北航学报
赠 阅

ISSN 1001-5965
CODEN BHHDE8

北京航空航天大學学报

JOURNAL OF BEIJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS

2018-04
Vol.44 No.4

北航学报
赠阅

北京航空航天大学学报

第 44 卷　第 4 期　（总第 302 期）　2018 年 4 月

目　次

高速柔性转子系统非线性振动响应特征分析 …… 洪杰，于欢，肖森，马艳红（653）
基于双模式驱动的飞行汽车起飞阶段动力匹配分析 …… 徐斌，田富刚（662）
基于稀疏傅里叶变换的快速捕获方法 …… 张春熹，李先慕，高爽（670）
气象因素对飞机进近飞行燃油效率的影响 …… 胡荣，吴文洁，陈琳，张菲菲（677）
BDS/GPS 组合导航接收机自主完好性监测算法 …… 王尔申，杨福霞，庞涛，曲萍萍，蓝晓宇，姜毅（684）
放飞过程中平流层飞艇运动与受力分析 …… 张泰华，姜鲁华，周江华（691）
基于 Rao-Blackwellized 蒙特卡罗数据关联的检测跟踪联合优化 …… 陈唯实，闫军，李敬（700）
轴向间距对转子叶片颤振特性的影响机理 …… 郑赟，余永博（709）
基于滚动时域优化的旋转弹解耦控制器设计 …… 陈伟，孙传杰，冯高鹏，拜云山（717）
面向装备 RUL 预测的平行仿真技术 …… 葛承垄，朱元昌，邸彦强（725）
基于输入成形的太阳能帆板自适应滑模控制 …… 周通，郭宏，徐金全（737）
基于模块化的缠绕机设计方法 …… 徐小明，张武翔，丁希仑（746）
行星三体引力摄动对卫星探测器大气制动的影响 …… 江玲，王悦，徐世杰（759）
RP-3 航空燃油中 CO_2 扩散系数实验分析 …… 李超越，冯诗愚，邵垒，潘俊，刘卫华（765）
基于 LSTM 循环神经网络的故障时间序列预测 …… 王鑫，吴际，刘超，杨海燕，杜艳丽，牛文生（772）
基于波导缝隙天线的 TEM 室频率扩展方法 …… 宋春江，冯骁尧，戴飞（785）
基于半监督集成学习的多核设计空间探索 …… 李丹丹，姚淑珍，王颖，王森章，谭火彬（792）
基于测试性 ***D*** 矩阵的多故障诊断与维修策略 …… 田恒，段富海，樊亮，桑勇，史萌（802）
基于拉普拉斯变换的空间目标碰撞概率计算方法 …… 霍俞蓉，李智，韩蕾（810）
月基平台对地观测数据传输链路方案设计及分析 …… 陈国强，阮智星，郭华东，刘广，丁翼星，张严心（820）
多模函数优化的改进花朵授粉算法 …… 郭庆，惠晓滨，张贾奎，李正欣（828）
基于自适应通信拓扑四旋翼无人机编队重构控制 …… 马思迁，董朝阳，马鸣宇，王青（841）
基于八叉树的简化构型三角片等值面削减算法 …… 徐雷，王华锋，潘海侠，林广艳，陈栎曦（851）
Ni 晶界上金属和非金属元素的相互作用 …… 方淑娟，石松鑫，祝令刚，周健（862）
高纯 V_2AlC 的无压烧结制备及表征 …… 高少华，缪奶华，周健（868）
Zn-Cr_2AlC 复合材料的制备与摩擦磨损性能 …… 李志，赵文月，缪奶华（874）
可吸收胶原膜的体内免疫反应评价 …… 张林，孙磊，徐梦洽，牛旭锋（879）
基于竞争博弈的多目标可靠性优化设计方法 …… 冯嘉珍，张建国，邱继伟（887）

期刊基本参数： CN 11-2625/V * 1956 * m * A4 * 242 * zh * P * ￥50.00 * 900 * 28 * 2018-04

（**编　辑**　娄　嘉　张　嵘　李　晶　张欣蔚　孙　芳　王艳梅）

JOURNAL OF BEIJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS

Vol. 44 No. 4 (Sum 302) April 2018

北航学报 赠阅

CONTENTS

Nonlinear vibration response characteristics of high-speed flexible rotor system
…… HONG Jie, YU Huan, XIAO Sen, MA Yanhong (653)
Power matching of flying cars during takeoff stage based on dual-mode driving
…… XU Bin, TIAN Fugang (662)
Fast acquisition methods based on sparse Fourier transform
…… ZHANG Chunxi, LI Xianmu, GAO Shuang (670)
Influence of meteorological factors on aircraft fuel efficiency in approach flight
…… HU Rong, WU Wenjie, CHEN Lin, ZHANG Feifei (677)
BDS/GPS combined navigation receiver autonomous integrity monitoring algorithm
…… WANG Ershen, YANG Fuxia, PANG Tao, QU Pingping, LAN Xiaoyu, JIANG Yi (684)
Kinematic and mechanical analysis on launch process of stratospheric airship
…… ZHANG Taihua, JIANG Luhua, ZHOU Jianghua (691)
Joint optimization of detection and tracking with Rao-Blackwellized Monte Carlo data association
…… CHEN Weishi, YAN Jun, LI Jing (700)
Influence mechanism of axial spacing on rotor blade flutter characteristics
…… ZHENG Yun, YU Yongbo (709)
Design of decoupling controller for spinning missile based on receding horizon optimal
…… CHEN Wei, SUN Chuanjie, FENG Gaopeng, BAI Yunshan (717)
Equipment RUL prediction oriented parallel simulation technology
…… GE Chenglong, ZHU Yuanchang, DI Yanqiang (725)
Adaptive sliding mode control of solar array with input shaping
…… ZHOU Tong, GUO Hong, XU Jinquan (737)
Modular design method for filament winding machine
…… XU Xiaoming, ZHANG Wuxiang, DING Xilun (746)
Effect of planet's third-body gravitational perturbation on aerobraking of moon's explorer
…… JIANG Ling, WANG Yue, XU Shijie (759)
Experimental analysis of carbon dioxide diffusion coefficient in RP-3 jet fuel
…… LI Chaoyue, FENG Shiyu, SHAO Lei, PAN Jun, LIU Weihua (765)
Exploring LSTM based recurrent neural network for failure time series prediction
…… WANG Xin, WU Ji, LIU Chao, YANG Haiyan, DU Yanli, NIU Wensheng (772)
Frequency extension method of TEM cells based on slotted waveguide antenna
…… SONG Chunjiang, FENG Xiaoyao, DAI Fei (785)
Multicore design space exploration via semi-supervised ensemble learning
…… LI Dandan, YAO Shuzhen, WANG Ying, WANG Senzhang, TAN Huobin (792)
Multiple fault diagnosis and maintenance strategy based on testability $\boldsymbol{D}$ matrix
…… TIAN Heng, DUAN Fuhai, FAN Liang, SANG Yong, SHI Meng (802)
Computation method of collision probability of space object based on Laplace transformation
…… HUO Yurong, LI Zhi, HAN Lei (810)
Moon-based platform data transmission link scheme design and analysis
…… CHEN Guoqiang, RUAN Zhixing, GUO Huadong, LIU Guang, DING Yixing, ZHANG Yanxin (820)
Improved flower pollination algorithm for multimodal function optimization
…… GUO Qing, HUI Xiaobin, ZHANG Jiakui, LI Zhengxin (828)
Formation reconfiguration control of quadrotor UAVs based on adaptive communication topology
…… MA Siqian, DONG Chaoyang, MA Mingyu, WANG Qing (841)
Octree based decimation algorithm for triangle isosurface using simplified patterns
…… XU Lei, WANG Huafeng, PAN Haixia, LIN Guangyan, CHEN Yuexi (851)
Interaction between metallic and nonmetallic elements on grain boundary of nickel
…… FANG Shujuan, SHI Songxin, ZHU Linggang, ZHOU Jian (862)
Synthesis and characterization of high purity V_2AlC prepared by pressureless sintering
…… GAO Shaohua, MIAO Naihua, ZHOU Jian (868)
Preparation and tribological properties of Zn-Cr_2AlC composites
…… LI Zhi, ZHAO Wenyue, MIAO Naihua (874)
Immunological response evaluation of absorbable collagen membrane in vivo
…… ZHANG Lin, SUN Lei, XU Menghan, NIU Xufeng (879)
Multi-objective reliability design optimization approach based on competition game
…… FENG Jiazhen, ZHANG Jianguo, QIU Jiwei (887)

北航学报 赠阅

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2017.0266

高速柔性转子系统非线性振动响应特征分析

洪杰[1,2], 于欢[1], 肖森[1], 马艳红[1,2,*]

(1. 北京航空航天大学 能源与动力工程学院, 北京 100083; 2. 先进航空发动机协同创新中心, 北京 100083)

摘 要: 针对高速柔性转子多支点支承的结构特点及转子动力特性设计的需要,分析松动支承对转子动力特性的影响,仿真研究得到多支点支承高速柔性转子系统的非线性振动响应特征。研究结果表明:工作在多阶临界转速以上的转子系统,存在松动支承时,工作中的柔性转子可能存在周期、拟周期、混沌运动。进而研究了松动支承位置、不平衡量、松动间隙等参数对多支点支承柔性转子振动响应的影响,分析结果为多支点支承高速柔性转子系统的动力学设计提供了理论方法。

关 键 词: 高速柔性转子; 松动支承; 非线性振动; 动力响应; 动力学设计
中图分类号: V231.96
文献标识码: A **文章编号**: 1001-5965(2018)04-0653-09

全下文载

航空发动机中,高速柔性转子为工作转速在弯曲振型临界转速之上的转子系统,该类转子系统往往采用多支点支承的方案,但受制造及装配工艺的限制,往往存在支承不同心问题。与地面设备相比,航空燃气涡轮发动机转子具有工作转速高、工作过程中除传递较大扭矩外还传递轴向力的特点,因此,多支点支承高速柔性转子系统在支承不同心激励下轴内部会产生交变应力,极易造成转子产生疲劳破坏。

因此,为减小支承不同心的影响,多支点支承的高速柔性转子可采取松动支承设计,即在支承结构中设计合理的间隙。但是,对于支承带有间隙的转子结构系统,转子振幅变化时,支承的约束作用会产生突变,使得转子系统支承刚度具有阶跃的非连续特征。而航空发动机主轴承工作环境极其恶劣,因此,松动支承引起的支承刚度非连续特征可能会使转子系统产生复杂的非线性响应,影响航空发动机转子系统的动力特性稳定。尤其对于高速柔性转子,由于该类转子系统质量、刚度分布极为不均,弯曲临界转速低,工作在超临界状态下,使得松动支承对高速柔性转子动力特性的不利影响更加突出[1]。而当转子支承系统的结构参数如松动支承的位置、松动间隙、不平衡量等设计合理时,转子系统非线性振动响应会受到抑制,所以多支点支承高速柔性转子采用松动支承设计时保持良好的动力特性,需对其非线性振动响应特征进行研究。

Wang[2]及 Hussien[3]等采用龙格库塔法对非线性油膜力及非线性支承共同作用下的 Jeffcott 转子的振动响应特征及其稳定性进行数值求解,将仿真结果和线性转子系统振动响应对比发现,支承刚度改变时,在一定转速范围内转子可产生拟周期运动;Goldman 和 Muszynska[4-6]考虑了支承结构的刚度、阻尼及切向摩擦,其研究结果表明松动支承激励下转子存在同步及次同步振动成分,同时存在高次谐波振动成分,一定工作状态下混沌运动可能产生;Karpenko 等[7-8]仅考虑了支承间隙及非连续支承刚度的影响,建立的多点法

收稿日期: 2017-04-28; **录用日期**: 2017-08-02; **网络出版时间**: 2017-10-18 11:38
网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20171018.1138.001.html
基金项目: 国家自然科学基金 (51575022)
* **通信作者**. E-mail: mayanh2002@163.com

引用格式: 洪杰, 于欢, 肖森, 等. 高速柔性转子系统非线性振动响应特征分析[J]. 北京航空航天大学学报, 2018, 44(4): 653-661. HONG J, YU H, XIAO S, et al. Nonlinear vibration response characteristics of high-speed flexible rotor system[J]. Journal of Beijing University of Aeronautics and Astronautics, 2018, 44(4): 653-661 (in Chinese).

北航学报 赠阅

数值求解结果表明,当转子与带有间隙的支承之间产生频繁碰撞时,可产生混沌运动;Kim、Noah[9]和 Ehrich[10-11]综合考虑了挤压油膜和支承间隙对转子振动响应的影响,分别通过数值仿真和实验证明了松动支承的间隙是导致带有支承松动的转子系统产生复杂的非线性振动响应特征的原因;Kim 和 Noah[12]采用谐波平衡法研究了带有间隙支承的转子结构系统的分岔特性及其稳定性,发现转子系统可通过 Hopt 分岔产生拟周期运动;Wiercigroch 等[13-14]利用 Jeffcott 转子研究了支承松动转子系统通往混沌的途径,发现间隙支承的转子结构系统可通过切分岔、倍周期分叉的途径通往混沌。

综上,目前针对松动支承激励下转子结构系统非线性振动响应特征的研究,计算模型多采用简单的 Jeffcott 转子,且没有综合考虑非连续支承刚度、切向摩擦及挤压油膜的影响,因而不能直接用于指导多支点支承高速柔性转子的松动支承设计。本文参考上述工作,基于松动支承的力学特征,研究得到非连续支承刚度产生机理,分析松动支承对柔性转子系统动力特性的影响,进而综合考虑阶跃的非连续支承刚度及油膜力的影响,建立带有松动支承的多支点支承高速柔性转子系统动力学模型,仿真研究得到了转子的非线性振动响应特征。进而根据支承结构参数对多支点支承高速柔性转子系统动力学特性的影响,对支承位置和支承刚度进行优化,以减弱非连续支承刚度对转子动力特性的影响,进而根据转子振动响应幅值随转速的变化设计松动间隙,使工作转速区域远离混沌响应区域,并具有一定的安全裕度,保证高速柔性转子采用松动支承设计时具有稳健的动力特性,从而为多支点支承高速柔性转子系统动力特性设计提供理论方法。

1 松动支承激励的力学特征及动力学模型

1.1 松动支承激励的力学特征

如图 1 所示(图中 o_{si}为各支点位置处转子轴线位置),工程中典型的松动支承设计是通过调整挤压油膜阻尼器(Squeeze Film Damper,SFD)的间隙,加大转子与支承之间的间隙,从而消除装配时支承不同心对转子的附加约束[1]。

图 2(a)为带有挤压油膜阻尼器的支承结构物理模型。转子运动过程中,支承刚度主要由油膜压力产生的油膜刚度及刚性支承的支承刚

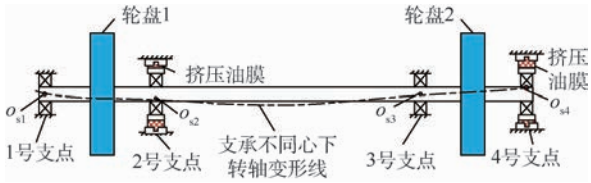
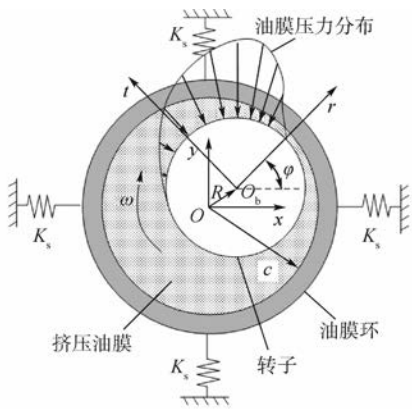
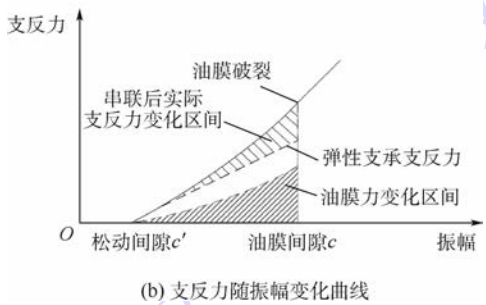


图 1 采用松动支承设计的多支点支承柔性转子模型

Fig. 1 Multi-supported flexible rotor model with bearing looseness design



(a) 带有挤压油膜阻尼器支承结构物理模型



(b) 支反力随振幅变化曲线

图 2 加大油膜间隙的挤压油膜阻尼器转子模型

Fig. 2 Model of SFD rotor system with increased oil film clearance

度 K_s 串联而成,由于油膜压力不同,因而在不同位置处挤压油膜力在径向方向 r 和切向方向 t 分量有所差异,因而当转子振动位移与水平方向的夹角 φ 不同时,在水平和垂直方向分量也会随之改变。对于油膜间隙较大的挤压油膜阻尼器,转子作微幅振动油膜不受挤压时,支承不起约束作用。转子振幅增加,转子与油膜环内表面产生挤压油膜作用时,会产生油膜力,为转子提供油膜刚度,此时支承本身刚度与油膜刚度为串联关系,油膜刚度起主要支承作用,在油膜受挤压的扩张区内,靠近间隙小的一端阻力大,故而油压升高大,靠近间隙大的一端阻力小,油压升高小,从而会形成如图 2(a)所示的油压分布。同时,由于油膜力除受振幅影响外,运动速度及进动角度也会改变油膜力大小,从而在振幅一定时,根据转子运动轨

迹的随机性，油膜力具有区间分布特征，如图 2(b)阴影所示。转子振幅继续增加与油膜环产生碰撞时，油膜会破裂，油膜刚度消失，此时支承本身刚度起主要支承作用。据此绘制支承支反力随振幅变化曲线，如图 2(b)所示，可以看出，松动支承可使转子支承刚度产生突变，从而支反力随振幅变化在油膜破裂点处导数不连续，具有非连续的力学特征，同时当转子与带间隙的支承之间产生频繁碰撞时，转子运动轨迹的随机性导致支承刚度具有区间分布特征，使转子系统具有非确定性特征。

支承刚度改变时，转子系统固有频率会随之改变，进而转子系统共振点会产生突变，使转子系统振动响应较为复杂。当转子动力特性对支承刚度敏感度较高时，支承松动可能会使转子产生混沌[1]。所以，高速柔性转子系统采用松动支承设计时，应对其支承方案及相应支承参数进行优化设计，以降低转子动力特性对支承刚度非连续变化的敏感度，使工作转速区域远离混沌响应区域并具有一定裕度，以保证转子动力特性稳定。

1.2　松动支承激励下多支点支承高速柔性转子系统的动力学模型

根据高速柔性转子多支点支承的特点以及松动支承的力学特征，建立如图 3 所示多支点支承高速柔性转子系统力学模型。O_1、O_2 为轮盘的几何中心，O_{1e}、O_{2e}为轮盘的质心，e_1、e_2 为质心到形心距离，轮盘质量分别为 m_1、m_2，轮盘处阻尼分别为 C_{d1}、C_{d2}。K_{si}和 C_{si}分别为各支点的支承刚度和阻尼，c_2、c_3 为油膜间隙，c_2'、c_3' 为松动间隙。

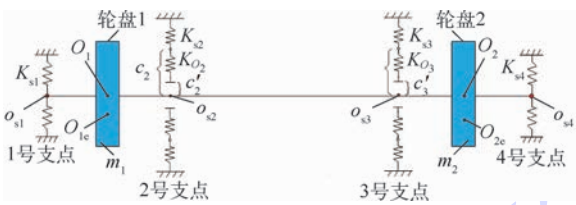


图 3　多支点支承高速柔性转子系统力学模型

Fig. 3　Mechanical model of multi-supported high-speed flexible rotor system

忽略陀螺力矩和扭转振动的影响，转子系统横向振动的运动微分方程为

$$\begin{cases} m_{s1}\ddot{x}_{s1} + C_{s1}\dot{x}_{s1} + k_1(x_{s1} - x_{d1}) + K_{s1}x_{s1} = 0 \\ m_{s1}\ddot{y}_{s1} + C_{s1}\dot{y}_{s1} + k_1(y_{s1} - y_{d1}) + K_{s1}y_{s1} = 0 \end{cases} \tag{1a}$$

$$\begin{cases} m_1\ddot{x}_{d1} + C_{d1}\dot{x}_{d1} + k_1(2x_{d1} - x_{s1} - x_{s2}) = \\ \quad m_1 e_1\omega^2\cos(\omega t) \\ m_1\ddot{y}_{d1} + C_{d1}\dot{y}_{d1} + k_1(2y_{d1} - y_{s1} - y_{s2}) = \\ \quad m_1 e_1\omega^2\sin(\omega t) \end{cases} \tag{1b}$$

$$\begin{cases} m_{s2}\ddot{x}_{s2} + C_{s2}\dot{x}_{s2} + k_1(x_{s2} - x_{d1}) + \\ \quad k_2(x_{s2} - x_{s3}) = F_{s2x} \\ m_{s2}\ddot{y}_{s2} + C_{s2}\dot{y}_{s2} + k_1(y_{s2} - y_{d1}) + \\ \quad k_2(y_{s2} - y_{s3}) = F_{s2y} \end{cases} \tag{1c}$$

$$\begin{cases} m_{s3}\ddot{x}_{s3} + C_{s3}\dot{x}_{s3} + k_2(x_{s3} - x_{s2}) + \\ \quad k_1(x_{s3} - x_{d2}) = F_{s3x} \\ m_{s3}\ddot{y}_{s3} + C_{s3}\dot{y}_{s3} + k_2(y_{s3} - y_{s2}) + \\ \quad k_1(y_{s3} - y_{d2}) = F_{s3y} \end{cases} \tag{1d}$$

$$\begin{cases} m_2\ddot{x}_{d2} + C_{d2}\dot{x}_{d2} + k_1(2x_{d2} - x_{s3} - x_{s4}) = \\ \quad m_2 e_2\omega^2\cos(\omega t) \\ m_2\ddot{y}_{d2} + C_{d2}\dot{y}_{d2} + k_1(2y_{d2} - y_{s3} - y_{s4}) = \\ \quad m_2 e_2\omega^2\sin(\omega t) \end{cases} \tag{1e}$$

$$\begin{cases} m_{s4}\ddot{x}_{s4} + C_{s4}\dot{x}_{s4} + k_1(x_{s4} - x_{d2}) + K_{s4}x_{s4} = 0 \\ m_{s4}\ddot{y}_{s4} + C_{s4}\dot{y}_{s4} + k_1(y_{s2} - y_{d2}) + K_{s4}y_{s4} = 0 \end{cases} \tag{1f}$$

式中：ω 为转子同步进动角速度；m_{si}为支承质量；k_i 为转子在轮盘 i 处的刚度；x_{si}、y_{si}为支承位置处转子振动位移分量；x_{di}、y_{di}为轮盘振动位移分量；F_{six}、F_{siy}为松动支承支反力；2、3 号支点采用支承松动设计，支承刚度为 K_{s2}、K_{s3}。支承的支反力随转子振幅变化具有非连续特征，其表达式为

$$F_{six} = \begin{cases} 0 & r_{si} \leqslant c_i' \\ f_{xi} & c_i' < r_{si} < c_i \\ -K_{si}(r_{si} - c_i)\cos\varphi_i + \mu K_{si}(r_{si} - c_i)\sin\varphi_i & \\ & r_{si} \geqslant c_i \end{cases} \tag{2}$$

$$F_{siy} = \begin{cases} 0 & r_{si} \leqslant c_i' \\ f_{yi} & c_i' < r_{si} < c_i \\ -K_{si}(r_{si} - c_i)\cos\varphi_i - \mu K_{si}(r_{si} - c_i)\sin\varphi_i & \\ & r_{si} \geqslant c_i \end{cases} \tag{3}$$

式中：f_{xi}、f_{yi}为油膜力；$r_{si} = \sqrt{x_{si}^2 + y_{si}^2}$为支点位置处的横向振幅；$\mu$ 为摩擦系数；φ_i 为转子振动位移与水平方向的夹角。

采用短轴承油膜理论，半油膜条件下油膜力的表达式为

$$\begin{cases} f_x = F_r\cos\varphi - F_t\sin\varphi \\ f_y = F_r\sin\varphi + F_t\cos\varphi \end{cases} \tag{4}$$

式中：

$$\begin{cases} \cos\varphi = \dfrac{x_{si}}{r_{si}} \\ \sin\varphi = \dfrac{y_{si}}{r_{si}} \\ F_r = \dot{r}_{si}I_2 + r_{si}\dot{\varphi}I_1 \\ F_t = \dot{r}_{si}I_1 + r_{si}\dot{\varphi}I_3 \end{cases} \tag{5}$$

北航学报 赠阅

其中:I_1、I_2、I_3 为 Sommerfeld 系数[15-16]。

2 松动支承激励下的转子系统动力学特性分析

2.1 非线性振动响应特征

根据松动支承激励下多支点支承高速柔性转子系统力学模型和动力学模型,参考涡轴发动机动力涡轮转子结构,选取结构的各参数如表 1、表 2所示。

改变转子转速,仿真计算得到松动支承激励下高速柔性转子系统振动响应,如图 4 ~ 图 7 所示。

从图 4、图 5 可以看出,松动支承激励下的柔性转子,其振动响应具有如下特征:

1) 对于带有间隙的 2 号支点位置处的振动响应,在 10 ~ 1 200 rad/s 转速范围内,随转速增加,其运动形式依次为周期—拟周期—混沌—拟周期运动。

表 1 结构参数取值

Table 1 Values of structural parameters

参 数	数 值
$m_1e_1/(\mathrm{g\cdot mm})$	10
$m_2e_2/(\mathrm{g\cdot mm})$	10
c_2'/mm	7×10^{-4}
c_4'/mm	4×10^{-4}
c_2-c_2'/mm	2×10^{-4}
c_4-c_4'/mm	2×10^{-4}
$k_1/(\mathrm{N\cdot m^{-1}})$	2×10^{5}
$k_2/(\mathrm{N\cdot m^{-1}})$	1×10^{4}

表 2 挤压油膜阻尼器参数取值

Table 2 Values of SFD parameters

参 数	滑油黏度/(Pa·s)	轴向承载长度/mm	轴承半径/mm
数 值	1×10^{-3}	90	40

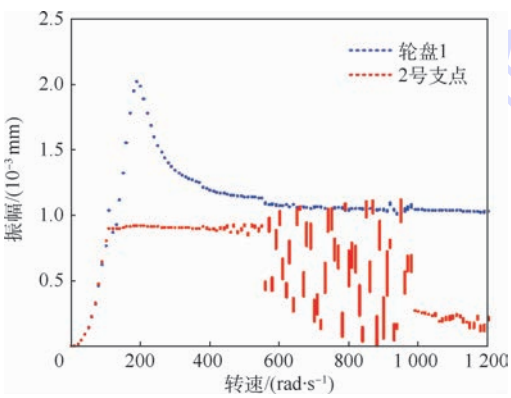


图 4 振幅随转速变化的分岔图

Fig. 4 Bifurcation diagram of vibration amplitude changing with rotational speed

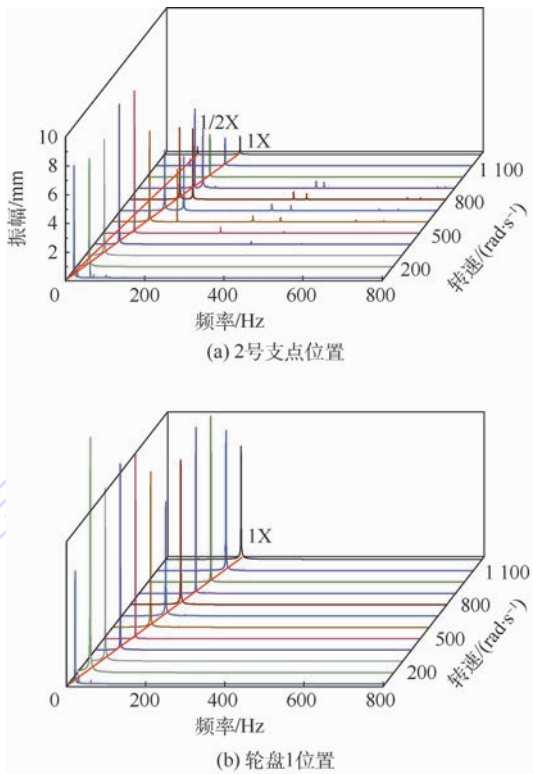


图 5 随转速变化的频域瀑布图

Fig. 5 Waterfall curve of frequency domain changing with rotational speed

如图 4 所示,当转速较低($\omega<100$ rad/s)时,由于转子振幅很小,松动支承几乎不起约束作用,轮盘 1 和 2 号支点振幅相同,此时转子处于两支点支承状态,为线性系统,频域响应仅为 1 倍频,转子做周期运动。

转子转速继续增加,向临界转速靠近,在不平衡力作用下,转子振幅增加使转子与松动支承之间的油膜受到挤压时,油膜的刚度和阻尼会抑制转子的振动,此时轮盘 1 和 2 号支点振幅出现差异,在非线性油膜力作用下,频域响应为 1 倍频和 3 倍频成分的叠加,转子产生拟周期运动。

转子转速继续增加,逐渐远离共振转速,不平衡激振力加大,当振幅下降至与松动支承之间产生频繁碰撞时,整个系统可看成 2 个系统的耦合:①固有非线性系统,即无不平衡激振作用下的松动支承转子系统;②不平衡激振力作用下的线性振动系统。其中,不平衡激振力可代表 2 个系统的耦合程度[16]。当不平衡激振力很小时,线性系统的振动很弱,对应非线性系统的作用也很弱。系统的运动便可近似看成是 2 个独立运动的叠加,即整个系统的运动是围绕平衡点之一的线性振动。此时根据初始条件不同,系统会绕不同的平衡点运动,从而在相平面上形成不同的流域。由于不平衡激振很小,振动很弱,系统的振动才不

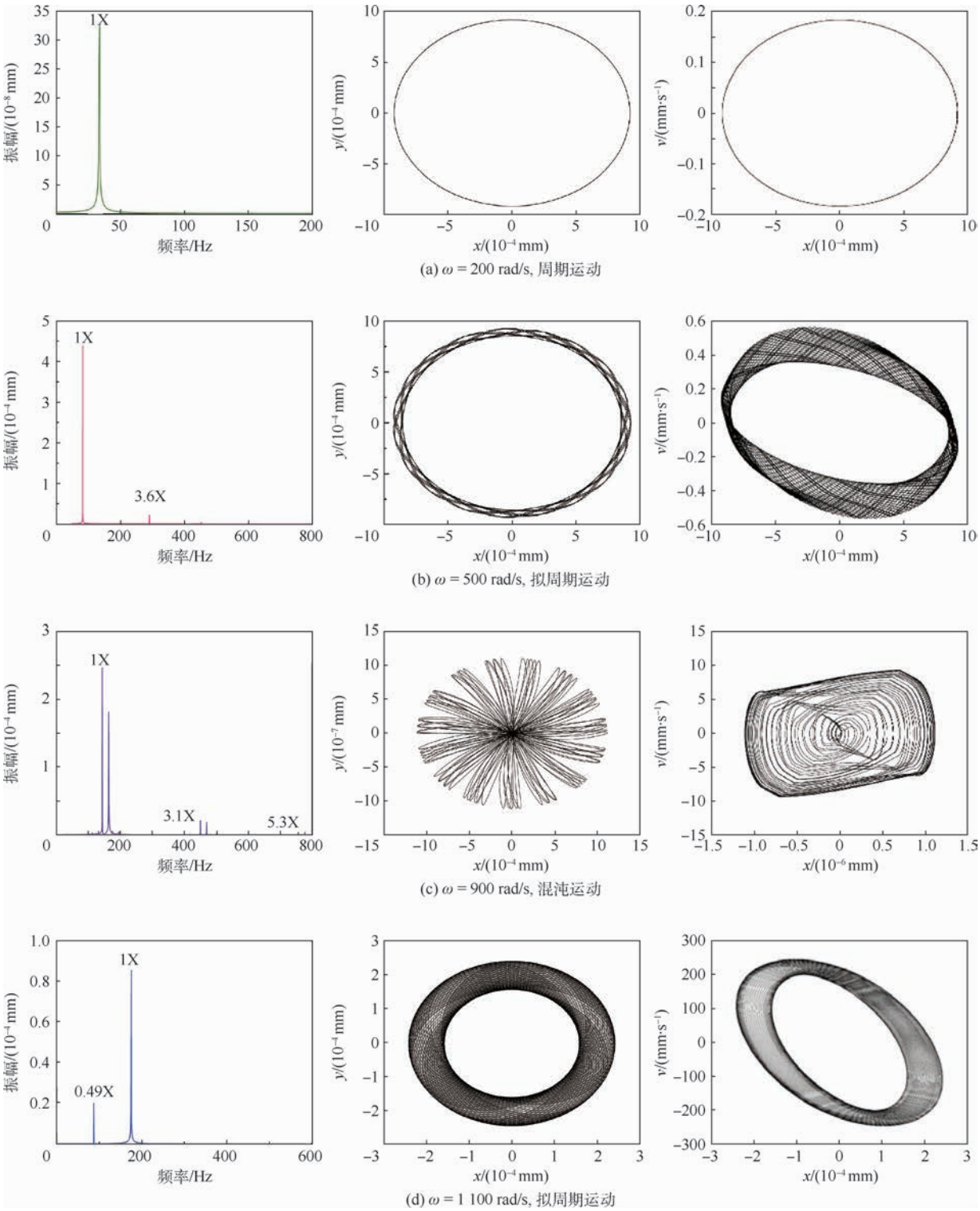


图 6　2 号支点位置转子的振动响应

Fig. 6　Vibration response of rotor at position of Support 2

会越出既定的流域而趋近于另一平衡点。

根据图 4 和图 6(c)的相图所示，当转速增加至 580 rad/s 左右时，受松动支承的影响，频域响应为 1 倍频、3 倍频、5 倍频及多种其他频率成分的叠加，运动形式为混沌运动。这是由于不平衡激振力增加，使得 2 个系统的相互影响增强到使振动可以超过初始流域的界限，从而可从初始的流域转到新的流域，将向其他平衡点靠近。同时，系统振幅也可能增大到超过新流域的界限而趋近于初始流域，此时系统会绕初始的平衡点运动。而系统在自发随机性的影响下会从一个流域进入另一个流域，系统在不同流域之间产生跳动而具有随机的特点，从而产生混沌运动[17-18]。

转速继续增加，不平衡激振力增大，使得线性

北航学报 赠阅

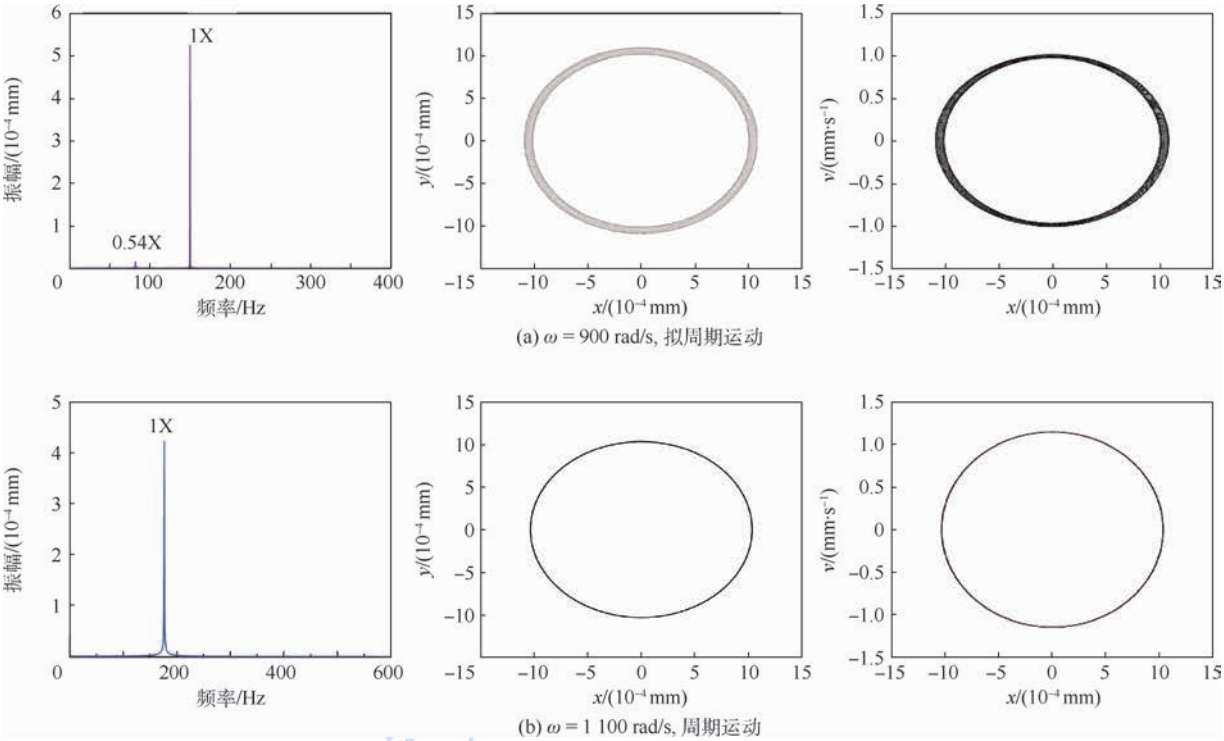


图 7 轮盘 1 位置转子的振动响应

Fig. 7 Vibration response of rotor at position of Disk1

系统的振动处于主导地位,振幅继续下降,在非线性油膜力作用下,转子运动状态为拟周期运动。

当不平衡激振力很大时,线性系统完全处于支配地位,松动支承引起的非线性影响极弱。此时,整个系统便按线性系统的运动方式运动,变为周期运动[16-17]。

2)图 7 为松动支承激励下轮盘的振动响应,运动形式主要为拟周期和周期运动。2 号支点位置处存在混沌运动时,对应轮盘振动响应以转速 1 倍频为主,同时包含其他频率成分,呈拟周期运动;2 号支点位置为拟周期和周期运动时,对应轮盘振动响应仅为 1 倍频,呈周期运动。

2.2 参数影响规律

为了研究松动间隙、非连续支承刚度、不平衡量大小及松动支承位置对转子响应的影响,以转子系统的加速过程为例,分析不同参数下的松动支承激励的转子系统响应分岔图。

2.2.1 松动支承位置影响

改变松动支承位置,将 2 号、4 号支点处设计松动间隙,其物理模型如图 8 所示。得到 2 号、3 号支点位置处及轮盘 2 位置处转子振幅随转速变化的分岔图。根据图 9,改变轮盘 2 位置处松动支承的位置,对 2 号支点位置处转子的振动响应影响较小,混沌响应区域稍有偏移。如图 10 所示,松动支承位置改变对轮盘2及其支承位置处

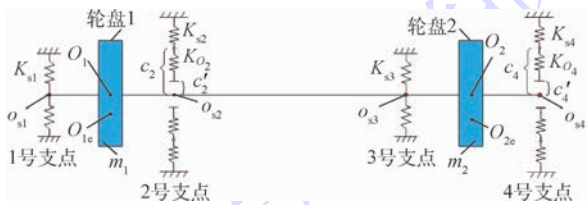


图 8 2 号、4 号支点采用松动支承设计的多支点支承柔性转子系统

Fig. 8 Multi-supported flexible rotor system with bearing clearance designed at Support 2 and Support 4

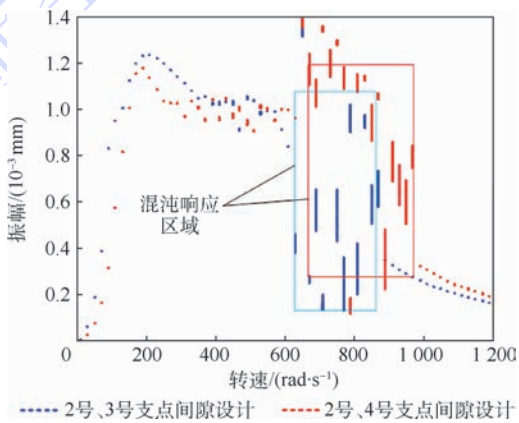


图 9 松动支承位置不同时 2 号支点振幅随转速变化的分岔图

Fig. 9 Bifurcation diagram of vibration amplitude of Support 2 changing with rotational speed when location of bearing looseness is different

振动响应影较大,4 号支点采用松动支承设计时,轮盘 2 在共振点处的峰值略有降低。根据图 11 可知,随转速增加,不平衡激振力增加使得转子与松动支承之间存在碰撞冲击时,松动支承位置处响应幅值对转速变化较敏感,因而在 2 号、3 号支点位置设计松动间隙时,3 号支点位置处转子响应幅值对转速变化较敏感。同时由于 3 号支点靠近转子质心,其振幅变化在非连续支承刚度影响下对转速变化敏感时,不利于转子动力特性稳定,因此,优先选择 2 号、4 号支点采用松动支承设计。

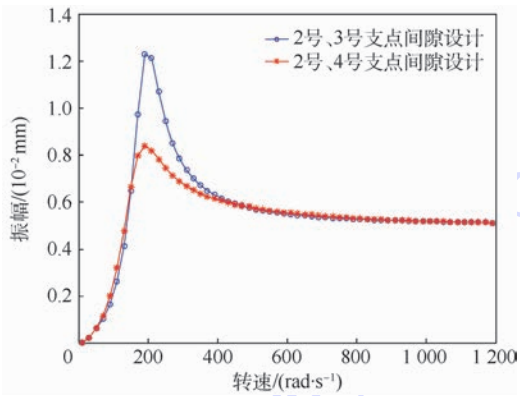


图 10　松动支承位置不同时轮盘 2 振幅随转速变化的分岔图

Fig. 10　Bifurcation diagram of vibration amplitude of Disk 2 changing with rotational speed when location of bearing looseness is different

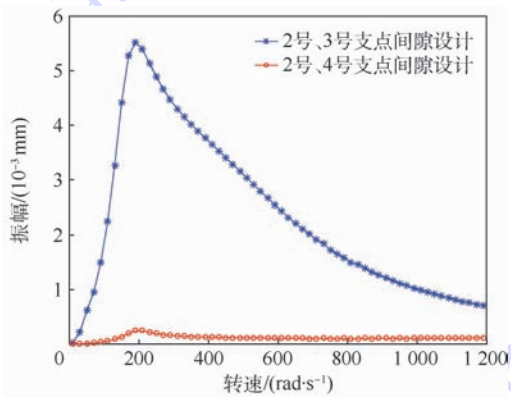


图 11　松动支承位置不同时 3 号支点振幅随转速变化的分岔图

Fig. 11　Bifurcation diagram of vibration amplitude of Support 3 changing with rotational speed when location of bearing looseness is different

2.2.2　不平衡量影响

如表 3 所示,改变轮盘 1、2 的不平衡量,得到轮盘 1 及 2 号支点位置处转子振幅随转速变化的分岔图,如图 12 所示。可以看出,转子不平衡量增加时,混沌响应的转速区域向高转速区偏移。

表 3　轮盘不平衡量取值

Table 3　Unbalance value of disk

不平衡量	m_1e_1/(g·mm)	m_2e_2/(g·mm)
小不平衡量	10	10
大不平衡量	50	50

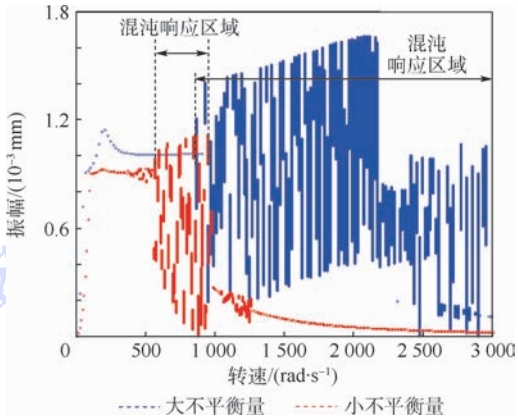


图 12　不平衡量不同时 2 号支点振幅随转速变化的分岔图

Fig. 12　Bifurcation diagram of vibration amplitude of Support 2 changing with rotational speed when unbalance value is different

这是由于当松动支承处的间隙保持不变时,只有当振幅达到 $0.4\times10^{-3}\sim0.9\times10^{-3}$ mm 时,2 号支点位置处转子才会与支承发生频繁碰撞,从而产生混沌运动。由于转子不平衡量增加时,转子振幅随转速变化会整体增加,因此与不平衡量小的情况相比转子在更高的转速范围内,2 号支点位置处转子的振幅才会减小至 $0.4\times10^{-3}\sim0.9\times10^{-3}$ mm,与支承产生频繁碰撞,从而产生混沌运动,且由于该振幅范围对应转速范围较宽,因此,对应混沌响应的转速范围较宽广。

2.2.3　非连续支承刚度影响

如表 4 所示,为研究支承刚度非连续的影响,改变松动支承位置处的支承刚度变化的区间范围,得到 2 号支点位置处转子振幅随转速变化的分岔图,如图 13 所示。可以看出,松动支承位置处支承刚度变化范围增大时,混响应区域稍有扩大。

表 4　支承刚度变化区间

Table 4　Variation range of bearing stiffness

支承刚度变化范围	K_{s2}/(N·m^{-1})	K_{s3}/(N·m^{-1})
支承刚度变化范围小	$0\sim1\times10^5$	$0\sim1\times10^5$
支承刚度变化范围大	$0\sim1\times10^7$	$0\sim1\times10^7$

2.2.4　松动间隙影响

如表 5 所示,改变松动支承位置处的间隙,得到 2 号支点位置处转子振幅随转速变化的分岔

北航学报 赠阅

图,如图 14 所示。可以看出,松动支承的间隙加大时,松动支承位置处转子振幅需增加至可与支承产生频繁碰撞才能引发混沌运动产生,所以产生混沌运动的转速区域向共振区域靠近,且混沌运动振幅随之增加。

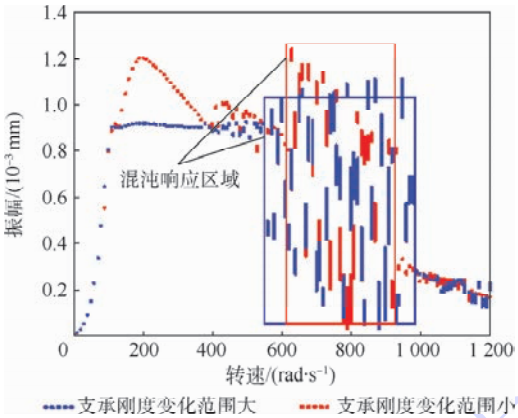


图 13 支承刚度不同时 2 号支点振幅随转速变化的分岔图

Fig. 13 Bifurcation diagram of vibration amplitude of Support 2 changing with rotational speed when bearing stiffness is different

表 5 松动间隙参数取值

Table 5 Parameter values of bearing clearance

松动间隙	c_2'/mm	c_4'/mm	c_2/mm	c_4/mm
小松动间隙	1.4×10^{-3}	8×10^{-4}	2×10^{-4}	2×10^{-4}
大松动间隙	7×10^{-3}	4×10^{-3}	2×10^{-3}	2×10^{-3}

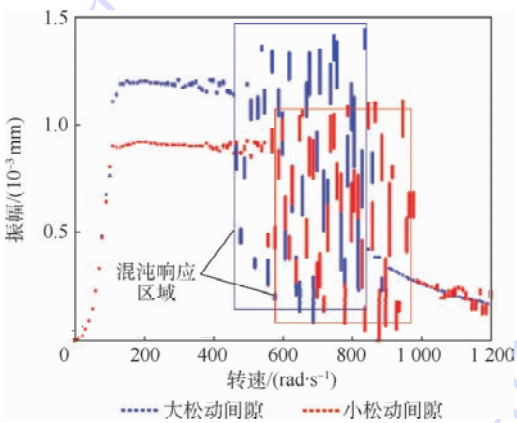


图 14 松动间隙不同时 2 号支点振幅随转速变化的分岔图

Fig. 14 Bifurcation diagram of vibration amplitude of Support 2 changing with rotational speed when bearing clearance is different

3 结 论

松动支承激励下多支点支承柔性转子的运动状态具有非线性特征,其非线性振动形式主要为拟周期和混沌运动:

1)在一定不平衡量下,转子与松动支承之间产生频繁碰撞时,松动支承位置处转子的运动形式为混沌运动。对应轮盘处运动形式为拟周期运动。

2)非线性油膜力作用下,松动支承位置处转子的运动形式为拟周期运动,对应轮盘处运动形式为周期运动。

参考文献(References)

[1] 于欢,马艳红,肖森,等. 高速柔性转子支承松动力学特征及动力特性[J]. 北京航空航天大学学报,2017,43(8):1677-1683.
YU H,MA Y H,XIAO S,et al. Mechanical and dynamic characteristics of bearing with looseness on high-speed flexible rotor [J]. Journal of Beijing University of Aeronautics and Astronautics,2017,43(8):1677-1683(in Chinese).

[2] WANG R,GUO X,WANG Y. Nonlinear analysis of rotor system supported by oil lubricated bearings subjected to base movements[J]. Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, 2016,230(4):167-172.

[3] HUSSIEN M,AL-WEDYAN H,TAHAT M S,et al. The behaviour of the Jeffcott rotor under a vibrating base of fluid film bearing[J]. Suranree Journal of Science and Technology,2008, 15(3):167-176.

[4] GOLDMAN P,MUSZYNSKA A. Chaotic behavior of rotor-stator systems with rubs[J]. Journal of Engineering for Gas Turbines and Power,1994,116(3):692-701.

[5] GOLDMAN P,MUSZYNKA A. Dynamic effects in mechanical structures with gap and impacting:Order and chaos[J]. Journal of Vibration and Acoustics,1994,116(4):541-547.

[6] GOLDMAN P,MUSZYNKA A. Chaotic response of unbalanced rotor/bearing/stator systems with looseness or rubs[J]. Chaos, Solitons & Fractals,1995,5(9):1682-1704.

[7] KARPENKO E V,WIERCIGROCH M,PAVLOVSKAIA E E,et al. Piecewise approximate analytical solutions for a Jeffcott rotor with a snubber ring[J]. International Journal of Mechanical Sciences,2002,44(3):475-488.

[8] KARPENKO E V,WIERCIGROCH M,CARTMELL M P. Regular and chaotic dynamics of a discontinuously nonlinear rotor system[J]. Chaos, Solitons & Fractals, 2002, 13(6):1231-1242.

[9] KIM Y B,NOAH S T. Quasi-periodic response and stability analysis for a non-linear Jeffcott rotor[J]. Journal of Sound and Vibration,1996,190(2):239-253.

[10] EHRICH F F. High order subharmonic response of high speed rotors in bearing clearance[J]. Journal of Vibration and Acoustics,1988,110(1):9-16.

[11] EHRICH F F. Subharmonic virbration of rotors in bearing clearance:ASME Paper No. 66-MD-1[R]. Ner York:ASME,1966.

[12] KIM Y B,NOAH S T. Bifurcation analysis for a modified Jeffcott rotor with bearing clearances[J]. Nonlinear Dynamics,

北航学报 赠阅

1990，1(3)：221-241.

[13] CHÁVEZ J P，WIERCIGROCH M. Bifurcation analysis of periodic orbits of a non-smooth Jeffcott rotor model[J]. Communications in Nonlinear Science & Numerical Simulation，2013，18(9)：2571-2580.

[14] WIERCIGROCH M，DE KRAKER B. Applied nonlinear dynamics and chaos of mechanical systems with discontinuities[M]. Singapore：World Scientific，2000：177-200.

[15] INAYAT-HUSSAIN J I，KANKI H，MUREITHI N W. On the bifurcations of a rigid rotor response in squeeze-film dampers[J]. Journal of Fluids & Structures，2003，17(3)：433-459.

[16] INAYAT-HUSSAIN J I，KANKI H，MUREITHI N W. Stability and bifurcation of a rigid rotor in cavitated squeeze-film dampers without centering springs[J]. Tribology International，2001，34(10)：689-702.

[17] 刘秉正，彭建华. 非线性动力学[M]. 北京：高等教育出版社，2004：132-135.
LIU B Z，PNEG J H. Nonlinear dynamics[M]. Beijing：Higher Education Press，2004：132-135(in Chinese).

[18] 赖志慧. 基于 Duffing 振子混沌和随机共振特性的微弱信号检测方法研究[D]. 天津：天津大学，2014.
LAI Z H. Weak-signal detection based on the chaotic and stochastic-resonance characteristics of Duffing oscillator[D]. Tianjin：Tianjin University，2014(in Chinese).

作者简介：

洪杰　男，博士，教授，博士生导师。主要研究方向：航空发动机转子动力学、航空发动机整机动力学、旋转机械振动控制、智能结构与新型阻尼材料等。

马艳红　女，博士，教授，博士生导师。主要研究方向：航空发动机整机动力学、旋转机械振动控制、智能结构与新型阻尼材料等。

Nonlinear vibration response characteristics of high-speed flexible rotor system

HONG Jie[1,2]，YU Huan[1]，XIAO Sen[1]，MA Yanhong[1,2,*]

(1. School of Energy and Power Engineering，Beijing University of Aeronautics and Astronautics，Beijing 100083，China；
2. Collaborative Innovation Center of Advanced Aero-Engine，Beijing 100083，China)

Abstract：According to the multi-supported structure feature of high-speed flexible rotor and with the demand of rotor dynamic design，the influence of the bearing looseness on the rotor dynamic was analyzed. The nonlinear vibration response characteristics of the high-speed flexible rotor system were obtained through simulation. The results show that when the working speed of rotor is above the critical speed，the periodic，quasi-periodic and chaos can exist when the flexible rotor working. Furthermore，the influence of location of bearing looseness，value of unbalance and bearing clearance on the vibration response of multi-supported flexible rotor was studied. The analysis results provide theoretical method for dynamic design of multi-supported high-speed flexible rotor system.

Keywords：high-speed flexible rotor；bearing looseness；nonlinear vibration；dynamic response；dynamic design

Received：2017-04-28；**Accepted**：2017-08-02；**Published online**：2017-10-18 11：38

URL：kns.cnki.net/kcms/detail/11.2625.V.20171018.1138.001.html

Foundation item：National Natural Science Foundation of China (51575022)

* **Corresponding author**. E-mail：mayanh2002@163.com

北京航空航天大学学报

北航学报 赠阅

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2017.0279

基于双模式驱动的飞行汽车起飞阶段动力匹配分析

全下文载

徐斌*，田富刚

（北京航空航天大学 交通科学与工程学院，北京 100083）

摘 要：针对折叠翼飞行汽车起飞阶段的动力匹配问题，研究了基本的动力控制策略，通过理论计算分析了起飞阶段双模式驱动特性，提出了最佳切换时刻的概念。对某型飞行汽车传动系统进行了动力匹配计算，并以该飞行汽车参数为基础，在 Simulink 中建立了不同工况下的行驶仿真模型，对起飞阶段行驶状态进行了仿真分析，给出了基于双模式驱动的动力匹配方案以及最佳切换时刻的选取原则。计算结果表明，通过采取双模式驱动，起飞加速时间缩短 22%，起飞滑跑距离缩短 13%。进一步对传动系统参数、整车设计参数以及发动机输出特性进行了优化分析，分析计算结果给出了各参数变化对起飞动力性能的影响。

关 键 词：飞行汽车；双模式驱动；最佳切换时刻；Simulink 仿真；参数分析
中图分类号：U469.79
文献标识码：A **文章编号**：1001-5965(2018)04-0662-08

飞行汽车的种类很多，固定翼飞行汽车因其较好的飞行性能和航程等优势，近年来逐渐成为国内外研究的重点[1-4]。对于采用固定翼的飞行汽车，其起飞过程时间很短，但作为地面行驶和飞行阶段之间的过渡过程，是飞行汽车应用中的关键阶段。

由于飞行汽车在地面行驶状态和飞行状态的驱动方式不同，因而飞行汽车动力系统的匹配和对应的驱动策略就很重要，然而近年来国内外鲜有针对性的成果。

本文针对某型飞行汽车，通过动力学分析和仿真建模，针对起飞阶段进行了动力匹配计算，分析了双模式驱动的优势，并在其基础上进行了传动系统优化分析。

1 双模式驱动与最佳切换时刻的定义

1.1 双模式驱动

对于固定翼飞行汽车[1-4]，起飞阶段的驱动力可以类似于飞机，由螺旋桨提供。然而螺旋桨的效率与进距比关系密切，在车速较低时螺旋桨提供的驱动力有限[1-2,5]。与之相对的，轮胎在这一阶段能够持续提供稳定的驱动力；这部分如果能够得到妥善利用，对于提高起飞阶段的加速性能大有裨益。如果能够充分利用轮胎在低速时提供的大驱动力，整个起飞阶段的加速性能都将得到提高。

当车速逐渐提升，进距比也随之增大，螺旋桨逐渐工作到高效率区；另一方面，升力的增长影响附着条件，轮胎能提供的驱动力越来越小；高速下螺旋桨由提供驱动力成为了必然选择。

综合考虑整个起飞阶段，当速度较低时由轮胎提供驱动力，使得整车能够更快的加速；当轮胎驱动力不足时改由螺旋桨驱动，直至达到起飞速度，即双模式驱动，较之单纯的螺旋桨驱动，更能充分利用发动机功率，起飞的加速时间和滑跑距

收稿日期：2017-05-05；**录用日期**：2017-06-30；**网络出版时间**：2017-09-13 11:46
网络出版地址：kns.cnki.net/kcms/detail/11.2625.V.20170913.1146.001.html
* **通信作者**. E-mail：xbacb@buaa.edu.cn

引用格式：*徐斌，田富刚. 基于双模式驱动的飞行汽车起飞阶段动力匹配分析[J]. 北京航空航天大学学报，2018，44(4)：662-669. XU B，TIAN F G. Power matching of flying cars during takeoff stage based on dual-mode driving[J]. Journal of Beijing University of Aeronautics and Astronautics，2018，44(4)：662-669 (in Chinese).*

北航学报 赠阅

离也都将得到优化。因而对于双模式驱动的研究具有重要意义。

1.2　最佳切换时刻

采用双模式驱动的控制策略可以充分利用发动机功率，2 种驱动模式切换时刻的选择成为关键。为了更好地分析这个问题，需要定义“最佳切换时刻”的概念，即对于采用双模式驱动起飞策略的飞行汽车，以在起飞加速过程中能够保持正常的行驶状态为前提，起飞过程中必然存在一个或几个时间点，在该时间点将发动机动力输出由轮胎转换至螺旋桨，可以使起飞过程加速时间、滑跑距离最短。将这一个或几个时间点定义为最佳切换时刻。本文的理论计算和优化过程都将以最佳切换时刻作为对象，后续将不再赘述。

2　动力学分析

2.1　双模式驱动行驶状态的分类

按照双模式驱动的思想，飞行汽车在运行过程中大致可以分为 3 种状态：

1）地面运行。机翼不展开，此时的受力情况与普通的汽车类似，发动机的扭矩通过轮胎转变为驱动车辆前进的驱动力。

2）起飞阶段。轮胎驱动，机翼展开，与地面运行状态的主要区别在于升力，主要出现在起飞滑跑的前半段。

3）螺旋桨驱动。机翼展开，受力情况与普通飞机类似，应用在起飞阶段及飞行阶段。

图 1 给出了飞行汽车运行过程中可能受到的所有力[6-7]。图中水平方向上：F_t 和 F_t'，分别为轮胎和螺旋桨提供的驱动力；飞行汽车为后驱，故轮胎驱动力作用于后轮；F_{f1} 和 F_{f2} 分别为前后轮胎受到的滚动阻力；F_w 为整车受到的迎风阻力。F_j 为整车受到的加速阻力。竖直方向上：G 为整车受到的重力；N_1、N_2 分别表示前轮、后轮受到的地面支持力；F_L 为机翼提供的升力。

本文在分析时，为简便忽略了一些实际问题的因素，对受力模型进行了如下简化：

1）认为前后轮载荷平均分配，忽略由于质心不平分轴距带来的一系列影响，该影响包括但不限于：后轮载荷所决定的附着力的限制、前后轮滚动阻力的大小、起飞过程中整车姿态变化导致的升阻力特性变化等。

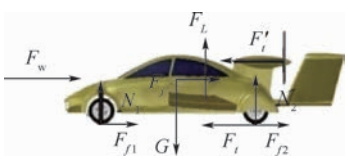


图 1　飞行汽车起飞阶段受力示意图[6-7]

Fig. 1　Force diagram of flying car during takeoff stage[6-7]

2）不针对力矩做分析，将车速达到起飞车速（升力、重力相平衡）作为分析的终点。

3）不考虑爬坡阻力 F_i 的影响，认为在起飞过程中沿着水平路面行驶。

2.2　飞行汽车动力学建模

本文参考一般车辆的受力分析特征[6-7]，建立的飞行汽车动力学模型为

水平方向：

$$F_t + F_t' = F_f + F_w + F_j \tag{1a}$$

竖直方向：

$$N = G - F_L \tag{1b}$$

将水平方向的各项计算过程展开，得

$$\frac{T_{tq} i_g i_0 \eta_T}{r} + C_T \rho_a D^4 n^2 = \left(G - \frac{1}{2} C_l A_l \rho_a v_a^2\right) f + \frac{1}{2} C_D A \rho_a v_a^2 + \delta m \frac{\mathrm{d}v_a}{\mathrm{d}t} \tag{2}$$

式中：T_{tq} 为发动机输出的转矩；i_g 为各挡位传动比；i_0 为主减速器传动比；η_T 为传动系统机械效率；r 为车轮半径；C_T、C_l 和 C_D 分别为螺旋桨拉力系数、机翼升力系数和迎风阻力系数；D 为螺旋桨直径；n 为螺旋桨转速；A_l 为机翼投影面积；f 为滚动阻力系数，根据路面状况的不同会有很大区别；A 为换算的迎风面积；ρ_a 为地面状况的空气密度，考虑到飞行汽车飞行高度不超过 2 000 m，在计算过程中认为空气密度在整个运行范围内变化不大；δ 为旋转质量换算系数，由经验公式求得[7]；m 为整车质量；v_a 为整车前进速度，在不考虑风速的情况下，整车相对于空气的速度等于整车的前进速度。

需要说明的是，3 种形式模式下的受力情况并不相同，实际计算中应用的受力模型如下：

1）地面运行。

$$\begin{cases} F_t = F_f + F_j + F_w \\ N = G \end{cases} \tag{3}$$

F_L 在这一阶段忽略。

2）起飞阶段。纯轮胎驱动。

$$\begin{cases} F_t = F_f + F_j + F_w \\ N = G - F_L \end{cases} \tag{4}$$

3）螺旋桨驱动。

$$\begin{cases} F_t' = F_f + F_j + F_w \\ N = G - F_L \end{cases} \tag{5}$$

北航学报 赠阅

3 动力匹配计算

根据 2.2 节所建立的动力学模型,对该飞行汽车进行动力匹配计算[8-13]。飞行汽车的基本参数如表 1 所示。

表 1 飞行汽车基本参数

Table 1 Main parameters of flying car

参 数	数 值
载荷 W/N	6 500
起飞速度/(km·h^{-1})	110
升阻比 K	10
迎风面积 A/m^2	2
车轮半径 r/m	0.52
螺旋桨传动比 i	2.43
螺旋桨直径 D/m	1.8
机翼投影面积 A_l/m^2	9

3.1 发动机选型计算

对于飞行汽车来说,发动机的选型由平飞时所需的功率作为重要参考。

飞行器巡航所需的推力为

$$F_T = \frac{Ph}{V} = \frac{W}{K} \tag{6}$$

式中:h 为稳定飞行的高度;V 为飞行速度。

提供这些推力所需的功率为

$$P = \frac{WV}{hK} \tag{7}$$

将飞行汽车的参数代入式(7),得 P = 50.56 kW。

本文选用的发动机外特性如图 2 所示。

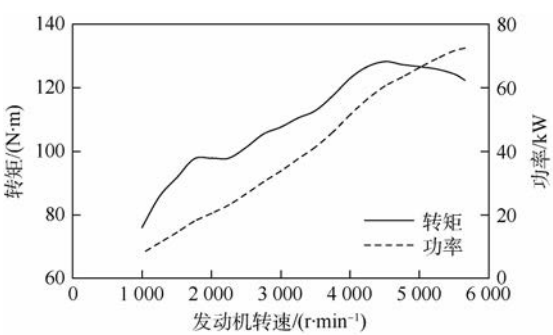


图 2 发动机外特性曲线

Fig. 2 External characteristic curves of engine

3.2 传动比计算

3.2.1 主减速器传动比

主减速器传动比根据最高挡(直接挡)下的最高车速进行计算,根据动力学模型:$F_t = F_f + F_w + F_j$,最高车速运行状态,加速阻力 F_j 为 0,则

$$\begin{cases} F_t = \dfrac{T_{tq}\eta_T}{r} i_0 \\ F_w = \dfrac{1}{2} C_D A r v_a^2 \\ v_a = \dfrac{n}{i_0} 2\pi r \\ F_f = Wf \end{cases} \tag{8}$$

利用发动机相关数据:最大转矩为 127 N·m,最大转矩转速为 5 500 r/min,即可求得主减速器传动比 i_0。

3.2.2 一挡传动比

一挡传动比根据最大爬坡度的要求,根据动力学模型:$F_t = F_f + F_w + F_j + F_i$,爬坡过程认为保持很低车速匀速前进,$F_w$、$F_f$ 均认为是 0,则

$$\frac{T_{tq} i_0 \eta_T}{r} i_1 = Wf\cos\theta + W\sin\theta \tag{9}$$

式中:θ 为爬坡坡度。

根据主减速器传动比及爬坡度要求,即可求得一挡传动比 i_1。

3.2.3 中间挡传动比

对于中间挡,当一挡和最高挡确定后,通常按照等比级数分配。同时,为避免各挡之间换挡困难,相邻两挡之间的传动比一般不超过 1.7~1.8。

根据以上原则,用于匹配选定发动机的传动系统传动比为:$i_0 = 3.2$;$i_1 = 2.4$;$i_2 = 1.8$;$i_3 = 1.34$;$i_4 = 1$。

4 双模式驱动行驶状态仿真

在节气门全开状态下,汽车原地起步连续换挡加速时间为

$$T(x) = \int_0^{v_{amin}} \frac{\delta W}{3.6g(F_t - F_f - F_w)} \mathrm{d}v_a + \sum_{n=1}^{k} \int_{v_{amin}}^{v_{amax}} \frac{\delta W}{3.6g(F_t - F_f - F_w)} \mathrm{d}v_a + \sum_{n=1}^{k-1} \left[\Delta t_m + \int_{v'_{an}}^{v_{an}} \frac{\delta W}{3.6g(F_t - F_f - F_w)} \mathrm{d}v_a \right] \tag{10}$$

式中:第 1 项为车辆原地起步阶段,即车速从 0 增加到最低稳定车速时所需的时间,第 2 项为从最低稳定车速连续换挡加速到最高车速所需要的时间,主要取决于换挡规律,第 3 项为换挡过程中所需要的时间;g 为重力加速度,计算时取为 9.8 m/s^2;Δt_m 为司机换挡所需的反应时间,根据经验取 0.2~0.5 s,本文在计算过程中取为 0.5 s。

以上述加速时间计算公式为基础,根据理论受力模型和参数选取,利用 Simulink 建立了模拟

北航学报 赠阅

行驶状态的仿真模型[14-16]。

仿真模型有四大部分，分别对应计算轮胎驱动力、螺旋桨驱动力、迎风阻力和滚动阻力，根据受力计算前进的加速度，进而通过积分求得整车的加速时间和滑跑距离。加速时间按照式(10)计算，仿真中计算连续换挡加速时间和换挡过程的延迟时间。需要说明的是，由于原地加速时间受离合器接合规律的影响，很难在仿真模型中体现，因此在仿真模型中，忽略了这部分加速时间。对于最低稳定车速的选取，参考同等整备质量的车辆。

5　双模式驱动控制及最佳切换时刻分析

5.1　最佳切换时刻的选取

2.1 节进行受力分析时提到，起飞阶段驱动与地面运行最大的区别在于升力的影响，升力随车速的提高而增大，受此影响轮胎与地面的附着力减小，轮胎能提供的驱动力也逐渐变小。当车速增大到某一值时，达到轮胎的打滑界限，如图 3 所示。

起飞滑跑开始阶段，螺旋桨的效率低，此时轮胎驱动力远高于对应发动机转速下螺旋桨提供的推力。随着发动机转速和车速的提升，螺旋桨提供的推力逐渐增大，而轮胎驱动力由于传动比以及升力的影响，呈现阶梯状减小的趋势。轮胎与螺旋桨所能提供的驱动力，在二挡末相等。

综合以上 2 个因素，初步确定在发动机由二挡切换到三挡时，脱开轮胎驱动，转为纯螺旋桨驱动。

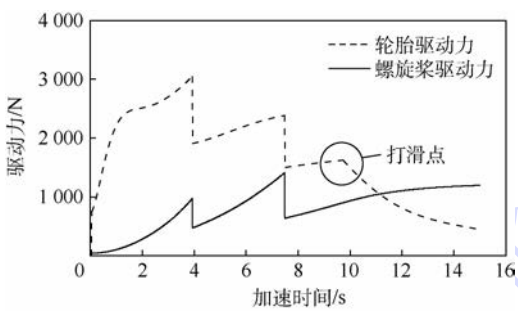


图 3　起飞阶段轮胎驱动力与对应发动机转速下螺旋桨驱动力

Fig. 3　Wheel driving force and propeller driving force under the same engine speed during takeoff stage

5.2　双模式驱动与纯螺旋桨驱动对比

在同样的工况下(初速度、目标速度、载荷等)，分别计算采取双模式驱动策略以及纯螺旋桨驱动 2 种起飞方式的加速时间和起飞滑跑距离，最终结果如图 4 所示。

加速时间及起飞滑跑距离计算结果如表 2 所示。

从表 2 中可以看出，采用双模式的驱动策略，可以有效提升起飞阶段的动力性能。

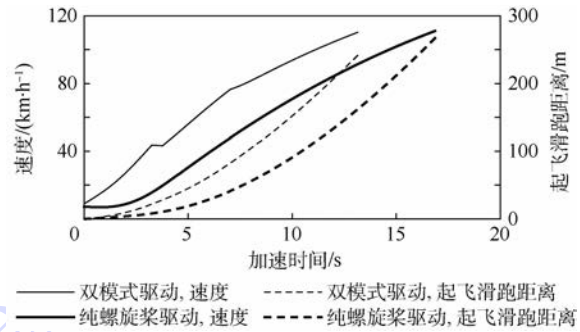


图 4　起飞过程加速曲线

Fig. 4　Accelerating curves during takeoff stage

表 2　双模式驱动与纯螺旋桨驱动结果对比

Table 2　Comparison of results of dual-mode driving and pure propeller driving

驱动方式	加速时间/s	起飞滑跑距离/m
纯螺旋桨驱动	16.87	276
双模式驱动	13.15	240
优化率/%	22.05	13.04

5.3　双模式驱动下各参数对起飞性能的影响

5.3.1　传 动 比

对于飞行汽车来说，不同的传动比会导致 2 个因素发生改变。以一挡传动比为例，较大的传动比使得：①一挡期间轮胎的驱动力更大；②一挡换二挡时的车速更低，使得飞行汽车在一挡加速持续时间更短。对于缩短起飞滑跑距离来说，前者有利，后者不利，因此从定性分析的角度很难判断传动比的变化对结果产生的影响。

通过仿真计算，一挡、二挡传动比变化对起飞性能的影响结果如图 5 所示。

从图 5(a)可以看出，当二挡传动比不变时，随着一挡传动比的增大，起飞滑跑距离呈现先递减后递增的变化趋势，即对于不同的二挡传动比，都有一个最佳的一挡传动比使得起飞滑跑距离最短。而且可以看出，随着二挡传动比的增大，该最佳一挡传动比值也逐渐增大。本文讨论的飞行汽车，当二挡传动比从 1.70 增大到 1.85 时，最佳的一挡传动比从 2.6 增大到 2.7。

从图 5(b)可以看出，当一挡传动比不变时，随着二挡传动比的增大，起飞滑跑距离呈现先递减后递增的变化趋势，即对于不同的一挡传动比，也存在一个最佳的二挡传动比使得起飞滑跑距离最短。这与一挡传动比对起飞滑跑距离的影响规

北航学报 赠阅

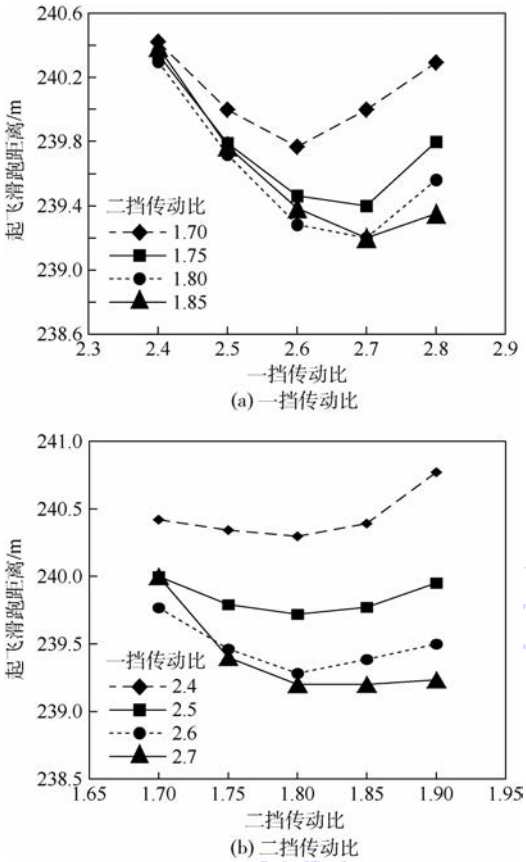


图 5 传动比对起飞滑跑距离的影响

Fig. 5 Influence of transmission ratio on takeoff running distance

律一致。

通过优化分析看出,一挡选择 2.7、二挡选择 1.80 是飞行汽车传动比的更优选择。然而,由于总体来说起飞滑跑时间比较短,传动比的优化对于缩短起飞滑跑距离产生的作用很有限(仅有 0.5%),所以在进行传动比的选择时,应以概念设计为核心,动力性要求可以适当弱化。

5.3.2 换挡转速

换挡转速的影响受发动机特性影响较大,对于本文选用的发动机,其峰值转矩出现在 4 800 r/min 前后,更高的换挡转速可以充分利用发动机的高转矩区;同时,轮胎提供的驱动力大于纯螺旋桨驱动力,换挡转速越高,则轮胎驱动的持续时间越长,相应的飞行汽车起飞滑跑距离越短。其影响具体如图 6 所示。

如图 6 所示,增大换挡转速能缩短约 5% 的加速时间和起飞滑跑距离,所以,在起飞阶段尽可能延迟换挡有益于提升动力性能。同时也应该注意到,转速越高意味着车速也会越快,有可能造成轮胎打滑,所以换挡转速不能无限制的提高。

5.3.3 整车设计参数

整车质量对加速时间和起飞滑跑距离的影响如图 7 所示。整车质量越重,所需加速时间越长,起飞滑跑距离越长。

整车所受的阻力包括滚动阻力和迎风阻力,其所占比重变化规律如图 8 所示。

随着车速的增大,迎风阻力逐渐成为主要因素,对于飞行汽车来说,从设计上减小迎风阻力也是提升动力性能的重要途径。

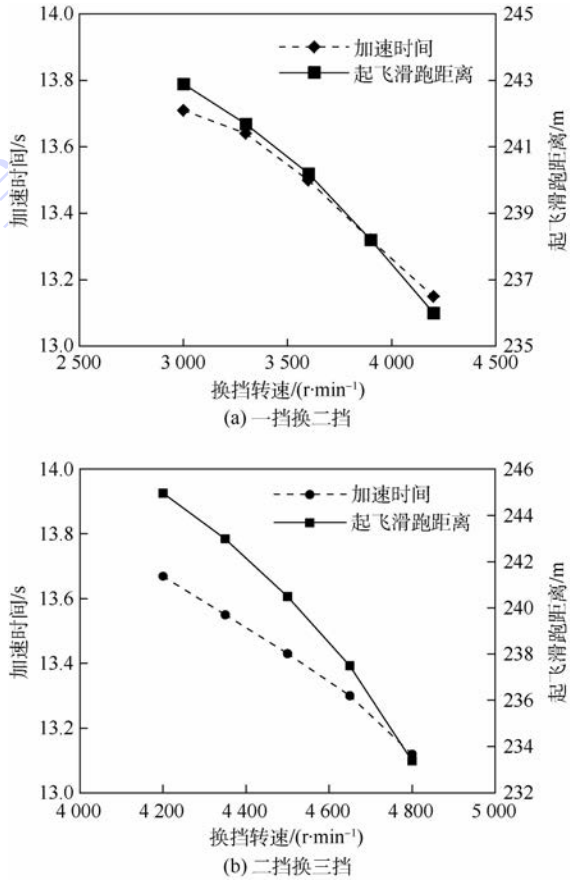


图 6 换挡转速对加速时间和起飞滑跑距离的影响

Fig. 6 Influence of shifting speed on acceleration time and takeoff running distance

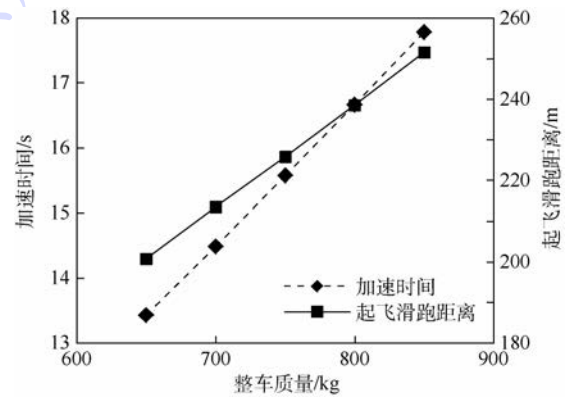


图 7 整车质量对加速时间和起飞滑跑距离的影响

Fig. 7 Influence of vehicle quality on acceleration time and takeoff running distance

北航学报 赠阅

整车质量为 800 kg 时阻力系数变动对起飞性能的影响如图 9 所示。阻力系数越大，起飞的加速时间越长，滑跑距离越大。

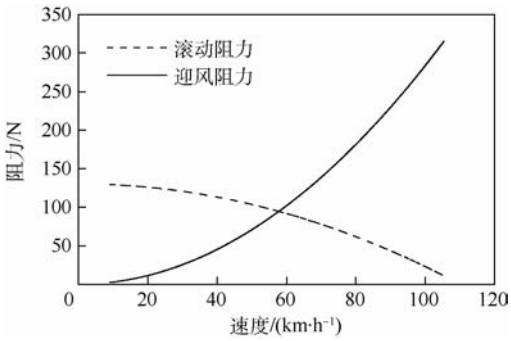


图 8　滚动阻力和迎风阻力与车速的关系

Fig. 8　Relationship between wheel and windward resistance and vehicle speed

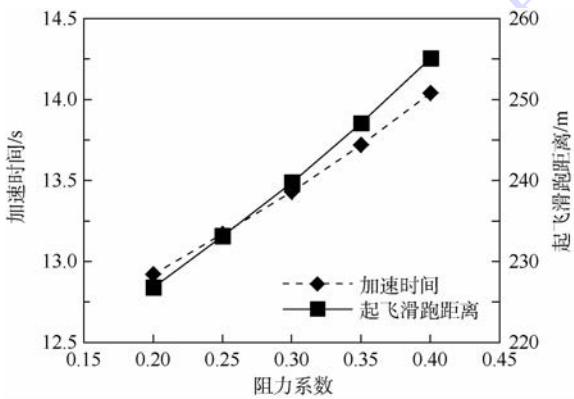


图 9　阻力系数对加速时间和起飞滑跑距离的影响

Fig. 9　Influence of drag coefficient on acceleration time and takeoff running distance

5.4　发动机特性对起飞性能的影响

路上行驶与空中飞行，不单单是 2 种运行状态，两者所处的环境条件、发动机工况、正常运行状态所需要的适应性系数等参数也有很大区别[17]。一般来说，乘用车的发动机经常工作在中低转速，而飞行器的发动机通常工作在高转速。本文选用的发动机转矩峰值出现在 4 800 r/min，高于一般的乘用车发动机。

选取适应性系数不同的发动机外特性曲线进行对比，分析发动机适应性系数对加速性能产生的影响。

用于计算的外特性曲线如图 10 所示。

分别以图 10 中 4 条特性曲线作为发动机外特性，按照双模式驱动的控制策略作为起飞策略进行计算，图 11、图 12 分别为加速过程中加速时间和起飞滑跑距离的对比。

由图 11、图 12 可以看到，当适应性系数增大时，起飞性能得到提升；适应性系数在2.2附近

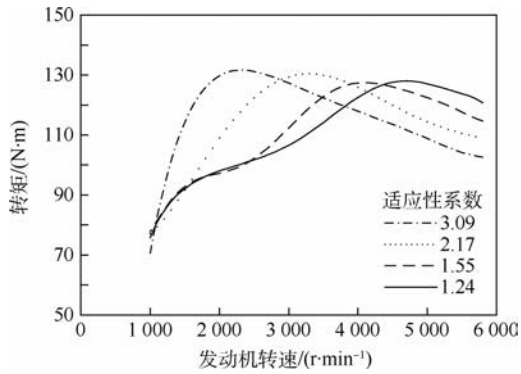


图 10　不同发动机转矩特性曲线

Fig. 10　Different engine torque characteristic curves

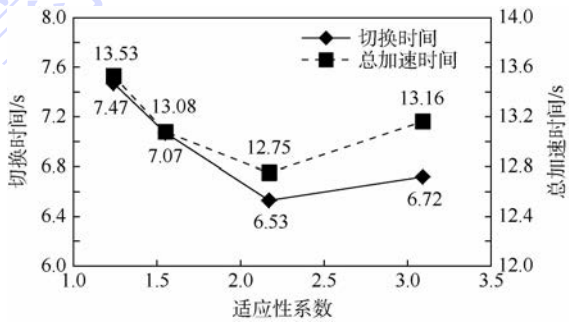


图 11　发动机适应性系数对切换时间和总加速时间的影响

Fig. 11　Influence of engine adaptability coefficient on switching time and take-off acceleration time

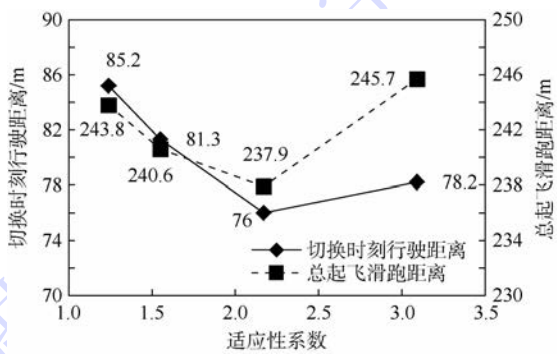


图 12　发动机适应性系数对切换时刻行驶距离和总起飞滑跑距离的影响

Fig. 12　Influence of engine adaptability coefficient on running distance at switching time and running distance of take-off stage

时，起飞滑跑距离最短，加速到起飞速度所需的时间也最少；在最高转矩和功率相差不大的情况下，匹配程度不同的发动机在切换前的加速时间和起飞滑跑距离甚至相差 10% 以上。分析原因，因为在起飞阶段，特别是轮胎驱动阶段发动机的常用工作转速在 2 000 ~ 4 000 r/min，过高或过低的峰值转速，都会因为扭矩特性的峰值区不能得到很好的利用而造成动力性能的恶化。当发动机的高

北航学报 赠阅

转矩区向低速区域移动时,发动机的转矩能够更好地被利用,故而加速性能得到了提升。综上,对于发动机外特性的考虑,需要综合 2 种情况:①如果侧重起飞机动性,则应该在满足飞行工况的前提条件下,尽可能地使发动机的高转矩区向低转速方向靠拢,提高适应性系数;②如果侧重飞行安全性,则应该使发动机的高转矩区域向飞行额定转速靠近,适应性系数适当减小,采取其他方式来提升起飞性能。

6 结 论

1)本文采用双模式的驱动策略,相比纯螺旋桨驱动,能有效提升起飞阶段的动力性能,缩短起飞滑跑阶段的时间和距离。

2)最佳切换时刻出现在轮胎驱动力和螺旋桨驱动力相等的位置;考虑到换挡产生的驱动力变化,以换挡点作为切换点有利于改善行驶稳定性。

3)在进行传动系统参数的选取时,应以起飞阶段轮胎驱动工况作为对象进行匹配和优化设计,尤其要注意的是,因为升力的影响,在进行挡位和换挡策略的设计时,要以避免车轮打滑作为边界条件。

4)对于飞行汽车,空气动力性能的影响较一般汽车而言更加明显,迎风阻力在起飞阶段成为主要阻力源。

5)与飞行汽车相匹配的发动机,适应性系数应根据起飞阶段发动机常用工况进行选取;与车用发动机不同,较大的适应性系数,不仅对提高起飞动力性无益,还会使飞行阶段工作恶化。

参考文献(References)

[1] 李根,马铁林,林海英,等. 飞行汽车研究进展及技术难点[C]//探索创新交流:中国航空学会青年科技论坛文集:第7集. 北京:航空工业出版社,2016:133-139.
LI G,MA T L,LIN H Y,et al. Research development and technical difficulties of flying car[C]//Exploration,Innovation,Exchange:Essays on the Youth Science and Technology Forum of China Aviation Society:Seventh. Beijing:Aeronautical Industry Press,2016:133-139(in Chinese).

[2] RAJASHEKARA K,WANG Q,MATSUSE K. Flying cars:Challenges and propulsion strategies[J]. IEEE Electrification Magazine,2016,4(1):46-57.

[3] 朱保利,程磊,吴恢鹏. 飞行汽车概念设计与气动特性分析[J]. 机械工程师,2014(5):87-89.
ZHU B L,CHENG L,WU H P. Conceptualdesign and aerodynamic characteristics analysis of flying car based on CFD[J]. Mechanical Engineer,2014(5):87-89(in Chinese).

[4] 王拖连,杨世文,薛姣,等. 陆空两用飞行汽车发展现状与展望[J]. 公路与汽运,2011(4):12-16.
WANG T L,YANG S W,XUE J,et al. The development status and prospect of airphibian flying car[J]. Highway and Automotive Applications,2011(4):12-16(in Chinese).

[5] 刘沛清. 空气螺旋桨理论及其应用[M]. 北京:北京航空航天大学出版社,2006.
LIU P Q. Air propeller theory and its application[M]. Beijing:Beihang University Press,2006(in Chinese).

[6] 邹晓敏,刘刚. 车辆道路行驶阻力的模拟及测量[J]. 质量与认证,2012(12):58-61.
ZOU X M,LIU G. Simulation and measurement of driving resistance of vehicle road[J]. China Quality Certification,2012(12):58-61(in Chinese).

[7] 李晓甫,赵克刚,黄向东,等. 汽车行驶阻力模型参数的确定[J]. 汽车工程,2011,33(8):645-648.
LI X F,ZHAO K G,HUANG X D,et al. Determination of parameters in vehicle driving resistance model[J]. Automotive Engineering,2011,33(8):645-648(in Chinese).

[8] 张杰,吴森堂. 一种变体飞行器的动力学建模与动态特性分析[J]. 北京航空航天大学学报,2015,41(1):58-64.
ZHANG J,WU S T. Dynamic modeling for a morphing aircraft and dynamic characteristics analysis[J]. Journal of Beijing University of Aeronautics and Astronautics,2015,4(1):58-64(in Chinese).

[9] 林学东. 汽车动力匹配技术[M]. 北京:中国水利水电出版社,2010.
LIN X D. Vehicle power matching technology[M]. Beijing:China Water and Power Press,2010(in Chinese).

[10] 杨连生. 内燃机性能及其与传动装置的优化匹配[M]. 北京:学术期刊出版社,1988.
YANG L S. Performance of internal combustion engine and its optimal matching with transmission[M]. Beijing:Academic Periodical Press,1988(in Chinese).

[11] 曲建清,宋辉. 无人机总体与起飞性能匹配性设计优化[J]. 飞行力学,2013,31(3):225-229.
QU J Q,SONG H. Design optimization of UAV configuration and takeoff performance[J]. Flight Dynamics,2013,31(3):225-229(in Chinese).

[12] ZHOU M,ZHAO L,ZHANG Y,et al. Pure electric vehicle power-train parameters matching based on vehicle performance[J]. International Journal of Control & Automation,2015,8(9):53-62.

[13] ZHANG L,HAO G,YANG X,et al. The electric vehicle power design and the matching characteristics analysis of the transmission system[J]. Telkomnika Indonesian Journal of Electrical Engineering,2013,11(11):6352-6357.

[14] 吴大卫,李寒冰,李书,等. 基于仿真模型的短距起飞性能优化[J]. 北京航空航天大学学报,2014,40(6):756-761.
WU D W,LI H B,LI S,et al. Short takeoff performance optimization based on simulation model[J]. Journal of Beijing University of Aeronautics and Astronautics,2014,40(6):756-761(in Cinese).

[15] 张威,张景海,隗海林,等. 汽车动力学仿真模型的发展[J]. 汽车技术,2003(2):1-4.

北航学报 赠阅

ZHANG W, ZHANG J H, HUAI H L, et al. Development of vehicle dynamic simulation model[J]. Automobile Technology, 2003(2):1-4(in Chinese).

[16] 冯超. 基于 Matlab/Simulink 的电动汽车仿真模型设计与应用[D]. 北京:中国科学院大学,2013.
FENG C. The simulation model design and application of the electric vehicle based on Matlab/Simulink[D]. Beijing: University of Chinese Academy of Sciences, 2013(in Chinese).

[17] 张雪文,徐明亮,杨欣. 汽油发动机的动力特性分析[J]. 机械制造与自动化,2012,41(1):47-49.
ZHANG X W, XU M L, YANG X. Dynamic characteristics of gasoline engines[J]. Machine Building Automation, 2012, 41(1):47-49(in Chinese).

作者简介:

徐斌　男,硕士,教授。主要研究方向:发动机涡轮增压及控制技术。

田富刚　男,硕士研究生。主要研究方向:飞行汽车概念设计及动力匹配。

Power matching of flying cars during takeoff stage based on dual-mode driving

XU Bin*, TIAN Fugang

(School of Transportation Science and Engineering, Beijing University of Aeronautics and Astronautics, Beijing 100083, China)

Abstract: To solve power-matching problem during takeoff stage of folding wing flying car, the basic dynamic control strategy was studied. The dual-mode driving characteristics during takeoff stage were analyzed through theoretical calculation, and the concept of the optimal switching time was proposed. Based on the basic parameters and results of power-matching calculation of the concept car, the driving state simulation models under different working conditions were established in Simulink. According to the driving state simulation models, the simulation analyses on driving state during takeoff stage were conducted, and the dual-mode driver selection principle and optimum matching scheme of switching time were obtained. The calculation results show that the takeoff acceleration time is reduced by 22% and the takeoff distance is shortened by 13% based on dual-mode driving. The optimization analyses on transmission and vehicle parameters and variation of output characteristics were further conducted, from which the effect of the parameters on power performance during takeoff stage was obtained.

Keywords: flying car; dual-mode driving; optimal switching time; Simulink simulation; parameter analysis

Received: 2017-05-05; **Accepted**: 2017-06-30; **Published online**: 2017-09-13 11:46

URL: kns.cnki.net/kcms/detail/11.2625.V.20170913.1146.001.html

* **Corresponding author**. E-mail: xbacb@buaa.edu.cn

北京航空航天大学学报

北航学报 赠阅

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2017.0280

基于稀疏傅里叶变换的快速捕获方法

全下文载

张春熹，李先慕，高爽*
（北京航空航天大学 仪器科学与光电工程学院，北京 100083）

摘 要：卫星信号捕获是接收机内基带信号处理的关键步骤，捕获的速度严重影响接收机首次定位的时间(TTFF)。在基于快速傅里叶变换(FFT)的并行码相位捕获(PCPS)方法基础上，将具有亚线性运算量的稀疏傅里叶变换(SFT)引入捕获过程中，提出了基于SFT的快速捕获方法，通过优化捕获过程中相关运算的效率，提高捕获速度。仿真结果表明，与传统的基于FFT的捕获方法相比，基于SFT的快速捕获方法运算效率提高到原来的2倍，更好地满足了卫星信号对快速捕获的要求。

关 键 词：卫星信号；快速捕获；稀疏傅里叶变换(SFT)；混叠；降采样
中图分类号：TN927
文献标识码：A **文章编号**：1001-5965(2018)04-0670-07

卫星信号捕获是卫星接收机内基带信号处理的第一步，其目的是完成可见卫星的确认并估算出接收信号码相位和载波频率的粗略值，为跟踪通道提供参数的初始化值。常用的捕获方法有基于时域的串行搜索、基于频域的并行频率搜索和基于频域的并行码相位搜索[1-2]。基于时域的串行搜索方法硬件实现复杂度低，但是其所需运算量大、花费时间长；基于频域的并行频率搜索在载波频率方向上进行并行搜索，而在码相位方向上进行串行搜索，捕获的运算量在载波频率方向减小，缩短了捕获时间；基于频域的并行码相位搜索是目前软件接收机平台中常用的捕获方法，通过快速傅里叶变换(Fast Fourier Transform, FFT)实现在码相位方向上的并行搜索，大幅度提高了卫星信号捕获的速度。

在基于频域的并行搜索方法中，FFT运算是实现卫星信号快速捕获的核心[3-4]，为了改善卫星信号捕获性能，产生了基4FFT和分裂基FFT算法[5-7]，通过优化FFT相关运算的效率，提高了捕获速度；但是这些优化的FFT算法的整体时间复杂度仍然是以点数、点数对数的乘积为渐近上限，在运算时间和捕获性能上并没有明显的改善。

由于卫星的伪随机码具有强自相关性，当本地复制伪码和输入信号的伪码相位接近时才能产生很强的相关值，其余搜索单元的相关值都接近于零以致可以忽略。因此对于某颗卫星的二维搜索而言，捕获的相关值具有稀疏的特性，从而可以将稀疏傅里叶变换(Sparse Fourier Transform, SFT)引入并行搜索过程中。

SFT是由MIT的Hassanieh等[8]于2012年提出，其核心思想是利用信号在频域中的稀疏特性，通过加窗—混叠—重构等步骤实现算法的亚线性处理。近年来，SFT在语音处理、医学成像和通信等领域得到应用[9-10]，而在卫星信号捕获中的应用主要是理论分析及仿真验证。文献[11]提出了利用SFT进行GPS信号捕获的方法，并进行了可行性分析，但是没有给出具体的捕获方案和结果。文献[12-13]提出了一种基于SFT的快速捕获与干扰抑制联合技术，简化了算法中快速傅里叶反变换(Inverse Fast Fourier Transform, IFFT)的

收稿日期：2017-05-05；**录用日期**：2017-07-13；**网络出版时间**：2017-09-15 17:09
网络出版地址：kns.cnki.net/kcms/detail/11.2625.V.20170915.1709.001.html
* **通信作者**. E-mail：gaoshuang@buaa.edu.cn

引用格式：*张春熹，李先慕，高爽. 基于稀疏傅里叶变换的快速捕获方法[J]. 北京航空航天大学学报，2018，44(4)：670-676.*
ZHANG C X, LI X M, GAO S. Fast acquisition methods based on sparse Fourier transform[J]. Journal of Beijing University of Aeronautics and Astronautics, 2018, 44(4): 670-676 (in Chinese).

北航学报 赠阅

计算量，但是没有简化输入信号与本地信号 FFT 算法的运算量。文献[14]利用 SFT 实现了 GPS 信号的快速捕获，并分析了降采样因子对捕获速度的提升，但是其对降采样因子的选取不太合理。

本文通过分析卫星信号捕获中的相关值具有稀疏的特性，在基于频域 FFT 的并行码相位的基础上，提出了基于 SFT 的快速捕获方法，利用实测的中频数据通过卫星软件接收机进行了捕获实验的验证，实验结果表明本文方法可以降低捕获运算的复杂度，提高捕获效率。

1　卫星信号捕获过程分析

卫星接收机中的信号捕获是一个三维搜索：卫星号（PRN 码）、码相位和载波频率[15]。当本地码和信号的码相位对齐的情况下，并且本地载波频率和输入信号的载波频率接近，才能产生很强的相关值。图 1 为某颗 GPS 卫星信号捕获成功时的各个搜索单元上的 IFFT 结果。很明显，在该卫星的二维搜索范围中出现了一个峰值，其不但明显高于其他各个搜索单元的积分值，而且也超过了捕获门限，可以说明对该卫星的信号已被成功捕获，同时也证明了捕获结果的输出值具有稀疏的特性。

基于 FFT 的并行码相位捕获方法通过 FFT 实现频域的循环相关，在码相位空间内实现并行搜索捕获功能，卫星信号捕获的效率得到极大的提高。由于 FFT 相关软件编程模块已经成熟并广泛应用，基于 FFT 的并行码相位捕获方法成为卫星软件接收机中常用的捕获方法，其实现原理框图如图 2 所示。

采用基于频域 FFT 并行码相位捕获方法，其实现的运算步骤如下：

步骤 1　将数字中频信号和本地载波发生器输出的正弦载波 Q_{NCO} 和余弦载波 I_{NCO} 混频，分

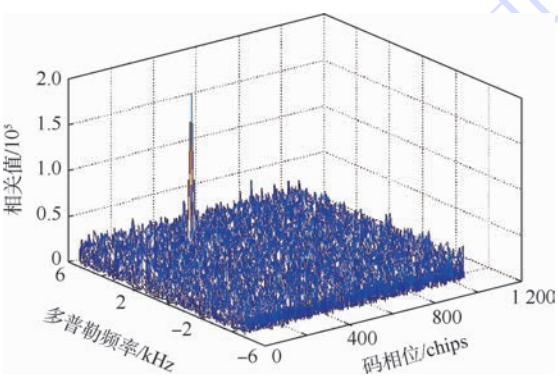


图 1　某颗 GPS 卫星捕获成功结果

Fig. 1　Successful acquisition result of a GPS satellite

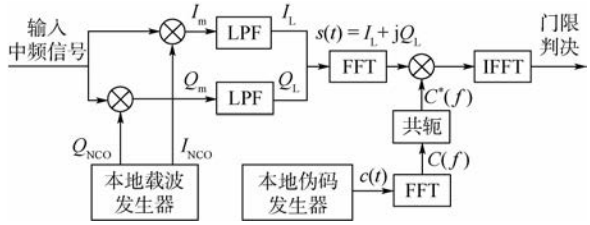


图 2　基于 FFT 的并行码相位捕获方法框图

Fig. 2　Block diagram of parallel code phase search method based on FFT

别得到同相 I 支路信号 I_{m} 和正交 Q 支路信号 Q_{m}，以 I 支路为实部、Q 支路为虚部得到基带的复信号 $s(t)$。

步骤 2　对步骤 1 中得到的基带复信号 $s(t)$ 做 FFT 变换。

步骤 3　对本地伪码发生器输出的伪码信号 $c(t)$ 做 FFT 得到 $C(f)$，并求复数共轭得到 $C^*(f)$。

步骤 4　将步骤 2 和步骤 3 得到的结果相乘并做 IFFT 变换。

步骤 5　对步骤 4 得到的 IFFT 结果取模，得到相关值，对相关值进行门限检测。如果相关值的峰值超过捕获门限值，则捕获到了信号，并得到了信号的码相位和载波频率值；反之，如果相关值的峰值不足够强，信号没有捕获成功，需要重新设置本地载波发生器的频率，重复步骤1 ~ 步骤 4。

上述方法在完成一个处理（2 次 FFT 和 1 次 IFFT）后能够得到全部伪码相位的结果，所以实现了伪码域的并行处理，而多普勒频率维度的搜索依然是串行的，即每次改变本地载波发生器的频点，则需要完成一次新的处理，即 2 次 FFT、1 次IFFT 和 N 次乘加运算。

设复数输入信号的长度为 N，其傅里叶变换运算量为 $\frac{N}{2}\mathrm{lb}\ N$ 次复数乘法和 $N\mathrm{lb}\ N$ 次复数加法，得到单次相关结果共需要 $N\left(1+\frac{3}{2}\mathrm{lb}\ N\right)$ 次复数乘法和 $3N\mathrm{lb}\ N$ 次复数加法。

2　稀疏傅里叶变换原理

SFT 是利用信号在频域中的稀疏特性，简化 FFT 运算过程的一种快速有效的算法。在信号具有稀疏性且关键信息处在较大谱峰上时，SFT 技术可以在获得与传统 FFT 相似的频谱分析效果的同时降低其运算复杂度。随着 FFT 点数的增大，SFT 具有的非线性运算复杂度特性将使其在运算方面的优势变得更加明显。

北航学报 赠阅

2.1 频域降采样

SFT 算法应用了信号在频域中的“稀疏”特性,利用频域降采样减少信号输入长度,提高 FFT 运算过程的效率。

设 $x(n)$ 是长度为 n 的离散序列,$X(k)$ 为其频域表示,其时域混叠结果为

$$y_i = \sum_{j=0}^{n/B-1} x_{i+Bj} \qquad i \in [0, B-1] \tag{1}$$

式中:整数 B 能整除 n。令 $p = n/B$,则有

$$Y(k) = X(kn/B) \qquad k \in [0, B-1] \tag{2}$$

由式(1)、式(2)可得,时域的混叠对应频域的降采样,反之,时域的降采样引起频域的混叠。图 3 阐释了这一对应关系。

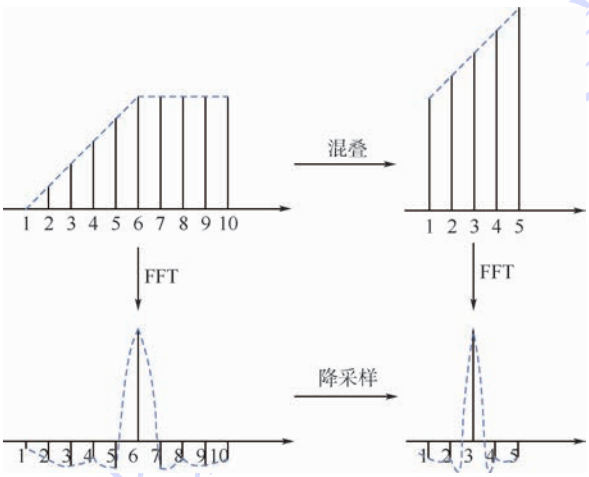


图 3 混叠和降采样示意图

Fig. 3 Schematic of aliasing and sub-sampling

2.2 稀疏傅里叶变换运算过程

SFT 包括频域重排、窗函数滤波、频域降采样、定位、估值及迭代等运算过程,如图 4 所示。

1)频域重排。取随机参数 σ、τ(τ 为奇数)对信号进行重排,$q(n) = x(\sigma n + \tau)$,则 $Q(\sigma k) = X(k) W_N^{-\tau k}$,$W_N = e^{-j2\pi/N}$ 为离散傅里叶变换的根因子。



图 4 SFT 运算过程

Fig. 4 Calculation process of SFT

2)窗函数滤波。将 $q(n)$ 通过滤波器后有:$y(n) = q(n) g(n)$。其中,窗函数滤波器通常选用多尔夫-切比雪夫滤波器或者理想的 box-car 滤波器,其特点满足过渡带陡峭、通带平滑、运算量小和大系数集中。

3)频域降采样。以因子 p 将 $y(n)$ 混叠,对混叠后信号作 B 点 FFT 运算得 $Z(k) = Y(kn/B)$。

4)哈希映射。定义哈希函数和偏移量满足:$h_{\sigma'}(k) = \text{round}(\sigma' kB/n)$,$o_{\sigma'}(k) = \sigma' k - h_{\sigma'}(k)(n/B)$,round(·)表示取整操作。

5)定位。将 $Z(k)$ 中最大的谱峰表示为集合 J,对于 J 中的每个元素,令集合 H 满足条件:$H = \{k \in [n] \mid h_{\sigma'}(k) \in J\}$。

6)估值。对于 $k \in H$,$X_i(k) = Z(h_{\sigma'}(k)) \cdot W_N^{\tau k} / G(o_{\sigma'}(k))$ 作为对应的估计值。其中,$G(o_{\sigma'}(k))$ 为 x 在映射 $o_{\sigma'}(k)$ 下平滑窗函数的谱。

3 基于稀疏傅里叶变换的快速捕获方法分析

3.1 稀疏傅里叶变换优化的捕获方法实现过程

通过第 2 节对 SFT 的分析,结合图 1 中捕获成功的结果,由于伪码信号的强自相关性特性,在 FFT 相关运算处理过程中,IFFT 的变换域具有稀疏的特性,因此可以利用 SFT 算法对 FFT 相关过程进行优化。

另外,输入中频信号的频域 $X(k)$ 并不具有稀疏特性,仅使用加窗—混叠—重构的方法不能使运算效率最大化提升。考虑利用 2.1 节中“时域的混叠对应频域的降采样”思想对这一过程进行优化。图 5 为基于 SFT 的快速捕获方法。

具体算法实现步骤如下:

步骤 1 将数字中频信号和本地载波发生器输出的正弦和余弦载波混频,得到基带的复信号。

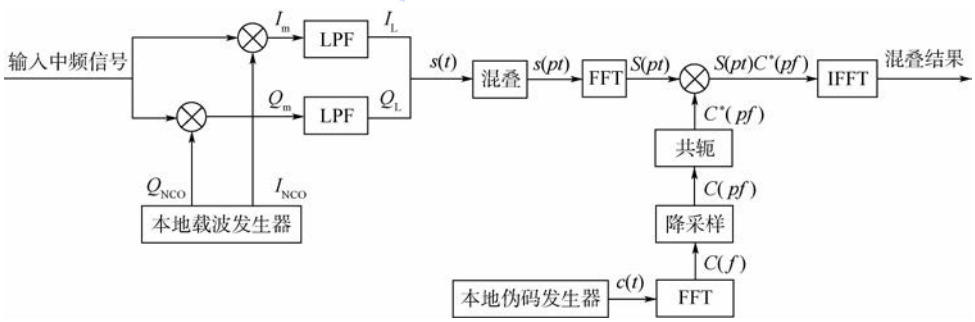


图 5 基于 SFT 的快速捕获方法示意图

Fig. 5 Schematic of fast acquisition method based on SFT

北航学报 赠阅

步骤 2　对步骤 1 得到的信号 $s(t)$ 以 N/p 求模的形式进行混叠,其中参数 $p=\sqrt{\mathrm{lb}\ N}$[11],得

$$s(pt)=\sum_{j=0}^{p-1}s(n+jN/p)$$

步骤 3　对混叠信号 $s(pt)$ 做 N/p 点 FFT。

步骤 4　对本地伪码做 N 点 FFT,输出结果为 $C(f)$。以因子 p 对其进行降采样,得 $C(pf)$,取其共轭 $C^*(pf)$。

步骤 5　将步骤 3 和步骤 4 的结果相乘,并将乘积做 IFFT 变换。

步骤 6　步骤 5 的 IFFT 输出结果为混叠结果,对其取模并进行门限判决。得出 p 个可能的谱峰位置,通过比较这 p 个坐标的相关值,得到相关值最大的坐标点,其对应的载波频率和码相位即为捕获结果。

3.2　时间复杂度分析

3.1 节的基于 SFT 捕获方法在完成一个处理后能够得到全部伪码相位的结果。步骤 2 混叠过程需要 N 次加法;步骤 3 降采样后 N/p 点 FFT 的运算量为 $(N/2p)\mathrm{lb}(N/p)$ 次复数乘法和 $(N/p)\cdot\mathrm{lb}(N/p)$ 次复数加法;步骤 4 对本地伪码信号的预处理为 N 点 FFT 运算量;步骤 5 相乘操作需要 N/p 次乘法,稀疏 IFFT 运算量同 N/p 点 FFT;步骤 6 鉴别唯一解的运算量为 $pN/p=N$。同基于 FFT 捕获方法对比,得到单个相关结果,共需要 $[N+(N/p)(1+\mathrm{lb}(N/p))+(N/2)\mathrm{lb}\ N]$ 次复数乘法和 $[N(1+\mathrm{lb}\ N)+2(N/p)\mathrm{lb}(N/p)]$ 次复数加法。其中,本地伪码发生器输出的伪码信号做 FFT 可以进行预处理,考虑其运算不占用算法时间。则采用 FFT 和 SFT 的捕获方法时间复杂度分别为 $O(N\mathrm{lb}\ N)$ 和 $O((N/p)\mathrm{lb}\ (N/p))$,由此可见基于 SFT 的快速捕获方法大大降低了相关的运算量,运算量减小到原来算法的 $p\mathrm{lb}\ p$ 倍,且随着数据长度 N 的增加,其效率提升更加明显。

4　验证实验及分析

为了验证基于 SFT 的快速捕获方法的效率,利用实测的中频数据对其进行测试。数据中频频率为 9.584 MHz,采样频率为 38.192 MHz。方法在基于 MATLAB 的软件接收机上实现[16],仿真实验对比了基于 FFT 的并行码相位捕获方法和基于 SFT 的快速捕获方法的捕获性能。

仿真实验中,频率搜索范围设置为 [-7, +7] kHz,频率搜索步长为 500 Hz。信号长度 T 分别为 1、2、5 ms,则 $p=\sqrt{\mathrm{lb}\ N}=3.9$,选取 p 值分别为 1、2、4;其中,$p=1$ 为采用基于 FFT 的并行码相位捕获方法捕获结果。

图 6 为信号长度 T 为 1 ms,降采样因子 $p=1$、2、4 时 PRN3 卫星捕获结果,表 1 为捕获实验结果对比。

由图 6 和表 1 可以看出,在信号长度为 1 ms 时,基于 FFT 的并行码相位捕获方法捕获结果为:伪码相位(采样点)为 $3.421\,2\times10^4$,多普勒频率为 2.5 kHz,运行时间约为 0.26 s。

降采样因子 $p=2$ 时,检测到的伪码相位(采样

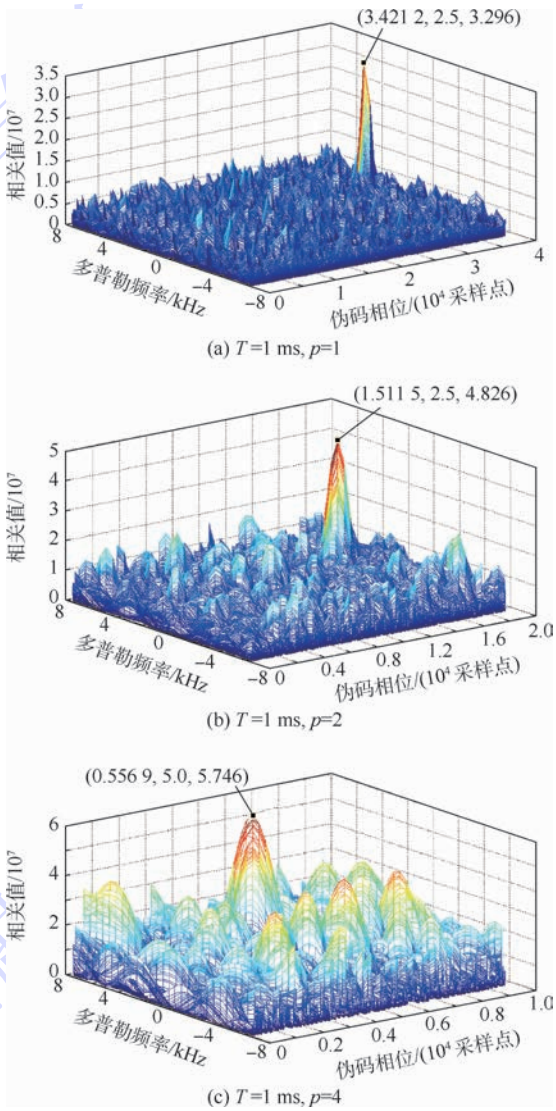


图 6　卫星捕获结果($T=1$ ms,$p=1,2,4$)

Fig. 6　Satellite acquisition results ($T=1$ ms, $p=1,2,4$)

表 1　卫星捕获实验结果($T=1$ ms)

Table 1　Satellite acquisition experimental results ($T=1$ ms)

p 值	伪码相位/(10^4 采样点)	多普勒频率/kHz	峰值/10^7	运算时间/s
1	3.421 2	2.5	3.296	0.262 481
2	1.511 5	2.5	4.826	0.143 569
4	0.556 9	5.0	5.746	0.093 134

北航学报 赠阅

点)为 1.5115×10^4,对应的 $p=2$ 个可能的偏移位置为 $N=1.5115\times10^4$、3.4211×10^4,分别计算这 2 个可能偏移位置各自的相关值,$N=3.4211\times10^4$ 点相关值远远大于 $N=1.5115\times10^4$ 点的相关值;多普勒频率为 2.5 kHz,算法运行时间约为 0.14 s,相对于基于 FFT 的并行码相位捕获方法,程序运行效率提高约 1.86 倍。

降采样因子 $p=4$ 时,检测到的伪码相位(采样点)为 0.5569×10^4,对应的 $p=4$ 个可能的偏移位置为 $N=0.5569\times10^4$、1.5117×10^4、2.4665×10^4、3.4213×10^4,分别计算这 4 个可能偏移位置各自的相关值,$N=3.4213\times10^4$ 点相关值远远大于其他 3 个点的相关值;多普勒频率为 5.0 kHz,算法运行时间约为 0.09 s,但是其多普勒频率估计误差较大,超出多普勒频率搜索步长。

为了验证运算效率随着数据长度 N 的增大而提升,图7给出了 $T=2$ ms的捕获实验结果,表2为 $T=2$ ms 时 $p=1$、2、4 的捕获实验结果对比。图 8 给出了 $T=5$ ms 的捕获实验结果,表 3 为 $T=5$ ms 时 $p=1$、2、4 的捕获实验结果对比。

由图 7、图 8 和表 2、表 3 可以得出,同 $T=1$ ms 时捕获结果分析一致,基于 SFT 的快速捕获方法可以实现成功捕获。$T=2$ ms 时,程序运行效率相对基于 FFT 的并行码相位捕获方法提升了 2.04 倍。$T=5$ ms 时,程序运行效率提升

表 2 卫星捕获实验结果($T=2$ ms)

Table 2 Satellite acquisition experimental results($T=2$ ms)

p 值	码相位/(10^4 采样点)	多普勒频率/kHz	峰值/10^7	运算时间/s
1	3.4213	2.5	6.810	0.475760
2	1.5117	2.5	9.224	0.235174
4	0.5568	4.5	9.288	0.116987

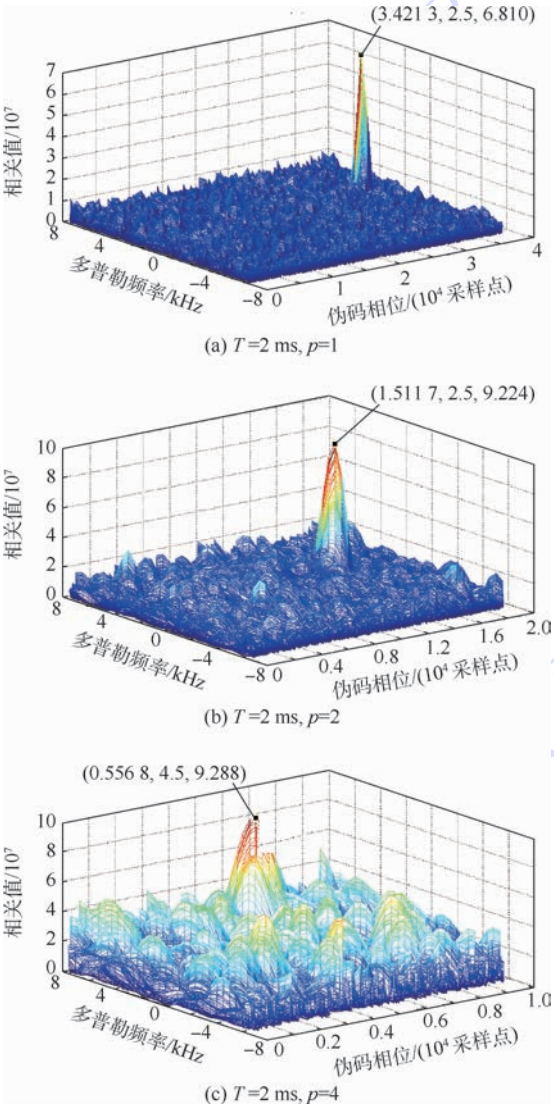


图 7 卫星捕获结果($T=2$ ms,$p=1,2,4$)

Fig. 7 Satellite acquisition results ($T=2$ ms, $p=1,2,4$)

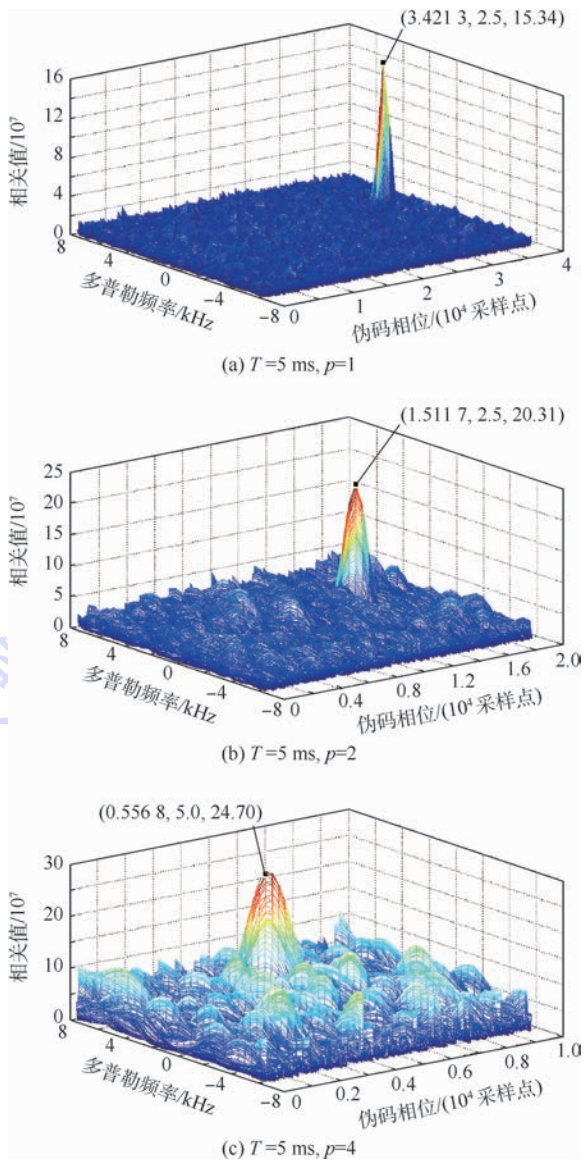


图 8 卫星捕获结果($T=5$ ms,$p=1,2,4$)

Fig. 8 Satellite acquisition results ($T=5$ ms, $p=1,2,4$)

北航学报 赠阅

表 3　卫星捕获实验结果（$T=5$ ms）

Table 3　Satellite acquisition experimental results ($T=5$ ms)

p 值	伪码相位/（10^4 采样点）	多普勒频率/kHz	峰值/10^7	运算时间/s
1	3.4213	2.5	15.34	1.205674
2	1.5117	2.5	20.31	0.546816
4	0.5568	5.0	24.70	0.274991

到原来的 2.2 倍。说明随着数据长度 N 的增大而运算效率提升更加明显。

5　结　论

本文在分析卫星捕获输出相关值结果具有“稀疏”特性的基础上，将稀疏傅里叶变换（SFT）引入基于 FFT 的并行码相位捕获方法中，提出了基于 SFT 的快速捕获方法，仿真结果表明：

1）本文方法可以实现卫星信号的快速捕获，与基于 FFT 的并行码相位捕获方法相比，算法性能提升到原来的 2 倍。

2）当参与运算的信号长度 N 增大时，算法的运算效率提升更加明显。$T=1$ ms 时，算法运算效率提升约 1.86 倍；$T=2$ ms 时，算法运算效率提升 2.04 倍；$T=5$ ms 时，算法运算效率提升 2.2 倍。

3）降采样因子 p 的取值要求：$p \leqslant \sqrt{\mathrm{lb}\ N}$，当 $p > \sqrt{\mathrm{lb}\ N}$，捕获结果误差较大。

为了更好地验证本文方法，需要利用不同载噪比的中频信号进行验证。

参考文献（References）

[1] 鲁郁. 北斗/GPS 双模软件接收机原理与实现技术[M]. 北京：电子工业出版社，2016：119-157.
LU Y. BDS/GPS dual-mode software receiver principles and realizing[M]. Beijing: Publishing House of Electronics Industry, 2016: 119-157 (in Chinese).

[2] 谢钢. GPS 原理与接收机设计[M]. 北京：电子工业出版社，2009：349-375.
XIE G. Principles of GPS and receiver design[M]. Beijing: Publishing House of Electronics Industry, 2009: 349-375 (in Chinese).

[3] COOLEY J W, TUKEY J W. An algorithm for the machine calculation of complex Fourier series[J]. Mathematics of Computation, 1965, 19(90): 297-301.

[4] RAO K R, KIM D N, HWANG J J. Fast Fourier transform-algorithms and applications[M]. Berlin: Springer, 2010: 15-40.

[5] AKOPIAN D. Fast FFT based GPS satellite acquisition methods [J]. IEE Proceedings: Radar Sonar and Navigation, 2005, 152(4): 277-286.

[6] SPANGENBERG S M, SCOTT I, MCLAUGHLIN S, et al. An FFT-based approach for fast acquisition in spread spectrum communication systems[J]. Wireless Personal Communications, 2000, 13(1-2): 27-55.

[7] SAGIRAJU P K, RAJU G V S, AKOPIAN D. Fast acquisition implementation for high sensitivity global positioning systems receivers based on joint and reduced space search[J]. IEE Proceedings: Radar Sonar and Navigation, 2008, 2(5): 376-387.

[8] HASSANIEH H, INDYK P, KATABI D, et al. Nearly optimal sparse Fourier transform[C]// Proceedings of the 44th ACM Symposium on Theory of Computing. New York: ACM, 2012: 563-578.

[9] 那美丽，周志刚，李霈霈. 基于稀疏傅里叶变换的低采样率带宽频谱感知[J]. 电子技术应用，2015，41(11)：85-88.
NA M L, ZHOU Z G, LI P P. Wideband spectrum sensing at low sampling rate based on the sparse Fourier transform[J]. Application of Electronic Technique, 2015, 41(11): 85-88 (in Chinese).

[10] 王雄. 基于稀疏傅里叶变换的水声快速解调算法研究[D]. 北京：北京理工大学，2015：39-42.
WANG X. Fast demodulation algorithm of underwater acoustic based on the sparse Fourier transform[D]. Beijing: Beijing Institute of Technology, 2015: 39-42 (in Chinese).

[11] HASSANIEH H, INDYK P, KATABI D, et al. FASTER GPS via the sparse Fourier transform[C]// Proceedings of the 18th Annual International Conference on Mobile Computing and Networking. New York: ACM, 2012: 353-364.

[12] MA Y, BU X Y, HAN H C, et al. Combined algorithm of acquisition and anti-jamming based on SFT[J]. Journal of Systems Engineering and Electronics, 2015, 26(3): 431-440.

[13] 龚巧娴. 基于 SFT 的快速捕获与干扰抑制联合算法研究[D]. 北京：北京理工大学，2015：20-32.
GONG Q X. Research on combined algorithm of acquisition and anti-jamming based on SFT[D]. Beijing: Beijing Institute of Technology, 2015: 20-32 (in Chinese).

[14] RAO M V G, RATNAM D V. Faster acquisition technique for software-defined GPS receivers[J]. Defence Science Journal, 2015, 65(1): 5-11.

[15] PARKINSON B, SPILKER J. Global positioning system: Theory and applications[M]. Reston: AIAA, 1996: 245-325.

[16] BORRE K, AKOS D M, BERTELSEN N. A software-defined GPS and Galileo receiver[M]. Berlin: Springer, 2007: 75-86.

作者简介：

张春熹　男，博士，教授，博士生导师。主要研究方向：光纤陀螺、光纤惯性导航系统、组合导航系统。

李先慕　男，博士研究生。主要研究方向：惯性/卫星组合导航技术。

高爽　女，博士，讲师。主要研究方向：惯性导航及光纤传感技术。

北航学报 赠阅

Fast acquisition methods based on sparse Fourier transform

ZHANG Chunxi, LI Xianmu, GAO Shuang*

(School of Instrumentation Science and Opto-electronics Engineering, Beijing University of Aeronautics and Astronautics, Beijing 100083, China)

Abstract: The acquisition of the satellite signals is the key step of baseband signal processing in the receiver. The speed of the acquisition has a serious effect on the time to first fix (TTFF). According to the parallel code phase search (PCPS) method based on fast Fourier transform (FFT), the sparse Fourier transform (SFT) which computes in sub-sampled time is applied to the PCPS algorithm to simplify the acquisition process. The fast acquisition method based on SFT is proposed. The faster acquisition is achieved by optimizing the efficiency of correlation operation. The simulation results show that the computation time of the proposed fast acquisition method based on SFT is 2 times faster than that of the conventional FFT-based acquisition method. And it can better meet the requirements of fast satellite signal acquisition.

Keywords: satellite signal; fast acquisition; sparse Fourier transform (SFT); aliasing; sub-sampling

北京航空航天大学学报

Received: 2017-05-05; **Accepted**: 2017-07-13; **Published online**: 2017-09-15 17:09

URL: kns.cnki.net/kcms/detail/11.2625.V.20170915.1709.001.html

* **Corresponding author**. E-mail: gaoshuang@buaa.edu.cn

北航学报 赠阅

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2017.0245

气象因素对飞机进近飞行燃油效率的影响

全下文载

胡荣*，吴文洁，陈琳，张菲菲
（南京航空航天大学 民航学院，南京 210016）

摘 要：气象因素对航空飞行意义重大。为了考察航空飞行的燃油效率，基于飞机性能数据库(BADA)模型，考虑气象因素，建立飞机燃油消耗的修正模型。以广州白云国际机场某进港航班为例，开展飞机进近飞行仿真试验，从燃油流量和燃油消耗量2个维度分别讨论气温、气压、风速变化对飞机燃油效率的影响。结果表明：气象因素与飞机燃油效率存在明显的相关性。当飞机飞行高度一定时，气温升高，燃油流量和燃油消耗量增大，燃油效率降低；气压增强，燃油流量无明显变化，燃油消耗量略有降低，燃油效率升高；风速增加，燃油流量和燃油消耗量先减小后增大，燃油效率先升后降，风速为4 m/s时燃油效率最高。当飞机飞行高度下降时，气温和气压升高，风速下降，燃油流量小幅度波动上升，燃油效率降低。最佳气象条件下，一次进近飞行能减少约3%的燃油消耗。研究结果对提高实际飞行的燃油效率有一定的参考意义。

关 键 词：气象因素；气温；气压；风速；燃油效率；燃油流量；燃油消耗量
中图分类号：V321.2+2；V241.7+2
文献标识码：A **文章编号**：1001-5965(2018)04-0677-07

航空运输是现代化快速、便捷高效的交通方式，为人们出行提供了极大便利。就目前而言，航空器性能不断提高，地面保障系统逐渐完善，但气象这类不可控因素仍对航空飞行有着巨大影响。在中国，由不利气象因素引发的重大飞行事故约占飞行事故总数的31%[1]，带来的人员和财产损失更是高达数亿元。由此可见，气象已经成为影响飞行安全与经济效益的重要因素。

众所周知，任何航空飞行活动都需在一定气象条件下进行。气象因素对航空飞行有着至关重要的影响，吸引了国内外众多学者对此开展研究。

就国内而言，不少学者研究了气象因素对飞行安全的影响。文献[2-3]结合近年来低能见度引发的民航安全事故，论述能见度对确保航空飞行安全的重要性，指出能见度可直接影响飞机的领航、起飞、着陆，对飞行安全影响显著。文献[4-5]从航空飞行安全与气象条件关联性出发，探讨能见度、大风、低云、雷雨、冰雪以及风切变等气象因素对飞行安全性的影响，提出不同气象条件下飞行的应对措施及保障飞行安全的参考建议。

同时，国内也有部分学者专注于研究气象与飞行性能效率的关系。文献[6]研究气象因素对飞机起降性能的影响，从气象和空气动力学知识入手，围绕温度、气压、风等因素变化对飞机起降滑跑距离产生的影响进行详细分析。文献[7]讨论气温、气压和空气密度与飞机千米耗油量的关系，建立最佳千米耗油量与经济高度的数学模型，并以B737-300为例对模型进行验证。文献[8-9]选取低温、大风、积冰和大雾天气，总结其种类、特点和变化规律，研究气象条件对签派放行油耗计算的影响。

此外，针对飞机燃油消耗的问题，国内外学者

收稿日期：2017-04-19；**录用日期**：2017-08-11；**网络出版时间**：2017-09-22 10:25
网络出版地址：kns.cnki.net/kcms/detail/11.2625.V.20170922.1025.004.html
基金项目：国家自然科学基金（71201082，71401072）；江苏省社会科学基金（15GLC001）
*通信作者. E-mail：hoorong@nuaa.edu.cn

引用格式：胡荣，吴文洁，陈琳，等. 气象因素对飞机进近飞行燃油效率的影响[J]. 北京航空航天大学学报，2018，44(4)：677-683.
HU R，WU W J，CHEN L，et al. Influence of meteorological factors on aircraft fuel efficiency in approach flight[J]. Journal of Beijing University of Aeronautics and Astronautics，2018，44(4)：677-683 (in Chinese).

北航学报 赠阅

也开展了大量的研究。文献[10-11]以飞机性能手册参数为依据,根据各飞行阶段性能特点,结合油耗历史数据,建立基于神经网络的燃油消耗估计模型及其动态修正方法,简化燃油消耗估算过程,提高燃油消耗估计精度。文献[12-14]考虑飞行中飞机的动能和势能变化,建立基于能量平衡原理的燃油消耗评估模型,利用真实飞行数据对达拉斯-沃斯堡国际机场飞机的燃油消耗总量进行估算。但以上研究过程均未考虑气象因素对燃油效率的影响。

尽管目前有关气象因素对航空飞行影响的研究已取得一定成果,但尚有不足:①多数研究关注气象因素对飞行安全的影响,较少研究其对飞行性能和经济效益的影响;②燃油建模研究多设在理想条件下,考虑气象因素引发飞机燃油效率变化的研究相对较少。

基于上述考虑,本文利用飞机性能数据库(Base of Aircraft Data,BADA)模型,根据气温、气压和风速修正理论,构建考虑气象因素的燃油消耗模型,模拟广州白云国际机场某进港航班的进近飞行过程,从燃油流量和燃油消耗量 2 个角度,探讨不同气象因素作用下飞机燃油效率的差异性,以期为不同气象条件下的航空飞行提供理论参考。

1 基本模型

1.1 BADA 模型

BADA 模型是由欧洲航空安全组织和航空器制造商协作开发的分析飞行动力学的模型,可确定飞机爬升、巡航和下降阶段的性能,包括与空气动力、发动机性能、质量、速度、高度、耗油率、大气环境等相关的飞机特性参数。BADA 模型可应用于飞行仿真、航迹预测、排放评估和油耗计算 4 个方面,且在飞行仿真度、复杂性和精确度上具有较大优势[15]。

BADA 模型由飞机性能模型、航空公司程序和飞机特征数据组成。飞机性能模型由全能量模型、气动模型、推力模型以及水平运动模型构成。其中,飞机性能模型的核心——全能量模型遵循能量守恒原理,认为作用于飞机的合外力做功等于飞机动能和势能的变化量,该模型反映了飞机运动中速度、推力和高度等参数的变化规律以及约束关系。以飞机下降飞行为例,建立如下模型:

$$(F - D)V_{\mathrm{TAS}} = m\frac{\mathrm{d}V_{\mathrm{TAS}}}{\mathrm{d}t}V_{\mathrm{TAS}} + mg\frac{\mathrm{d}h}{\mathrm{d}t} \tag{1}$$

$$\sin\gamma = -\frac{1}{V_{\mathrm{TAS}}}\cdot\frac{\mathrm{d}h}{\mathrm{d}t} \tag{2}$$

式中:F 为发动机推力;D 为飞机所受阻力;V_{TAS} 为真空速度;m 为飞机质量;g 为重力加速度;h 为飞行高度;t 为飞行时间;γ 为下降航迹角。

飞行中飞机受升力、重力、阻力及推力的综合作用,直接影响其飞行状态,根据气动模型和推力模型,可计算飞机的阻力和推力为

$$D = \frac{C_{D1}\rho S V_{\mathrm{G}}^2}{2} + \frac{2C_{D2}m^2g^2}{\rho S V_{\mathrm{G}}^2} \tag{3}$$

$$F = m\frac{\mathrm{d}V_{\mathrm{TAS}}}{\mathrm{d}t} + D - mg\sin\gamma \tag{4}$$

式中:C_{D1}、C_{D2} 为与机型有关的阻力系数;ρ 为空气密度;S 为机翼面积;V_{G} 为地速。

飞机飞行过程中,直线运动受风向和风速影响,参考水平运动模型,通过偏流角和风向角,可计算飞机相对地面的运动速度为

$$V_{\mathrm{G}} = V_{\mathrm{TAS}}\cos\varphi + V_{\mathrm{W}}\cos\omega \tag{5}$$

式中:V_{W} 为风速;φ 为偏流角;ω 为风向角。

依据飞机性能参数,参照质量守恒原理,认为飞行中飞机质量减少量即为其消耗的燃油量,获取单位推力的燃油消耗率,计算飞机燃油流量:

$$\eta = C_{\mathrm{f1}}\left(1 + \frac{V_{\mathrm{G}}}{C_{\mathrm{f2}}}\right) \tag{6}$$

$$f = \eta C_T F \tag{7}$$

式中:η 为单位推力的燃油消耗率;C_{f1}、C_{f2} 为油耗参数;f 为燃油流量;C_T 为推力等级。

1.2 气象换算模型

飞机在空中飞行受大气合外力作用,气温、气压和风速是大气最基本的物理状态,也是影响飞机飞行性能的重要因素。国际标准大气条件下,海平面绝对温度为 288.15 K,海平面绝对压强为 101 325 Pa,气温和气压随高度呈线性变化的换算原理如下:

$$T = T_{\mathrm{G}} - 0.006h \tag{8}$$

$$\theta = \frac{T}{T_{\mathrm{ISA}}} \tag{9}$$

$$P = P_{\mathrm{G}} - 12.12h \tag{10}$$

$$\delta = \frac{P}{P_{\mathrm{ISA}}} \tag{11}$$

式中:T 为实际大气条件下不同高度的气温;T_{G} 为实际条件下的场面气温;T_{ISA} 为标准大气条件的气温;θ 为实际与标准大气条件下的气温之比;P 为实际大气条件下不同高度的气压;P_{G} 为实际条件下的场面气压;P_{ISA} 为标准大气条件的气压;δ 为实际与标准大气条件下的气压之比。

在大气边界层内,风速和高度具有相关性。

北航学报 赠阅

中国气象部门一般会提供距离地面 10 m 高度定时观测的风速资料,不同高度的风速采用乘幂律公式修正计算:

$$V_W = V_R\left(\frac{h}{h_R}\right)^{\lambda} \tag{12}$$

式中:V_R 为参考的观测风速;h_R 为距地参考高度,取值为 10 m;λ 为稳定度参数,考虑机场周边建筑特征,取值为 0.10[16]。

1.3　燃油消耗模型

不同气象因素会对飞机燃油消耗产生不同影响,利用气象换算模型对国际民航组织(International Civil Aviation Organization,ICAO)标准燃油消耗模型进行修正计算:

$$f_M = \frac{f}{\delta}\theta^{3.8}e^{0.2Ma^2} \tag{13}$$

$$Q = n\int_0^{t_T} f_M \mathrm{d}t \tag{14}$$

式中:f_M 为单发修正燃油流量;Ma 为飞行马赫数;Q 为飞机总燃油消耗量;t_T 为飞行总时间;n 为发动机数量。

2　仿真试验

为了解气象因素对飞机燃油效率的影响效果,基于第 1 节模型对飞机进近飞行过程进行仿真试验,利用 MATLAB 软件进行数据处理,并对仿真结果进行分析和讨论。

2.1　试验数据

以广州白云国际机场 2016 年 9 月 26 日某进港航班为例:中国某航空公司采用 A319 执飞越南河内—中国广州的国际航班,10 时 57 分从该机场 IDUMA S 方向开始,沿 IDUMA、SHL 进场点,下降至 GG422 完成进近过程,进近历时约 8 min。进近的水平雷达轨迹和三维空间轨迹如图 1 和图 2 所示。

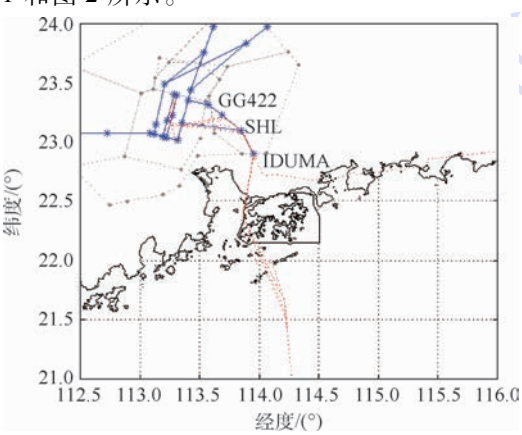


图 1　水平雷达轨迹

Fig. 1　Horizontal radar trajectory

通过实际调研,结合实时气象预报,获取当天广州白云国际机场的基本气象信息,如表 1 所示。

下面以该航班为例,采用燃油流量和燃油消耗量作为评价燃油效率的标准,通过设置气温、气压和风速等气象因素的不同情景,研究气象因素对飞机燃油效率的影响。

基于此分别开展气温试验、气压试验和风速试验。假定气温为 20,23,…,35℃,气压为 100 000,100 300,…,101 500 Pa,风速为 0,2,…,10 m/s。依据上述仿真飞行试验,获取燃油流量、飞行时间和高度等仿真数据,讨论气温、气压、风速与燃油流量、燃油消耗量的变化规律,分析气象因素对飞机燃油效率的影响。

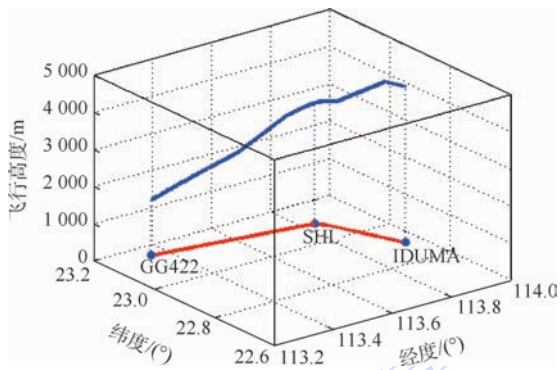


图 2　三维空间轨迹

Fig. 2　Three-dimensional spatial trajectory

表 1　广州白云国际机场 2016 年 9 月 26 日 10:00—12:00 气象实况

Table 1　Actual weather of 10:00 – 12:00 at Guangzhou Baiyun Airport on September 26, 2016

天　气	气压/Pa	气温/℃	风　向	风速/(m·s^{-1})
阴	100 300	23	西北风	4

2.2　气温影响分析

物理学知识表明,对流层中气温和高度有严格的线性关系,平均气温垂直递减率为 0.006℃/m,即高度每下降 100 m,气温平均升高约 0.6℃。根据 6 组气温试验数据,分别绘制进近飞行过程中燃油流量、燃油消耗量与气温的关系示意图,分别如图 3 和图 4 所示。

当飞行高度相同时,气温升高,燃油流量缓慢增加;当飞行高度不断下降时,气温随之升高,理论上燃油流量应当持续增加,但当飞机飞越 SHL 进场点时,燃油流量出现先减小后增大的波动。其波动原因主要是:为满足 SHL 进场点的高度和方向要求,飞机飞行速度先减小后增大,推力先减小后增大。尽管飞行高度下降,气温升高,但气温远不如推力和飞行速度对燃油流量的影响大,因

北航学报 赠阅

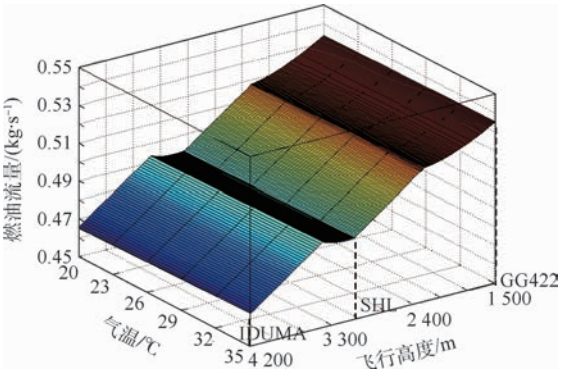


图 3 气温-飞行高度-燃油流量关系

Fig. 3 Relationship among air temperature, flight height and fuel flow

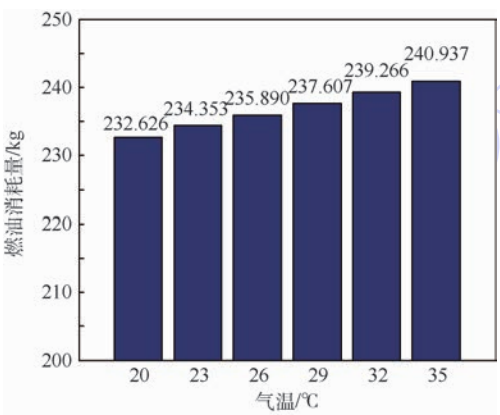


图 4 气温-燃油消耗量对比

Fig. 4 Fuel consumption comparison at different air temperatures

而造成燃油流量先降后升的波动情况。

当机场场面气压和风速不变时,随着气温升高,燃油流量缓慢增加,由于飞行时间相同,使得燃油消耗量逐渐增加,气温和燃油消耗量符合线性递增关系,近似达到 0.57 kg/℃的递增率。

综合来看,在气压和风速不变的前提下,气温升高,燃油流量和燃油消耗量增大,意味着飞机增加了航空燃油的使用,因而燃油效率降低。

2.3 气压影响分析

在飞机进近飞行过程中,飞行高度下降,大气压力升高。标准大气条件下的气压垂直递减率为 12.12 Pa/m,相当于高度每下降 100 m,气压平均升高约 1 212 Pa。根据气压试验,绘制气压和燃油效率的关系图,如图 5 和图 6 所示。

当飞行高度相同时,气压升高,燃油流量无明显变化;当飞行高度下降时,气压上升,燃油流量波动上升。这是因为气压对燃油流量表现出和气温类似的变化规律,当飞机飞越 SHL 进场点时,为满足该点的高度限制及航向偏转要求,导致燃油流量先减小后增大。

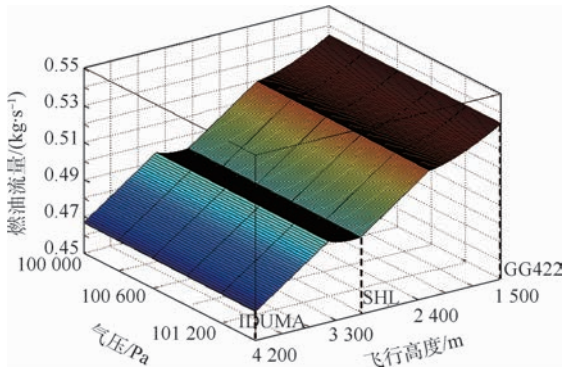


图 5 气压-飞行高度-燃油流量关系

Fig. 5 Relationship among air pressure, flight height and fuel flow

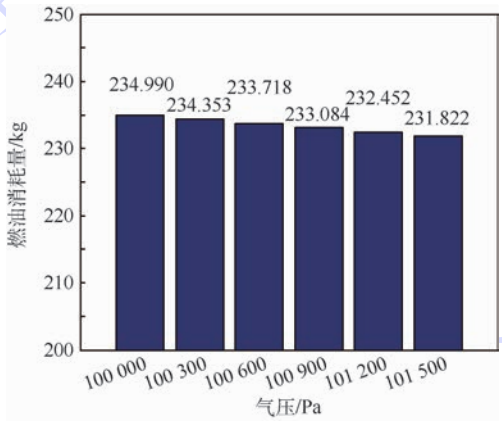


图 6 气压-燃油消耗量对比

Fig. 6 Fuel consumption comparison at different air pressure

当机场场面气温和风速不变时,随着气压升高,燃油流量无明显变化,但飞行时间稍有缩短,使得燃油消耗量逐渐减少,气压和燃油消耗量符合线性递减关系,近似满足 21 g/Pa 的递减率。

综上所述,在气温和风速不变时,气压增强,燃油流量无明显变化,燃油消耗量略有降低,表明飞机节省航空燃油的使用,故而燃油效率升高。

2.4 风速影响分析

飞机一般采用逆风起降,各型号飞机起飞、爬升、进近、着陆的顶风、侧风、顺风限制标准必须严格按照规定执行,风速和风向对飞机运行效率影响显著。依据风速试验,可得风速与燃油流量、燃油消耗量的变化规律,如图 7 和图 8 所示。

当飞行高度相同时,随着风速增加,燃油流量先减小后增大,风速为 4 m/s 时燃油流量最小;当飞行高度下降时,风速下降,燃油流量小幅度波动上升,波动原因是飞机飞越 SHL 进场点时,发动机推力和飞行速度均发生变化。

当机场场面气温和气压不变时,随着风速增加,燃油消耗量先减小后增大,存在最佳风速(即最低点 4 m/s)。这是因为在无风或微风时,飞

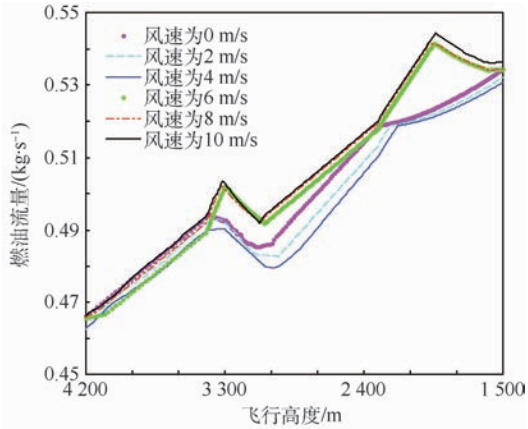


图 7　风速-飞行高度-燃油流量关系

Fig. 7　Relationship among wind speed, flight height and fuel flow

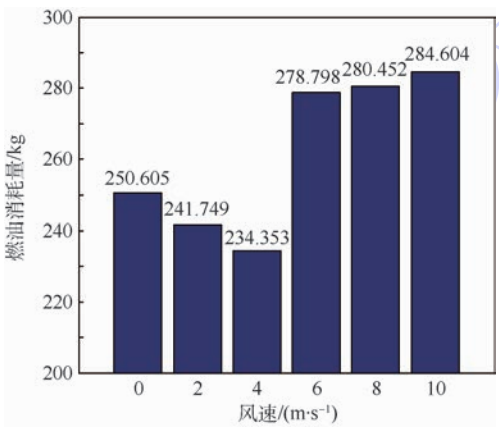


图 8　风速-燃油消耗量对比

Fig. 8　Fuel consumption comparison at different wind speeds

机不能有效借助风力,完全靠燃油提供动力实现进近飞行,因此不能真正省油;而在风速较大时,风力对飞机降落造成干扰,压差阻力和诱导阻力增大,可能使飞机出现摇晃、颠簸等现象,飞机需克服风力实现进近,故而也不能有效省油。

总而言之,在气温和气压不变时,风速增加,燃油流量和燃油消耗量均先减小后增大,因而燃油效率先升后降,且 4 m/s 的风速对飞机进近飞行最为有利,燃油效率相对最高。

2.5　省油效果

2.2 节 ~2.4 节试验结果表明,气象因素与飞机燃油效率存在明显的相关性,气温 20℃,气压 101 500 Pa 和风速 4 m/s 时的燃油效率最高,最佳气象与实际气象条件下的省油效果对比如表 2 所示。

由上述分析可知,在本文给定的最佳气象条件下,一次进近飞行大约可减少 3% 的燃油消耗,减少了燃油成本,从而增加航空公司运营收益。按照 2016 年广州白云国际机场年起降 435 231 架次估算,全年可节约 2 877.7 t 燃油,同时依据 2016 年航空燃油价格 3 383 元/t 估算,可为航空公司节省约 973.5 万元燃油成本。

表 2　实际气象和最佳气象的省油效果对比

Table 2　Fuel saving comparison between actual meteorological factors and optimal meteorological factors

气　象	气温/℃	气压/Pa	风速/(m·s⁻¹)	燃油消耗量/kg	省油量/kg	省油比例/%
实际气象	23	100 300	4	234.353		
最佳气象	20	101 500	4	227.741	6.612	2.82

3　结　论

1) 当飞机的飞行高度一定时,随着气温升高,燃油效率降低;随着气压增强,燃油效率升高;随着风速增加,燃油效率先升高后降低,在风速 4 m/s时燃油效率最高。

2) 当飞行高度不断下降时,气温和气压均升高,风速减小,燃油流量小幅度波动上升,燃油效率降低。

3) 研究案例表明,最佳气象的省油效果明显,一次进近飞行能减少约 3% 的燃油消耗,可为航空公司一年减少油耗约 2 877.7 t,节约燃油成本约 973.5 万元。

建议机场运控部门加强气象情报告知服务,合理利用气象信息,及时把握最佳气象条件来实现飞机的起降。

建议在保证飞行安全的前提下,协调飞行驾驶、空中管制和签派放行配置,提高不同气象条件下飞行的燃油效率,最大限度减少飞机燃油消耗,节约燃油成本,增加航空运营的经济收益。

参考文献 (References)

[1] 黄晓栋. 气象条件对于民航飞行安全的影响[J]. 科技与企业,2015,24(17):180.
HUANG X D. Effects of meteorological conditions on civil aviation flight safety[J]. Science and Technology,2015,24(17):180(in Chinese).

