

基于自适应动态规划的多设定值跟踪控制方法

李晓理[†], 刘德馨, 贾超, 丁大伟

(北京科技大学 自动化学院, 北京 100083; 北京科技大学 钢铁流程先进控制教育部重点实验室, 北京 100083)

摘要: 针对一类典型的带有控制约束的非线性离散时间系统, 提出了一种基于自适应动态规划(adaptive dynamic programming, ADP)算法的多设定值跟踪控制方法, 并对其收敛性和稳定性做了严格分析. 在ADP迭代跟踪控制的基础上, 根据多模型控制的思想, 设置阶梯状的参考轨迹, 使得系统状态逐渐地跟踪到最终设定值, 保证了系统的稳定性, 极大地减小超调量, 加快了响应时间, 改善控制品质; 同时由于控制器约束的存在, 引入非二次型的性能指标函数, 使得控制量始终在有界的范围内变化. 最后对仿真结果进行了分析, 结果表明了此方法的可行性和有效性.

关键词: 非线性系统; 离散时间; 自适应动态规划; 多模型

中图分类号: TP273

文献标识码: A

Multiple set-points tracking control method based on adaptive dynamic programming

LI Xiao-li[†], LIU De-xin, JIA Chao, DING Da-wei

(School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China;
Key Laboratory of Advanced Control of Iron and Steel Process (Ministry of Education),
University of Science and Technology Beijing, Beijing 100083, China)

Abstract: To deal with a class of nonlinear discrete time systems with control constraints, we propose a multiple set-points tracking control strategy based on adaptive dynamic programming (ADP), and rigorously prove the convergence and stability of the system. According to the idea of multiple-model control, a ladder-shaped reference trajectory is designed. In the entire control process, system states track the final value of the reference trajectory step by step, ensuring the stability of the system and significantly reducing the overshoot and response-time. A non-quadratic performance functional is introduced to tackle the control constraints by consistently restricting the control inputs in a bounded range. Simulation results are shown and analyzed, demonstrating the feasibility and validity of the proposed optimal tracking control scheme.

Key words: nonlinear systems; discrete time; adaptive dynamic programming (ADP); multiple-model

1 引言(Introduction)

自适应动态规划(adaptive dynamic programming, ADP)的本质是根据强化学习原理, 模拟人脑通过环境反馈进行学习, 被认为是一种非常接近人脑智能的方法^[1]. 其主要原理是利用函数近似结构, 如神经网络, 来逼近动态规划中的性能指标函数和控制策略, 从而满足最优性原理. 自适应动态规划有效地解决了传统动态规划计算中“维数灾”的难题, 为求解非线性离散系统最优控制提供了一种切实可行的理论和方法. 自1991年Werbos P J^[2]首次提出ADP算法的概念以来, 这种方法受到国内外专家学者的广泛关注, 并取得了丰硕的成果. Al-Tamimi A, Lewis F L等提出了离散系统的ADP算法, 并证明了迭代控制的稳定性和迭代性能指标的收敛性^[3-4]. Abu-Khalaf,

Ferrari S等讨论了带有控制器约束的ADP算法^[5-6]. Zhang H G等用ADP算法成功解决了最优跟踪控制问题^[7]. Azhmyakov V等将动态规划与鲁棒最优控制结合起来, 开辟了新的理论方向^[8].

由于非线性系统的复杂性和多样性, 其控制问题一直是控制理论中的难点, 人们往往期望控制系统满足两方面要求: 首先, 在保证系统稳定性的情况下, 输入和状态有界, 并且输出值能够精确地跟踪期望设定值. 其二, 提高控制性能, 在减小超调量的同时, 缩短调节时间. 这在实际控制中很难达到. 带控制器约束的ADP算法能够把控制量的波动限定在一个范围内, 保证了系统的稳定性, 但却常常导致很大的超调量, 控制效果不好.

近几年来, 多模型控制作为强有力的非线性控制

收稿日期: 2012-10-18; 收修稿日期: 2013-01-28.

[†]通信作者. E-mail: lixiaoli@hotmail.com.

基金项目: 新世纪优秀人才支持计划资助项目(NCET-11-0578); 国家自然科学基金资助项目(61074055); 中央高校基本科研业务费专项资金资助项目(FRF-TP-12-005B).

方法成为了控制工程界的研究热点,其思想是利用多个简单模型逼近非线性系统的不确定性,再根据多个简单模型设计多个控制器分而治之,进而构成多模型控制算法,从而获得更好的控制品质^[9].

本文针对一类典型的非线性离散时间系统,把ADP跟踪控制算法和多模型控制器设计思想相结合,建立多个阶梯状的参考设定值,设计控制器使得系统输出阶梯状地接近最终跟踪期望设定值,既保证系统的稳定性,又极大的减小了超调量.同时,由于控制器约束的存在,控制量始终保持有界.仿真实验的结果表明,此方法能有效的改善系统的瞬态响应,并把控制量的波动限定在一个很小的范围内,快速得到最优控制律.

2 基于ADP算法的带有控制器约束的非线性离散时间系统跟踪控制(ADP algorithm based tracking control of nonlinear discrete-time system with control constraints)

2.1 问题描述(Problem formulation)

考虑如下一类带有控制器约束的仿射非线性离散系统:

$$x(k+1) = f(x(k)) + g(x(k))u(k), \quad k \geq 0, \quad (1)$$

其中: $x(k) \in \mathbb{R}^n$ 是状态向量, $f(x(k)) \in \mathbb{R}^n$, $g(x(k)) \in \mathbb{R}^{n \times m}$ 是Lipschitz连续函数. 假设系统在集合 $\Omega \subset \mathbb{R}^n$ 上是可控的, 且 $\Omega_u = \{u(k) | u(k) = [u_1(k), u_2(k), \dots, u_m(k)]^T \in \mathbb{R}^m\}$.

