

# 弹道导弹机动能力受限反拦截博弈弹道优化方法

游勇安,孙瑞胜\*,刘宣廷,陈伟  
(南京理工大学 能源与动力工程学院,江苏 南京 210094)

**摘要:**针对突防-拦截过程中存在的机动能力受限问题,提出了一种机动能力受限下基于自适应动态规划的博弈弹道优化策略求解方法。通过建立仿射非线性微分博弈模型,并考虑到机动能力受限,设计了含积分形式的控制能量项的性能指标函数;基于微分博弈理论推导了博弈鞍点控制策略,并基于自适应动态规划算法设计一个评价网络对微分博弈策略进行近似逼近求解,给出了神经网络权值自适应更新律,并对其稳定性进行了推导证明。仿真结果表明,所提出的策略求解方法能够在机动能力受限的情况下实现反拦截效果,精确打击敌方目标。

**关键词:**弹道导弹;机动能力受限;微分博弈;自适应动态规划

**中图分类号:**TJ765.3   **文献标志码:**A   **文章编号:**1001-2486(2026)01-238-09

## Ballistic missile maneuverability limited anti-interception game trajectory optimization method

YOU Yongan, SUN Ruisheng\*, LIU Xuanting, CHEN Wei  
(School of Energy and Power Engineering, Nanjing University of Science and Technology, Nanjing 210094, China)

**Abstract:** Aiming at the problem of maneuverability limitation in the process of penetration and interception, a game trajectory optimization strategy solution based on adaptive dynamic programming was proposed under the condition of limited maneuverability. By establishing an affine nonlinear differential game model and considering the limited maneuverability, the performance index function of the control energy term with integral form was designed. The saddle point control strategy of the game was derived based on the differential game theory, and an evaluation network was designed based on the adaptive dynamic programming algorithm to approximate the solution of the differential game strategy. The weight adaptive updating law of the neural network was given and its stability was proved. Simulation results show that the proposed strategy solving method can achieve anti-interception effect and accurately strike enemy targets under the circumstance of limited maneuverability.

**Keywords:** ballistic missile; maneuverability limited; differential game; adaptive dynamic programming

弹道导弹具有速度快、射程远、精度高、威力大、突防能力强等特点,如何对来袭弹道导弹实施有效拦截一直是各军事强国防空能力建设的重点<sup>[1]</sup>。在此背景下,突防能力已逐渐成为衡量弹道导弹性能的核心指标<sup>[2]</sup>。突防弹道优化问题也可以看作是一种双方竞争对抗过程,双方通过采取不同攻防策略寻求自身收益的最大化。由于突防满足博弈问题的特性,国内外的学者们开始利用博弈理论来研究导弹的突防弹道优化问题。其中,微分博弈法最受学者们青睐<sup>[3-8]</sup>。

微分博弈是一种无限策略型博弈,这种无限策略集对应的策略空间主要受最大机动能力等限制,一般以控制约束的形式在模型中体现。考虑到战场环境的复杂多变性,加之各类约束条件的存在,这无疑给微分博弈问题的求解增添了重重困难,使之面临更高的要求和挑战。因此,相应的先进控制方法和现代控制理论也被应用于微分博弈问题从而提高系统的稳定性和鲁棒性<sup>[9]</sup>。文献[10-13]均考虑了机动能力受限情况下的突防微分博弈问题,但是其考虑的模型是线性化模

收稿日期:2025-03-13

基金项目:国家自然科学基金资助项目(12572403);水下信息与控制重点实验室资助项目(2022-JCJQ-LB-030-09)

第一作者:游勇安(2000—),男,江西丰城人,博士研究生,E-mail:yoyongan\_njust@163.com

\*通信作者:孙瑞胜(1978—),男,江苏盐城人,教授,博士,博士生导师,E-mail:rs.sun@njust.edu.cn

引用格式:游勇安,孙瑞胜,刘宣廷,等.弹道导弹机动能力受限反拦截博弈弹道优化方法[J].国防科技大学学报,2026,48(1):238-246.

Citation:YOU Y A, SUN R S, LIU X T, et al. Ballistic missile maneuverability limited anti-interception game trajectory optimization method[J]. Journal of National University of Defense Technology, 2026, 48(1): 238-246.

型,过于简单,不符合弹道导弹突防问题的非线性、强耦合性。文献[14-16]针对非线性系统的微分博弈问题,结合自适应动态规划算法进行在线求解,但是其并没有在策略求解过程中考虑机动能力受限问题,这对系统的稳定性及鲁棒性带来了很大的挑战。Gao 等<sup>[17]</sup>针对一类具有约束输入的导弹系统,设计了一种滑模控制器,然后基于零和微分博弈理论的自适应动态规划算法进行求解。Zhao 等<sup>[18]</sup>针对具有不匹配干扰的输入受限非线性系统,利用自适应动态规划算法设计滑模容错控制策略,并采用经验回放技术更新评价指标权重。Wang 等<sup>[19]</sup>针对具有状态约束的高超声速飞行器,提出了一种基于自适应动态规划算法的自适应容错控制设计方法,以提高系统的跟踪性能和容错能力。从上述研究中能够发现,由于增加了输入约束,问题的求解难度进一步增加,且所设计的求解算法也更加复杂。

针对以上问题,本文建立了仿射非线性微分博弈模型,设计了对控制能量项进行积分约束的性能指标函数,并基于微分博弈理论和自适应动态规划算法对微分博弈策略进行在线逼近,给出了神经网络权值自适应更新律,并对其稳定性进行了推导证明,通过仿真算例来验证所提模型和算法的有效性。

1 模型建立

1.1 非线性微分博弈模型

图 1 给出了博弈问题的突防场景。在突防场景中,我方成员包括弹道导弹,敌方成员包括两枚拦截导弹(拦截导弹 1,2)和目标。将攻防对抗过程划分为突防阶段 1、突防阶段 2 和打击阶段,并假设弹道导弹在突防阶段 1 结束时摆脱拦截导弹

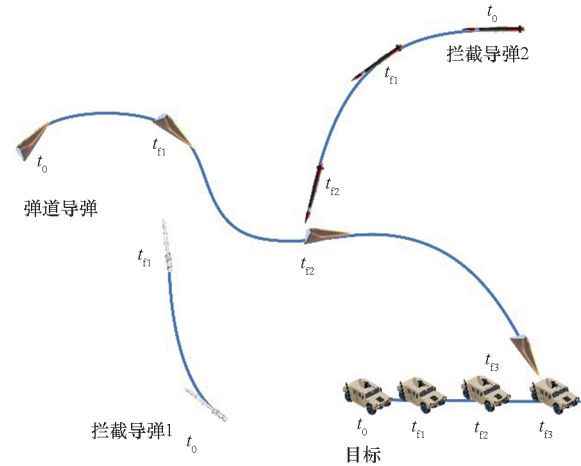


图 1 突防场景

Fig. 1 Penetration scenario

1,在突防阶段 2 结束时摆脱拦截导弹 2。

由于不同阶段的模型类似,为缩减篇幅,下文将仅推导突防阶段 1 的微分博弈策略。

在地面坐标系水平面内描述弹道导弹(M)、拦截导弹 1(I1)、拦截导弹 2(I2)和目标(T)的运动,相对运动方程组为:

$$\dot{X} = AX + BU_{I1} + CU_{I2} + DU_M \tag{1}$$

式中, $X$  为状态向量, $U_{I1}$ 、 $U_{I2}$ 、 $U_M$  分别为拦截导弹 1、拦截导弹 2、弹道导弹的加速度向量, $A$ 、 $B$ 、 $C$ 、 $D$  为系数矩阵。

将弹道导弹、拦截导弹 1、拦截导弹 2 在弹道坐标系下的加速度  $U_{2i} = [a_{x2i}, a_{z2i}]^T (i = \{M, I1, I2\})$  作为控制向量,则由弹道坐标系和地面坐标系的转换关系可得:

$$\begin{bmatrix} a_{xi} \\ a_{zi} \end{bmatrix} = L^{-1}(\psi_{vi}) \begin{bmatrix} a_{x2i} \\ a_{z2i} \end{bmatrix}, i = \{M, I1, I2\} \tag{2}$$

式中, $L(\psi_v) = \begin{bmatrix} \cos\psi_v & -\sin\psi_v \\ \sin\psi_v & \cos\psi_v \end{bmatrix}$ ,  $\psi_v$  为弹道偏角。

则有:

$$\dot{X} = AX + BL^{-1}U_{I1} + CL^{-1}U_{I2} + DL^{-1}U_M \tag{3}$$

令  $B_L = BL^{-1}$ 、 $C_L = CL^{-1}$ 、 $D_L = DL^{-1}$ , 则:

$$\begin{aligned} \dot{X} &= AX + B_L U_{I1} + C_L U_{I2} + D_L U_M \\ &= f(X) + \sum_{i=1}^3 g_i(X) u_i \end{aligned} \tag{4}$$