[2] 乌日恒,丰建军. 浅谈能见度对飞行的影响[J]. 科技风,2011,24(21):56.
WU R H,FENG J J. Probe into effect of visibility on flight[J]. Technology Wind,2011,24(21):56(in Chinese).

[3] 张序,刘雪涛,周川,等. 低能见度条件下飞行事故分析及运行控制研究[J]. 失效分析与预防,2014,9(6):383-387.
ZHANG X,LIU X T,ZHOU C,et al. Analysis of flight accidents and operational control under low visibility condition[J]. Failure Analysis and Prevention,2014,9(6):383-387(in Chinese).

[4] 张建荣. 气象条件对飞机及其飞行的影响分析[J]. 航空科

北航学报 赠阅

学技术,2014,25(5):54-56.
ZHANG J R. Impact analysis of weather conditions on aircraft and its flight[J]. Aeronautical Science & Technology,2014,25(5):54-56(in Chinese).

[5] 杨春风,王荣,李新泉. 影响航空飞行安全的气象要素探讨[J]. 现代农业科技,2010,22(1):297-298.
YANG C F,WANG R,LI X Q. Study on meteorological factors affecting aviation flight safety[J]. Modern Agricultural Science and Technology,2010,22(1):297-298(in Chinese).

[6] 张传林. 气象因素对飞机起降滑跑距离的影响[J]. 科技风,2011,24(17):137.
ZHANG C L. Effect on meteorological factors in aircraft taxiing distance[J]. Technology Wind,2011,24(17):137(in Chinese).

[7] 王永忠,蔡廷剑. 气象因素对飞机千米耗油量影响的探讨[J]. 高原山地气象研究,2004,24(3):44-46.
WANG Y Z,CAI T J. Discussion on the influence of meteorological factors on aircraft fuel consumption[J]. Plateau and Mountain Meteorology Research,2004,24(3):44-46(in Chinese).

[8] 何宁. 恶劣气象条件下的航空运行与签派放行[D]. 兰州:兰州大学,2013:35-41.
HE N. Dispatch release and aviation operations under severe weather conditions[D]. Lanzhou:Lanzhou University,2013:35-41(in Chinese).

[9] 曹力,贾倩茜,舒平,等. 面向签派的飞行燃油消耗估计方法[J]. 中国民航飞行学院学报,2015,26(1):24-28.
CAO L,JIA Q Q,SHU P,et al. Flight fuel consumption evaluation method for dispatch business[J]. Journal of Civil Aviation Flight University of China,2015,26(1):24-28(in Chinese).

[10] 刘婧. 基于飞行数据分析的飞机燃油估计模型[D]. 南京:南京航空航天大学,2010:26-44.
LIU J. The aircraft fuel estimation model based on flight data analysis[D]. Nanjing:Nanjing University of Aeronautics and Astronautics,2010:26-44(in Chinese).

[11] 刘芳. 下降阶段的飞机油耗优化建模方法研究[D]. 南京:南京航空航天大学,2012:16-23.
LIU F. Research on optimization modeling of aircraft fuel consumption in terminal descent[D]. Nanjing:Nanjing University of Aeronautics and Astronautics,2012:16-23(in Chinese).

[12] COLLINS B P. Estimation of aircraft fuel consumption[J]. Journal of Aircraft,2015,19(11):969-975.

[13] NIKOLERIS T,GUPTA G,KISTLER M. Detailed estimation of fuel consumption and emissions during aircraft taxi operations at Dallas/Fort Worth International Airport[J]. Transportation Research Part D Transport & Environment,2011,16(4):302-308.

[14] SINGH V,SHARMA S K. Fuel consumption optimization in air transport:A review,classification,critique,simple meta-analysis,and future research implications[J]. European Transport Research Review,2015,7(2):1-24.

[15] 张军峰,蒋海行,武晓光,等. 基于 BADA 及航空器意图的四维航迹预测[J]. 西南交通大学学报,2014,49(3):553-558.
ZHANG J F,JIANG H X,WU X G,et al. 4D trajectory prediction based on BADA and aircraft intent[J]. Journal of Southwest Jiaotong University,2014,49(3):553-558(in Chinese).

[16] 李鹏,田景奎. 不同下垫面近地层风速廓线特征[J]. 资源科学,2011,33(10):2005-2010.
LI P,TIAN J K. Characteristics of surface layer wind speed profiles over different underlying surfaces[J]. Resources Science,2011,33(10):2005-2010(in Chinese).

作者简介:

胡荣 男,博士,副教授,硕士生导师。主要研究方向:民航可持续发展、民航经济分析。

吴文洁 女,硕士研究生。主要研究方向:绿色民航。

陈琳 女,硕士研究生。主要研究方向:交通运输规划与管理。

张菲菲 女,硕士研究生。主要研究方向:交通运输规划与管理。

北京航空航天大学学报

北航学报 赠阅

Influence of meteorological factors on aircraft fuel efficiency in approach flight

HU Rong*, WU Wenjie, CHEN Lin, ZHANG Feifei

(College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)

Abstract: Meteorological factors are significant to aviation flight. To investigate the fuel efficiency in aviation flight, base of aircraft data (BADA) model was combined with meteorological factors, and the aircraft fuel consumption correction model was built. Taking the example of one inbound flight approach at Guangzhou Baiyun Airport, the flight simulation tests were developed and the influence of the change in air temperature, air pressure and wind speed on aircraft fuel efficiency was discussed from two aspects: fuel flow and fuel consumption. The results show that the correlation between meteorological factors and fuel efficiency is obvious. When aircraft flight altitude is certain, if air temperature rises, fuel flow and fuel consumption increase, and fuel efficiency decreases; if air pressure rises, fuel flow changes little but fuel consumption slightly decreases, and fuel efficiency increases; if wind speed rises, fuel flow and fuel consumption firstly decrease and then increase, fuel efficiency increases first and decreases afterwards, and fuel efficiency is the highest when wind speed is 4 m/s. If flight altitude reduces, air temperature and air pressure increase but wind speed increases, fuel flow increases with a slight fluctuation, and fuel efficiency decreases. Fuel consumption in approach can be reduced by about 3% under the best meteorological factors. The study results are significant for improving the fuel efficiency in actual flight.

Keywords: meteorological factors; air temperature; air pressure; wind speed; fuel efficiency; fuel flow; fuel consumption

Received: 2017-04-19; **Accepted**: 2017-08-11; **Published online**: 2017-09-22 10:25

URL: kns.cnki.net/kcms/detail/11.2625.V.20170922.1025.004.html

Foundation items: National Natural Science Foundation of China (71201082, 71401072); Social Science Foundation of Jiangsu Province (15GLC001)

* **Corresponding author**. E-mail: hoorong@nuaa.edu.cn

北京航空航天大学学报

北航学报 赠阅

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2017.0277

BDS/GPS 组合导航接收机自主完好性监测算法

王尔申[1,*], 杨福霞[1], 庞涛[1], 曲萍萍[1], 蓝晓宇[1], 姜毅[2]

(1. 沈阳航空航天大学 电子信息工程学院, 沈阳 110136; 2. 大连海事大学 信息科学技术学院, 大连 116026)

摘 要: 为使接收机自主完好性监测(RAIM)技术应用于民航垂直引导进近(APV)飞行阶段成为可能,研究了 BDS/GPS 组合导航 RAIM 算法。提出了一种基于 BDS/GPS 定位解最优加权平均解的算法,结合最优加权平均解与 BDS/GPS 定位解的关系建立检验统计量,根据最大允许的虚警率计算检验门限,实现对故障所在卫星导航系统的检测,并采用加权最小二乘残差法对故障进行检测与识别。研究结果对多星座组合卫星导航系统应用于民航 APV 飞行阶段的导航具有一定的参考意义。

关 键 词: 接收机自主完好性监测(RAIM); 北斗卫星导航系统(BDS); 组合导航; 垂直引导进近(APV); 加权平均

中图分类号: V241.6; TN967.1

文献标识码: A **文章编号**: 1001-5965(2018)04-0684-07

全下文载

接收机自主完好性监测(RAIM)算法源于民航用户对卫星导航系统可靠性的苛刻要求[1]。目前,卫星导航完好性监测技术主要有 3 种:卫星自主完好性监测(SAIM)、利用接收机内部冗余信息的 RAIM 和外部辅助增强系统[2]。相对于其他 2 种完好性监测技术,RAIM 是利用接收机的超定解对卫星测量值进行一致性校验的技术,其更能保证用户对告警时间的要求。早期的 RAIM 算法大多基于 GPS 单星座,可以满足民航航路到非精密进近阶段的完好性要求。目前,全球卫星导航系统(GNSS)发展为多星座多频点,参与定位的卫星成倍增加,接收机具有更多的冗余观测信息,可有效地减少 RAIM 空洞,也使 RAIM 技术可应用于一些对完好性性能要求更高的应用,如民航飞行中垂直引导进近(APV)Ⅰ、Ⅱ等阶段。

多星座 RAIM 算法是指使用 2 个或 2 个以上的卫星导航系统的冗余观测信息进行一致性校验,以检测和识别故障。单星座系统中至少观测到 5 颗卫星才能进行故障检测,考虑到不同星座的时间系统之间的偏差,对于多星座需要更多冗余观测量才能进行故障检测。文献[3]对 RAIM 的可用性预测进行了分析。文献[4-6]提出了针对多系统多故障情况下的完好性监测算法,其中文献[4]对 GPS/Galileo/GLONASS 组合导航在 APV-Ⅰ、APV-Ⅱ阶段利用 RAIM 算法进行了分析。Ene 等[7-8]分析了 GPS/Galileo 组合导航在 APV 阶段的 RAIM 可用性。文献[9]对 GPS/Galileo 组合导航 RAIM 算法中的截止高度角、用户等效距离误差等问题进行了分析。徐君毅等[10]对 APV 阶段的垂直保护水平在全球的分布情况进行了研究。Lee 等[11]提出了最优加权平均解(OWAS)算法,国内的研究人员也对此算法进行了扩展,郭靖等[12]采用新的分组方法,满足双星座双故障的应用场景,王式太[13]设计了基于 M

收稿日期: 2017-05-03; **录用日期**: 2017-05-19; **网络出版时间**: 2017-09-26 14:37

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20170926.1437.001.html

基金项目: 国家自然科学基金(61571309,61101161); 中央高校基本科研业务费专项资金(3132016317); 辽宁省"百千万人才工程"; 辽宁省高等学校优秀人才支持计划(LR2016069); 辽宁省教育厅科学技术研究项目(L201716); 航空科学基金(2015ZC54010)

* **通信作者**. E-mail: wes2016@126.com

引用格式: 王尔申, 杨福霞, 庞涛, 等. BDS/GPS 组合导航接收机自主完好性监测算法[J]. 北京航空航天大学学报, 2018, 44(4): 684-690. WANG E S, YANG F X, PANG T, et al. BDS/GPS combined navigation receiver autonomous integrity monitoring algorithm[J]. Journal of Beijing University of Aeronautics and Astronautics, 2018, 44(4): 684-690 (in Chinese).

估计的 RAIM 粗差探测和剔除算法。其中，OWAS 算法能够方便地兼容不同的星座，避免了不同系统间时间系统、坐标系统转换带来的误差影响，具有处理速度快、可用性高的特点，因此，本文采用 OWAS 算法对 BDS/GPS 组合导航应用于 APV 阶段进行了研究。

OWAS 算法是将卫星按星座分为 2 组，监测同时发生在同一星座中的多星故障，但此算法不能有效地对双星座同时发生故障进行检测。本文采用 OWAS 算法，通过优化加权因子，提高算法的可用性，并对 OWAS 算法进行了改进。通过实测数据验证了 OWAS 算法对 APV 阶段的可用性和双星座双故障的检测。

1　OWAS RAIM 算法

OWAS RAIM 算法原理如图 1 所示。

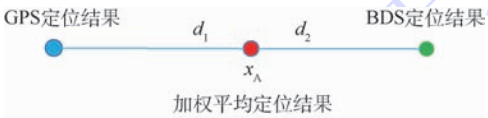


图 1　OWAS RAIM 算法原理

Fig. 1　OWAS RAIM algorithm principle

加权平均后的定位结果为

$$\boldsymbol{x}_A = r\boldsymbol{x}_1 + (1-r)\boldsymbol{x}_2 \tag{1}$$

式中：加权因子 $0 \leqslant r \leqslant 1$；$\boldsymbol{x}_n$（$n=1,2$）分别为 BDS 和 GPS 的定位结果。

图 1 中：d_n（$n=1,2$）分别为 $\boldsymbol{x}_n$ 与 $\boldsymbol{x}_A$ 的间隔距离。对于加权平均解，要求以 0.999 的概率满足：$|\boldsymbol{x}_A - \boldsymbol{x}| \leqslant |\boldsymbol{x}_A - \boldsymbol{x}_n| + |\boldsymbol{x}_n - \boldsymbol{x}| \leqslant T + a_n$。其中：$T$（$n=1,2$）为检验门限，$a_n$ 为无故障假设条件下定位误差上限。所以，无故障卫星时，每个导航系统内的保护水平为：$\mathrm{PL}_n = T + a_n$，$\mathrm{PL} = \max(\mathrm{PL}_1, \mathrm{PL}_2)$。OWAS 算法在只考虑垂直方向上的定位精度与完好性要求时，可以归纳为如下问题：

$$\min(\max(\mathrm{VPL}_1, \mathrm{VPL}_2)) \tag{2}$$

式中：VPL 为垂直保护级别。

同时，满足以下条件时，RAIM 算法可用：

$$\sigma_V \leqslant \sigma_{V\text{-}max} \tag{3}$$

$$d_{major} \leqslant d_{major\text{-}max} \tag{4}$$

$$\mathrm{VPL} \leqslant \mathrm{VAL} \tag{5}$$

$$\mathrm{HPL} \leqslant \mathrm{HAL} \tag{6}$$

式中：σ_V 为垂直方向上的定位误差标准差；$\sigma_{V\text{-}max}$ 为垂直方向上的定位误差标准差的最大值；d_{major} 为最终定位解水平方向上的定位误差标准差；$d_{major\text{-}max}$ 为飞行阶段规定的 95% 水平方向上的定位误差标准差的最大值；VAL 为垂直保护门限；HPL 为水平保护级别；HAL 为水平保护门限[10]。

σ_V 和 d_{major} 的标准由以下 2 个操作需求推导：

1）95% 的导航系统误差（Navigation System Errors，NSE）边界：$1.96\sigma_V = 4$ m，$1.96 d_{major} = 16$ m（1.96 对应于高斯分布假设的 95% 分布）。

2）NSE 边界满足无故障条件完好性可用性时：$K_V \sigma_V \leqslant$ VAL（垂直方向），$K_H d_{major} \leqslant$ HAL（水平方向），当 $P_r\{\mathrm{HMI}\} = 10^{-8}/\mathrm{approach}$（$P_r$ 为概率，HMI 为危险误导信息，approach 为进近）时，$K_V = 5.73$，$K_H = 6.47$[10]。

$\sigma_{V\text{-}max}$ 和 $d_{major\text{-}max}$ 的值分别为

$$\sigma_{V\text{-}max} = \min(4\ \mathrm{m}/1.96, \mathrm{VAL}/K_V)$$

$$d_{major\text{-}max} = \min(16\ \mathrm{m}/1.96, \mathrm{HAL}/K_H)$$

1.1　协方差矩阵

OWAS 算法为

$$\boldsymbol{x}_A = \frac{w_1\boldsymbol{x}_1 + w_2\boldsymbol{x}_2}{w_1 + w_2} = \frac{w_1}{w_1 + w_2}\boldsymbol{x}_1 + \frac{w_2}{w_1 + w_2}\boldsymbol{x}_2 = r\boldsymbol{x}_1 + (1-r)\boldsymbol{x}_2 \tag{7}$$

式中：w_1 为 GPS 定位解在最终解中所占权重；w_2 为 BDS 定位解在最终解中所占权重；$r \equiv \dfrac{w_1}{w_1 + w_2}$，且 $0 \leqslant r \leqslant 1$。

OWAS 算法的位置误差与加权因子的关系如下：

$$\Delta\boldsymbol{x}_A = r\Delta\boldsymbol{x}_1 + (1-r)\Delta\boldsymbol{x}_2 \tag{8}$$

式中：$\Delta\boldsymbol{x}_1$ 为 GPS 定位解与真实位置之间的误差；$\Delta\boldsymbol{x}_2$ 为 BDS 定位解与真实位置之间的误差；$\Delta\boldsymbol{x}_A$ 为最终加权平均定位解与真实位置之间的误差。

加权最小二乘定位解的位置误差[14]表达式如下：

$$\Delta\boldsymbol{x}_n = (\boldsymbol{G}_n^{\mathrm{T}}\boldsymbol{W}_n\boldsymbol{G}_n)^{-1}\boldsymbol{G}_n^{\mathrm{T}}\boldsymbol{W}_n\Delta\boldsymbol{R}_n \tag{9}$$

$$\boldsymbol{P}_n \equiv E(\Delta\boldsymbol{x}_n\Delta\boldsymbol{x}_n^{\mathrm{T}}) = (\boldsymbol{G}_n^{\mathrm{T}}\boldsymbol{W}_n\boldsymbol{G}_n)^{-1} \tag{10}$$

式中：$\boldsymbol{G}$ 为伪距误差与接收机位置及钟差的关联矩阵；$\boldsymbol{W}$ 为加权矩阵；$\Delta\boldsymbol{R}$ 为伪距误差矢量；$\boldsymbol{P}_n$ 为定位误差的协方差矩阵（$n=1,2$ 分别表示 GPS 和 BDS）。

最优解和各导航系统定位解误差互相关表达式如下：

$$E(\Delta\boldsymbol{x}_A\Delta\boldsymbol{x}_1^{\mathrm{T}}) = r\boldsymbol{P}_1 + (1-r)E(\Delta\boldsymbol{x}_2\Delta\boldsymbol{x}_1^{\mathrm{T}}) \tag{11}$$

$$E(\Delta\boldsymbol{x}_A\Delta\boldsymbol{x}_2^{\mathrm{T}}) = rE(\Delta\boldsymbol{x}_1\Delta\boldsymbol{x}_2^{\mathrm{T}}) + (1-r)\boldsymbol{P}_2 \tag{12}$$

假设 GPS 与 BDS 伪距误差不相关，则 $E(\Delta\boldsymbol{R}_2 \cdot \Delta\boldsymbol{R}_1^{\mathrm{T}}) = 0$，$E(\Delta\boldsymbol{R}_1\Delta\boldsymbol{R}_2^{\mathrm{T}}) = 0$。由此，得

$$E(\Delta\boldsymbol{x}_2\Delta\boldsymbol{x}_1^{\mathrm{T}}) = (\boldsymbol{G}_2^{\mathrm{T}}\boldsymbol{W}_2\boldsymbol{G}_2)^{-1}\boldsymbol{G}_2^{\mathrm{T}}\boldsymbol{W}_2E(\Delta\boldsymbol{R}_2\Delta\boldsymbol{R}_1^{\mathrm{T}}) \cdot \boldsymbol{W}_1\boldsymbol{G}_1(\boldsymbol{G}_1^{\mathrm{T}}\boldsymbol{W}_1\boldsymbol{G}_1)^{-1} = 0 \tag{13}$$

北航学报 赠阅

$$E(\Delta \boldsymbol{x}_1 \Delta \boldsymbol{x}_2^{\mathrm{T}}) = (\boldsymbol{G}_1^{\mathrm{T}} \boldsymbol{W}_1 \boldsymbol{G}_1)^{-1} \boldsymbol{G}_1^{\mathrm{T}} \boldsymbol{W}_1 \cdot E(\Delta \boldsymbol{R}_1 \Delta \boldsymbol{R}_2^{\mathrm{T}}) \boldsymbol{W}_2 \boldsymbol{G}_2 (\boldsymbol{G}_2^{\mathrm{T}} \boldsymbol{W}_2 \boldsymbol{G}_2)^{-1} = 0 \tag{14}$$

因此，$E(\Delta \boldsymbol{x}_{\mathrm{A}} \Delta \boldsymbol{x}_1^{\mathrm{T}}) = r\boldsymbol{P}_1$，$E(\Delta \boldsymbol{x}_{\mathrm{A}} \Delta \boldsymbol{x}_2^{\mathrm{T}}) = (1-r)\boldsymbol{P}_2$，即

$$E(\Delta \boldsymbol{x}_1 \Delta \boldsymbol{x}_{\mathrm{A}}^{\mathrm{T}}) = r\boldsymbol{P}_1 \tag{15}$$

$$E(\Delta \boldsymbol{x}_2 \Delta \boldsymbol{x}_{\mathrm{A}}^{\mathrm{T}}) = (1-r)\boldsymbol{P}_2 \tag{16}$$

同理，加权平均解 $\boldsymbol{x}_{\mathrm{A}}$ 的定位误差协方差矩阵 $\boldsymbol{P}_{\mathrm{A}}$ 如下：

$$\boldsymbol{P}_{\mathrm{A}} = E(\Delta \boldsymbol{x}_{\mathrm{A}} \Delta \boldsymbol{x}_{\mathrm{A}}^{\mathrm{T}}) = r^2 \boldsymbol{P}_1 + (1-r)^2 \boldsymbol{P}_2 \tag{17}$$

协方差矩阵 $\boldsymbol{J}_n$ 为

$$\boldsymbol{J}_n \equiv E((\Delta \boldsymbol{x}_n - \Delta \boldsymbol{x}_{\mathrm{A}})(\Delta \boldsymbol{x}_n - \Delta \boldsymbol{x}_{\mathrm{A}})^{\mathrm{T}}) = E(\Delta \boldsymbol{x}_n \Delta \boldsymbol{x}_n^{\mathrm{T}} - \Delta \boldsymbol{x}_{\mathrm{A}} \Delta \boldsymbol{x}_n^{\mathrm{T}} - \Delta \boldsymbol{x}_n \Delta \boldsymbol{x}_{\mathrm{A}}^{\mathrm{T}} + \Delta \boldsymbol{x}_{\mathrm{A}} \Delta \boldsymbol{x}_{\mathrm{A}}^{\mathrm{T}}) \tag{18}$$

将式(10)、式(11)、式(15)、式(17)代入式(18)中，可得

$$\boldsymbol{J}_1 = (1-r)^2(\boldsymbol{P}_1 + \boldsymbol{P}_2) = (1-r)^2 \boldsymbol{P}_{\mathrm{S}} \tag{19}$$

式中：$\boldsymbol{P}_{\mathrm{S}} = \boldsymbol{P}_1 + \boldsymbol{P}_2$。

同理

$$\boldsymbol{J}_2 = E((\Delta \boldsymbol{x}_2 - \Delta \boldsymbol{x}_{\mathrm{A}})(\Delta \boldsymbol{x}_2 - \Delta \boldsymbol{x}_{\mathrm{A}})^{\mathrm{T}}) = r^2(\boldsymbol{P}_1 + \boldsymbol{P}_2) = r^2 \boldsymbol{P}_{\mathrm{S}} \tag{20}$$

1.1.1 加权因子确定保护级别

令

$$K_{\mathrm{fd}} \equiv Q^{-1}\left(\frac{P_{\mathrm{fd}}}{4}\right) \tag{21}$$

$$K_{\mathrm{md}} \equiv Q^{-1}(P_{\mathrm{md}}) \tag{22}$$

式中：Q^{-1} 为标准单边累加分布函数的逆函数；P_{fd} 为虚警率(参考国际民航不同进近阶段导航性能要求，虚警率与连续性保持一致。虚警率除以 4 是因为总的允许虚警率在垂直和水平方向上平均分配[10])；P_{md} 为最大允许漏警率(完好性风险等于漏警率与卫星故障率的乘积)。

令

$$\sigma_{\mathrm{V},1} \equiv \sqrt{\boldsymbol{P}_1(3,3)} \tag{23}$$

$$\sigma_{\mathrm{V},2} \equiv \sqrt{\boldsymbol{P}_2(3,3)} \tag{24}$$

$$\sigma_{\mathrm{V,S}} \equiv \sqrt{\boldsymbol{P}_{\mathrm{S}}(3,3)} = \sqrt{\boldsymbol{P}_1(3,3) + \boldsymbol{P}_2(3,3)} \tag{25}$$

由式(19)和式(20)可以得出

$$\sigma_{\mathrm{dV},1} \equiv \sqrt{\boldsymbol{J}_1(3,3)} = (1-r)\sqrt{\boldsymbol{P}_{\mathrm{S}}(3,3)} = (1-r)\sigma_{\mathrm{V,S}} \tag{26}$$

$$\sigma_{\mathrm{dV},2} \equiv \sqrt{\boldsymbol{J}_2(3,3)} = r\sqrt{\boldsymbol{P}_{\mathrm{S}}(3,3)} = r\sigma_{\mathrm{V,S}} \tag{27}$$

VPL_1 和 VPL_2 分别为

$$\mathrm{VPL}_1 = T_{\mathrm{V},1} + a_{\mathrm{V},1} = K_{\mathrm{fd}}\sigma_{\mathrm{dV},1} + K_{\mathrm{md}}\sigma_{\mathrm{V},1} = K_{\mathrm{fd}}(1-r)\sigma_{\mathrm{V,S}} + K_{\mathrm{md}}\sigma_{\mathrm{V},1} \tag{28}$$

$$\mathrm{VPL}_2 = T_{\mathrm{V},2} + a_{\mathrm{V},2} = K_{\mathrm{fd}}\sigma_{\mathrm{dV},2} + K_{\mathrm{md}}\sigma_{\mathrm{V},2} = K_{\mathrm{fd}} r\sigma_{\mathrm{V,S}} + K_{\mathrm{md}}\sigma_{\mathrm{V},2} \tag{29}$$

由图 2 可以看出，VPL_1 是加权因子的一元线性递减函数；VPL_2 是加权因子的一元线性递增函数。在 OWAS 算法中，$\mathrm{VPL} = \max(\mathrm{VPL}_1, \mathrm{VPL}_2)$，当 $\mathrm{VPL}_1 = \mathrm{VPL}_2$ 时，得到最小化的 VPL，此时可得

$$r_{\text{min-VPL}} = \frac{1}{2} + \frac{1}{2} \cdot \frac{K_{\mathrm{md}}}{K_{\mathrm{fd}}} \cdot \frac{\sigma_{\mathrm{V},1} - \sigma_{\mathrm{V},2}}{\sigma_{\mathrm{V,S}}} \tag{30}$$

$$\mathrm{VPL} = \frac{1}{2} K_{\mathrm{fd}} \sigma_{\mathrm{V,S}} + \frac{1}{2} K_{\mathrm{md}}(\sigma_{\mathrm{V},1} + \sigma_{\mathrm{V},2}) \tag{31}$$

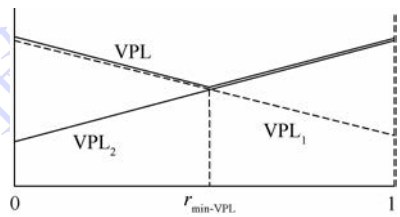


图 2 VPL_1、VPL_2 与加权因子的函数

Fig. 2 VPL_1 and VPL_2 as a function of weight factor

1.1.2 加权因子确定误差精度 σ_{V} 和 d_{major}

OWAS 位置误差的协方差矩阵 $\boldsymbol{P}_{\mathrm{A}} = E(\Delta \boldsymbol{x}_{\mathrm{A}} \Delta \boldsymbol{x}_{\mathrm{A}}^{\mathrm{T}}) = r^2 \boldsymbol{P}_1 + (1-r)^2 \boldsymbol{P}_2$。

σ_{V} 和 d_{major} 为

$$\sigma_{\mathrm{V}} = \sqrt{\boldsymbol{P}_{\mathrm{A}}(3,3)} \tag{32}$$

$$d_{\mathrm{major}} = \left\{ \frac{\boldsymbol{P}_{\mathrm{A}}(1,1) + \boldsymbol{P}_{\mathrm{A}}(2,2)}{2} + \left[\left(\frac{\boldsymbol{P}_{\mathrm{A}}(1,1) - \boldsymbol{P}_{\mathrm{A}}(2,2)}{2} \right)^2 + (\boldsymbol{P}_{\mathrm{A}}(1,2))^2 \right]^{\frac{1}{2}} \right\}^{\frac{1}{2}} \tag{33}$$

由式(3)、式(4)可知，RAIM 算法可用要满足：

$\sigma_{\mathrm{V}} \leqslant \sigma_{\text{V-max}}$

$d_{\mathrm{major}} \leqslant d_{\text{major-max}}$

1.1.3 确定最优加权因子

OWAS 最重要的是在降低保护水平的条件下满足精度的要求。

如图 3 所示，$r_{\text{min-VPL}}$ 表示垂直保护水平最小时的加权因子，但此时不满足 σ_{V} 的标准，而 $r_{\text{min-}\sigma_{\mathrm{V}}}$ 满足最小化 σ_{V}，却不满足 $\mathrm{VPL} \leqslant \mathrm{VAL}$，$r_{\text{best-int}}$ 表示在保证精度要求下修正加权因子 $r_{\text{min-VPL}}$ 来最小化 VPL，$r_{\text{best-acc}}$ 表示通过修正最小化精度加权因子 σ_{V} 保证可用性的要求。

APV 飞行阶段时需更大可能的保证算法的可用性，故采用 $r_{\text{best-int}}$ 作为最终加权因子，则

$$\sigma_{\mathrm{V}} \leqslant \sigma_{\text{V-max}} \rightarrow \boldsymbol{P}_{\mathrm{A}}(3,3) \leqslant \sigma_{\text{V-max}}^2 \rightarrow r^2 \boldsymbol{P}_1(3,3) + (1-r)^2 \boldsymbol{P}_2(3,3) - \sigma_{\text{V-max}}^2 \leqslant 0 \tag{34}$$

北航学报 赠阅

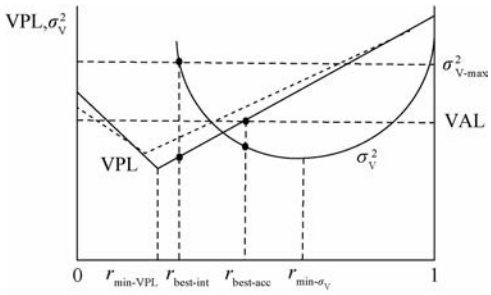


图 3　加权因子的计算

Fig. 3　Determination of weight factor

式(23)和式(24)代入到式(34)中，可得

$$(\sigma_{V,1}^2+\sigma_{V,2}^2)r^2-2\sigma_{V,2}^2 r-(\sigma_{V\text{-max}}^2-\sigma_{V,2}^2)\leqslant 0 \tag{35}$$

由式(35)计算可得 2 个解，如下：

$$r_1=\frac{\sigma_{V,2}^2-\sqrt{\sigma_{V,2}^4+(\sigma_{V,1}^2+\sigma_{V,2}^2)(\sigma_{V\text{-max}}^2-\sigma_{V,2}^2)}}{\sigma_{V,1}^2+\sigma_{V,2}^2} \tag{36}$$

$$r_2=\frac{\sigma_{V,2}^2+\sqrt{\sigma_{V,2}^4+(\sigma_{V,1}^2+\sigma_{V,2}^2)(\sigma_{V\text{-max}}^2-\sigma_{V,2}^2)}}{\sigma_{V,1}^2+\sigma_{V,2}^2} \tag{37}$$

由于 $r_1\leqslant r\leqslant r_2$，且 $0\leqslant r\leqslant 1$，所以

$$r_{acc,min}\leqslant r\leqslant r_{acc,max} \tag{38}$$

式中：$r_{acc,min}=\max(r_1,0)$；$r_{acc,max}=\min(r_2,1)$。

当 $r_{acc,min}\leqslant r_{min\text{-}VPL}\leqslant r_{acc,max}$ 时，加权因子 r 如下：

$$r=r_{min\text{-}VPL} \tag{39}$$

当 $r_{min\text{-}VPL}\leqslant r_{acc,min}$ 时，加权因子 r 如下：

$$r=r_{acc,min} \tag{40}$$

当 $r_{acc,max}\leqslant r_{min\text{-}VPL}$ 时，加权因子 r 如下：

$$r=r_{acc,max} \tag{41}$$

1.2　故障检测与识别

OWAS RAIM 算法是以 GPS、BDS 到 $\boldsymbol{x}_A$ 的距离作为故障检测的检验统计量[15-16]，按照国际民航组织(ICAO)要求的性能标准得到检验门限。当双星座同时出现故障时，2 个检验统计量均会在故障产生的时刻超过检验门限。当单星座产生故障时，故障星座的定位解使得 OWAS 与无故障星座的距离增加，容易产生误警，此时采用加权最小二乘残差的故障检测对 OWAS 定位域的故障检测进行辅助，完成单星座故障的检测。

1.2.1　故障检测

GNSS 伪距观测线性化方程为

$$\Delta\boldsymbol{R}=\boldsymbol{G}\boldsymbol{x}+\boldsymbol{\varepsilon} \tag{42}$$

式中：$\boldsymbol{\varepsilon}$ 为观测噪声矢量。

对式(42)运用加权最小二乘法求解，得到的解为

$$\boldsymbol{x}_{wls}=(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{G})^{-1}\boldsymbol{G}^{\mathrm{T}}\boldsymbol{W}\Delta\boldsymbol{R}=\boldsymbol{B}\Delta\boldsymbol{R} \tag{43}$$

式中：$\boldsymbol{B}=(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{G})^{-1}\boldsymbol{G}^{\mathrm{T}}\boldsymbol{W}$。

伪距残差矢量 $\boldsymbol{v}$ 为

$$\boldsymbol{v}=\Delta\boldsymbol{R}-\boldsymbol{G}\boldsymbol{x}_{wls}=(\boldsymbol{I}_n-\boldsymbol{G}(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{G})^{-1}\boldsymbol{G}^{\mathrm{T}}\boldsymbol{W})\Delta\boldsymbol{R}=\boldsymbol{C}\Delta\boldsymbol{R} \tag{44}$$

矩阵 $\boldsymbol{C}$ 的协因数矩阵为

$$\boldsymbol{Q}=\boldsymbol{W}^{-1}-\boldsymbol{G}(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{G})^{-1}\boldsymbol{G}^{\mathrm{T}}$$

伪距残差平方和为

$$\mathrm{SSE}_{\mathrm{WLS}}=\boldsymbol{v}^{\mathrm{T}}\boldsymbol{v}=\Delta\boldsymbol{R}^{\mathrm{T}}\boldsymbol{W}\Delta\boldsymbol{R} \tag{45}$$

在卫星导航系统正常工作的情况下，各卫星伪距残差都较小，伪距残差平方和也较小；当某个测量伪距存在较大偏差时，伪距残差平方和会变大。假设卫星导航系统正常工作时，观测噪声 $\boldsymbol{\varepsilon}$ 中的各个分量是均值为 0、方差为 σ^2 的正态分布随机误差，且它们相互独立。依据统计分布理论，若 $\boldsymbol{\varepsilon}$ 的均值为 0，$\mathrm{SSE}_{\mathrm{WLS}}$ 服从自由度为 $(n-4)$ 的 χ^2 分布；若 $\boldsymbol{\varepsilon}$ 的均值不为 0，则 $\mathrm{SSE}_{\mathrm{WLS}}$ 服从自由度为 $(n-4)$ 的非中心化 χ^2 分布。对 $\mathrm{SSE}_{\mathrm{WLS}}$ 作二元假设：

1) 无故障假设 $H_0: E(\boldsymbol{\varepsilon})=0$，则 $\mathrm{SSE}_{\mathrm{WLS}}\sim\chi^2(n-4)$。

2) 有故障假设 $H_1: E(\boldsymbol{\varepsilon})\neq 0$，则 $\mathrm{SSE}_{\mathrm{WLS}}\sim\chi^2(\lambda,n-4)$。

$$P_{\mathrm{r}}(\mathrm{SSE}_{\mathrm{WLS}}>t_{\mathrm{WLS}})=1-\int_0^{t_{\mathrm{WLS}}}f_{\chi^2(n-4)}(x)\,\mathrm{d}x=\int_{t_{\mathrm{WLS}}}^{\infty}f_{\chi^2(n-4)}(x)\,\mathrm{d}x=P_{\mathrm{FA}} \tag{46}$$

由式(46)可以解得 t_{WLS} 值。

1.2.2　故障识别

根据残差和观测误差的关系式，定义故障识别的检验统计量为

$$d_i=\frac{|\boldsymbol{v}_i|}{\sqrt{\boldsymbol{Q}_{ii}}} \tag{47}$$

式中：$i=1,2,\cdots,n$，并且 n 为可见卫星数目；$\boldsymbol{v}_i$ 为利用加权最小二乘法求得的伪距残差；$\boldsymbol{Q}_{ii}$ 为矩阵 $\boldsymbol{Q}$ 的第 i 行第 i 列；$d_i\sim N(0,1)$。

对统计量 d_i 做二元假设：

1) H_0(无故障)：$E(\boldsymbol{\varepsilon}_i)=0$，则 $d_i\sim N(0,1)$。

2) H_1(有故障)：$E(\boldsymbol{\varepsilon}_i)\neq 0$，则 $d_i\sim N(\delta_i,1)$。

其中，δ_i 为统计量偏移参数。给定总体的误警率 P_{FA}，可计算出对应的检验门限 $T_{1\,\mathrm{WLS}}$。计算如下：

$$P_{\mathrm{r}}(d_i>T_{1\,\mathrm{WLS}})=\int_{T_{1\,\mathrm{WLS}}}^{\infty}f_{N(0,1)}(x)\,\mathrm{d}x=P_{\mathrm{FA}}/n \tag{48}$$

由式(48)可解得门限值 $T_{1\,\mathrm{WLS}}$。

北航学报 赠阅

设 $d_i = |v_i| / \sqrt{Q_{ii}}$ 为检验统计量，对应的检验门限值为 $T_{1\,\mathrm{WLS}}$。比较检验统计量 d_i 与检验门限 $T_{1\,\mathrm{WLS}}$ 进行卫星故障识别。

2 算法验证与结果分析

采用 IGS(International GPS Service)2016 年 7 月 31 日 0 时的 BDS、GPS 系统 RINEX 数据对本文算法进行仿真验证。

在 GPS 可见星中的第 16 颗卫星、BDS 可见星中的第 7 颗卫星中，分别在 500 ~ 900 s 时刻人为加入 30、50、70 m 伪距偏差。

图 4(a)、(b)分别为未加伪距偏差时 GPS 和 BDS 的垂向定位域检验统计量。图 5(a)、(b)分别为在 500 s 开始加入 30、50、70 m 伪距偏差时GPS、BDS的检验统计量。可知，当加入伪距偏差时，检验统计量值会发生跳变；未加入伪距偏差时，检验统计量值波动较小。

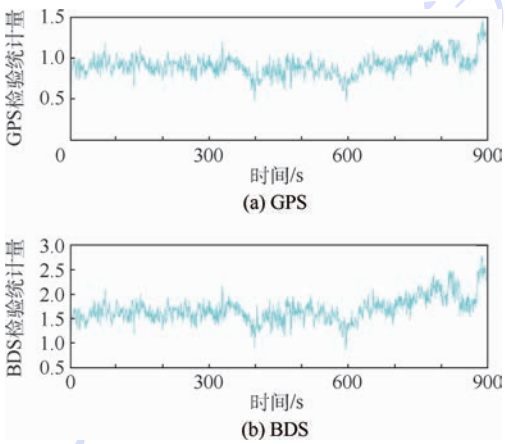


图 4 BDS/GPS 未加入伪距偏差的检验统计量

Fig. 4 BDS/GPS test statistics with no pseudo-range bias

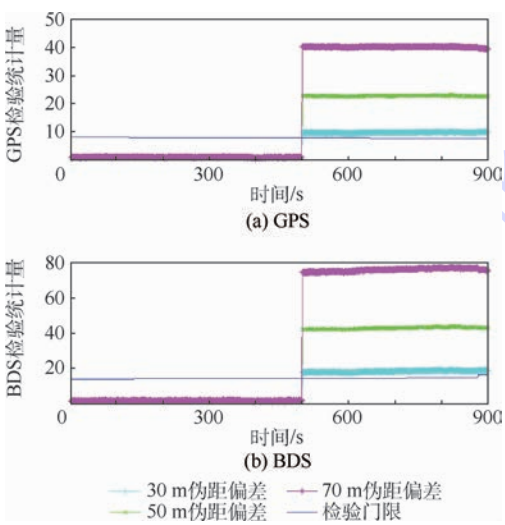


图 5 BDS/GPS 加入不同伪距偏差的检验统计量

Fig. 5 BDS/GPS test statistics with different pseudo-range bias

双故障卫星检测结果如图 6 所示。

由图 6 可知，在 GPS 和 BDS 系统的第 16 颗卫星和第 7 颗卫星中加入 50 m 伪距偏差，检验统计量大于检验门限，可以对故障进行检测。

双星座双故障算法识别率结果如图 7 所示。可知，当伪距偏差值较小时，不能识别出故障卫星；当故障卫星伪距偏差在 30 m 以上时，BDS/GPS 能识别全部故障。

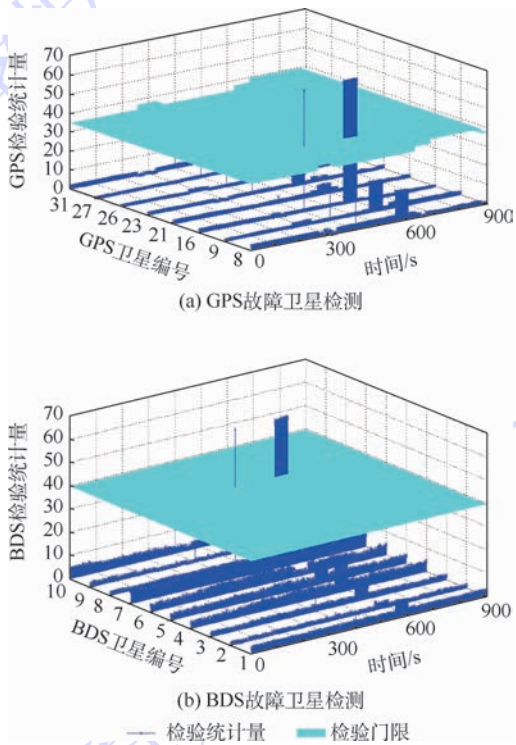


图 6 BDS/GPS 故障卫星检测

Fig. 6 BDS fault satellite detection

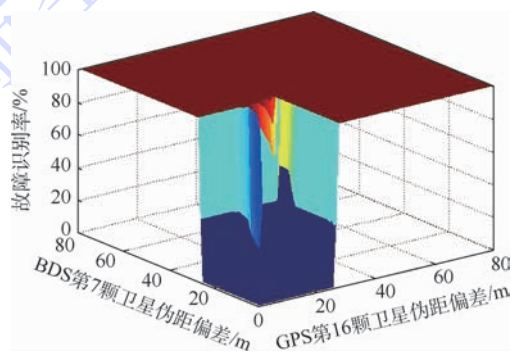


图 7 故障识别率

Fig. 7 Fault identification rate

3 结 论

本文研究了一种 BDS/GPS 组合导航 RAIM 算法，对算法进行了详细的理论分析，并利用 IGS

北航学报 赠阅

网站数据对算法进行验证。

1）本文算法对 BDS/GPS 双卫星故障检测与识别是可用的。

2）当伪距偏差在 30 m 以上时可有效地检测与识别。

研究结果对多星座组合导航接收机自主完好性监测算法研究有一定的参考价值。

参考文献（References）

[1] BROCARD P, JULIEN O, MABILLEAU M. Autonomous integrity monitoring proposal for critical rail applications[C]// Proceedings of the 28th International Technical Meeting of the Satellite Division of the Institute of Navigation. Washington, D. C.: ION, 2015: 706-734.

[2] 徐肖豪，杨传森，刘瑞华. GNSS 用户端自主完好性监测研究综述[J]. 航空学报，2013，34(3)：451-463.
XU X H, YANG C S, LIU R H. Review and prospect of GNSS receiver autonomous integrity monitoring[J]. Acta Aeronautica et Astronautica Sinica, 2013, 34(3): 451-463(in Chinese).

[3] 吕宗平，倪育德，陈君，等. 基于 MHSS 算法的 ARAIM 完好性和可用性预测[J]. 中国空间科学技术，2017，37(1)：41-48.
LYU Z P, NI Y D, CHEN J, et al. Prediction of ARAIM integrity and availability based on MHSS algorithm[J]. Chinese Space Science and Techonlogy, 2017, 37(1): 41-48(in Chinese).

[4] HEWITSON S, WANG J L. GNSS receiver autonomous integrity monitoring (RAIM) performance analysis[J]. GPS Solutions, 2006, 10(3): 155-170.

[5] OBER P B, HARRIMAN D. On the use of multi-constellation-RAIM for aircraft approaches[C]// Proceedings of the 19th International Technical Meeting of the Satellite Division of the Institute of Navigation. Washington, D. C.: ION, 2006: 2587-2596.

[6] 蒋虎，袁运斌，王海涛，等. 精密进近阶段的多系统 GNSS 组合 RAIM 可用性算法及分析[J]. 中国空间科学技术，2016，36(3)：32-40.
JIANG H, YUAN Y B, WANG H T, et al. Multi-GNSS RAIM availability algorithms and analysis for precise approach[J]. Chinese Space Science and Technology, 2016, 36(3): 32-40(in Chinese).

[7] ENE A. Further development of Galileo-GPS RAIM for vertical guidance[C]// Proceedings of the 19th International Technical Mceting of the Satellite Division of the Institute of Navigation. Washington, D. C.: ION, 2006: 2597-2607.

[8] ENE A, BLANCH J, WALTER T. Galileo-GPS RAIM for vertical guidance[C]// Proceedings of the 2006 National Technical Meeting of the Institute of Navigation. Washington, D. C.: ION, 2006: 432-440.

[9] MARTINEAU A, MACABIAU C, MABILLEAU M. GNSS RAIM assumptions for vertically guided approaches[C]// Proceedings of the 22nd International Technical Meeting of the Satellite Division of the Institute of Navigation. Washington, D. C.: ION, 2009: 2791-2803.

[10] 徐君毅，杨元喜，李金龙，等. COMPASS 与其他 GNSS 组合导航完好性分析[J]. 中国科学：地球科学，2013，43(10)：1632-1642.
XU J Y, YANG Y X, LI J L, et al. Integrity analysis of COMPASS and other GNSS combined navigation[J]. Science China: Earth Sciences, 2013, 43(10): 1632-1642(in Chinese).

[11] LEE Y C, BRAFF R, FERNOW J P, et al. GPS and Galileo with RAIM or WAAS for vertically guided approaches[C]// Proceedings of the 18th International Technical Meeting of the Satellite Division of the Institute of Navigation. Washington, D. C.: ION, 2005: 1801-1825.

[12] 郭靖，崔晓伟，陆明泉，等. 支持垂直引导进近的多星座 RAIM 算法[J]. 清华大学学报(自然科学版)，2011，52(2)：157-160.
GUO J, CUI X W, LU M Q, et al. Multi-constellation RAIM for approach with vertical guidance[J]. Journal of Tsinghua University(Science and Technology), 2011, 52(2): 157-160(in Chinese).

[13] 王式太. 基于 M 估计的 GNSS 接收机自主完好性监测研究[D]. 北京：中国地质大学，2017：157-160.
WANG S T. Multi-constellation RAIM for approach with vertical guidance[D]. Beijing: China University of Geosciences, 2017: 157-160(in Chinese).

[14] LEE Y C. Investigation of extending receiver autonomous integrity monitoring (RAIM) to combined use of Galileo and modernized GPS[C]// Proceedings of the 17th International Technical Meeting of the Satellite Division of the Institute of Navigation. Washington, D. C.: ION, 2004: 1691-1698.

[15] HWANG P Y, BROWN R G. RAIM-FDE revisited: A new breakthrough in availability performance with NIORAIM(novel integrity-optimized RAIM)[J]. Navigation, 2006, 53(1): 41-51.

[16] KAPLAN E, HEGARTY C. Understanding GPS: Principles and application[M]. 2nd ed. Boston: Artech House, 2006: 39-45.

作者简介：

王尔申　男，博士，教授。主要研究方向：卫星导航、卫星导航信号性能评估。

北航学报 赠阅

BDS/GPS combined navigation receiver autonomous integrity monitoring algorithm

WANG Ershen[1,*], YANG Fuxia[1], PANG Tao[1], QU Pingping[1], LAN Xiaoyu[1], JIANG Yi[2]

(1. School of Electronic and Information Engineering, Shenyang Aerospace University, Shenyang 110136, China;
2. Information Science and Technology College, Dalian Maritime University, Dalian 116026, China)

Abstract: In order to make receiver autonomous integrity monitoring (RAIM) technique be applied to approach with vertical guidance (APV) flight phase in civil aviation navigation, the BDS/GPS combined navigation RAIM algorithm is studied. An algorithm based on the optimal weighted average solution of BDS/GPS position solution is proposed. Combined with the relationship between the optimal weighted average solution and BDS/GPS position solution, a test statistic is established. The test threshold is obtained according to the maximum allowable false alarm probability, and then the satellite navigation system fault is detected. Through the weighted least squares residual method, the satellite fault is detected and identified. The results are instructive for the study of multi-constellation combined satellite navigation system applied to the navigation in civil aviation APV flight phase.

Keywords: receiver autonomous integrity monitoring (RAIM); BeiDou navigation satellite system (BDS); combined navigation; approach with vertical guidance (APV); weighted average

北京航空航天大学学报

Received: 2017-05-03; **Accepted**: 2017-05-19; **Published online**: 2017-09-26 14:37

URL: kns.cnki.net/kcms/detail/11.2625.V.20170926.1437.001.html

Foundation items: National Natural Science Foundation of China (61571309, 61101161); the Fundamental Research Funds for the Central Universities (3132016317); Liaoning BaiQianWan Talents Program; Program for Liaoning Excellent Talents in University (LR2016069); Scientific Study Project for Liaoning Province Ministry of Education (L201716); Aeronautical Science Foundation of China (2015ZC54010)

* **Corresponding author**. E-mail: wes2016@126.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2017.0227

放飞过程中平流层飞艇运动与受力分析

张泰华[1,2], 姜鲁华[1,3,*], 周江华[1]

(1. 中国科学院光电研究院, 北京 100094; 2. 中国科学院大学, 北京 100049; 3. 中国科学院高能物理研究所, 北京 100049)

摘 要: 平流层飞艇放飞方式是其安全起飞的先决条件。本文对平流层飞艇放飞过程动力学响应建立了力学模型,提出了解析求解方法,开展了定量分析研究。依据影响平流层飞艇放飞过程的关键因素,对单氦气囊结构以及多氦气囊结构的平流层飞艇放飞过程进行了动力学分析,将单氦气囊结构飞艇动力学响应的定量分析结果与飞行试验过程中获得的数据进行对比,验证了分析方法的准确性,为进一步优化放飞过程的操作提供了依据。

关 键 词: 平流层飞艇; 放飞过程; 动力学分析; 单氦气囊结构; 多氦气囊结构

中图分类号: V274

文献标识码: A **文章编号**: 1001-5965(2018)04-0691-09

由于平流层在大气环境、气象条件和高度等多方面的综合特点,在平流层部署飞艇进行长期环境监测、通信中继和对地观测成为研究热点[1]。平流层飞艇的工作高度比卫星低,电波传播时延小,自由空间损耗小,造价相对较低。同时,平流层飞艇还有较强的抗毁生存能力。因此,平流层飞艇效费比将高于其他亚轨道飞行器和低轨卫星,在区域监控和天-地、天-空信息交互方面优势明显,无论是在军用,还是科学及民用领域,都具有极其广阔的应用前景[2-3]。在国际上,美国和日本等国进行了一些探索性试验,美国 HiSentinel 系列[4-5]和 HALE-D[6-7]飞艇曾进行了一些关键技术探索和飞行试验,日本 GTS 系列[8]和 SPF[9]不带动力飞艇也曾在升空试验中成功到达了平流层高度,这些工作为平流层飞艇技术的发展积累了一定的经验。

从解除地面设备对飞艇的约束到飞艇离开地面的过程即为飞艇的放飞过程。将体积庞大的平流层飞艇从艇库转运至放飞场,并尽快实施放飞决定了整个飞行试验是否能顺利实施,一些平流层飞艇在放飞阶段即被迫推迟或取消飞行试验[6-7,9]。平流层飞艇放飞后,由静止开始加速上升,艇体内氦气窜动引起浮心和重心变化导致飞艇运动复杂,庞大的体积和巨大的表面积导致了气动与附加质量的叠加,影响了飞艇的动力响应。位于柔性艇体上设备在放飞过程中的过载及其对柔性艇体的应力集中进一步增加了放飞过程的不确定因素和风险。

随着飞行试验的开展,国内外对平流层飞艇放飞过程也进行了大量研究。Selby[10]对平流层飞艇在不同工况下放飞及放飞后的特性做了较为深入的研究,分别分析了在顺风和逆风情况下实施放飞时所受的气动力及放飞风险;赵攀峰等[11]以常规布局形式的平流层飞艇为例,分析了飞艇放飞、回收的一般步骤,对上升、下降过程做了数值仿真分析;郭虓等[12]分析了飞艇基本热力学行为,研究了艇体及内部气体的能量方程并建立了详细的飞艇动力学和运动学模型,通过对不同场景的问题进行最优化轨迹求解,评估了飞艇在上升过程中,太阳能辐射和风场对热交换的影响,进

收稿日期: 2017-04-13; **录用日期**: 2017-05-22; **网络出版时间**: 2017-07-05 09:33
网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20170705.0933.001.html
基金项目: 国家"863"计划(2013AA122904); 国家科技重大专项
* **通信作者**. E-mail: jianglh@ihep.ac.cn

引用格式: *张泰华, 姜鲁华, 周江华. 放飞过程中平流层飞艇运动与受力分析[J]. 北京航空航天大学学报, 2018, 44(4): 691-699. ZHANG T H, JIANG L H, ZHOU J H. Kinematic and mechanical analysis on launch process of stratospheric airship[J]. Journal of Beijing University of Aeronautics and Astronautics, 2018, 44(4): 691-699 (in Chinese).*

北航学报 赠阅

而将不同时间点放飞对飞艇的影响进行了评估；吴雷等[13]分析了飞艇在放飞过程中因其所处环境的巨大变化导致飞艇内部气体质量和分布情况的变化情况，给出了能基本反映这一变化过程的飞艇放飞段动力学模型。总的来说，这些研究主要对飞艇放飞时机的选择及模型的求解进行了定性的分析，还没有开展较为系统的研究。

本文分析了放飞过程飞艇动力学响应特征，计算飞艇完成放飞所需的时间，以及在此期间飞艇的仰角、角速度和地面支撑力等变化。将计算结果与飞行试验比较，为平流层飞艇放飞过程提供了解析求解方法，也为相关操作及地面辅助措施的选择提供参考。本文在计算过程中忽略地面风对放飞过程的影响，主要分析放飞过程中的艇体气动阻力。

1 平流层飞艇囊体结构形式

平流层飞艇通常选择在 20 km 左右的平流层底部飞行或驻留，平流层大气稀薄，要依靠浮力保持稳定驻空，需要飞艇具有较大的体积，一般在数万至数十万立方米之间，飞艇的长度在百米量级[14]。平流层飞艇在地面时，由于大气密度较高，只需在飞艇内充入较少的氦气就能提供满足所需的升空浮力，艇囊 90% 以上的空间中充满空气以使飞艇保持设计外形。一般而言，平流层飞艇的艇囊可分为单氦气囊和多氦气囊 2 种形式，如图 1 所示。

单氦气囊结构中，艇囊中的氦气都在一个氦气囊中，与空气囊之间只有一个副气囊隔层，在飞艇纵向一般对气体无约束。日本于 2003 年 8 月成功实施了飞行试验的 SPF-1 飞艇即采用了单氦气囊结构[9]，飞艇前点解除约束后，飞艇加速抬头升空，放飞照片如图 2 所示。

多氦气囊结构中，艇囊中的氦气和(或)空气被分割成了 2 个以上的单独空间，以限制其在飞艇纵向的窜动，防止飞艇姿态变化超出可控范围。为了便于在飞艇上升过程中排气顺畅和保持压差稳定，也可将氦气囊或空气囊进行局部联通。

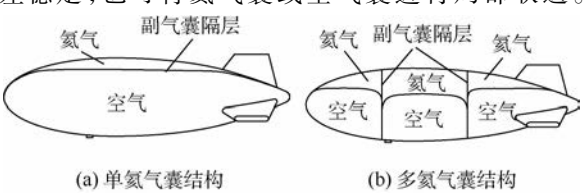


图 1 平流层飞艇氦气囊结构示意图

Fig. 1 Schematic of helium-envelope structure of stratospheric airship

美国于 2011 年 7 月进行了“HALE-D”平流层飞艇飞行试验[6-7]（见图 3），该飞艇解除两侧约束后，飞艇姿态平稳，推测其囊体结构采用了多氦气囊结构形式。

本文对平流层飞艇采用 2 种氦气囊结构形式的放飞过程进行分析，并将分析得到的单氦气囊结构的定量结果与飞行试验过程中获得的数据进行对比，验证分析计算方法的正确性，也为进一步优化放飞过程的操作提供依据。



图 2 SPF-1 飞艇放飞(日本)

Fig. 2 Launch of SPF-1 airship (Japan)



图 3 HALE-D 飞艇放飞(美国)

Fig. 3 Launch of HALE-D airship (USA)

2 单氦气囊结构放飞过程分析

根据单氦气囊结构的特点和安全可靠放飞的要求，一种典型的平流层飞艇放飞过程如图 4 所示。放飞前，飞艇接近于水平放置，艇体两侧各用 2 根绳索拉住，并使艇体后部坐落在支撑垫上，艇体处于小仰角状态；放飞时，解除艇体两端约束，在浮力作用下，艇体快速抬头，仰角迅速增大，当艇体前部升到一定高度时，后部脱离支撑垫，飞

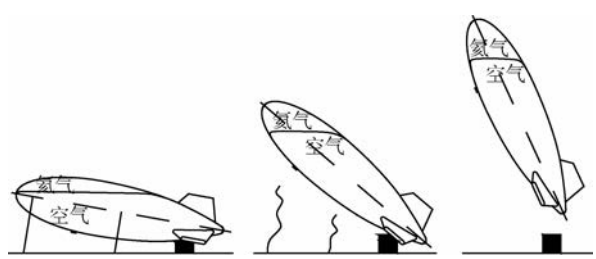


图 4 一种单氦气囊结构飞艇的放飞过程

Fig. 4 Launch process of airship with single helium envelope structure

艇以大仰角状态起飞离地,放飞过程完成。

2.1　飞艇受力

飞艇依靠内部充气达到一定的正压,以保证其设计外形并使其具有一定的刚度。据此,假设飞艇为刚体,建立如图 5 所示坐标系。OXY 坐标系为地面坐标系,$O'xy$ 为艇基坐标系,以飞艇体心为坐标原点,$O'x$ 轴为飞艇纵轴,$O'z$ 轴垂直 $O'xy$平面,坐标轴符合右手法则。

图 5 中:$\boldsymbol{T}_1$ 和 $\boldsymbol{T}_2$ 分别表示地面设备对飞艇约束力;$\boldsymbol{F}_{\mathrm{b}}$ 为艇囊内氦气总浮力;N_x、N_y 为飞艇所受的接触力;$\boldsymbol{G}$ 为艇体结构重力。

为使问题具有普遍性,假设飞艇在放飞前初始仰角为 φ_0。从后往前依次解除两端拉绳 $\boldsymbol{T}_2$、$\boldsymbol{T}_1$ 的约束,飞艇前部逐渐升起。在此过程中的某一时刻,飞艇受力和角度如图 6 所示。

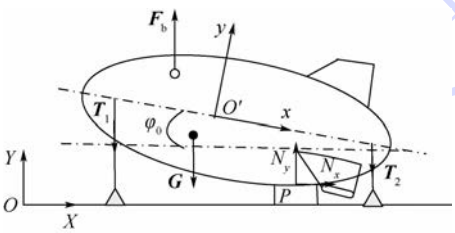


图 5　飞艇放飞前受力示意图

Fig. 5　Force diagram before airship launch

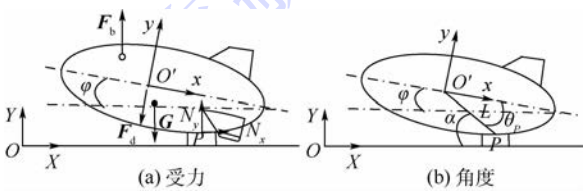


图 6　飞艇放飞过程中受力与角度示意图

Fig. 6　Schematic of force status and angle during airship launch process

地面坐标系与艇体坐标系的关系为

$$\begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \tag{1}$$

2.2　动力学方程

飞艇放飞前系统静止,力系平衡,有

$$\boldsymbol{F}_{\mathrm{b}} + \boldsymbol{G} + \sum_{i=1}^{2} \boldsymbol{T}_i + \boldsymbol{N} + \boldsymbol{F}_{\mathrm{d}} = \boldsymbol{0} \tag{2}$$

式中:$\boldsymbol{F}_{\mathrm{b}}$、$\boldsymbol{F}_{\mathrm{d}}$ 和 $\boldsymbol{N}$ 分别为浮力、气动阻力和腹部支撑力。放飞后,在体心坐标系上建立方程,可得矢量形式的动力学方程[15]为

$$\begin{bmatrix} m\boldsymbol{E} + \boldsymbol{m}' & -m[\boldsymbol{r}_G\times] \\ m[\boldsymbol{r}_G\times] & \boldsymbol{I}_o + \boldsymbol{I}_o' \end{bmatrix} \begin{bmatrix} \dot{\boldsymbol{V}}_o \\ \dot{\boldsymbol{\omega}} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\omega}\times[(m\boldsymbol{E}+\boldsymbol{m}')\boldsymbol{V}_o] + m\boldsymbol{\omega}\times(\boldsymbol{\omega}\times\boldsymbol{r}_G) \\ m\boldsymbol{r}_G\times(\boldsymbol{\omega}\times\boldsymbol{V}_o) + [\boldsymbol{\omega}\times](\boldsymbol{I}_o+\boldsymbol{I}_o')\boldsymbol{\omega} \end{bmatrix} = \begin{bmatrix} \boldsymbol{F}_{\mathrm{s}} \\ \boldsymbol{M}_{\mathrm{s}} \end{bmatrix} \tag{3}$$

式中:$\boldsymbol{F}_{\mathrm{s}}$、$\boldsymbol{M}_{\mathrm{s}}$ 分别为浮力、重力、气动阻力、腹部支撑力的合力与合力矩;m 为飞艇闭系统质量;$\boldsymbol{E}$ 为单位矩阵;$\boldsymbol{I}_o$ 为飞艇闭系统相对体心的转动惯量;$\boldsymbol{m}'$和 $\boldsymbol{I}_o'$ 分别为飞艇附加质量和附加惯量矩阵;$\boldsymbol{V}_o$ 和 $\boldsymbol{\omega}$ 分别为飞艇(相对地面)的速度和转动角速度;$\boldsymbol{r}_G$ 为体心至质心的矢径。

展开后,可得到 6 自由度标量形式的动力学方程。本文着重考虑飞艇发放过程纵剖面内的运动,因此抽取其纵向 3 自由度标量形式的动力学方程如下:

$$\begin{bmatrix} m_x & 0 & -my_{\mathrm{a}} \\ 0 & m_y & mx_{\mathrm{a}} \\ -my_{\mathrm{a}} & mx_{\mathrm{a}} & J_z \end{bmatrix} \begin{bmatrix} \dot{V}_x \\ \dot{V}_y \\ \dot{\omega}_z \end{bmatrix} + \omega_z \begin{bmatrix} -(m_yV_y + m\omega_z x_{\mathrm{a}}) \\ m_xV_x - m\omega_z y_{\mathrm{a}} \\ m(V_x x_{\mathrm{a}} + V_y y_{\mathrm{a}}) \end{bmatrix} = \begin{bmatrix} N_x + F_x \\ N_y + F_y \\ M_N + M_z \end{bmatrix} \tag{4}$$

式中:M_N 为飞艇所受的接触力矩;F_x、F_y 和 M_z 分别为浮力、重力、气动阻力产生的合力分量和合力矩;$(x_{\mathrm{a}}, y_{\mathrm{a}})$为飞艇系统总质心。

式(4)变换后得

$$\begin{cases} N_x = m_x\dot{V}_x - my_{\mathrm{a}}\dot{\omega}_z - \omega_z(m_yV_y + m\omega_z x_{\mathrm{a}}) - F_x \\ N_y = m_y\dot{V}_y + mx_{\mathrm{a}}\dot{\omega}_z + \omega_z(m_xV_x - m\omega_z y_{\mathrm{a}}) - F_y \\ M_N = -my_{\mathrm{a}}\dot{V}_x + mx_{\mathrm{a}}\dot{V}_y + J\dot{\omega}_z + m\omega_z(x_{\mathrm{a}}V_x + y_{\mathrm{a}}V_y) - M_z \end{cases} \tag{5}$$

式中:J 为飞艇转动惯量。

$$\begin{cases} m_x = m + m_{\mathrm{Ax}} \\ m_y = m + m_{\mathrm{Ay}} \\ J_z = I_z + I_{\mathrm{Az}} \end{cases} \tag{6}$$

其中:m_{Ax}、m_{Ay}和 I_z、I_{Az}分别为附加质量和惯性矩。

运动学方程:

$$\dot{\varphi} = -\omega_z \tag{7}$$

式中:φ 为飞艇仰角。

$$\begin{bmatrix} \dot{X}_o \\ \dot{Y}_o \end{bmatrix} = \begin{bmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{bmatrix} \begin{bmatrix} V_x \\ V_y \end{bmatrix} \tag{8}$$

2.3　约束条件

在放飞过程中,假设飞艇与支撑点 P 之间无相对滑动,飞艇质心与支撑点 P 之间距离为 L。当飞艇仰角为 φ 时,图 6 中各角度间关系为

$$\alpha = \varphi + \theta_P \tag{9}$$

将飞艇姿态变化过程中质心近似为不变,则

北航学报 赠阅

θ_P 为常数,对式(9)求导,有

$$\dot{\alpha} = \dot{\varphi} = -\omega_z \tag{10}$$

由此可知,飞艇轴向角速度 $\dot{\alpha}$ 与质心绕支撑点 P 旋转角速度 $\dot{\varphi}$ 相同,为计算方便,推导过程将角速度统一用 ω_z 求解。根据支撑点 P,可得

$$\begin{cases} V_x = L\omega_z \sin\theta_P \\ V_y = -L\omega_z \cos\theta_P \end{cases} \tag{11}$$

对式(11)求导,可得

$$\begin{cases} \dot{V}_x = L\dot{\omega}_z \sin\theta_P \\ \dot{V}_y = -L\dot{\omega}_z \cos\theta_P \end{cases} \tag{12}$$

将式(11)、式(12)代入式(5),得

$$\begin{cases} N_x = (m_x L\sin\theta_P - m y_a)\dot{\omega}_z + \\ \quad (m_y L\cos\theta_P - m x_a)\omega_z^2 - F_x \\ N_y = (-m_y L\cos\theta_P + m x_a)\dot{\omega}_z + \\ \quad (m_x L\sin\theta_P - m y_a)\omega_z^2 - F_y \\ M_N = (-m y_a L\sin\theta_P - m x_a L\cos\theta_P + J_z)\dot{\omega}_z + \\ \quad mL(x_a\sin\theta_P - y_a\cos\theta_P)\omega_z^2 - M_z \end{cases} \tag{13}$$

2.4 力和力矩

根据飞艇受力,将飞艇受到的气动阻力分解为沿 x 和 y 2 个方向,根据飞艇外形仿真得到这 2 个方向气动阻力系数[16],有

$$M_N = L(N_y\cos\theta_P - N_x\sin\theta_P) \tag{14}$$

$$\begin{cases} F_x = -(F_b - G)\sin\varphi + F_{dx} = \\ \quad -F_f\sin\varphi + F_{dx} \\ F_y = (F_b - G)\cos\varphi - F_{dy} = F_f\cos\varphi - F_{dy} \\ M_z = F_b(x_b\cos\varphi + y_b\sin\varphi) - \\ \quad mg(x_a\cos\varphi + y_a\sin\varphi) \end{cases} \tag{15}$$

式中:(x_b, y_b) 为飞艇浮心;F_f 为净浮力;F_{dx} 和 F_{dy} 为飞艇气动阻力。

根据飞艇气动阻力公式[17],将式(11)代入得

$$\begin{cases} F_{dx} = 0.5 C_{dx}\rho\nabla^{\frac{2}{3}} L^2\omega_z^2\sin^2\theta_P \\ F_{dy} = 0.5 C_{dy}\rho\nabla^{\frac{2}{3}} L^2\omega_z^2\cos^2\theta_P \end{cases} \tag{16}$$

式中:ρ 为空气密度;∇ 为飞艇体积;C_{dx} 和 C_{dy} 分别为飞艇在 x 和 y 方向的阻力系数。

由式(13)~式(16)得到解除约束后的动力学微分方程

$$f_1\dot{\omega}_z + f_2\omega_z^2 + f_3 = 0 \tag{17}$$

式中:

$$\begin{aligned} f_1 &= -L^2(m_y\cos^2\theta_P + m_x\sin^2\theta_P) + \\ &\quad 2mL(x_a\cos\theta_P + y_a\sin\theta_P) - J_z \end{aligned} \tag{18}$$

$$\begin{aligned} f_2 &= L^2[(m_x - m_y)\cos\theta_p\sin\theta_p + \\ &\quad 0.5\rho\nabla^{\frac{2}{3}} L(C_{dy}\cos^3\theta_P + C_{dx}\sin^3\theta_P)] \end{aligned} \tag{19}$$

$$\begin{aligned} f_3 &= F_b(x_b\cos\varphi + y_b\sin\varphi) - F_f L(\cos\varphi\cos\theta_P + \\ &\quad \sin\varphi\sin\theta_P) - mg(x_a\cos\varphi + y_a\sin\varphi) \end{aligned} \tag{20}$$

式(17)与式(7)、式(8)一起构成了方程的解。支撑点 P 所受的约束力,可由式(13)和式(15)得到,转换到地面坐标系为

$$N_X = N_x\cos\varphi + N_y\sin\varphi \tag{21}$$

$$N_Y = -N_x\sin\varphi + N_y\cos\varphi \tag{22}$$

2.5 辅助方程

飞艇系统总质量为

$$\begin{aligned} m &= m_s + m_{He} + m_{air} = \\ &\quad m_s + m_{He} + \rho(\nabla - \nabla_{He}) = \\ &\quad m_s + \rho\nabla + m_{He}(1 - c) \end{aligned} \tag{23}$$

式中:$c = \rho/\rho_{He}$,ρ_{He} 为氦气密度;m_s、m_{He} 和 m_{air} 分别为艇体结构、氦气和空气的质量。系统总质心为

$$m\begin{bmatrix} x_a \\ y_a \end{bmatrix} = m_s\begin{bmatrix} x_s \\ y_s \end{bmatrix} + m_{He}(1 - c)\begin{bmatrix} x_b \\ y_b \end{bmatrix} \tag{24}$$

式中:(x_s, y_s) 为艇体结构质心。总惯量为

$$I_z = I_s + I_{He} + I_{air} = I_s + \rho\left[\left(\frac{1}{c} - 1\right)I_{zb} + I_{zt}\right] \tag{25}$$

式中:I_{He}、I_{air} 和 I_s 分别为氦气、空气和艇体的结构惯量;I_{zt} 为艇体几何惯量;I_{zb} 为氦气部分几何惯量。附加质量[18]和惯性矩为

$$\begin{cases} m_{Ax} = k_1\rho\nabla \\ m_{Ay} = k_2\rho\nabla \\ I_{Az} = k_3\rho I_{zt} \end{cases} \tag{26}$$

式中:k_1、k_2 和 k_3 为附加质量系数。

2.6 工况分析

1)飞艇放飞瞬间

在飞艇解除约束瞬间,由于 $\omega_z = 0$,由式(17)可求得此时飞艇初始角加速度:

$$\dot{\omega}_z = -\frac{f_3}{f_1} \tag{27}$$

根据艇上设备位置,进而可得艇上设备放飞时受到的初始过载。

2)飞艇离地

因支撑点 P 无法提供拉力,由式(22)可得,当 $N_Y < 0$ 时,飞艇离开支撑点 P,完成放飞。

3 多氦气囊结构放飞过程分析

多氦气囊结构中,艇囊内设置多个氦气囊,极大地限制了因飞艇姿态变化而导致的氦气窜动,也严格控制飞艇浮心变化。在飞艇放飞升空过程中,飞艇姿态一直处于可控状态。实施放飞时,飞艇一般小仰角或水平状态离地,其飞行过程如

北航学报 赠阅

图 7所示。

在实施这种结构飞艇的放飞过程中,飞艇浮心与重心保持不变,飞艇的运动形式将大大简化,解除约束后,飞艇加速上升。飞艇在放飞前后的约束及受力状态如图 8 所示。

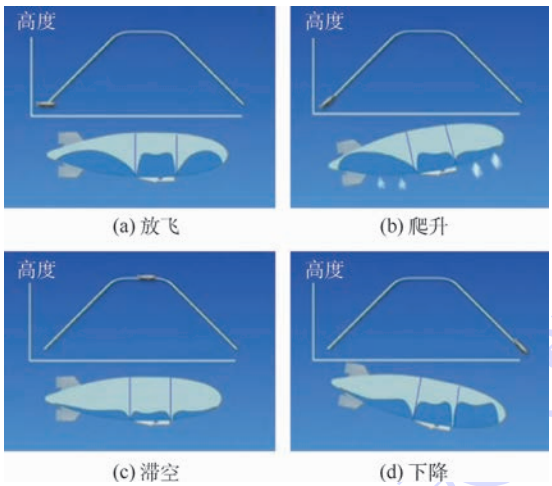


图 7　多氦气囊结构飞艇的飞行过程

Fig. 7　Flight process of airship with multi-helium envelope structure

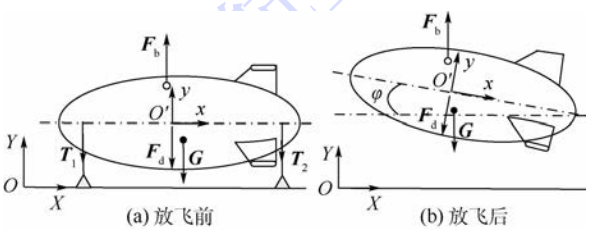


图 8　多氦气囊结构飞艇放飞前后受力

Fig. 8　Force status for multi-helium envelope structure before and after airship launch

解除约束后,在不考虑水平来流的情况下,飞艇将垂直上升,即

$$V_x = \dot{V}_x = 0 \tag{28}$$

根据飞艇动力学方程,由式(4)变换后得

$$\begin{cases} \dot{V}_y = \dfrac{F_y}{m_y} - \dfrac{m}{m_y} x_a \dot{\omega}_z + \dfrac{m}{m_y} y_a \omega_z^2 \\ M_z = m x_a \dot{V}_y + J \dot{\omega}_z + m V_y y_a \omega_z \end{cases} \tag{29}$$

将式(29)中第 1 式代入第 2 式,整理得

$$\left(-\frac{m^2}{m_y} x_a^2 + J \right) \dot{\omega}_z + \frac{m^2}{m_y} x_a y_a \omega_z^2 + m V_y y_a \omega_z = M_z - m x_a \frac{F_y}{m_y} \tag{30}$$

在上升过程中的气动阻力,由式(16)有

$$F_{dy} = 0.5 C_{dy} \rho \nabla^{\frac{2}{3}} V_y^2 \tag{31}$$

代入式(15),得

$$\begin{cases} F_y = F_f \cos\varphi - 0.5 C_{dy} \rho \nabla^{\frac{2}{3}} V_y^2 \\ M_z = F_b (x_b \cos\varphi + y_b \sin\varphi) - mg(x_a \cos\varphi + y_a \sin\varphi) \end{cases} \tag{32}$$

式中:g 为重力加速度。

将式(32)代入式(30),得

$$\left(-\frac{m^2}{m_y} x_a^2 + J \right) \dot{\omega}_z + \frac{m^2}{m_y} x_a y_a \omega_z^2 + m V_y y_a \omega_z = (x_b \cos\varphi + y_b \sin\varphi) F_b - (x_a \cos\varphi + y_a \sin\varphi) mg - \frac{m}{m_y} x_a (F_f \cos\varphi - 0.5 C_{dy} \rho \nabla^{\frac{2}{3}} V_y^2) \tag{33}$$

式(33)与式(7)、式(8)一起构成了方程的解。

对于稳定上升中的飞艇而言,有

$$\omega_z = \dot{\omega}_z = F_y = M_z = 0 \tag{34}$$

代入式(32),得

$$\begin{cases} F_f \cos\varphi - 0.5 C_{dy} \rho \nabla^{\frac{2}{3}} V_y^2 = 0 \\ F_b (x_b \cos\varphi + y_b \sin\varphi) - mg(x_a \cos\varphi + y_a \sin\varphi) = 0 \end{cases} \tag{35}$$

由式(35)即可求得稳定上升时的升速,也可根据设计升速,求出飞艇所需净浮力。

4　算　例

4.1　单氦气囊结构

算例以中国科学院光电研究院设计的某一平流层试验飞艇为对象,该飞艇采用单氦气囊结构,成功完成了放飞和飞行试验。该飞艇的一些主要参数包括:$\nabla = 1.06 \times 10^4\ \mathrm{m}^3$,$m_s = 1\,090\ \mathrm{kg}$,$m_{He} = 192\ \mathrm{kg}$,$m = 1.295 \times 10^4\ \mathrm{kg}$,$(x_s, y_s) = (0.31, -2.8)$,$I_s = 5.277 \times 10^5\ \mathrm{kg \cdot m^2}$,$I_{zt} = 2.664 \times 10^6\ \mathrm{kg \cdot m^2}$,$c = \rho/\rho_{He} = 7.236$,$L = 29.4\ \mathrm{m}$,$\theta_P = 8.2°$,$\varphi_0 = 9°$,阻力系数 $C_{dx} = 0.07$ 和 $C_{dy} = 1.46$均由数值仿真得到[19],根据旋转椭球体估算得到附加质量系数$k_1 = 0.081\,173$、$k_2 = 0.860\,33$ 和 $k_3 = 0.609\,43$。(x_b, y_b) 和 I_{zb}随飞艇仰角变化可通过插值得到,如图 9 所示。

根据上述分析,得到支撑点 P 受到的作用力 (N_X, N_Y)随时间变化曲线如图 10 所示。

飞艇在放飞过程中,质心水平和垂直速度的变化如图 11 所示,飞艇仰角随时间变化的计算结果和实测数据对比如图 12 所示,角速度随时间变化的计算结果和实测数据的对比如图 13 所示。从图 12 和图 13 的对比曲线可以看出,飞艇实际放飞中解除约束的响应比计算分析结果明显迟滞,这主要是因为在分析中将飞艇是近似为刚体,

北航学报 赠阅

而实测角速度值较为发散是飞艇上的角度传感器的测量误差所致,但在变化趋势与计算结果一致,能够满足工程应用需求。

通过上述分析计算,得到飞艇整个放飞过程中的姿态变化如图 14 所示。

计算表明,飞艇放飞后 17.9 s,$N_Y<0$,此后飞艇离开支撑点,也意味着飞艇完成放飞离开地面,此时飞艇仰角为 43.8°。在飞艇放飞过程中,支撑点水平作用力最大$N_{X\max}=6.183$ kN,支撑点

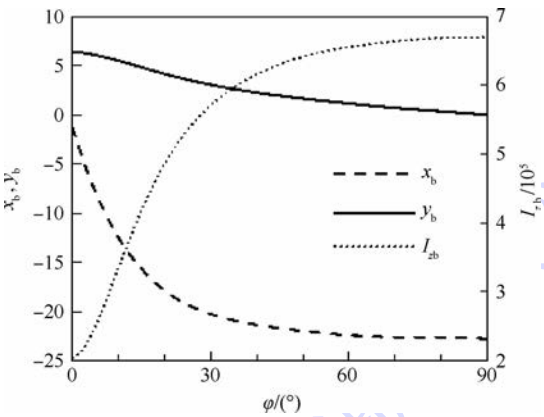


图 9 参数 x_b,y_b 和 I_{zb} 随飞艇仰角变化示意图

Fig. 9 Fluctuation of parameter x_b, y_b and I_{zb} with airship elevation angle

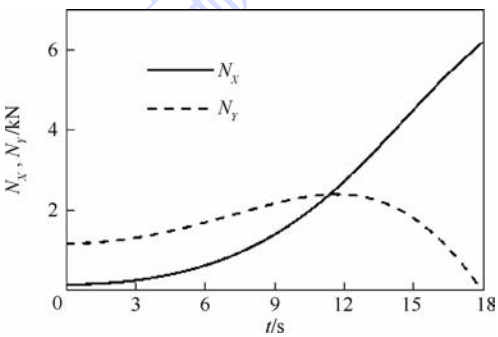


图 10 飞艇放飞过程 N_X,N_Y 变化

Fig. 10 Fluctuation of N_X, N_Y during airship launch process

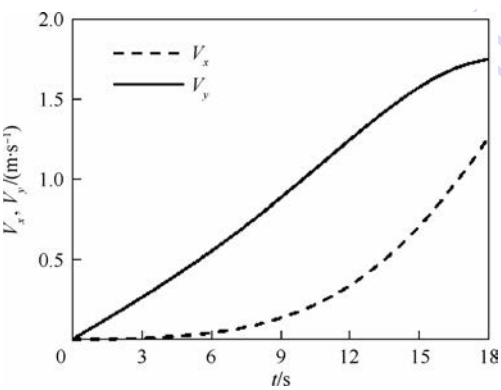


图 11 飞艇放飞过程中质心速度变化

Fig. 11 Fluctuation of center-of-mass velocity during airship launch process

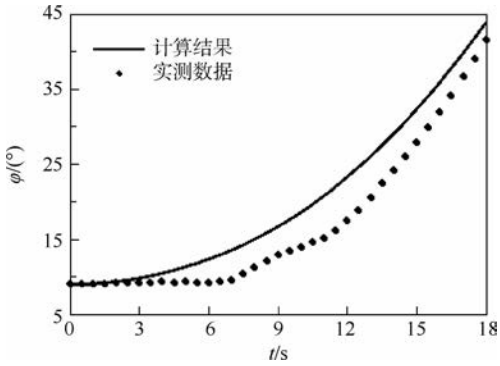


图 12 飞艇仰角计算结果与实测数据对比

Fig. 12 Comparison of elevation angle of airship between calculation results and flight test data

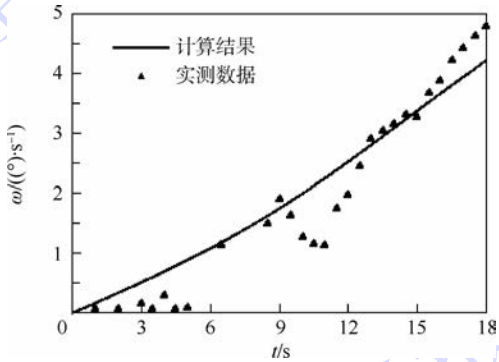


图 13 飞艇角速度计算结果与实测数据对比

Fig. 13 Comparison of angular velocity of airship between calculation results and flight test data

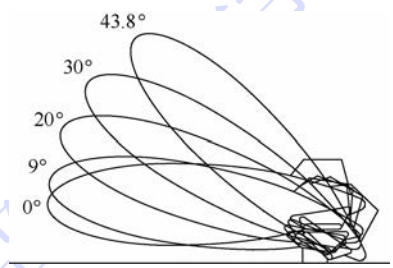


图 14 飞艇放飞过程中的姿态变化

Fig. 14 Change of airship attitude during launch process

垂直作用力最大 $N_{Y\max}=2.405$ kN。

4.2 多氦气囊结构

如果算例所示飞艇采用多氦气囊结构形式,解除约束后飞艇重心和浮心都保持不变,利用第 3 节分析,可以得到飞艇解除约束后的仰角和升速变化如图 15 所示。

由图 15 可知,如果算例中的飞艇采用多氦气囊结构,放飞后稳定升速为 1.56 m/s。在放飞过程中,解除约束后,飞艇升速在 3 ~ 5 min 内趋于稳定,仰角在 21° ~ 5°之间进行宽幅波动后趋于稳定。增大浮重比对飞艇仰角影响较小,但能显著增大稳定升速。如果平流层飞艇要快速升空,并安全穿过对流层顶疾风区,需增大浮力,以增大

飞艇稳定升速。为此,以浮重比 η 作为输入参数,其定义为

$$\eta = \frac{F_b}{m_s g} \tag{36}$$

不同的浮重比,飞艇稳定升空时的仰角和升速如表 1 所示。

由表 1 计算结果可知,飞艇浮重比增大,飞艇稳定升速增大,但对飞艇稳定仰角影响较小。选取浮重比 η 分别为 1.1、1.5 和 1.8 三种情况下,飞艇放飞后的升速变化如图 16 所示。

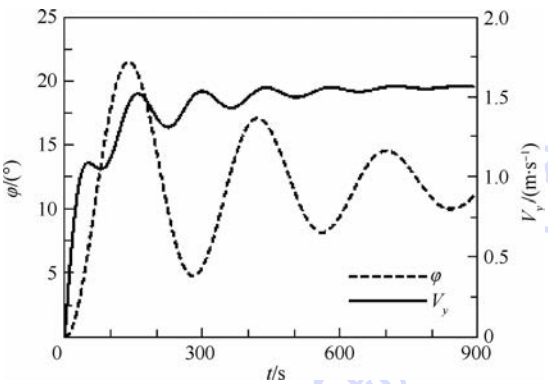


图 15　放飞后飞艇仰角与升速变化

Fig. 15　Fluctuation of pitch angle and rising velocity of airship after launch

表 1　不同浮重比飞艇稳定仰角和升速

Table 1　Stable pitch angle and rising velocity of airship for different buoyancy-weight ratios

浮 重 比	仰角/(°)	升速/(m·s⁻¹)
1.1	11.13	1.56
1.2	11.41	2.20
1.3	11.56	2.69
1.4	11.66	3.11
1.5	11.72	3.47
1.6	11.77	3.81
1.7	11.81	4.11
1.8	11.84	4.39
1.9	11.86	4.66
2.0	11.88	4.91

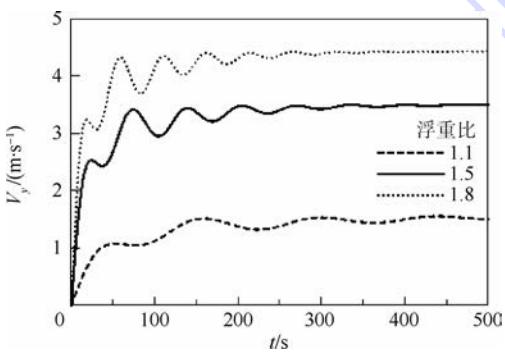


图 16　3 种典型浮重比下飞艇升速变化

Fig. 16　Fluctuation of rising velocity of airship for three typical buoyancy-weight ratios

4.3　两种结构形式比较

通过上述分析可知,氦气囊结构设计对飞艇放飞和升空影响很大,不同的氦气囊结构形式决定了飞艇放飞方式。在飞艇升空过程中,升速是一个非常重要的技术指标,以算例中的飞艇为例,不同的设计升速与所需的浮重比 η 如图 17 所示。

采用单氦气囊和多氦气囊结构形式,各有利弊,不同形式的氦气囊结构,决定了不同的放飞和升空过程。在飞艇到达设计高度,进行平飞阶段后,无论采用哪种氦气囊结构形式,其浮重比 η 都趋近于 1,即达到浮重平衡状态。在飞艇上升过程中,随着高度增加,不但要排出空气囊中的空气以确保艇囊压差在安全范围内,也要逐渐排出多余氦气,以确保飞艇最终的浮重平衡。两种氦气囊结构的对比如表 2 所示。

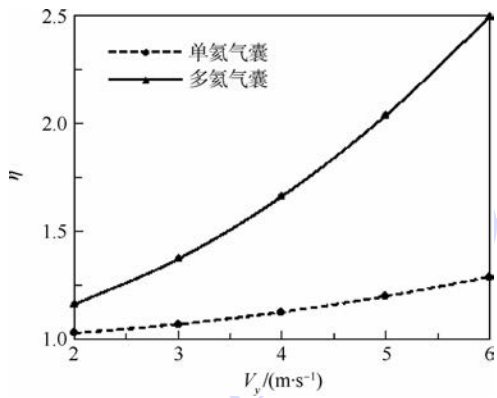


图 17　两种氦气囊结构升速与浮重比关系

Fig. 17　Comparison of rising velocity and buoyancy-weight ratio between two kinds of helium envelope structure

表 2　两种氦气囊结构对比

Table 2　Comparison between two kinds of helium envelope structure

比较项目	单氦气囊结构	多氦气囊结构
氦气分布	同一容腔	多个容腔
飞艇浮心变化	很大,不可控	很小,受控
升空姿态	大仰角	小仰角,可选择
放飞时飞艇姿态	加速抬头,尾部下顿	仰角不变
排气方式	尾部集中排气	分段排气
升速/(m·s⁻¹)	5 ~ 10	2 ~ 5
浮重比	小	大
副气囊结构	简单	复杂
加工工艺	简单	复杂
压控	简单	复杂
放飞形式	复杂	简单
典型案例	日本 SPF-1	美国 HALE-D

5　结　论

通过对平流层飞艇在放飞过程中的动力学响

北航学报 赠阅

应分析,结果表明:

1)在单氦气囊结构形式的放飞过程中,飞艇受到的支撑点水平作用力持续增大,支撑点垂直作用力先增大后减小,在放飞后段飞艇尾部必将发生滑动,若采用硬连接方式限制尾部滑动,应充分考虑柔性艇尾的承载能力,确保柔性艇尾的安全。

2)在多氦气囊结构形式的放飞过程中,放飞后升速从零开始剧烈波动后收敛较快,仰角变化剧烈,收敛较慢,飞艇以稳定升速升空后,仰角变化逐渐收敛。增大浮重比对飞艇仰角影响较小,但能显著增大稳定升速。

3)通过两种氦气囊结构形式对比,其各有优缺点:采用单氦气囊结构形式浮重比较小,可以大仰角快速升空,但仰角变化不可逆不可控,放飞方式复杂;采用多氦气囊结构形式升空所需净浮力较大,仰角实时可控可调,压控与排气复杂,放飞方式简单。

4)通过与飞行试验数据比较,本文方法能够满足工程应用需求,为平流层飞艇放飞方式的选择和评估提供了理论分析途径。根据平流层飞艇放飞场实际条件,可对后续的研究进一步细化或区分不同工况:地面风甚至非定常风对飞艇放飞的影响,飞艇行进中放飞动力学,以及将飞艇支撑点代以实际中的苫布或气垫等,这些分析都可基于本文的计算分析,进行部分方程或条件的增减而完成。

参考文献(References)

[1] LIAO L, PASTERNAK I. A review of airship structural research and development[J]. Progress in Aerospace Sciences, 2009, 45(4-5): 83-96.

[2] STOCKBRIDGE C, CERUTI A, MARZOCCA P. Airship research and development in the areas of design, structures, dynamics and energy systems[J]. International Journal of Aeronautical & Space Sciences, 2012, 13(2): 170-187.

[3] WILSON J R. A new era for airships[J]. Aerospace America, 2004, 42(5): 27-31.

[4] STEVE S. The HiSentinel airship[C]// 7th AIAA Aviation Technology, Integration and Operations Conference. Reston: AIAA, 2007.

[5] SMITH I, LEE M, FORTNEBERRY M, et al. HiSentinel80: Flight of a high altitude airship[C]// 11th AIAA Aviation Technology, Integration, and Operations Conference. Reston: AIAA, 2011.

[6] Lockheed Martin Space System Company High altitude airship [EB/OL]. (2014-11-10)[2017-04-13]. http://www.lockheedmartin.com/us/products/lighter-than-air-vehicles.

[7] GAO. Future aerostat and airship investment decisions drive oversight and coordination needs: GAO-13-81[R]. Washington, D. C.: GAO, 2012.

[8] JAXA. Ground-to-stratosphere flight test and evaluation of materials and structure for stratospheric airship test vehicle: JAXA-RM-04-012[R]. Tokyo: JAXA, 2004: 7-35.

[9] SHUNICHI O, NOBORU S. R&D status of RFC technology for SPF airship in Japan[C]// 9th Annual International Energy Conversion Engineering Conference. Reston: AIAA, 2011.

[10] SELBY C. High altitude airship station keeping and launch model development using output from numerical weather prediction models[D]. West Lafayette: Purdue University, 2008.

[11] 赵攀峰,王永林,刘传超. 平流层飞艇放飞、回收过程初步分析[J]. 航空科学技术, 2007(4): 24-29.
ZHAO P F, WANG Y L, LIU C C. Preparatory analyse on the release and recovery course of stratosphere airship[J]. Aeronautical Science and Technology, 2007(4): 24-29 (in Chinese).

[12] 郭旎,祝明,武哲. 综合热力学模型的平流层飞艇上升轨迹优化[J]. 北京航空航天大学学报, 2012, 38(10): 1346-1351.
GUO X, ZHU M, WU Z. Ascent trajectory optimization for stratospheric airships with thermal effects[J]. Journal of Beijing University of Aeronautics and Astronautics, 2012, 38(10): 1346-1351 (in Chinese).

[13] 吴雷,李勇,梁栋. 平流层飞艇放飞段动力学建模[C]// 第25届中国控制会议, 2006: 546-550.
WU L, LI Y, LIANG D. Dynamics modeling on the launch process for a stratospheric airship[C]// Proceeding of the 25th Chinese Control Conference, 2006: 546-550 (in Chinese).

[14] BLACKINGTON E. United States air force, schriever air force base: AIAA-2003-6005[R]. Reston: AIAA, 2003.

[15] 周江华. 平流层飞艇运动控制律与定点控制律设计[D]. 北京:中国科学院空间科学与应用中心, 2009: 22-26.
ZHOU J H. Control law design for motion control and station-keeping control of stratospheric platform airship[D]. Beijing: Center for Space Science and Applied Research, Chinese Academy of Sciences, 2009: 22-26 (in Chinese).

[16] CUI Y X, YANG Y C, ZHOU J H, et al. Numerical aerodynamic investigations on stratospheric airships of different tail configurations[C]// IEEE Aerospace Conference. Piscataway, NJ: IEEE Press, 2015: 3-4.

[17] KHOURY G, GILLELT J. Airship technology[M]. Cambridge: Cambridge University Press, 1999: 20-23.

[18] LI Y W, NATHON M. Modeling and simulation of airship dynamics[J]. Journal of Guidance, Control, and Dynamics, 2007, 30(6): 1691-1700.

[19] 张泰华,姜鲁华,张冬辉,等. 临近空间飞艇艇库外约束及稳定性分析[J]. 中国空间科学技术, 2016, 36(5): 72-80.
ZHANG T H, JIANG L H, ZHANG D H, et al. Constraint and stability analysis of near space airship outside hangar[J]. Chinese Space Science and Technology, 2016, 36(5): 72-80 (in Chinese).

北航学报 赠阅

作者简介:

张泰华　男,博士,高级工程师。主要研究方向:浮空器结构设计、发放技术。

姜鲁华　男,博士,研究员,博士生导师。主要研究方向:浮空器总体设计。

周江华　男,博士,研究员,博士生导师。主要研究方向:飞艇飞行控制。

Kinematic and mechanical analysis on launch process of stratospheric airship

ZHANG Taihua[1,2], JIANG Luhua[1,3,*], ZHOU Jianghua[1]

(1. Academy of Opto-Electronics, Chinese Academy of Sciences, Beijing 100094, China;
2. University of Chinese Academy of Sciences, Beijing 100049, China;
3. Institute of High Energy Physics, Chinese Academy of Sciences, Beijing 100049, China)

Abstract: Launch mode of the stratospheric airship is the prerequisite factor to safely reach the target altitude. In this paper, a dynamics model for the launch process of stratospheric airship is firstly established, and the analytical solution method is put forward, and the quantitative analysis is carried out. Secondly, according to the typical issues that influence the launch process, the dynamic analysis of both the single-helium and multi-helium envelope structure are carried out, including dynamics response and force status. Furthermore, the numerical results of a single helium envelope structure are obtained and then compared with experimental data. The result verified the accuracy of the analytical method, which can also provide the basis for launch process of the airship and the design of the launch equipment.

Keywords: stratospheric airship; launch process; dynamics analysis; single helium envelope structure; multi-helium envelope structure

北京航空航天大学学报

Received: 2017-04-13; **Accepted**: 2017-05-22; **Published online**: 2017-07-05 09:33

URL: kns.cnki.net/kcms/detail/11.2625.V.20170705.0933.001.html

Foundation items: National High-tech Research and Development Program of China (2013AA122904); National Science and Technology Major Project

* **Corresponding author**. E-mail: jianglh@ihep.ac.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2017.0228

基于 Rao-Blackwellized 蒙特卡罗数据关联的检测跟踪联合优化

陈唯实*, 闫军, 李敬
(中国民航科学技术研究院 机场研究所, 北京 100028)

摘 要: 提出了基于 Rao-Blackwellized 蒙特卡罗数据关联的雷达目标检测跟踪联合优化算法。Rao-Blackwellization 方法将单目标跟踪与数据关联分开处理,将序贯蒙特卡罗方法(粒子滤波)用于数据关联,实现杂波与虚警量测中的多目标跟踪。同时,根据粒子的分布范围确定波门大小。在考虑粒子权重的前提下,利用检测单元与所有粒子的相对位置对检测门限进行修正,提高检测率。将本文算法与已经实现的基于空域特性的杂波抑制算法相结合,分别应用于仿真数据、S 波段相参与非相参雷达实测数据。实验结果表明,本文算法能够在粒子数较少的情况下,实现对小弱目标的检测与跟踪。

关 键 词: 数据关联; 雷达; 目标; 检测; 跟踪
中图分类号: TN959.1
文献标识码: A **文章编号**: 1001-5965(2018)04-0700-09

"低慢小"目标的探测是一个世界性难题,困难主要源于以下 3 个方面[1-2]:①飞行高度"低",一般在 1 000 m 以下,地面雷达观测时会有大量杂波进入接收机,尤其在建筑密集的城市环境中,探测难度更大。②飞行速度"慢",一般小于 200 km/h,其回波信号处于杂波主瓣区,容易与鸟群等低速杂波混淆。为防止过多虚警,雷达往往设置速度阈值,导致低速的低空小目标被滤除。③雷达散射截面积(RCS)"小",目标信噪比低,雷达反射信号忽隐忽现,目标不易被发现与识别。随着无人机技术的日渐成熟,其作为一种典型的"低慢小"目标,严重威胁航班安全运行[3-4]。

雷达系统通常包括检测和跟踪 2 个部分。其中,检测部分将目标从噪声和杂波环境中提取出来,获得目标的速度、位置等状态信息;跟踪部分对目标下一刻的状态进行预测,进而形成稳定航迹。在常规雷达系统中,检测与跟踪部分通常分别考虑和设计[5-6]。实际上,二者是相互影响的。检测是跟踪的前提,良好的检测是保证跟踪性能的基础;同时,跟踪能进一步完善检测结果,利用跟踪获得的目标动态特征能够提高检测能力[7-8]。因此,本文提出一种基于 Rao-Blackwellized 蒙特卡罗数据关联的检测跟踪联合优化算法,对雷达目标检测与跟踪进行交互处理,将跟踪结果反馈给检测部分,提高雷达系统对低空小目标的检测能力,同时保证下一步的精确跟踪。

1 检测跟踪联合优化技术

雷达系统在完成目标检测之后,通常将波门内超过一定阈值的量测数据通过跟踪算法进行关联[9]。如将跟踪结果反馈至检测中心,进而根据该反馈信息调整波门内的检测阈值,改善检测结

收稿日期: 2017-04-13; **录用日期**: 2017-07-21; **网络出版时间**: 2017-09-14 09:03
网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20170914.0903.001.html
基金项目: 国家自然科学基金委员会-中国民航局民航联合研究基金(U1633122); 国家重点研发计划(2016YFC0800406)
* **通信作者**. E-mail: chenwsh@mail.castc.org.cn

引用格式: 陈唯实, 闫军, 李敬. 基于 Rao-Blackwellized 蒙特卡罗数据关联的检测跟踪联合优化[J]. 北京航空航天大学学报, 2018, 44(4): 700-708. CHEN W S, YAN J, LI J. Joint optimization of detection and tracking with Rao-Blackwellized Monte Carlo data association[J]. Journal of Beijing University of Aeronautics and Astronautics, 2018, 44(4): 700-708 (in Chinese).

北航学报 赠阅

果,此类算法即可称为检测跟踪联合优化算法。王云奇[10]结合目标、环境等先验知识的辅助,量化分析检测性能与跟踪性能的关系,实现检测和跟踪的耦合。刘红亮、严俊坤等[11-12]提出一种航迹恒虚警的目标检测跟踪一体化算法,根据帧虚警概率调整预测波门内的检测门限,完成目标检测与跟踪的联合优化过程。闫学昭[13]采用 DSP 等硬件编程方法,搭建了雷达目标检测与跟踪交互处理模块,将跟踪结果反馈至检测系统,在提高检测准确率的前提下实现了精确跟踪。

综上所述,这些算法的共同点在于基于一定规则降低波门内的检测阈值,以牺牲波门内局部虚警率为代价提升局部目标的检测率[11]。可见,波门的设置至关重要,如果波门面积(体积)过小,则可能将检测目标排除在外;如果波门面积(体积)过大,则可能引入过多虚警,导致计算机过载。针对此缺点,本文提出一种基于 Rao-Blackwellized 蒙特卡罗数据关联的检测跟踪联合优化算法。该算法采用序贯蒙特卡罗(Sequential Monte Carlo,SMC)(又称为粒子滤波)方法实现多目标跟踪,由于粒子能够用于表示多种不同的数据关联假设,基于 SMC 方法的多目标跟踪算法可视为多假设跟踪(Multiple Hypothesis Tracking,MHT)的推广[14]。此外,Rao-Blackwellizion 方法的应用能提高算法的准确率和效率,采用卡尔曼滤波或扩展卡尔曼滤波等方法估计目标状态,将 SMC 方法仅用于数据关联估计,使得联合后验分布由混合高斯表示,比纯粹的 SMC 方法具有更小的方差。同时,本文算法能够根据粒子的分布范围确定波门大小,在考虑粒子权重的前提下,利用检测单元与所有粒子的相对位置对检测门限进行修正。

2　基于 SMC 的检测跟踪联合优化算法

本文提出一种基于 Rao-Blackwellized 蒙特卡罗数据关联的检测跟踪联合优化算法,包括检测和跟踪 2 个部分,将粒子滤波跟踪获取的目标状态信息反馈至检测部分,以修正相关检测单元的检测门限,实现检测与跟踪的联合优化,算法流程如图 1 所示,详细步骤如下:

步骤 1　粒子滤波目标状态预估。

以一组 N 个粒子表示目标的当前状态 $\{w_{k-1}^{(i)},c_{k-1}^{(i)},\boldsymbol{m}_{k-1}^{(i)},\boldsymbol{P}_{k-1}^{(i)}:i=1,2,\cdots,N\}$,基于每个粒子 $i=1,2,\cdots,N$ 在 $k-1$ 时刻的预估值 $\boldsymbol{m}_{k-1}^{(i)}$ 和

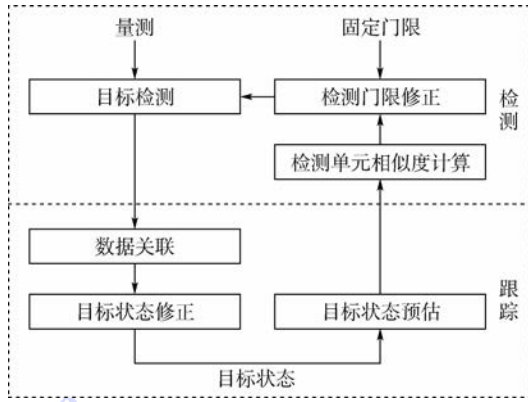


图 1　目标检测跟踪联合优化算法流程图

Fig. 1　Flowchart of joint optimization algorithm for target detection and tracking

协方差 $\boldsymbol{P}_{k-1}^{(i)}$ 对 k 时刻的目标状态进行卡尔曼预估:

$$\begin{cases}\boldsymbol{m}_{k|k-1}^{(i)}=\boldsymbol{A}_{k-1}(c_{k-1}^{(i)})\boldsymbol{m}_{k-1}^{(i)}\\ \boldsymbol{P}_{k|k-1}^{(i)}=\boldsymbol{A}_{k-1}(c_{k-1}^{(i)})\boldsymbol{P}_{k-1}^{-(i)}\boldsymbol{A}_{k-1}^{\mathrm{T}}(c_{k-1}^{(i)})+\boldsymbol{Q}_{k-1}(c_{k-1}^{(i)})\end{cases}\tag{1}$$

式中:$c_{k-1}^{(i)}$ 为数据关联指数;$\boldsymbol{A}_{k-1}$ 和 $\boldsymbol{Q}_{k-1}$ 分别为 $k-1$ 时刻的转换矩阵和过程噪声矩阵;$\boldsymbol{m}_{k|k-1}^{(i)}$ 和 $\boldsymbol{P}_{k|k-1}^{(i)}$ 分别为粒子 i 在 k 时刻的预估状态和协方差。

计算每个粒子新的非归一化权重为

$$(w_k^{(i)})^* \propto (w_{k-1}^{(i)})^* \frac{p(\boldsymbol{y}_k|\boldsymbol{m}_k^{(i)})p(\boldsymbol{m}_k^{(i)}|\boldsymbol{m}_{k-1}^{(i)})}{\pi(\boldsymbol{m}_k|\boldsymbol{m}_{k-1}^{(i)},\boldsymbol{y}_{1:k})}\tag{2}$$

式中:$w_{k-1}^{(i)}$ 为粒子权重;$\boldsymbol{y}_k$ 为 k 时刻的量测;$\boldsymbol{y}_{1:k}$ 为 $1\sim k$ 时刻的所有量测;$\pi(\boldsymbol{m}_k|\boldsymbol{m}_{k-1}^{(i)},\boldsymbol{y}_{1:k})$ 为前一时刻的重要性分布;$p(\boldsymbol{m}_k^{(i)}|\boldsymbol{m}_{k-1}^{(i)})$ 为目标的动态模型,描述了目标的随机动态;$p(\boldsymbol{y}_k|\boldsymbol{m}_k^{(i)})$ 为测量模型,描述了基于目标当前状态的测量值分布。

进行权重归一化处理:

$$w_k^{(i)}=\frac{(w_k^{(i)})^*}{\sum_{i=1}^{N}(w_k^{(i)})^*}\tag{3}$$

步骤 2　检测单元相似度计算。

基于高斯概率分布函数计算检测单元 $\boldsymbol{u}_k$ 与每个粒子预估位置的空间相似度:

$$p^{(i)}(\boldsymbol{u}_k)=N(\boldsymbol{u}_k|\boldsymbol{H}_k(c_k^{(i)})\boldsymbol{m}_{k|k-1}^{(i)},\ \boldsymbol{H}_k(c_k^{(i)})\boldsymbol{P}_{k|k-1}^{(i)}\boldsymbol{H}_k^{\mathrm{T}}(c_k^{(i)})+\boldsymbol{R}_k(c_k^{(i)}))\tag{4}$$

式中:$\boldsymbol{H}_k$ 和 $\boldsymbol{R}_k$ 分别为 k 时刻的量测模型矩阵和量测噪声矩阵;$N(\cdot)$ 表示检测单元与粒子预估位置的相似度。

结合每个粒子的权重,得到每个检测单元与所有粒子群预估位置的空间相似度为

北航学报 赠阅

$$p(\boldsymbol{u}_k)=\sum_{i=1}^{N}(p^{(i)}(\boldsymbol{u}_k)w_k^{(i)}) \tag{5}$$

步骤 3 检测门限修正。

利用步骤 2 计算所得的每个检测单元与所有粒子群预估位置的空间相似度，对每个检测单元的固定检测门限 θ 进行修正[15-16]，修正后的检测门限表示为

$$t(\boldsymbol{u}_k)=\theta e^{-\gamma p(\boldsymbol{u}_k)} \tag{6}$$

式中：γ 为门限修正强度参数。

该模型的合理性在于：对于与粒子预估位置相距较远的检测单元，其为目标量测的概率较小，则 $p(\boldsymbol{u}_k)$ 值趋近于 0，$e^{-\gamma p(\boldsymbol{u}_k)}$ 值趋近于 1，以至于不对阈值进行修正；对于与粒子预估位置相距较近的检测单元，其为目标量测的概率增大，则 $p(\boldsymbol{u}_k)$ 值增大，且 $0<e^{-\gamma p(\boldsymbol{u}_k)}<1$。参数 γ 能够调节该指数函数模型的非线性属性，较大的 γ 值能够强化模型对检测门限的影响；相反，较小的 γ 值会弱化以上影响。在目标检测过程中，对于小弱目标，一般会设置较低的检测门限，通过跟踪算法剔除杂波和虚警，并通过本文算法进一步降低预估区域的检测门限；对于回波较强的目标，可适当提高检测门限，以减少虚警数量，提高跟踪效率。

步骤 4 目标检测。

根据步骤 3 计算所得的检测门限，由式(7)判断检测单元中是否存在目标：

$$E(\boldsymbol{u}_k)=\begin{cases}1 & \mathrm{rcs}(\boldsymbol{u}_k)\geqslant t(\boldsymbol{u}_k)\\ 0 & \text{其他}\end{cases} \tag{7}$$

式中：$\mathrm{rcs}(\boldsymbol{u}_k)$ 为检测单元 $\boldsymbol{u}_k$ 处的量测回波强度值，若该值大于或等于检测门限，则数据确认为目标($E(\boldsymbol{u}_k)=1$)，由 $\boldsymbol{y}_k$ 表示，反之则无目标($E(\boldsymbol{u}_k)=0$)。

步骤 5 数据关联。

假设马尔可夫模型为 m 次，则数据关联指数 c_k 取决于先前的 m 个关联结果 $c_{k-m:k-1}$，假设虚警在量测空间 V 中均匀分布，卡尔曼滤波的量测相似度计算如下：

$$p(\boldsymbol{y}_k\mid c_k,\boldsymbol{y}_{1:k-1},c_{1:k-1}^{(i)})=\begin{cases}1/V & c_k=0\\ \mathrm{KF}_{\mathrm{lh}}(\boldsymbol{y}_k,\boldsymbol{m}_{j,k\mid k-1}^{(i)},\boldsymbol{P}_{j,k\mid k-1}^{(i)},\boldsymbol{H}_{j,k},\boldsymbol{R}_{j,k}) & c_k=j\end{cases} \tag{8}$$

式中：$j=1,2,\cdots,T$，T 为目标数目；$\mathrm{KF}_{\mathrm{lh}}(\cdot)$ 代表卡尔曼滤波器测量相似度估计；$\boldsymbol{H}_{j,k}$ 和 $\boldsymbol{R}_{j,k}$ 分别为目标 j 的量测模型矩阵和量测噪声矩阵。

对于 $j=1,2,\cdots,T$，有

$$[\boldsymbol{m}_{j,k\mid k-1}^{(i)},\boldsymbol{P}_{j,k\mid k-1}^{(i)}]=\mathrm{KF}_{\mathrm{p}}(\boldsymbol{m}_{j,k-1}^{(i)},\boldsymbol{P}_{j,k-1}^{(i)},\boldsymbol{A}_{j,k-1},\boldsymbol{Q}_{j,k-1}) \tag{9}$$

式中：$\mathrm{KF}_{\mathrm{p}}(\cdot)$ 代表卡尔曼滤波的预估部分；$\boldsymbol{m}_{j,k-1}^{(i)}$ 和 $\boldsymbol{P}_{j,k-1}^{(i)}$ 分别为粒子 i 目标 j 的预估值和协方差；$\boldsymbol{A}_{j,k-1}$ 和 $\boldsymbol{Q}_{j,k-1}$ 分别为动态模型转换矩阵和目标 j 的过程噪声矩阵。

数据关联结果由最优重要性分布决定，其概率分布结果计算如下：

1）计算非归一化噪声关联概率：

$$\hat{\pi}_0^{(i)}=p(\boldsymbol{y}_k\mid c_k^{(i)}=0,\boldsymbol{y}_{1:k-1},c_{1:k-1}^{(i)})\cdot p(c_k^{(i)}=0\mid c_{k-m:k-1}^{(i)}) \tag{10}$$

式中：$\hat{\pi}_0^{(i)}$ 为粒子 i 的归一化噪声关联概率；$p(\cdot\mid\cdot)$ 表示条件概率。

2）为每个目标 $j=1,2,\cdots,T$ 计算非归一化目标关联概率：

$$\hat{\pi}_j^{(i)}=p(\boldsymbol{y}_k\mid c_k^{(i)}=j,\boldsymbol{y}_{1:k-1},c_{1:k-1}^{(i)})\cdot p(c_k^{(i)}=j\mid c_{k-m:k-1}^{(i)}) \tag{11}$$

式中：$\hat{\pi}_j^{(i)}$ 为粒子 i 归一化的目标 j 的关联概率。

3）归一化重要性分布：

$$\pi_j^{(i)}=\frac{\hat{\pi}_j^{(i)}}{\sum_{j'=0}^{T}\hat{\pi}_{j'}^{(i)}}\qquad j=0,1,\cdots,T \tag{12}$$

4）概率 $\hat{\pi}_j^{(i)}(j=0,1,\cdots,T)$ 代表每个粒子中的离散概率分布，其中数据关联指数 $c_k^{(i)}$($i=1,2,\cdots,N$)采样如下：$c_k^{(i)}=0$ 的概率为 $\pi_0^{(i)}$；$c_k^{(i)}=1$ 的概率为 $\pi_1^{(i)}$；$c_k^{(i)}=2$ 的概率为 $\pi_2^{(i)}$；…；$c_k^{(i)}=T$ 的概率为 $\pi_T^{(i)}$。

步骤 6 粒子滤波目标状态修正。

基于数据关联结果 $c_k^{(i)}$ 对每个粒子进行卡尔曼滤波更新：

$$\begin{cases}\boldsymbol{v}_k^{(i)}=\boldsymbol{y}_k-\boldsymbol{H}_k(c_k^{(i)})\boldsymbol{m}_{k\mid k-1}^{(i)}\\ \boldsymbol{S}_k^{(i)}=\boldsymbol{H}_k(c_k^{(i)})\boldsymbol{P}_{k\mid k-1}^{(i)}\boldsymbol{H}_k^{\mathrm{T}}(c_k^{(i)})+\boldsymbol{R}_k(c_k^{(i)})\\ \boldsymbol{K}_k^{(i)}=\boldsymbol{P}_{k\mid k-1}^{(i)}\boldsymbol{H}_k^{\mathrm{T}}(c_k^{(i)})(\boldsymbol{S}_k^{(i)})^{-1}\\ \boldsymbol{m}_k^{(i)}=\boldsymbol{m}_{k\mid k-1}^{(i)}+\boldsymbol{K}_k^{(i)}\boldsymbol{v}_k^{(i)}\\ \boldsymbol{P}_k^{(i)}=\boldsymbol{P}_{k\mid k-1}^{(i)}-\boldsymbol{K}_k^{(i)}\boldsymbol{S}_k^{(i)}(\boldsymbol{K}_k^{(i)})^{\mathrm{T}}\end{cases} \tag{13}$$

式中：$\boldsymbol{m}_k^{(i)}$ 和 $\boldsymbol{P}_k^{(i)}$ 分别为粒子 i 在 k 时刻更新后的估计值和协方差；$\boldsymbol{v}_k^{(i)}$、$\boldsymbol{S}_k^{(i)}$ 和 $\boldsymbol{K}_k^{(i)}$ 为过程变量。

重新计算粒子权重并估计有效粒子数：

$$n_{\mathrm{eff}}=\frac{1}{\sum_{i=1}^{N}(w_k^{(i)})^2} \tag{14}$$

如有效粒子数过低(如 $n_{\mathrm{eff}}<N/4$)，则进行重采样。

3 实验结果分析

本节分别针对仿真数据和雷达实测数据，评

北航学报 赠阅

价本文算法的有效性，评价指标包括目标数 N_d、虚警数 N_{fa}、检测率 P_d、虚警率 P_{fa}、工作特征（Receiver Operator Characteristic，ROC）曲线、均方根误差（Root-Mean-Square Error，RMSE）等。

3.1　仿真数据

根据雷达目标散射原理，RCS 随目标与雷达的相对角度变化，具有起伏特性。雷达回波起伏与 RCS 值相关，雷达与目标间角度的变化导致雷达回波强度变化，最终决定雷达图像中像素灰度值的大小。同时，低空雷达具有较宽的垂直波束，接收的回波中容易混入地杂波而导致虚警。本节的仿真数据以瑞利分布模拟地杂波，以 4 类 χ^2 分布模拟低空目标的 RCS 概率密度分布[17]。

模拟低空目标的 4 类 χ^2 分布概率密度函数为

$$p(\sigma)=\frac{k}{(k-1)!\bar{\sigma}}\left(\frac{k\sigma}{\bar{\sigma}}\right)^{k-1}\exp\left(-\frac{k\sigma}{\bar{\sigma}}\right) \tag{15}$$

式中：$\bar{\sigma}$ 为 RCS 平均值；σ 为 RCS 随机变量；$k=1,2,3,4$ 代表 4 类分布的双自由度数值。

模拟地杂波的瑞利分布概率密度函数为

$$p(\sigma)=\frac{\sigma}{b^2}\exp\left(-\frac{\sigma^2}{2b^2}\right) \tag{16}$$

式中：b 为瑞利系数。

以离散维纳过程速度模型建立杂波环境中的单目标运动模型，其中目标的状态可以写为

$$\boldsymbol{x}_k=(x_k\quad y_k\quad \dot{x}_k\quad \dot{y}_k)^{\mathrm{T}} \tag{17}$$

式中：(x_k,y_k) 为二维直角坐标系中的目标位置；$(\dot{x}_k,\dot{y}_k)$ 为速度。

离散动态可以表示为线性、时不变构造方程：

$$\boldsymbol{x}_k=\begin{bmatrix}1&0&\Delta t&0\\0&1&0&\Delta t\\0&0&1&0\\0&0&0&1\end{bmatrix}\boldsymbol{x}_{k-1}+\boldsymbol{q}_{k-1} \tag{18}$$

式中：$\boldsymbol{q}_{k-1}$ 为离散高斯过程白噪声，矩特征为

$$\begin{cases}E(\boldsymbol{q}_{k-1})=\boldsymbol{0}\\ E(\boldsymbol{q}_{k-1}\boldsymbol{q}_{k-1}^{\mathrm{T}})=\begin{bmatrix}\frac{1}{3}\Delta t^3&0&\frac{1}{2}\Delta t^2&0\\0&\frac{1}{3}\Delta t^3&0&\frac{1}{2}\Delta t^2\\\frac{1}{2}\Delta t^2&0&\Delta t&0\\0&\frac{1}{2}\Delta t^2&0&\Delta t\end{bmatrix}q\end{cases} \tag{19}$$

其中：时间步长设定为 $\Delta t=0.1$；过程噪声的功率谱密度设定为 $q=0.1$。

为构造杂波测量环境，如果测量值为噪声，则数据关联指数 c_k 设定为 0，如果为真实目标则设定为 1。杂波测量值均匀地分布在空间 $[-5,5]\times[-4,4]$ 中。真实目标的测量模型与附加的高斯噪声成线性关系。因此，可以将联合测量相似度表示为

$$p(\boldsymbol{y}_k\mid\boldsymbol{x}_k,c_k)=\begin{cases}1/80 & c_k=0\\ N(\boldsymbol{y}_k\mid\boldsymbol{H}\boldsymbol{x}_k,\boldsymbol{R}) & c_k=1\end{cases} \tag{20}$$

量测噪声矩阵为

$$\boldsymbol{R}=\begin{bmatrix}0.05&0\\0&0.05\end{bmatrix} \tag{21}$$

数据关联的先验值完全独立，写为

$$p(c_k)=\begin{cases}p_c & c_k=0\\ 1-p_c & c_k=1\end{cases} \tag{22}$$

说明量测为杂波的概率是 p_c，量测为目标的概率是 $1-p_c$。p_c 代表了杂波在量测中所占的比例，该值越大，跟踪的难度也越大。

如图 2 所示，经过 240 步仿真，目标从（-4，-0.2）出发，在 0～1.0 s 以速度（1，0）匀速运动，在 1.1～3.0 s 完成右转弯，在 3.1～3.5 s 以速度（0，-1）匀速运动，在 3.6～5.5 s 完成左转弯，在 5.6～8.0 s 以速度（1，0）匀速运动，在 8.1～10.5 s 完成左转弯，在 10.6～14.0 s 以速度（0，1）匀速运动，在 14.1～16.5 s 完成左转弯，在 16.6～19.0 s 以速度（-1，0）匀速运动，在 19.1～21.0 s 完成左转弯，在 21.1～21.5 s 以速度（0，-1）匀速运动，在 21.6～23.5 s 完成右转弯，在 23.6～24.0 s 以速度（-1，0）匀速运动，直至结束。杂波概率为 0.3。针对 4 类目标 RCS 模型，按照图2中的雷达目标运动轨迹，经过1 000次

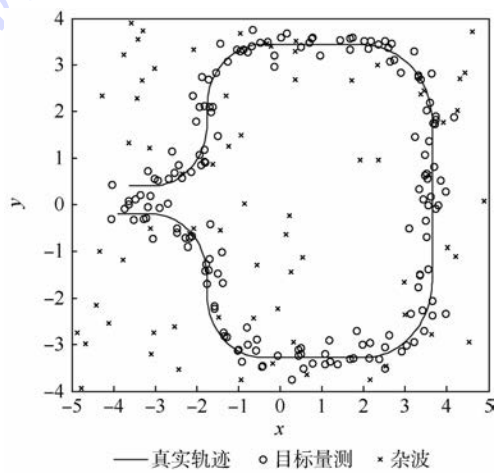


图 2　雷达目标仿真运动轨迹

Fig. 2　Simulation of radar target motion trajectory

北航学报 赠阅

蒙特卡罗仿真并取平均值,图 3 给出不同杂波概率条件下(p_c = 0.1, 0.3, 0.5)的 ROC 曲线和 RMSE 曲线,对比了采用本文算法对检测门限进行修正前后的跟踪结果,粒子数为 10。

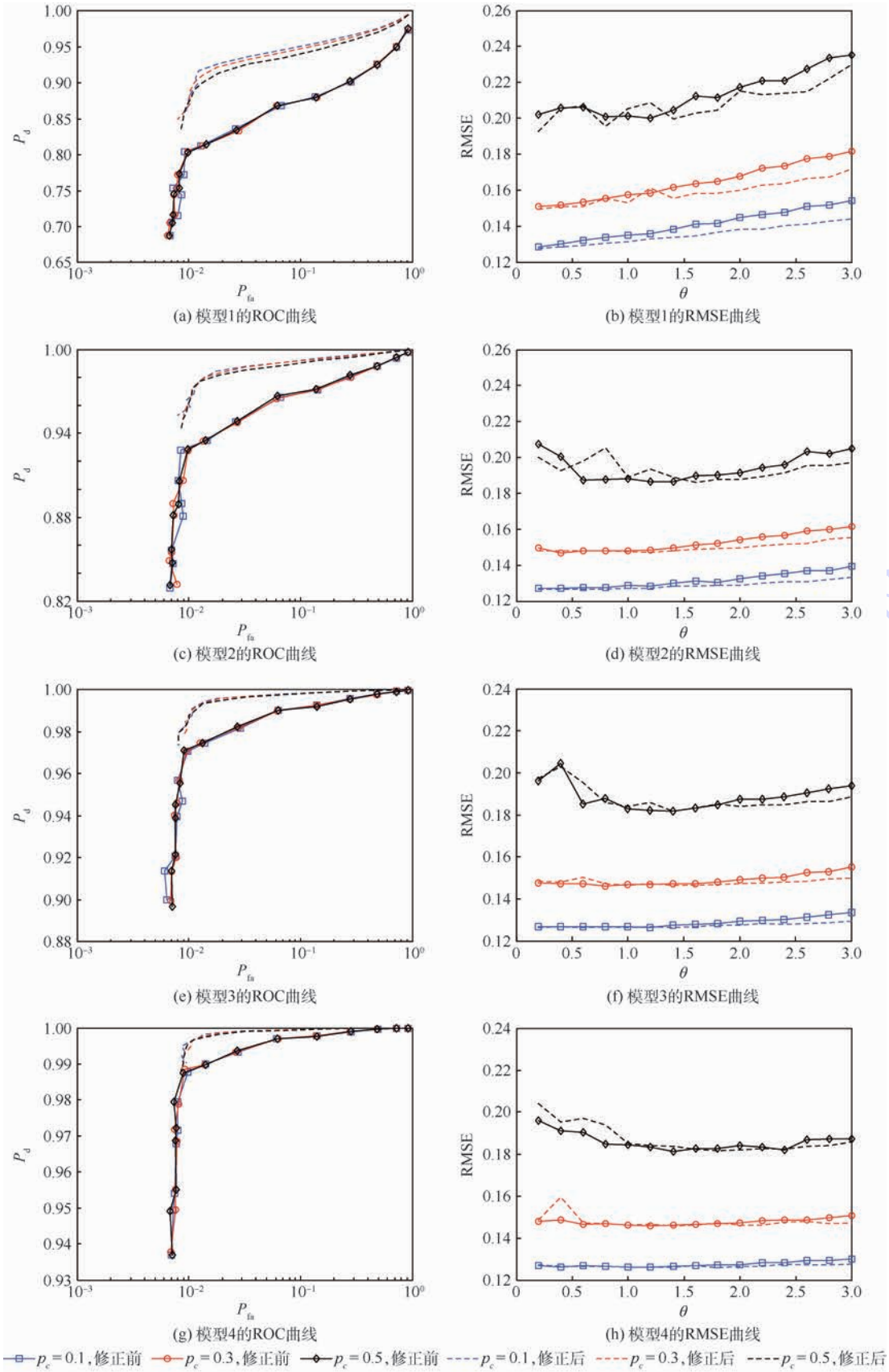


图 3 雷达目标仿真运动跟踪效果

Fig. 3 Simulation results of radar target motion tracking

北航学报 赠阅

由图 3 中 4 类模型的 ROC 曲线可见,本文算法均明显改善了 4 类目标模型的检测效果,其中,模型 1 的检测效果稍差,模型 2 ~ 模型 4 的检测效果逐渐改善,这是由目标的散射特性决定的。具体而言,模型 1 的典型目标为前向观察的小型喷气飞机,模型 2 的典型目标为螺旋桨推进飞机、直升机等,模型 3 的典型目标为喷气飞机、大型民用客机,模型 4 的典型目标为侧向观测的导弹与高速飞行体等。显然,模型 1 的检测难度最大,模型 2 ~ 模型 4 的检测难度逐渐降低。由于目标检测阈值在修正前后均采用了数据关联算法剔除杂波,p_c 值对 ROC 曲线的影响不大。由图 3 中 4 类模型的 RMSE 曲线可见,p_c 值对 RMSE 曲线的作用明显,当 p_c 值提高时,RMSE 明显提高,跟踪效果变差。同时发现,当杂波比例较低时(p_c = 0.1, 0.3),量测中的虚警较少,跟踪精度(RMSE)较高,此时,本文算法利用跟踪结果对检测门限进行反馈修正,能够进一步提高跟踪精度。但是,当杂波比例较高时(p_c = 0.5),杂波和虚警在量测中的比例达到 50%,此时由于目标量测较少,导致跟踪精度(RMSE)降低。由实验结果可见,如果阈值设置过低(θ < 1.5),检测门限经过修正后的跟踪结果反而可能进一步恶化,这可能是由于过低的阈值引入部分虚警导致的。因此,当虚警比例较高时,建议适当提高阈值(θ > 1.5),采用检测门限反馈修正算法能够在一定程度上改善跟踪结果,确保算法收敛。

针对 4 类目标 RCS 模型,表 1 ~ 表 4 给出了粒子数 N 为 10、50 和 100 的情况下,本文算法给定不同分割阈值时的 RMSE 值,杂波比例均设定为 p_c = 0.1。通过对以上 4 组数据的分析,粒子数对检测结果影响不大,粒子数较多时的检测结果有时甚至稍逊于粒子数较少的时候。其原因在于本文算法采用 Rao-Blackwellization 方法将单目标跟踪与数据关联分开处理,将SMC(粒子滤波)方法用于数据关联,采用较少的粒子数,实现杂波与虚警量测中的多目标跟踪,可视为 MHT 的推广。Rao-Blackwellized 粒子滤波的理论基础是某些滤波方程可以闭合的形式计算,其他采用蒙特卡罗采样,而不是对所有方程都采用采样方法。Rao-Blackwell 的思想实现了较小的估计方差,其可视为用无穷集合去替代有限集合,往往能得到更为准确的结果。

表 1　不同粒子数情况下模型 1 的 RMSE 值

Table 1　RMSE values of Model 1 with different numbers of particles

θ	RMSE		
	N = 10	N = 50	N = 100
0.5	0.1287	0.1289	0.1294
1.0	0.1314	0.1313	0.1318
1.5	0.1342	0.1336	0.1347
2.0	0.1375	0.1380	0.1375
2.5	0.1406	0.1409	0.1408
3.0	0.1437	0.1441	0.1446

表 2　不同粒子数情况下模型 2 的 RMSE 值

Table 2　RMSE values of Model 2 with different numbers of particles

θ	RMSE		
	N = 10	N = 50	N = 100
0.5	0.1266	0.1265	0.1265
1.0	0.1273	0.1270	0.1267
1.5	0.1282	0.1279	0.1280
2.0	0.1291	0.1296	0.1291
2.5	0.1311	0.1305	0.1307
3.0	0.1325	0.1328	0.1329

表 3　不同粒子数情况下模型 3 的 RMSE 值

Table 3　RMSE values of Model 3 with different numbers of particles

θ	RMSE		
	N = 10	N = 50	N = 100
0.5	0.1266	0.1264	0.1264
1.0	0.1264	0.1265	0.1264
1.5	0.1268	0.1259	0.1267
2.0	0.1270	0.1272	0.1271
2.5	0.1284	0.1283	0.1282
3.0	0.1291	0.1296	0.1285

表 4　不同粒子数情况下模型 4 的 RMSE 值

Table 4　RMSE values of Model 4 with different numbers of particles

θ	RMSE		
	N = 10	N = 50	N = 100
0.5	0.1266	0.1269	0.1265
1.0	0.1262	0.1258	0.1263
1.5	0.1266	0.1265	0.1262
2.0	0.1263	0.1265	0.1260
2.5	0.1272	0.1266	0.1272
3.0	0.1275	0.1275	0.1275

3.2　实测数据

基于一组 S 波段非相参雷达数据和一组 S 波段相参雷达数据,在前期研究成果的基础上,采用本文算法进行处理,并与现有经典算法进行对比分析,2 组数据的采样周期均为 2.5 s。第 1 组测试数据为 S 波段非相参雷达采集的图像序列,图中目标为一架波音 737 客机,分辨率 1 024 ×

北航学报 赠阅

1 024,量程 22 km,共 120 帧。图 4 给出了某帧图像的检测跟踪结果。其中,图 4(a)在原始图像中标定了目标位置,分别采用固定阈值(Fixed Threshold,FT)、单元平均恒虚警(Cell Average-Constant False Alarm Rate,CA-CFAR)和本文算法进行目标检测;图 4(b)为目标跟踪结果。实验结果表明,目标检测阈值设置为 $\theta=70$ 时,FT 算法在引入 228 个虚警的前提下能够检测到目标,CA-CFAR 算法不仅能检测到目标,且将虚警数降为 25 个,基于前期研究成果[18],结合本文算法,粒子数为 10,能够检测到小弱目标,且将虚警数减少到 2 个。图 4(b)为检测阈值设定为 $\theta=120$ 时,本文算法的目标跟踪结果,此时 120 帧图像的目标检测累积之和为 $N_d=119$,$N_{fa}=3$,所有信息叠加在雷达背景图像上,3 个虚警位置疑为未形成轨迹的低空未知目标。

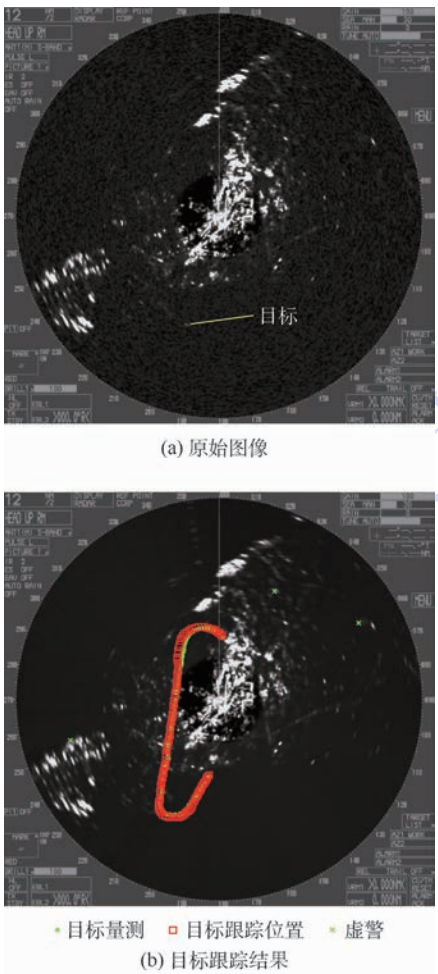


图 4 S 波段非相参雷达目标检测与跟踪结果

Fig. 4 Target detection and tracking results of S-band incoherent radar

为比较各类算法的鲁棒性,表 5 给出了设置不同分割阈值的情况下,各类算法处理 120 帧雷达图像获取的目标数和虚警数总和,包括 FT、CA-CFAR、最小选择恒虚警(Small of-Constant False Alarm Rate,SO-CFAR)和本文算法。形态学处理(Morphological Processing, MP)是一种经典的图像处理算法,对于剔除雷达图像中由单个像素组成的杂波信息,效果良好[19]。此处,将 MP 算法作为所有算法的后处理,将“腐蚀”和“膨胀”处理相结合,既能剔除虚警,又能将属于同一目标的多个量测区域重新联通,避免将目标信息误认为虚警。由表 5 可见,当检测到的目标数为 114 时,FT 的虚警数为 3 539,CA-CFAR 的虚警数为 2 469,而 SO-CFAR 的虚警数为 3 541。本文算法在设定不同阈值时,几乎能检测到所有目标,虚警数最小值仅为 1。总体而言,本文算法最优,CA-CFAR 算法优于 FT 算法,SO-CFAR 算法甚至略逊于 FT 算法,其原因在于 SO-CFAR 算法通常将局部检测阈值设置过低,容易导致较高虚警率。

第 2 组测试数据在某雷达测试外场,由 S 波段相参雷达采集的轻小型无人机目标数据,该

表 5 S 波段非相参雷达目标检测结果对比

Table 5 Comparison of target detection results of S-band incoherent radar

θ	FT + MP		CA-CFAR + MP		SO-CFAR + MP		本文算法 + MP	
	N_d	N_{fa}	N_d	N_{fa}	N_d	N_{fa}	N_d	N_{fa}
30	119	16 030	114	3 320	119	16 053	120	3 504
40	119	9 357	114	2 469	119	9 368	120	1 032
50	116	5 674	113	1 886	116	5 679	120	329
60	114	3 539	111	1 492	114	3 541	120	105
70	112	2 077	109	1 124	112	2 077	120	36
80	112	1 197	110	774	112	1 197	119	16
90	109	645	107	473	109	645	119	10
100	108	331	107	284	108	331	119	6
110	106	147	105	136	106	147	119	3
120	104	44	104	41	104	44	119	1

北航学报 赠阅

无人机目标沿正北偏西方向逐渐远离雷达飞行，量程 6 km。门限 θ 为目标 RCS 值，典型无人机目标的 RCS 值很小，属于小弱目标。

图 5 中，采样周期为 360 个。表 6 对比了设定不同 θ 值时的检测结果。本文算法将粒子数设定为 10，能够在检测到 339 个无人机目标的情况下将虚警数降低到 17。相参雷达已经剔除了静止的背景地物回波，且对虚警具有一定的抑制作用，本文算法的作用主要在于跟踪目标的同时提高对小弱目标的检测能力。

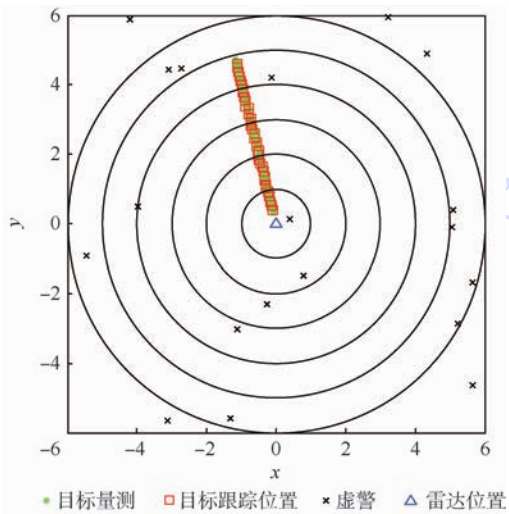


图 5　S 波段相参雷达无人机目标检测与跟踪结果

Fig. 5　UAV target detection and tracking results of S-band coherent radar

表 6　S 波段相参雷达无人机目标检测结果对比

Table 6　Comparison of UAV target detection results of S-band coherent radar

θ	FT 算法		本文算法	
	N_d	N_{fa}	N_d	N_{fa}
0.01	327	21	339	21
0.02	273	21	339	21
0.03	201	17	339	17

4　结　论

本文结合 SMC 方法计算的粒子范围确定波门大小，在考虑粒子权重的前提下，利用检测单元与所有粒子的相对位置对检测门限进行修正，提高对小弱目标的检测能力。通过将本文算法与其他经典算法进行对比，得出以下结论：

1）由仿真结果可见，本文算法对于各种类型雷达目标的检测结果均有明显改善，在杂波比例较低时，对跟踪精度也有一定提高。

2）本文算法分别对每个目标进行卡尔曼滤波预估、更新和测量相似度估计，比纯粹的粒子滤波方法效率更高，以较少的粒子数就能实现对目标的精确跟踪。

3）将本文算法与前期研究成果相结合，并应用于 S 波段非相参雷达采集的图像序列，检测结果表明，本文算法能够改善局部区域内的目标检测结果，优于其他经典算法。

4）将本文算法应用于 S 波段相参雷达数据，能够提高对低空无人机目标的检测跟踪能力。

参考文献（References）

[1] 韩崇昭，朱洪艳，段战胜，等. 多源信息融合[M]. 北京：清华大学出版社，2006：320-365.
HAN C Z，ZHU H Y，DUAN Z S，et al. Multi-source information fusion[M]. Beijing：Tsinghua University Press，2006：320-365 (in Chinese).