定义如下状态误差 $e(k)$:

$$e(k) = x(k) - \eta(k), \quad (2)$$

其中 $\eta(k)$ 是参考轨迹, 且 $\eta(k) \in \mathbb{R}^n$.

给出稳态输入控制 $u_e(k) = g^{-1}(\eta(k))(\eta(k+1) - f(\eta(k)))$, 且 $g^{-1}(\cdot)$ 为 $g(\cdot)$ 的伪逆矩阵. 考虑 $v(k) = u(k) - u_e(k)$, 则有

$$\begin{aligned} e(k+1) &= x(k+1) - \eta(k+1) = \\ &= f(x(k)) + g(x(k))u(k) - \eta(k+1) = \\ &= f(e(k) + \eta(k)) + g(e(k) + \eta(k)) \\ &\quad (v(k) + u_e(k)) - \eta(k+1). \end{aligned} \quad (3)$$

设定 $|v(k)| \leq \bar{v}$, $\bar{v}$ 是控制器约束的饱和界, 且定义 $\bar{U} = \text{diag}\{\bar{v}_1, \bar{v}_2, \dots, \bar{v}_m\}$.

对于带有控制器约束的离散系统最优跟踪控制问题, 其目的是寻找最优控制律, 使得系统(3)渐近稳定, 并且能够使性能指标函数最小. 为了求解最优跟踪控制问题, 下面的假设是必须的:

假设 1 对于系统(3), 存在控制 $v(k)$, 满足

1) $v(k)$ 在 Ω_v 域上是连续的, 如果 $e(k) = 0$, 那么

$v(k) = 0$;

2) $v(k)$ 稳定系统(3).

令性能指标函数为

$$\begin{aligned} J(e(k), v(k)) &= \\ &= \sum_{n=k}^{\infty} \{Q(e(n)) + W(v(n))\} = \\ &= Q(e(k)) + W(v(k)) + \\ &= \sum_{n=k+1}^{\infty} \{Q(e(n)) + W(v(n))\} = \\ &= Q(e(k)) + W(v(k)) + \\ &= J(e(k+1), v(k+1)), \end{aligned} \quad (4)$$

其中 $Q(e(k))$ 和 $W(v(k))$ 是正定函数, 取

$$\begin{cases} Q(e(k)) = e^T(k)Qe(k), \\ W(v(k)) = 2 \int_0^{v(k)} \varphi^{-T}(\bar{U}^{-1}s)\bar{U}Rds. \end{cases} \quad (5)$$

这里 Q 和 R 取对角正定矩阵, $s \in \mathbb{R}^m$, $\varphi \in \mathbb{R}^m$. 定义 $\varphi^{-T} = (\varphi^{-1})^T$, 且

$$\begin{aligned} \varphi^{-1}(v(k)) &= \\ &= [\varphi^{-1}(v_1(k)) \quad \varphi^{-1}(v_2(k)) \quad \dots \quad \varphi^{-1}(v_m(k))]^T, \end{aligned}$$

$\varphi(\cdot)$ 是一个有界的单调递增奇函数, 满足 $|\phi(\cdot)| \leq 1$, 其一阶导数的界是一个常数. 例如, 双曲正切函数 $\varphi(\cdot) = \tanh(\cdot)$. $\varphi^{-1}(\cdot)$ 表示 $\varphi(\cdot)$ 的反函数, 通过上面的定义, $\varphi^{-1}(\cdot)$ 为单调的奇函数且 R 是正定的, 所以 $W(v(k))$ 也是正定的^[10].

令 $J^*(e(k))$ 代表最优性能指标函数, 根据贝尔曼最优性原理, $J^*(e(k))$ 满足下面的HJB方程:

$$\begin{aligned} J^*(e(k)) &= \min_{v(k)} \{Q(e(k)) + W(v(k)) + \\ &= J^*(e(k+1))\}. \end{aligned} \quad (6)$$

最优控制律 $v^*(k)$ 满足

$$\begin{aligned} v^*(k) &= \arg \min_{v(k)} \{Q(e(k)) + W(v(k)) + \\ &= J^*(e(k+1))\}. \end{aligned} \quad (7)$$

令 $J^*(e(k))$ 对 $v(k)$ 的偏导数为零:

$$\begin{aligned} \frac{\partial J^*(e(k))}{\partial v(k)} &= \\ &= 2\bar{U}R\varphi^{-1}(\bar{U}^{-1}v(k)) + \frac{\partial e(k+1)}{\partial v(k)} \times \frac{\partial J^*(e(k+1))}{\partial e(k+1)} = \\ &= 2\bar{U}R\varphi^{-1}(\bar{U}^{-1}v(k)) + g^T(e(k)) \frac{\partial J^*(e(k+1))}{\partial e(k+1)} = 0. \end{aligned}$$

求得最优控制律 $v^*(k)$:

$$v^*(k) = \bar{U}\varphi\left(-\frac{1}{2}(\bar{U}R)^{-1}g^T(e(k))\frac{\partial J^*(e(k+1))}{\partial e(k+1)}\right). \quad (8)$$

综上所述, 带有控制器约束的非线性离散系统跟

踪控制问题, 转化成为求解最优性能指标 $J^*(e(k))$ 的问题. 本文用迭代ADP算法来求解.

