Research on PMSM Current Decoupling Control Based on Deep Reinforcement Learning TD3-PI *

WANG Yilin, LIU Huibo*
(School of Automation and Electrical Engineering, Inner Mongolia University of Science and Technology, Baotou Inner Mongolia 014010, China)

Abstract: In response to the issue of cross-coupling of *dq*-axis currents in permanent magnet synchronous motor vector control systems, a novel current decoupling control strategy that integrates the twin delayed deep deterministic policy gradient (TD3) algorithm and proportional-integral control is proposed. Initially, an analysis is conducted on the mathematical model and voltage equations of the PMSM. Subsequently, the state-action space, reward-punishment function, and network structure of the TD3 agent are designed, enabling the agent to effectively recognize and learn the complex non-linear characteristics of the current loop using system data. The reward-punishment function, taking into account current tracking error, coupling effect, and the stability of the agent's output, allows the trained agent to provide real-time coupling compensation voltage based on the current state, thus optimizing the current decoupling performance. Finally, simulation experiments validate the effectiveness of the TD3-PI control strategy under complex conditions such as high-speed operation, sudden load changes, and inductance parameter mismatch. The strategy significantly enhances the dynamic tracking performance and control precision of *dq*-axis currents, demonstrating strong robustness.

Key words: deep reinforcement learning; twin delayed deep deterministic policy gradient; permanent magnet synchronous motor; current control; cross coupling
EEACC: 7220; 8310D **doi**: 10.3969/j.issn.1005-9490.2026.04.018

基于深度强化学习 TD3-PI 的 PMSM 电流解耦控制研究 *

王艺霖, 刘慧博*
(内蒙古科技大学自动化与电气工程学院, 内蒙古 包头 014010)

摘 要: 针对永磁同步电机矢量控制系统中 *dq* 轴电流交叉耦合的问题, 提出一种融合双延迟深度确定性策略梯度(TD3)算法与 PI 控制的新型电流解耦控制策略。首先, 对永磁同步电机数学模型与电压方程进行分析。其次, 设计了 TD3 智能体的状态动作空间、奖惩函数及网络结构, 智能体能够利用系统数据有效识别、学习电流环复杂非线性特性。奖惩函数综合考虑了电流跟踪误差、耦合效应及智能体输出的稳定性, 使训练后的智能体能够根据电流状态实时提供耦合补偿电压, 优化电流解耦性能。最后, 仿真模拟试验验证了 TD3-PI 控制策略在高转速、负载突变及电感参数失配等复杂工况下的有效性, 该策略显著提升了 *dq* 轴电流的动态跟踪性能和控制精度, 具有较强鲁棒性。

关键词: 深度强化学习; 双延迟确定性策略梯度; 永磁同步电机; 电流控制; 交叉耦合
中图分类号: TM351; TP391 **文献标识码**: A **文章编号**: 1005-9490(2026)04-0829-08

永磁同步电机(Permanent Magnet Synchronous Machine, PMSM)具有体积小、功率密度高、控制性能好等优点, 在工业机器人、航空航天以及电动汽车等领域得到广泛应用[1]。PMSM 多采用矢量控制技术, 基于矢量变换和旋转坐标系将定子电流分解为 *d* 轴励磁分量和 *q* 轴转矩分量, 实现了电机励磁和转矩的独立控制[2]。然而, *d*、*q* 轴电压方程中存在交叉耦合项, 在电机高速运行、负载突变或电感参数失配等工况下, 耦合项会引入不确定性较强扰动, 影响电机性能和控制精度。

为了降低耦合项的影响, 学界提出了多种解耦控制策略。电流反馈解耦控制[3]利用反馈电流和转速直接计算并补偿耦合电压, 该策略易于实现, 但极度依赖精确的电机模型。双 PI 解耦控制[4]引入

项目来源: 国家自然科学基金项目(62063027)
收稿日期: 2024-02-28 **修改日期**: 2024-04-22

附加交叉耦合 PI 控制器以实现解耦,但参数整定与系统稳定性之间难以平衡。内模解耦控制[5]虽然提高了解耦效果,但其性能仍旧依赖于精确的电机模型参数,并且在调节过程中存在欠阻尼振荡。复矢量解耦控制[6]在动态解耦方面表现较好,但对参数变化敏感。干扰观测器解耦[7]相较对角化解耦具有更好的动态性能,但其效能受滤波器带宽限制。扩张状态观测器解耦[8]通过观测系统状态并估计总扰动实现动态补偿,减少了对模型参数的依赖,但非线性函数的引入增加了观测器参数的整定难度。这些解耦控制策略在一定程度上缓解了耦合效应,但各自均存在局限性。

深度强化学习(Deep Reinforcement Learning, DRL)结合了深度学习的数据解析能力和强化学习的决策优化能力,能够在不依赖精确系统模型的情况下自适应地解决复杂非线性问题,适合 PMSM 控制系统这类动态复杂场景的优化与控制[9]。文献[10]验证了基于 Actor-Critic 架构的 DRL 算法在电机电流控制中的可行性。文献[11]采用深度确定性策略梯度(Deep Deterministic Policy Gradient, DDPG)[12]设计级联速度电流控制器,增强了 PMSM 系统对参数不确定性和负载扰动的鲁棒性。文献[13]提出基于 DDPG 算法的速度控制策略,实现了控制器参数自整定与转矩电流的自适应补偿,克服了转动惯量变化和转矩扰动的影响,尽管 DDPG 算法在训练时存在超参数敏感和价值估计过高的问题。