式中, $f(X)$  和  $g_i(X)$  均为关于状态量  $X$  的非线性连续函数。式(4)即为水平面内的非线性微分博弈模型,属于一种典型的仿射非线性系统。

1.2 性能指标函数

微分博弈问题本质上是一种连续、动态的冲突竞争行为<sup>[9]</sup>。根据微分博弈理论,突防双方应选取合适的控制策略以最大化或最小化给定的性能指标函数,从而使自身利益最大化,同时要考虑机动能力受限并保证系统的稳定性。

为此,定义性能指标函数为:

$$J = \frac{1}{2} \int_0^{t_f} [X^T Q X + U_{I1}^T W_{I1} U_{I1} + U_{I2}^T W_{I2} U_{I2} - u(U_M)] dt \tag{5}$$

式中: $Q = \begin{bmatrix} I_{2 \times 2} & 0_{2 \times 2} & 0_{2 \times 2} & 0_{2 \times 2} & 0_{2 \times 2} & 0_{2 \times 2} \\ 0_{2 \times 2} & I_{2 \times 2} & 0_{2 \times 2} & 0_{2 \times 2} & 0_{2 \times 2} & 0_{2 \times 2} \\ 0_{2 \times 2} & 0_{2 \times 2} & -I_{2 \times 2} & 0_{2 \times 2} & 0_{2 \times 2} & 0_{2 \times 2} \\ 0_{2 \times 2} & 0_{2 \times 2} & 0_{2 \times 2} & 0_{2 \times 2} & 0_{2 \times 2} & 0_{2 \times 2} \\ 0_{2 \times 2} & 0_{2 \times 2} & 0_{2 \times 2} & 0_{2 \times 2} & 0_{2 \times 2} & 0_{2 \times 2} \\ 0_{2 \times 2} & 0_{2 \times 2} & 0_{2 \times 2} & 0_{2 \times 2} & 0_{2 \times 2} & 0_{2 \times 2} \end{bmatrix};$

$W_{11} = \begin{bmatrix} w_{11x} & 0 \\ 0 & w_{11z} \end{bmatrix}$ 、 $W_{12} = \begin{bmatrix} w_{12x} & 0 \\ 0 & w_{12z} \end{bmatrix}$  分别为拦截导弹 1、拦截导弹 2 控制能量的权重矩阵,假设  $w_{11x} = w_{11z} = w_{11}$ ,  $w_{12x} = w_{12z} = w_{12}$ ;  $t_f$  为突防阶段博弈结束时刻;  $u(U_M)$  为处理机动能力受限问题选择的泛函。

考虑到机动能力受限问题,性能指标函数中弹道导弹的控制能量项设计为如式(6)所示的积分形式,而不是二次型形式,从而对机动能力受限问题进行有效处理。

$$u(U_M) = 2 \int_0^{U_M} \left[ \sigma \operatorname{arctanh} \left( \frac{\tau}{\sigma} \right) \right]^T W_M d\tau \quad (6)$$

式中:  $\tau$  为积分变量,表示弹道导弹加速度的中间增量;  $W_M = \begin{bmatrix} w_{Mx} & 0 \\ 0 & w_{Mz} \end{bmatrix}$  为弹道导弹控制能量的权重矩阵,  $w_{Mx} = w_{Mz} = w_M$ ;  $\sigma$  为待设计增益,是在数值上不大于导弹最大加速度的无量纲量。处理受限问题时常用的符号函数  $\operatorname{sgn}(\cdot)$  是不连续的,这会影响系统的性能,因此使用连续的反双曲正切函数  $\operatorname{arctanh}(\cdot)$  代替。

拦截导弹 1 希望最小化  $J$ , 拦截导弹 2 希望最小化  $J$ , 弹道导弹希望最大化  $J$ 。

微分博弈模型为:

$$\begin{aligned} \max_{U_M} \min_{U_{11}, U_{12}} & \frac{1}{2} \int_0^{t_f} [X^T Q X + U_{11}^T W_{11} U_{11} + \\ & U_{12}^T W_{12} U_{12} - u(U_M)] dt \\ \text{s. t.} \quad & \dot{X} = A X + B_L U_{11} + C_L U_{12} + D_L U_M \end{aligned} \quad (7)$$

2 弹道优化策略

2.1 基本思想

2.1.1 自适应动态规划的基本原理

自适应动态规划 (adaptive dynamic programming, ADP) 是通过设计强化学习中的“执行-评价”结构,利用函数近似方法,完成对非线性偏微分方程的近似求解。ADP 算法的基本结构如图 2 所示。

从图 2 可以看出,ADP 算法主要包括动态系统、执行网络、评价网络三部分。其中:动态系统通常是已知的系统状态微分方程;执行网络相当于控制器,根据评价网络反馈的信息生成相应的控制策略;评价网络则用于逼近最优性能指标函数。执行网络和评价网络构成了智能体系统。

2.1.2 机动能力受限下自适应动态规划思路

对于仿射非线性系统式(4),评价网络可以代替执行网络来实现控制器的作用。因此,ADP

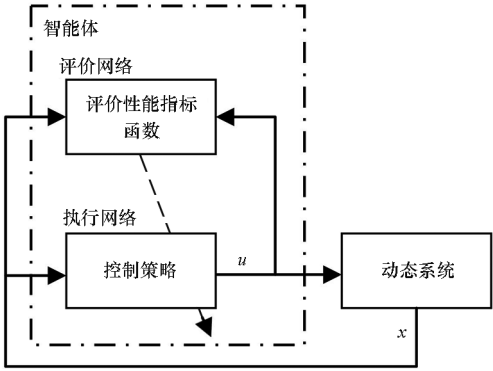


图 2 ADP 算法基本结构  
Fig. 2 ADP algorithm basic structure

算法结构可以简化为单评价网络结构,如图 3 所示。算法流程见算法 1。

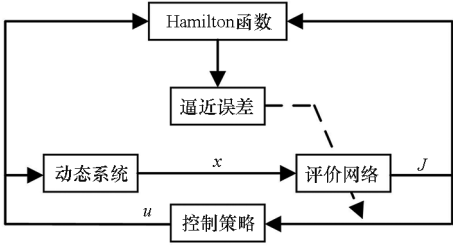


图 3 受限下基于 ADP 算法的学习过程  
Fig. 3 Learning process based on ADP algorithm under constraints

算法 1 受限下 ADP 算法

Alg. 1 ADP algorithm under constraints

输入: 初始状态 state0, 性能指标函数参数 params, 终止条件 threshold  
输出: 控制策略 strategy

初始化系统状态 state = state0  
初始化性能指标函数 J  
设计评价网络近似结构, 并确定相应参数  
while true do  
  评价网络计算近似性能指标函数 J  
  计算控制策略 strategy  
  if 满足终止条件 then  
    break  
  end if  
end while

2.2 策略设计

2.2.1 神经网络逼近

构造 Hamilton 函数:

$$H = \frac{1}{2} [X^T Q X + U_{11}^T W_{11} U_{11} + U_{12}^T W_{12} U_{12} - u(U_M)] +$$

$$\boldsymbol{\lambda}^T(\boldsymbol{A}\boldsymbol{X}+\boldsymbol{B}_L\boldsymbol{U}_\Pi+\boldsymbol{C}_L\boldsymbol{U}_{12}+\boldsymbol{D}_L\boldsymbol{U}_M) \tag{8}$$

式中: $\boldsymbol{\lambda}$  为伴随变量, $\boldsymbol{\lambda}=\nabla_{\boldsymbol{X}}J=\frac{\partial J}{\partial \boldsymbol{X}}$ 。

根据纳什均衡原理,该微分博弈应满足 HJB 方程:

$$0=H(\boldsymbol{X},\boldsymbol{U}_\Pi^*,\boldsymbol{U}_{12}^*,\boldsymbol{U}_M^*) \tag{9}$$

假设 HJB 方程式(9)的解存在且唯一,则微分博弈策略可根据极值条件 $\frac{\partial H}{\partial \boldsymbol{U}_\Pi},\frac{\partial H}{\partial \boldsymbol{U}_{12}},\frac{\partial H}{\partial \boldsymbol{U}_M}$ 求得:

$$\begin{cases} \boldsymbol{U}_\Pi^*=-\boldsymbol{W}_\Pi^{-1}\boldsymbol{B}_L^T\nabla_{\boldsymbol{X}}J \\ \boldsymbol{U}_{12}^*=-\boldsymbol{W}_{12}^{-1}\boldsymbol{C}_L^T\nabla_{\boldsymbol{X}}J \\ \boldsymbol{U}_M^*=\sigma\tanh(\boldsymbol{G}^*) \end{cases} \tag{10}$$

