DOI: 10.7641/CTA.2014.41029

一种新型附加学习控制器及电力系统应用实例

郭文涛†, 梅生伟, 刘　锋

(清华大学 电机系 电力系统国家重点实验室, 北京 100084)

摘要: 维纳的《控制论》和钱学森的《工程控制论》共同奠定了经典控制理论的基础. 在此基础之上, 现代控制理论在性能优化和处理不确定性等方面对经典控制理论作了进一步的发展. 本文提出一种融合经典控制与现代控制的控制方法, 将属于现代控制的近似动态规划学习控制器以并联的方式附加到经典控制器上, 从而形成一类新型附加学习控制器. 该控制器采用基于策略迭代近似动态规划的训练算法和基于最小二乘的代价函数逼近算法, 从而具有高策略搜索效率和高样本利用效率, 其中动作依赖代价函数的引入使得在线学习不依赖系统模型. 总之, 所提附加学习控制器一方面融合了已有经典控制器的先验知识, 另一方面为已有经典控制器提供了可设计的目标函数和自趋优机制. 论文进一步从理论上严格证明了所提附加学习控制方法的稳定性和收敛性. 针对双馈风电场暂态无功控制问题的仿真研究验证了所提附加学习控制器的正确性和方法的有效性.

关键词: 附加学习控制; 在线; 优化; 自适应; 近似动态规划

中图分类号: TP181　　**文献标识码**: A

A novel supplementary learning controller and its application in power systems

GUO Wen-tao†, MEI Sheng-wei, LIU Feng

(State Key Laboratory of Power Systems, Department of Electrical Engineering, Tsinghua University, Beijing 100084, China)

Abstract: “Cybernetics” by Norbert Wiener and “Engineering Cybernetics” by Tsien Hsue-shen have laid a solid foundation for classical control theory. Based on the classical control theory, modern control theory makes advance in optimizing performance and handling uncertainties. In this paper, a supplementary learning controller is proposed as a method to incorporate the classical control theory and the modern control theory, which adds a supplementary learning controller based on approximate dynamic programming on an existing classical controller. Policy iteration approximated dynamic programming algorithm and least squares method are employed as the training algorithm, which enjoys the policy-search efficiency of policy iteration and the data-utilization efficiency of least squares. Action dependent cost function is introduced to make the online learning model-free. By using such a supplementary learning controller, the prior knowledge of the existing classical controller can be fully utilized. On the other hand, the supplementary learning controller can optimize the performance of the closed-loop system. Furthermore, stability and convergence of the proposed method is proved rigorously. Simulation studies on reactive power control of doubly-fed induction generators (DFIGs) based wind farm validate the proposed supplementary learning controller.

Key words: supplementary learning control; online; optimization; adaptation; approximate dynamic programming

1　引言(Introduction)

进入20世纪后, 现代工业特别是军事工业的快速发展推动了自动调节和系统稳定性等方面的实践与研究工作. 在总结大量工程实践的基础上, 维纳于1948年完成了《控制论》一书, 将控制论定义为“在动物和机器中控制和通讯的科学”[1]. 《控制论》一书从哲学高度建立了自动控制的认识论和方法论. 但对于科技工作者特别是一线工程人员来说, 该书的论述过于抽象, 难以直接应用于工程实践. 为解决这一问题, 钱学森于1954年出版了《工程控制论》, 系统地发展了能应用于工程实际的自动控制理论与方法, 成为继维纳《控制论》之后自动控制领域的又一里程碑式成果.

《控制论》和《工程控制论》奠定了经典控制理论的基础. 经典控制理论主要研究线性、时不变、单输入单输出的控制问题, 以Nyquist判据和Routh判据为稳定性分析方法, 以幅相频率特性图和根轨迹图为设计方法, 以比例–积分–微分(proportional-integral-

收稿日期: 2014−11−04; 录用日期: 2014−12−18.

†通信作者. E-mail: gwt329@gmail.com.

基金项目: 国家自然科学基金资助项目(51377092); 国家重点基础研究专项经费资助项目(2012CB215103); 国家自然科学基金创新研究群体科学基金资助项目(51321005).

derivative, PID)控制为主要调节方法[2]. PID控制器具有结构简单, 参数物理意义明确, 便于现场调试等优点. 在当前控制工程实践中, 基于经典控制理论设计的PID控制器是应用最为广泛的一类控制器. 据估计, 工程中超过90%的控制器都是PID控制器[3].

随着工业技术的发展, 工程实践对控制理论与方法提出了更高的要求, 如处理多输入多输出问题、性能优化、处理不确定性等. 为适应工业需要, 控制理论界在经典控制的基础上发展了多种新型控制理论, 统称为现代控制理论. 其中, 线性最优控制[4]和非线性最优控制[5]基于准确的被控系统数学模型, 可以优化系统的控制性能; 鲁棒控制[6]基于标称系统的数学模型, 可以抑制干扰对闭环系统性能的影响; 自适应控制[7]则给出了在线处理不确定性的机制.

众所周知, Bellman的动态规划原理是最优控制的充分必要条件. 但经典动态规划随状态空间维数、策略空间维数、决策阶段数的增加会产生“维数灾”问题. 为此, 学者们自20世纪70年代以来开展了近似动态规划(approximate dynamic programming, ADP)[8–9]理论的研究. 近似动态规划评价现有控制策略的控制效果并给出评价函数, 进一步根据评价函数更新控制策略, 如此反复迭代直到收敛到最优控制策略[10–14]. 在近似动态规划中, 神经网络等函数逼近结构被用来近似代价函数和控制策略, 从而降低了存储和计算负担. 近似动态规划兼具最优控制和自适应控制的优点, 在电力系统控制等领域取得了一些应用成果[15–18], 是一类很有潜力的现代控制理论.

值得一提的是, 经典控制和现代控制具有密切的联系. 如文[19]所指出, 经典控制是现代控制的基础, 现代控制则是经典控制的发展. 而将二者有机结合, 则是发展和提高未来控制技术水平的必然选择[19]. 循此思路, 本文提出一种将经典控制与现代控制融合的控制方法, 将属于现代控制方法的近似动态规划学习控制器以并联的方式附加到经典控制器上, 在不改变原控制器结构的前提下, 赋予原控制器在线学习的能力, 实现原控制器的在线自趋优和自适应. 提出这一方法的出发点, 是已有的经典控制器往往经过精心设计, 并已在实际运行中得到检验, 抛弃已有控制器的先验知识既是一种浪费, 也不容易被工程一线所接受. 与常规的直接用近似动态规划控制取代原控制器的方法不同, 本文所提方法将近似动态规划的在线学习能力与原控制器的先验知识有机结合, 一方面原控制器通过附加学习控制器获得在线自趋优和自适应的能力, 在运行中持续提高控制性能, 另一方面近似动态规划附加控制充分利用已有控制器的先验知识, 为从理论上保证学习的稳定性与收敛性提供支撑.