虽然 DRL 算法已在 PMSM 伺服控制系统领域应用中取得了较好的研究成果,但对电流解耦控制方面的研究仍相对有限。因此,本文提出一种结合双延迟深度确定性策略梯度(Twin Delayed Deep Deterministic Policy Gradient, TD3)[14]算法和 PI 控制的新型电流解耦控制策略 TD3-PI,旨在抑制电流耦合现象并提升电流控制性能。TD3-PI 的架构不仅促进了智能体对耦合现象的深层理解,还降低了算法训练的复杂度。通过设计激励电流控制性能提升的奖惩函数,使 TD3-PI 控制器能够在复杂工况下有效补偿 d、q 轴电压,优化电流动态解耦过程,提高跟踪精度。仿真试验的结果验证了该策略的有效性。

1 永磁同步电机数学模型

本文以 PMSM 为研究对象,基于理想假设:考虑绕组对称性、温度恒定,忽略铁芯饱和及涡流和磁滞损耗的影响。在同步旋转 $d-q$ 坐标系下,PMSM 电压方程可表示为:

$$\begin{cases} u_d = R_s i_d + L_d \dfrac{\mathrm{d}i_d}{\mathrm{d}t} - \omega_e L_q i_q \\ u_q = R_s i_q + L_q \dfrac{\mathrm{d}i_q}{\mathrm{d}t} + \omega_e L_d i_d + \omega_e \psi_f \end{cases} \tag{1}$$

式中:u_d、u_q 为 d、q 轴定子电压;i_d、i_q 为 d、q 轴定子电流;L_d、L_q 为 d、q 轴电感;R_s 为定子电阻;ω_e 为转子角速度;ψ_f 为永磁体磁链。

对式(1)进行 Laplace 变换可得:

$$\begin{bmatrix} u_d \\ u_q \end{bmatrix} = \begin{bmatrix} sL_d + R_s & -\omega_e L_q \\ \omega_e L_d & sL_q + R_s \end{bmatrix} \begin{bmatrix} i_d \\ i_q \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \omega_e \psi_f \tag{2}$$

根据式(2)绘制 PMSM 电流环控制框图(图 1)。可以看出,$d-q$ 坐标变换虽然简化了 PMSM 的控制过程,但仍然存在系统固有的强耦合特性,耦合量的大小受电机转速,d、q 轴电流及电感参数的影响。特别是在电机高速运行或发生负载突变情况下,动态耦合尤为明显,这对电流控制性能产生了不利影响。

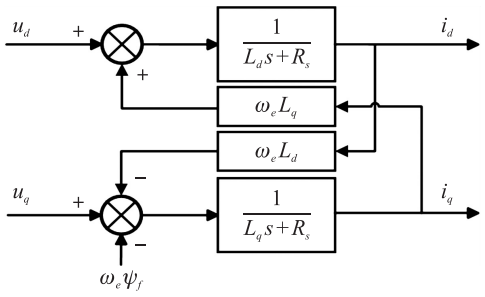


图 1 PMSM 电流环框图

2 深度强化学习 TD3 算法

2.1 强化学习理论基础

强化学习是一种机器学习方法,主要基于马尔可夫决策过程来描述智能体与环境之间的交互[15]。在每个决策时间点 t,智能体根据当前环境状态 s_t 和策略 $\pi(a_t|s_t)$ 采取动作 a_t,随即接收环境反馈的奖励 r_t,并转移到新状态 s_{t+1}。这一交互过程如图 2 所示。

强化学习的核心目标是通过计算回报函数和价值函数来最大化累积奖励的期望。定义回报函数 R_t

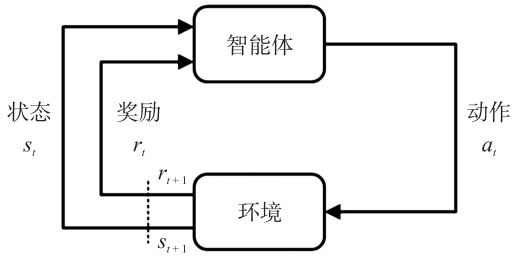


图 2 智能体与环境交互示意图

为从当前状态 s_t 开始的未来奖励总和,可表示为:

$$R_t=\sum_{k=0}^{\infty}\gamma^k r_{t+k} \tag{3}$$

式中:$\gamma\in[0,1]$ 为折扣因子,权衡了即时奖励与未来奖励的重要性。定义价值函数 Q^π 是在状态 s_t 采取策略 π 并执行动作 a_t 可获得的累积奖励期望,即:

$$Q^\pi(s_t,a_t)=E_\pi[R_t|s_t,a_t] \tag{4}$$

将其写成 Bellman 期望方程的形式:

$$\begin{aligned}Q^\pi(s_t,a_t)&=E_\pi[R_t|s_t,a_t]=\\&E_\pi[r_t+\gamma r_{t+1}+\gamma^2 r_{t+2}+\cdots|s_t,a_t]=\\&E_\pi[r_t|s_t,a_t]+\gamma E_\pi[r_{t+1}+\gamma r_{t+2}+\cdots|s_t,a_t]=\\&r_t(s_t,a_t)+\gamma E_\pi[R_{t+1}|s_t,a_t]=\\&E_\pi[r_t+\gamma Q^\pi(s_{t+1},a_{t+1})|s_t,a_t]\end{aligned} \tag{5}$$