2.2 ADP算法的推导(Derivation of ADP algorithm)

ADP的原理是通过利用函数近似结构神经网络, 来近似动态规划中的性能指标函数 $J(e(k))$, 从而获得最优性能指标函数 $J^*(e(k))$ 和最优控制律 $v^*(k)$ 以满足最优性原理.

首先, 对于任意给定的初始状态误差 $e(k)$, 令初始的性能指标函数 $J_0(\cdot) = 0$, 注意到这个初始性能指标函数不一定是最优的. 然后求得相应的控制律为

$$v_0(k) = \arg \min_{v(k)} \{Q(e(k)) + W(v(k)) + J_0(e(k+1))\}. \quad (9)$$

继续更新性能指标函数

$$J_1(e(k)) = Q(e(k)) + W(v_0(k)) + J_0(e(k+1)), \quad (10)$$

其中

$$e(k+1) = f(e(k) + \eta(k)) + g(e(k) + \eta(k))(v_0(k) + u_e(k)) - \eta(k+1). \quad (11)$$

然后, 对于 $i=1, 2, \dots$, ADP算法在 $v_i(k)$ 和 $J_{i+1}(e(k))$ 之间反复迭代, 可得到如下迭代公式:

$$v_i(k) = \arg \min_{v(k)} \{Q(e(k)) + W(v(k)) + J_i(e(k+1))\}, \quad (12)$$

$$J_{i+1}(e(k)) = Q(e(k)) + W(v_i(k)) + J_i(e(k+1)), \quad (13)$$

$$e(k+1) = f(e(k) + \eta(k)) + g(e(k) + \eta(k))(v_i(k) + u_e(k)) - \eta(k+1). \quad (14)$$

引理 1^[2] 对方程(6)–(7), 当 i 趋近于 ∞ 时, $v_i(k)$ 收敛到最优控制律 $v^*(k)$, 性能指标函数 $J_{i+1}(e(k))$ 收敛到离散HJB方程的最优解 $J^*(e^*(k))$, 且满足性能指标函数 $J^*(e^*(k))$ 是正定的.

定理 1^[11] 由引理1, 最优控制律 $v^*(k)$ 使得系统(1)渐近稳定.

证 最优性能指标 $J^*(e^*(k))$ 表达成式(6)的形式为

$$J^*(e^*(k)) = Q(e^*(k)) + W(v^*(k)) + J(e^*(k+1)),$$

则可得

$$J^*(e^*(k+1)) - J^*(e^*(k)) = -\{Q(e^*(k)) + W(v^*(k))\} \leq 0. \quad (15)$$

根据Lyapunov函数的定义, $J^*(e^*(k))$ 为Lyapunov函数, 系统是渐近稳定的.

2.3 ADP算法的神经网络实现(Neural networks implementation of ADP algorithm)

为了便于实现ADP算法, 采用径向基函数(radial basis function, RBF)神经网络来近似 $J_i(e(k))$ 和 $v_i(k)$. 假设神经网络隐含层神经元的个数是 l , 输入层和隐含层之间的权值矩阵是 V , 隐含层和输出层之间的权值矩阵是 W , 隐含层激活函数采用高斯函数 $\phi(\cdot) \in \mathbb{R}^l$, 输出层采用线性激活函数. 在本文中, 需要两个神经网络, 它们分别是评判网络和执行网络.

2.3.1 评判网络(Critic network)

评判网络被用来近似性能指标函数, 其输入层和隐含层之间的权值矩阵是 V_c , 隐含层和输出层之间的权值矩阵是 W_c , 输入为 $e(k)$, 输出表示成

$$\hat{J}_i(e(k)) = (W_{c_i}(k))^T \phi((V_{c_i}(k))^T e(k)). \quad (16)$$

目标函数为

$$J_i(e(k)) = Q(e(k)) + W(v_{i-1}(k)) + \hat{J}_{i-1}(e(k+1)). \quad (17)$$

定义评判网络的误差函数

$$ec_i(k) = \hat{J}_i(e(k)) - J_i(e(k)). \quad (18)$$

评判网络的目标是使得下面函数极小化:

$$Ec_i(k) = \frac{1}{2}(ec_i(k))^2. \quad (19)$$

因此根据梯度下降法, 评判网络的权值更新法则表示为

$$\begin{cases} W_{c_{i+1}}(k) = W_{c_i}(k) + \Delta W_{c_i}(k), \\ \Delta W_{c_i}(k) = \alpha_c \left[-\frac{\partial Ec_i(k)}{\partial W_{c_i}(k)} \right], \\ \frac{\partial Ec_i(k)}{\partial W_{c_i}(k)} = \frac{\partial Ec_i(k)}{\partial \hat{J}_i(e(k))} \frac{\partial \hat{J}_i(e(k))}{\partial W_{c_i}(k)}, \end{cases} \quad (20)$$

其中 α_c 是评判网络的学习率.