本文组织结构如下: 第2节对最优附加控制问题进行建模. 第3节给出基于策略迭代近似动态规划和动作依赖代价函数的附加学习控制算法, 包括策略评价和策略改善的在线实现. 第4节严格证明了所提基于近似动态规划的附加学习控制方法的稳定性和收敛性. 第5节以双馈风电场暂态无功控制为例对所提附加控制方法进行仿真测试. 第6节给出本文结论.

2 问题建模(Problem formulation)

考虑一般的离散时间非线性系统:

$$x_{k+1} = \bar{F}(x_k, u_k),\ k = 0, 1, 2, \cdots, \tag{1}$$

其中: $x_k \in \mathbb{R}^n$ 是状态向量, $u_k \in \mathbb{R}^m$ 是控制向量, $\bar{F}(x_k, u_k) : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ 是系统动态.

本文假设系统(1)存在初始控制器$u_{\mathrm{o},k} = u_{\mathrm{o}}(x_k)$. 在初始控制器作用下, 系统能保持稳定, 但性能有待进一步改善. 为改善初始控制器的性能, 将附加控制器$u_{\mathrm{s},k} = u_{\mathrm{s}}(x_k)$并联在初始控制器上. 考虑附加控制器的作用, 系统(1)的控制输入为

$$u_k = u_{\mathrm{o},k} + u_{\mathrm{s},k} = u_{\mathrm{o}}(x_k) + u_{\mathrm{s}}(x_k). \tag{2}$$

附加控制器的设计目标, 是在稳定系统(1)的前提下优化一定的性能指标. k时间的瞬时代价函数通常定义为

$$U(x_k, u_{\mathrm{s},k}) = W(x_k) + u_{\mathrm{s},k}^{\mathrm{T}} R u_{\mathrm{s},k},\ k = 0, 1, 2, \cdots, \tag{3}$$

其中: $W(x_k) : \mathbb{R}^n \to \mathbb{R}$是$x_k$的正定函数, $R \in \mathbb{R}^{m\times m}$为正定矩阵. 在瞬时代价函数(3)中, 与状态$x_k$有关的代价反映在$W(x_k)$中, 控制代价反映在$u_{\mathrm{s},k}^{\mathrm{T}} R u_{\mathrm{s},k}$中.

附加控制器$u_{\mathrm{s},k} = u_{\mathrm{s}}(x_k)$的优化目标,通常为附加控制器作用下系统从状态$x_k$出发的总瞬时代价, 或动作依赖代价函数

$$Q^{u_{\mathrm{s}}}(x_k, u_{\mathrm{s},k}) = U(x_k, u_{\mathrm{s},k}) + \sum_{i=k+1}^{\infty} U(x_i, u_{\mathrm{s}}(x_i)). \tag{4}$$

最优附加控制, 就是寻找附加控制器$u_{\mathrm{s},k} = u_{\mathrm{s}}(x_k)$, 以稳定系统(1)并最小化代价函数(4). 最优代价函数定义为

$$Q^*(x_k, u_{\mathrm{s},k}) = \min_{u_{\mathrm{s}}} Q^{u_{\mathrm{s}}}(x_k, u_{\mathrm{s},k}). \tag{5}$$

根据Bellman最优性准则[20], 最优代价函数满足如下的Bellman最优性方程:

$$Q^*(x_k, u_{\mathrm{s},k}) = U(x_k, u_{\mathrm{s},k}) + \min_{u_{\mathrm{s},k+1}} Q^*(x_{k+1}, u_{\mathrm{s},k+1}). \tag{6}$$

最优附加控制器可由下式解出:

$$u_{\mathrm{s}}^*(x_k) = \arg\min_{u_{\mathrm{s},k}} Q^*(x_k, u_{\mathrm{s},k}). \tag{7}$$

在数学上, 离散时间非线性系统的Bellman最优性方程(6)没有一般的解析解法. 动态规划从理论上可以

求解这一多阶段决策问题, 但在应用中存在难以克服的“维数灾”问题. 为了解决这一问题, 本文采用近似动态规划方法在线训练附加控制器以优化原控制器的控制性能.

为方便下文的讨论, 本文将在给定初始控制器$u_{\mathrm{o},k}=u_{\mathrm{o}}(x_k)$和附加控制信号$u_{\mathrm{s},k}$作用下的系统方程记为如下形式:

$$x_{k+1}=F(x_k,u_{\mathrm{s},k})=\bar{F}(x_k,u_{\mathrm{s},k}+u_{\mathrm{o}}(x_k)). \quad (8)$$

3 基于ADP的附加学习控制(Supplementary learning control based on ADP)

3.1 附加学习控制的整体框架(Framework of supplementary learning control)

本文采用图1所示的ADP附加学习控制结构实现对初始控制器的在线优化, 赋予原控制器以在线优化和在线适应的能力. 具体地, 基于ADP的附加学习控制结构包括两个部分, 评价体和执行体, 其中评价体用于评价当前附加控制器的性能, 执行体实际上为附加控制器, 在评价体的指导下输出附加控制信号以改善原控制器的性能.

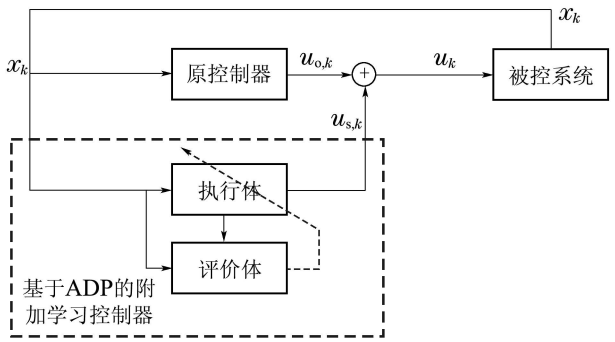

图 1 基于ADP的附加学习控制的结构图

Fig. 1 Schematic of supplementary learning control based on ADP

考虑到策略迭代[21]近似动态规划算法具有策略搜索效率高和收敛速度快的优势, 策略迭代近似动态规划算法被用于图1所示的ADP附加学习控制结构的学习算法. 策略迭代算法要求初始附加控制器为容许控制(admissible control)[22].

定义 1(容许控制)　附加控制器$u_{\mathrm{s},k}=u_{\mathrm{s}}(x_k)$是系统(8)关于代价函数(4)的容许控制, 若$u_{\mathrm{s},k}=u_{\mathrm{s}}(x_k)$在$\mathbb{R}^n$上连续, $u_{\mathrm{s}}(0)=0$, $u_{\mathrm{s},k}=u_{\mathrm{s}}(x_k)$能稳定系统(8)且对应的代价函数(4)对任意$x_k\in\mathbb{R}^n$是有限的.

注 1　在控制工程实践中, 原控制器通常能使系统稳定且代价函数有限. 在这种情形下, 使用本文提出附加学习控制方法, 可以保留原控制器的先验知识, 只需令初始附加控制器为零输出控制器, 即可满足初始容许控制的要求.