其中: $\boldsymbol{G}^*=\frac{1}{\sigma}\boldsymbol{W}_M^{-1}\boldsymbol{D}_L^T\nabla_{\boldsymbol{X}}J$ 。

此时,将 $\boldsymbol{U}_M^*$ 代入 $u(\boldsymbol{U}_M)$ ,可得:

$$u(\boldsymbol{U}_M)=2\sigma\cdot(\nabla_{\boldsymbol{X}}J)^T\boldsymbol{D}_L\tanh(\boldsymbol{G}^*)+\sigma^2\sum_{i=1}^2W_{Mi}\ln[1-\tanh^2(G_i^*)] \tag{11}$$

式中: $G_i^*$ 表示 $\boldsymbol{G}^*$ 的第 $i$ 个分量。

联立式(8)、式(10)、式(11),HJB 方程(9)可改写为:

$$\begin{aligned} 0=&\frac{1}{2}\boldsymbol{X}^T\boldsymbol{Q}\boldsymbol{X}-\frac{1}{2}(\nabla_{\boldsymbol{X}}J)^T\boldsymbol{B}_L\boldsymbol{W}_\Pi^{-1}\boldsymbol{B}_L^T\nabla_{\boldsymbol{X}}J- \\ &\frac{1}{2}(\nabla_{\boldsymbol{X}}J)^T\boldsymbol{C}_L\boldsymbol{W}_{12}^{-1}\boldsymbol{C}_L^T\nabla_{\boldsymbol{X}}J+(\nabla_{\boldsymbol{X}}J)^T\boldsymbol{A}\boldsymbol{X}+ \\ &\sigma^2\sum_{i=1}^2W_{Mi}\ln[1-\tanh^2(G_i^*)] \end{aligned} \tag{12}$$

式(12)为关于 $\nabla_{\boldsymbol{X}}J$ 的时变非线性偏微分方程,难以求得解析解。在此,将结合自适应动态规划算法设计一个评价网络在线逼近性能指标函数和最优微分博弈策略,并完成对 HJB 方程的近似求解。

为此,先给出如下合理性假设:

**假设 1** 针对模型(7)和微分博弈策略(10),存在一个连续可微的 Lyapunov 函数 $J_s(\boldsymbol{x})$ 满足 $\dot{J}_s(\boldsymbol{X})=(\nabla_{\boldsymbol{X}}J_s)^T(\boldsymbol{A}\boldsymbol{X}+\boldsymbol{B}_L\boldsymbol{U}_\Pi^*+\boldsymbol{C}_L\boldsymbol{U}_{12}^*+\boldsymbol{D}_L\boldsymbol{U}_M^*)<0$ 。那么,一定存在一个正定矩阵 $\boldsymbol{A}\in\mathbf{R}$ 使得式(13)成立。

$$\dot{J}_s(\boldsymbol{X})<-(\nabla_{\boldsymbol{X}}J_s)^T\boldsymbol{A}\nabla_{\boldsymbol{X}}J_s \tag{13}$$

设计的评价网络结构为:

$$J=\boldsymbol{W}^T\boldsymbol{S}(\boldsymbol{X})+\boldsymbol{\varepsilon}(\boldsymbol{X}) \tag{14}$$

式中: $\boldsymbol{W}^*$ 为神经网络的理想权重向量; $\boldsymbol{S}(\boldsymbol{X})$ 为神经网络的激活函数; $\boldsymbol{\varepsilon}(\boldsymbol{X})$ 为神经网络的有界逼近误差, $\boldsymbol{\varepsilon}(\boldsymbol{X})$ 满足 $|\boldsymbol{\varepsilon}(\boldsymbol{X})|\leq\delta$ ,其中 $\delta$ 为正常数。从而可以得到 $J$ 的梯度为 $\nabla_{\boldsymbol{X}}J=(\nabla\boldsymbol{S})^T\boldsymbol{W}^*+\nabla_{\boldsymbol{X}}\boldsymbol{\varepsilon}$ 。

将式(14)代入 $\boldsymbol{G}^*$ ,可得:

$$\boldsymbol{G}^*=\frac{1}{\sigma}\boldsymbol{W}_M^{-1}\boldsymbol{D}_L^T(\nabla\boldsymbol{S})^T\boldsymbol{W}^*+\frac{1}{\sigma}\boldsymbol{W}_M^{-1}\boldsymbol{D}_L^T\nabla_{\boldsymbol{X}}\boldsymbol{\varepsilon} \tag{15}$$

令 $\boldsymbol{\psi}=\frac{1}{\sigma}\boldsymbol{W}_M^{-1}\boldsymbol{D}_L^T(\nabla\boldsymbol{S})^T\boldsymbol{W}^*$ ,则式(10)可改

写为:

$$\begin{cases} \boldsymbol{U}_\Pi^*=-\boldsymbol{W}_\Pi^{-1}\boldsymbol{B}_L^T(\nabla\boldsymbol{S})^T\boldsymbol{W}^*-\boldsymbol{\varepsilon}_\Pi^* \\ \boldsymbol{U}_{12}^*=-\boldsymbol{W}_{12}^{-1}\boldsymbol{C}_L^T(\nabla\boldsymbol{S})^T\boldsymbol{W}^*-\boldsymbol{\varepsilon}_{12}^* \\ \boldsymbol{U}_M^*=\sigma\tanh(\boldsymbol{\psi})-\boldsymbol{\varepsilon}_M^* \end{cases} \tag{16}$$

其中, $\boldsymbol{\varepsilon}_\Pi^*=\boldsymbol{W}_\Pi^{-1}\boldsymbol{B}_L^T\nabla_{\boldsymbol{X}}\boldsymbol{\varepsilon}$ , $\boldsymbol{\varepsilon}_{12}^*=\boldsymbol{W}_{12}^{-1}\boldsymbol{C}_L^T\nabla_{\boldsymbol{X}}\boldsymbol{\varepsilon}$ , $\boldsymbol{\varepsilon}_M^*=-\sum_{i=1}^2W_{Mi}[1-\tanh^2(\xi_i)]\boldsymbol{D}_L^T\nabla_{\boldsymbol{X}}\boldsymbol{\varepsilon}(\xi_i\in[\psi_i,G_i^*])$ ,下标 $i$ 表示对应向量的第 $i$ 个分量。

综上所述,HJB 方程(12)简化为:

$$\begin{aligned} 0=&\frac{1}{2}\boldsymbol{X}^T\boldsymbol{Q}\boldsymbol{X}+\boldsymbol{W}^{*T}\nabla\boldsymbol{S}\boldsymbol{A}\boldsymbol{X}+\boldsymbol{\varepsilon}_{HJB}+ \\ &\sigma^2\sum_{i=1}^2W_{Mi}\ln[1-\tanh^2(\psi_i)]- \\ &\frac{1}{2}\boldsymbol{W}^{*T}\boldsymbol{B}_\Pi\boldsymbol{W}^*-\frac{1}{2}\boldsymbol{W}^{*T}\boldsymbol{B}_{12}\boldsymbol{W}^* \end{aligned} \tag{17}$$

式中: $\boldsymbol{B}_\Pi=\nabla\boldsymbol{S}\boldsymbol{A}_\Pi(\nabla\boldsymbol{S})^T$ , $\boldsymbol{A}_\Pi=\boldsymbol{B}_L\boldsymbol{W}_\Pi^{-1}\boldsymbol{B}_L^T$ ;  $\boldsymbol{B}_{12}=\nabla\boldsymbol{S}\boldsymbol{A}_{12}(\nabla\boldsymbol{S})^T$ , $\boldsymbol{A}_{12}=\boldsymbol{C}_L\boldsymbol{W}_{12}^{-1}\boldsymbol{C}_L^T$ ;  $\psi_i$ 表示向量 $\boldsymbol{\psi}$ 的第 $i$ 个分量; $\boldsymbol{\varepsilon}_{HJB}$ 是由神经网络的有界逼近误差 $\boldsymbol{\varepsilon}(\boldsymbol{X})$ 引起的残差项,满足

$$\begin{aligned} \boldsymbol{\varepsilon}_{HJB}=&\sigma^2\sum_{i=1}^2W_{Mi}\frac{1}{\eta_{2i}}[-2\tanh(\eta_{1i})][1- \\ &\tanh^2(\eta_{1i})]\varphi_i-\frac{1}{2}(\nabla_{\boldsymbol{X}}\boldsymbol{\varepsilon})^T\boldsymbol{A}_\Pi[2(\nabla\boldsymbol{S})^T\boldsymbol{W}^*+\nabla_{\boldsymbol{X}}\boldsymbol{\varepsilon}]- \\ &\frac{1}{2}(\nabla_{\boldsymbol{X}}\boldsymbol{\varepsilon})^T\boldsymbol{A}_{12}[2(\nabla\boldsymbol{S})^T\boldsymbol{W}^*+\nabla_{\boldsymbol{X}}\boldsymbol{\varepsilon}]+(\nabla_{\boldsymbol{X}}\boldsymbol{\varepsilon})^T\boldsymbol{A}\boldsymbol{X} \end{aligned}$$