[2] ASLAN M S，SARANL A. Threshold optimization for tracking a nonmaneuvering target[J]. IEEE Transactions on Aerospace and Electronic Systems，2011，37(2)：2844-2859.

[3] 吕信明. 军用无人机的发展及对策[J]. 国防科技，2013，34(1)：5-7.
LV X M. Military UAV development and countermeasures[J]. National Defense Science & Technology，2013，34(1)：5-7 (in Chinese).

[4] 何煦. 对低空小型无人机的对抗方法研究[J]. 中国新通信，2016，18(15)：147-149.
HE X. Research on countermeasures against low-altitude small UAV[J]. China New Telecommunications，2016，18(15)：147-149 (in Chinese).

[5] RICHARDS M A. Fundamentals of radar signal processing[M]. New York：McGraw-Hill，2005：1-15.

[6] SKOLNIK M I. Introduction to radar systems[M]. New York：McGraw-Hill，2002：5-10.

[7] BAR-SHALOM Y，LI X R，KIRUBARAJAN T. Estimation with application to tracking and navigation：Theory algorithms and software[M]. New York：John Wiley & Sons，Inc.，2001：35-42.

[8] LI X R，JILKOV V P. Survey of maneuvering target tracking，Part V：Multiple-model methods[J]. IEEE Transactions on Aerospace and Electronic Systems，2005，41(4)：1255-1321.

[9] BLACKMAN S S. Multiple hypothesis tracking for multiple target tracking[J]. IEEE Aerospace and Electronic Systems Magazine，2004，19(1-2)：5-18.

[10] 王云奇. 知识辅助检测跟踪一体化算法研究[D]. 成都：电子科技大学，2014：4-12.
WANG Y Q. Research of detection-tracking-integration algorithm with knowledge-aided[D]. Chengdu：University of Electronic Science and Technology of China，2014：4-12 (in Chinese).

[11] 刘红亮，周生华，刘宏伟，等. 一种航迹恒虚警的目标检测跟踪一体化算法[J]. 电子与信息学报，2016，38(5)：1072-1078.
LIU H L，ZHOU S H，LIU H W，et al. An integrated target detection and tracking algorithm with constant track false alarm

北航学报 赠阅

rate[J]. Journal of Electronics & Information Technology, 2016,38(5):1072-1078(in Chinese).

[12] 严俊坤,刘红亮,戴奉周,等.一种具有恒虚警性质的检测跟踪联合处理算法[J].电子与信息学报,2014,36(11):2666-2671.
YAN J K,LIU H L,DAI F Z,et al. Joint detection and tracking processing algorithm with constant false alarm rate property [J]. Journal of Electronics & Information Technology,2014,36(11):2666-2671(in Chinese).

[13] 闫学昭.雷达目标检测与跟踪交互处理[D].大连:大连海事大学,2011:35-40.
YAN X Z. Radar target detection and tracking interactive processing[D]. Dalian: Dalian Maritime University, 2011: 35-40 (in Chinese).

[14] KOKKALA J,SÄRKKÄ S. Combining particle MCMC with Rao-Blackwellized Monte Carlo data association for parameter estimation in multiple target tracking[J]. Digital Signal Processing,2015,47:84-95.

[15] MCHUGH J M,KONRAD J,SALIGRAMA V,et al. Foreground-adaptive background subtraction[J]. IEEE Signal Processing Letters,2009,16(5):390-393.

[16] STAUFFER C,GRIMSON E. Learning patterns of activity using real-time tracking[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2000,22(8):747-757.

[17] 黄培康,殷红成,许小剑.雷达目标特性[M].北京:电子工业出版社,2005:114.
HUANG P K, YIN H C, XU X J. Radar target characteristics [M]. Beijing: Publishing House of Electronics Industry, 2005: 114(in Chinese).

[18] 陈唯实,李敬.基于空域特性的低空空域雷达目标检测[J].航空学报,2015,36(9):3060-3068.
CHEN W S,LI J. Radar target detection in low-altitude airspace with spatial features[J]. Acta Aeronautica et Astronautica Sinica,2015,36(9):3060-3068(in Chinese).

[19] GONZALEZ R C, WOODS R E, EDDINS S L. Digital image processing using MATLAB [M]. Englewood Cliffs: Prentice Hall,2009:261-268.

作者简介:

陈唯实 男,博士,高级工程师。主要研究方向:低空空域安全监视、雷达目标检测与跟踪、机场安全运行技术。

Joint optimization of detection and tracking with Rao-Blackwellized Monte Carlo data association

CHEN Weishi*, YAN Jun, LI Jing

(Airport Research Institute, China Academy of Civil Aviation Science and Technology, Beijing 100028, China)

Abstract: A joint optimization algorithm was proposed for radar target detection and tracking with Rao-Blackwellized Monte Carlo data association. Rao-Blackwellization made the separation of single target tracking and data association, where the data association was solved by the sequential Monte Carlo method (particle filtering), leading to the multiple target tracking in the environment of clutter and false alarm measurements. Meanwhile, the size of the wave gate depended on the distribution range of particles. Under the consideration of the particle weights, the detection threshold was modified with the relative position of the detection units to all the particles, improving the detection rate. Finally, combined with the algorithm for clutter suppression with spatial features achieved in the previous research, the proposed algorithm was applied to the simulated data as well as the ground-truth data collected by the S-band incoherent and coherent radars. It is demonstrated that the proposed algorithm can realize the detection and tracking of small targets with relatively small number of particles.

Keywords: data association; radar; target; detection; tracking

Received: 2017-04-13; **Accepted**: 2017-07-21; **Published online**: 2017-09-14 09:03

URL: kns.cnki.net/kcms/detail/11.2625.V.20170914.0903.001.html

Foundation items: Joint Research Foundation of National Natural Science Foundation of China (NSFC) and Civil Aviation Administration of China (CAAC) (U1633122); National Key R&D Program (2016YFC0800406)

* **Corresponding author**. E-mail: chenwsh@mail.castc.org.cn

北航学报 赠阅

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2017.0251

轴向间距对转子叶片颤振特性的影响机理

郑赟[1,2,*]，余永博[1,2]

（1. 先进航空发动机协同创新中心，北京 100083； 2. 北京航空航天大学 能源与动力工程学院，北京 100083）

摘 要：使用自行开发的非定常流固耦合数值模拟程序，研究了上游叶排影响转子叶片颤振特性的机理，采用影响系数法分析轴向间距影响转子气动弹性稳定性的规律。结果表明：在协调叶栅中，叶片吸力面相邻的叶片振动对转子叶片气动阻尼的大小起决定性作用，其影响甚至超过振动叶片本身的影响；多排环境中，导叶（IGV）对转子叶片气动阻尼最小值的影响最大，并使其对应的节径增大；相邻叶片振动引起的通道变化抑制了导叶对非定常压力波的反射作用；随着轴向间距的减小，导叶对非定常压力波的反射作用减弱了非定常压力波的周向衰减，从而增大了叶片振动的非定常影响范围；在多排环境中使用影响系数法需要测量更多的叶片才能得到较为准确的气动阻尼。

关 键 词：轴向间距；影响系数法；气动弹性稳定性；颤振；气动阻尼

中图分类号：V211.6

文献标识码：A **文章编号**：1001-5965(2018)04-0709-08

全下文载

当前航空发动机高推重比的设计要求导致更轻的结构、更高的级负荷和更紧凑的级间距，这些趋势都使得气动弹性稳定性问题在新的机型上越来越突出[1]。叶片颤振是叶片振动的一种，涉及到复杂的流固耦合作用，颤振一旦发生，经常会导致叶片大幅度的剧烈振动从而使其在很短的时间内断裂，造成结构巨大的破坏。

近些年来，国内外研究者对这种与颤振相关的非定常流动现象做过大量的参数化研究，包括叶片间相位角（Inter Blade Phase Angle，IBPA）、折合频率、振动模态、叶尖间隙、激波等。Lane[2]根据行波理论定义的叶片间相位角，描述振动的叶片间存在的相互气动干涉，是叶盘振动节径的表征量。Carta 和 St Hilaire[3-4]通过实验研究了叶片间相位角、迎角、折合频率等参数对发动机叶片气动弹性稳定性的影响。另外，叶片振动模态对颤振边界的影响很大，张小伟等[5]采用能量法发现振动模态通过影响叶片表面的非定常气动力使气动弹性稳定性发生改变。Bell 和 He[6]进行了三维涡轮叶栅弯曲振动下的非定常气动响应实验，首次研究了叶尖间隙对叶表非定常压力的影响，结果表明非定常压力幅值随间隙的变化规律一致。Isomura 和 Giles[7]指出通道激波对压力面有失稳作用，对吸力面有稳定作用；激波是否处于脱体位置是诱发颤振的重要条件。综上所述，大多数的叶片颤振研究都局限在单转子模型基础上。

随着对叶片颤振问题认识的加深，研究发现由于没有考虑上下游叶排对转子叶片非定常流动的影响，仅对单转子进行叶片颤振特性分析，有时会对叶片气弹失稳的算例得出气弹稳定的相反结论[8]，所以叶排干涉与颤振之间关联性研究也逐渐兴起。Culver 和 Liu[9]研究结果表明加入导叶（Inlet Guide Vanes，IGV）之后跨声速压气机叶片扭转振型的气动弹性稳定性下降。而 Huang 等[10]对低压涡轮叶片扭转振型的颤振研究表明，上游静子叶排对转子气动弹性稳定性在不同节径

收稿日期：2017-04-25；**录用日期**：2017-08-11；**网络出版时间**：2017-09-04 10:31
网络出版地址：kns.cnki.net/kcms/detail/11.2625.V.20170904.1031.001.html
* **通信作者**. E-mail：zheng_yun@buaa.edu.cn

引用格式：*郑赟，余永博. 轴向间距对转子叶片颤振特性的影响机理[J]. 北京航空航天大学学报，2018，44(4)：709-716.*
ZHENG Y，YU Y B. Influence mechanism of axial spacing on rotor blade flutter characteristics[J]. Journal of Beijing University of Aeronautics and Astronautics，2018，44(4)：709-716 (in Chinese).

北航学报 赠阅

下影响规律不一致。杨慧和李振鹏[11]用流固耦合法研究了转静干涉对三维压气机气动弹性稳定性的影响,结果表明导叶引起的非定常压力波反射使得弯曲振型的稳定性增强。上述结果表明上下游叶排对叶片气动弹性稳定性有着显著的影响,从而轴向间距成为学者研究上下游叶排影响的关键因素。Li 和 He[12]使用单通道模型研究了跨音压气机气动弹性稳定性受转静叶排轴向间距的影响规律,结果表明转静叶排存在一个最优间距值,静子引起的干涉效应使得转子的稳定性最佳。杨慧和沈真[13]用能量法研究了上下游叶排对二维压气机转子弯曲颤振特性的影响,结果表明压气机转子上下游叶排在小的轴向间距下,使转子气动弹性稳定性降低。上述方法模型虽然给出了上下游叶排及对转子叶片颤振特性的影响以及最佳轴向间距的存在,但没有分析出转子叶片气动阻尼主要来源以及转子叶片表面非定常气动功变化的规律。影响系数法[14]是一种基于小振幅振动线性叠加的方法,量化每个叶片对叶栅整体非定常气动响应的贡献。北京航空航天大学杨慧等[15]采用影响系数法考虑叶尖间隙变化对压气机线性叶栅进行了气动弹性实验。实验结果表明:随着叶尖间隙的增加,间隙流对振动叶片有失稳影响,间隙流在叶栅稳定性最差时影响最显著。

本文使用影响系数法,考虑上游叶排轴向间距变化,分析弯曲振动的涡轮转子的气动弹性稳定性。首先,对转子叶片表面非定常压力进行傅里叶变换,分析轴向间距影响气动阻尼的规律;其次,对影响系数法在多排算例中的适用性进行了分析;最后,分析叶片表面非定常压力幅值和相位影响因素,进一步解释轴向间距影响转子叶片最小气动阻尼系数的机理,加深对叶片颤振问题的认识。

1 数值方法

1.1 控制方程

流场的数值模拟采用了自行开发程序 HGAE,求解动边界和动态变形网格下的守恒型积分形式的三维非定常可压缩雷诺平均 Navier-Stokes 方程,其形式为

$$\frac{\partial}{\partial t}\int_{\Omega}\boldsymbol{q}\mathrm{d}\Omega+\oint_{\partial\Omega}(\boldsymbol{F}-w\boldsymbol{q})\hat{\boldsymbol{n}}\mathrm{d}A=\oint_{\Omega}\boldsymbol{S}\mathrm{d}\Omega \tag{1}$$

式中:$\partial\Omega$、Ω 和 $\hat{\boldsymbol{n}}$ 分别为控制体某物理时刻的表面面积、体积和表面外法向;$\boldsymbol{q}$ 为守恒型流体状态向量;A 为表面面积;w 为运动控制体表面的网格速度;$\boldsymbol{F}$ 为对流和耗散通量项;$\boldsymbol{S}$ 为外力、坐标系运动等对守恒律贡献的原项矢量;t 为时间。使用非结构化的有限体积法进行空间离散,使用 2 阶精度 ROE 迎风格式进行流通量计算,使用 Spalart-Allmaras 一方程湍流计算模型。该程序在气动弹性领域的数值验证可以参见文献[16]。

1.2 影响系数法

影响系数法是通过振动中间一个叶片来获得振动叶栅所有叶片间相位角下非定常气动响应数据的方法,简单高效,但不适用于非线性大分离流,有其局限性。其基本原理为:

1)协调叶栅所有叶片都是振动的,当不考虑叶片间相位角影响时,所有叶片都是一样的,某个叶片表面的非定常气动力相当于所有振动叶片对其影响的线性叠加;图 1 是由 $2m+1$ 个叶片组成的涡轮单转子叶片排(图 1 中 $m=1$),当所有叶片都振动时,参考叶片 0(表示为 b0)上无量纲非定常影响系数记为 C_{f0}(此处以压力 $\tilde{P}$ 表征,叶片间相位角为 σ,所有叶片都振动时,对于叶片 b0,叶片 bn 的振动对其的影响系数记为 $C_f(n,0)$,ω 为叶片振动频率,$\boldsymbol{h}$ 为叶片 b0 的位移,$\boldsymbol{h}_0$ 为叶片 b0 振动的幅值,安装角 $\gamma=56.65°$,振动方向为 60.4°。

$$\tilde{P}=C_{f0}\mathrm{e}^{\mathrm{i}\omega t}=\sum_{n=-m}^{m}C_f(n,0)\mathrm{e}^{\mathrm{i}n\sigma}\mathrm{e}^{\mathrm{i}\omega t} \tag{2}$$

因此获得叶片 b0 上由 $2m+1$ 个振动叶片耦合后的非定常影响系数

$$C_{f0}=\sum_{n=-m}^{m}C_f(n,0)\mathrm{e}^{\mathrm{i}n\sigma} \tag{3}$$

2)只有叶片 b0 振动时,参考叶片 b0 的振动对叶片 bn 的影响系数,记为 $C_f(n,0)$,系数可以通过实验或计算得出

$$C_f(-n,0)=C_f(0,n) \tag{4}$$

进一步得到

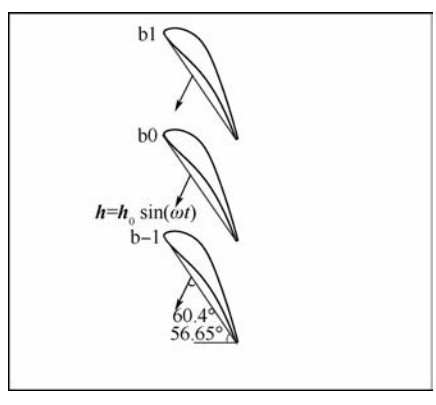


图 1 转子叶栅几何示意图

Fig. 1 Geometric sketch map of rotor cascade

$$C_{f0}=\sum_{n=-m}^{m}C_f(0,-n)\mathrm{e}^{\mathrm{i}n\sigma} \tag{5}$$

此方法的要点如下:非定常压力影响系数 $C_f(n,0)$ 称为“直接项”,其通过量化每个叶片对整个叶栅非定常气动响应的贡献,来体现每个叶片的相对重要性。而影响系数 C_{f0} 称为“耦合项”,为各个叶片气动阻尼之和。分解的“直接项”可以认识每个叶片的相对重要性,而“耦合项”能判断颤振起点,可以得到强迫响应中的阻尼项。由于叶片间相位角仅仅出现在后处理中不是计算参数,大大地缩减了获得所有叶片间相位角下振动叶栅气动阻尼的工作量。对影响系数法具体介绍参见文献[17]。

2　算例说明及定常验证

本文使用影响系数法进行气动弹性稳定性分析,转子叶栅共 21 个叶片(即 $m=10$),其几何叶型选自气动弹性标准算例 4[18],中间叶片 b0 的振动参数选自其二维亚声速工况[19]。标准算例 4 是三维的跨声速高负荷涡轮叶栅,叶片具有较大的厚度和弯角,容易产生气流分离,是现代典型的涡轮叶片造型,具有完整的实验数据,便于数值验证。实验采用了所有叶片全部振动的行波法,转子叶栅为协调叶栅,振动方向如图 1 所示,其基本实验参数见表 1。

为了研究轴向间距对转子气动弹性稳定性的影响,在转子上游加入了类似薄平板的导叶。图 2为导叶-转子叶片单通道网格示意图,为了满足程序的计算要求,计算网格为边界层加密的非结构网格,保证第 1 层网格的 y^+ 在 1.5 以下,周期性边界采用了对接型网格设置。在转静交界面处,定常、非定常计算分别采用掺混面模型和滑移面模型。

本文选取了 7 个不同的轴向间距 Δx 以弦长 c 无量纲化(即 $\Delta x/c$),分别为 10%、15%、30%、50%、70%、90%、150%,轴向间距改变时转子通道的网格保持不变,针对不同间距生成导叶通道的网格。表 2 为算例的边界条件,在计算时为了保证不同轴向间距下转子叶片表面的定常气动载荷基本一致,对转子出口的静压进行了微调。

表 1　叶片几何和实验参数

Table 1　Blade geometry and experimental parameters

参　数	数　值
弦长/mm	72
栅距/mm	56.25
安装角/(°)	56.65
振幅/(10^{-4} m)	3
振动方向/(°)	60.4
振动频率/Hz	149
折合频率	0.1187

在非定常颤振计算之前,进行了定常验证。图 3 为转子叶片表面定常压力系数 C_p 计算值和实验值的对比图,x 为网格坐标。计算结果与实验数据基本符合,仅在压力面尾缘和吸力面 0.3 倍弦长处存在细微差异。

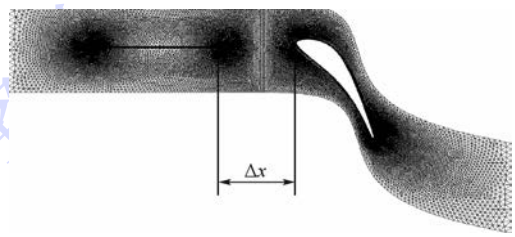


图 2　导叶-转子叶片单通道网格

Fig. 2　IGV-rotor blade single-passage grids

表 2　边界条件

Table 2　Boundary conditions

参　数	数　值
进口总压/kPa	160.9
进口总温/K	317.8
出口静压/kPa	101.3
出口气流角/(°)	-71.5
出口等熵马赫数	0.85

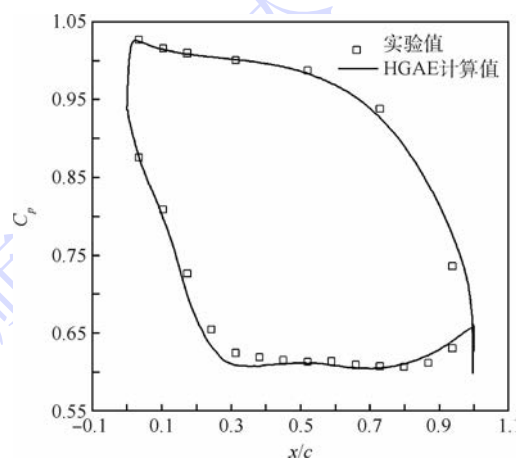


图 3　叶片表面定常压力系数分布

Fig. 3　Distribution of steady pressure coefficient on blade surface

3　计算结果

非定常颤振分析以收敛的定常计算结果作为初场。取转子叶片振动周期为 1 个周期,每周期包含 200 个时间步,时间步长为 $3.355\,704\,698\times10^{-5}$ s。

北航学报 赠阅

3.1 不同轴向间距对气动阻尼的影响

为了研究轴向间距对转子气动弹性稳定性的影响,首先进行了单转子颤振计算。图 4 为不同轴向间距下转子叶片气动阻尼系数 D 随叶片间相位角σ 的变化,从图中可以看出:单转子的最小气动阻尼值与实验值基本一致,说明影响系数法能准确预测最小气动阻尼系数以及其对应的叶片间相位角;轴向间距对气动阻尼系数的影响范围主要在气动弹性最不稳定状态对应的叶片间相位角附近;随着轴向间距的减小,最小气动阻尼系数对应的叶片间相位角明显左移(见图 4 中红色箭头),即导叶使得气动弹性最不稳定点的节径增大;在最佳轴向间距下,导叶的影响可以使转子气动弹性稳定性提高 32.04%,在工程应用中轴向间距对气动弹性稳定性的影响不可忽略。

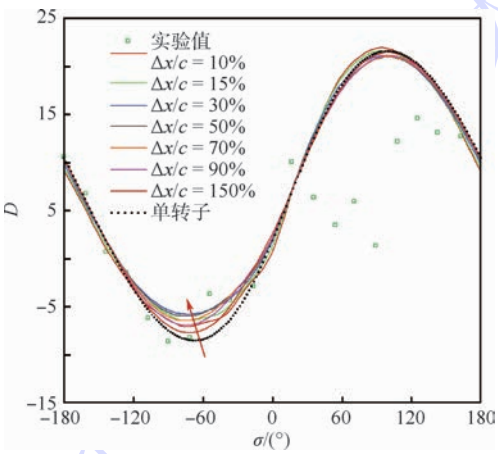


图 4 不同轴向间距下的气动阻尼系数

Fig. 4 Aerodynamic damping coefficient with different axial spacing

3.2 确定影响转子叶片气动阻尼的关键叶片

通常研究人员最关注气动阻尼系数最小值处对应的叶片间相位角。图 5 给出了单转子叶排在叶片间相位角为 -72°时各叶片非定常气动阻尼系数分量,各分量之和为影响系数耦合项 C_{f0}(见 1.2 节),图中 b-1 和 b0 气动阻尼系数分量最大,是转子叶片气动阻尼系数贡献最大的部分。可以看出,气动阻尼系数分量的数值随着与振动叶片周向距离的增加指数级的减小,比如 b-2 的气动阻尼系数大小只有 b0 的 4.27%。

图 6 为有导叶时 b-2 ~ b2 气动阻尼系数分量随 Δx 变化情况(孤立转子是以 $\Delta x/c = 180\%$ 来表示)。从图 6 可以看出:添加导叶之后,b-1 和 b0 仍是构成气动阻尼系数的主要部分;影响系数 $C_f(0,-1)$(表示在影响系数法中,叶片 b0 的振动对 b-1 气动阻尼系数的影响,见 1.2 节)的绝对值明显大于 $C_f(0,0)$,并且影响系数 $C_f(0,-1)$ 使得转子气动阻尼系数由正变为负,直接改变了转子气动弹性稳定性。上述结果表明:系数 $C_f(1,0)$(表示在协调叶栅中,叶片 b1 的振动对 b0 气动阻尼系数的影响,见 1.2 节)对转子气动阻尼系数的大小起决定性作用,其影响甚至超过振动叶片 b0 本身的影响 $C_f(0,0)$。

由于 b0 叶片振动产生的非定常扰动沿周向的快速衰减,使用影响系数法时,只需选取有限的叶片数目。为了检验影响系数法在多排算例中的适用性,图 7 给出了不同叶片数下转子在叶片间相位角为 -72°时气动阻尼系数随 Δx 的变化规律(孤立转子是以 $\Delta x/c = 180\%$ 来表示),b-1 ~ b1 表示 b-1、b0、b1 3 个叶片气动阻尼系数叠加。由图 7可知:随着轴向间距的减小,气动阻尼系数先

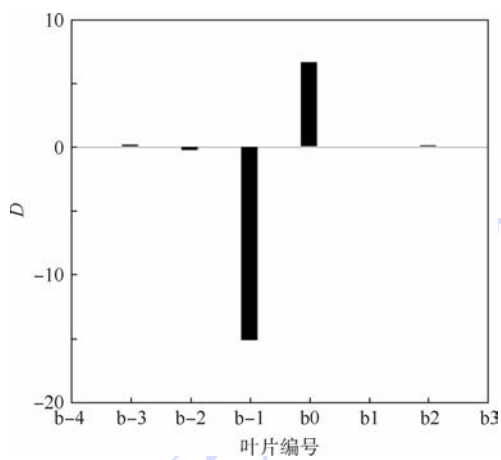


图 5 叶片间相位角为 -72°时,单转子各叶片气动阻尼系数分量

Fig. 5 Each blade's aerodynamic damping coefficient component of single-rotor at inter blade phase angle equals to -72°

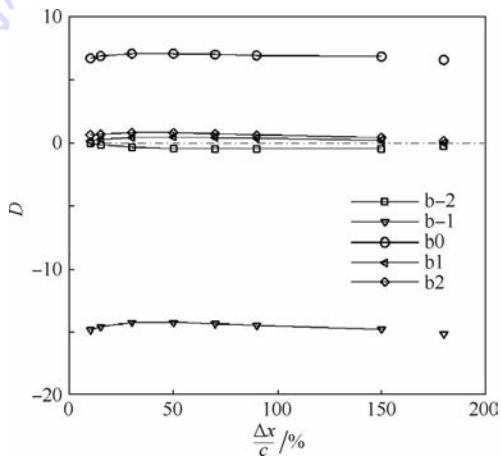


图 6 不同轴向间距下气动阻尼系数分量

Fig. 6 Aerodynamic damping coefficient component with different axial spacing

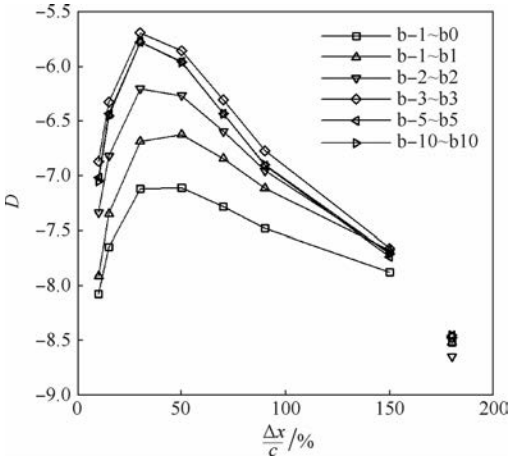


图 7　不同叶片数下气动阻尼系数

Fig. 7　Aerodynamic damping coefficient with different blade numbers

增加后减小,转子叶排在 $\Delta x/c = 30\%$ 时气动弹性稳定性最好,说明导叶-转子之间存在一个最佳的轴向间距,这和文献[12]中的结论一致,上游导叶对转子叶片表面非定常压力波的反射,导致转子叶片的气动阻尼系数随轴向间距改变发生变化;b -1和 b0 气动阻尼系数分量的叠加已经能基本预测导叶对转子气动阻尼系数的影响趋势,更多的叶片数只是使得转子气动阻尼系数误差更小。

表 3 为在叶片间相位角为 -72°时不同叶片数下转子气动阻尼系数相对误差。从表 3 可以看出:在 $\Delta x/c \geqslant 90\%$ 时,b0 及其相邻叶片就可以较为准确得出转子叶排气动阻尼系数;随着 Δx 的减小,导叶在转子叶片上引起的非定常气动响应增强,b-1 和 b0 气动阻尼系数分量相对于转子叶排气动阻尼系数所占的比重有所下降,表明导叶对非定常压力波的反射作用减弱了非定常压力波的周向衰减;在 $\Delta x/c < 90\%$ 时,b0 及其相邻叶片得出转子气动阻尼系数相对误差明显增大,需要测量更多的叶片才能得到较为准确的气动阻尼系数。为了进一步弄清楚轴向间距影响转子气动阻尼系数的机理,需要对 b0 和 b-1 叶片表面非定常气动力及其做功情况进行分析。

表 3　不同叶片数下气动阻尼系数相对误差

Table 3　Relative error of aerodynamic damping coefficients with different blade numbers　%

叶片编号	$\Delta x/c = 10\%$	$\Delta x/c = 15\%$	$\Delta x/c = 30\%$	$\Delta x/c = 50\%$	$\Delta x/c = 70\%$	$\Delta x/c = 90\%$	$\Delta x/c = 150\%$	单转子
b-1 ~ b0	14.58	18.50	23.31	19.25	13.24	8.40	2.21	0.77
b-1 ~ b1	12.27	13.84	15.77	11.08	6.39	3.09	0.37	0.76
b-2 ~ b2	4.08	5.64	7.42	5.10	2.51	0.81	0.13	2.30
b-3 ~ b3	2.57	1.99	1.44	1.76	1.97	1.84	0.63	0.24
b-5 ~ b5	0.44	0.34	0.03	0.09	0.01	0.11	0.34	0.02

3.3　b0 与 b-1 非定常压力和气动功分布

对 b0 与 b-1 来说,叶片表面非定常压力响应来源主要有两部分:一部分是由于 b0 振动引起的非定常压力响应,另一部分来源于上游导叶对转子叶片表面非定常压力波的反射。

图 8(a)为 b0 表面无量纲非定常气动功 W 的分布图,可以看出:振动叶片本身的吸力面(SS)前部和压力面(PS)做负功,吸力面后半部分做正功;做功最大位置发生在吸力面前缘;受导叶的影响,随 Δx 的减小,做功变化主要发生在吸力面前部和压力面。由于转子叶片表面气动功的大小是由叶片表面非定常气动力的一阶幅值和相位决定的[13],需要进一步分析 b0 表面非定常压力响应一阶幅值和相位变化情况。

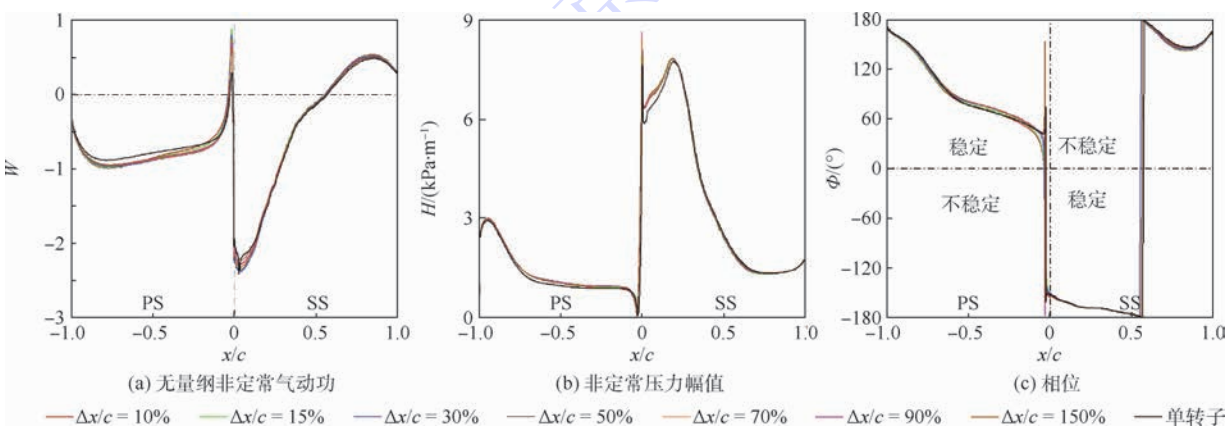


图 8　b0 表面无量纲非定常气动功、压力幅值和相位

Fig. 8　Non-dimensional unsteady aerodynamic work, pressure amplitude and phase on b0 surface

北航学报 赠阅

图8(b)、(c)为b0表面非定常压力响应一阶幅值 H 和相位 Φ 分布图,可以看出:b0一阶幅值在吸力面前部和吸力面后半部分较大,原因是在图1中b0振动必然导致b0与b-1、b1之间部分流道宽度发生周期性改变,通过计算发现b0的振动主要影响b-1、b0吸力面前半部分与b0、b1压力面后半部;幅值最大位置位于吸力面前部,原因是这里不仅为图3中压力梯度最大的区域而且是b0振动直接影响区域;随轴向间距的变化,导叶使b0压力面幅值和相位都有改变,改变主要集中在图2中转子受导叶影响区域(吸力面前部和压力面)。同时这些幅值和相位改变说明转子表面压力波在上游导叶发生反射,反射波与转子叶片表面压力波叠加,从而导致图8中b0表面气动功变化。

图9(a)为b-1表面无量纲非定常气动功分布图,可以看出:气动力做功主要集中在吸力面前部;在b-1吸力面前部,轴向间距引起叶片表面气动功的变化只占叶片表面气动功很小的一部分,说明振动引起的通道周期性变化对转子气动阻尼系数的影响远大于导叶对其的影响。结合图8(a)和图9(a)可以看出:b-1吸力面做的正功明显大于b0做的负功,从而使转子气动阻尼系数为负,并与图6相对应。

图9(b)、(c)为b-1表面非定常压力响应一阶幅值和相位分布图,可以看出:幅值较大区域集中在吸力面前部,原因是这部分不仅为压力梯度最大的区域而且是b0振动引起的通道周期性变化区域;随轴向间距的变化,导叶对压力面幅值和相位都有影响,而压力面由于不受叶片振动直接影响变化较大,说明相邻叶片振动引起的通道变化抑制了导叶对非定常压力波的反射作用。

对比图8(b)和图9(b)不难发现:b0与b-1吸力面幅值基本一致,但它们在图6气动阻尼系数分量却相差甚远,从相位图中不难发现是由于b0与b-1吸力面相位角的不同导致的。

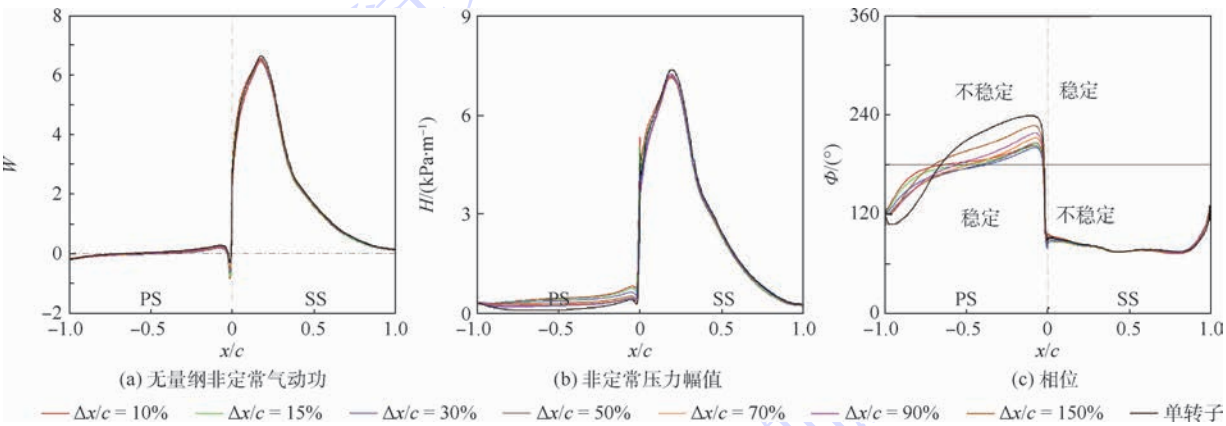


图9 b-1表面无量纲非定常气动功、压力幅值和相位

Fig. 9 Non-dimensional unsteady aerodynamic work, pressure amplitude and phase on b-1 surface

4 结 论

本文使用影响系数法研究了不同轴向间距下上游叶排对转子气动弹性稳定性的影响,得到以下结论:

1)导叶轴向间距对每个节径下的转子叶片气动阻尼的影响程度不同,主要影响范围在气动弹性最不稳定状态对应的节径附近,气动阻尼系数最小值处影响最大。随着轴向间距的减小,转子叶片气动阻尼系数最小值先增加后减小,并使得气动弹性最不稳定点对应的节径增大。

2)在协调叶栅中,叶片b1振动对b0影响 $C_f(1,0)$,对转子气动阻尼的大小起决定性作用,其影响甚至超过振动叶片b0本身的影响 $C_f(0,0)$。研究发现两叶片吸力面上的相位差异是造成这种影响结果的直接原因。

3)随着轴向间距的减小,导叶对非定常压力波的反射作用减弱了非定常压力波的周向衰减,这说明导叶的存在使得转子叶片振动的非定常影响范围明显增大。为了获得较为准确的转子叶片气动阻尼值,使用影响系数法时需要包括更多的叶片数。

4)在协调叶栅中,相邻叶片振动引起的通道变化对气动阻尼的影响抑制了导叶对非定常压力波的反射作用。

参考文献(References)

[1] 陈懋章. 风扇/压气机技术发展和对今后工作的建议[J]. 航空动力学报,2002,17(1):1-15.
CHEN M Z. Development of fan/compressor techniques and suggestions on further researches[J]. Journal of Aerospace

北航学报 赠阅

Power,2002,17(1): 1-15(in Chinese).

[2] LANE F. System mode shapes in the flutter of compressor blade rows[J]. Journal of the Aeronautical Sciences,1956,23(1): 54-66.

[3] CARTA F O,ST HILAIRE A O. Experimentally determined stability parameters of a subsonic cascade oscillating near stall [J]. Journal of Engineering for Power,1978,100(1): 111-120.

[4] CARTA F O,ST HILAIRE A O. Effect of inter-blade phase angle and incidence angle on cascade pitching stability[J]. Journal of Engineering for Power,1980,102(2):391-396.

[5] 张小伟,王延荣,徐可宁. 叶轮机械叶片颤振的影响参数[J]. 航空动力学报,2011,26(7):1558-1562.
ZHANG X W,WANG Y R,XU K N. Effects of parameters on blade flutter in turbomachinery[J]. Journal of Aerospace Power,2011,26(7):1558-1562(in Chinese).

[6] BELL D L,HE L. Three dimensional unstea-dy flow for an oscillating turbine blade and the influence of tip leakage[C]// ASME 1998 International Gas Turbine and Aeroengine Congress and Exhibition. New York: ASME,1998: V005T14A049.

[7] ISOMURA K,GILES M B. A numerical study of flutter in a transonic fan[J]. Journal of Turbomachinery,1998,120(3): 500-507.

[8] HSU K,HOYNIAK D,ANAND M S. Full annulus multi-row flutter analyses[C]// ASME Turbo Expo 2012: Turbine Technical Conference and Exposition. New York: ASME,2012: 1453-1462.

[9] CULVER R,LIU F. Mixing-plane method for flutter computation in multi-stage turbomachines: AIAA-2009-862[R]. Reston: AIAA,2009.

[10] HUANG X Q,HE L,BELL D L. Influence of upstream stator on rotor flutter stability in a low pressure steam turbine stage[J]. Proceedings of the Institution of Mechanical Engineers,Part A: Journal of Power and Energy,2006,220(1): 25-35.

[11] 杨慧,李振鹏. 转静干涉对转子叶片颤振特性的影响[J]. 北京航空航天大学学报,2016,42(2): 258-264.
YANG H,LI Z P. Influence of rotor-stator interaction on rotor blade flutter characteristics[J]. Journal of Beijing University of Aeronautics and Astronautics,2016,42(2):258-264(in Chinese).

[12] LI H D,HE L. Blade aerodynamic damping variation with rotor-stator gap: A computational study using single-passage approach [J]. Journal of Turbomachinery,2005,127(3):573-579.

[13] 杨慧,沈真. 上下游干涉对转子叶片颤振特性的影响[J]. 航空动力学报,2016,31(5):1170-1177.
YANG H,SHEN Z. Influence of up- and downstream blade rows on the rotor blade flutter characteristics[J]. Journal of Aerospace Power,2016,31(5):1170-1177(in Chinese).

[14] HANAMURA Y,TANAKA H,YAMAGUCHI K. A simplified method to measure unsteady forces acting on the vibrating blades in cascade[J]. Bulletin of the JSME,1980,23(180): 880-887.

[15] 杨慧,何力,王延荣. 压气机线性振荡叶栅气弹试验研究(二):叶尖间隙的影响[J]. 航空学报,2008,24(4): 804-810.
YANG H,HE L,WANG Y R. Experimental study on aeroelasticity in linear oscillating compressor cascade. Part Ⅱ: Tip-clearance effect[J]. Journal of Turbomachinery,2008,24(4):804-810(in Chinese).

[16] 郑赟. 基于非结构网格的气动弹性数值方法研究[J]. 航空动力学报,2009,24(9):2069-2077.
ZHENG Y. Computationalaeroelasticity with an unstructured grid method[J]. Journal of Turbomachinery,2009,24(9): 2069-2077(in Chinese).

[17] 杨慧,郑赟. 测量叶轮机振动叶片表面非定常气动响应的实验技术-影响系数法[J]. 实验流体力学,2011,25(3): 71-77.
YANG H,ZHENG Y. Experimental technique for measuring unsteady aerodynamics of vibrating blade in turbomachinery influence coefficient method[J]. Journal of Experiments in Fluid Mechanics,2011,25(3):71-77(in Chinese).

[18] BLCS A,FRANSSON T H. Aeroelasticity in turbomachines: Comparison of theoretical and experimental cascade results: LTT-CONF-1986-001[R]. Lausanne: Éole Polytechnique Fédérale de Lausanne,1986.

[19] FRANSSON T H,VERDON J M. Updated report on standard configurations for unsteady flow through vibrating axial flow turbomachine cascades: LTT-CONF-1991-005[R]. Lausanne: Éole Polytechnique Fédérale de Lausanne,1991.

作者简介:

郑赟　男,博士,讲师。主要研究方向:计算流体力学、叶轮机械气动弹性数值模拟。

余永博　男,硕士研究生。主要研究方向:叶轮机械气动弹性数值模拟。

北京航空航天大学学报

北航学报 赠阅

Influence mechanism of axial spacing on rotor blade flutter characteristics

ZHENG Yun[1,2,*], YU Yongbo[1,2]

(1. Collaborative Innovation Center for Advanced Aero-Engine, Beijing 100083, China;
2. School of Energy and Power Engineering, Beijing University of Aeronautics and Astronautics, Beijing 100083, China)

Abstract: The mechanism of the upstream blade row affecting the flutter characteristics of the rotor blade is studied, using the self-developed unsteady fluid-solid coupling numerical simulation program. The influence coefficient method is used to analyze the effect of axial spacing on the aeroelastic stability of rotor. The results show that the neighboring blade vibration, which adjacent to the certain blade's suction side, determines the value of rotor aerodynamic damping in the tuned cascade, and its effect is even larger than the certain blade vibration itself; in the multi-row environment, inlet guide vanes (IGV) mainly affects the minimum aerodynamic damping of rotor blade, and IGV increases the nodal diameter value of the most unstable point of aeroelasticity; the channel change caused by blade vibration restrains the reflection of IGV on unsteady pressure wave; with the decrease of axial spacing, the range of unsteady influence of blade vibration is obviously increased, since the effect of IGV on the unsteady pressure wave attenuates the circumferential attenuation of unsteady pressure wave; the influence coefficient method needs to measure more blades to obtain more accurate aerodynamic damping in the multi-row environment.

Keywords: axial spacing; influence coefficient method; aeroelastic stability; flutter; aerodynamic damping

北京航空航天大学学报

Received: 2017-04-25; **Accepted**: 2017-08-11; **Published online**: 2017-09-04 10:31
URL: kns.cnki.net/kcms/detail/11.2625.V.20170904.1031.001.html
* **Corresponding author**. E-mail: zheng_yun@buaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2017.0271

基于滚动时域优化的旋转弹解耦控制器设计

陈伟*, 孙传杰, 冯高鹏, 拜云山
(中国工程物理研究院 总体工程研究所, 绵阳 621900)

摘 要: 旋转弹在飞行过程中会受到外界干扰和不确定性的影响,并且存在气动交联、惯性交联和控制交联,为了实现稳定飞行,有必要进行解耦控制器设计。提出了一种基于滚动时域优化(RHO)的解耦控制方法。将旋转弹和舵机系统的动力学模型用状态空间形式描述,基于旋转弹、指令滤波器、积分跟踪误差的状态方程得到一个增广状态方程。采用基于指令滤波器的滚动时域优化方法进行控制量解算,根据系统输出与指令信号之间的差值实时调节控制器增益,实现旋转弹解耦控制。从加速度控制仿真结果可以看出,所设计的控制系统基本不受转速、建模误差和外界干扰的影响,具有较高的鲁棒性。

关 键 词: 旋转弹; 解耦控制; 滚动时域优化(RHO); 加速度控制; 指令滤波器
中图分类号: V249.1
文献标识码: A **文章编号**: 1001-5965(2018)04-0717-08

旋转弹在飞行过程中绕自身纵轴连续滚转,在无控条件下可有效降低气动不对称、结构不对称和推力偏心等扰动因素所带来的弹道散布,提高落点精度,在制导控制条件下可省掉滚转控制回路,简化控制系统组成,因而旋转体制为众多弹箭类武器系统所采用。但是,弹体的旋转也使旋转弹在空气动力学特性、飞行力学特性、控制理论与方法等方面明显有别于非旋转弹,并带来一些特殊问题,如马格努斯效应、陀螺效应等,使得旋转弹在俯仰和偏航方向的运动存在交叉耦合作用[1]。在旋转弹控制系统设计中,需要考虑各种耦合因素,不仅包括以马格努斯效应为表现形式的气动交联、以陀螺效应为表现形式的惯性交联,而且包括舵机系统延迟引起的控制交联。在复杂的飞行环境下弹体自身会受到各种不确定性因素的干扰,这都给旋转弹控制系统设计带来了较大挑战[2-3]。

相关学者在旋转弹控制系统设计领域开展了较为深入研究,并取得了一定的研究成果。文献[4-5]分析了双通道控制旋转弹的各种耦合特性,并采用前馈补偿解耦方法实现了旋转弹基于过载驾驶仪的静态解耦控制。文献[6]采取对角占优解耦控制方法设计了弹体动力学环节以及执行机构动力学环节的解耦控制器。上述中基于前馈补偿解耦和对角占优解耦方法需要对模型有较为精确的认识,而且所设计的补偿器并不能确保系统在全频域段具有较好的解耦效果,从而学者将现代控制系统设计方法应用于旋转弹控制系统设计。文献[7]采用动态逆方法进行姿态控制器设计,该方法可有效处理动力学中的非线性因素。文献[8-10]采用鲁棒 H_∞ 控制方法进行旋转弹控制器设计,具有良好的鲁棒性和自适应能力。由于滚动时域优化(RHO)方法,又被称为模型预测控制,对模型精度要求不高,且具有较好的解耦控制能力[11],其被学者应用于导弹控制领域[12-13],并取得了较好的控制效果。本文在文献[11]的

收稿日期: 2017-05-02; **录用日期**: 2017-06-09; **网络出版时间**: 2017-08-31 14:32
网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20170831.1432.006.html
基金项目: 国家自然科学基金(U143010033)
* **通信作者**. E-mail: chenweifenglaile@163.com

引用格式: 陈伟, 孙传杰, 冯高鹏, 等. 基于滚动时域优化的旋转弹解耦控制器设计[J]. 北京航空航天大学学报, 2018, 44(4): 717-724. CHEN W, SUN C J, FENG G P, et al. Design of decoupling controller for spinning missile based on receding horizon optimal [J]. Journal of Beijing University of Aeronautics and Astronautics, 2018, 44(4): 717-724 (in Chinese).

北航学报 赠阅

基础上,提出一种基于指令滤波器的 RHO 控制方法,并将其应用于旋转弹解耦控制,在不同的弹体旋转速度下取得了较好的解耦控制效果,确保了飞行指令的稳定跟踪。

1 旋转弹动力学模型

由于旋转弹在飞行过程中以一定角速度绕其纵轴连续滚转,为了便于分析,引入准弹体坐标系,建立旋转弹动力学模型。准弹体坐标系中旋转弹的质心运动动力学方程组和导弹绕质心转动的动力学方程组分别为[14]

$$\begin{cases} m\dfrac{\mathrm{d}V}{\mathrm{d}t}=P\cos\beta\cos\alpha-X-mg\cos\psi_{\mathrm{V}}\sin\theta \\ mV\cos\psi_{\mathrm{V}}\dfrac{\mathrm{d}\theta}{\mathrm{d}t}=P(\cos\gamma_{\mathrm{V}}^{*}\cos\beta\sin\alpha+ \\ \quad \sin\gamma_{\mathrm{V}}^{*}\sin\beta)+(Y-Y_{\mathrm{m}})\cos\gamma_{\mathrm{V}}^{*}- \\ \quad (Z-Z_{\mathrm{m}})\sin\gamma_{\mathrm{V}}^{*}-mg\cos\theta \\ -mV\dfrac{\mathrm{d}\psi_{\mathrm{V}}}{\mathrm{d}t}=P(\sin\gamma_{\mathrm{V}}^{*}\cos\beta\sin\alpha-\cos\gamma_{\mathrm{V}}^{*}\sin\beta)+ \\ \quad (Y-Y_{\mathrm{m}})\sin\gamma_{\mathrm{V}}^{*}+(Z-Z_{\mathrm{m}})\cos\gamma_{\mathrm{V}}^{*}- \\ \quad mg\sin\psi_{\mathrm{V}}\sin\theta \end{cases} \tag{1}$$

$$\begin{cases} J_x\dfrac{\mathrm{d}\omega_x}{\mathrm{d}t}=M_x^{\sigma_x}\sigma_x+M_x^{\omega_x}\omega_x \\ J_y\dfrac{\mathrm{d}\omega_y}{\mathrm{d}t}=M_y^{\beta}\beta+M_y^{\sigma_y}\sigma_y+M_y^{\omega_y}\omega_y+M_y''\dfrac{\dot{\gamma}D}{V}\alpha- \\ \quad (J_x-J_z)\omega_x\omega_z-J_z\omega_z\dot{\gamma} \\ J_z\dfrac{\mathrm{d}\omega_z}{\mathrm{d}t}=M_z^{\alpha}\alpha+M_z^{\sigma_z}\sigma_z+M_z^{\omega_z}\omega_z+M_z''\dfrac{\dot{\gamma}D}{V}\beta- \\ \quad (J_y-J_x)\omega_y\omega_x+J_y\omega_y\dot{\gamma} \end{cases} \tag{2}$$

式中:m 为旋转弹质量;g 为重力加速度;V 为飞行速度;$\dot{\gamma}$ 为旋转速度;D 为弹体直径;P 为推力;X 为阻力;Y 为升力;Z 为侧向力;Y_{m} 为沿着升力方向的马格努斯力;Z_{m} 为沿着侧向力方向的马格努斯力;α 为迎角;β 为侧滑角;γ_{V}^{*} 为速度准倾斜角;θ 为弹道倾角;ψ_{V} 为弹道偏角;ω_x、ω_y、ω_z 为旋转弹转动角速度;J_x、J_y、J_z 为旋转弹转动惯量;$M_x^{\sigma_x}$ 为滚转力矩对尾翼安装角的偏导数;$M_x^{\omega_x}$ 为滚转力矩对弹体绕纵轴旋转角速度的偏导数;M_y^{β} 为偏航力矩对侧滑角 β 的一阶偏导数;$M_y^{\sigma_y}$ 为偏航力矩对偏航舵偏角 σ_y 的一阶偏导数;$M_y^{\omega_y}$ 为偏航力矩对弹体转动角速度 ω_y 的一阶偏导数;M_y'' 为偏航力矩对弹体无因次旋转角速度 $\dot{\gamma}D/V$ 和迎角 α 的二阶偏导数;M_z^{α} 为俯仰力矩对迎角 α 的一阶偏导数;$M_z^{\sigma_z}$ 为俯仰力矩对俯仰舵偏角 σ_z 的一阶偏导数;$M_z^{\omega_z}$ 为俯仰力矩对弹体转动角速度 ω_z 的一阶偏导数;M_z'' 为俯仰力矩对弹体无因次旋转角速度 $\dot{\gamma}D/V$ 和侧滑角 β 的二阶偏导数。

为了简化控制系统设计,需要对弹体动力学方程组(1)和(2)进行简化,不失一般性,作如下假设:

1)假设在一小段飞行过程中,弹体的速度不变,弹体的转速不变。

2)假设在控制系统设计的相关研究中,忽略重力作用。

3)假设迎角 α、侧滑角 β、偏航角 ψ、弹道偏角 ψ_{V}、偏航角微分 $\dot{\psi}$ 和俯仰角微分 $\dot{\vartheta}$ 为小量,它们之间的乘积为零。

4)假设控制舵产生的升力和弹体受到的总升力相比是小量。

当重点考虑弹体短周期运动时,可假定速度的方向不变,而只考虑弹轴的摆动运动,即假定弹道倾角 θ 和弹道偏角 ψ_{V} 在弹体运动短周期内基本保持不变,可近似为零,并根据小角度假设条件,有如下近似关系式成立[1]:

$$\begin{cases} \vartheta\approx\alpha \\ \psi\approx\beta \\ \gamma_{\mathrm{V}}^{*}\approx0 \end{cases} \tag{3}$$

$$\begin{cases} \omega_x\approx\dot{\gamma} \\ \omega_y\approx\dot{\psi} \\ \omega_z\approx\dot{\vartheta} \end{cases} \tag{4}$$

将式(3)和式(4)代入式(1)和式(2)整理化简后可得如下所示旋转弹动力学模型:

$$\begin{cases} \dot{\vartheta}=\omega_z \\ \dot{\psi}=\omega_y \\ \dot{\omega}_z=\dfrac{M_z^{\alpha}}{J_z}\vartheta+\dfrac{M_z''}{J_z}\cdot\dfrac{\dot{\gamma}D}{V}\psi+\dfrac{M_z^{\omega_z}}{J_z}\omega_z+ \\ \quad \dfrac{J_x}{J_z}\dot{\gamma}\omega_y+\dfrac{M_z^{\sigma_z}}{J_z}\sigma_z \\ \dot{\omega}_y=\dfrac{M_y''}{J_y}\cdot\dfrac{\dot{\gamma}D}{V}\vartheta+\dfrac{M_y^{\beta}}{J_y}\psi-\dfrac{J_x}{J_y}\dot{\gamma}\omega_z+ \\ \quad \dfrac{M_y^{\omega_y}}{J_y}\omega_y+\dfrac{M_y^{\sigma_y}}{J_y}\sigma_y \end{cases} \tag{5}$$

对于轴对称弹体,有 $J_y=J_z$,$M_z^{\alpha}=M_y^{\beta}$,$M_z''=M_y''$,$M_z^{\omega_z}=M_y^{\omega_y}$,$M_z^{\sigma_z}=M_y^{\sigma_y}$。

令法向加速度 a_y、侧向加速度 a_z 为系统输出,根据法向加速度和侧向加速度的定义有

北航学报 赠阅

$$\begin{cases} a_y = V\left(\dfrac{P + qSc_y^{\alpha}}{mV}\alpha - \dfrac{qSc''_{ym}}{mV}\cdot\dfrac{\dot{\gamma}D}{V}\beta\right) \\ a_z = -V\left(\dfrac{qSc''_{zm}}{mV}\cdot\dfrac{\dot{\gamma}D}{V}\alpha + \dfrac{P + qSc_z^{\beta}}{mV}\beta\right) \end{cases} \tag{6}$$

式中：q 为动压；S 为机翼参考面积；c_y^{α} 为升力系数对迎角 α 的导数；c''_{ym} 为升力系数对弹体无因次旋转角速度 $\dot{\gamma}D/V$ 和侧滑角 β 的二阶偏导数；c''_{zm} 为侧力系数对弹体无因次旋转角速度 $\dot{\gamma}D/V$ 和迎角 α 的二阶偏导数；c_z^{β} 为侧力系数对侧滑角 β 的导数。

根据式(3)，a_y 和 a_z 可由式(7)近似得到

$$\begin{cases} a_y = \dfrac{P + qSc_y^{\alpha}}{m}\vartheta - \dfrac{qSc''_{ym}}{m}\cdot\dfrac{\dot{\gamma}D}{V}\psi \\ a_z = -\dfrac{qSc''_{zm}}{m}\cdot\dfrac{\dot{\gamma}D}{V}\vartheta - \dfrac{P + qSc_z^{\beta}}{m}\psi \end{cases} \tag{7}$$

考虑建模误差和外界干扰，将式(5)和式(7)写成如下状态空间形式：

$$\begin{cases} \dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{\sigma} + \boldsymbol{b}_0 \\ \boldsymbol{y} = \boldsymbol{C}\boldsymbol{x} \end{cases} \tag{8}$$

式中：$\boldsymbol{x} = [\vartheta \quad \psi \quad \omega_z \quad \omega_y]^{\mathrm{T}}$ 为旋转弹状态向量；$\boldsymbol{y}$ 为旋转弹测量输出；$\boldsymbol{\sigma} = [\sigma_z \quad \sigma_y]^{\mathrm{T}}$ 为控制输入；$\boldsymbol{b}_0$ 为模型不确定项；$\boldsymbol{A}$、$\boldsymbol{B}$、$\boldsymbol{C}$ 的具体表达式为

$$\boldsymbol{A} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ \dfrac{M_z^{\alpha}}{J_z} & \dfrac{M''_z}{J_z}\cdot\dfrac{\dot{\gamma}D}{V} & \dfrac{M_z^{\omega_z}}{J_z} & \dfrac{J_x}{J_z}\dot{\gamma} \\ \dfrac{M''_y}{J_y}\cdot\dfrac{\dot{\gamma}D}{V} & \dfrac{M_y^{\beta}}{J_y} & -\dfrac{J_x}{J_y}\dot{\gamma} & \dfrac{M_y^{\omega_y}}{J_y} \end{bmatrix}$$

$$\boldsymbol{B} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ \dfrac{M_z^{\sigma_z}}{J_z} & 0 \\ 0 & \dfrac{M_y^{\sigma_y}}{J_y} \end{bmatrix}$$

$$\boldsymbol{C} = \begin{bmatrix} \dfrac{P + qSc_y^{\alpha}}{m} & -\dfrac{qSc''_{ym}}{m}\cdot\dfrac{\dot{\gamma}D}{V} & 0 & 0 \\ -\dfrac{qSc''_{zm}}{m}\cdot\dfrac{\dot{\gamma}D}{V} & -\dfrac{P + qSc_z^{\beta}}{m} & 0 & 0 \end{bmatrix}$$

2　旋转条件下舵机系统动力学模型

舵机系统的指令执行框图如图 1 所示。

弹体的控制指令 σ_{cy}、σ_{cz} 形成于非旋转的准弹体坐标系下，根据传感器测量得到此时弹体相对于准弹体坐标系的滚转角 γ_d，将 σ_{cz}、σ_{cy} 进行分解，得到弹体坐标系下的控制指令 σ_{c1}、σ_{c2}，σ_{c1}、σ_{c2} 经舵机驱动舵面偏转，从而得到弹体坐标系下

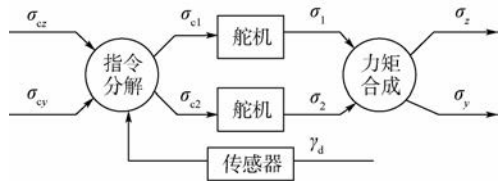


图 1　舵机系统的指令执行框图

Fig. 1　Command execution block diagram of servo system

舵面偏转角 σ_1、σ_2 和控制力矩，最后再将 σ_1、σ_2 和控制力矩合成到准弹体坐标系下的舵机响应 σ_z、σ_y 和控制力矩。

舵机响应传递函数为

$$G_d(s) = \frac{k_s}{T_s^2 s^2 + 2\mu_s T_s s + 1} \tag{9}$$

式中：k_s 为舵机系统增益；T_s 为舵机系统时间常数；μ_s 为舵机系统阻尼。

根据图 1 所示的指令执行框图和舵机响应传递函数，可得舵机系统的传递函数矩阵为[1]

$$\boldsymbol{G}_{\sigma}(s) = k_s\begin{bmatrix} \cos(\tau\dot{\gamma}) & \sin(\tau\dot{\gamma}) \\ -\sin(\tau\dot{\gamma}) & \cos(\tau\dot{\gamma}) \end{bmatrix}\cdot \begin{bmatrix} G_s(s) & G_{sco}(s) \\ -G_{sco}(s) & G_s(s) \end{bmatrix} \tag{10}$$

式中：τ 为从滚转角 γ 的测量到控制指令的生成并分解输入到实际舵机系统所需的时间；$G_s(s)$ 和 $G_{sco}(s)$ 分别为前向通道传递函数和耦合通道传递函数，其表达式分别为

$$G_s(s) = (T_s^2 s^2 + 2\mu_s T_s s + 1 - T_s^2\dot{\gamma}^2)/[(T_s^2 s^2 + 2\mu_s T_s s + 1 - T_s^2\dot{\gamma}^2)^2 + (2T_s^2\dot{\gamma}s + 2\mu_s T_s\dot{\gamma})^2]$$

$$G_{sco}(s) = (2T_s^2\dot{\gamma}s + 2\mu_s T_s\dot{\gamma})/[(T_s^2 s^2 + 2\mu_s T_s s + 1 - T_s^2\dot{\gamma}^2)^2 + (2T_s^2\dot{\gamma}s + 2\mu_s T_s\dot{\gamma})^2]$$

可以看出，旋转弹旋转导致控制通道出现耦合，随着转速 $\dot{\gamma}$ 的提高，前向通道传递函数 $G_s(s)$ 的增益将减小，耦合通道传递函数 $G_{sco}(s)$ 的增益将增大，控制通道耦合加剧。

当忽略转速变化且仅考虑舵机系统在常值指令输入下的稳态输出时，则准弹体系中舵机系统的动力学模型可简化为

$$\boldsymbol{\sigma} = k_s k_r\begin{bmatrix} \cos\gamma_d & \sin\gamma_d \\ -\sin\gamma_d & \cos\gamma_d \end{bmatrix}\boldsymbol{\sigma}_c \tag{11}$$

式中：$\boldsymbol{\sigma}_c = [\sigma_{cz} \quad \sigma_{cy}]^{\mathrm{T}}$ 为准弹体坐标系中控制指令；k_r 表征了舵机系统由于弹体滚转出现的幅值衰减特性；$\gamma_d = \tau\dot{\gamma} + \gamma_c$ 为舵机系统总延迟角，γ_c 为由于舵机响应滞后造成的动态相位滞后角度。k_r、γ_c 的表达式如下所示，具体推导过程见文献[1]。

北航学报 赠阅

$$k_r=\frac{1}{\sqrt{(1-T_s^2\dot{\gamma}^2)^2+(2\mu_s T_s\dot{\gamma})^2}}$$

$$\gamma_c=\arccos\frac{1-T_s^2\dot{\gamma}^2}{\sqrt{(1-T_s^2\dot{\gamma}^2)^2+(2\mu_s T_s\dot{\gamma})^2}}$$

将式(11)代入式(8)可得考虑舵机系统动态特性的旋转弹动力学模型

$$\begin{cases}\dot{\boldsymbol{x}}=\boldsymbol{A}\boldsymbol{x}+\boldsymbol{B}_c\boldsymbol{\sigma}_c+\boldsymbol{b}_0\\ \boldsymbol{y}=\boldsymbol{C}\boldsymbol{x}\end{cases}\tag{12}$$

式中:$\boldsymbol{B}_c$ 表达式为

$$\boldsymbol{B}_c=k_s k_r\boldsymbol{B}\begin{bmatrix}\cos\gamma_d & \sin\gamma_d\\ -\sin\gamma_d & \cos\gamma_d\end{bmatrix}$$

3 解耦控制器设计

采用以下指令滤波器[15]得到加速度指令信号:

$$\begin{cases}\dfrac{a_{yc}}{a_{yc0}}=\dfrac{\omega_{ay}^2}{s^2+2\xi_{ay}\omega_{ay}s+\omega_{ay}^2}\\ \dfrac{a_{zc}}{a_{zc0}}=\dfrac{\omega_{az}^2}{s^2+2\xi_{az}\omega_{az}s+\omega_{az}^2}\end{cases}\tag{13}$$

式中:a_{yc}和 a_{zc}分别为法向加速度指令和侧向加速度指令;a_{yc0}和 a_{zc0}分别为法向加速度指令滤波器和侧向加速度指令滤波器的输入信号;ξ_{ay}和 ξ_{az}分别为法向加速度指令滤波器和侧向加速度指令滤波器的阻尼;ω_{ay}和 ω_{az}分别为法向加速度指令滤波器和侧向加速度指令滤波器的自然频率。

将指令滤波器式(13)写成状态空间的形式

$$\begin{cases}\dot{\boldsymbol{x}}_r=\boldsymbol{A}_r\boldsymbol{x}_r+\boldsymbol{B}_r\boldsymbol{r}_{c0}\\ \boldsymbol{y}_r=\boldsymbol{C}_r\boldsymbol{x}_r\end{cases}\tag{14}$$

式中:指令滤波器状态向量 $\boldsymbol{x}_r=[a_{yc}\quad\dot{a}_{yc}\quad a_{zc}\quad\dot{a}_{zc}]^{\mathrm{T}}$;指令滤波器输出 $\boldsymbol{y}_r=[a_{yc}\quad a_{zc}]^{\mathrm{T}}$;指令滤波器输入 $\boldsymbol{r}_{c0}=[a_{yc0}\quad a_{zc0}]^{\mathrm{T}}$;指令滤波器状态矩阵 $\boldsymbol{A}_r$、控制矩阵 $\boldsymbol{B}_r$、输出矩阵 $\boldsymbol{C}_r$ 的表达式为

$$\boldsymbol{A}_r=\begin{bmatrix}-2\xi_{ay}\omega_{ay} & 1 & 0 & 0\\ -\omega_{ay}^2 & 0 & 0 & 0\\ 0 & 0 & -2\xi_{az}\omega_{az} & 1\\ 0 & 0 & -\omega_{az}^2 & 0\end{bmatrix}$$

$$\boldsymbol{B}_r=\begin{bmatrix}0 & 0\\ \omega_{ay}^2 & 0\\ 0 & 0\\ 0 & \omega_{az}^2\end{bmatrix}$$

$$\boldsymbol{C}_r=\begin{bmatrix}1 & 0 & 0 & 0\\ 0 & 0 & 1 & 0\end{bmatrix}$$

RHO 优化控制的性能指标如下所示[16]:

$$\min_{\boldsymbol{\sigma}_c}J_R=\frac{1}{2}\int_{t_0}^{t_f}[(\boldsymbol{y}_r-\boldsymbol{y})^{\mathrm{T}}\boldsymbol{Q}_p(\boldsymbol{y}_r-\boldsymbol{y})+\boldsymbol{y}_I^{\mathrm{T}}\boldsymbol{Q}_I\boldsymbol{y}_I+\boldsymbol{\sigma}_c^{\mathrm{T}}\boldsymbol{R}_r\boldsymbol{\sigma}_c]\,\mathrm{d}t\tag{15}$$

式中:t_f 和 t_0 分别为区间的上界和下界;$\boldsymbol{Q}_p$ 和 $\boldsymbol{Q}_I$ 分别为跟踪误差和跟踪误差积分的加权阵;$\boldsymbol{R}_r$ 为控制量的加权阵;$\boldsymbol{y}_I$ 为积分误差,可由下式得到

$$\begin{cases}\dot{\boldsymbol{x}}_I=\boldsymbol{y}_r-\boldsymbol{y}\\ \boldsymbol{y}_I=\boldsymbol{x}_I\end{cases}\tag{16}$$

综合旋转弹线性状态方程、指令滤波器状态方程和积分误差方程得到增广状态方程为

$$\begin{bmatrix}\dot{\boldsymbol{x}}_r\\ \dot{\boldsymbol{x}}\\ \dot{\boldsymbol{x}}_I\end{bmatrix}=\begin{bmatrix}\boldsymbol{A}_r & \boldsymbol{0} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{A} & \boldsymbol{0}\\ \boldsymbol{C}_r & -\boldsymbol{C} & \boldsymbol{0}\end{bmatrix}\begin{bmatrix}\boldsymbol{x}_r\\ \boldsymbol{x}\\ \boldsymbol{x}_I\end{bmatrix}+\begin{bmatrix}\boldsymbol{0}\\ \boldsymbol{B}_c\\ \boldsymbol{0}\end{bmatrix}\boldsymbol{\sigma}_c+\begin{bmatrix}\boldsymbol{B}_r\boldsymbol{r}_{c0}\\ \boldsymbol{0}\\ \boldsymbol{0}\end{bmatrix}+\begin{bmatrix}\boldsymbol{0}\\ \boldsymbol{b}_0\\ \boldsymbol{0}\end{bmatrix}\tag{17}$$

将式(17)写成对应的如下形式:

$$\dot{\boldsymbol{x}}_R=\boldsymbol{A}_R\boldsymbol{x}_R+\boldsymbol{B}_R\boldsymbol{\sigma}_c+\boldsymbol{b}_R+\boldsymbol{b}\tag{18}$$

式中:$\boldsymbol{A}_R$、$\boldsymbol{B}_R$、$\boldsymbol{b}_R$、$\boldsymbol{b}$ 表达式分别为

$$\boldsymbol{A}_R=\begin{bmatrix}\boldsymbol{A}_r & \boldsymbol{0} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{A} & \boldsymbol{0}\\ \boldsymbol{C}_r & -\boldsymbol{C} & \boldsymbol{0}\end{bmatrix}$$

$$\boldsymbol{B}_R=\begin{bmatrix}\boldsymbol{0}\\ \boldsymbol{B}_c\\ \boldsymbol{0}\end{bmatrix}$$

$$\boldsymbol{b}_R=\begin{bmatrix}\boldsymbol{B}_r\boldsymbol{r}_{c0}\\ \boldsymbol{0}\\ \boldsymbol{0}\end{bmatrix}$$

$$\boldsymbol{b}=\begin{bmatrix}\boldsymbol{0}\\ \boldsymbol{b}_0\\ \boldsymbol{0}\end{bmatrix}$$

基于增广状态方程式(18)将性能指标式(15)重新写成

$$\min_{\boldsymbol{\sigma}_c}J_R=\frac{1}{2}(\boldsymbol{x}_R^{\mathrm{T}}\boldsymbol{Q}_R\boldsymbol{x}_R+\boldsymbol{\sigma}_c^{\mathrm{T}}\boldsymbol{R}_r\boldsymbol{\sigma}_c)\,\mathrm{d}t\tag{19}$$

式中:

$$\boldsymbol{Q}_R=\begin{bmatrix}\boldsymbol{C}_r^{\mathrm{T}}\boldsymbol{Q}_p\boldsymbol{C}_r & -\boldsymbol{C}_r^{\mathrm{T}}\boldsymbol{Q}_p\boldsymbol{C} & \boldsymbol{0}\\ -\boldsymbol{C}^{\mathrm{T}}\boldsymbol{Q}_p\boldsymbol{C}_r & \boldsymbol{C}^{\mathrm{T}}\boldsymbol{Q}_p\boldsymbol{C} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{Q}_I\end{bmatrix}$$

对控制量进行设计,先构造如下 Hamilton 函数[17]:

$$H(\boldsymbol{x}_R,\boldsymbol{\sigma}_c,\boldsymbol{\lambda}_R,t)=\frac{1}{2}\boldsymbol{x}_R^{\mathrm{T}}\boldsymbol{Q}_R\boldsymbol{x}_R+\frac{1}{2}\boldsymbol{\sigma}_c^{\mathrm{T}}\boldsymbol{R}_r\boldsymbol{\sigma}_c+(\boldsymbol{A}_R\boldsymbol{x}_R+\boldsymbol{B}_R\boldsymbol{\sigma}_c+\boldsymbol{b}_R+\boldsymbol{b})^{\mathrm{T}}\boldsymbol{\lambda}_R\tag{20}$$

式中:$\boldsymbol{\lambda}_R$ 为时变拉格朗日乘子向量。

北航学报 赠阅

$\boldsymbol{\sigma}_c$ 使得 Hamilton 函数值最小的必要条件为 $\partial H/\partial \boldsymbol{\sigma}_c = 0$，有

$$\frac{\partial H}{\partial \boldsymbol{\sigma}_c}(\boldsymbol{x}_R, \boldsymbol{\sigma}_c, \boldsymbol{\lambda}_R, t) = \boldsymbol{R}_r \boldsymbol{\sigma}_c + \boldsymbol{B}_R^T \boldsymbol{\lambda}_R \tag{21}$$

由于

$$\frac{\partial^2 H}{\partial \boldsymbol{\sigma}_c^2}(\boldsymbol{x}_R, \boldsymbol{\sigma}_c, \boldsymbol{\lambda}_R, t) = \boldsymbol{R}_r \tag{22}$$

式中：$\boldsymbol{R}_r$ 为正定矩阵；$\boldsymbol{H}$ 为 $\boldsymbol{\sigma}_c$ 的二次型，则满足式(21)的 $\boldsymbol{\sigma}_c$ 可使 H 值最小。求式(21)的解为

$$\boldsymbol{\sigma}_c = -\boldsymbol{R}_r^{-1} \boldsymbol{B}_R^T \boldsymbol{\lambda}_R \tag{23}$$

将时变拉格朗日乘子向量 $\boldsymbol{\lambda}_R$ 设计成状态 $\boldsymbol{x}_R$ 的线性组合，有

$$\boldsymbol{\lambda}_R = \boldsymbol{K}_R \boldsymbol{x}_R + \boldsymbol{k}_R \tag{24}$$

式中：$\boldsymbol{K}_R$ 和 $\boldsymbol{k}_R$ 为参数变量。

对于时变拉格朗日乘子向量 λ_R 有如下等式成立[18]：

$$-\dot{\boldsymbol{\lambda}}_R = \frac{\partial}{\partial \boldsymbol{x}_R}\left(\frac{1}{2}\boldsymbol{x}_R^T \boldsymbol{Q}_R \boldsymbol{x}_R + \frac{1}{2}\boldsymbol{\sigma}_c^T \boldsymbol{R}_r \boldsymbol{\sigma}_c\right) + \frac{\partial}{\partial \boldsymbol{x}_R}(\boldsymbol{A}_R \boldsymbol{x}_R + \boldsymbol{B}_R \boldsymbol{\sigma}_c + \boldsymbol{b}_R + \boldsymbol{b})^T \boldsymbol{\lambda}_R \tag{25}$$

整理式(25)可得

$$\dot{\boldsymbol{\lambda}}_R = -\boldsymbol{Q}_R \boldsymbol{x}_R - \boldsymbol{A}_R^T \boldsymbol{\lambda}_R \tag{26}$$

将式(24)两边对时间 t 求导得

$$\dot{\boldsymbol{\lambda}}_R = \dot{\boldsymbol{K}}_R \boldsymbol{x}_R + \boldsymbol{K}_R(\boldsymbol{A}_R \boldsymbol{x}_R + \boldsymbol{B}_R \boldsymbol{\sigma}_c + \boldsymbol{b}_R + \boldsymbol{b}) + \dot{\boldsymbol{k}}_R \tag{27}$$

将式(23)代入式(27)得

$$\dot{\boldsymbol{\lambda}}_R = \dot{\boldsymbol{K}}_R \boldsymbol{x}_R + \boldsymbol{K}_R \boldsymbol{A}_R \boldsymbol{x}_R - \boldsymbol{K}_R \boldsymbol{B}_R \boldsymbol{R}_r^{-1} \boldsymbol{B}_R^T \boldsymbol{K}_R \boldsymbol{x}_R - \boldsymbol{K}_R \boldsymbol{B}_R \boldsymbol{R}_r^{-1} \boldsymbol{B}_R^T \boldsymbol{k}_R + \boldsymbol{K}_R(\boldsymbol{b}_R + \boldsymbol{b}) + \dot{\boldsymbol{k}}_R \tag{28}$$

根据式(26)和式(28)得

$$-\boldsymbol{Q}_R \boldsymbol{x}_R - \boldsymbol{A}_R^T(\boldsymbol{K}_R \boldsymbol{x}_R + \boldsymbol{k}_R) = \dot{\boldsymbol{K}}_R \boldsymbol{x}_R + \boldsymbol{K}_R \boldsymbol{A}_R \boldsymbol{x}_R + \dot{\boldsymbol{k}}_R - \boldsymbol{K}_R \boldsymbol{B}_R \boldsymbol{R}_r^{-1} \boldsymbol{B}_R^T \boldsymbol{K}_R \boldsymbol{x}_R - \boldsymbol{K}_R \boldsymbol{B}_R \boldsymbol{R}_r^{-1} \boldsymbol{B}_R^T \boldsymbol{k}_R + \boldsymbol{K}_R(\boldsymbol{b}_R + \boldsymbol{b}) \tag{29}$$

整理式(29)得

$$(\dot{\boldsymbol{K}}_R + \boldsymbol{K}_R \boldsymbol{A}_R + \boldsymbol{A}_R^T \boldsymbol{K}_R - \boldsymbol{K}_R \boldsymbol{B}_R \boldsymbol{R}_r^{-1} \boldsymbol{B}_R^T \boldsymbol{K}_R + \boldsymbol{Q}_R)\boldsymbol{x}_R + \dot{\boldsymbol{k}}_R + \boldsymbol{A}_R^T \boldsymbol{k}_R - \boldsymbol{K}_R \boldsymbol{B}_R \boldsymbol{R}_r^{-1} \boldsymbol{B}_R^T \boldsymbol{k}_R + \boldsymbol{K}_R(\boldsymbol{b}_R + \boldsymbol{b}) = \boldsymbol{0} \tag{30}$$

由式(30)对于任意的状态向量 $\boldsymbol{x}_R$ 都成立，则可以得到如下 2 个黎卡提微分方程：

$$\dot{\boldsymbol{K}}_R = -\boldsymbol{K}_R \boldsymbol{A}_R - \boldsymbol{A}_R^T \boldsymbol{K}_R + \boldsymbol{K}_R \boldsymbol{B}_R \boldsymbol{R}_r^{-1} \boldsymbol{B}_R^T \boldsymbol{K}_R - \boldsymbol{Q}_R \tag{31}$$

$$\dot{\boldsymbol{k}}_R = -\boldsymbol{K}_R(\boldsymbol{b}_R + \boldsymbol{b}) - \boldsymbol{A}_R^T \boldsymbol{k}_R + \boldsymbol{K}_R \boldsymbol{B}_R \boldsymbol{R}_r^{-1} \boldsymbol{B}_R^T \boldsymbol{k}_R \tag{32}$$

通过求解式(31)和式(32)所示的黎卡提微分方程得到参数变量 $\boldsymbol{K}_R$ 和 $\boldsymbol{k}_R$。由于式(32)中模型不确定项 $\boldsymbol{b}$ 未知，且 RHO 方法对模型精度要求不高，可用以下黎卡提微分方程近似计算 $\boldsymbol{k}_R$[11]：

$$\dot{\boldsymbol{k}}_R = -\boldsymbol{K}_R \boldsymbol{b}_R - \boldsymbol{A}_R^T \boldsymbol{k}_R + \boldsymbol{K}_R \boldsymbol{B}_R \boldsymbol{R}_r^{-1} \boldsymbol{B}_R^T \boldsymbol{k}_R \tag{33}$$

由于 $\boldsymbol{K}_R$ 和 $\boldsymbol{k}_R$ 在有限时域区间[t_0　t_f]边界 t_f 可以稳定在任意一个常值上，认为 $\boldsymbol{K}_R$ 和 $\boldsymbol{k}_R$ 在时域边界 t_f 为 0。在有限时域内对式(31)和式(33)进行反向积分，得到 $\boldsymbol{K}_R(t_0)$ 和 $\boldsymbol{k}_R(t_0)$[11]。通过式(34)计算得到控制器增益

$$[\boldsymbol{F}_R \quad \boldsymbol{f}_R] = -\boldsymbol{R}_r^{-1} \boldsymbol{B}_R^T [\boldsymbol{K}_R(t_0) \quad \boldsymbol{k}_R(t_0)] \tag{34}$$

式中：$\boldsymbol{F}_R = [\boldsymbol{F}_r \quad \boldsymbol{F}_x \quad \boldsymbol{F}_I]$ 为控制器增益，$\boldsymbol{F}_r$、$\boldsymbol{F}_x$ 和 $\boldsymbol{F}_I$ 分别为对应指令滤波器状态、旋转弹状态和积分误差状态的控制器增益；$\boldsymbol{f}_R$ 为针对建模误差的补偿控制器增益。

基于 RHO 控制器结构如图 2 所示。通过式(35)计算得到控制量

$$\boldsymbol{\sigma}_c = \boldsymbol{F}_R \boldsymbol{x}_R + \boldsymbol{f}_R \tag{35}$$

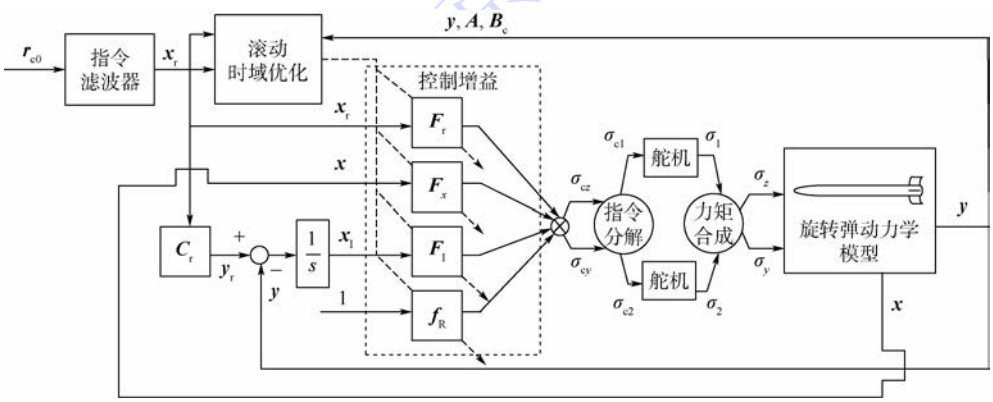


图 2　基于 RHO 的控制系统结构

Fig. 2　Structure of control system based on RHO

北航学报 赠阅

4 仿真分析

为了验证控制系统的鲁棒性，在原有气动参数基础上加上15%作为建模误差，并在旋转弹准弹体坐标俯仰通道施加 $30\sin(2\pi t)$ N·m 的干扰信号。指令滤波器参数为

$$\boldsymbol{A}_{\mathrm{r}}=\begin{bmatrix}-20 & 1 & 0 & 0\\ -100 & 0 & 0 & 0\\ 0 & 0 & -20 & 1\\ 0 & 0 & -100 & 0\end{bmatrix}$$

$$\boldsymbol{B}_{\mathrm{r}}=\begin{bmatrix}0 & 0\\ 100 & 0\\ 0 & 0\\ 0 & 100\end{bmatrix}$$

$$\boldsymbol{C}_{\mathrm{r}}=\begin{bmatrix}1 & 0 & 0 & 0\\ 0 & 0 & 1 & 0\end{bmatrix}$$

RHO 控制器设计参数为

$$\boldsymbol{Q}_{\mathrm{p}}=\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$$

$$\boldsymbol{Q}_{\mathrm{I}}=\begin{bmatrix}100 & 0\\ 0 & 100\end{bmatrix}$$

$$\boldsymbol{R}_{\mathrm{r}}=\begin{bmatrix}0.1 & 0\\ 0 & 0.1\end{bmatrix}$$

旋转弹质量 $m=465$ kg，机翼参考面积 $S=0.0709\ \mathrm{m}^2$，弹体直径 $D=0.3$ m，舵机系统增益 $k_{\mathrm{s}}=10$；舵机系统时间常数 $T_{\mathrm{s}}=0.016$；舵机系统阻尼 $\mu_{\mathrm{s}}=0.5$；飞行速度 $V=1\,200$ m/s。

法向加速度指令为幅值 100 $\mathrm{m/s^2}$ 的阶跃信号，侧向加速度指令为0，仿真结果如图3所示，分别给出了转速 $\dot{\gamma}$ 在2、4、10 r/s 条件下的响应曲线。图3(a)、(b)分别为法向加速度和侧向加速度响应曲线，能够较好地跟踪指令信号。图3(c)、(d)分别为准弹体坐标系下俯仰舵偏角和偏航舵偏角响应曲线，在指令跟踪段俯仰舵以小幅振荡调整来抑制外界干扰的影响，随着转速 $\dot{\gamma}$ 的提高，所需用来抑制侧向加速度的偏航舵偏角越大。从仿真结果可以看出，所设计的控制系统基本不受转速、建模误差和外界干扰的影响，具有较高的鲁棒性。

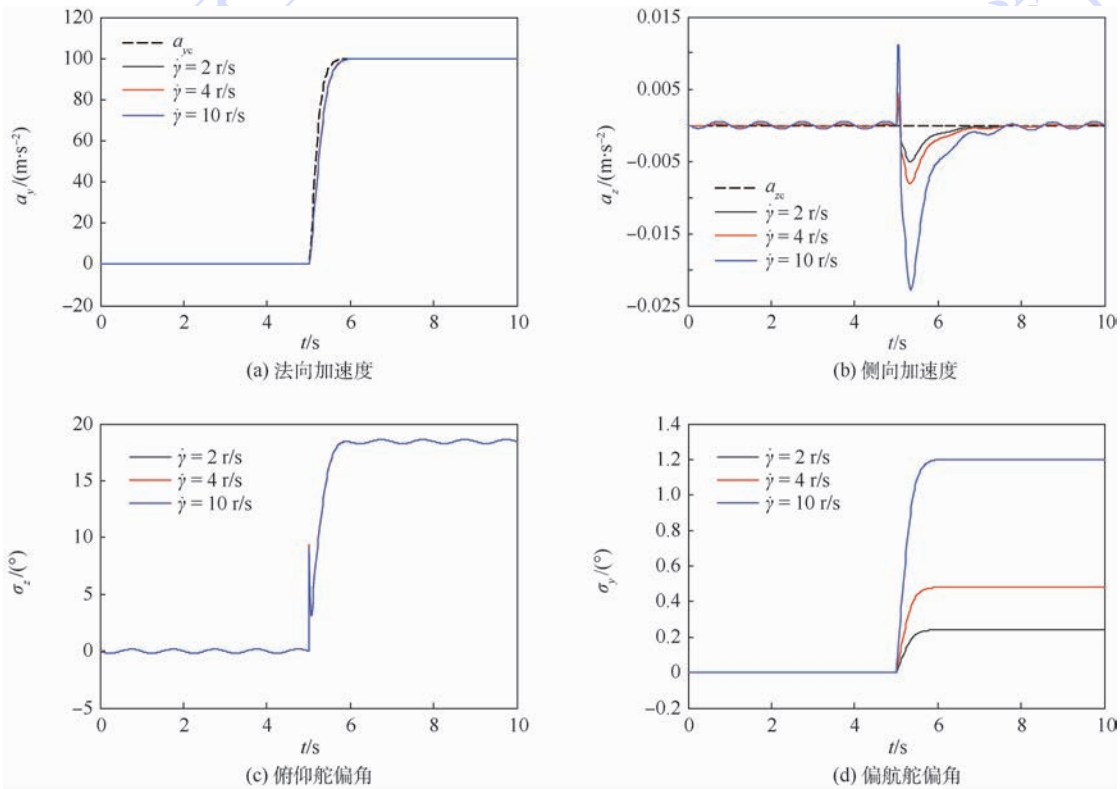


图3 基于 RHO 控制器的仿真结果

Fig. 3 Simulation results based on RHO controller

5 结 论

1）旋转弹旋转导致控制通道出现耦合，随着转速的提高，舵机系统前向通道传递函数的增益将减小，舵机系统耦合通道传递函数的增益将增大，控制通道耦合加剧。

北航学报 赠阅

2）在进行基于指令滤波器的 RHO 控制器设计中，将控制量设计为增广状态向量的线性组合。控制量中包含增广状态向量的部分充分利用反馈信息以达到保障基本稳定飞行的目的；截距控制量用来抑制建模误差和外界干扰的影响。

3）本文设计的 RHO 控制器在滚动的时域区间内实时对控制参数进行更新，有效抑制了建模误差和外界干扰的影响，使得系统输出能够较好地跟踪指令信号，实现解耦控制。

参考文献（References）

[1] 杨树兴，赵良玉，闫晓勇. 旋转弹动态稳定性理论[M]. 北京：国防工业出版社，2014：1-20.
YANG S X，ZHAO L Y，YAN X Y. Dynamic stability theory of spinning projectile[M]. Beijing：National Defence Industry Press，2014：1-20(in Chinese).

[2] 周伟. 旋转弹动态稳定性与鲁棒变增益控制[D]. 北京：北京理工大学，2016：1-30.
ZHOU W. Dynamic stability and robust variable gain control of spinning projectile[D]. Beijing：Beijing Institute of Technology，2016：1-30(in Chinese).

[3] ZHOU W，YANG S X，DONG J L. Coning motion instability of spinning missiles induced by hinge moment[J]. Aerospace Science and Technology，2013，30(1)：239-245.

[4] 陈罗婧，刘莉，于剑桥. 双通道控制旋转导弹自动驾驶仪解耦控制研究[J]. 北京理工大学学报，2008，28(1)：11-14.
CHEN L J，LIU L，YU J Q. Decoupling control of a double-channel control rolling missile autopilot[J]. Transaction of Beijing Institute of Technology，2008，28(1)：11-14(in Chinese).

[5] 陈罗婧，刘莉，于剑桥. 双通道控制旋转导弹自动驾驶仪回路数学变换及其耦合性分析[J]. 北京理工大学学报，2007，27(10)：847-850.
CHEN L J，LIU L，YU J Q. Transform and coupling analysis of double-channel control rolling missile autopilot loop[J]. Transaction of Beijing Institute of Technology，2007，27(10)：847-850(in Chinese).

[6] 袁天保，刘新建，秦子增. 自旋弹弹导弹动力学与控制[J]. 宇航学报，2006，27(2)：217-220.
YUAN T B，LIU X J，QIN Z Z. Dynamic and control of spinning ballistic missile[J]. Journal of Astronautics，2006，27(2)：217-220(in Chinese).

[7] 闫晓勇，李克勇，杨树兴. 基于动态逆理论的自旋弹控制方法[J]. 弹箭与制导学报，2009，29(5)：83-86.
YAN X Y，LI K Y，YANG S X. The control method of spinning missile based on dynamic inversion[J]. Journal of Projectiles，Rockets，Missiles and Guidance，2009，29(5)：83-86(in Chinese).

[8] THEODOULIS S，WERNERT P. Flight control for a class of 155 mm spin-stabilized projectiles with course correction fuse[C]// AIAA Guidance，Navigation，and Control Conference. Reston：AIAA，2011：241-245.

[9] THEODOULIS S，GASSMANN V，WERNERT P. Guidance and control design for a class of spin-stabilized fin-controlled projectiles[J]. Journal of Guidance，Control，and Dynamics，2013，36(2)：517-531.

[10] THEODOULIS S，SEVE F，WERNERT P. Robust gain-scheduled autopilot design for spin-stabilized projectiles with a course-correction fuze[J]. Aerospace Science and Technology，2015，42：477-489.

[11] WARD D G，BARRON R L. A self-designing receding horizon optimal flight controller[C]// Proceedings of the American Control Conference. Washington，D. C.：Barron Associates，1995：3490-3494.

[12] FAWZY M，ABOELELA M A S，RHMAN O A E，et al. Design of missile control system using model predictive control[J]. The Online Journal on Computer Science and Information Technology，2011，1(3)：64-70.

[13] BACHTIAR V. Multi-objective design of model predictive control and its application in missile autopilot and guidance[D]. Melbourne：University of Melbourne，2016.

[14] 钱杏芳，林瑞雄，赵亚男. 导弹飞行力学[M]. 北京：北京理工大学出版社，2006：60-70.
QIAN X F，LIN R X，ZHAO Y N. Missile flight aerodynamics[M]. Beijing：Beijing Institute of Technology Press，2006：60-70(in Chinese).

[15] SONNEVELDT L，CHU Q P，MULDER J A. Nonlinear flight control design using constrained adaptive backstepping[J]. Journal of Guidance，Control，and Dynamics，2007，30(2)：322-335.

[16] 陈伟，冯高鹏，邓坤. 基于滚动时域优化的变后掠角飞机修正控制律设计[J]. 南京航空航天大学学报，2017，49(1)：96-104.
CHEN W，FENG G P，DENG K. Design of retrofit control law based on receding horizon optimal technique for variable sweep wing aircraft[J]. Journal of Nanjing University of Aeronautics & Astronautics，2017，49(1)：96-104(in Chinese).

[17] KIRK D E. Optimal control theory an introduction[M]. New York：Dover Puchblications，2004：90-93.

[18] TODOROV E. Optimal control theory[M]. Massachusetts：MIT Press，2006：1-15.

作者简介：

陈伟　男，博士，工程师。主要研究方向：飞行器建模与智能控制。

北航学报 赠阅

Design of decoupling controller for spinning missile based on receding horizon optimization

CHEN Wei*, SUN Chuanjie, FENG Gaopeng, BAI Yunshan

(Institute of Systems Engineering, China Academy of Engineering Physics, Mianyang 621900, China)

Abstract: The spinning missiles are subjected to external disturbances and uncertainties during flight, and there are the aerodynamic cross-linking, inertial cross-linking and control cross-linking. To realize stable flight, it is necessary to design decoupling controller. Therefore, a decoupling control method based on the receding horizon optimization (RHO) was proposed. The kinematic models of the spinning missile and the servo system were expressed in the form of state space, and the augmented equations of state were obtained based on the equations of state of the spinning missile, the command filters and the integral of the tracking error. The control value was calculated using the RHO based on the command filter, and in order to realize the decoupling control of the spinning missile, the controller gains were adjusted in real time according to the difference between the system output and the command signal. Through the acceleration control simulation results, it can be seen that the designed control system is basically not affected by the spinning rate, the modeling errors and the external disturbances, which has high robustness.

Keywords: spinning missile; decoupling control; receding horizon optimization (RHO); acceleration control; command filter

北京航空航天大学学报

Received: 2017-05-02; **Accepted**: 2017-06-09; **Published online**: 2017-08-31 14:32

URL: kns.cnki.net/kcms/detail/11.2625.V.20170831.1432.006.html

Foundation item: National Natural Science Foundation of China (U143010033)

* **Corresponding author**. E-mail: chenweifenglaile@163.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2017.0262

面向装备 RUL 预测的平行仿真技术

葛承垒, 朱元昌, 邸彦强*
(军械工程学院 电子与光学工程系, 石家庄 050003)

摘 要: 装备平行仿真是系统建模与仿真领域的新兴仿真技术,已经成为研究热点。在装备维修保障领域中,分析了装备剩余寿命(RUL)预测存在的突出问题,即模型参数固定、不具备自适应演化能力,成为阻碍实现装备剩余寿命自适应预测的首要因素。结合装备平行仿真理论,在建模分析的基础上提出了面向装备剩余寿命预测的平行仿真框架,该框架以 Wiener 状态空间模型为基础仿真模型,在动态注入的装备退化观测数据驱动下,利用期望最大化(EM)算法在线更新模型参数,并利用卡尔曼滤波(KF)算法实现仿真输出数据与观测数据的同化(DA),从而实现仿真模型动态演化,使得仿真输出不断逼近装备真实退化状态,为准确预测剩余寿命提供高逼真度仿真模型和数据输出。以某轴承性能退化数据为数据驱动源,对该框架进行了验证,仿真结果表明平行仿真方法能准确仿真装备性能退化过程,在提高预测精度的基础上实现了装备剩余寿命的自适应预测,有力证明了平行仿真方法的可行性和有效性。

关 键 词: 平行仿真; 模型演化; 剩余寿命(RUL); 数据同化(DA); 参数估计
中图分类号: TP391.9
文献标识码: A **文章编号**: 1001-5965(2018)04-0725-12

在以往仿真技术中,仿真系统与真实系统的关系在结构上是分离的、在运行上是离线的、在地位/角色上是辅助的。随着应用需求的变化,仿真系统和真实系统开始出现融合的趋势,如嵌入式仿真[1]、在线仿真[2]等概念。而且,近年来在复杂系统和自然现象等研究领域[3]出现了虚实共生和平行执行[4]的概念,强调仿真系统/人工系统与真实系统之间在结构上相互依存、时间上并行执行(不一定同步)。在装备作战运用和维修保障领域中,实时决策[5]、在线规划、健康管理、精确维修等需求日益迫切,建立与装备共生、平行的仿真系统是解决此类问题的一种新途径,在此背景下,装备平行仿真技术应运而生。

以装备维修保障领域为例,武器装备结构的复杂性和运行工况条件的时变性使得装备故障率和维修保障难度增加,装备一旦发生故障,不仅影响武器装备作战效能的发挥,而且停机时间长、停机费用高,这就要求对装备进行精确维修。传统维修方式如修复性维修和定时维修存在“过维修”和“欠维修”的问题,已不能满足精确维修需要。随着状态监测技术的发展,基于状态的维修(Condition Based Maintenance, CBM)能有效减少停机时间、降低维修费用、提高装备可用度,已成为实现装备精确维修的重要手段。CBM 是指通过内部传感器或外部监测设备获得装备状态信息,据此对装备健康状态进行评价,为维修决策提供依据。其中,剩余寿命(Remaining Useful Life, RUL)预测是 CBM 的关键环节[6]。RUL 是指装备从当前时刻至发生故障时刻的时间长度,准确的 RUL 预测是维修规划的前提和依据,且根据实

收稿日期: 2017-04-26; **录用日期**: 2017-07-07; **网络出版时间**: 2017-09-15 17:09
网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20170915.1709.002.html
基金项目: 装备预研基金重点项目(9140A04020115JB34011)
* **通信作者**. E-mail: 1049084176@qq.com

引用格式: 葛承垒, 朱元昌, 邸彦强. 面向装备 RUL 预测的平行仿真技术[J]. 北京航空航天大学学报, 2018, 44(4): 725-736.
GE C L, ZHU Y C, DI Y Q. Equipment RUL prediction oriented parallel simulation technology[J]. Journal of Beijing University of Aeronautics and Astronautics, 2018, 44(4): 725-736 (in Chinese).

北航学报 赠阅

时监测数据自适应修正 RUL 预测结果成为重要研究方向[7]。当前的 RUL 预测方法中模型不具备与装备的双向交互能力,不支持模型参数自适应调整和模型输出自适应校正,模型参数是已知的或者根据历史数据以离线估计的方式得到,成为阻碍实现 RUL 自适应预测的重要因素。

装备平行仿真技术为解决 RUL 自适应预测问题提供了有效技术途径,针对装备 RUL 自适应预测需求,利用平行仿真技术可实现预测模型演化,不断逼近装备真实退化状态,实现 RUL 的在线、实时、自适应预测,称此技术为面向装备 RUL 预测的平行仿真技术。

1 装备平行仿真理论

装备平行仿真是系统建模与仿真领域的新兴技术,已经成为研究前沿和热点。装备平行仿真是一种新的仿真应用模式,旨在将仿真系统和武器装备通过双向交互同时运行在一起,仿真系统以在线的方式实时获取装备信息,用于在线修正仿真模型,仿真结果可以动态地反馈给装备并影响武器装备的运行,从而提升武器装备的自身性能,以及运用和保障效能[8]。以这种模式运行的仿真系统称之为平行仿真系统,平行仿真系统是对装备的仿真,包括武器装备全部或部分环节、特性的仿真模型。

在装备平行仿真中,实际武器装备和平行仿真系统通过传感器、执行器进行双向交互。传感器给平行仿真系统提供实际武器装备信息,执行器使得平行仿真系统对实际武器装备执行控制等操作。一些实际武器装备或其子系统并不可控,此时执行器不一定存在。装备平行仿真示意图如图 1 所示。由传感器提供的实际武器装备信息可以划分为 2 类,即在时刻 t 的实际武器可观测装备状态信息 S_t 和行为信息 B_t。

S_t 反映了实际武器装备在给定时刻多个内部状态变量的取值,可以将实际武器装备状态划分为已知状态 S_a 和未知状态 S_{ua}。由于武器装备的复杂性,很难获得装备的完全状态信息 $S_t = S_a \cup S_{ua}$。例如,获取装备的完全状态信息可能会消耗大量时间或者经过密集计算,当平行仿真系统利用这些信息时可能已经出现信息过时的情况。除此之外,一些状态变量能够利用传感器进行观测,而一些状态变量却不可观测。B_t 反映了状态信息中不能反映的可计量的实际武器装备信息。在面向装备 RUL 预测的平行仿真中,主要涉及状态信息 S_t。

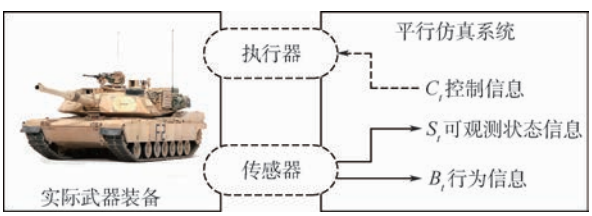


图 1 装备平行仿真示意图

Fig. 1 Schematic of equipment parallel simulation

实际武器装备能够受控于平行仿真系统发送的控制信息 C_t,根据控制信息的执行方式,可以划分为自动控制信息 C_t^{A} 和手动控制信息 C_t^{M}。由于控制信息的获取需要经过平行仿真系统一定时间的计算或推理,因此 C_t 是离散控制信息。控制信息 C_t 只在一定受限时间内有效,经过一定时间后可能是无效信息,此时有效的控制信息可能就变为 C_{t+1}。在面向装备 RUL 预测的平行仿真中,平行仿真系统的反馈为 RUL 及其概率密度函数(Probability Density Function,PDF),但 RUL 及其 PDF 都属于隐含控制量,没有直接体现在 C_t 中,RUL、PDF 反馈给装备维护人员后,由装备维护人员根据反馈结果和维修经验进行维修决策,进而利用执行器对实际武器装备或其子系统进行修复、更换等视情维修操作。

虽然平行仿真中的“平行”在英文中翻译为“parallel”,但与并行仿真中“parallel”有本质不同,装备平行仿真中的“平行”与平行系统理论中的“平行”具有相同的内涵,是指实际系统和仿真系统的平行互动关系。装备平行仿真的理论缘起与平行系统理论[9-10]、动态数据驱动应用系统(Dynamic Data Driven Application System,DDDAS)[11-13]、共生仿真[14]、在线仿真等理论范式密切相关,文献[8]已作详细综述,但是也存在明显差异。平行系统理论强调基于 Agent 的建模,DDDAS、共生仿真、在线仿真中实际系统和仿真系统不存在平行关系,共生仿真强调利用假设分析(What-If Analysis,WIA)方法进行多脚本仿真,在线仿真是与离线仿真相对的概念,实际系统和仿真系统之间的连接是其基本特征。在以上相关技术中,仿真模型侧重于一次性构建,属于相对固定和特定的模型,具体表现在模型参数固定、不具备数据同化能力,即仿真模型不具备演化能力[15]。平行仿真的建模技术属于演化建模或适应性建模,与以往仿真技术相比,模型演化是平行仿真技术在建模方式上的核心内涵和主要技术特征。平行仿真系统通过模型演化使得仿真模型输出不断逼近装备真实状态,提高仿真系统预测的

北航学报 赠阅

准确性。此外,平行仿真技术还具有虚实共生、数据驱动、高效运行的特点[8]。

目前,国内仿真领域的专家学者正在逐步开展装备平行仿真技术研究,主要面向装备指挥决策支持领域[16]、维修保障领域[17]及航天控制仿真领域,针对指挥决策支持的平行仿真研究相对较多,如毛少杰[18-19]、邱晓刚[5]、窦林涛[20]等,目前处于框架建立和关键技术讨论阶段,在航天控制仿真领域中,王会霞[21]讨论了平行仿真技术的特点,给出了其在导弹控制系统仿真验证中的应用框架,本文的研究针对的是装备维修保障领域中平行仿真技术研究,且主要关注演化建模技术。

2　面向 RUL 预测的平行仿真框架

面向装备 RUL 预测的平行仿真是指构建与实际武器装备平行运行的仿真系统即平行仿真系统,实际武器装备和平行仿真系统通过特定接口互联在一起,并利用状态监测技术以在线形式实时感知装备退化状态以驱动仿真模型状态和参数更新,实现仿真模型演化,得到更为准确的 RUL 预测结果,并将预测结果反馈给装备维护人员,为维修决策提供实时数据支持,提升实际武器装备的保障效能。

2.1　建模分析

模型是仿真的基础,建立恰当的模型是面向装备 RUL 预测平行仿真的前提。模型的恰当主要表现在 3 个方面:一是所建立模型应该符合建模意图,即利于 RUL 预测;二是所建立的模型应该易于进行状态估计;三是所建立模型应该能反映装备潜在的特性和规律。传统的寿命预测、可靠性评估等理论注重的是对装备失效数据的研究,相关的理论方法已经十分完善,这些分析方法都是建立在大量失效数据的基础之上的。然而随着制造工艺的提高,产品寿命延长、可靠性提高,很难得到装备的失效数据,即便采用加速寿命试验方法也存在试验周期长、试验成本高等突出问题。装备完成特定功能是由其性能参数所表征的,受环境应力、负荷的影响,装备性能参数随时间退化是不可避免的,性能退化是导致装备故障的主要原因,故装备故障机理可通过性能退化过程进行解释分析。相较于失效数据,与装备健康状态相关的随机退化信号是可以通过状态监测技术获得的,如振动监测和油液分析,充分挖掘和利用这些退化数据,能够有效分析装备的可靠性和 RUL。因此,基于性能退化建模是首选建模方向。

为便于描述装备动态变化的退化状态并对装备退化状态进行估计,宜建立装备退化状态空间模型(State Space Model,SSM),包括状态方程和观测方程。然而,受退化数据有限的制约,直接建立装备退化 SSM 十分困难。考虑到运行环境、内部结构对装备性能退化的影响具有随机性和不确定性,利用随机过程模型进行装备退化建模分析是比较合理的选择。基于随机过程模型的装备性能退化建模通过利用随机过程拟合装备某一性能退化量的变化轨迹来描述设备退化的趋势,以此来预测设备达到失效阈值的时间。Wiener 过程不仅能描述具有增加或减小趋势的非单调退化过程,而且具有良好的数学性质,即容易得到 RUL 的 PDF 解析解,有利于实时预测,因此本文选用 Wiener 过程。故将 SSM 和 Wiener 过程结合在一起构建装备退化 SSM 既有利于装备退化状态估计,也有利于刻画装备退化规律。本文所提的演化建模框架是以基于 Wiener SSM(WSSM)的性能退化建模为基础的。

2.2　演化建模框架

面向装备 RUL 预测的平行仿真示意图如图 2 所示。其具体过程为:利用传感器等监测设备对武器装备退化状态进行实时监测并得到监测数据,经虚实接口传输到平行仿真系统中,平行仿真系统首先对装备退化状态进行在线感知,主要涉及数据降噪、缺失数据处理等数据预处理技术,得到反映装备退化状态的特征量,并与 WSSM 输出不断进行数据同化,实现装备退化状态的跟踪循环,结合参数在线估计算法对 WSSM 中未知参数进行实时估计,实现模型参数演化,不断逼近装备真实退化状态,进而得到演化后的 WSSM,并利用装备失效阈值实时预测装备 RUL 及其 PDF,平行仿真系统将预测结果反馈给装备维护人员,以此作为视情维修决策的重要依据,利用执行器对武器装备进行备件更换等维修操作,从而提高装备保障效能。

数据同化(Data Assimilation,DA)和参数在线估计是实现 WSSM 演化的关键。由于装备退化数据的动态注入,平行仿真系统必须具备数据同化能力,以提高仿真系统的适应能力和预测能力。数据同化将最新观测数据引入 WSSM 中,减少 WSSM 噪声的影响,通过不断校正 WSSM 输出结果使得 WSSM 预测轨迹更加接近真实情况,常用的数据同化算法有卡尔曼滤波(Kalman Filter,KF)[22]、粒子滤波(Particle Filter,PF)等。参数在线估计算法用于演化 WSSM 未知参数,期望最大化(Expectation Maximum,EM)算法是目前常用

北航学报 赠阅

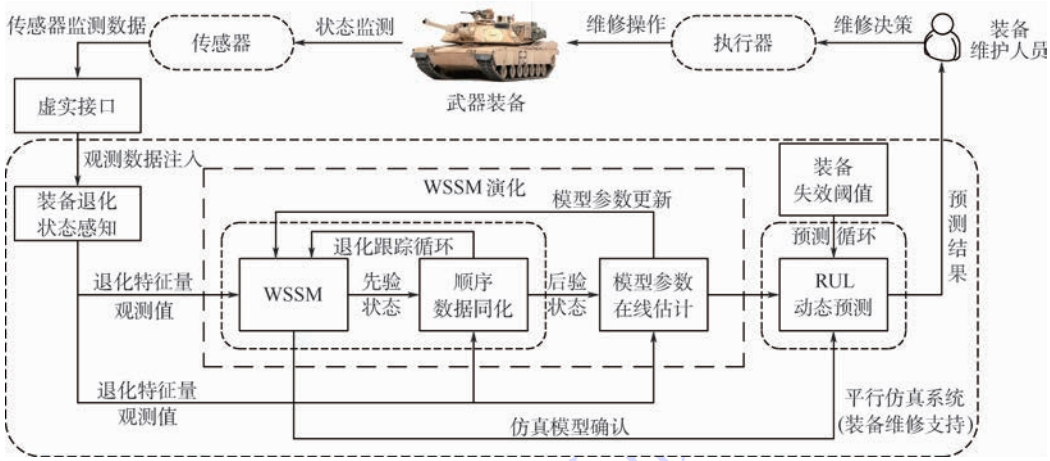


图 2 面向装备 RUL 预测的平行仿真示意图

Fig. 2 Schematic of equipment RUL prediction oriented parallel simulation

的用于估计 SSM 未知参数的算法。由于此演化建模框架以 WSSM 为基础,使其还考虑了历史监测数据和监测噪声对 RUL 的影响,使得 RUL 预测结果更为准确合理。

2.3 基于 WSSM 的装备性能退化建模

2.3.1 Wiener 过程与 MLE-IG 法

Wiener 过程可记为

$$x(t) = x(0) + \eta t + \sigma B(t) \tag{1}$$

式中:$x(0)$ 为初始退化状态,一般设为 0;η、σ 为未知参数,η 为漂移系数,σ 为扩散系数;$B(t)$ 为标准 Brownian 运动,且服从均值为 0、方差为 t 的正态分布,即 $B(t) \sim N(0,t)$。设装备失效阈值为 w,则装备剩余寿命 T 定义为退化过程首次通过失效阈值的时间:

$$T(w) = \inf\{t: x(t) \geqslant w \mid x(0) < w\} \tag{2}$$

Wiener 过程具有以下性质:

1) Wiener 过程属于无后效性的齐次 Markov 过程。

2) 在不同时间段内,$x(t)$ 的增量 Δx_i 相互独立。

3) 独立增量 Δx_i 服从均值为 $\eta\Delta t_i$、方差为 $\sigma^2\Delta t_i$ 的正态分布,即 $\Delta x_i \sim N(\eta\Delta t_i, \sigma^2\Delta t_i)$。

Wiener 过程的数学期望 $E(x(t)) = \eta t$ 是时间 t 的线性函数,漂移系数是反映退化过程的重要参数,Wiener 过程的方差 $\mathrm{var}(x(t)) = \sigma^2 t$ 表征了退化过程在时刻 t 的不确定性,因此对参数 η、σ^2 的实时估计是实现准确 RUL 预测的前提。由于 $E(x(t)) = \eta t$ 是线性形式,故 Wiener 过程适宜于描述线性退化系统。以往对 Wiener 过程中参数 η、σ^2 的估计利用极大似然估计(Maximum Likelihood Estimation, MLE)方法,由于 $\Delta x_i \sim N(\eta\Delta t_i, \sigma^2\Delta t_i)$,则 Δx_i 的 PDF 为

$$f(\Delta x_i; \eta\Delta t_i, \sigma^2\Delta t_i) = \frac{1}{\sqrt{2\pi\sigma^2\Delta t_i}} \exp\left(-\frac{(\Delta x_i - \eta\Delta t_i)^2}{2\sigma^2\Delta t_i}\right) \tag{3}$$

则其似然函数为

$$L(\eta, \sigma^2) = \prod_{i=1}^{n} f(\Delta x_i; \eta\Delta t_i, \sigma^2\Delta t_i) = \prod_{i=1}^{n}\left(\frac{1}{\sqrt{2\pi\sigma^2\Delta t_i}}\right) \exp\left(-\sum_{i=1}^{n}\frac{(\Delta x_i - \eta\Delta t_i)^2}{2\sigma^2\Delta t_i}\right) \tag{4}$$

对似然函数取对数,得其对数似然函数

$$\ln L(\eta, \sigma^2) = \ln \prod_{i=1}^{n} f(\Delta x_i; \eta\Delta t_i, \sigma^2\Delta t_i) = \sum_{i=1}^{n}\ln\left(\frac{1}{\sqrt{2\pi\sigma^2\Delta t_i}}\right) - \sum_{i=1}^{n}\frac{(\Delta x_i - \eta\Delta t_i)^2}{2\sigma^2\Delta t_i} \tag{5}$$

对数似然函数分别对 η、σ^2 求偏导数,并令偏导数为 0,即

$$\begin{cases} \dfrac{\partial \ln L(\eta, \sigma^2)}{\partial \eta} = 0 \\ \dfrac{\partial \ln L(\eta, \sigma^2)}{\partial \sigma^2} = 0 \end{cases} \tag{6}$$

可求得参数 η、σ^2 的极大似然估计值:

$$\begin{cases} \hat{\eta} = \dfrac{\sum_{i=1}^{n}\Delta x_i}{\sum_{i=1}^{n}\Delta t_i} \\ \hat{\sigma}^2 = \dfrac{1}{n}\sum_{i=1}^{n}\dfrac{(\Delta x_i - \hat{\eta}\Delta t_i)^2}{\Delta t_i} \end{cases} \tag{7}$$

得到参数估计值 $\hat{\eta}$、$\hat{\sigma}$ 后,若装备在时刻 t_i 的退化量为 $x(t_i)$,根据 Wiener 过程首达时分布 $X \sim \mathrm{IG}((w - x(t_i))/\hat{\eta}, (w - x(t_i))^2/\hat{\sigma}^2)$ 即服从逆高斯(IG)分布,则此时装备 RUL 的 PDF、累积分布函数(Cumulative Distribution Function,

北航学报 赠阅

CDF)为

$$f(t)=\frac{w-x(t_i)}{\sqrt{2\pi\hat{\sigma}^2t^3}}\exp\left(-\frac{(w-x(t_i)-\hat{\eta}t)^2}{2\hat{\sigma}^2t}\right) \tag{8}$$

$$F(t)=1-\Phi\left(\frac{w-x(t_i)-\hat{\eta}t}{\hat{\sigma}\sqrt{t}}\right)+\exp\left(\frac{2\hat{\eta}(w-x(t_i))}{\hat{\sigma}^2}\right)\Phi\left(\frac{-(w-x(t_i))-\hat{\eta}t}{\hat{\sigma}\sqrt{t}}\right) \tag{9}$$

称此传统方法为 MLE-IG 法。注意到,MLE-IG 法无法实现模型状态、参数的演化,且未考虑监测噪声和历史数据的影响,得到的 RUL 预测结果必然存在较大误差。

2.3.2　Wiener 状态空间模型

为考虑历史数据的影响和漂移系数的不确定性,将漂移系数 η 扩展成退化漂移状态 η_k,即

$$\eta_k=\eta_{k-1}+\nu_k \tag{10}$$

式中:$\nu_k\sim N(0,\varepsilon^2)$,标准差 ε 表征了 η_k 的不确定性。

装备退化状态监测中受传感器测量精度以及装备运行工况的影响,精确的退化状态数据无法直接测量得到,考虑测量噪声的影响,观测方程可写为

$$y_k=x_k+\pi_k \tag{11}$$

式中:$\pi_k\sim N(0,\phi^2)$,ϕ^2 为测量噪声的方差。得到模型

$$\begin{cases}\eta_k=\eta_{k-1}+\nu_k\\ x(t)=x(0)+\eta t+\sigma B(t)\\ y_k=x_k+\pi_k\end{cases} \tag{12}$$

进一步,通过对 $x(t)$ 进行 Euler 离散化,可以得到 WSSM 为

$$\begin{cases}\eta_k=\eta_{k-1}+\nu_k\\ x_k=x_{k-1}+\eta_{k-1}\tau_k+\sigma B(\tau_k)\\ y_k=x_k+\pi_k\end{cases} \tag{13}$$

式中:$\tau_k=t_k-t_{k-1}$ 为传感器采样间隔,并假设初始退化状态 x_0、初始退化漂移状态 η_0、漂移状态噪声 $\nu_1,\nu_2,\cdots,\nu_k$,以及 Brownian 运动 $B(\tau_1)$,$B(\tau_2),\cdots,B(\tau_k)$ 和传感器监测噪声 $\pi_1,\pi_2,\cdots,\pi_k$ 相互独立。方程包含 2 个状态,即退化状态 x 和漂移状态 η。由式(13)所知,二者存在相关性,故难以同时进行估计。为解决此问题,可将 x、η 合成为一个扩展状态 z,即 $z_k=(\eta_k,x_k)^{\mathrm{T}}$,WSSM 可进一步记为

$$\begin{cases}\boldsymbol{z}_k=\boldsymbol{F}_k\boldsymbol{z}_{k-1}+\boldsymbol{w}_k\\ y_k=\boldsymbol{H}\boldsymbol{z}_k+\pi_k\end{cases} \tag{14}$$

式中:$\boldsymbol{F}_k=\begin{bmatrix}1&0\\ \tau_k&1\end{bmatrix}$;$\boldsymbol{w}_k\sim N(0,\boldsymbol{Q}_k)$,$\boldsymbol{Q}_k=\begin{bmatrix}\varepsilon^2&0\\ 0&\sigma^2\tau_k\end{bmatrix}$;$\boldsymbol{H}=[0\quad 1]$。由于传感器一般采用等间隔采样,故 τ_k 是常数,因此 $\boldsymbol{F}_k$ 和 $\boldsymbol{Q}_k$ 的下标可以忽略。WSSM 最终记为

$$\begin{cases}\boldsymbol{z}_k=\boldsymbol{F}\boldsymbol{z}_{k-1}+\boldsymbol{w}_k\\ y_k=\boldsymbol{H}\boldsymbol{z}_k+\pi_k\end{cases} \tag{15}$$

3　基于 KF-EM 算法的 WSSM 演化

3.1　基于 KF 算法的 WSSM 输出校正

作为动态数据驱动的仿真理论,平行仿真能提高 RUL 预测精度的关键在于不断地将实际装备退化状态观测数据与 WSSM 输出结果进行数据同化,即校正装备退化 SSM 输出。数据同化算法是联系观测数据和 WSSM 的桥梁,是实现退化状态预测的基础。本文利用 KF 算法实现数据同化,KF 能在状态线性变化和误差高斯分布情况下得到状态最优估计,其实质是对装备退化状态进行跟踪。利用 KF 实现 WSSM 输出校正主要包括预测和更新 2 个步骤。

步骤 1　预测

利用时刻 $k-1$ 扩展状态 z_{k-1} 的后验估计 $\hat{z}_{k-1|k-1}$ 及协方差 $\boldsymbol{P}_{k-1|k-1}$ 得到当前 k 时刻扩展状态 z_k 的先验估计 $\hat{z}_{k|k-1}$ 及协方差 $\boldsymbol{P}_{k|k-1}$,即

$$\hat{z}_{k|k-1}=\boldsymbol{F}\hat{z}_{k-1|k-1} \tag{16}$$

$$\boldsymbol{P}_{k|k-1}=\boldsymbol{F}\boldsymbol{P}_{k-1|k-1}\boldsymbol{F}^{\mathrm{T}}+\boldsymbol{Q} \tag{17}$$

$$\boldsymbol{P}_{k-1|k-1}=E[(z_{k-1|k-1}-\hat{z}_{k-1|k-1})(z_{k-1|k-1}-\hat{z}_{k-1|k-1})^{\mathrm{T}}] \tag{18}$$

步骤 2　更新

利用 k 时刻扩展状态 z_k 的先验估计 $\hat{z}_{k|k-1}$ 及协方差 $\boldsymbol{P}_{k|k-1}$ 和观测值 y_k 得到扩展状态 z_k 的后验估计,即

$$\tilde{y}_k=y_k-\boldsymbol{H}\hat{z}_{k|k-1} \tag{19}$$

$$S_k=\boldsymbol{H}\boldsymbol{P}_{k|k-1}\boldsymbol{H}^{\mathrm{T}}+\phi^2 \tag{20}$$

$$\boldsymbol{K}_k=\boldsymbol{P}_{k|k-1}\boldsymbol{H}^{\mathrm{T}}S_k^{-1} \tag{21}$$

$$\hat{z}_{k|k}=\hat{z}_{k|k-1}+\boldsymbol{K}_k\tilde{y}_k \tag{22}$$

$$\boldsymbol{P}_{k|k}=(\boldsymbol{I}-\boldsymbol{K}_k\boldsymbol{H})\boldsymbol{P}_{k|k-1} \tag{23}$$

式中:$\tilde{y}_k$ 为新息;S_k 为 $\tilde{y}_k$ 的协方差;$\boldsymbol{K}_k$ 为卡尔曼增益;$\hat{z}_{k|k}$ 和 $\boldsymbol{P}_{k|k}$ 分别为时刻 k 扩展状态后验估计的均值和协方差矩阵。当得到新的量测值 y_{k+1} 时,根据式(16)~式(23)就可以得到 $k+1$ 时刻扩展状态的后验估计。

3.2　基于 EM 算法的 WSSM 参数演化

WSSM 中未知参数 $\boldsymbol{\theta}$ 为 $\boldsymbol{\mu}_0$、$\boldsymbol{\Sigma}_0$、$\boldsymbol{Q}$、ϕ,$\boldsymbol{\mu}_0$ 和

北航学报 赠阅

$\boldsymbol{\Sigma}_0$ 分别为初始隐藏状态 z_0 的数学期望和协方差矩阵，即 $z_0 \sim N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$。$\boldsymbol{\theta}$ 为装备退化特性和规律的重要特征，但是在应用中这些参数是未知且可变的，需要根据装备退化数据对 $\boldsymbol{\theta}$ 进行在线估计。实际上，$\boldsymbol{\theta}$ 的初始值可通过同一型号装备的历史数据估计得到，然而同一型号装备也存在差异，加之装备运行环境等因素的影响，利用同一型号历史数据估计未知参数存在很大误差，因此根据装备退化状态数据对参数 $\boldsymbol{\theta}$ 进行在线估计能有效提高平行仿真系统预测的准确性。MLE 是常用的估计算法，然而由于 WSSM 中存在隐藏状态，MLE 并不适用。EM 算法是一种能有效估计含隐藏状态模型参数的迭代算法，平行仿真系统利用 EM 算法可对 WSSM 中未知参数 $\boldsymbol{\theta}$ 进行实时、在线估计。定义隐藏状态和测量数据分别为 $\boldsymbol{Z}_k = (z_1, z_2, \cdots, z_k)$、$\boldsymbol{Y}_k = (y_1, y_2, \cdots, y_k)$。

在采样时刻 k、EM 算法第 l 步时，参数 $\hat{\boldsymbol{\theta}}_k^{(l)}$ 可用式(24)估计

$$\hat{\boldsymbol{\theta}}_k^{(l)} = \arg\max_{\boldsymbol{\theta}} \psi(\boldsymbol{\theta} \mid \hat{\boldsymbol{\theta}}_k^{(l-1)}) = \arg\max_{\boldsymbol{\theta}} E_{\boldsymbol{Z}_k \mid \boldsymbol{Y}_k, \hat{\boldsymbol{\theta}}_k^{(l-1)}}(\varphi_k(\boldsymbol{\theta})) \tag{24}$$

式中：$\varphi_k(\boldsymbol{\theta})$ 为 $\boldsymbol{Z}_k$、$\boldsymbol{Y}_k$ 的联合对数似然函数；$\psi(\boldsymbol{\theta} \mid \hat{\boldsymbol{\theta}}_k^{(l-1)})$ 为 $\varphi_k(\boldsymbol{\theta})$ 关于 $\boldsymbol{Z}_k$ 以 $\boldsymbol{Y}_k$ 和 $\hat{\boldsymbol{\theta}}_k^{(l-1)}$ 为条件的数学期望。具体地，式(24)可分解为 E 步和 M 步。

1) E 步

根据 Bayesian 定理和 Markov 性质，$\varphi_k(\boldsymbol{\theta})$ 可展开为

$$\begin{aligned}\varphi_k(\boldsymbol{\theta}) &= \ln p(\boldsymbol{Z}_k, \boldsymbol{Y}_k \mid \boldsymbol{\theta}) = \ln p(z_0 \mid \boldsymbol{\theta}) + \sum_{i=1}^{k} \ln p(z_k \mid z_{k-1}, \boldsymbol{\theta}) + \sum_{i=1}^{k} \ln p(y_k \mid z_k, \boldsymbol{\theta}) \propto \\ &\ln|\boldsymbol{\Sigma}_0| + (z_0 - \boldsymbol{\mu}_0)^{\mathrm{T}} \boldsymbol{\Sigma}_0^{-1} (z_0 - \boldsymbol{\mu}_0) + k\ln\phi^2 + \frac{1}{\phi^2}\sum_{i=1}^{k}(y_i - \boldsymbol{H}z_i)^2 + k\ln|\boldsymbol{Q}| + \\ &\sum_{i=1}^{k}(z_i - \boldsymbol{F}z_{i-1})^{\mathrm{T}} \boldsymbol{Q}^{-1}(z_i - \boldsymbol{F}z_{i-1})\end{aligned} \tag{25}$$

将式(25)代入 $\psi(\boldsymbol{\theta} \mid \hat{\boldsymbol{\theta}}_k^{(l-1)})$ 可得

$$\begin{aligned}\psi(\boldsymbol{\theta} \mid \hat{\boldsymbol{\theta}}_k^{(l-1)}) \propto\ & \ln|\boldsymbol{\Sigma}_0| + E[(z_0 - \boldsymbol{\mu}_0)\boldsymbol{\Sigma}_0^{-1}(z_0 - \boldsymbol{\mu}_0)^{\mathrm{T}}] + k\ln\phi^2 + \frac{1}{\phi^2}\sum_{i=1}^{k}E(y_i - \boldsymbol{H}z_i)^2 + \\ & k\ln|\boldsymbol{Q}| + \sum_{i=1}^{k}E[(z_i - \boldsymbol{F}z_{i-1})\boldsymbol{Q}^{-1}(z_i - \boldsymbol{F}z_{i-1})^{\mathrm{T}}]\end{aligned} \tag{26}$$

进一步推导，有

$$\begin{aligned}\psi(\boldsymbol{\theta} \mid \hat{\boldsymbol{\theta}}_k^{(l-1)}) \propto\ & k\ln\phi^2 + k\ln|\boldsymbol{Q}| + \frac{1}{\phi^2}\sum_{i=1}^{k}E(y_i - \boldsymbol{H}z_i)^2 + \mathrm{tr}\{\boldsymbol{\Sigma}_0^{-1}E[(z_0 - \boldsymbol{\mu}_0)(z_0 - \boldsymbol{\mu}_0)^{\mathrm{T}}]\} + \ln|\boldsymbol{\Sigma}_0| + \\ & \mathrm{tr}\{\boldsymbol{Q}^{-1}E[(z_i - \boldsymbol{F}z_{i-1})(z_i - \boldsymbol{F}z_{i-1})^{\mathrm{T}}]\}\end{aligned} \tag{27}$$

式中：tr 为矩阵迹运算。为计算式(27)，需首先计算 $\hat{z}_{i\mid k} = E(z_i \mid \boldsymbol{Y}_k, \hat{\boldsymbol{\theta}}_k^{(l-1)})$、$\boldsymbol{P}_{i\mid k} = \mathrm{cov}(z_i \mid \boldsymbol{Y}_k, \hat{\boldsymbol{\theta}}_k^{(l-1)})$、$\boldsymbol{P}_{i,i-1\mid k} = \mathrm{cov}(z_i, z_{i-1} \mid \boldsymbol{Y}_k, \hat{\boldsymbol{\theta}}_k^{(l-1)})$。这里采用 Rauch-Tung-Striebel(RTS)平滑算法计算上述 3 个未知量，RTS 平滑算法利用 KF 算法得到的最优滤波信息作为滤波初值，按照时间逆序作平滑处理，属于固定区间最优平滑，实质是一种后向滤波。

平滑增益为

$$\boldsymbol{C}_i = \boldsymbol{P}_{i\mid i}\boldsymbol{F}^{\mathrm{T}}\boldsymbol{P}_{i+1\mid i}^{-1} \tag{28}$$

平滑状态向量为

$$\hat{z}_{i\mid k} = E(z_i \mid \boldsymbol{Y}_k, \hat{\boldsymbol{\theta}}_k^{(l-1)}) = \hat{z}_{i\mid i} + \boldsymbol{C}_i(\hat{z}_{i+1\mid k} - \hat{z}_{i+1\mid i}) \tag{29}$$

平滑状态向量的协方差矩阵为

$$\boldsymbol{P}_{i\mid k} = \mathrm{cov}(z_i \mid \boldsymbol{Y}_k, \hat{\boldsymbol{\theta}}_k^{(l-1)}) = \boldsymbol{P}_{i\mid i} + \boldsymbol{C}_i(\boldsymbol{P}_{i+1\mid k} - \boldsymbol{P}_{i+1\mid i})\boldsymbol{C}_i^{\mathrm{T}} \tag{30}$$

$$\boldsymbol{P}_{k,k-1\mid k} = (\boldsymbol{I} - \boldsymbol{K}_k\boldsymbol{H})\boldsymbol{F}\boldsymbol{P}_{k-1\mid k-1} \tag{31}$$

$$\boldsymbol{P}_{i,i-1\mid k} = \boldsymbol{P}_{i\mid i}\boldsymbol{C}_{i-1}^{\mathrm{T}} + \boldsymbol{C}_i(\boldsymbol{P}_{i+1,i\mid k} - \boldsymbol{F}\boldsymbol{P}_{i\mid i})\boldsymbol{C}_{i-1}^{\mathrm{T}} \tag{32}$$

根据式(28)～式(32)，式(27)中的 3 项数学期望可记为

$$\boldsymbol{\Gamma}_{1,i} = E(y_i - \boldsymbol{H}z_i)^2 = (y_i - \boldsymbol{H}\hat{z}_{i\mid k})^2 + \boldsymbol{H}\boldsymbol{P}_{i\mid k}\boldsymbol{H}^{\mathrm{T}} \tag{33}$$

$$\boldsymbol{\Gamma}_2 = E[(z_0 - \boldsymbol{\mu}_0)(z_0 - \boldsymbol{\mu}_0)^{\mathrm{T}}] = \boldsymbol{P}_{0\mid k} + (\hat{z}_{0\mid k} - \boldsymbol{\mu}_0)(\hat{z}_{0\mid k} - \boldsymbol{\mu}_0)^{\mathrm{T}} \tag{34}$$

$$\begin{aligned}\boldsymbol{\Gamma}_{3,i} = E[(z_i - \boldsymbol{F}z_{i-1})(z_i - \boldsymbol{F}z_{i-1})^{\mathrm{T}}] =\ & \boldsymbol{P}_{i\mid k} - \boldsymbol{P}_{i,i-1\mid k}\boldsymbol{F}^{\mathrm{T}} - \boldsymbol{F}\boldsymbol{P}_{i,i-1\mid k}^{\mathrm{T}} + \boldsymbol{F}\boldsymbol{P}_{i-1\mid k}\boldsymbol{F}^{\mathrm{T}} + \\ & (\hat{z}_{i\mid k} - \boldsymbol{F}\hat{z}_{i-1\mid k})(\hat{z}_{i\mid k} - \boldsymbol{F}\hat{z}_{i-1\mid k})^{\mathrm{T}}\end{aligned} \tag{35}$$

根据式(33)～式(35)，可得联合对数似然函数的数学期望

$$\psi(\boldsymbol{\theta} \mid \hat{\boldsymbol{\theta}}_k^{(l-1)}) \propto \ln|\boldsymbol{\Sigma}_0| + k\ln\phi^2 + k\ln|\boldsymbol{Q}| + \frac{1}{\phi^2}\sum_{i=1}^{k}\boldsymbol{\Gamma}_{1,i} + \mathrm{tr}\{\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Gamma}_2\} + \mathrm{tr}\left\{\boldsymbol{Q}^{-1}\sum_{i=1}^{k}\boldsymbol{\Gamma}_{3,i}\right\} \tag{36}$$

2) M 步

EM 算法第 l 步时未知参数 $\boldsymbol{\theta}$ 的估计可利用对式(36)取偏导数并令偏导数为 0 求得，即

$$\frac{\partial\psi(\boldsymbol{\theta} \mid \hat{\boldsymbol{\theta}}_k^{(l-1)})}{\partial\boldsymbol{\theta}} = 0 \tag{37}$$

求解可得参数 $\boldsymbol{\theta}$ 的在线估计值，即

北航学报 赠阅

$$\begin{cases}\boldsymbol{\mu}_0 = \hat{\boldsymbol{z}}_{0\mid k}\\ \boldsymbol{\Sigma}_0 = \boldsymbol{P}_{0\mid k}\\ \boldsymbol{Q} = \dfrac{1}{k}\sum\limits_{i=1}^{k}\boldsymbol{\Gamma}_{3,i}\\ \phi^2 = \dfrac{1}{k}\sum\limits_{i=1}^{k}\Gamma_{1,i}\end{cases} \tag{38}$$

至此,得到第 l 步的未知参数估计值即完成一次迭代 $\hat{\boldsymbol{\theta}}_k^{(l-1)}\rightarrow\hat{\boldsymbol{\theta}}_k^{(l)}$ 后,将此估计值代入 KF 方程和 RTS 平滑算法更新隐藏状态,然后再执行EM 算法,直至所有未知参数收敛至稳定值为止,即满足 $|\psi(\hat{\boldsymbol{\theta}}_k^{(l)}\mid\hat{\boldsymbol{\theta}}_k^{(l-1)},\boldsymbol{Y}_k)-\psi(\hat{\boldsymbol{\theta}}_k^{(l-1)}\mid\hat{\boldsymbol{\theta}}_k^{(l-1)},\boldsymbol{Y}_k)|<\varepsilon_1$ 或 $|\hat{\boldsymbol{\theta}}_k^{(l)}-\hat{\boldsymbol{\theta}}_k^{(l-1)}|<\varepsilon_2$ 时停止迭代(ε_1、ε_2 为较小的正数)。

4　RUL 实时预测

根据式(2)可知,RUL 定义为从当前时刻至退化状态首次到达失效阈值 w 的时间间隔。由于 x_k 和 η_k 都是随机变量,所以 RUL 的 PDF 不再服从 IG 分布,为实时预测 RUL 的 PDF,下面进行 k 时刻 RUL 计算的推导。根据 Wiener 过程性质知,WSSM 中以 x_k、η_k 为条件的 RUL 的 PDF 服从 IG 分布,即

$$(T_k\mid\eta_k,x_k,\boldsymbol{Y}_k)\sim\mathrm{IG}\left(\frac{w-x_k}{\eta_k},\frac{(w-x_k)^2}{\sigma^2}\right) \tag{39}$$

式中:T_k 为第 k 个监测时刻的剩余寿命,其 PDF 可以写为

$$f(T_k\mid\eta_k,x_k,\boldsymbol{Y}_k)=\sqrt{\frac{(w-x_k)^2}{2\pi T_k^3\sigma^2}}\exp\left(-\frac{(w-\eta_kT_k-x_k)^2}{2\sigma^2T_k}\right) \tag{40}$$

由于 $\boldsymbol{z}_k=(\eta_k,x_k)^{\mathrm{T}}$ 是双变量,易知其服从二维正态分布 $\boldsymbol{z}_k\sim N(\hat{\boldsymbol{z}}_{k\mid k},\boldsymbol{P}_{k\mid k})$,即

$$\begin{pmatrix}\eta_k\\ x_k\end{pmatrix}\sim N\left(\begin{pmatrix}\hat{\eta}_{k\mid k}\\ \hat{x}_{k\mid k}\end{pmatrix},\begin{pmatrix}\sigma_\eta^2 & \kappa\\ \kappa & \sigma_x^2\end{pmatrix}\right)$$

其 PDF 可展开为

$$f(\eta_k,x_k)=\frac{1}{2\pi\sigma_\eta\sigma_x\sqrt{1-\rho^2}}\cdot\exp\left(-\frac{1}{2\sqrt{1-\rho^2}}\left[\left(\frac{\eta_k-\hat{\eta}_{k\mid k}}{\sigma_\eta}\right)^2-\frac{2\rho(\eta_k-\hat{\eta}_{k\mid k})(x_k-\hat{x}_{k\mid k})}{\sigma_\eta\sigma_x}+\left(\frac{x_k-\hat{x}_{k\mid k}}{\sigma_x}\right)^2\right]\right) \tag{41}$$

式中:$\rho=\kappa/\sigma_\eta\sigma_x$ 为 x_k 和 η_k 的相关系数,$\kappa=\mathrm{cov}(\eta_k,x_k)$ 为 x_k 和 η_k 的协方差。根据 Bayesian 定理和全概率公式,求解积分可得 k 时刻 RUL 的 PDF 解析表达式为

$$\begin{aligned}f(T_k\mid\boldsymbol{Y}_k)&=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}f(T_k\mid\eta_k,x_k,\boldsymbol{Y}_k)f(\eta_k,x_k)\,\mathrm{d}\eta_k\mathrm{d}x_k=[(\sigma_\eta^2T_k+\sigma^2+\sigma_\eta\sigma_x)(\omega-\hat{x}_{k\mid k})+(\sigma_\eta T_k\rho+\sigma_x)\sigma_x\hat{\eta}_{k\mid k}]/\\&\sqrt{2\pi(\sigma_x^2+\sigma_\eta^2T_k^2+\sigma^2T_k+2\sigma_\eta\sigma_xT_k\rho)^3}\cdot\exp\left(-\frac{(\omega-T_k\hat{\eta}_{k\mid k}-\hat{x}_{k\mid k})^2}{\sigma_x^2+\sigma_\eta^2T_k^2+\sigma^2T_k+2\sigma_\eta\sigma_xT_k\rho}\right)\end{aligned} \tag{42}$$

根据式(42),平行仿真系统能够实时计算 RUL 的 PDF,为视情维修决策提供数据支撑。

5　应用实例

为验证面向装备 RUL 预测的平行仿真框架的可行性和有效性,尤其是对平行仿真系统中模型演化方法即数据同化算法和参数在线估计算法及 RUL 预测的验证,基于开源数据对本文方法进行验证。

5.1　开源退化数据与特征提取

作为数据驱动的模型演化范式,装备性能退化数据是面向装备 RUL 预测平行仿真的研究基础。本文拟采用美国 NASA 官方网站提供的 NSF I/UCR 中心的 IMS(Intelligent Maintenance Systems)轴承全寿命试验数据[23]进行方法验证。此权威数据已被维修保障领域研究人员广泛用于相关理论和方法的验证。试验数据采样频率 20 kHz,数据长度 20 480,采集时间范围 2004-02-12T10:32:39—2004-02-19T06:22:39,每 10 min 采集 1 次,8 d 内共收集到 984 组数据,时间分布为连续的 164 h。在全寿命周期试验后期,轴承 1 出现了外沟槽故障。

为兼顾退化特征量对退化过程的敏感性和平稳性,选择振动信号的均方根值(Root Mean Square,RMS)作为退化特征量,其计算公式为

$$I_{\mathrm{RMS}}=\sqrt{\frac{1}{N}\sum_{i=1}^{N}e_i^2} \tag{43}$$

式中:N 为采样点数;e_i 为第 i 个采样点的振动加速度。轴承 1 的振动信号均方根值如图 3 所示。轴承 1 的 RMS 在第 532 个监测时刻发生明显变化,在第 980 个监测时刻均方根值最大,随后装备发生外沟槽故障,因此将第 532 个监测时刻作为退化状态估计的起点,失效阈值为第 980 个监测时刻对应的均方根值 0.725 mm/s^2,即 $w=0.725$。

北航学报 赠阅

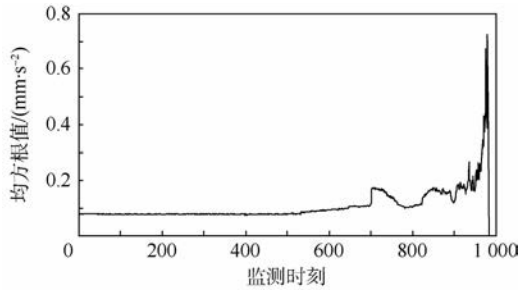


图 3 全寿命试验中轴承 1 的均方根值

Fig. 3 RMS of the 1st bearing in life test

5.2 WSSM 演化与 RUL 预测

WSSM 参数初始设置为 $x_0=0$、$\eta_0=0.01$、$\varepsilon=0.01$、$\sigma=0.01$、$\tau=1$、$\phi=0.02$。图 4 为仿真退化状态与实际观测退化状态的对比图，基于数据同化和参数在线估计的 WSSM 演化方法能有效仿真轴承的退化过程。图 4、图 5 均以第 532 个监测时刻为起点。

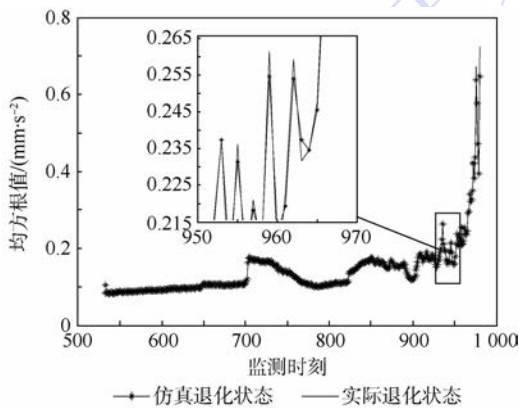


图 4 退化状态对比

Fig. 4 Comparison of degradation state

为量化退化状态对比结果，利用均方根误差（Root Mean Square Error，RMSE）来进行度量，RMSE 计算公式为

$$I_{\mathrm{RMSE}}=\sqrt{\frac{1}{m}\sum_{k=1}^{m}\left(\frac{y_k-\hat{x}_{k|k}}{y_k}\right)^2} \tag{44}$$

式中：m 为监测时刻数目。经计算，仿真退化状态与实际观测退化状态的均方根误差仅为 1.891%，充分说明平行仿真方法能有效建模和仿真轴承的性能退化过程。相应地，随着轴承性能退化数据的动态注入，平行仿真系统利用数据同化算法和 EM 算法对参数向量 $\boldsymbol{\theta}$ 进行在线演化，在第 532 ~ 980 个监测时刻的在线演化结果如图 5 所示。

图 5 结果表明，模型参数随着观测退化数据的积累能很快收敛，在第 700 个监测时刻附近和第 930 个监测时刻后，由于轴承退化状态的变化较为剧烈，使得参数发生明显变化，这也正体现

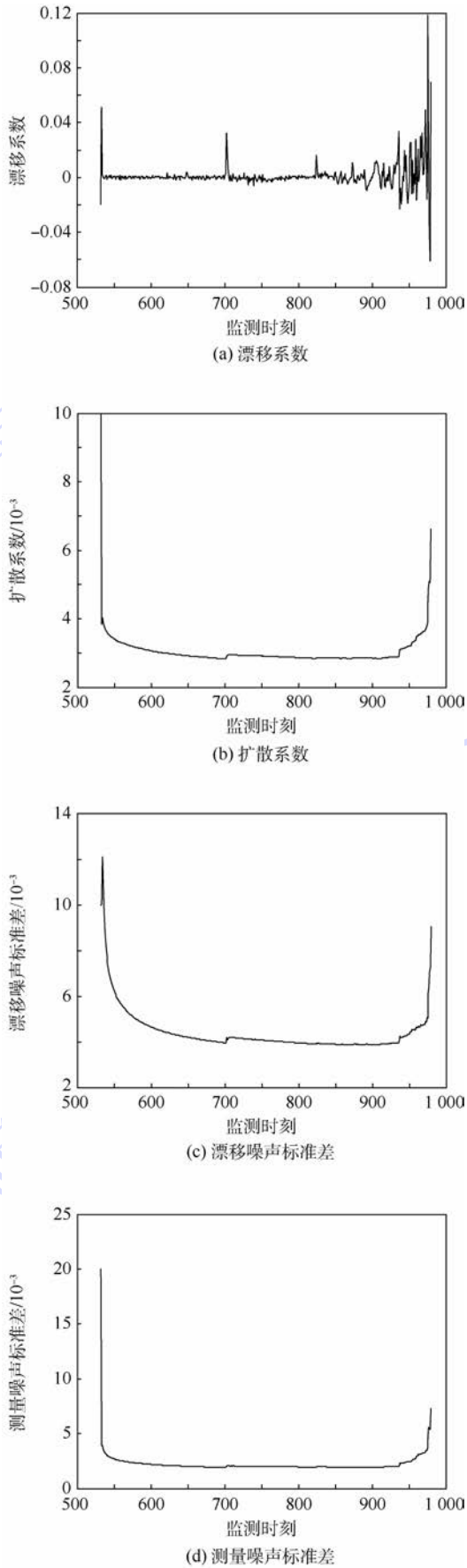


图 5 WSSM 未知参数在线演化

Fig. 5 Online evolution of unknown parameters for WSSM

北航学报 赠阅

了平行仿真中模型演化的重要性。在每一个监测时刻,一旦模型中的参数实现在线演化,对应地 RUL 的 PDF 就能通过式(42)计算得到。图 6 给出了6 个不同监测时刻上预测的 RUL 的 PDF 曲线。

图 6 中,在每一个监测时刻,实际的剩余寿命均落在剩余寿命概率密度曲线的范围内,实际剩余寿命位于 PDF 峰值对应的剩余寿命附近,而且随着轴承性能退化数据的动态注入和累积,剩余寿命的 PDF 越来越尖锐,右偏态特性越小,说明模型参数越来越准确,剩余寿命预测的不确定性不断降低,实现了轴承剩余寿命的在线、实时、自适应预测,能为装备维护人员维修决策提供重要依据。本文所用的 Wiener 过程描述的是一维退化量情形,针对多维退化量下 RUL 预测问题,本文方法仍然有较强适用性。若多特征量相互独立,可分别使用本文方法进行 RUL 预测,根据 RUL 预测的最小值安排维修决策;若多特征量相关,可先利用主成分分析(Principal Component Analysis,PCA)法得到若干反映原多特征量的独立主成分及其对应权值,进行加权求和后,得到一个无量纲的综合特征量,对此综合特征量利用本文方法仍然可以得到满意的 RUL 预测结果。

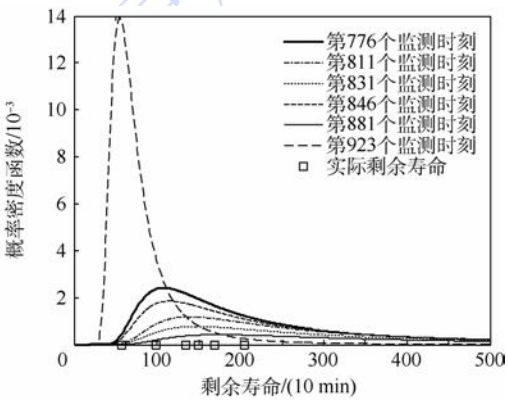


图 6　不同监测时刻上预测的 RUL 的 PDF

Fig. 6　PDF of RUL predicted at different monitoring time points

5.3　不同方法的比较

利用本次试验得到的轴承退化数据,采用 MLE 法离线估计漂移系数和扩散系数,得到 $\hat{\eta}=0.0014$、$\hat{\sigma}=0.0236$。本文方法和 MLE-IG 法得到的 6 个不同监测时刻(776、811、831、846、881、923)上 RUL 预测对比如图 7 所示。虽然 2 种方法都能将实际剩余寿命包含在变化范围内,但由于传统的 MLE-IG 法中利用同类装备的历史数据来估计漂移系数和扩散系数,一旦完成估计就不再改变,不能利用实时监测数据演化模型参数,

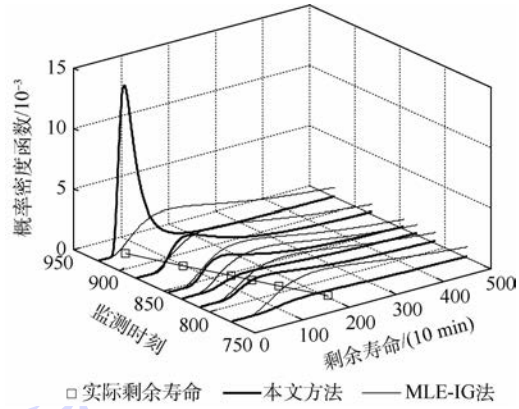


图 7　MLE-IG 法与本文方法的对比

Fig. 7　Comparison between MLE-IG method and proposed method

所以得到的剩余寿命 PDF 更加分散,右偏程度很大,表明预测的不确定性很强。对比结果表明,由于利用实时退化数据驱动 WSSM 演化,本文的方法能有效减小 RUL 预测的不确定性、提高 RUL 预测精度。

MLE-IG 法中得到第 k 个监测时刻的剩余寿命期望值 $E(T_k)=(w-y_k)/\hat{\eta}$,本文方法得到第 k 个监测时刻的剩余寿命期望值 $E(T_k)=\int_0^{\infty}T_k f(T_k\mid \boldsymbol{Y}_k)\mathrm{d}T_k$,经计算可得 2 种方法预测的剩余寿命期望值与轴承实际剩余寿命对比如图 8 所示。从定性角度看,MLE-IG 法预测的剩余寿命期望值在初始监测时刻精度较高,由于无法利用轴承最新的退化观测数据校正模型参数,使得随着时间的推移预测结果与实际剩余寿命之间的差距越来越大,而且预测的剩余寿命远高于实际剩余寿命,使之发生“欠维修”,这就意味着容易导致装备停机,造成不必要损失,甚至造成重大事故。本文采用的平行仿真方法在初始监测时刻预测精度较低,这是由于平行仿真系统获取的轴承退化观测数据较少引起的,随着时间推移,轴承的退化观测数据不断动态注入到平行仿真系统中,驱动 WSSM 演化,校正 WSSM 参数和输出,从而提高 RUL 预测精度,实现剩余寿命的在线、实时、自适应预测,为装备精确维修提供预测数据支持。为进一步量化对比结果,在每个监测时刻,定义损失函数为实际 RUL 与预测 RUL 的期望均方误差(Mean Square Error,MSE),即

$$M_k=E[(T_k-\tilde{T}_k)^2]=\int_0^{\infty}(T_k-\tilde{T}_k)^2 f(T_k\mid \boldsymbol{Y}_k)\mathrm{d}T_k \tag{45}$$

式中:$\tilde{T}_k$ 为第 k 个监测时刻的实际剩余寿命。进一步可定义总均方差(Total Mean Square Error,

北航学报 赠阅

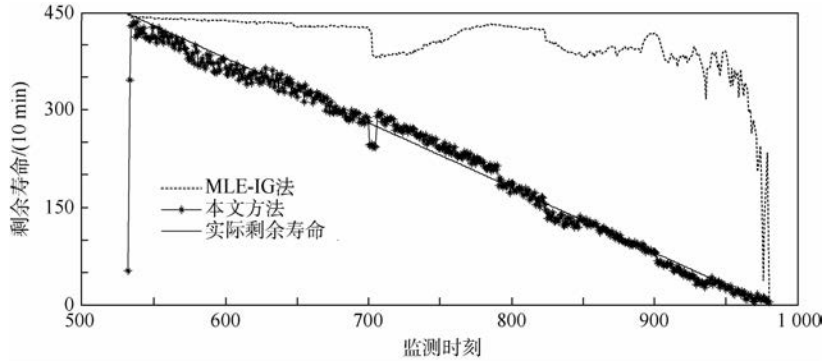


图 8 预测的剩余寿命期望值与实际剩余寿命对比

Fig. 8 Comparison between expectation of predictive RUL and bearing actual RUL

TMSE)为所有测量点 MSE 之和,即

$$I_{\mathrm{TMSE}}=\sum_{k=1}^{m}E[(T_k-\tilde{T}_k)^2]=$$

$$\sum_{k=1}^{m}\int_0^{\infty}(T_k-\tilde{T}_k)^2 f(T_k\mid \boldsymbol{Y}_k)\,\mathrm{d}T_k \tag{46}$$

总均方差越小表明 RUL 预测结果越准确,MLE-IG 法和本文方法的总均方差分别为 13.345 5、0.352 2,说明利用平行仿真方法预测 RUL 的精度远高于传统方法,充分证明了本文方法的有效性。

6 结 论

1) 本文率先将装备平行仿真技术引入装备维修保障领域,以虚实互动的思想解决装备剩余寿命自适应预测问题,提出了面向装备(RUL)预测的平行仿真框架,并以某轴承开源退化数据为数据驱动源进行了实例验证。

2) 实例仿真结果表明,平行仿真方法能有效建模并以高逼真度仿真装备的退化过程,仿真退化状态与实际观测退化状态的均方根误差仅为 1.891%。

3) 平行仿真方法不仅考虑了利用实时动态注入的退化数据演化 WSSM,包括基于 KF 算法的数据同化和基于 EM 算法的参数在线估计,同时考虑了历史监测数据和监测噪声对 RUL 的影响,使得 RUL 预测结果更为准确合理。

4) 较于传统的 MLE-IG 法,基于平行仿真方法预测装备 RUL 具有 PDF 右偏特性小、PDF 峰值接近实际 RUL 的特点,能有效提高装备 RUL 预测精度,总均方差为 0.352 2,远远小于 MLE-IG 法的 13.345 5,实现了装备 RUL 的自适应预测,同时利用平行仿真方法预测装备 RUL 还具有在线、实时的特点。

本文方法还存在一定局限性。①WSSM 适宜仿真装备性能线性退化的情况,在退化过程非线性很强情况下,预测精度会明显降低,这就需要研究面向非线性退化装备 RUL 预测的平行仿真技术;②受装备复杂度和环境因素等影响,装备性能退化过程具有较强的不确定性,呈现出多阶段性的特点,这就要求利用实时退化数据自适应选择适宜仿真模型,即需要研究模型自适应更替技术。

致谢 感谢北京仿真中心航天系统仿真重点实验室蔡继红研究员和北京航空航天大学自动化科学与电气工程学院宋晓教授对本文的贡献。

参考文献(References)

[1] 李伟,李宗毅,翟中敏. 嵌入式仿真技术及其在军事上的应用[J]. 系统仿真学报,2009,21(21):6968-6971.
LI W,LI Z Y,ZHAI Z M. Embedded simulation and its application in military[J]. Journal of System Simulation,2009,21(21):6968-6971(in Chinese).