Bellman 期望方程揭示了当前状态与未来状态之间的关联,为 Q-learning 算法利用时间差分(Time Difference,TD)方法迭代更新动作价值函数,逼近最优策略 π^* 奠定了基础。这一过程可表示为:

$$\begin{cases}\Delta Q(s_t,a_t)=\alpha[r_t+\gamma\max\limits_a Q(s_{t+1},a)-Q(s_t,a_t)]\\ \pi^*=\arg\max\limits_a Q^\pi(s_t,a)\end{cases} \tag{6}$$

式中:α 为学习率;$r_t+\gamma\max\limits_a Q(s_{t+1},a)$ 为 TD 目标。

随着强化学习问题状态空间维度的增加,Q 学习算法遇到了挑战。深度 Q 学习网络[16]利用深度神经网络来拟合价值函数,解决了高维状态空间的问题,但对连续动作空间的问题仍难以处理。

2.2　深度确定性策略梯度

DDPG 算法具有 Actor-Critic 架构,利用深度神经网络近似表示策略与价值函数。其中,Actor 策略网络 $\mu(s;\theta)$ 负责直接生成动作,而 Critic 价值网络 $q(s,a;w)$ 则评估这些动作在给定状态下的价值。

为了增强算法的稳定性及训练效率,DDPG 引入了经验回放和目标网络技术。经验回放机制是将智能体与环境的交互数据以元组(s_j,a_j,r_j,s_{j+1})形式存储于回放缓冲区(replay buffer),并从中随机抽取小批量样本(mini-batch)用于训练,降低了样本间的相关性并加快了学习进度。目标网络采用与当前网络结构相同但参数更新较慢的网络来计算 TD 目标,平滑目标值的变化,以降低价值估计误差,并增强学习过程的稳定性。

在 DDPG 训练过程中,价值网络 $q(s,a;w)$ 的更新依赖于目标策略网络 $\mu(s;\theta^-)$ 和目标价值网络 $q(s,a;w^-)$ 计算得到的 TD 目标 y_j。通过梯度下降法优化损失函数 $L(w)$,以实现网络参数 w 的更新。具体可表示为:

$$y_j=r_j+\gamma\cdot q(s_{j+1},a_{j+1};w^-)|a_{j+1}=\mu(s_{j+1};\theta^-) \tag{7}$$

$$\begin{cases}L(w)=M^{-1}\sum_j[q(s_j,a_j;w)-y_j]^2\\ \nabla_w L(w)=M^{-1}\{[q(s_j,a_j;w)-y_j]\cdot\\ \quad\nabla_w q(s_j,a_j;w)]\}\end{cases} \tag{8}$$

式中:M 为 mini-batch 的个数。策略网络 $\mu(s;\theta)$ 的更新则依靠价值网络提供的信息来不断改进参数 θ 优化策略,可表示为:

$$\nabla_\theta J(\theta)\approx M^{-1}\sum_j\nabla_a q(s,a;w)|_{s=s_j,a=\mu(s_j)}\cdot\nabla_\theta\mu(s;\theta)|_{s_j} \tag{9}$$

目标网络的参数采用软更新方式,可表示为:

$$\begin{cases}w^-\leftarrow\tau w+(1-\tau)w^-\\ \theta^-\leftarrow\tau\theta+(1-\tau)\theta^-\end{cases} \tag{10}$$

式中:τ 为平滑因子。

然而,DDPG 算法在实际应用中仍存在价值估计过高和误差累积传播的问题,且对超参数敏感。

2.3　双延迟深度确定性策略梯度

TD3 算法是 DDPG 的改进版本,通过引入截断双 Q 学习、目标策略平滑处理和延迟策略更新三大关键技术,有效缓解了价值估计过高的问题,并提升了算法性能。TD3 算法的流程结构如图 3 所示。

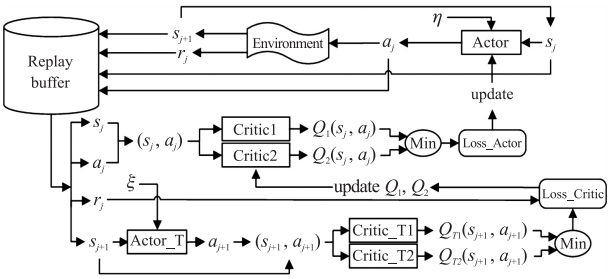


图 3　TD3 算法流程结构图

截断双 Q 学习是 TD3 算法创新性地采用了两个独立的 Critic 价值网络来评估动作价值,并取两者估计中较小的值来计算 TD 目标 y_j,用于价值网络的更新,以此来削弱价值网络的高估问题。TD3 算法的 TD 目标 y_j 可表示为:

$$y_j=r_j+\gamma\cdot\min_{i=1,2}q_i(s_{j+1},a_{j+1};w_i^-)|a_{j+1}=\mu(s_{j+1};\theta^-)+\xi \tag{11}$$

目标策略平滑处理是通过在 Actor 目标策略网络输出的动作上添加截断正态分布的高斯噪声 ξ,正则化策略网络的输出,使价值估计更平滑,进而增强了策略网络训练的稳定性和策略的鲁棒性。