策略迭代算法包括两个步骤: 一是通过评价体实现的策略评价, 二是在评价体指导下通过执行体实现的策略改善. 这两个步骤交替进行直到算法收敛.

步骤 1　策略评价.

策略评价根据附加控制器应用于被控系统后的在线响应确定附加控制器的代价函数. 策略评价需求解满足如下方程的函数$Q^{(i)}(x_k,u_{\mathrm{s},k})$:

$$\begin{aligned}&Q^{(i)}(x_k,u_{\mathrm{s},k})=\\&U(x_k,u_{\mathrm{s},k})+Q^{(i)}(x_{k+1},u_{\mathrm{s}}^{(i)}(x_{k+1})),\\&i=0,1,2,\cdots,\ \forall x_k,\ \forall u_{\mathrm{s},k},\end{aligned}\quad (9)$$

其中: i是迭代次数指标, $u_{\mathrm{s}}^{(i)}(x_k)$是第$i$次迭代中的附加控制器, 而$Q^{(i)}(x_k,u_{\mathrm{s},k})$是对应于附加控制器$u_{\mathrm{s}}^{(i)}(x_k)$的如式(4)所示的代价函数, $Q^{(i)}(x_k,u_{\mathrm{s},k})$是$x_k$与$u_{\mathrm{s},k}$的正定函数.

步骤 2　策略改善.

策略改善是根据策略评价中得到的代价函数$Q^{(i)}(x_k,u_{\mathrm{s},k})$, 确定第$(i+1)$次迭代中的附加控制器$u_{\mathrm{s}}^{(i+1)}(x_k)$:

$$\begin{aligned}u_{\mathrm{s}}^{(i+1)}(x_k)=&\arg\min_{u_{\mathrm{s},k}}Q^{(i)}(x_k,u_{\mathrm{s},k}),\\&i=0,1,2,\cdots,\forall x_k.\end{aligned}\quad (10)$$

式(9)−(10)描述了策略迭代算法的一般形式. 下面本文介绍如何在线实现策略评价和策略改善.

3.2 策略评价的在线实现(Online implementation of policy evaluation)

为避免对所有状态和控制对应的代价函数进行遍历, 近似动态规划算法采用一定的函数逼近结构来近似代价函数. 式(11)所示的参数线性逼近结构是最常采用的一种逼近结构:

$$\begin{aligned}Q(x_k,u_{\mathrm{s},k})=&\sum_{i=1}^{L}w_i\varphi_i(x_k,u_{\mathrm{s},k})=\\&W^{\mathrm{T}}\phi(x_k,u_{\mathrm{s},k}),\end{aligned}\quad (11)$$

其中: $W\in\mathbb{R}^L$是逼近结构的权重向量, $\phi(x_k,u_{\mathrm{s},k}):\mathbb{R}^n\times\mathbb{R}^m\to\mathbb{R}^L$是基函数向量. 基函数$\phi(x_k,u_{\mathrm{s},k})$可以是多项式函数和径向基函数等.

将代价函数逼近式(11)代入策略评价式(9)可得

$$\begin{aligned}&W^{(i)\mathrm{T}}(\phi(x_k,u_{\mathrm{s},k})-\phi(x_{k+1},u_{\mathrm{s}}^{(i)}(x_{k+1})))=\\&U(x_k,u_{\mathrm{s},k}),\end{aligned}\quad (12)$$

其中$W^{(i)}$是第i次迭代中的代价函数的参数. 考虑到最小二乘算法对样本的高利用效率, 式(12)采用批量最小二乘算法进行计算. 首先采集N个样本数据, 包括$\phi(x_j,u_{\mathrm{s},j})-\phi(x_{j+1},u_{\mathrm{s}}^{(i)}(x_{j+1}))$和$U(x_j,u_{\mathrm{s},j})$, 其中$j=k,k+1,\cdots,k+N$, 通常$N\geqslant L$. 将这些样本按行排列并记为$\psi_k\in\mathbb{R}^{L\times N}$和$\mu_k\in\mathbb{R}^{1\times N}$, 则$W^{(i)}$可由下式计算得到:

$$W^{(i)} = (\psi_k^{\dagger})^{\mathrm{T}}\mu_k^{\mathrm{T}}, \tag{13}$$

其中$\psi_k^{\dagger}$为ψ_k的伪逆或Moore-Penrose逆[23]. 式(12)也可用递归最小二乘计算[24].

3.3 策略改善的在线实现(Online implementation of policy improvement)

在求解得到代价函数$Q^{(i)}(x_k, u_{\mathrm{s},k})$后, 附加控制器可根据式(10)进行更新. 注意第$(i+1)$次迭代中的附加控制器$u_{\mathrm{s}}^{(i+1)}(x_k)$完全由第$i$次迭代的代价函数$Q^{(i)}(x_k, u_{\mathrm{s},k})$决定, 而与系统动态方程无关, 因此实现了算法的无模型化.

进一步, 本文考虑一种特例, 代价函数逼近结构为如下式所示的二次函数:

$$Q^{(i)}(x_k, u_{\mathrm{s},k}) = \begin{bmatrix} x_k \\ u_{\mathrm{s},k} \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} S_{\mathrm{xx}}^{(i)} & S_{\mathrm{xu}}^{(i)} \\ S_{\mathrm{ux}}^{(i)} & S_{\mathrm{uu}}^{(i)} \end{bmatrix} \begin{bmatrix} x_k \\ u_{\mathrm{s},k} \end{bmatrix}, \tag{14}$$

其中: $S_{\mathrm{xx}}^{(i)} \in \mathbb{R}^{n\times n}$, $S_{\mathrm{xu}}^{(i)} \in \mathbb{R}^{n\times m}$, $S_{\mathrm{ux}}^{(i)} = S_{\mathrm{xu}}^{(i)\mathrm{T}}$, $S_{\mathrm{uu}}^{(i)} \in \mathbb{R}^{m\times m}$是第$i$次迭代中代价函数的参数. 在这种代价函数逼近结构下, 策略改善式(10)具有解析解

$$u_{\mathrm{s}}^{(i+1)}(x_k) = -[S_{\mathrm{uu}}^{(i)}]^{-1}S_{\mathrm{ux}}^{(i)}x_k, \tag{15}$$

也就是说, 附加控制器是线性状态反馈控制器.

4 稳定性和收敛性分析(Stability and convergence analysis)

本节分析所提基于近似动态规划的附加学习控制方法的稳定性和收敛性. 首先, 假设如下条件成立.

假设 1 系统(1)是可控的且系统动态$\bar{F}(x_k, u_k)$对x_k和u_k是Lipschitz连续的.

假设 2 在控制$u_k = 0$的情况下, $x_k = 0$是系统(1)的平衡点, 即$\bar{F}(0, 0) = 0$.

