(c) 最坏扰动增益 L 的收敛曲线

图 2 MFOPRL 训练过程的迭代收敛曲线

图 2 (a) 为迭代黎卡提方程解 P 的收敛曲线，在第 11 次迭代时，满足 $\|P^i - P^*\| = 4.1513 \times 10^{-9}$ ，

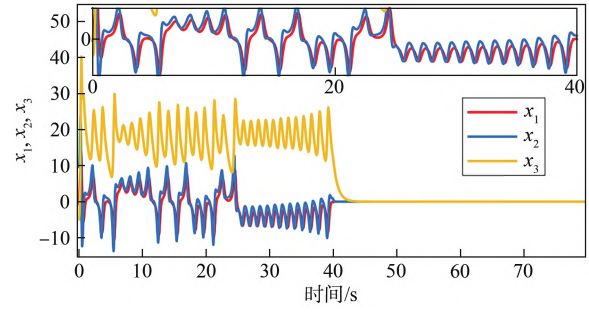


图 3 PMSM 的混沌现象在 MFOPRL 作用下的状态稳定结果

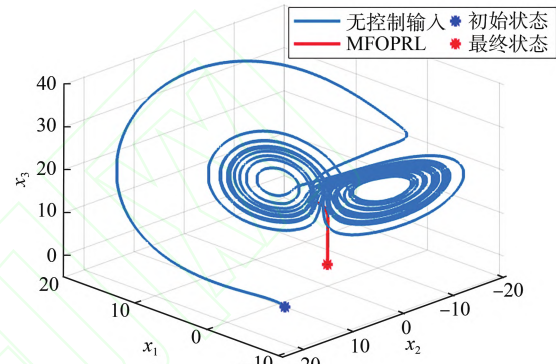
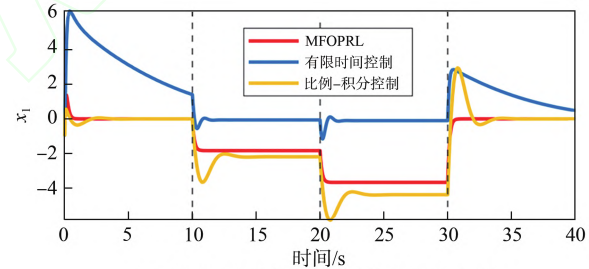
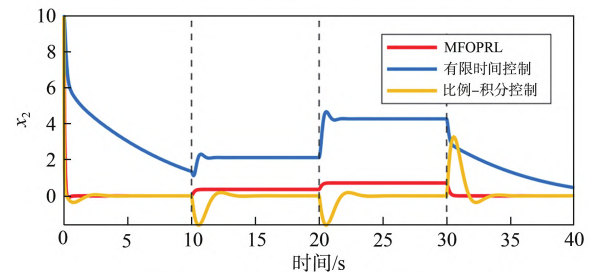


图 4 PMSM 混沌系统的奇异吸引子



(a) 状态 x_1 在时变不确定负载下的变化曲线



(b) 状态 x_2 在时变不确定负载下的变化曲线

图 5 不同控制方法作用下状态变量变化曲线对比

即收敛到黎卡提方程 (9) 的唯一解 P^* 。图 2 (b) 为最优控制器迭代增益 K 的收敛曲线，在第 11 次迭代时， $\|K^i - K^*\| = 5.6220 \times 10^{-10}$ ，即为最优控制器增益。图 2 (c) 为零和博弈下，考虑的最坏扰动增益 L 的收敛曲线， $\|L^i - L^*\| = 5.6220 \times 10^{-10} = 4.0 \times 10^{-3}$ ，此时 L^i 收敛到最坏扰动增益 L^* 。由以上

分析得出, 通过执行异策略, 在探索阶段收集到足够的电机混沌状态运行数据, 在无系统模型的前提下, 采用 MFOPRL 迭代训练得到最优解的 \mathbf{P}^* 、 \mathbf{K}^* 和 \mathbf{L}^* 。