延迟更新策略是通过降低策略网络的更新频率,以确保在 Critic 价值网络稳定后更有效地指导 Actor 策略网络更新,提升算法的收敛速度和整体性能。

TD3 算法的网络更新机制依然采用 DDPG 的梯

度方法,这里不再赘述。

总的来说,TD3 算法能够高效处理 PMSM 控制系统中如转速、电流、电压等复杂连续状态动作空间,并基于交互数据自适应地优化控制策略。此外,其目标策略平滑机制能够应对如电流控制过程中的突变和非线性特性,有助于实现更精准、稳定的电流解耦控制效果。

3 TD3-PI 电流控制器设计

3.1 传统 PI 电流控制器设计

在 PMSM 矢量 PI 控制系统中,通常采用两个闭环控制回路:速度外环和电流内环。速度外环利用 PI 控制器根据转速误差调节输出转矩,以实现转速的精确控制。电流内环则通过两个独立的 PI 控制器针对 d 轴和 q 轴的电流误差,输出驱动 PMSM 的电压矢量 u_d^* 和 u_q^*,从而确保电流控制的准确性。

传统 PI 电流控制器的设计为:

$$u_{d,q}^* = K_p(i_{d,q}^* - i_{d,q}) + K_i \int (i_{d,q}^* - i_{d,q})\,\mathrm{d}t \quad (12)$$

式中:$i_{d,q}^*$和 $i_{d,q}$分别为 d 轴和 q 轴的电流给定信号及电流反馈值;K_p 和 K_i 分别为比例和积分增益。

传统 PI 电流控制器结构简单、易于实现,但在其设计过程中忽视了电机耦合项对控制性能的影响,在处理电流耦合及应对电机非线性动态特性方面存在局限。

3.2 TD3 智能体电压补偿器设计

在 TD3-PI 电流解耦控制策略中,TD3 智能体将提供 d、q 轴动态补偿电压 V_d、V_q,以抑制 d、q 轴间的耦合影响,并适应各种运行条件下的动态和非线性特征。通过整合 TD3 智能体产生的补偿电压与 PI 电流调节器的输出,形成了新的电压矢量信号 u_{dn}^*和 u_{qn}^*。这种架构既利用了深度强化学习的优势,又保留了 PI 控制方法的稳定性与可靠性,同时促进了 TD3 智能体的快速学习。具有新型 TD3-PI 架构电流控制器的 PMSM 矢量控制系统结构如图 4 所示。

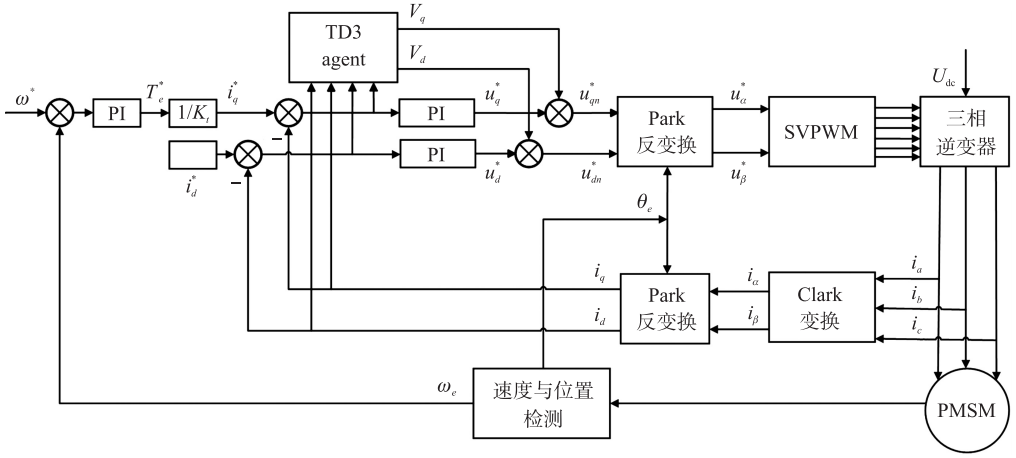


图 4 新型 TD3-PI 矢量控制系统结构图

3.2.1 观测状态及动作选取

观测状态的合理选取对智能体准确评估当前环境状态和系统性能至关重要。本文选取 d、q 轴反馈电流及电流偏差信号作为观测状态,即 $S=\{e_{id}, e_{iq}, i_d, i_q\}$,以全面捕捉电流的实时跟踪性能。智能体的决策动作由二维连续补偿电压矢量 V_d、V_q 构成,即 $A=\{V_d, V_q\}$。

3.2.2 奖惩函数设计

奖惩函数是引导 TD3 智能体学习的关键,其设计应简洁明了,便于智能体能够快速理解并实现有效学习。在针对 d、q 轴电流跟踪目标的背景下,奖惩函数综合考虑了电流跟踪误差、耦合效应和动作输出的稳定性三个因素,具体设计为:

$$R = -(k_1 \cdot e_{iq}^2 + k_2 \cdot e_{id}^2 + \lambda \cdot |e_{id} \cdot e_{iq}| + \chi \cdot \sum_j u_{j,t-1}^2) \quad (13)$$