假设 3 原控制器$u_k = u_{\mathrm{o}}(x_k)$是系统(1)的容许控制, 且满足$u_{\mathrm{o}}(0) = 0$; 附加控制器$u_{\mathrm{s},k} = u_{\mathrm{s}}(x_k)$满足$u_{\mathrm{s}}(0) = 0$.

引入附加学习控制器后系统的稳定性由定理1给出.

定 理 1 令假设1–3成立. 迭代代价函数$Q^{(i)}(x_k, u_{\mathrm{s},k})$和迭代附加控制器$u_{\mathrm{s}}^{(i+1)}(x_k)$分别由策略评价式(9)和策略改善式(10) 得到. 则对$i = 0, 1, 2, \cdots$, 在迭代附加控制器$u_{\mathrm{s}}^{(i)}(x_k)$作用下, 系统(8)渐近稳定.

证 若$i = 0$, 根据初始条件, $u_{\mathrm{s}}^{(0)}(x_k)$是系统(8)的容许控制, 则必有$u_{\mathrm{s}}^{(0)}(x_k)$使系统(8)渐近稳定.

若$i \geqslant 1$, 在策略评价式(9)中, 取迭代次数为$i-1$并令$u_{\mathrm{s},k} = u_{\mathrm{s}}^{(i)}(x_k)$, 有

$$\begin{aligned} &Q^{(i-1)}(x_k, u_{\mathrm{s}}^{(i)}(x_k)) = \\ &U(x_k, u_{\mathrm{s}}^{(i)}(x_k)) + \\ &Q^{(i-1)}(F(x_k, u_{\mathrm{s}}^{(i)}(x_k)), u_{\mathrm{s}}^{(i-1)}(F(x_k, u_{\mathrm{s}}^{(i)}(x_k)))). \end{aligned} \tag{16}$$

根据策略改善式(10), 得

$$Q^{(i-1)}(x_k, u_{\mathrm{s}}^{(i)}(x_k)) \leqslant Q^{(i-1)}(x_k, u_{\mathrm{s}}^{(i-1)}(x_k)). \tag{17}$$

联立式(16)和式(17), 有

$$\begin{aligned} &Q^{(i-1)}(F(x_k, u_{\mathrm{s}}^{(i)}(x_k)), u_{\mathrm{s}}^{(i-1)}(F(x_k, u_{\mathrm{s}}^{(i)}(x_k)))) - \\ &Q^{(i-1)}(x_k, u_{\mathrm{s}}^{(i-1)}(x_k)) \leqslant \\ &-U(x_k, u_{\mathrm{s}}^{(i)}(x_k)). \end{aligned} \tag{18}$$

由于$U(x_k, u_{\mathrm{s},k})$正定, 由式(18)可知

$$\begin{aligned} &Q^{(i-1)}(F(x_k, u_{\mathrm{s}}^{(i)}(x_k)), u_{\mathrm{s}}^{(i-1)}(F(x_k, u_{\mathrm{s}}^{(i)}(x_k)))) < \\ &Q^{(i-1)}(x_k, u_{\mathrm{s}}^{(i-1)}(x_k)) \end{aligned} \tag{19}$$

对任意$x_k \neq 0$成立. 另, $Q^{(i-1)}(x_k, u_{\mathrm{s}}^{(i-1)}(x_k))$为$x_k$的正定函数, 因此$Q^{(i-1)}(x_k, u_{\mathrm{s}}^{(i-1)}(x_k))$为系统(8) 在附加控制器$u_{\mathrm{s}}^{(i)}(x_k)$作用下的Lyapunov函数, 系统(8)在附加控制器$u_{\mathrm{s}}^{(i)}(x_k)$作用下渐近稳定. 证毕.

根据定理1, 附加学习控制器在学习过程中不会导致系统失稳, 这是附加学习控制器可以实现在线应用的重要前提. 下面本文分析所提附加学习控制方法的收敛性.

引理 1 令假设1–3成立. 迭代代价函数$Q^{(i)}(x_k, u_{\mathrm{s},k})$和迭代附加控制器$u_{\mathrm{s}}^{(i+1)}(x_k)$分别由策略评价式(9)和策略改善式(10)得到, 则对$i = 0, 1, 2, \cdots$, $\forall x_k$, 有$Q^{(i+1)}(x_k, u_{\mathrm{s}}^{(i+1)}(x_k)) \leqslant Q^{(i)}(x_k, u_{\mathrm{s}}^{(i)}(x_k))$.

证 根据策略评价式(9), 有

$$\begin{aligned} &Q^{(i+1)}(x_k, u_{\mathrm{s}}^{(i+1)}(x_k)) = \\ &U(x_k, u_{\mathrm{s}}^{(i+1)}(x_k)) + Q^{(i)}(x_{k+1}, u_{\mathrm{s}}^{(i)}(x_{k+1})) + \\ &Q^{(i+1)}(x_{k+1}, u_{\mathrm{s}}^{(i+1)}(x_{k+1})) - \\ &Q^{(i)}(x_{k+1}, u_{\mathrm{s}}^{(i)}(x_{k+1})) = \\ &Q^{(i)}(x_k, u_{\mathrm{s}}^{(i+1)}(x_k)) + Q^{(i+1)}(x_{k+1}, u_{\mathrm{s}}^{(i+1)}(x_{k+1})) - \\ &Q^{(i)}(x_{k+1}, u_{\mathrm{s}}^{(i)}(x_{k+1})). \end{aligned} \tag{20}$$

代入策略改善式(10), 得

$$\begin{aligned} &Q^{(i+1)}(x_k, u_{\mathrm{s}}^{(i+1)}(x_k)) \leqslant \\ &Q^{(i)}(x_k, u_{\mathrm{s}}^{(i)}(x_k)) + Q^{(i+1)}(x_{k+1}, u_{\mathrm{s}}^{(i+1)}(x_{k+1})) - \\ &Q^{(i)}(x_{k+1}, u_{\mathrm{s}}^{(i)}(x_{k+1})). \end{aligned} \tag{21}$$

进一步, 有

$$\begin{aligned} &Q^{(i+1)}(x_k, u_{\mathrm{s}}^{(i+1)}(x_k)) - Q^{(i)}(x_k, u_{\mathrm{s}}^{(i)}(x_k)) \leqslant \\ &Q^{(i+1)}(x_{k+1}, u_{\mathrm{s}}^{(i+1)}(x_{k+1})) - \\ &Q^{(i)}(x_{k+1}, u_{\mathrm{s}}^{(i)}(x_{k+1})) \leqslant \cdots \leqslant \\ &Q^{(i+1)}(x_{k+N}, u_{\mathrm{s}}^{(i+1)}(x_{k+N})) - \end{aligned}$$

$$Q^{(i)}(x_{k+N}, u_{\rm s}^{(i)}(x_{k+N})), \tag{22}$$

其中N为任意正整数. 根据定理1, $u_{\rm s}^{(i+1)}(x_k)$为渐近稳定控制器, 则$N \to \infty$时$x_{k+N} \to 0$. 在式(22)中, 令$N \to \infty$, 则有

$$Q^{(i+1)}(x_k, u_{\rm s}^{(i+1)}(x_k)) - Q^{(i)}(x_k, u_{\rm s}^{(i)}(x_k)) \leqslant 0. \tag{23}$$

证毕.