[2] DAVIS W J. On-line simulation:Need and evolving research requirements[M]// BANKS J. Handbook of simulation:Principles,methodology,advances,applications and practices. Hoboken:Wiley,1998:465-516.

[3] AYDT H,TURNER S J,CAI W T,et al. Research issues in symbiotic simulation[C]// Proceedings of the 2010 Winter Simulation Conference. Piscataway,NJ:IEEE Press,2010:1213-1222.

[4] CHEN B. KD-ACP:A software framework for social computing in emergency management[J]. Mathematical Problems in Engineering,2014,21(3):35-38.

[5] 邱晓刚,胡艮胜. 面向辅助决策的平行系统思考[J]. 指挥与控制学报,2016,2(3):230-233.
QIU X G,HU G S. The consideration of parallel system for decision support[J]. Journal of Command and Control,2016,2(3):230-233(in Chinese).

[6] GEBRAEEL N,ELWANY A,PAN J. Residual life predictions in the absence of prior degradation knowledge[J]. IEEE Transactions on Reliability,2009,58(1):106-117.

[7] SI X S,WANG W B,HU C H,et al. Remaining useful life esti-

北航学报 赠阅

mation-A review on the statistical data driven approaches[J]. European Journal of Operational Research, 2011, 213(1): 1-14.

[8] 葛承垄,朱元昌,邸彦强,等. 装备平行仿真技术的基础理论问题[J]. 系统工程与电子技术,2017,39(5):1169-1177.
GE C L, ZHU Y C, DI Y Q, et al. Basic theoretical issues of equipment parallel simulation technology[J]. System Engineering and Electronics, 2017, 39(5): 1169-1177(in Chinese).

[9] 王飞跃. 平行系统方法与复杂系统的管理和控制[J]. 控制与决策,2004,19(5):485-489.
WANG F Y. Parallel system methods for management and control of complex systems[J]. Control & Decision, 2004, 19(5): 485-489(in Chinese).

[10] WANG F Y. Parallel control and management for intelligent transportation systems: Concepts, architectures, and applications[J]. IEEE Transactions on Intelligent Transportation Systems, 2010, 11(3): 630-638.

[11] DAREMA F. Dynamic data driven application systems[EB/OL]. [2016-07-01]. http://www.cise.nsf.gov/dddas.

[12] HUANG Y, VERBRAECK A. A dynamic data-driven approach for rail transport system simulation[C]// Proceedings of the 2009 Winter Simulation Conference. Piscataway, NJ: IEEE Press, 2009: 2553-2562.

[13] GU F. Dynamic data driven application system for wildfire spread simulation[D]. Atlanta: Georgia State University, 2010.

[14] FUJIMOTO R, LUNCEFORD D, PAGE E, et al. Grand challenges for modeling and simulation: 350[R]. Schloss Dagstuhl: Parallel/Distributed Simulation Working Group, 2002.

[15] 葛承垄,朱元昌,邸彦强,等. 装备平行仿真理论框架研究[J]. 指挥与控制学报,2017,3(1):48-56.
GE C L, ZHU Y C, DI Y Q, et al. Theoretical framework for equipment parallel simulation[J]. Journal of Command and Control, 2017, 3(1): 48-56(in Chinese).

[16] 陈淼,孙强,王东科. 平行系统在反导指挥控制中的应用[J]. 指挥控制与仿真,2017,39(1):70-72.
CHEN M, SUN Q, WANG D K. Application of parallel system method in command and control for anti-missile system[J]. Command Control & Simulation, 2017, 39(1): 70-72(in Chinese).

[17] 葛承垄,朱元昌,邸彦强,等. 装备精确维修平行仿真系统及关键技术研究[J]. 现代防御技术,2016,44(6):162-168.
GE C L, ZHU Y C, DI Y Q, et al. Research on parallel simulation system and essential technology for equipment precision maintenance[J]. Modern Defence Technology, 2016, 44(6): 162-168(in Chinese).

[18] 毛少杰,周芳,楚威,等. 面向指挥决策支持的平行仿真系统研究[J]. 指挥与控制学报,2016,2(4):315-321.
MAO S J, ZHOU F, CHU W, et al. Parallel simulation systems for command and decision support[J]. Journal of Command and Control, 2016, 2(4): 315-321(in Chinese).

[19] 孙黎阳,楚威,毛少杰,等. 面向 C^4ISR 系统决策支持的平行仿真框架[J]. 指挥信息系统与技术,2015,6(3):56-61.
SUN L Y, CHU W, MAO S J, et al. Parallel simulation framework for C^4ISR system command decision support[J]. Command Information System and Technology, 2015, 6(3): 56-61(in Chinese).

[20] 窦林涛,初阳,周玉芳,等. 平行仿真技术在指控系统中的应用构想[J]. 指挥控制与仿真,2017,39(1):62-69.
DOU L T, CHU Y, ZHOU Y F, et al. Conception of the application of parallel simulation technology in command and control system[J]. Command Control & Simulation, 2017, 39(1): 62-69(in Chinese).

[21] 王会霞. 平行仿真技术研究[J]. 航天控制,2016,34(6):64-67.
WANG H X. Research on parallel simulation technology[J]. Aerospace Control, 2016, 34(6): 64-67(in Chinese).

[22] KALMAN R E. A new approach to linear filtering and prediction problems[J]. Transactions of the ASME-Journal of Basic Engineering, 1960, 82(1): 35-45.

[23] MORTADA M. Diagnosis of rotor bearings using logical analysis of data[J]. Journal of Quality in Maintenance Engineering, 2011, 17(4): 371-397.

作者简介:

葛承垄　男,博士研究生。主要研究方向:装备平行仿真及其应用。

朱元昌　男,博士,教授。主要研究方向:系统仿真。

邸彦强　男,博士,副教授。主要研究方向:系统仿真。

北京航空航天大学学报

北航学报 赠阅

Equipment RUL prediction oriented parallel simulation technology

GE Chenglong, ZHU Yuanchang, DI Yanqiang*

(Department of Electronic and Optics Engineering, Ordnance Engineering College, Shijiazhuang 050003, China)

Abstract: As an emerging simulation technology in the field of system modeling & simulation, equipment parallel simulation has become research emphasis. In the field of equipment maintenance support, the outstanding problem of equipment remaining useful life(RUL) prediction is analyzed, i.e., the stable model parameters without self-evolution ability, which has become the primary factor that hinders adaptive prediction of equipment remaining useful life. Combined with parallel systems theory, equipment remaining useful life prediction oriented parallel simulation framework is proposed on the basis of modeling analysis and Wiener state space model is taken as the basic simulation model in the framework. Driven by the dynamic implanted equipment degradation observation data, the model parameters are updated online by using expectation maximum (EM) algorithm and the data assimilation (DA) between simulation outputs and observation data is executed by using Kalman filter(KF), so as to realize dynamic evolution of the simulation model. The simulation model evolution which makes the simulation outputs close to equipment real degradation state provides high fidelity model and data for equipment remaining useful life prediction accurately. The framework is verified by the performance degradation data of a bearing. The simulation results show that the parallel simulation method can accurately simulate the equipment performance degradation process and the adaptive prediction of equipment remaining useful life is realized on the basis of the improved prediction accuracy, proving the feasibility and effectiveness of parallel simulation method.

Keywords: parallel simulation; model evolution; remaining useful life(RUL); data assimilation(DA); parameter estimation

Received: 2017-04-26; **Accepted**: 2017-07-07; **Published online**: 2017-09-15 17:09

URL: kns.cnki.net/kcms/detail/11.2625.V.20170915.1709.002.html

Foundation item: Key Project of Equipment Pre-Research Foundation of China (9140A04020115JB34011)

* **Corresponding author**. E-mail: 1049084176@qq.com

北京航空航天大学学报

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2017.0276

基于输入成形的太阳能帆板自适应滑模控制

周通, 郭宏, 徐金全*
(北京航空航天大学 自动化科学与电气工程学院, 北京 100083)

摘 要: 为提高太阳能帆板驱动系统(SADS)的角位置控制性能和抑制太阳能帆板的柔性振动,提出了一种自适应滑模控制(ASMC)与输入成形技术相结合的控制策略。该控制策略通过自适应滑模控制保证了系统在不确定性影响下的一致有界性和渐进一致有界性,从而提高了太阳能帆板驱动系统的角位置控制性能。同时,通过基于参考模型的输入成形器(IS)规划了指令轨迹,进而抑制了太阳能帆板的柔性振动。仿真结果表明了控制策略的有效性。

关 键 词: 输入成形; 自适应滑模控制(ASMC); 太阳能帆板; 驱动系统; 柔性振动
中图分类号: V442
文献标识码: A **文章编号**: 1001-5965(2018)04-0737-09

作为在轨空间站的重要组件,太阳能帆板担负着为整个空间站供电的任务。为提高太阳能转化效率,太阳能帆板驱动系统(SADS)的角位置控制吸引了很多学者的深入研究。然而,太阳能帆板是柔性体,其动态特性包含了大量与刚体运动相耦合的弱阻尼柔性模态。该特性极易在系统的对日定向过程中引起太阳能帆板的柔性振动,进而影响空间站的运动与控制。因此,寻求一种角位置控制和振动抑制方法对于太阳能帆板驱动系统显得尤为重要。

针对太阳能帆板的振动抑制问题,国内外学者做了大量研究。采用智能材料进行主动控制是一种有效抑制振动的方法。在基于智能材料的控制方法方面,提出了刚度控制[1]以及正位置反馈控制[2-3]。然而,这类方法需要在原结构上附加一系列的传感器和作动器,增加了系统的质量,同时降低了可靠性,因此该主动控制方法在空间系统的实际应用中受到一定限制[4]。另一种抑制振动的方法是输入成形,该方法将期望指令与脉冲序列在时域中进行卷积,所形成的新指令作为系统输入来控制系统的运动以消除柔性模态的振动[5]。文献[6]将输入成形器(IS)应用于太阳能帆板驱动系统的振动抑制中,提升了系统的工作性能。然而,输入成形器的设计依赖于太阳能帆板驱动系统的数学模型,由于系统存在不确定性,该数学模型难以准确建立[7]。不仅如此,输入成形器是基于线性系统理论设计的,无法直接应用于非线性系统[8-9]。

太阳能帆板驱动系统的角位置控制同样被广泛研究。文献[10]针对太阳能帆板驱动系统提出一种 H_∞ 鲁棒控制。文献[11-12]和文献[13-14]针对刚柔耦合系统分别设计了自适应控制和滑模控制,文献[15]基于耗散理论提出一种主动稳定器。以上控制方法都能在完成指定的角位置阶跃响应的同时抑制柔性振动,并且缩短了过渡过程时间,提高了系统性能对参数摄动与外部干扰的鲁棒性。然而,由于缺少对指令轨迹的设计,使得系统在瞬态过程中存在较大的柔性

收稿日期: 2017-05-02; **录用日期**: 2017-08-01; **网络出版时间**: 2017-09-25 10:01
网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20170925.1001.001.html
基金项目: 航空科学基金(2016ZC51025)
* **通信作者**. E-mail: xujinquan@buaa.edu.cn

引用格式: *周通, 郭宏, 徐金全. 基于输入成形的太阳能帆板自适应滑模控制[J]. 北京航空航天大学学报, 2018, 44(4): 737-745. ZHOU T, GUO H, XU J Q. Adaptive sliding mode control of solar array with input shaping[J]. Journal of Beijing University of Aeronautics and Astronautics, 2018, 44(4): 737-745 (in Chinese).*

北航学报 赠阅

振动。文献[16-18]针对刚柔耦合系统分别提出了带自适应律和不带自适应律的滑模控制,并且采用基于参考模型的输入成形器设计了系统的指令轨迹。该方法通过滑模控制令太阳能帆板驱动系统表现得像参考模型。这种方法的参考模型属于多维度系统,为获得输入成形器预期的振动抑制效果,滑模控制在保证系统角位置和角速度跟踪参考轨迹的同时还要保证系统的模态坐标跟踪参考模型的模态坐标,增加了控制设计的复杂性和计算成本。

基于上述讨论,本文提出一种自适应滑模控制(ASMC)与输入成形技术相结合的控制策略。该控制策略通过自适应滑模控制保证了系统在不确定性影响下的一致有界性和渐进一致有界性,从而提高了太阳能帆板驱动系统的角位置控制性能。同时,通过基于参考模型的输入成形器规划了系统的指令轨迹,进而抑制了太阳能帆板的柔性振动。本文首先建立了太阳能帆板驱动系统的数学模型,其次提出了包含输入成形器、参考模型和自适应滑模控制的控制策略,最后给出了仿真结果。

1 太阳能帆板驱动系统数学建模

太阳能帆板驱动系统包含伺服电机、减速器、太阳能帆板等,其简化结构如图 1 所示。驱动系统的轮毂上固结了 2 个对称配置的柔性帆板,电机产生力矩驱动柔性帆板旋转[19]。将轮毂视为刚体且将太阳能帆板视为欧拉-伯努利梁,利用假设模态法,可以得到帆板的横向弯曲挠度表达式如下:

$$w(x,t)=\sum_{i=1}^{n}\phi_i(x)q_i(t)=\boldsymbol{\Phi}(x)\boldsymbol{q}(t) \tag{1}$$

式中:$\boldsymbol{\Phi}(x)=[\phi_1(x)\quad\phi_2(x)\quad\cdots\quad\phi_n(x)]$,$\phi_i(x)$为横向振动的第$i$阶模态振型函数;$\boldsymbol{q}(t)=[q_1(t)\quad q_2(t)\quad\cdots\quad q_n(t)]^{\mathrm{T}}$,$q_i(t)$为横向振动的第 i 阶模态坐标。利用拉格朗日方法,可以得到太阳能帆板驱动系统的动力学表达式如下[20]:

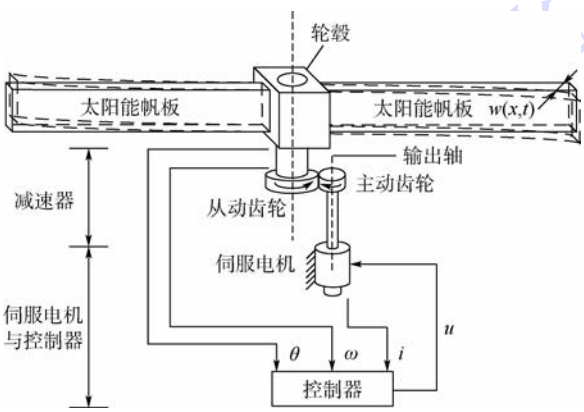


图 1 太阳能帆板驱动系统结构

Fig. 1 Structure of SADS

$$\omega=\frac{\mathrm{d}\theta}{\mathrm{d}t} \tag{2}$$

$$(J+\Delta J)\frac{\mathrm{d}\omega}{\mathrm{d}t}=T_{\mathrm{c}}-\sum_{i=1}^{n}F_i\frac{\mathrm{d}^2q_i(t)}{\mathrm{d}t^2}+T_{\mathrm{d}}(\theta,\omega,t) \tag{3}$$

$$F_i\frac{\mathrm{d}\omega}{\mathrm{d}t}+\frac{\mathrm{d}^2q_i(t)}{\mathrm{d}t^2}+2\xi_i\omega_i\frac{\mathrm{d}q_i(t)}{\mathrm{d}t}+\omega_i^2q_i(t)=0\quad i=1,2,\cdots,n \tag{4}$$

式中:θ 为系统角位置;ω 为系统角速度;J 为伺服电机、减速器、轮毂和太阳能帆板的总转动惯量;ΔJ 为转动惯量的不确定性;F_i 为第 i 阶模态的耦合系数;T_{c} 为作用在减速器输出轴上的驱动力矩;$T_{\mathrm{d}}(\theta,\omega,t)$为外界干扰力矩(如摩擦力矩);$\xi_i$ 为第 i 阶模态的阻尼比;ω_i 为第 i 阶模态的角频率;$-\sum_{i=1}^{n}F_i\frac{\mathrm{d}^2q_i(t)}{\mathrm{d}t^2}$ 为太阳能帆板横向振动对轮毂产生的耦合力矩。

需要指出的是,式(2)和式(3)构成太阳能帆板驱动系统的刚体子系统,式(4)为柔性体子系统。根据式(2)和式(3)可知,刚体子系统包含非线性项 $T_{\mathrm{d}}(\theta,\omega,t)$,是一个非线性系统。根据式(4)可知,柔性体子系统是线性的。

当表贴式永磁同步电机(SPMSM)作为系统的伺服电机时,电机的转矩方程和电压方程可以写为[21]

$$\begin{cases}T_{\mathrm{c}}=nk_{\mathrm{T}}i_{\mathrm{q}}\\ u_{\mathrm{d}}=Ri_{\mathrm{d}}+L\dfrac{\mathrm{d}i_{\mathrm{d}}}{\mathrm{d}t}-pn\omega Li_{\mathrm{q}}\\ u_{\mathrm{q}}=Ri_{\mathrm{q}}+L\dfrac{\mathrm{d}i_{\mathrm{q}}}{\mathrm{d}t}+k_{\mathrm{T}}n\omega+pn\omega Li_{\mathrm{d}}\end{cases} \tag{5}$$

式中:u_{d}、u_{q}、i_{d}、i_{q}、R 和 L 分别表示电机在 d-q 轴系下的定子电压、定子电流、定子电阻和定子电感;n 为减速器传动比;p 为电机的极对数;k_{T} 为电机的转矩常数。

2 控制系统设计

本文所设计的太阳能帆板驱动系统采用位置-电流双闭环控制结构,如图 2 所示,θ_{r}^{*}、θ_{r} 和 θ_{m} 分别为角位置期望指令、成形后的角位置指令和参考模型的角位置;ω_{m}为参考模型的角速度;i_{d}^{*} 和 i_{q}^{*} 分别为电机的 d 轴和 q 轴电流指令。其中,位置环采用所提出的包含输入成形器、参考模型和自适应滑模控制的控制策略,电流环采用 $i_{\mathrm{d}}=0$

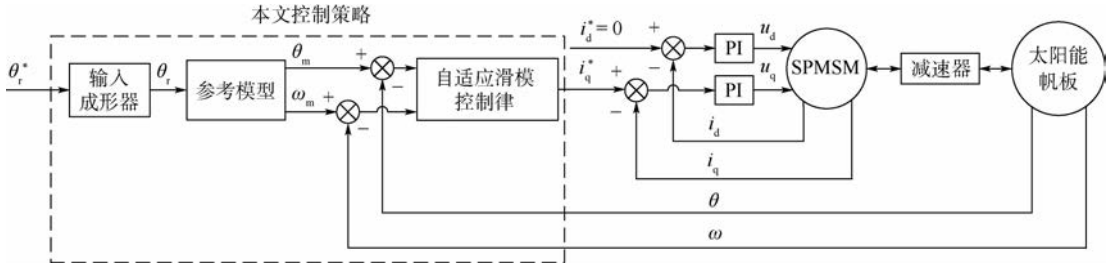


图 2　基于本文控制策略的太阳能帆板驱动系统控制结构

Fig. 2　Control structure of SADS based on proposed control strategy

的比例-积分(PI)控制策略[22]。由于在实际系统中电流环带宽远大于位置环带宽,在位置环的控制设计中,电流环的影响将被忽略。

本文的动机是通过输入成形器来抑制系统的柔性振动。然而,输入成形器是基于线性系统理论设计的,它不能直接应用于非线性系统。为了解决这个问题,本文提出了一个线性的参考模型,并利用自适应滑模控制使得太阳能帆板驱动系统的刚体子系统式(2)~式(3)表现得像这个线性的参考模型。因此,闭环的太阳能帆板驱动系统可以由这个线性的参考模型和线性的柔性体子系统式(4)来代表。也就是说,闭环的太阳能帆板驱动系统表现得像一个线性系统。在此基础上,输入成形器可以根据线性的参考模型和线性的柔性体子系统式(4)来设计。

2.1　参考模型设计

由于在实际系统中耦合力矩 $-\sum_{i=1}^{n} F_i \frac{\mathrm{d}^2 q_i(t)}{\mathrm{d}t^2}$ 是有界的,耦合力矩可以被看作减速器输出轴上的一个干扰力矩。在此基础上,将刚体子系统式(2)~式(3)改写为

$$\begin{cases} \dfrac{\mathrm{d}\theta}{\mathrm{d}t} = \omega \\ \dfrac{\mathrm{d}\omega}{\mathrm{d}t} = \dfrac{1}{J} T_{\mathrm{c}} + D(t) \end{cases} \tag{6}$$

式中:

$$D(t) = \frac{1}{J}\left(T_{\mathrm{d}}(\theta,\omega,t) - \sum_{i=1}^{n} F_i \frac{\mathrm{d}^2 q_i(t)}{\mathrm{d}t^2} - \Delta J \frac{\mathrm{d}\omega}{\mathrm{d}t} \right) \tag{7}$$

对于实际系统,外界干扰力矩 $T_{\mathrm{d}}(\theta,\omega,t)$ 和由转动惯量不确定引起的摄动项 $-\Delta J \dfrac{\mathrm{d}\omega}{\mathrm{d}t}$ 也是有界的[23-24]。因此,$D(t)$ 是有界的。

根据以上分析,针对标称系统

$$\begin{cases} \dfrac{\mathrm{d}\theta}{\mathrm{d}t} = \omega \\ \dfrac{\mathrm{d}\omega}{\mathrm{d}t} = \dfrac{1}{J} T_{\mathrm{c}} \end{cases} \tag{8}$$

设计如下比例-微分(PD)控制:

$$T_{\mathrm{c}} = K_{\mathrm{P}}(\theta_{\mathrm{r}} - \theta_{\mathrm{m}}) + K_{\mathrm{D}}(\omega_{\mathrm{r}} - \omega_{\mathrm{m}}) \tag{9}$$

式中:角速度指令 $\omega_{\mathrm{r}} = 0$;K_{P} 为比例系数;K_{D} 为微分系数。将式(9)代入式(8),可以得到参考模型如下:

$$\begin{cases} \dfrac{\mathrm{d}\theta_{\mathrm{m}}}{\mathrm{d}t} = \omega_{\mathrm{m}} \\ \dfrac{\mathrm{d}\omega_{\mathrm{m}}}{\mathrm{d}t} = \dfrac{1}{J}\left[K_{\mathrm{P}}(\theta_{\mathrm{r}} - \theta_{\mathrm{m}}) - K_{\mathrm{D}}\omega_{\mathrm{m}} \right] \end{cases} \tag{10}$$

2.2　输入成形器设计

为了抑制系统的柔性振动,本节针对闭环太阳能帆板驱动系统设计了输入成形器。

输入成形技术的基本原理如下[25]:将脉冲序列与任意的输入指令相卷积,所形成的新指令作为系统输入来控制系统的运动,如图 3 所示。相比于原指令,所形成的新指令将引起更小的柔性振动。其中,脉冲序列称为输入成形器。

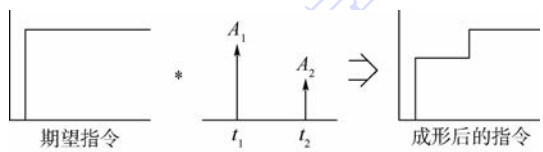


图 3　输入成形器原理图

Fig. 3　Schematic diagram of IS

由于零振动(ZV)输入成形器具有结构简单的特点,本文采用考虑前两阶模态的具有 4 个时间脉冲的零振动输入成形器,其表达式如下[26]:

$$h(t) = \sum_{i=1}^{4} A_i \delta(t - t_i) \tag{11}$$

式中:

$$\begin{cases} A_1 = \dfrac{1}{(1+K_1)(1+K_2)} & t_1 = 0 \\ A_2 = \dfrac{K_2}{(1+K_1)(1+K_2)} & t_2 = \dfrac{\pi}{\omega_{\mathrm{n2}}\sqrt{1-\zeta_2^2}} \\ A_3 = \dfrac{K_1}{(1+K_1)(1+K_2)} & t_3 = \dfrac{\pi}{\omega_{\mathrm{n1}}\sqrt{1-\zeta_1^2}} \\ A_4 = \dfrac{K_1 K_2}{(1+K_1)(1+K_2)} & \\ & t_4 = \dfrac{\pi}{\omega_{\mathrm{n1}}\sqrt{1-\zeta_1^2}} + \dfrac{\pi}{\omega_{\mathrm{n2}}\sqrt{1-\zeta_2^2}} \end{cases} \tag{12}$$

北航学报 赠阅

其中：A_i 和 $t_i(i=1,2,3,4)$ 分别为脉冲的幅值和作用时间；ω_{ni} 和 $\zeta_i(i=1,2)$ 分别为闭环太阳能帆板驱动系统的固有频率和阻尼比，并将在本节下文中推导。

$$\begin{cases} K_1 = \exp\left(-\dfrac{\pi\zeta_1}{\sqrt{1-\zeta_1^2}}\right) \\ K_2 = \exp\left(-\dfrac{\pi\zeta_2}{\sqrt{1-\zeta_2^2}}\right) \end{cases} \tag{13}$$

由于刚体子系统式(2)～式(3)表现得像参考模型式(10)，结合柔性体子系统式(4)，闭环太阳能帆板驱动系统的运动方程可以表示为

$$\begin{cases} \omega = \dfrac{\mathrm{d}\theta}{\mathrm{d}t} \\ J\dfrac{\mathrm{d}\omega}{\mathrm{d}t} = K_{\mathrm{P}}(\theta_{\mathrm{r}}-\theta) - K_{\mathrm{D}}\omega \\ F_i\dfrac{\mathrm{d}\omega}{\mathrm{d}t} + \dfrac{\mathrm{d}^2 q_i(t)}{\mathrm{d}t^2} + 2\xi_i\omega_i\dfrac{\mathrm{d}q_i(t)}{\mathrm{d}t} + \omega_i^2 q_i(t) = 0 \\ \quad i = 1,2,\cdots,n \end{cases} \tag{14}$$

由于在实际系统中，高阶模态不易被激发，因此通常对模态进行截断，本文将取系统前两阶模态(即 $n=2$)进行研究。在此基础上，式(14)可以改写成形式如下的状态方程：

$$\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}u \tag{15}$$

式中：

$$\boldsymbol{A} = \begin{bmatrix} -\dfrac{K_{\mathrm{D}}}{J} & 0 & 0 & -\dfrac{K_{\mathrm{P}}}{J} & 0 & 0 \\ \dfrac{F_1K_{\mathrm{D}}}{J} & -2\xi_1\omega_1 & 0 & \dfrac{F_1K_{\mathrm{P}}}{J} & -\omega_1^2 & 0 \\ \dfrac{F_2K_{\mathrm{D}}}{J} & 0 & -2\xi_2\omega_2 & \dfrac{F_2K_{\mathrm{P}}}{J} & 0 & -\omega_2^2 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \tag{16}$$

$$\boldsymbol{B} = \begin{bmatrix} \dfrac{K_{\mathrm{P}}}{J} & -\dfrac{F_1K_{\mathrm{P}}}{J} & -\dfrac{F_2K_{\mathrm{P}}}{J} & 0 & 0 & 0 \end{bmatrix}^{\mathrm{T}} \tag{17}$$

状态向量为

$$\boldsymbol{x} = \begin{bmatrix} \omega & \dfrac{\mathrm{d}q_1(t)}{\mathrm{d}t} & \dfrac{\mathrm{d}q_2(t)}{\mathrm{d}t} & \theta & q_1 & q_2 \end{bmatrix}^{\mathrm{T}} \tag{18}$$

输入为

$$u = \theta_{\mathrm{r}} \tag{19}$$

根据式(16)，可得矩阵 $\boldsymbol{A}$ 的复共轭特征值为 $-\omega_1\xi_1 \pm \omega_1\sqrt{1-\xi_1^2}\,\mathrm{i}$ 和 $-\omega_2\xi_2 \pm \omega_2\sqrt{1-\xi_2^2}\,\mathrm{i}$。

因此闭环系统式(14)的固有频率 $\omega_{ni}=\omega_i$，阻尼比 $\zeta_i=\xi_i(i=1,2)$。

2.3 自适应滑模控制设计

为了保证刚体子系统式(2)～式(3)表现得像参考模型式(10)并改善角位置控制性能，本节设计了自适应滑模控制。

首先，对系统作如下假设：

假设 1 存在正常数 $\bar{D}$，使得 $\bar{D} > |D(t)|$。

定义角位置误差 $e_\theta = \theta_{\mathrm{m}} - \theta$ 和角速度误差 $e_\omega = \omega_{\mathrm{m}} - \omega$，并结合式(6)和式(10)可得

$$\frac{\mathrm{d}e_\theta}{\mathrm{d}t} = e_\omega \tag{20}$$

$$\frac{\mathrm{d}e_\omega}{\mathrm{d}t} = \frac{K_{\mathrm{P}}\theta_{\mathrm{r}}}{J} - \frac{K_{\mathrm{P}}\theta_{\mathrm{m}}}{J} - \frac{K_{\mathrm{D}}\omega_{\mathrm{m}}}{J} - D(t) - \frac{T_{\mathrm{c}}}{J} \tag{21}$$

定义滑模函数为

$$S = \boldsymbol{G}\boldsymbol{x}_{\mathrm{e}} = \begin{bmatrix} k & 1 \end{bmatrix}\begin{bmatrix} e_\theta \\ e_\omega \end{bmatrix} \tag{22}$$

式中：$\boldsymbol{G} = [k \quad 1]$；$\boldsymbol{x}_{\mathrm{e}} = [e_\theta \quad e_\omega]^{\mathrm{T}}$；$k$ 为正常数。

因此，自适应滑模控制可以设计为

$$T_{\mathrm{c}}^* = Jke_\omega + K_{\mathrm{P}}\theta_{\mathrm{r}} - K_{\mathrm{P}}\theta_{\mathrm{m}} - K_{\mathrm{D}}\omega_{\mathrm{m}} + J\hat{D}\mathrm{sgn}(S) \tag{23}$$

式中：$\hat{D}$ 为 $\bar{D}$ 的估计值，其更新律为

$$\dot{\hat{D}} = \kappa(|S| - \chi\hat{D}) \tag{24}$$

其中：κ、χ 均为正常数。根据式(24)，易知 $\hat{D} \geqslant 0$。

下面定理给出了所设计的自适应滑模控制对系统稳定性的影响。

定理 1 考虑控制系统式(20)～式(21)，在满足假设 1 的条件下，所设计的控制律式(23)～式(24)能够保证系统满足以下性能。

1) 一致有界：对于任意 $\tau>0$，存在 $z(\tau)<\infty$，使得当 $\|\boldsymbol{\sigma}(t_0)\| \leqslant \tau$ 时，对于任意 $t \geqslant t_0$，$\|\boldsymbol{\sigma}(t)\| \leqslant z(\tau)$。

2) 渐进一致有界：对于任意 $\tau>0$，$\|\boldsymbol{\sigma}(t_0)\| \leqslant \tau$，存在 $\underline{z}>0$，使得对于任意 $\bar{z} > \underline{z}$，当 $t \geqslant t_0 + T(\bar{z},\tau)$，且 $T(\bar{z},\tau)<\infty$ 时，$\|\boldsymbol{\sigma}(t)\| \leqslant \bar{z}$。

证明 本文利用 Lyapunov minimax 方法证明上述定理。

首先，选择控制系统的 Lyapunov 函数为

$$V = \frac{1}{2}S^2 + \frac{1}{2\kappa}(\hat{D}-\bar{D})^2 \tag{25}$$

对 Lyapunov 函数进行一次求导并代入式(20)～式(22)，可得

$$\dot{V} = S\dot{S} + \frac{1}{\kappa}(\hat{D}-\bar{D})\dot{\hat{D}} = S\left(k\frac{\mathrm{d}e_\theta}{\mathrm{d}t} + \frac{\mathrm{d}e_\omega}{\mathrm{d}t}\right) + \frac{1}{\kappa}(\hat{D}-\bar{D})\dot{\hat{D}} =$$

北航学报 赠阅

$$S\left(ke_\omega+\frac{K_P\theta_r}{J}-\frac{K_P\theta_m}{J}-\frac{K_D\omega_m}{J}-D(t)-\frac{T_c}{J}\right)+\frac{1}{\kappa}(\hat{D}-\bar{D})\dot{\hat{D}} \tag{26}$$

将式(23)和式(24)代入式(26)可得

$$\dot{V}=S(-D(t)-\hat{D}\mathrm{sgn}(S))+\frac{1}{\kappa}(\hat{D}-\bar{D})\kappa(|S|-\chi\hat{D})=-SD(t)-S\hat{D}\mathrm{sgn}(S)+\hat{D}|S|-\chi\hat{D}^2-|S|\bar{D}+\chi\bar{D}\hat{D}=-SD(t)-|S|\hat{D}+\hat{D}|S|-\chi\hat{D}^2-|S|\hat{D}+\chi\bar{D}\hat{D}=-SD(t)-\chi\hat{D}^2-|S|\bar{D}+\chi\bar{D}\hat{D}\leqslant|S||D(t)|-\chi\hat{D}^2-|S|\bar{D}+\chi\bar{D}\hat{D}=\underbrace{-|S|(\bar{D}-|D(t)|)-\chi\left(\hat{D}-\frac{1}{2}\bar{D}\right)^2}_{\leqslant 0}+\underbrace{\frac{1}{4}\chi\bar{D}^2}_{>0} \tag{27}$$

根据式(27)和文献[27-28]的结论可知，e_θ 和 e_ω 是一致有界和渐进一致有界的。因此，本文所设计的自适应滑模控制律式(23)～式(24)可以保证系统式(20)～式(21)在不确定性影响下的一致有界性和渐进一致有界性。　**证毕**

需要指出的是，根据自适应律式(24)可知，当滑模函数 S 趋近于 0，$\hat{D}$ 也趋近于 0。因此，自适应律将实时减小滑模切换增益 $J\hat{D}$，进而降低滑模抖振。为了进一步降低滑模抖振，在实际应用中可以将符号函数替换为饱和函数。

3　仿真结果

为了验证本文所提出的控制策略的有效性，本节以国际空间站的太阳能帆板驱动系统为例进行了数值仿真。系统的主要参数如表 1 所示[29-30]。

本文用振动能量衡量振动的强弱，并将其定义为

$$E=\sum_{i=1}^{n}\left(\frac{\mathrm{d}q_i(t)}{\mathrm{d}t}\right)^2+\sum_{i=1}^{n}\omega_i^2q_i^2 \tag{28}$$

假设外界干扰力矩 $T_d(t)$ 由随机干扰力矩 $T_r(t)$ 和摩擦力矩 $T_f(t)$ 构成，则

$$T_d(t)=T_r(t)+T_f(t) \tag{29}$$

式中：

$$T_f(t)=\begin{cases}-T_s\mathrm{sgn}(\omega) & \omega\neq 0\\ -T_u & \omega=0\ \text{且}\ |T_u|\leqslant T_{max}\\ -T_{max}\mathrm{sgn}(T_u) & \text{其他}\end{cases}$$

表 1　太阳能帆板驱动系统仿真参数[29-30]

Table 1　Simulation parameters of SADS[29-30]

参　数	数　值
定子电阻/Ω	28
定子电感/H	0.134
极对数	12
转矩常数/(N·m·A^{-1})	9.22
额定电流/A	1.4
额定功率/W	22
传动比	325
最大静摩擦力矩/(N·m)	404.54
第一阶模态耦合系数	188.7
第二阶模态耦合系数	30.1
第一阶模态阻尼比	0.01
第二阶模态阻尼比	0.01
第一阶模态角频率/(rad·s^{-1})	1.789
第二阶模态角频率/(rad·s^{-1})	11.21
转动惯量/(kg·m^2)	1.7×10^6
滑动摩擦力矩/(N·m)	324.31

$$T_u=T_c+T_r(t)-\sum_{i=1}^{n}F_i\frac{\mathrm{d}^2q_i(t)}{\mathrm{d}t^2}$$

式中：T_u 为作用在减速器输出轴上的主动力矩之和。根据表 1 可知，最大静摩擦力矩 T_{max} 为 404.54 N·m，滑动摩擦力矩 T_s 为 324.31 N·m。将随机干扰力矩 $T_r(t)$ 的上界设为 85 N·m[31]。此外，将转动惯量不确定性表示为

$$\Delta J=0.1J\sin(4t)+0.1J\sin(8t) \tag{30}$$

为了说明本文所提控制策略的有效性，现对比例-积分-微分控制策略(PID)、自适应滑模控制策略、自适应滑模控制与输入成形相结合的控制策略这 3 种情况进行仿真。其中，PID 控制采用位置-转速-电流三闭环控制结构，参数通过根轨迹方法设计；ASMC 的参数通过经验选取；控制器参数如表 2 所示。

为了公平对比，电流环控制参数和随机干扰力矩 $T_r(t)$ 波形(如图 4 所示)在 3 种仿真情况中保持不变，3 个控制器采样周期均设定为 80 μs，系统的前向通道和反馈通道的时间延迟均设定为 10 ms。

图 5 为角位置阶跃给定情况下分别采用 PID 控制、ASMC 和 ASMC + IS 时的位置响应波形图。由图 5 可见，采用 ASMC + IS 时，系统调节时间为 150.7 s，超调量为 0.002 9 rad，超调过后，角位置轨线进入与位置指令最大偏差为 0.000 1 rad 的区域。采用 ASMC 时，超调量和最大偏差与采用 ASMC + IS 时相同，调节时间降低至 149.6 s。采用 PID 控制时，调节时间和超调量分别为216.5 s 和 0.23 rad，当角位置轨线进入2% 误差带后，轨

北航学报 赠阅

表 2 3 种控制器的仿真参数

Table 2 Simulation parameters of three controllers

控制器	参数	数值
PID 控制	比例系数(位置环)	0.075
	积分系数(位置环)	0.000 8
	微分系数(位置环)	2
	比例系数(速度环)	33.3
	积分系数(速度环)	0.1
	微分系数(速度环)	0
	比例系数(电流环)	500
	积分系数(电流环)	6 000
	微分系数(电流环)	0
ASMC	K_P	1 530
	K_D	89 760
	k	3
	κ	150
	χ	3
IS	A_1	0.26
	A_2	0.25
	A_3	0.25
	A_4	0.24
	t_1/s	0
	t_2/s	0.28
	t_3/s	1.76
	t_4/s	2.03

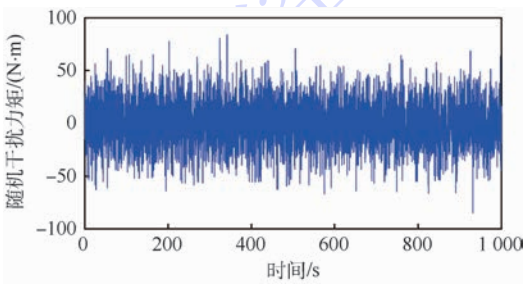


图 4 随机干扰力矩波形

Fig. 4 Random disturbance torque waveform

线与位置指令最大偏差为 0.004 5 rad。

图 6 为采用 3 种控制策略时的角速度响应波形图。由图 6 可见,采用 ASMC + IS 时,转速峰值为 0.012 rad/s;采用 ASMC 和 PID 控制时,转速峰值分别为 0.012 和 0.041 rad/s。

图 7 为采用 3 种控制策略时的振动能量波形图。由图 7 可见,采用 ASMC + IS 时,振动能量峰值为 0.01 J;而采用 ASMC 和 PID 控制策略时,振动能量峰值分别为 0.033 和 0.235 J。

图 8 为采用 3 种控制策略时的驱动力矩波形图。由图 8 可见,采用 ASMC + IS 时,驱动力矩峰值为 1 771 N · m;而采用 ASMC 和 PID 控制策略时,驱动力矩峰值分别为 1 878 和 4 195 N · m。

综上可知,相比于 PID 控制,ASMC + IS 有效改善了调节时间、超调量、跟踪误差、转速峰值和驱动力矩峰值,并将振动能量峰值降低了95.7%。

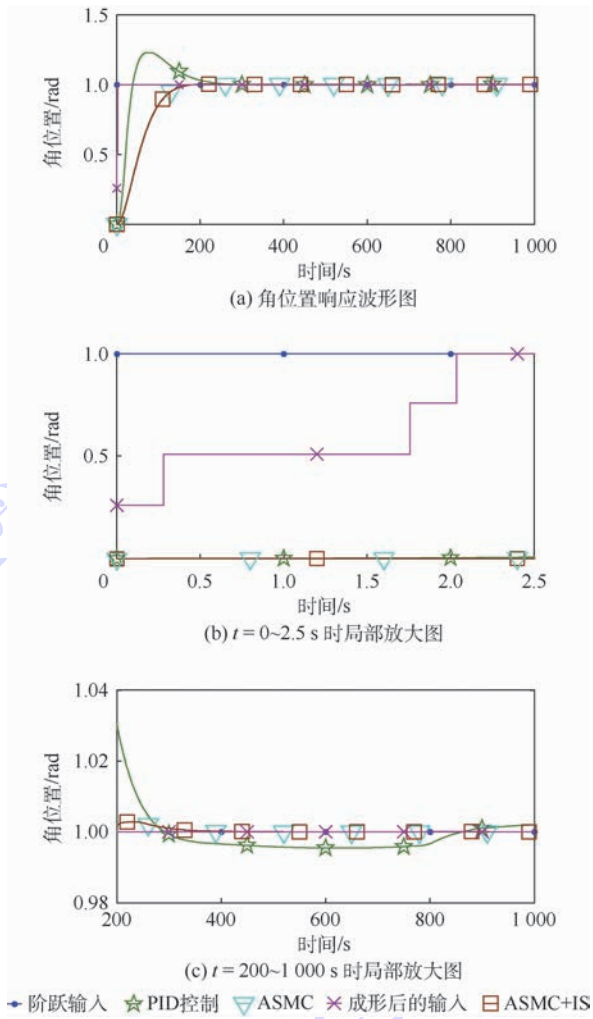


图 5 角位置响应波形图及其局部放大图

Fig. 5 Oscillogram of time response of angular position and its partial enlarged views

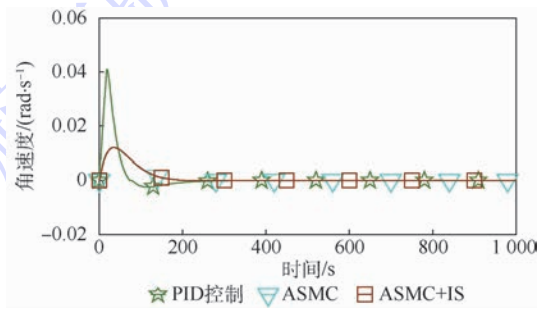


图 6 角速度响应波形图

Fig. 6 Oscillogram of time response of angular velocity

因此,ASMC + IS 比 PID 控制具有更强的角位置控制能力和振动抑制能力。相比于 ASMC,ASMC + IS 在超调量、跟踪误差和转速峰值方面具有相同的性能。尽管采用 ASMC + IS 时调节时间增加了 0.67%,但是振动能量和驱动力矩峰值分别降低了69.7% 和5.7%。因此,在角位置控制和振动抑制方面,ASMC + IS 比 ASMC 具有更

北航学报 赠阅

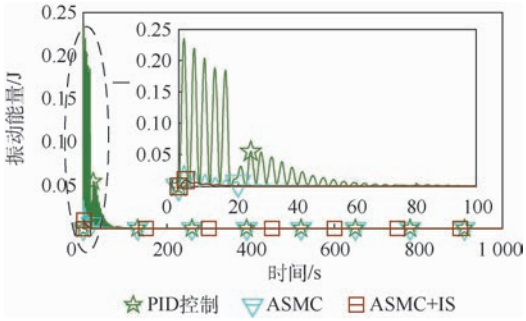


图 7　振动能量波形图

Fig. 7　Oscillogram of vibration energy

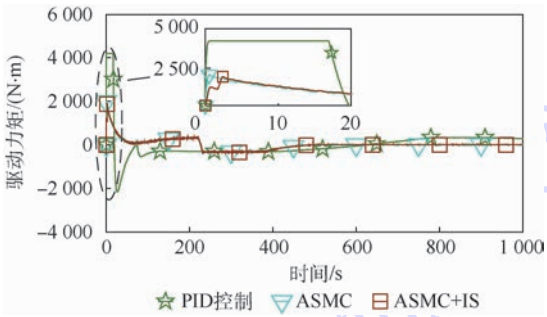


图 8　驱动力矩波形图

Fig. 8　Oscillogram of driving torque

好的综合表现。

4　结　论

1）为提高太阳能帆板驱动系统的角位置控制性能和抑制太阳能帆板的柔性振动，本文提出了一种自适应滑模控制与输入成形技术相结合的控制策略。

2）理论分析表明，本文控制策略可以保证系统在不确定性影响下的一致有界性和渐进一致有界性。

3）仿真结果表明，相比于 PID 控制，本文所提出的控制策略 ASMC + IS 在调节时间、超调量和跟踪误差方面具有更好的性能，并且降低了振动能量，有效抑制了柔性振动；相比于 ASMC，ASMC + IS 同样有效降低了振动能量，而其他性能基本相同。因此，ASMC + IS 能在保证系统角位置控制性能的同时抑制太阳能帆板的柔性振动。

参考文献（References）

[1] JOSE S, GOPALAKRISHNAN E, TANGIRALA A K, et al. Stiffness control of cylindrical shells under axial compression using piezocomposite actuators-An experimental investigation[J]. Mechanics of Advanced Materials and Structures, 2017, 24(1): 16-26.

[2] SHIN C, HONG C, JEONG W B, et al. Active vibration control of plates using positive position feedback control with PZT actuators[J]. Noise Control Engineering Journal, 2016, 64(2): 279-289.

[3] OMIDI E, MAHMOODI S N. Novel hybrid positive feedback control for active vibration suppression in flexible structure [C]//2014 American Control Conference. Piscataway, NJ: IEEE Press, 2014: 2723-2728.

[4] GASBARRI P, SABATINI M, LEONANGELI N, et al. Flexibility issues in discrete on-off actuated spacecraft: Numerical and experimental tests[J]. Acta Astronautica, 2014, 101: 81-97.

[5] GASBARRI P, MONTI R, SABATINI M. Very large space structures: Non-linear control and robustness to structural uncertainties[J]. Acta Astronautica, 2014, 93: 252-265.

[6] NA S, TANG G, CHEN L. Vibration reduction of flexible solar array during orbital maneuver [J]. Aircraft Engineering and Aerospace Technology, 2014, 86(2): 155-164.

[7] PAI M. Closed-loop input shaping control of vibration in flexible structures via adaptive sliding mode control[J]. Shock and Vibration, 2012, 19(2): 221-233.

[8] SINGHOSE W, PORTER L, KENISON M, et al. Effects of hoisting on the input shaping control of gantry cranes[J]. Control Engineering Practice, 2000, 8(10): 1159-1165.

[9] SINGER N, SEERING W. Preshaping command inputs to reduce system vibration [M]. Cambridge: MIT Press, 1988: 76-82.

[10] LU D, LIU Y. Singular formalism and admissible control of spacecraft with rotating flexible solar array[J]. Chinese Journal of Aeronautics, 2014, 27(1): 136-144.

[11] LEE K W, SINGH S N. L1 adaptive control of flexible spacecraft despite disturbances[J]. Acta Astronautica, 2012, 80: 24-35.

[12] HU Q. Robust adaptive attitude tracking control with L2-gain performance and vibration reduction of an orbiting flexible spacecraft[J]. Journal of Dynamic Systems, Measurement and Control, Transactions of the ASME, 2011, 133(1): 011009.

[13] PUKDEBOON C. Adaptive-gain second-order sliding mode control of attitude tracking of flexible spacecraft[J]. Mathematical Problems in Engineering, 2014, 2014(5): 693-697.

[14] CHU M, CHEN G, JIA Q, et al. Simultaneous positioning and non-minimum phase vibration suppression of slewing flexible-link manipulator using only joint actuator[J]. Journal of Vibration and Control, 2014, 20(10): 1488-1497.

[15] ALIPOUR K, ZARAFSHAN P, EBRAHIMI A. Dynamics modeling and attitude control of a flexible space system with active stabilizers[J]. Nonlinear Dynamics, 2016, 84(4): 2535-2545.

[16] PAI M. Discrete-time dynamic output feedback input shaping control of vibration in uncertain time-delay flexible structures [J]. Applied Mathematics and Computation, 2015, 250: 675-688.

[17] PAI M. Robust input shaping control for multi-mode flexible structures using neuro-sliding mode output feedback control [J]. Journal of the Franklin Institute, 2012, 349(3): 1283-1303.

[18] HU Q. Robust adaptive sliding mode attitude control and vibra-

北航学报 赠阅

tion damping of flexible spacecraft subject to unknown disturbance and uncertainty[J]. Transactions of the Institute of Measurement and Control,2012,34(4):436-447.

[19] XU W,MENG D,CHEN Y,et al. Dynamics modeling and analysis of a flexible-base space robot for capturing large flexible spacecraft[J]. Multibody System Dynamics,2014,32(3):357-401.

[20] 白圣建,黄新生. 快速机动大型挠性航天器的动力学建模[J]. 航空学报,2009,30(10):1985-1992.
BAI S J, HUANG X S. Dynamic modeling of large flexible spacecraft undergoing fast maneuvering[J]. Acta Aeronautica et Astronautica Sinica,2009,30(10):1985-1992(in Chinese).

[21] LI J,LI S,CHEN X. Adaptive speed control of a PMSM servo system using an RBFN disturbance observer[J]. Transactions of the Institute of Measurement and Control,2012,34(5):615-626.

[22] LIU H,LI S. Speed control for PMSM servo system using predictive functional control and extended state observer[J]. IEEE Transactions on Industrial Electronics,2012,59(2):1171-1183.

[23] LIANG H,SUN Z,WANG J. Robust decentralized attitude control of spacecraft formations under time-varying topologies, model uncertainties and disturbances[J]. Acta Astronautica,2012,81(2):445-455.

[24] WU S,RADICE G,GAO Y,et al. Quaternion-based finite time control for spacecraft attitude tracking[J]. Acta Astronautica,2011,69(1-2):48-58.

[25] HU Q. Input shaping and variable structure control for simultaneous precision positioning and vibration reduction of flexible spacecraft with saturation compensation[J]. Journal of Sound and Vibration,2008,318(1-2):18-35.

[26] SUNG Y G,SINGHOSE W E. Robustness analysis of input shaping commands for two-mode flexible systems[J]. IET Control Theory and Applications,2009,3(6):722-730.

[27] CORLESS M J,LEITMANN G. Continuous state feedback guaranteeing uniform ultimate boundedness for uncertain dynamic systems[J]. IEEE Transactions on Automatic Control,1981,26(5):1139-1144.

[28] IOANNO P A,KOKOTOVIC P V. Robust redesign of adaptive control[J]. IEEE Transactions on Automatic Control,1984,29(3):202-211.

[29] LIM T W,COOPER P A,AYERS J K. Structural dynamic interaction with solar tracking control for evolutionary space station concepts[C]// Proceedings of the 33rd Structures,Structural Dynamics and Materials Conference. Reston:AIAA,1992:2108-2117.

[30] BOUCHER R L. Mechanically induced g-jitter from space station rotary joints:JSC-CN-6110[R]. Washington,D. C.:NASA,2000.

[31] BOUCHER R L. Identification and mitigation of low frequency vibration sources on space station[C]// Dynamics Specialists Conference. Reston:AIAA,1996:451-462.

作者简介:

周通 男,博士研究生。主要研究方向:太阳能帆板驱动系统的高精度控制、特种电机设计与驱动控制。

郭宏 男,博士,教授,博士生导师。主要研究方向:机载电气系统设计、特种电机设计及其驱动控制。

徐金全 男,博士,讲师。主要研究方向:高可靠电机设计及其驱动控制。

北京航空航天大学学报

北航学报 赠阅

Adaptive sliding mode control of solar array with input shaping

ZHOU Tong, GUO Hong, XU Jinquan*

(School of Automation Science and Electrical Engineering, Beijing University of Aeronautics and Astronautics, Beijing 100083, China)

Abstract: This paper proposes a control strategy which combines adaptive sliding mode control (ASMC) with input shaping technology for the solar array drive system (SADS) to improve the angular position control performance and suppress the flexible vibration. To improve the angular position control performance, ASMC is introduced, which is able to guarantee the uniform boundedness and uniform ultimate boundedness, regardless of the uncertainty. The command trajectory is planned by the input shaper (IS) based on the reference model, which suppresses the flexible vibration of solar array. The simulation results verify the validity of the proposed control strategy.

Keywords: input shaping; adaptive sliding mode control (ASMC); solar array; drive system; flexible vibration

北京航空航天大学学报

Received: 2017-05-02; **Accepted**: 2017-08-01; **Published online**: 2017-09-25 10:01

URL: kns.cnki.net/kcms/detail/11.2625.V.20170925.1001.001.html

Foundation item: Aeronautical Science Foundation of China (2016ZC51025)

* **Corresponding author**. E-mail: xujinquan@buaa.edu.cn

北航学报 赠阅

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2017.0233

全下文载

基于模块化的缠绕机设计方法

徐小明，张武翔*，丁希仑

（北京航空航天大学 机械工程及自动化学院，北京 100083）

摘 要：缠绕机具有生产效率高、产品质量稳定等特点，是碳纤维复合材料成型的关键工艺装备。针对产品小批量、定制化的生产需求，提出了针对缠绕机结构和控制系统的模块化设计方法（MDM），拓展缠绕机的功能多样性。将缠绕机的结构部件进行功能分解和关联强度分析，采用组遗传算法（GGA）将部件聚类为标准化的模块，根据产品需求，以产品性能最好和成本最低为目标建立模块配置模型，基于快速分类的非支配遗传算法（NSGA-Ⅱ）求解多目标优化模型完成机械结构模块化配置。提出基于现场总线的分布式网络控制器结构，将控制器的接口标准化和网络化，根据机械结构的模块化配置实现控制器的快速重构。基于模型组件对象（COM）技术将软件模块设计为 COM 组件，采用 k 近邻（kNN）方法进行控制模式分类，并进行 COM 组件的重构，控制软件动态解析控制模式并管理 COM 组件的状态转移关系，从而实现软件的快速重构。对结构、控制器和软件模块化方法的研究能够实现缠绕机的快速重构，拓展缠绕机的功能。

关 键 词：缠绕机；模块化方法；多目标优化；分布式控制；模型组件对象（COM）技术
中图分类号：V261.97
文献标识码：A **文章编号**：1001-5965(2018)04-0746-13

碳纤维复合材料具有超高模、耐腐蚀、低热膨胀系数和可设计性强等性能，可以制备出热膨胀系数为零的承力结构件[1]，在航空航天等领域中得到广泛应用[2]。缠绕成型是进行大批量生产的有效工艺方法，缠绕机是实现该工艺的核心装备。随着碳纤维复合材料制品的应用范围越来越广泛，客户需求呈现出多样化和定制化的特点，客户在多品种、小批量、高质量方面的要求使缠绕成型制造等高端制造面临着严峻的挑战[3]。

模块化设计方法（Modular Design Method，MDM）是实现产品快速设计的有效手段，通过对产品进行功能分解，从而划分出一系列的通用化模块，建立产品的模块化体系。当客户的需求发生变化时，通过具有特定功能的模块的选择和组合重构系统功能，从而实现产品的快速响应设计[4]。采用模块化设计方法可以实现高效产品升级、降低复杂性、降低费用、实现快速产品设计[5]。

在机械系统的模块化设计方法中，Stone 等[6]以客户需求程度作为衡量标准，通过定量框架的方法量化功能与客户需求之间的关系，提出一种启发式的模块分类方法。Kreng 和 Lee[7]以模块内组件的功能和物理联系最大、模块内组件的模块驱动力（MDF）的相似性最大为目标进行模块划分。Wei 等[8]以模块之间的内部关联最大、模块之间的互相关联最小、系统的整体稳定性最大

收稿日期：2017-04-17；**录用日期**：2017-05-26；**网络出版时间**：2017-07-11 11:47
网络出版地址：kns.cnki.net/kcms/detail/11.2625.V.20170711.1147.006.html
基金项目：国家自然科学基金（51575018）
* **通信作者**. E-mail：zhangwuxiang@buaa.edu.cn

引用格式：徐小明，张武翔，丁希仑. 基于模块化的缠绕机设计方法[J]. 北京航空航天大学学报，2018，44(4)：746-758.
XU X M，ZHANG W X，DING X L. Modular design method for filament winding machine[J]. Journal of Beijing University of Aeronautics and Astronautics，2018，44(4)：746-758 (in Chinese).

北航学报 赠阅

为目标采用多目标优化求解方法进行模块划分。Jose 和 Tollenaere[9]将模块化设计方法分为 5 类:聚类方法[10],图论和矩阵分割方法[11],数学规划方法[12],人工智能方法[13-14],遗传算法和启发式方法[15]。

陆良等[16]采用模块化方法将汽车的硬件控制电路和软件算法划分为标准模块,通过模块的合理组合完成汽车控制电器的软硬件设计。Pritschow 等[17]提出控制系统的模块化设计方法应满足标准化、可扩展性、互换性和实效性要求,并基于控制模型库实现控制系统重构。何岭松等[18]基于模型组件对象(Compent Object Model,COM)技术提出软件总线结构的装配式可重构虚拟仪器系统,实现测试系统的在线升级和功能调整。

当前的研究缺乏针对机械结构、控制器和控制软件的模块化总体设计方法的研究。本文针对缠绕机的结构组成特点,提出适应不同缠绕工艺(干法、湿法和半干法)的模块化重构方法,拓展缠绕机的功能多样性。

1　总体设计

缠绕机的模块化设计方法包括模块划分和模块配置方法。模块划分将缠绕机的组件根据功能和结构聚类为标准模块,并根据功能和工艺的不同设计模块实例,模块实例之间具有标准化和通用化的接口;根据产品的需求变化,将不同功能的实例进行增减和替换,以实现产品功能的快速重构。

缠绕机机电系统的模块化设计模型如图 1 所示,分别建立模块化的机械结构、控制器和控制软件。机械结构的模块化配置方法是控制器和软件配置的基础;控制器根据机械结构的模块配置结果选择对应的传感模块和控制模块,并通过网络接口连接到现场总线;控制软件根据控制模块的连接状态对控制模式进行分类,解析每个控制模块之间的状态转移关系,并调用对应的软件模块完成软件的功能重构。

机械系统的模块划分和模块配置是进行系统模块化设计方法的基础,控制器通过机电接口与机械结构连接,其模块划分和配置粒度与机械结构保持一致,控制软件通过现场总线与控制器进行连接和数据交换。因此,缠绕机的模块化设计问题分为:

1)机械结构的模块划分和模块配置问题。

2)控制器的接口标准化方法和控制系统的组合重构问题。

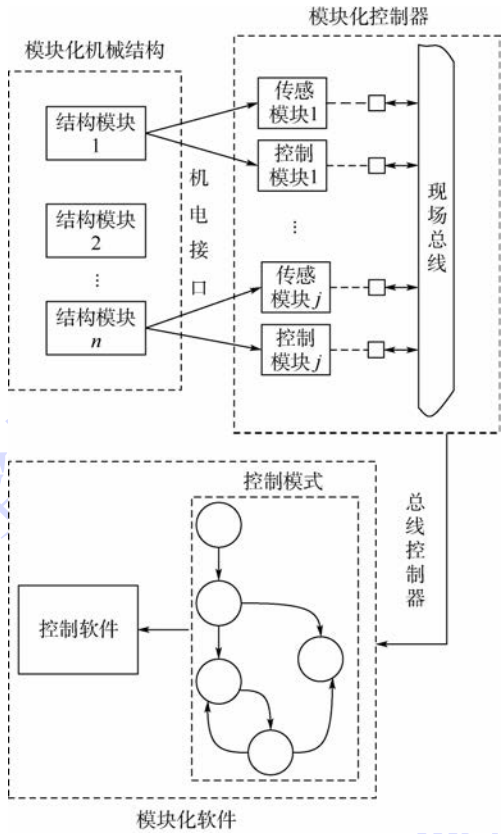


图 1　缠绕机机电系统模块化设计模型

Fig. 1　Modular design model for mechatronic system of winding machine

3)控制软件的控制模式划分和软件模块重构问题。

本文针对上述 3 个问题进行研究,分别提出相应的模块化设计方法。

2　模块划分

2.1　机械结构模块划分

2.1.1　模块划分数学模型

模块划分将缠绕机的独立组件根据相互关联强度和功能相似度聚类为功能模块。首先分析组件之间的关联强度,然后依据划分准则建立模块划分的数学模型,求解模型得到组件的聚类结果。从功能关联和全生命周期的稳定性角度,模块划分的准则为:①同一个模块内部组件之间的功能关联、几何关联最大;②产品全生命周期的稳定性最大,即影响产品全生命周期的模块驱动力的组件应该划分在同一个模块中。

根据准则①,组件之间的功能和几何位置关联强度矩阵为

北航学报 赠阅

$$\boldsymbol{F}_{\mathrm{f/g}}=\begin{bmatrix}1 & F_{\mathrm{f/g},12} & \cdots & F_{\mathrm{f/g},1i} & F_{\mathrm{f/g},1j} & \cdots & F_{\mathrm{f/g},1(n-1)} & F_{\mathrm{f/g},1n}\\ F_{\mathrm{f/g},21} & 1 & \cdots & F_{\mathrm{f/g},2i} & F_{\mathrm{f/g},2j} & \cdots & F_{\mathrm{f/g},2(n-1)} & F_{\mathrm{f/g},2n}\\ \vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots\\ F_{\mathrm{f/g},i1} & F_{\mathrm{f/g},i2} & \cdots & 1 & F_{\mathrm{f/g},ij} & \cdots & F_{\mathrm{f/g},i(n-1)} & F_{\mathrm{f/g},in}\\ F_{\mathrm{f/g},j1} & F_{\mathrm{f/g},j2} & \cdots & F_{\mathrm{f/g},ji} & 1 & \cdots & F_{\mathrm{f/g},j(n-1)} & F_{\mathrm{f/g},jn}\\ \vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots\\ F_{\mathrm{f/g},(n-1)1} & F_{\mathrm{f/g},(n-1)2} & \cdots & F_{\mathrm{f/g},(n-1)i} & F_{\mathrm{f/g},(n-1)j} & \cdots & 1 & F_{\mathrm{f/g},(n-1)n}\\ F_{\mathrm{f/g},n1} & F_{\mathrm{f/g},n2} & \cdots & F_{\mathrm{f/g},ni} & F_{\mathrm{f/g},nj} & \cdots & F_{\mathrm{f/g},n(n-1)} & 1\end{bmatrix} \tag{1}$$

式中：n 为系统中的组件的数量；$F_{\mathrm{f},ij}$ 为组件 i 与组件 j 之间的功能关联强度；$F_{\mathrm{g},ij}$ 为组件 i 与组件 j 之间的几何位置关联强度，其值 9、3、1 和 0 分别表示强关联、一般关联、弱关联和无关联，相同组件关于其自身的关联强度为 1。根据 $\boldsymbol{F}_{\mathrm{f}}$ 和 $\boldsymbol{F}_{\mathrm{g}}$ 可以得到描述组件之间总体功能关系的综合关联矩阵 $\boldsymbol{Y}$

$$\boldsymbol{Y}=\omega_{\mathrm{f}}\boldsymbol{F}_{\mathrm{f}}+\omega_{\mathrm{g}}\boldsymbol{F}_{\mathrm{g}} \tag{2}$$

式中：ω_{f} 为功能关联的权重值；ω_{g} 为几何位置关联的权重值，其关系满足 $\omega_{\mathrm{f}}+\omega_{\mathrm{g}}=1$。

除了组件之间的功能和几何位置关联，缠绕机在全生命周期的稳定性是影响生产效率、产品质量的关键因素，因此应分析影响稳定性的关键设计因素，即模块驱动力。在模块划分过程中，对于模块驱动力有相似贡献度的组件应该划分在一个模块中，一个模块内的模块驱动力的更高相似性会使模块化的系统更加稳定。组件与模块驱动力的关联矩阵为

$$\boldsymbol{F}_{\mathrm{s}}=\begin{bmatrix}f_{11} & \cdots & f_{1n}\\ \vdots & & \vdots\\ f_{m1} & \cdots & f_{mn}\end{bmatrix} \tag{3}$$

式中：f_{ij} 为第 i 个模块驱动力与第 j 个组件之间的关联程度，其值 9、3、1 和 0 分别表示强关联、中等关联、弱关联和无关联。

由 K 个模块组成的系统可以用矩阵 $\boldsymbol{G}=[g_{jk}]$ 表示，其中：

$$g_{jk}=\begin{cases}1 & \text{组件 } j \text{ 属于组件 } k\\ 0 & \text{组件 } j \text{ 不属于组件 } k\end{cases}$$

根据组件的综合关联矩阵，所有模块的组件之间的总聚合度为

$$I_{\mathrm{C}}(\boldsymbol{Y})=\frac{\sum_{k=1}^{K}\left[\left(\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}y_{ij}g_{ik}g_{jk}\right)G_k\right]}{C_{\max}} \tag{4}$$

式中：y_{ij} 为组件 i 与组件 j 之间的关于功能和几何位置的综合关联强度；G_k 为第 k 个模块的相似度系数；$C_{\max}$ 为模块之间的最大关联度之和。

$$\begin{cases}G_k=\dfrac{\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}y_{ij}g_{ik}g_{jk}}{\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}y_{\max}g_{ik}g_{jk}}\\ C_{\max}=\sum_{k=1}^{K}\left(\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}y_{\max}g_{ik}g_{jk}\right)\end{cases} \tag{5}$$

式中：$y_{\max}$ 为组件之间关于功能和几何位置的综合关联强度最大值，即 $y_{\max}=9$。

根据组件与模块驱动力的关联矩阵，模块内部组件之间的相似度可以表示为

$$I_{\mathrm{R}}(\boldsymbol{Y})=\frac{\sum_{k=1}^{K}\left[\sum_{i=1}^{m}\omega_i E_{ik}\left(1-\dfrac{S_{ik}}{S_{ik}^{\max}}\right)\right]}{K} \tag{6}$$

式中：ω_i 为第 i 个模块驱动力的权重，通过层次分析法进行定性和定量分析得到；E_{ik} 为第 i 个模块驱动力在第 k 个模块中的熵；S_{ik} 为在第 k 个模块中第 i 个模块驱动力的不相似性的平方和；$S_{ik}^{\max}$ 为第 k 个模块中第 i 个模块驱动力的不相似性的平方和的最大值。

$$E_{ik}=\frac{-1}{\ln c_k}\sum_{j=1}^{c_k}\left[\frac{f_{ij}}{\sum_{j=1}^{c_k}f_{ij}}\ln\left(\frac{f_{ij}}{\sum_{j=1}^{c_k}f_{ij}}\right)\right] \tag{7}$$

$$\begin{cases}S_{ik}=\sum_{j=1}^{c_k}(f_{ij}-\bar{f}_i)^2\\ S_{ik}^{\max}=c_k\left(\dfrac{f_{\max}}{2}\right)^2\end{cases} \tag{8}$$

建立模块划分的目标函数为

$$\max\ Z=\omega_{\mathrm{r}}I_{\mathrm{R}}(\boldsymbol{Y})+\omega_{\mathrm{c}}I_{\mathrm{C}}(\boldsymbol{Y}) \tag{9}$$

其约束条件为

$$\begin{cases}\omega_{\mathrm{r}}+\omega_{\mathrm{c}}=1\\ \sum_{k=1}^{K}g_{jk}=1 \quad j=1,2,\cdots,n\\ \sum_{j=1}^{n}g_{jk}=c_k \quad k=1,2,\cdots,K\\ \sum_{k=1}^{K}c_k=n \quad K_{\min}\leqslant K\leqslant K_{\max}\end{cases} \tag{10}$$

2.1.2 模型求解

根据所建立的数学模型，模块划分问题转化

为最优化求解问题,在满足约束条件的前提下寻找的最大 Z 值。Z 值越大,模块划分的结果关于功能和稳定性最大。由于模块划分过程中模块的数量和模块中的组件数量是变化的,因此采用本文采用经过改进的组遗传算法(Grouping Genetic Algorithm,GGA)进行模块划分的聚类求解。GGA 的求解流程如图 2 所示,包括基因编码,交叉、变异、遗传等操作,与常规遗传算法不同的是,GGA 以组为单位进行上述操作。根据遗传算法的求解结果,将不同的组件聚类为模块。

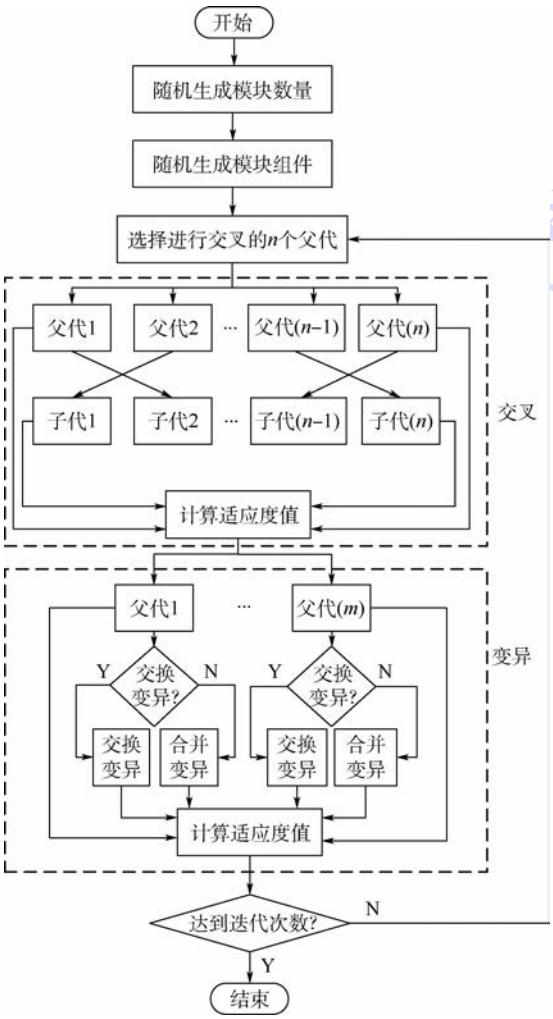


图 2　GGA 求解流程

Fig. 2　Solving process for GGA

2.2　控制器和软件模块设计

控制器模块划分的粒度与机械结构保持一致,根据机械结构与控制器之间的接口可以完成控制器的模块划分。为了实现控制系统的重构功能,采用具有标准接口的分布式现场总线控制系统结构。控制器模块均作为独立的网络节点,通过标准化的网络接口连接到现场总线实现控制网络。

缠绕机的控制器主要分为:基于模拟信号的控制器和基于总线通信的控制器,针对 2 类控制器,分别设计标准化的与主控制器之间的接口,以便在完成机械结构重构以后实现控制系统的快速重构。

对于模拟量输入输出、数字量输入输出类型的传感器和驱动器,其接口设计如图 3 所示,包括通信接口、电源接口、信号输入/输出接口和调试接口,通信接口用于向现场总线发送和接收传感器信号和控制指令,用于接收传感器信号时,微处理器采集传感器的信号,根据配置文件将传感器信号转换为物理信号,通过通信接口上传到控制软件;用于控制驱动信号时,微处理器接收现场总线的控制指令,并根据配置文件转换为相应的驱动信号驱动电机运行。

用于协议转换的控制器接口设计如图 4 所示,控制器的主要功能是将现场总线的控制指令转换为电机驱动器的控制指令,通过协议转换电路实现通信端口的电平转换和控制指令编码。

根据控制器的硬件组成和工艺参数要求,缠绕机的软件模块主要包括:通信管理模块、温度控制模块、张力控制模块和运行状态检测模块。根据不同的控制算法和输入输出条件分别设计不同功能的软件模块。

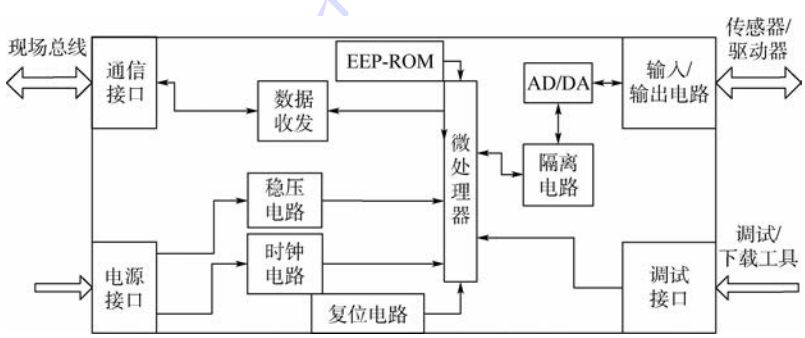


图 3　模拟量/数字量控制器接口设计

Fig. 3　Interface design for analog/digital controller

北航学报 赠阅

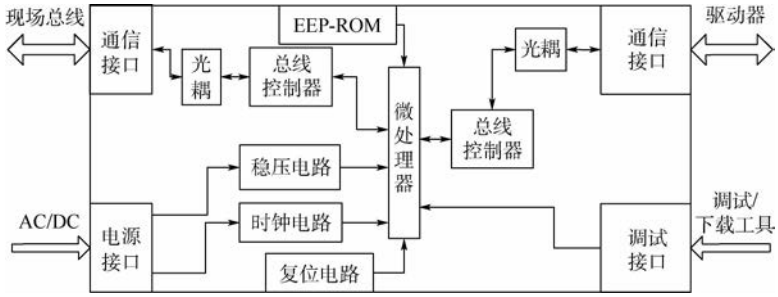


图 4 通信协议转换控制器接口设计

Fig. 4 Interface design for communication protocols converter

3 模块配置

3.1 机械系统模块配置

工艺参数调节系统的性能矢量为

$$\boldsymbol{P}=(P_1,P_2,\cdots,P_r,\cdots,P_R)^{\mathrm{T}}$$

采用层次分析法,其权重为

$$\boldsymbol{W}_{\mathrm{P}}=(w_{\mathrm{P1}},w_{\mathrm{P2}},\cdots,w_{\mathrm{P}R})^{\mathrm{T}}$$

模块实例和产品每一项性能的相关度矩阵 $\boldsymbol{M}_{\mathrm{p}}$ 可以表示为

$$\boldsymbol{M}_{\mathrm{p}}=\begin{bmatrix}\tau_{11,1} & \tau_{11,2} & \cdots & \tau_{11,k} & \cdots & \tau_{11,R}\\ \tau_{12,1} & \tau_{12,2} & \cdots & \tau_{12,k} & \cdots & \tau_{12,R}\\ \vdots & \vdots & & \vdots & & \vdots\\ \tau_{ij,1} & \tau_{ij,2} & \cdots & \tau_{ij,k} & \cdots & \tau_{ij,R}\\ \vdots & \vdots & & \vdots & & \vdots\\ \tau_{MN_M,1} & \tau_{MN_M,2} & \cdots & \tau_{MN_M,k} & \cdots & \tau_{MN_M,R}\end{bmatrix} \tag{11}$$

式中:$\tau_{ij,k}$ 为第 i 个模块的第 j 个实例与第 k 项性能的相关度。

客户对于所定制的产品需求矢量为

$$\boldsymbol{C}_{\mathrm{r}}=(C_1,C_2,\cdots,C_q,\cdots,C_Q)^{\mathrm{T}}$$

采用层次分析法,其权重为

$$\boldsymbol{W}_{\mathrm{C}}=(w_{\mathrm{C1}},w_{\mathrm{C2}},\cdots,w_{\mathrm{C}Q})^{\mathrm{T}}$$

产品需求和系统性能的相关度矩阵 $\boldsymbol{M}_{\mathrm{C}}$ 可以表示为

$$\boldsymbol{M}_{\mathrm{C}}=\begin{bmatrix}\gamma_{11} & \gamma_{12} & \cdots & \gamma_{1j} & \cdots & \gamma_{1Q}\\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2j} & \cdots & \gamma_{2Q}\\ \vdots & \vdots & & \vdots & & \vdots\\ \gamma_{i1} & \gamma_{i2} & \cdots & \gamma_{ij} & \cdots & \gamma_{iQ}\\ \vdots & \vdots & & \vdots & & \vdots\\ \gamma_{R1} & \gamma_{R2} & \cdots & \gamma_{Rj} & \cdots & \gamma_{RQ}\end{bmatrix} \tag{12}$$

式中:γ_{ij} 为第 i 项系统性能和第 j 项产品需求之间的相关度。

根据相关度矩阵和权重值,针对需求矢量 $\boldsymbol{C}_{\mathrm{r}}$ 的系统性能为

$$F=\sum_{i=1}^{M}\sum_{j=1}^{N_i}\left\{\sigma_{ij}\sum_{k=1}^{R}\left[\tau_{ij,k}w_{\mathrm{P}k}\sum_{m=1}^{Q}(\gamma_{km}w_{\mathrm{C}m})\right]\right\} \tag{13}$$

式中:F 为系统性能的值,值越大表示系统的性能越好;σ_{ij} 为二元决策变量,表示模块实例在配置产品的存在性,$\sigma_{ij}=1$ 表示第 i 个模块的第 j 个实例被配置在产品中,$\sigma_{ij}=0$ 表示第 i 个模块的第 j 个实例不被配置在产品中。

除了产品性能以外,生产成本是缠绕机系统设计过程中的另外一个考虑因素。第 i 个模块中第 j 个实例的成本为

$$C_{ij}=C_{\mathrm{M}}+C_{\mathrm{E}}+C_{\mathrm{r}}C_{\mathrm{R}} \tag{14}$$

式中:C_{M} 为生产过程中的材料成本;C_{E} 为生产过程中的维护成本,包括轴承、电机、传感器等在高温环境中工作的维护成本;C_{R} 为设备维护清洗时间;C_{r} 为单位时间的人工成本。

由此可以得到产品配置的生产成本为

$$C=\sum_{i=1}^{M}\sum_{j=1}^{N_i}(\sigma_{ij}C_{ij})+C_{\mathrm{a}}\left(\sum_{i=1}^{M}\sum_{j=1}^{N_i}\sigma_{ij}-1\right) \tag{15}$$

式中:C_{a} 为模块的装配重组成本。由此,可以得到模块化产品配置的目标函数为

$$\begin{cases}\max & F\\ \min & C\end{cases}$$

约束条件为

$$\begin{cases}\sum_{i=1}^{M_{\mathrm{CO}}}\sigma_{ij}=M_{\mathrm{CO}}\\ \sum_{j=1}^{M_i}\sigma_{ij}=1\\ \sum_{i=1}^{M_{\mathrm{O}}}\sigma_{ij}=M_{\mathrm{O}}\\ \sum_{j=1}^{M_k}\sigma_{kj}\leqslant 1\\ C\leqslant C_{\max}\end{cases} \tag{16}$$

式中:M_{CO} 为核心模块的数量,每个核心模块必须选择一个实例;M_i 为第 i 个模块的实例数量;M_{O} 为可选模块的数量,每个可选模块最多选择一个实例;M_k 为第 k 个模块的实例数量;$C_{\max}$ 为客户所

能接受的最大生产成本。

为了解决模块配置中的多目标优化问题,采用基于快速分类的非支配遗传算法(NSGA-Ⅱ)[19]搜索 Pareto 前沿。适应度的值由非支配排序级别确定,该排序级别使用快速非支配排序机制和不需要外部参数的拥挤距离以避免对于目标的偏好,并且使用精英策略以防止优秀个体的丢失。

3.2　控制系统模块配置

根据机械结构的模块配置,根据机电接口可以确定控制器的重构形式,并通过现场总线重构组成新的控制网络。控制系统需要根据所连接的网络节点,进行控制模式识别和软件重构。控制系统的模块化配置方法如图 5 所示。控制器作为独立的网络节点通过网络接口连接到上行总线和下行总线,并通过网络控制器与控制软件交换数据,控制模块具有唯一的 ID 标识。控制软件首先通过 ID 标识扫描控制模块的连接状态,根据预先定义的控制模式数据库和控制模块的特征进行控制模式分类。首先根据控制特征搜索数据库中的相同项,如果数据库中不存在相同项,采用 k 近邻(kNN)方法进行控制模式分类。控制软件解析控制模式脚本,采用基于 COM 组件的重构方法对控制器之间的连接关系和数据交换进行动态重构,进而进行软件重构生成软件控制系统。最后通过白盒测试和黑盒测试对控制系统功能的完整性和组件状态转移关系的正确性进行验证,通过测试后形成功能完整的工艺参数调节系统的控制系统重构。

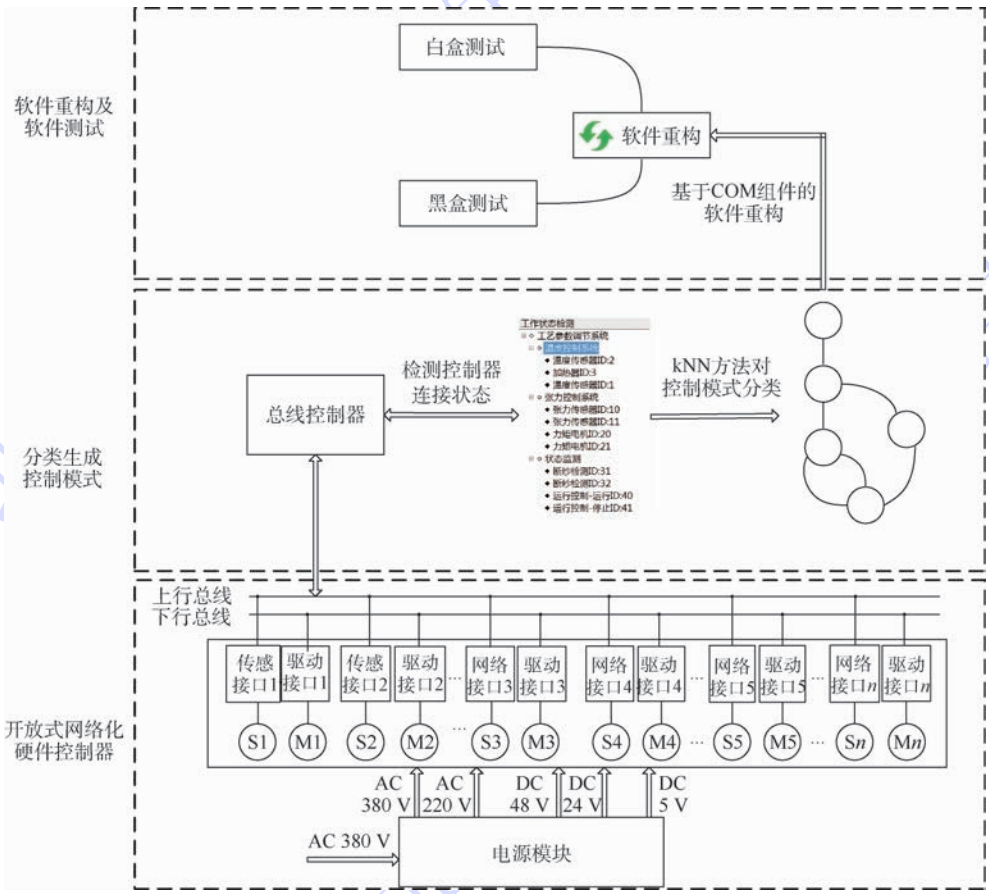


图 5　控制系统的模块化配置方法

Fig. 5　Modular configuration method for control system

3.2.1　控制模式设计

在缠绕机工艺参数调节系统中,工艺流程和传感/驱动组件的不同,导致其控制模式各不相同,应根据控制器的重构形式和连接数量,确定其控制模式。有限状态机通过节点和事件描述离散事件之间的连接关系,可以用来标识控制器之间的执行状态,即控制模式。湿法缠绕的控制模式如图 6 所示。

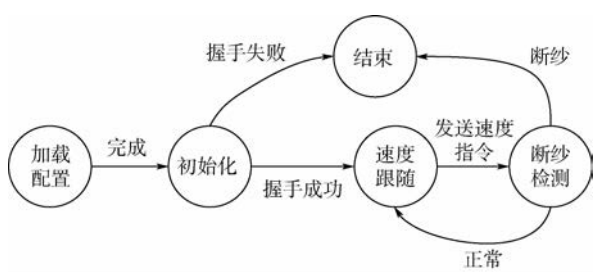


图 6　湿法缠绕控制模式

Fig. 6　Control pattern for wet winding

北航学报 赠阅

3.2.2 控制模式分类方法

采用 kNN 方法根据控制器的连接状态进行控制模式分类。由 n 个特征组成的向量 $\boldsymbol{P}_i=(p_{i1},p_{i2},\cdots,p_{in})$ 表示一种控制模式，控制模式的训练样本集合为 $\boldsymbol{Q}=\{(\boldsymbol{P}_i,L_i),i=1,2,\cdots,K\}$，其中 K 为控制模式的样本数量，L_i 为控制模式类型标签，$i=1,2,\cdots,M$，M 为控制模式类型的数量。为了进一步提高 kNN 方法的准确度，采用加权 kNN 方法在基本 kNN 方法的基础上对 k 个类型标签进行相似度计算，其具体的计算流程为：

1）$\boldsymbol{P}_x$ 为待分类的控制模式的向量，L_x 为控制模式的标签。首先计算训练样本集合 $\boldsymbol{Q}$ 中的向量与 $\boldsymbol{P}_x$ 之间的距离

$$d_{xi}=\left[\sum_{m=1}^{n}(p_{im}-p_{xm})^2\right]^{1/2} \tag{17}$$

将距离排序，取距离最近的 $k+1$ 个近邻样本，近邻样本的集合为 $\boldsymbol{Q}'=\{(\boldsymbol{P}'_i,L'_i),i=1,2,\cdots,k+1\}$。

2）取集合 $\boldsymbol{Q}'$的最大样本距离，即第 $k+1$ 个样本与 $\boldsymbol{P}_x$ 的距离将前 k 个样本距离进行归一化处理。

$$D_{xi}=\frac{\left[\sum_{m=1}^{n}(p'_{im}-p_{xm})^2\right]^{1/2}}{\left[\sum_{m=1}^{n}(p'_{(k+1)m}-p_{xm})^2\right]^{1/2}} \tag{18}$$

3）引入高斯核函数，将归一化的 $\boldsymbol{P}_x$ 与 k 个近邻样本的距离转化为样本之间的相似概率 p_{xi}

$$p_{xi}=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{D_{xi}}{2}} \tag{19}$$

4）根据近邻样本的相似概率，求出 $\boldsymbol{P}_x$ 为控制模式 L_i 的概率为

$$p_{Li}=\frac{1}{\sum_{m=1}^{k}p_{xi}}p_{xi}\sigma \tag{20}$$

式中：σ 为二值函数

$$\sigma=\begin{cases}1 & L'_i=L_i\\ 0 & L'_i\neq L_i\end{cases} \tag{21}$$

根据式(20)求出 $\boldsymbol{P}_x$ 为控制模式为 L_i 的概率，取概率的最大值即 $\boldsymbol{P}_x$ 的控制模式，

$$L_x=\arg\max(p_{L_i}) \tag{22}$$

3.2.3 加权 kNN 方法准确率分析

根据已知的控制特征和控制模式生成训练样本集合 $\boldsymbol{Q}$，$\boldsymbol{Q}$ 中共包含 1569 个训练样本，在 MATLAB 中进行分类准确性分析。针对控制特征向量 $\boldsymbol{P}_x$，首先在 $\boldsymbol{Q}$ 中检索相同样本，保证 $\boldsymbol{P}_x\subseteq\boldsymbol{Q}$ 时准确率达到 100%，如果不存在，则采用加权 kNN 方法进行分类。在 1 569 个训练样本中分别随机挑选 100 和 200 个样本作为测试样本，测试不同 k 取值下加权 kNN 方法的准确率，测试结果如表 1 所示。

从表 1 可以看出，加权 kNN 方法的准确率较高，且对样本数量和 k 值具有较好的鲁棒性。然而由于加权 kNN 分类方法仍然存在一定的错误率，因此，在完成控制系统和软件重构后，应通过一定的测试用例测试控制模式的分类准确性。

表 1 不同测试样本数量和 k 值的加权 kNN 方法准确率

Table 1 Woighted kNN method accuracy for different test sample sizes and values of k

样本数量	加权 kNN 方法准确率/%		
	$k=5$	$k=7$	$k=9$
100	94	91.3	93.3
200	92.7	92.3	91.5

3.2.4 软件重构方法

控制软件解析控制模式脚本，根据模块之间的状态转移顺序动态调用不同的软件模块。软件重构模型如图 7 所示，先将软件功能进行模块化设计，减小模块之间的耦合和数据交换，并将软件模块设计为 COM 组件，定义组件接口。COM 组件作为客户组件，软件控制系统的主控模块作为服务器对客户组件进行管理，根据控制模式的不同，服务器进行客户组件的添加、删除和通信等操作，从而实现操作模式的动态构建。客户组件之间的连接关系和执行顺序存储在控制模式脚本中，服务器通过解析控制模式脚本对客户组件之间的执行顺序进行控制。

COM 组件的通信包括客户与服务器的通信以及客户与客户之间的通信，通信模型如图8

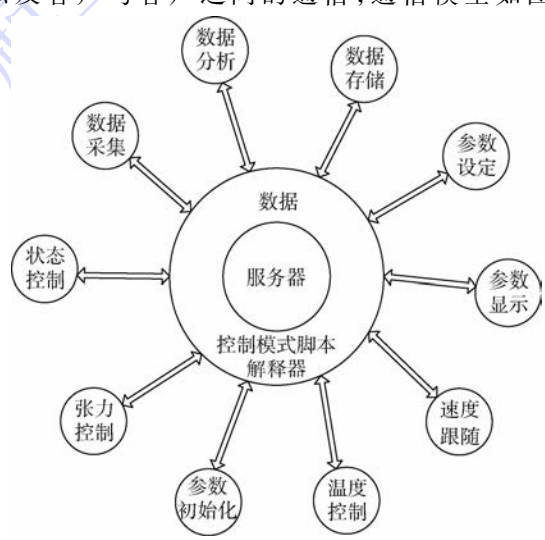


图 7 软件重构模型

Fig. 7 Reconfiguration model for software

北航学报 赠阅

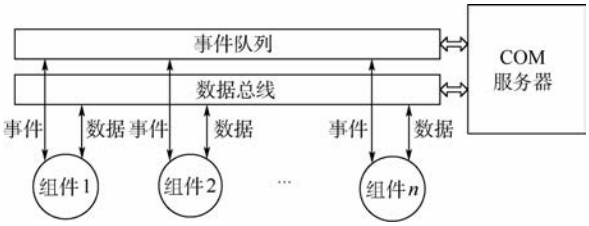


图 8　COM 组件通信模型

Fig. 8　Communication model for COM component

所示。COM 服务器通过事件队列管理客户组件的事件消息,通过软件数据总线管理数据的访问,数据通过数据类型变量保存在数据总线上。根据数据传递方向的不同,客户组件与服务器之间的通信方式分为:服务器调用客户组件并将参数传递给客户组件;客户组件将运算结果等数据发送至服务器以供其他组件调用。服务器调用客户时,通过客户组件提供的访问接口调用相应的方法,客户组件从数据总线读取相关的数据;客户组件运算完成后,将运算结果发送至数据总线,同时生成事件消息通知服务器。

客户组件之间的通信通过事件槽技术实现,而不采用组件之间直接通信的方式,以减小组件之间的耦合,便于控制软件的重构。当组件 1 需要读取组件 2 的数据时,组件 1 向服务器发送请求数据读取事件,服务器接收到该事件以后从组件 2 读取数据,并存储至数据总线,然后服务器将数据写入组件 1。

4　案例分析

根据本文所提出的模块划分和模块配置方法,针对不同的产品需求(天线支撑杆、预浸料、筒形壳体)进行案例分析研究。首先采用模块划分方法将组件聚类为具有标准接口的模块,并设计不同功能的模块实例;然后根据产品需求的变化进行模块选择,完成机械结构和控制系统的功能重构。

4.1　模块划分

根据第 2 节所提出的模块划分方法,采用 MATLAB 软件设计优化算法。根据首先缠绕机由多辊系组成的特点,将缠绕机以辊轮为基本单位分为 30 个独立的组件,建立组件的综合关联矩阵 $\boldsymbol{Y}$;然后确定影响全寿命周期稳定性的模块驱动力包括:运动可靠性和平稳性、技术提高和技术革新,加工质量保证,快速服务和维修,纤维磨损,组件寿命。确定组件和模块驱动力关联矩阵 $\boldsymbol{F}_s$,并采用层次分析法求解模块驱动力的权重为 $\boldsymbol{\omega}=(0.269\,9,0.203\,1,0.051\,9,0.071\,6,0.333\,5,0.070\,0)$。根据模块划分数学模型,采用 GGA 进行组件参数,GGA 的参数如下:初始种群大小为 200,迭代次数为 100,交叉概率为 0.4,变异概率为 0.2,其中交换操作概率为 0.5,分割和合并操作概率为 0.5,稳定性相关的模块驱动力矩阵的权重值 $\omega_r=0.5$,与功能和几何位置相关的关系矩阵的权重值 $\omega_c=0.5$,模块数量最小值 $K_{\min}=3$,模块数量最大值 $K_{\max}=27$。采用 GGA 的适应度函数值随迭代次数的变化结果如图 9 所示。

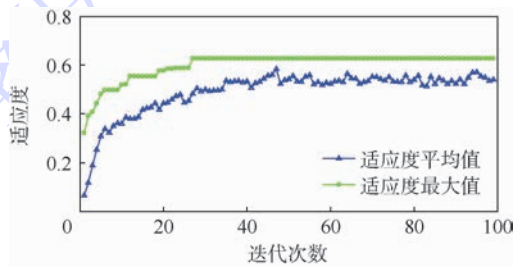


图 9　GGA 优化结果

Fig. 9　GGA optimization results

从图 9 可以看出,GGA 在迭代次数为28 时收敛到目标函数的最大值,近似最优解的值为 $Z=0.627\,2$。根据 GGA 的优化结果,最终聚类得到 7 个模块:放卷模块、单束张力测量模块、浸胶模块、含胶量调整模块、集束张力调整模块、速度和集束张力测量模块、丝嘴集束模块。对每一个模块设计不同的功能实例,放卷模块根据放卷力矩施加方式和是否改变纤维传动方向设计 6 个实例;单束张力测量模块根据不同响应速度设计 2 个实例;浸胶模块根据浸胶工艺的不同设计 3 个实例;含胶量调整模块根据含胶量调节方式设计 2 个实例;集束张力调整模块根据张力施加大小设计 6 个实例;速度和集束张力测量模块根据测量形式设计 3 个实例;丝嘴集束模块根据展纱功能设计 2 个实例。实例之间的辊轮采用相同的安装接口。

4.2　天线支撑杆缠绕的模块配置

采用多丝束高模量碳纤维进行双工位天线支撑杆的同步缠绕,所定制的产品需求为:缠绕工艺为热熔法,放卷辊数量为 10,胶液类型为氰酸酯树脂,碳纤维型号为 M55J,缠绕线型为 45° 螺旋缠绕。

4.2.1　机械系统配置

工艺参数调节系统的性能要求包括:温度,纤维磨损,参数控制精度,设备体积,维护性,纤维展开宽度,运动的平稳性和稳定性,生产效率。通过分析,可以得到每一个模块实例和产品性能的相

北航学报 赠阅

关度矩阵 $\boldsymbol{M}_{\mathrm{P}}$、产品需求和系统性能的相关度矩阵 $\boldsymbol{M}_{\mathrm{C}}$，并采用层次分析法分别得到产品需求的权重值 $\boldsymbol{W}_{\mathrm{C}}$ 和系统性能的权重值 $\boldsymbol{W}_{\mathrm{P}}$。根据式(14)，可以计算每一个模块实例的成本。根据模块配置数学模型，采用 NSGA-Ⅱ 进行多目标优化求解，优化参数设置如下：初始种群大小为 60，迭代次数为 100，交叉概率为 0.6，变异概率为 0.02。成本和性能的变化曲线如图 10 所示。

从图 10 可以看出，当迭代次数大于30 时，成本和性能均收敛到稳定值，说明已经收敛到最优边界。根据优化结果，得到各个模块实例的配置方案，将实例之间进行组合得到模块实例的连接关系如图 11 所示。力矩电机放卷装置和称重传感器单束张力测量装置组成闭环反馈系统控制纤维的张力保持稳定，并使 10 束纤维张力保持均匀，然后10束纤维经过浸胶模块完成浸胶过程，

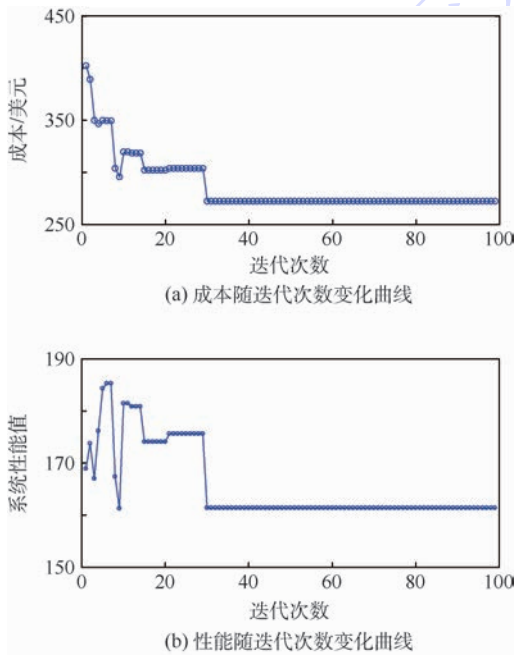


图 10 天线支撑杆缠绕的 NSGA-Ⅱ 优化结果

Fig. 10 NSGA-Ⅱ optimization results for aerial studdle winding

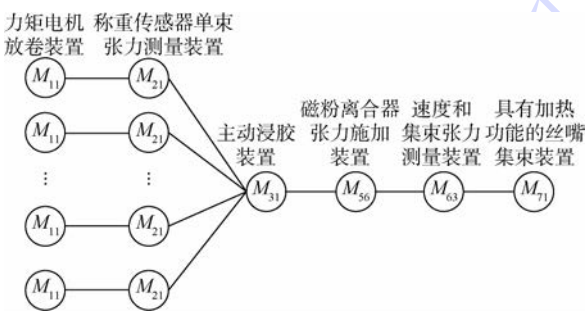


图 11 天线支撑杆缠绕模块之间的拓扑关系

Fig. 11 Topological relationship between modules for aerial studdle winding

浸胶辊在设定的温度下实现含胶量调整，磁粉离合器为集束纤维增大张力，并将速度和集束张力测量模块作为反馈装置控制 10 束纤维的总张力，最后通过丝嘴集束模块缠绕到芯模表面。

4.2.2 控制系统配置

根据机械系统的模块配置结果，将对应的控制器连接到现场总线，控制系统检测控制器的连接状态，根据控制特征采用加权 kNN 方法进行控制模式分类。在缠绕机控制系统中，控制模式的划分特征为：放卷形式，放卷数量，断纱检测，单束张力测量，单束张力施加，磁粉离合器数量，浸胶工艺，集束张力测量，速度传感器，温度控制。根据缠绕机的工艺特点，控制模式库为：湿法速度跟随，干法/半干法速度跟随，干法/半干法小张力控制，干法/半干法速度跟随和张力控制，干法/半干法多丝束张力分布式控制。分别采用有限状态机模型描述这 5 种控制模式。控制系统根据控制器的连接状态得到控制系统的控制特征，然后采用加权 kNN 方法进行分类，最终得到的控制模式为干法/半干法多丝束张力分布式控制模式，如图 12所示。首先加载系统配置，完成控制器和数据总线的初始化，然后进行辊轮的温度控制，并检测温度是否异常，最后达到设定温度以后进行芯模缠绕，系统运行过程中分别进行单束张力控制、集束张力和速度控制，并检测温度、断纱和运行状态。

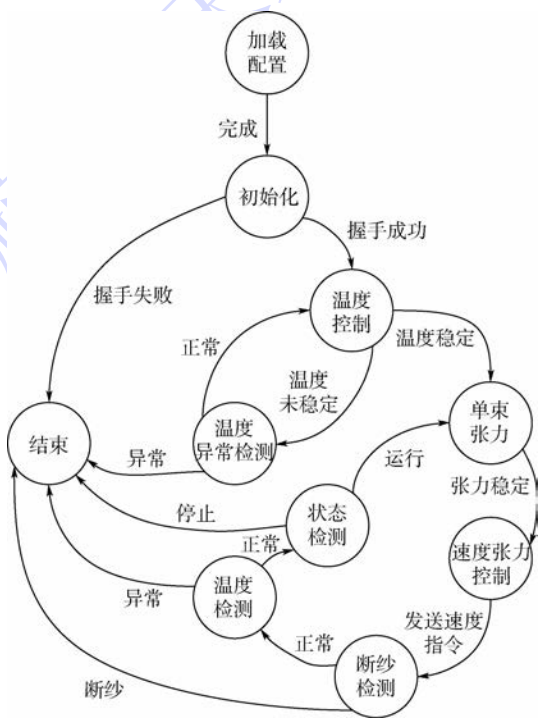


图 12 天线支撑杆缠绕的控制模式

Fig. 12 Control pattern for aerial studdle winding

北航学报 赠阅

4.2.3　系统测试

控制软件根据所生成的控制模式解析 COM 组件之间的状态转移关系完成控制软件的重构。然而由于加权 kNN 方法存在一定的错误率,因此采用白盒测试和黑盒测试方法进行系统测试以保证模块化系统重构后的质量和可靠性。

白盒测试方法在软件重构阶段的主要作用是根据控制模式之间的状态转移关系,对控制系统的所有路径进行测试。该测试方法需要保证各个组件的所有执行路径全覆盖;并且针对逻辑判断条件完成测试。采用白盒测试中的路径覆盖测试方法遍历控制模式中的所有路径,根据系统的期望输出结果和实际输出结果判断组件之间的状态转移关系是否正确。

黑盒测试方法采用等价类方法验证系统的功能完整性。工艺参数调节装置的功能包括:浸胶/传导/胶槽等辊轮的加热功能,速度跟随功能,张力控制功能,状态监测功能;系统的输入条件为:温度设定值,速度设定值,张力设定值,断纱状态;系统输出为:加热器功率,辊轮温度,辊轮转速,系统张力,系统运行状态。采用等价类划分方法将输入划分等价类,并生成等价类表,根据系统的期望输出结果和实际输出结果比对判断控制系统的功能是否完整。

4.2.4　结果分析

所提出的产品需求为 10 束高模量碳纤维使用氰酸酯树脂进行小尺寸杆件缠绕,其工艺参数要求温度控制精度高、含胶量均匀、10 束纤维张力分布均匀。采用模块化方法生成的模块配置方案中,高模量纤维易磨损,力矩电机放卷装置和称重传感器单束张力测量装置组成闭环反馈系统施加较小的传输张力,并控制每束张力保持稳定,从而使 10 束纤维之间的张力分布均匀。磁粉离合器辊轮、速度和集束张力测量模块能够在浸胶前施加较大的纤维张力并保持稳定。浸胶模块具有含胶量调整装置和温控装置,可以实现纤维的含胶量调整。浸胶模块、集束张力调整模块和丝嘴集束模块均具备加热功能,可以满足氰酸酯树脂的温度控制要求。控制软件重构生成的控制模式实现系统温度控制、单束张力控制、速度和张力控制,使温度和张力保持稳定。因此本文模块化缠绕机设计方法能够根据产品的需求快速实现功能重构,并满足缠绕制品的工艺参数要求。

4.3　预浸料缠绕的模块配置

湿法缠绕生产预浸料的工艺过程简单,生产效率高,所定制的预浸料产品需求为:缠绕工艺为湿法;放卷辊数量为 2;胶液类型为环氧树脂;碳纤维型号为 T800;缠绕线型为 90°缠绕,缠绕模具为圆筒,直径为 1 800 mm。

在模块化配置方法中,模块实例和产品性能的相关度矩阵 $\boldsymbol{M}_{\mathrm{P}}$ 和模块实例的成本均保持不变。根据预浸料的需求计算产品需求和系统性能的相关度矩阵 $\boldsymbol{M}_{\mathrm{C}}$,并采用层次分析法分别得到产品需求的权重值 $\boldsymbol{W}_{\mathrm{C}}$ 和系统性能的权重值 $\boldsymbol{W}_{\mathrm{P}}$。在湿法缠绕工艺中,$\boldsymbol{W}_{\mathrm{C}}$ = (0.284 5, 0.132 4, 0.105 3, 0.169 8, 0.308 0), $\boldsymbol{W}_{\mathrm{P}}$ = (0.021 0, 0.079 8, 0.056 1, 0.107 3, 0.216 7, 0.197 4, 0.095 4, 0.226 3),采用 NSGA-Ⅱ 进行多目标优化求解得到性能和成本的变化曲线如图 13 所示。

从图 13 可以看出,当迭代次数大于 37 时,成本和性能均收敛到稳定值。根据优化的模块实例的配置方案,将实例之间进行组合得到模块实例的连接关系如图 14 所示,2 束碳纤维通过摩擦带放卷装置施加一定的纤维张力,通过浸胶装置实现树脂和纤维的混合,然后通过含胶量调整装置和展纱装置调节展纱宽度后缠绕到芯模表面。

根据机械结构的模块化配置结果将控制器连接到现场总线,根据控制特征采用 kNN 方法得到的控制模式为:湿法速度跟随。如图 6 所示,控制软件根据所生成的控制模式解析 COM 组件之间的状态转移关系完成控制软件的重构。湿法缠绕时控制系统的主要功能为控制浸胶辊的速

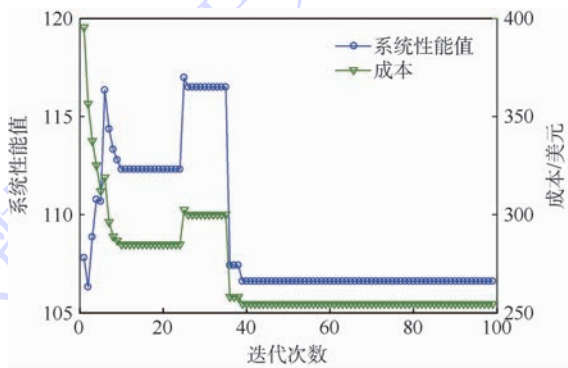


图 13　湿法缠绕的 NSGA-Ⅱ 优化结果

Fig. 13　NSGA-Ⅱ optimization results for wet winding

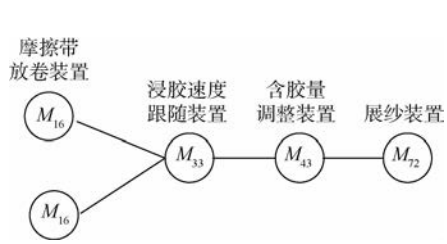


图 14　湿法缠绕模块之间的拓扑关系

Fig. 14　Topological relationship between modules for wet winding

北航学报 赠阅

度跟随芯模的缠绕速度。

4.4 筒形壳体缠绕的模块配置

筒形壳体采用高强度纤维和氰酸酯树脂在较大的张力下进行缠绕成型，所定制的筒形壳体的产品需求为：缠绕工艺为热熔法，放卷辊数量为1，胶液类型为氰酸酯树脂，碳纤维型号为 T800，缠绕线型为 30°螺旋缠绕。在筒形壳体缠绕工艺中，$\boldsymbol{W}_{\mathrm{C}}$ = (0.390 8, 0.102 7, 0.197 2, 0.204 3, 0.105 0)，$\boldsymbol{W}_{\mathrm{P}}$ = (0.127 7, 0.236 1, 0.246 3, 0.053 5, 0.065 1, 0.058 0, 0.157 01, 0.056 4)，采用 NSGA-Ⅱ进行多目标优化求解得到性能和成本的变化曲线如图 15 所示。

从图 15 中可以看出，当迭代次数大于 22 时，成本和性能均收敛到稳定值。根据优化的模块实例的配置方案，将实例之间进行组合得到模块实例的连接关系如图 16 所示，放卷装置通过力矩电机施加较小的张力以减小传送过程中的纤维磨损，浸胶装置之前通过单束张力施加装置施加设定的纤维张力，并通过调节浸胶辊的电机转速实现速度跟随和张力控制，在丝嘴处对树脂进行加热以保持黏度，最终缠绕到芯模表面。

根据机械结构的模块化配置结果将控制器连接到现场总线，根据控制特征采用加权 kNN 方法得到的控制模式为：干法/半干法速度跟随和张力控制，如图 17 所示，控制软件根据所生成的控制模式解析 COM 组件之间的状态转移关系完

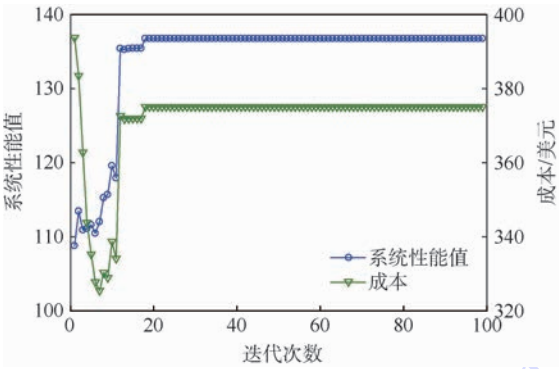


图 15 筒形壳体缠绕的 NSGA-Ⅱ优化结果

Fig. 15 NSGA-Ⅱ optimization results for cylindrical shell winding

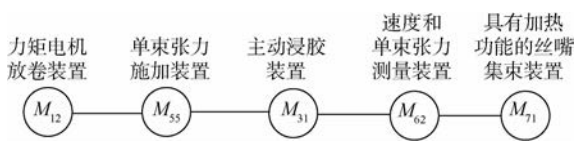


图 16 筒形壳体缠绕模块之间的拓扑关系

Fig. 16 Topological relationship between modules for cylindrical shell winding

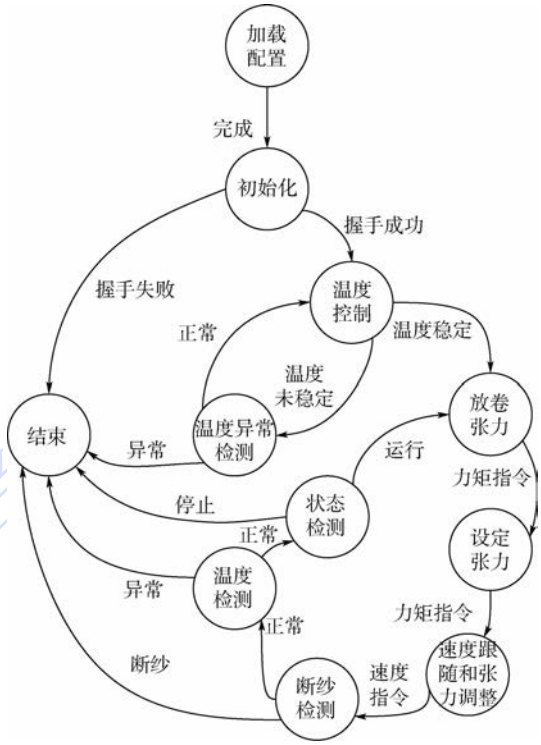


图 17 筒形壳体缠绕的控制模式

Fig. 17 Control pattern for cylindrical shell winding

成控制软件的重构。首先完成控制器和数据总线的初始化；然后进行辊轮的温度控制，并检测温度是否异常，达到设定温度以后，根据设定的张力值控制放卷力矩电机和磁粉离合器的转矩施加纤维张力；最后调节浸胶辊的转速跟随芯模的缠绕速度并调整传送张力使之保持稳定，在运行过程中检测断纱、温度系统的运行是否正常。

通过对天线支撑杆、预浸料、筒形壳体缠绕的模块划分和模块配置方法的分析中可以看出，本文模块化设计方法能够根据不同的产品需求进行不同功能模块实例的最优配置，并实现机械结构、控制器和控制软件的功能重构，使缠绕机能够满足不同的缠绕工艺要求，拓展了缠绕机的功能多样性。

5 结 论

本文提出缠绕机机械结构、控制器和控制软件的模块化设计方法，得出：

1）通过定义组件之间的关联矩阵，采用 GGA 和 NSGA-Ⅱ能够有效实现模块划分和模块重新配置。

2）基于现场总线的开放式控制系统结构可以实现控制器的快速重构，基于 COM 组件技术、有限状态机可以实现软件的动态重构。

3）采用模块化设计方法可以快速响应客户

北航学报 赠阅

需求的变化,满足不同的缠绕工艺要求,实现缠绕机功能的快速重构。

参考文献(References)

[1] 顾轶卓,李敏,李艳霞,等. 飞行器结构用复合材料制造技术与工艺理论进展[J]. 航空学报,2015,36(8):2773-2797.
GU Y Z,LI M,LI Y X,et al. Progress on manufacturing technology and process theory of aircraft composite structure[J]. Acta Aeronautica et Astronautica Sinica,2015,36(8):2773-2797(in Chinese).

[2] SOUTIS C. Carbon fiber reinforced plastics in aircraft construction[J]. Materials Science and Engineering:A,2005,412(1-2):171-176.

[3] PIRAN F A S,LACERDA D P,ANTUNES J A V,et al. Modularization strategy:Analysis of published articles on production and operations management(1999 to 2013)[J]. The International Journal of Advanced Manufacturing Technology,2016,86(1-4):507-519.

[4] HUANG C C. Overview of modular product development[J]. Proceedings of the National Science Council,2000,24(3):149-165.

[5] LAU A K W,YAM R C M,TANG E. The impacts of product modularity on competitive capabilities and performance:An empirical study[J]. International Journal of Production Economics,2007,105(1):1-20.

[6] STONE R B,WOOD K L,CRAWFORD R H. A heuristic method for identifying modules for product architectures[J]. Design Studies,2000,21(1):5-31.

[7] KRENG V B,LEE T. Modular product design with grouping genetic algorithm—A case study[J]. Computers & Industrial Engineering,2004,46(3):443-460.

[8] WEI W,LIU A,LU S C Y,et al. A multi-principle module identification method for product platform design[J]. Journal of Zhejiang University-Science A(Applied Physics & Engineering),2015,16(1):1-10.

[9] JOSE A,TOLLENAERE M. Modular and platform methods for product family design:Literature analysis[J]. Journal of Intelligent Manufacturing,2005,16(3):371-390.

[10] KIMURA F,KATO S,HATA T,et al. Product modularization for parts reuse in inverse manufacturing[J]. CIRP Annals-Manufacturing Technology,2001,50(1):89-92.

[11] HUANG C C,KUSIAK A. Modularity in design of products and systems[J]. IEEE Transactions on Systems,Man,and Cybernetics-Part A:Systems and Humans,1998,28(1):66-77.

[12] 高飞,肖刚,潘双夏,等. 产品功能模块划分方法[J]. 机械工程学报,2007,43(5):29-35.
GAO F,XIAO G,PAN S X,et al. Method of product function module partition[J]. Chinese Journal of Mechanical Engineering,2007,43(5):29-35(in Chinese).

[13] FAN B,QI G,HU X,et al. A network methodology for structure-oriented modular product platform planning[J]. Journal of Intelligent Manufacturing,2015,26(3):553-570.

[14] PANDREMENOS J,CHRYSSOLOURIS G. A neural network approach for the development of modular product architectures[J]. International Journal of Computer Integrated,2011,24(10):879-887.

[15] KRENG V B,LEE T P. MPD with grouping genetic algorithm:A case study[J]. Computers and Industrial Engineering,2004,46(3):443-460.

[16] 陆良,杨殿阁,顾铮珉,等. 采用模块化思想的汽车电器智能化设计方法[J]. 西安交通大学学报,2010,44(5):111-115.
LU L,YANG D G,GU Z M,et al. Designing strategy for smart vehicle electrical/electronic devices based on modularization[J]. Journal of Xi'an Jiaotong University,2010,44(5):111-115(in Chinese).

[17] PRITSCHOW G,KIRCHER C,KREMER M,et al. Control systems for RMS and methods of their reconfiguration[M]. Berlin:Springer,2006:195-211.

[18] 何岭松,张登攀,赖红. 可重构虚拟仪器系统[J]. 机械工程学报,2005,41(9):78-81.
HE L S,ZHANG D P,LAI H. Reconfigurable virtual instrument system[J]. Chinese Journal of Mechanical Engineering,2005,41(9):78-81(in Chinese).

[19] DEB K,PRATAP A,AGARWAL S,et al. A fast and elitist multiobjective genetic algorithm:NSGA-Ⅱ[J]. IEEE Transactions on Evolutionary Computation,2002,6(2):182-197.

作者简介:

徐小明　男,博士研究生。主要研究方向:复合材料缠绕成型技术。

张武翔　男,博士,副教授。主要研究方向:机构学与机器人学。

丁希仑　男,博士,教授,博士生导师。主要研究方向:空间机器人与空间机构学、仿生机器人、智能控制与自动化检测技术。

北京航空航天大学学报

北航学报 赠阅

Modular design method for filament winding machine

XU Xiaoming, ZHANG Wuxiang*, DING Xilun

(School of Mechanical Engineering and Automation, Beijing University of Aeronautics and Astronautics, Beijing 100083, China)

Abstract: Winding machine is the core equipment for filament winding with the features of high production efficiency and stable product quality. To meet the requirements of customized and small batch production, the modular design method (MDM) is proposed to expand the function of winding machine. The system structure is analyzed, relation matrices between components are established, and the grouping genetic algorithm (GGA) is then employed to conduct modular optimization to cluster components into standard modules. Multi-objective optimization method based on non-dominated sorting genetic algorithm Ⅱ (NSGA-Ⅱ) is proposed to create a complete system by combining instances with consideration of performance and cost simultaneously. The modular design method for the control system is presented based on distributed network controller for a bus system, and the network interface of the controller is standardized as independent function module, and then the rapid configuration of the controller is achieved according to the modular configuration of the mechanical structure. The k-nearest neighbor (kNN) method is used to classify the control mode by detecting the connection state of the module and the dynamic reconfiguration method based on component object model (COM) component to realize the state transition sequence and data exchange of the modules. The research on the modularization of mechanical structure, controller and software can realize the rapid reconfiguration of the winding machine, and expand the function of the winding machine.

Keywords: filament winding machine; modular method; multi-objective optimization; distributed control; component object model (COM) technology

Received: 2017-04-17; **Accepted**: 2017-05-26; **Published online**: 2017-07-11 11:47

URL: kns.cnki.net/kcms/detail/11.2625.V.20170711.1147.006.html

Foundation item: National Natural Science Foundation of China (51575018)

* **Corresponding author**. E-mail: zhangwuxiang@buaa.edu.cn

北航学报 赠阅

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2017.0249

行星三体引力摄动对卫星探测器大气制动的影响

全下文载

江玲，王悦*，徐世杰

（北京航空航天大学 宇航学院，北京 100083）

摘 要：考虑行星引力在其卫星探测器大气制动过程中的显著摄动，建立了基于 Milankovitch 参数的平均轨道动力学模型并对土卫六探测器进行仿真。首先，将轨道参数转换为无奇异的 Milankovitch 参数，考虑探测卫星的大气阻力、扁率摄动以及行星引力摄动，建立了半解析轨道方程。其次，以土卫六探测器为对象，选择不同的土星初始方位角进行有大气和无大气情况下的数值仿真，并进行比较分析。结果表明，土星初始方位角的选择会引起土卫六大气制动轨道偏心率和近拱点高度在不同范围内震荡，极大地影响大气制动效果。

关 键 词：大气制动；三体引力摄动；Milankovitch 参数；平均轨道要素；半解析轨道运动方程

中图分类号：V412.4+.1

文献标识码：A **文章编号**：1001-5965(2018)04-0759-06

大气制动是一种气动力辅助的航天器变轨方式[1]。将转移轨道的近拱点置于环绕天体的大气内，利用大气阻力降低航天器的轨道速度，通过多次穿越大气，使轨道远拱点高度逐步降低至目标高度，此时在远拱点执行脉冲变轨，即可实现环绕目标天体的任务轨道。这个过程所消耗的推进剂质量远小于直接变轨消耗的推进剂，可以大量节省深空探测器宝贵的燃料，从而延长寿命周期。到目前为止，已经有 3 次火星探测任务成功使用了大气制动技术。2016 年欧洲太空局发射的 ExoMars任务的火星微量气体轨道器（TGO）目前正在实施大气制动。应用结果表明大气制动技术对于深空探测任务的价值极大[2]。在未来也是深空探测技术的重点组成部分。

近年来，随着航天技术的发展，大气制动在轨道建模、优化和控制等问题的研究越来越深入。大量的研究集中在大气层内飞行轨迹的优化和制导方面，如在满足热流的条件下使得探测器消耗的燃料最少[3-4]，为防止热流过大对近拱点高度进行控制[5]，在气动参数变化的情况下能跟踪设计轨迹等[6-7]。还有一部分研究涉及到轨道建模，如近拱点选择对大气制动过程的影响[8]，大气制动期间的气动模拟方法[9-10]，时间消耗与探测器、轨道参数的关系[11]，以及旋转和静止大气环境下的轨道特性[12]等。但都没有涉及到气动力辅助变轨期间三体引力的影响。

众所周知，太阳系中最有科学价值的天体是木卫二、土卫六和土卫二，而土卫六是太阳系中唯一拥有浓密大气层的卫星，同地球一样富含氮气，成为当前国际深空探测的热点之一[13]。“卡西尼-惠更斯”计划的顺利进行更推动了对土卫六的探测活动。土卫六具有稠密的大气，在未来探测任务中可以在其附近实施大气制动以节约变轨的燃料消耗。

在过去火星等天体的大气制动轨道研究中，三体引力摄动比较微弱，或摄动引起的偏心率/近

收稿日期：2017-04-25；**录用日期**：2017-06-16；**网络出版时间**：2017-09-05 10:06
网络出版地址：kns.cnki.net/kcms/detail/11.2625.V.20170905.1006.001.html
*通信作者. E-mail：ywang@buaa.edu.cn

引用格式：江玲，王悦，徐世杰. 行星三体引力摄动对卫星探测器大气制动的影响[J]. 北京航空航天大学学报，2018，44(4)：759-764. JIANG L，WANG Y，XU S J. Effect of planet's third-body gravitational perturbation on aerobraking of moon's explorer[J]. Journal of Beijing University of Aeronautics and Astronautics，2018，44(4)：759-764(in Chinese).