PMSM 的混沌现象在 MFOPRL 作用下的状态稳定结果如图 3 所示。前 40 s 执行随机异策略, 收集的数据满足定理 2 的条件, 状态曲线呈现一定的波动性。第 40 s 开始, 执行训练的 MFOPRL 最优控制器, 状态快速收敛到零。PMSM 混沌系统的奇异吸引子如图 4 所示, 蓝线为随机异策略收集数据的状态曲线, 红线为训练的 MFOPRL 作用下的电机运行状态, 最后收敛到零点。

4.2 对比实验 1: 负载扰动不确定

为模拟 PMSM 在混沌状态运行中可能存在的负载扰动, 本文设置了对比实验, 将 MFOPRL 方法与线性二次型调节器 (linear quadratic regulator, LQR)、有限时间控制器^[12] (finite-time controller, FT) 和比例积分控制器 (portion-integral controller, PI) 等控制方法进行对比。具体的控制器形式如下所示:

$$u_{\text{FT}} = x_2(t) - \rho |x_2|^\lambda \operatorname{sgn}(x_2) - (\gamma x_2 x_1 - \alpha x_1^2 - x_3^2 + \alpha x_2 x_1 + \rho |x_3|^{\lambda+1} + \rho |x_1|^{\lambda+1}) \frac{\operatorname{sgn}(x_2)}{|x_2|} \quad (29)$$

$$u_{\text{PI}} = \mathbf{K}_p \mathbf{x} + \mathbf{K}_I \int_0^t \mathbf{x}(\tau) d\tau \quad (30)$$

其中, FT 控制器参数 $\rho=0.2$ 、 $\lambda=0.7$, PI 控制器参数 $k_p=k_I=30$ 。为测试不同控制器对于未知负载扰动的抗干扰性, 设置如下阶段性变化负载:

$$d_L = \begin{cases} 12, 10\text{s} < t < 20\text{s} \\ 24, 20\text{s} < t < 30\text{s} \end{cases} \quad (31)$$

不同控制方法作用下状态变量变化曲线对比如图 5 所示。图 5 (a) 和图 5 (b) 分别为状态 x_1 和 x_2 在时变不确定负载下的变化曲线。相对 PI 和 FT 两种方法, MFOPRL 方法的收敛速度最快, 超调量最小。对比实验 1 的实验结果见表 2。表 2 列举了对比实验 1 中不同控制方法的累积资源消耗损失值 $\int_{t_0}^{t_f} (\mathbf{x}^T \mathbf{Q} \mathbf{x} + \mathbf{u}^T \mathbf{R} \mathbf{u}) dt$ 以及最大的控制输入量。为了综合评估控制性能, 本文比较了不同方法在实验 1 (含负载扰动) 条件下的累计损失值, 该值反映了整个过程中的跟踪误差和控制消耗。由表 2 可知, 本文提出的 MFOPRL 方法的累计损失值 ($1.002 7 \times$

10^7) 相较于传统 FT 方法 ($1.644 9 \times 10^7$) 降低了约 39.04%。这主要得益于强化学习组件的在线学习和自适应优化能力, 使其能够更有效地补偿未知扰动和模型不确定性, 从而显著降低系统的总体控制成本。MFOPRL 方法在 3 种方法中的累计损失值最小, 而在其状态收敛速度显著快于 FT 方法的优势下, 其最大输入 $|u|$ 几乎与 FT 方法相同, 这说明在负载扰动不确定的条件下, MFOPRL 方法的控制性能显著优于 PI 和 FT 两种方法。

表 2 对比实验 1 的实验结果

对比项	MFOPRL	FT	PI
累计损失值	$1.002 7 \times 10^7$	$1.644 9 \times 10^7$	$1.772 7 \times 10^7$
最大输入 $ u $	126.382 9	122.898 0	335.410 0