2.3.2 执行网络(Action network)

执行网络用来得到控制律, 其输入层和隐含层之间的权值矩阵是 V_a , 隐含层和输出层之间的权值矩阵是 W_a , 输入为 $e(k)$, 输出表示成

$$\hat{v}_i(k) = (W_{a_i}(k))^T \phi((V_{a_i}(k))^T e(k)). \quad (21)$$

目标函数为

$$v_i(k) = \arg \min_{v(k)} \{Q(e(k)) + W(v(k)) + \hat{J}_i(e(k+1))\}. \quad (22)$$

定义执行网络的误差函数

$$ea_i(k) = \hat{v}_i(k) - v_i(k). \quad (23)$$

执行网络的目标是使得下面函数极小化:

$$Ea_i(k) = \frac{1}{2}(ea_i(k))^2. \quad (24)$$

执行网络的权值更新法则与评判相似. 根据梯度下降法, 可以得到

$$\begin{cases} Wa_{i+1}(k) = Wa_i(k) + \Delta Wa_i(k), \\ \Delta Wa_i(k) = \alpha_a \left[-\frac{\partial Ea_i(k)}{\partial Wa_i(k)} \right], \\ \frac{\partial Ea_i(k)}{\partial Wa_i(k)} = \frac{\partial Ea_i(k)}{\partial \hat{v}_i(k)} \frac{\partial \hat{v}_i(k)}{\partial Wa_i(k)}, \end{cases} \quad (25)$$

其中 α_a 是执行网络的学习率.

给出初始状态误差, 执行网络、评判网络1和评判网络2的初始权值, 给定迭代上限 $i_{\max}$, 对于第 i 次迭代, 算法流程如下:

1) 根据式(21), 由执行网络得到 $\hat{v}_i(k)$; 再根据式(16), 由评判网络2得到 $\hat{J}_i(e(k))$.

2) 用 $\hat{v}_i(k)$ 代替 $v_i(k)$, 把 $\hat{v}_i(k)$ 和 $e(k)$ 代入系统方程(14), 得到下一时刻的状态误差 $e(k+1)$;

3) 再根据式(16), 由评判网络1得到 $\hat{J}_i(e(k+1))$, 并反馈给执行网络, 根据式(25)调整权值, 更新为 $Wa_{i+1}(k)$;

4) 根据式(20)调整评判网络2的权值, 更新为 $Wc_{i+1}(k)$, 并赋值给评判网络1.

第 i 次迭代计算完毕, $i = i + 1$, 若 $i \leq i_{\max}$ 返回步骤1). 否则迭代结束.

整个迭代算法的结构如图1所示.

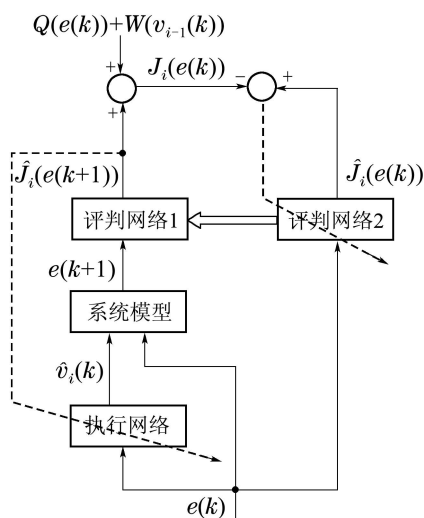


图1 ADP算法结构图

Fig. 1 The structure diagram of the ADP algorithm

3 基于ADP算法的多设定值控制器设计 (Design of multiple set-points controller based on ADP)

对于仿射非线性离散系统, 设计稳定的、具有很好控制品质的控制器往往很难做到. ADP方法把神经网络与动态规划方法相结合, 设计非线性仿射系统的控制器, 可以保证在输入有界的情况下, 使得被控对象的状态跟踪设定值. 但是控制器不能保证系统的动态品质, 尤其是瞬态过程, 常常导致超调很大.

将ADP算法与多模型控制相结合, 在控制量有界的前提下设计多设定值跟踪控制器, 使被控对象的状态从一个设定值跟踪到另一个设定值, 逐渐的接近最终的跟踪目标. 这种思想, 席裕庚教授在1996年的非线性系统预测控制的多模型方法里就曾经提出^[12], 后在相应的文献里又多次出现在实际应用当中. 这种方法是将非线性被控对象的输出设定值阶梯化, 再对不同的设定值其附近寻找系统的工作平衡点, 并在平衡点处将非线性系统线性化, 据此设计多个线性模型, 然后基于多个线性模型设计控制器, 阶段性的跟踪阶梯状的参考设定值, 最终跟踪到期望设定值. 但由于线性模型的逼近误差, 多线性模型控制器的切换常常导致系统不稳定, 并且对被控对象的输入输出特性有一定要求.

本文保留了多模型控制阶梯状设定值的思路, 考虑到ADP控制方法输入有界的优良性质, 不会由于设定值的微小变化使得输出有大的波动. 采用ADP控制器, 阶梯式的、渐近地跟踪设定值, 减少超调, 改善控制品质.

考虑下面的带有控制器约束的非线性离散时间系统:

$$x(k+1) = f(x(k)) + g(x(k))u(k), \quad k \geq 0,$$

其中:

$$f(x) = \begin{bmatrix} -0.8x_2(k) \\ \sin(0.8x_1(k) - x_2(k)) + 1.8x_2(k) \end{bmatrix},$$

$$g(x) = \begin{bmatrix} 0 \\ -x_2(k) \end{bmatrix}.$$

设控制器约束为 $\bar{v} = 0.2$, 取式(5)中 $Q = I_2$, $R = 0.5$, 初始状态 $x(0) = \begin{bmatrix} x_1(0) \\ x_2(0) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$.