$\varphi_i$ 表示向量 $\boldsymbol{\varphi}$ 的第 $i$ 个分量。

但是,由于理想权重通常是未知的,上述 HJB 方程依然无法求解。因此,构建评价网络的在线输出 $\boldsymbol{W}_e$ 估计理想权重 $\boldsymbol{W}^*$ ,式(14)的近似结构为:

$$\hat{J}=\boldsymbol{W}_e^T\boldsymbol{S}(\boldsymbol{X}) \tag{18}$$

式中, $\hat{J}$ 为对性能指标函数 $J^*$ 的估计。则其梯度为 $\nabla_{\boldsymbol{X}}\hat{J}=(\nabla\boldsymbol{S})^T\boldsymbol{W}_e$ 。

将式(18)代入式(10),获得近似的微分博弈策略:

$$\begin{cases} \hat{\boldsymbol{U}}_\Pi=-\boldsymbol{W}_\Pi^{-1}\boldsymbol{B}_L^T(\nabla\boldsymbol{S})^T\boldsymbol{W}_e \\ \hat{\boldsymbol{U}}_{12}=-\boldsymbol{W}_{12}^{-1}\boldsymbol{C}_L^T(\nabla\boldsymbol{S})^T\boldsymbol{W}_e \\ \hat{\boldsymbol{U}}_M=\sigma\tanh(\hat{\boldsymbol{G}}) \end{cases} \tag{19}$$

其中: $\hat{\boldsymbol{G}}=\frac{1}{\sigma}\boldsymbol{W}_M^{-1}\boldsymbol{D}_L^T(\nabla\boldsymbol{S})^T\boldsymbol{W}_e$ 。

## 2.2.2 神经网络权值自适应更新律

将式(19)重新代入式(12),可得:

$$0 = \frac{1}{2} \mathbf{X}^T \mathbf{Q} \mathbf{X} + \mathbf{W}_c^T \nabla \mathbf{S} \mathbf{A} \mathbf{X} + \sigma^2 \sum_{i=1}^2 W_{Mi} \ln[1 - \tanh^2(G_i^*)] - \frac{1}{2} \mathbf{W}_c^T \mathbf{B}_{11} \mathbf{W}_c - \frac{1}{2} \mathbf{W}_c^T \mathbf{B}_{12} \mathbf{W}_c \triangleq e_H \quad (20)$$

对比式(17)和式(20)可知,神经网络节点趋近于无穷时,由神经网络残差引起的逼近误差  $e_H$  将趋近于 0。因此,选取评价网络的误差函数为  $E_H = \frac{1}{2} e_H^T e_H$ 。

对此,需要设计一个评价网络的权值更新律,使得逼近误差  $e_H$  趋近于 0。通过梯度下降法,设计评价网络的权值更新律为:

$$\begin{aligned} \dot{\mathbf{W}}_c = & -\alpha_c \frac{\beta_c}{m_{\beta c}^2} e_H + \\ & \alpha_c \sum (X, \hat{U}_{11}, \hat{U}_{12}, \hat{U}_M) \nabla S(A_1 - A_2) \nabla_X J_s + \\ & \alpha_c \left[ \sigma \nabla SD_L P \frac{\beta_c}{m_{\beta c}^2} + \frac{1}{2} \frac{\beta_c}{m_{\beta c}^2} \mathbf{W}_c^T (\mathbf{B}_{11} + \mathbf{B}_{12}) - \right. \\ & \left. \left( Y_2 - Y_1 \frac{\beta_c}{m_{\beta c}^2} \right) \right] \mathbf{W}_c \end{aligned} \quad (21)$$

式中:  $\alpha_c > 0$  表示评价网络的学习率;  $m_{\beta c} = 1 + \beta_c^T \beta_c$ ,  $\beta_c = \nabla S(\mathbf{A} \mathbf{X} + \mathbf{B}_L \hat{U}_{11} + \mathbf{C}_L \hat{U}_{12} + \mathbf{D}_L \hat{U}_M)$ ;  $\mathbf{A}_2 = \mathbf{A}_{11} + \mathbf{A}_{12}$ ,  $\mathbf{A}_1 = \mathbf{D}_L \mathbf{W}_M^{-1} (\mathbf{I} - \mathbf{I}_G) \mathbf{D}_L^T$  且  $\mathbf{I}_G = \text{diag}(\tanh^2 \hat{G}_i)$ ;  $\mathbf{P} = -\tanh(\hat{G}) - \text{sgn}(\hat{G})$ ;  $J_s$  为满足假设 1 的 Lyapunov 函数;  $\sum (X, \hat{U}_{11}, \hat{U}_{12}, \hat{U}_M)$  定义为

$$\sum (X, \hat{U}_{11}, \hat{U}_{12}, \hat{U}_M) = \begin{cases} 0, (\nabla_X J_s)^T (\mathbf{A} \mathbf{X} + \mathbf{B}_L \hat{U}_{11} + \mathbf{C}_L \hat{U}_{12} + \mathbf{D}_L \hat{U}_M) < 0 \\ 1, \text{其他} \end{cases}$$

至此,权值更新律设计完成。

## 2.2.3 稳定性证明

在进行稳定性证明前,给出如下假设:

**假设 2** 评价网络的理想权值  $\mathbf{W}^*$  与逼近误差偏导数  $\nabla \varepsilon$  的范数均有界,即存在大于 0 的常数使得  $\|\mathbf{W}^*\| < b_w$ ,  $\|\nabla_X \varepsilon\| < b_\varepsilon$  成立。考虑到  $\nabla_X \varepsilon$  有界,则式(16)中的评价网络残差项  $\varepsilon_{11}^*$ 、 $\varepsilon_{12}^*$ 、 $\varepsilon_M^*$  均有界,满足  $\|\varepsilon_{11}^*\| < b_{11}$ 、 $\|\varepsilon_{12}^*\| < b_{12}$ 、 $\|\varepsilon_M^*\| < b_M$ 。

基于以上分析,可以得出如下定理:

**定理 1** 针对式(7),当假设 1、假设 2 成立时,设计微分博弈策略和评价网络的权值更新律分别如式(19)、式(21)所示,则闭环系统的状态和权值估计误差满足最终一致有界稳定。其中,

权值误差定义为  $\mathbf{W}^* = \mathbf{W}_c + \tilde{\mathbf{W}}_c$ 。

证明:选取 Lyapunov 函数:

$$L_x(t) = J_s(X) + \frac{1}{2} \tilde{\mathbf{W}}_c^T \alpha_c^{-1} \dot{\tilde{\mathbf{W}}}_c \quad (22)$$

则  $L_x(t)$  关于时间求导可得:

$$\dot{L}_x(t) = (\nabla_X J_s)^T (\mathbf{A} \mathbf{X} + \mathbf{B}_L \hat{U}_{11} + \mathbf{C}_L \hat{U}_{12} + \mathbf{D}_L \hat{U}_M) + \tilde{\mathbf{W}}_c^T \alpha_c^{-1} \dot{\tilde{\mathbf{W}}}_c \quad (23)$$

在继续证明之前,先做如下的处理:

将式(20)和式(17)作差,可推得:

$$\begin{aligned} e_H = & -\tilde{\mathbf{W}}_c^T \nabla \mathbf{S} \mathbf{A} \mathbf{X} + 2\sigma^2 \mathbf{W}_M [\boldsymbol{\psi}^T \text{sgn}(\boldsymbol{\psi}) - \hat{\mathbf{G}} \text{sgn}(\hat{\mathbf{G}})] + \\ & 2\sigma^2 \sum_{i=1}^2 W_{Mi} \frac{1 + \exp[-2\psi_i \text{sgn}(\psi_i)]}{1 + \exp[-2\hat{G}_i \text{sgn}(\hat{G}_i)]} - \\ & \tilde{\mathbf{W}}_c^T \nabla \mathbf{S} \mathbf{B}_L \hat{U}_{11} - \tilde{\mathbf{W}}_c^T \nabla \mathbf{S} \mathbf{C}_L \hat{U}_{12} + \\ & \frac{1}{2} \tilde{\mathbf{W}}_c^T \mathbf{B}_{11} \tilde{\mathbf{W}}_c + \frac{1}{2} \tilde{\mathbf{W}}_c^T \mathbf{B}_{12} \tilde{\mathbf{W}}_c - \varepsilon_{HJB} \end{aligned} \quad (24)$$