式中:e_{id}、e_{iq}分别为 d 轴和 q 轴电流跟踪偏差;$|e_{id} \cdot e_{iq}|$项反映了 d 轴和 q 轴电流之间的相互作用,该项增大表明耦合效应加剧;$u_{j,t-1}$为前一个时间步长的动作输出;k_1、k_2、λ 和 χ 为权重因子。

该奖惩函数激励了 TD3 智能体着重于减少电流误差、减轻耦合效应、保持动作输出的平滑性,并在此基础上进一步提升整体控制性能。在训练时,将 k_1 和 k_2 设为 1,λ 设为 0.2,χ 设为 0.000 1,以实现均衡的奖惩效果。

3.3 TD3 智能体的训练过程

在 MATLAB/Simulink R2021b 环境中,搭建了 PMSM 双闭环矢量控制系统模型,作为 TD3 智能体训练的基础平台。为兼顾速度控制的约束和外部干扰的影响,保留了速度控制外环,以确保控制策略的

整体稳定性与有效性。系统参数配置如下:速度 PI 控制器 $K_p = 0.147$, $K_i = 0.003\ 8$,抗饱和系数 $K_c = 0.003\ 8$,输出限幅为±3 T_e;电流 PI 控制器 $K_p = 30$, $K_i = 8\ 800$,抗饱和系数 $K_c = R_s/L_{d,q}$,输出限幅为 $\pm U_{dc}/\sqrt{3}$;PMSM 和逆变器参数见表 1。

表 1　PMSM 和逆变器参数

参数	值
额定功率 P_e/W	750
定子电阻 R_s/Ω	1.4
定子电感 L_d、L_q/mH	4.81
极对数 n_p	5
额定转矩 T_e/(N·m)	2.39
额定转速 ω_N/(r/min)	3 000
转矩系数 K_t	0.4448
转动惯量 J/(kg·m^2)	0.000 78
转子磁链 ψ_f/Wb	0.059 3
母线电压 U_{dc}/V	311
开关频率 f_{sw}/Hz	10 000

利用 Reinforcement Learning 工具箱搭建了图 5 所示的 TD3 智能体结构框架。通过编写 M 脚本文件,调用 Deep Learning 工具箱,建立了网络通信接口和 Actor-Critic 网络结构,并设定了超参数。其中,神经网络采用全连接层,训练超参数详见表 2。在训练过程中,初始化不同的给定转速、转矩电流和电感值,以模拟多种工况条件,全面训练 TD3 智能体。

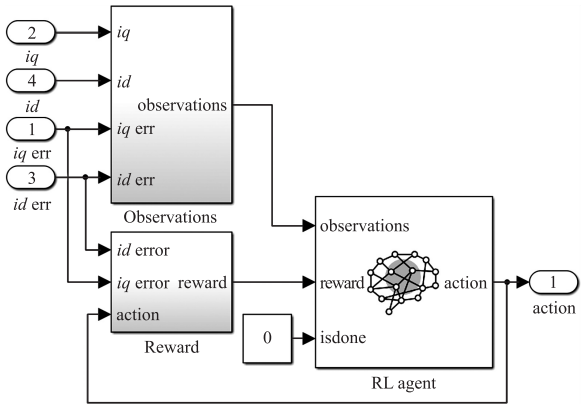


图 5　TD3 智能体搭建图

图 6 所示为 TD3 智能体训练过程的累积奖励变化曲线,反映了智能体的训练效果及性能。训练初期,由于采取随机策略和策略更新的延迟,智能体表现较差,遭受较大惩罚。随着训练的推进,智能体通过不断探索和优化策略,显著提高了补偿电压的输出效率、电流跟踪的准确性及解耦能力,使奖励值逐渐增高。在训练约至第 70 回合时,智能体的奖励值稳定在约−117,标志着其控制策略已趋于收敛,能够有效降低电流跟踪偏差和抑制耦合效应,实现了与 PI 电流调节器的高效协同控制。当训练达到预设的 100 回合后,将 TD3 智能体的最终参数进行保存,从而作为一个精确的离线模型,其决策完全基于策略网络,不再进行网络参数的更新,确保了在线控制过程的高度稳定性和可靠性。

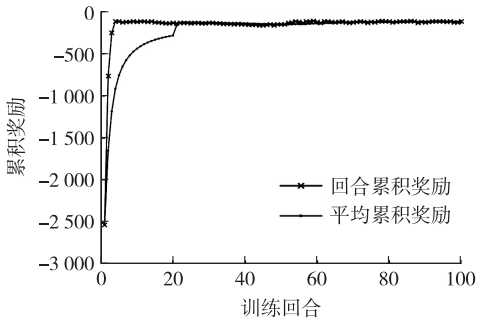


图 6　TD3 智能体训练过程累积奖励曲线

表 2　Actor-Critic 网络的训练超参数

超参数	值
Actor 隐层神经元	100/50
Critic 隐层神经元	200/150
Actor/Critic 学习率	0.0001/0.0001
Actor 激活函数	ReLU | tanh
Critic 激活函数	ReLU |/
折扣系数 γ	0.99
采样时间	0.000 1
经验缓冲区大小	250 000
小批量梯度样本数 M	256
目标平滑因子 τ	0.005
最大训练回合	100
单回合最大步长	2 000
噪声 η 方差/衰减率	0.05/0.002
噪声 ξ 方差/衰减率	0.1/0.001