选择评判网络和执行网络都为2个输入、8个隐层神经元和1个输出的2-8-1结构模型, 初始权值均设置为 $[-0.1, 0.1]$, 学习率 $\alpha_c = \alpha_a = 0.1$, 每次迭代循环1000次, 设状态 $x_1(k)$ 的参考轨迹 $\eta_1(k)$ 是在-0.16和0之间波动的矩形波, 状态 $x_2(k)$ 的参考轨迹 $\eta_2(k)$

是在0.2和0之间波动的矩形波, 则其状态轨迹曲线如图2所示。

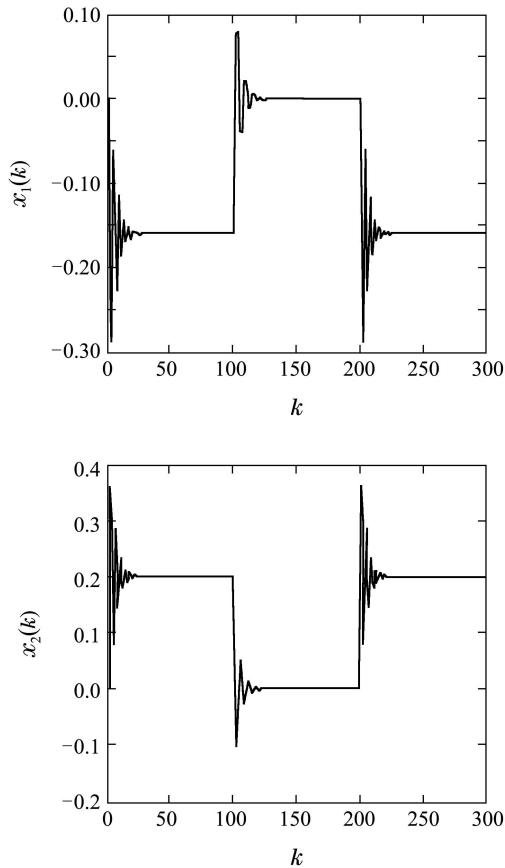


图2 状态跟踪曲线

Fig. 2 The state tracking curves

由图2可看出, 对于设定的跟踪, 超调量还是比较大的, 接近100%, 且稳定时间较长. 采用多模型控制的思想^[13], 设置阶梯状的参考轨迹, 使得系统状态 $x(k)$ 阶梯状地跟踪设定值, 最终渐近地达到稳态.

首先, 建立2个设定值, 即 $x_1(k)$ 先跟踪-0.08, 而后跟踪到-0.16; 然后先跟踪-0.08, 再跟踪到0; 即曲线设定值 $0 \sim -0.08 \sim -0.16$ 之间过渡. 同理, $x_2(k)$ 在 $0 \sim 0.1 \sim 0.2$ 之间过渡. 则状态曲线图如图3所示.

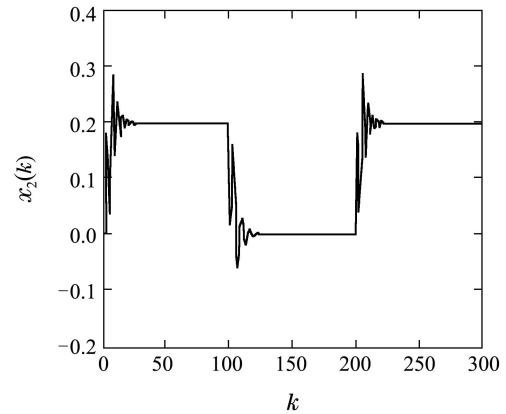
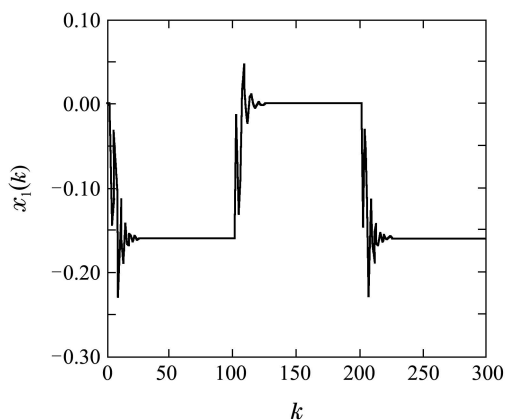


图3 2个阶梯形设定值的状态跟踪曲线

Fig. 3 The state tracking curves of 2 steps set-points

若建立3个设定值, $x_1(k)$ 在 $0 \sim -0.05 \sim -0.11 \sim -0.16$ 之间过渡. $x_2(k)$ 在 $0 \sim 0.07 \sim 0.14 \sim 0.2$ 之间过渡. 则状态曲线如图4所示.