引理1给出了所提附加学习控制方法的一个重要性质, 即代价函数随迭代次数是单调递减的. 每经过一次迭代学习, 控制器的性能都单调改善. 即使迭代学习尚未收敛, 中间过程中的附加控制器的性能也优于初始控制器. 这是支持所提附加学习控制方法在线应用的另一个性质. 进一步, 由引理1可证明如下的收敛性定理.

定理 2　令假设1–3成立. 迭代代价函数$Q^{(i)}(x_k, u_{{\rm s},k})$和迭代附加控制器$u_{\rm s}^{(i+1)}(x_k)$分别由策略评价式(9)和策略改善式(10)得到, 最优代价函数$Q^*(x_k, u_{\rm s}^*(x_k))$和最优附加控制器$u_{\rm s}^*(x_k)$由式(6)–(7)定义. 则当迭代次数$i \to \infty$时, 迭代代价函数$Q^{(i)}(x_k, u_{\rm s}^{(i)}(x_k))$趋于最优代价函数$Q^*(x_k, u_{\rm s}^*(x_k))$, 迭代附加控制器$u_{\rm s}^{(i)}(x_k)$趋于最优附加控制器$u_{\rm s}^*(x_k)$.

证　为了证明该定理, 本文首先证明$Q^{(i)}(x_k, u_{\rm s}^{(i)}(x_k)) \geqslant Q^*(x_k, u_{\rm s}^*(x_k))$.

由式(6)–(7)可得

$$\begin{aligned} &Q^*(x_k, u_{\rm s}^*(x_k)) = \\ &\min_{u_{{\rm s},k}}\{U(x_k, u_{{\rm s},k}) + \min_{u_{{\rm s},k+1}} Q^*(x_{k+1}, u_{{\rm s},k+1})\} = \\ &\cdots = \\ &\min_{u_{\rm s}} \sum_{i=k}^{\infty} U(x_i, u_{\rm s}(x_i)). \end{aligned} \tag{24}$$

另一方面, 由策略评价式(9)可得

$$\begin{aligned} &Q^{(i)}(x_k, u_{\rm s}^{(i)}(x_k)) = \\ &U(x_k, u_{\rm s}^{(i)}(x_k)) + Q^{(i)}(x_{k+1}, u_{\rm s}^{(i)}(x_{k+1})) = \\ &\cdots = \\ &\sum_{i=k}^{\infty} U(x_i, u_{\rm s}^{(i)}(x_i)). \end{aligned} \tag{25}$$

比较式(24)和式(25)可得$Q^{(i)}(x_k, u_{\rm s}^{(i)}(x_k)) \geqslant Q^*(x_k, u_{\rm s}^*(x_k))$.

第2步证明当迭代次数$i \to \infty$时,

$$Q^{(i)}(x_k, u_{\rm s}^{(i)}(x_k)) \leqslant Q^*(x_k, u_{\rm s}^*(x_k)).$$

根据式(20)、策略评价式(9)、策略改善式(10), 有

$$\begin{aligned} &Q^{(i+1)}(x_k, u_{\rm s}^{(i+1)}(x_k)) \leqslant \\ &Q^{(i)}(x_k, u_{\rm s}^*(x_k)) + Q^{(i+1)}(x_{k+1}, u_{\rm s}^{(i+1)}(x_{k+1})) - \\ &Q^{(i)}(x_{k+1}, u_{\rm s}^{(i)}(x_{k+1})) = \\ &U(x_k, u_{\rm s}^*(x_k)) + Q^{(i)}(x_{k+1}, u_{\rm s}^{(i)}(x_{k+1})) + \\ &Q^{(i+1)}(x_{k+1}, u_{\rm s}^{(i+1)}(x_{k+1})) - \\ &Q^{(i)}(x_{k+1}, u_{\rm s}^{(i)}(x_{k+1})). \end{aligned} \tag{26}$$

根据引理1, 式(26)可推得

$$\begin{aligned} &Q^{(i+1)}(x_k, u_{\rm s}^{(i+1)}(x_k)) \leqslant \\ &U(x_k, u_{\rm s}^*(x_k)) + Q^{(i)}(x_{k+1}, u_{\rm s}^{(i)}(x_{k+1})) \leqslant \\ &\cdots \leqslant \\ &\sum_{j=k}^{k+i} U(x_j, u_{\rm s}^*(x_j)) + Q^{(0)}(x_{k+i+1}, u_{\rm s}^{(0)}(x_{k+i+1})). \end{aligned} \tag{27}$$

由于$u_{\rm s}^{(0)}(x_k)$是渐近稳定控制, $Q^{(0)}(x_k, u_{\rm s}^{(0)}(x_k))$是正定函数, 故$i \to \infty$时, $Q^{(0)}(x_{k+i+1}, u_{\rm s}^{(0)}(x_{k+i+1})) \to 0$. 故$i \to \infty$时由式(27)可得

$$Q^{(i)}(x_k, u_{\rm s}^{(i)}(x_k)) \leqslant Q^*(x_k, u_{\rm s}^*(x_k)).$$

综上所述, 当迭代次数$i \to \infty$时, 迭代代价函数$Q^{(i)}(x_k, u_{\rm s}^{(i)}(x_k)) \to Q^*(x_k, u_{\rm s}^*(x_k))$. 相应地, 迭代附加控制器$u_{\rm s}^{(i)}(x_k) \to u_{\rm s}^*(x_k)$.　　证毕.

定理2证明了随着迭代学习的进行, 迭代代价函数收敛到最优代价函数, 迭代附加控制器收敛到最优附加控制器. 事实上, 由于策略迭代算法的高策略搜索效率和最小二乘方法的高样本利用效率, 附加学习控制器的收敛速度非常快.

5　风电场附加无功控制应用(Reactive power control of DFIG wind farm)

为测试所提附加学习控制方法的在线优化能力和对不确定性的适应能力, 本节以双馈风机(DFIG)风电场的附加无功控制为例进行仿真测试.

5.1　测试电力系统(Test power system)

测试电力系统如图2所示, 包括3个地理区域. 区域1为发电中心, 发电机G1建模为理想电压源, 发电机G2建模为含详细励磁和调速系统模型的水电机组; 区域3为负荷中心, 本地发电机G3建模为含详细励磁和调速系统模型的火电机组; 区域2位于区域1和区域3之间, 主要承担输电任务, 其中在节点12接入400 MW的双馈风电场. 系统详细建模和参数可参考文[25–26].