根据  $\boldsymbol{\psi}$  和  $\hat{\mathbf{G}}$  的定义,可得:

$$\begin{aligned} & 2\sigma^2 \mathbf{W}_M [\boldsymbol{\psi}^T \text{sgn}(\boldsymbol{\psi}) - \hat{\mathbf{G}} \text{sgn}(\hat{\mathbf{G}})] \\ = & \sigma \tilde{\mathbf{W}}_c^T \nabla \mathbf{S} \mathbf{D}_L \text{sgn}(\hat{\mathbf{G}}) + \sigma \tilde{\mathbf{W}}_c^T \nabla \mathbf{S} \mathbf{D}_L [\text{sgn}(\boldsymbol{\psi}) - \text{sgn}(\hat{\mathbf{G}})] \end{aligned} \quad (25)$$

则  $e_H$  最终表示为:

$$\begin{aligned} e_H = & -\tilde{\mathbf{W}}_c^T \{ \nabla \mathbf{S} (\mathbf{A} \mathbf{X} + \mathbf{B}_L \hat{U}_{11} + \mathbf{C}_L \hat{U}_{12} + \mathbf{D}_L \hat{U}_M) + \\ & \sigma \nabla \mathbf{S} \mathbf{D}_L [-\tanh(\hat{\mathbf{G}}) - \text{sgn}(\hat{\mathbf{G}})] \} + \\ & \frac{1}{2} \tilde{\mathbf{W}}_c^T \mathbf{B}_{11} \tilde{\mathbf{W}}_c + \frac{1}{2} \tilde{\mathbf{W}}_c^T \mathbf{B}_{12} \tilde{\mathbf{W}}_c + \Delta e_H \end{aligned} \quad (26)$$

其中:

$$\begin{aligned} \Delta e_H = & \sigma \tilde{\mathbf{W}}_c^T \nabla \mathbf{S} \mathbf{D}_L [\text{sgn}(\boldsymbol{\psi}) - \text{sgn}(\hat{\mathbf{G}})] + \\ & 2\sigma^2 \sum_{i=1}^2 W_{Mi} \frac{1 + \exp[-2\psi_i \text{sgn}(\psi_i)]}{1 + \exp[-2\hat{G}_i \text{sgn}(\hat{G}_i)]} - \varepsilon_{HJB} \end{aligned}$$

根据式(21),考虑到  $\tilde{\mathbf{W}}_c = \mathbf{W}^* - \mathbf{W}_c$ ,可得:

$$\begin{aligned} & \tilde{\mathbf{W}}_c^T \alpha_c^{-1} \dot{\tilde{\mathbf{W}}}_c \\ = & -\tilde{\mathbf{W}}_c^T \frac{\beta_c \beta_c^T}{m_{\beta c}^2} \tilde{\mathbf{W}}_c + \Delta e_H \frac{\beta_c^T}{m_{\beta c}^2} \tilde{\mathbf{W}}_c - \\ & \sigma \tilde{\mathbf{W}}_c^T \nabla \mathbf{S} \mathbf{D}_L P \frac{\beta_c^T}{m_{\beta c}^2} \mathbf{W}^* + \\ & \frac{1}{2} \tilde{\mathbf{W}}_c^T \frac{\beta_c}{m_{\beta c}^2} \mathbf{W}^{*T} (\mathbf{B}_{11} + \mathbf{B}_{12}) \mathbf{W}^* - \\ & \tilde{\mathbf{W}}_c^T \frac{\beta_c}{m_{\beta c}^2} \mathbf{W}^{*T} (\mathbf{B}_{11} + \mathbf{B}_{12}) (\mathbf{W}^* - \tilde{\mathbf{W}}_c) - \\ & \sum (X, \hat{U}_{11}, \hat{U}_{12}, \hat{U}_M) (\nabla_X J_s)^T (A_1 - A_2) (\nabla S)^T \tilde{\mathbf{W}}_c + \\ & \tilde{\mathbf{W}}_c^T \left( Y_2 - Y_1 \frac{\beta_c^T}{m_{\beta c}^2} \right) \mathbf{W}_c \end{aligned} \quad (27)$$

式中,  $\tilde{\mathbf{W}}_c^T \left( Y_2 - Y_1 \frac{\beta_c^T}{m_{\beta c}^2} \right) \mathbf{W}_c$  可整理为:

$$\begin{aligned}
& \tilde{\mathbf{W}}_c^T \left( \mathbf{Y}_2 - \mathbf{Y}_1 \frac{\boldsymbol{\beta}_c^T}{m_{\beta c}} \right) \mathbf{W}_c \\
&= \tilde{\mathbf{W}}_c^T \mathbf{Y}_2 \mathbf{W}^* - \tilde{\mathbf{W}}_c^T \mathbf{Y}_2 \tilde{\mathbf{W}}_c - \tilde{\mathbf{W}}_c^T \mathbf{Y}_1 \frac{\boldsymbol{\beta}_c^T}{m_{\beta c}} \mathbf{W}^* + \tilde{\mathbf{W}}_c^T \mathbf{Y}_1 \frac{\boldsymbol{\beta}_c^T}{m_{\beta c}} \tilde{\mathbf{W}}_c
\end{aligned} \quad (28)$$

综上所述,式(23)最终可整理为:

$$\begin{aligned}
\dot{\mathbf{L}}_x(t) &= -\tilde{\mathbf{W}}_c^T \frac{\boldsymbol{\beta}_c \boldsymbol{\beta}_c^T}{m_{\beta c}^2} \tilde{\mathbf{W}}_c - \tilde{\mathbf{W}}_c^T \mathbf{Y}_2 \tilde{\mathbf{W}}_c + \\
& \tilde{\mathbf{W}}_c^T \frac{\boldsymbol{\beta}_c}{m_{\beta c}} \left[ \mathbf{Y}_1 + \frac{1}{m_{\beta c}} \mathbf{W}^{*T} (\mathbf{B}_{11} + \mathbf{B}_{12}) \right] \tilde{\mathbf{W}}_c + \\
& \tilde{\mathbf{W}}_c^T \frac{\boldsymbol{\beta}_c}{m_{\beta c}} \left[ -\frac{1}{2m_{\beta c}} \mathbf{W}^{*T} (\mathbf{B}_{11} + \mathbf{B}_{12}) \mathbf{W}^* + \frac{1}{m_{\beta c}} \Delta e_H \right] + \\
& (\nabla_X J_s)^T (\mathbf{A}\mathbf{X} + \mathbf{B}_L \hat{\mathbf{U}}_{11} + \mathbf{C}_L \hat{\mathbf{U}}_{12} + \mathbf{D}_L \hat{\mathbf{U}}_M) + \\
& \tilde{\mathbf{W}}_c^T \left( \mathbf{Y}_2 \mathbf{W}^* - \mathbf{Y}_1 \frac{\boldsymbol{\beta}_c^T}{m_{\beta c}} \mathbf{W}^* - \sigma \nabla S \mathbf{D}_L \mathbf{P} \frac{\boldsymbol{\beta}_c^T}{m_{\beta c}^2} \mathbf{W}^* \right) - \\
& \sum (X, \hat{\mathbf{U}}_{11}, \hat{\mathbf{U}}_{12}, \hat{\mathbf{U}}_M) (\nabla_X J_s)^T (\mathbf{A}_1 - \mathbf{A}_2) (\nabla S)^T \tilde{\mathbf{W}}_c
\end{aligned} \quad (29)$$