4　系统仿真及分析

为验证所提出的 TD3-PI 控制策略的有效性和解耦性能,在 MATLAB/Simulink 环境下进行仿真实验,并与传统 PI 控制和电流反馈解耦控制进行对比分析。系统参数配置与 3.3 节中一致,仿真采用离散化处理,以模拟实际的数字控制系统。

仿真实验中,PMSM 采用 $i_d^* = 0$ 控制方式,通过改变转矩电流 i_q 来观察励磁电流 i_d 的波动。运行工况为:电机以 3 000 r/min 额定转速空载启动并保持高速运行,在 0.5 s 时引入负载突变,模拟负载转矩从 0 N·m 突变至额定转矩 2.39 N·m 的工况。同时,在相同运行条件下,研究了实际电感值为 50%标称电感值($L_d = 0.5\hat{L}_d$, $L_q = 0.5\hat{L}_q$)和 150%标称电感值($L_d = 1.5\hat{L}_d$, $L_q = 1.5\hat{L}_q$)时的电流动态响应,以验证不同控制策略在电感参数失配情况下的解耦性能。

图 7~图 9 展示了实际电感值为标称值时不同控制策略下 d、q 轴电流波形图。

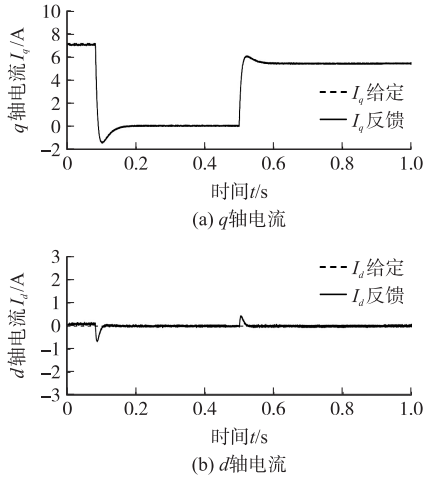


图 7 $L_d=\hat{L}_d$,$L_q=\hat{L}_q$ PI 控制 d、q 轴电流波形

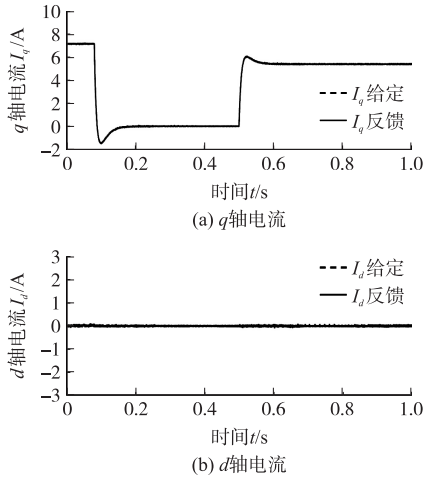


图 8 $L_d=\hat{L}_d$,$L_q=\hat{L}_q$ 电流反馈解耦控制 d、q 轴电流波形

表 3 给出了突加 2.39 N·m 负载转矩下 d、q 轴电流稳态时的跟踪偏差标准差。可以看出,转矩电流 i_q 在 0.08 s 从 7.2 A 迅速降至 0.025 A,在 0.5 s 时上升至 0.415 A。在这一过程中,由于电机高速启动,耦合效应导致电机反电动势迅速增加,PI 控制和 TD3-PI 均出现了较小的跟踪误差。PI 控制的励磁电流 i_d 在 0.08 s 突降至−0.691 A,在 0.5 s 时突增至 0.457 A,呈现出明显的“尖峰”波动,反映出较强的耦合效应。相比之下,TD3-PI 控制明显降低了励磁电流 i_d 的波动,其在 0.08 s 降至−0.195 A,在 0.5 s 时增至 0.143 A。相较于 PI 控制,TD3-PI 在 0.08 s 和 0.5 s 时间点上的励磁电流 i_d 波动分别降低了 71.78%和 68.7%,有效缓解了耦合效应。电流反馈解耦控制准确补偿了耦合量,励磁电流 i_d 基本无波动,实现了电流的动态解耦。同时,表 3 表明了在电机突加 2.39 N·m 负载转矩后,TD3-PI 控制的电流偏差标准差最小,反映了其电流纹波相对最小,证明了 TD3-PI 控制策略在稳态控制精度上的优势。