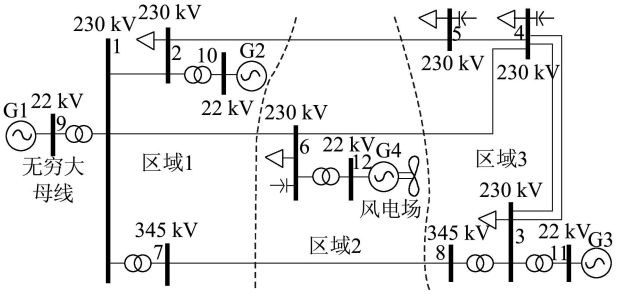

图 2　测试电力系统单线图

Fig. 2　Single-line diagram of the benchmark power system

5.2 风电场附加无功控制器设计(Design of reactive power control)

双馈风电场对系统中的短路故障非常敏感. 短路故障会引起风电场公共接入点电压跌落, 进而导致风电机组输入机械功率和输出电磁功率的不平衡, 在定子回路产生出大电流并在转子回路感应出大电流, 进一步激发风电场输出有功功率振荡. 为解决上述问题, 对双馈风电场进行暂态无功控制是一条可行的路径[27]. 双馈风电场暂态无功控制的思路是通过故障期间和故障后的无功支撑与调制, 补偿风电场公共接入点(PCC)电压跌落, 并抑制故障后风电场的有功功率振荡. 然而部分已有的风电场暂态无功研究依赖于STATCOM等动态无功补偿设备[26], 未能充分利用双馈风电机组自身的暂态无功控制能力. 文[28]提出双馈风电场为邻近的定速恒频风电场提供暂态无功支撑, 文[29]提出一种综合考虑正常运行和故障运行的双馈风电机组无功控制方法, 但文[28–29]所提控制器缺少性能优化能力和适应不确定性能力. 为克服已有研究的不足, 本文利用所提附加学习控制方法设计双馈风电场的附加无功控制.

双馈风电机组的无功功率由转子侧变流器的无功控制环进行控制, 通常以输出一定的无功功率为控制目标. 本文所提出的风电场附加无功控制, 是在转子侧变流器无功控制环上并联一个暂态无功调制信号$\Delta Q_{\rm s}^*$, 实现补偿故障期间公共接入点电压跌落和抑制故障后风电场的有功功率振荡的目标. 附加无功控制器的输入输出关系如图3所示, 输入变量为$x=[\Delta V_6, \Delta P_{\rm g4}]^{\rm T}$, 其中: ΔV_6为节点6的电压偏差, $\Delta P_{\rm g4}$为风电场输出有功功率偏差, 故障前风电场稳态输出功率$P_{\rm g4,ref}=300\,\text{MW}$, 故障前公共接入点稳态电压$V_{6,\rm ref}=1.02\,\text{pu}$, $V_{\rm B}=230\,\text{kV}$, $P_{\rm B}=400\,\text{MW}$, $Q_{\rm B}=40\,\text{MVar}$.

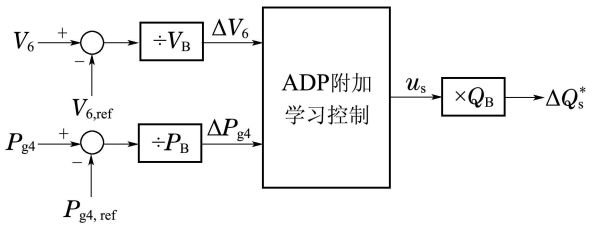

图 3 附加无功控制器的输入输出关系

Fig. 3 Input/output of the supplementary reactive power control

附加无功控制的瞬时代价函数定义为

$$U(x_k, u_{{\rm s},k}) = R_{\rm V}\Delta V_{6,k}^2 + R_{\rm P}\Delta P_{{\rm g4},k}^2 + R_{\rm u}u_{{\rm s},k}^2, \tag{28}$$

其中:

$$R_{\rm V}=R_{\rm P}=1,000,\ R_{\rm u}=10.$$

式(28)包含了对公共接入点电压偏差、风电场输出功率偏差和控制代价的限制.

在基于策略迭代近似动态规划的学习算法中, 本文采用式(14)所示的二次函数来逼近代价函数. 根据3.3节的分析, 当代价函数为二次函数时, 附加控制器具有线性反馈形式

$$u_{\rm s} = K_{\rm V}\Delta V_6 + K_{\rm P}\Delta P_{\rm g4}. \tag{29}$$

基于策略迭代近似动态规划的学习算法如第3.2节和第3.3节所述. 在每次迭代中, 对系统施加外界激励, 如风速变化、短路故障等, 使系统偏离平衡状态运行, 采集系统的暂态响应作为学习的样本. 在本应用中, 采样周期取1 ms, 策略评价在$N=3000$个样本上进行, 即每3 s完成一次策略评价和策略改善的迭代.

5.3 算例 1: 最优性测试(Case 1: optimization test)

算例1验证所提附加学习控制方法的在线优化能力. 由图4可见, 附加无功控制器的反馈系数在10次迭代学习内收敛. 进一步将学习收敛后的附加无功控制器应用于风电场, 与无附加无功控制时的系统时域响应进行对比. 仿真情景为节点1发生持续150 ms的AB两相对地短路故障, 风电场公共接入点电压, 输出有功功率, 风机转子电流分别如图5–7所示.