定义  $\mathbf{z}^T = [\tilde{\mathbf{W}}_c^T, \tilde{\mathbf{W}}_c^T \boldsymbol{\beta}_c]$ ,  $\bar{\boldsymbol{\beta}}_c = \frac{\boldsymbol{\beta}_c}{m_{\beta c}}$ , 可得:

$$\begin{aligned}
\dot{\mathbf{L}}_x(t) &= -\mathbf{z}^T \mathbf{M} \mathbf{z} + \mathbf{z}^T \mathbf{N} + \\
& (\nabla_X J_s)^T (\mathbf{A}\mathbf{X} + \mathbf{B}_L \hat{\mathbf{U}}_{11} + \mathbf{C}_L \hat{\mathbf{U}}_{12} + \mathbf{D}_L \hat{\mathbf{U}}_M) - \\
& \sum (X, \hat{\mathbf{U}}_{11}, \hat{\mathbf{U}}_{12}, \hat{\mathbf{U}}_M) (\nabla_X J_s)^T (\mathbf{A}_1 - \mathbf{A}_2) (\nabla S)^T \tilde{\mathbf{W}}_c \\
& \leq -\lambda \min(\mathbf{M}) \|\mathbf{z}\|^2 + \|\mathbf{N}\| \|\mathbf{z}\| + \\
& (\nabla_X J_s)^T (\mathbf{A}\mathbf{X} + \mathbf{B}_L \hat{\mathbf{U}}_{11} + \mathbf{C}_L \hat{\mathbf{U}}_{12} + \mathbf{D}_L \hat{\mathbf{U}}_M) - \\
& \sum (X, \hat{\mathbf{U}}_{11}, \hat{\mathbf{U}}_{12}, \hat{\mathbf{U}}_M) (\nabla_X J_s)^T (\mathbf{A}_1 - \mathbf{A}_2) (\nabla S)^T \tilde{\mathbf{W}}_c
\end{aligned} \quad (30)$$

式中:

$$\begin{aligned}
\mathbf{M} &= \begin{bmatrix} \mathbf{Y}_2 & m_{12} \\ m_{12} & \mathbf{I} \end{bmatrix} \\
\mathbf{N} &= \begin{bmatrix} \mathbf{Y}_2 \mathbf{W}^* - \mathbf{Y}_1 \bar{\boldsymbol{\beta}}^T \mathbf{W}^* - \Gamma \\ -\frac{1}{2m_{\beta c}} \mathbf{W}^{*T} (\mathbf{B}_{11} + \mathbf{B}_{12}) \mathbf{W}^* + \frac{1}{m_{\beta c}} \Delta e_H \end{bmatrix} \\
m_{12} &= -\frac{1}{2} \left[ \mathbf{Y}_1 + \frac{1}{m_{\beta c}} \mathbf{W}^{*T} (\mathbf{B}_{11} + \mathbf{B}_{12}) \right] \\
\Gamma &= \sigma \nabla S \mathbf{D}_L \mathbf{P} \frac{\bar{\boldsymbol{\beta}}^T}{m_{\beta c}} \mathbf{W}^*
\end{aligned}$$

$-\lambda \min(\cdot)$  表示对应矩阵的最小特征值。

考虑到式(30)的取值和  $\sum (X, \hat{\mathbf{U}}_{11}, \hat{\mathbf{U}}_{12}, \hat{\mathbf{U}}_M)$  有关,下面根据  $\sum (X, \hat{\mathbf{U}}_{11}, \hat{\mathbf{U}}_{12}, \hat{\mathbf{U}}_M)$  的取值分两种情况讨论:

情况1:  $\sum (X, \hat{\mathbf{U}}_{11}, \hat{\mathbf{U}}_{12}, \hat{\mathbf{U}}_M) = 0$ 。此时  $\nabla_X J_s$  满足  $(\nabla_X J_s)^T (\mathbf{A}\mathbf{X} + \mathbf{B}_L \hat{\mathbf{U}}_{11} + \mathbf{C}_L \hat{\mathbf{U}}_{12} + \mathbf{D}_L \hat{\mathbf{U}}_M) < 0$ , 则式(30)可改写为:

$$\dot{\mathbf{L}}_x(t) \leq -\lambda \min(\mathbf{M}) \left( \|\mathbf{z}\| - \frac{\|\mathbf{N}\|}{2\lambda \min(\mathbf{M})} \right)^2 + \frac{\|\mathbf{N}\|^2}{4\lambda \min(\mathbf{M})} \quad (31)$$

此时,若式(32)成立,则  $\dot{\mathbf{L}}_x(t) < 0$ 。

$$\|\mathbf{z}\| > \frac{\|\mathbf{N}\|}{\lambda \min(\mathbf{M})} \quad (32)$$

情况2:  $\sum (X, \hat{\mathbf{U}}_{11}, \hat{\mathbf{U}}_{12}, \hat{\mathbf{U}}_M) = 1$ 。此时,在评价网络的学习过程中闭环信号不再有界,考虑如式(33)所示的泰勒级数展开式。

$$\tanh(\boldsymbol{\psi}) - \tanh(\hat{\mathbf{G}}) = \frac{1}{\sigma} \mathbf{W}_M^{-1} (\mathbf{I} - \mathbf{I}_G) \mathbf{D}_L^T (\nabla S)^T \tilde{\mathbf{W}}_c + b_o \quad (33)$$

式中:  $b_o = o(\boldsymbol{\psi} - \hat{\mathbf{G}})^2$  为  $(\boldsymbol{\psi} - \hat{\mathbf{G}})^2$  的高阶无穷小。

又有  $\nabla_X \boldsymbol{\varepsilon} = -(\nabla S)^T \tilde{\mathbf{W}}_c$ , 此时,式(30)转换为:

$$\begin{aligned}
\dot{\mathbf{L}}_x(t) &\leq -\lambda \min(\mathbf{M}) \|\mathbf{z}\|^2 + \|\mathbf{N}\| \|\mathbf{z}\| + \\
& (\nabla_X J_s)^T (\mathbf{A}\mathbf{X} + \mathbf{B}_L \mathbf{U}_{11}^* + \mathbf{C}_L \mathbf{U}_{12}^* + \mathbf{D}_L \mathbf{U}_M^*) + \\
& \sigma b_o (\nabla_X J_s)^T \mathbf{D}_L - (\nabla_X J_s)^T \mathbf{D}_L \boldsymbol{\varepsilon}_M^* - \\
& (\nabla_X J_s)^T \mathbf{B}_L \boldsymbol{\varepsilon}_{11}^* - (\nabla_X J_s)^T \mathbf{C}_L \boldsymbol{\varepsilon}_{12}^*
\end{aligned} \quad (34)$$

当假设1成立时,即

$\dot{J}_s(\mathbf{X}) < -(\nabla_X J_s)^T \mathbf{A} \nabla_X J_s$ , 则式(34)简化为:

$$\begin{aligned}
\dot{\mathbf{L}}_x(t) &\leq -\lambda \min(\mathbf{M}) \left( \|\mathbf{z}\| - \frac{\|\mathbf{N}\|}{2\lambda \min(\mathbf{M})} \right)^2 - \\
& \lambda \min(\mathbf{A}) \left( \|\nabla_X J_s\| - \frac{b_1}{2\lambda \min(\mathbf{A})} \right)^2 + b_2
\end{aligned} \quad (35)$$

式中:  $b_1 = \sigma b_o b_d - b_d b_M - b_B b_{11} - b_C b_{12}$ ,  $b_d, b_M, b_B, b_C, b_{11}, b_{12}$  均为有界复合常数;  $b_2 = \frac{\|\mathbf{N}\|^2}{4\lambda \min(\mathbf{M})} + \frac{b_1^2}{4\lambda \min(\mathbf{A})}$ 。

因此,若要保证  $\dot{\mathbf{L}}_x(t) < 0$ , 必须满足:

$$\|\mathbf{z}\| > \frac{\|\mathbf{N}\|}{2\lambda \min(\mathbf{M})} + \sqrt{\frac{b_2}{\lambda \min(\mathbf{M})}} \quad (36)$$

或者

$$\|\nabla_X J_s\| > \frac{b_1}{2\lambda \min(\mathbf{A})} + \sqrt{\frac{b_2}{\lambda \min(\mathbf{A})}} \quad (37)$$

通过上述证明,可以保证权值更新误差  $\tilde{\mathbf{W}}_c$  是半全局一致最终有界的,定理1证毕。

### 3 仿真实验

为验证本文所提出的机动能力受限下基于自适应动态规划算法的博弈弹道优化方法的正确性和有效性,设置如下仿真场景,并将仿真结果与比例导引法、非机动能力受限下自适应动态规划算法(在策略设计时未考虑机动能力受限,仅在仿

真中对加速度进行限幅处理)进行对比研究。由于在打击阶段不存在拦截导弹,故采用比例导引打击目标。

3.1 仿真条件

各仿真参数如下:

突防阶段 1 控制能量权重  $w_{i11}=0.1$ 、 $w_{i12}=0.1$ 、 $w_{i1M}=0.08$ ;突防阶段 2 控制能量权重  $w_{i212}=0.88$ 、 $w_{i2M}=0.1$ ;初始时间  $t_0=0$  s;拦截导弹 1、拦截导弹 2 和弹道导弹在弹道系  $x$ 、 $z$  方向上的最大加速度分别取  $0\text{ m/s}^2$ 、 $60\text{ m/s}^2$ 。