4.3 对比实验 2: 模型不确定

由于 PMSM 的运行状况易受外界环境条件变化而变化, 混沌动力学模型中的参数 σ 和 γ 可能因此出现一定的漂移情况。为验证 MFOPRL 方法在无模型信息前提下的控制性能, 本节用正态分布模拟系统模型参数出现的不确定性变化, 设置 $\sigma \sim N(\sigma_0, 3^2)$ 、 $\gamma \sim N(\gamma_0, 10^2)$, 其中 $\sigma_0=5.46$, $\gamma_0=20$, 并设置 LQR 对比实验:

$$\mathbf{u}_{\text{LQR}} = -\mathbf{K}_{\text{LQR}} \mathbf{x} \quad (32)$$

其中, 增益 $\mathbf{K}_{\text{LQR}} = [23.11, 14.95, 0; 0, 0, 0.41]$, 并进行多次重复对比实验。

不同方法的平均状态收敛曲线对比如图 6 所示。图 6 为在模型具有不确定性时测试 MFOPRL 和 LQR 的状态变量区域。图 6 (a) 和图 6 (b) 分别为状态 x_1 和 x_2 在多次重复测试下的平均状态收敛曲线。MFOPRL 方法相较 LQR 方法的平均状态收敛曲线的收敛速度更快, 最大最小状态之间的可能出现状态区域面积更小, MFOPRL 方法相较 LQR 方法在模型具有一定不确定性下的性能更优异, 控制器适应性更强。

MFOPRL 和 LQR 方法状态收敛时间的分布情况对比如图 7 所示。由图 7 可知, MFOPRL 方法的平均收敛时间相较 LQR 方法的更小, MFOPRL 方法拟合得到的正态分布曲线的跨度相较 LQR 方法更小, 拟合得到的方差更小, MFOPRL 方法的收敛时间更加集中。对比实验 2 的实验结果见表 3, 表 3 对比了 LQR 和 MFOPRL 方法的平均收敛时间、收敛时间方差、收敛成功率以及 $|x_1|$ 最大值。由表 3

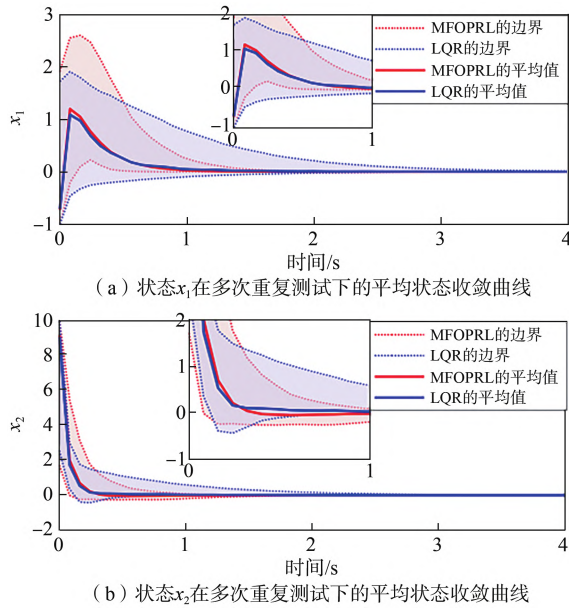


图6 不同方法的平均状态收敛曲线对比

可知，MFOPRL方法的成功率（91.176 5%）相较于传统LQR方法（82.352 9%），性能提升了约10.71%，这进一步证明了本文所提方法在不同评价指标下的有效性。相较于LQR方法，本文提出的MFOPRL方法对PMSM的混沌系统出现不确定性的适应性更强，对环境变化导致的系统参数漂移的鲁棒性也更强。