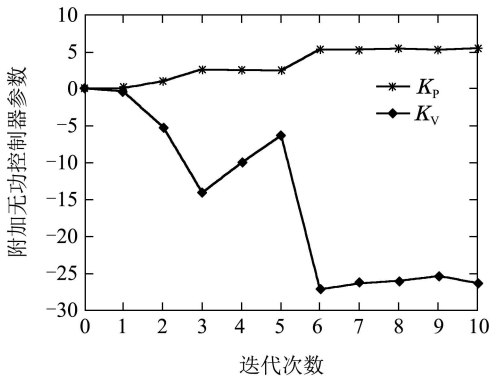

图 4 附加无功控制器的反馈系数: 算例1

Fig. 4 Feedback gain of supplementary controller: case 1

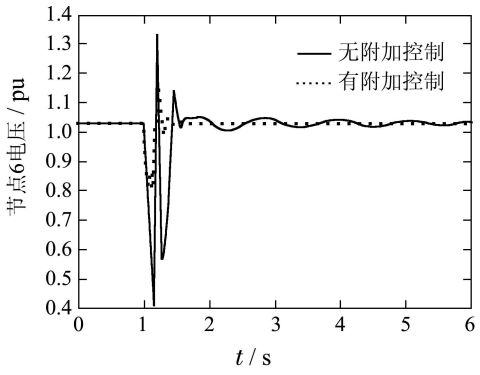

图 5 节点6电压: 算例1

Fig. 5 Voltage of bus 6: case 1

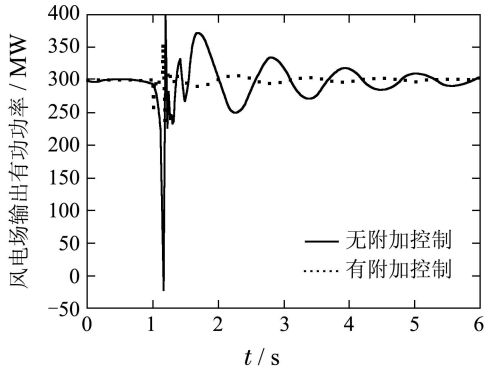

图 6 风电场输出有功功率: 算例1

Fig. 6 Output active power of wind farm: case 1

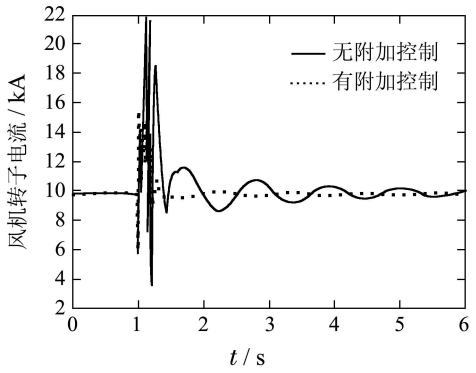

图 7 风机转子电流: 算例1

Fig. 7 Rotor current of DFIG: case 1

由图5可见, 引入附加无功控制后, 公共接入点在故障期间的最低电压由0.41 pu提高到0.81 pu, 且故障后电压得到快速恢复. 由图6可见, 附加无功控制能有效抑制故障后的风电场输出功率振荡, 有利于增强系统稳定性. 由图7可见, 附加无功控制的引入使得转子的峰值电流由21.85 kA降为15.45 kA, 且大电流持续时间缩短, 故障后转子电流振荡得到抑制. 综上, 附加无功控制的引入改善了风电场的动态响应性能, 同时增强了系统稳定性.

5.4 算例 2: 适应性测试(Case 2: adaptation test)

算例2验证所提附加学习控制的适应能力. 假设在算例1的学习完成之后, 节点6的本地负荷减少25%. 在这种变化下, 算例1得到的附加无功控制器的反馈系数对于新的系统不再是最优的. 为优化控制性能, 本文用附加学习控制方法进行在线优化. 由图8可见在线学习在10次迭代内收敛. 图9–11比较了附加控制器更新前后系统的时域响应, 仿真情景为节点1发生持续100 ms的三相对地短路故障. 由图9–11可见, 进行自适应学习后, 风电场公共接入点电压在故障后快速恢复, 输出功率的振荡得到显著抑制, 风机转子电流的峰值有所降低, 故障后转子电流振荡得到明显抑制. 仿真结果验证了所提附加学习控制具有在线自适应能力.

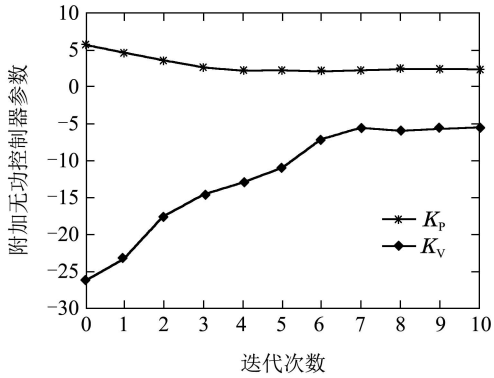

图 8 附加无功控制器的反馈系数: 算例2

Fig. 8 Feedback gain of supplementary controller: case 2

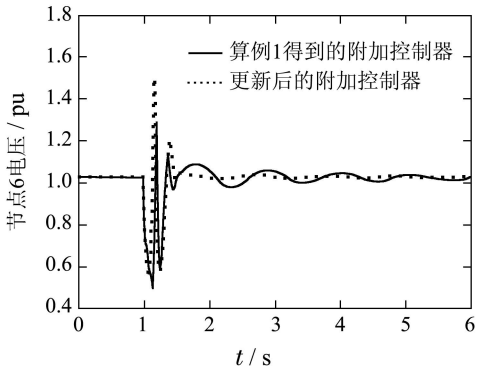

图 9 节点6电压: 算例2

Fig. 9 Voltage of bus 6: case 2

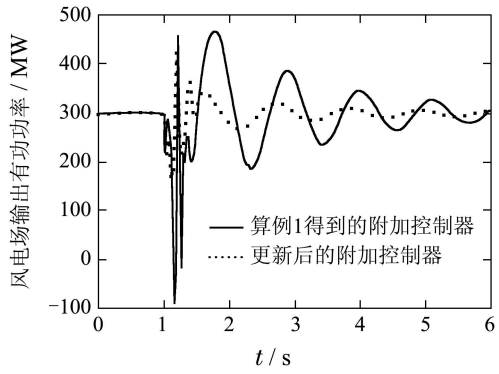

图 10 风电场输出有功功率: 算例2

Fig. 10 Output active power of wind farm: case 2

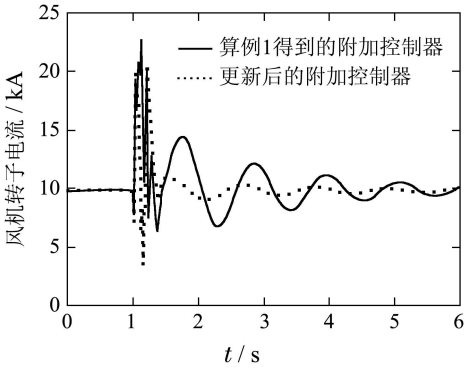

图 11 风机转子电流: 算例2

Fig. 11 Rotor current of DFIG: case 2

6 结论(Conclusions)

本文提出一种基于近似动态规划的附加学习控制设计方法, 不同于用近似动态规划控制器代替原控制器的通常做法, 该方法一方面保留了原控制器的先验知识, 另一方面利用附加控制器的在线学习能力优化闭环系统的控制性能. 所提方法具有: 1) 稳定性, 即学习过程不会导致系统失稳; 2) 收敛性, 学习最终单调收敛到最优附加控制器, 且由于策略迭代的高策略搜索效率和最小二乘的高样本利用效率, 算法收敛速度快; 3) 无需被控系统的数学模型, 这意味着附加学习控制器能对系统模型的改变进行自适应优化. 以上性质使得所提方法适合于在线优化控制器性能和在线适应不确定性, 在可再生能源大规模接入条件下的智能电网控制[30]等领域具有很好的工程应用前景.

参考文献(References):

[1] 郑应平. 钱学森与控制论 [J]. 中国工程科学, 2001, 3(10): 7 – 12.
(ZHENG Yingping. Qian Xuesen and engineering cybernetics [J]. *Engineering Science*, 2001, 3(10): 7 – 12.)