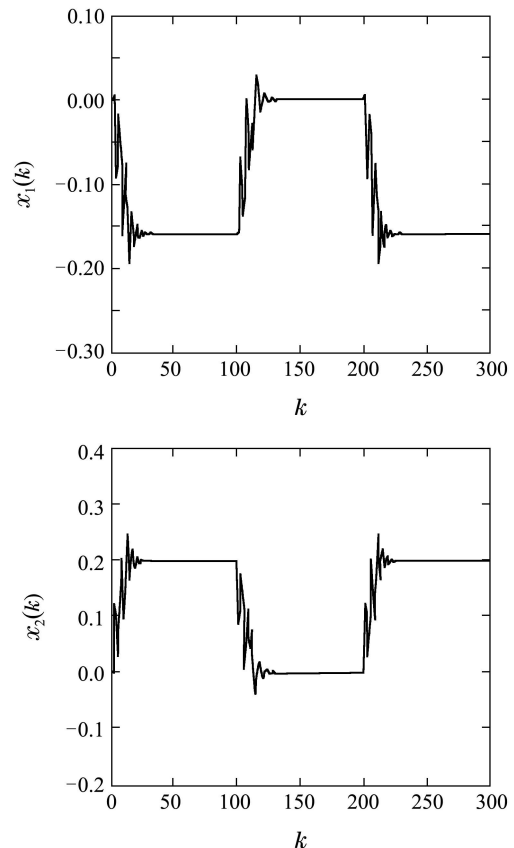


图4 3个阶梯形设定值的状态跟踪曲线

Fig. 4 The state tracking curves of 3 steps set-points

对比图2至图4可看出, 随着设定值的增加, 在调节时间相差不大的情况下, 状态轨迹曲线的超调量越来越小, 跟踪的更平稳, 系统也越稳定.

图5为分别采用1个、2个、3个设定值时, 控制量 $u(k)$ 的轨迹, 从图中可以看出, 由于控制器约束的存在, 实际控制量与稳态控制之间的差值 $v(k)$ 的幅值始终被限制在 $[-0.2 \ 0.2]$ 之间, 控制输入不会有大的波动, 保证了系统的稳定性.

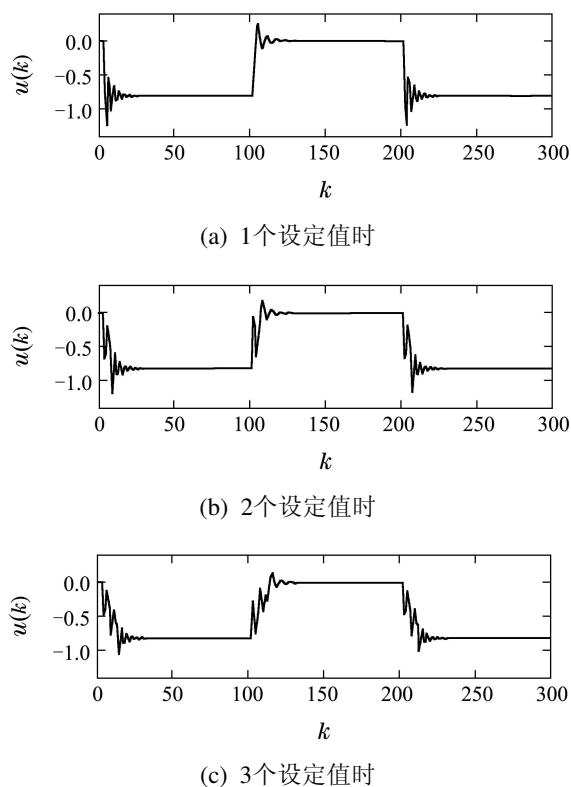


图5 控制量对比图

Fig. 5 The comparison of controlling quantities

在状态的初始值和最终期望跟踪的设定值 $\eta(k)$ 之间取 N 个中间设定值,设计阶梯状的参考轨迹,通常要考虑以下两个问题:若 N 的选取过大,即中间设定值过多,可以减少超调,但会使得系统的稳定时间过长,同时输出波动较大;若 N 过小,中间设定值过少,第 $N-1$ 个设定值与最终设定值 $\eta(k)$ 之间有较大的距离,仍然会出现超调较大的结果.因此,设计多个设定值, N 的选取往往针对以上两个因素折中处理,通常 N 取值2或3,也可适当增大.对于响应较快的控制系统,可在状态的初始值和最终设定值之间等间隔的取多个中间设定值,防止状态值在短时间内增长过快;而对于响应较慢的控制系统,可在最终设定值的60%~80%处和最终的设定值之间设立多个中间设定值.

4 化工过程连续搅拌釜的ADP控制(ADP control of continuous stirred tank reactor)

化工过程连续搅拌釜(continuous stirred tank reactor, CSTR)是关联严重、非线性强的控制对象.被控对象的变量包括生成物的浓度、反应器的温度和反应器冷却装置的温度等等,十分复杂.

对于CSTR反应过程的控制是一个典型的过程控制问题,很多文献都提出过针对此类系统的控制器设计方法,其中就有多模型控制^[14-15].首先求解系统的平衡点曲线(近似为一个S型的曲线),在平衡点

附近建立多个线性模型,并在线性模型附近建立过渡型的设定值,使得系统状态阶梯式的上升.但由于线性模型与系统模型存在误差,模型的多少,模型的选取,控制系统的鲁棒性,控制器的切换过程,或控制量的跳变都会导致控制效果不好.

本文将具有控制器约束的ADP控制应用于CSTR的控制,即采用阶梯式设定值,同时只是针对原非线性系统设计ADP控制器,由于控制输入的上限约束,不会带来多模控制切换过程中的输入值大的跳变,保证了控制系统的稳定性,同时控制品质得到提高.

CSTR反应过程可由如下的离散时间状态方程和输出方程来描述:

$$\begin{aligned} x_1(k+1) &= 0.8x_1(k) + 0.2D_a(1-x_1(k)) \exp\left[\frac{\gamma x_2(k)}{\gamma + x_2(k)}\right], \\ x_2(k+1) &= 0.8x_1(k) + 0.2BD_a(1-x_1(k)) \exp\left[\frac{\gamma x_2(k)}{\gamma + x_2(k)}\right] + 0.2\beta D_a(u(k) - x_2(k)), \\ y(k) &= x_2(k), \end{aligned}$$

其中:状态变量 $x_1(k)$ 为生成物的浓度, $x_2(k)$ 为反应器的温度,输入量 $u(k)$ 为反应器冷却装置的温度.在一定的反应条件下,取常量参数 $D_a=0.072$, $\gamma=20$, $B=8$, $\beta=0.3$.