不同方法状态收敛时间的分布情况对比如图8

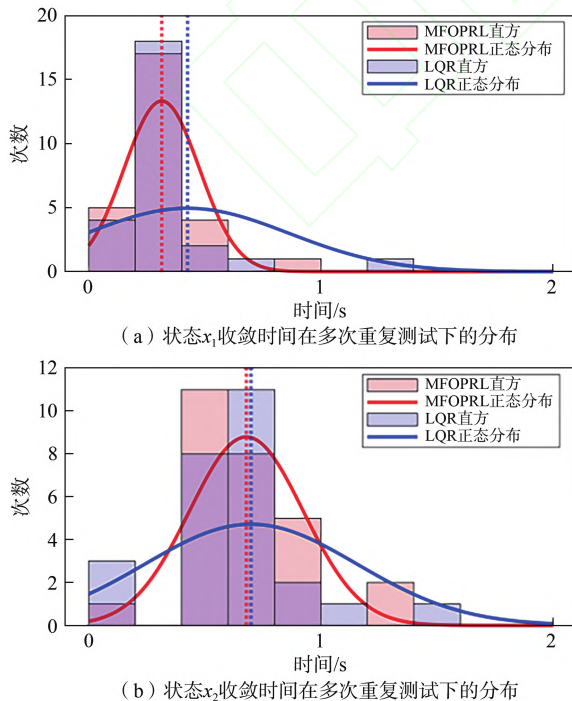


图7 MFOPRL和LQR方法状态收敛时间的分布情况对比

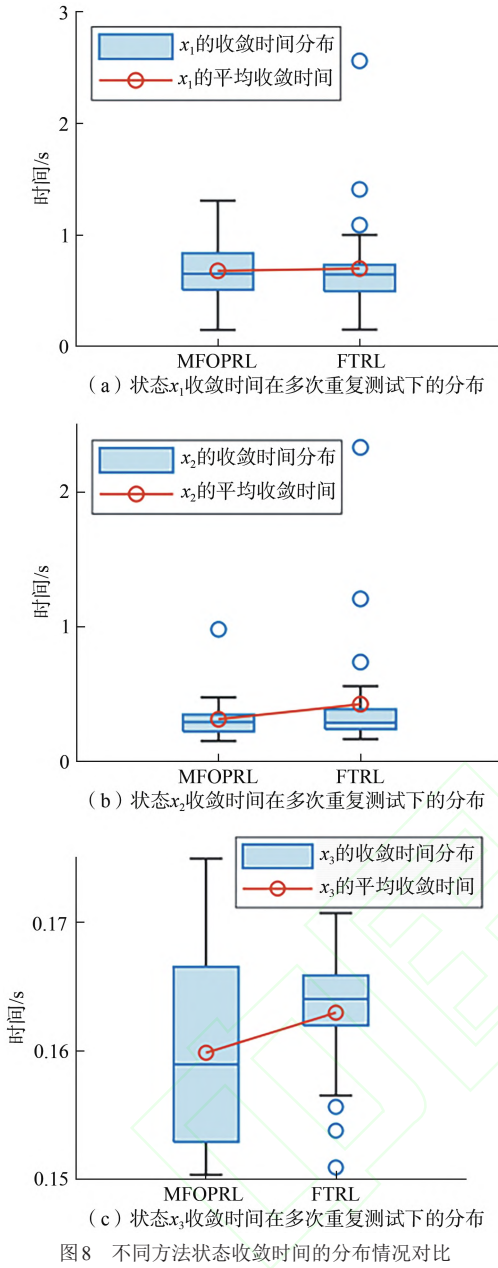
表3 对比实验2的实验结果

对比项	MFOPRL	LQR
x_1	$N(0.599\ 11, 0.152\ 94^2)$	$N(0.689\ 47, 0.322\ 23^2)$
x_2	$N(0.269\ 02, 0.095\ 34^2)$	$N(0.439\ 55, 0.263\ 46^2)$
x_3	$N(0.163\ 41, 0.007\ 33^2)$	$N(0.162\ 63, 0.006\ 27^2)$
成功率	91.176 5%	82.352 9%