北航学报 赠阅

拱点震荡周期与大气制动的时间相比很长,三体摄动基本被忽略。然而,土卫六附近的探测器轨道受到土星周期为 8 d(土卫六轨道周期的一半)的显著三体引力摄动[14],引起偏心率/近拱点的大幅震荡,会对探测器所处环境的大气密度产生显著影响,可能会加速大气制动的过程,也可能导致大气制动失败或探测器坠毁。因此,有必要对土星三体引力摄动对土卫六附近大气制动的影响进行详细研究。

本文采用基于 Milankovitch 参数的半解析轨道方程来研究探测器在环绕土卫六大气制动过程中土星三体引力摄动的影响。首先,考虑扁率摄动、三体引力摄动以及大气阻力,基于无奇异的 Milankovitch 参数建立了半解析轨道模型。然后,通过数值计算分析土星初始方位角选择对土卫六大气制动轨道运动的影响,为地外天体大气制动变轨方案提供参考。

1 轨道动力学模型及分析

1.1 轨道动力学建模

对于大气制动的轨道来说,大气阻力与三体引力摄动均为长期影响,适用于平均轨道要素建模。将长周期和短周期摄动隔离开,就得到摄动影响和轨道变化的直接关系。而 Milankovitch 轨道要素具有简洁的向量形式和对物理关系的显式表达,没有奇异,有利于摄动影响的讨论[15]。

大气制动过程中,探测器围绕卫星运行,将反复穿越卫星的大气层,同时卫星绕行星运动。卫星、探测器和行星的几何关系示意图如图 1 所示,$\boldsymbol{p}$为卫星自转轴方向矢量,$\boldsymbol{H}$为探测器轨道角动量矢量,$\boldsymbol{e}$ 为探测器轨道偏心率矢量,$\boldsymbol{d}_S$ 为行星相对于卫星的位置矢量,θ 为行星相对于轨道拱线的方位角,$\boldsymbol{r}$ 和 $\boldsymbol{v}$ 分别为探测器的位置和速度向量。

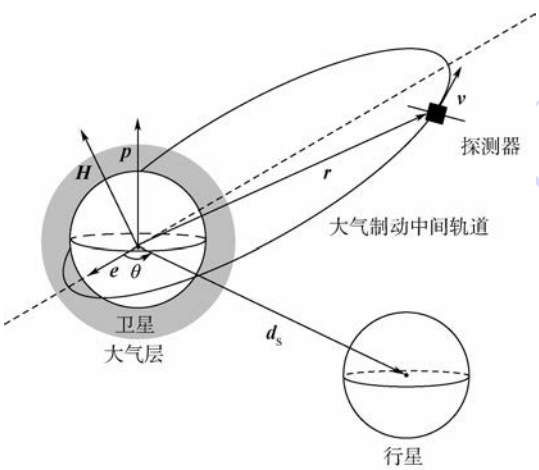


图 1 大气制动过程中卫星、探测器和行星的几何关系示意图

Fig. 1 Sketch map of geometric relation among moon, explorer and planet during aerobraking

将角动量 $\boldsymbol{H}$ 和偏心率矢量 $\boldsymbol{e}$ 写成位置 $\boldsymbol{r}$ 和速度 $\boldsymbol{v}$ 运算的形式

$$\boldsymbol{H}=\boldsymbol{r}\times\boldsymbol{v} \tag{1}$$

$$\boldsymbol{e}=\frac{1}{\mu}\boldsymbol{v}\times\boldsymbol{r}\times\boldsymbol{v}-\frac{\boldsymbol{r}}{r} \tag{2}$$

式中:μ 为卫星的引力常数;r 为探测器距卫星的距离。

对 $\boldsymbol{H}$ 和 $\boldsymbol{e}$ 建立高斯方程,并且将其平均化,有[15]

$$\dot{\overline{\boldsymbol{H}}}=\frac{1}{2\pi}\int_0^{2\pi}\dot{\boldsymbol{H}}\mathrm{d}M=\frac{1}{2\pi}\int_0^{2\pi}\boldsymbol{r}\times\boldsymbol{a}_{\mathrm{p}}\mathrm{d}M \tag{3}$$

$$\dot{\overline{\boldsymbol{e}}}=\frac{1}{2\pi}\int_0^{2\pi}\dot{\boldsymbol{e}}\mathrm{d}M=\frac{1}{2\pi}\int_0^{2\pi}\frac{1}{\mu}[\boldsymbol{v}\times(\boldsymbol{r}\times\boldsymbol{a}_{\mathrm{p}})-\boldsymbol{H}\times\boldsymbol{a}_{\mathrm{p}}]\mathrm{d}M \tag{4}$$

式中:M 为平近点角;$\boldsymbol{a}_{\mathrm{p}}=\boldsymbol{a}_{J2}+\boldsymbol{a}_S+\boldsymbol{a}_{\mathrm{atm}}$为摄动力,包括扁率摄动力 $\boldsymbol{a}_{J2}$、行星三体引力摄动力 $\boldsymbol{a}_S$ 和大气阻力 $\boldsymbol{a}_{\mathrm{atm}}$,定义分别为[15-16]

$$\boldsymbol{a}_{J2}=\frac{3\mu J_2R_{\mathrm{E}}^2}{2r^4}\{[1-5(\hat{\boldsymbol{r}}\hat{\boldsymbol{p}})^2]\hat{\boldsymbol{r}}+2(\hat{\boldsymbol{r}}\hat{\boldsymbol{p}})\hat{\boldsymbol{p}}\} \tag{5}$$

$$\boldsymbol{a}_S=-\mu_S\left[\frac{\boldsymbol{r}-\boldsymbol{d}_S}{|\boldsymbol{r}-\boldsymbol{d}_S|}+\frac{\boldsymbol{d}_S}{|\boldsymbol{d}_S|^3}\right] \tag{6}$$

$$\boldsymbol{a}_{\mathrm{atm}}=-\frac{1}{2}\rho\delta|\boldsymbol{v}-\boldsymbol{v}_{\mathrm{atm}}|(\boldsymbol{v}-\boldsymbol{v}_{\mathrm{atm}}) \tag{7}$$

$$\rho=\rho_{\mathrm{p0}}\exp\left(-\frac{r-r_{\mathrm{p0}}}{H_{\mathrm{atm}}}\right) \tag{8}$$

式中:J_2 为卫星扁率;R_{E} 为卫星平均半径;$\hat{\boldsymbol{r}}$ 和 $\hat{\boldsymbol{p}}$ 分别为位置和卫星自转轴单位向量;μ_S 为行星引力常数;ρ 为大气密度,此处采用指数模型;$\delta=SC_D/m$ 为探测器的弹道系数,S、C_D、m 分别为探测器的有效面积、气动系数和质量;$\boldsymbol{v}_{\mathrm{atm}}$为大气速度矢量;$\rho_{\mathrm{p0}}$为 r_{p0}处的大气密度;H_{atm}为大气标高。

将式(5)~式(8)代入式(3)和式(4)经过运算化简后可得到与文献[15]类似的结果:

$$\dot{\overline{\boldsymbol{H}}}=\dot{\overline{\boldsymbol{H}}}_{J2}+\dot{\overline{\boldsymbol{H}}}_S+\dot{\overline{\boldsymbol{H}}}_{\mathrm{atm}}=\frac{3a^2\mu_S}{2|\boldsymbol{d}_S|^3}[5(\hat{\boldsymbol{d}}_S\boldsymbol{e})\boldsymbol{e}\times\hat{\boldsymbol{d}}_S-(\hat{\boldsymbol{d}}_S\boldsymbol{h})\boldsymbol{h}\times\hat{\boldsymbol{d}}_S]-\frac{1}{2}\delta\sqrt{\frac{\mu(1-e^2)}{2a\pi z}}\rho_{\mathrm{p0}}\exp\left(\frac{r_{\mathrm{p0}}-r_{\mathrm{p}}}{H_{\mathrm{atm}}}\right)(1+K_1)\boldsymbol{H}-\frac{3\mu J_2R_{\mathrm{E}}^2}{2a^3|\boldsymbol{h}|^5}(\hat{\boldsymbol{p}}\boldsymbol{h})\hat{\boldsymbol{p}}\times\boldsymbol{h} \tag{9}$$

$$\dot{\overline{\boldsymbol{e}}}=\dot{\overline{\boldsymbol{e}}}_{J2}+\dot{\overline{\boldsymbol{e}}}_S+\dot{\overline{\boldsymbol{e}}}_{\mathrm{atm}}=\frac{3\mu_S}{2n|\boldsymbol{d}_S|^3}[5(\hat{\boldsymbol{d}}_S\boldsymbol{e})\boldsymbol{h}\times$$

北航学报 赠阅

$$\hat{\boldsymbol{d}}_S - (\hat{\boldsymbol{d}}_S\boldsymbol{h})\boldsymbol{e}\times\hat{\boldsymbol{d}}_S - 2\boldsymbol{h}\times\boldsymbol{e}] - \frac{3nJ_2R_E^2}{4a^2|\boldsymbol{h}|^5}\left\{\left[1-\frac{5}{|\boldsymbol{h}|^2}(\hat{\boldsymbol{p}}\boldsymbol{h})^2\right]\boldsymbol{h}\times\boldsymbol{e} + 2(\hat{\boldsymbol{p}}\cdot\boldsymbol{h})\hat{\boldsymbol{p}}\times\boldsymbol{e}\right\} - \delta\frac{1+e}{a\sqrt{2\pi z}}\rho_{p0}\cdot \exp\left(\frac{r_{p0}-r_p}{\boldsymbol{H}_{atm}}\right)(1+K_2)\boldsymbol{H}\hat{\boldsymbol{e}} \tag{10}$$

式中：$\hat{\boldsymbol{d}}_S$ 和 $\hat{\boldsymbol{e}}$ 分别为 $\boldsymbol{d}_S$ 和 $\boldsymbol{e}$ 的单位向量，$\boldsymbol{e}=e\hat{\boldsymbol{e}}$，$e$ 为偏心率；$\boldsymbol{h}=\boldsymbol{H}\sqrt{\mu a}$，$a$ 为轨道半长轴；$z=ae/H_{atm}$；$K_1=\dfrac{1+3e^2}{8z(1-e^2)}$；$n=\sqrt{\mu/a^3}$ 为轨道运动的平均角速度；$K_2=\dfrac{3e^2-4e-3}{8z(1-e^2)}$。

1.2　摄动力影响分析

考虑式(9)、式(10)等号右边第 1 项 $\dot{\boldsymbol{H}}_{J2}$ 和 $\dot{\boldsymbol{e}}_{J2}$，均与 $\boldsymbol{e}$ 垂直，因此扁率只会影响轨道平面和半长轴进动，而不影响轨道大小变化。

同样地，考虑式(9)、式(10)等号右边第 3 项 $\dot{\boldsymbol{H}}_{atm}$ 和 $\dot{\boldsymbol{e}}_{atm}$，大气阻力只会引起半长轴和偏心率的减小，而不影响轨道平面方向变化。

而行星三体引力摄动的分析就比较复杂。由 $\dot{\boldsymbol{e}}_S$ 可以看到偏心率变化的大小和符号由土星相对于椭圆轨道拱线的方位角 $\theta=\langle\boldsymbol{d}_S,\boldsymbol{e}\rangle$（见图 1）决定。以土星为例，该依赖关系可近似由正弦曲线描述，如图 2 所示。

土星的运动导致其相对于轨道拱线的方位角不断增大，从而导致椭圆轨道的偏心率发生震荡。由图 2 可知，该震荡的周期为土卫六轨道周期的一半，且震荡的初始相位由土星的初始方位角决定：若土星初始方位角为 0°，则初始偏心率为震荡的最高点；而若土星初始方位角为 90°，则初始偏心率为震荡的最低点。

因此土星三体引力摄动会引起轨道偏心率的震荡，从而使轨道的近拱点抬高或者降低。对于有大气的天体来说，这影响着轨道的寿命。而对于大气制动过程来说，近拱点的抬高或降低意味着探测器穿越大气层时利用大气制动效果的减弱或者加强。

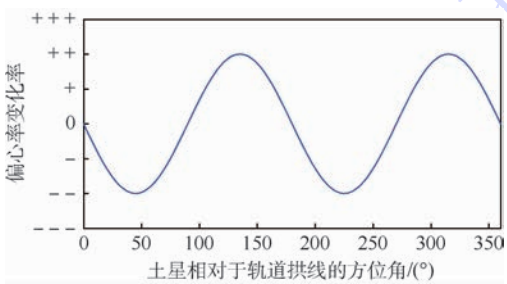


图 2　偏心率变化率与土星相对于轨道拱线的方位角近似关系

Fig. 2　Approximation of eccentricity variation with respect to azimuth between Saturn and apse line

2　数值仿真及分析

按照轨道模型式(9)、式(10)，本文给出一组不同初始方位角情况下的土卫六探测器大气制动轨道数值仿真，并分析土星三体引力摄动对大气制动过程的影响。大气制动可以节省的燃料消耗主要与轨道的初始和结束状态相对目标轨道的关系有关，不在本文的讨论范围，因此在仿真中并未考虑。仿真参数如表 1 所示，其中大气密度参数根据文献[17]数据拟合得到。

表 1 中轨道根数与 Milankovitch 参数的转换关系可参见文献[15]。仿真时长为 200 d，仿真步长为 10 s，并设置偏心率和近拱点高度阈值分别为 0.008 和 100 km，如果偏心率或轨道近拱点高度达到此值，中断仿真。

下面暂不考虑大气阻力，给出不同土星初始方位角时的轨道运动如图 3 所示。

从图 3 可以发现，不同的土星初始方位角将导致明显不同的偏心率运动范围和近拱点高度运动范围。当土星初始方位角从 30°增加到 50°时，近拱点高度震荡中心明显变小。这意味着如果在有大气的情况下，初始方位角为 50°的轨道将在穿越大气过程中经历更高的大气密度，减速效果更强烈。

表 1　仿真参数

Table 1　Simulation parameters

参　数	数　值
土卫六引力系数/($km^3\cdot s^{-2}$)	2.2187×10^{10}
土卫六平均半径/km	2575.5
扁率	5.0446×10^{-6}
土星引力系数/($km^3\cdot s^{-2}$)	3.792098×10^{7}
土卫六距土星距离/km	1.22187×10^{6}
土卫六偏心率	0
黄赤交角/(°)	0
探测器面值比/($m^2\cdot kg^{-1}$)	0.01
初始近拱点高度/km	2000
初始近拱点处大气密度/($kg\cdot km^{-3}$)	3.162278×10^{-8}
标高/km	57.906
探测器阻力系数	2.2
初始远拱点高度/km	20000
初始轨道倾角/(°)	20
初始升交点赤经/(°)	0
近拱点幅角/(°)	0

北航学报 赠阅

考虑大气阻力后,不同土星初始方位角情况下的轨道运动如图 4 所示。

从图 4 可以发现,在三体引力摄动和大气阻力的双重作用下,不同土星初始方位角的轨道偏心率运动和近拱点高度运动范围和趋势明显不同。图 4(a)表明在土星初始方位角为 50°时,轨道偏心率很快达到了结束阈值,而 30°情况下时大气阻力作用并不明显。从图 4(b)可以看出土星初始方位角为 50°时近拱点高度震荡谷值最低,气动作用最明显,同时近拱点高度震荡的峰值也降低,这是由于气动阻力的作用,轨道半长轴也在减小,相应地近拱点也会降低。图 4(c)显示随着时间的变化,大气制动作用越来越快。考虑大气制动最关注的指标热流率$\dot{Q}=0.5\rho v^3$,此处 v 为探测器速度大小,结合表 2 可以看到,土星初始方位角为 50°的情况下,虽然大气制动经历的时间最短,但是热流率远高于其他情况,这将指导大气制动过程初始方位的选择。同时可以看到,远拱点和近拱点高度的波动状态,与之前在二体运

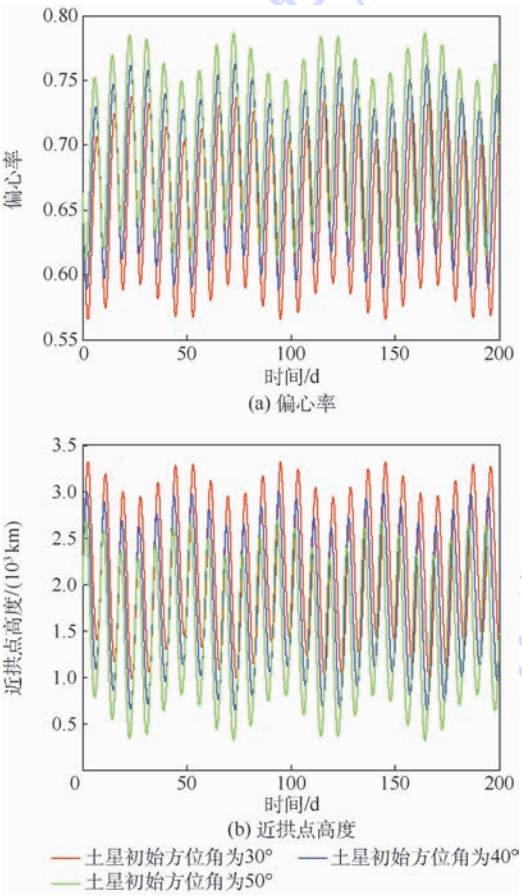


图 3 不考虑大气阻力,土星初始方位角分别为 30°、40°、50°时轨道偏心率、近拱点高度曲线

Fig. 3 Curves of eccentricity and periapsis height with initial azimuth of Saturn choices of 30°, 40° and 50°, when atmospheric drag is excluded

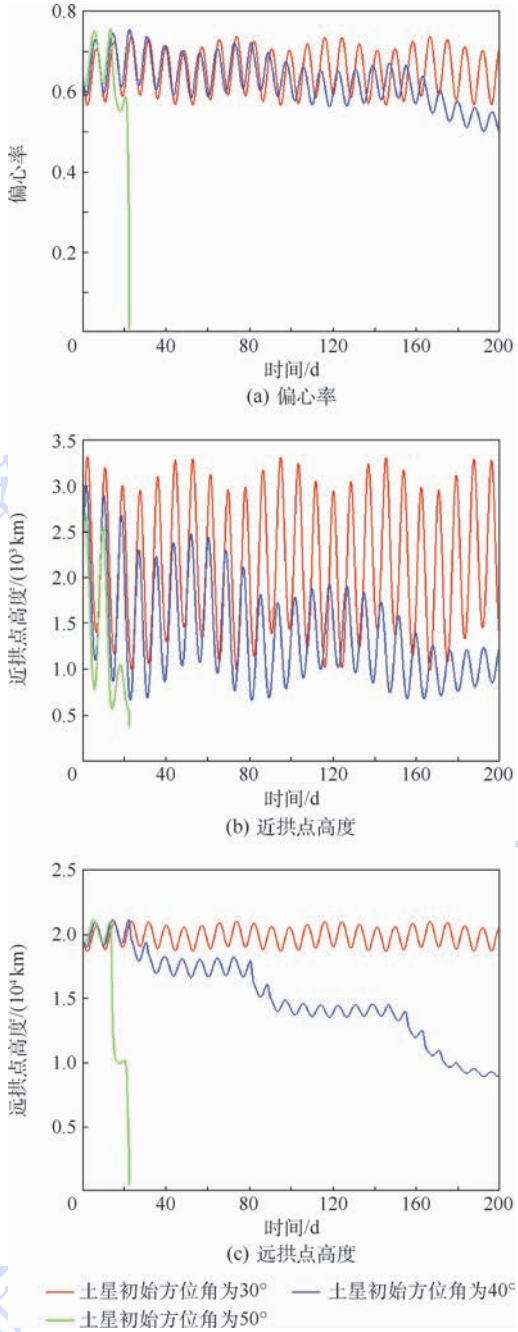


图 4 土星初始方位角分别为 30°、40°、50°时大气制动过程中轨道偏心率、近拱点高度和远拱点高度曲线

Fig. 4 Curves of eccentricity, periapsis height and apoapsis height with initial azimuth of Saturn choices of 30°, 40° and 50°, during aerobraking

表 2 不同土星初始方位角下大气制动参数

Table 2 Aerobraking parameters at different initial azimuth of Saturn

土星初始方位角/(°)	仿真结束时的半长轴/km	最大热流率/(W·m^{-2})	大气制动经历时间/d
30	13 559	0.134 5	200
40	7 596.2	38.19	200
50	2 694.5	3 331.7	22.66

北航学报 赠阅

动假设下的大气制动过程中单调下降的趋势不同，将影响大气制动退出条件以及燃料消耗。

在大气制动过程中，如果不考虑土星三体引力摄动，相同的轨道初始条件，期间不加入远拱点机动降低近拱点高度，大气制动过程将超过数十年。

3　结　论

通过轨道动力学建模和数值仿真讨论了土星三体引力对土卫六大气制动过程的影响，仿真结果表明：

1）土星的三体引力摄动对土卫六大气制动过程的影响较大，会极大地影响大气制动过程。在土卫六大气制动设计中不能忽略土星三体引力摄动。

2）三体引力摄动对大气制动的影响主要取决于土星的初始方位角。初始方位角的不同导致椭圆轨道的偏心率震荡范围的不同，从而引起近拱点高度震荡范围的变化，导致不同的大气制动效果。

3）本文采用了 Milankovitch 参数和平均轨道要素的方法来计算大气制动过程，拓宽了大气制动的仿真手段，为未来进一步开展地外天体气动力辅助变轨技术的研究奠定了良好的基础。

参考文献（References）

[1] MUNK M M, POWELL R W. Aeroassist technology planning for exploration[J]. Advances in the Astronautical Sciences, 2000, 105(2): 1073-1083.

[2] SPENCER D A, TOLSON R. Aerobraking cost and risk decisions [J]. Journal of Spacecraft and Rockets, 2007, 44(6): 1285-1293.

[3] RAO A V, TANG S, HALLMAN W P. Numerical optimization study of multiple-pass aeroassisted orbital transfer[J]. Optimal Control Applications and Methods, 2002, 23(4): 215-238.

[4] 吴德隆，王小军. 航天器气动力辅助变轨动力学与最优控制[M]. 北京：中国宇航出版社，2006：10.
WU D L, WANG X J. Aeroassisted orbit transfer dynamics and optimal control for spacecraft[M]. Beijing: China Astronautics Publishing House, 2006: 10(in Chinese).

[5] 季英良，朱宏玉，杨博. 利用气动力的大气制动过程中近心点高度控制[J]. 北京航空航天大学学报，2015，41(3)：517-522.
JI Y L, ZHU H Y, YANG B. Perigee altitude control using aerodynamic force during aerobraking[J]. Journal of Beijing University of Aeronautics and Astronautics, 2015, 41(3): 517-522 (in Chinese).

[6] JITS R Y, WALBERG G D. Blended control, predictor-corrector guidance algorithm: An enabling technology for mars aerocapture[J]. Acta Astronautica, 2004, 54(6): 385-398.

[7] LU P, CERIMELE C J, TIGGES M A, et al. Optimal aerocapture guidance [J]. Journal of Guidance, Control, and Dynamics, 2015, 38(4): 553-565.

[8] 陈统，徐世杰. 火星轨道大气制动策略研究[C]// 全国第十四届空间及运动体控制技术学术年会论文集. 北京：中国自动化学会空间及运动体控制专业委员会，2010：129-134.
CHEN T, XU S J. Research on mars orbit aerobraking strategy [C]// Proceedings of the 14th National Academic Conference on Space- and Moving-body Control Technology. Beijing: Space and Moving-body Control Committee of Chinese Association of Automation, 2010: 129-134(in Chinese).

[9] KUMAR M, TEWARI A. Trajectory and attitude simulation for mars aerocapture and aerobraking[J]. Journal of Spacecraft and Rockets, 2006, 43(3): 585-593.

[10] 张文普，韩波，张成义. 大气制动期间探测器的气动特性和轨道计算[J]. 应用数学和力学，2010，31(9)：1016-1026.
ZHANG W P, HAN B, ZHANG C Y. Spacecraft aerodynamics and trajectory simulation during aerobraking[J]. Applied Mathematics and Mechanics, 2010, 31(9): 1016-1026 (in Chinese).

[11] 周垂红，刘林. 利用行星大气制动形成环绕型目标轨道的耗时问题[J]. 飞行器测控学报，2013，32(5)：438-443.
ZHOU C H, LIU L. Time needed to use aerobraking to insert planetary low orbiters[J]. Journal of Spacecraft TT&C Technology, 2013, 32(5): 438-443(in Chinese).

[12] 吕敬，张明明，龚胜平. 旋转大气下火星探测器轨道捕获[J]. 北京航空航天大学学报，2013，39(3)：315-319.
LV J, ZHANG M M, GONG S P. Aerocapture period under rotating atmospheric environment for mars vehicle[J]. Journal of Beijing University of Aeronautics and Astronautics, 2013, 39(3): 315-319(in Chinese).

[13] 吴耀，姚伟，吕晓辰，等. 一种用于土卫六探测的热机浮空器性能分析[J]. 宇航学报，2016，37(2)：223-228.
WU Y, YAO W, LV X C, et al. Performance analysis of a heat engine aeroboat for titan exploration[J]. Journal of Astronautics, 2016, 37(2): 223-228(in Chinese).

[14] VAN HOOLST T, RAMBAUX N, KARATEKIN Ö, et al. The effect of gravitational and pressure torques on Titan's length-of-day variations[J]. ICARUS, 2009, 200(1): 256-264.

[15] WANG Y, GURFIL P. Dynamical modeling and lifetime analysis of geostationary transfer orbits[J]. Acta Astronautica, 2016, 128: 262-276.

[16] ROSENGREN A J, SCHEERES D J. Long-term dynamics of high area-to-mass ratio objects in high-earth orbit[J]. Advances in Space Research, 2013, 52(8): 1545-1560.

[17] WAITE J H, BELL J, LORENZ R, et al. A model of variability in Titan's atmospheric structure[J]. Planetary and Space Science, 2013, 86: 45-56.

作者简介：

江玲　女，博士研究生。主要研究方向：航天器动力学与控制。

王悦　男，博士，副教授。主要研究方向：航天飞行动力学、深空探测、空间碎片减缓。

徐世杰　男，博士，教授，博士生导师。主要研究方向：航天器动力学与控制、鲁棒控制和深空探测。

北航学报 赠阅

Effect of planet's third-body gravitational perturbation on aerobraking of moon's explorer

JIANG Ling, WANG Yue*, XU Shijie

(School of Astronautics, Beijing University of Aeronautics and Astronautics, Beijing 100083, China)

Abstract: An averaged orbital dynamics model for a natural satellite aerobraking was established and an Titan explorer was simulated to study the notable planet's third-body gravitational perturbation during the process. Firstly, non-singular Milankovitch elements were introduced to represent the orbital motion. A semi-analytical orbital equation was obtained considering moon's atmospheric drag, oblateness perturbation, and planet's gravitational perturbation. Secondly, taking Titan explorer as an example, simulations and analysis were carried out with different azimuth choices between Saturn and apse line when atmospheric drag was excluded and included. The results show that different initial azimuth of Saturn with respect to the apsidal line would cause that eccentricity and periapsis height oscillate in different intervals, which will ultimately influence the aerobraking result.

Keywords: aerobraking; third-body gravitational perturbation; Milankovitch elements; averaged orbital elements; semi-analytical orbital motion equation

北京航空航天大学学报

Received: 2017-04-25; **Accepted**: 2017-06-16; **Published online**: 2017-09-05 10:06

URL: kns.cnki.net/kcms/detail/11.2625.V.20170905.1006.001.html

* **Corresponding author**. E-mail: ywang@buaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2017.0265

RP-3 航空燃油中 CO_2 扩散系数实验分析

李超越[1], 冯诗愚[1,*], 邵垒[1], 潘俊[2], 刘卫华[1]

(1. 南京航空航天大学 航空宇航学院 飞行器环境控制与生命保障工业和信息化部重点实验室, 南京 210016;
2. 中航工业南京机电液压工程研究中心 航空机电系统综合航空科技重点实验室, 南京 211106)

摘 要: CO_2 在航空燃油中的质量扩散规律对飞机油箱惰化系统的研究极其重要。搭建压力降法实验装置测量 CO_2 在 RP-3 航空燃油中的扩散系数,实验测试了 -20、0、20、40 和 60℃恒温条件下的 CO_2 气相空间压力随时间变化关系。根据 Fick 定律建立了容器中二维扩散方程,并采用数值解法,设定扩散系数值,求解气体在航空燃油中的浓度分布,根据质量守恒和实际气体状态方程可得到 CO_2 气相空间压力,并与实验记录的气体压力进行比较。以扩散系数为自变量,推导了实验与理论计算误差函数,并采用 Husain 单一变量搜索法,使误差函数值最小,可得到扩散系数最优解。研究还显示 CO_2 在 RP-3 航空燃油中的扩散系数随温度升高而增大,满足 Arrhenius 方程。

关 键 词: CO_2; RP-3 航空燃油; 压力降法; 传质; 扩散系数; 优化
中图分类号: V312; TQ021.4
文献标识码: A **文章编号**: 1001-5965(2018)04-0765-07

来源于绿色惰化的 3C 惰化系统(low temperature Controllable oxygen Consumed Catalytic inerting system)中,CO_2 和 N_2 作为惰性化气体返回油箱气相空间,降低油箱氧气浓度,达到安全惰化的目的[1-3]。而气体从燃油中逸出或向燃油中溶解的传质过程会改变油箱上部气相空间可燃混合物各组分的浓度,直接影响惰化系统的设计和优化。采用 Henry 定律来描述气体在石油制品中(包括燃油)的平衡溶解度是最为常见的方式,但是平衡溶解度仅能反映达到平衡时传质的量,无法根据不同的传质速率,给出达到此传质量所需的时间。扩散系数表征了气体在燃油中溶解逸出的速率,是计算一定时间内气体物质的量变化的基本物理量,因此了解气体在航空燃油中的真实扩散系数对油箱惰化系统的选型和设计具有重要意义。通常石油制品中 CO_2 溶解度远大于 N_2[4],因此掌握 CO_2 在航空燃油中的扩散系数对 3C 惰化系统的研究尤为关键。

气体在液体中扩散系数研究常见于石油化工行业,包括其向原油及多孔介质扩散传质的报道[5-8],且研究以实验测定为主。测量气体在不同温度和压力下的扩散系数有直接和间接测量两大类方法[9]。直接测量主要是混合物成分分析法[10-11],其通过扩散过程中质谱分析原理,分析气液混合物中气体浓度变化而得到气体扩散系数值,但成分分析法存在实验花费大、分析过程繁杂、对扩散过程产生影响等问题,因此使用并不普遍;间接测量有压力降法[12-15]、恒定压力测量法[6,16]、激光全息干涉测量法[17-19]等。相较于其他几种实验方法,压力降法由于精确度高、代价

收稿日期: 2017-04-27; **录用日期**: 2017-08-31; **网络出版时间**: 2017-09-22 10:27
网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20170922.1027.005.html
基金项目: 航空科学基金(20132852040); 江苏省研究生科研与实践创新计划项目(KYCX17_0279); 中航工业技术创新基金(2014D60931R); 中央高校基本科研业务费专项资金; 江苏高校优势学科建设工程基金
* **通信作者**. E-mail: shiyuf@nuaa.edu.cn

引用格式: *李超越, 冯诗愚, 邵垒, 等. RP-3 航空燃油中 CO_2 扩散系数实验分析[J]. 北京航空航天大学学报, 2018, 44(4): 765-771. LI C Y, FENG S Y, SHAO L, et al. Experimental analysis of carbon dioxide diffusion coefficient in RP-3 jet fuel[J]. Journal of Beijing University of Aeronautics and Astronautics, 2018, 44(4): 765-771 (in Chinese).*

北航学报 赠阅

低、时间短而得到广泛应用[20]。

压力降法的基本原理是:在恒温条件下密闭容器中,气体分子扩散溶解于液体后,气相空间物质减少,气体压力降低。当气体扩散系数增大,则压力降落速率快,反之则压力降低速率慢。

扩散系数与扩散体系中物质化学组成密切相关,不能简单地采用其他液体中的扩散系数进行类比或简化,如文献[8,10,21-24]测量了 CH_4 和 CO_2 在不同油品中的扩散系数,实验结果显示,不同条件下气体扩散系数差异较大。目前国内外文献中均无法查询到 CO_2 在国产 RP-3 航空燃油中的扩散系数,这对 3C 惰化系统的精准设计十分不利。有鉴于此,本文搭建了压力降法实验平台,测量了 CO_2 在 RP-3 航空燃油溶解时气相空间压力的变化。扩散系数的大小会造成压力变化速度不同,因此可根据压力来间接得到扩散系数,一般的做法是假设满足一维扩散条件,即认为扩散的垂直方向无限大,这样在忽略高次项的条件下,可显式得到扩散系数与压力的关系[8,10,12,22]。实际的容器直径总是有限的,因此无法满足一维扩散假设,上述解析解法存在一定误差,有鉴于此,本文采用数值解法,并认为扩散是二维的,但是数值解法中气体压力与扩散系数为隐式关系,需在一定数值范围内依次假设扩散系数值并根据 Fick 扩散方程、质量守恒和气体状态方程计算扩散过程中气体压力变化,并与实验测量压力值进行比较,定义计算与实验压力误差函数,当压力误差函数取值最小时可认为此时对应扩散系数为最优扩散系数值。数值解法减少了解析方程中的简化假设,并建立二维扩散模型求解扩散系数值,更加符合实际扩散过程,由此通过数值差分方法及误差分析法使计算结果更加精确。

1 实验材料和方法

压力降法通过测量密封不锈钢容器内与 RP-3航空燃油直接接触的 CO_2 的压力变化来计算气体扩散系数,实验装置如图 1 所示。RP-3 航空燃油由容器底部注入,燃油高度 41.5 mm,气体由高压气瓶降压后从容器上方引入,直到达到实验开始所需压力,气相高度 78.5 mm,扩散容器截面直径 70 mm。工质充注结束后关闭管路阀门,实验开始,并利用压力传感器和数据采集仪记录气体压力随时间变化。为保持实验过程中温度恒定,整个实验过程在恒温水浴中进行。

为减少实验误差,实验开始前须对实验装置进行气密性检测,防止因泄漏而导致气体压力降

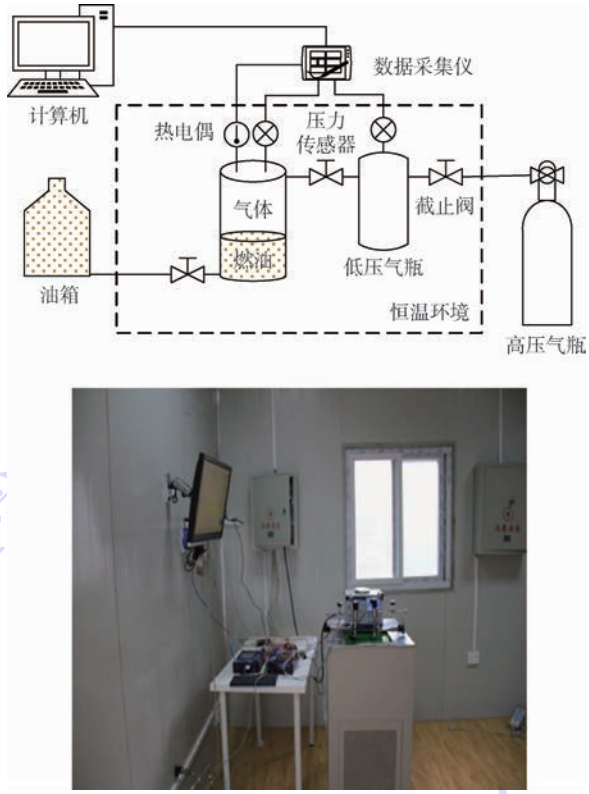


图 1 压力降法实验装置图

Fig. 1 Schematic diagram of pressure-decay experiment apparatus

低,保证实验可靠性。实验中所用 CO_2 纯度为 99.99%,RP-3 航空燃油由航空机电系统综合航空科技重点实验室提供。主要实验设备包含热电偶 T、压力传感器(量程 0 ~ 0.5 MPa)、恒温水浴 DC-3030,其精度分别为 ±0.1℃、F.S. ±0.1%、±0.1℃。

2 扩散模型

气体扩散过程如图 2 所示。对气-液扩散进行数学分析须选择合适的扩散方程并对实验模型作出如下假设:

1) CO_2 气相空间不存在浓度梯度。

2) 扩散过程为二维各向同性扩散。

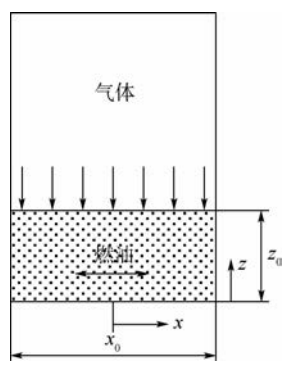


图 2 扩散模型示意图

Fig. 2 Schematic diagram of diffusion model

3）气-液边界处航空燃油中 CO_2 浓度为各时刻气体压力条件下对应的饱和溶解度[7]。

4）扩散系数为与浓度无关的常数[8]。

5）由于在实验温度条件下扩散体系中燃油蒸气压力相对于 CO_2 气体压力极低，且当温度恒定时燃油蒸气压力保持不变，因此实验过程中可忽略燃油蒸气压力对实验的影响[15]。

6）扩散体系中不发生化学反应[25]。

根据 Fick 第一定律，二维扩散可表示为

$$\begin{cases}\dfrac{\mathrm{d}c_z}{\mathrm{d}t}+v\dfrac{\mathrm{d}c_z}{\mathrm{d}z}=D\dfrac{\mathrm{d}^2c_z}{\mathrm{d}z^2}\\ \dfrac{\mathrm{d}c_x}{\mathrm{d}t}+v\dfrac{\mathrm{d}c_x}{\mathrm{d}x}=D\dfrac{\mathrm{d}^2c_x}{\mathrm{d}x^2}\end{cases}\tag{1}$$

式中：c 为气体在液体中浓度，$\mathrm{mol/m^3}$；t 为时间，s；v 为流体速度，m/s；z 为液体高度位置，m；x 为液体水平方向位置，m；D 为扩散系数，$\mathrm{m^2/s}$。

静止条件下密闭扩散容器内燃油流速可忽略不计，则式(1)可表示为

$$\begin{cases}\dfrac{\mathrm{d}c_z}{\mathrm{d}t}=D\dfrac{\mathrm{d}^2c_z}{\mathrm{d}z^2}\\ \dfrac{\mathrm{d}c_x}{\mathrm{d}t}=D\dfrac{\mathrm{d}^2c_x}{\mathrm{d}x^2}\end{cases}\tag{2}$$

根据本节假设 3)，气-液边界扩散浓度为 t 时刻气体压力为 p 时所对应平衡溶解度，则对应扩散模型边界条件为

初始条件：

$$c=0\qquad 0\leqslant z\leqslant z_0,t=0\tag{3}$$

边界 1：

$$c(z,t)=c(p)\qquad z=z_0,t>0\tag{4}$$

边界 2：

$$\frac{\mathrm{d}c(z,t)}{\mathrm{d}z}=0\qquad z=0,t>0\tag{5}$$

边界 3：

$$\frac{\mathrm{d}c(x,t)}{\mathrm{d}x}=0\qquad x=-\frac{x_0}{2},t>0\tag{6}$$

边界 4：

$$\frac{\mathrm{d}c(x,t)}{\mathrm{d}x}=0\qquad x=\frac{x_0}{2},t>0\tag{7}$$

式中：z_0 为液体总高度，m；x_0 为液体水平方向总长度，m；p 为气体压力，Pa。

气体在石油制品中溶解度可用 ASTM D2780—92[4] 所提供的阿斯特瓦尔德系数表示。随着扩散进行，CO_2 溶解于 RP-3 航空燃油，气体压力降低，根据阿斯特瓦尔德系数可求得在不同时刻压力变化时 CO_2 在液相中溶解度，并以此作为式(4)在气-液界面处边界条件，运用中心差分法，假定扩散系数值求解扩散方程式(2)，计算可得燃油中不同位置在不同时刻气体浓度分布，由此可知液相中气体总的物质的量。根据质量守恒定律，扩散过程中气体溶解于燃油的物质的量与气体自身减少的物质的量相等。根据实际气体状态方程，在一定温度和压力下气体物质的量可表示为

$$n_{\mathrm{g}}=\frac{pV_{\mathrm{g}}}{ZRT}\tag{8}$$

式中：n_{g} 为气体物质的量，mol；V_{g} 为气体体积，$\mathrm{m^3}$；Z 为气体压缩因子；气体常数 $R=8.314\ \mathrm{J/(mol\cdot K)}$；$T$ 为温度，K。

实验过程中气体初始压力为 500 ~ 600 kPa，且扩散过程中压力变化范围不大，可认为气体压缩因子为常数，则实验过程中 CO_2 扩散导致气体压力变化为

$$\Delta p=\frac{\Delta n_{\mathrm{g}}ZRT}{V_{\mathrm{g}}}\tag{9}$$

式中：Δp 为气体压力变化，Pa；Δn_{g} 为气体物质的量变化，mol。

因此，容器内气体压力可表示为

$$p=p_{\mathrm{i}}-\frac{\Delta n_{\mathrm{g}}ZRT}{V_{\mathrm{g}}}\tag{10}$$

式中：p_{i} 为气体初始压力，Pa。

3　数值优化

首先进行不同温度下 CO_2 在 RP-3 航空燃油中压力降测量并记录。在与实验相同初始压力条件下，假定扩散系数值并结合上述扩散模型进行气相空间压力计算，与实验记录压力进行比较得到实验与理论计算气体压力平均误差函数。以扩散系数为独立变量，在一定范围内求得气体压力误差函数值，当误差函数取值最小时所对应的扩散系数即为最优解。

压力平均误差函数 Δp_{ave} 表示在扩散过程中时刻 $t_j(j=0,1,\cdots,m)$ 时理论计算压力与实验压力差值的平方平均数，单位为 Pa。平均误差函数可表示为

$$\Delta p_{\mathrm{ave}}=\sqrt{\frac{\sum_{j=1}^{m}|\Delta p_j|^2_{t=t_j}}{m}}=\sqrt{\frac{\sum_{i=1}^{m}|p_{\mathrm{cal}}(t)-p_{\mathrm{exp}}(t)|^2_{t=t_j}}{m}}\tag{11}$$

式中：Δp_j 为各时刻压力差值；p_{cal} 为计算压力，Pa；p_{exp} 为实验压力，Pa。

扩散初始条件确定后，扩散系数值为求解压

北航学报 赠阅

力变化唯一变量。根据 Husain 的单一变量搜索法[26],在预设扩散系数取值范围内,可求得扩散系数最优解。由文献[23,27]可知,气体在液体中扩散系数值取值范围一般为 $1\times10^{-11}\sim1\times10^{-6}$ m^2/s,在此范围内误差函数为单峰值函数,因此存在特定扩散系数使扩散过程中理论计算压力与实验记录压力误差最小,从而达到求解目的。

4 实验结果及分析

实验测量并记录了 -20、0、20、40、60℃ 5 组不同温度下 CO_2 在 CO_2-RP-3 航空燃油体系的压力降分布。扩散过程中每隔 4 s 时间选取压力值进行理论与实验压力对比。选择扩散系数计算步长为 0.1 m^2/s 进行求解,如图 3 所示为压力平均误差函数 Δp_{ave} 随扩散系数变化关系。从图 3 中可看出在扩散系数取值范围内 $f(D,\ \Delta p_{ave})$ 为一单峰函数,Δp_{ave} 存在最小值,与本文设想一致。

从图 3 中可看出在特定温度下,当 Δp_{ave} 取最小值 $\Delta p_{ave,min}$ 时,对应唯一确定的扩散系数值 D,如表 1 所示,且扩散系数步长的选取对最优扩散系数值产生的偏差仅为 0.9% ~1.6%,可满足要求。表 1 中所得扩散系数值即为 CO_2 在 RP-3 航空燃油中的最优解。

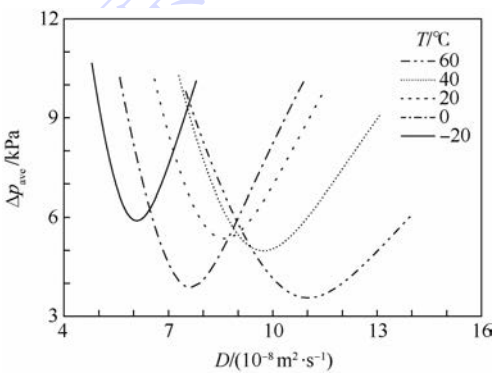


图 3 CO_2-RP-3 燃油体系不同温度下压力平均误差随扩散系数变化关系

Fig. 3 Variation of average pressure difference with diffusion coefficient under different temperature for CO_2-RP-3 jet fuel system

表 1 扩散系数最优解

Table 1 Optimum solution of diffusion coefficient

T/℃	$\Delta p_{ave,min}$/kPa	D/(10^{-8} m^2 · s^{-1})
-20	5.89	6.1
0	3.87	7.6
20	5.35	8.6
40	4.98	9.7
60	3.59	11.1

为验证最优扩散系数解能更准确地描述实际扩散过程,将所求扩散系数值 D 代入扩散方程模型,由此可计算得到扩散体系中气体压力降分布曲线。以 60℃扩散为例,取 3 组不同扩散系数值分别进行计算,其中一组为所求最优扩散系数 D,另外 2 组为任意大于或小于 D 的值,理论计算与实验记录压力变化对比如图 4 所示。由图 4 可知扩散系数取值不同时计算与实验压力之间平均误差发生变化,运用最优扩散系数值计算气体压力与实验记录之间误差最小,由此表明数值优化方法可以更加准确地求解 CO_2 在 RP-3 航空燃油中的扩散系数。

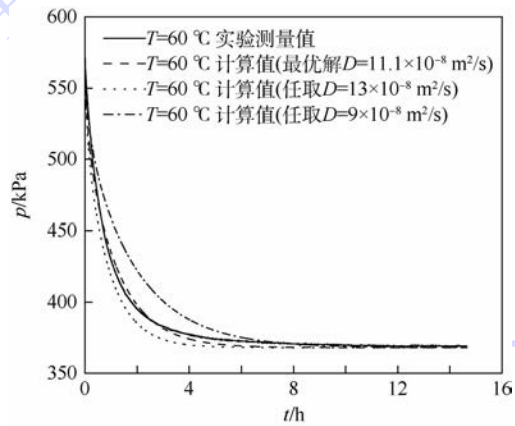


图 4 计算压力降与实验数据比较

Fig. 4 Comparison of calculated pressure decay with experimental data

从表 1 所求扩散系数值可知,CO_2 在 RP-3 航空燃油扩散体系中扩散系数随温度升高而增大,这是因为温度升高,气体分子热运动加剧,导致分子扩散速率加快。而且温度升高导致液体黏度降低,气体扩散系数增大[28]。当温度由 -20℃增大到 60℃时,扩散系数值增大了 81.97%。在工程应用中,扩散系数随温度升高而增大,可用 Arrhenius 方程描述:

$$D = A\exp\left(-\frac{E}{RT}\right) \tag{12}$$

式中:A 为数值拟合常数,m^2/(s · K);E 为表观活化能,J/mol。

对扩散系数与温度进行数值拟合,如图 5 所示,由最小二乘法拟合结果可知,CO_2 在 RP-3 航空燃油扩散体系中 A 为 6.8×10^{-7} m^2/(s · K),E 为 5 040.2 J/mol。

对本文实验系统和实验结果进行误差分析,以 60℃时 CO_2 扩散为例,在相同实验系统和初始条件下进行 3 次重复性实验,记录气体压力降过程,如图6所示。在相同条件下实验结果基本相同,运

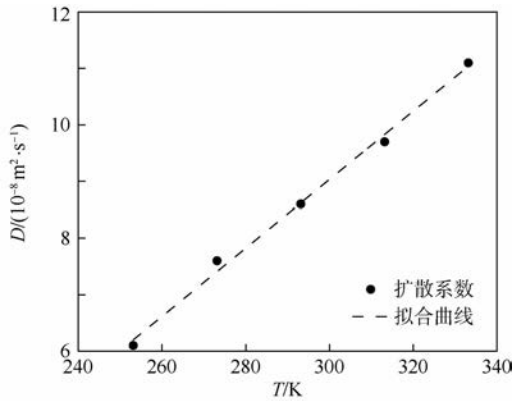


图 5　扩散系数随温度变化关系

Fig. 5　Relationship between diffusion coefficient and temperature

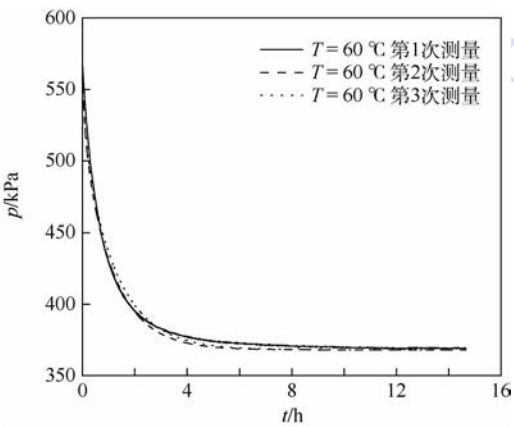


图 6　实验测量不同时刻压力变化

Fig. 6　Experimental measurement of variation of pressure versus time

用第 2、3 节方法求解扩散系数方程式(2)和误差函数方程式(3)，所得扩散系数分别为 10.9×10^{-8}、11.1×10^{-8}、11.7×10^{-8} m^2/s。本文所得扩散系数值与平均扩散系数值误差仅为 1.15%，因此在实验误差允许范围内，本实验系统和实验结果具有可靠性。

5　结　论

1）采用压力降法进行扩散系数实验测量，构建二维扩散模型，运用数值解法和误差分析能更加准确地描述实际扩散过程。

2）气体扩散理论模型和实验测量结果之间存在一定误差，但运用数值优化方法求解扩散系数可使偏差最小。

3）随着温度的升高，扩散系数值增大，当温度从 -20℃ 升高到 60℃ 时，扩散系数值增大了 81.97%，在不同温度条件下 CO_2 在 RP-3 航空燃油中的扩散系数可以用 Arrhenius 方程表示。通过对扩散系数的数值拟合，可了解更大温度范围内气体扩散系数值，为机载惰化系统的设计提供必要的数据支持。

参考文献（References）

[1] JOHNSON R W, ZAKI R, YATES S F. Advanced carbon dioxide fuel tank inerting system: US 7905259B2[P]. 2011-03-15.

[2] LIMAYE S, KOENIG D. Catalytic reactive component reduction system and methods for the use thereof: US 7694916B2[P]. 2010-04-13.

[3] REYNOLDS T L, EKLUND T I, HAACK G A. Onboard inert gas generation system onboard oxygen gas generation system (OBIGGS/OBOGS) study: Part Ⅱ: Gas separation technology-state of the art: NASA/CR-2001-210950[R]. Washington, D. C.: NASA, 2001.

[4] ASTM Committee. Standard test method for estimation of solubility of gases in petroleum liquids: ASTM D2780—92[S]. West Conshohocken: ASTM International, 2002.

[5] 王斐，邵晓红，汪文川. CH_4 和 CO_2 在活性炭微球内扩散系数的测定[J]. 化工学报，2006，57(8)：1891-1896.
WANG F, SHAO X H, WANG W C. Measurements of diffusion coefficients for methane and carbon dioxide in activated mesocarbon microbeads[J]. Journal of Chemical Industry and Engineering, 2006, 57(8): 1891-1896(in Chinese).

[6] ETMINAN S R, MAINI B B, CHEN Z X, et al. Constant-pressure technique for gas diffusivity and solubility measurements in heavy oil and bitumen[J]. Energy & Fuels, 2010, 24(1): 533-549.

[7] BEHZADFAR E, HATZIKIRIAKOS S G. Diffusivity of CO_2 in bitumen: Pressure-decay measurements coupled with rheometry [J]. Energy & Fuels, 2014, 28(2): 1304-1311.

[8] ZHANG Y P, HYNDMAN C L, MAINI B B. Measurement of gas diffusivity in heavy oils[J]. Journal of Petroleum Science and Engineering, 1999, 25(1-2): 37-47.

[9] RIAZI M R, WHITSON C H. Estimating diffusion coefficients of dense fluids[J]. Industrial & Engineering Chemistry Process Design & Development, 1993, 32(12): 3081-3088.

[10] UNATRAKARN D, ASGHARI K, CONDOR J. Experimental studies of CO_2 and CH_4 diffusion coefficient in bulk oil and porous media[J]. Energy Procedia, 2011, 4: 2170-2177.

[11] 严小伟，单奕彬，王靖岱. 反相气相色谱法测定小分子溶剂在聚乙烯粒子中的无限稀释扩散系数[J]. 化工学报，2007，58(8)：1917-1925.
YAN X W, SHAN Y B, WANG J D. Measurement of infinite dilution diffusion coefficients of small molecule solvents in nascent polyethylene particles by inverse gas chromatography[J]. Journal of Chemical Industry and Engineering, 2007, 58(8): 1917-1925 (in Chinese).

[12] CIVAN F, RASMUSSEN M L. Analysis and interpretation of gas diffusion in quiescent reservoir, drilling, and completion fluids: Equilibrium vs. non-equilibrium models[C] // SPE Annual Technical Conference and Exhibition. Richardson: SPE, 2003: 1-13.

[13] SHEIKHA H, DARVISH P M, MEHROTRA A K. Development of graphical methods for estimating the diffusivity coefficient of

北航学报 赠阅

gases in bitumen from pressure-decay data[J]. Energy & Fuels, 2005,19(5):2041-2049.

[14] JAMIALAHMADI M,EMADI M,STEINHAGEN H M. Diffusion coefficients of methane in liquid hydrocarbons at high pressure and temperature[J]. Journal of Petroleum Science and Engineering,2006,53(1-2):47-60.

[15] FARAJZADEH R,ZITHA P L J,BRUINING J. Enhanced mass transfer of CO_2 into water:Experiment and modeling[J]. Industrial & Engineering Chemistry Process Design & Development, 2009,48:6423-6431.

[16] DING C,FAN Y. Measurement of diffusion coefficients of air in silicone oil and in hydraulic oil[J]. Chinese Journal of Chemical Engineering,2011,19(2):205-211.

[17] BOCHNER N,PIPMAN J. A simple method of determining diffusion constants by holographic interferometry[J]. Journal of Physics D:Applied Physics,1976,9:1825-1830.

[18] 朱春英,马友光,赵长伟. 氨基酸在水溶液中的扩散系数[J]. 化工学报,2005,56(2):243-245.
ZHU C Y,MA Y G,ZHAO C W. Diffusion coefficients of amino acids in aqueous solutions[J]. Journal of Chemical Industry and Engineering,2005,56(2):243-245(in Chinese).

[19] COLOMBANI J,BERT J. Holographic interferometry for the study of liquids[J]. Journal of Molecular Liquids,2007,134(1-3):8-14.

[20] GHOLAMI Y,AZIN R,FATEHI R,et al. Suggesting a numerical pressure-decay method for determining CO_2 diffusion coefficient in water[J]. Journal of Molecular Liquids,2015,211:31-39.

[21] NGUYEN T A,ALI S M F. Role of diffusion and gravity segregation in oil recovery by the immiscible carbon dioxide wag process [C]// Conference: 6. UNITAR International Conference on Heavy Crude and Tar Sands on Fueling for a Clean and Safe Environment. New York:UNITAR,1995:393-403.

[22] THARANIVASAN A K,YANG C D,GU Y G. Comparison of three different interface mass transfer models used in the experimental measurement of solvent diffusivity in heavy oil[J]. Journal of Petroleum Science and Engineering,2004,44(3-4):269-282.

[23] RONGY L,HAUGEN K B,FIROOZABADI A. Mixing from Fickian diffusion and natural convection in binary non-equilibrium fluid phases[J] AIChE Journal,2012,58(5):1336-1345.

[24] SCHMIDT T,LESHCHYSHYN T H,PUTTAGUNTA V R. Diffusion of carbon dioxide into Athabasca bitumen[C]//33rd Annual Technical Meeting of the Petroleum Society of CIM. Richardson:SPE,1982:82-100.

[25] GHOLAMI Y,AZIN R,FATEHI R,et al. Prediction of carbon dioxide dissolution in bulk water under isothermal pressure decay at different boundary conditions[J]. Journal of Molecular Liquids,2015,202:3-33.

[26] HUSAIN A G K. Optimization techniques for chemical engineers [M]. Noida:Macmillan of India,1976:63-65.

[27] UPRETI S R,MEHROTRA A K. Diffusivity of CO_2, CH_4, C_2H_6 and N_2 in Athabasca bitumen[J]. The Canadian Journal of Chemical Engineering,2002,80(1):116-125.

[28] NAPOLEON O,UMESL R P D. Predicting diffusion coefficients in nonpolar solvents[J]. Industrial & Engineering Chemistry Process Design & Development,1981,20(4):662-665.

作者简介:

李超越 男,博士研究生。主要研究方向:飞机燃油箱惰性化技术。

冯诗愚 男,博士,副教授。主要研究方向:油箱惰性化和飞机燃油系统。

北京航空航天大学学报

Experimental analysis of carbon dioxide diffusion coefficient in RP-3 jet fuel

LI Chaoyue[1], FENG Shiyu[1,*], SHAO Lei[1], PAN Jun[2], LIU Weihua[1]

1. Key Laboratory of Aircraft Environment Control and Life Support of MIIT, College of Aerospace Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China; 2. Aviation Key Laboratory of Science and Technology on Aero Electromechanical System Integration, AVIC Nanjing Engineering Institute of Aircraft Systems, Nanjing 211106, China)

Abstract: The rule of mass diffusion of CO_2 in jet fuel is an important consideration in the research of the aircraft fuel tank inerting system. A pressure-decay test apparatus was constructed to measure the diffusion coefficient of carbon dioxide in RP-3 jet fuel by monitoring the pressure variation at constant temperature of −20, 0, 20, 40, and 60℃, respectively. The two-dimensional diffusion equation in the hermetic container was derived based on Fick's law, and the concentration distribution was determined via the numerical method with the assumption that diffusion coefficient was known. Then pressure on ullage was predicted by utilizing mass conservation and real gas state equation and the calculating result was compared with the experimental data. By adopting the diffusion coefficient as the independent variable, the error functions as the experimental and theoretical data were derived. The optimum solution to the diffusion coefficient was obtained by Husain single variable search method to minimize the error. The study also reveals that the diffusion coefficient increases with the rise of the temperature, and the Arrhenius equation could be employed to correlate the diffusion coefficient and temperature.

Keywords: carbon dioxide; RP-3 jet fuel; pressure-decay method; mass transfer; diffusion coefficient; optimization

Received: 2017-04-27; **Accepted**: 2017-08-31; **Published online**: 2017-09-22 10:27

URL: kns.cnki.net/kcms/detail/11.2625.V.20170922.1027.005.html

Foundation items: Aeronautical Science Foundation of China (20132852040); Postgraduate Research & Practice Innovation Program of Jiangsu Province (KYCX17_0279); Technology Innovation Foundation of AVIC (2014D60931R); the Fundamental Research Funds for the Central Universities; Priority Academic Program Development of Jiangsu Higher Education Institutions

* **Corresponding author**. E-mail: shiyuf@nuaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2017.0285

基于 LSTM 循环神经网络的故障时间序列预测

王鑫[1], 吴际[1,*], 刘超[1], 杨海燕[1], 杜艳丽[2], 牛文生[1,3]

(1. 北京航空航天大学 计算机学院, 北京 100083; 2. 丰台职业教育中心学校, 北京 100076;
3. 中航工业西安航空计算技术研究所, 西安 710068)

摘 要: 有效地预测使用阶段的故障数据对于合理制定可靠性计划以及开展可靠性维护活动等具有重要的指导意义。从复杂系统的历史故障数据出发,提出了一种基于长短期记忆(LSTM)循环神经网络的故障时间序列预测方法,包括网络结构设计、网络训练和预测过程实现算法等,进一步以预测误差最小为目标,提出了一种基于多层网格搜索的 LSTM 预测模型参数优选算法,通过与多种典型时间序列预测模型的实验对比,验证了所提出的 LSTM 预测模型及其参数优选算法在故障时间序列分析中具有很强的适用性和更高的准确性。

关 键 词: 长短期记忆(LSTM)模型; 循环神经网络; 故障时间序列预测; 多层网格搜索; 深度学习

中图分类号: O213.2; V37; TP18

文献标识码: A **文章编号**: 1001-5965(2018)04-0772-13

对于有高可靠性和安全性需要的复杂系统,有效地预测使用阶段的可靠性指标是十分重要的。目前,已有众多方法用来解决可靠性预测问题,这些方法大致可以分为3类[1]:①基于故障机理(Physics-of-Failure, PoF)的方法,PoF 是一种根据故障发生的内在机制和根本原因进行间接预测的方法;②数据驱动(Data-Driven, DD)的方法,DD 是一种应用统计学或者机器学习等技术手段对可靠性指标进行直接预测的方法;③融合的方法,这种方法是一种 PoF 和 DD 相结合的方法。近年来,数据驱动的方法由于其便捷性和高效性等特点,在实际可靠性预测中的应用日渐广泛[2-3]。

故障时间序列,作为一个重要的可靠性指标,能够展示故障的动态演化过程,并且已经被多种数据驱动的方法预测,比如自回归移动平均(Autoregressive Integrated Moving Average, ARIMA)[4]、奇异谱分析(Singular Spectrum Analysis, SSA)[5]、支持向量回归(Support Vector machines Regression, SVR)[6]、人工神经网络(Artificial Neural Network, ANN)[7]等。由于可靠性数据通常不易获取,已有的研究大多是面向组件级故障数据的,比如文献[7]提供的柴油机涡轮增压器和汽车发动机故障时间序列。这些数据表现为规则性很强的曲线形态并且能够被单一模型很好地拟合和预测。然而,对于系统级故障数据,比如文献[3]提供的民航飞机整机故障时间序列,由于其复杂且不规则的曲线形态,已有的单一模型很难达到理想的预测效果[8]。为此,文献[8]提出了一种基于 SSA 和 SVR 的混合方法,该方法首先从原始数据中提取故障特征成分然后分别建模和预测,得到了比单一模型更好的实验结果。然而,混合方

收稿日期: 2017-05-08; **录用日期**: 2017-08-11; **网络出版时间**: 2017-10-13 14:05
网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20171013.1405.003.html
基金项目: 中国民用航空专项研究项目(MJ-S-2013-10); 国防科工局技术基础项目(JSZL2014601B008); 国家自然科学基金(61602237)

* **通信作者**. E-mail: wuji@buaa.edu.cn

引用格式: 王鑫, 吴际, 刘超, 等. 基于LSTM 循环神经网络的故障时间序列预测[J]. 北京航空航天大学学报, 2018, 44(4): 772-784. WANG X, WU J, LIU C, et al. Exploring LSTM based recurrent neural network for failure time series prediction[J]. Journal of Beijing University of Aeronautics and Astronautics, 2018, 44(4): 772-784 (in Chinese).

法的构建过程复杂、人工依赖性强,不利于在实际中推广和使用。

近年来,随着深度学习技术的不断发展,一些深度学习模型逐渐被应用到时序数据的研究中。深度学习模型是一种拥有多个非线性映射层级的深度神经网络模型,能够对输入信号逐层抽象并提取特征,挖掘出更深层次的潜在规律[9]。在众多深度学习模型中,循环神经网络(Recurrent Neural Network, RNN)将时序的概念引入到网络结构设计中,使其在时序数据分析中表现出更强的适应性。在众多 RNN 的变体中,长短期记忆(Long Short-Term Memory, LSTM)模型[10]弥补了 RNN 的梯度消失和梯度爆炸、长期记忆能力不足等问题,使得循环神经网络能够真正有效地利用长距离的时序信息。LSTM 模型在不同领域的时序数据研究中已有不少成功的应用案例,包括文字语言相关的语言建模、语音识别、机器翻译[11],多媒体相关的音频和视频数据分析、图片标题建模[12-13],道路运输相关的交通流速预测[14],以及医学相关的蛋白质二级结构序列预测[15]等。然而,在可靠性领域,LSTM 模型的应用非常有限,特别是对于故障时间序列预测这一研究问题,目前还未发现相关研究。

本文针对系统级故障时间序列数据,提出了一种基于 LSTM 循环神经网络的预测方法,包括 3 层(输入层、隐藏层和输出层)网络结构的详细设计,以及网络训练和网络预测的实现算法等。在此基础上,以预测误差最小为目标,进一步提出了一种基于多层网格搜索的 LSTM 预测模型参数优选算法。应用文献[3]提供的民航飞机故障数据展开实验,并与 Holt-Winters、ARIMA 等多种时间序列预测模型进行实验对比。实验结果展示了所提出的 LSTM 预测模型及其参数优选算法在故障时间序列预测中的优越性能。

1　相关理论和技术

本节简要介绍 LSTM 模型,包括前向计算方法、基于时间的反向传播(Back Propagation Through Time, BPTT)算法、Adam 参数优化算法,以及相关的 RNN、GRU 模型。

对于给定序列 $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)$,应用一个标准的 RNN 模型[16](如图 1 所示),可以通过迭代式(1)~式(2)计算出一个隐藏层序列 $\boldsymbol{h}=(h_1,h_2,\cdots,h_n)$和一个输出序列 $\boldsymbol{y}=(y_1,y_2,\cdots,y_n)$。

$$\boldsymbol{h}_t=f_a(\boldsymbol{W}_{xh}\boldsymbol{x}_t+\boldsymbol{W}_{hh}\boldsymbol{h}_{t-1}+\boldsymbol{b}_h) \tag{1}$$

$$\boldsymbol{y}_t=\boldsymbol{W}_{hy}\boldsymbol{h}_t+\boldsymbol{b}_y \tag{2}$$

式中:$\boldsymbol{W}$ 为权重系数矩阵(比如 $\boldsymbol{W}_{xh}$ 表示输入层到隐藏层的权重系数矩阵);$\boldsymbol{b}$ 为偏置向量(比如 $\boldsymbol{b}_h$ 表示隐藏层的偏置向量);f_a 为激活函数(比如 tanh 函数);下标 t 表示时刻。

尽管 RNN 能够有效地处理非线性时间序列,但是仍然存在以下 2 个问题[14]:①由于梯度消失和梯度爆炸的问题,RNN 不能处理延迟过长的时间序列;②训练 RNN 模型需要预先确定延迟窗口长度,然而实际应用中很难自动地获取这一参数的最优值。由此,LSTM 模型应用而生。LSTM 模型是将隐藏层的 RNN 细胞替换为 LSTM 细胞,使其具有长期记忆能力。经过不断地演化,目前应用最为广泛的 LSTM 模型细胞结构[16]如图 2 所示,$\boldsymbol{z}$ 为输入模块,其前向计算方法可以表示为

$$\boldsymbol{i}_t=\sigma(\boldsymbol{W}_{xi}\boldsymbol{x}_t+\boldsymbol{W}_{hi}\boldsymbol{h}_{t-1}+\boldsymbol{W}_{ci}\boldsymbol{c}_{t-1}+\boldsymbol{b}_i) \tag{3}$$

$$\boldsymbol{f}_t=\sigma(\boldsymbol{W}_{xf}\boldsymbol{x}_t+\boldsymbol{W}_{hf}\boldsymbol{h}_{t-1}+\boldsymbol{W}_{cf}\boldsymbol{c}_{t-1}+\boldsymbol{b}_f) \tag{4}$$

$$\boldsymbol{c}_t=\boldsymbol{f}_t\boldsymbol{c}_{t-1}+\boldsymbol{i}_t\tanh(\boldsymbol{W}_{xc}\boldsymbol{x}_t+\boldsymbol{W}_{hc}\boldsymbol{h}_{t-1}+\boldsymbol{b}_c) \tag{5}$$

$$\boldsymbol{o}_t=\sigma(\boldsymbol{W}_{xo}\boldsymbol{x}_t+\boldsymbol{W}_{ho}\boldsymbol{h}_{t-1}+\boldsymbol{W}_{co}\boldsymbol{c}_t+\boldsymbol{b}_o) \tag{6}$$

$$\boldsymbol{h}_t=\boldsymbol{o}_t\tanh(\boldsymbol{c}_t) \tag{7}$$

式中:$\boldsymbol{i}$、$\boldsymbol{f}$、$\boldsymbol{c}$、$\boldsymbol{o}$分别为输入门、遗忘门、细胞状态、

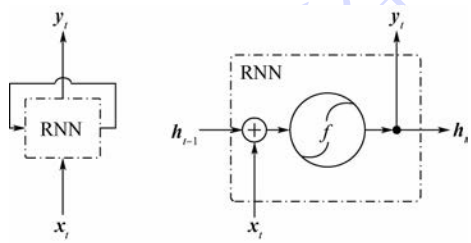


图 1　RNN 模型及隐藏层细胞结构

Fig. 1　RNN model and cell structure in hidden layer

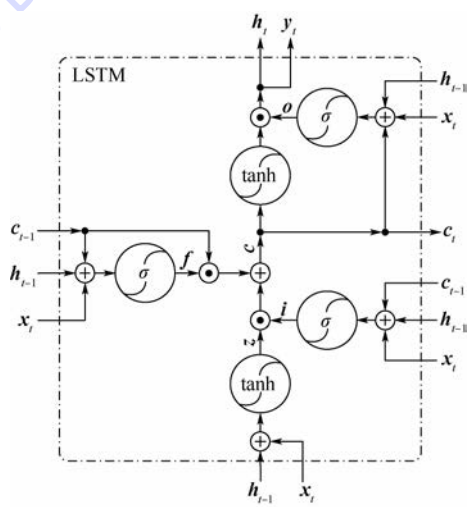


图 2　LSTM 隐藏层细胞结构

Fig. 2　LSTM cell structure in hidden layer

北航学报 赠阅

输出门;$\boldsymbol{W}$ 和 $\boldsymbol{b}$ 分别为对应的权重系数矩阵和偏置项;σ 和 tanh 分别为 sigmoid 和双曲正切激活函数。LSTM 模型训练过程采用的是与经典的反向传播(Back Propagation, BP)算法原理类似的 BPTT 算法[17],大致可以分为 4 个步骤:①按照前向计算方法(式(3)~式(7))计算 LSTM 细胞的输出值;②反向计算每个 LSTM 细胞的误差项,包括按时间和网络层级 2 个反向传播方向;③根据相应的误差项,计算每个权重的梯度;④应用基于梯度的优化算法更新权重。

基于梯度的优化算法种类众多,比如随机梯度下降(Stochastic Gradient Descent, SGD)[18]、AdaGrad[19]、RMSProp[20]等算法。本文选用的是文献[21]提出的适应性动量估计(Adaptive moment estimation, Adam)算法。Adam 优化算法是一种有效的基于梯度的随机优化方法,该算法融合了 AdaGrad 和 RMSProp 算法的优势,能够对不同参数计算适应性学习率并且占用较少的存储资源。相比于其他随机优化方法,Adam 算法在实际应用中整体表现更优[21]。

此外,LSTM 模型演化出了很多变体,其中最成功的一种是文献[22]提到的门限循环单元(Gated Recurrent Unit, GRU)。GRU 模型是 LSTM 模型的简化版本,但是保留了 LSTM 模型的长期记忆能力,其主要变动是将 LSTM 细胞中的输入门、遗忘门、输出门替换为更新门和重置门,并将细胞状态和输出 2 个向量合二为一。在实际应用中,GRU 模型与 LSTM 模型具有很强的可比性[22]。

2 研究方法

根据系统级故障时间序列数据的特点,结合第 1 节介绍的相关理论和技术,本节给出基于 LSTM 模型的故障时间序列预测方法,以及对应的基于多层网格搜索的参数优选算法。

2.1 基于 LSTM 的故障时间序列预测

考虑到单变量故障时间序列有限样本点的数据特征,以及循环神经网络从简的设计原则,本文构建 LSTM 预测模型的整体框架如图 3 所示,包括输入层、隐藏层、输出层、网络训练以及网络预测 5 个功能模块。输入层负责对原始故障时间序列进行初步处理以满足网络输入要求,隐藏层采用图 2 表示的 LSTM 细胞搭建单层循环神经网络,输出层提供预测结果,网络训练采用第 1 节提到的 Adam 优化算法,网络预测采用迭代的方法逐点预测。

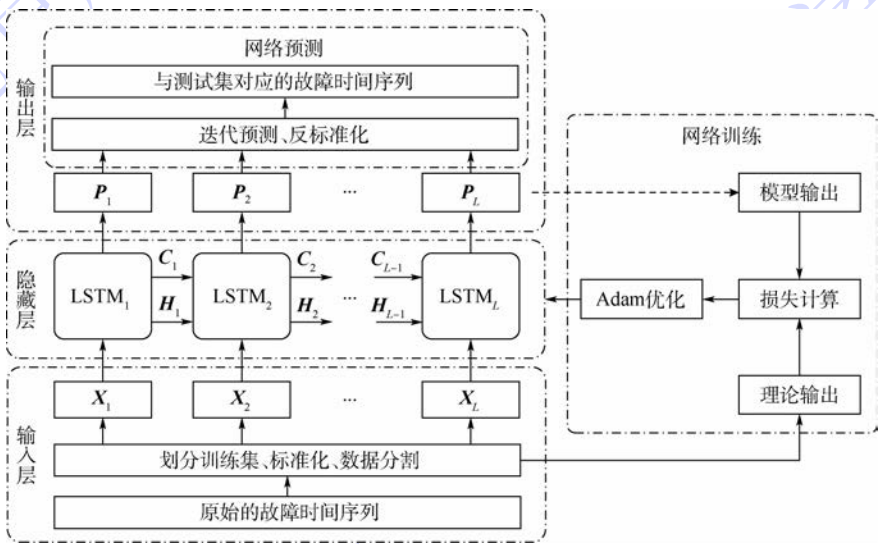


图 3 基于 LSTM 的故障时间序列预测框架

Fig. 3 LSTM based framework for failure time series prediction

2.1.1 网络训练

网络训练主要以隐藏层为研究对象。首先在输入层中,定义原始故障时间序列为 $\boldsymbol{F}_0=\{f_1,f_2,\cdots,f_n\}$,则划分的训练集和测试集可以表示为 $\boldsymbol{F}_{\mathrm{tr}}=\{f_1,f_2,\cdots,f_m\}$ 和 $\boldsymbol{F}_{\mathrm{te}}=\{f_{m+1},f_{m+2},\cdots,f_n\}$,满足约束条件 $m<n$ 和 $m,n\in\mathbf{N}$。然后对训练集中的元素 f_t 进行标准化,采用经典的 z-score 标准化公式(均值为 0,标准差为 1,表示为 zscore),标准化后的训练集可以表示为

$$\boldsymbol{F}'_{\mathrm{tr}}=\{f'_1,f'_2,\cdots,f'_m\} \tag{8}$$

$$f'_t=\left(f_t-\sum_{t=1}^{n}f_t/n\right)\sqrt{\sum_{t=1}^{n}\left(f_t-\sum_{t=1}^{n}f_t/n\right)^2/n}$$
$$1\leqslant t\leqslant m,t\in\mathbf{N} \tag{9}$$

为了适应隐藏层输入的特点,应用数据分割

的方法对 $\boldsymbol{F}'_{\mathrm{tr}}$进行处理,设定分割窗口长度取值为 L,则分割后的模型输入为

$$\boldsymbol{X}=\{\boldsymbol{X}_1,\boldsymbol{X}_2,\cdots,\boldsymbol{X}_L\} \tag{10}$$

$$\boldsymbol{X}_p=\{f'_p,f'_{p+1},\cdots,f'_{m-L+p-1}\}$$
$$1\leqslant p\leqslant L;p,L\in\mathbf{N} \tag{11}$$

对应的理论输出为

$$\boldsymbol{Y}=\{\boldsymbol{Y}_1,\boldsymbol{Y}_2,\cdots,\boldsymbol{Y}_L\} \tag{12}$$

$$\boldsymbol{Y}_p=\{f'_{p+1},f'_{p+2},\cdots,f'_{m-L+p}\} \tag{13}$$

接下来,将 $\boldsymbol{X}$ 输入隐藏层。从图 3 可以看出,隐藏层包含 L 个按前后时刻连接的同构 LSTM 细胞,$\boldsymbol{X}$ 经过隐藏层后的输出可以表示为

$$\boldsymbol{P}=\{\boldsymbol{P}_1,\boldsymbol{P}_2,\cdots,\boldsymbol{P}_L\} \tag{14}$$

$$\boldsymbol{P}_p=\mathrm{LSTM}_{\mathrm{forward}}(\boldsymbol{X}_p,\boldsymbol{C}_{p-1},\boldsymbol{H}_{p-1}) \tag{15}$$

式中:$\boldsymbol{C}_{p-1}$和 $\boldsymbol{H}_{p-1}$分别为前一个 LSTM 细胞的状态和输出;$\mathrm{LSTM}_{\mathrm{forward}}$表示第 1 节提到的 LSTM 细胞前向计算方法(式(3)~式(7))。设定细胞状态向量大小为 S_{state},则 $\boldsymbol{C}_{p-1}$和 $\boldsymbol{H}_{p-1}$2 个向量的大小均为 S_{state}。可以看出,隐藏层输出 $\boldsymbol{P}$、模型输入 $\boldsymbol{X}$ 和理论输出 $\boldsymbol{Y}$ 都是维度为 $(m-L,L)$ 的二维数组。选用均方误差作为误差计算公式,训练过程的损失函数可以定义为

$$\mathrm{loss}=\sum_{i=1}^{L(m-L)}(p_i-y_i)^2/[L(m-L)] \tag{16}$$

设定损失函数最小为优化目标,给定网络初始化的随机种子数 seed、学习率 η 以及训练步数 steps,应用 Adam 优化算法不断更新网络权重,进而得到最终的隐藏层网络。

2.1.2　网络预测

本小节应用训练好的 LSTM 网络(表示为 $\mathrm{LSTM}^*_{\mathrm{net}}$)进行预测。预测过程采用迭代的方法。首先,理论输出 $\boldsymbol{Y}$ 的最后一行数据为

$$\boldsymbol{Y}_f=\{f'_{m-L+1},f'_{m-L+2},\cdots,f'_m\} \tag{17}$$

将 $\boldsymbol{Y}_f$ 输入$\mathrm{LSTM}^*_{\mathrm{net}}$,输出结果可以表示为

$$\boldsymbol{P}_f=\mathrm{LSTM}^*_{\mathrm{net}}(\boldsymbol{Y}_f)=\{p_{m-L+2},p_{m-L+3},\cdots,p_{m+1}\} \tag{18}$$

则 $m+1$ 时刻的预测值为 p_{m+1}。然后,将 $\boldsymbol{Y}_f$ 的最后 $L-1$ 个数据点和 p_{m+1}合并为新的一行数据

$$\boldsymbol{Y}_{f+1}=\{f'_{m-L+2},f'_{m-L+3},\cdots,p_{m+1}\} \tag{19}$$

将 $\boldsymbol{Y}_{f+1}$输入$\mathrm{LSTM}^*_{\mathrm{net}}$,则 $m+2$ 时刻的预测值为 p_{m+2},依次类推,得到的预测序列为

$$\boldsymbol{P}_{\mathrm{o}}=\{p_{m+1},p_{m+2},\cdots,p_n\} \tag{20}$$

接下来,通过对 $\boldsymbol{P}_{\mathrm{o}}$ 进行 z-score 反标准化(表示为 de_zscore),得到最终的与测试集 $\boldsymbol{F}_{\mathrm{te}}$对应的预测序列为

$$\boldsymbol{P}_{\mathrm{te}}=\mathrm{de_zscore}(\boldsymbol{P}_{\mathrm{o}})=\{p^*_{m+1},p^*_{m+2},\cdots,p^*_n\} \tag{21}$$

$$p^*_k=p_k\sqrt{\sum_{t=1}^{n}\left(f_t-\sum_{t=1}^{n}f_t/n\right)^2/n}+\sum_{t=1}^{n}f_t/n$$
$$m+1\leqslant k\leqslant n,k\in\mathbf{N} \tag{22}$$

类似地,将 $\boldsymbol{X}$ 的每一行作为模型输入可以得到与训练集 $\boldsymbol{F}_{\mathrm{tr}}$对应的拟合序列 $\boldsymbol{P}_{\mathrm{tr}}$。最后,通过计算 $\boldsymbol{F}_{\mathrm{tr}}$和 $\boldsymbol{P}_{\mathrm{tr}}$,以及 $\boldsymbol{F}_{\mathrm{te}}$和 $\boldsymbol{P}_{\mathrm{te}}$的偏差定量地给出模型的拟合和预测精度。

总的来说,基于 LSTM 的故障时间序列模型训练和预测算法概括如下:

算法 1　训练并预测 LSTM 故障时间序列模型

输入:$\boldsymbol{F}_{\mathrm{o}},m,L,S_{\mathrm{state}},\mathrm{seed},\mathrm{steps},\eta$。

输出:与测试集对应的预测序列以及模型精度。

get $\boldsymbol{F}_{\mathrm{tr}},\boldsymbol{F}_{\mathrm{te}}$ from $\boldsymbol{F}_{\mathrm{o}}$ by m
$\boldsymbol{F}'_{\mathrm{tr}}=\mathrm{zscore}(\boldsymbol{F}_{\mathrm{tr}})$
get $\boldsymbol{X},\boldsymbol{Y}$ from $\boldsymbol{F}'_{\mathrm{tr}}$ by L
create $\mathrm{LSTM}_{\mathrm{cell}}$ by S_{state}
connect $\mathrm{LSTM}_{\mathrm{net}}$ by $\mathrm{LSTM}_{\mathrm{cell}}$ and L
initialize $\mathrm{LSTM}_{\mathrm{net}}$ by seed
for each step in 1:steps
  $\boldsymbol{P}=\mathrm{LSTM}_{\mathrm{forward}}(\boldsymbol{X})$
  $\mathrm{loss}=\sum_{i=1}^{L(m-L)}(p_i-y_i)^2/[L(m-L)]$
  update $\mathrm{LSTM}_{\mathrm{net}}$ by Adam with loss and η
get $\mathrm{LSTM}^*_{\mathrm{net}}$
for each j in 0:$(n-m-1)$
  $\boldsymbol{P}_{f+j}=\mathrm{LSTM}^*_{\mathrm{net}}(\boldsymbol{Y}_{f+j})$
  append $\boldsymbol{P}_{\mathrm{o}}$ with $\boldsymbol{P}_{f+j}[-1]$
$\boldsymbol{P}_{\mathrm{te}}=\mathrm{de_zscore}(\boldsymbol{P}_{\mathrm{o}})$
error measure $\varepsilon(\boldsymbol{P}_{\mathrm{te}},\boldsymbol{F}_{\mathrm{te}}),\varepsilon_{\mathrm{e}}(\boldsymbol{P}_{\mathrm{tr}},\boldsymbol{F}_{\mathrm{tr}})$

其中:训练过程涉及输入层、隐藏层、输出层、网络训练 4 个模块,预测过程主要涉及输出层模块;$\mathrm{LSTM}_{\mathrm{cell}}$表示 LSTM 隐藏层细胞(如图 2 所示);$\mathrm{LSTM}_{\mathrm{net}}$表示 LSTM 隐藏层网络(如图 3 所示);$\varepsilon_{\mathrm{e}}$为误差度量函数。

2.2　基于多层网格搜索的 LSTM 预测模型参数优选

在构建上述 LSTM 预测模型中,涉及到众多参数,其中以分割窗口长度 L、状态向量大小 S_{state} 和学习率 η 最为关键[16]。为了达到更好的预测效果,本文采用网格搜索的方法对这 3 个参数进行优选。相比与其他的超参数优化方法(比如遗传算法[23]、随机搜索算法[16]、粒子群算法[24],贝叶斯算法[25]等),网格搜索是一种简单实用、容易并行计算且计算耗时可控的优化方法[26],能够很好地满足故障时间序列预测的任务需求和实验要求。参数优选的依据是测试集全部测试点上的预

北航学报 赠阅

测精度最高,即预测误差 $\varepsilon(\boldsymbol{P}_{te},\boldsymbol{F}_{te})$ 最小,目标函数可以表示为

$$\min\ \varepsilon(\boldsymbol{P}_{te},\boldsymbol{F}_{te})$$
$$\text{s.t.}\begin{cases}2\leqslant L\leqslant L_{max}\leqslant m,\text{step}_L\mid L\\ 2\leqslant S_{state}\leqslant S_{max},\text{step}_{state}\mid S_{state}\\ \eta\in\{\eta_1,\eta_2,\cdots,\eta_r\},\text{step}_\eta\mid r\\ L,S_{state},r,\text{step}_L,\text{step}_{state},\text{step}_\eta\in\mathbf{N}\end{cases}\tag{23}$$

式中:step_L、step_{state} 和 step_η 分别为对应参数的网格搜索步长。L、S_{state} 和 η 这 3 个参数构成了一个三维搜索空间,可以通过多层网格搜索算法(算法 2)获取最优参数组合。搜索过程主要包括 3 层,从内到外分别对 S_{state}、L 和 η 进行网格搜索。首先,固定随机种子数 seed 和训练步数 steps,根据式(23)预设 3 个参数的取值范围(为了降低模型复杂度,分别将 L_{max} 和 S_{max} 控制在较小的取值);然后,分别遍历 3 个参数的取值范围,在最内层训练并预测 LSTM 故障时间序列模型(如算法 2 所示,表示为 $\text{LSTM}_{predict}$),保存对应的模型参数和模型精度;最后,对所有保存的结果按照预测精度由高到低排序,则最前面的参数组合即优选的模型参数。

算法 2 LSTM 预测模型参数优选

输入:$\boldsymbol{F}_o$, seed, steps, m, L, S_{state}, η, step_L, step_{state}, step_η。

输出:测试集上预测误差较低的参数组合。

```
predefine values of seed, steps
predefine value ranges of L, S_state, η
for each η in η_1, η_2, ⋯, η_r by step_η
    for each L in 2:L_max by step_L
        for each S_state in 2:S_max by step_state
            execute LSTM_predict(F_o, m, L, S_state, seed, steps, η)
            append results with L, S_state, η, ε(P_te, F_te), ε(P_tr, F_tr)
end (for); end (for); end (for)
return results ranked by ε(P_te, F_te)
```

3 实验验证

本节结合一个航空领域的工业案例,应用第 2 节提出的 LSTM 预测模型及其参数优选算法展开实验验证。

3.1 实验准备

首先介绍实验中所选用的故障数据集,与本文提出的 LSTM 模型相对比的其他时间序列预测模型,评价不同预测模型优劣的精度度量指标,以及实验运行平台和软硬件环境配置。

3.1.1 数据集

实验数据采用文献[3]提供的系统级故障时间序列数据集。该数据集包含两架正在运营的波音 737 飞机 18 年(1997—2014 年)的故障记录,通过整理后得到 2 个包含 216 个数据点的月度故障时间序列数据。故障时间序列如图 4 所示,分别对应 A 飞机和 B 飞机 2 个数据源,横坐标为年份,纵坐标为月度故障数。从图中可以看出,这 2 个系统级故障数据相比于文献[7]提供的组件级故障数据表现出更为复杂的曲线形态。此外,对于复杂系统来说,一定时期内发生的故障数可以视为系统的平均故障率[27],用公式表示为 $\lambda_a=\sum_{i=1}^{N}f_i\Big/\sum_{i=1}^{N}t_i=1/\text{MTBF}$,其中 f_i 和 t_i 分别为第 i 次使用周期中发生的故障数和工作时长,N 为指定时期内的工作次数,MTBF(Mean Time Between Failure)为平均故障间隔时间。因此,本文选择故障数这一重要的可靠性指标作为实验对象,并设定前 17 年的 204 个数据点作为训练集,第 18 年的 12 个数据点作为测试集。

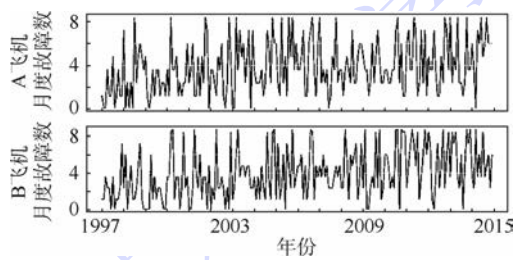


图 4 A、B 飞机的月度故障时间序列数据

Fig. 4 Monthly failure time series data for Aircraft A and Aircraft B

3.1.2 对比模型

除了第 1 节提到的 RNN 和 GRU 模型,本文将 LSTM 模型与以下 5 种时间序列预测模型进行实验对比。

1) Holt-Winters 模型

Holt-Winters 又称 3 次指数平滑,是一种能够处理含有趋势性和周期性成分的时间序列分析方法[28]。其思想是利用历史数据的不同特征成分(水平、趋势和季节)来递推当前数据。Holt-Winters 模型的重要参数是与特征成分对应的 3 个平滑系数,即 α、β 和 γ,取值均为 0 到 1 之间,且越靠近 1 则预测结果越依赖于近期观测值。此外,根据季节性成分在递推公式中的不同组织形式,Holt-Winters 模型又分为加法和乘法 2 种类型(分

别表示为 Holt-Winters[a] 和 Holt-Winters[m])。在实际使用中,这 2 种类型均需要计算不同特征成分的初始值,并根据均方误差最小来计算 3 个平滑系数。

2)自回归移动平均

ARIMA 是时间序列分析的经典理论和方法,其模型可以表示为 ARIMA(p,d,q),其中 p、d、q 分别为自回归项数、差分次数、移动平均项数[29]。在实际使用中,这 3 个参数可以通过观测自相关函数(Auto Correlation Function, ACF)和偏自相关函数(Partial Auto Correlation Function, PACF)确定,也可以通过计算 AIC(Akaike Information Criterion)或 BIC(Bayesian Information Criterion)值取其最小来确定。本文采用了文献[30]提出的自动化方法建立 ARIMA 模型。

3)奇异谱分析

SSA 是一种时域和频域相结合的非参数方法,可以用于处理非线性、非平稳以及包含噪声的时间序列,其核心思想是提取序列中的有效成分建模和预测[31]。SSA 包括分解和重构 2 个过程,其中分解过程又包括嵌入和奇异值分解(Singular Value Decomposition,SVD)2 个子过程,重构过程又包括分组和对角平均 2 个子过程。在实际使用中,需要确定的 2 个参数为嵌入子过程的窗口长度 L_{ssa}和分组子过程的分组类别 G_{ssa}。本文根据文献[32]提出的方法确定这 2 个参数的取值范围并取其最大值。此外,SSA 包含递归和向量 2 种预测方法(分别表示为 SSA[r] 和 SSA[v]),其中向量预测方法拥有更好的稳定性但是需要消耗更多的计算资源。

4)多元线性回归

多元线性回归(Multiple Linear Regression, MLR)是一种广泛应用于预测任务的多因素分析方法[33]。MLR 用于时间序列预测的模型可以表示为 $Y_t = a_0 + a_1 Y_{t-1} + a_2 Y_{t-2} + \cdots + a_k Y_{t-k} + e$,其中:$Y_t$ 为 t 时刻的预测值,$a_1, a_2, \cdots, a_k$ 为不同历史时刻数据 $Y_{t-1}, Y_{t-2}, \cdots, Y_{t-k}$ 对应的回归系数,a_0 和 e 分别为偏置项和误差项[34]。当连续的历史时刻数据被选为多因素变量时,k 也可以称为窗口长度(表示为 L_{mlr})。本文选定连续 2 年的月度故障数据点个数 24 作为 k 的取值。

5)支持向量回归

SVR 是一种可以用于时间序列预测的机器学习方法[35]。这种方法通过一个非线性核函数将多维输入映射到更高维度的特征空间后执行回归运算,进而得到与输出指标的非线性映射关系。本文选择常用的高斯径向基函数(Gaussian Radial Basis Function, RBF)作为非线性核函数,并采用文献[36]提出的启发式方法计算核函数参数 σ_{svr} 和误差边界 ε。此外,SVR 模型的惩罚因子 C 和窗口长度 L_{svr}分别设置为 3 和 24。

3.1.3　度量指标

本文从以下 2 个方面评价所有的预测模型:计算耗时和模型精度。对于计算耗时,本文统计了每个模型在构建过程中消耗的时间;对于模型精度,本文选择均方根误差(Root Mean Square Error, RMSE)作为度量标准。RMSE 的计算公式可以表示为 $\mathrm{RMSE} = \sqrt{\sum_{t=1}^{T} (y_t - f_t)^2 / T}$,其中,$f_t$ 和 y_t 分别为故障时间序列在 t 时刻的观测值和模型输出值,T 为数据点个数。本文通过对训练集和测试集分别计算 RMSE 值来定量地评价模型的拟合和预测精度。

3.1.4　平台和环境

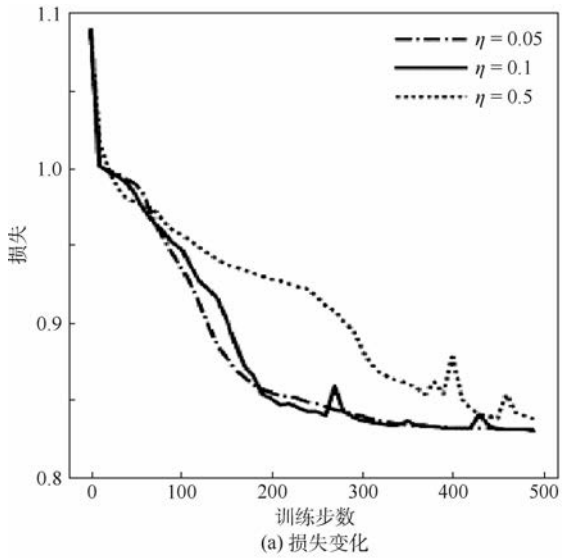
实验所使用计算机的配置如下:处理器为英特尔酷睿 Duo CPU i5-6500,CPU 频率为 3.20 GHz 和 3.19 GHz;内存为 4.00 GB;操作系统为 Windows 10(64 位);程序设计语言为 Python 3.5.2(64 位)和 R 3.3.3(64 位);集成开发环境为 PyCharm Community Edition 2016.3.2 和 Rstudio 0.99.903。程序设计过程中,RNN、LSTM 和 GRU 模型由 Python 的 tensorflow 0.12.0rc0[37] 程序包实现,Holt-Winters、ARIMA、SSA、MLR 和 SVR 模型由 R 的 stats 3.3.3[38]、forecast 8.0[39]、Rssa 0.14[40] 和 rminer 1.4.2[41] 程序包实现。

3.2　实验结果

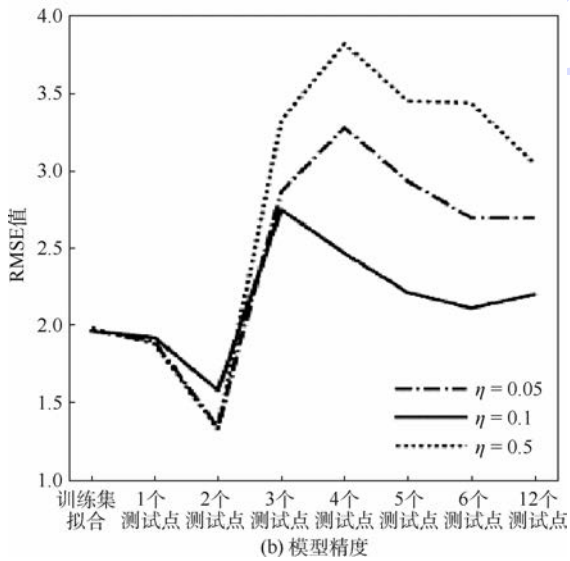
首先,以飞机 A 为例,应用 2.1 节提出的方法对标准化后的故障时间序列训练集建立 LSTM 预测模型。这里初步根据经验确定模型参数,分割窗口长度 L 取最小值 2,状态向量大小 S_{state}取半年的月度故障数据点个数 6,随机种子数seed = 1,训练步数 steps = 500。图 5 展示了相同参数下不同学习率($\eta = 0.05, 0.1, 0.5$)训练 LSTM 模型的损失变化和模型精度。可以看出:当 $\eta = 0.05$ 和 $\eta = 0.1$ 时,最终获得的损失较小(0.83 左右);3 个学习率对应的训练集拟合精度基本相同(RMSE 值在 2.0 附近);在 3、4、5、6 和 12 个测试点上,$\eta = 0.1$ 的预测精度最高(RMSE 值最低)。因此,本文选定 $\eta = 0.1$ 训练 LSTM 模型。

为了验证 LSTM 模型在不同类型循环神经网络中的优势,本文将 LSTM 模型的隐藏层细胞替换为 RNN 和 GRU 结构,并以相同参数进行实验。

北航学报 赠阅



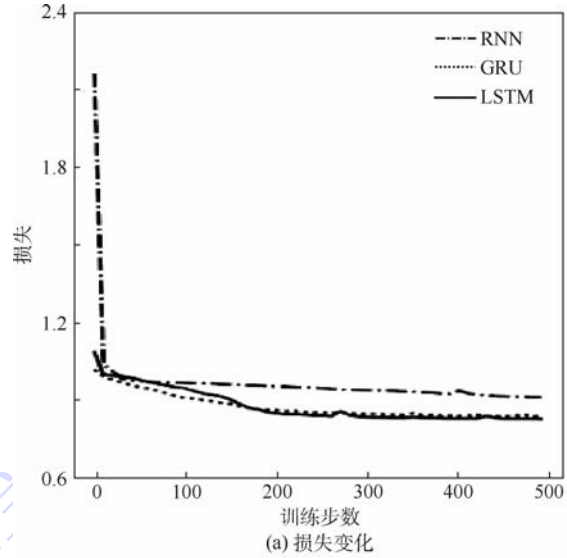
(a) 损失变化



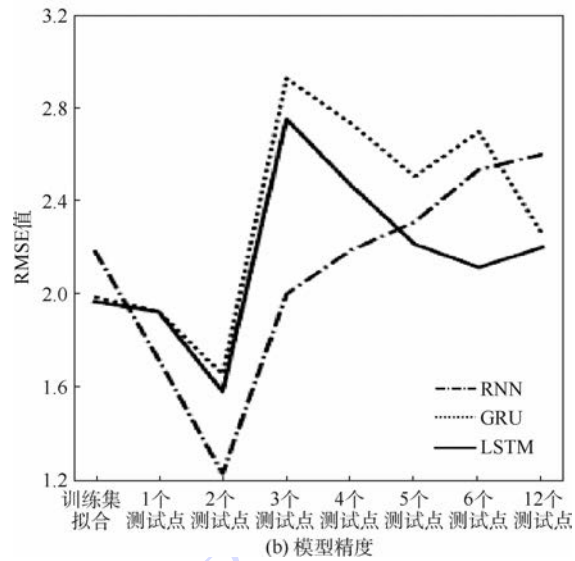
(b) 模型精度

图 5 不同学习率的损失变化和模型精度对比(A 飞机)

Fig. 5 Comparison of loss change and model accuracy with different learning rates (Aircraft A)



(a) 损失变化



(b) 模型精度

图 6 不同隐藏层细胞的损失变化和模型精度对比(学习率 $\eta = 0.1$,A 飞机)

Fig. 6 Comparison of loss change and model accuracy with different hidden layer cells (learning rate $\eta = 0.1$, Aircraft A)

实验结果如图 6 所示。从图中可以看出:LSTM 和 GRU 的损失变化相似并且都优于 RNN,对应的训练集拟合精度也较高;在 5、6 和 12 个测试点上,LSTM 的预测精度要高于 RNN 和 GRU。

为了进一步验证 LSTM 模型的应用效果,本文采用不同的时间序列预测模型进行对比,实验结果如表 1 所示。从表中可以看出:LSTM 模型的拟合精度要低于 SSA 和 SVR 模型,但是高于其他 6 种模型;LSTM 模型的整体预测精度较高,在 6 和 12 个测试点上的预测精度最高(RMSE 值分别达到了 2.109 和 2.196);LSTM 模型在该参数组合下的计算耗时要少于 ARIMA 模型,但是多于其他 8 种模型。

对于 B 飞机,实验流程与 A 飞机类似。这里根据经验调整模型参数,令状态向量大小 S_{state} 仍然取 6,分割窗口长度 L 取一年的月度故障数据点个数 12,随机种子数 seed = 100,训练步数 steps = 1 000。图 7 展示了相同参数下不同学习率($\eta = 0.01, 0.03, 0.05$)训练 LSTM 模型的损失变化和模型精度。可以看出:3 个学习率对应的训练集拟合精度基本相同(RMSE 值在 1.2 附近);$\eta = 0.05$ 时存在一定的过拟合现象,导致预测精度波动较大;在 6 和 12 个测试点上,$\eta = 0.03$ 的预测精度最高(RMSE 值最低)。因此,本文选定 $\eta = 0.03$ 训练 LSTM 模型。

接下来,本文替换隐藏层细胞为 RNN 和 GRU 结构并进行实验,结果如图 8 所示。可以看出,LSTM 模型无论是在损失变化还是整体模型

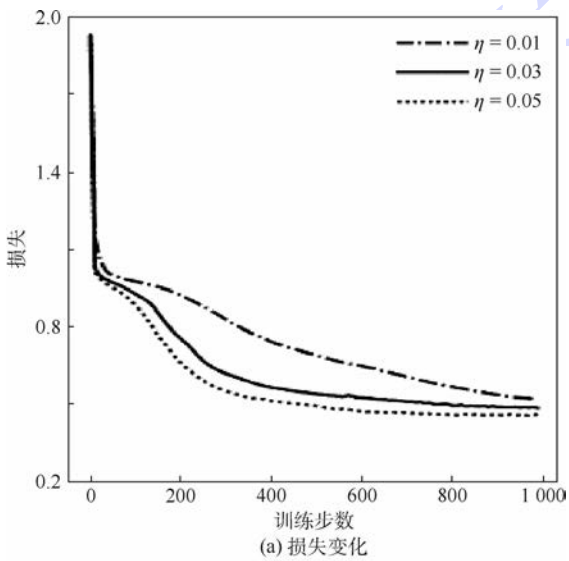
北航学报 赠阅

表 1　不同预测模型实验结果对比(A 飞机)

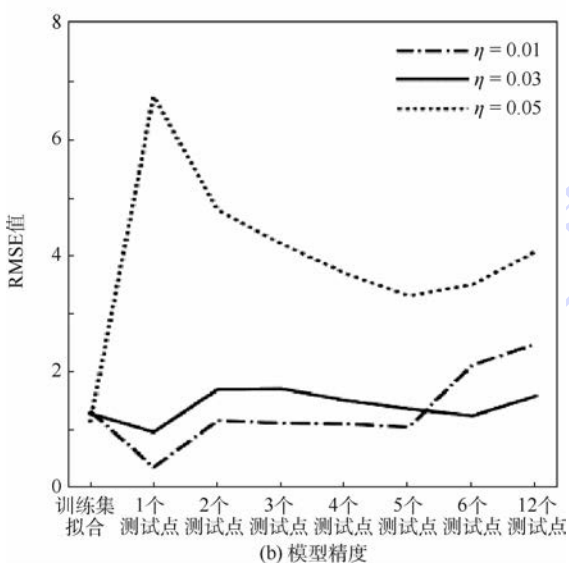
Table 1　Experimental results for different prediction models (Aircraft A)

模　型	模型参数	训练集拟合 RMSE 值	测试集预测 RMSE 值					耗时/s
			1 个测试点	2 个测试点	3 个测试点	6 个测试点	12 个测试点	
Holt-Winters[a]	$\alpha=0.044,\beta=0.073,\gamma=0.223$	2.617	0.088	0.069	2.278	2.882	2.595	0.02
Holt-Winters[m]	$\alpha=0,\beta=0,\gamma=0.6$	3.199	1.704	1.205	2.837	3.179	3.066	0.02
ARIMA	$p=2,d=1,q=2$	2.329	1.475	1.224	2.712	2.832	2.509	1.53
SSA[r]	$L_{ssa}=96,G_{ssa}=\text{list}(1;50)$	0.770	2.487	1.781	1.871	2.437	2.622	0.02
SSA[v]	$L_{ssa}=96,G_{ssa}=\text{list}(1;50)$	0.770	2.509	1.843	2.175	2.500	2.295	0.02
MLR	$L_{mlr}=24$	2.221	2.490	1.773	2.602	2.617	2.381	0.02
SVR	$L_{svr}=24,C=3,\varepsilon=0.259,\sigma_{svr}=0.023$	1.167	1.740	1.526	1.967	2.139	2.321	0.03
RNN	$L=2,S_{state}=6,\text{seed}=1,\text{steps}=500,\eta=0.1$	2.183	1.716	1.225	1.995	2.528	2.595	0.61
GRU	$L=2,S_{state}=6,\text{seed}=1,\text{steps}=500,\eta=0.1$	1.982	1.921	1.651	2.920	2.691	2.248	0.78
LSTM	$L=2,S_{state}=6,\text{seed}=1,\text{steps}=500,\eta=0.1$	1.962	1.919	1.577	2.745	2.109	2.196	0.81

注:最小 RMSE 值和最小耗时由下划线标记。



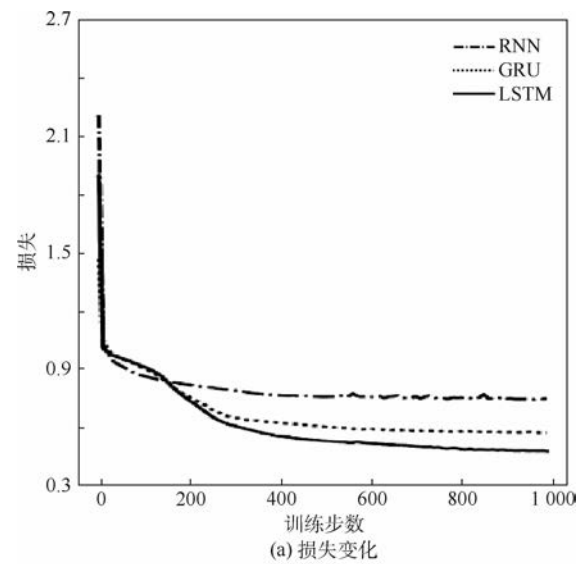
(a) 损失变化



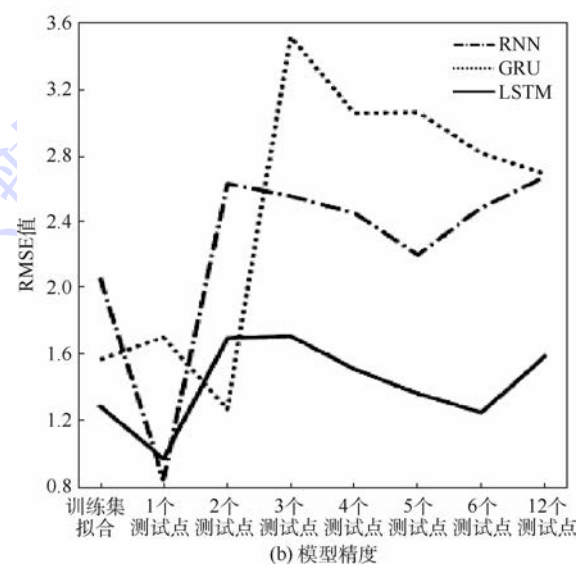
(b) 模型精度

图 7　不同学习率的损失变化和模型精度对比(B 飞机)

Fig. 7　Comparison of loss change and model accuracy with different learning rates (Aircraft B)



(a) 损失变化



(b) 模型精度

图 8　不同隐藏层细胞的损失变化和模型精度对比(学习率 $\eta=0.03$,B 飞机)

Fig. 8　Comparison of loss change and model accuracy with different hidden layer cells (learning rate $\eta=0.03$, Aircraft B)

北航学报 赠阅

精度上都优于 RNN 和 GRU 模型。图 9 展示了 LSTM 模型的拟合(图 9(a))和预测(图 9(b))结果,其中黑色线条分别代表原始故障时间序列数据划分的训练集和测试集,灰色线条分别代表 LSTM 模型在训练集和测试集上的拟合序列和预测序列。从图中可以看出,在 12 个测试点上,LSTM 模型很好地跟踪了真实的故障数据,达到了比较理想的预测效果。

最后,LSTM 模型和不同时间序列预测模型的实验对比结果如表 2 所示。从表中可以看出:LSTM模型的拟合精度要低于SSA模型,但是高于其他 7 种模型;LSTM 模型的整体预测精度较高,在 3、6、12 个测试点上的预测精度最高(RMSE 值分别达到了 1.703、1.237 和 1.580);RNN、GRU 和 LSTM 模型在该参数组合下的计算耗时要明显多于其他 7 种模型,并且以 LSTM 模型耗时最多。

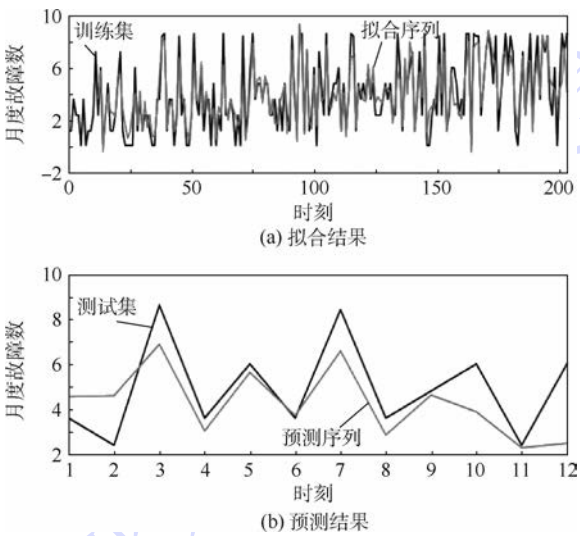


图 9 LSTM 模型的拟合和预测结果
(学习率 $\eta=0.03$,B 飞机)

Fig. 9 Fitting and forecasting results with LSTM model (learning rate $\eta=0.03$, Aircraft B)

3.3 参数优选

在 3.2 节的实验中,LSTM 模型的参数取值主要是通过经验来确定的。本节应用 2.2 节提到的多层网格搜索算法,对 LSTM 模型的 3 个关键参数进行优选。首先,固定非关键参数取值:随机种子数 seed = 1,训练步数 steps = 500;然后,设定 3 个参数的取值范围:分割窗口长度 $L\in\{2,3,\cdots,24\}$,状态向量大小 $S_{\text{state}}\in\{2,3,\cdots,24\}$,学习率 $\eta\in\{0.001,0.003,0.005,0.01,0.03,0.05\}$,其中 L 和 S_{state} 的搜索步长为 1;最后,设置目标函数为 12 个测试点上预测精度最高(RMSE 值最小),应用 2.2 节中的算法 2 进行网格搜索。

图 10 和图 11 分别展示了针对 2 个数据源(A 飞机和 B 飞机)建立 2.1 节提到的 LSTM 预测模型的参数搜索结果。在每个子图中,横坐标为分割窗口长度 L,纵坐标为状态向量大小 S_{state},Rm 为最小 RMSE 值;不同子图对应学习率 η 的不同取值;网格中的方块面积越大、颜色越深表示 RMSE 值越小。从图 10 和图 11 中可以看出,当 L 和 S_{state} 取值较小时更容易获得较高的预测精度。表 3 和表 4 分别列出了针对 2 个数据源的前 5 组最优参数组合以及对应的模型精度。表 1 和表 2 对比可知,优选后的 LSTM 模型精度明显提高。

表 2 不同预测模型实验结果对比(B 飞机)

Table 2 Experimental results for different prediction models (Aircraft B)

模型	模型参数	训练集拟合 RMSE 值	测试集预测 RMSE 值					耗时/s
			1 个测试点	2 个测试点	3 个测试点	6 个测试点	12 个测试点	
Holt-Winters[a]	$\alpha=0.011,\beta=0.210,\gamma=0.191$	2.907	1.749	3.605	3.353	2.609	2.474	0.02
Holt-Winters[m]	$\alpha=0,\beta=0,\gamma=0.438$	3.231	1.837	4.252	3.807	2.950	2.816	0.02
ARIMA	$p=4,d=1,q=1$	2.565	1.719	2.079	2.566	2.091	2.021	1.77
SSA[r]	$L_{\text{ssa}}=96,G_{\text{ssa}}=\text{list}(1:50)$	0.853	0.178	2.142	1.768	4.289	5.023	0.02
SSA[v]	$L_{\text{ssa}}=96,G_{\text{ssa}}=\text{list}(1:50)$	0.853	0.730	1.686	1.987	2.904	3.161	0.02
MLR	$L_{\text{mlr}}=24$	2.547	1.729	2.026	2.912	2.418	2.360	0.02
SVR	$L_{\text{svr}}=24,C=3,\varepsilon=0.252,\sigma_{\text{svr}}=0.023$	1.353	0.241	1.374	2.893	2.278	2.121	0.03
RNN	$L=12,S_{\text{state}}=6,\text{seed}=100,\text{steps}=1\,000,\eta=0.03$	2.058	0.828	2.630	2.556	2.484	2.671	3.13
GRU	$L=12,S_{\text{state}}=6,\text{seed}=100,\text{steps}=1\,000,\eta=0.03$	1.559	1.696	1.257	3.525	2.815	2.690	5.36
LSTM	$L=12,S_{\text{state}}=6,\text{seed}=100,\text{steps}=1\,000,\eta=0.03$	1.276	0.956	1.691	1.703	1.237	1.580	5.64

注:最小 RMSE 值和最小耗时由下划线标记。

北航学报 赠阅

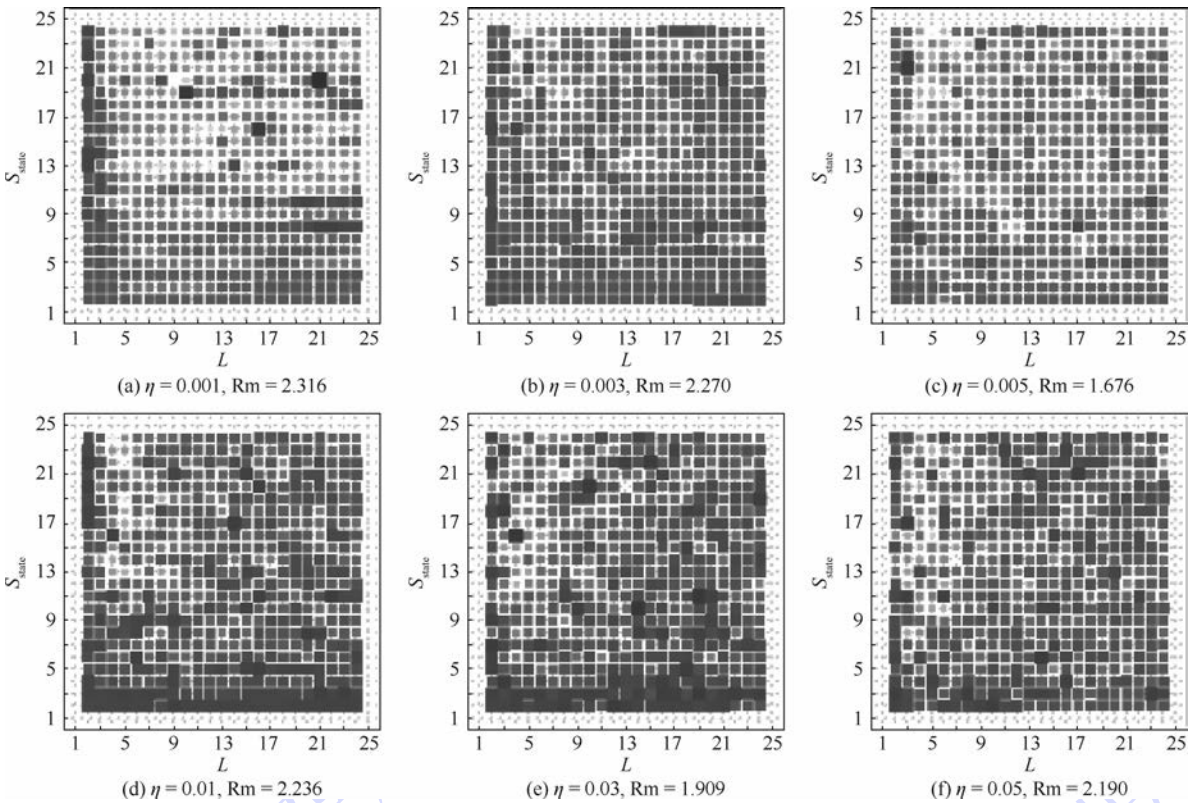


图 10　LSTM 模型 3 参数多层网格搜索结果(A 飞机)

Fig. 10　Multilayer grid search results for three parameters of LSTM model (Aircraft A)

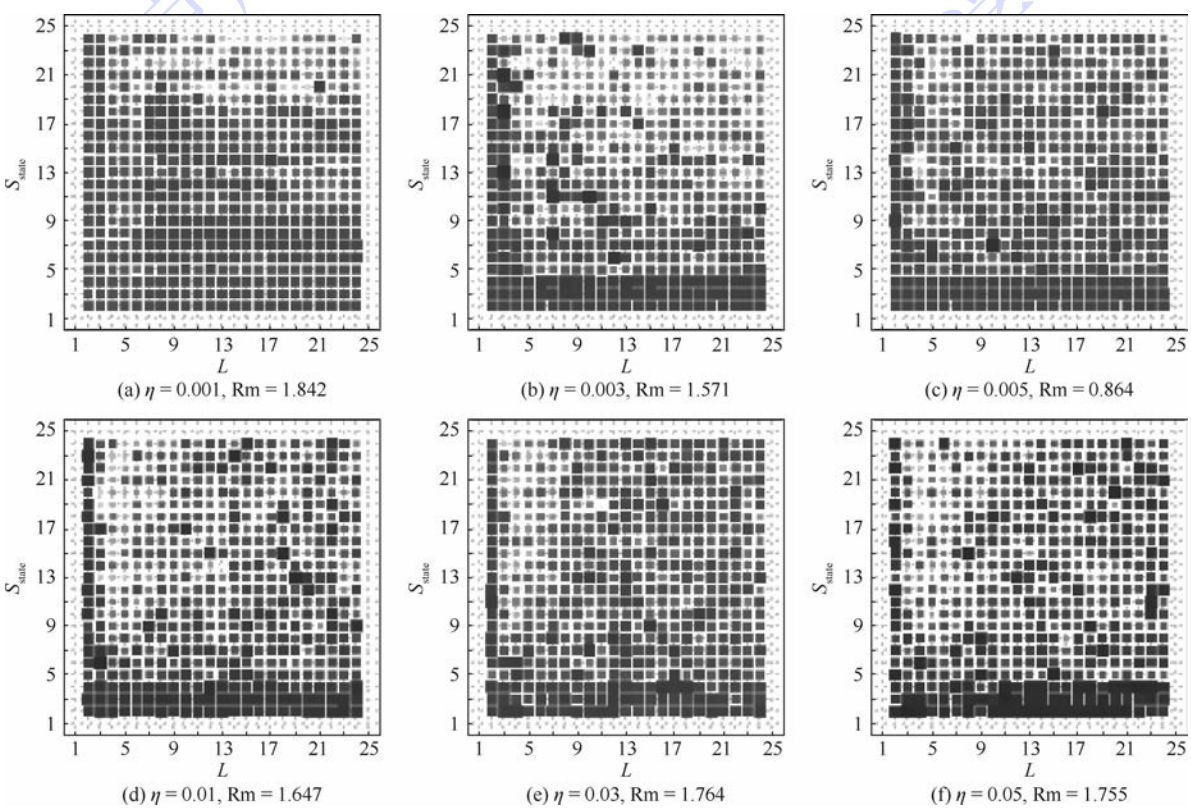


图 11　LSTM 模型 3 参数多层网格搜索结果(B 飞机)

Fig. 11　Multilayer grid search results for three parameters of LSTM model (Aircraft B)

北航学报 赠阅

表 3 LSTM 模型前 5 组最优参数组合以及对应的模型精度(A 飞机)

Table 3 The first five groups of optimal parameters andcorresponding model accuracy for LSTM model (Aircraft A)

排名	模型参数			训练集拟合 RMSE 值	测试集预测 RMSE 值					耗时/s
	L	S_{state}	η		1 个测试点	2 个测试点	3 个测试点	6 个测试点	12 个测试点	
1	3	21	0.005	1.261	0.694	0.921	1.261	1.154	1.676	1.56
2	14	10	0.03	0.321	2.539	1.834	2.506	2.390	1.909	3.63
3	17	8	0.005	1.311	1.923	1.824	2.137	2.004	2.041	3.94
4	19	11	0.03	0.289	0.054	0.762	1.290	2.058	2.061	4.84
5	4	16	0.03	0.584	3.860	2.759	2.395	1.991	2.081	1.66

表 4 LSTM 模型前 5 组最优参数组合以及对应的模型精度(B 飞机)

Table 4 The first five groups of optimal parameters and corresponding model accuracy for LSTM model (Aircraft B)

排名	模型参数			训练集拟合 RMSE 值	测试集预测 RMSE 值					耗时/s
	L	S_{state}	η		1 个测试点	2 个测试点	3 个测试点	6 个测试点	12 个测试点	
1	10	7	0.005	1.794	1.703	1.209	1.005	0.942	0.864	2.55
2	3	18	0.003	2.288	0.833	1.398	1.162	1.227	1.571	1.44
3	19	18	0.005	0.945	0.495	1.804	2.206	2.093	1.636	5.95
4	3	13	0.003	1.978	0.306	1.400	1.517	1.300	1.647	1.31
5	3	6	0.01	2.056	0.563	1.436	1.440	1.182	1.647	0.97

4 结 论

本文提出了基于 LSTM 循环神经网络的系统级故障时间序列预测方法,包括对 LSTM 模型的训练、预测以及参数优选等内容。实验验证表明:

1) 与典型的时间序列预测模型相比,LSTM 模型的拟合和预测性能整体更优。

2) LSTM 模型在训练过程中的损失变化和模型精度对学习率的取值较为敏感,过低或过高的学习率可能会导致欠拟合或过拟合问题,影响模型的预测性能。

3) 与其他类型的循环神经网络(RNN 和 GRU)相比,LSTM 模型的拟合和预测精度整体更高,但是训练过程的耗时也更多。

4) 基于多层网格搜索的参数优选算法效果显著,特别是对于第 2 个数据源(B 飞机),在测试点 12 上的预测精度(RMSE 值)达到了0.864,而文献[8]中提出的混合模型其最优 RMSE 值仅为 1.879。

总的来说,本文验证了 LSTM 模型在可靠性预测领域中的适用性,扩展了深度学习技术的应用范畴。基于目前的工作,后续可以展开进一步研究:比如扩展隐藏层层数,检验多隐藏层 LSTM 网络结构的应用效果;或者从众多 LSTM 模型参数入手,寻求更有效的参数优化方法。此外,本文是从历史数据出发,应用数据驱动的技术逆向建立预测模型。下一步,可以从可靠性相关的领域知识出发,应用提取出来的关键特征和要素正向研究可靠性预测方法。

致谢 感谢任健老师在实验数据方面提供的支持,感谢王森章博士在语言方面给予的帮助,感谢评阅论文的各位专家。

参考文献 (References)

[1] VICHARE N M, PECHT M G. Prognostics and health management of electronics[J]. IEEE Transactions on Components & Packaging Technologies, 2006, 29(1): 222-229.

[2] SAPANKEVYCH N I, SANKAR R. Time series prediction using support vector machines: A survey[J]. IEEE Computational Intelligence Magazine, 2009, 4(2): 24-38.

[3] 王鑫, 吴际, 刘超, 等. 奇异谱分析在故障时间序列分析中的应用[J]. 北京航空航天大学学报, 2016, 42(11): 2321-2331.
WANG X, WU J, LIU C, et al. Application of singular spectrum analysis for failure time series[J]. Journal of Beijing University of Aeronautics and Astronautics, 2016, 42(11): 2321-2331(in Chinese).

[4] 李瑞莹, 康锐. 基于 ARMA 模型的故障率预测方法研究[J]. 系统工程与电子技术, 2008, 30(8): 1588-1591.
LI R Y, KANG R. Research on failure rate forecasting method based on ARMA model[J]. Systems Engineering and Electronics, 2008, 30(8): 1588-1591(in Chinese).

[5] ROCCO S C M. Singular spectrum analysis and forecasting of failure time series[J]. Reliability Engineering & System Safety, 2013, 114(6): 126-136.

[6] MOURA M D C, ZIO E, LINS I D, et al. Failure and reliability prediction by support vector machines regression of time series data[J]. Reliability Engineering & System Safety, 2011, 96(11): 1527-1534.

[7] XU K, XIE M, TANG L C, et al. Application of neural networks

北航学报 赠阅

in forecasting engine systems reliability[J]. Applied Soft Computing,2003,2(4):255-268.

[8] WANG X,WU J,LIU C,et al. A hybrid model based on singular spectrum analysis and support vector machines regression for failure time series prediction[J]. Quality & Reliability Engineering International,2016,32(8):2717-2738.

[9] LECUN Y,BENGIO Y,HINTON G. Deep learning[J]. Nature,2015,521(7553):436-444.

[10] GRAVES A. Long short-term memory[M]. Berlin:Springer,2012:1735-1780.

[11] SRIVASTAVA N,MANSIMOV E,SALAKHUTDINOV R. Unsupervised learning of video representations using LSTMs[C]// Proceedings of the 32nd International Conference on Machine Learning. Lille:JMLR W&CP,2015:843-852.

[12] DONAHUE J,HENDRICKS L A,ROHRBACH M,et al. Long-term recurrent convolutional networks for visual recognition and description[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence,2015,39(4):677-691.

[13] VINYALS O,TOSHEV A,BENGIO S,et al. Show and tell:A neural image caption generator[C]// IEEE Conference on Computer Vision and Pattern Recognition. Piscataway,NJ:IEEE Press,2015:3156-3164.

[14] MA X,TAO Z,WANG Y,et al. Long short-term memory neural network for traffic speed prediction using remote microwave sensor data[J]. Transportation Research Part C Emerging Technologies,2015,54:187-197.

[15] HANSON J,YANG Y,PALIWAL K,et al. Improving protein disorder prediction by deep bidirectional long short-term memory recurrent neural networks[J]. Bioinformatics,2017,33(5):685.

[16] GREFF K,SRIVASTAVA R K,KOUTNIK J,et al. LSTM:A search space odyssey[J]. IEEE Transactions on Neural Networks & Learning Systems,2016,PP(99):1-11.

[17] GRAVES A,SCHMIDHUBER J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures[J]. Neural Networks,2005,18(5-6):602.

[18] AMARI S I. Backpropagation and stochastic gradient descent method[J]. Neurocomputing,1993,5(4-5):185-196.

[19] DUCHI J,HAZAN E,SINGER Y. Adaptive subgradient methods for online learning and stochastic optimization[J]. Journal of Machine Learning Research,2011,12(7):257-269.

[20] YEUNG S,RUSSAKOVSKY O,NING J,et al. Every moment counts:Dense detailed labeling of actions in complex videos[J]. International Journal of Computer Vision,2017(8):1-15.

[21] KINGMA D P,BA J. Adam:A method for stochastic optimization[C]// ICLR 2015,2015:1-15.

[22] CHUNG J,GULCEHRE C,CHO K H,et al. Empirical evaluation of gated recurrent neural networks on sequence modeling[C]// NIPS 2014 Deep Learning and Representation Learning Workshop,2014:1-9.

[23] CHEN P W,WANG J Y,LEE H M. Model selection of SVMs using GA approach[C]// IEEE International Joint Conference on Neural Networks. Piscataway,NJ:IEEE Press,2004:2035-2040.

[24] BRATTON D,KENNEDY J. Defining a standard for particle swarm optimization[C]// IEEE Swarm Intelligence Symposium. Piscataway,NJ:IEEE Press,2007:120-127.

[25] SNOEK J,LAROCHELLE H,ADAMS R P. Practical Bayesian optimization of machine learning algorithms[C]// International Conference on Neural Information Processing Systems. Lake Tahoe:Curran Associates Inc.,2012:2951-2959.

[26] HSU C W,CHANG C C,LIN C J. A practical guide to support vector classification[EB/OL]. (2016-05-19)[2017-03-20]. https://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf.

[27] SU C,JIN Q,FU Y. Correlation analysis for wind speed and failure rate of wind turbines using time series approach[J]. Journal of Renewable & Sustainable Energy,2012,4(3):687-700.

[28] CHATFIELD C. The Holt-Winters forecasting procedure[J]. Journal of the Royal Statistical Society,1978,27(3):264-279.

[29] BARTHOLOMEW D J. Time series analysis forecasting and control[J]. Journal of the Operational Research Society,1971,22(2):199-201.

[30] HYNDMAN R J,KHANDAKAR Y. Automatic time series forecasting:The forecast package for R[J]. Journal of Statistical Software,2008,27(3):1-22.

[31] VAUTARD R,YIOU P,GHIL M. Singular-spectrum analysis:A toolkit for short,noisy chaotic signals[J]. Physica D-Nonlinear Phenomena,1992,58(1-4):95-126.

[32] GOLYANDINA N,KOROBEYNIKOV A. Basic singular spectrum analysis and forecasting with R[J]. Computational Statistics & Data Analysis,2014,71(1):934-954.

[33] NIKOLOPOULOS K,GOODWIN P,PATELIS A,et al. Forecasting with cue information:A comparison of multiple regression with alternative forecasting approaches[J]. European Journal of Operational Research,2007,180(1):354-368.

[34] BIANCO V,MANCA O,NARDINI S. Electricity consumption forecasting in Italy using linear regression models[J]. Energy,2009,34(9):1413-1421.

[35] BRERETON R G,LLOYD G R. Support vector machines for classification and regression[J]. Analyst,2010,135(2):230-267.

[36] CHERKASSKY V,MA Y. Practical selection of SVM parameters and noise estimation for SVM regression[J]. Neural Networks,2004,17(1):113-126.

[37] Google. tensorflow rminer 1.4.2[EB/OL]. [2017-03-20]. https://www.tensorflow.org/versions/r0.12.

[38] R Core Team. The R project for statistical computing[EB/OL]. [2017-03-20]. https://www.r-project.org.

[39] HYNDMAN R. robjhyndman/forecast[EB/OL]. [2017-03-20]. https://github.com/robjhyndman/forecast.

[40] KOROBEYNIKOV A. asl/rssa[EB/OL]. [2017-03-20]. https://github.com/asl/rssa.

[41] PAULO C. rminer:Data mining classification and regression methods[EB/OL]. [2017-03-20]. https://cran.r-project.org/package=rminer.

作者简介:

王鑫　男,博士研究生。主要研究方向:数据驱动技术。

北航学报 赠阅

吴际 男,博士,副教授,硕士生导师。主要研究方向:模型驱动、软件可靠性分析。

刘超 男,博士,教授,博士生导师。主要研究方向:软件工程、软件测试。

Exploring LSTM based recurrent neural network for failure time series prediction

WANG Xin[1], WU Ji[1,*], LIU Chao[1], YANG Haiyan[1], DU Yanli[2], NIU Wensheng[1,3]

(1. School of Computer Science and Engineering, Beijing University of Aeronautics and Astronautics, Beijing 100083, China;
2. Fengtai Vocational Education Central School, Beijing 100076, China;
3. Aeronautical Computing Technique Research Institute, Aviation Industry Corporation of China, Xi'an 710068, China)

Abstract: Effectively forecasting the failure data in the usage stage is essential to reasonably make reliability plans and carry out reliability maintaining activities. Beginning with the historical failure data of complex system, a long short-term memory (LSTM) based recurrent neural network for failure time series prediction is presented, in which the design of network structure, the procedures and algorithms of network training and forecasting are involved. Furthermore, a multilayer grid search algorithm is proposed to optimize the parameters of LSTM prediction model. The experimental results are compared with various typical time series prediction models, and validate that the proposed LSTM prediction model and the corresponding parameter optimization algorithm have strong adaptiveness and higher accuracy in failure time series prediction.

Keywords: long short-term memory (LSTM) model; recurrent neural network; failure time series prediction; multilayer grid search; deep learning

北京航空航天大学学报

Received: 2017-05-08; **Accepted**: 2017-08-11; **Published online**: 2017-10-13 14:05

URL: kns.cnki.net/kcms/detail/11.2625.V.20171013.1405.003.html

Foundation items: China Civil Aviation Special Research Project (MJ-S-2013-10); Technology Foundation Program of the National Defense Technology Industry Ministry (JSZL2014601B008); National Natural Science Foundation of China (61602237)

* **Corresponding author**. E-mail: wuji@buaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2017.0296

基于波导缝隙天线的 TEM 室频率扩展方法

宋春江[1,2], 冯骁尧[3], 戴飞[3,*]

(1. 清华大学 工程物理系, 北京 100084; 2. 北京系统工程研究所, 北京 100101;
3. 北京航空航天大学 电子信息工程学院, 北京 100083)

摘 要: 为突破实际测试当中的局限,研究了横电磁波传输(TEM)室频率范围的扩展方法。运用电磁场与微波技术相关原理,通过分析 TEM 室壳体开缝对高次模的影响,利用壳体表面电流分布变化规律和波导缝隙天线原理重新解释了开缝对高次模的抑制作用,进而提出了一种 TEM 室壳体表面开缝的设计方法。通过电磁场数值仿真,验证了抑制高次模工程方法的有效性,并对其控制参数和约束条件的准确性进行了分析评估;通过加工实物并进行实测,进一步验证了新方法的效果。仿真与测试结果表明,开缝抑制高次模的工程方法能够在不减小测试空间与影响主模传输的前提下,将测试带宽扩展了 42.9%。

关 键 词: 横电磁波传输(TEM)室; 电磁兼容; 高次模; 工作带宽; 波导缝隙天线
中图分类号: TM937.1
文献标识码: A **文章编号**: 1001-5965(2018)04-0785-07

横电磁波传输(TEM)室是一种重要的电磁兼容测试设备,最初由 Crawford 于 1974 年提出[1]。TEM 室基于变形的同轴线,内部可传输 TEM 波。TEM 室由主传输段和过渡段两部分组成,被试品置于主传输段可测得其电磁发射特性。按照电磁场理论,为避免出现高次模,同轴线上限使用频率与其截面的最大尺寸成反比:截面尺寸越大,上限使用频率越低。因此,TEM 室的测试空间与频率范围是一对矛盾的指标,无法同时满足。

国内外学者在扩展 TEM 室的测试空间和频带范围方面开展了许多研究,进行了一些有益尝试。Arezoomand 等借助电磁场数值仿真工具 CST 和 HFSS,针对标准 TEM 室提出了 2 种改进内导体结构以实现良好匹配的方法[2]。戴飞、宋春江等提出了利用对称结构构建二元 TEM 室扩展测试空间的方法[3-4]。Wen 等研究了带状线 TEM 室结构及其工作频段的扩展方法[5]。

在不改变几何尺寸的前提下,扩展使用频段一直是 TEM 室的研究热点,其核心是抑制小室产生高次模。TEM 室前几阶高次模是 TE 模式,此模式严重影响了小室内部的场强分布,限制了小室的工作带宽[6]。Walters 和 Leat 提出了内导体开缝切断高次模表面电流以抑制非对称小室 TE_{111} 谐振模式的方法[7]。Deng 等提出了一系列扩展 TEM 室工作带宽的方法,包括壳体和隔膜开缝(切断高次模表面电流从而抑制高次模)、放置磁环或铁氧体元件(在高次模频率产生高阻抗抑制电流传输)以及敷设吸波材料(通过降低高次模品质因数抑制高次模)等[8-9]。

上述文献也指出一些使用上的约束和待改进之处,如壳体开缝虽然扩展工作带宽,但会对外产生辐射,影响小室的屏蔽性能,必须将小室置于内贴敷吸波材料的屏蔽箱中使用;在小室内壁敷设吸波材料压缩了有效的测试空间,会削弱主模的

收稿日期: 2017-05-10; **录用日期**: 2017-06-09; **网络出版时间**: 2017-08-03 10:57
网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20170803.1057.001.html
基金项目: 国家自然科学基金 (61571027, 61521091, 61427803)
* **通信作者**. E-mail: daphige@buaa.edu.cn

引用格式: *宋春江, 冯骁尧, 戴飞. 基于波导缝隙天线的 TEM 室频率扩展方法[J]. 北京航空航天大学学报, 2018, 44(4): 785-791. SONG C J, FENG X Y, DAI F. Frequency extension method of TEM cells based on slotted waveguide antenna[J]. Journal of Beijing University of Aeronautics and Astronautics, 2018, 44(4): 785-791 (in Chinese).*

北航学报 赠阅

场强等。虽然国内外学者通过实验或者数值仿真针对壳体开缝的效果进行了研究，但未给出开缝的设计方法。为此，本文通过分析壳体表面电流的分布变化，根据波导缝隙天线理论，进一步解释了开缝对高次模的抑制作用，并提出一种不同于文献[8-9]的开缝方案。

1 壳体表面开缝扩展频率范围原理

首先通过 HFSS 电磁场数值仿真软件建立 TEM 室模型，考察端口 S 参数、小室内电磁场的模式分布以及壳体电流的特点。建立的 TEM 室模型结构尺寸如图 1 所示，总长度是 $l_1=300$ mm，主传输段的长宽高分别是 $l_2=150$ mm，$a=150$ mm 和 $b=100$ mm，内部隔板的宽度与厚度分别是 $w=114$ mm 与 $t=1$ mm，坐标系为空间直角坐标系，其输入输出端口 S 参数仿真结果如图 2 所示。在 $f=1.68$ GHz 之前的低频段，小室的 S_{11} 保持在 −15 dB以下，说明输入端口能量的反射很小，阻抗匹配情况良好；S_{21} 基本保持在 −0.5 dB 以上，说明输入输出端口之间能量的传输情况良好，损耗非常微小。也就是说，小室的工作带宽是 1.68 GHz。$f=1.68$ GHz 处，TEM 室的 S_{11} 接近 0，而 S_{21} 很小，大部分能量都被反射回去，此处即为小室第一阶高次模产生的频点，由电场和磁场的仿真结果也验证了第一阶高次模是 TE 模式。

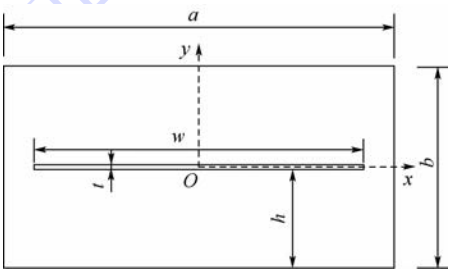


图 1 TEM 室结构尺寸

Fig. 1 Structure size of TEM cell

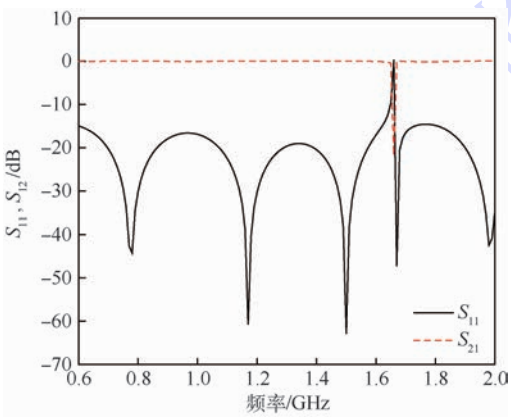


图 2 TEM 室 S 参数仿真结果

Fig. 2 Simulation results of S parameter of TEM cell

TEM 室主模传输的是横电磁波，其电场和磁场方向均垂直于电磁波的传输方向（z 轴），而电流和磁场方向遵循右手螺旋定则，因此壳体表面电流沿 z 轴方向流动（如图 3 所示）[10]。TEM 室的第一阶高次模是 TE 模式，其电场方向仍然垂直于 z 轴；磁场含有垂直于 z 轴方向的分量，但主要分量沿着 z 轴的纵向（如图 4 所示）；壳体表面电流方向（如图 5 所示）与磁场方向垂直，其主要分量垂直于 z 轴[11]，符合麦克斯韦方程组。E、H、J 分别为电场强度、磁场强度和电流密度。

根据微波理论[11]，沿着表面电流方向开缝，不会破坏波导内部的传播模式；而垂直表面电流方向开缝，会破坏传播模式。扩展工作频带的首要目标是抑制第一阶高次模，而第一阶高次模与主模在壳体表面电流方向不一致（如图 3 和图 5 所示），因此在壳体表面沿 z 轴平行方向开缝，只会切断高次模的电流，不会影响 TEM 模电流。这种开缝方式破坏了高次模的边界条件，但对主模的影响可以忽略，能够对高次模起到很好的抑

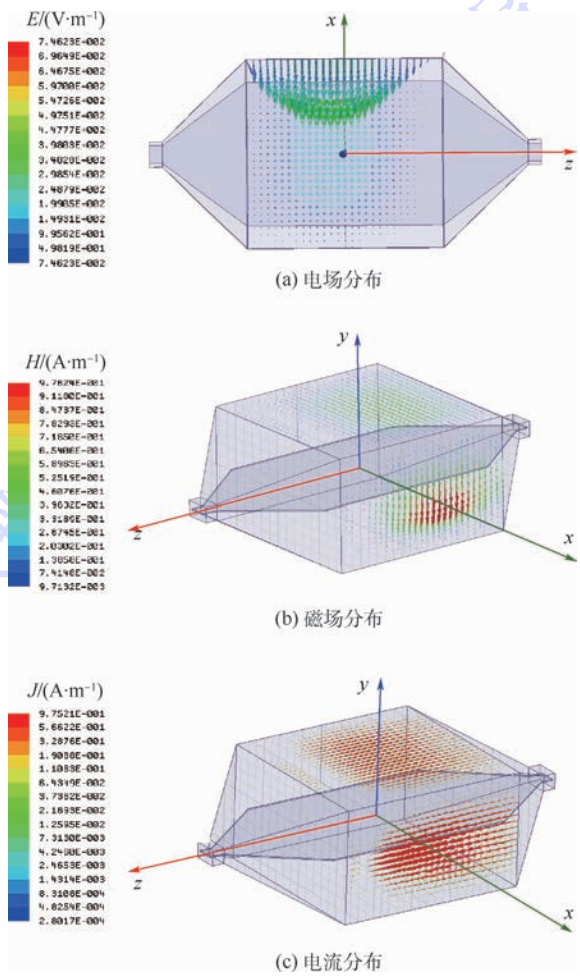
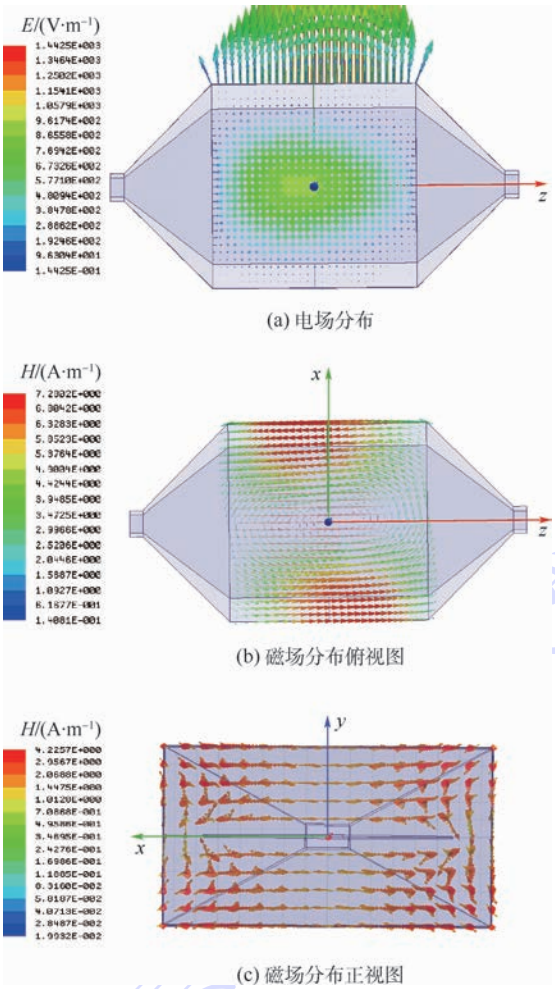


图 3 主模电磁场和电流分布

Fig. 3 Electromagnetic field and current distribution of main mode

北航学报 赠阅



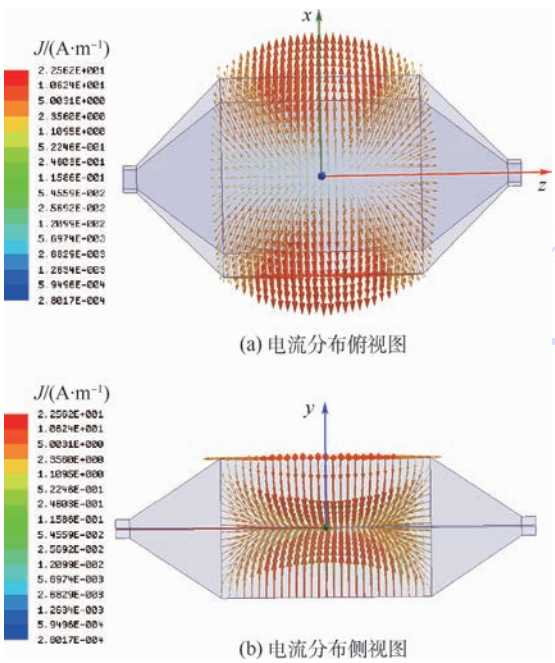
(a) 电场分布

(b) 磁场分布俯视图

(c) 磁场分布正视图

图 4　第一阶高次模电磁场分布

Fig. 4　Electromagnetic field distribution of first-order higher order mode



(a) 电流分布俯视图

(b) 电流分布侧视图

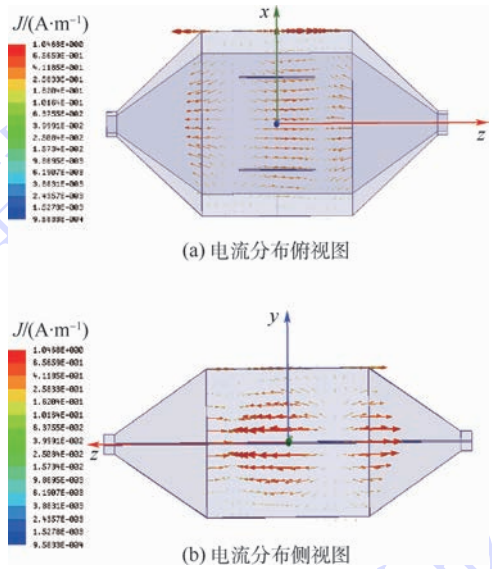
图 5　第一阶高次模电流分布

Fig. 5　Current distribution of first-order higher order mode

制作用。

开缝后表面电流变化情况如图 6 所示，横向表面电流被切断后明显减弱，进而抑制了高次模。

如图 6 所示，本文设计了与文献[8-9]不同的开缝方案，即仅在主传输段表面部分开缝。文献[8-9]的方案沿着传输方向在整个 TEM 室壳体表面开缝，而且开缝数量也明显多于本文方案。



(a) 电流分布俯视图

(b) 电流分布侧视图

图 6　TEM 室开缝后 1.68 GHz 电流分布

Fig. 6　1.68 GHz current distribution of slotted TEM cell

2　TEM 室壳体表面的开缝设计

文献[8-9]的开缝方案是以切断高次模的表面电流抑制高次模传输。根据微波理论，金属表面电流被切割将形成电磁辐射。因此，在 TEM 室主传输段的壳体上开缝将形成类似波导缝隙天线的结构。可根据波导缝隙天线理论[12-13]，优化开缝方式以提高辐射效率，从而在切断高次模电流的基础上，通过增强高次模的辐射，进一步抑制 TEM 室高次模的传输。

传统矩形波导缝隙天线内的主模是 TE_{10} 模式，其开缝方式如图 7 所示[14]。其中，方式 1 在波导宽边纵向开缝，称之为宽边并联缝隙，与本文

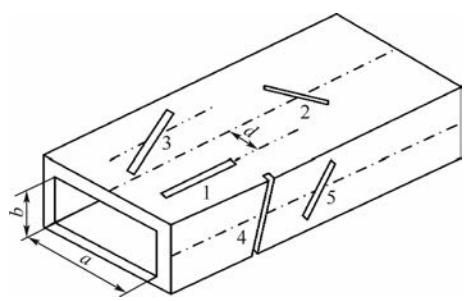


图 7　波导缝隙天线的形式[14]

Fig. 7　Types of slotted waveguide antenna[14]

北航学报 赠阅

的开缝方式类似;且当并联缝隙的长度取工作波长的一半时,辐射效率较高。因此,为使开缝抑制高次模的效果最好,应使 TEM 室缝隙天线的效果最强,开缝长度应为第一阶高次模截止波长的一半左右,缝隙宽度应远小于缝隙的长度,使其不超过缝隙长度的 1/10。

考虑到 TEM 室主传输段外壳整体的立体结构与矩形波导颇为相似,参考矩形波导的基本原理,将主传输段视为矩形波导与内导体隔板的组合体。参考黄志洵对小室内前几阶高次模变化规律的总结[15],小室内部场分布将是 TEM 模式与多个 TE 模式的混合,如图 4 所示的仿真结果。

从仿真结果可以看出,TEM 室第一阶高次模的结构与矩形波导 TE_{11} 模式具有很高的相似性,因此可以依据矩形波导高次模频率计算公式对小室主要高次模产生频率 f_c 进行估计:

$$f_c = \frac{c}{2}\sqrt{\left(\frac{m}{a}\right)^2 + \left(\frac{n}{b}\right)^2} \tag{1}$$

式中:c 为真空中的光速。

代入 $m = n = 1$ 计算后可得 $f_c = 1.8$ GHz,与仿真结果的 1.68 GHz 较为吻合。仿真结果略低的原因是,虽然 TEM 室的内导体隔板宽度一般略小于主传输段宽边的长度,难以形成 TE_{10} 模式,但是仍然可能存在 TE_{02} 与 TE_{20} 模式。TE_{11}、TE_{02}、TE_{20} 模式对场的贡献度受到内导体隔板宽度的影响,其规律如图 8 所示。

在 TEM 室壳体上下面(宽边)各开 2 条缝隙,每一边的 2 条缝隙沿各自面的中轴线对称分布,并沿电磁波传输方向(z 轴)位于小室的正中。根据 TE_{11} 模的截止频率为 1.8 GHz,选择 TEM 室壳体表面开缝长度为 83 mm,约为半个截止波长(如图 9 所示);取缝隙宽度为 1 mm;缝隙与壳体宽边中轴线的距离(即图 7 中的 d)取 41 mm,约为 1/4 个截止波长。

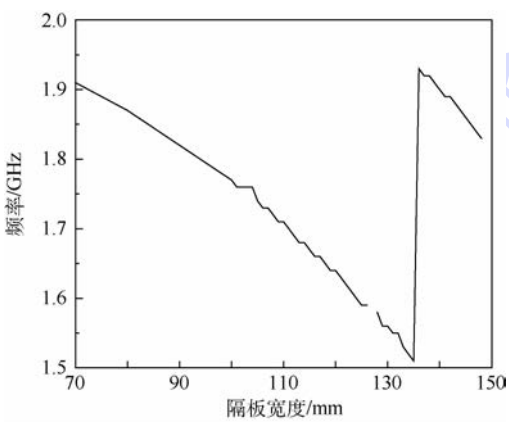


图 8 第一阶高次模频率随隔板宽度变化规律

Fig. 8 Variation of first-order higher order mode frequency with width of septum

对于 TE_{11} 模式,壳体上下左右表面中轴线上的电流沿中轴线(z 轴)方向传输,根据波导缝隙天线理论[11],在中轴线上纵向开缝不会切断电流,也不会形成辐射。为比较垂直和沿着高次模电流方向开缝对高次模的影响,分别建立了左右侧面(短边)中间纵向不开缝(见图 9)和开缝的模型。仿真结果显示左右侧面中轴线上是否开缝对高次模的影响不大,与理论相符。

仿真分析的结果如图 10 所示,本文的 TEM 室壳体表面开缝方案对第一阶高次模具有良好的抑制作用,使得小室的上限频率由 1.68 GHz 扩展到了约 2 GHz。

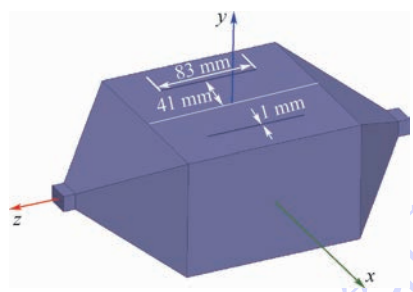


图 9 抑制高次模的开缝方式

Fig. 9 Slotting method to suppress higher order mode

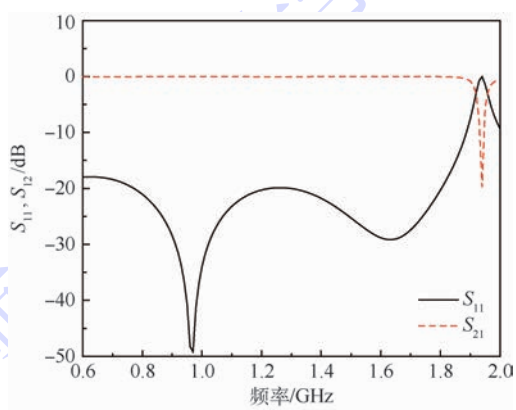


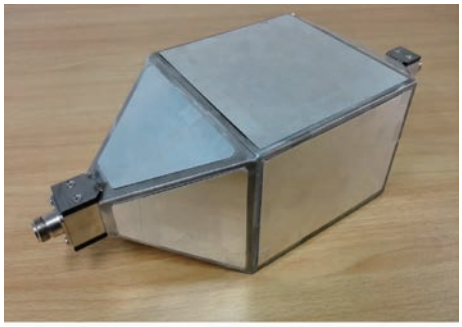
图 10 TEM 室开缝后 S 参数仿真结果

Fig. 10 Simulation results of S parameter of slotted TEM cell

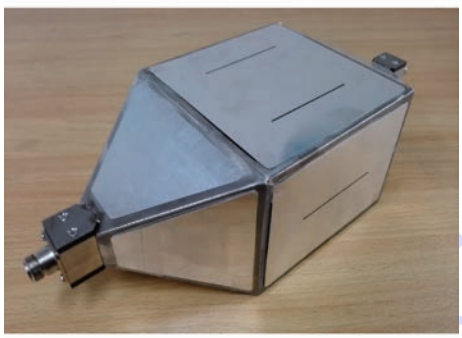
3 实物测试与验证

根据仿真模型设计加工的 TEM 室实物如图 11 所示,小室主体结构框架选用不锈钢材料,整体直接焊接成型,壳体表面选取铝合金材料制作成薄板,有利于拆卸以及高次模开缝研究,两头选取 N 型射频连接器与外界实现电连接/馈电。实物测试如图 12 所示。

北航学报 赠阅



(a) 未开缝TEM室



(b) 壳体开缝TEM室

图 11　TEM 室实物图

Fig. 11　Photos of TEM cell

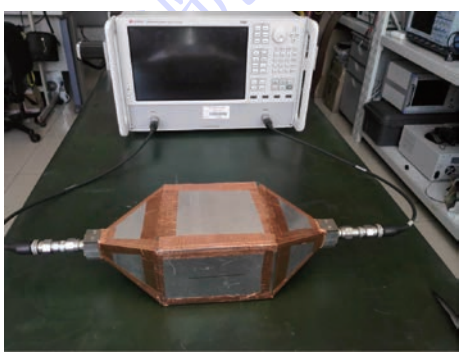


图 12　TEM 室实物测试图

Fig. 12　Testing photo of TEM cell

TEM 室输入与输出端口之间的 S 参数实物测试结果如图 13 所示,实测结果与仿真结果趋势一致,第一阶高次模的频率实测值与仿真结果吻合。

对比图 13 全开缝与上下表面开缝、左右侧面开缝与未开缝的测试结果表明,只有切割电流才能抑制高次模。采用本文提出的壳体表面开缝新方法,TEM 室的工作带宽实测达 2.4 GHz,优于仿真结果,比未开缝的 TEM 室提升 42.9%。

本文方法在扩展频段的效果上略优于文献[8-9]综合采用多种处理方法后的结果,但方案简单且开缝数量远少于文献[8-9]中的方案,机械结构更稳定,更易加工。

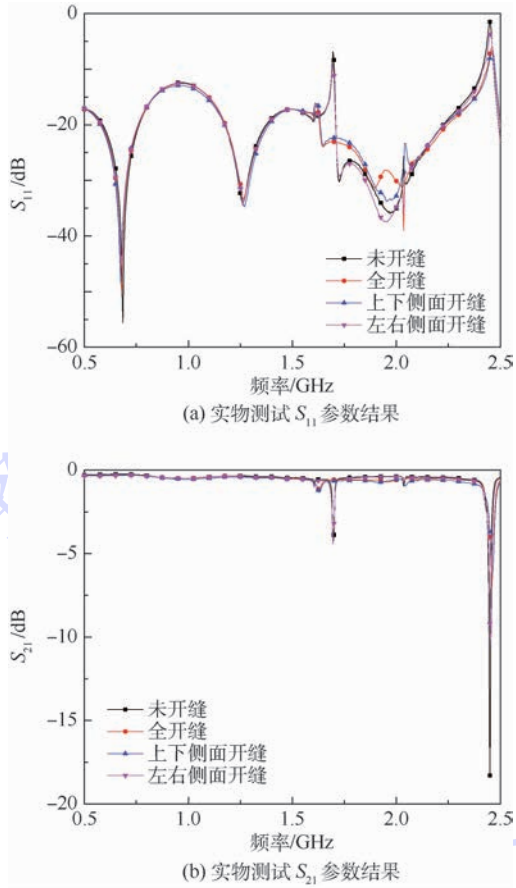


图 13　实物测试 S 参数结果

Fig. 13　S parameter results of real cell test

4　结　论

通过分析 TEM 室壳体表面电流分布的变化规律,根据波导缝隙天线理论,重新解释了开缝对高次模的抑制作用,进而提出一种新的壳体表面开缝抑制高次模、扩展 TEM 室测试频率范围的方法:

1）第一阶高次模以 TE_{11} 为主,可以 TEM 室主传输段的结构尺寸确定 TE_{11} 模式的截止频率。

2）在主传输段上下表面纵向开缝,缝隙长度选取为 TE_{11} 模式截止波长的一半左右,缝隙宽度应远小于缝隙长度,缝隙间距约半个截止波长。

3）应避免在上下左右表面的中间开缝。

采用本文方法,仅在主传输段上下表面纵向开 4 条缝隙时,以 TEM 室端口的 S_{11} 参数小于 −10 dB为基准,上限测试频率提高了 42.9%。同时,本文方法在结构和加工上更具优势。后续,将进一步研究缝隙的各种阻抗加载方式,以减小开缝造成的电磁泄漏。

致谢　感谢刘伟博士在 TEM 室实测过程中

北航学报 赠阅

的大力支持与帮助。

参考文献（References）

[1] CRAWFORD M L. Generation of standard EM fields using TEM transmission cells[J]. IEEE Transactions on Electromagnetic Compatibility, 1974, 16(4): 189-195.