仿真使用的神经网络为单隐层神经网络,突防阶段 1 输入层为 12 个状态量,隐含层中含有 12 个神经元;突防阶段 2 输入层为 8 个状态量,隐含层中含有 8 个神经元;神经网络相关参数均为无量纲量,设计如下:

1)突防阶段 1:学习率  $\alpha_{c1}=-3\times 10^{-6}$ ,增益  $\sigma_{i1}=60$ ,设计参数  $Y_{i11}=3$ , $Y_{2i1}=30\times [1,1,1,1,1,1,1,1,1,1,1,1]^T$ 。

2)突防阶段 2:学习率  $\alpha_{c2}=-3\times 10^{-6}$ ,增益  $\sigma_{i2}=60$ ,设计参数  $Y_{i12}=3$ , $Y_{2i2}=30\times [1,1,1,1,1,1,1,1,1,1,1,1]^T$ 。

需要注意的是,由于模型中状态量  $\mathbf{X}$  的各分量的量级差异较大,在使用激活函数之前需要进行归一化处理。

攻防双方的博弈初始条件如表 1 所示。

表 1 攻防双方的博弈初始条件

Tab. 1 Initial condition of the game between attack and defense			
成员	初始位置/ km	初始速度/ (m/s)	初始弹道 偏角/(°)
弹道导弹	(0, -5)	1 500	0
拦截导弹 1	(40, 0)	1 800	-172.87
拦截导弹 2	(50, 0)	2 200	-172.29
目标	(75, 0)	20	0

3.2 仿真结果

本节将在数值仿真中验证所设计的机动能力受限下自适应动态规划算法,并与比例导引法、非机动能力受限下自适应动态规划算法的仿真结果进行对比,仿真结果如图 4~7 所示。为了充分体现本文所提方法的突防有效性,还将设置两个策略组合,并通过数值打靶的方式计算弹道导弹的突防成功概率,仿真结果如图 8、表 2 所示。

由图 4、图 5 可知,在本文提出的突防方法下,弹道导弹很好地完成了突防任务。具体来说,在面对拦截导弹 1、拦截导弹 2 的攻击时,弹道导弹采取了适当的机动来躲避拦截导弹的进攻,从而实现成功突防。在成功躲避拦截导弹后,也完成了打击地面目标的任务。而在比例导引法下,弹道导弹并没有考虑逃逸机动,且拦截导弹的初始位置和初始速度更具优势,使得弹道导弹无法成功突防。

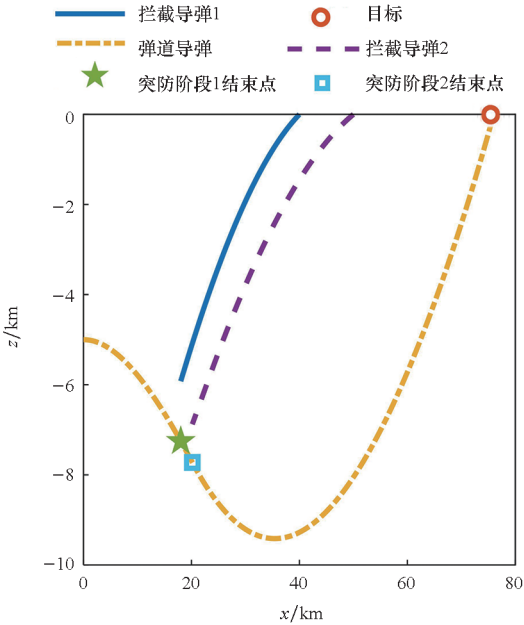


图 4 机动能力受限下攻防双方在  $xOz$  平面的飞行轨迹

Fig. 4 Flight trajectories of attack and defense in  $xOz$  plane with limited maneuverability

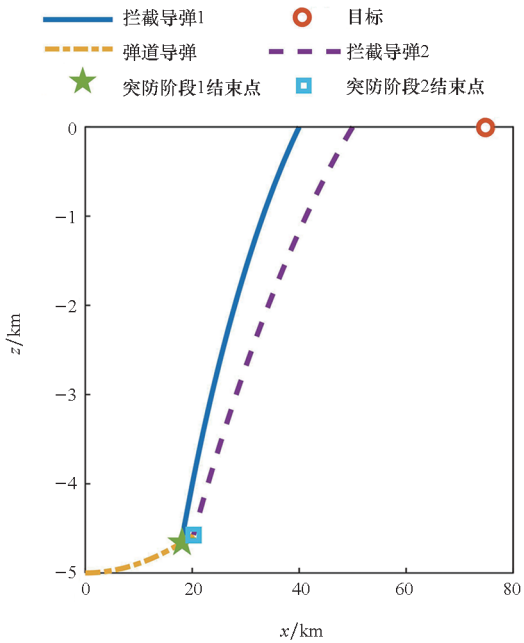


图 5 比例导引法下攻防双方在  $xOz$  平面的飞行轨迹

Fig. 5 Flight trajectories of attack and defense in  $xOz$  plane with proportional navigation method



从弹道导弹的加速度曲线图6可以看出,相较于非机动能力受限下自适应动态规划算法,在所设计的控制算法下,弹道导弹的控制输入始终处于机动能力受限范围内,这表明所设计的算法实现了对输入饱和现象的有效处理。

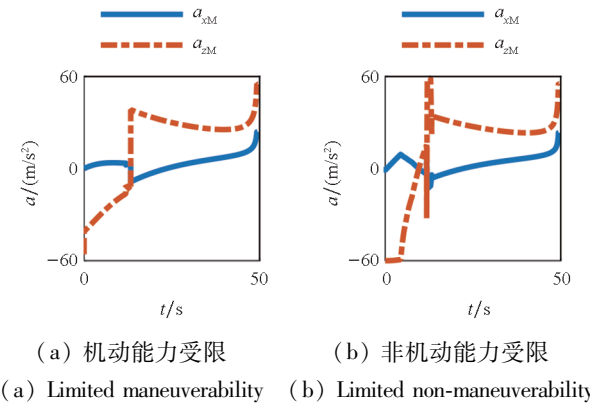


图6 弹道导弹加速度曲线

Fig.6 Ballistic missile acceleration curve

从拦截导弹的加速度曲线图7可以看出,相较于非机动能力受限下的自适应动态规划算法,在所设计的控制算法下,拦截导弹的控制输入也更小。

为系统评估本文所提方法的突防性能,设置了两种策略组合进行对比仿真。基准策略组合(1号):攻防双方均采用经典的比例导引法。待考察策略组合(2号):弹道导弹采用本文所提的方法,而拦截导弹仍采用比例导引法。通过调整弹道导弹的初始位置,设置110组数值打靶仿真工况。弹道导弹 $x$ 方向初始位置的区间为 $-20 \sim 20$  km,步长为4 km; $z$ 方向初始位置的区间为 $-10 \sim 10$  km,步长为2 km,不包括0 km。图8中,突防成功表示弹道导弹能够摆脱拦截导弹,并成功打中目标;反之,便是突防失败。

从图8能够看出,在策略组合1下,弹道导弹

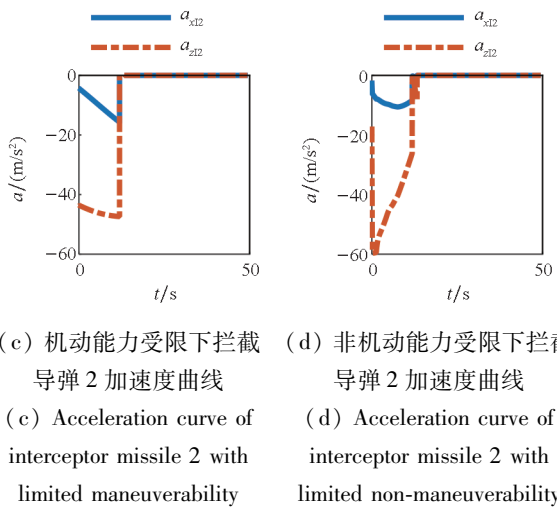
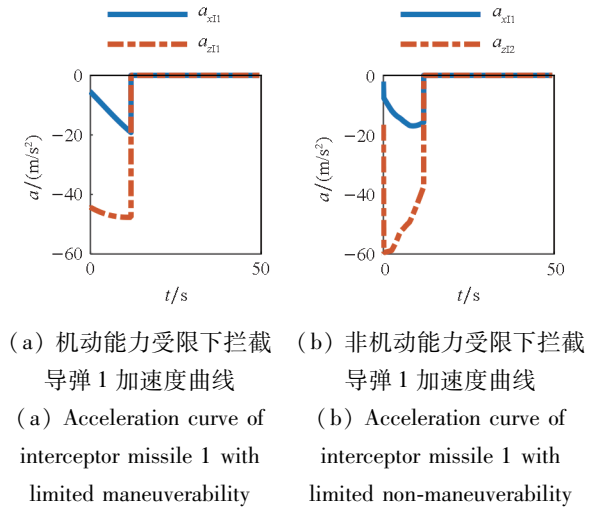