所示。图8通过箱形图展示了本文提出的MFOPRL方法与基于模型的有限时间强化学习（FTRL）^[33-34]方法在多次重复实验中状态 x_1 、 x_2 、 x_3 收敛时间的分布对比，其中FTRL的神经网络基元设置为 $[x_1^{1+\alpha}, x_2^{1+\alpha}, x_3^{1+\alpha}]^T$ ，其中 $\alpha=0.8$ ，初始权重为 $1+\text{rand}(3, 1)$ 。从图8中可以清晰地看出，MFOPRL方法在收敛速度和稳定性方面均优于FTRL方法。具体而言，对于状态 x_1 、 x_2 ，MFOPRL方法的中位数收敛时间明显更短，且其四分位数间距（interquartile range, IQR）显著小于FTRL方法，这表明MFOPRL方法不仅收敛更快，而且在多次实验中表现出更高的一致性和鲁棒性，尤其在 x_2 状态上优势尤为突出。虽然两种方法在 x_3 状态上的收敛都非常快且稳定，但MFOPRL方法仍然展现出略微更快的收敛速度和更集中的分布。总体而言，该对比结果有力证明了MFOPRL方法相较于基于模型的FTRL方法在提升系统状态收敛效率和稳定性方面的优越性。

5 结束语

本文首先针对PMSM的混沌动力学进行建模，设计前馈补偿项消去模型中的非线性耦合项，接着针对可能出现的不稳定未知外部负载扰动建立了零和博弈，设计通过迭代黎卡提方程求解的抗干扰最优控制器。在此基础上，针对电机可能受环境变化影响造成的控制模型参数偏移，引入强化学习的思想，设计了一种基于电机历史运行数据训练的无模型异策略强化学习方法MFOPRL，通过数据驱动求解具有鲁棒性的最优控制器。最后通过数值仿真，验证了在外界负载扰动不确定的情况下，MFOPRL控制相较FI和PI的资源利用效率更高，稳定速度更快；在电机模型参数不确定的情况下，MFOPRL方法相较LQR控制器的收敛速度更快，混沌现象的稳定抑制成功率更高。



附录 式(18)和式(19)的推导

本附录提供了文中式(18)和式(19)的详细推导过程,这些表达式利用克罗内克积(\otimes)和向量化算子(vec)将标量表达式转换为便于参数估计的形式。

1 式(18)的推导

式(18)旨在转换项 $(u + K^j x_E)^T RK^{j+1} x_E$ 。令 $e = u + K^j x_E$, 则有:

$$(u + K^j x_E)^T RK^{j+1} x_E = e^T RK^{j+1} x_E$$

本文利用克罗内克积恒等式 $b^T Ca = (a^T \otimes b^T)$

$\text{vec}(C)$ 。令 $b = e$, $C = RK^{j+1}$, $a = x_E$, 则有:

$$e^T RK^{j+1} x_E = (x_E^T \otimes e^T) \text{vec}(RK^{j+1})$$

接下来,利用恒等式 $\text{vec}(ABC) = (C^T \otimes A) \text{vec}(B)$ 来处理 $\text{vec}(RK^{j+1})$ 。假设 $x_E \in \mathbb{R}^2$, $u \in \mathbb{R}^2$, $K^{j+1} \in \mathbb{R}^{2 \times 3}$, $R \in \mathbb{R}^{2 \times 2}$ 。令 $A = R$, $B = K^{j+1}$, $C = I_3$ ($n=3$ 维单位阵), 则有:

$$\begin{aligned} \text{vec}(RK^{j+1}) &= \text{vec}(RK^{j+1} I_n) \\ &= (I_n^T \otimes R) \text{vec}(K^{j+1}) \\ &= (I_n \otimes R) \text{vec}(K^{j+1}) \end{aligned}$$