[2] AREZOOMAND M, MEYBODI M K, NOORI N. Design of a TEM cell using both multi-step and piecewise linear tapering [C]// Proceedings of the 2016 8th International Symposium on Telecommunications (IST). Piscataway, NJ: IEEE Press, 2016: 571-574.

[3] DAI F, WANG M, SU D L. A design of new twin TEM cells [C]// IEEE 2005 International Symposium on Microwave, Antenna, Propagation and EMC Technologies for Wireless Communications Proceedings. Piscataway, NJ: IEEE Press, 2005: 10-13.

[4] SONG C J, FENG X Y. A new design and implementation of expanding testing space of a transverse electromagnetic cell[C]// The 9th International Conference on Microwave and Millimeter Wave Technology. Piscataway, NJ: IEEE Press, 2016: 967-969.

[5] WEN L, YALIN G, JIN L. Three new strip-line TEM cells in EMC test[C]// Proceedings of the 2016 IEEE International Conference on Electronic Information and Communication Technology (ICEICT). Piscataway, NJ: IEEE Press, 2016: 497-500.

[6] HILL D A. Bandwidth limitations of TEM cells due to resonances[J]. Journal of Microwave Power, 1983, 18(2): 181-195.

[7] WALTERS A J, LEAT C. Control of the TE_{111} resonance in an asymmetric TEM cell[J]. IEEE Transactions on Electromagnetic Compatibility, 2008, 50(2): 431-434.

[8] DENG S W, POMMERENKE D, HUBING T. Mode suppressed TEM cell design for high frequency IC measurements[C]// IEEE International Symposium on Electromagnetic Compatibility. Piscataway, NJ: IEEE Press, 2007: 1-6.

[9] DENG S W, POMMERENKE D, HUBING T. An experimental investigation of higher order mode suppression in TEM cells [J]. IEEE Transactions on Electromagnetic Compatibility, 2008, 50(2): 416-419.

[10] 苏东林, 陈爱新, 谢树果, 等. 电磁场与电磁波[M]. 北京: 高等教育出版社, 2008: 349-356.
SU D L, CHEN A X, XIE S G, et al. Electromagnetic field and electromagnetic wave[M]. Beijing: Higher Education Press, 2008: 349-356(in Chinese).

[11] 全绍辉. 微波技术基础[M]. 北京: 高等教育出版社, 2011: 155-157.
QUAN S H. The basis of microwave technology[M]. Beijing: Higher Education Press, 2011: 155-157(in Chinese).

[12] 卢万铮. 天线原理与技术[M]. 西安: 西安电子科技大学出版社, 2004: 221-224.
LU W Z. Antenna principle and technology[M]. Xi'an: Xidian University Press, 2004: 221-224(in Chinese).

[13] 李晓川. 波导缝隙天线极化特性分析与关键技术研究[D]. 哈尔滨: 哈尔滨工程大学, 2012: 5-11.
LI X C. Analysis and research of the polarization characteristic and the key technology for the waveguide slot antenna[D]. Harbin: Harbin Engineering University, 2012: 5-11 (in Chinese).

[14] 林昌禄. 天线工程手册[M]. 北京: 电子工业出版社, 2002: 272-273.
LIN C L. Antenna engineering handbook[M]. Beijing: Publishing House of Electronics Industry, 2002: 272-273(in Chinese).

[15] 黄志洵. 矩形结构传输线理论进展[J]. 光纤与电缆及其应用技术, 1989(6): 12-17.
HUANG Z X. Theoretical development of rectangular structure transmission line[J]. Optical Fiber & Electric Cable and Their Applications, 1989(6): 12-17(in Chinese).

作者简介：

宋春江 男，硕士研究生，助理研究员。主要研究方向：电磁兼容、雷达。

冯骁尧 男，硕士研究生。主要研究方向：电磁兼容。

戴飞 男，博士，副教授。主要研究方向：电磁兼容。

北京航空航天大学学报

北航学报 赠阅

Frequency extension method of TEM cells based on slotted waveguide antenna

SONG Chunjiang[1,2], FENG Xiaoyao[3], DAI Fei[3,*]

(1. Department of Engineering Physics, Tsinghua University, Beijing 100084, China;
2. Institute of Beijing System Engineering, Beijing 100101, China;
3. School of Electronic and Information Engineering, Beijing University of Aeronautics and Astronautics, Beijing 100083, China)

Abstract: In order to break through limitatiions in actual testing, the extension method of frequency range of a transverse electromagnetic (TEM) cell is studied. Using the theory of electromagnetic field and microwave technology, the influence of adding slots to the cell on the higher order modes is analyzed, and the suppressing effect of slots on the higher order modes is reinterpreted by summarizing the variation of the distribution of surface current. According to the principle of slotted waveguide antenna, a new method has been proposed to design the slotted surface of the TEM cell. Along with the numerical simulation of electromagnetic field, the paper validates the effectiveness of an engineering method to suppress higher order modes and evaluates the accuracy of the control parameters and the constraint conditions. The effect of the new method is further verified by processing a real cell and its measurement. The simulation and test results show that, without reducing the test space and affecting the main mode, the engineering method of slotted cell can extend the bandwidth by 42.9% by suppressing higher order modes.

Keywords: transverse electromagnetic (TEM) cell; electromagnetic compatibility; higher order mode; working bandwidth; slotted waveguide antenna

北京航空航天大学学报

Received: 2017-05-10; **Accepted**: 2017-06-09; **Published online**: 2017-08-03 10:57
URL: kns.cnki.net/kcms/detail/11.2625.V.20170803.1057.001.html
Foundation items: National Natural Science Foundation of China (61571027, 61521091, 61427803)
* **Corresponding author**. E-mail: daphige@buaa.edu.cn

北航学报 赠阅

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2017.0297

基于半监督集成学习的多核设计空间探索

李丹丹[1], 姚淑珍[1,*], 王颖[2], 王森章[3], 谭火彬[4]

(1. 北京航空航天大学 计算机学院, 北京 100083; 2. 中国科学院计算技术研究所 计算机体系结构国家重点实验室, 北京 100190; 3. 南京航空航天大学 计算机科学与技术学院, 南京 211106; 4. 北京航空航天大学 软件学院, 北京 100083)

摘 要: 随着处理器的系统结构日趋复杂,设计空间呈指数式增长,并且软件模拟技术极为费时,成为处理器设计的重要挑战。提出了一种结合集成学习和半监督学习技术的高效设计空间探索方法。具体而言,该方法包括2个阶段:使用均匀随机采样方法从处理器设计空间中选择一小组具有代表性的设计点,通过模拟获得性能响应,从而组成训练数据集;提出基于半监督学习的AdaBoost(SSLBoost)模型预测未模拟的样本配置的响应,从而搜索最优的处理器设计配置。实验结果表明,与现有的基于人工神经网络和支持向量机(SVM)的有监督预测模型相比,SSLBoost模型能够使用更少的模拟样本构建出不差于现有方法性能的预测模型;而当模拟样本数量相同时,SSLBoost模型的预测精度更高。

关 键 词: 设计空间探索; 半监督学习; 集成学习; AdaBoost; 微处理器; 预测模型
中图分类号: TP302
文献标识码: A **文章编号**: 1001-5965(2018)04-0792-10

全下文载

随着集成电路工艺的进步以及系统结构设计的复杂度不断提高,特别是随着多核技术的发展,处理器设计相关的参数变得越来越多,从而使得处理器的设计空间呈指数式增长;且处理器的工作负载通常由大量具有不同特性的应用程序组成。为满足处理器性能、功耗和设计成本等方面的要求和限制,系统结构设计师面临着更加严峻的挑战:从庞大的设计空间中寻找满足约束条件的最优参数组合,即是设计空间探索。为解决该问题,系统结构设计师通常在制造芯片前,利用软件模拟技术来模拟一部分具有代表性的工作负载来探索处理器的设计空间,进而评估处理器各种参数组合的性能。然而,周期精准的模拟器的低速问题(比真实处理器运算速度慢了3~5个数量级),使得对设计空间中每种可能的设计参数组合都进行模拟是非常耗时且不可行的。导致的结果是,设计师被迫减少评估的参数配置数量:通常基于设计师对某些参数相对重要性的经验以及功率预算等来确定需要模拟的参数组合。但此类方法有以下缺点:

1) 针对不同的应用程序,微体系结构设计参数对性能或功耗的影响程度是不同的,因此,设计师的经验可能是不准确的。

2) 缺乏统计的严谨性,得到的结论可能不正确。

3) 缺乏对各个参数的重要性及意义、参数之间的相互作用等问题的探讨。

为了解决以上问题,近年来的相关的研究主要关注如下3个方向:分析模型、快速的模拟技术和基于机器学习技术构建预测模型的方法。理论上,分析模型用来准确表征处理器性能和各种微体系结构参数之间的关系,能够消除对耗时模拟

收稿日期: 2017-05-11; **录用日期**: 2017-06-16; **网络出版时间**: 2017-08-31 14:35
网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20170831.1435.007.html
基金项目: 航空科学基金 (2013ZC51023)
* **通信作者**. E-mail: szyao@buaa.edu.cn

引用格式: *李丹丹, 姚淑珍, 王颖, 等. 基于半监督集成学习的多核设计空间探索[J]. 北京航空航天大学学报, 2018, 44(4): 792-801. LI D D, YAO S Z, WANG Y, et al. Multicore design space exploration via semi-supervised ensemble learning[J]. Journal of Beijing University of Aeronautics and Astronautics, 2018, 44(4): 792-801 (in Chinese).*

北航学报 赠阅

的需要。然而,现有的处理器性能分析模型技术基于几个简化的假设,仅对少量的微结构参数建模[1-2]。并且,分析模型需要大量的专业领域知识,不具备通用性。因此,此类模型缺乏在处理器设计周期使用的准确性和灵活性。快速的模拟技术主要通过减少模拟指令来加速模拟,比较著名的有 SimPoint[3] 和 SMARTS[4],能够极大地加快单个模拟的速度。这些方法在一定程度上减少了设计空间探索中的模拟成本,但设计空间中的配置数量太过于庞大,仅仅通过减少单个配置模拟的时间非常有限。随着机器学习技术的日益发展以及其在数据挖掘、社交网络分析等领域[5-6]的广泛应用,近年来,统计机器学习算法也被引入到系统结构设计领域中[7-20]。首先通过对设计空间中的一小部分配置进行模拟,然后根据模拟结果组成训练数据集,并构建机器学习预测模型来预测设计空间中未模拟的配置性能或功耗响应,从而极大地减少设计空间探索中的模拟成本。这种方法取得了较好的实验结果,往往和快速模拟技术 SimPoint 结合起来进行设计空间探索。例如,Joseph 等提出使用线性模型建模处理器性能与微系统结构参数的关系[7]。然而,在实际应用中,处理器的响应结果与体系结构设计参数的关系并非只有简单的线性关系。Joseph 等又提出基于径向基函数(RBF)的非线性回归模型预测处理器性能[8]。Lee 等在线性模型的基础上,提出了采用样条函数来刻画两者之间的非线性关系[9-10]。Ïpek 等提出了基于人工神经网络(Artificial Neural Network, ANN)的预测模型[11]。郭崎等提出基于模型树的预测模型[12]。庞九凤等提出基于支持向量机(SVM)的预测模型[13]。Palermo 等提出结合实验设计和响应面模型技术来识别满足系统级约束的 Pareto 解集合[14]。2015 年,Palermo 等又提出基于监督的高水平分析的谱感知 Pareto 迭代细化方法,以寻找设计空间中满足多约束的解集合[15]。针对单一预测模型的预测结果抖动的问题,郭崎等于 2015 年提出一种结合神经网络模型、SVM 和模型树 3 种预测模型的元模型,从而提高了预测模型的鲁棒性[16]。2016 年,考虑到随机采样训练样本的缺点,Li 等提出基于主动学习采样的设计空间探索方法[17]。针对通用处理器的设计而言,为了避免对每个不同的程序都构建一个预测模型,Khan、Dubach 和 Li 等分别提出交叉应用的预测模型[18-20]。由于模拟太过耗时且设计空间过于庞大,对于设计空间探索来讲,关键是如何利用一个足够小的样本集来构建预测模型,准确学习参数和响应之间的关系。传统方法多数都是基于有监督学习的预测模型,通常需要模拟大量的配置来提高预测模型的精度。为了减少模拟成本(次数),一个直观的考虑是基于少量已标记的样本集,标记设计空间中未模拟(即没有标记的)的配置(样本),从而达到扩充训练数据集的目的,进而提升预测模型的精度。此外,集成学习技术能够通过结合多个基本预测模型,来提升模型的预测精度。因此,本文的目标是利用半监督学习和集成学习技术尽量减少模拟的次数并保证预测模型的精度。

本文首先提出采用集成学习算法 AdaBoost 提升人工神经网络算法的预测精度和鲁棒性。然后结合 AdaBoost 算法的特点,提出一种基于半监督学习的 AdaBoost(Semi-Supervised Learning based AdaBoost,SSLBoost)模型,利用未标记样本来扩展训练数据集,进一步提升预测模型的精度。最后采用 SSLBoost 模型对未模拟配置的性能进行预测,从而搜索到最佳设计参数组合。

1　基于半监督集成学习的预测模型

本文结合半监督学习和集成学习技术,以更少的模拟成本,来提升设计空间模型的精度。采用 AdaBoost 模型提升 ANN 的鲁棒性和预测精度,并在此基础上,提出一种半监督集成学习算法。图 1 给出了本文提出的设计空间探索框架,主要分为 2 个阶段:训练数据集采样阶段和预测模型阶段。在采样阶段,首先利用基于均匀随

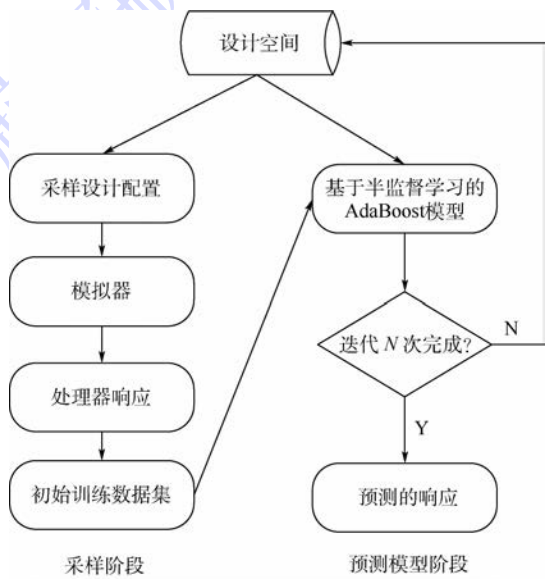


图 1　基于半监督集成学习的设计空间探索框架

Fig. 1　Design space exploration framework based on semi-supervised ensemble learning

北航学报 赠阅

机采样方法从处理器设计空间中选取一部分设计配置(设计参数组合,如核的数目、Cache 大小和发射宽度等的设计结构组合),利用模拟器进行模拟得到相应的性能响应(执行时间);然后将这些性能响应与设计参数共同组成初始训练数据集。在预测模型阶段,利用第 1 个阶段的训练数据集构建基于半监督学习的 AdaBoost 模型,迭代地对未标记样本进行标记,从而扩大训练数据集的规模,更加精确地预测设计空间中未模拟的设计结构的性能。

1.1 AdaBoost. RT 模型

考虑到预测模型的鲁棒性和准确性,本文采用集成算法 AdaBoost 中的一个变种 AdaBoost. RT 算法[21] 构建设计空间探索的预测模型。AdaBoost. RT 是 Shrestha 和 Solomatine 在 2006 年提出一种基于回归问题的提升算法[21],和其他提升算法一致,通过对多个弱学习器集成并提升为强的学习器。本文采用 ANN 作为弱学习器。

设 $L=\{(\boldsymbol{x}_1,y_1),(\boldsymbol{x}_2,y_2),\cdots,(\boldsymbol{x}_m,y_m)\}$ 为训练集,m 为模拟配置的数量,$\boldsymbol{x}_m$ 为由设计空间所有设计参数的取值组成的向量,例如,设计参数有频率、发射宽度、核数目等,$\boldsymbol{x}_1=(2,6,2,\cdots)$ 就表示 $\boldsymbol{x}_1$ 是一个频率大小为 2 GHz,发射宽度为 6,核的数目为 2 的一个设计配置。y_i 为模拟相应配置的处理器性能响应,如常用性能指标执行时间。$(\boldsymbol{x}_i,y_i)$ 即配置与模拟的性能组成的向量,在机器学习领域被称为一个已标记样本,组成的已标记样本集合被称为训练数据集。AdaBoost. RT 模型迭代地训练多个弱预测器 $h_1,h_2,\cdots,h_T$,这些弱预测器的线性组合作为最终的预测模型 H。算法的初始,弱预测器都是由 m 个带相同权重 $D_1(i)=1/m$ 的样本训练得到的。ANN 弱预测器的性能是使用训练数据的输出估计值和真实值的相对误差来评价的。即 ANN 的错误率由一个预设的相对误差阈值 ϕ 来计算的,ϕ 用来区分预测是正确或错误的。若一个配置的相对误差(Absolute Relative Error, $E_t(i)$)大于阈值 ϕ,那么就认为该配置的预测响应是错误的,否则就是正确的。错误预测的数量可以用来计算误差率 ε_t,计算公式如下:

$$\varepsilon_t=\sum_{i:E_t(i)>\phi}D_t(i)E_t(i)=\left|\frac{h_t(\boldsymbol{x}_i)-y_i}{y_i}\right| \tag{1}$$

设置 $\beta_t=\varepsilon_t^n$,这里 n 为幂系数(例如线性、平方或立方)。计算更新样本的权重(D_t 为第 t 次迭代的权重),给予误差率高的样本更高的权重。AdaBoost 模型通过调整每个已标记样本的权重来得到多个弱预测器。在第 i 次迭代后,AdaBoost. RT 模型更新关于每个样本的分布 $D_{t+1}(i)$ 为

$$D_{t+1}(i)=\begin{cases}\dfrac{D_t(i)}{Z_t}\beta_t & E_t(i)\leqslant\phi\\[2ex] \dfrac{D_t(i)}{Z_t} & \text{其他}\end{cases} \tag{2}$$

式中:Z_t 为一个标准化因子。在 T 次迭代以后,如式(3)所示,得到最终的回归模型 H,为每个输入配置 $\boldsymbol{x}$ 分配一个响应值 $H(\boldsymbol{x})$,即预测模型对未模拟配置的性能预测值。

$$H(\boldsymbol{x})=\sum_{t=1}^{T}\left(\left(\ln\frac{1}{\beta_t}\right)h_t(\boldsymbol{x})\Big/\sum_{t=1}^{T}\left(\ln\frac{1}{\beta_t}\right)\right) \tag{3}$$

1.2 基于半监督学习的 AdaBoost 模型

基于有监督学习的预测模型往往需要模拟大量的配置,才能构建出比较精确的预测模型,模拟成本较高。为了减少模拟次数,一个直观的考虑是利用已有的、大量未模拟的设计配置(也称为未标记样本)来提高预测模型的精度。即采用半监督学习技术,根据少量已经标记的样本为基础,选取置信度较高的未标记样本进行“伪”标记,来增加训练数据集。

因此,本文提出了 SSLBoost 模型。其基本思想是结合半监督学习和集成学习技术,一方面基于已标记训练样本,利用多个学习器集成地提升为更强的学习器,另一方面,通过自训练(self-training)的半监督学习技术利用未标记样本增加训练数据集的规模,进一步提升模型的预测精度。即将 AdaBoost 算法引入到半监督学习的过程当中,通过多次迭代训练来提高预测模型的精度。

图 2 给出了该算法流程图,首先利用已标记样本集 L 训练 AdaBoost 模型;然后从未标记样本集 U 中随机选择一部分组成未标记样本池 P,利用 AdaBoost 对未标记样本池 P 中的样本进行预测,根据 AdaBoost 中基本学习器(ANN)对样本的预测结果,对样本进行置信度评估,选择出置

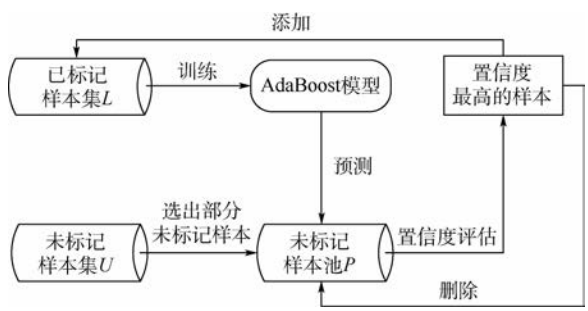


图 2 基于半监督学习的 AdaBoost 模型流程图

Fig. 2 Flowchart of AdaBoost model based on semi-supervised learning

信度最高的样本;利用预测结果对其标记(本文称为“伪”标记),添加到训练数据集中,同时在未标记样本池中删除该样本。重复迭代该过程,利用未标记样本来增加训练样本的数量,从而提升预测模型的精度。

在半监督学习技术中,一个关键的问题就是样本的置信度评估,其直接关系到算法是否选择出有效的未标记样本,影响算法的预测精度[22]。相比半监督分类学习,半监督回归学习中的样本置信度评估更难。分类问题中可以通过比较未标记样本属于不同类别的概率来评估,但回归问题的类别标签是连续的实值,很难找到这样的估计概率。而本文的设计空间探索模型对性能的预测为回归问题。

本文通过 2 个指标对未标记样本进行置信度评估。①一致性(consistency)指标:AdaBoost 模型中基本学习器预测未标记样本分歧越小(即一致性越高)的样本置信度越高;②误差下降指标:将 AdaBoost 模型预测结果对未标记样本进行“伪”标记,添加到训练样本集中,更新 AdaBoost 模型,对已标记样本集进行预测,相比没有添加该“伪”标记样本时,使对已标记样本集的预测误差减小的样本,置信度越高。也就是说,如果多个基本学习器对一个未模拟的设计配置预测得出比较一致、分歧很小的性能结果,则可以认为该配置的预测结果是相对稳定且准确的。若同时将该配置加入训练数据集中,使得预测模型对已模拟设计配置的性能预测精度更高,那么该配置就是置信度较高的样本。

本文采用变异系数(coefficient of variation)对未标记样本的一致性进行评估。变异系数为标准差与平均数的比值称,记为 C,反应了数据在单位均值上的离散程度。变异系数值越低,代表设计配置的预测结果分歧越小,一致性越高。

算法 1 给出了 SSLBoost 算法的伪代码。该算法的目标是迭代地依次扩大已标记的数据集,提升预测模型的精度。在训练过程的每一次迭代,每个学习器预测未标记样本池中的样本,选择出置信度高的样本加入到训练数据集中。在算法开始,由初始训练集初始化 AdaBoost. RT 模型 H。设 $h_1,h_2,\cdots,h_T$ 为模型 H 的 T 个 ANN,因为 AdaBoost 中的每个 ANN 对未模拟的配置样本预测的结果不同,变异系数越小代表这 T 个 ANN 对该样本的预测一致性越高。第 i 个未标记配置 $\boldsymbol{x}_i$ 的 C 值由式(4)来计算:

$$\begin{cases} C(\boldsymbol{x}_i)=\dfrac{\sigma_i}{\mu_i} \\ \sigma_i=\sqrt{\dfrac{1}{T}\sum\limits_{j=1}^{T}(h_j(\boldsymbol{x}_i)-\mu_i)^2} \\ \mu_i=\dfrac{1}{T}\sum\limits_{j=1}^{T}h_j(\boldsymbol{x}_i) \end{cases} \tag{4}$$

式中:μ_i 为 AdaBoost 中的所有 ANN 对第 i 个未模拟设计配置 $\boldsymbol{x}_i$ 预测结果的平均值;σ_i 为对 $\boldsymbol{x}_i$ 的所有预测结果的标准差。对未标记样本池中的样本预测,并基于 C 值进行升序排序。从 C 值最小的样本开始,将模型 H 对 $\boldsymbol{x}_i$ 的预测结果对其进行“伪”标记,添加到训练数据集中,记为 L',更新预测模型为 H',预测原训练样本集 L 的误差 $e_{L'}$,相比没有加入该“伪”标记样本时构建的模型 H 预测 L 的误差 e_L。如果 $e_{L'}<e_L$,即该未标记样本使得模型对已标记样本的预测误差变小,则将该“伪”标记的样本加入到训练数据集中,否则,对 C 值排在后面的未标记样本进行“伪”标记,迭代重复以上操作,直到遇到一个能够使得模型对原标记样本预测误差变小的未标记样本,并将其加入到训练集中;利用更新后的训练集重新构建 AdaBoost 模型,从未标记设计空间中随机选一个填充未标记样本池 P,然后进入下一轮迭代。这里误差的计算采用平均绝对误差(Relative Mean Absolute Error,RMAE)为误差指标,定义 $\mathrm{RMAE}=\dfrac{1}{n}\sum\limits_{i=1}^{n}\left(\dfrac{y_i-y_i'}{y_i}\right)\times 100\%$,其中,$y_i$ 为模拟结果的处理器性能响应(执行时间),y_i' 为预测的性能响应值,n 为预测的样本数(算法中为已模拟的配置数目)。最后,输出 AdaBoost 模型对所有未模拟的设计配置的预测结果。

基于 AdaBoost 中所有 ANN 对同一个未标记样本池中的样本预测,对能够达成比较一致的预测结果的、且能够使得模型对已标记样本的预测误差下降的样本,基本上可以断定该样本的预测响应值是非常接近真实模拟响应值的,则可以省去不必要的模拟,将其添加到训练数据集,进而提升 AdaBoost 模型的预测精度。

算法 1　SSLBoost 算法

输入:已标记的配置集合 L,所有未模拟的配置集合 U。

输出:未模拟的配置 $\boldsymbol{x}$ 的预测性能值。

创建未标记的样本池 P,随机从 U 中选择 p 个未标记样本到 P。

利用 L 训练出 AdaBoost. RT 模型 H。

北航学报 赠阅

迭代 K 次:

1 H 预测 P 中每一个未标记的配置,对于每个未标记样本 $\boldsymbol{x}_i \in P$,获取预测结果 $h_j(\boldsymbol{x}_i)$。

2 计算未标记样本池 P 中每个未标记样本的变异系数 C,基于 C 值对 P 中样本升序排序

$$\begin{cases} C(\boldsymbol{x}_i) = \dfrac{\sigma_i}{\mu_i} \\ \sigma_i = \sqrt{\dfrac{1}{T}\sum_{j=1}^{T}(h_j(\boldsymbol{x}_i) - \mu_i)^2} \\ \mu_i = \dfrac{1}{T}\sum_{j=1}^{T} h_j(\boldsymbol{x}_i) \end{cases}$$

3 利用模型 H 预测 L,记录百分比误差 e_L。

4 for $n = 1 : p$

选择 C 值排名为 n 的样本 $\boldsymbol{x}_n$,将 H 对 $\boldsymbol{x}_n$ 的预测结果 $y_n = H(\boldsymbol{x}_n)$ 作为"伪"标记赋予该样本 $(\boldsymbol{x}_n, y_n)$,加入 L 形成新的集合 $L' \leftarrow L \cup \{(\boldsymbol{x}_n, y_n)\}$,更新模型为 H',预测 L,记录百分比误差 $e_{L'}$。

If $e_{L'} < e_L$

样本 $\boldsymbol{x}_n$ 是置信度最高的样本,记录 $\pi \leftarrow (\boldsymbol{x}_n, y_n)$

跳出循环;

5 //更新 L 和 P

$P \leftarrow P - \pi$; $L \leftarrow L \cup \pi$

6 利用新的集合 L 重新构建 Adaboost 模型 H。

7 从 U 中随机挑选出来的样本来补充未标记样本池 P,使其大小为 p。

8 输出预测结果 $H(\boldsymbol{x})$。

2 实验方法与环境

2.1 模拟器及基准测试程序

本文采用科研领域广泛使用的周期精准的模拟器 GEM5[23] 模拟应用程序在多核处理器多种配置之上的性能。由于多核处理器主要关注并行程序的执行性能,因此本文选取了 PARSEC(The Princeton Application Repository for Shared-memory Computers)基准测试集[24],一个多线程应用程序组成的测试程序集。该程序集代表了未来运行在片上多核系统中的共享内存应用程序的发展趋势。本文从中选取了 10 个程序来对设计空间进行评估,包括 blackscholes、bodytrack、canneal、dedup、facesim、ferret、fluidanimate、freqmine、streamcluster 和 vips 等。由于基准程序中需要模拟的动态指令过多,本文采用了快速模拟技术 SimPoint 来模拟 1 亿条有代表性的指令,来节省每次模拟的模拟时间。本文采用执行时间作为性能响应。

2.2 处理器设计空间

表 1 给出了本文需要探索的多核处理器设计空间,涉及了 10 个主要与存储系统密切相关的设计参数。这些设计参数的不同组合构成的设计空间包含了超过 419 万个(4 194 304 个)不同的设计配置。

本文采用均匀随机采样的方法对设计空间分别采集了 2 200 个设计进行模拟,从中随机选择 200 个样本作为训练数据集,剩余 2 000 个样本作为测试数据集。本文将构建的预测模型对测试数据集进行预测的误差作为对模型准确性的评估。

表 1 多核设计空间

Table 1 Multicore design space

设计参数	设计参数取值	数 量
Band Width/(GB·s^{-1})	8 ~ 64:8 +	8
Frequency/GHz	1 ~ 4.5: 0.5 +	8
Issue Width	1,2,4,8	4
Number of Cores	1,2,4,8	4
L2 Cache Size/MB	2,4,8,16	4
L2 Cache Block Size/B	16,32,64,128	4
L2 Cache Associativity	2,4,8,16	4
L2 Cache MSHR	32 ~ 256:2 *	4
L1 Dcache/KB	16,32,64,128	4
L1 ICache/KB	16,32,64,128	4

3 模型评估与比较

3.1 评估指标

为评估 SSLBoost 模型的性能,本文采用 1.2 节中提到的相对平均绝对误差 RMAE 和平均方误差(Mean Squared Error,MSE)作为评估指标,其定义 $\mathrm{MSE} = \dfrac{1}{n}\sum_{i=1}^{n}(y_i - y_i')^2$。RMAE 能够评估模型的误差百分比,而 MSE 能够更加精确地评估模型的准确度。RMAE 和 MSE 越低,意味着预测模型越高效。

3.2 预测精度评估

为了评估本文所提的基于半监督集成学习的设计空间探索方法对多核处理器建模的有效性,本文主要针对以下 3 个方面进行实验评估:①对 SSLBoost 模型中的集成学习算法 AdaBoost 的有效性进行评估,通过与 ANN 性能进行比较,来验证 AdaBoost 能否有效提升 ANN 的性能。②对 SSLBoost 模型中半监督学习算法的有效性进行评估,通过对比 SSLBoost 模型和有监督学习的 Ada-

北航学报 赠阅

Boost 模型的性能,来验证半监督学习算法的有效性。本文所采用的对未标记样本进行置信度评估的 2 个指标,是半监督集成学习算法的主流成熟技术,具有广泛的应用[22,25-26]。其中,一致性指标能够保证多个基本学习器 ANN 对样本预测的确定性,确定性越高即分歧越小的样本的预测准确率越高,属于预测器已经学习到的知识[22-25];误差下降指标也是半监督学习中增加样本的一个普遍采用的评价指标[26]。③将 SSLBoost 的预测结果与当前主流的设计空间探索方法进行了比较,包括基于 ANN[11] 和基于 SVM[13] 的预测模型。对于 ANN,本文实验中的参数设置与文献[11]中的设置一致,即包含一个 16 个神经元的隐含层,learning rate 为 0.001,而 momentum 为 0.5。对于 SVM,本文采用 LIBSVM[27] 中所提供的参数选择工具来对每个程序分别自动搜索最优的训练参数,以得到精确的预测模型。本文的 SSLBoost 模型中的基本学习器 ANN 设置的是双隐含层,每层 8 个神经元,其他参数设置与文献[11]中的一致。ANN 基本学习器设为 20 个。半监督迭代次数 K 设为 100,也就是说,本文训练样本集经过半监督学习后,被扩大到 300 的规模。鉴于训练样本集是随机采样的,对所有方法运行 10 次取预测误差的平均值作为模型的预测误差。

图 3(a)、(b)分别给出了在多核处理器设计场景下,上述方法对于性能的预测误差 RMAE 和 MSE 的对比结果,Avg 为平均值。从图 3 可以看出,SSLBoost 模型的 RMAE 分别比 SVM、ANN 和 AdaBoost 模型平均降低了 7.6%、5.2% 和1.5%。尤其是较难预测的程序 streamcluster,SSLBoost 与 SVM、ANN 和 AdaBoost 模型相比,RMAE 平均降低了 14%、10% 和 4.2%。图 3(b)展示了以 ANN 的预测误差作为标准,几种方法经过归一化处理后的 MSE 对比。可以看出,SVM、ANN 和 AdaBoost 模型的 MSE 平均是 SSLBoost 模型的 4.9 倍、3.1 倍和 1.5 倍。

以上实验结果表明,在同样的模拟次数(200)开销下,本文提出的设计空间探索方法可以显著地提升预测模型精度。与 ANN 相比,有监督的 AdaBoost 算法能够明显提升 ANN 的预测精度,并且鲁棒性较好,表明了本文所采用的集成学习算法的有效性。而 SSLBoost 模型是结合了半监督学习算法的 AdaBoost 模型,因此 SSLBoost 模型的预测精度高于原始的 AdaBoost 模型,则表明了本文提出的半监督学习算法的有效性。SSLBoost 不但利用了有标记样本对单个样本精确描述的优势,而且发挥了无标记样本对训练数据集整体描述的重要作用,从而使训练出的模型具有更好的泛化性能,充分体现了半监督学习的优势。

为了评估半监督学习的迭代过程中预测模

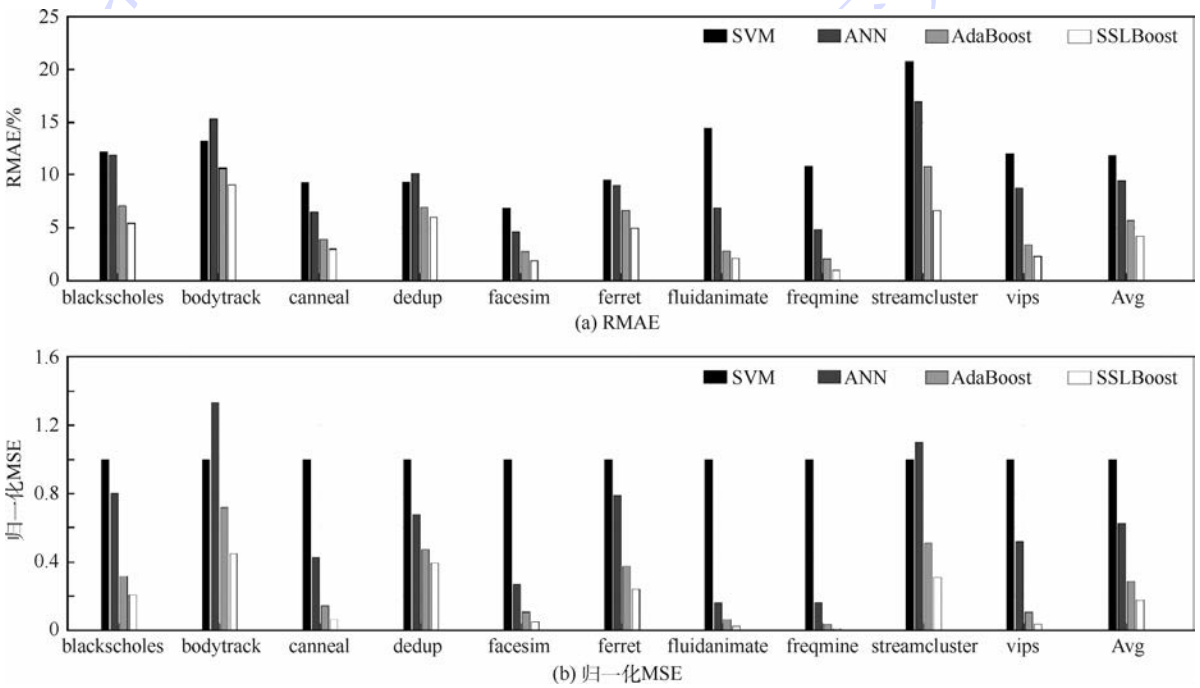


图 3　在多核设计场景下不同方法的预测精度对比

Fig. 3　Prediction accuracy comparsion of different methods in multicore design scenarios

北航学报 赠阅

型的精度变化，本文以基准程序 vips 为例，图 4(a)给出预测模型的 RMAE 随着迭代过程的变化；图 4(b)给出将迭代初始的预测误差 MSE 作为标准进行归一化后，MSE 随着迭代过程的变化。RMAE 基本是随着迭代次数的增长均匀逐渐下降的，MSE 在迭代次数大于 10 时，下降幅度较大，之后趋于平缓。由于初始训练样本集的限制，从未模拟配置中学习到的知识是有限的，当逐渐用尽能够提升预测模型精度的优秀未标记样本后，再选择的置信度高的未标记样本，对于训练数据集来说是冗余的。

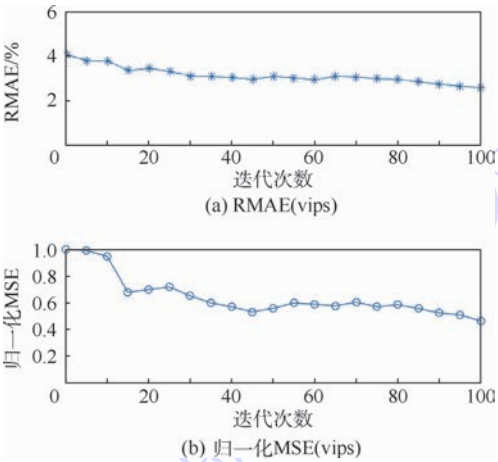


图 4 多核设计场景下 SSLBoost 模型关于不同数量的训练迭代次数的预测精度

Fig. 4 Prediction accuracy of SSLBoost model with respect to different numbers of training iterations in multicore design scenarios

3.3 模拟次数

由 3.2 节表明，本文提出的 SSLBoost 模型的预测精度在配置模拟次数相同的情况下优于其他 2 种方法(ANN 和 SVM)。本节则评估 ANN、SVM 分别需要多少训练模拟样本才能达到与 SSLBoost 模型相同的预测精度。

对于每个基准程序，和 3.2 节相同，采用 200 个均匀随机采样的训练样本来构建 SSLBoost 模型、ANN 和 SVM。从 2.2 节提到的2 000 个测试样本集中均匀随机采样选出 300 个样本，作为 ANN 和 SVM 的备用训练集。为公平起见，剩余的 1 700 个样本作为 3 种方法的共同测试样本集。然后从备用训练数据集中随机选择 10 个样本加入到训练数据集中，重新构建 ANN 和 SVM，检验 ANN 和 SVM 的预测精度是否达到 SSLBoost 的预测精度。重复这个过程，利用 10 个样本作为增长步长，直到 ANN 和 SVM 达到与 SSLBoost 相同的预测精度，或者用尽备用训练数据集里的样本。

从表 2 可以看出，ANN 和 SVM 分别需要 191%、228 + % 的模拟样本，才能达到和 SSLBoost 模型相当的预测精度。“ + ”表示对于一些基准程序，ANN 或 SVM 用尽了所有备用训练样本(300)，还是不能达到 SSLBoost 模型的预测精度。总之，SSLBoost 模型能够节省大量的训练模拟。

表 2 为达到 SSLBoost 模型相同的预测精度，ANN 和 SVM 所需要的模拟配置(训练样本)数量

Table 2 Numbers of simulated configurations (training examples) required by ANN and SVM to achieve the same level of prediction accuracy as SSLBoost model

基准程序	配置数量	
	ANN	SVM
blackscholes	360	500 +
bodytrack	350	300
canneal	320	500 +
dedup	370	270
facesim	440	500 +
ferret	310	480
fluidanimate	430	500 +
freqmine	470	500 +
streamcluster	400	500 +
vips	370	500 +
平均	382	455 +

3.4 设计空间探索时间

通常来说，模拟 10^8 条指令需要 10 min 以上的模拟的时间。基于本文的训练数据集，SSLBoost 的训练时间从 0.5 ~ 1 h 不等(随基准程序变化)。虽然比 ANN 和 SVM 的训练开销时间(秒级)要长，但是节省了更多的模拟次数，即减少了大量的模拟时间开销。实际在工业实践中，每个程序的全部动态指令数目通常都有数 10^9、甚至 10^{10} 条。在这种情况下，SSLBoost 模型的训练开销与模拟开销相比基本上可以忽略不计。因此，3.3 节的模拟次数比较，也表明了 SSLBoost 模型能够极大地减少设计空间探索时间。

3.5 探索更优配置

通常来说，除非将整个设计空间中的配置都模拟了，才能找到最优的配置，但这是不可行的。一个折中方案是对比由 SSLBoost 模型预测出的最优配置和由现有设计空间探索方法 ANN 和 SVM 预测的最优配置相对比。具体来讲，从整个设计空间中随机采样 3 000 个配置，然后利用这 3 种方法来预测这 3 000 个配置的性能响应，从而寻找到其中的最优配置。最后模拟 3 种方法找到的最优配置，并直接比较其真实的模拟性能响应。以基准程序 vips 为例，表 3 展示了 SSLBoost 模型和其他2种模型预测的最优配置以及预测精度

北航学报 赠阅

表 3　各种模型预测的最优配置对比(vips)

Table 3　Predicted optimal configurations of different models(vips)

设计参数及性能	SSLBoost 模型	ANN	SVM
Band Width/(GB·s^{-1})	64	8	40
Frequency/GHz	4.5	4.5	4
Issue/Fetch/Commit Width	4	2	4
Number of Cores	1	1	2
L2 Cache Size/MB	4	2	2
L2 Cache Block Size/B	64	128	64
L2 Cache Associativity	8	2	2
L2 Cache MSHR	128	64	128
L1 Dcache/KB	128	128	64
L1 ICache/KB	64	16	16
真实模拟性能/ms	11.52	11.78	13.68
预测性能/ms	11.12	10.27	10.42
预测误差/%	3.5	12.9	23.8

(RMAE)。这里 ANN、SVM 的训练样本集和 SSLBoost模型的初始训练集大小都为 200，测试样本集合大小为2 000。从表 3 可以看出，与其他 2 种方法相比，SSLBoost 模型寻找的最优配置的性能最好，并且，SSLBoost 模型对该配置预测的性能响应非常接近这个配置的真实模拟性能，仅仅 3.5% 的预测误差。而其他 2 个方法得到的最优配置性能都较低，且预测精度也较低(误差为 12.9%、23.8%)。因此，在探索更优配置方面，SSLBoost 模型无论在准确性和还是有效性上都优于 ANN 和 SVM。

4　结　论

本文提出了一种有效且高效的结合集成学习和半监督学习算法的多核设计空间探索框架。实验结果表明：

1）采用集成学习算法 AdaBoost 模型来集成 ANN 的预测模型能够有效地提升 ANN 的预测精度。

2）半监督集成学习模型 SSLBoost 模型利用未标记配置来扩充训练样本集，在相同的模拟代价下，相比现有的基于 ANN 和 SVM 的探索方法，能够极大地降低模型的预测误差。

3）相比基于 ANN 和 SVM 的方法，SSLBoost 模型能够使用更少的模拟样本来达到相当的预测精度。

因此，本文所提出的多核设计空间探索方法 SSLBoost 模型比现有的探索方法更加高效、准确。

参考文献（References）

[1] NOONBURG D B, SHEN J P. Theoretical modeling of superscalar processor performance[C]// Proceeding of International Symposium on Microarchitecture. New York: ACM, 1994: 52-62.

[2] KARKHANIS T S, SMITH J E. Automated design of application specific superscalar processors: An analytical approach[C]// Proceedings of the 34th International Symposium on Computer Architecture. New York: ACM, 2007: 402-411.

[3] HAMERLY G, PERELMAN E, CALDER B. How to use SimPoint to pick simulation points[J]. ACM Sigmetrics Performance Evaluation Review, 2004, 31(4): 25-30.

[4] WUNDERLICH R E, WENISCH T F, FALSAFI B, et al. SMARTS: Accelerating microarchitecture simulation via rigorous statistical sampling[C]// Proceedings of the 30th Annual International Symposium on Computer Architecture. New York: ACM, 2003: 84-97.

[5] WANG S, HU X, YU P S, et al. MMRate: Inferring multi-aspect diffusion networks with multi-pattern cascades[C]// ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2014: 1246-1255.

[6] WANG S, LI Z, CHAO W, et al. Applying adaptive over-sampling technique based on data density and cost-sensitive SVM to imbalanced learning[C]// International Symposium on Neural Networks. Piscataway, NJ: IEEE Press, 2012: 1-8.

[7] JOSEPH P J, VASWANI K, THAZHUTHAVEETIL M J. Construction and use of linear regression models for processor performance analysis[C]// Proceedings of the 12th International Symposium on High-Performance Computer Architecture. Piscataway, NJ: IEEE Press, 2006: 99-108.

[8] JOSEPH P J, VASWANI K, THAZHUTHAVEETIL M J. A predictive performance model for superscalar processors[C]// Proceedings of the 39th Annual IEEE/ACM International Symposium on Microarchitecture. Piscataway, NJ: IEEE Press, 2006: 161-170.

[9] LEE B C, BROOKS D M. Accurate and efficient regression modeling for microarchitectural performance and power prediction[C]// Proceedings of 12th International Conference on Architectural Support for Programming Language and Operating Systems. New York: ACM, 2006: 185-194.

[10] LEE B C, COLLINS J, WANG H, et al. CPR: Composable performance regression for scalable multiprocessor models[C]// Proceedings of the 41 st Annual IEEE/ACM International Symposium on Microarchitecture. Piscataway, NJ: IEEE Press, 2008: 270-281.

[11] ÏPEK E, MCKEE S A, CARUANA R, et al. Efficiently exploring architectural design spaces via predictive modeling[C]// Proceedings of 12th International Conference on Architectural Support for Programming Language and Operating Systems. New York: ACM, 2006: 195-206.

[12] 郭崎，陈天石，陈云霁. 基于模型树的多核设计空间探索技术[J]. 计算机辅助设计与图形学学报，2012，24(6)：710-720.
GUO Q, CHEN T S, CHEN Y J. Model tree based multi-core design space exploration[J]. Journal of Computer-Aided Design & Computer Graphics, 2012, 24(6): 710-720(in Chinese).

[13] PANG J F, LI X F, XIE J S, et al. Microarchitectural design

北航学报 赠阅

space exploration via support vector machine[J]. Acta Scientiarum Naturalium Universitatis Pekinensis,2010,46(1):55-63.

[14] PALERMO G, SILVANO C, ZACCARIA V. ReSPIR: A response surface-based Pareto iterative refinement for application-specific design space exploration[J]. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 2009,28(12):1816-1829.

[15] XYDIS S,PALERMO G,ZACCARIA V,et al. SPIRIT:Spectral-aware Pareto iterative refinement optimization for supervised high-level synthesis[J]. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 2015, 34(1): 155-159.

[16] GUO Q,CHEN T,ZHOU Z H,et al. Robust design space modeling[J].ACM Transactions on Design Automation of Electronic Systems,2015:20(2):18.

[17] LI D,YAO S,LIU Y H,et al. Efficient design space exploration via statistical sampling and AdaBoost learning[C]// Design Automation Conference. New York:ACM,2016:1-6.

[18] KHAN S,XEKALAKIS P,CAVAZOS J,et al. Using predictive-modeling for cross-program design space exploration in multicore systems[C]// Proceedings of the 22nd International Conference on Parallel Architectures and Compilation Techniques. Piscataway,NJ:IEEE Press,2007:327-338.

[19] DUBACH C,JONES T,OBOYLE M. Microarchitectural design space exploration using an architecture-centric approach[C]// Proceedings of the 40th Annual IEEE/ACM International Symposium on Microarchitecture. Piscataway, NJ: IEEE Press, 2007:262-271.

[20] LI D,WANG S,YAO S,et al. Efficient design space exploration by knowledge transfer[C]// Eleventh IEEE/ACM/IFIP International Conference on Hardware/software Codesign and System Synthesis. New York:ACM,2016:1-10.

[21] SHRESTHA D L,SOLOMATINE D P. Experiments with AdaBoost. RT, an improved boosting scheme for regression[J]. Neural Computation,2006,18(7):1678-1710.

[22] ZHOU Z H,LI M. Semi-supervised learning by disagreement[J]. Knowledge and Information Systems, 2010, 24(3): 415-439.

[23] BINKERT N,BECKMANN B,BLACK G,et al. The gem5 simulator[J]. ACM SIGARCH Computer Architecture News,2011, 39(2):1-7.

[24] BIENIA C,KUMAR S,SINGH J P,et al. The PARSEC benchmark suite: Characterization and architectural implications [C]// Proceedings of the 17th International Conference on Parallel Architecture and Compilation Techniques. New York: ACM,2008:72-81.

[25] HAMED V, RONG J, ANIL K. Semi-supervised boosting for multi-class classification[C]// European Conference on Principles of Data Mining and Knowledge Discovery,2008:522-537.

[26] ZHOU Z H,LI M. Semi-supervised regression with co-training [C]// Proceedings of the 19th International Joint Conference on Artificial Intelligence. New York:ACM,2005:908-913.

[27] CHANG C C,LIN C J. LIBSVM:A library for support vector machines[J]. ACM Transactions on Intelligent Systems and Technology,2011,2(3):27-1-27-27.

作者简介：

李丹丹 女，博士研究生。主要研究方向：处理器设计、机器学习、软件工程技术。

姚淑珍 女，博士，教授，博士生导师。主要研究方向：先进软件工程技术、形式化方法、Petri 网理论。

王颖 男，博士，讲师。主要研究方向：存储系统、节能加速器、容错体系结构。

王森章 男，博士，副研究员。主要研究方向：数据挖掘、社交网络分析、大数据。

谭火彬 男，博士，讲师。主要研究方向：软件工程、软件建模。

北京航空航天大学学报

北航学报 赠阅

Multicore design space exploration via semi-supervised ensemble learning

LI Dandan[1], YAO Shuzhen[1,*], WANG Ying[2], WANG Senzhang[3], TAN Huobin[4]

(1. School of Computer Science and Engineering, Beijing University of Aeronautics and Astronautics, Beijing 100083, China;
2. State Key Laboratory of Computer Architecture, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China;
3. College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China;
4. School of Software, Beijing University of Aeronautics and Astronautics, Beijing 100083, China)

Abstract: With the increasing complexity of microprocessor architecture, the design space is growing exponentially and the software simulation technology is extremely time-consuming. Design space exploration becomes one major challenge when processors are designed. The paper proposed an efficient design space exploration method combining semi-supervised learning and ensemble learning techniques. Specifically, it includes two phases: uniform random sampling method is firstly employed to select a small set of representative design points, and then simulation is conducted with the points to constitute the training set; semi-supervised learning based AdaBoost (SSLBoost) model is further proposed to predict the responses of the configurations that have not been simulated. Then the optimal processor design configuration is found. The experimental results demonstrate that compared with the prediction models based on the existing artificial neural network and support vector machine (SVM), the proposed SSLBoost model can build a comparable accurate model using fewer simulations. When the number of simulation examples is fixed, the prediction accuracy of SSLBoost model is higher.

Keywords: design space exploration; semi-supervised learning; ensemble learning; AdaBoost; microprocessor; predictive model

Received: 2017-05-11; **Accepted**: 2017-06-16; **Published online**: 2017-08-31 14:35
URL: kns.cnki.net/kcms/detail/11.2625.V.20170831.1435.007.html
Foundation item: Aeronautical Science Foundation of China (2013ZC51023)
* **Corresponding author**. E-mail: szyao@buaa.edu.cn

北京航空航天大学学报

北航学报 赠阅

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2017.0225

基于测试性 *D* 矩阵的多故障诊断与维修策略

全下文载

田恒[1], 段富海[1,*], 樊亮[2], 桑勇[1], 史萌[2]

(1. 大连理工大学 机械工程学院, 大连 116024;
2. 中国航空工业集团公司西安飞行自动控制研究所 惯性技术航空科技重点实验室, 西安 710065)

摘 要: 针对复杂的多故障诊断问题以及多故障直接处理方法实现的难点,在测试性 ***D*** 矩阵基础上,提出了一种基于单故障化的多故障诊断与维修策略(MFDMSTS)。首先,在多故障假设下引入析取运算,定义了可隔离单故障和可隔离多故障,据此定义将多故障转化为单故障,并将转化的单故障与测试集组成新的 ***D*** 矩阵;然后,运用单故障诊断算法处理新的 ***D*** 矩阵,得到最优诊断树;最后,针对诊断树的不同叶子节点,提出了多故障诊断与维修策略。实例验算表明:MFDMSTS 能降低平均诊断费用和平均诊断步数,并大幅降低误修率。

关 键 词: 测试性 *D* 矩阵; 多故障诊断; 析取运算; 多故障单故障化; 维修策略
中图分类号: TP206.3
文献标识码: A **文章编号**: 1001-5965(2018)04-0802-08

随着装备对测试性要求的不断提高,故障诊断技术研究日益深入,其被广泛应用在航空、航天和兵器等高精尖的新武器装备故障诊断上。目前大多数故障诊断是在单故障假设下进行的,然而在实际工程中,并非所有的故障都是单个发生的,有些是随机组合或者耦合出现,单故障假设在实际应用中会遇到许多问题。因此研究人员逐渐开展了多故障诊断研究,目前多故障诊断问题是国内外故障诊断研究的热点[1]。

多故障的处理方法可大致分为多故障直接处理方法和扩展单故障方法。多故障直接处理方法较多,如连可等[2]基于小波变换和多类支持向量机(SVM)提出了一种多故障诊断方法;杨鹏等[3]运用布尔逻辑运算生成完备最小碰集,然后用基于信息量启发式算法生成近似最优的多故障隔离策略;赵亮亮等[4]提出一种基于多色集与概念格的多故障推理方法;Kundu 等[5]通过粒子群算法和多故障注入实现多故障诊断;Issury 等[6]基于一阶逻辑公式和布尔代数工具,利用冲突和诊断的概念创建最小诊断集的方式解决多故障隔离问题;Bartyś[7-9]提出了基于多故障可隔离矩阵和多故障弱隔离结构剩余集,寻找最优故障隔离结构的方法以及基于二值诊断矩阵和动态二值矩阵进行多故障的诊断和隔离方法。多故障直接处理方法实现的难点在于故障组合的不确定性,无法判断故障组合方式就很难确定最优的诊断算法;另外多故障组合非常多,且以指数形式的增加,多故障直接处理方法必须考虑如何处理“爆炸”的故障组合。扩展单故障的多故障处理方法,如 Shakeri 等[10]提出基于扩展单故障和碰集的 4 种多故障诊断策略,并按照“测试—维修”方式诊断多故障;王子玲等[11]基于扩展单故障策略,应用附加测试来隔离隐藏故障,并调用单故障策略来隔离冒充故障,从而实现多故障诊断和隔离。扩展单故障的多故障处理方法,单故障测试阶段最优,却无法保证后续扩展阶段是最优的测试序列。多故

收稿日期: 2017-04-13; **录用日期**: 2017-05-19; **网络出版时间**: 2017-06-30 09:02
网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20170630.0902.004.html
基金项目: 航空科学基金 (20150863003)
* **通信作者**. E-mail: duanfh@dlut.edu.cn

引用格式: 田恒, 段富海, 樊亮, 等. 基于测试性 ***D*** 矩阵的多故障诊断与维修策略[J]. 北京航空航天大学学报, 2018, 44(4): 802-809. TIAN H, DUAN F H, FAN L, et al. Multiple fault diagnosis and maintenance strategy based on testability ***D*** matrix [J]. Journal of Beijing University of Aeronautics and Astronautics, 2018, 44(4): 802-809 (in Chinese).

障算法仍存在诸多问题，但单故障算法已相对成熟，如贪心算法[12]和 AO* 算法[13]等。

综合上述分析，本文基于测试性 **D** 矩阵，研究运用单故障算法处理多故障问题的新方法。首先基于测试性 **D** 矩阵元素意义的变化引入析取运算，根据析取运算提出几个基本定义，运用定义和多故障规律判断多故障组合；然后完成多故障单故障化，并将新的单故障集合与原测试集组成新的 **D** 矩阵，运用单故障算法对新 **D** 矩阵进行处理，得到最优诊断序列，从而得到多故障诊断的最优诊断树；最后将对应的单故障转化为多故障，对多故障集合提出维修策略。

1　析取运算

为便于叙述，引入 $T=\{t_1,t_2,\cdots,t_n\}$ 表示测试集，$C=\{c_1,c_2,\cdots,c_n\}$ 表示对应测试的成本；$F=\{f_1,f_2,\cdots,f_m\}$ 为单故障集合；$P=\{p(f_1),p(f_2),\cdots,p(f_m)\}$ 为对应单故障的概率；$\boldsymbol{D}=[d_{ij}]_{m\times n}$ 为故障和测试的依赖矩阵，即 **D** 矩阵。本文用 $\mathrm{eval}(f_i,t_j)$ 表示故障与测试之间的依赖关系：$d_{ij}=\mathrm{eval}(f_i,t_j)$。如果 t_j 可以检测 f_i，$d_{ij}=1$；否则 $d_{ij}=0$。

D 矩阵的行和列可用向量形式表示：

行向量 $\boldsymbol{V}_i^f=[\mathrm{eval}(f_i,t_1)\quad \mathrm{eval}(f_i,t_2)\ \cdots\ \mathrm{eval}(f_i,t_n)]=[\mathrm{eval}(f_i,t_j)]_{1\times n}(1\leqslant j\leqslant n)$，简记为 $\boldsymbol{V}_i^f=[\mathrm{eval}(f_i,T)]$。其中，$\boldsymbol{V}^f$ 表示行向量，下标 i 表示第 i 行。

列向量 $\boldsymbol{V}_j^t=[\mathrm{eval}(f_1,t_j)\quad \mathrm{eval}(f_2,t_j)\ \cdots\ \mathrm{eval}(f_m,t_j)]^{\mathrm{T}}=[\mathrm{eval}(f_i,t_j)]_{1\times m}^{\mathrm{T}}(1\leqslant i\leqslant m)$，简记为 $\boldsymbol{V}_j^t=[\mathrm{eval}(F,t_j)]^{\mathrm{T}}$。其中，$\boldsymbol{V}^t$ 表示列向量，下标 j 表示第 j 列。

多故障假设下，**D** 矩阵元素 d_{ij} 的意义与单故障假设下不同[14]，多故障下 1 和 0 存在“遮掩”。例如若 t_j 的测试结果为 1 时，故障集 F 中除了满足 $\mathrm{eval}(f_i,t_j)=1$ 的故障 f_i 可能出现之外，其他的故障虽然满足 $[\mathrm{eval}((F/f_i),t_j)]^{\mathrm{T}}=[0]^{\mathrm{T}}$，但也有可能发生故障。

基于上述分析，引入析取运算。析取运算符为“$\vee$”，运算关系如下：

1）**D** 矩阵元素之间的析取运算。设 $\mathrm{eval}(f_i,t_j)=0$，$\mathrm{eval}(f_i,t_r)=0$，$\mathrm{eval}(f_i,t_k)=1$，$\mathrm{eval}(f_i,t_q)=1$，其中 $t_j,t_r,t_k,t_q\in T$，则：

$$\mathrm{eval}(f_i,t_j)\vee\mathrm{eval}(f_i,t_r)=0$$

$$\mathrm{eval}(f_i,t_j)\vee\mathrm{eval}(f_i,t_k)=1$$

$$\mathrm{eval}(f_i,t_k)\vee\mathrm{eval}(f_i,t_q)=1$$

2）假设 $t_s,t_w\in T$，$f_i,f_j\in F$，则测试向量 $\boldsymbol{V}_s^t$ 和 $\boldsymbol{V}_w^t$ 之间，故障向量 $\boldsymbol{V}_i^f$ 和 $\boldsymbol{V}_j^f$ 之间的析取运算式分别为

$$\boldsymbol{V}_s^t\vee\boldsymbol{V}_w^t=[\mathrm{eval}(F,t_j)]^{\mathrm{T}}\vee[\mathrm{eval}(F,t_w)]^{\mathrm{T}}=[(\mathrm{eval}(f_i,t_s)\vee\mathrm{eval}(f_i,t_w)]^{\mathrm{T}}$$

$$\boldsymbol{V}_i^f\vee\boldsymbol{V}_j^f=[\mathrm{eval}(f_i,T)]\vee[\mathrm{eval}(f_j,T)]=[(\mathrm{eval}(f_i,t_w)\vee\mathrm{eval}(f_j,t_w)]$$

3）测试向量和故障向量连续析取运算式分别为

$$\boldsymbol{V}_1^t\vee\boldsymbol{V}_2^t\vee\cdots\vee\boldsymbol{V}_n^t=\sum_{i=1}^{n}\vee\boldsymbol{V}_i^t$$

$$\boldsymbol{V}_1^f\vee\boldsymbol{V}_2^f\vee\cdots\vee\boldsymbol{V}_m^f=\sum_{i=1}^{m}\vee\boldsymbol{V}_m^f$$

2　多故障组合的基本定义与规律

多故障单故障化的难点在于多故障的判断，多故障的组合是随机的，对于有 m 个单故障的系统，将会有 2^m-1 个组合故障（不含系统无故障状态 f_0），并且以指数的形式增加，这就给多故障诊断带来困难。**D** 矩阵中的测试集有限，能被诊断隔离的故障只是少数，因此从可诊断隔离的故障入手，通过判断故障或者故障集能否被隔离，实现可隔离故障的单故障化。在多故障假设下，单故障、无故障与多故障组合均有可能发生。基于以上分析，提出几个基本定义，并应用文献[10]中具有典型意义的某电路系统 **D** 矩阵（见表 1）进行说明，无故障状态 f_0 的概率为 0.62。

表 1　某电路系统的 D 矩阵[10]

Table 1　D matrix of a circuit system[10]

故障状态	测试					概率
	t_1	t_2	t_3	t_4	t_5	
f_1	0	1	0	0	1	0.014
f_2	0	0	1	1	0	0.027
f_3	1	0	0	1	1	0.125
f_4	1	1	0	0	0	0.068
f_5	1	1	1	1	0	0.146

2.1　单故障的基本定义

单故障假设下，隔离测试和诊断测试共同作用，不分主次[15]；多故障假设下，故障更多的是依赖隔离测试，诊断测试则是起辅助判断作用。假设测试结果 100% 可靠。

定义 1　在 $\boldsymbol{D}_{m\times n}$ 中，对于故障 f_i，如果存在测试集 $X\subset T$，$\forall t_j\in X$，满足 $\sum_{t_j\in X}\vee\boldsymbol{V}_j^t=[w_q]^{\mathrm{T}}$，$w_q=\begin{cases}0 & q=i\\ 1 & q\neq i\end{cases}$，$1\leqslant q\leqslant m$，则称 f_i 为可隔离单故障。

北航学报 赠阅

满足定义 1 的测试集 X 称为 f_i 隔离测试集，由 X 组成的集合记为 F_{IT}^i，称为故障 f_i 隔离测试总集。满足 $\mathrm{eval}(f_i,t_k)=1$ 的测试 t_k，称为 f_i 的诊断测试。

例如表 1 中的 f_5，没有满足定义 1 的测试集，那么它不存在故障隔离测试集，也不能被隔离。f_1 存在故障隔离测试总集 $F_{\mathrm{IT}}^1=\{\{t_3,t_1\},\{t_4,t_1\}\}$，则 f_1 是可隔离单故障。

如果故障 f_i 存在故障隔离测试集，则能被隔离，否则只能够隔离到一组包含 f_i 的模糊集。f_0 只能通过隔离测试进行判断，其故障隔离测试集必定存在。

定义 2 在 $\boldsymbol{D}_{m\times n}$ 中，如果存在测试集 $Z\subset T$，$\forall t_j\in Z$，满足 $\sum_{t_j\in Z}\vee\boldsymbol{V}_j^t=[1]_{1\times m}$，则称 Z 为 f_0 故障隔离测试集。由 Z 组成的集合记为 F_{IT}^0，称为故障 f_0 隔离测试总集。

2.2 多故障组合的基本定义

定义 3 在 $\boldsymbol{D}_{m\times n}$ 中，若存在测试集 $Y\subset T$，$\forall t_k\in Y$，$\sum_{\forall t_k\in Y}\vee\boldsymbol{V}_k^t=[w_q]^{\mathrm{T}}$，满足

$$w_q=\begin{cases}0 & \forall f_i\in M_{\mathrm{FS}},\mathrm{eval}(f_i,t_k)\\ 1 & \forall f_j\notin M_{\mathrm{FS}},\mathrm{eval}(f_j,t_k)\end{cases},1\leqslant q\leqslant m$$

则称 M_{FS} 为可隔离多故障。

测试集 Y 称为多故障 M_{FS} 的隔离测试集，由测试集 Y 成组的集合记为 $F_{\mathrm{IT}}(M_{\mathrm{FS}})$，称为多故障 M_{FS} 的故障隔离测试总集。

例如表 1 中 f_3 和 f_5 组成的多故障，没有满足定义 3 的测试集，那么它不存在故障隔离测试集，也不能被隔离。f_1 和 f_4 组成的多故障存在故障隔离测试总集 $F_{\mathrm{IT}}(f_1,f_4)=\{\{t_3,t_4\},\{t_4\}\}$，则 $M_{\mathrm{FS}}(f_1,f_4)$ 是可隔离多故障。

对于组合故障 M_{fi}，若存在 $M_{fj}(i\neq j)$ 满足 $\sum_{f_i\in M_{fi}}\vee\boldsymbol{V}_i^f=\sum_{f_j\in M_{fj}}\vee\boldsymbol{V}_j^f$，则称 M_{fi} 和 M_{fj} 征兆相同。

2.3 多故障规律

单故障假设下，能被诊断隔离的故障均可以在诊断树叶子节点上显示。多故障也有相似的规律：

1）可隔离单故障和可隔离多故障（统称为可隔离故障，记为 I_{FS}），最终会显示在诊断树的叶子节点上。

2）显示在叶子节点上的多故障，并不一定表示被完全隔离，其中会隐含一些故障，由于它们之间的“征兆”相同，共用一个叶子节点。

3）所有故障组合都会在诊断树上出现，但有的故障被其他故障掩盖，无法分辨出。被掩盖的故障可能是隐含故障、伪故障，甚至单故障元素。

另外，多故障 F 是特例，表示集合 $F=\{f_1,f_2,\cdots,f_m\}$ 中的所有故障同时发生。F 在叶子节点上显示，却不存在隔离测试集。这是因为整体上看多故障 F 是唯一的故障，对任意测试 t_j，若不通过则表示 F 故障，通过则是 f_0，因此不用隔离测试也可以隔离 F。

3 基于单故障化的多故障诊断与维修策略

基于单故障化的多故障诊断与维修策略（Multiple Fault Diagnosis and Maintenance Strategy by Translating into Single fault，MFDMSTS）分为单故障化和维修诊断 2 部分。单故障化阶段考虑多故障单故障化以及测试序列寻优，维修诊断阶段考虑故障集内部的组合以及维修策略。

3.1 单故障化

步骤 1 确定可隔离单故障和隔离测试集。根据定义 1 和定义 2 确定可隔离的单故障、无故障 f_0、F_{IT}^i 和 F_{IT}^0。

步骤 2 确定可隔离多故障集和隔离测试集。根据定义 3 确定可隔离的多故障集合 M_{FS} 和 $F_{\mathrm{IT}}(M_{\mathrm{FS}})$。

步骤 3 根据征兆相同的确定可隔离故障集内部的组合数，且不用考虑隐含故障和伪故障等关系。

步骤 4 计算可隔离故障集的概率。由步骤 1 和 2 得到一组可隔离故障集合 $F_{\mathrm{S}}=\{F_{\mathrm{S}}^1,F_{\mathrm{S}}^2,\cdots,F_{\mathrm{S}}^z\}\cup f_0$，$p(F_{\mathrm{S}}^i)$ 概率计算式为

$$p(F_{\mathrm{S}}^i)=\sum_{k=1}^{|F_{\mathrm{S}}^i|}\prod_{f_j\in M_{fk}}p(f_j)\tag{1}$$

式中：$|F_{\mathrm{S}}^i|$ 为 F_{S}^i 包含的故障组合数；M_{fk} 为 F_{S}^i 的一组故障组合；f_j 为 M_{fk} 的组成元素。

步骤 5 可隔离故障集的概率归一化。步骤 4 的计算结果会造成整体故障概率大于 1，那么需要对故障概率进行归一化处理，概率归一化计算式为

$$p(s_0)=1\Big/\left(1+\sum_{k=1}^{z}\frac{p(F_{\mathrm{S}}^k)}{1-p(F_{\mathrm{S}}^k)}\right)\tag{2}$$

$$p(s_i)=\frac{p(F_{\mathrm{S}}^i)}{1-p(F_{\mathrm{S}}^i)}\Big/\left(1+\sum_{k=1}^{z}\frac{p(F_{\mathrm{S}}^k)}{1-p(F_{\mathrm{S}}^k)}\right)\tag{3}$$

式中：$p(s_0)$ 为无故障状态的概率；z 为集合 F_{S} 的基数；$p(s_i)$ 为 F_{S}^i 归一化后的概率。

步骤 6 可隔离故障集与测试集 T 组成新的 $\boldsymbol{D}$ 矩阵，令 s_0 表示 f_0，s_i 表示 F_{S}^i，将 s_0、$S=\{s_1,$

北航学报 赠阅

$s_2,\cdots,s_z\}$ 与测试集 T 组成新的 $\boldsymbol{D}_{(z+1)\times n}$ 矩阵。

步骤 7　用单故障算法处理新 $\boldsymbol{D}$ 矩阵，得到相应的诊断树，计算诊断树的各项指标，如平均诊断费用、故障诊断率和隔离率等[16]。

3.2　维修诊断

完成系统的单故障化并得到最优诊断树后，需要对诊断树的各个叶子节点提出合适的维修策略，基于叶子节点内组合数的不同，分为 2 类维修方式：

1）组合数为 1 的叶子节点，表示该叶子节点故障元素是确定唯一的，只需对该叶子节点的所有故障元素进行更换或者维修。

2）组合数大于 1 的叶子节点，表示该叶子节点可能存在隐含故障或者伪故障。应该按照叶子节点内部的所有单故障元素之间的概率大小排序，先依次对这些单故障元素进行“测试—维修”[10]。

例如某叶子节点内部包含多故障 $\{f_5\}$，$\{f_2,f_5\}$，第 1 个维修的集合为 $\{f_5\}$，第 2 个维修的集合为 $\{f_2,f_5\}$，维修 $\{f_5\}$ 后，第 2 个维修内容只需维修 f_2 即可。这种方法适用于非耦合故障系统，能有效提高维修的效率。但该方法会引起误修，因此提出误修率的计算公式：

$$M_{\mathrm{R}}(F_{\mathrm{S}}^i)=\frac{\left[p(F_{\mathrm{S}}^i)-\left(\sum_{j=1}^{|R(F_{\mathrm{S}}^i)|}\prod_{f_k\in R(F_{\mathrm{S}}^i)_j}p(f_k)\right)\right]p(s_i)}{p(F_{\mathrm{S}}^i)} \tag{4}$$

式中：$M_{\mathrm{R}}(F_{\mathrm{S}}^i)$ 为可隔离故障集 F_{S}^i 的误修率；$R(F_{\mathrm{S}}^i)=\{R(F_{\mathrm{S}}^i)_1,R(F_{\mathrm{S}}^i)_2,\cdots,R(F_{\mathrm{S}}^i)_h\}$ 为可隔离故障 F_{S}^i 维修诊断树的叶子节点集，$R(F_{\mathrm{S}}^i)_j$ 为第 j 个叶子节点所完成维修的故障集。

维修诊断树中的每个叶子节点可能包含有多个故障集，每个维修诊断树的叶子节点仅有一个故障集被维修，并且该集合包含该叶子节点所有的故障组合情形。

4　实例计算

4.1　电路系统实例

4.1.1　故障诊断及结果对比

在多故障假设下，按照 3.1 节步骤 1～5 对表 1 系统单故障化处理，得到 I_{FS}、对应的组合数、F_{IT}、可隔离故障集的概率和归一化后的概率，如表 2 所示。

根据步骤 6，将 s_0、$S=\{s_1,s_2,\cdots,s_{10}\}$、$p(s_i)$ 和测试集 $T=\{t_1,t_2,t_3,t_4,t_5\}$ 组成新的 $\boldsymbol{D}_{11\times5}$，如表 3 所示。

根据步骤 7，采用文献[13]中的 AO* 算法处理表 3 的新 $\boldsymbol{D}$ 矩阵，得到诊断树，如图 1 所示。

MFDMSTS 方法的故障诊断隔离和维修诊断是分开的。运用文献[10]的 Sure3 算法对表 1 进行多故障处理，为使结果有可比性，按照诊断隔离和“测试—维修”2 种方式计算，结果分别用 Sure3* 和 Sure3 表示，并且计算的过程用本文归一化后的概率。通过计算平均诊断费用 C_{ost}、平均诊断步数 N_{D}、故障诊断率 FDR 和故障隔离率 FIR 来评价 MFDMSTS，计算式分别为

$$C_{\mathrm{ost}}=\sum_{i=0}^{z}p(s_i)\sum_{t_j\in T_i}c_j \tag{5}$$

$$N_{\mathrm{D}}=\sum_{i=0}^{z}\frac{|T_i|}{1+z} \tag{6}$$

$$\mathrm{FDR}=\sum_{i=0}^{z}p(s_i) \tag{7}$$

表 2　多故障单故障化的结果

Table 2　Results of translating multiple faults into single fault

I_{FS}	组合数	F_{IT}	故障状态	概率 $p(F_{\mathrm{S}}^i)$	归一化概率 $p(s_i)$
f_0	1	$\{\{t_2,t_4\},\{t_1,t_2,t_3\},\cdots,\{t_1,t_2,t_3,t_4,t_5\}\}$	s_0		0.671 3
f_1	1	$\{\{t_4,t_1\},\{t_3,t_1\}\}$	s_1	0.014 000	0.009 5
f_2	1	$\{\{t_2,t_5\},\{t_2,t_1\},\{t_1,t_5\}\}$	s_2	0.027 000	0.018 6
f_3	1	$\{\{t_2,t_3\}\}$	s_3	0.125 000	0.095 9
f_4	1	$\{\{t_3,t_5\},\{t_4,t_5\}\}$	s_4	0.068 000	0.049 0
f_1f_2	1	$\{\{t_1\}\}$	s_5	0.000 378	0.000 3
f_2f_3	1	$\{\{t_2\}\}$	s_6	0.003 400	0.002 3
f_1f_4	1	$\{\{t_3,t_4\},\{t_4\}\}$	s_7	0.000 950	0.000 6
$f_1f_3f_4$	3	$\{\{t_3\}\}$	s_8	0.010 400	0.007 0
$f_2f_4f_5$	5	$\{\{t_5\}\}$	s_9	0.162 000	0.129 8
$f_1f_2f_3f_4f_5$	16		s_{10}	0.022 800	0.015 7
总和	32				1.000 0

北航学报 赠阅

表 3 MFDMSTS 得到的新 D 矩阵

Table 3 New D matrix obtained by MFDMSTS

故障状态	测试					概率
	t_1	t_2	t_3	t_4	t_5	
s_0	0	0	0	0	0	0.671 3
s_1	0	1	0	0	1	0.009 5
s_2	0	0	1	1	0	0.018 6
s_3	1	0	0	1	1	0.095 9
s_4	1	1	0	0	0	0.049 0
s_5	0	1	1	1	1	0.000 3
s_6	1	0	1	1	1	0.002 3
s_7	1	1	0	0	1	0.000 6
s_8	1	1	0	1	1	0.007 0
s_9	1	1	1	1	0	0.129 8
s_{10}	1	1	1	1	1	0.015 7

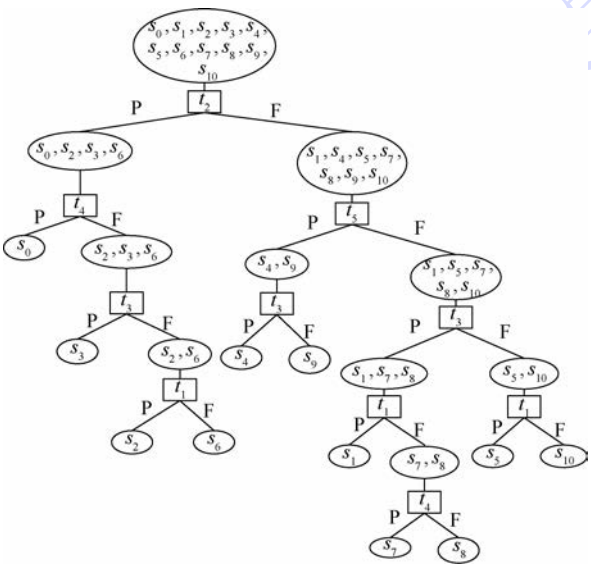


图 1 基于 MFDMSTS 和 AO* 的新 D 矩阵诊断树

Fig. 1 Diagnostic tree of new D matrix based on MFDMSTS and AO*

$$\mathrm{FIR}=\sum_{|s_i|=1}p(s_i) \tag{8}$$

式中：C_{ost} 为平均诊断费用，c_j 为对应 t_j 的费用；N_D 为平均诊断步数；$\boldsymbol{T}_i$ 为 s_i 的测试序列，$|\boldsymbol{T}_i|$ 为 $\boldsymbol{T}_i$ 的长度；$|s_i|$ 为故障 s_i 内的组合数。

假设 $c_j=1$，由式(5) ~ 式(8)得 MFDMSTS、Sure3* 和 Sure3 的相应结果，如表 4 所示。

由表 4 可知，MFDMSTS 能有效降低多故障的 C_{ost}，因为单故障化可以对所有可隔离故障集进行测试序列的寻优计算，而 Sure3* 与 Sure3 只是对单故障假设下的诊断树和测试序列进行扩展，忽略了一些概率较大的故障集的顺序；Sure3* 与 Sure3 的 C_{ost} 相差不大是因为维修对象的概率较小，对成本影响较小；在故障诊断阶段，由于Sure3* 和 MFDMSTS 不存在“测试—维修”，

表 4 MFDMSTS、Sure3* 和 Sure3 的诊断结果对比

Table 4 Comparison of diagnosis results among MFDMSTS, Sure3* and Sure3

算法	C_{ost}	N_D	FDR/%	FIR/%
MFDMSTS	2.390	3.727	100	84.75
Sure3*	2.515	3.727	100	84.75
Sure3	2.528	4.154	100	85.46

Sure3* 和 MFDMSTS 的 FIR 会比较低，两者的 N_D 也会较小。

在 2 G，2.6 GHz 的 PC 机上运用 MATLAB 运行程序，该电路系统的多故障转化为单故障的计算时间为 0.077 0 s。

4.1.2 维修策略

完成故障的诊断后，将对诊断树的叶子节点提出合适的维修策略。本文运用诊断树的形式表现测试诊断过程，用 $\{N|L\}$ 的格式表示维修诊断树的节点含义，其中 N 表示节点内的组合故障数，L 表示所包含的故障状态。诊断树叶子节点的维修策略以及与 Sure3 算法的维修结果对比，如表 5 所示，图 2 和图 3 分别为叶子节点 s_9 和 s_{10} 的维修诊断树。

$A_0=\{5\,|f_2,f_4,f_5\}$，$A_1=\{2\,|f_0,f_2\}$，$A_2=\{3\,|f_2,f_4\}$，$A_3=\{1\,|f_0\}$，$A_4=\{1\,|f_2,f_5\}$，$A_5=\{1\,|f_4,f_5\}$，$A_6=\{2\,|f_2,f_4,f_5\}$。叶子节点为 A_3、A_4、A_5 和 A_6，分别表示维修 f_5 f_2f_5 f_4f_5 和 $f_2f_4f_5$。虽然只有 4 个多故障被维修，但 $f_2f_4f_5$ 维修过程中包含 f_2f_4，因此 5 个故障组合被维修策略所包含。

$B_0=\{16\,|f_1,f_2,f_3,f_4,f_5\}$，$B_1=\{2\,|f_0,f_2\}$，$B_2=\{14\,|f_1,f_2,f_4\}$，$B_3=\{1\,|f_0\}$，$B_4=\{1\,|f_2\}$，$B_5=\{6\,|f_0,f_2\}$，$B_6=\{8\,|f_0,f_2,f_4\}$，$B_7=\{2\,|f_0\}$，$B_8=\{4\,|f_0,f_2\}$，$B_9=\{3\,|f_0\}$，$B_{10}=\{5\,|f_0,f_2\}$。叶子节点为 B_3、B_4、B_7、B_8、B_9 和 B_{10}，分别表示 f_3 f_5、$f_2f_3f_5$ $f_1f_3f_5$ $f_1f_2f_3f_5$ $f_1f_3f_4f_5$ 和 $f_1f_2f_3f_4f_5$ 的故障组合被维修。至此，16 个故障组合被维修策略所包含。

由表 5 可知，总共有 31 个故障组合被维修，包含了多故障假设下所有的故障总数，所有叶子节点的误修率均在 0.3% 以下，但 Sure3 算法中有叶子节点的误修率高达 12.95%。

4.2 惯导二次电源实例

表 6 为某型惯导二次电源的 D 矩阵，共有 9 个故障模式(F_1 ~ F_9)，无故障 F_0 概率为 0.74。多故障假设下将会有 511 个故障模式。

现用 MFDMSTS 对其进行诊断维修，可得到 41 个可隔离故障，即诊断树有 42 个叶子结点(限于篇幅，省略故障诊断树和叶子结点的维修树)。

北航学报 赠阅

表 5　MFDMSTS 与 Sure3 的维修策略对比

Table 5　Comparison of maintenance strategy between MFDMSTS and Sure3

多故障状态	组合数	s_i 中的组合故障	MFDMSTS		Sure3	
			维修策略	误修率/%	维修策略	误修率/%
s_1, s_2, s_3, s_4, s_5, s_6, s_7	1	f_1, f_2, f_3, f_4, f_1f_4, f_2f_3	直接维修检测的故障集	0	直接维修检的故障集	0
s_8	3	f_3f_4, f_1f_3, $f_1f_3f_4$	维修 f_3f_4,测试,维修 f_1,结束	0.12	维修 f_3,测试,维修 f_1,测试,维修 f_4	0
s_9	5	f_5, f_4f_5, f_2f_5, f_2f_4, $f_2f_4f_5$	图 2	0.15	维修 $f_2f_4f_5$	12.95
s_{10}	16	f_3f_5, f_1f_5, $f_3f_4f_5$, …, $f_1f_2f_3f_4f_5$	图 3	0.26	维修 $f_1f_2f_3f_4f_5$	1.57

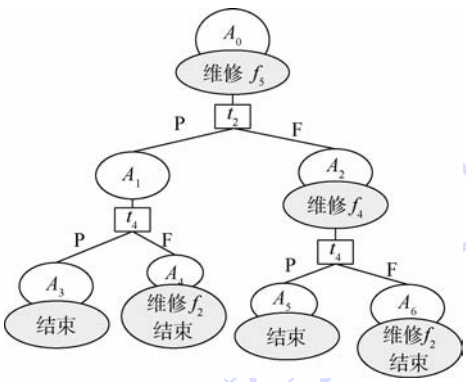


图 2　叶子节点 s_9 的维修诊断树

Fig. 2　Maintenance diagnostic tree of leaf node s_9

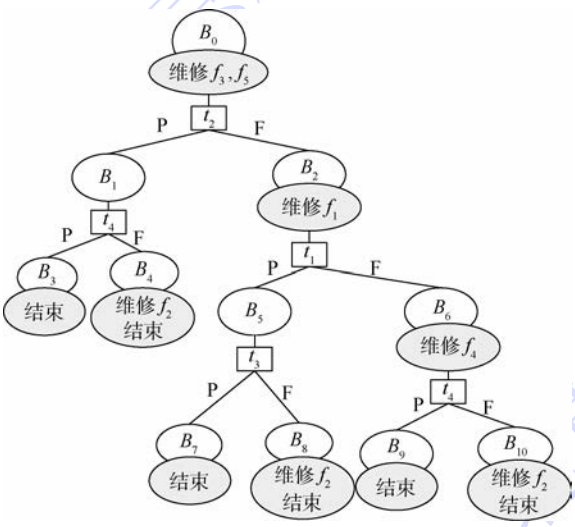


图 3　叶子节点 s_{10} 的维修诊断树

Fig. 3　Maintenance diagnostic tree of leaf node s_{10}

同时应用 TEAMS 软件对惯导二次电源进行建模仿真并将两者结果进行对比(其中单故障算法为 Rollout 算法)。假设 $c_j=1$,通过计算惯导二次电源诊断树的 C_{ost}、N_D、FDR 和 FIR 评价诊断结果。由于误修率是对诊断树的叶子节点而言,但是叶子节点过多,因此引入平均误修率($M_{R\text{-}avg}$)对诊断树所有叶子结点的维修树进行评价。

表 6　某型惯导二次电源的 *D* 矩阵

Table 6　*D* matrix for secondary electrical power of an inertial navigation

故障状态	测试						概率
	T_1	T_2	T_3	T_4	T_5	T_6	
F_1	1	1	1	0	1	1	0.063
F_2	0	0	1	0	0	0	0.021
F_3	0	1	0	0	0	0	0.013
F_4	0	0	1	1	0	0	0.037
F_5	0	0	0	0	0	1	0.017
F_6	0	0	0	1	0	0	0.023
F_7	0	0	0	0	1	1	0.034
F_8	1	1	0	0	0	0	0.022
F_9	0	0	1	0	1	0	0.030

$$M_{R\text{-}avg}=\frac{\sum_{i=1}^{z}M_R(F_S^i)}{z} \tag{9}$$

式中:z 为可隔离故障个数。

由表 7 可知,在 C_{ost}、N_D、FDR 和 FIR 方面,MFDMSTS 的结果与 TEAMS 仿真的结果一致;但是 TEAMS 没有维修诊断,因此没有误修率的结果,而 MFDMSTS 的诊断结果中存在 11 组合数为 1 的叶子节点和 12 个组合数为 2 的叶子节点,这些节点的误修率全为 0;同时其余 19 个叶子节点中有 7 个叶子节点的故障概率为 10^{-5},误修率可近似为 0,另外有 10 个叶子节点的误修率约等于 0,因此总体的误修率比较小,约为 0.012 8%。

在上述运行环境下该型惯导二次电源的多故障转化为单故障的计算时间为 0.115 0 s。

表 7　MFDMSTS 与 TEAMS 仿真分析结果对比

Table 7　Comparison of analysis results between MFDMSTS and TEAMS simulation

算法	C_{ost}	N_D	FDR/%	FIR/%	$M_{R\text{-}avg}$/%
MFDMSTS	4.22	5.45	100	82.69	0.012 8
TEAMS	4.22	5.45	100	82.69	

北航学报 赠阅

5 结 论

1）基于 $\boldsymbol{D}$ 矩阵元素 d_{ij} 的含义、有限的测试，提出了可隔离故障、隔离测试集定义，然后基于隔离测试集和多故障的规律，提出了 MFDMSTS。

2）MFDMSTS 在故障诊断阶段简化了故障之间的关系如隐含故障、伪故障等，获得了最优的多故障诊断树，与 Sure3 算法相比能有效降低平均诊断费用和平均诊断步数；该策略的诊断结果与 TEAMS 软件仿真的结果一致。

MFDMSTS 的故障维修阶段是基于叶子节点的内部故障组合以及概率提出的细致维修策略，在维修范围包含所有可能的故障组合基础上，大幅降低了故障的误修率。

3）本文方法可将大量的多故障转化为数目较少的单故障，被计算处理的故障数目得到了减少，通过对转化的单故障进行故障诊断，并对相应的叶子节点维修，故障维修的精度得以提高。

参考文献（References）

[1] 张可，周东华，柴毅. 复合故障诊断技术综述[J]. 控制理论与应用，2015，32(9)：1143-1157.
ZHANG K，ZHOU D H，CHAI Y. Review of multiple fault diagnosis methods[J]. Control Theory & Applications，2015，32(9)：1143-1157(in Chinese).

[2] 连可，王厚军，龙兵. 基于 SVM 的模拟电子系统多故障诊断研究[J]. 仪器仪表学报，2007，28(6)：1029-1034.
LIAN K，WANG H J，LONG B. Study on SVM based analog electronic system multiple fault diagnosis[J]. Chinese Journal of Scientific Instrument，2007，28(6)：1029-1034(in Chinese).

[3] 杨鹏，邱静，刘冠军. 多故障诊断策略优化生成技术研究[J]. 兵工学报，2008，29(11)：1379-1383.
YANG P，QIU J，LIU G J. Research on optimization generation technology for multiple fault diagnostic strategy[J]. Acta Armamentarii，2008，29(11)：1379-1383(in Chinese).

[4] 赵亮亮，肖明清，盛晟，等. 基于多色集与概念格的多故障推理研究[J]. 仪器仪表学报，2014，35(12)：2690-2699.
ZHAO L L，XIAO M Q，SHENG S，et al. Study on reasoning of multi-failure based on polychromatic sets and concept lattice[J]. Chinese Journal of Scientific Instrument，2014，35(12)：2690-2699(in Chinese).

[5] KUNDU S，CHATTOPADHYAY S，SENGUPTA I，et al. Multiple fault diagnosis based on multiple fault simulation using particle swarm optimization[C]// 2011 24th Annual Conference on VLSI Design. Piscataway，NJ：IEEE Press，2011：364-369.

[6] ISSURY I，HENRY D，CHARBONNEL C，et al. A Boolean algebraic-based solution for multiple fault diagnosis：Application to a spatial mission[J]. Aerospace Science and Technology，2013，28(1)：214-226.

[7] BARTYŚ M. Multiple fault isolation algorithm based on binary diagnostic matrix[J].Advances in Intelligent Systems and Computing，2014，230：441-452.

[8] BARTYŚ M. Diagnosing multiple faults with the dynamic binary matrix[J]. IFAC-Papers Online，2015，48(21)：1297-1302.

[9] BARTYŚ M. The weak isolability of the structure of binary residuals of multiple faults[J]. Advances in Intelligent Systems and Computing，2016，393：3-11.

[10] SHAKERI M，RAGHAVAN V，PATTIPATI K R，et al. Sequential testing algorithms for multiple fault diagnosis[J]. IEEE Transactions on Systems，Man，and Cybernetics，Part A：Systems and Humans，2000，30(1)：1-14.

[11] 王子玲，许爱强，王文双，等. 基于扩展单故障策略的多故障诊断算法[J]. 海军航空工程学院学报，2009，24(6)：695-698.
WANG Z L，XU A Q，WANG W S，et al. An extended single-fault algorithm for multiple fault diagnosis[J]. Journal of Naval Aeronautical and Astronautical University，2009，24(6)：695-698(in Chinese).

[12] YANG C L，YAN J H，LONG B，et al. A novel test optimizing algorithm for sequential fault diagnosis[J]. Microelectronics Journal，2014，45(6)：719-722.

[13] PATTIPATI K R，ALEXANDRISIS M G. Application of heuristic search and information theory to sequential fault diagnosis[J]. IEEE Transactions on Systems，Man，and Cybernetics，1990，20(4)：872-887.

[14] SHEPPARD J W，BUTCHER S G W. A formal analysis of fault diagnosis with $\boldsymbol{D}$-matrices[J]. Journal of Electronic Testing：Theory and Applications，2007，23(4)：309-322.

[15] TSAI Y T，HSU Y Y. A study of function-based diagnosis strategy and testability analysis for a system[J]. Proceedings of the Institution of Mechanical Engineers，Part C：Journal of Mechanical Engineering Science，2012，226(1)：273-282.

[16] 石君友. 测试性设计分析与验证[M]. 北京：国防工业出版社，2011.
SHI J Y. Testability design analysis and verification[M]. Beijing：National Defense Industry Press，2011(in Chinese).

作者简介：

田恒 男，博士研究生。主要研究方向：故障诊断与测试性分析。

段富海 男，教授，博士生导师。主要研究方向：可靠性与测试性的设计分析。

北航学报 赠阅

Multiple fault diagnosis and maintenance strategy based on testability *D* matrix

TIAN Heng[1], DUAN Fuhai[1,*], FAN Liang[2], SANG Yong[1], SHI Meng[2]

(1. School of Mechanical Engineering, Dalian University of Technology, Dalian 116024, China;
2. Aviation Key Laboratory of Science and Technology on Inertia, AVIC Xi'an Flight Automatic Control Research Institute, Xi'an 710065, China)

Abstract: In view of the complicated multiple fault diagnosis problems and difficulties in the realization of multiple fault direct processing algorithm, on the basis of testability ***D*** matrix and single fault algorithm, a novel multiple fault diagnosis and maintenance strategy based on translating into single fault (MFDMSTS) is introduced. Firstly, under the assumption of multiple faults, disjunctive operation is introduced, and isolable single fault and isolable multiple faults are defined. According to the definitions, single fault, which composes a new ***D*** matrix with test set, is translated into from multiple faults. Secondly, single fault diagnosis algorithm is used to process the new ***D*** matrix, and the optimal diagnostic tree is obtained. Finally, a further multiple fault diagnosis and maintenance strategy is proposed for different leaf nodes in the diagnostic tree. Experimental verification demonstrates that MFDMSTS is an effective way to reduce average diagnostic cost and average diagnostic steps, and to greatly reduce the rate of maintenance error.

Keywords: testability *D* matrix; multiple fault diagnosis; disjunctive operation; translating multiple faults into single fault; maintenance strategy

北京航空航天大学学报

Received: 2017-04-13; **Accepted**: 2017-05-19; **Published online**: 2017-06-30 09:02
URL: kns.cnki.net/kcms/detail/11.2625.V.20170630.0902.004.html
Foundation item: Aeronautical Science Foundation of China (20150863003)
* **Corresponding author**. E-mail: duanfh@dlut.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2017.0263

基于拉普拉斯变换的空间目标碰撞概率计算方法

霍俞蓉[1], 李智[2,*], 韩蕾[2]
(1. 装备学院 航天装备系, 北京 101416; 2. 装备学院 航天指挥系, 北京 101416)

摘 要: 碰撞概率是空间碎片碰撞预警中评判碰撞发生可能性大小的重要依据,对航天器机动规避具有重要意义。基于拉普拉斯变换的碰撞概率计算方法利用拉普拉斯变换以及幂级数的定义,推导了在短时间接近情况下碰撞概率的幂级数表达式,分析了碰撞概率的截断误差并确定了在不同精度要求下的幂级数项数。针对2009年美俄卫星的碰撞事件,将基于拉普拉斯变换的碰撞概率计算结果与Chan方法、Monte Carlo方法的计算结果进行比较,验证了基于拉普拉斯变换方法在计算精度上的优势。

关 键 词: 空间碎片; 碰撞预警; 碰撞概率; 拉普拉斯变换; 幂级数
中图分类号: V412.4
文献标识码: A **文章编号**: 1001-5965(2018)04-0810-10

由于人类的航天活动日益增多,空间中在轨的空间目标数量不断增大,因此航天器与空间碎片的碰撞风险越来越大。为了保证航天任务顺利进行,保证航天员、航天器与空间环境的安全,空间碎片碰撞预警已成为航天任务设计与执行过程中必须进行的重要工作。

碰撞概率是目前国内外广泛使用的判断2个空间目标发生碰撞的可能性大小的重要判据。当前,国内外对碰撞概率方法的理论研究已经相对成熟。文献[1]采用Monte Carlo方法得到了碰撞概率的计算结果;Foster和Estes[2]、Patera[3]和Alfano[4]提出了不同的碰撞概率数值积分模型;Chan[5-6]提出了一种将积分式进行近似的解析方法。

Monte Carlo方法以及Foster和Estes[2]提出的数值方法可以得到精度较高的碰撞概率结果但是计算速度较慢。Chan[5-6]解析方法计算速度快,但精度劣于数值方法。解析方法中比较经典的为Chan方法,根据文献[7]中对该方法的精度分析,在概率密度函数(Probability Density Function,PDF)接近圆分布时,能够较好地近似碰撞概率的真实值。但在PDF分布椭圆的长短轴之比较大时,误差较大,并且由于该方法对积分域进行了近似,导致近似积分模型与原积分模型产生了无法准确衡量的偏差。

针对上述方法中的不足,本文在参考文献[8]的基础上,基于拉普拉斯变换的相关理论,推导了碰撞概率的幂级数表达式;研究了幂级数表达式的截断误差计算方法;讨论了在不同精度需求下幂级数项数的取值;通过对2009年美俄卫星碰撞事件的仿真计算,将基于拉普拉斯变换方法的碰撞概率计算结果与Chan方法、Monte Carlo方法的计算结果进行对比,验证了基于拉普拉斯变换方法在计算精度上的优势。

1 碰撞概率计算的基本方法

1.1 碰撞概率计算假设

碰撞概率的计算不仅涉及到接近时刻(Time

收稿日期: 2017-04-26; **录用日期**: 2017-07-21; **网络出版时间**: 2017-09-14 16:04
网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20170914.1604.003.html
基金项目: 国家"863"计划(2015SQ704102)
* **通信作者**. E-mail: lizhizys000@163.com

引用格式: *霍俞蓉, 李智, 韩蕾. 基于拉普拉斯变换的空间目标碰撞概率计算方法[J]. 北京航空航天大学学报, 2018, 44(4): 810-819. HUO Y R, LI Z, HAN L. Computation method of collision probability of space object based on Laplace transformation [J]. Journal of Beijing University of Aeronautics and Astronautics, 2018, 44(4): 810-819 (in Chinese).*

北航学报 赠阅

of Closet Approach, TCA)和接近距离(Miss Distance, MD),还涉及到轨道的误差协方差信息。由于 2 个空间目标在碰撞时所受的力比较复杂,误差数据会根据目标运动状态的改变发生变化,因此,为了简化碰撞概率的计算过程,得到较准确的碰撞概率值,本文做出以下假设[9-11]:

1)由于 2 个目标的相遇时间非常短,可将相遇期间内 2 个目标的相对运动视为速度恒定的线性运动。

2)在碰撞时,目标的相对速度非常大,并且碰撞时间很短,因此,可以忽略 2 个目标速度的不确定性。

3)2 个目标的位置误差互不关联。

4)在 2 个目标间的相对运动过程中,相遇时间非常短,因此认为误差椭球在相遇期间内保持不变。

5)2 个空间目标均等效为半径已知的球体。

6)由于 2 个目标的误差都是随机的,可以认为 2 个目标的位置误差都服从三维高斯分布。

当 2 个目标间的距离小于其等效半径之和时,被判定为发生碰撞。

1.2　相遇坐标系与位置误差的投影

1.2.1　相遇坐标系

令在 TCA 处,与 2 个目标的相对速度矢量垂直并且经过相对距离矢量的平面为相遇平面。在相遇平面内建立对应的相遇坐标系,即可消去速度方向的误差并且进行碰撞概率计算。相遇坐标系以主目标 O_1 为坐标原点,x_e 轴指向从目标 O_2 在相遇平面内的投影点,z_e 轴指向 2 个目标间相对速度的方向,y_e 轴在相遇平面内垂直于 x_e 轴,相遇坐标系如图 1 所示。

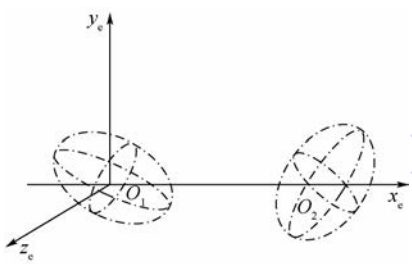


图 1　相遇坐标系示意图

Fig. 1　Schematic diagram of encounter coordinate system

3 个轴的单位矢量可表示为

$$\begin{cases} \boldsymbol{i}_e = \dfrac{\boldsymbol{\rho}_{TCA}}{|\boldsymbol{\rho}_{TCA}|} \\ \boldsymbol{k}_e = \dfrac{\boldsymbol{v}_{rel}}{|\boldsymbol{v}_{rel}|} \\ \boldsymbol{j}_e = \boldsymbol{k}_e \times \boldsymbol{i}_e \end{cases} \tag{1}$$

式中:$\boldsymbol{\rho}_{TCA}$ 和 $\boldsymbol{v}_{rel}$ 分别为 TCA 处目标间的相对距离矢量和相对速度矢量。

1.2.2　位置误差的投影

通过误差分析方法即可得到 2 个空间目标的误差协方差矩阵。将 2 个目标的误差椭球、相对运动状态矢量分别投影到相遇平面内,由于 2 个目标的随机位置矢量相互独立且满足三维高斯分布,2 个目标在相遇平面内的位置矢量满足三维高斯分布,相对位置矢量也是满足高斯分布的随机矢量。因此,在接近时刻处,令 2 个空间目标在相遇平面内的位置矢量分别为 $\boldsymbol{r}_{e1}$ 和 $\boldsymbol{r}_{e2}$,则相遇平面内的目标间相对位置矢量 $\boldsymbol{r}_{erel}$ 的均值为

$$E(\boldsymbol{r}_{erel}) = E(\boldsymbol{r}_{e1} - \boldsymbol{r}_{e2}) = [\boldsymbol{\rho}_{TCA} \quad 0]^T \tag{2}$$

目标间相对位置矢量的误差方差矩阵为

$$\begin{aligned} \mathrm{var}(\boldsymbol{r}_{erel}) &= \mathrm{var}(\boldsymbol{r}_{e1} - \boldsymbol{r}_{e2}) = \\ &\mathrm{var}(\boldsymbol{r}_{e1}) + \mathrm{var}(\boldsymbol{r}_{e2}) - 2\mathrm{cov}(\boldsymbol{r}_{e1}, \boldsymbol{r}_{e2}) = \\ &\mathrm{var}(\boldsymbol{r}_{e1}) + \mathrm{var}(\boldsymbol{r}_{e2}) \end{aligned} \tag{3}$$

通过式(2)和式(3),即可将 2 个目标的位置误差椭球投影到相遇平面内并形成联合误差椭圆域。若将 2 个目标看作是半径分别为 R_1 和 R_2 的均匀球体,则可以根据 2 个目标的半径大小形成联合球体,将联合球体投影到相遇平面内形成联合圆域,圆域的有效半径为 $R = R_1 + R_2$。图 2 为相遇坐标系内的联合圆域和联合误差椭圆域示意图。

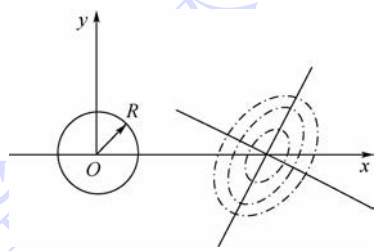


图 2　联合圆域和联合误差椭圆域示意图

Fig. 2　Schematic diagram of combined disk and combined error ellipse

1.3　碰撞概率计算的简化

在定义了相遇坐标系后,碰撞概率的计算就被简化为一个对 PDF 的二重积分过程,PDF 表示为

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y}\exp\left\{-\frac{1}{2}\left[\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2}\right]\right\} \tag{4}$$

式中:μ_x 和 μ_y 分别为联合误差椭圆沿 x 和 y 方向的期望值;σ_x 和 σ_y 分别为联合误差椭圆沿 x 和 y 方向的标准差。

碰撞概率 P_c 的计算式为

北航学报 赠阅

$$P_c = \iint\limits_{x^2+y^2\leqslant R^2} f(x,y)\,\mathrm{d}x\mathrm{d}y \tag{5}$$

2 基于拉普拉斯变换的碰撞概率计算模型

2.1 计算流程

基于拉普拉斯变换的碰撞概率计算主要有3个步骤:

1)将由式(5)表示的二维PDF积分公式重写为$g(\lambda)$,其中$\lambda = R^2$,函数$g(\lambda)$的拉普拉斯变换L_g在闭合域内进行。将L_g进行泰勒展开,然后使用拉普拉斯逆变换将原始的积分函数写成幂级数的形式。

2)计算得到幂级数的相关系数。

3)得到碰撞概率计算结果。

从数值计算的角度考虑,直接计算$g(\lambda)$值是比较困难的。即使$g(\lambda)$的级数展开式是收敛的,但是由于每一项的大小接近、符号相反,会造成在一定精度下,级数的每一项加起来的和为0,结果为无效数字。在该情况下,对于较大数值的λ而言,级数展开是没有意义的[12]。因此,在运算前应设置一个函数ψ,使用$\psi \cdot g$替换g,以避免幂级数包含无用项,并且防止经过展开后的幂级数项在相加后出现高消去现象。

2.2 拉普拉斯变换与幂级数表达式

2.2.1 拉普拉斯变换

在进行拉普拉斯变换之前将碰撞概率改写为$P_c = g(R^2)$的形式,函数$g:\mathbf{R}^+ \mapsto \mathbf{R}^+$定义为

$$g(\lambda) = \frac{1}{2\pi\sigma_x\sigma_y}\iint\limits_{x^2+y^2\leqslant\lambda}\exp\left\{-\frac{1}{2}\left[\frac{(x-\mu_x)^2}{\sigma_x^2}+\frac{(y-\mu_y)^2}{\sigma_y^2}\right]\right\}\mathrm{d}x\mathrm{d}y \tag{6}$$

文献[13]阐述了选择ψ的方法,从拉普拉斯变换的角度考虑,本文使用的函数ψ的形式为$\psi = \mathrm{e}^{p\lambda}$,其中$p$为待定因子。

2.1节因为把原二重积分公式重写为了函数$g(\lambda)$,所以若要求解原积分公式的拉普拉斯变换,首先需求得函数$g(\lambda)$的拉普拉斯变换。由拉普拉斯变换定理可得函数$g(\lambda)$的拉普拉斯变换为

$$L_g(s) = \int_0^{+\infty} g(\lambda)\exp(-s\lambda)\,\mathrm{d}\lambda = \frac{1}{2\pi\sigma_x\sigma_y}\int_0^{+\infty}\iint\limits_{x^2+y^2\leqslant\lambda}\exp\left\{-s\lambda-\frac{1}{2}\left[\frac{(x-\mu_x)^2}{\sigma_x^2}+\frac{(y-\mu_y)^2}{\sigma_y^2}\right]\right\}\mathrm{d}x\mathrm{d}y\mathrm{d}\lambda = \frac{1}{2\pi\sigma_x\sigma_y s}\int_{R^2}\exp\left[-sx^2-\frac{1}{2}\cdot\frac{(x-\mu_x)^2}{\sigma_x^2}\right]\cdot\exp\left[-sy^2-\frac{1}{2}\cdot\frac{(y-\mu_y)^2}{\sigma_y^2}\right]\mathrm{d}x\mathrm{d}y \tag{7}$$

式中:s为复数变量。

根据富比尼定理[14]:

$$\int_{R^2} f(x,y)\,\mathrm{d}x\mathrm{d}y = \int_{-\infty}^{+\infty}\left(\int_{-\infty}^{+\infty} f(x,y)\,\mathrm{d}x\right)\mathrm{d}y \tag{8}$$

以及当$f(x,y) = a(x)b(y)$时有

$$\int_{R^2} f(x,y)\,\mathrm{d}x\mathrm{d}y = \int_{-\infty}^{+\infty} a(x)\,\mathrm{d}x\int_{-\infty}^{+\infty} b(y)\,\mathrm{d}y \tag{9}$$

可得$L_g(s)$的最终表达式为

$$L_g(s) = \frac{\exp\left[-s\left(\frac{\mu_x^2}{2s\sigma_x^2+1}+\frac{\mu_y^2}{2s\sigma_y^2+1}\right)\right]}{s\sqrt{2s\sigma_x^2+1}\sqrt{2s\sigma_y^2+1}} \tag{10}$$

定义函数$h:\mathbf{R}^+ \mapsto \mathbf{R}^+$为如下形式:

$$h(\lambda) = \mathrm{e}^{p\lambda}g(\lambda) \tag{11}$$

式中:$p \in \mathbf{R}^+$。则h的拉普拉斯变换L_h可表示为

$$L_h(s) = \int_0^{+\infty} g(\lambda)\exp(p\lambda)\exp(-s\lambda)\,\mathrm{d}\lambda = \int_0^{+\infty} g(\lambda)\exp(-(s-p)\lambda)\,\mathrm{d}\lambda = L_g(s-p) = \exp\left\{-(s-p)\left[\frac{\mu_x^2}{2(s-p)\sigma_x^2+1}+\frac{\mu_y^2}{2(s-p)\sigma_y^2+1}\right]\right\}\Big/\left[(s-p)\cdot\sqrt{2(s-p)\sigma_x^2+1}\sqrt{2(s-p)\sigma_y^2+1}\right] \tag{12}$$

式中:$|s| > p$,将式(12)转化为

$$L_h(s) = \frac{a_0\exp\left(\frac{Q_x}{s-p\eta}+\frac{Q_y}{s}\right)}{\sqrt{s(s-p\eta)}(s-p)} \tag{13}$$

式中:η、Q_x、Q_y、a_0各项被定义为

$$\begin{cases}\eta = 1-\dfrac{\sigma_y^2}{\sigma_x^2}\\ Q_x = \dfrac{\mu_x^2}{4\sigma_x^4}\\ Q_y = \dfrac{\mu_y^2}{4\sigma_y^4}\\ a_0 = \dfrac{1}{2\sigma_x\sigma_y}\exp\left[-\dfrac{1}{2}\left(\dfrac{\mu_x^2}{\sigma_x^2}+\dfrac{\mu_y^2}{\sigma_y^2}\right)\right]\end{cases} \tag{14}$$

北航学报 赠阅

为了得到积分公式的幂级数表达式，可将 $L_h(s)$ 进行泰勒展开，然后对该泰勒展开式的每一项进行拉普拉斯逆变换，再通过一系列计算得到积分公式的幂级数表达式。

2.2.2　幂级数表达式

令 $L_h(s^{-1})$ 的值定义为

$$L_h(s^{-1}) = \frac{s^2 a_0 \exp\left(sQ_y + \dfrac{Q_x}{1-p\eta s}\right)}{\sqrt{1-p\eta s}(1-ps)} \tag{15}$$

定义 $H_h(s)=s^{-2}L_h(s^{-1})$，则 $H_h(s)$ 的导数为

$$\frac{\mathrm{d}H_h}{\mathrm{d}s} = \frac{a_0\exp\left(sQ_y+\dfrac{Q_x}{1-p\eta s}\right)}{(1-p\eta s)(1-ps)^2}\cdot$$

$$\frac{\left[Q_y+\dfrac{Q_x p\eta}{(1-p\eta s)^2}\right]\sqrt{1-p\eta s}(1-ps)}{(1-p\eta s)(1-ps)^2} -$$

$$\left(\frac{1-ps}{2\sqrt{1-p\eta s}} - p\sqrt{1-p\eta s}\right)a_0\exp\left(sQ_y+\frac{Q_x}{1-p\eta s}\right)\Big/\left[(1-p\eta s)(1-ps)^2\right] =$$

$$H_h(s)\left[Q_y+\frac{Q_x p\eta}{(1-p\eta s)^2}-\frac{1}{2(1-p\eta s)}+\frac{p}{1-ps}\right] = H_h(s)u(s) \tag{16}$$

由 $H_h(0)=a_0$ 以及式(16)可知，$H_h(s)$ 在复平面内是复可微的，根据幂级数的定义[14]，$H_h(s)$ 可展开为幂级数的形式。幂级数的求解复杂，因此使用了 Maple 数学工具(世界上最为通用的数学工具之一，如 MATLAB、Mathematica 等，可在 www.maplesoft.com 下载)对其进行系数的求值。根据 $H_h(s)=s^{-2}L_h(s^{-1})$，可知 $L_h(s)$ 的幂级数形式为

$$L_h(s) = \frac{1}{s^2}\sum_{k=0}^{\infty}a_k\frac{1}{s^k} = \sum_{k=0}^{\infty}\frac{a_k}{s^{k+2}} \tag{17}$$

式中：a_k 为系数。将式(17)中的每一项进行拉普拉斯逆变换，即可得到积分公式的幂级数形式。根据有理函数拉普拉斯变换与逆变换列表中的

$$\frac{1}{s^n} = \frac{1}{(n-1)!}t^{n-1} \tag{18}$$

式中：$n=0,1,\cdots$。碰撞概率积分公式的幂级数形式可表示为

$$P_c = g(\lambda) = h(\lambda)\mathrm{e}^{-p\lambda} = \mathrm{e}^{-p\lambda}L_h^{-1}(s) = \mathrm{e}^{-pR^2}\sum_{k=0}^{+\infty}\frac{a_k R^{2(k+1)}}{(k+1)!} = \mathrm{e}^{-pR^2}\sum_{k=0}^{+\infty}c_k \tag{19}$$

式中：

$$\begin{cases}
c_0 = a_0R^2\\
c_1 = \dfrac{a_0R^4}{2}\left[p\left(\dfrac{\eta}{2}+1\right)+Q_x+Q_y\right]\\
c_2 = \dfrac{a_0R^6}{12}\left\{\left[p\left(\dfrac{\eta}{2}+1\right)+Q_x+Q_y\right]^2+p^2\left(\dfrac{\eta^2}{2}+1\right)+2p\eta Q_x\right\}\\
c_3 = \dfrac{3a_0R^8}{144}\left\{\left[p\left(\dfrac{\eta}{2}+1\right)+Q_x+Q_y\right]\left[p^2\left(\dfrac{\eta^2}{2}+1\right)+2p\eta Q_x\right]\right\}+\dfrac{a_0R^8}{144}\left\{\left[p\left(\dfrac{\eta}{2}+1\right)+Q_x+Q_y\right]^3+2\left[p^3\left(\dfrac{\eta^3}{2}+1\right)+3p^2\eta^2Q_x\right]\right\}\\
c_{k+4} = -\dfrac{R^8p^3\eta^2Q_y}{(k+2)(k+3)(k+4)^2(k+5)}c_k+\dfrac{R^6p^2\eta\left[p\eta\left(k+\dfrac{5}{2}\right)+2Q_y\left(\dfrac{\eta}{2}+1\right)\right]}{(k+3)(k+4)^2(k+5)}c_{k+1}-\\
\quad R^4p\left[p\eta\left(\dfrac{\eta}{2}+1\right)(2k+5)+\eta\left(2Q_y+\dfrac{3p}{2}\right)+Q_x+Q_y\right]c_{k+2}\Big/\left[(k+4)^2(k+5)\right]+\\
\quad R^2\left[p(2\eta+1)(k+3)+p\left(\dfrac{\eta}{2}+1\right)+Q_x+Q_y\right]c_{k+3}\Big/\left[(k+4)(k+5)\right]
\end{cases} \tag{20}$$

根据文献[13]中级数展开的定义以及文献[8]，函数 ψ 中的因子 p 可表示为

$$p = \frac{1}{\sigma_x^2} \tag{21}$$

3　截断误差与幂级数项数的确定

3.1　截断误差

假设取幂级数的前 n 项和作为碰撞概率的值，那么就会存在截断误差。令幂级数前 n 项的和表示为 P_c'，则截断误差 S_n 为

$$S_n = P_c - P_c' = \mathrm{e}^{-pR^2}\sum_{k=n}^{+\infty}\frac{a_k R^{2(k+1)}}{(k+1)!} \tag{22}$$

从式(22)可看出，截断误差在 n 以及 R 确定的情况下只与 a_k 有关。假设有 2 个变量 $\bar{a}_k$ 以及

北航学报 赠阅

$\tilde{a}_k$ 分别为

$$\begin{cases}\bar{a}_k = a_0 p^k \left(1 + \dfrac{\eta}{2} + \dfrac{Q_x + Q_y}{p}\right)^k \\ \tilde{a}_k = a_0 p^k\end{cases} \tag{23}$$

通过计算可以得到以下关系：

$$\frac{a_k}{\tilde{a}_k} \geqslant 1 \geqslant \frac{a_k}{\bar{a}_k} \tag{24}$$

根据式(20)和式(24)可得 $\bar{a}_k$、$\tilde{a}_k$ 和 a_k 的关系为 $\bar{a}_k \geqslant a_k \geqslant \tilde{a}_k$，因此截断误差的取值为

$$\begin{cases} S_n \geqslant \dfrac{a_0 \mathrm{e}^{-pR^2} (pR^2)^{n+1}}{p(n+1)!} \\ S_n \leqslant \dfrac{a_0 \left[R^2 p\left(1 + \dfrac{\eta}{2} + Q_x + Q_y\right)\right]^{n+1}}{\left[p\left(1 + \dfrac{\eta}{2} + \dfrac{Q_x + Q_y}{p}\right)\right](n+1)!} \cdot \\ \qquad \dfrac{\exp\left(R^2\left(\dfrac{p\eta}{2} + Q_x + Q_y\right)\right)}{\left[p\left(1 + \dfrac{\eta}{2} + \dfrac{Q_x + Q_y}{p}\right)\right](n+1)!} = \hat{S} \end{cases} \tag{25}$$

3.2 幂级数项数的确定

国际上通用的规避机动概率黄色门限值为 10^{-5}、红色门限值为 10^{-4}[15]。在计算中，可以将碰撞概率的门限值设为 D_d，为了保证碰撞概率的计算精度至少保持在$\lceil -\lg D_\mathrm{d} \rceil$位，截断误差应小于或等于门限值，即 $\hat{S} \leqslant D_\mathrm{d}$。显然，上述需求依赖于幂级数项数 n 的值。

将截断误差 $\hat{S}$ 写为

$$\hat{S} = \frac{a_0 \mathrm{e}^{-pR^2} (pR^2)^{n+1}}{p(n+1)!} \cdot \frac{\left(1 + \dfrac{\eta}{2} + \dfrac{Q_x + Q_y}{p}\right)^{n+1}}{1 + \dfrac{\eta}{2} + \dfrac{Q_x + Q_y}{p}} \cdot \frac{\exp\left(pR^2\left(1 + \dfrac{\eta}{2} + \dfrac{Q_x + Q_y}{p}\right)\right)}{1 + \dfrac{\eta}{2} + \dfrac{Q_x + Q_y}{p}} = \frac{a_0 \mathrm{e}^{-pR^2} (pR^2)^{n+1}}{p(n+1)!} \left(1 + \frac{\eta}{2} + \frac{Q_x + Q_y}{p}\right)^n \cdot \exp\left(pR^2\left(1 + \frac{\eta}{2} + \frac{Q_x + Q_y}{p}\right)\right) \tag{26}$$

由于斯特林公式(Stirling's Approximation)对 $n!$ 的定义为

$$n! \approx \sqrt{2\pi n}\left(\frac{n}{\mathrm{e}}\right)^n \tag{27}$$

则 $\hat{S}$ 可写为

$$\hat{S} \approx \left[\frac{\mathrm{e}pR^2\left(1 + \dfrac{\eta}{2} + \dfrac{Q_x + Q_y}{p}\right)}{n+1}\right]^{n+1} \cdot \frac{a_0 \exp\left(pR^2\left(\dfrac{\eta}{2} + \dfrac{Q_x + Q_y}{p}\right)\right)}{p\sqrt{2\pi(n+1)}\left(1 + \dfrac{\eta}{2} + \dfrac{Q_x + Q_y}{p}\right)} < \left[\mathrm{e}pR^2\left(1 + \frac{\eta}{2} + \frac{Q_x + Q_y}{p}\right)/2\right]^{n+1} \cdot \frac{a_0 \exp\left(pR^2\left(\dfrac{\eta}{2} + \dfrac{Q_x + Q_y}{p}\right)\right)}{p\sqrt{2\pi(n+1)}\left(1 + \dfrac{\eta}{2} + \dfrac{Q_x + Q_y}{p}\right)} \tag{28}$$

令 $\hat{S} \leqslant D_\mathrm{d}$，通过计算可知

$$\mathrm{e}pR^2\left(1 + \frac{\eta}{2} + \frac{Q_x + Q_y}{p}\right) < 1 \tag{29}$$

则可得到如下关系式：

$$\left(\frac{1}{2}\right)^{n+1} a_0 \exp\left(pR^2\left(\frac{\eta}{2} + \frac{Q_x + Q_y}{p}\right)\right) \Big/ \left[p\sqrt{2\pi(n+1)}\left(1 + \frac{\eta}{2} + \frac{Q_x + Q_y}{p}\right)\right] \leqslant D_\mathrm{d} \Rightarrow$$
$$2^{n+1} \geqslant a_0 \exp\left(pR^2\left(\frac{\eta}{2} + \frac{Q_x + Q_y}{p}\right)\right) \Big/ \left[D_\mathrm{d} p\sqrt{2\pi(n+1)}\left(1 + \frac{\eta}{2} + \frac{Q_x + Q_y}{p}\right)\right] \Rightarrow$$
$$n + 1 \geqslant \mathrm{lb}\left(a_0 \exp\left(pR^2\left(\frac{\eta}{2} + \frac{Q_x + Q_y}{p}\right)\right) \Big/ \left[D_\mathrm{d} p\sqrt{2\pi(n+1)}\left(1 + \frac{\eta}{2} + \frac{Q_x + Q_y}{p}\right)\right]\right) \tag{30}$$

令 $n + 1 = n_1$，由于 $n_1 \geqslant 1$，则式(30)中的最后一个不等式可写为

$$\mathrm{lb}\frac{a_0 \exp(pR^2 A)}{D_\mathrm{d} p\sqrt{2\pi n_1}(1 + A)} \geqslant 1 \Rightarrow \frac{a_0 \exp(pR^2 A)}{D_\mathrm{d} p\sqrt{2\pi}(1 + A)} \cdot \frac{1}{\sqrt{n_1}} \geqslant 2 \Rightarrow \sqrt{n_1} \leqslant \frac{a_0 \exp(pR^2 A)}{2D_\mathrm{d} p\sqrt{2\pi}(1 + A)} \Rightarrow n_1 \leqslant \left[\frac{a_0 \exp(pR^2 A)}{2D_\mathrm{d} p\sqrt{2\pi}(1 + A)}\right]^2 \tag{31}$$

式中：$A = \eta/2 + (Q_x + Q_y)/p$。通过计算，幂级数项数 n 的表达式为

$$\begin{cases} n_1 = 2\left\lceil\left(\dfrac{a_0 \exp(pR^2 A)}{2D_\mathrm{d} p\sqrt{2\pi}(1 + A)}\right)^2\right\rceil \\ n_2 = \left\lceil \mathrm{lb}\dfrac{a_0 \exp(pR^2 A)}{D_\mathrm{d} p\sqrt{2\pi n_1}(1 + A)}\right\rceil \\ n = \max\{n_1, n_2\} - 1 \end{cases} \tag{32}$$

若将 D_d 的值取为 $10^{-20} \sim 10^{0}$，则 n 的计算结果如图 3 所示。

北航学报 赠阅

若将 2 个目标的联合有效半径 R 的值取设为 5、10、15、50、100、200、500 和 1000 m,则 n 在不同联合有效半径以及不同概率门限值下的计算结果如图 4 所示。

从图 3 和图 4 结果可以看出,使用基于拉普拉斯变换的碰撞概率方法可以根据所需门限值(也作为计算精度的需求)的不同,得到不同的幂级数项数。从图 4 可知,当 R 在 1 000 m 以内、计算精度需求为 0 ~ ⌈ $-\lg 10^{-5}$ ⌉时,幂级数项数 n 不超过 7,随着计算精度的需求增大,项数逐渐增大。当 R 为 1 000 m、计算精度需求为 0 ~ ⌈ $-\lg 10^{-5}$ ⌉时,幂级数项数 n 不超过 13。国际上通用的规避机动概率黄色门限值为 10^{-5},碰撞概率的计算精度可设置为⌈ $-\lg 10^{-5}$ ⌉,因此,基于拉普拉斯的变换方法明显减少了计算量,同时保证了碰撞概率的计算精度。

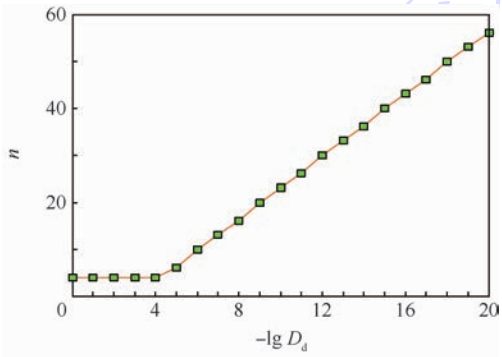


图 3　幂级数项数随概率门限值的变化

Fig. 3　Variation of number of terms of power series with probability threshold

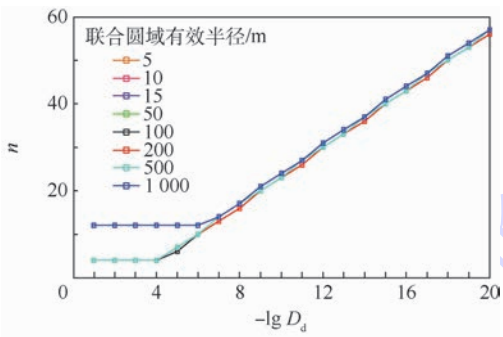


图 4　幂级数项数随概率门限值以及联合圆域有效半径的变化

Fig. 4　Variation of number of terms of power series with probability threshold and radius of combined disk

4　碰撞实例仿真验证

本文以 2009 年 2 月 10 日,美俄卫星发生的碰撞实例作为仿真算例。其中,美国卫星为铱星公司的 Iridium-33(ID:24946),俄罗斯失效卫星为 Cosmos-2251(ID:22675)。通过接近分析得到,TCA[5] 为 2009 年 2 月 10 日 16:55:59.8 094 031,MD[5] 为 0.578 479 931 6 km。美俄卫星在 TCA 的运动状态如表 1 和表 2 所示。

表 1　主、从目标的位置参数

Table 1　Position parameters of primary and secondary object　　km

目　标	x	y	z
主目标	-1 457.353 760	1 589.546 912	6 814.195 621
从目标	-1 457.572 590	1 589.024 470	6 814.313 123

表 2　主、从目标的速度参数

Table 2　Velocity parameters of primary and secondary object　　km/s

目　标	v_x	v_y	v_z
主目标	-7.001 700	-2.439 510	-0.926 295
从目标	3.578 697	-6.172 823	2.200 328

4.1　不同误差椭球分布下的碰撞概率变化

令 2 个空间目标的半径都为 5 m,假设 2 个目标的位置误差椭球大小相等,误差椭球沿星基坐标系 RSW 的坐标轴 R、S、W 方向的比值 $\sigma_R:\sigma_S:\sigma_W$ 不同时,Chan 方法与基于拉普拉斯变换方法计算出的概率随 σ_S 变化的情况如图 5 所示。

从图 5 可以看出,误差值对碰撞概率的计算结果影响较大。使用基于拉普拉斯变换方法的计算结果曲线与 Chan 方法的计算结果曲线基本重合,初步证明了基于拉普拉斯变换的碰撞概率计算方法是正确的并且符合初步的碰撞预警的精度需求。

4.2　不同方法的碰撞概率计算结果比较

通过误差分析,美俄卫星 R、S、W 方向的位置误差标准差如表 3 所示。

令 2 个目标的半径都为 5 m,通过计算可得,在 TCA 处,2 个目标在相遇平面内的 3σ 联合误差椭圆和碰撞圆域,如图 6 所示。

对于 Chan 方法,取无穷级数的前 10 项以近似 PDF 积分式的值 P_k,计算过程中,碰撞概率无穷级数形式的前 10 项结果如表 4 所示。

当使用基于拉普拉斯变换的方法对 2 个目标进行碰撞概率计算时,将幂级数的项数取值为 $k=1,2,\cdots,10$,则得到的碰撞概率值 P_k 和截断误差 S_k 如表 5 所示。

当无穷级数和幂级数项数分别取 10 时,Monte Carlo 方法、Chan 方法以及使用基于拉普拉斯变换方法的碰撞概率计算结果如表 6 所示。

从表 6 可以看出,基于拉普拉斯变换方法与 Chan

北航学报 赠阅

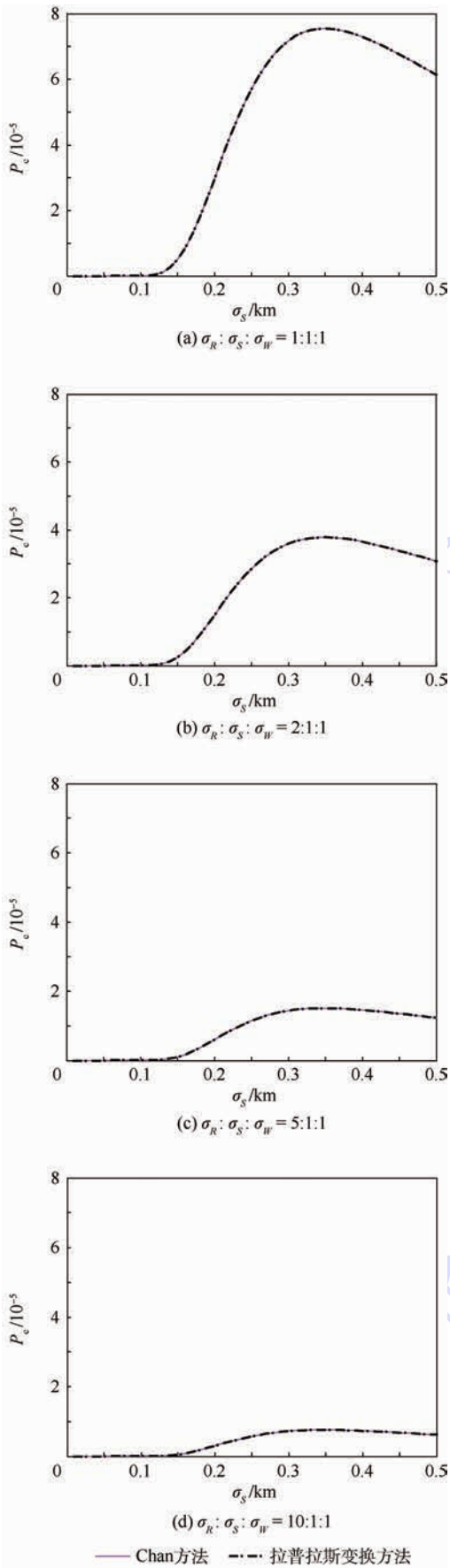


图 5 不同球形分布下碰撞概率的变化趋势

Fig. 5 Change trend of collision probability under different spherical distribution

表 3 主、从目标在 *RSW* 坐标系下的位置误差标准差

Table 3 Standard deviation of position error of primary and secondary object in *RSW* coordinate system

目 标	σ_R/km	σ_S/km	σ_W/km
主目标	0.231 207	0.206 188 5	0.071 975
从目标	0.036 323 4	0.410 206 9	0.034 114

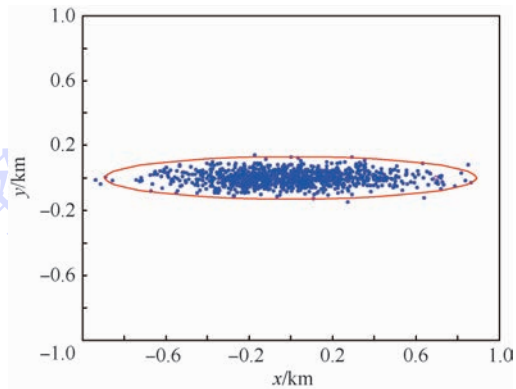


图 6 3σ 联合误差椭圆和碰撞圆域

Fig. 6 3σ combined error ellipse and collision disk

表 4 Chan 方法下 P_k 的计算结果

Table 4 Calculation results of P_k by Chan method

k	$P_k/10^{-4}$
1	1.817 439 461 411 834
2	1.828 466 082 330 511
3	1.828 488 389 607 843
4	1.828 488 412 175 072
5	1.828 488 412 188 786
6	1.828 488 412 188 801
7	1.828 488 412 188 806
8	1.828 488 412 188 808
9	1.828 488 412 188 809
10	1.828 488 412 188 809

表 5 基于拉普拉斯变换方法下 P_k 和 S_k 的计算结果

Table 5 Calculation results of P_k and S_k based on Laplace transformation method

k	$P_k/10^{-4}$	S_k
1	1.772 573 705 611 427	$8.291\,431\,627\,403\,867\times10^{-7}$
2	1.815 996 175 328 694	$2.031\,139\,453\,405\,921\times10^{-8}$
3	1.817 048 378 909 693	$3.317\,101\,046\,456\,115\times10^{-10}$
4	1.817 052 720 190 974	$4.062\,926\,108\,033\,688\times10^{-12}$
5	1.817 053 331 566 511	$3.981\,155\,431\,361\,421\times10^{-14}$
6	1.817 053 372 236 674	$3.250\,858\,729\,893\,972\times10^{-16}$
7	1.817 053 375 065 217	$2.275\,308\,404\,578\,399\times10^{-18}$
8	1.817 053 375 264 924	$1.393\,447\,138\,227\,291\times10^{-20}$
9	1.817 053 375 279 203	$7.585\,569\,687\,077\,649\times10^{-23}$
10	1.817 053 375 280 234	$3.716\,451\,044\,970\,535\times10^{-25}$

北航学报 赠阅

方法、Monte Carlo 方法所计算得到的 P_c 都大于红色门限值,说明在该接近点发生碰撞的可能性非常大,进一步证明了基于拉普拉斯变换的碰撞概率计算方法是正确有效的。表 7 列出了 Chan 方法、基于拉普拉斯变换方法所计算的碰撞概率与 Monte Carlo 方法所计算的碰撞概率的相对误差。

基于拉普拉斯变换方法的碰撞概率与 Chan 方法、Monte Carlo 方法的相对误差分别为 0.625 4%、0.078 9%,而 Chan 方法与 Monte Carlo 方法的相对误差为 0.549 9%。结果表明,基于拉普拉斯变换的碰撞概率计算精度不仅能够满足风险评估的要求,而且在项数取 10 时比 Chan 方法的计算精度高。

表 6　Monte Carlo 方法、Chan 方法、基于拉普拉斯变换方法的碰撞概率计算结果

Table 6　Collision probability calculation results of Monte Carlo, Chan and based on Laplace transformation methods

方　法	碰撞概率
Monte Carlo	0.000 181 848 841 218 880
Chan	0.000 182 848 841 218 880 9
拉普拉斯变换	0.000 181 705 337 528 023 4

表 7　Chan 方法、基于拉普拉斯变换方法与 Monte Carlo 方法的碰撞概率相对误差

Table 7　Relative error of collision probability between Monte Carlo method and Chan, based on Laplace transformation method

方　法	相对误差/%
Chan	0.549 9
拉普拉斯变换	0.078 9

同时,随着幂级数项数 n 的取值增大,基于拉普拉斯变换方法计算的碰撞概率的截断误差随之减小。当 n 的取值小于 5 时,每一个项数所对应的碰撞概率结果相差较大,而在 n 大于 5 时,其结果基本没有变化。因此,对于实例 1,幂级数项数 n 取 5,即可满足初步的碰撞预警要求。需要注意的是,由于 2 种方法的模型不同,当 Chan 方法和拉普拉斯变换方法的级数同时取首项时,Chan 方法的计算结果精度要比拉普拉斯方法高,但却不满足碰撞风险评估的要求,所以为了尽可能的提高计算精度,项数 n 的取值应选择较大的值。

4.3　计算精度需求对幂级数项数的影响分析

为了分析计算精度需求对幂级数项数 n 及碰撞概率的影响,现将$\lceil -\lg D_d \rceil$取 4、5、6 三个值,计算对应的项数 n、碰撞概率以及截断误差。表 8 列出了$\lceil -\lg D_d \rceil$取值不同时,对应的幂级数项数 n 的取值、相应的截断误差以及碰撞概率结果。

从表 8 可以看出,随着计算精度要求不断提高,幂级数项数 n 的值不断增大,截断误差不断减小,碰撞概率值也更加精确。可以预见,只需将幂级数项数设置为较大的值,碰撞概率的计算结果将会足够精确,但是为了提高计算速度,可根据黄色门限值大小来确定精度需求,从而计算幂级数项数的值。

表 8　精确位数 $-\lg D_d$ 不同时所需的幂级数项数、碰撞概率和截断误差

Table 8　Number of terms of power series, collision probability and truncation error with different exact digits $-\lg D_d$

$-\lg D_d$	n	P_c	S_n
4	3	0.000 181 704 837 890 969 3	$9.122\,027\,877\,754\,318 \times 10^{-10}$
5	6	0.000 181 705 337 223 667 4	$5.108\,492\,289\,833\,384 \times 10^{-16}$
6	9	0.000 181 705 337 527 920 3	$8.344\,126\,655\,785\,414 \times 10^{-23}$

4.4　基于拉普拉斯变换方法与 Chan 方法的计算时间比较

Monte Carlo 方法是数值方法之一,由于数值方法没有简化,所以精度高,但是计算速度慢。因此本节只对解析方法中的 Chan 方法和拉普拉斯变换方法的计算时间进行实验比较,以验证拉普拉斯变换方法在计算速度上的优势。

依旧以美俄卫星碰撞实例为算例,以表 1 ~ 表 3 中的运动状态与误差数据作为 2 个目标的参数数据,在实验环境相同的条件下,当幂级数项数 n 取不同的值时,Chan 方法和拉普拉斯变换方法所花费的计算时间比较结果如图 7 所示。

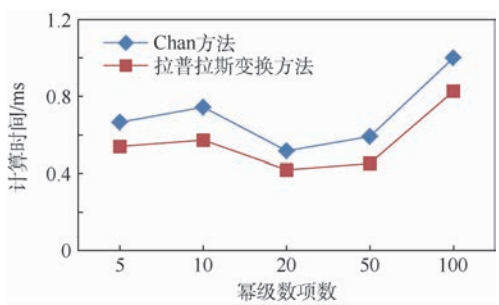


图 7　Chan 方法和拉普拉斯变换方法的计算时间结果

Fig. 7　Results of computation time of Chan and Laplace transformation methods

通过计算,充分证明了基于拉普拉斯的变换方法与 Chan 方法相比,在速度上有所提高。基于拉普拉斯变换的方法是对原积分模型直接进行拉

北航学报 赠阅

普拉斯变换和泰勒展开,从而获得幂级数表达式的碰撞概率计算方法,虽然没有对积分域进行近似,但却是一种解析方法,因此兼具了精度和速度方面的优势。

5 结 论

基于拉普拉斯变换的碰撞概率计算方法是一种解析方法,该方法避免了对二维积分近似过程中的误差,因此,兼具计算精度和速度上的优势。

1)基于拉普拉斯变换的碰撞概率计算方法将概率密度函数的积分公式转换到复平面内进行计算,再通过拉普拉斯逆变换并使用 Maple 数学工具将原积分公式表示为幂级数形式,这使得原本在实数域中计算较复杂的问题变得简单和易理解。

2)通过对幂级数项数的确定,可以在保证计算精度的条件下,提高碰撞概率的计算速度。

3)针对美俄卫星碰撞实例,将基于拉普拉斯变换方法的碰撞概率计算结果与 Chan 方法和高精度的基于 Monte Carlo 数值方法的概率计算结果进行比较。结果表明,基于拉普拉斯变换的碰撞概率计算方法与 Chan 方法相比,能够避免积分域的近似,在计算精度上有所提高,能够满足碰撞预警的精度要求。

参考文献(References)

[1] ALFANO S. Satellite conjunction Monte Carlo analysis[J]. Advances in the Astronautical Sciences,2009,134:2007-2024.

[2] FOSTER J L,ESTES H S. A parametric analysis of orbital debris collision probability and maneuver rate for space vehicles: NASA/JSC-25898[R]. Houston: NASA Johnson Space Flight Center,1992.

[3] PATERA R P. General method for calculating satellite collision probability[J]. Journal of Guidance, Control, and Dynamics, 2001,24(4):716-722.