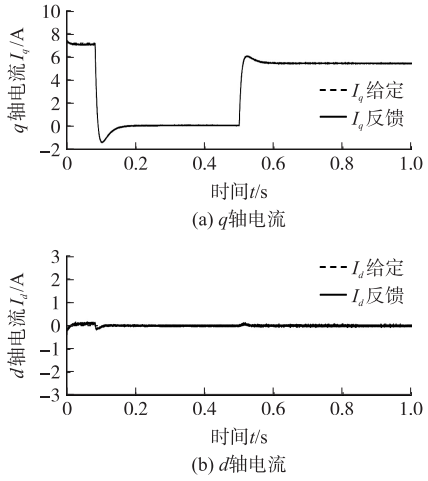


图 9 $L_d=\hat{L}_d$,$L_q=\hat{L}_q$ TD3-PI 控制 d、q 轴电流波形

表 3 突加额定负载转矩下电流稳态跟踪偏差标准差

Std	PI	电流反馈	TD3-PI
e_{iq}	0.043 6	0.042 6	0.042 4
e_{id}	0.030 9	0.030 8	0.030 4

图 10~图 12 展示了在电感参数失配的情况下,三种控制策略的 d、q 轴电流波形。可以发现,在实际电感值降低后,d、q 轴电流纹波明显增大,同时由

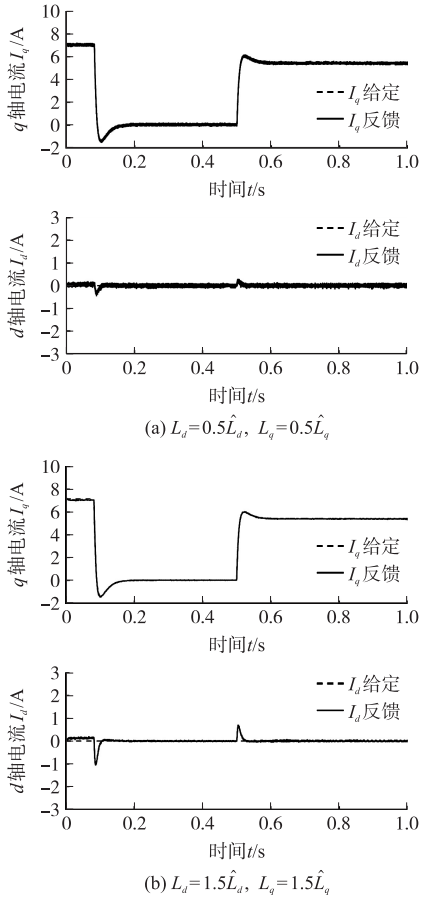
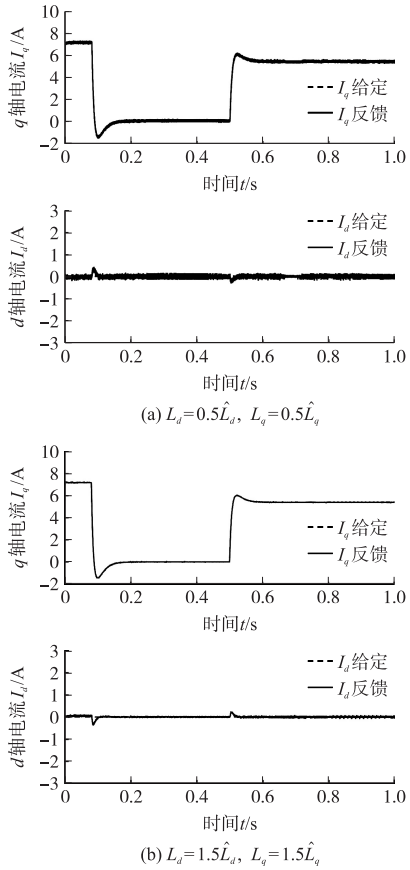


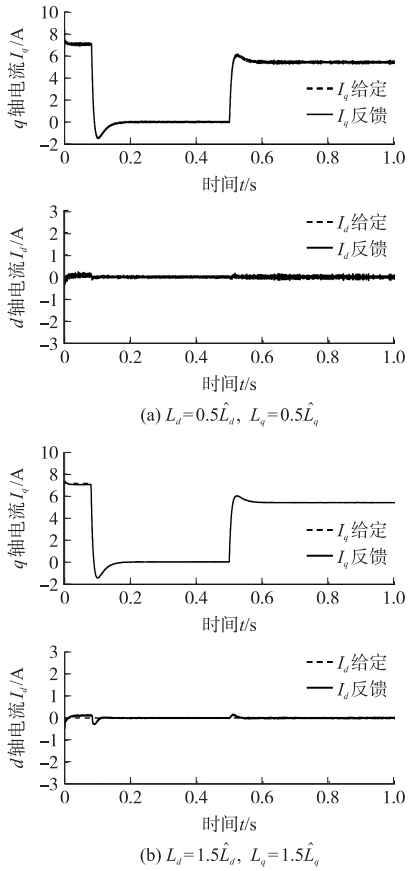
图 10 PI 控制电感参数失配时 d、q 轴电流波形



(a) $L_d=0.5\hat{L}_d$, $L_q=0.5\hat{L}_q$

(b) $L_d=1.5\hat{L}_d$, $L_q=1.5\hat{L}_q$

图 11　电流反馈解耦控制电感参数失配时 d、q 轴电流波形



(a) $L_d=0.5\hat{L}_d$, $L_q=0.5\hat{L}_q$

(b) $L_d=1.5\hat{L}_d$, $L_q=1.5\hat{L}_q$

图 12　TD3-PI 控制电感参数失配时 d、q 轴电流波形

于电感值减小导致耦合电压降低,相应的耦合效应也有所减弱,d 轴电流波动减小;相反,电感值增加时,电流纹波减小,耦合效应增强,d 轴电流波动加剧。

在实际电感值为 50%和 150%标称电感下,当转矩电流 i_q 突变时,PI 控制的励磁电流 i_d 波动峰值分别为−0.469 A、0.298 A 和−1.073 A、0.715 A;而电流反馈解耦控制因电感参数失配无法精确补偿耦合量,其解耦性能明显下降,励磁电流 i_d 的波动峰值分别为 0.442 A、−0.314 A 和−0.38 A、0.243 A;相比之下,TD3-PI 控制的励磁电流 i_d 的波动均明显小于传统 PI 控制和电流反馈解耦控制,其波动峰值分别为−0.148 A、0.136 A 和−0.311 A、0.2 A。与 PI 控制相比,TD3-PI 在 50%和 150%标称电感下的励磁电流 i_d 波动峰值分别减小了 68.44%、54.36% 和 71.01%、66.01%。与电流反馈解耦控制相比,TD3-PI 在 50%和 150%标称电感下的励磁电流 i_d 波动峰值分别减小了 66.52%、56.69%和 18.16%、17.69%。