将此结果代回,得到:

$$e^T RK^{j+1} x_E = (x_E^T \otimes e^T) (I_n \otimes R) \text{vec}(K^{j+1})$$

因此,式(18)推导完成。

2 式(19)的推导与说明

式(19)涉及项 $\gamma^2 d^T L^T L d$ 。利用恒等式 $b^T Ca = (a^T \otimes b^T) \text{vec}(C)$, 令 $b = d$, $C = L^T L$, $a = d$, 得到:

$$d^T (L^T L) d = (d^T \otimes d^T) \text{vec}(L^T L)$$

因此,

$$\gamma^2 d^T L^T L d = \gamma^2 (d^T \otimes d^T) \text{vec}(L^T L)$$

因此,式(19)推导完成。

参考文献:

- [1] 刘凌, 司杰文, 林起联. 支持向量机预测可变参数的永磁同步电机快速终端滑模控制[J]. 西安交通大学学报, 2021, 55(6): 53-60.
LIU L, SI J W, LIN Q L. A fast terminal sliding mode control of permanent magnet synchronous motor with variable parameters predicted by SVM[J]. Journal of Xi'an Jiaotong University, 2021, 55(6): 53-60.
- [2] ZHAO J G, YANG C Y, GAO W N, et al. Reinforcement learning and optimal control of PMSM speed servo system[J]. IEEE Transactions on Industrial Electronics, 2023, 70(8): 8305-8313.
- [3] 刘思源, 刘凌, 靳东松. 永磁同步电机新型有效磁链固定时滑模扰动辨识策略[J]. 西安交通大学学报, 2023, 57(9): 144-151.
LIU S Y, LIU L, JIN D S. A novel active flux fixed-time sliding mode disturbance identification method for permanent magnet synchronous motor[J]. Journal of Xi'an Jiaotong University, 2023, 57(9): 144-151.
- [4] 邵保领, 刘凌, 胡铭遥. 面向永磁同步电机负载转矩观测的自抗扰改进滑模控制[J]. 西安交通大学学报, 2024, 58(3): 204-211.
SHAO B L, LIU L, HU M Y. Improved sliding mode control with disturbance rejection for load torque observation of permanent magnet synchronous motors[J]. Journal of Xi'an Jiaotong University, 2024, 58(3): 204-211.
- [5] 李政, 胡广大, 崔家瑞, 等. 永磁同步电机调速系统的积分型滑模变结构控制[J]. 中国电机工程学报, 2014, 34(3): 431-437.
LI Z, HU G D, CUI J R, et al. Sliding-mode variable structure control with integral action for permanent magnet synchronous motor[J]. Proceedings of the CSEE, 2014, 34(3): 431-437.