[4] ALFANO S. Satellite collision probability enhancements[J]. Journal of Guidance,Control,and Dynamics,2006,29(3):588-592.

[5] CHAN F K. Collision probability analysis for earth orbiting satellites[J]. Advances in the Astronautically Sciences, 1997 (96):1033-1048.

[6] CHAN F K. Spacecraft collision probability[M]. El Segundo, CA: Aerospace Press,2008.

[7] 陈磊,韩蕾,白显宗,等. 空间目标轨道力学与误差分析[M]. 长沙:国防科技大学出版社,2010:178-180.
CHEN L,HAN L,BAI X Z,et al. Orbit target orbit mechanics and error analysis[J]. Changsha: National University of Defense Technology Press,2010:178-180(in Chinese).

[8] SERRA R,ARZELIER D,JOLDES M,et al. Fast and accurate computation of orbital collision probability for short-term encounters[J]. Journal of Guidance, Control, and Dynamics, 2016,39(5):1-13.

[9] ALFRIEND K T. AKELLAM R,FRISBEE J,et al. Probability of collision error analysis[J]. Space Debris, 1999, 1(1): 21-35.

[10] AKELLA M R,ALFRIEND K T. Probability of collision between space objects[J]. Journal of Guidance, Control, and Dynamics,2000,23(5):769-772.

[11] ALFANO S. A numerical implementation of spherical object collision probability[J]. Journal of the Astronautical Sciences, 2005,53(1):103-109.

[12] CHEVILLARD S,MEZZAROBBA M. Multiple-precision evaluation of the Airy Ai function with reduced cancellation[C]// 21st IEEE Symposium on Computer Arithmetic (ARITH). Piscataway,NJ:IEEE Press,2013:175-182.

[13] GAWRONSKI W,MÜLLER J,REINHARD M. Reduced cancellation in the evaluation of entire functions and applications to the error function[J]. SIAM Journal on Numerical Analysis, 2007,45(6):2564-2576.

[14] HACKBUSCH W,SCHWARZ H R. Teubner-taschenbuch der mathematik[M]. Berlin:Springer,2013:595.

[15] 李甲龙,熊建宁,许晓丽,等. 碰撞风险评估标准适用性分析[J]. 天文学报,2014,55(5):404-414.
LI J L,XIONG J N,XU X L,et al. A research on adaptability of collision criteria[J]. Acta Astronomica Sinica,2014,55(5): 404-414(in Chinese)

作者简介:

霍俞蓉 女,博士研究生。主要研究方向:空间碎片碰撞预警。

李智 男,博士,教授,博士生导师。主要研究方向:空间态势感知、空间环境。

韩蕾 女,博士,讲师。主要研究方向:航天器飞行动力学。

北航学报 赠阅

Computation method of collision probability of space object based on Laplace transformation

HUO Yurong[1], LI Zhi[2,*], HAN Lei[2]

(1. Department of Space Equipment, Equipment Academy, Beijing 101416, China;
2. Department of Space Command, Equipment Academy, Beijing 101416, China)

Abstract: The collision probability of space object is an important information for judging whether a collision occurs in the space debris collision warning, and is important for the maneuvering avoidance of the spacecraft. Based on the Laplace transformation and the definition of the power series, the collision probability calculation method and the collision probability expression expressed by the power series in the short-term encounter were discussed. The truncation error in the form of power series was determined and the number of terms of power series was analyzed under different precision requirements. The results of collision probability calculation based on Laplace transformation are compared with those of Chan method and Monte Carlo method for 2009 US-Russian satellite collision event. The validity of Laplace transformation method and the advantage of computing accuracy are verified.

Keywords: space debris; collision warning; collision probability; Laplace transformation; power series

北京航空航天大学学报

Received: 2017-04-26; **Accepted**: 2017-07-21; **Published online**: 2017-09-14 16:04
URL: kns.cnki.net/kcms/detail/11.2625.V.20170914.1604.003.html
Foundation item: National High-tech Research and Development Program of China (2015SQ704102)
* **Corresponding author**. E-mail: lizhizys000@163.com

北航学报 赠阅

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2017.0290

月基平台对地观测数据传输链路方案设计及分析

陈国强[1,2], 阮智星[1,*], 郭华东[1], 刘广[1], 丁翼星[1], 张严心[2]

(1. 中国科学院 遥感与数字地球研究所, 北京 100094; 2. 北京交通大学 电子信息工程学院, 北京 100044)

摘 要: 数据传输及处理能力是月基平台构建中的一个重要问题,如何高效准确地传输海量对地观测数据至地球供后续研究是开展月基对地观测的关键环节。通过 STK 和 MATLAB 软件联合仿真,模拟月基平台对地观测数据传输链路,首次提出适用于月基平台的下行链路通信方案:通过构建中纬度地球站、最小间隔经度值为 40°的 2 颗中继卫星组的设计方案,可以最大程度地实现全天候、无时断的信号传输,满足下行链路接收端获取足够强度和低误码率的信息,从而保障月基平台的运行。

关 键 词: 月基平台; 对地观测; 深空通信; 数据传输; 链路

中图分类号: TN958; V443.2

文献标识码: A **文章编号**: 1001-5965(2018)04-0820-08

全下文载

随着中国航天技术和探月工程的不断发展,利用月球作为观测平台开展对地观测的理念,已成为国内外月球开发研究中的前沿思路之一[1-2]。建立良好的对地观测平台有助于更加全面地了解和认识地球,相较于传统的星载平台,月基平台具有更远的观测距离(3.8×10^5 km),能够覆盖更宽的地球表面;具有更大的平台搭建空间(直径 3 460 km),可在月球表面区域集中搭建多个平台,不必如星载平台考虑星间最小角度限制。月球绕地公转周期与其自转周期之间的相等关系,使月球在绕地公转时,总有固定的半个球面始终朝向地球,故相较于多数的星载平台,可以实现全天候、周期性、可变观测视角的对地观测和通信[3]。不同于星载平台存在的光压摄动、地球扁率摄动和大气阻力摄动等问题,月基平台不需要定期校正平台的位置;相较于以人造轨道卫星为基础的星载平台,月基平台的寿命也长于前者,且可以省去与前者之间相互通信的星历校准环节。如上所述,搭建以月球为基础的对地观测平台对高效开展地球宏观现象的科学研究具有重要意义[4],并且在可预见的未来,待技术和理论成熟后,可以将其观测对象转变成除地球以外的其他宇宙深空对象。

月地之间的远距离和月基平台对地观测的海量数据给月基平台的数据传输带来了巨大的挑战[5]。月基平台对地球进行观测和通信时,由于电磁波的传输距离遥远,不可避免会因为太空辐射、太阳风和日冕[6]等因素而产生较大的深空自由空间损耗;且由于传输距离遥远,难以避免发散问题,电磁波的波长越大,发散问题越严重,目标接收到的信号越弱。同时,月基平台为了满足地球宏观现象的观测能力,需要尽可能搭设多源传感器,实现长时间周期的大尺度观测。观测过程将记录较大数据量的遥感影像,这将对月基平台的数据传输能力提出更高的要求。因此,在设计月基平台对地传输链路方案时,需要尽量满足月地充足的数据通信时间和较高的数据传输质量的设计要求,从而保障月基平台的进一步发展。月基平台对地观测的相关研究具有开创性。本文旨在为月基平台对地观测系统建设提供支持出发,

收稿日期: 2017-05-08; **录用日期**: 2017-08-02; **网络出版时间**: 2017-09-26 14:45
网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20170926.1445.002.html
基金项目: 国家自然科学基金(41590852, 41590853, 41501403); 中国科学院前沿科学重点研究计划(QYZDY-SSW-DQC026)
* **通信作者**. E-mail: ruanzx@radi.ac.cn

引用格式: 陈国强, 阮智星, 郭华东, 等. 月基平台对地观测数据传输链路方案设计及分析[J]. 北京航空航天大学学报, 2018, 44(4): 820-827. CHEN G Q, RUAN Z X, GUO H D, et al. Moon-based platform data transmission link scheme design and analysis [J]. Journal of Beijing University of Aeronautics and Astronautics, 2018, 44(4): 820-827 (in Chinese).

北航学报 赠阅

重点满足下行链路接收端能够得到强度更大和误码率更低的有用信号，通过 STK 和 MATLAB 软件联合仿真，构建并讨论可能的月基平台对地观测数据传输方案，结合月基平台需求分析链路性能[7-8]，从而提出适应于月基平台对地的传输方案。

1　数据传输链路中的重要参数

月基平台传输通信链路的设计和卫星通信链路相似，影响传输的因素包括接收端和发射端之间的角度关系和传输过程的相关损耗等。在评价链路性能时，常用链路的信噪比和误码率等参数进行分析。

1.1　发射站相对地球站的仰角和方位角

随着中国航天和通信事业的发展，点波束天线以其灵活性好、增益高和安全保密性强等特点，在各类通信卫星上得到了越来越广泛的应用。由于空间通信的主体相距较远，点波束天线的指向对角度更加敏感，天线指向的微小偏差，可能使点波束完全偏离原覆盖区，从而直接导致通信中断[9-10]。研究地球站与发射站（如卫星等）的几何关系，主要是求算地球站观测发射站的仰角和方位角等参数，以此为星地链路计算提供依据。因此，对于空间数据传输，发射端与接收端的仰角和方位角对信号最终能否被传输至关重要。在月基平台对地观测中，若存在转发中继卫星，则卫星相对地球站的仰角和方位角将影响卫星发射机对地球站的覆盖增益[10]，继而影响所获取信号的强度等参数。如图 1 和图 2 所示，可以算出仰角和方位角。

若已知地球站的经度 λ_P 和纬度 ϕ_P，卫星星下点经度 λ_S 和纬度 ϕ_S，卫星的高度 h，地球站的位置 P，星下点的位置 G，北极方向 N. P，依次可求出地球站到卫星的仰角 E 和方位角 A。

地球站到卫星之间的仰角为

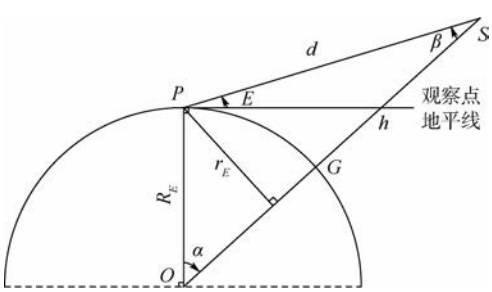


图 1　地球站至卫星的仰角

Fig. 1　Elevation of earth station to satellite

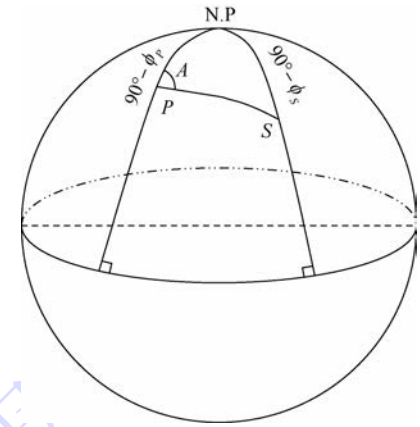


图 2　地球站至卫星的方位角

Fig. 2　Azimuth of earth station to satellite

$$E=\arctan\left(\frac{\cos\alpha-\dfrac{R_E}{R_E+h}}{\sin\alpha}\right) \tag{1}$$

式中：R_E 为地球站点处与地心的距离；α 为卫星星下覆盖区对应地球半中心角，其满足如下关系式：

$$\cos\alpha=\cos(\lambda_P-\lambda_S)\cos\phi_S\cos\phi_P+\sin\phi_S\sin\phi_P \tag{2}$$

地球站到发射端之间的方位角为

$$A=\arctan(\sin(\lambda_P-\lambda_S)\cos\phi_S/(\cos(\lambda_P-\lambda_S)\cos\phi_S\cos\phi_P-\sin\phi_S\sin\phi_P)) \tag{3}$$

若 $\phi_S=0$，则卫星可视为地球静止轨道（Geostationary Earth Orbit，GEO）卫星[11]。

1.2　传输损耗

信号在自由空间传播时，衰减主要来自自由空间的传播损耗 L_S：

$$L_S=\left(\frac{4\pi d}{\lambda}\right)^2=\left(\frac{4\pi d}{c}\right)^2 \tag{4}$$

$$L_S=32.44+20\lg d+20\lg f \tag{5}$$

式中：d 为传输距离，km；f 为载波频率，MHz；λ 为信号波长；c 为信号传播速度，一般取为光速。式(4) L_S 的单位为 W，式(5) L_S 的单位为 dB。

在卫星通信中，除了空间传输损耗之外，还有其他的损耗项，如雨衰、大气吸收损耗和极化损耗等。图 3 为卫星链路中信号传输过程中引发的损耗类型，其中 EIRP 为等效全向辐射功率。

1.3　链路评价参数

1.3.1　信 噪 比

信号的信噪比通常是输出信号的功率与同时输出的噪声功率的比，常用分贝数（dB）表示，设备的信噪比越高表明它产生的噪声越少。一般来说，信噪比越大，混在信号里的噪声越小，信号的质量越高，否则相反[12]。

北航学报 赠阅

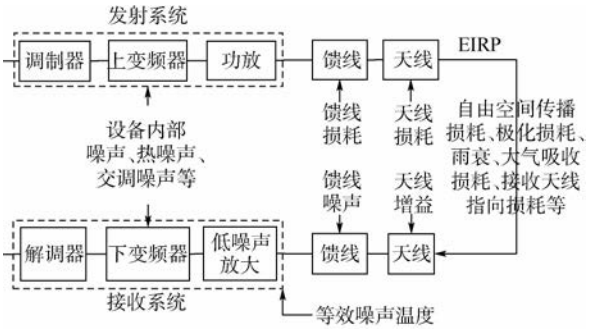


图 3 卫星链路组成环节

Fig. 3 Segments of satellite link

1.3.2 误码率

信号误码率指的是在传输的码元总数中发生差错的码元数所占的比例(平均值),也指信号的差错概率。影响信号误码率的因子很多,通常以通信链路的信噪比大小作为判断误码率高低的指标[12]。对于军事通信下行链路,按照业务类型可分为:电话、传真、数据、图像和综合业务,其对应的误码率分别不应高于 10^{-4}、10^{-4}、10^{-6}、10^{-6}和 10^{-6};对于美国静止轨道卫星 DIRECTV-14,其下行遥测链路误码率不应高于 10^{-6}[12]。当信号误码率为 10^{-6}时,表明在信息传输过程中每十万位信息中仅存在一两位错误,对于实际的卫星传输链路,该数量级可体现其良好稳定的通信性能。因此在后文的中继卫星下行链路信噪比阈值中,可以将其设置成误码率为 10^{-6}时对应的信噪比(Eb/No),此时信噪比一般取为 10 dB。

2 数据传输链路方案设计

目前的月地通信主要面向探月工程相关的需求,包括月球探测数据、地球大尺度影像和月地控制命令等信息的传送,所涉及的数据量较小。月基平台在满足地球宏观现象的观测需求的基础上,数据传输链路设计更加重视月地链接时效性和链路质量。月地之间的数据传输包含2 种方式:①月地之间直接进行数据通信;②在月基平台与地球站之间的下行链路通信中,加入若干中继卫星,以期达到对电磁波信号的降噪和功率增强的处理。月基对地观测通信业务包含月基平台至中继卫星的深空通信操作业务和中继卫星至地球站的近地空间通信业务。在实际运用中,中继卫星通常选为地球静止轨道 GEO 卫星。本文将针对月地链接方式开展链路方案模拟和对比分析,提出适应于月基平台的最优设计思路。

2.1 月基平台数据传输链路基本参数

2.1.1 通信收发端站点位置和相关参数

在月球与地球的相互作用中,地球在月面的投影位置始终以近似椭圆的形状绕月面 0°N,0°W运动。若以上述点近似为地球在月球的固定投影点,则月基平台发射天线的指向近似垂直于月面并指向地球。

本文选取了同经度范围,不同纬度的 3 个地球站,其位置等参数如表 1 所示。

链路方案中月球站天线直径设置为 4 m,地球站为大口径直径 30 m 的天线;星载接收机发射机天线直径分别设置为 3.7 m 和 2 m。整个下行链路的传输频率设置为 14.5 GHz,即系统所选用的频段为 Ku 频段。

表 1 位置参数

Table 1 Position parameter

类 型	名 称	位 置
地球站	Llat	16°52′N,112°20′E
	Mlat	39°54′N,116°23′E
	Hlat	73°15′N,115°05′E
月球站	MBPoint	0°S,0°E

2.1.2 传输损耗参数

雨衰是指降雨引起的电波传播损耗,是由雨滴和雾对微波能量的吸收和散射产生的,其大小与雨量和电波穿过雨区的有效传输距离有关,并且随着频率的增大而增大[12-13],本文选取的雨衰参数文件为通用的 ITU-R P618-9。

大气损耗包括大气吸收损耗和大气折射(散焦)损耗。其中,无线电波往返大气时,在大气层传播中会受到包括电离层中自由电子、离子、对流层中氧分子和水蒸气分子等的吸收,由此产生大气吸收损耗[14]。本文选取的大气吸收损耗参数文件为 Simple Satcom。而大气折射损耗是由于大气层中的大气密度因高度的变化而不同,使电磁波在传输路径上出现弯曲的现象,引起天线波束扩散,继而产生折射损耗。无线电波往返大气层时,在大气层传播中会产生吸收损耗。

2.1.3 数据传输编码方式

本文在仿真过程中,选择较默认编码方式 BPSK 更为有效的 QPSK 编码方式。BPSK 编码方式在一个符号的持续时间 T 内能够传送 1 个比特的信息(+1 或者 −1),当需要传送比特 +1 时,就发送正余弦波;当需要传送比特 −1 时,就发送负余弦波。而 QPSK 调制技术能够在一个符号周期内传送 2 个比特的信息。QPSK 编码方式不仅利用了同相(in-phase)分量,还利用了正交(quadrature)分量。同相分量和正交分量彼此正交,因此两者可以被结合在一起而不会相互干扰。所以,QPSK 编码方式可以在第 1 个 BPSK 信号上叠

北航学报 赠阅

加与之正交的第 2 个 BPSK 信号,而不产生任何干扰。故 QPSK 调制技术能够成倍地提高 BPSK 的频带利用效率,因为在 T 时间内还能够传送另一个比特。

2.2　月地直传通信方案

月基平台与地球通信,其中一种方式为直接通信,不通过任何中继卫星进行转发,图 4 为直接传输示意图。

在同一个自然月内,月球在地球上投影落点分为北半球、南半球和赤道位置附近,而赤道位置附近又可以分为由北向南通过赤道及由南向北通过赤道。本文按照月球在地球上投影落点位置,模拟了月地关系这 4 种不同时间段的链路场景。图5表征了各个不同时期不同地球站接收信号的信噪比的变化,通过信噪比的变化可以得出信号误码率的变化情况。在场景的每一个阶段中,时长为 2 d,每次间隔的步长为 60 s。通过场景模拟可知,若不考虑对最高误码率限制的最低信噪比值,对地观测系统的收发端一般均有不可见的时期,且在不同的可见时间段内,信噪比值会出现较大的振荡。



图 4　直接传输示意图

Fig. 4　Schematic of direct transmission

当月球的投影点位于地球北半球时候,高纬度地球站的可见性相较于中低纬度的地球站高。因为从空间位置上,地球站与地心的连线与月地轴心连线的夹角为锐角;并且此时中、低纬度地区可见情况高度重合,包括可见时间及接收端信噪比值大小。其他不同时期的地球站有时会有非锐角出现。当月球的投影点位于地球南半球时候,高纬度的地球站已经变得不可见,也就不能完成月地通信任务。通过叠加图 5(c)和图 5(d),可以得出月球投影落于赤道时,此时高、中、低纬度的地球站的可见性差别最小,同一性最高。表 2 为各个阶段不同纬度的地球站对月基平台的可见时长占对应阶段总时长的百分比。

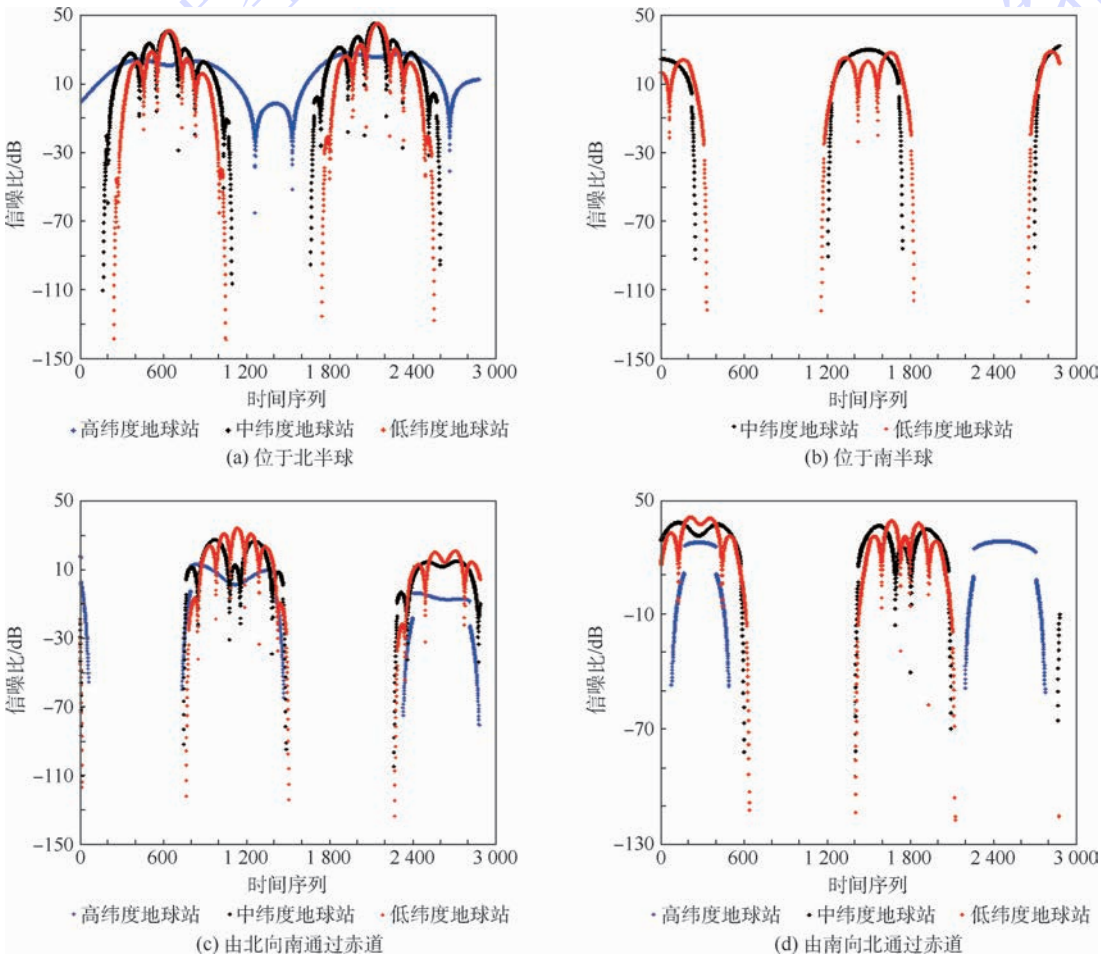


图 5　直传模式月球投影

Fig. 5　Lunar shadow in direct transmission mode

北航学报 赠阅

但实际情况下对于月基平台通信,地球站可见不一定就能够正常有效地进行通信。两端能够进行通信的基本条件是其间链路的信噪比满足最低要求。当月基平台与地球站可见时,地球站可能处于月基平台波束覆盖的边缘,链路的损耗过大,致使接收端接收到的信号强度过小、误码率过高,使得接收到的信号质量大大降低,甚至成为无用的信号,所以通信链路传输的编码应不大于其最大的误码率值,对于月球至地球通信的下行链路而言,如 1.3.2 节所述,按照最高误码率为 10^{-6},即信噪比值 Eb/No 不应低于 10 dB 计,可得到有效通信概率,如表 3 所示。

由表 3 可以看出,当考虑基本的信噪比约束后,月地系统的可通信时间窗口会被压缩,若将误码率的值进一步降低,时间窗口会进一步缩短,很明显,上述各种情况的信噪比峰值均价不超过 40 dB,若最低 Eb/No 阈值为 40 dB,则月地的通信将变得不可能,这种情况不满足于可靠、安全和实时传输数据的要求。因此月基平台与地球站的直接通信不可避免会遇到信号强度衰减、传播码数的错误丢失、信号的延迟和通信两端可见性等问题,并不是最优的链路方案。

表 2 直传模式地球站可见性概率

Table 2 Visibility probability of earth station in direct transmission mode %

地球站	月球投影位置			
	南半球	北半球	由北向南过赤道	由南向北过赤道
Llat	42.94	56.12	46.76	48.11
Mlat	34.03	64.66	46.76	45.58
Hlat	0	100	47.73	34.55

表 3 直传模式地球站有效通信概率

Table 3 Effective communication probability of earth station in direct transmission mode %

地球站	月球投影位置			
	南半球	北半球	由北向南过赤道	由南向北过赤道
Llat	29.41	38.96	30.74	39.71
Mlat	29.28	49.11	29.03	41.89
Hlat	0	72.17	5.48	23.81

2.3 月地中继通信方案

为了有效解决月地直接通信时出现的可见性和链路信噪比最低要求的问题,本文引入中继卫星对由月球传入地球的信号进行中继转发的思路。中继卫星一般是地球静止轨道卫星,是地球同步轨道卫星的特例[15]。该颗卫星定点于同经度范围的 3 个地球站的上空,也可以定点于其他位置。一般而言,当卫星与地球站的纬度相对位置已经固定时,两者经度相差越小,信号在传输过程经过的大气层路径或越短,大气损耗和多径损耗也会降低。

图 6 为一种月地中继传输示意图,通过布置在月基平台上的传感器,可以将发射机的信号向着中继卫星的方向传输。

图 7 基于一颗中继卫星的链路方案,模拟了月球绕地球周期内 4 个时间段的数据传输信噪比变化情况。对于加有一颗中继卫星时,月地系统基本上可以实现全天候的可见度,只是当月球的投影向地球赤道靠近时,系统会出现小时间范围的不可见,此时系统也就不能正常通信。

总体而言,中纬度地区地球站较高低纬度地球站的通信效果好,且此时三者曲线没有较大的起伏,与没有添加中继卫星相比,保证了信号传输的一致性、避免了较大的振荡。

在地球站可见性概率方面,如表 4 所示,添加了中继卫星后的地球站可见性概率明显优于地月直接通信。

为了使全天候的信号传输通信成为可能,可以在 2.2 节中继卫星 A 之外引进另一颗中继卫星 B,其位置定点于沿卫星绕地转动方向相反的某处,模拟的链路方案如图 8 所示。

此时在 A 的通信时间窗口之外的通信盲区有 2 种可供选择的链路:①月基平台、中继卫星 B、中继卫星 A、地球站;②月基平台、中继卫星 B、地球站。在确定 2 颗卫星的星间角度的时候,应当留有一定的裕度。该方案 2 颗中继卫星经度值相差 40°,可以充分保证了当卫星 A 进入信号传输盲区后,设备有充分的时间调整响应;待卫星离开信号传输盲区后,设备不至于立刻切换而改变链路对象。各卫星的参数保持一致,卫星 B 对卫星 A 的信号进行二次放大、降噪处理,使下行链路的传输信号比一次处理更能拟合源信号。

对上述 2 颗中继卫星进行仿真,可以发现地球高、中、低纬度地球站的信噪比随时间的变化



图 6 中继传输示意图

Fig. 6 Schematic diagram of relay transmission

北航学报 赠阅

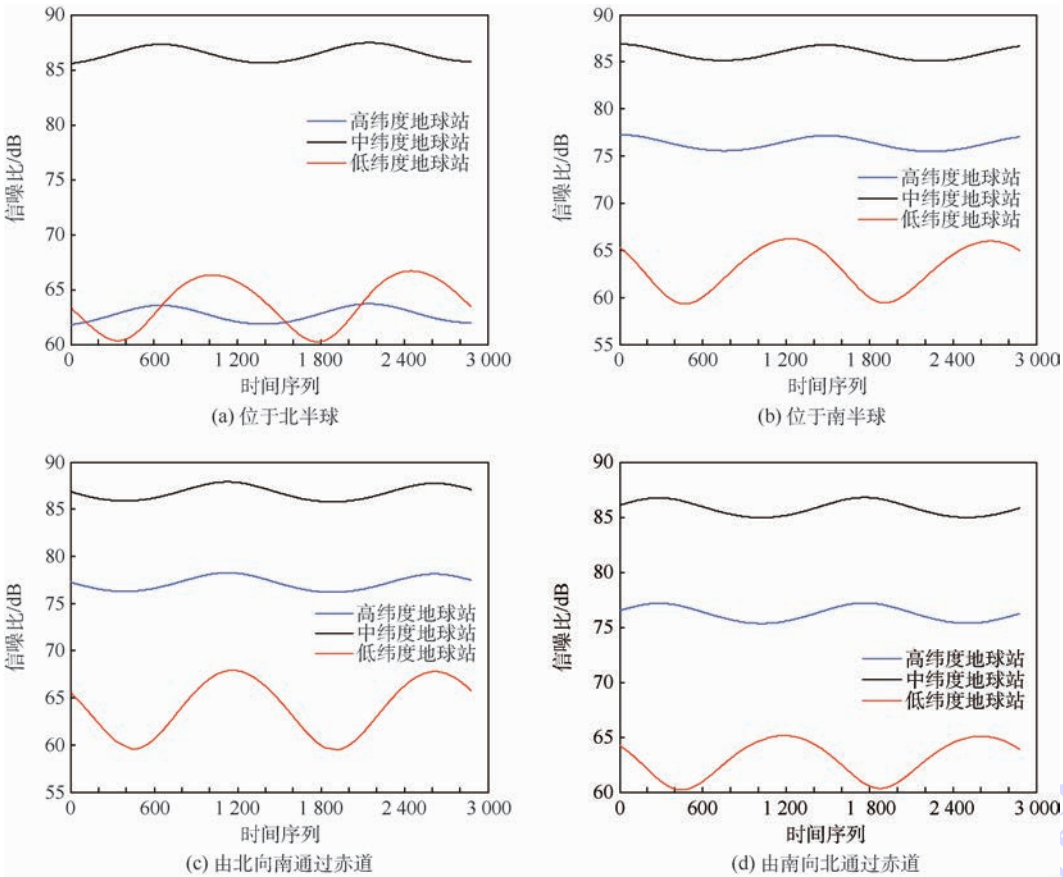


图 7　单颗中继卫星模式月球投影

Fig. 7　Lunar shadow in single relay satellite mode

曲线高度重合,故此只在此表示出中纬度地球站的 Eb/No 随时间的变化趋势图。

图7(c)、图7(d)和图9模拟可知,不同模式

表 4　单颗中继卫星模式地球站可见性概率

Table 4　Visibility probability of earth station in single relay satellite mode　%

地球站	月球投影位置			
	南半球	北半球	由北向南过赤道	由南向北过赤道
Llat	100	100	94.76	95.59
Mlat	100	100	94.76	95.59
Hlat	100	100	94.76	95.59

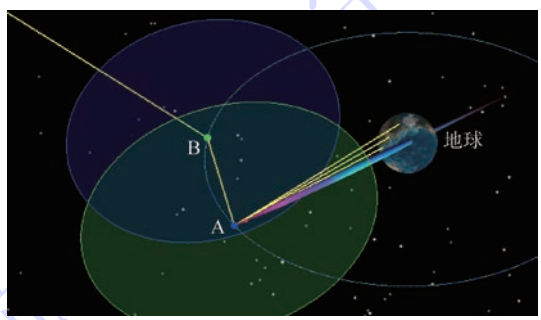


图 8　2 颗中继卫星组传输示意图

Fig. 8　Schematic diagram of transmission of two relay satellites

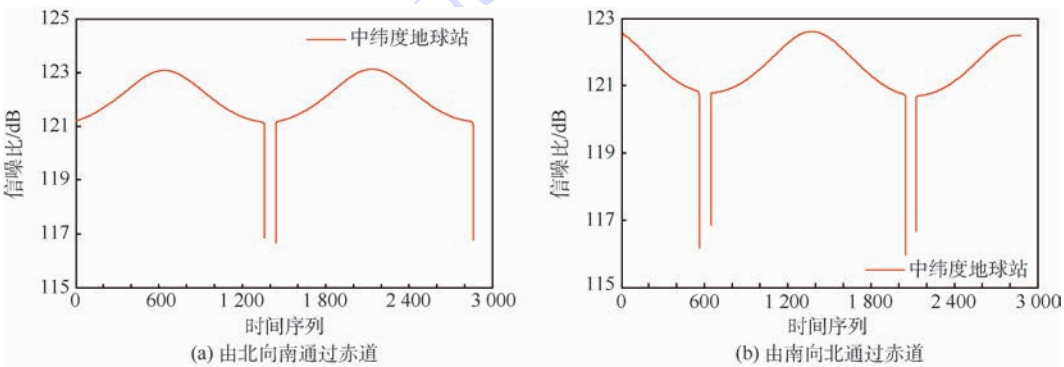


图 9　2 颗中继卫星模式月球投影

Fig. 9　Lunar shadow in two relay satellites mode

北航学报 赠阅

下信号有效通信窗口不一样,故此可以在中继卫星通信将要进入通信盲区时,改变链路的路径,从而实现全天候,无时断的信号传输。

3 结 论

月基平台的相关研究是极具开创性的工作,对于全面深刻认识月基平台建设的可行性及关键技术具有重要意义。本文利用 STK 和 MATLAB 模拟技术,基于深空通信和卫星数据传输等基本理论,仿真月基数据传输链路,首次提出了适用于月基对地观测平台的下行链路通信方案。分析认为月地直传通信方案存在传输两端可见性低、下行链路通信周期内信噪比较低且不稳定,将无法满足月基平台的观测和后续分析需求。本文提出的月地下行链路中继通信方案,在仿真条件下认为设置中纬度地面接收站、最小间隔经度值为 40°的 2 颗中继卫星,可以很好地实现全天候、无时断的信号传输,保障月基对地观测平台的运行。

月基数据传输方案的构建将进一步开展月地接收、发射端参数和信号链路参数等研究,期待在未来若干年内,能够真正地实现以月球为一个大型集中的基站而对地球、甚至其他宇宙对象进行观测研究。

参考文献(References)

[1] NASA. Lunar exploration objectives, version 1 [EB/OL]. (2016-12-01) [2017-05-08]. https://www.nasa.gov/pdf/163560main_LunarExplorationObjectives.pdf.

[2] 欧阳自远. 我国月球探测的总体科学目标与发展战略[J]. 地球科学进展,2004,19(3):351-358.
OUYANG Z Y. China's overall scientific goal and development strategy of lunar exploration[J]. Advances about Earth Science,2004,19(3):351-358(in Chinese).

[3] 郭华东. 地球系统空间观测:从科学卫星到月基平台[J]. 遥感学报,2016,20(5):716-723.
GUO H D. Spatial observation of earth system: From science satellite to moon platform[J]. Journal of Remote Sensing,2016,20(5):716-723(in Chinese).

[4] 丁翼星,郭华东,刘广. 面向全球变化探测的月基对地观测覆盖性能分析[J]. 湖南大学学报(自然科学版),2014,41(10):96-102.
DING Y X,GUO H D,LIU G. Coverage performance analysis of earth observation from lunar base for global change detection[J]. Journal of Hunan University (Natural Science),2014,41(10):96-102(in Chinese).

[5] 郭华东,丁翼星,刘广,等. 面向全球变化探测的月基成像雷达概念研究[J]. 中国科学:地球科学,2013,43(11):1760-1769.
GUO H D,DING Y X,LIU G,et al. Research on moon's imaging radar concept for global changes detections[J]. Chinese Science: Earth Science,2013,43(11):1760-1769(in Chinese).

[6] 娄岩,赵义武,佟首峰,等. 太阳直射对 GEO 卫星和地面站通信的影响[J]. 计算机仿真,2015,32(3):234-237.
LOU Y,ZHAO Y W,TONG S F,et al. Solar radiation impact on GEO satellite to ground laser communication[J]. Computer Simulation,2015,32(3):234-237(in Chinese).

[7] 李博,叶晖,张宏伟. 基于 STK/MATLAB 接口的卫星通信链路研究[J]. 无线电通信技术,2016,42(6):37-40.
LI B,YE H,ZHANG H W. Dynamical satellite link budget analysis based on STK/MATLAB interface[J]. Radio Communications Technology,2016,42(6):37-40(in Chinese).

[8] 李国军,霍德聪. 利用 STK 计算卫星通信链路余量[J]. 空间电子技术,2012(1):68-72.
LI G J,HUO D C. Satellite communication link budget computation using STK[J]. Space Electronic Technology,2012(1):68-72(in Chinese).

[9] 韩湘,郭新哲,屈会鹏. GEO 通信卫星姿态对点波束天线指向影响分析[J]. 飞行器测控学报,2015,34(1):77-82.
HAN X,GUO X Z,QU H P. Impact of attitude movement on spot beam pointing for GEO communication satellites[J]. Journal of Spacecraft TT&C Technology,2015,34(1):77-82(in Chinese).

[10] 鲍凯,徐慨,项顺祥. 基于 STK 的点波束覆盖分析与仿真[J]. 通信技术,2013,46(1):17-19.
BAO K,XU K,XIANG S X. Analysis and simulation of spot beam coverage based on STK[J]. Communications Technology,2013,46(1):17-19(in Chinese).

[11] 周帆,潘成胜,常小凯. 基于 STK 的卫星通信视景仿真[J]. 沈阳理工大学学报,2007,26(5):25-28.
ZHOU F,PAN C S,CHANG X K. Visual simulation for satellite communication based on STK[J]. Transactions of Shenyang Ligong University,2007,26(5):25-28(in Chinese).

[12] 闵士权. 卫星通信系统工程设计与应用[M]. 北京:电子工业出版社,2015:268-280.
MIN S Q. Design and application of satellite communication system[M]. BeiJing: Electronic Industry Press,2015:268-280(in Chinese).

[13] 丁溯泉,张波,刘世勇. STK 使用技巧及载人航天工程应用[M]. 北京:国防工业出版社,2016:54-61.
DING S Q,ZHANG B,LIU S Y. STK's using skills and applications on manned space engineering[M]. Beijing: National Defense Industry Press,2016:54-61(in Chinese).

[14] 赵业福,李进华. 无线电跟踪测量系统[M]. 北京:国防工业出版社,2001:56-70.
ZHAO Y F,LI J H. Radio's tracking & measurement system[M]. Beijing: National Defense Industry Press,2001:56-70(in Chinese).

[15] 方炎申,顾中舜,陈英武. 中继卫星与用户航天器之间星间链路的研究[J]. 计算机工程与设计,2005,26(8):2064-2066.
FANG Y S,GU Z S,CHEN Y W. Research on inter-satellite links between relay satellites and user spacecrafts[J]. Computer Engineering and Design,2005,26(8):2064-2066(in Chinese).

北航学报 赠阅

作者简介:

陈国强　男,硕士研究生。主要研究方向:交通信息工程。

阮智星　女,博士,助理研究员。主要研究方向:雷达遥感、月基对地观测。

Moon-based platform data transmission link scheme design and analysis

CHEN Guoqiang[1,2], RUAN Zhixing[1,*], GUO Huadong[1], LIU Guang[1], DING Yixing[1], ZHANG Yanxin[2]

(1. Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100094, China;
2. School of Electronic & Information Engineering, Beijing Jiaotong University, Beijing 100044, China)

Abstract: The ability of data transmission and processing is an important factor in the construction of the moon-based platform. How to transmit the observation data to the earth station efficiently and accurately is a critical step to carry out the further researches. Through the joint simulation of STK and MATLAB, we simulate data transmission link from Moon-based platform to Earth and put forward a downlink communication scheme for the platform. It has been proved that the scheme of two relay satellites with 40 degrees apart and of utilizing the middle latitude earth station can receive reliable information all the time. It could make the downlink receiver get information with enough power and low bit error rate, so as to ensure the operation of the moon-based earth observation platform.

Keywords: moon-based platform; earth observation; deep space communication; data transmission; link

北京航空航天大学学报

Received: 2017-05-08; **Accepted**: 2017-08-02; **Published online**: 2017-09-26 14:45

URL: kns.cnki.net/kcms/detail/11.2625.V.20170926.1445.002.html

Foundation items: National Natural Science Foundation of China (41590852, 41590853, 41501403); Key Program of the Chinese Academy of Sciences (QYZDY-SSW-DQC026)

* **Corresponding author**. E-mail: ruanzx@radi.ac.cn

北航学报 赠阅

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2017.0240

多模函数优化的改进花朵授粉算法

郭庆，惠晓滨*，张贾奎，李正欣
（空军工程大学 装备管理与安全工程学院，西安 710051）

摘 要：为了探讨花朵授粉算法(FPA)在解算多模函数优化问题中存在的不足，通过定义种群多样性及差异性指标，定性分析了FPA在多模复杂函数优化中的寻优缺点。基于模拟退火思想优化全局授粉过程，并利用Nelder-Mead单纯形搜索技术对花朵局部授粉进行重构，提出一种新的花朵授粉寻优架构。仿真结果表明，相对于基本的FPA、布谷鸟算法、萤火虫算法，改进花朵授粉算法能够有效避免陷入局部最优，具备优异的全局勘探和局部开采能力，对多模优化问题具有一定优势。

关 键 词：花朵授粉算法(FPA)；模拟退火；Nelder-Mead单纯形法；多模函数优化
中图分类号：TP181
文献标识码：A **文章编号**：1001-5965(2018)04-0828-13

全下文载

花朵授粉算法(Flower Pollination Algorithm，FPA)是由英国剑桥大学学者Yang于2012年提出的，其基本思想来源于对自然界花朵自花授粉、异花授粉的模拟，是一种新的元启发式群智能随机优化技术[1]。之后，Yang等[2-3]在FPA的基础上模拟花朵多配子的形式提出了多目标的FPA(Multi-Objective Flower Pollination Algorithm，MOFPA)。该算法的寻优框架主要由Levy飞行的全局勘探机制、概率调控全局勘探与局部开采的平衡机制以及贪婪式的进化机制组成。FPA的参数相对较少、容易调控等优势使得该算法在一些领域得到了广泛的应用。在经济领域，Al-Ma'Shumah等[4]应用该算法解决市场周期价值问题，Prathiba[5]和Dubey[6]等应用该算法解决经济调度问题；在工程领域，Abdelaziz等应用该算法解决电力调度问题[7]，Sharawi等[8]应用该算法解决无线传感器网络优化问题；在医学领域，Emary等[9]应用该算法解决视网膜血管分割问题等。

FPA与经典的粒子群优化(PSO)算法、人工蜂群算法、布谷鸟算法等仿生算法相同，也存在易陷入局部最优、易早熟和后期收敛速度慢等典型缺点。为了克服这些缺点，国外许多学者(中国当前研究较少)在FPA的基础上进行一些针对性的研究与创新。如Wang和Zhou[10]引入邻域搜索策略、维度评价及更改策略和动态切换概率策略对算法进行改进，提高算法在高维函数优化问题的全局勘探与开发能力；Abdel-Raouf等[11]针对定积分的求解问题结合混沌优化技术提出一种改进花朵授粉算法(IFPCH)，针对数独问题并结合和声搜索算法提出一种混合花朵授粉算法(FPCHS)[12]，从不同侧面改善FPA的收敛性能；El-Henawy和Ismail[13]将粒子群优化算法与FPA相融合提出一种混合算法，并将其应用到求解整数规划问题及带约束的全局优化问题[14]；Łukasik和Kowalski[15]对FPA进行更加详尽的描述以及对其所涉及的参数进行了研究，并与经典的粒子群优化算法进行了对比，得出FPA前期收敛速度不如粒子群优化算法，但其后期种群的多样性要

收稿日期：2017-04-18；**录用日期**：2017-05-19；**网络出版时间**：2017-06-19 15:53
网络出版地址：kns.cnki.net/kcms/detail/11.2625.V.20170619.1553.005.html
基金项目：国家自然科学基金(61502521)
*通信作者. E-mail：zibai4991@qq.com

引用格式：郭庆，惠晓滨，张贾奎，等. 多模函数优化的改进花朵授粉算法[J]. 北京航空航天大学学报，2018，44(4)：828-840.
GUO Q，HUI X B，ZHANG J K，et al. Improved flower pollination algorithm for multimodal function optimization[J]. Journal of Beijing University of Aeronautics and Astronautics，2018，44(4)：828-840(in Chinese).

优于粒子群优化算法的结论。

上述的各种研究改进算法,在某些方面提高了 FPA 的寻优性能,但 FPA 在其他领域仍存在一些不足,需要不断地改进与创新。如在多模复杂函数寻优问题上,FPA 的全局勘探与局部开采能力表现不是很理想。为此,本文针对这一问题进行研究,首先从单峰函数寻优入手探究 FPA 的寻优机理,再结合多峰函数寻优分析 FPA 在多模寻优时存在的问题,最后结合模拟退火(Simulated Annealing,SA)及 Nelder-Mead 单纯形法提出一种改进的花朵授粉法——NS-FPA。

1　花朵授粉算法及其寻优分析

1.1　花朵授粉算法

Yang[1] 通过对自然界显性花朵授粉的分析与研究,将花朵授粉分为两大基本形式:生物异花授粉与非生物自花授粉。并依据花朵的常性与授粉行为抽象出以下 4 条规则:

1) 生物异花授粉被考虑为算法的全局勘探行为,并由传粉者携带花粉通过 Levy 飞行的机制实现全局授粉。

2) 非生物自花授粉被视为算法的局部开采行为,或称局部授粉。

3) 花朵的常性可以被认为是繁衍概率,它与两朵参与授粉花朵的相似性呈正比例关系。

4) 花朵的全局授粉与局部授粉通过转换概率 $p\in[0,1]$ 进行调节。由于物理上的邻近性和风等因素的影响,在整个授粉活动中,转换概率 p 是一个非常重要的参数。文献[1]中对该参数的试验研究认为,取 $p=0.8$ 更利于算法寻优。

实际情况下,每棵显花植物拥有许多花朵,并且每朵花都有成千上万的花朵配子。为了算法的实现,Yang[1] 假设每棵显花植物仅有一朵花,每朵花仅包含一个配子,每个配子被视为解空间中的一个候选解。FPA 寻优的伪代码如下:

Code:FPA( )

Objective:$\min f(\boldsymbol{x})/\max f(\boldsymbol{x}),\boldsymbol{x}=(\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_d)$

Input:N, $K_{\max}$, p, λ

随机初始化 N 个候选解 $\boldsymbol{x}_0^1$, $\boldsymbol{x}_0^2,\cdots,\boldsymbol{x}_0^N$

计算相应的函数值 y_0^1, $y_0^2,\cdots,y_0^N$;

计算最优解 $\boldsymbol{x}^{\text{best}}$、最优值 y^{best};

while($t<K_{\max}$)

　for $i=1,2,\cdots,N$ do

　　if rand $<p$ then

　　依据式(1)执行全局授粉;

　　else

　　依据式(2)执行局部授粉;

　　endif

　　计算子代花粉的函数值;

　　更新花粉的位置;

　endfor

　更新最优解 $\boldsymbol{x}^{\text{best}}$、最优值 y^{best};

endwhile

Output:$\boldsymbol{x}^{\text{best}}$, y^{best}

$$\boldsymbol{x}_i^{t+1}=\boldsymbol{x}_i^t+L(\boldsymbol{x}_i^t-\boldsymbol{x}^{\text{best}}) \tag{1}$$

$$\begin{cases}\boldsymbol{x}_i^{t+1}=\boldsymbol{x}_i^t+\xi(\boldsymbol{x}_j^t-\boldsymbol{x}_k^t)\\ \xi\sim U(0,1)\end{cases} \tag{2}$$

式中:$\boldsymbol{x}_i^{t+1}$ 和 $\boldsymbol{x}_i^t$ 分别为第 $t+1$ 和 t 代的候选解;$\boldsymbol{x}_j^t$ 和 $\boldsymbol{x}_k^t$ 为相同植物种类的不同花粉配子;ξ 为局部授粉系数;U 为均匀分布;L 为飞行步长,服从 Levy 分布:

$$L\sim\frac{\lambda\Gamma(\lambda)\sin(\pi\lambda/2)}{\pi}\cdot\frac{1}{s^{1+\lambda}}\quad s\gg s_0>0 \tag{3}$$

其中:$\Gamma(\lambda)$ 为伽马函数,$\lambda=1.5$;s 为步长;s_0 为最小步长。文献[1]中采用 Mantegna's 算法生成 Levy 飞行步长 L。

1.2　寻优分析

为了定性分析 FPA 在寻优过程中花粉种群多样性的变化及寻优特性,本文针对 FPA 定义如下 2 个基本参数。

定义 1　多样性(diversity)是指每一代花粉的离散程度,简记为 Div。则第 t 代离散程度 Div^t 计算式为

$$\text{Div}^t=\frac{\sqrt{\sum_{i=1}^{N}\|\boldsymbol{x}_i^t-\bar{\boldsymbol{x}}^t\|^2}}{N\max_i(\|\boldsymbol{x}_i^t-\bar{\boldsymbol{x}}^t\|)} \tag{4}$$

$$\bar{\boldsymbol{x}}^t=\sum_{i=1}^{N}\boldsymbol{x}_i^t/N \tag{5}$$

式中:N 为花粉种群规模;$\bar{\boldsymbol{x}}^t$ 为第 t 代花粉在解空间中分布的几何中心。

定义 2　差异性(difference)是指花粉种群子代与父代的差异程度,简记为 Dif。则第 t 代差异程度 Dif^t 计算式为

$$\text{Dif}^t=\frac{\sqrt{\sum_{i=1}^{N}\|\boldsymbol{x}_i^t-\boldsymbol{x}_i^{t-1}\|^2}}{NR} \tag{6}$$

式中:R 为花粉最大移动半径(即解空间的半径)。

在 FPA 寻优过程中,通过对花粉多样性 Div 与差异性 Dif 的联合刻画以分析花粉种群多样性

北航学报 赠阅

的动态变化,再结合寻优迭代过程中花粉位置的动态变化分析 FPA 的寻优特点。

首先,从单峰函数寻优入手。图 1 为二维 Sphere 函数的 3D 图,该函数是典型的单峰函数。

依据参考文献[1,3]的算法流程及参数设定(花粉数量取 30)对该 Sphere 函数进行 100 次迭代寻优,其多样性 Div 与差异性 Dif 变化曲线如图 2所示。

图 3 为 FPA 进化过程中花粉配子在解空间内的动态分布图。图中,从第 1、10、30、50、80、100 代花粉种群在解空间中分布,可以直观的看出随着种群的进化,花粉种群在解空间的分布趋于集中,直至全部集中在全局最优位置;而反映在图 2 中的则是花粉种群的多样性指标 Div 随着进化代次的增加基本呈衰减走势,这说明花粉的空间多样性在不断锐减。从图 3 中还能够看出,第 50 代以后,花粉种群基本都已聚集到全局最优值附近;表现在图 2中的则是差异性指标 Dif 在第 50 代之后波动较小且基本趋于 0,这说明算法后期花粉在广阔解空间区域的全局勘探能力基本丧失,这对多模复杂函数寻优是极其不利的,很可能会致使算法陷入局部最优。另外,单峰函数不会存在陷入到局部最优的情况,能够集中体现算法的局部开采能力,将图 3 中第 100 代的花粉分布图进行局部放大,如图4(红星代表最小值点)所示。可以看出当 x、y 坐标放大到 0.1 等级时,花粉已经远离最小值点,这说明 FPA 的局部开采能力较为薄弱。

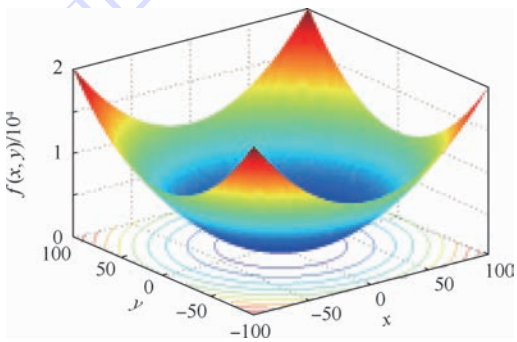


图 1 Sphere 函数 3D 图

Fig. 1 3D figure of Sphere function

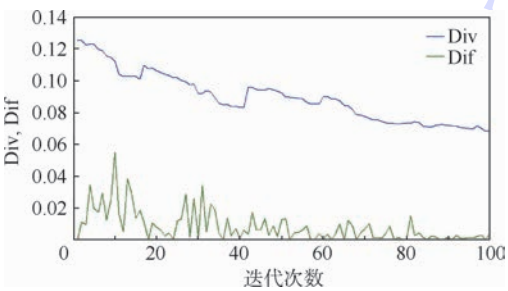


图 2 Sphere 函数 FPA 寻优参数变化曲线

Fig. 2 Parameter variation curves of Sphere function using FPA optimization

其次,对多模函数进行寻优,分析 FPA 对多模复杂函数的寻优缺点。图 5 为 Bridge 函数的 3D 图,该函数峰值较多,是较为复杂的多模函数。同样依据文献[1,3]对该函数进行 100 次迭代寻优实验,给出参数变化曲线(见图 6)及花粉种群的空间分布图(见图 7)。

图 7 中,单从第 1、10、30、50、80、100 代花粉种群在解空间中的分布看,每代的花粉种群分布

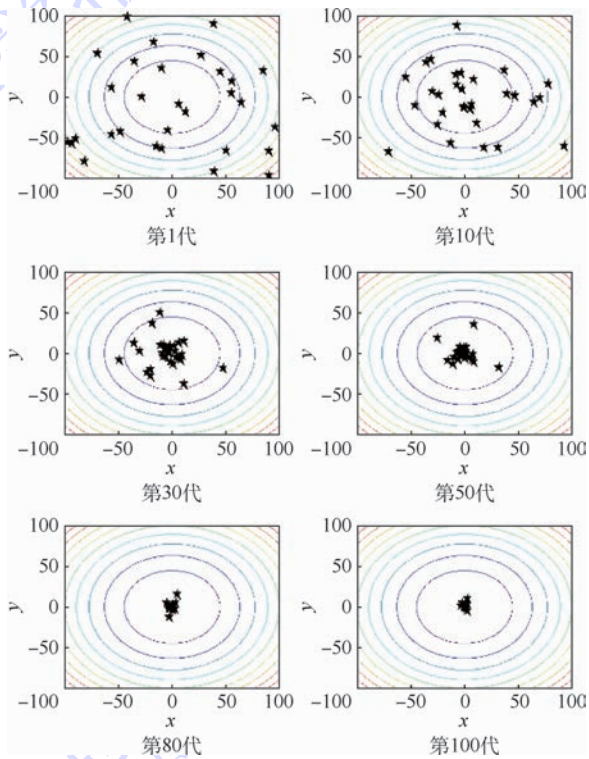


图 3 Sphere 函数 FPA 寻优花粉分布

Fig. 3 Pollen distribution of Sphere function using FPA optimization

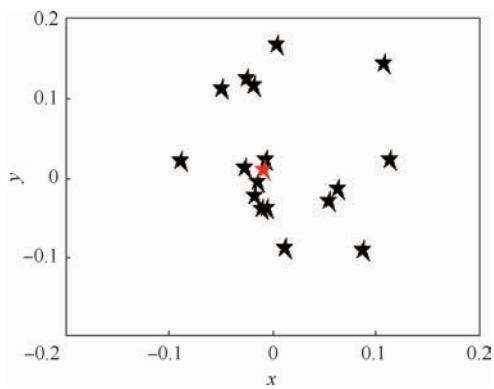


图 4 图 3 第 100 代的局部放大

Fig. 4 Partial enlargement of Fig. 3 the 100th generation

北航学报 赠阅

都较为松散,没有一部分花粉集中的现象。再对比第 1、10、30、50、80、100 代花粉的空间分布,可以看出第50、80、100代的花粉种群在解空间的位置分布基本一致。从图 6 可以看出多样性指标 Div 在第 50 代之后无波动、基本趋于平稳,差异性指标 Dif 在第 50 代之后基本为 0、无明显的波动。这说明 50 代之后大部分花粉已陷入到局部最优位置无法进行位置更新,这使得算法的全局勘探能力下降,以至于无法探索到全局最优位置,证实了上述的可能性。同时,也检验了 2 个参数指标(Div、Dif)联合分析的合理性与有效性。

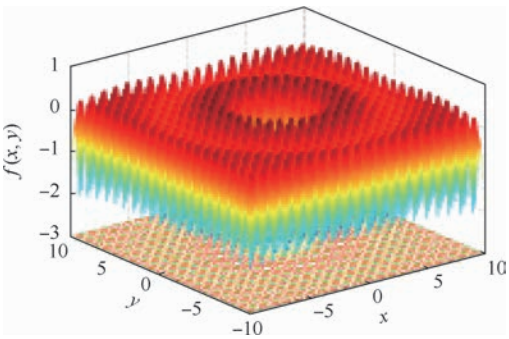


图 5　Bridge 函数 3D 图

Fig. 5　3D figure of Bridge function

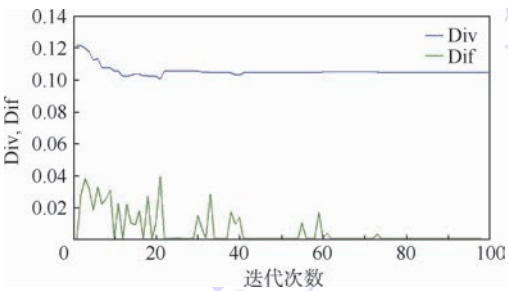


图 6　Bridge 函数 FPA 寻优参数变化曲线

Fig. 6　Parameter variation curves of Bridge function using FPA optimization

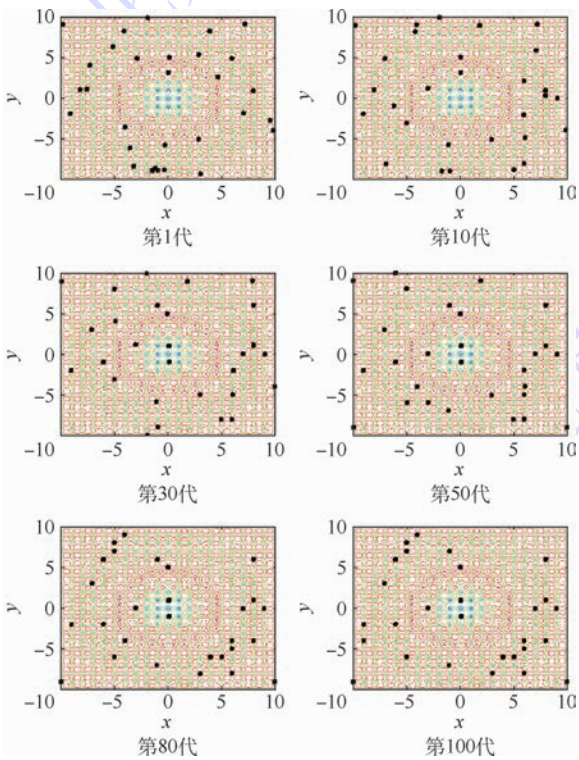


图 7　Bridge 函数 FPA 寻优花粉分布

Fig. 7　Pollen distribution of Bridge function using FPA optimization

最后结合 FPA 的具体流程,笔者认为造成 FPA 后期花粉差异性为 0 的主要原因在于 FPA 的贪婪式进化机制(子代的函数值比父代优才能进化),使算法贪婪过度,长期无法进行位置更新,从而导致多数的花粉陷入到局部最优位置,致使 FPA 的全局开拓能力削弱。理想状态的种群分布状况应该是:部分种群在不断聚集,即算法的局部开采性;而另一部分种群在全局空间内不断的探索,即算法的全局勘探性。而表现在 Div、Dif 上应当是:种群多样性的保持及算法后期差异性的延续(不为 0)。

2　改进策略及改进的花朵授粉算法

2.1　改进策略

针对 FPA 在多模复杂函数寻优的缺点,对算法进行改进。在全局授粉过程中引入模拟退火的思想,采取 Metropolis 准则对新状态进行更新。模拟退火的 Metropolis 准则[16]是依概率的形式对新状态进行更新,在 FPA 的全局授粉中引入该准则,将改善算法的过度贪婪情况,从而提高全局勘探能力。另外,为了进一步提高 FPA 的局部开采能力,结合改进的 Nelder-Mead 单纯形法对 FPA 的局部授粉行为进行重构。

Nelder-Mead 单纯形法是由 Nelder 和 Mead[17]于 1965 年提出的一种用于优化多维无约束问题的搜索算法。此算法通过在 D 维空间中取 $D+1$ 个点构成一个单纯形,计算其顶点函数值;然后对最差点进行内部压缩、外部压缩、映射和扩展等操作,以获得更好的点取代最差点,重构单纯形,从而不断的逼近最优点。此算法是一种直接的搜索算法,其寻优不受目标函数的连续性与可导性影响,具备精细的局部开采能力。为此,本文类比该算法的优势对 FPA 的局部授粉进行重构,以增强 FPA 的收敛精度。重构的局部授粉过程如下:

首先,利用花粉 i 及其周围最近的 $m-1$ 个花粉的位置信息构建单纯形(见图 8),并根据式(7)计算单纯形的中心:

北航学报 赠阅

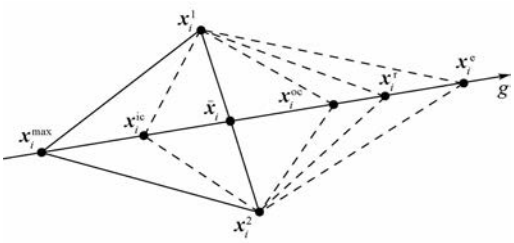


图 8 Nelder-Mead 单纯形法

Fig. 8 Nelder-Mead simplex method

$$\bar{\boldsymbol{x}}_i = \left(\sum_{j=1}^{m} \boldsymbol{x}_i^j - \max_j(\boldsymbol{x}_i^j) \right) \Big/ (m-1) \tag{7}$$

$$\begin{cases} \boldsymbol{x}_i^*(\boldsymbol{u}) = \bar{\boldsymbol{x}}_i + \boldsymbol{u}(\bar{\boldsymbol{x}}_i - \max_j(\boldsymbol{x}_j)) \\ \boldsymbol{u} = [u_{\mathrm{ic}} \quad u_{\mathrm{oc}} \quad u_{\mathrm{r}} \quad u_{\mathrm{e}}] \\ u_{\mathrm{ic}} = -\alpha^2, u_{\mathrm{oc}} = \alpha, u_{\mathrm{r}} = 1, u_{\mathrm{e}} = 1/\alpha \end{cases} \tag{8}$$

然后,根据式(8)分别对单纯形进行内部压缩、外部压缩、映射和扩展操作,计算 4 种操作所对应的位置坐标 $\boldsymbol{x}_i^{\mathrm{ic}}$、$\boldsymbol{x}_i^{\mathrm{oc}}$、$\boldsymbol{x}_i^{\mathrm{r}}$、$\boldsymbol{x}_i^{\mathrm{e}}$,并进行越界处理(依据文献[18]的分析与证明,这里取 $\alpha=0.618$)。

最后,计算 4 个位置的目标函数值 $y_i^*(\boldsymbol{u}) = f(\boldsymbol{x}_i^*(\boldsymbol{u}))$ 并与花粉 i 的目标函数值 y_i 进行对比,若比其优,将对花粉 i 进行位置更新。

2.2 改进的花朵授粉算法

结合改进策略及基本 FPA 的计算流程,给出 NS-FPA 的基本实现步骤(以寻最小值为例)及其流程图(见图 9)。

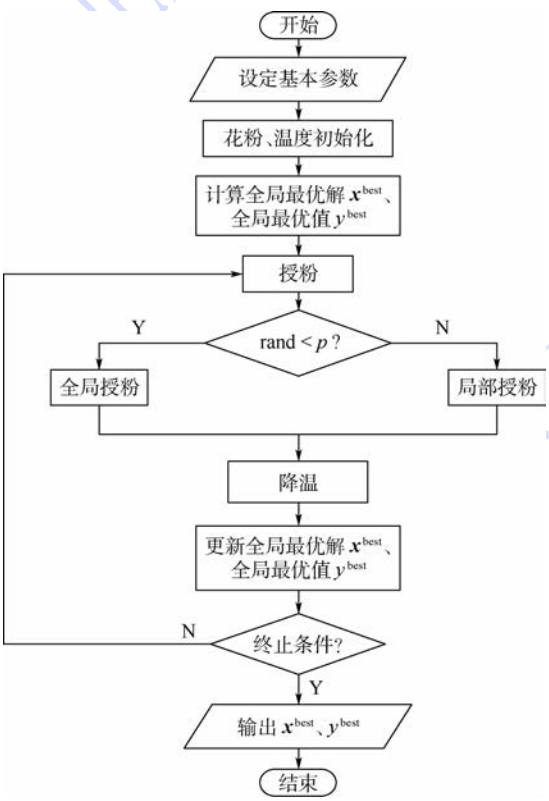


图 9 NS-FPA 流程

Fig. 9 Flowchart of NS-FPA

步骤 1 初始参数,设定花粉种群规模 N、最大迭代次数 $K_{\max}$、转换概率 p、初温概率 p_0、温度衰减系数 γ 等。

步骤 2 在解空间里随机初始化花粉位置 $\boldsymbol{x}^0$,并计算目标函数值 y^0;初始温度 T_0 可由式(9)计算:

$$T_0 = \frac{\min_N(y^0) - \max_N(y^0)}{\ln p_0} \tag{9}$$

式中:p_0 通常取 0.7 ~ 0.9[19]。

步骤 3 计算当前的全局目标函数最小值,并记录对应得最优解 $\boldsymbol{x}^{\mathrm{best}}$。

步骤 4 若条件(rand < p)成立,转步骤 5;否则,转步骤 6。

步骤 5 全局授粉:依据式(1)计算子代花粉位置,并计算目标函数值;由式(10)计算可接受概率 p_{r};若条件(rand < p_{r})成立,则进行位置更新。

$$p_{\mathrm{r}} = \exp\left[\frac{y_i^{t+1} - y_i^t}{T(\min_N(y^t) - \max_N(y^t))} \right] \tag{10}$$

步骤 6 局部授粉:依据 2.1 节中重构的局部授粉过程进行。

步骤 7 降温:按式(11)进行降温操作:

$$T = T_0 \gamma^t \qquad t = 1, 2, \cdots, K_{\max} \tag{11}$$

式中:γ 通常取 0.7 ~ 0.99[19]。

步骤 8 计算全局目标函数值的最小值,并更新最优解 $\boldsymbol{x}^{\mathrm{best}}$。

步骤 9 若满足终止条件,则输出全局最优解;否则,转至步骤 4。

3 仿真计算

为了验证 NS-FPA 的有效性,首先结合 1.2 节对 NS-FPA 进行定性仿真分析,验证算法改进后的全局勘探能力;其次,选取多个多模复杂函数进行寻优仿真实验,定量分析 NS-FPA 的收敛速度与收敛精度,从而分析 NS-FPA 的局部开采能力及综合性能;然后,对 NS-FPA 中新增的 2 个参数 p_0、γ 进行探讨,以确定其取值范围;最后,对比 NS-FPA、FPA 的时间复杂度,探讨 NS-FPA 的缺失。

3.1 定性仿真分析

为了与 1.2 节形成鲜明对比,本节同样针对二维 Sphere 函数及 Bridge 函数采用 NS-FPA 在同样参数下进行 100 次迭代寻优,并给出寻优过程中花粉多样性 Div 指标与差异性 Dif 指标变化曲线及花粉的动态分布图(见图 10 ~ 图 14)以定性

分析 NS-FPA 的寻优特性。

从图 11 和图 14 中可以直观地看出：无论对 Sphere 函数还是 Bridge 函数寻优，第 10、30、50、80、100 代花粉种群在空间的分布均呈现一部分种群聚集、另一部分种群松散的特征（由于第 1 代是随机初始化，因此整体较为松散、种群差异性最大）。另外，各代之间的种群分布除全局最优值附近的集中种群分布相似，这是说明部分花粉正在逐步向全局最优位置聚拢，体现了算法的导向性；分散的那部分种群分布有较大的差异，这是说明另外一部分花粉在不断的开拓新区域，体现了种群的全局勘探性。而从图 10 和图 13 中则可以看出：种群的多样性指标 Div 及差异性指标 Dif 在算法的 1～10 代（起始阶段）有略微的下降，因为这是算法由完全的随机性逐渐地向算法兼顾随机性与导向性转变的过程；第 10 代之后 Div 与 Dif基本呈平稳的波动状态，无整体的锐减趋势。

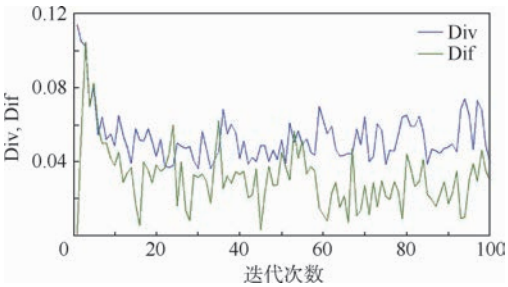


图 10　Sphere 函数 NS-FPA 寻优参数变化曲线

Fig. 10　Parameter variation curves of Sphere function using NS-FPA optimization

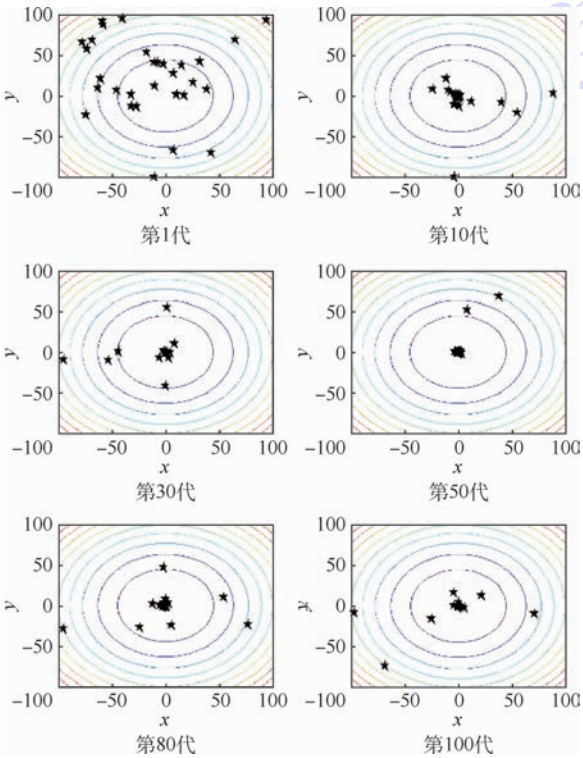


图 11　Sphere 函数 NS-FPA 寻优花粉分布

Fig. 11　Pollen distribution of Sphere function using NS-FPA optimization

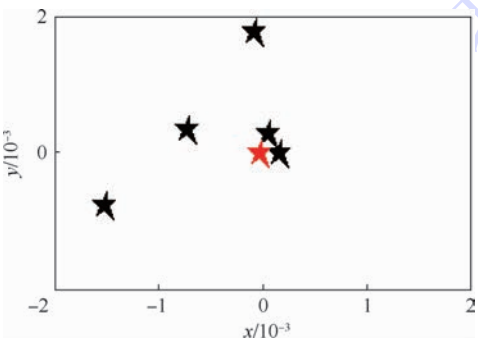


图 12　图 11 第 100 代的局部放大

Fig. 12　Partial enlargement of Fig. 11 the 100th generation

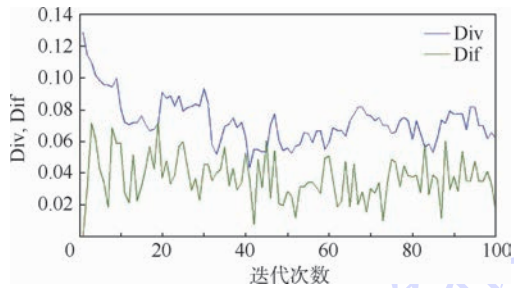


图 13　Bridge 函数 NS-FPA 寻优参数变化曲线

Fig. 13　Parameter variation curves of Bridge function using NS-FPA optimization

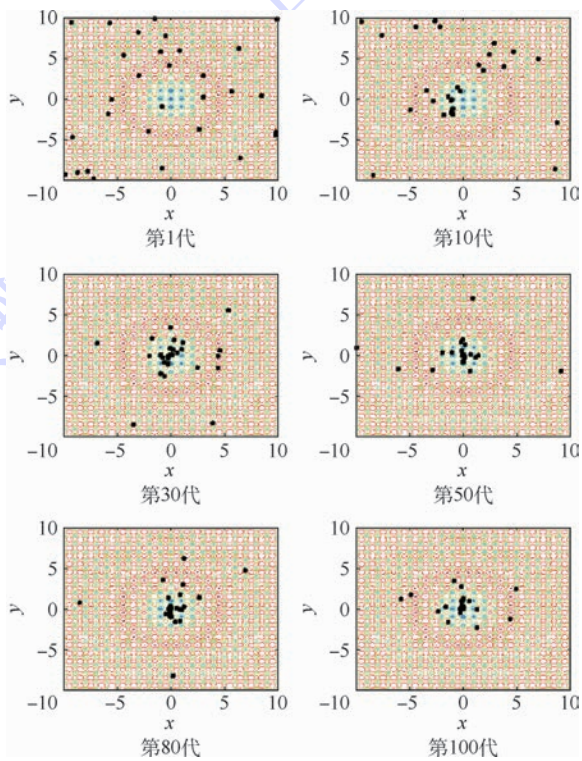


图 14　Bridge 函数 NS-FPA 寻优花粉分布

Fig. 14　Pollen distribution of Bridge function using NS-FPA optimization

北航学报 赠阅

这说明 NS-FPA 在算法运行的中后期能够很好的保持花粉种群的多样性与差异性,从而增加花粉跳出局部最优的几率,这对多模函数寻优是有利的。

与图 2 ~ 图 4、图 6 和图 7 相对比,可以得出以下结论:改进的 FPA 在多模函数寻优过程中能够保证花粉的多样性及差异性,从而能够避免算法陷入局部最优,有效地保证了 FPA 的全局勘探能力,基本达到了 1.2 节中所述的种群分布的理想状态。最后,结合图 4 和图 12 可看出 NS-FPA 的局部开采能力得到了明显的改善。

3.2 定量仿真分析

选择 15 个多模复杂测试函数、1 个单峰函数进行仿真实验并与 FPA、布谷鸟算法(CS)[20]、萤火虫算法(FA)[21]进行对比(见表 1)。16 个标准测试函数的基本信息如表 2 所示。为了更加直观地表现测试函数的复杂性,图 1(F1)、图 5(F11)和图 15给出了F1 ~ F16 的二维函数 3D 图。

通过图 15 再结合图 1 和图 5,16 个测试函数的多模复杂程度有所差异,但基本可以分为 3 类:F1(单峰),F2 ~ F8(多模),F9 ~ F16(超多模)。为了保证评价的客观性,在测试中分别对 16 个函数进行 50 次仿真计算,用 50 次仿真结果的统计特性对对比结果进行分析。依据文献[1,22]分别对 4 种寻优算法进行参数设定,如表 1 所示。其仿真对比结果如表 3 所示。

表 1 4 个算法的参数设置

Table 1 Parameter setup for four algorithms

算 法	参数设置
FPA	转换概率 $p=0.8$;参数 $\lambda=1.5$
CS	发现概率 $p_a=0.25$;步长调节量 $\alpha=0.01$;参数 $\beta=1.5$
FA	步长因子 $\alpha=0.25$;吸引度因子 $\beta=0.2$;光强吸收强度 $\gamma=1$
NS-FPA	转换概率 $p=0.8$;温度衰减系数 $\gamma=0.7$;初温概率 $p_0=0.8$;参数 $\lambda=1.5$

注:公共参数:种群规模 $N=50$;最大迭代次数 $K_{\max}$,F1 ~ F13 取 100,F14 ~ F16 取 10 000。

表 2 16 个标准测试函数

Table 2 16 standard test functions

代 号	函 数	表 达 式	解 空 间	最 小 值
F1	Sphere	$f(\boldsymbol{X})=\sum_{i=1}^{D}x_i^2$	$[-100,100]^2$	0
F2	Brid	$f(\boldsymbol{X})=\sin x\mathrm{e}^{(1-\cos y)^2}+\cos y\mathrm{e}^{(1-\sin x)^2}+(x-y)^2$	$[-2\pi,2\pi]^2$	-106.764 5
F3	Roots	$f(\boldsymbol{X})=-[1+\lvert(x+y\mathrm{i})^6-1\rvert]^{-1}$	$[-2,2]^2$	-1
F4	Bohachevsky	$f(\boldsymbol{X})=x^2+2y^2-0.3\cos(3\pi x)-0.4\cos(4\pi y)+0.7$	$[-100,100]^2$	0
F5	Eggcrate	$f(\boldsymbol{X})=x^2+y^2+25(\sin^2 x+\sin^2 y)$	$[-10,10]^2$	0
F6	Guichi-f4	$f(\boldsymbol{X})=-x\sin(4\pi x)+y\sin(4\pi y+\pi)$	$[-2,2]^2$	-3.253 9
F7	Cross-in-Tray	$f(\boldsymbol{X})=\frac{-1}{10\,000}\left(\left\lvert\sin x\sin y\exp\left\lvert100-\frac{\sqrt{x^2+y^2}}{\pi}\right\rvert\right\rvert+1\right)^{0.1}$	$[-10,10]^2$	-2.062 61
F8	Holder Table	$f(\boldsymbol{X})=-\left\lvert\sin x\cos y\exp\left\lvert1-\frac{\sqrt{x^2+y^2}}{\pi}\right\rvert\right\rvert$	$[-10,10]^2$	-19.208 5
F9	Drop-Wave	$f(\boldsymbol{X})=-\left(1+\cos\left(12\sqrt{x^2+y^2}\right)\right)\Big/[0.5(x^2+y^2)+2]$	$[-5.12,5.12]^2$	-1
F10	Levy	$f(\boldsymbol{X})=\sin^2(3\pi x)+(x-1)^2(1+\sin^2(3\pi y))+(y-1)^2(1+\sin^2(2\pi y))$	$[-10,10]^2$	0
F11	Bridge	$f(\boldsymbol{X})=-\sin\sqrt{x^2+y^2}\Big/\sqrt{x^2+y^2}-\exp((\cos(2\pi x)+\cos(2\pi y))/2)+0.712\,9$	$[-10,10]^2$	-3.005 4
F12	Shubert	$f(\boldsymbol{X})=\sum_{i=1}^{5}i\cos(i+(i+1)x)\sum_{i=1}^{5}i\cos(i+(i+1)y)$	$[-50,50]^2$	-186.730 9
F13	Rastrigin	$f(\boldsymbol{X})=10D+\sum_{i=1}^{D}(x_i^2-10\cos(2\pi x_i))$	$[-20,20]^4$	0
F14	Ackley	$f(\boldsymbol{X})=-20\exp\left(-0.2\sqrt{\sum_{i=1}^{D}x_i^2/D}\right)-\exp\left(\sum_{i=1}^{D}\cos(2\pi x_i)/D\right)+20-e$	$[-32,32]^{30}$	0
F15	Griewank	$f(\boldsymbol{X})=\sum_{i=1}^{D}x_i^2/400-\prod_{i=1}^{D}\cos(x_i/\sqrt{i})+1$	$[-600,600]^{40}$	0
F16	Vincent	$f(\boldsymbol{X})=-\sum_{i=1}^{D}\sin(10\ln x_i)$	$[0.25,10]^{50}$	-50

北航学报 赠阅

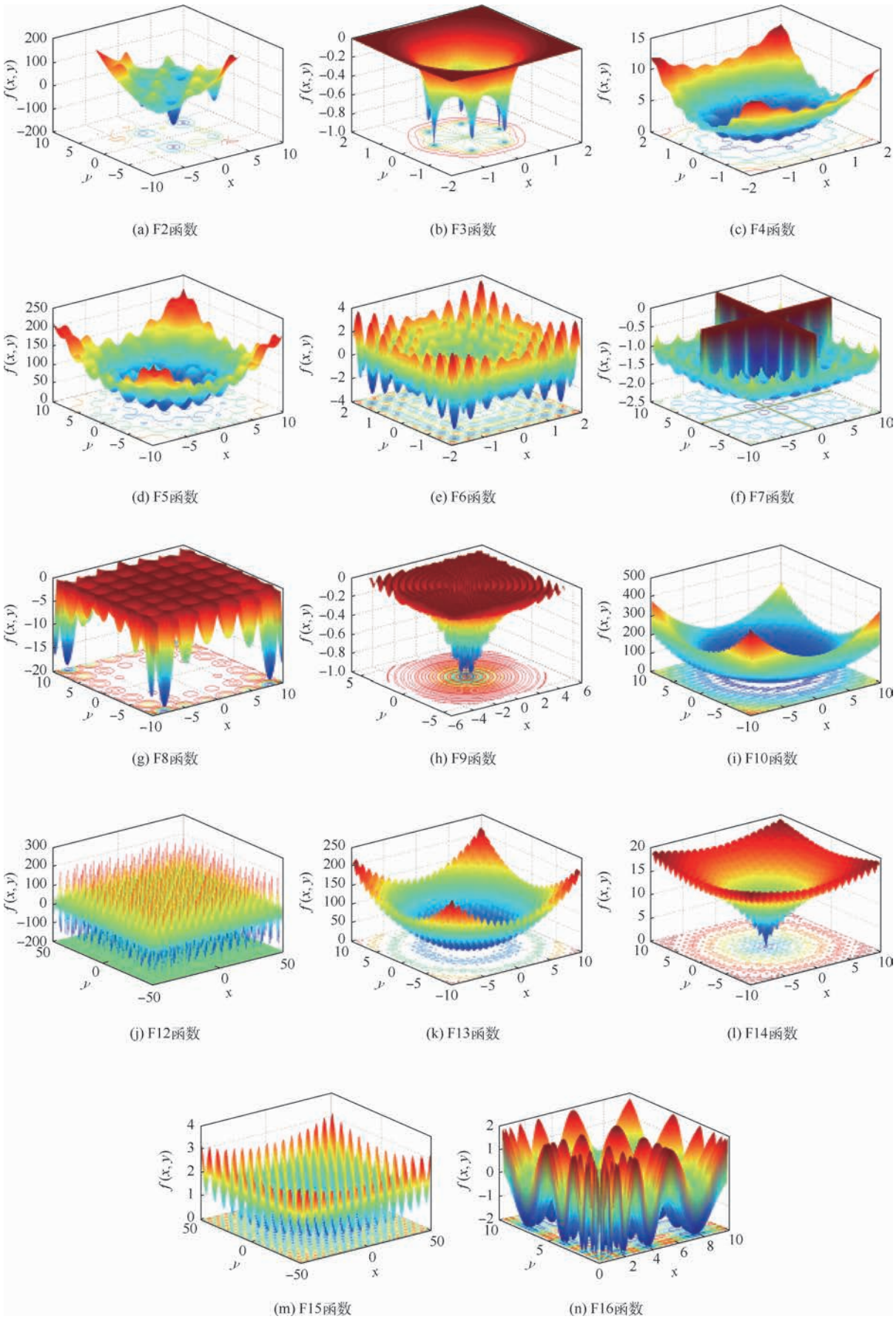


图 15　二维函数 3D 图

Fig. 15　3D figures of 2D functions

北航学报 赠阅

表 3 16 个标准测试函数优化结果对比

Table 3 Comparison of optimization results of 16 standard test functions

函数代号	FPA			CS		
	最优值	平均值	标准差	最优值	平均值	标准差
F1	2.82	5.42	1.63	9.85×10^{-7}	2.37×10^{-3}	5.49×10^{-4}
F2	−106.764 5	−106.762 7	2.24×10^{-3}	−106.764 5	−106.757 0	1.64×10^{-3}
F3	−0.999 8	−0.997 0	1.88×10^{-3}	−0.999 7	−0.994 5	1.21×10^{-3}
F4	6.46×10^{-5}	2.46×10^{-2}	3.49×10^{-2}	2.56×10^{-1}	1.81	3.97×10^{-1}
F5	2.09×10^{-6}	2.41×10^{-3}	5.07×10^{-3}	1.88×10^{-3}	0.53	1.24×10^{-1}
F6	−3.253 9	−3.253 7	3.20×10^{-4}	−3.253 9	−3.253 3	1.29×10^{-4}
F7	−2.062 6	−2.062 3	7.46×10^{-5}	−2.062 6	−2.062 5	2.11×10^{-5}
F8	−19.208 5	−19.208 3	1.73×10^{-4}	−19.208 5	−19.208 2	7.18×10^{-5}
F9	−0.993 3	−0.947 3	1.77×10^{-2}	−0.986 5	−0.936 2	1.77×10^{-2}
F10	1.24×10^{-5}	2.93×10^{-2}	3.70×10^{-2}	4.06×10^{-6}	5.60×10^{-3}	1.40×10^{-3}
F11	−3.004 1	−2.877 1	8.56×10^{-2}	−2.928 0	−2.651 2	7.64×10^{-2}
F12	−186.730 7	−186.712 5	2.68×10^{-2}	−186.730 8	−186.709 8	6.13×10^{-3}
F13	6.60	15.36	5.01	2.93	8.05	1.46
F14	3.87	6.34	1.17	4.66	8.35	0.92
F15	4.64	11.29	2.66	2.73	6.97	1.21
F16	−49.61	−49.22	0.16	−49.84	−49.74	9.33×10^{-2}

函数代号	FA			NS-FPA		
	最优值	平均值	标准差	最优值	平均值	标准差
F1	8.63×10^{-11}	5.79×10^{-9}	5.66×10^{-9}	3.88×10^{-9}	1.82×10^{-5}	2.89×10^{-5}
F2	−106.764 5	−104.939 3	5.23	−106.764 5	−106.764 5	4.76×10^{-6}
F3	−1.000 0	−0.999 8	1.06×10^{-4}	−1.000 0	−0.999 9	7.81×10^{-5}
F4	6.70×10^{-7}	4.88×10^{-4}	8.66×10^{-4}	8.32×10^{-9}	5.00×10^{-5}	3.94×10^{-5}
F5	8.34×10^{-8}	3.59×10^{-6}	5.29×10^{-6}	5.15×10^{-9}	1.88×10^{-7}	1.70×10^{-7}
F6	−3.253 9	−3.253 1	8.20×10^{-4}	−3.253 9	−3.253 9	5.93×10^{-6}
F7	−2.062 6	−2.062 6	1.77×10^{-4}	−2.062 6	−2.062 6	1.30×10^{-8}
F8	−19.208 5	−19.207 9	9.73×10^{-4}	−19.208 5	−19.208 5	4.50×10^{-10}
F9	−1.000 0	−0.994 9	1.75×10^{-2}	−1.000 0	−0.999 9	1.99×10^{-4}
F10	4.68×10^{-9}	1.41×10^{-7}	2.10×10^{-7}	1.26×10^{-10}	6.79×10^{-9}	7.70×10^{-9}
F11	−3.005 4	−2.933 6	0.20	−3.005 4	−3.005 1	3.18×10^{-4}
F12	−186.730 9	−177.764 8	13.30	−186.730 9	−186.730 8	5.50×10^{-4}
F13	2.38×10^{-6}	1.33	1.41	0	1.18×10^{-7}	7.42×10^{-7}
F14	1.80×10^{-3}	3.83×10^{-3}	1.12×10^{-3}	6.46×10^{-12}	1.62×10^{-7}	1.09×10^{-6}
F15	3.00×10^{-4}	7.47×10^{-2}	0.19	2.97×10^{-4}	1.32×10^{-3}	7.53×10^{-4}
F16	−49.87	−9.62	9.83×10^{-2}	−50.00	−49.99	2.51×10^{-4}

首先,分别从多模、超多模函数的寻优结果中对比 FPA 与 NS-FPA 的寻优特性。表 3 中,从多模函数 F2 ~ F8 的寻优对比结果中可以看出,F2 函数:FPA 50 次寻优的最优值能够达到精确解,而平均值却只能够达到小数点后 2 位;NS-FPA 的最优值及平均值都能够达到精确解;再看标准差,FPA 值达到 10^{-3},而 NS-FPA 达到 10^{-6},比 FPA 高 3 个数量级,这说明 NS-FPA 要更稳定。F3 函数:FPA 的最优值精度达到 10^{-3},平均值精度达到 10^{-2};而NS-FPA的最优值为精确解,平均值精度达到 10^{-4},标准差比 FPA 高 2 个数量级。F4 函数:FPA 的最优值精度达到 10^{-5},平均值精度达到 10^{-2};而 NS-FPA 的最优值精度达到 10^{-9},平均值精度达到 10^{-5},标准差比 FPA 高 3 个数量级。F5 函数:FPA 的最优值精度达到 10^{-6},平均值精度达到 10^{-3};而 NS-FPA 的最优值精度达到 10^{-9},平均值精度达到 10^{-7},标准差比 FPA 高 4 个数量级。F6 函数:FPA 的最优值能够达到精确解,平均值精度达到 10^{-3};而 NS-FPA 的最优值及平均值都达到精确解,标准差比 FPA 高 2 个数量级。F7 函数:FPA 的最优值能够达到精确解,平均值精度达到 10^{-3};而 NS-FPA 的最优值及平均值都达到精确解,标准差比 FPA 高 3 个数量级。F8 函数:FPA 的最优值能够达到精确解,平

均值精度达到 10^{-3};而 NS-FPA 的最优值及平均值都达到精确解,标准差比 FPA 高 6 个数量级。

在超多模函数 F9 ~ F16 的寻优对比结果中,F9 函数:FPA 的最优值精度达到 10^{-2},平均值精度达到 10^{-1};而 NS-FPA 的最优值达到精确解,平均值精度达到 10^{-4},标准差比 FPA 高 2 个数量级。F10 函数:FPA 的最优值精度达到 10^{-5},平均值精度达到 10^{-2};而 NS-FPA 的最优值精度达到 10^{-10},平均值精度达到 10^{-9},标准差比 FPA 高 7 个数量级。F11 函数:FPA 的最优值精度达到 10^{-2},平均值精度较差;而 NS-FPA 的最优值达到精确解,平均值精度达到 10^{-3},标准差比 FPA 高 2 个数量级。F12 函数:FPA 的最优值精度达到 10^{-3},平均值精度为 10^{-1};而 NS-FPA 的最优值达到精确解,平均值精度达到 10^{-4},标准差比 FPA 高 2 个数量级。对于高维超多模函数 F13 ~ F16,FPA 基本失去了寻优精度,而 NS-FPA 依旧能够寻优且平均寻优精度均达到 10^{-3} 以上,标准差也达到 10^{-4} 以上,这验证了改进方案的有效性。

从整体上看,NS-FPA 的寻优精度要比 FPA 至少高出 2 个数量级,寻优稳定性(标准差)高出 2 ~ 7 个数量级,这说明相对于 FPA,NS-FPA 能够更好的处理多峰寻优,且具有较高的稳定性。

其次,与 Yang 等[20-21]近几年提出的布谷鸟算法、萤火虫算法进行对比。表 3 中,对于多模函数 F2 ~ F8 的寻优,可以看出 FPA 与 CS 的最优值、平均值、标准差几乎相当。除 F2 函数外,FA 的最优值、平均值、标准差要略微优于 FPA 与 CS,但 NS-FPA 的评价指标都要优于 FPA、CS、FA,至少高2 个数量级的计算精度。对于超多模函数 F9 ~ F16,FPA 与 CS 的寻优效果相当;F8 ~ F9,FA 的寻优效果要劣于 CS 与 FPA;F10 ~ F16,FA 的寻优效果要优于 CS 与 FPA;但 NS-FPA 的寻优效果都要明显优于 FPA、CS、FA。综合来看,NS-FPA 对多模寻优的处理要更优于 FPA、CS、FA。

最后,图 16 给出 F11 函数 50 次优化结果对比,从图中可以直观地看出 NS-FPA 与 FPA、CS、FA 相比寻优结果波动性较小、更加稳定,这说明 NS-FPA 的鲁棒性较好,也间接证实了模拟退火策略能够使得 FPA 成功避免陷入局部最优,从而保证了算法的全局勘探性。图 17 为 F11 函数优化过程进化曲线对比,从图中可以看出 NS-FPA 的寻优结果更靠近理论值 -3,并且收敛速度也要快于 FPA、CS、FA。为了直观地体现各个算法的收敛精度,图18以对数坐标轴的形式给出F13函数优化过程进化曲线对比,从图中可以看出 NS-FPA 的终止计算精度达到了 10^{-15},并且仍有下降的趋势,这体现 NS-FPA 优秀的局部开采能力,也进一步证实了 NS-FPA 的有效性。

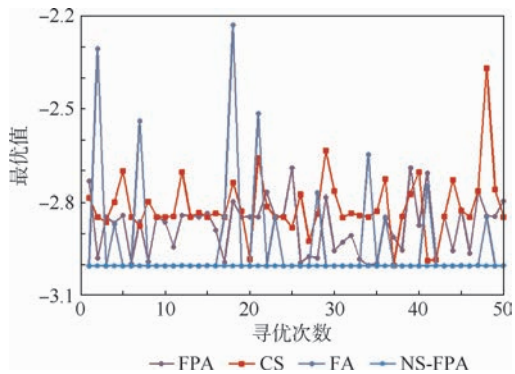


图 16　F11 函数 50 次寻优结果对比

Fig. 16　Comparison of 50 times' optimization results of F11 function

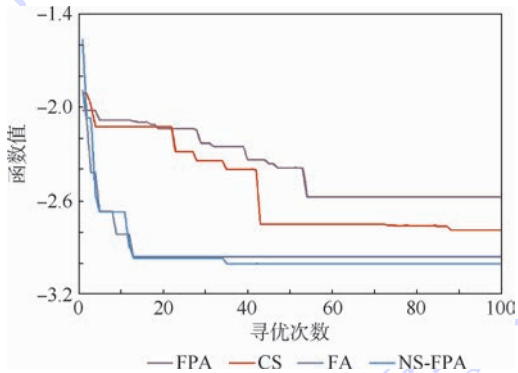


图 17　F11 函数寻优进化曲线

Fig. 17　Optimization evolution curves of F11 function

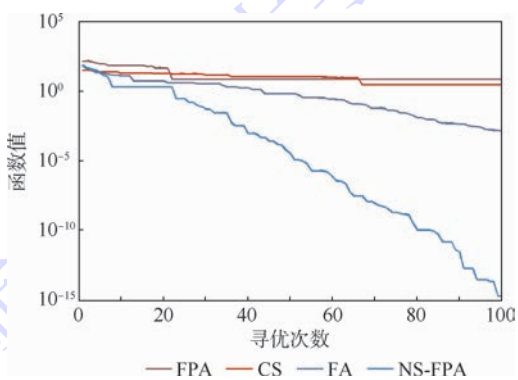


图 18　F13 函数寻优进化曲线

Fig. 18　Optimization evolution curves of F13 function

3.3　NS-FPA 的参数分析

尽管 NS-FPA 的寻优效果较好,与 FPA 相比,NS-FPA 多了 2 个基本参数初温概率 p_0、温度衰减系数 γ。为了讨论其对 NS-FPA 寻优性能的影响,本节以超多模函数 F9 为例,进行仿真计算实验。

首先,讨论初温概率 p_0 对算法性能的影响。

北航学报 赠阅

实验时其他参数不改变,设定不同的 p_0 后,分别对 F9 函数进行 50 次寻优计算,统计结果如表 4 所示(左侧三栏)。从表 4 中可以看出随着 p_0 的增大,算法的平均值不断逼近理论解,且标准差不断变小(标准差越小,算法的鲁棒性越好),直至 p_0 取 0.7 时,达到最优,之后有所下降。当 p_0 取 0.6~0.9 时,平均值、标准差达到高于 10^{-4} 的精度。因此,在寻优计算时建议 p_0 取 0.6~0.9。

其次,讨论温度衰减系数 γ 对算法性能的影响。同样,在实验时其他参数不改变,设定不同的 γ 后,分别对 F9 函数进行 50 次寻优计算,统计结果如表 4 所示(右侧三栏)。从表 4 中可以看出,随着 γ 增大,算法的平均值和标准差逐渐变小,直至 $\gamma=0.8$ 时,达到最小,之后有所增加。当 γ 取 0.7~0.9 时,平均值、标准差达到高于 10^{-4} 的精度。因此,在寻优计算时建议 γ 取 0.7~0.9。

表 4 不同 p_0、γ 参数下的 NS-FPA 寻优结果

Tabel 4 Optimization results of NS-FPA with different p_0 and γ

p_0	平均值	标准差	γ	平均值	标准差
0.1	-0.99808	9.20×10^{-3}	0.1	-0.99670	1.28×10^{-2}
0.2	-0.99844	9.19×10^{-3}	0.2	-0.99911	4.04×10^{-3}
0.3	-0.99861	8.57×10^{-3}	0.3	-0.99915	3.27×10^{-3}
0.4	-0.99914	4.14×10^{-3}	0.4	-0.99965	1.40×10^{-3}
0.5	-0.99935	2.73×10^{-3}	0.5	-0.99972	1.21×10^{-3}
0.6	-0.99981	9.14×10^{-4}	0.6	-0.99978	1.14×10^{-3}
0.7	-0.99999	6.52×10^{-5}	0.7	-0.99983	9.39×10^{-4}
0.8	-0.99993	9.59×10^{-5}	0.8	-0.99998	7.47×10^{-5}
0.9	-0.99986	7.12×10^{-4}	0.9	-0.99989	6.15×10^{-4}

3.4 NS-FPA 的时间复杂度分析

与 FPA 相比,NS-FPA 的局部授粉较为复杂,在算法复杂度上要大于 FPA。为直观地对比 2 种算法的时间复杂度,本节从 3.2 节选择 6 个测试函数分别用 FPA、NS-FPA 在同一平台上单独进行 50 次寻优实验,统计 50 次实验的平均耗时,结果如表 5 所示。

表5中,NS-FPA的平均耗时要大于FPA,且平均高出 32.49%。再考虑到 3.2 节中,对于 FPA 不能求解的多模函数,NS-FPA 却具有较好的寻优性能,笔者认为在时间上多出 32.49% 的消耗是值得的,它增强了算法的求解性能。

表 5 NS-FPA 与 FPA 的平均消耗时间对比

Tabel 5 Comparison of mean consumption time between NS-FPA and FPA

函数代号	平均耗时/s		NS-FPA 与 FPA 平均耗时的百分比
	NS-FPA	FPA	
F5	0.28	0.20	142.56
F6	0.28	0.20	142.76
F11	0.31	0.22	137.15
F13	0.28	0.22	126.89
F14	33.56	27.22	123.31
F15	34.99	28.63	122.24

4 结 论

本文针对多模复杂函数优化问题对 FPA 进行研究,首先提出了种群多样性指标与差异性指标,定性分析了基本 FPA 的寻优缺陷;然后,基于模拟退火的思想解决算法过度贪婪的问题,再结合 Nelder-Mead 单纯形法对局部授粉过程进行重构,提出一种改进的花朵授粉算法——NS-FPA。通过定性、定量的方式对其进行仿真对比分析得出以下结论:

1) 基于模拟退火思想的全局授粉行为能够避免 FPA 陷入局部最优,改善了 FPA 的全局勘探能力,解决了算法早熟的问题。

2) 重构的局部授粉行为改善了算法的局部开采能力,大大提升了算法的收敛速度与收敛精度。

3) NS-FPA 可以用来解决多模复杂优化问题。

4) NS-FPA 在寻优时需要比 FPA 多消耗 32.49% 的时间。

参考文献 (References)

[1] YANG X S. Flower pollination algorithm for global optimization [C]// International Conference on Unconventional Computation and Natural Computation. Berlin: Springer, 2012: 240-249.

[2] YANG X S, KARAMANOGLU M, HE X. Multi-objective flower algorithm for optimization [J]. Procedia Computer Science, 2014, 18(1): 861-868.

[3] YANG X S, KARAMANOGLU M, HE X. Flower pollination algorithm: A novel approach for multiobjective optimization [J]. Engineering Optimization, 2014, 46(9): 194-195.

[4] AL-MA'SHUMAH F, PERMANA D, SIDARTO K A. Solving inverse problem for Markov chain model of customer lifetime value using flower pollination algorithm [J]. Journal of Molecular Structure, 2015, 1692(1): 7-11.

[5] PRATHIBA R, MOSES M B, SAKTHIVEL S. Flower pollination algorithm applied for different economic load dispatch problems [J]. International Journal of Engineering & Technology, 2014, 6(2): 1009-1016.

[6] DUBEY H M, PANDIT M, PANIGRAHI B K. Hybrid flower pollination algorithm with time-varying fuzzy selection mechanism for wind integrated multi-objective dynamic economic dispatch [J]. Renewable Energy, 2015, 83: 188-202.

[7] ABDELAZIZ A Y, ALI E S, ELAZIM S M A. Implementation of flower pollination algorithm for solving economic load dispatch

北航学报 赠阅

and combined economic emission dispatch problems in power systems[J]. Energy,2016,101:506-518.

[8] SHARAWI M,EMARY E,IMANE A,et al. Flower pollination optimization algorithm for wireless sensor network lifetime global optimization[J]. Applied Soft Computing,2014,4(3):2231-2307.

[9] EMARY E,ZAWBAA H M,HASSANIEN A E,et al. Retinal vessel segmentation based on flower pollination search algorithm [M]. Berlin:Springer,2014:93-100.

[10] WANG R,ZHOU Y Q. Flower pollination algorithm with dimension by dimension improvement[J]. Mathematical Problems in Engineering,2014,2014:481791.

[11] ABDEL-RAOUF O, ABDEL-BASET M, EL-HENAWY I. An improved flower pollination algorithm with chaos[J]. International Journal of Education & Management Engineering,2014,4(2):1-8.

[12] ABDEL-RAOUF O,EL-HENAWY I,ABDEL-BASET M. A novel hybrid flower pollination algorithm with chaotic harmony search for solving sudoku puzzles[J]. International Journal of Engineering Trends & Technology,2014,6(3):126-132.

[13] EL-HENAWY I,ISMAIL M. An improved chaotic flower pollination algorithm for solving large integer programming problems [J]. International Journal of Digital Content Technology & Its Applic,2014,8(3):72-79.

[14] ABDEL-RAOUF O,ABDEL-BASET M,EL-HENAWY I. A new hybrid flower pollination algorithm for solving constrained global optimization problems[J]. International Journal of Applied Operational Research,2014,3(2):21-28.

[15] ŁUKASIK S,KOWALSKI P A. Study of flower pollination algorithm for continuous optimization[M]. Berlin:Springer,2015:451-459.

[16] 傅文渊,凌朝东. 布朗运动模拟退火算法[J]. 计算机学报,2014,37(6):1301-1308.
FU W Y,LING C D. Brownian motion based simulated annealing algorithm[J]. Chinese Journal of Computers,2014,37(6):1301-1308(in Chinese).

[17] NELDER J A,MEAD R A. A simplex method for function minimization[J]. The Computer Journal,1965,7(4):308-313.

[18] NAZARETH L,TSENG P. Gilding the lily:A variant of the Nelder-Mead algorithm based on golden-section search[J]. Computational Optimization & Applications,2002,22(1):133-144.

[19] YANG X S. Nature-inspired optimization algorithms[M]. Beckington:Luniver Press,2014:67-75.

[20] YANG X S,DEB S. Cuckoo search:Recent advances and applications[J]. Neural Computing and Applications,2014,24(1):169-174.

[21] YANG X S,HE X. Firefly algorithm:Recent advances and applications[J]. International Journal of Swarm Intelligence,2013,1(1):36-50.

[22] KAIPA K N,GHOSE D. Glowworm swarm optimization:Theory,algorithms,and applications[M]. Berlin:Springer,2017:91-131.

作者简介:

郭庆　男,硕士研究生。主要研究方向:信息系统工程与智能计算等。

惠晓滨　男,博士,教授,博士生导师。主要研究方向:复杂系统建模与仿真、数据分析与智能决策。

张贾奎　男,硕士研究生。主要研究方向:信息系统工程与智能计算等。

李正欣　男,博士,讲师。主要研究方向:复杂系统建模与仿真、数据分析与智能决策。

北京航空航天大学学报

北航学报 赠阅

Improved flower pollination algorithm for multimodal function optimization

GUO Qing, HUI Xiaobin*, ZHANG Jiakui, LI Zhengxin

(Equipment Management and Safety Engineering College, Air Force Engineering University, Xi'an 710051, China)

Abstract: In order to discuss the defects of flower pollination algorithm (FPA) in solving multimodal optimization problems, the optimal disadvantages of flower pollination algorithm in multimodal function optimization were qualitatively analyzed by defining population diversity and difference index. And then a new framework of FPA was constructed by optimizing the global pollination process based on the simulated annealing idea and using Nelder-Mead simplex search method to reconstruct the local pollination process. The simulation results show that the improved flower pollination algorithm can effectively avoid falling into local optimum and has better global exploration and local exploitation abilities, which has advantages to solve multimodal function optimization, compared with primary flower pollination algorithm, cuckoo search algorithm and firefly algorithm.

Keywords: flower pollination algorithm (FPA); simulated annealing; Nelder-Mead simplex method; multimodal function optimization

北京航空航天大学学报

Received: 2017-04-18; **Accepted**: 2017-05-19; **Published online**: 2017-06-19 15:53

URL: kns.cnki.net/kcms/detail/11.2625.V.20170619.1553.005.html

Foundation item: National Natural Science Foundation of China (61502521)

* **Corresponding author**. E-mail: zibai4991@qq.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2017.0281

基于自适应通信拓扑四旋翼无人机编队重构控制

马思迁[1], 董朝阳[1,*], 马鸣宇[1], 王青[2]
(1. 北京航空航天大学 航空科学与工程学院, 北京 100083; 2. 北京航空航天大学 自动化科学与电气工程学院, 北京 100083)

摘 要: 针对四旋翼无人机编队重构协同控制问题,基于切换通信拓扑结构的积分滑模控制(ISMC)方法进行了研究。根据四旋翼无人机间的通信拓扑关系以及编队重构特性,建立了四旋翼无人机编队模型。针对编队重构过程中可能出现的通信可靠性问题,提出了通信拓扑切换条件,结合积分滑模控制理论对切换通信拓扑条件下的协同控制器进行设计,并结合切换系统理论对编队系统稳定性进行了证明。仿真结果表明,无人机编队系统在编队重构过程中采用切换通信拓扑结构以及滑模控制方法能保证系统的稳定性,验证了方法的有效性。

关 键 词: 飞行编队控制; 有向拓扑; 自适应; 切换系统; 积分滑模; 四旋翼无人机
中图分类号: V249; TP273
文献标识码: A **文章编号**: 1001-5965(2018)04-0841-10

近年来,多个体协调控制问题成为众多领域的研究热点[1-2],无人机编队作为其重要的应用方向吸引了大量的关注[3-4]。相比于单个无人机,无人机编队在执行任务时间消耗、任务完成效率、带载能力以及容错性上都具有显著优势[5-6]。在军事领域,无人机编队能根据不同的任务需求调整结构,提高对战场环境的适应性;在民用领域,无人机编队以其高容错性低成本高效地完成任务,有更广阔的应用空间[7-8]。

当前国内外对无人机编队的编队形式以及控制方法有了广泛的研究。文献[9]提出了基于一致性理论的编队控制方法,文献[10]参考鸽群行为,以领航者-跟随者作为其编队飞行控制结构。以上文献采用了领航者-跟随者的结构树形拓扑,人为规定领航者的存在大大降低了编队系统的容错性。文献[11]通过设计微分几何控制律避免了无人机编队重构过程中可能存在的相互碰撞。文献[12-13]分别提出了无人机编队重构控制律使飞行编队快速稳定地收敛至目标状态。文献[9-13]均未考虑编队重构过程中状态变化对编队中通信结构的影响。在通常情况下,协同编队的实现是基于编队节点间稳定的通讯连接,无人机间通信受地形、相对距离及工作频率等多种因素的影响[14-15],因此,在无人机编队重构过程中采用固定的通信拓扑结构而导致信号失真的加剧会对协同编队系统造成不良影响。

采用切换的通信拓扑结构能有效地解决上述问题[16-17]。切换通信拓扑结构使得在编队重构过程中编队间通信结构也随之发生变化,建立可靠通信连接并舍弃不可靠通信连接,从而保证存在相互通信的无人机间通信的可靠性。文献[18]针对节点缺失的编队结构自修复控制方法进行了研究。文献[19]提出了一种基于高度的导航率切换条件。受以上方法的启发,本文将

收稿日期: 2017-05-05; **录用日期**: 2017-08-01; **网络出版时间**: 2017-09-22 10:32
网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20170922.1032.006.html
基金项目: 国家自然科学基金(61374012)
* **通信作者**. E-mail: dongchaoyang@buaa.edu.cn

引用格式: 马思迁, 董朝阳, 马鸣宇, 等. 基于自适应通信拓扑四旋翼无人机编队重构控制[J]. 北京航空航天大学学报, 2018, 44(4): 841-850. MA S Q, DONG C Y, MA M Y, et al. Formation reconfiguration control of quadrotor UAVs based on adaptive communication topology[J]. Journal of Beijing University of Aeronautics and Astronautics, 2018, 44(4): 841-850 (in Chinese).

北航学报 赠阅

切换通信拓扑结构应用于无人机编队重构的过程中,并提出了基于编队无人机间相对位置、通信干扰以及通信频率等多种因素的切换条件,增强了无人机编队在队形及外界环境改变时的通信效率与可靠性。考虑到滑模控制理论对变结构体良好的适用性[20-21],采用积分滑模相关理论设计了针对四旋翼无人机编队的控制器。

本文首先建立编队系统模型,包括四旋翼无人机建模、编队重构描述及切换通信拓扑描述;然后针对四旋翼无人机编队模型设计了编队系统相应的积分滑模控制器,并采用李雅普诺夫函数法对滑模控制器作用下的系统的稳定性进行了分析;最后通过 MATLAB 仿真验证了方法的有效性。

1 问题描述

1.1 四旋翼无人机建模

四旋翼无人机为具有 4 个输入和 6 个状态自由度的欠驱动系统。以相互垂直的一对旋翼臂分别作为四旋翼无人机的机体坐标系 x 轴与 y 轴,取机体坐标系 z 轴向上为正,四旋翼无人机结构图及其对应姿态角定义如图 1 所示。

忽略地球曲率对四旋翼无人机的影响,考虑四旋翼无人机的旋转 3 个姿态角间的相互耦合,根据拉格朗日方程得到无人机在机体轴系下的 6 自由度运动、动力学模型可描述为[22]

$$\begin{cases} \dot{\boldsymbol{P}} = \boldsymbol{V} \\ m\dot{\boldsymbol{V}} = U_1\boldsymbol{R}\boldsymbol{e}_3 - mg\boldsymbol{e}_3 \\ \dot{\boldsymbol{\Theta}} = \boldsymbol{\Omega} \\ \boldsymbol{J}\dot{\boldsymbol{\Omega}} = \boldsymbol{\Gamma} - \boldsymbol{F}_c\boldsymbol{\Omega} \end{cases} \tag{1}$$

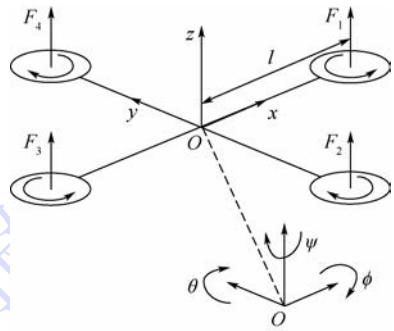


图 1 四旋翼无人机坐标系及姿态角

Fig. 1 Quadrotor UAV coordinate system and attitude angle

式中:$\boldsymbol{V} = [v_x \quad v_y \quad v_z]^{\mathrm{T}}$ 为速度矢量;$\boldsymbol{P} = [x \quad y \quad z]^{\mathrm{T}}$ 为无人机惯性坐标系中质心位置;m 为飞行器质量;g 为重力加速度;$\boldsymbol{\Theta} = [\phi \quad \theta \quad \psi]^{\mathrm{T}}$ 为无人机的旋转姿态角;$\boldsymbol{\Omega} = [p \quad q \quad r]^{\mathrm{T}}$ 为无人机的转动角速度;$\boldsymbol{e}_3 = [0 \quad 0 \quad 1]^{\mathrm{T}}$ 为竖直方向单位向量;U_1 为四旋翼无人机升力输入,通过 4 个旋翼转速同时增加或减小实现;$\boldsymbol{\Gamma}$ 为四旋翼无人机系统旋转力矩,通过 4 个旋翼差动实现。

$\boldsymbol{R}$ 为从机体坐标系到惯性坐标系的转换关系,表达式为

$$\boldsymbol{R} = \begin{bmatrix} \cos\theta\cos\psi & \sin\theta\sin\phi\cos\psi - \cos\phi\sin\psi & \cos\phi\sin\theta\cos\psi + \sin\phi\sin\psi \\ \cos\theta\sin\psi & \sin\phi\sin\theta\sin\psi + \cos\phi\cos\psi & \cos\phi\sin\theta\sin\psi - \sin\phi\cos\psi \\ -\sin\theta & \sin\phi\cos\theta & \cos\phi\cos\theta \end{bmatrix} \tag{2}$$

I_{xx}、I_{yy} 及 I_{zz} 为刚体绕 3 个惯性主轴的转动惯量,$\boldsymbol{J}$ 为刚体惯性张量 $\boldsymbol{I} = \mathrm{diag}(I_{xx}, I_{yy}, I_{zz})$ 在惯性坐标系中的表示,表达式为

$$\boldsymbol{J} = \begin{bmatrix} I_{xx} & 0 & -I_{xx}\sin\theta \\ 0 & I_{yy}\cos^2\phi + I_{zz}\sin^2\phi & J_{23} \\ -I_{xx}\sin\theta & (I_{yy} - I_{zz})\sin\phi\cos\phi\cos\theta & J_{33} \end{bmatrix} \tag{3}$$

式中:$J_{33} = I_{xx}\sin^2\theta + I_{yy}\sin^2\phi\cos^2\theta + I_{zz}\cdot\cos^2\phi\cos^2\theta$;$J_{23} = (I_{yy} - I_{zz})\sin\phi\cos\phi\cos\theta$。

$\boldsymbol{F}_c$ 为科里奥利力以及惯性离心力项,可表示为

$$\boldsymbol{F}_c = \dot{\boldsymbol{J}} - \frac{1}{2}\cdot\frac{\partial}{\partial\boldsymbol{\Theta}}(\dot{\boldsymbol{\Theta}}^{\mathrm{T}}\boldsymbol{J}) \tag{4}$$

1.2 编队重构描述

1.2.1 编队描述

对于三维空间中的任意节点上的四旋翼无人机在三维空间中的位置可由其位置矢量描述。设 $\boldsymbol{r}_j$ 为编队中节点第 i 和 j 个无人机相对于地面惯性坐标系 $O'x'y'z'$ 中的某一点 O'' 的位置矢量,$\boldsymbol{r}_{ij} = \boldsymbol{r}_i - \boldsymbol{r}_j$ 为两节点其相对位置矢量,则其在三维空间中的位置矢量关系如图 2 所示。

则对于给定的四旋翼无人机编队,可由特定各个编队节点无人机的相对位置矢量 $\boldsymbol{r}_{ij}$ 进行描述,$\boldsymbol{P}_i$ 为第 i 个四旋翼无人机的位置矢量,t 为时间。若 n 个无人机满足:

$$\lim_{t\to\infty}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\left\|\boldsymbol{P}_i - \boldsymbol{r}_i - (\boldsymbol{P}_j - \boldsymbol{r}_j)\right\| =$$

北航学报 赠阅

$$\lim_{t\to\infty}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\|\boldsymbol{P}_i-\boldsymbol{P}_j-\boldsymbol{r}_{ij}\|=0 \tag{5}$$

则称这 n 个无人机能组成稳定编队。

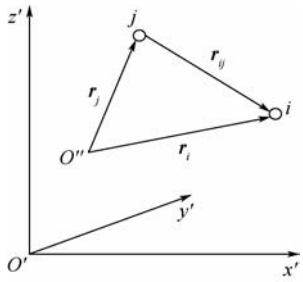


图 2　四旋翼无人机相对位置矢量关系

Fig. 2　Relative position vector relationship of quadrotor UAVs

1.2.2　编队重构描述

在飞行编队执行任务的过程中，由于各种不同的实际需要，需要根据实际情况改变原有的稳定编队形式，实时进行编队重构，则描述编队的相对位置矢量 $\tilde{\boldsymbol{r}}_{ij}=\tilde{\boldsymbol{r}}_{ij}(t)$ 可写作一个实时变化的矢量。本文旨在设计积分滑模控制律，实现四旋翼无人机编队对重构编队队形进行跟踪的目的。若这 n 个四旋翼无人机，满足：

$$\lim_{t\to\infty}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\|\boldsymbol{P}_i-\tilde{\boldsymbol{r}}_i(t)-(\boldsymbol{P}_j-\tilde{\boldsymbol{r}}_j(t))\|=$$

$$\lim_{t\to\infty}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\|\boldsymbol{P}_i-\boldsymbol{P}_j-\tilde{\boldsymbol{r}}_{ij}(t)\|=0 \tag{6}$$

则称这 n 个无人机能对重构编队队形进行跟踪。

1.3　自适应切换通信拓扑

考虑到在无人机飞行编队重构的过程中，采用固定通信拓扑结构会造成实际所需的通信距离与最大通信距离的限制矛盾。本文采用自适应切换通信拓扑的通信结构，旨在针对编队重构过程中的各节点无人机相对位置变化过程中，采用最佳的拓扑通信结构保证各节点无人机的正常通信。

设 $\mathcal{G}=(\boldsymbol{v},\boldsymbol{\varepsilon},\boldsymbol{\mathcal{A}})$ 为一个 n 节点的有向加权图，其中 $\boldsymbol{v}=\{\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_n\}$ 为节点集合，$\boldsymbol{\varepsilon}\subseteq\boldsymbol{v}\times\boldsymbol{v}$ 为边集，$\boldsymbol{\mathcal{A}}=(a_{ij})_{n\times n}$ 为权重邻接矩阵。可将一个 n 节点四旋翼无人机编队间的有向信息传输抽象成一个 n 节点的有向加权图。若 $a_{ij}>0$，则表示边 $(j,i)\in\boldsymbol{\varepsilon}$，即第 i 个节点无人机能从第 j 个节点无人机获取状态信息；若 $a_{ij}=0$，则表示第 i 个节点无人机无法从第 j 个节点无人机获取状态信息。由此定义节点 i 的邻居集为 $N_i=\{j\in\boldsymbol{v}(j,i)\in\boldsymbol{\varepsilon}\}$。定义度矩阵 $\boldsymbol{\mathcal{D}}=\mathrm{diag}\{d_1,d_2,\cdots,d_n\}$，其中 $d_i=\sum_{j\in N_i}a_{ij}$。则可定义 $\mathcal{G}$ 的 Laplacian 矩阵 $\boldsymbol{\mathcal{L}}=\boldsymbol{\mathcal{D}}-\boldsymbol{\mathcal{A}}$，若 $\mathcal{G}$ 为平衡有向图，则 $\mathrm{rank}(\boldsymbol{\mathcal{L}})=n-1$，且有 $n-1$ 个正特征值。

定义 1　第 i 个节点无人机从第 j 个节点无人机获取状态信息通信代价函数 $W(\boldsymbol{v}_i,\boldsymbol{v}_j)$，其与通信距离、通信传输介质、通信频率以及通信过程中信号失真等影响通信可靠性的相关。通信距离越远、通信过程中信号失真越严重，$W(\boldsymbol{v}_i,\boldsymbol{v}_j)$ 取值越大。

将各个节点之间建立单向通信连接的条件集合记为 C_{ij}，Λ_1 为节点之间建立单向通信连接阈值。当 $(\boldsymbol{v}_i,\boldsymbol{v}_j)\in C_{ij}=\{(\boldsymbol{v}_i,\boldsymbol{v}_j)\mid W(\boldsymbol{v}_i,\boldsymbol{v}_j)\leqslant\Lambda_1\}$ 时，建立第 i 个节点无人机从第 j 个节点无人机获取状态信息的通信连接，即 $a_{ij}=0\rightarrow a_{ij}>0$。

将各个节点之间断开单向通信连接的条件集合记为 $\tilde{C}_{ij}$，Λ_2 为节点之间断开单向通信连接阈值。当 $(\boldsymbol{v}_i,\boldsymbol{v}_j)\in\tilde{C}_{ij}=\{(\boldsymbol{v}_i,\boldsymbol{v}_j)\mid W(\boldsymbol{v}_i,\boldsymbol{v}_j)\geqslant\Lambda_2\}$ 时，切断第 i 个节点无人机从第 j 个节点无人机获取状态信息的通信连接，即 $a_{ij}>0\rightarrow a_{ij}=0$。为防止在切换点附近的节点状态参数摄动造成通信系统结构的高频切换，取 $\Lambda_2>\Lambda_1$。

则完整的无人机编队系统通信拓扑切换指令的集合可被定义为各个节点之间单向通信连接的建立与切断通信连接条件的集合 C 可定义为

$$C=\bigcup_{\boldsymbol{v}_i,\boldsymbol{v}_j\in\boldsymbol{v}}(C_{ij}\cup\tilde{C}_{ij}) \tag{7}$$

根据飞行编队在重构过程中各个节点无人机相对位置的变化，编队系统采用与之相匹配的通信拓扑结构。这一过程可以描述为，当飞行编队任意两节点无人机间的通信代价函数 $W(\boldsymbol{v}_i,\boldsymbol{v}_j)$ 达到切换条件，即 $\underset{\boldsymbol{v}_i,\boldsymbol{v}_j\in\boldsymbol{v}}{\forall}(\boldsymbol{v}_i,\boldsymbol{v}_j)\in C$ 时，飞行编队通信自动更新其通信拓扑模式，即 $\boldsymbol{\mathcal{L}}(\mathcal{G}_k)\rightarrow\boldsymbol{\mathcal{L}}(\mathcal{G}_{k+1})$。

2　基于积分滑模控制的协同编队控制器设计及稳定性证明

本文采用积分滑模控制器，四旋翼无人机为具有 4 个输入和 6 个状态自由度的典型欠驱动系统，因此只能设计滑模控制器至多保证 4 个状态自由度的跟踪，其余状态自由度设计控制器保证其稳定。考虑到编队飞行的实际情况，本文通过设计滑模控制器保证 3 个平动状态自由度对给定指令进行跟踪以保证协同编队队形，其余 3 个转动状态自由度则设计控制器保证其稳定。

北航学报 赠阅

2.1 协同编队控制器设计

2.1.1 位置协同控制器设计

为保证系统滑动模态下不降阶,且具有较高的鲁棒性,设计积分滑模面为

$$\boldsymbol{s}_{\mathrm{p}}=\boldsymbol{V}_{\mathrm{e}}+\lambda_{\mathrm{p}}\int\boldsymbol{V}_{\mathrm{e}}\mathrm{d}t \tag{8}$$

式中:λ_{p} 为决定积分滑模面的一个常值;$\boldsymbol{V}_{\mathrm{e}}$ 为速度跟踪误差,且有

$$\boldsymbol{V}_{\mathrm{e}}=\begin{bmatrix}\sum\limits_{j\in N_1}(a_{1j}(\boldsymbol{V}_1-\boldsymbol{V}_j-\dot{\tilde{\boldsymbol{r}}}_{1j}(t)))\\ \sum\limits_{j\in N_2}(a_{2j}(\boldsymbol{V}_2-\boldsymbol{V}_j-\dot{\tilde{\boldsymbol{r}}}_{2j}(t)))\\ \vdots\\ \sum\limits_{j\in N_n}(a_{nj}(\boldsymbol{V}_n-\boldsymbol{V}_j-\dot{\tilde{\boldsymbol{r}}}_{nj}(t)))\end{bmatrix}=$$

$$[\boldsymbol{\mathcal{L}}(\mathcal{G}_k)\otimes\boldsymbol{I}_3]\cdot\begin{bmatrix}\boldsymbol{V}_1\\ \boldsymbol{V}_2\\ \vdots\\ \boldsymbol{V}_n\end{bmatrix}-\begin{bmatrix}\sum\limits_{j\in N_1}(a_{1j}\dot{\tilde{\boldsymbol{r}}}_{1j}(t))\\ \sum\limits_{j\in N_2}(a_{2j}\dot{\tilde{\boldsymbol{r}}}_{2j}(t))\\ \vdots\\ \sum\limits_{j\in N_n}(a_{nj}\dot{\tilde{\boldsymbol{r}}}_{nj}(t))\end{bmatrix} \tag{9}$$

式中:设 $\boldsymbol{V}_i=[v_{xi}\quad v_{yi}\quad v_{zi}]^{\mathrm{T}}$ 为飞行编队中第 i 个节点无人机飞行速度矢量;$\boldsymbol{I}_3$ 为3 阶单位阵。

考虑对于每一个特定的通信拓扑结构,单独的设计各个通信拓扑情况下的控制器,则有

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{\mathcal{L}}(\mathcal{G}_k)=\boldsymbol{0} \tag{10}$$

则对 $\boldsymbol{s}_{\mathrm{p}}=\boldsymbol{V}_{\mathrm{e}}+\lambda_{\mathrm{p}}\int\boldsymbol{V}_{\mathrm{e}}\mathrm{d}t$ 求导可得

$$\dot{\boldsymbol{s}}_{\mathrm{p}}=\dot{\boldsymbol{V}}_{\mathrm{e}}+\boldsymbol{\lambda}_{\mathrm{p}}\boldsymbol{V}_{\mathrm{e}}=[\boldsymbol{\mathcal{L}}(\mathcal{G}_k)\otimes\boldsymbol{I}_3]\cdot$$

$$\begin{bmatrix}\dot{\boldsymbol{V}}_1\\ \dot{\boldsymbol{V}}_2\\ \vdots\\ \dot{\boldsymbol{V}}_n\end{bmatrix}-\begin{bmatrix}\sum\limits_{j\in N_1}(a_{1j}\ddot{\tilde{\boldsymbol{r}}}_{1j}(t))\\ \sum\limits_{j\in N_2}(a_{2j}\ddot{\tilde{\boldsymbol{r}}}_{2j}(t))\\ \vdots\\ \sum\limits_{j\in N_n}(a_{nj}\ddot{\tilde{\boldsymbol{r}}}_{nj}(t))\end{bmatrix}+$$

$$\lambda_{\mathrm{p}}\left([\boldsymbol{\mathcal{L}}(\mathcal{G}_k)\otimes\boldsymbol{I}_3]\cdot\begin{bmatrix}\boldsymbol{V}_1\\ \boldsymbol{V}_2\\ \vdots\\ \boldsymbol{V}_n\end{bmatrix}-\begin{bmatrix}\sum\limits_{j\in N_1}(a_{1j}\dot{\tilde{\boldsymbol{r}}}_{1j}(t))\\ \sum\limits_{j\in N_2}(a_{2j}\dot{\tilde{\boldsymbol{r}}}_{2j}(t))\\ \vdots\\ \sum\limits_{j\in N_n}(a_{nj}\dot{\tilde{\boldsymbol{r}}}_{nj}(t))\end{bmatrix}\right) \tag{11}$$

取虚拟位置控制律:

$$\boldsymbol{u}_{\mathrm{p}}=[(\boldsymbol{u}_{\mathrm{p1}})^{\mathrm{T}}\quad(\boldsymbol{u}_{\mathrm{p2}})^{\mathrm{T}}\quad\cdots\quad(\boldsymbol{u}_{\mathrm{pn}})^{\mathrm{T}}]^{\mathrm{T}} \tag{12}$$

式中:$\boldsymbol{u}_{\mathrm{p}i}=[u_{\mathrm{p}xi}\quad u_{\mathrm{p}yi}\quad u_{\mathrm{p}zi}]^{\mathrm{T}}=U_{i1}\boldsymbol{R}_i\boldsymbol{e}_3$ 为第 i 个四旋翼无人机虚拟控制输入,U_{i1} 为第 i 个无人机总推力输入。

令 $\dot{\boldsymbol{s}}_{\mathrm{p}}=0$,由于 $\boldsymbol{\mathcal{L}}(\mathcal{G}_k)$ 为一个 $n-1$ 阶的非满秩矩阵,导出滑动模态下的等效控制律:

$$\boldsymbol{u}_{\mathrm{eq}}=\boldsymbol{M}\boldsymbol{g}(\boldsymbol{1}\otimes\boldsymbol{e}_3)+\boldsymbol{M}[\boldsymbol{\mathcal{L}}(\mathcal{G}_k)\otimes\boldsymbol{I}_3]^{+}\cdot$$

$$\left(-\lambda_{\mathrm{p}}[\boldsymbol{\mathcal{L}}(\mathcal{G}_k)\otimes\boldsymbol{I}_3]\cdot\begin{bmatrix}\boldsymbol{V}_1\\ \boldsymbol{V}_2\\ \vdots\\ \boldsymbol{V}_n\end{bmatrix}+\begin{bmatrix}\sum\limits_{j\in N_1}a_{1j}(\ddot{\tilde{\boldsymbol{r}}}_{1j}(t)+\lambda_{\mathrm{p}}\dot{\tilde{\boldsymbol{r}}}_{1j}(t))\\ \sum\limits_{j\in N_2}a_{2j}(\ddot{\tilde{\boldsymbol{r}}}_{2j}(t)+\lambda_{\mathrm{p}}\dot{\tilde{\boldsymbol{r}}}_{2j}(t))\\ \vdots\\ \sum\limits_{j\in N_n}a_{nj}(\ddot{\tilde{\boldsymbol{r}}}_{nj}(t)+\lambda_{\mathrm{p}}\dot{\tilde{\boldsymbol{r}}}_{nj}(t))\end{bmatrix}\right) \tag{13}$$

式中:$[\boldsymbol{\mathcal{L}}(\mathcal{G}_k)\otimes\boldsymbol{I}_3]^{+}$ 为 $\boldsymbol{\mathcal{L}}(\mathcal{G}_k)\otimes\boldsymbol{I}_3$ 广义逆;$\boldsymbol{M}=\mathrm{diag}(m_1,m_2,\cdots,m_n)\otimes\boldsymbol{I}_3$ 为质量矩阵与单位阵的克罗内克积;$\boldsymbol{1}^{\mathrm{T}}=[1\quad 1\quad\cdots\quad 1]_{1\times n}$。

为保证系统鲁棒性,获得协同编队虚拟位置控制律:

$$\boldsymbol{u}_{\mathrm{p}}=\boldsymbol{u}_{\mathrm{eq}}-\kappa_{\mathrm{p}}\cdot\mathrm{sgn}(\boldsymbol{s}_{\mathrm{p}}) \tag{14}$$

式中:κ_{p} 为决定外环收敛速度的常数,且满足 $\kappa_{\mathrm{p}}>4U_{1\max}$,$U_{1\max}=\max\{U_{11}^{\lim},U_{12}^{\lim},\cdots,U_{1n}^{\lim}\}$,$U_{1i}^{\lim}$ 为第 i 个四旋翼无人机升力最大输入限制。

2.1.2 姿态控制器设计

根据各个节点无人机虚拟控制输入 $\boldsymbol{u}_{\mathrm{p}i}$,分别计算其总推力输入 U_{i1} 和姿态子系统中间指令姿态角信号 $\boldsymbol{\Theta}_{\mathrm{d}i}=[\phi_{\mathrm{d}i}\quad\theta_{\mathrm{d}i}\quad\psi_{\mathrm{d}i}]^{\mathrm{T}}$。由关系 $\boldsymbol{u}_{\mathrm{p}i}=U_{i1}\cdot\boldsymbol{R}_i\boldsymbol{e}_3$,展开 $\boldsymbol{u}_{\mathrm{p}i}$ 则有

$$\begin{cases}u_{\mathrm{p}xi}=U_{i1}(\cos\phi_i\sin\theta_i\cos\psi_i+\sin\phi_i\sin\psi_i)\\ u_{\mathrm{p}yi}=U_{i1}(\cos\phi_i\sin\theta_i\sin\psi_i-\sin\phi_i\cos\psi_i)\\ u_{\mathrm{p}zi}=U_{i1}\cos\phi_i\cos\theta_i\end{cases} \tag{15}$$

则总推力输入 U_{i1} 和姿态子系统中间指令姿态角信号 $\boldsymbol{\Theta}_{\mathrm{d}i}=[\phi_{\mathrm{d}i}\quad\theta_{\mathrm{d}i}\quad\psi_{\mathrm{d}i}]^{\mathrm{T}}$ 可表示为

$$\begin{cases}\theta_{\mathrm{d}i}=\arctan\left(\dfrac{u_{\mathrm{p}xi}\cos\psi_{\mathrm{d}i}+u_{\mathrm{p}yi}\sin\psi_{\mathrm{d}i}}{u_{\mathrm{p}zi}}\right)\\ \phi_{\mathrm{d}i}=\arctan\left(\cos\theta_{\mathrm{d}i}\dfrac{u_{\mathrm{p}xi}\sin\psi_{\mathrm{d}i}-u_{\mathrm{p}yi}\cos\psi_{\mathrm{d}i}}{u_{\mathrm{p}zi}}\right)\\ U_{i1}=\dfrac{u_{\mathrm{p}zi}}{\cos\phi_{\mathrm{d}i}\cos\theta_{\mathrm{d}i}}\end{cases} \tag{16}$$

对于第 i 个节点无人机,根据其中间指令姿

北航学报 赠阅

态角信号 $\boldsymbol{\Theta}_{\mathrm{di}}$ 设计控制输入转矩 $\boldsymbol{\Gamma}_i$。设计积分滑模面：

$$\boldsymbol{s}_{ai}=\boldsymbol{\Omega}_{ei}+\lambda_{ai}\int\boldsymbol{\Omega}_{ei}\mathrm{d}t \tag{17}$$

式中：$\boldsymbol{\Omega}_{ei}=\boldsymbol{\Omega}_i-\dot{\boldsymbol{\Theta}}_{\mathrm{di}}$ 为第 i 个节点无人机的个各轴角度偏差；λ_{ai} 为决定积分滑模面的一个常值。

令 $\dot{\boldsymbol{s}}_{ai}=\boldsymbol{0}$，导出滑动模态下的等效控制律：

$$\boldsymbol{\Gamma}_{\mathrm{eq}i}=\boldsymbol{J}_i\cdot(-\lambda_{ai}\boldsymbol{\Omega}_i+\lambda_{ai}\dot{\boldsymbol{\Theta}}_{\mathrm{di}}+\ddot{\boldsymbol{\Theta}}_{\mathrm{di}})+\boldsymbol{F}_{ci}\boldsymbol{\Omega}_i \tag{18}$$

为保证系统内环收敛速度高于外环收敛速度，获得协同编队第 i 个节点无人机控制输入转矩控制律：

$$\boldsymbol{\Gamma}_i=\boldsymbol{\Gamma}_{\mathrm{eq}i}-\kappa_{ai}\cdot\mathrm{sgn}(\boldsymbol{s}_{ai})-c_{2i}\boldsymbol{s}_{ai} \tag{19}$$

式中：$\kappa_{ai}>0$，$c_{2i}>0$ 均为决定第 i 个节点无人机内环收敛速度的控制器参数。

2.2　稳定性证明

在给出主要定理之前，有如下引理成立：

引理 1　若 $\boldsymbol{Ax}=\boldsymbol{b}$ 有解，则必有特解：$\boldsymbol{x}_0=\boldsymbol{A}^{+}\boldsymbol{b}$，使得 $\boldsymbol{Ax}_0=\boldsymbol{b}$。且 $\boldsymbol{x}_0$ 为极小范数解。

证明　令 $\boldsymbol{Ax}=\boldsymbol{b}$ 解为 $\boldsymbol{x}_1$，使得 $\boldsymbol{Ax}_1=\boldsymbol{b}$

$$\boldsymbol{Ax}_0=\boldsymbol{AA}^{+}\boldsymbol{b}=\boldsymbol{AA}^{+}\boldsymbol{Ax}_1=\boldsymbol{Ax}_1=\boldsymbol{b} \tag{20}$$

$\boldsymbol{x}_0=\boldsymbol{A}^{+}\boldsymbol{b}$ 为 $\boldsymbol{Ax}=\boldsymbol{b}$ 特解，考虑内积：

$$(\boldsymbol{Ax}_0-\boldsymbol{b},\boldsymbol{Ax})=(\boldsymbol{Ax})^{\mathrm{H}}(\boldsymbol{Ax}_0-\boldsymbol{b})=\boldsymbol{x}^{\mathrm{H}}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{Ax}_0-\boldsymbol{A}^{\mathrm{H}}\boldsymbol{b})=\boldsymbol{0} \tag{21}$$

令 $\boldsymbol{x}=\boldsymbol{Y}$ 为 $\boldsymbol{Ax}=\boldsymbol{0}$ 的通解，则 $\boldsymbol{A}(\boldsymbol{x}_0+\boldsymbol{Y})=\boldsymbol{b}$。故 $\boldsymbol{x}=\boldsymbol{x}_0+\boldsymbol{Y}$ 为 $\boldsymbol{Ax}=\boldsymbol{b}$ 通解。

$\forall$ 通解：$\boldsymbol{x}=\boldsymbol{x}_0+\boldsymbol{Y}$，则有 $\boldsymbol{x}_0\perp\boldsymbol{Y}$，其范数关系可表示为

$$\|\boldsymbol{x}\|_2=\|\boldsymbol{x}_0+\boldsymbol{Y}\|_2\geqslant\|\boldsymbol{x}_0\|_2 \tag{22}$$

证毕

定理 1　考虑 N 个四旋翼无人机组成的系统，每个飞行器的模型由式描述。假定 N 个四旋翼无人机间的有限种形式的通信拓扑均为平衡强连接拓扑结构。那么，根据设计的滑模控制律及切换拓扑条件，多无人机系统能够在队形重构的过程中保证其稳定性，且保证控制输入最小。

证明　对于定理 1 的证明分为 2 步：首先，证明多无人机系统能够在队形重构的过程中保证其稳定性；然后，对控制输入最小进行证明。

1）多无人机系统编队重构稳定性证明

对于四旋翼无人机编队系统，取各个切换通信拓扑状态的公共李雅普诺夫函数：

$$W=W_{\mathrm{p}}+W_{\mathrm{a}} \tag{23}$$

式中：$W_{\mathrm{a}}=\sum_{v_i\in v}(W_{ai})=\sum_{v_i\in v}\left(\frac{1}{2}\boldsymbol{s}_{ai}^{\mathrm{T}}\boldsymbol{K}_{ai}\boldsymbol{s}_{ai}\right)$，$W_{ai}$ 为第 i 个节点无人机姿态李雅普诺夫函数分量，$\boldsymbol{K}_{ai}$ 为 3×3 常值矩阵；$W_{\mathrm{p}}=\frac{1}{2}\boldsymbol{s}_{\mathrm{p}}^{\mathrm{T}}\boldsymbol{K}_{\mathrm{p}}\boldsymbol{s}_{\mathrm{p}}$，$\boldsymbol{s}_{\mathrm{p}}$ 为位置滑模面，$\boldsymbol{K}_{\mathrm{p}}$ 为 $3n\times3n$ 常值矩阵。

为说明系统稳定性，需证明 $W>0$，$\dot{W}<0$。

a）证明 $\dot{W}_{\mathrm{p}}<0$

取 $\boldsymbol{K}_{\mathrm{p}}=\boldsymbol{I}_{3n\times3n}$，则有 $\dot{W}_{\mathrm{p}}=\boldsymbol{s}_{\mathrm{p}}^{\mathrm{T}}\dot{\boldsymbol{s}}_{\mathrm{p}}$，可得

$$\dot{\boldsymbol{s}}_{\mathrm{p}}=\dot{\boldsymbol{V}}_{\mathrm{e}}+\lambda_{\mathrm{p}}\boldsymbol{V}_{\mathrm{e}}=[\boldsymbol{\mathcal{L}}(\mathcal{G}_k)\otimes\boldsymbol{I}_3]\cdot\begin{bmatrix}\dot{\boldsymbol{V}}_1\\\dot{\boldsymbol{V}}_2\\\vdots\\\dot{\boldsymbol{V}}_n\end{bmatrix}-\begin{bmatrix}\sum_{j\in N_1}(a_{1j}\ddot{\tilde{\boldsymbol{r}}}_{1j}(t))\\\sum_{j\in N_2}(a_{2j}\ddot{\tilde{\boldsymbol{r}}}_{2j}(t))\\\vdots\\\sum_{j\in N_n}(a_{nj}\ddot{\tilde{\boldsymbol{r}}}_{nj}(t))\end{bmatrix}+\lambda_{\mathrm{p}}\left([\boldsymbol{\mathcal{L}}(\mathcal{G}_k)\otimes\boldsymbol{I}_3]\cdot\begin{bmatrix}\boldsymbol{V}_1\\\boldsymbol{V}_2\\\vdots\\\boldsymbol{V}_n\end{bmatrix}-\begin{bmatrix}\sum_{j\in N_1}(a_{1j}\dot{\tilde{\boldsymbol{r}}}_{1j}(t))\\\sum_{j\in N_2}(a_{2j}\dot{\tilde{\boldsymbol{r}}}_{2j}(t))\\\vdots\\\sum_{j\in N_n}(a_{nj}\dot{\tilde{\boldsymbol{r}}}_{nj}(t))\end{bmatrix}\right) \tag{24}$$

代入式（11）无人机动力学方程及协同编队虚拟位置控制律：

$$\dot{\boldsymbol{s}}_{\mathrm{p}}=[\boldsymbol{\mathcal{L}}(\mathcal{G}_k)\otimes\boldsymbol{I}_3]\cdot\boldsymbol{M}^{-1}\boldsymbol{M}\cdot[\boldsymbol{\mathcal{L}}(\mathcal{G}_k)\otimes\boldsymbol{I}_3]^{+}\cdot\left(-\lambda_{\mathrm{p}}[\boldsymbol{\mathcal{L}}(\mathcal{G}_k)\otimes\boldsymbol{I}_3]\cdot\begin{bmatrix}\boldsymbol{V}_1\\\boldsymbol{V}_2\\\vdots\\\boldsymbol{V}_n\end{bmatrix}+\begin{bmatrix}\sum_{j\in N_1}(a_{1j}(\ddot{\tilde{\boldsymbol{r}}}_{1j}(t)+\lambda_{\mathrm{p}}\dot{\tilde{\boldsymbol{r}}}_{1j}(t)))\\\sum_{j\in N_2}(a_{2j}(\ddot{\tilde{\boldsymbol{r}}}_{2j}(t)+\lambda_{\mathrm{p}}\dot{\tilde{\boldsymbol{r}}}_{2j}(t)))\\\vdots\\\sum_{j\in N_n}(a_{nj}(\ddot{\tilde{\boldsymbol{r}}}_{nj}(t)+\lambda_{\mathrm{p}}\dot{\tilde{\boldsymbol{r}}}_{nj}(t)))\end{bmatrix}\right)-[\boldsymbol{\mathcal{L}}(\mathcal{G}_k)\otimes\boldsymbol{I}_3]\cdot\kappa_{\mathrm{p}}\cdot\mathrm{sgn}(\boldsymbol{s}_{\mathrm{p}})-$$

北航学报 赠阅

$$\begin{bmatrix}\sum_{j\in N_1}(a_{1j}\ddot{\tilde{r}}_{1j}(t))\\ \sum_{j\in N_2}(a_{2j}\ddot{\tilde{r}}_{2j}(t))\\ \vdots\\ \sum_{j\in N_n}(a_{nj}\ddot{\tilde{r}}_{nj}(t))\end{bmatrix}+[\boldsymbol{\mathcal{L}}(\mathcal{G}_k)\otimes\boldsymbol{I}_3]\cdot$$

$$\begin{bmatrix}U_{11}(\boldsymbol{R}_1-\boldsymbol{R}_{d1})\cdot\boldsymbol{e}_3\\ U_{21}(\boldsymbol{R}_2-\boldsymbol{R}_{d2})\cdot\boldsymbol{e}_3\\ \vdots\\ U_{n1}(\boldsymbol{R}_n-\boldsymbol{R}_{dn})\cdot\boldsymbol{e}_3\end{bmatrix}+\lambda_p\boldsymbol{V}_e \tag{25}$$