图7 拦截导弹加速度曲线

Fig.7 Intercept missile acceleration curve

的 $x$ 方向初始位置距离拦截导弹越近,留给拦截导弹的反应距离越短,能够突防成功时的 $z$ 方向初始位置区间就越大。而在策略组合2下, $z$ 方向初始距离绝对值过大,会突防失败,这是由于弹道导弹在拦截阶段进行躲避机动,在打击阶段时距离目标较远,无法打中目标。

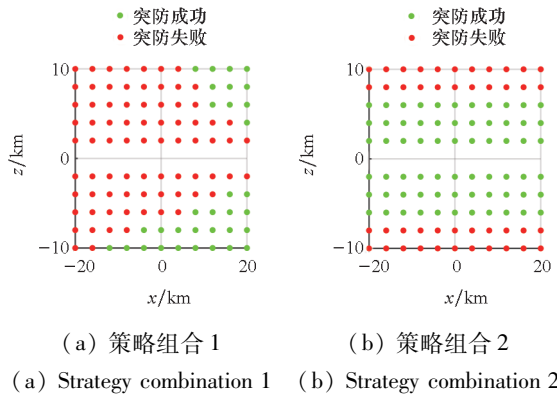


图8 不同策略组合下的数值打靶结果

Fig.8 Numerical target shooting results under different strategy combinations

表2 不同策略组合下的突防成功概率

Fig.2 The probability of successful penetration under different strategy combinations

组合序号	突防成功 工况数	突防失败 工况数	突防成功 概率/%
1	33	77	30
2	66	44	60

从表2能够看出,策略组合2相比策略组合1的突防成功概率提升了30%,这表明了本文所提方法能够有效地提升“一红突二蓝”场景下弹道导弹的突防成功概率。



4 结论

本文针对机动能力受限下微分博弈问题的数值求解研究,设计了考虑机动能力受限的性能指标函数以及微分博弈模型,设计了一种机动能力受限下的博弈策略,并引入自适应动态规划算法在线求解该问题得到微分博弈数值策略。所生成的博弈策略考虑了机动能力受限问题并有效实现弹道导弹的突防和打击地面高价值目标的任务。仿真结果表明了所提出的“一红突二蓝”场景下考虑机动能力受限的微分博弈问题的自适应动态规划求解方法的有效性。

参考文献 (References)

[1] 赵蒙, 端军红, 王明宇, 等. 大气层外弹道导弹中段拦截弹道规划[J]. 兵工学报, 2022, 43(7): 1589 – 1595.  
ZHAO M, DUAN J H, WANG M Y, et al. Trajectory planning for exoatmospheric midcourse ballistic missile interception[J]. Acta Armamentarii, 2022, 43(7): 1589 – 1595. (in Chinese)

[2] 桂航, 孙瑞胜, 刘宣廷. 固体弹道导弹大气层外弹道飞行时间控制方法[J]. 宇航学报, 2023, 44(7): 1042 – 1052.  
GUI H, SUN R S, LIU X T. Control method for extra-atmospheric flight time of solid ballistic missiles[J]. Journal of Astronautics, 2023, 44(7): 1042 – 1052. (in Chinese)

[3] WANG C G, CHEN D H, LIAO W H. Interactive multiple-model learning filter for spacecraft pursuit-evasion game strategy Switch based on long short-term memory network[J]. Aerospace, 2024, 11(11): 894.

[4] 程涛, 周浩, 董晓飞, 等. 多飞行器突防打击一体化微分对策制导律设计[J]. 北京航空航天大学学报, 2022, 48(5): 898 – 909.  
CHENG T, ZHOU H, DONG X F, et al. Differential game guidance law design for integration of penetration and strike of multiple flight vehicles[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(5): 898 – 909. (in Chinese)

[5] 张乘铭, 朱彦伟, 杨乐平, 等. 航天器交会型轨道追逃策略的滚动时域优化[J]. 国防科技大学学报, 2024, 46(3): 21 – 29.  
ZHANG C M, ZHU Y W, YANG L P, et al. Receding horizon optimization for spacecraft pursuit-evasion strategy in rendezvous[J]. Journal of National University of Defense Technology, 2024, 46(3): 21 – 29. (in Chinese)

[6] LIANG H Z, LI Z, WU J Z, et al. Optimal guidance laws for a hypersonic multiplayer pursuit-evasion game based on a differential game strategy[J]. Aerospace, 2022, 9(2): 97.

[7] XI A X, CAI Y L, DENG Y F, et al. Zero-sum differential game guidance law for missile interception engagement via neuro-dynamic programming [J]. Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering, 2023, 237(14): 3352 – 3366.

[8] CHEN J Q, SUN R S, LU Y. Cooperative game penetration guidance for multiple hypersonic vehicles under safety critical framework [J]. Chinese Journal of Aeronautics, 2024,

37(1): 247 – 255.

[9] 陈必露, 刘春生, 高煜欣. 输入受限下的鲁棒微分博弈拦截制导律设计[J]. 飞行力学, 2022, 40(3): 57 – 64.  
CHEN B L, LIU C S, GAO Y X. Robust differential game guidance law design of interception system with input constraints[J]. Flight Dynamics, 2022, 40(3): 57 – 64. (in Chinese)

[10] JA'AFARU A B, IBRAGIMOV G. On some pursuit and evasion differential game problems for an infinite number of first-order differential equations [J]. Journal of Applied Mathematics, 2012, 2012(1): 717124.

[11] 黄洁, 张友安, 赵育良, 等. 基于终端落角和加速度约束的末制导律设计[J]. 飞行力学, 2020, 38(3): 52 – 56, 62.  
HUANG J, ZHANG Y A, ZHAO Y L, et al. Design of terminal guidance law based on terminal fall angle and acceleration constraints[J]. Flight Dynamics, 2020, 38(3): 52 – 56, 62. (in Chinese)

[12] 赵亮博, 徐玮, 朱广生, 等. 再入飞行器中制导段微分对策式主动机动方法研究[J]. 导弹与航天运载技术(中英文), 2022(6): 83 – 89.  
ZHAO L B, XU W, ZHU G S, et al. Autonomous anti-interception penetration based on differential games for reentry vehicle on mid-course gliding guidance [J]. Missiles and Space Vehicles, 2022(6): 83 – 89. (in Chinese)

[13] 梁丽, 王佳楠, 王春彦, 等. 机动能力受限的目标 – 攻击 – 防御定性微分对策[J]. 指挥与控制学报, 2022, 8(4): 471 – 476.  
LIANG L, WANG J N, WANG C Y, et al. Target-attacking-defending qualitative differential game with limited maneuverability[J]. Journal of Command and Control, 2022, 8(4): 471 – 476. (in Chinese)

[14] MING Z Y, ZHANG H G, LIANG Y L, et al. Nonzero-sum differential games of continuous-time nonlinear systems with uniformly ultimately  $\varepsilon$ -bounded by adaptive dynamic programming [J]. Applied Mathematics and Computation, 2022, 430: 127248.

[15] 王子瑶, 唐胜景, 郭杰, 等. 高超声速攻防博弈自适应微分对策三维制导[J]. 兵工学报, 2023, 44(8): 2342 – 2353.  
WANG Z Y, TANG S J, GUO J, et al. Adaptive 3-dimensional differential game guidance for hypersonic attack and defense[J]. Acta Armamentarii, 2023, 44(8): 2342 – 2353. (in Chinese)

[16] 陈畅文. 再入航天器的离轨再入轨迹优化与控制[D]. 哈尔滨: 哈尔滨工业大学, 2023.  
CHEN C W. Optimization and control of de-orbit re-entry trajectory of re-entry spacecraft[D]. Harbin: Harbin Institute of Technology, 2023. (in Chinese)

[17] GAO Y X, LIU C S, JIANG S, et al. Zero-sum differential games-based fast adaptive robust optimal sliding mode control design for uncertain missile autopilot with constrained input[J]. International Journal of Control, 2022, 95(7): 1789 – 1801.

[18] ZHAO H, WANG H Q, NIU B, et al. Event-triggered fault-tolerant control for input-constrained nonlinear systems with mismatched disturbances via adaptive dynamic programming[J]. Neural Networks, 2023, 164: 508 – 520.

[19] WANG L, QI R Y, JIANG B. Adaptive fault-tolerant optimal control for hypersonic vehicles with state constraints based on adaptive dynamic programming [J]. Journal of the Franklin Institute, 2024, 361(8): 106833.