- [6] 毛北行, 王东晓. 分数阶永磁同步电机混沌系统自适应滑模同步[J]. 浙江大学学报(理学版), 2023, 50(5): 564-568.
MAO B X, WANG D X. Self-adaptive sliding mode synchronization of fractional-order permanent magnet synchronization motor chaotic systems[J]. Journal of Zhejiang University (Science Edition), 2023, 50(5): 564-568.
- [7] 王江彬, 刘崇新. 4 阶混沌电力系统的协同控制方法[J]. 西安交通大学学报, 2020, 54(1): 26-31.
WANG J B, LIU C X. Synergetic control method for four-dimensional chaotic power system[J]. Journal of Xi'an Jiaotong University, 2020, 54(1): 26-31.
- [8] 李宁洲, 邱思旋, 卫晓娟, 等. 一类永磁同步电机混沌运动的 GWO-RBFNN 双参数协同智能优化控制[J]. 哈尔滨理工大学学报, 2024, 29(3): 28-36.
LI N Z, QIU S X, WEI X J, et al. GWO-RBFNN dual-parameter collaborative intelligent optimal control of chaotic motion of a class of permanent magnet synchronous motor[J]. Journal of Harbin University of Science and Technology, 2024, 29(3): 28-36.
- [9] REN H P, LIU D. Nonlinear feedback control of chaos in permanent magnet synchronous motor[J]. IEEE Transactions on Circuits and Systems II: Express Briefs, 2006, 53(1): 45-50.
- [10] WEI Q, WANG X Y, HU X P. Optimal control for permanent magnet synchronous motor[J]. Journal of Vibration and Control, 2014, 20(8): 1176-1184.
- [11] DO T D. Optimal control design for chaos suppression of PM synchronous motors[C]//Proceedings of the 2016 2nd International Conference on Control Science and Systems Engineering (ICCSSE). Piscataway: IEEE Press, 2016: 88-92.
- [12] HOU Y Y. Finite-time chaos suppression of permanent magnet synchronous motor systems[J]. Entropy, 2014, 16(4): 2234-2243.
- [13] YAN T F, LIU Q Y, DOU B W, et al. An adaptive dynamic programming method for torque ripple minimization of PMSM[J]. Journal of Industrial & Management Optimization, 2021, 17(2): 827-839.
- [14] 陈强, 南余荣, 邢科新. 基于扩张状态观测器的永磁同步电机混沌系统自适应滑模控制[J]. 物理学报, 2014, 63(22): 117-124.
CHEN Q, NAN Y R, XING K X. Adaptive sliding-mode control of chaotic permanent magnet synchronous motor system based on extended state observer[J]. Acta Physica Sinica, 2014, 63(22): 117-124.
- [15] ZHANG J, SUN J H, GU H, et al. Control of PMSM chaos using backstepping-based adaptive fuzzy method in the presence of uncertainty and disturbance[J]. Systems Science & Control Engineering, 2021, 9(1): 673-688.
- [16] 谢东燊, 杨俊华, 熊锋俊, 等. 永磁直线同步电机解耦自适应滑模混沌控制[J]. 计算机仿真, 2019, 36(5): 263-268.
XIE D S, YANG J H, XIONG F J, et al. Decoupling adaptive sliding mode chaos control of permanent magnet linear synchronous motor[J]. Computer Simulation, 2019, 36(5): 263-268.
- [17] 毛北行. 分数阶整数阶多混沌系统的自适应滑模同步[J]. 中山大学学报(自然科学版), 2020, 59(4): 128-133.
MAO B X. Self-adaptive sliding mode synchronization of fractional-order and integer-order uncertain multi-chaotic systems[J]. Acta Scientiarum Naturalium Universitatis Sunyatseni, 2020, 59(4): 128-133.
- [18] ZHANG J X, WANG S L, ZHOU P, et al. Novel prescribed performance-tangent barrier Lyapunov function for neural adaptive control of the chaotic PMSM system by backstepping[J]. International Journal of Electrical Power & Energy Systems, 2020, 121: 105991.
- [19] YIN X, SHE J H, LIU Z T, et al. Chaos suppression in speed control for permanent-magnet-synchronous-motor drive system[J]. Journal of the Franklin Institute, 2020, 357(18): 13283-13303.
- [20] SAHOO A, XU H, JAGANNATHAN S. Near optimal event-triggered control of nonlinear discrete-time systems using neurodynamic programming[J]. IEEE Transactions on Neural Networks and Learning Systems, 2016, 27(9): 1801-1815.