化成仿射形式

$$x(k+1) = f(x(k)) + g(x(k))u(k), \quad k \geq 0,$$

其中:

$$\begin{aligned} f(x(k)) &= \begin{bmatrix} 0.8x_1(k) + 0.0144(1-x_1(k)) \exp\left[\frac{\gamma x_2(k)}{\gamma + x_2(k)}\right] \\ 0.74x_2(k) + 0.1152(1-x_1(k)) \exp\left[\frac{\gamma x_2(k)}{\gamma + x_2(k)}\right] \end{bmatrix}, \\ g(x(k)) &= \begin{bmatrix} 0 \\ 0.06 \end{bmatrix}. \end{aligned}$$

假设抗饱和和控制器约束为 $\bar{v}=5$.取式(5)中, $Q=I_2$, $R=0.5$;初始状态

$$x(0) = \begin{bmatrix} x_1(0) \\ x_2(0) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

选择评判网络和执行网络都为2个输入,8个隐层神经元和1个输出的2-8-1结构模型,初始权值均设置为 $[-0.1, 0.1]$,学习率 $\alpha_c = \alpha_a = 0.1$,每次迭代循环1000次.

首先考虑单设定值的情况,系统直接跟踪到 $y(k) = x_2(k) = 5.5$,既 $\eta(k) = 5.5$, $0 \leq k \leq 100$,经过 $k=100$ 步后,其状态轨迹曲线如图6所示.

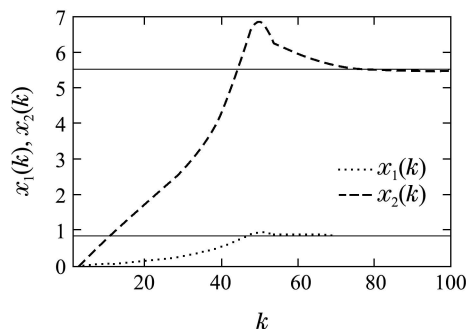


图6 单设定值状态跟踪曲线

Fig. 6 The state tracking curves of one set-points

可看出, 系统的响应时间较长, 取4.5(最终设定值的80%)为第一个设定值, 之后在4.5和最终的设定值5.5之间, 设立阶梯状的参考轨迹, 即取式(2)中的 $\eta(k)$ 为

$$\eta(k) = \begin{cases} 4.5, & 0 \leq k < 30, \\ 5, & 31 \leq k < 60, \\ 5.5, & k \geq 61. \end{cases}$$

使得输出逐渐跟踪上最终值5.5, 其状态轨迹曲线如图7所示。

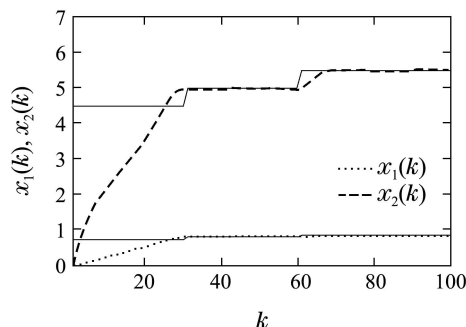


图7 多设定值状态跟踪曲线

Fig. 7 The state tracking curves of multiple set-points

两图比较可看出, 对于输出量 $y(k) = x_2(k)$, 单设定值控制方法需要的调节时间约为75步, 多设定值控制需要70步; 在超调量方面, 单设定值控制的峰值为6.5, 多设定值控制仅在跟踪第一个阶梯时出现了0.5的超调, 就全局而言, 输出量没有超过最终值5.5. 因此, 在调节时间相差不大的情况下, 多设定值控制能大大减小系统的超调, 实现更平稳的跟踪。

图8是系统的控制律误差 $u(k)$ 的轨迹, 其中点线为单设定值控制律, 虚线为多设定值控制律. 由于控制器约束的存在, 二者的控制律误差 $v(k)$ 都在 $\bar{v} = 5$ 以内, 但多设定值控制的误差更小。

图9为式(4)中性能指标 J 的曲线, 同样地, 点线为单设定值控制作用下的结果, 虚线为多设定值控制作用下的结果, 可看出, 多设定值同样优于单设定值控制方法。

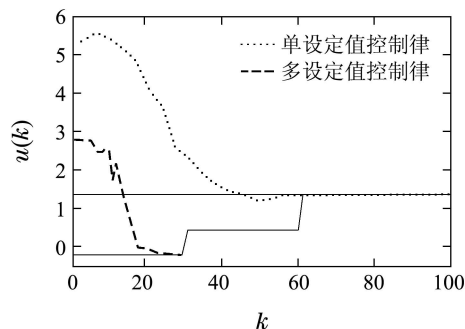


图8 控制量曲线

Fig. 8 The curves of controlling quantities

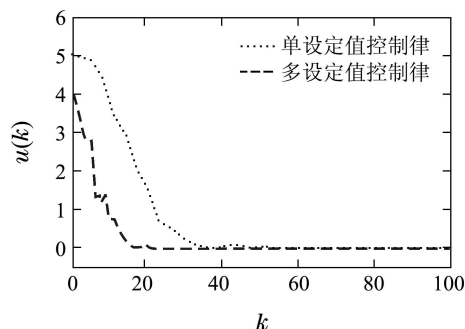


图9 性能指标函数曲线

Fig. 9 The curves of performance indexes

5 结论(Conclusions)