可见,TD3-PI 控制在电感参数失配时,仍能够有效对耦合电压进行动态补偿,具有良好的解耦能力及鲁棒性,这得益于 TD3 智能体在训练阶段对电感参数失配情况的学习与适应。仿真实验结果证明了本文所提方法在提升电流解耦性能方面的有效性。

5　结语

本文提出了一种基于 TD3-PI 的新型 PMSM 电流解耦控制策略。通过利用 TD3 算法的自适应调节能力,对电流控制过程中的耦合效应进行动态补偿,提高了解耦性能和控制精度。仿真结果表明,与传统 PI 控制、电流反馈解耦控制相比,TD3-PI 控制策略在减少电流波动、提升稳态控制精度以及增强电感参数失配条件下的系统鲁棒性方面具有一定优势。此研究不仅证明了深度强化学习技术在 PMSM 控制领域的潜力和优势,也为解决其他类型电机控制问题提供了新的思路和方法。

参考文献:

[1]　寇宝泉,赵晓坤,张浩泉,等. 永磁同步电机电磁结构及磁场调节技术的综述分析[J]. 中国电机工程学报,2021,41(20):7126−7141.

[2]　付兴贺,顾胜东,熊嘉鑫. 永磁同步电机交直轴电 流解耦控制方法综述[J]. 中国电机工程学报,2024,44(1):314−332.

[3]　Lorenz R D,Lawson D B. Performance of Feed Forward

Current Regulators for Field Oriented Induction Machine Controllers [C]//1986 Annual Meeting Industry Applications Society, Denver, CO, USA, IEEE, 1986: 99-105.

[4] Zhu H, Xiao X, Li Y D. PI Type Dynamic Decoupling Control Scheme for PMSM High Speed Operation[C]// 2010 Twenty-Fifth Annual IEEE Applied Power Electronics Conference and Exposition (APEC), Palm Spring, CA, USA, IEEE, 2010: 1736-1739.

[5] Harnefors L, Nee H P. Model-Based Current Control of AC Machines Using the Internal Model Control Method [J]. IEEE Transactions on Industry Applications, 1998, 34(1): 133-141.

[6] Briz F, Degner M W, Lorenz R D. Analysis and Design of Current Regulators Using Complex Vectors [J]. IEEE Transactions on Industry Applications, 2000, 36(3): 817-825.

[7] Fu X, He H, Xu Y, et al. A Strongly Robust and Easy-Tuned Current Controller for PMSM Considering Parameters Variation[J]. IEEE Access, 2020, 8: 44228-44238.

[8] 杨淑英, 王玉柱, 储昭晗, 等. 基于增益连续扩张状态观测器的永磁同步电机电流解耦控制[J]. 中国电机工程学报, 2020, 40(6): 1985-1997.

[9] 曹凯, 朱勇, 高强, 等. 深度强化学习在自动控制领域研究现状与展望[J]. 排灌机械工程学报, 2023, 41(6): 638-648.

[10] Schenke M, Kirchgässner W, Wallscheid O. Controller Design for Electrical Drives by Deep Reinforcement Learning: A Proof of Concept[J]. IEEE Transactions on Industrial Informatics, 2019, 16(7): 4650-4658.

[11] Bhattacharjee S, Halder S, Yan Y, et al. Real-Time Sil Validation of a Novel Pmsm Control Based on Deep Deterministic Policy Gradient Scheme for Electrified Vehicles[J]. IEEE Transactions on Power Electronics, 2022, 37(8): 9000-9011.

[12] Lillicrap T P, Hunt J J, Pritzel A, et al. Continuous Control with Deep Reinforcement Learning [EB/OL]. (2019-07-05). https://arxiv.org/abs/1509.02971.

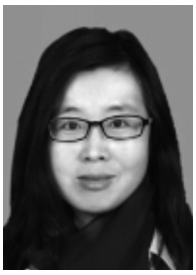
[13] Chen P, He Z, Chen C, et al. Control Strategy of Speed Servo Systems Based on Deep Reinforcement Learning [J]. Algorithms, 2018, 11(5): 65.

[14] Fujimoto S, Hoof H, Meger D. Addressing Function Approximation Error in Actor-Critic Methods [C]//35th International Conference on Machine Learning, Stockholm, Sweden, 2018: 1587-1596.

[15] 王琦, 杨毅远, 江季. Easy RL: 强化学习教程[M]. 北京: 人民邮电出版社, 2022: 33-66.

[16] Mnih V, Kavukcuoglu K, Silver D, et al. Playing Atari with Deep Reinforcement Learning[EB/OL]. (2023-12-19). https://arxiv.org/abs/1312.5602.

王艺霖(1998—),男,硕士研究生,研究方向为永磁同步电机控制与应用,merlynwong@163.com;



刘慧博(1972—),女,通讯作者,博士,教授,研究方向为伺服电机高性能控制,1294851515@qq.com。