- [21] 廖泽华, 梁子钰, 周天民, 等. 基于平行控制的离散非线性系统的事件触发近似最优控制[J]. 智能科学与技术学报, 2021, 3(4): 435-443.
LIAO Z H, LIANG Z Y, ZHOU T M, et al. Parallel control-based event-driven approximate optimal control of discrete-time nonlinear systems[J]. Chinese Journal of Intelligent Science and Technology, 2021, 3(4): 435-443.
- [22] JIANG Y, JIANG Z P. Computational adaptive optimal control for continuous-time linear systems with completely unknown dynamics[J]. Automatica, 2012, 48(10): 2699-2704.
- [23] LU J W, WEI Q L, ZHOU T M, et al. Event-triggered near-optimal control for unknown discrete-time nonlinear systems using parallel control[J]. IEEE Transactions on Cybernetics, 2023, 53(3): 1890-1904.
- [24] LU J W, WEI Q L, WANG F Y. Parallel control for nonzero-sum games with completely unknown nonlinear dynamics via reinforcement learning[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2025, 55(4): 2884-2896.
- [25] PING Z W, JIA Y J, XIONG B G, et al. Optimal output regulation for PMSM speed servo system using approximate dynamic programming[J]. Science China Information Sciences, 2023, 66(7): 170206.
- [26] EL-SOUSY F F M, AMIN M M, AL-DURRA A. Adaptive optimal tracking control via actor-critic-identifier based adaptive dynamic programming for permanent-magnet synchronous motor drive system[J]. IEEE Transactions on Industry Applications, 2021, 57(6): 6577-6591.
- [27] FAN Z X, LI S H, LIU R J. ADP-based optimal control for systems with mismatched disturbances: a PMSM application[J]. IEEE Transactions on Circuits and Systems II: Express Briefs, 2023, 70(6): 2057-2061.
- [28] SHI T Y, ZOU W C, GUO J, et al. Adaptive speed regulation for permanent magnet synchronous motor systems with speed and current constraints[J]. IEEE Transactions on Circuits and Systems II: Express Briefs, 2024, 71(4): 2079-2083.
- [29] EQRA N, VATANKHAH R, EGHTEHAD M. A novel adaptive multi-critic based separated-states neuro-fuzzy controller: architecture and application to chaos control[J]. ISA Transactions, 2021, 111: 57-70.
- [30] 夏丽娜, 李擎, 宋睿卓, 等. 未知异构多智能体系统无模型自适应动态规划同步控制[J]. 智能科学与技术学报, 2021, 3(4): 444-448.
XIA L N, LI Q, SONG R Z, et al. Synchronization control of unknown heterogeneous multi-agent system via model-free adaptive dynamic programming[J]. Chinese Journal of Intelligent Science and Technology, 2021, 3(4): 444-448.
- [31] 李金娜, 程薇燃. 基于强化学习的数据驱动多智能体系统最优一致性综述[J]. 智能科学与技术学报, 2020, 2(4): 327-340.
LI J N, CHENG W R. An overview of optimal consensus for data driven multi-agent system based on reinforcement learning[J]. Chinese Journal of Intelligent Science and Technology, 2020, 2(4): 327-340.
- [32] 李涛, 魏庆来. 基于深度强化学习的智能暖气温度控制系统[J]. 智能科学与技术学报, 2020, 2(4): 348-353.
LI T, WEI Q L. Intelligent heating temperature control system based on deep reinforcement learning[J]. Chinese Journal of Intelligent Science and Technology, 2020, 2(4): 348-353.
- [33] ZHANG L J, CHEN Y. Finite-time adaptive dynamic programming for

affine-form nonlinear systems[J]. IEEE Transactions on Neural Networks and Learning Systems, 2025, 36(2): 3573-3586.

- [34] TAN J, XUE S, GUAN Q, et al. Finite-time safe reinforcement learning control of multi-player nonzero-sum game for quadcopter systems[J]. Information Sciences, 2025, 712: 122117.

[作者简介]



谭浚楷 (2001-), 男, 西安交通大学电气工程学院硕士生, 主要研究方向为自适应动态规划与强化学习、逆强化学习等。



薛霜思 (1992-), 男, 博士, 西安交通大学电气工程学院助理教授, 主要研究方向为数据驱动网络控制、智能系统控制。



郭子航 (2001-), 男, 西安交通大学电气工程学院硕士生, 主要研究方向为自适应神经网络控制、多智能体控制等。



曹晖 (1978-), 男, 博士, 西安交通大学电气工程学院教授, 主要研究方向为数据驱动系统、智能系统控制。