本文主要研究了基于ADP算法的非线性离散时间仿射系统的最优控制问题, 采用多模型控制中多设定值控制的思想, 使得系统状态的参考轨迹成阶梯状, 逐渐跟踪到最终设定值; 同时由于控制器约束的存在, 把控制量限定在合理的范围内. 仿真实验达到了期望的控制效果, 极大的提高了系统的动态响应特征和控制品质。

参考文献(References):

- [1] 赵冬斌, 刘德荣, 易建强. 基于自适应动态规划的城市交通信号优化控制方法综述 [J]. 自动化学报, 2009, 35(6): 676 – 681.
(ZHAO Dongbin, LIU Derong, YI Jianqiang. An overview on the adaptive dynamic programming based urban city traffic signal optimal control [J]. *Acta Automatica Sinica*, 2009, 35(6): 676 – 681.)
- [2] WERBOS P J. A menu of designs for reinforcement learning over time [M] // MILLER W T, SUTTON R S, WERBOS P J. *Neural Networks for Control*. Cambridge, USA: MIT Press, 1995.
- [3] AL-TAMIMI A, LEWIS F L. Discrete-time nonlinear HJB solution using approximate dynamic programming: convergence proof [C] // *Proceedings of IEEE International Symposium on Approximate Dynamic Programming and Reinforcement Learning*. Honolulu: IEEE, 2007: 240 – 247.
- [4] AL-TAMIMI A, LEWIS F L, ABU-KHALAF M. Discrete-time nonlinear HJB solution using approximate dynamic programming: convergence proof [J]. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 2008, 38(4): 943 – 949.
- [5] ABU-KHALAF M, LEWIS F L. Nearly optimal state feedback control of constrained nonlinear systems using a neural networks HJB approach [J]. *Annual Reviews in Control*, 2004, 28(2): 239 – 251.

- [6] FERRARI S, STECK J E, CHANDRAMOHAN R. Adaptive feed-back control by constrained approximate dynamic programming [J]. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 2008, 38(4): 982 – 987.
- [7] ZHANG H G, LUO Y H, LIU D R. Neural-network-based near-optimal control for a class of discrete-time affine nonlinear systems with control constraints [J]. *IEEE Transactions on Neural Networks*, 2009, 20(9): 1490 – 1503.
- [8] AZHMYAKOV V, BOLTYANSKI V G, POZNYAK A S. On the dynamic programming approach to multi-model robust optimal control problems [C] // *Proceedings of IEEE International Symposium on American Control Conference*. Washington: IEEE, 2008: 4468 – 4473.
- [9] 王伟, 李晓理. 多模型自适应控制 [M]. 北京: 科学出版社, 2001. (WANG Wei, LI Xiaoli. *Multiple Model Adaptive Control* [M]. Beijing: Science Press, 2001.)
- [10] LYSHEVSKI S E. Nonlinear discrete-time systems: Constrained optimization and application of nonquadratic costs [C] // *Proceedings of IEEE International Symposium on American Control Conference*. Philadelphia: IEEE, 1998: 3699 – 3703.
- [11] ZHANG H G, SONG R Z, WEI Q L, et, al. Optimal tracking control for a class of nonlinear discrete-time systems with time delays based on heuristic dynamic programming [J]. *IEEE Transactions on Neural Networks*, 2011, 22(12): 1851 – 1862.
- [12] 席裕庚, 王凡. 非线性系统预测控制的多模型方法 [J]. 自动化学报, 1996, 22(4): 456 – 461. (XI Yugeng, WANG Fan. Nonlinear multi-model predictive control [J]. *Acta Automatica Sinica*, 1996, 22(4): 456 – 461.)
- [13] 李晓理, 石龙辉, 丁大伟. 水箱液位系统多模型控制方法 [J]. 控制理论与应用, 2011, 28(3): 370 – 374. (LI Xiaoli, SHI Longhui, DING Dawei. Multi-model control for water level system [J]. *Control Theory & Applications*, 2011, 28(3): 370 – 374.)
- [14] 孙维, 李晓理, 王伟. 基于多模型的非线性系统自适应最小方差控制 [J]. 控制理论与应用, 2002, 19(4): 815 – 819. (SUN Wei, LI Xiaoli, WANG Wei. Multiple model based adaptive minimum variance control of nonlinear system [J]. *Control Theory & Applications*, 2002, 19(4): 815 – 819.)
- [15] 李晓理, 任威, 胡广大. 连续搅拌釜的模糊多模型自适应控制 [J]. 系统仿真学报, 2009, 21(7): 1959 – 1961. (LI Xiaoli, REN Wei, HU Guangda. Multiple model adaptive control of CSTR [J]. *Journal of System Simulation*, 2009, 21(7): 1959 – 1961.)

作者简介:

李晓理 (1971–), 男, 教授, 博士生导师, 从事自适应控制和智能控制等研究, E-mail: lixiaoli@hotmail.com;

刘德馨 (1987–), 女, 博士研究生, 从事神经网络和智能控制等研究, E-mail: bjldxin@126.com;

贾超 (1988–), 男, 硕士研究生, 从事神经网络和智能控制等研究, E-mail: jiachaocharles@yahoo.com.cn;

丁大伟 (1980–), 男, 副教授, 从事鲁棒控制与切换控制等研究, E-mail: ddaweiauto@163.com.