[2] 卢强, 梅生伟. 现代电力系统控制评述—清华大学电力系统国家重点实验室相关科研工作缩影及展望 [J]. 系统科学与数学, 2012, 32(10): 1207 – 1225.
(LU Qiang, MEI Shengwei. Review of modern power system control–epitome and prospect of related research in state key lab of power systems, Tsinghua University [J]. *Journal of Systems Science and Mathematical Sciences*. 2012, 32(10): 1207 – 1225)

[3] KNOSPE C. PID control [J]. *IEEE Control Systems Magazine*, 2006, 26(1): 30 – 31.

[4] LEWIS F L, BRABIE D, SYRMOS V L. *Optimal Control* [M]. Hoboken, NJ: John Wiley and Sons, 2012.

[5] LU Q, SUN Y, MEI S. *Nonlinear Control Systems and Power System Dynamics* [M]. Norwell, MA: Kluwer, 2001.

[6] ZHOU K, DOYLE J C. *Essentials of Robust Control* [M]. Upper Saddle River, NJ: Prentice Hall, 1998.

[7] TAO G. *Adaptive Control Design and Analysis* [M]. Hoboken, NJ: John Wiley and Sons, 2003.

[8] SI J, BARTO A G, POWELL W B, et al. *Handbook of Learning and Approximate Dynamic Programming: Scaling up to the Real World* [M]. Piscataway, NJ: Wiley-IEEE Press, 2004.

[9] LEWIS F L, VRABIE D, VAMVOUDAKIS K G. Reinforcement learning and feedback control: using natural decision methods to design optimal adaptive controllers [J]. *IEEE Control Systems Magazine*, 2012, 32(6): 76 – 105.

[10] WERBOS P J. ADP: the key direction for future research in intelligent control and understanding brain intelligence [J]. *IEEE Systems, Man, and Cybernetics, Part B*, 2008, 38(4): 898 – 900.

[11] WANG F Y, ZHANG H, LIU D. Adaptive dynamic programming: an introduction [J]. *IEEE Computational Intelligence Magazine*, 2009, 4(2): 39 – 47.

[12] BERTSEKAS D P, TSITSKKLIS J N. *Neura-Dynamic Programming* [M]. Belmont, MA: Athena Scientific, 1996.

[13] SUTTON R S, BARTO A G. *Reinforcement Learning: An Introduction* [M]. Cambridge, MA: MIT Press, 1998.

[14] PROKHOROV D V, WUNSCH D C. Adaptive critic designs [J]. *IEEE Transactions on Neural Networks*, 1997, 8(5): 997 – 1007.

[15] GUO W, LIU F, SI J, et al. Incorporating approximate dynamic programming-based parameter tuning into PD-type virtual inertia control of DFIGs [C] //*Proceedings of 2013 International Joint Conference on Neural Networks*. Piscataway, NJ: IEEE. 2013: 1 – 8.

[16] GUO W, LIU F, SI J, et al. Online adaptation of controller parameters based on approximate dynamic programming [C] //*Proceedings of 2014 International Joint Conference on Neural Networks*. Piscataway, NJ: IEEE. 2014: 256 – 262.

[17] GUO W, SI J, LIU F, et al. Policy iteration approximate dynamic programming using volterra series based actor [C] //*Proceedings of 2014 International Joint Conference on Neural Networks*. Piscataway, NJ: IEEE. 2014: 249 – 255.

[18] GUO W, LIU F, MEI S, et al. Approximate dynamic programming based supplementary frequency control of thermal generators in power systems with large-scale renewable generation integration [C] //*Proceedings of 2014 IEEE PES General Meeting*. Piscataway, NJ: IEEE. 2014: 1 – 5.

[19] 魏韡, 梅生伟, 张雪敏. 先进控制理论在电力系统中的应用: 综述及展望 [J]. 电力系统保护与控制, 2013, 41(12): 143 – 153.
(WEI Wei, MEI Shengwei, ZHANG Xuemin. Review of advanced control theory and application in power system in power system protection and control [J]. *Power System Protection and Control*, 2013, 41(12): 143 – 153.)

[20] BELLMAN R E. *Dynamic Programming* [M]. Princeton, NJ: Princeton University Press, 1957.

[21] HOWARD R A. *Dynamic Programming and Markov Processes* [M]. Cambridge, MA: MIT Press, 1960.

[22] BEARD R W, SARIDIS G N, WEN J T. Galerkin approximations of the generalized Hamilton-Jacobi-Bellman equation [J]. *Automatica*, 1997, 33(12): 2159 – 2177.

[23] KATSIKIS V N, PAPPAS D, PETRALIAS A. An improved method for the computation of the Moore-Penrose inverse matrix [J]. *Applied Mathematics and Computation*, 2011, 217(23): 9828 – 9834.

[24] GEIST M, PIETQUIN O. Algorithmic survey of parametric value function approximation [J]. *IEEE Transactions on Neural Network Learning Systems*, 2013, 24(6): 845 – 867.

[25] JIANG S, ANNAKKAGE U D, GOLE A M. A platform for validation of FACTS models [J]. *IEEE Transactions on Power Delivery*, 2006, 21(1): 484 – 491.

[26] QIAO W, HARLEY R G, VENAYAGAMOORTHY G K. Coordinated reactive power control of a large wind farm and a STATCOM using heuristic dynamic programming [J]. *IEEE Transactions on Energy Conversion*, 2009, 24(2): 493 – 503.

[27] GUO W, LIU F, HE D, et al. Reactive power control of DFIG wind farm using online supplementary learning controller based on approximate dynamic programming [C] //*Proceedings of 2014 International Joint Conference on Neural Networks*. Piscataway, NJ: IEEE, 2014: 1453 – 1460.

[28] FOSTER S, XU L, FOX B. Coordinated reactive power control for facilitating fault ride through of doubly fed induction generator- and fixed speed induction generator-based wind farms [J]. *IET Renewable Power Generation*, 2010, 4(2): 128 – 138.

[29] KAY1Kć1 M, MILANOCIć J V. Reactive power control strategies for DFIG-based plants [J]. *IEEE Transactions on Energy Conversion*, 2007, 22(2): 389 – 396.

[30] 梅生伟, 朱建全. 智能电网中的若干数学与控制科学问题及其展望 [J]. 自动化学报, 2013, 39(2): 119 – 131.
(MEI Shengwei, ZHU Jianquan. Mathematical and control scientific issues of smart grid and its prospects [J]. *Acta Automatica Sinica*, 2013, 39(2): 119 – 131.)

作者简介:

郭文涛 (1987–), 男, 博士研究生, 研究方向为近似动态规划理论及其在智能电网控制中的应用, E-mail: gwt329@gmail.com;

梅生伟 (1964–), 男, 博士, 教授, 博士生导师, 长江学者, 研究方向为电力系统优化运行、非线性控制、复杂系统;

刘 锋 (1977–), 男, 博士, 副教授, 研究方向为电力系统稳定与控制、博弈论和机器学习理论在智能电网中的应用.