式中：$\boldsymbol{R}_{di}$为第 i 个节点以虚拟姿态角为欧拉角的坐标转换关系。

又

$$\boldsymbol{V}_e=[\boldsymbol{\mathcal{L}}(\mathcal{G}_k)\otimes\boldsymbol{I}_3]\cdot\begin{bmatrix}\boldsymbol{V}_1\\ \boldsymbol{V}_2\\ \vdots\\ \boldsymbol{V}_n\end{bmatrix}-$$

$$\begin{bmatrix}\sum_{j\in N_1}(a_{1j}\dot{\tilde{r}}_{1j}(t))\\ \sum_{j\in N_1}(a_{2j}\dot{\tilde{r}}_{2j}(t))\\ \vdots\\ \sum_{j\in N_n}(a_{nj}\dot{\tilde{r}}_{nj}(t))\end{bmatrix}$$

为方便书写，定义以下变换：

$$\boldsymbol{A}=\boldsymbol{\mathcal{L}}(\mathcal{G}_k)\otimes\boldsymbol{I}_3 \tag{26}$$

$$\boldsymbol{b}=-\lambda_p[\boldsymbol{\mathcal{L}}(\mathcal{G}_k)\otimes\boldsymbol{I}_3]\cdot\begin{bmatrix}\boldsymbol{V}_1\\ \boldsymbol{V}_2\\ \vdots\\ \boldsymbol{V}_n\end{bmatrix}+$$

$$\begin{bmatrix}\sum_{j\in N_1}(a_{1j}(\ddot{\tilde{r}}_{1j}(t)+\lambda_p\dot{\tilde{r}}_{1j}(t)))\\ \sum_{j\in N_2}(a_{2j}(\ddot{\tilde{r}}_{2j}(t)+\lambda_p\dot{\tilde{r}}_{2j}(t)))\\ \vdots\\ \sum_{j\in N_n}(a_{nj}(\ddot{\tilde{r}}_{nj}(t)+\lambda_p\dot{\tilde{r}}_{nj}(t)))\end{bmatrix} \tag{27}$$

$$\boldsymbol{\delta}=\begin{bmatrix}U_{11}(\boldsymbol{R}_1-\boldsymbol{R}_{d1})\cdot\boldsymbol{e}_3\\ U_{21}(\boldsymbol{R}_2-\boldsymbol{R}_{d2})\cdot\boldsymbol{e}_3\\ \vdots\\ U_{n1}(\boldsymbol{R}_n-\boldsymbol{R}_{dn})\cdot\boldsymbol{e}_3\end{bmatrix} \tag{28}$$

则有 $\dot{\boldsymbol{s}}_p=\boldsymbol{A}\boldsymbol{A}^+\boldsymbol{b}-\boldsymbol{A}\cdot[\kappa_p\cdot\mathrm{sgn}(\boldsymbol{s}_p)-\boldsymbol{\delta}]-\boldsymbol{b}$，根据引理 1，$\boldsymbol{x}_0=\boldsymbol{A}^+\boldsymbol{b}$ 为方程极小范数解。

$$\dot{\boldsymbol{s}}_p=-[\boldsymbol{\mathcal{L}}(\mathcal{G}_k)\otimes\boldsymbol{I}_3]\cdot[\kappa_p\cdot\mathrm{sgn}(\boldsymbol{s}_p)-\boldsymbol{\delta}] \tag{29}$$

又 $\forall\boldsymbol{v}_1\in\boldsymbol{v}$

$$\boldsymbol{R}_i\cdot\boldsymbol{e}_3=\begin{bmatrix}\cos\phi_i\sin\theta_i\cos\psi_i+\sin\phi_i\sin\psi_i\\ \cos\phi_i\sin\theta_i\sin\psi_i-\sin\phi_i\cos\psi_i\\ \cos\phi_i\cos\theta_i\end{bmatrix} \tag{30}$$

$$\boldsymbol{R}_{di}\cdot\boldsymbol{e}_3=\begin{bmatrix}\cos\phi_{di}\sin\theta_{di}\cos\psi_{di}+\sin\phi_{di}\sin\psi_{di}\\ \cos\phi_{di}\sin\theta_{di}\sin\psi_{di}-\sin\phi_{di}\cos\psi_{di}\\ \cos\phi_{di}\cos\theta_{di}\end{bmatrix} \tag{31}$$

$$|(\boldsymbol{R}_i-\boldsymbol{R}_{di})\cdot\boldsymbol{e}_3|\leqslant|\boldsymbol{R}_i\boldsymbol{e}_3|+|\boldsymbol{R}_{di}\boldsymbol{e}_3|\leqslant\begin{bmatrix}4\\4\\2\end{bmatrix} \tag{32}$$

根据定义，$\kappa_p>4U_{1max}$。

$$\dot{\boldsymbol{s}}_p=-[\boldsymbol{\mathcal{L}}(\mathcal{G}_k)\otimes\boldsymbol{I}_3][(\kappa_p-4U_{1max})\mathrm{sgn}(\boldsymbol{s}_p)+(4U_{1max}\cdot\mathrm{sgn}(\boldsymbol{s}_p)-\boldsymbol{\delta})] \tag{33}$$

编队中可能的通信拓扑结构均为平衡强连接拓扑结构，$\boldsymbol{\mathcal{L}}(\mathcal{G}_k)$始终为 $n-1$ 阶矩阵，且有 $n-1$ 个正特征值。即 $\mathrm{rank}[\boldsymbol{\mathcal{L}}(\mathcal{G}_k)\otimes\boldsymbol{I}_3]=3n-3$，且有 $3n-3$ 个正特征值。

$$\dot{W}_p=-\boldsymbol{s}_p^{\mathrm{T}}[\boldsymbol{\mathcal{L}}(\mathcal{G}_k)\otimes\boldsymbol{I}_3]\cdot\{(\kappa_p-4U_{1max})\cdot\mathrm{sgn}(\boldsymbol{s}_p)+[4U_{1max}\cdot\mathrm{sgn}(\boldsymbol{s}_p)-\boldsymbol{\delta}]\}<-(\kappa_p-4U_{1max})\cdot\boldsymbol{s}_p^{\mathrm{T}}[\boldsymbol{\mathcal{L}}(\mathcal{G}_k)\otimes\boldsymbol{I}_3]\cdot\mathrm{sgn}(\boldsymbol{s}_p)<0 \tag{34}$$

b) $\forall\boldsymbol{v}_i\in\boldsymbol{v}$，证明 $\dot{W}_{ai}<0$

$\forall\boldsymbol{v}_i\in\boldsymbol{v}$，$W_{ai}=\frac{1}{2}\boldsymbol{s}_{ai}^{\mathrm{T}}\boldsymbol{K}_{ai}\boldsymbol{s}_{ai}$，取 $\boldsymbol{K}_{ai}=\boldsymbol{I}_6$ 则 $\dot{W}_{ai}$ 可写为

$$\dot{W}_{ai}=\boldsymbol{s}_{ai}^{\mathrm{T}}\dot{\boldsymbol{s}}_{ai}=\boldsymbol{s}_{ai}^{\mathrm{T}}\cdot[\dot{\boldsymbol{\Omega}}_i-\ddot{\boldsymbol{\Theta}}_{di}+\lambda_{ai}(\boldsymbol{\Omega}_i-\dot{\boldsymbol{\Theta}}_{di})] \tag{35}$$

又 $\boldsymbol{J}_i\dot{\boldsymbol{\Omega}}_i=\boldsymbol{\Gamma}_i-\boldsymbol{C}_i\boldsymbol{\Omega}_i$，$\boldsymbol{J}_i$ 为刚体惯性张量 $\boldsymbol{I}_i=\mathrm{diag}(I_{xx_i},I_{yy_i},I_{zz_i})$ 在惯性坐标系中的表示，$\boldsymbol{J}_i$ 特征值均为正值，故 $\boldsymbol{J}_i$ 的逆存在。

$$\dot{\boldsymbol{\Omega}}_i=\boldsymbol{J}_i^{-1}\cdot(\boldsymbol{\Gamma}_i-\boldsymbol{C}_i\boldsymbol{\Omega}_i) \tag{36}$$

代入第 i 个节点无人机控制输入转矩控制律则有

$$\dot{W}_{ai}=\boldsymbol{s}_{ai}^{\mathrm{T}}\dot{\boldsymbol{s}}_{ai}=\boldsymbol{s}_{ai}^{\mathrm{T}}\cdot\{\boldsymbol{J}_i^{-1}\cdot[\boldsymbol{J}_i\cdot(-\lambda_{ai}\dot{\boldsymbol{\Omega}}_i+\lambda_{ai}\ddot{\boldsymbol{\Theta}}_{di}+\dddot{\boldsymbol{\Theta}}_{di})-\kappa_{ai}\cdot\mathrm{sgn}(\boldsymbol{s}_{ai})-c_{2i}\boldsymbol{s}_{ai}]-\dddot{\boldsymbol{\Theta}}_{di}+\lambda_{ai}(\dot{\boldsymbol{\Omega}}_i-\ddot{\boldsymbol{\Theta}}_{di})\}=-\boldsymbol{s}_{ai}^{\mathrm{T}}\cdot\boldsymbol{J}_i^{-1}\cdot(\kappa_{ai}\cdot\mathrm{sgn}(\boldsymbol{s}_{ai})+c_{2i}\boldsymbol{s}_{ai})<0 \tag{37}$$

由此可得 $\dot{W}_{ai}<0$。

c) 综合 a)和 b)证明可得

$$W>0,\dot{W}<0 \tag{38}$$

编队系统稳定性得证。

2) 控制输入最小证明

对于一切能满足 $\dot{\boldsymbol{s}}_p=\boldsymbol{0}$ 的虚拟等效控制量组

北航学报 赠阅

成的集合 $C_{u_{peq}}$ 中的元素:

$$\boldsymbol{u}_{p}^{-}=[(\boldsymbol{u}_{p1}^{-})^{T}\quad(\boldsymbol{u}_{p2}^{-})^{T}\quad\cdots\quad(\boldsymbol{u}_{pn}^{-})^{T}]^{T}\tag{39}$$

式中:$\boldsymbol{u}_{pi}^{-}=[u_{pxi}^{-}\quad u_{pyi}^{-}\quad u_{pzi}^{-}]^{T}=U_{i1}\boldsymbol{R}_{i}\boldsymbol{e}_{3}$ 为第 i 个四旋翼无人机可能的虚拟控制输入。

根据 $\boldsymbol{s}_{p}$ 定义可得,代入动力学方程式(1)及 $\boldsymbol{u}_{p}^{-}$ 定义式(39),根据等效虚拟控制率求解定义 $\dot{\boldsymbol{s}}_{p}=\boldsymbol{0}$ 则有

$$\dot{\boldsymbol{s}}_{p}=[\boldsymbol{\mathcal{L}}(\mathcal{G}_{k})\otimes\boldsymbol{I}_{3}]\cdot\boldsymbol{M}^{-1}\cdot[\boldsymbol{u}_{p}^{-}-\boldsymbol{M}g(\boldsymbol{1}\otimes\boldsymbol{e}_{3})]-\left(-\lambda_{p}[\boldsymbol{\mathcal{L}}(\mathcal{G}_{k})\otimes\boldsymbol{I}_{3}]\cdot\begin{bmatrix}\boldsymbol{V}_{1}\\\boldsymbol{V}_{2}\\\vdots\\\boldsymbol{V}_{n}\end{bmatrix}+\begin{bmatrix}\sum_{j\in N_{1}}(a_{1j}(\ddot{\tilde{\boldsymbol{r}}}_{1j}(t)+\lambda_{p}\dot{\tilde{\boldsymbol{r}}}_{1j}(t)))\\\sum_{j\in N_{2}}(a_{2j}(\ddot{\tilde{\boldsymbol{r}}}_{2j}(t)+\lambda_{p}\dot{\tilde{\boldsymbol{r}}}_{2j}(t)))\\\vdots\\\sum_{j\in N_{n}}(a_{nj}(\ddot{\tilde{\boldsymbol{r}}}_{nj}(t)+\lambda_{p}\dot{\tilde{\boldsymbol{r}}}_{nj}(t)))\end{bmatrix}\right)=\boldsymbol{0}\tag{40}$$

为方便书写代入变换式(26)和式(27)则有:

$$\boldsymbol{A}\cdot\boldsymbol{M}^{-1}\cdot[\boldsymbol{u}_{p}^{-}-\boldsymbol{M}g(\boldsymbol{1}\otimes\boldsymbol{e}_{3})]-\boldsymbol{b}=\boldsymbol{0}\tag{41}$$

即

$$\boldsymbol{A}\cdot\boldsymbol{M}^{-1}\cdot\boldsymbol{u}_{p}^{-}=\boldsymbol{A}\boldsymbol{M}^{-1}\boldsymbol{M}g(\boldsymbol{1}\otimes\boldsymbol{e}_{3})+\boldsymbol{b}\tag{42}$$

可能的通信拓扑结构均为平衡强连接拓扑结构,$\boldsymbol{\mathcal{L}}(\mathcal{G}_{k})$ 始终为 $n-1$ 阶矩阵,且有 $n-1$ 个正特征值。即 $\operatorname{rank}\boldsymbol{A}=\operatorname{rank}[\boldsymbol{\mathcal{L}}(\mathcal{G}_{k})\otimes\boldsymbol{I}_{3}]=3n-3$,且有 $3n-3$ 个正特征值。因此,方程有无限组解。

若设 $\boldsymbol{A}\cdot\boldsymbol{M}^{-1}\cdot[\boldsymbol{u}_{p}^{-}-\boldsymbol{M}g(\boldsymbol{1}\otimes\boldsymbol{e}_{3})]=\boldsymbol{0}$ 的通解为 $\boldsymbol{u}_{p}^{-}=\boldsymbol{u}_{\delta}$。

根据引理 1,方程组的解集 $C_{u_{peq}}$ 中的元素满足 $\boldsymbol{u}_{p}^{-}=\boldsymbol{u}_{eq}+\boldsymbol{u}_{\delta}$,且有 $\boldsymbol{u}_{eq}\perp\boldsymbol{u}_{\delta}$。

其范数关系可表示为

$$\|\boldsymbol{u}_{p}^{-}\|_{2}=\|\boldsymbol{u}_{eq}+\boldsymbol{u}_{\delta}\|_{2}\geqslant\|\boldsymbol{u}_{eq}\|_{2}\tag{43}$$

式中:$\boldsymbol{u}_{eq}$ 定义由式(13)给出,为本文所使用的等效虚拟控制率。

考察第 i 个四旋翼无人机升力 U_{i1},关系 $\boldsymbol{u}_{pi}^{-}=[u_{pxi}^{-}\quad u_{pyi}^{-}\quad u_{pzi}^{-}]^{T}=U_{i1}\boldsymbol{R}_{i}\boldsymbol{e}_{3}$,对等式两边同取模可得

$$U_{i1}=\|\boldsymbol{u}_{pi}^{-}\|_{2}\tag{44}$$

结合式(43),控制输入最小得证。　　**证毕**

3　仿真与分析

为了验证本文方法的有效性,在 MATLAB 环境下进行仿真验证。考虑含有 5 个四旋翼无人机的编队系统,各无人机初始参数如表 1所示。

只考虑无人机间的距离对无人机间通信的影响,选取状态信息通信代价函数 $W(\boldsymbol{v}_{i},\boldsymbol{v}_{j})=\|\tilde{\boldsymbol{r}}_{ij}(t)\|_{2}$,切换阈值 $\Lambda_{1}=3.5$,$\Lambda_{2}=5.5$。当各个节点之间存在通信时,设置对应拓扑权重如表 2所示,行编号为发射端编号,列编号为接收端编号。

期望队形如图 3 所示,当 $0<t\leqslant 9$ s 时,$\boldsymbol{r}_{12}=[0\quad 0\quad 2]^{T}$,$\boldsymbol{r}_{23}=[0\quad 0\quad 2]^{T}$,$\boldsymbol{r}_{34}=[0\quad 0\quad 2]^{T}$,$\boldsymbol{r}_{45}=[0\quad 0\quad 2]^{T}$。

当 $t>9$ s 时,$\boldsymbol{r}_{12}=[0\quad 3\quad 0]^{T}$,$\boldsymbol{r}_{13}=[3\quad 1.5\quad 0]^{T}$,$\boldsymbol{r}_{14}=[1\quad 1.5\quad -2]^{T}$,$\boldsymbol{r}_{15}=[1\quad 1.5\quad 2]^{T}$。

四旋翼无人机基本参数 $I_{xx}=0.004\ \text{kg}\cdot\text{m}^{2}$,$I_{yy}=0.004\ \text{kg}\cdot\text{m}^{2}$,$I_{zz}=0.008\ \text{kg}\cdot\text{m}^{2}$,$m=3$ kg。

图4给出了四旋翼无人机空间飞行轨迹、升力

表 1　无人机初始状态

Table 1　Initial conditions of UAV

无人机编号	位置/m	速度/(m·s^{-1})
1	[3,1,50]	[-0.1,0,1]
2	[-1,3,50]	[-0.3,-0.1,1]
3	[2,-1,50]	[-0.1,-0.1,1]
4	[1,1,50]	[-0.1,0,1]
5	[0,0,52]	[0,-0.1,0]

表 2　有向拓扑权重

Table 2　Weight of directed topology

无人机编号	1	2	3	4	5
1		0.77	0.56	0.69	0.27
2	0.78		0.74	0.66	0.53
3	0.63	0.75		0.73	0.62
4	0.55	0.68	0.80		0.82
5	0.31	0.49	0.61	0.74	

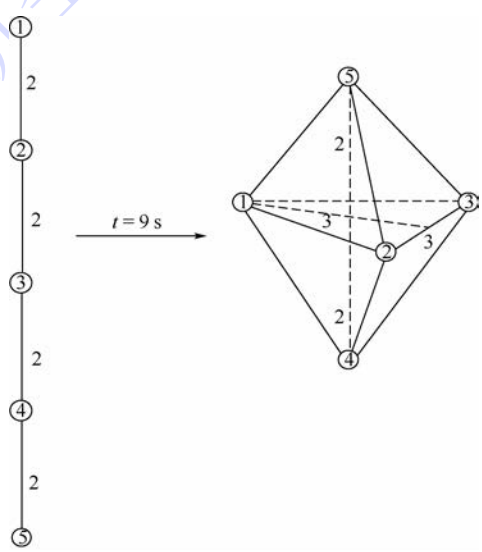


图 3　四旋翼无人机期望队形

Fig. 3　Desired formation of quadrotor UAVs

北航学报 赠阅

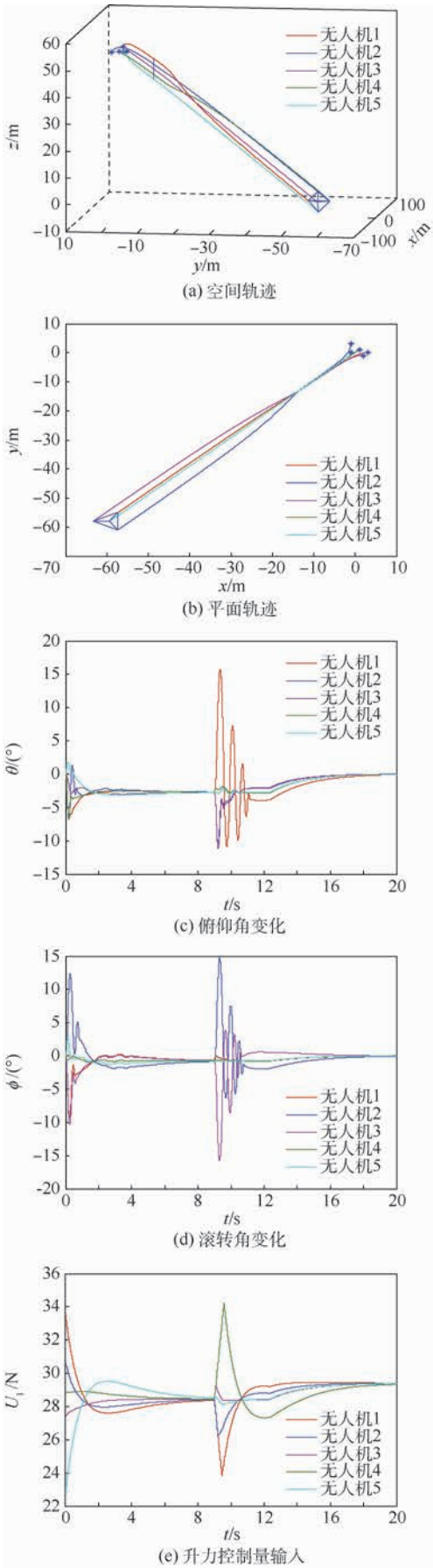


图 4 四旋翼无人机编队重构仿真结果

Fig. 4 Reconfiguration simulation results of quadrotor UAV formation

控制量以及相应的参数变化。从四旋翼无人机编队的空间飞行轨迹和平面飞行轨迹可以看出，在切换拓扑的通信协议下采用积分滑模控制方法能生成给定编队并完成编队重构，验证了方法的有效性。

图 4(c)和(d)为四旋翼无人机俯仰角和滚转角的变化，从图中可以看出，四旋翼无人机在编队重构及拓扑切换的过程中，能连续平滑的完成编队的重构并保持良好的一致性。

在编队重构过程中，其对应的拓扑结构切换可表示为图 5 所示的形式。可以看出，在状态信息通信代价函数 $W(\boldsymbol{v}_i,\boldsymbol{v}_j)=\|\tilde{\boldsymbol{r}}_{ij}(t)\|_2$ 的作用下，通信拓扑结构随着四旋翼无人机相对位置的变化而改变，并在这样的切换通信拓扑结构下，编队系统可以保持系统良好的一致性。

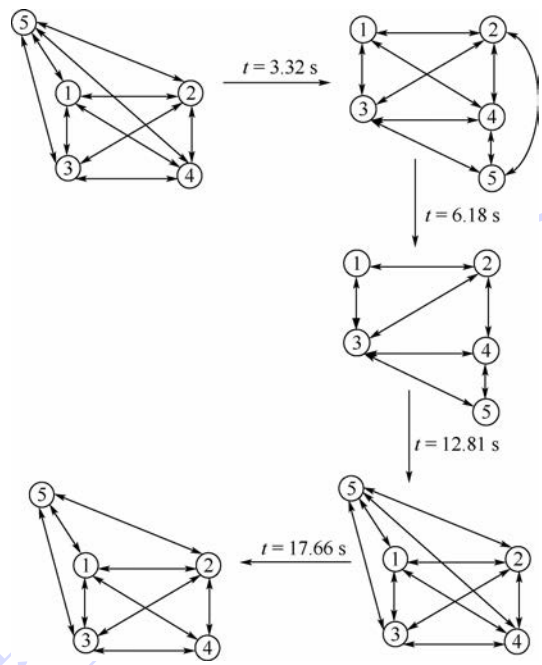


图 5 四旋翼无人机编队通信拓扑结构

Fig. 5 Communication topology of quadrotor UAV formation

4 结 论

本文提出了基于有向切换通信拓扑的四旋翼无人机编队重构的积分滑模控制方法，并验证了方法的有效性，其具体结论如下：

1）通过引入代价函数，使无人机通信拓扑结构能随无人机编队结构发生变化，削弱了由于信息传输环境变化对通信造成的影响。

2）采用无领航者的编队通信拓扑结构，有效的防止了领航者-跟随者编队结构中领航者失效引发整个编队系统崩溃的情况。

3）积分滑模控制方法可在整个编队重构以

及拓扑切换的过程中保持其滑模面不发生改变，从而保证了控制器结构的一致性。

4）仿真结果算例表明，无人机编队系统在编队重构过程中采用切换通信拓扑结构以及滑模控制方法能保证系统的稳定性，验证了方法的有效性。

参考文献（References）

[1] 闵海波，刘源，王仕成，等. 多个体协调控制问题综述[J]. 自动化学报，2012，38(10)：1558-1570.
MIN H B，LIU Y，WANG S C，et al. An overview on coordination control problem of multi-agent system[J]. Acta Automatica Sinica，2012，38(10)：1558-1570(in Chinese).

[2] 关永强，纪志坚，张霖，等. 多智能体系统能控性研究进展[J]. 控制理论与应用，2015，32(4)：421-431.
GUAN Y Q，JI Z J，ZHANG L，et al. Recent developments on controllability of multi-agent systems[J]. Control Theory & Applications，2015，32(4)：421-431(in Chinese).

[3] 樊琼剑，杨忠，方挺，等. 多无人机协同编队飞行控制研究现状[J]. 航空学报，2009，30(4)：683-691.
FAN Q J，YANG Z，FANG T，et al l. Research status of coordinated formation flight control for multi-UAVs[J]. Acta Aeronautica et Astronautica Sinica，2009，30(4)：683-691(in Chinese).

[4] OH K K，PARK M C，AHN H S. A survey of multi-agent formation control[J]. Automatica，2015，53：424-440.

[5] 周绍磊，康宇航，史贤俊，等. 基于 RQPSO-DMPC 的多无人机编队自主重构方法[J]. 北京航空航天大学学报，2017，43(10)：1960-1971.
ZHOU S L，KANG Y H，SHI X J，et al. Autonomous reconfiguration control method for multi-UAVs formation based on RQPSO-DMPC[J]. Journal of Beijing University of Aeronautics and Astronautics，2017，43(10)：1960-1971(in Chinese).

[6] 马鸣宇，董朝阳，王青，等. 基于事件驱动的多飞行器编队协同控制[J]. 北京航空航天大学学报，2017，43(3)：506-515.
MA M Y，DONG C Y，WANG Q，et al. Event-based cooperative control of multiple fight vehicles formation[J]. Journal of Beijing University of Aeronautics and Astronautics，2017，43(3)：506-515(in Chinese).

[7] 刘跃峰，张安. 有人机/无人机编队协同任务分配方法[J]. 系统工程与电子技术，2010，32(3)：584-588.
LIU Y F，ZHANG A. Cooperative task assignment method of manned/unmanned aerial vehicle formation[J]. Systems Engineering and Electronics，2010，32(3)：584-588(in Chinese).

[8] 李远. 多 UAV 协同任务资源分配与编队轨迹优化方法研究[D]. 长沙：国防科技大学，2011：1-14.
LI Y. Research on resources allocation and formation trajectories optimization for multiple UAVs cooperation mission[D]. Changsha：National University of Defense Technology，2011：1-14(in Chinese).

[9] 邢关生，杜春燕，宗群，等. 基于一致性的小型四旋翼机群自主编队分布式运动规划[J]. 控制与决策，2014，29(11)：2081-2084.
XING G S，DU C Y，ZONG Q，et al. Consensus-based distributed motion planning for autonomous formation of miniature quadrotor groups[J]. Control and Decision，2014，29(11)：2081-2084(in Chinese).

[10] 邱华鑫，段海滨，范彦铭. 基于鸽群行为机制的多无人机自主编队[J]. 控制理论与应用，2015，32(10)：1298-1304.
QIU H X，DUAN H B，FAN Y M. Multiple unmanned aerial vehicle autonomous formation based on the behavior mechanism in pigeon flocks[J]. Control Theory & Applications，2015，32(10)：1298-1304(in Chinese).

[11] JOONGBO S，YOUDAN K，TSOURDOS A，et al. Multiple UAV formation reconfiguration with collision avoidance guidance via different geometry concept[C]//28th International Congress of the Aeronautical Sciences. Birsbane：ICAS，2012.

[12] 茹常见，魏瑞轩，戴静，等. 基于纳什议价的无人机编队自主重构控制方法[J]. 自动化学报，2013，39(8)：1349-1359.
RU C J，WEI R X，DAI J，et al. Autonomous reconfiguration control method for UAV's formation based on nash bargain[J]. Acta Automatica Sinica，2013，39(8)：1349-1359(in Chinese).

[13] FERNANDEZ C A，BELMONTE L M，MORALES R，et al. Generalized proportional integral control for an unmanned quadrotor system[J]. International Journal of Advanced Robotic Systems，2015，2：1-14.

[14] 郝妍娜，洪志良. 基于 MCU 和 nRF905 的低功耗远距离无线传输系统[J]. 集成电路应用，2007(8)：44-47.
HAO Y N，HONG Z L. A low-power long-range wireless data transmission platform based on MCU and nRF905 chip[J]. Application of Integrated Circuits，2007(8)：44-47(in Chinese).

[15] 彭颖，王高才，黄书强，等. 移动网络中基于最优停止理论的数据传输能耗优化策略[J]. 计算机学报，2016，39(6)：1162-1175.
PENG Y，WANG G G，HUANG S Q，et al. An energy consumption optimization strategy for data transmission based on optimal stopping theory in mobile networks[J]. Chinese Journal of Computer，2016，39(6)：1162-1175(in Chinese).

[16] RERA O S，RICHARD M M. Consensus problems in networks of agents with switching topology and time-delays[J]. IEEE Transactions on Automatic Control，2004，49(9)：1520-1533.

[17] LIU X M，LIN H，CHEN B M. A graph-theoretic characterization of structural controllability for multi-agent system with switching topology[C]//Joint 48th IEEE Conference on Decision and Control and 28th Chinese Control Conference. Piscataway，NJ：IEEE Press，2009：7012-7017.

[18] 张飞，陈卫东. 移动机器人编队自修复的切换拓扑控制[J]. 控制理论与应用，2010，27(3)：289-295.
ZHANG F，CHEN W D. Switched topology control for self-healing of mobile robot formation[J]. Control Theory & Applications，2010，27(3)：289-295(in Chinese).

[19] DANIEL R W，ROBBERTO F. Terminal guidance for lunar landing and retargeting using a hybrid control strategy[J]. Journal of Guidance，Control，and Dynamics，2016，39(5)：1168-1172.

[20] GALZI D，SHTESSEL Y. Closed-coupled formation flight control using quasi-continuous high-order sliding-mode [C]//

北航学报 赠阅

American Control Conference. Piscataway, NJ: IEEE Press, 2007:1799-1804.

[21] 刘金琨. 滑模变结构控制 MATLAB 仿真[M]. 2 版. 北京:清华大学出版社,2012:211-325.
LIU J K. Sliding mode control design and MATLAB simulation [M]. 2nd ed. Beijing: Tsinghua University Press, 2012: 211-325(in Chinese).

[22] RAFFO G V, ORTEGA M G, RUBIO F R. An integral predictive/nonlinear H_∞ control structure for a quadrotor helicopter [J]. Automatica, 2010, 46(1): 29-39.

作者简介:

马思迁 男,硕士研究生。主要研究方向:飞行器协同控制。

董朝阳 男,博士,教授,博士生导师。主要研究方向:飞行器设计,导航、制导与控制。

Formation reconfiguration control of quadrotor UAVs based on adaptive communication topology

MA Siqian[1], DONG Chaoyang[1,*], MA Mingyu[1], WANG Qing[2]

(1. School of Aeronautic Science and Engineering, Beijing University of Aeronautics and Astronautics, Beijing 100083, China;
2. School of Automation Science and Electrical Engineering, Beijing University of Aeronautics and Astronautics, Beijing 100083, China)

Abstract: This paper focuses on the cooperative control problem in formation reconfiguration of multiple flight vehicles, and the method of integral sliding mode control (ISMC) based on switching communication topology is investigated. The multiple flight vehicle system is modeled involving the connecting of communication topology and the character of formation reconfiguration of quadrotor UAV. The switching condition of the communication topology is raised considering the possible communication reliability problem in the process of formation reconfiguration. The cooperative controller is designed by ISMC method under the condition of switching topology communication, and its stability is proved adopting switching system theory. The simulation results show that the stability of UAV formation system can be ensured with ISMC method and switching communication topology in the process of formation reconfiguration, and illustrate the effectiveness of the proposed method.

Keywords: flight formation control; directed topology; adaptive; switching system; integral sliding mode; quadrotor UAV

Received: 2017-05-05; **Accepted**: 2017-08-01; **Published online**: 2017-09-22 10:32
URL: kns.cnki.net/kcms/detail/11.2625.V.20170922.1032.006.html
Foundation item: National Natural Science Foundation of China (61374012)
* **Corresponding author**. E-mail: dongchaoyang@buaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2017.0282

基于八叉树的简化构型三角片等值面削减算法

徐雷，王华锋*，潘海侠，林广艳，陈栎曦
（北京航空航天大学 软件学院，北京 100083）

摘 要：采用简化构型的SMC算法相比标准MC算法能够有效减少构成等值面的三角片的数量，但因为其仅是在体元内部的简化，所以不能较好地利用数据集表面局部形态特征。针对这一问题进一步提出OSMC算法。其根据简化构型的特点，首先采用八叉树结构组织体元，然后采用自底向上的合并策略合并节点，最后实现局部区域三角片合并。实验证明：OSMC算法能够实现比SMC算法更多的三角片削减，尤其对于具有较多平坦区域的数据集效果显著，其对公开数据集数据的平均削减率为55.1%，而SMC算法为29.7%，在面对高分辨率的地质数据时其最高削减率达到了80%，平均也超过了50%，同时OSMC算法能够更好地适应数据集分辨率的增长。

关 键 词：等值面；MC算法；SMC算法；八叉树；超体元
中图分类号：TP391
文献标识码：A **文章编号**：1001-5965(2018)04-0851-11

可视化技术是诸多数据密集型领域的一种重要研究手段。近年来，等值面算法在三维数据场可视化方面得到了广泛地应用和发展。Lorensen和Cline[1]于1987年提出的MC(Marching Cubes)算法作为等值面提取最为流行的经典算法，其原理简单易于实现，自从被提出以来，得到了业界广泛地使用，包括生物化学[2]、生物医学[3-4]、数字雕刻[5]、环境科学[6]、自然现象呈现[7]以及可视化算法分析[8]等。但经典的MC算法也有一些不足之处，如存在拓扑结构的二义性[9]、对高分辨率数据集计算效率低[10-11]及产生的结果集规模庞大[12]等问题。

近年来，针对MC算法的改进层出不穷。本文着眼于MC算法产生结果集规模庞大这一问题，介绍了现阶段主要的改进方法，并且介绍了一种基于简化构型的SMC(Simplified Marching Cubes)算法。根据简化构型的特点，针对SMC算法的两方面问题提出了一种改进的OSMC(Octree based Simplified Marching Cubes)算法，依次介绍了算法的八叉树构建、八叉树收缩和三角片提取3步。在实验与讨论的部分通过对MC算法、SMC算法、OSMC算法三者实验数据的对比，从表面形态、三角片削减率和分辨率增长3个方面说明OSMC算法的优势并探讨了其原因。

1 相关研究成果

近年来，越来越多的高分辨率数据集被投入使用。MC算法产生庞大的三角片数量给后期的渲染造成了很大压力，同时也加重了存储负担。不少研究人员针对这一问题，提出了诸多解决方案。总体上看，这一问题的解决方案大致分为两类。

第一类是对生成的网格数据采取网格削减算法。如Schroeder等[13]于1992年提出的渐进网

收稿日期：2017-05-05；**录用日期**：2017-09-01；**网络出版时间**：2017-10-13 13:58
网络出版地址：kns.cnki.net/kcms/detail/11.2625.V.20171013.1358.002.html
* **通信作者**. E-mail：wanghuafeng@buaa.edu.cn

引用格式：徐雷，王华锋，潘海侠，等. 基于八叉树的简化构型三角片等值面削减算法[J]. 北京航空航天大学学报，2018，44(4)：851-861. XU L，WANG H F，PAN H X，et al. Octree based decimation algorithm for triangle isosurface using simplified patterns [J]. Journal of Beijing University of Aeronautics and Astronautics，2018，44(4)：851-861 (in Chinese).

北航学报 赠阅

格算法与 Garland 和 Heckbert[14] 于 1997 年提出的二次误差边折叠算法。这类算法的输入是多边形网格数据,是在规模较大的网格数据已经被生成之后再采用一定方式来删除顶点面片等几何元素,是一种事后补救方法。这类方法通常处理费时,并且建立辅助结构会占据大量存储空间,并不是针对本文提出问题的理想解决方案,因此不作过多探讨。

第二类是采取避免生成较多三角片的表面重建方法。相比于事后削减的方法,这一类方法利用了图像数据的特点,效率更高。Montani 等[15] 提出了一种离散化的 DiscMC 算法,这种算法通过将插值点选择在体元边的中点,提高了相邻体元间的三角形共面的概率。然后再将共面的三角形合并成平面多边形,以此来减少三角片。其在 2000 年又进一步改进了此方法,提高了 DiscMC 算法对等值面的逼近程度[16]。Shu 等[17] 于 1995 年提出一种自适应的 MC 算法,这种算法在执行过程中不断地重复对体数据划分子体,直到在子体内寻找到相对于等值面的误差在一定范围内逼近表面。这样就能够实现使用较少的三角片来表示等值面。Shekhar 等[18] 于 1996 年针对 Shu 等[17] 的方法削减程度不足的问题又提出了一种使用了八叉树结构削减三角片的算法,其中使用了表面追踪技术,避免了处理空体元。在体元合并中评估误差,保证误差在给定阈值之内,最后修补缝隙、生成三角片。八叉树是在表面重建算法中重要的一种数据结构,能够用于代表隐函数[19-20] 和自适应提取等值面[21-22]。

上述几种方法均是利用了等值面局部形态特点来进行三角片削减,Cui 和 Liu[23] 于 2008 年与 Vignoles 等[24] 于 2011 年各自提出了一种不同于上述几种方法简化思路的算法。2 种算法实现细节有所不同,但其核心思想一致,都使用了一种简化后的体元构型,为方便起见将文献[13-14]提出的方法统一称作 SMC 算法。使用简化后的构型,减少了部分体元中提取三角片的数量,从而削减了三角片数,同时提高了计算效率,并且不会存在二义性问题。SMC 算法在对低分辨率数据使用时产生的表面与 MC 算法差别显著,逼近程度低于 MC 算法,故未能有广泛应用。但随着近年来高分辨率数据得到越来越多的应用,简化构型的优势便体现出来。在高分辨率数据上,SMC 算法和 MC 算法产生的表面的形态差别变得几乎可以忽略不计,故其为 MC 算法的合理替代方案,并且广泛应用于实践之中。但由于简化构型仅是在体元内部的简化,无法考虑等值面局部形态特征,因而 SMC 算法存在较大的改进空间。主要体现在如下 2 个方面:

1)对于部分数据集,SMC 算法削减率较低。原因在于并非所有的简化构型都比 MC 算法构型具有更少的三角片数,若数据场内有相当比例活动体元正属于这类构型,则 SMC 算法无法削减这些体元的三角片数。

2)无法很好地应对分辨率的提高。原因在于体数据分辨率提高后体元数量会大大增加,仅是体元内部的简化无法应对体元总数的增加带来的影响。

本文提出的 OSMC 算法,既利用了简化构型的优点,同时也考虑了等值面局部形态特征。OSMC 算法与 SMC 算法产生的表面形状完全相同,但三角片数量更少。并且对于具有较多平坦区域的数据集,OSMC 算法能有效地弥补 SMC 算法削减程度不足的问题,同时还能够较好地应对数据集分辨率的增长。

2 OSMC 算法

2.1 简化构型

SMC 算法中的简化构型是基于 MC 算法构型做出的简化模型。

假设等值面 $s_c=\{(x,y,z):F(x,y,z)=c\}$,其中 s_c 为等值面集合,(x,y,z) 为点的坐标,F 为映射函数,c 为等值面的值。若等值面穿过了2 个相邻体素所构成的边,体素位置坐标分别为 $V_l(x_l,y_l,z_l)$ 和 $V_h(x_h,y_h,z_h)$,相应体素值分别为 $I_l<c$ 和 $I_h\geqslant c$,在 MC 算法中插值点的位置应采用式(1)计算:

$$V=\frac{I_h-c}{I_h-I_l}V_l+\frac{c-I_l}{I_h-I_l}V_h \tag{1}$$

若将插值点的位置总取为 V_h,即

$$V=V_h \tag{2}$$

则会使一部分三角形的顶点位置重合,如此操作,便会导致生成的三角形产生退化。对 MC 算法的所有基本构型采取式(2)的操作即形成简化构型,即 SMC 算法构型。图 1 和图 2 中对比显示了 15 种 MC 算法构型与其对应的 SMC 算法构型。可以看出在简化构型中一部分三角形在顶点移动后退化消失,导致多种构型产生的三角形减少甚至消失,也直接导致最终产生的等值面的三角形的数量会相应减少。

将基于简化构型提取出三角片所构成的表面成为 SMC 表面。SMC 表面相比于 MC 表面有如

北航学报 赠阅

下 3 个特点:①SMC 表面的所有三角片的顶点都是体素点;②组成 SMC 表面的三角片只有 3 种可能形状和有限种法线方向;③与 MC 表面相比,SMC 表面位置略向高于等值面的方向偏移,形成一定的系统误差,此误差随着数据集分辨率的提高而减少。

OSMC 算法是基于八叉树的特性来实现局部共面三角片的合并。OSMC 算法一共分为3 步:八叉树构建、八叉树收缩以及三角片提取,流程如图 3 所示。2.2 ~2.4 节分别对这 3 个步骤进行详细说明。

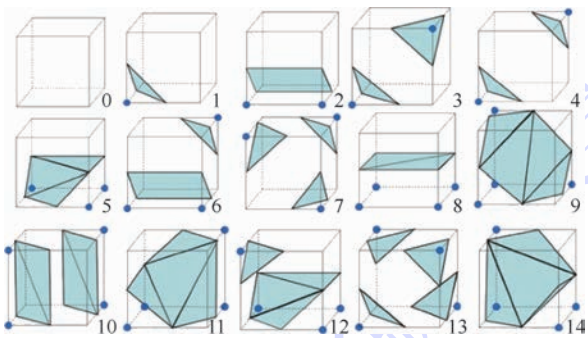


图 1　MC 算法构型

Fig. 1　MC algorithm configuration

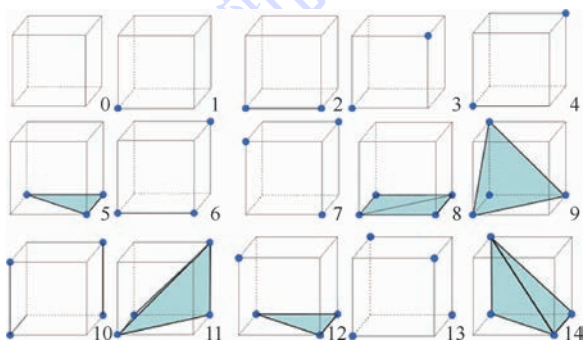


图 2　SMC 算法构型

Fig. 2　SMC algorithm configuration

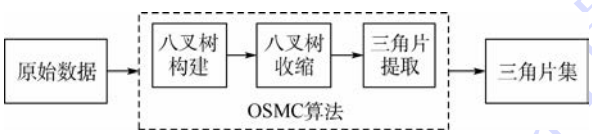


图 3　OSMC 算法流程

Fig. 3　Flowchart of OSMC algorithm

2.2　八叉树构建

2.2.1　活动体元寻找方法

无论是 MC 算法还是 SMC 算法,插值和提取三角片等计算都是针对活动体元。活动体元的 8 个顶点体素根据其体素值大小共有 256 种有限的情况,每一种情况即是一个体元状态(cube configuration)。若令体元 C 的 8 个端点体素值依次为 $I_0, I_1, \cdots, I_7$,则体元状态的计算方法如下:

$$B_i = \begin{cases} 0 & I_i < \text{isovalue} \\ 2^i & I_i < \text{isovalue} \end{cases} \tag{3}$$

$$\text{config}(C) = B_0 \vee B_1 \cdots \vee B_7 \tag{4}$$

式中:B_i 为体元状态;isovalue 为等值面的值;I_i 为体素值;∨ 符号代表位运算或。

config(C)不为 0 或 255 的体元即为活动体元。标准 MC 算法是通过全维度扫描所有体元,逐个求取体元状态 config(C)来寻找活动体元,而往往活动体元只占全部体元总数的很小一部分,这导致算法近 80% 的计算时间花费在访问非活动体元上。为了提高时间效率,不少算法采用其他方式去寻找活动体元,例如文献[21]采用八叉树组织图像数据,文献[18]的表面追踪。但这些方法都有一定的局限性,往往只适用于某些有特殊性的场合。例如文献[21]的方法主要针对一次预处理、多次重建的应用场合;而文献[18]的方法必须人工选择种子体元作为输入参数。一般场合下,数据场体素值分布情况应当是完全未知的,任何一个体元都必须至少被遍历到一次才能获取其体元状态。所以为了不失一般性,OSMC 算法采用遍历所有体元的方式寻找活动体元。实际上,任何方法只要能利用实际应用场合的特点找到活动体元,都可以替代全部遍历的方式。采用何种方式取决于对算法时间效率和通用性的权衡。

2.2.2　为活动体元创建节点

对每一个活动体元,需要为其创建对应的节点。OSMC 算法采用指针式八叉树,其任意节点 N 均包含父节点 p 和 8 个子节点 $c_0, c_1, \cdots, c_7$ 的指针以及一个数据域 data,为了让每个节点总能指代自然数表示的立方体范围,八叉树对空间按 2 的幂划分。因此 OSMC 算法的八叉树 T 所指代的空间尺寸应为大于体数据最长维的最小的 2 的幂。设数据场为:DF = $\{(x,y,z), x \in [0,w], y \in [0,h], z \in [0,d]\}$,则 T 的深度计算式为

$$n = \text{depth}(T) = \lceil \text{lb}(\max\{w+1, h+1, d+1\}) \rceil \tag{5}$$

则 T 实际所能表示的最大空间包围盒 $R(T)$ 在 3 个坐标轴上范围均为 $[0, 2^n - 1]$。对 T 中任意非叶节点,其子节点对应着将父节点空间范围各个维度等分之后的子立方体空间范围。依据这样的划分方式,利用节点之间的父子关系即可以求出 N 所指代的空间包围盒 $R(N)$。对于叶子节点,$R(N)$的 3 个维度的上下界均相等,即有

$$R(N_{\text{leaf}}) = \{(x_{\min}, y_{\min}, z_{\min})\} \tag{6}$$

因而叶节点 N_{leaf} 可以与一个三维坐标对应。

北航学报 赠阅

树的初始状态只包含一个根节点。每当找到一个活动体元 $C(x,y,z)$,为其建立叶节点需要确定其在各层的层内索引 k_i,$0 \leqslant i \leqslant n-1$。层内索引反映了从根节点建立叶节点的路径,利用八叉树空间划分与体元坐标二进制位的对应关系,能够求得层内索引。层内索引的求法如下:

$$k_i = 4 \times [x \gg (n-i) \wedge 1] + 2 \times [y \gg (n-i) \wedge 1] + 1 \times [z \gg (n-i) \wedge 1] \tag{7}$$

式中:符号"∧"代表位运算与;符号"≫"代表位运算逻辑右移。

本步骤执行到对所有活动体元都有对应的节点为止。经过此步形成的八叉树有如下 2 个特点:①八叉树的所有叶节点与活动体元坐标一一对应。②所有的叶子节点均处在同一深度。

2.3 八叉树收缩

八叉树收缩是一个自底向上迭代的过程。从八叉树 depth(T) -1 层节点开始,尝试将子节点的信息归并到父节点,相应节点内的体元合并成为更大的"超体元"(super cell),若成功则删除子节点。这一步骤的关键是判断节点能够收缩的条件。为此首先提出"共面构型"的概念,共面构型是指满足如下任意条件之一的体元构型:

1)该体元为空体元。

2)该体元构型内含有三角片且三角片均在同一平面。

3)该体元构型内不含三角片,但其互补构型内含有三角片且三角片均在同一平面。

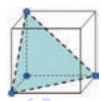
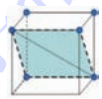
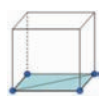
图 1 和图 2 中符合条件 2)的构型均为第 8、9 号构型;符合条件 3)的为第 1、2 号构型;其余 10 种构型为非共面构型。

OSMC 算法收缩节点的基本原理是:对八叉树上的一个节点 N,若其子节点 $c_0,c_1,\cdots,c_7$ 中分别提取的三角片构成集合 T_c,T_c 中所有三角片均共面且有相同的法向量,则此三角片集合组成的形状等价于对 N 提取三角片集合 T_N 所组成的形状,且一定有 Card(T_N) ≤ Card(T_c) 成立。这样使用更少的三角片组成相同平面区域就能实现三角片数量的削减。显然只有属于共面构型的 c_0,$c_1,\cdots,c_7$ 才能满足此条件,共面构型为有限种可枚举的情况,因此上述原理不难被证明。这样节点收缩的问题便转化为子节点中提取的三角片是否共面的问题。

在空间解析几何中,平面方程一般形式为 $Ax+By+C'z=D$,三元组(A,B,C')是决定法线方向的参数。而 D 值反映了其空间位置。2 个三角片若共面,则其一定具有相同的法线方向和 D 值。通过 2.1 节对构型的分析可以得知,共面构型内三角片所在的平面方程形式只可能是如表 1 的 13 种可能。

表 1 平面方程类型

Table 1 Types of plane equation

构 型	三角片类型	平面方程类型
	正三角形	$x+y+z=c$ $x+y-z=c$ $x-y-z=c$ $x-y+z=c$
	一般直角三角形	$x+y=c$ $x-y=c$ $y+z=c$ $y-z=c$ $x+z=c$ $x-z=c$
	等腰直角三角形	$x=c$ $y=c$ $z=c$

对于位置在数据场原点的共面构型的体元 $C_0(0,0,0)$,三角片的方程 $A_0x+B_0y+C_0z=D_0$ 是由体元状态 config(C_0)唯一决定的。因为方程形式只有有限种,所以可以建立两者的函数映射关系:方程:config(C)→(A_c,B_c,C_c,D_c)和法向量:config(C)→(A_c,B_c,C_c)。一般地,对于共面构型体元 $C(x_c,y_c,z_c)$,其三角片所在平面方程相对 C_0 存在一个平移关系,方程解析式为

$$A_c(x-x_c)+B_c(y-y_c)+C_c(z-z_c)=D_c \tag{8}$$

令 $D=A_cx_c+B_cy_c+C_cz_c+D_c$,则四元组($A_c$,$B_c$,$C_c$,$D_c$)就能够反映体元 C 中三角片相对于整个数据场 F 的平面方程。又由于具有映射关系:方程:config(C)→(A_c,B_c,C_c),所以二元组 < config(C),D > 即能充分反映体元的三角片位置信息。八叉树的所有叶节点,由于和活动体元一一对应,所以其形式为

$$N_{\text{leaf}} = \{p, c_0, c_1, \cdots, c_7, < \text{config}(C), D >\} \tag{9}$$

综上所述,八叉树 T 上任意一个节点 N 能收缩,其子节点 $c_0,c_1,\cdots,c_7$ 必须满足如下条件:

1)若 N 为叶节点,非空 c_i 为共面构型。

2)若 N 不为叶节点,非空 c_i 为收缩成功的节点。

3)对于任意非空的 $c_i,c_j,i\neq j$,有

$$\text{normal}(\text{config}(c_i)) = \text{normal}(\text{config}(c_j)) \tag{10}$$

4)对于任意非空的 $c_i,c_j,i\neq j$,有 $D_i=D_j$。

收缩操作首先对倒数第 2 层节点执行,每成功执行一次收缩操作,当前节点记录下的从子节点继承来的 D 值,并利用 config(c_0),config(c_1),…,

北航学报 赠阅

config(c_7)计算 config(N),同时删掉子节点。当某个节点收缩操作失败后,其父节点也不再进行收缩,直到所有的节点都不能再收缩后结束。可以采用 FIFO 的容器实现这样自底向上的迭代操作。

2.4　三角片提取

八叉树 T 收缩完成后,叶节点不再都处于同一深度,一部分节点对应单体元,即处在 depth(T)层的节点。而在 depth(T)层之上的叶节点对应的是“超体元”。超体元即是收缩了下层节点后形成的广义体元。对单体元和超体元,提取三角片的方法有所不同。OSMC 算法本步骤采用程序遍历八叉树叶节点,依次判断节点属于单个体元还是超体元。对其分别用不同的方法求取三角片集合 T_N,直到所有叶节点都被访问为止。

2.4.1　对单体元的三角片提取

节点 N 对应单个体元即说明 N 的父节点收缩失败。由 2.1 节对 SMC 算法的分析可知其生成三角片的顶点都位于体素点上。对 N 的三角片提取可直接使用 SMC 算法的查找表,根据体元状态索引即可获取三角片体元顶点的组成方式,而三角片顶点的位置可由节点空间范围 $R(N)$ 确定,这样就得到三角形集合 T_N,即

$$(\text{smctable}(\text{config}(N)),R(N)) \rightarrow T_N \tag{11}$$

2.4.2　对超体元的三角片提取

超体元节点 N 中虽然三角片的顶点依然是体素点,但不一定是超体元的 8 个顶点。如图 4 所示。

从该超体元对应的八叉树节点数据域中可以获得三角片空间位置信息 $<\text{config}(N),D_N>$,则三角片所在的平面方程参数 $<A_N,B_N,C_N,D_N>$ 便可以求取。因为三角片的顶点实在超体元边上,需要使用标准 MC 算法的查找表来确定被穿过的体元边集合 E_N 和三角片的集合 T_N 的信息,即

$$E_N = \text{mctable}(\text{config}(N)) \tag{12}$$

$$(\text{config}(N),E_N) \rightarrow T_N \tag{13}$$

之后再采用空间解析几何的方法,通过 E_N 中每一条体元边 e_i 的直线方程与平面方程 $A_Nx + B_Ny + C_Nz = D_N$ 联立求解三角片顶点位置,公式如下:

$$\boldsymbol{V}_i = \begin{cases} \left(x_i, y_i, \dfrac{D_N - A_Nx_i - B_Ny_i}{C}\right) & e_i \text{ 平行于 } x \text{ 轴} \\ \left(\dfrac{D_N - B_Ny_i - C_Nz_i}{A}, y_i, z_i\right) & e_i \text{ 平行于 } y \text{ 轴} \\ \left(x_i, \dfrac{D_N - A_Nx_i - C_Nz_i}{B}, z_i\right) & e_i \text{ 平行于 } z \text{ 轴} \end{cases} \tag{14}$$

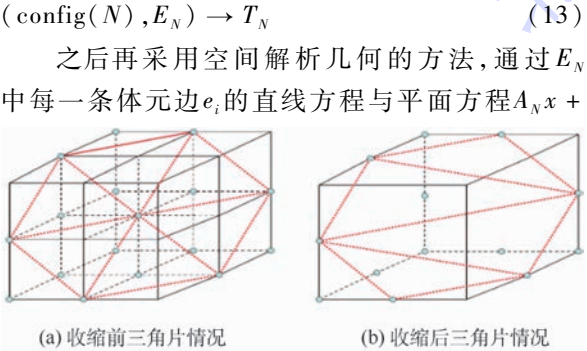


图 4　超体元提取三角片

Fig. 4　Triangle extraction for super cell

3　实验与讨论

本节将通过实验证明 OSMC 算法的优越性。本文算法采用 C + + 实现,实验硬件平台为 Intel(R) Xeon(R) X5650 6 核 CPU,内存 32 GB,操作系统为 64 位 Windows 7。所有实验均在相同的软硬件环境下进行。

3.1　公开数据集实验

OSMC 算法被应用于公开数据集上进行测试。图 5 为 MC 算法、SMC 算法和 OSMC 算法分别对尺寸为 256 像素 × 256 像素 × 128 像素的 Engine数据集重建表面后形成的模型预览图。可以看出,表面的整体形态上三者区别并不明显,而将图像的局部进行放大后,如图 6 所示,可以看出 MC 表面与 SMC 表面细节上略有差别,而 OSMC 表面与 SMC 表面是完全一致的。也就是 OSMC 表面对等值面的逼近误差与 SMC 算法是相同的。该误差产生的原因是因为使用了 SMC 算法简化构型,在分辨率不算高的 Engine 数据上能够在细节上体现出与 MC 表面的区别。

图 7 为 MC 算法、SMC 算法和 OSMC 算法在 Engine 数据集局部位置三角片集合的构成情况。从对比图中可以看出:SMC 算法在形状不规则的区域的三角片密度低于 MC 算法,但在平坦区域的三角片密度和 MC 算法一样;OSMC 算法在平坦区域合并了大部分共面三角片,使角片密度

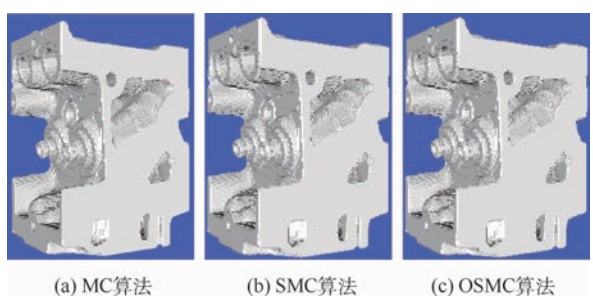


图 5　3 种算法对 Engine 数据产生的表面全视图

Fig. 5　Isosurface full view of three algorithms on Engine data set

北航学报 赠阅

降低,而在形状不规则区域三角片密度和 SMC 算法相同。综合起来看,OSMC 算法相比于 MC 算法,既在形状不规则区域具有较低三角片密度,也在平坦区域有更少三角片密度,不过细节部分略有损失。而与 SMC 算法比较而言,细节部分保存完好,更显优势。

表 2 列举了 9 组数据集分别在 MC 算法、SMC 算法、OSMC 算法下重建表面生成的三角片总数,以及 OSMC 算法对 MC、SMC 算法的削减率对比。可以看出,对于不同的数据集,算法的削减程度有所差别。SMC 算法在对总体形态不规则程度较高的数据集上相比 MC 算法有较好的削减率,例如 Skull、Lobster 数据集。OSMC 算法的削减程度则相对较低,但也依然超过了 SMC 算法。在具有较大比例规则表面的数据集上,OSMC 算法表现非常出色,在 SMC 算法的基础上依然有显著的削减,例如 Engine、Fan、Table 和 Teapot 数据集,并且可以很好地保留 SMC 算法的所有细节。

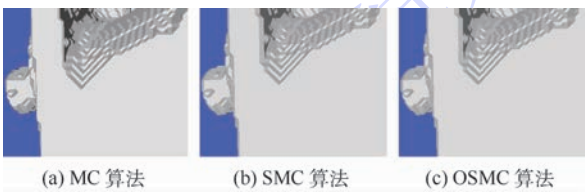


图 6 细节表面图

Fig. 6 Detailed surface

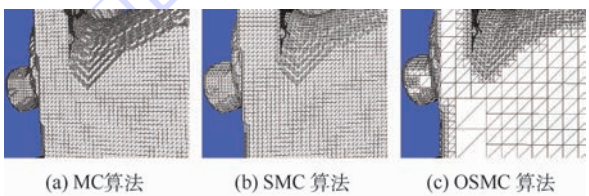


图 7 表面三角片构成情况

Fig. 7 Composition of surface triangle

表 2 部分数据集结果三角片数及削减率的对比

Table 2 Comparison between result of triangles numbers and reduction rate of part data set

数据集	三角片数			削减率/%	
	MC 算法	SMC 算法	OSMC 算法	OSMC 较 SMC 算法	OSMC 较 MC 算法
Engine	622 764	432 370	259 479	-40.0	-58.3
Fan	508 928	370 256	179 179	-51.6	-64.8
Table	475 548	439 692	90 073	-79.5	-81.1
Teapot	745 152	508 798	304 008	-40.2	-59.2
Bonsai	644 340	435 994	372 986	-14.5	-42.1
Skull	1 842 404	1 114 208	1 033 229	-7.3	-43.9
Backpack	3 985 836	2 876 044	2 066 527	-28.1	-48.2
Phantom	7 457 316	5 296 868	3 557 633	-32.8	-52.3
Lobster	1 245 556	721 914	632 738	-12.4	-49.2

3.2 人类肺部结节 CT 数据实验

将 OSMC 算法应用于真实的人类肺部结节 CT 数据,数据的最大分辨率为 55 像素 ×55 像素 ×120 像素。

如图 8 所示,人类肺部结节形状不规则,细节部分很多。细节部分对于诊断病情却至关重要,而 OSMC 算法在这方面表现得很好,对细节刻画很仔细,能够生成和 SMC 算法完全一样的网格。

表 3 列出了针对不同病人的完全不同的肺部结节 CT 数据,以及 MC、SMC 和 OSMC 算法生成的三角片的个数。可以看出,对于不同的肺部结节数据,尽管 OSMC 与 SMC 算法生成的网格形状是完全一样的,但是 OSMC 算法生成的三角片更少,最多能少 20% 左右。

图 9 展示了图 8 中肺部结节上不平滑的一个小触角处生成的网格形状。可以看出,OSMC 和 SMC 算法生成的形状完全一样,但是在 SMC 算法生成的形状一致的几个相邻三角片处,OSMC算

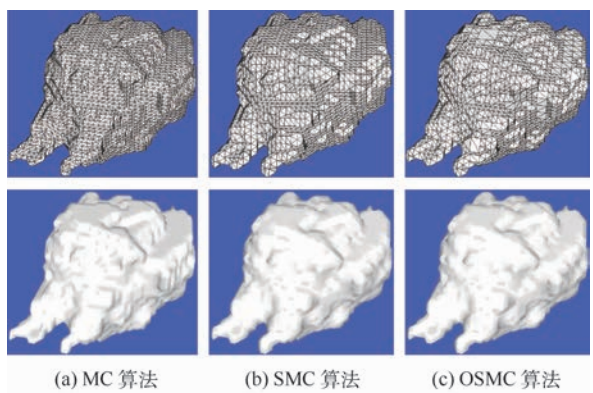


图 8 3 种算法人体肺部结节数据产生的网格和表面全视图

Fig. 8 Mesh and isosurface full view of three algorithms on human pulmonary nodule data set

表 3 人体肺部结节数据结果三角片数及削减率的对比

Table 3 Comparison between result of triangle numbers and reduction rate of human pulmonary nodules data set

数据集	三角片数			削减率/%	
	MC 算法	SMC 算法	OSMC 算法	OSMC 较 SMC 算法	OSMC 较 MC 算法
肺部结节 1	11 652	7 524	6 498	-13.64	-44.23
肺部结节 2	1 992	1 366	1 160	-15.08	-41.77
肺部结节 3	3 704	2 564	2 012	-21.53	-45.68
肺部结节 4	17 360	10 744	9 198	-14.39	-47.02
肺部结节 5	16 212	10 098	8 623	-14.61	-46.81
肺部结节 6	25 052	16 258	13 583	-16.45	-45.78
肺部结节 7	44 344	25 638	24 213	-5.56	-45.40
肺部结节 8	19 972	11 714	11 350	-3.11	-43.17
总计				-10.80	-45.37

北航学报 赠阅

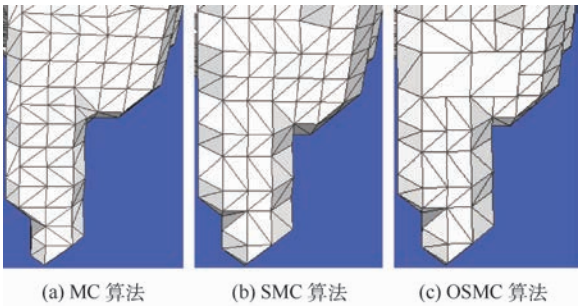


图 9　3 种算法人体肺部结节 CT 数据结果表面细节三角片构成情况

Fig. 9　Triangles mesh part view of three algorithms on CT human pulmonary nodule data set

法能够合并简化网格,使生成的三角片数量减少。

综上,OSMC 算法在面对等值面较粗糙细节较多的数据时,能够完整地保留等值面的所有细节部分,不会改变生成的等值面的形状,并且在相对平滑的部分完成合并简化,有效地减少了生成的三角片的数量。

3.3　地质数据实验

笔者还在真实的地质断面分离数据上测试了 OSMC 算法,其中最大数据的分辨率为 1 000 像素 ×1 000像素 ×500 像素。图 10 展示了 MC、SMC 和 OSMC 算法对同一个地质断面数据生成的等值面,由于分辨率很高,可以看出,OSMC 算法生成的结果与 MC 算法很相近,与 SMC 算法完全一样。

表 4 列出了在 7 个不同的地质断面分离数据上使用这 3 个算法生成的等值面的三角片个数以及相对于标准 MC、SMC 算法,OSMC 算法生成三角片的数量的减少比率。显然,OSMC 算法在此类数据上表现非常好,相对于 MC 和 SMC 算法,OSMC 算法生成的三角片数量都远少于前者,减少比率最高可以达到 80%,平均都在 50% 以上。

图 11 展示了 OSMC 算法之所以表现的如此优异的原因。因为地质断面数据在其外部边缘处非常平滑,并且数据量巨大,传统的 MC 和 SMC 算法对每一个小的区域都要生成一个三角片。但其实这些三角片都是共面并且完全一样的,而 OSMC 算法可以将这些很多共面的小的三角片合并成一个大的三角片。这样既没有改变等值面的形状,并且大大减少了处于边缘的平滑的部分三角片数量。MC 算法需要生成几万乃至几十万个三角片,但是 OSMC 算法只需要几百或者几千个就可以了。

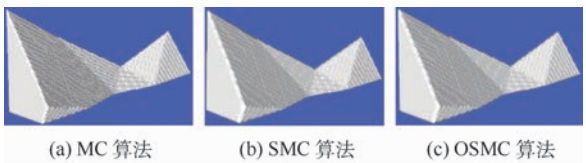


图 10　3 种算法对地质体数据产生的表面全视图

Fig. 10　Isosurface full view of three algorithms on geological data set

图 12 展示了在地质断面部分 MC、SMC 和 OSMC 算法生成的等值面的结果。可以看出,OSMC 算法生成的等值面与 SMC 算法完全一样,没有丢失任何原本的断面信息,并且在局部共面的地方进行了合并简化,使生成的三角片相应减少。

综上可以看出,本文 OSMC 算法在这种局部相对平滑的数据表现非常好。一方面,其生成的等值面的三角片数量较 MC 和 SMC 算法有很大比例的减少;另一方面,在需要精细刻画的局部,

表 4　地质体数据结果三角片数及削减率的对比

Table 4　Comparison between result of triangle numbers and reduction rate of geological data set

数据集	三角片数			削减率/%	
	MC 算法	SMC 算法	OSMC 算法	OSMC 较 SMC 算法	OSMC 较 MC 算法
地质体 1	4 216 160	3 068 372	1 569 472	-48.85	-62.77
地质体 2	3 570 748	3 044 366	1 123 834	-63.09	-68.53
地质体 3	3 998 268	3 018 910	1 068 565	-64.60	-73.27
地质体 4	7 187 412	5 691 550	1 902 811	-66.57	-73.53
地质体 5	2 374 620	1 788 710	764 733	-57.28	-67.80
地质体 6	3 348 048	2 714 290	639 024	-76.46	-80.91
地质体 7	33 408	262 02	14 492	-44.70	-56.62
总计				-63.40	-71.35

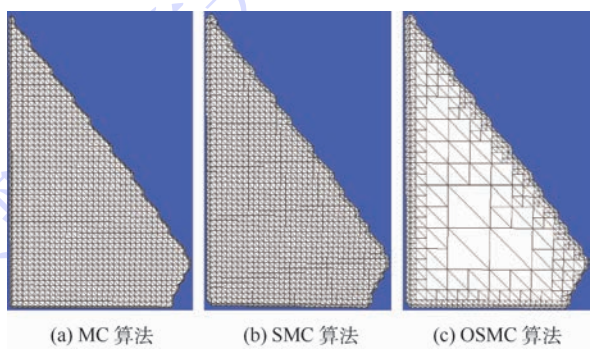


图 11　地质体数据集上平滑部分的三角片构成情况

Fig. 11　Triangle mesh of flat area of geological data set

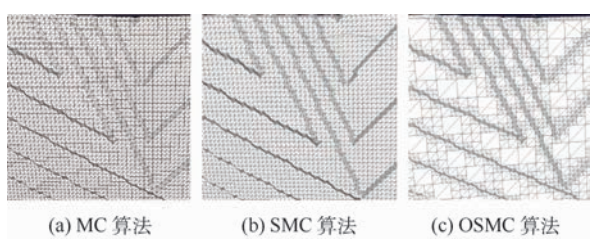


图 12　地质体数据集上非平滑部分的三角片构成情况

Fig. 12　Triangle mesh of uneven area of geological data set

北航学报 赠阅

OSMC 算法没有因为合并而改变原本等值面的形状,也就没有丢失任何信息。

3.4 分辨率增长下的实验

表 5 和表 6 分别列举了 Table 数据集和四面体数据集中 MC 算法、SMC 算法和 OSMC 算法在数据集分辨率增加为原来的 1、2、3、4 倍时,三角片数量的增长情况。可以看出,当数据集分辨率变为原来的 n 倍后,MC 算法和 SMC 算法的三角片数量变为接近原来的 n^2 倍。而 OSMC 算法则以低于 n^2 的速度增长。图 13 和图 14 中通过曲线来反映增长趋势,由于 MC 算法和 SMC 算法增长倍率相差很小,故曲线几乎合并为一条,而 OSMC 算法的倍率曲线则明显较低。其原因在于 OSMC 算法在八叉树收缩的步骤中合并了局部平坦区域中的体元,在分辨率提升的情况下,虽然这些区域的体元密度增加,但仍然会合并成相同的超体元。

从图 13 和图 14 中还可以看出,对于具有更大比例平坦表面的数据集,如四面体数据集,其增长曲线相比 Table 数据集更低。根据 OSMC 算法的特点不难推断,若数据集不规则表面比例增大,则 OSMC 算法增长曲线会向 MC 和 SMC 算法的增长曲线靠近。这也说明 OSMC 算法削减程度与数据集表面的形态特征有关。

表 5 Table 数据集 4 种分辨率下实验结果

Table 5 Results under four resolutions for Table data set

分辨率/(像素×像素×像素)	MC 算法		SMC 算法		OSMC 算法	
	三角片数	与 1 倍分辨率比值	三角片数	与 1 倍分辨率比值	三角片数	与 1 倍分辨率比值
197×70×101	111 652	1.00	103 056	1.00	7 028	1.00
394×140×203	475 548	4.26	439 692	4.27	10 196	3.49
591×210×304	1 045 120	9.36	955 408	9.27	17 464	5.84
788×280×406	1 920 724	17.20	1 786 762	17.34	25 334	10.33

表 6 四面体数据集 4 种分辨率下实验结果

Table 6 Experimental results under four resolutions for tetrahedral data set

分辨率/(像素×像素×像素)	MC 算法		SMC 算法		OSMC 算法	
	三角片数	与 1 倍分辨率比值	三角片数	与 1 倍分辨率比值	三角片数	与 1 倍分辨率比值
100×100×100	55 868	1.00	37 438	1.00	7 028	1.00
200×200×200	224 648	4.02	150 152	4.01	10 196	1.45
300×300×300	513 332	9.19	343 384	9.17	17 464	2.48
400×400×400	914 936	16.38	610 738	16.31	25 334	3.60

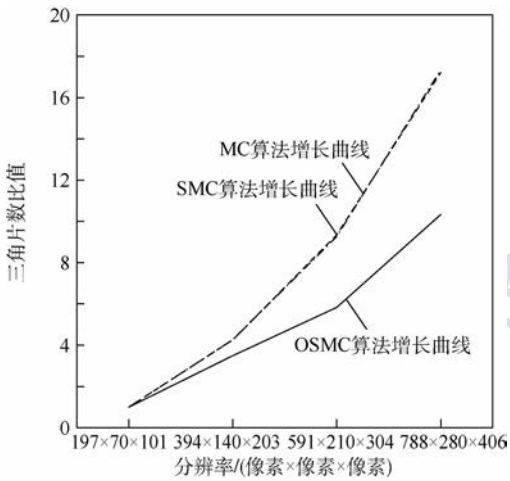


图 13 Table 数据增长曲线

Fig. 13 Table data growth curves

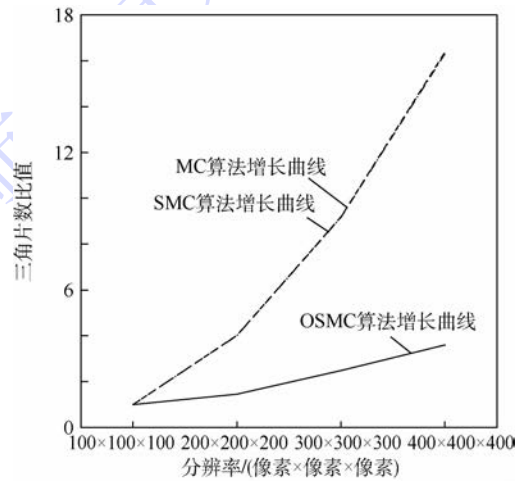


图 14 四面体数据增长曲线

Fig. 14 Tetrahedral data growth curves

4 结 论

本文在简化构型等值面的基础上提出了普适通用的 OSMC 算法,使用八叉树结构合并了等值面局部的部分共面三角片。

1) OSMC 算法相对于 SMC 算法能进一步减少三角片的数量,同时保证生成表面的形态与其完全一致。

2) OSMC 算法的削减效果在具有较多平坦区域的数据集上表现显著,能有效弥补 SMC 算法对这类数据集削减程度不足的问题。

3) OSMC 算法对实验数据的平均削减率为 55.1%,高于 SMC 算法的 29.7%,在面对高分辨率的地质数据时其最高削减率达到了 80%,平均也超过了 50%,效果尤为明显。

4) 在数据集分辨率逐渐增大的情况下,OSMC算法的结果集规模会以低于 MC 算法和 SMC 算法的速率增长。

OSMC 算法也存在如下 2 方面的问题:一是合并体元的条件设置局限在共面构型上,限制了削减率的进一步提升。二是八叉树的划分方式可能会把共面三角片分在非兄弟节点上,这样的情况在收缩时无法合并。所以下一步的研究方向是改善算法的合并条件以及八叉树的收缩过程,进一步增加算法的削减率。

参考文献 (References)

[1] LORENSEN W E, CLINE H E. Marching cubes: A high resolution 3D surface construction algorithm[C]// ACM Siggraph Computer Graphics. New York: ACM, 1987, 21(4): 163-169.

[2] HEIMBACH I, RHIEM F, BEULE F, et al. pyMolDyn: Identification, structure, and properties of cavities/vacancies in condensed matter and molecules[J]. Journal of Computational Chemistry, 2017, 38(6): 389-394.

[3] YIM P J, VASBINDER G B C, HO V B, et al. Isosurfaces as deformable models for magnetic resonance angiography[J]. IEEE Transactions on Medical Imaging, 2003, 22(7): 875-881.

[4] BRONSON J R, SASTRY S P, LEVINE J A, et al. Adaptive and unstructured mesh cleaving[J]. Procedia Engineering, 2014, 82: 266-278.

[5] FERLEY E, CANI M P, GASCUEL J D. Practical volumetric sculpting[J]. The Visual Computer, 2000, 16(8): 469-480.

[6] STEIN R, SHIH A M, BAKER M P, et al. Scientific visualization of water quality in the Chesapeake bay[C]// IEEE Visualization Conference. Piscataway, NJ: IEEE Press, 2000: 509-512.

[7] TREMBILSKI A. Two methods for cloud visualisation from weather simulation data[J]. The Visual Computer, 2001, 17(3): 179-184.

[8] KIM K, WITTENBRINK C M, PANG A. Data level comparison of surface classification and gradient filters[C]// Volume Graphics 2001. Berlin: Springer, 2001: 19-34.

[9] NIELSON G M, HAMANN B. The asymptotic decider: Resolving the ambiguity in marching cubes[C]// Proceedings of the 2nd Conference on Visualization' 91. Piscataway, NJ: IEEE Press, 1991: 83-91.

[10] RECK F, DACHSBACHER C, GROSSO R, et al. Realtime isosurface extraction with graphics hardware[C]// Eurographics. Geneva: Eurographics, 2004: 1-4.

[11] JOHANSSON G, CARR H. Accelerating marching cubes with graphics hardware[C]// Proceedings of the 2006 Conference of the Center for Advanced Studies on Collaborative Research. Armonk: IBM Corporation, 2006, 6: 39.

[12] NEWMAN T S, YI H. A survey of the marching cubes algorithm[J]. Computers & Graphics, 2006, 30(5): 854-879.

[13] SCHROEDER W J, ZARGE J A, LORENSEN W E. Decimation of triangle meshes[C]// ACM Siggraph Computer Graphics. New York: ACM, 1992, 26(2): 65-70.

[14] GARLAND M, HECKBERT P S. Surface simplification using quadric error metrics[C]// Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques. New York: ACM Press/Addison-Wesley Publishing Co., 1997: 209-216.

[15] MONTANI C, SCATENI R, SCOPIGNO R. Discretized marching cubes[C]// Proceedings of the Conference on Visualization' 94. Piscataway, NJ: IEEE Press, 1994: 281-287.

[16] MONTANI C, SCATENI R, SCOPIGNO R. Decreasing isosurface complexity via discrete fitting[J]. Computer Aided Geometric Design, 2000, 17(3): 207-232.

[17] SHU R, ZHOU C, KANKANHALLI M S. Adaptive marching cubes[J]. The Visual Computer, 1995, 11(4): 202-217.

[18] SHEKHAR R, FAYYAD E, YAGEL R, et al. Octree-based decimation of marching cubes surfaces[C]// 7th Annual IEEE Conference on Visualization. Piscataway, NJ: IEEE Press, 1996: 335-342.

[19] OHTAKE Y, BELYAEV A, ALEXA M, et al. Multi-level partition of unity implicits[C]// ACM Siggraph 2005 Courses. New York: ACM, 2005: 173.

[20] KAZHDAN M, HOPPE H. Screened poisson surface reconstruction[J]. ACM Transactions on Graphics (TOG), 2013, 32(3): 29.

[21] WILHELMS J, VAN GELDER A. Octrees for faster isosurface generation[J]. ACM Transactions on Graphics (TOG), 1992, 11(3): 201-227.

[22] WESTERMANN R, KOBBELT L, ERTL T. Real-time exploration of regular volume data by adaptive reconstruction of isosurfaces[J]. The Visual Computer, 1999, 15(2): 100-111.

[23] CUI S H, LIU J. Simplified patterns for extracting the isosurfaces of solid objects[J]. Image and Vision Computing, 2008, 26(2): 174-186.

[24] VIGNOLES G L, DONIAS M, MULAT C, et al. Simplified marching cubes: An efficient discretization scheme for simulations of deposition/ablation in complex media[J]. Computational Materials Science, 2011, 50(3): 893-902.

作者简介:

徐雷　男,硕士研究生。主要研究方向:模式识别、机器学习。

王华锋　男,博士,副教授,硕士生导师。主要研究方向:图像处理、模式识别、机器学习。

北航学报 赠阅

附录 A:生成 SMC 简化构型查找表

为快速处理体元,需要为 SMC 简化构型构造有 256 个入口的查找表,SMC 查找表的构建基于 MC 算法的查找表。首先需要对体元的各个顶点的边按图 A1 的方式编号。

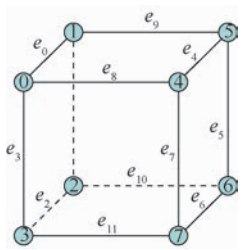


图 A1 体元顶点与边

Fig. A1 Number of vertices and edges of cell

MC 算法查找表共有 256 行,每一行是由形式为{ $<e_0,e_1,e_2>$, $<e_3,e_4,e_5>$,…}的三元组构成,对于任意一种体元状态,可从中获取三角形顶点所在边的索引和三角形的组合方式 $<e_i,e_j,e_k>$。而对于 SMC 算法,由于三角片的顶点不在位于体元边上,而是体元顶点上,那么单个三角片就不再使用 3 条边索引表示,而是使用 3 个顶点索引表示。即: $<V_i,V_j,V_k>$,从标准 MC 算法查找表构建 SMC 算法查找表的步骤如下:

算法 SMC 构型查找表生成算法

输入:MC 构型查找表 T 输出。

输出:SMC 构型查找表 ST。

```
Begin
    For cube state i from 0 to 255
        Get triangles from T[i]
        For each triangle
            Get the three edge indices
            Get two vertex indices for each edge
            Caculate three intersected points for the three edge
            If(the three vertices are not equal) then
                Find the cube vertex where the three intersected points locate
                Record the cube vertex index to ST[i]
End
```

附录 B:生成三角片平面方程查找表

在八叉树收缩时需要根据三角片方程和法向量来合并体元,这些方程和法线的参数在算法执行时动态计算会造成较大的开销,所以也需要像 MC 算法一样构造查找表来提升效率。构造 Equation表基于简化构型,需要利用附录 A 构造的 SMC 查找表。

Equation 表同样有 256 个入口,每个入口对应三角片方程参数(A,B,C,D)。具体构造方法是遍历 256 种体元构型,判断其中三角片是否为共面构型,若为共面构型,判断其方程类型为表 1 中的哪一种,根据不同方程类型和体元顶点坐标反求方程参数。伪代码如下:

算法 三角片方程查找表生成算法

输入:SMC 构型查找表 T 输出。

输出: 三角片方程查找表 Eq。

```
Begin
    For cube state i from 0 to 255
      if( i is coplanar configration )
          Get vertices V on the plane using T[i]
          Get the triangle t using T[i]
          If(t is equilateral triangle)
            Solve (A,B,C,D) using V
          If(t is isosceles right-angle triangle)
              Find the plane which t is pararell with
          Solve (A,B,C,D) using V
          If(t is non-isosceles right-angle triangle)
          Find the axis t is pararell with
          Solve (A,B,C,D) using V
          Set (A,B,C,D) to Eq[i]
      else
          Set Eq[i] to invalid value
End
```

Octree based decimation algorithm for triangle isosurface using simplified patterns

XU Lei, WANG Huafeng*, PAN Haixia, LIN Guangyan, CHEN Yuexi

(School of Software, Beijing University of Aeronautics and Astronautics, Beijing 100083, China)

Abstract: It is universally acknowledged that SMC based on simplified patterns extracts less triangles than the standard MC. Because only in-cube decimation was exploited, SMC is not able to take full advantage of local features of isosurfaces. Based on this observation, a new method named OSMC is presented in this paper. Based on characteristics of simplified configuration, OSMC first use octree structure to organize cells as nodes, then merge the nodes from bottom to top, and finally achieve local area triangles merging. The experimental results illustrate that the proposed method does further decimation than SMC, especially for datasets with large flat areas. The proposed method achieves an average reduction rate up to 55.1%, while the average reduction rate for SMC is 29.7%. The reduction rate reaches 80% at the highest and it is above 50% in average when OSMC is used on high-resolution geological dataset. Moreover, the new method is more adaptive to the increment of the dataset resolution.

Keywords: isosurface; MC algorithm; SMC algorithm; octree; super cell

北京航空航天大学学报

Received: 2017-05-05; **Accepted**: 2017-09-01; **Published online**: 2017-10-13 13:58

URL: kns.cnki.net/kcms/detail/11.2625.V.20171013.1358.002.html

* **Corresponding author**. E-mail: wanghuafeng@buaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2017.0293

Ni 晶界上金属和非金属元素的相互作用

方淑娟，石松鑫，祝令刚*，周健

（北京航空航天大学 材料科学与工程学院，北京 100083）

摘 要：采用基于密度泛函理论的第一性原理计算方法，研究了常用合金化元素与非金属元素在晶界(GB)的偏析及其相互作用。选取2种类型的晶界结构进行研究：密排的晶界和较疏松的晶界。溶解能计算表明，金属和非金属元素在Σ5晶界上均有明显的偏析行为，而在Σ3晶界上的偏析现象不明显。通过计算常用过渡族合金化元素与非金属杂质C、H、O、N、B在各自最稳定位置的相互作用能，发现Ru、Re、W和Ta对O有强烈的排斥作用，因此对晶界抗氧化有益处；Ta排斥H，有抗晶界氢脆的效果。通过对金属元素与非金属元素在晶界上相互作用的系统研究，为Ni合金的晶界工程提供有价值的参考。

关 键 词：晶界(GB)偏析；镍基高温合金；氢脆；点缺陷；第一性原理计算

中图分类号：V252.2；O483；O77⁺1

文献标识码：A **文章编号**：1001-5965(2018)04-0862-06

自20世纪40年代以来，高温合金经过几十年的发展与变革，已成为工业建设高速高效、安全可靠的重要保障。航空工业领域主要使用的是微观结构下界面稳定的单晶高温合金。民用工业的生产发展则是由多晶高温合金支撑起来的，通常具有较低的成本，便于大面积推广使用。在高温强度、耐损伤性、组织稳定性和使用寿命等方面，民用工业对材料性能的要求有时甚至更为苛刻。与单晶高温合金相比，多晶高温合金的研究价值和应用前景同样不可忽视，但是需要进一步提高其性能[1-2]。镍基高温合金在先进航空发动机所用材料中可占一半以上，结构特性优越，具有更大的改性空间。其中，镍基多晶高温合金广泛应用于核反应堆结构材料及涡轮盘、叶片等涡轮燃气机的热端部件中，在高温、高压、高机械应力、腐蚀、辐射等严苛的环境条件下长期稳定工作，质量控制成为重中之重。

晶界(GB)本身就相当于一个非常复杂的缺陷，高温变形条件下，结构的失效往往从晶界处发生。由于晶界结构与晶粒内不同，存在多种缺陷及应力场，一些杂质元素倾向于在晶界处偏聚和扩散，形成置换原子或者溶解间隙原子。这些点缺陷参与扩散过程，能够充当质量传递的介质，极具研究意义。例如，晶界氧化[3-4]和氢元素引起的脆化[5]将分别导致晶间应力腐蚀开裂和合金的脆性断裂。中国开发出的低偏析高温合金系列，降低了S、P和Si等杂质的含量，使承温能力比原型合金提高20～25℃，采用Mg偏聚晶界的微合金化技术，显著提高了合金的强度和塑性[1]。

已有大量的理论实验研究关于合金化元素在Ni中的偏析[4,6-8]和元素对Ni的作用[9-11]，表明化学成分是影响晶界性质并决定材料性能的最重要的因素，元素在晶界上的偏析是改变化学成分的普遍途径。精确控制合金成分，有效去除杂质，或者将含量控制在较低水平，是提升高温合金性能的关键环节，对于旨在优化晶界性能的晶界工程

收稿日期：2017-05-09；**录用日期**：2017-06-12；**网络出版时间**：2017-06-29 11:59
网络出版地址：kns.cnki.net/kcms/detail/11.2625.V.20170629.1159.002.html
基金项目：国家自然科学基金（51401009，51571008）
*通信作者. E-mail：lgzhu7@buaa.edu.cn

引用格式：方淑娟，石松鑫，祝令刚，等. Ni晶界上金属和非金属元素的相互作用[J]. 北京航空航天大学学报，2018，44(4)：862-867. FANG S J，SHI S X，ZHU L G，et al. Interaction between metallic and nonmetallic elements on grain boundary of nickel[J]. Journal of Beijing University of Aeronautics and Astronautics，2018，44(4)：862-867(lin Chinese).

有着重要意义。俄歇谱仪分析显示,微量元素偏聚在晶界上会脆化金属[12]。Bagot 等[13]通过原子力探针扫描(atom probe tomography)技术研究 Ni 基合金 RR1000,在晶界上发现了 Cr、Mo 和 B 等元素,并且检测到了 C 元素的富集,尽管不如 B 元素的含量显著。理论研究方面,Razumovskiy 等[14]使用 Rice-Thomson-Wang 模型研究了 4d 和 5d 过渡族金属元素对典型的大角度 Ni 晶界结合力的影响,然后在粉末冶金的 Ni 合金中加入理论研究筛选出的元素 Zr、Hf、Nb、Ta 和 B,计算结果与实验符合得很好,通过调整化学成分提升了合金的力学性能。Všianská 和 Šob[15]发现偏析的 sp 轨道杂化元素极大降低了 Ni 原子的磁性,Ga、Te、Sb、In 和 Sn 等元素使 Ni 脆化。对于轻质元素 H,Stefano 等[16]发现它容易被晶界吸引,以晶界为通道快速扩散,在晶界上的偏析则不显著。

目前,理论研究主要集中在单个元素的偏析行为及对合金的强化/脆化作用,关注共偏析影响和多种元素相互作用的研究还非常少。对于偏析元素相互作用的研究,可以为合金设计者提供有价值的信息,保护合金不被有害元素侵袭,提高合金纯洁度,保持有益元素持续发挥作用,改善合金性能,延长使用寿命。另外,利用第一性原理计算方法,实现了研究单一杂质、避免其他成分干扰的构想,弥补了实验方法的短板,从而揭示原子尺度上的作用机制,为多种元素共同作用的研究打下基础[17]。

本文着眼于元素之间的相互作用,采用基于密度泛函理论的第一性原理计算方法,选取 2 种典型的 Ni 晶界结构,研究了 14 种常用过渡族合金化金属元素和 C、H、O、N 和 B 5 种非金属元素在 Ni 晶界上的偏析规律。

1　计算方法

本文采用维也纳大学开发的软件包 VASP(Vienna Ab initio Simulation Package)[18]完成计算。利用 GGA-PW91 广义梯度近似泛函[19]来描述交换关联能,使用 PAW 势[20]处理电子与核的相互作用,Ni 最外层 16 个电子均作为价电子处理。平面波截断能为 400 eV。在倒易空间中使用 Monkhorst-Pack 方法确定 k 点。Σ3 晶界超胞的 k 点网格为 3×3×1,Σ5 晶界超胞的 k 点网格为 2×2×1。所有 Ni 相关的计算都设置了自旋极化。能量收敛判据为 2 个电子步之间总能变化小于 10^{-5} eV,原子间作用力小于 0.03 eV/(10^{-10} nm)。

基于重合位置点阵(CSL),本文选取 2 个典型[15,21-22]的 Ni 晶界结构,密排的晶界和含有较多孔隙的晶界(以下分别简写为 Σ3 和 Σ5 晶界)进行研究。采用 Java 应用 Grain Boundary Studio[23]构建晶界初始原子结构,依据计算出的体材料的晶格常数,生成 3D 周期性模型,用于后续第一性原理计算。

结构优化中所有原子都得到了松弛。为了忽略周期性结构中 2 个相邻晶界之间的相互作用引起的计算误差,同时控制因体系结构增大而附加的计算量,需要确定合适的超胞大小。最终选取的 Σ3 晶界包含 12 层原子,晶格常数为 $a = 7.42 \times 10^{-10}$ nm, $b = 7.42 \times 10^{-10}$ nm, $c = 24.23 \times 10^{-10}$ nm,共 108 个原子。Σ5 晶界包含 20 层原子,晶格常数为 $a = 7.00 \times 10^{-10}$ nm, $b = 7.82 \times 10^{-10}$ nm, $c = 15.64 \times 10^{-10}$ nm,共 80 个原子。晶界的能量跟文献[16]作了对比,证明超胞大小是合适的。

为了描述晶界结构的稳定性,定义晶界能 γ_{GB} 的计算公式为

$$\gamma_{GB} = \frac{E_{GB} - E_{bulk}}{2A} \tag{1}$$

式中:E_{GB} 为晶界超胞的总能;E_{bulk} 为相同原子数块体 Ni 的总能;A 为充分弛豫后的界面面积。取因数 2 表示一个超胞中包含了 2 个晶界。γ_{GB} 越小,晶粒的结合强度越高。

定义溶解能 E^{sol} 来描述元素溶于合金的能力,其计算公式为

$$E^{sol} = E^{tot}_{Ni+M/X} - E^{tot}_{Ni} - E_{M/X} \tag{2}$$

式中:$E^{tot}_{Ni+M/X}$ 为含有金属添加元素 M 或非金属添加元素 X 的晶界超胞总能;E^{tot}_{Ni} 为只含有 Ni 的超胞总能;$E_{M/X}$ 为单个 M/X 原子在块体材料中的能量。

合金化原子 M 和非金属原子 X 之间相互作用能 ΔE_{MX} 的计算公式[24]为

$$\Delta E_{MX} = (E_{Ni+M-X} + E_{Ni}) - (E_{Ni+M} + E_{Ni+X}) \tag{3}$$

式中:E_{Ni+M-X} 为包含 M-X 原子对的 Ni 超胞总能;E_{Ni} 为纯 Ni 超胞的总能;E_{Ni+M} 和 E_{Ni+X} 分别为包含一个金属 M 原子和一个非金属 X 原子的超胞总能。当 ΔE_{MX} 为负值,M 和 X 之间互相吸引;当 ΔE_{MX} 为正值,M 和 X 之间互相排斥。也就是说,金属 M 原子对轻质元素 X 的偏析有促进或者抑制作用。

2　结果与讨论

2.1　晶界结构

构建的 2 个 Ni 晶界结构沿[001]方向的视图

北航学报 赠阅

如图 1 所示。虚线表示晶界平面,数字由小到大依次标出离界面(0 层)由近到远,一直到块体区域的原子层编号。

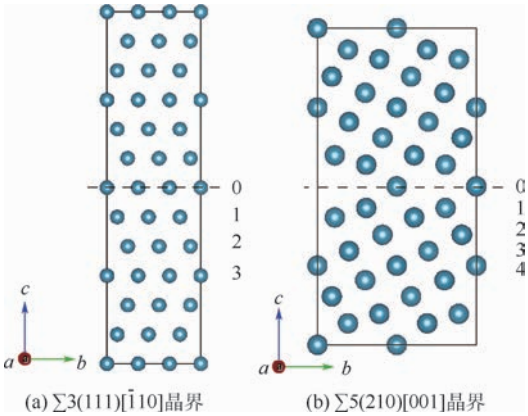


图 1 2 种晶界弛豫后沿[001]方向结构

Fig. 1 Two relaxed grain boundary structures viewed along [001] direction

从图 1 可以看出,Σ3 晶界平面上原子排布较密,Σ5 晶界上原子排布疏松,含有较大孔洞。由晶界面上原子数 N 与晶界面积 A 的比值,$\frac{N_{\Sigma3}}{A_{\Sigma3}} = 0.252 \times 10^{20}\ \mathrm{nm}^{-2}$,$\frac{N_{\Sigma5}}{A_{\Sigma5}} = 0.073 \times 10^{20}\ \mathrm{nm}^{-2}$,也可看出 Σ3 晶界上原子密度更大。计算得到的 Σ3 和 Σ5 晶界的能量分别是 0.16 J/m^2 和 1.59 J/m^2,与文献[16]中 0.18 J/m^2 和 1.29 J/m^2 的计算结果近似。Σ3 晶界能量低,显示出与体材料结构的相似性。

2.2 合金化金属元素的偏析行为

晶界上的易偏析元素会对合金的结构与性能产生不同的影响[5,8-9]。有益微量元素是冶炼过程中有意加入的,如稀土元素以及 Mg、Ca、Ba、Zr 和 Hf 等,它们改善原子间的键合状态,加强界面结合力,提升合金的高温强度,同时抑制含有害元素的低熔点化合物或共晶产物在晶界生成,阻碍有害杂质向晶界偏析。有害元素来自炼钢的原材料和冶炼过程中的环境及设备污染,可降低晶界结合力[2]。

Ni 合金中常用的合金化元素有 Cr、Co、Mo、Al、Fe、Ti、Mn、Cu、Ta、Zr、Hf、W、Re 和 Ru 等,均包含在本文研究范围内。其中,Cr、W、Mo、Ta、Re 和 Ru 都是有效的固溶强化元素[2]。由于合金化元素原子半径较大,不易占据间隙位置,通常替代 Ni 原子。

各类合金化元素在 Σ3 和 Σ5 晶界不同位置处的溶解能计算结果如图2(a)和图2(b)所示。

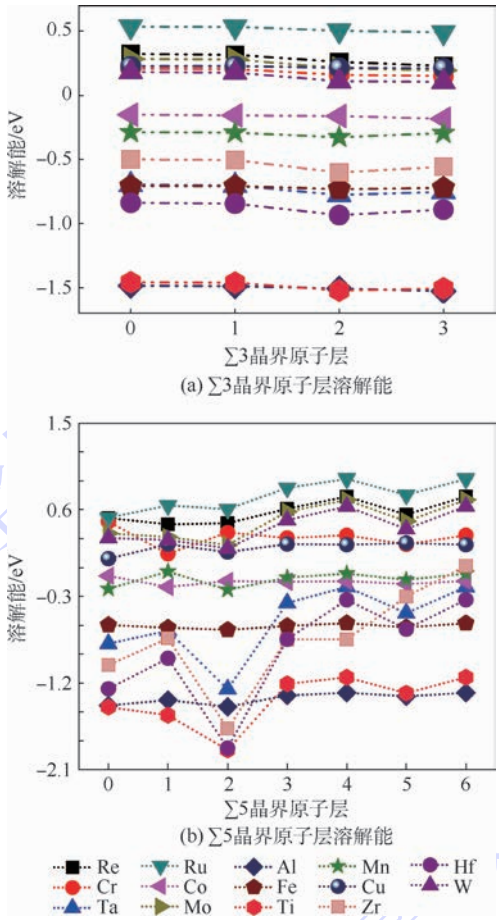


图 2 溶解能随合金化元素占晶界不同原子层的变化

Fig. 2 Variation of solution energy with alloying elements doped at different layers

总的来说,在距离晶界较远的地方,即近似块体的区域,溶解能趋于平衡。2 种晶界的溶解能曲线各不相同。Σ3 晶界的曲线趋于水平,表明合金化元素没有明显的偏析倾向,在体材料区域较稳定。而 Σ5 晶界的能量曲线波动比较明显,大多数元素倾向于向晶界偏析,并呈现出 2 种不同的趋势,原子尺寸是主要的影响因素[1,8]:比 Ni 尺寸更大的原子,如 Ta、Al、Zr、Hf 和 Ti,在晶界的次近邻原子层上有最稳定的取代位置;尺寸略小的原子,如 Cr、Co 和 Re 等,在晶界平面的最近邻原子层有最稳定的位置。

2.3 非金属元素的偏析行为

高温合金对杂质含量的控制有严格要求。气体元素 H、O 和 N 严重损害高温合金的力学性能。C 元素能强化晶界,应用于几乎所有的高温合金中。B 元素能够增大界面原子间结合力,减少晶界缺陷,析出少量颗粒相[1-2]。

B 元素的溶解度大于 C 元素,小于替代式合金化元素,往往认为它在晶界空位处偏析,属于非平衡偏析性质。本文通过计算验证了这一点。B 元素在间隙位置的溶解能是 -0.89 eV,低于其在

北航学报 赠阅

替代位置的能量 5.16 eV。

对于 H、O 和 N 元素,选用单个分子计算能量,对于 B 和 C 元素,则在晶体结构数据库 ICSD 中寻找常温下稳定的晶体构型计算能量。对于基态是顺磁的 H_2 和 O_2 分子,在计算中开启自旋极化。非金属元素单个原子能量计算的截断能均取 400 eV,与晶界计算的截断能保持一致。各项计算中 k 点的数量均经过校正,确保能代表性地描述各类结构。得到的 H、C、O、N 和 B 单个原子的能量分别为 -1.12、-9.10、-4.90、-8.33 和 -6.69 eV。

选取间隙原子位置时,首先在晶界平面的孔洞上取若干个点,其中 Σ5 晶界选取的位置如图 3 所示。至于 Σ3 晶界,由于周期性和对称性,平面上的孔洞位置取一个即可,不于图中再作表述。

对应于图 3,非金属杂质 Σ5 晶界孔洞处的溶解能 E^{sol}(下标为各元素)计算结果如表 1 所示。可以看出,在对称位置 5 和 6,各个非金属原子都达到了最低能量,即为孔洞处的最稳定位置。

根据得到的非金属杂质位于晶界附近一直到块体区域的四面体和八面体间隙的能量,研究

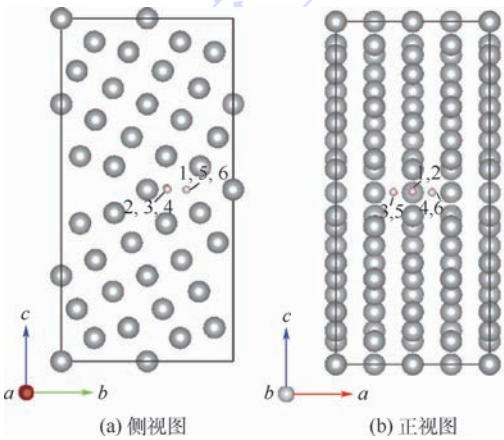


图 3　Σ5 晶界平面上间隙位置

Fig. 3　Doping sites at Σ5 GB interface plane

表 1　单个非金属原子占据 Σ5 晶界孔洞处不同位置的溶解能

Table 1　Solution energy of Σ5 grain boundary containing one nonmetallic atom at different sites on hollow of GB plane　eV

Σ5 晶界间隙位置	E_H^{sol}	E_C^{sol}	E_O^{sol}	E_N^{sol}	E_B^{sol}
1	-1.693	0.747	-1.165	-1.146	-0.899
2	-1.691	0.583	-0.880	-1.683	-0.898
3	-1.736	0.745	-1.539	-1.459	-0.899
4	-1.736	0.744	-0.968	-1.459	-0.899
5	-1.736	0.744	-1.541	-1.461	-0.900
6	-1.736	0.743	-1.541	-1.462	-0.900

元素在结构中的整体偏析倾向。表 2 列出了间隙原子在晶界上和远离晶界的块体区域溶解能之差,o 代表八面体间隙,t 代表四面体间隙。差值为负则表示元素倾向于偏析在晶界上。与金属元素在晶界中的偏析趋势相似,所有非金属元素都倾向于偏聚在 Σ5 晶界上,而 Σ3 晶界吸引到的元素寥寥无几。在 Σ3 晶界中能发现的规律是,元素更倾向于块体区域中八面体间隙的位置。回到偏析现象明显的 Σ5 晶界,C 和 B 表现出了相似的偏析行为,它们都驻留在靠近界面的位置。这一现象与实验中的发现相吻合[13,25]。

表 2　间隙原子在晶界和块体处的溶解能差值

Table 2　Solution energy differences of interstitial atoms at grain boundary and bulk　eV

晶界	间隙	H	C	O	N	B
Σ3	o	0.280	0.038	0.562	1.235	-0.012
	t	0.008	0.067	0.078	0.073	0.070
Σ5	o	-0.281	-1.519	-2.111	-0.650	-1.857
	t	-0.196	-0.474	-0.314	0.154	-0.394

2.4　元素在 Ni 晶界上的共偏析作用

加入多种微量元素,意在通过元素之间的协同作用获得更好的强韧化效果[11]。由于 Σ5 晶界上元素的偏析行为比 Σ3 晶界上明显得多,本节仅对 Σ5 晶界进行元素相互作用的研究。选取有明显偏析倾向的金属元素 Ru、Re、W、Zr、Hf 和 Ta,以及 2.3 节的 5 种非金属元素,将其分别置于各自最稳定的位置,如图 4 所示,数字 1 ~ 5 代表晶界附近金属原子可能占据的原子层,由图 2 可知 3 号为稳定位置,非金属原子在 6 号位置,计算相互作用能。

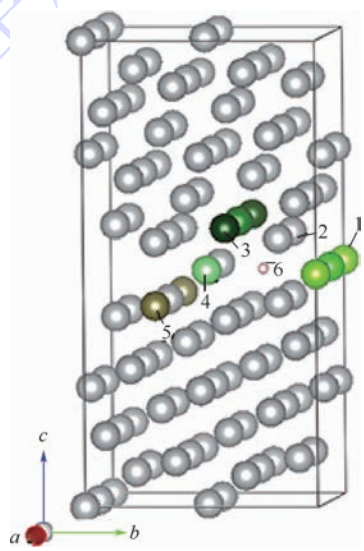


图 4　金属与非金属元素在晶界上各自最稳定的位置

Fig. 4　The most stable positions of metallic and nonmetallic elements on grain boundaries

北航学报 赠阅

相互作用能的计算结果如图 5 所示。由图中可知,金属元素对 C 和 H 的作用有相似之处:略微吸引 C,较强地排斥 H。这意味着所有元素都可以促进 C 在晶界的偏析,而抑制 H 的偏析。对于 O 和 N 来说,Ru、Re、W、Ta 与 O 的相互作用能绝对值大,且为正值,体现出强烈的排斥作用,推知 Ru、Re、W 和 Ta 可以提升晶界的抗氧化性。然而它们对 N 的排斥作用比较弱。对于常用的晶界强化元素 B,Ru、Re 和 W 能够产生吸引作用,而 Zr、Hf 和 Ta 对其是排斥作用。Ta 对 H 的排斥作用强,可在晶界抗氢脆方面发挥有益作用。

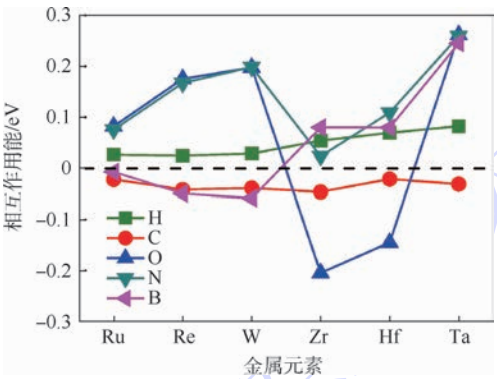


图 5 强偏析金属元素与非金属元素在 Σ5 晶界上的相互作用能

Fig. 5 Interaction energies between metallic elements with significant segregation tendency and nonmetallic elements on Σ5 grain boundary

3 结 论

1)通过第一性原理计算研究了金属元素(14 种过渡族合金化元素)和非金属元素(C、H、O、N 和 B)在晶界上的的偏析行为,以及它们的相互作用。在本文所研究的元素当中,密排结构、晶界能低的 Σ3 晶界没有体现出明显偏析现象。对于含有较多孔隙、晶界能较高的 Σ5 晶界,可以总结出以下偏析规律:Cu、Mn 和 Ru 倾向于占据在晶界平面,而 Ta、Al、Zr、Hf 和 Ti 在晶界的第二近邻原子层有更稳定的位置,Cr、Co 和 Re 的最稳定位置在晶界的第一近邻原子层。非金属元素倾向于占据在晶界上的孔洞位置。

2)合金化元素与非金属杂质的相互作用研究结果表明,Zr 和 Hf 显示出对 O 强烈的吸引作用,Ta 对 O、N 和 B 有强烈的排斥作用。以计算结果为基础,可以推想,Ru、Re、W 和 Ta 能够抑制晶界的氧化作用,Ta 可以有效地阻止氢脆效应,是良好的抑制杂质偏析的元素。以上结论可以为 Ni 合金的晶界工程提供参考。

参考文献(References)

[1] 黄乾尧,李汉康. 高温合金[M]. 北京:冶金工业出版社,2000:3.
HUANG Q Y, LI H K. Superalloys[M]. Beijing: Metallurgical Industry Press, 2000:3(in Chinese).

[2] 郭建亭. 高温合金材料学[M]. 北京:科学出版社,2008:4.
GUO J T. Materials science and engineering for superalloys[M]. Beijing: Science Press, 2008:4(in Chinese).

[3] VISKARI L H, RNQVIST M, MOORE K L, et al. Intergranularcrack tip oxidation in a Ni-base superalloy[J]. Acta Materialia, 2013, 61(10):3630-3639.

[4] ALEXANDROV V, SUSHKO M L, SCHREIBER D K, et al. Ab initio modeling of bulk and intragranular diffusion in Ni alloys[J]. Journal of Physical Chemistry Letters, 2015, 6(9):1618-1623.

[5] BECHTLE S, KUMAR M, SOMERDAY B P, et al. Grain-boundary engineering markedly reduces susceptibility to intergra-nular hydrogen embrittlement in metallic materials[J]. Acta Materialia, 2009, 57(14):4148-4157.

[6] ANGELO J E, MOODY N R, BASKES M I. Trapping of hydrogen to lattice defects in nickel[J]. Modelling & Simulation in Materials Science & Engineering, 1995, 3(3):289-307.

[7] DINGREVILLE R, BERBENNI S. On the interaction of solutes with grain boundaries[J]. Acta Materialia, 2016, 104:237-49.

[8] 刘文冠. 基于第一性原理的镍基合金晶界脆化机理的理论研究[D]:北京:中国科学院大学,2014:37.
LIU W G. First-principles study of intergranular embrittlement in Ni-based alloy[D]: Beijing: University of Chinese Academy of Sciences, 2014:37(in Chinese).

[9] SINGH N, TALAPATRA A, JUNKAEW A, et al. Effect of ternary additions to structural properties of NiTi alloys[J]. Computational Materials Science, 2016, 112:347-355.

[10] AI C, LI S, LIANG Y, et al. Influence of Mo and Ta additions on solidification behavior of Ni_3Al single crystal alloys[J]. Progress in Natural Science: Materials International, 2015, 25(4):353-360.

[11] ZHAO W, SUN Z, GONG S. Synergistic effect of co-alloying elements on site preferences and elastic properties of Ni_3Al: A first-principles study[J]. Intermetallics, 2015, 65:75-80.

[12] 周自强. 晶界研究的现状和发展[J]. 北京航空航天大学学报,1989,3(3):117-124.
ZHOU Z Q. Recent progress and development on grain boundary research[J]. Journal of Beijing University of Aeronautics and Astronautics, 1989, 3(3):117-124(in Chinese).

[13] BAGOT P A J, SILK O B W, DOUGLAS J O, et al. An atom probe tomography study of site preference and partitioning in a nickel-based superalloy[J]. Acta Materialia, 2017, 125:156-165.

[14] RAZUMOVSKIY V I, LOZOVOI A Y, RAZUMOVSKII I M. First-principles-aided design of a new Ni-base superalloy: Influence of transition metal alloying elements on grain boundary and bulk cohesion[J]. Acta Materialia, 2015, 82:369-377.

[15] VŠIANSKÁ M, ŠOB M. The effect of segregated sp-impurities on

grain-boundary and surface structure, magnetism and embrittlement in nickel[J]. Progress in Materials Science, 2011, 56(6): 817-840.

[16] STEFANO D D, MROVEC M, ELSÄSSER C. First-principles investigation of hydrogen trapping and diffusion at grain boundaries in nickel[J]. Acta Materialia, 2015, 98: 306-312.

[17] ZHANG S, KONTSEVOI O Y, FREEMAN A J, et al. First priciples investigation of zinc-induced embrittlement in an aluminum grain boundary[J]. Acta Materialia, 2011, 59(15): 6155-6167.

[18] KRESSE G, FURTHMÜLLER J. Efficiency of Ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set[J]. Computational Materials Science, 1996, 6(1): 15-50.

[19] ERNZERHOF M. Generalized gradient approximation made simple[J]. Physical Review Letters, 1996, 77(18): 3865.

[20] PERDEW J P, CHEVARY J, VOSKO S, et al. Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation[J]. Physical Review B, 1992, 46(11): 6671.

[21] LIU W, HAN H, REN C, et al. Effects of rare-earth on the cohesion of Ni Σ5 (012) grain boundary from first-principles calculations[J]. Computational Materials Science, 2015, 96: 374-378.

[22] XIA S, ZHOU B X, CHEN W J, et al. Effects of strain and annealing processes on the distribution of Σ3 boundaries in a Ni-based superalloy[J]. Scripta Materialia, 2006, 54(12): 2019-2022.

[23] OGAWA H. GBstudio: A builder software on periodic models of CSL boundaries for molecular simulation[J]. Materials Transactions, 2006, 47(11): 2706-2710.

[24] 祝令刚. Ti 和 Nb 合金中氧化问题的理论研究[D]. 沈阳：中国科学院金属研究所，2013：84.
ZHU L G. Theoretical study on the oxidation of Ti and Nb-based alloys[D]. Shenyang: Insititute of Metal Research, Chinese Academy of Sciences, 2013: 84(in Chinese).

[25] ALAM T, FELFER P J, CHATURVEDI M, et al. Segregation of B, P, and C in the Ni-based superalloy, inconel 718[J]. Metallurgical and Materials Transactions A, 2012, 43(7): 2183-2191.

作者简介：

方淑娟　女，硕士研究生。主要研究方向：镍基高温合金。

祝令刚　男，博士，讲师。主要研究方向：材料的力学性能、缺陷结构以及材料中的物质扩散。

Interaction between metallic and nonmetallic elements on grain boundary of nickel

FANG Shujuan, SHI Songxin, ZHU Linggang*, ZHOU Jian

(School of Materials Science and Engineering, Beijing University of Aeronautics and Astronautics, Beijing 100083, China)

Abstract: First-principles calculations based on density functional theory are used to study the segregation of metallic and nonmetallic elements as well as their interactions on the grain boundary (GB) of nickel. Two typical GBs are constructed: close-packed GB and quite open GB. According to the solution energy calculation, both the metallic and nonmetallic elements show strong segregation behavior on the Σ5 GB, while for the Σ3 GB, the segregation is not so significant. The interaction energies between the common transition alloying elements and nonmetallic elements C, H, O, N and B are calculated when they occupy the most energy-favorable positions. It is found that Ru, Re, W and Ta exhibit strong repulsion to O, indicating beneficial effects for the oxidation resistance behavior; Ta shows strong repulsive interaction to H, which should be helpful for the inhibition of the hydrogen-embrittlement. The present work is useful for the grain boundary engineering of Ni-based superalloy by systematically studying the interaction between metallic and nonmetallic elements at grain boundaries.

Keywords: grain boundary (GB) segregation; Ni-based superalloy; hydrogen-embrittlement; point defect; first-principles calculation

Received: 2017-05-09; **Accepted**: 2017-06-12; **Published online**: 2017-06-29 11:59

URL: kns.cnki.net/kcms/detail/11.2625.V.20170629.1159.002.html

Foundation items: National Natural Science Foundation of China (51401009, 51571008)

* **Corresponding author**. E-mail: lgzhu7@buaa.edu.cn

北京航空航天大学学报

北航学报 赠阅

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2017.0295

高纯 V_2AlC 的无压烧结制备及表征

全下文载

高少华，缪奶华，周健*
（北京航空航天大学 材料科学与工程学院，北京 100083）

摘 要：采用无压烧结的方法，以 V、Al、C 混合粉末为原料制备 V_2AlC 粉体材料。通过不同烧结温度下物相的演变过程对反应路径进行研究，同时探究了烧结助剂 NaF 对烧结过程的影响。实验结果表明，在 1 300～1 500℃温度区间内 V_3Al_2、C 和 VC 发生反应生成 V_2AlC 相，且无压烧结制备高纯 V_2AlC 的最佳工艺为 1 500℃保温 2 h，元素摩尔配比为 V∶Al∶C=2∶1.2∶1。此外，烧结助剂 NaF 的使用加快了反应过程，并使得反应温度降至 1 400℃。实验得到的高纯度、颗粒尺寸分布适中（40～100 μm）的 V_2AlC 可用做提高材料耐磨性的增强体以及二维材料 V_2C 的前驱体。

关 键 词：MAX 相；V_2AlC；无压烧结；反应路径；烧结助剂
中图分类号：V254.2；TQ127.1
文献标识码：A **文章编号**：1001-5965(2018)04-0868-06

MAX 相是一类具有六方晶体结构（$P6_3/mmc$）的三元过渡金属碳化物或氮化物，其中 M 为过渡族元素，A 为ⅢA 或者ⅣA 族元素，X 为 C 或 N 元素[1-2]。目前，实验发现的 MAX 相有 70 多种。基于材料的成键以及结构特点，这类化合物兼具金属和陶瓷的优良特性[3-5]。V_2AlC 作为 MAX 相的一种，拥有低密度（4.07 g/cm^3）[2]，良好的抗氧化性（500℃以下）[6]，高的热导率（48 W/(m·K)）[7]，良好的断裂韧性以及抗热震性[8]等特点。因此，V_2AlC 可用于耐磨材料，金属基复合材料的增强相材料，加热元件，喷嘴等方面。此外，V_2AlC 粉末与 HF 反应获得的 V_2C 二维材料在锂离子电池、电容器和催化剂载体等方面具有极大的应用价值[9-12]。

V_2AlC 广泛应用的前提是制备成本低。目前，用于合成 V_2AlC 的方法包括热等静压（Hot Isostatic Pressing，HIP）[6]、热压（Hot Pressing，HP）[8]和放电等离子体烧结（Spark Plasma Sintering，SPS）等[13-14]，这些方法主要用于制备块状的 V_2AlC。然而，作为一种便捷、高效的制备方法，无压烧结制备 V_2AlC 粉体材料还没有详细的报道。本文通过无压烧结制备出高纯的 V_2AlC 粉体材料，讨论了 V_2AlC 的反应机理，不同成分配比对纯度的影响以及烧结助剂对烧结温度的影响。

1 实 验

实验所用的原料为 V、Al 和 C 粉末（99.9%，300 目）。粉末按照 V∶Al∶C=2∶x∶1（x=1.0，1.1，1.2，1.3）的摩尔配比进行称量，然后将称量好的粉末放入真空球磨罐中球磨 10 h。球磨前将罐内气体置换为氩气。混合均匀的粉末放置于刚玉坩埚中，并用坩埚盖密封。无压烧结在真空管式烧结炉中进行，升温速率为 5℃/min，并通氩气进行保护。为了研究反应过程中物相的演变过程，确定最佳的烧结温度，对 2V/1.2Al/C 混合粉末进行 700、900、1 100、1 300、1 400、1 500℃下保

收稿日期：2017-05-10；**录用日期**：2017-06-30；**网络出版时间**：2017-09-13 14:34
网络出版地址：kns.cnki.net/kcms/detail/11.2625.V.20170913.1434.002.html
基金项目：国家自然科学基金（51571008）
* **通信作者**. E-mail：jzhou@buaa.edu.cn

引用格式：高少华，缪奶华，周健. 高纯 V_2AlC 的无压烧结制备及表征[J]. 北京航空航天大学学报，2018，44(4)：868-873.
GAO S H，MIAO N H，ZHOU J. Synthesis and characterization of high purity V_2AlC prepared by pressureless sintering[J]. Journal of Beijing University of Aeronautics and Astronautics，2018，44(4)：868-873 (in Chinese).

温 2 h 的烧结实验。由于加热过程中 Al 的挥发损失,Al 的用量对产物纯度存在显著的影响。因此,在最佳的烧结温度条件下进行不同成分配比的实验以制备出高纯的 V_2AlC 材料。此外,高的烧结温度对无压烧结制备 V_2AlC 提出了更严苛的要求,为了降低烧结温度,以质量分数 4% NaF 作为烧结助剂加入到原始粉末中,进行烧结过程的对比研究[15]。

采用 X 射线衍射仪(X-ray diffraction,XRD,型号为 Rigaku D/max2500PC)对烧结后的样品进行物相分析,并通过参比强度(Relative Intensity Ratio,RIR)方法对主相进行定量分析。通过扫描电子显微镜(Scanning Electron Microscope,SEM,型号为 JSM-6010)观察粉末样品的颗粒形貌。此外,通过激光粒度分析仪(Laser particle size analyzer,型号为 Mastersizer 3000E)分析烧结样品的颗粒尺寸分布。

2　结果与讨论

图 1 为在 700、900、1 100、1 300、1 400、1 500℃,保温 2 h 条件下烧结样品的 XRD 图谱。随着烧结温度的升高,物相发生一系列的转变,不同温度下物相组成如表 1 所示。当烧结温度为 700℃时,烧结样品中的物相除原始粉末元素外,产生了 2 种新的物相 $Al_{45}V_7$ 和 Al_3V。生成新物相的反应过程如下:

$$7V + 45Al \rightarrow Al_{45}V_7 \quad (1)$$

$$V + 3Al \rightarrow Al_3V \quad (2)$$

当温度升至 900℃,Al 以及 $Al_{45}V_7$ 相消失,同时出现新的物相 Al_8V_5 和 V_3Al_2,这与热压工

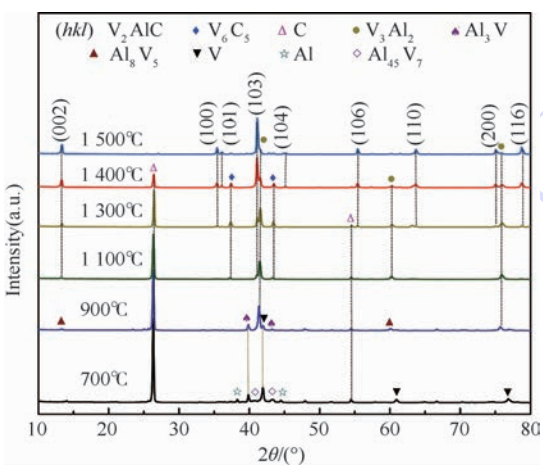


图 1　不同温度下无压烧结 2V/1.2Al/C 混合粉末的 XRD 图谱

Fig. 1　XRD patterns of 2V/1.2Al/C mixed powders pressureless sintered at different temperatures

表 1　不同温度下的物相组成

Table 1　Phase composition at different temperatures

温度/℃	物相组成
700	V, Al, C, $Al_{45}V_7$, Al_3V
900	V, C, Al_3V, Al_8V_5, V_3Al_2
1 100	C, V_3Al_2, V_2AlC, V_6C_5
1 300	C, V_3Al_2, V_2AlC, V_6C_5
1 400	V_2AlC, V_3Al_2, V_6C_5, C
1 500	V_2AlC, V_3Al_2

艺过程中存在稳定的 Al_8V_5 相不同[8]。其中 V_3Al_2 相不是一种稳定物相,这是由于混合粉末中 V、Al 元素的含量远远偏离 Al_8V_5 物相的元素配比,在高温条件下 Al_8V_5 相与剩余的 V 发生反应,形成 V_3Al_2 固溶体。这一结果与 V-Al 二元相图一致[16]。在 900℃下可能发生的反应如下:

$$7V + 8Al_3V \rightarrow 3Al_8V_5 \quad (3)$$

$$7V + Al_8V_5 \rightarrow 4V_3Al_2 \quad (4)$$

当温度升高到 1 100℃,Al_3V 和 Al_8V_5 物相消失,同时出现 V_2AlC 和 V_6C_5 相,说明 V_2AlC 相出现的起始温度在 900 ~ 1 100℃。结合 V-C 二元相图可以得到,V_6C_5 相为室温物相,在高温状态下以 VC 形式出现[17]。对比发现,在 1 300℃与 1100℃条件下烧结样品的物相组成相同,但是,随着温度的升高 C 的峰减弱,V_2AlC 的峰增强,说明在该过程中,V_3Al_2,C 和 VC 共同反应生成 V_2AlC 相。该阶段的反应过程可能为

$$2C + V_3Al_2 \rightarrow V_2AlC + VC + Al \quad (5)$$

$$VC + V_3Al_2 + C \rightarrow 2V_2AlC \quad (6)$$

当烧结温度升至 1 400℃时,没有新的物相生成,C 的峰进一步减弱,V_2AlC 峰进一步增强,说明反应过程同式(5)和式(6)。在 1 500℃条件下,样品中包含主要物相 V_2AlC 和微量的 V_3Al_2,同时 C 和 V_6C_5 相消失,说明反应已完全,可以确定 1 500℃为合适的反应温度。

不同温度下的烧结实验得到最佳的烧结温度为 1 500℃。烧结过程中 Al 的损失比较严重,为了获得高纯的 V_2AlC 材料,避免碳化物杂质的出现,在最佳烧结温度下进行了不同 Al 摩尔配比的烧结实验。原材料中元素摩尔配比为 V ∶ Al ∶ C = 2 ∶ x ∶ 1(x = 1.0,1.1,1.2,1.3)。烧结试样的 XRD 图谱如图 2 所示。由图 2 可知 V_3Al_2、V_6C_5 与 V_2AlC 相共同存在。随着铝元素相对含量的增加,V_6C_5 的含量先降低后增加,在摩尔配比为 V ∶ Al ∶ C = 2 ∶ 1.2 ∶ 1 时达到最低,说明为了得到高纯的 V_2AlC 相,最佳的元素配比为 V ∶ Al ∶ C = 2 ∶ 1.2 ∶ 1。

北航学报 赠阅

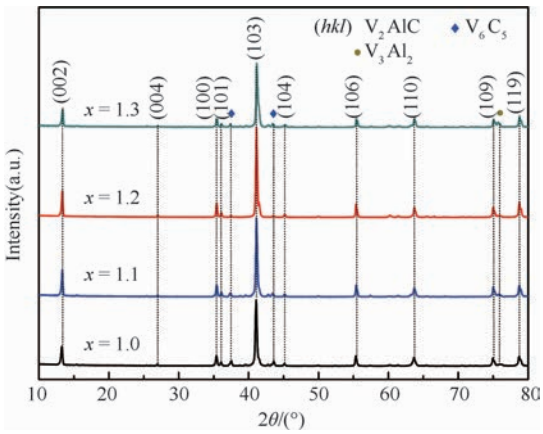


图 2 不同元素配比(2V ∶ xAl ∶ C)的烧结样品的 XRD 图谱

Fig. 2 XRD patterns of sintered samples with different molar ratios of 2V ∶ xAl ∶ C

实验通过 RIR 方法对各物相的相对含量进行测定。由于主相 V_2AlC 的 RIR 值未知,质量分数为 30% 的 α-Al_2O_3 微粉作为内标物质被掺入到烧结后的粉末样品中。V_3Al_2 的(110)晶面和 V_6C_5 的$(11\bar{2})$晶面衍射峰作为分析对象,得到两者的相对含量。由于烧结试样中没有非晶相的存在,因此剩余物相为主相 V_2AlC。经过归一化处理得到在最佳制备工艺条件下(1 500℃,保温 2 h,元素组成为2V ∶ 1.2Al ∶ C),主相 V_2AlC 的质量分数达到 95.3%。

V_2AlC 的合成需要 1 500℃左右的高温,极大地限制了 V_2AlC 的制备和应用。为了降低反应温度,将 4% NaF 作为烧结助剂加入到混合粉末中,并进行了不同温度下的烧结实验,烧结温度分别为 1 300 、1 350 和 1 400℃。图 3 为加入烧结助剂 NaF 前后在 1 300℃保温 2 h 的烧结样品的 XRD 图谱。对比各物相的特征峰发现,加入 NaF 后,C 的衍射峰明显减弱,中间物相 V_6C_5 的衍射峰增强,同时目标产物 V_2AlC 的特征峰明显增强。不同温度下烧结样品的 XRD 图谱如图 4 所示。通过图 4 可以看出在 1 400℃条件下,反应物 C 的衍射峰消失,主要物相为 V_2AlC,说明反应已基本完成。上述实验结果表明,烧结助剂 NaF 的使用加快了反应速率,同时将烧结温度从 1 500℃降至 1 400℃,原因是由于高温条件下熔融的 NaF 促进了传质过程。通过物相分析发现,新的物相 V_4AlC_3 与 V_2AlC 相共同存在(见图 4)。Hu 等[8] 的报道指出 V_4AlC_3 相可能由 V_2AlC 与 VC 在高温条件下反应生成,反应方程式如下:

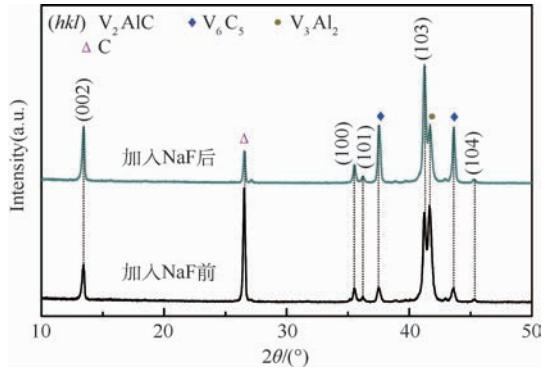


图 3 加入烧结助剂 NaF 前后烧结样品的 XRD 图谱(烧结温度为 1 300℃,保温时间 2 h)

Fig. 3 XRD patterns of samples before and after adding NaF (sintered at 1 300℃ for 2 h)

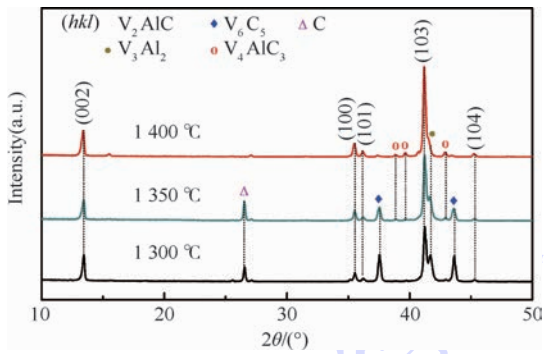


图 4 元素摩尔配比 V ∶ Al ∶ C = 2 ∶ 1.2 ∶ 1,加入烧结助剂的混合粉末在不同温度下烧结试样的 XRD 图谱

Fig. 4 XRD patterns of 2V/1.2Al/C mixed powders with NaF sintered at different temperatures

$$V_2AlC + 2VC \rightarrow V_4AlC_3 \quad (7)$$

与 V 同族的 Ta 元素,其对应的 Ta_2AlC 在 1 650℃高温下稳定性降低,同样存在着部分 Ta_2AlC 相向 Ta_4AlC_3 相转化的现象[18]。由于烧结过程中存在 Al 的挥发损失,即使使用过量的 Al,在样品表面仍存在一定含量的 VC。VC 的存在加速了反应式(7)的进行。为了进一步验证 V_2AlC 与 V_4AlC_3 的相对稳定性关系,对反应式(7)进行验证。通过第一性原理计算对各物相的晶格参数和总能量进行合理的优化,作为进一步分析的基础,结果见表 2。第一性原理计算是基于密度泛函理论,使用 Perdew 等提出的广义梯度近似法(GGA-PBE)进行的[19]。优化后各物相的晶格参数与 XRD 数据相一致,数据误差较小,进一步验证了优化的准确性。定义反应后各物质的总能减去反应前各物质的总能为反应的形成能,如方程式(8)所示:

$$\Delta E = E_{tot}(V_4AlC_3) - E_{tot}(V_2AlC) - 2E_{tot}(VC) \quad (8)$$

式中：E_{tot}为物质的总能。由于反应会向能量降低的方向进行，若形成能为负值，在一定的条件下反应向正向进行。计算得到反应形成能为 -0.34 eV，说明反应式(7)在高温条件下能够自发进行，因此 V_4AlC_3 的出现可能是由 V_2AlC 与 VC 反应产生的。

图 5 为通过激光粒度分析仪测得的 V_2AlC 的颗粒尺寸分布。通过分析发现无压烧结获得的样品的颗粒尺寸主要集中在 40 ~ 100 μm 之间。结合 SEM 照片(图 6(a))不难发现，部分颗粒尺寸达到 150 μm 是小颗粒凝聚造成的。图 6(b)为烧结样品经过破碎处理后的形貌。观察发现，在颗粒边缘处出现清晰的层片状纹理。同时，V_2AlC 材料在受到外力作用后，材料发生扭折产生分层(图 6(c))，显示出 MAX 相独特的层片状结构特征以及弹性各向异性。综上，实验获得的具有高的纯度，适中的颗粒尺寸的 V_2AlC 粉末对于 V_2AlC 作为耐磨材料以及 V_2C 二维材料前驱体材料等的应用具有重要意义。

表 2　各化合物的晶格参数及总能

Table 2　Lattice parameters and total energies of different compounds

化合物	$a=b$/nm		c/nm		总能/eV
	计算	XRD	计算	XRD	
V_2AlC	0.291 2	0.292 4	1.310 7	1.320 4	-32.847
VC	0.415 5		0.415 5		-18.945
V_4AlC_3	0.291 9	0.293 1[20]	2.274 4	2.271 9[20]	-71.085

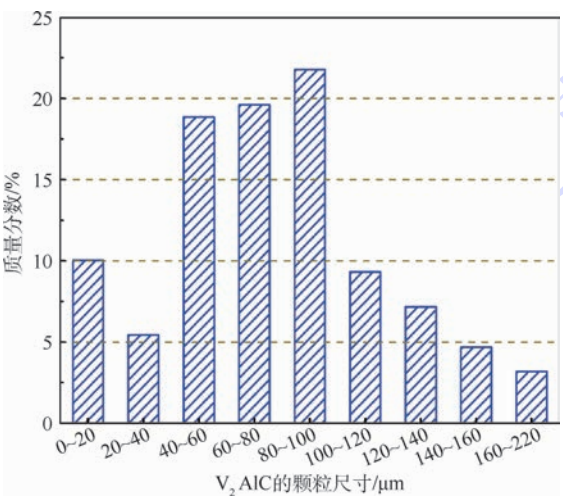
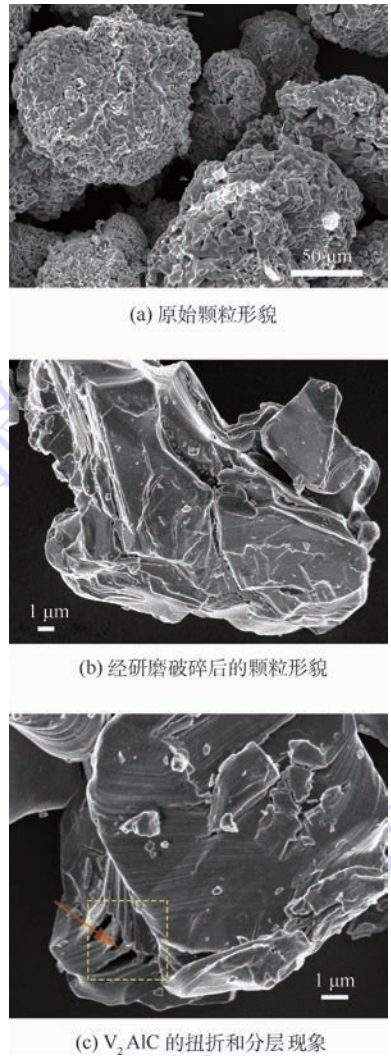


图 5　V_2AlC 的颗粒尺寸分布柱状图

Fig. 5　Histogram of size distribution of V_2AlC powders



(a) 原始颗粒形貌

(b) 经研磨破碎后的颗粒形貌

(c) V_2AlC 的扭折和分层现象

图 6　V_2AlC 的 SEM 照片

Fig. 6　SEM photograph of V_2AlC powders

3　结　论

1) 本文以 V、Al 和 C 混合粉末为原材料通过无压烧结的方法成功制备出 V_2AlC 粉末。通过对反应机理的研究发现，在 1 300 ~ 1 500℃温度区间，V_3Al_2、C 和 VC 发生反应生成 V_2AlC 相。

2) 烧结温度和 Al 含量对产物纯度有较大的影响。以 V ∶ Al ∶ C = 2 ∶ 1.2 ∶ 1 的摩尔配比的混合粉末在 1500℃下保温 2 h 可以得到高纯度的 V_2AlC(纯度达到 95.3%)。

3) 烧结助剂 NaF 可以显著加快反应过程并将反应温度降低至约 1 400℃。

4) V_2AlC 粉末(尺寸分布在 40 ~ 100 μm)适合用作耐磨材料以及 V_2C 二维材料的前驱体材料。

参考文献(References)

[1] WANG J Y, ZHOU Y C. Recent progress in theoretical predic-

北航学报 赠阅

tion, preparation, and characterization of layered ternary transition-metal carbides[J]. Annual Review of Materials Research, 2009, 39(39): 415-443.

[2] BARSOUM M W. The $M_{N+1}AX_N$ phases: A new class of solids: T hermodynamically stable nanolaminates[J]. Progress in Solid State Chemistry, 2000, 28(1): 201-281.

[3] SUN Z M, MUSIC D, AHUJA R, et al. Bonding and classification of nanolayered ternary carbides[J]. Physical Review B, 2004, 70(9): 092102.

[4] DAHLQVIST M, ALLING B R, ROS N J. Stability trends of MAX phases from first principles[J]. Physical Review B, 2010, 81(22): 220102.

[5] SUN Z M. Progress in research and development on MAX phases: A family of layered ternary compounds[J]. International Materials Reviews, 2011, 56(3): 143-166.

[6] GUPTA S, BARSOUM M W. Synthesis and oxidation of V_2AlC and $(Ti_{0.5}, V_{0.5})_2AlC$ in air[J]. Journal of the Electrochemical Society, 2004, 151(2): D24-D29.

[7] HETTINGER J, LOFLAND S, FINKEL P, et al. Electrical transport, thermal transport, and elastic properties of M_2AlC (M = Ti, Cr, Nb, and V)[J]. Physical Review B, 2005, 72(11): 115-120.

[8] HU C F, HE L F, LIU M Y, et al. In situ reaction synthesis and mechanical properties of V_2AlC[J]. Journal of the American Ceramic Society, 2008, 91(12): 4029-4035.

[9] NAGUIB M, MOCHALIN V N, BARSOUM M W, et al. 25th anniversary article: MXenes: A new family of two-dimensional materials[J]. Advanced Materials, 2014, 26(7): 992-1005.

[10] CHEN J, CHEN K, TONG D Y, et al. CO_2 and temperature dual responsive "Smart" MXene phases[J]. Chemical Communications, 2015, 51(2): 314-317.

[11] WANG L, YUAN L Y, CHEN K, et al. Loading actinides in multi-layered structures for nuclear waste treatment: The first case study of uranium capture with vanadium carbide MXene [J]. ACS Applied Materials & Interfaces, 2016, 8(25): 16396-16403.

[12] CHEN J, XIAO P, GU J C, et al. A smart hybrid system of Au nanoparticle immobilized PDMAEMA brushes for thermally adjustable catalysis[J]. Chemical Communications, 2014, 50(10): 1212-1214.

[13] SHI L, OUISSE T, SARIGIANNIDOU E, et al. Synthesis of single crystals of V_2AlC phase by high-temperature solution growth and slow cooling technique[J]. Acta Materialia, 2015, 83: 304-309.

[14] LI X, LI X, LIANG B Y. Synthesis of phase purity V_2AlC via self-propagation high temperature sintering[J]. International Journal of Materials Research, 2013, 104(6): 603-605.

[15] SUN Z M, ZHANG Y, ZHOU Y C. Synthesis of Ti_3SiC_2 powders by a solid-liquid reaction process[J]. Scripta Materialia, 1999, 41(1): 61-66.

[16] HALLSTEDT B. Thermodynamic evaluation of the Al-V-C system[J]. Calphad-Computer Coupling of Phase Diagrams and Thermochemistry, 2013, 41(6): 156-159.

[17] LIPATNIKOV V, GUSEV A, ETTMAYER P, et al. Phase transformations in non-stoichiometric vanadium carbide[J]. Journal of Physics: Condensed Matter, 1999, 11: 163-184.

[18] HU C F, ZHANG J, BAO Y W, et al. In-situ reaction synthesis and decomposition of Ta_2AlC[J]. International Journal of Materials Research, 2008, 99(1): 8-13.

[19] PERDEW J P, BURKE K, ERNZERHOF M. Generalized gradient approximation made simple[J]. Physical Review Letters, 1996, 77(18): 3865-3868.

[20] HU C F, ZHANG J, WANG J M, et al. Crystal structure of V_4AlC_3: A new layered ternary carbide[J]. Journal of the American Ceramic Society, 2008, 91(2): 636-639.

作者简介:

高少华 男,硕士研究生。主要研究方向:MAX 相材料的制备及应用。

周健 男,博士,副教授,硕士生导师。主要研究方向:金属间化合物、热电材料、计算材料。

北京航空航天大学学报

北航学报 赠阅

Synthesis and characterization of high purity V_2AlC prepared by pressureless sintering

GAO Shaohua, MIAO Naihua, ZHOU Jian*

(School of Materials Science and Engineering, Beijing University of Aeronautics and Astronautics, Beijing 100083, China)

Abstract: In this paper, we reported the synthesis of high purity V_2AlC by pressureless sintering with V, Al and C mixed powders. The reaction paths were studied and discussed according to the phase transformation at different sintering temperatures. Besides, the effect of sintering aid NaF on singtering process was stuied. The experimental results show that V_2AlC was synthesized by the reaction of V_3Al_2, VC and C within the temperature range of 1 300 – 1 500℃ and that high purity V_2AlC could be obtained with the optimized molar ratio of V ∶ Al ∶ C = 2 ∶ 1.2 ∶ 1 at 1 500℃ for 2 h. In addition, the use of sintering aid NaF could greatly promote the reaction process and reduce the reaction temperature to 1 400℃. Based on the high purity and suitable size distribution (40 ~ 100 μm), the synthesized powders can be easily used as the reinforcement material or V_2C precursor material.

Keywords: MAX phase; V_2AlC; pressureless sintering; reaction path; sintering aid

北京航空航天大学学报

Received: 2017-05-10; **Accepted**: 2017-06-30; **Published online**: 2017-09-13 14:34

URL: kns.cnki.net/kcms/detail/11.2625.V.20170913.1434.002.html

Foundation item: National Natural Science Foundation of China (51571008)

* **Corresponding author**. E-mail: jzhou@buaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2017.0298

Zn-Cr_2AlC 复合材料的制备与摩擦磨损性能

李志，赵文月，缪奶华*
（北京航空航天大学 材料科学与工程学院，北京 100083）

摘 要：为改善金属 Zn 的摩擦磨损性能，采用热压法制备 Cr_2AlC 陶瓷颗粒增强 Zn 基复合材料，并研究了 Cr_2AlC 质量分数对复合材料的金相组织、维氏硬度、相对密度及摩擦磨损性能的影响。结果表明，复合材料的硬度随着 Cr_2AlC 质量分数的适量增加而明显升高。当 Cr_2AlC 的质量分数达到 20% 时，复合材料的硬度是纯 Zn 的 1.52 倍。摩擦磨损实验表明，Cr_2AlC 颗粒的引入，可显著改善复合材料的摩擦磨损性能，摩擦系数由纯 Zn 的 0.75 降到 Zn-20% Cr_2AlC 的 0.65，Zn-30% Cr_2AlC 的磨损率相比纯 Zn 下降了 80.54%。分析磨损表面形貌，得出其磨损类型为磨粒磨损和剥层磨损。

关 键 词：Cr_2AlC；Zn 基复合材料；维氏硬度；组织形貌；摩擦磨损性能
中图分类号：TB333
文献标识码：A **文章编号**：1001-5965(2018)04-0874-05

锌合金具有优良的机械性能、摩擦学性能、优异的铸造性及环境友好等优点[1-6]，可以代替铸造铝合金、轴承青铜、铸铁、塑料甚至钢材等，用于制造在中等使用温度条件下操作的摩擦元件。而锌合金存在熔点较低、耐热性较差和抗蠕变强度低等问题[7-8]，使其应用范围受限。因此研究人员采用不同的增强材料（颗粒、晶须或短纤维），以期获得更强的机械性能和摩擦学性能。传统的增强相陶瓷颗粒与基体之间的界面结合性较差、热膨胀系数差异较大，使得其性能仍然不太理想。最近，兼具金属与陶瓷优异特性的三元层状化合物 MAX（$M_{n+1}AX_n$ 的缩写，其中 M 代表过渡族金属元素，A 代表第三和第四主族元素，X 代表 C 和/或 N，$n=1,2,3$）相[9-11]增强相的使用，可有效提高金属基复合材料的性能。Gupta 等[12]用热压法制备了 Ti_3SiC_2 增强 Zn 基复合材料，结果表明，在基体中添加一定量的 Ti_3SiC_2 颗粒可提升材料的摩擦磨损性能，但对力学性能几乎没有增益效果；Li 等[13]利用无压烧结结合热压烧结的方法制备了 Ti_3AlC_2/ZA27 复合材料，实验得到材料的拉伸强度和弯曲强度分别为 335 MPa 和570 MPa，力学性能明显改善，这得益于增强相颗粒与基体之间的良好界面结合以及晶粒的细化。Cr_2AlC[14-15]与 MAX 相[9]体系其他常用作增强相的化合物都具有相近的弹性模量和硬度值，但 Cr_2AlC 的热膨胀系数更大，与 Zn 更匹配，因此 Cr_2AlC 可能更适合作为 Zn 基复合材料的增强相。

本实验首先采用无压烧结法合成了高纯的 Cr_2AlC 粉末，然后采用热压烧结法合成了 Zn-Cr_2AlC 复合材料，并表征了复合材料的基本物理性能，进一步研究了 Cr_2AlC 颗粒的含量对 Zn-Cr_2AlC 复合材料的摩擦磨损性能的影响。

1 实 验

采用无压烧结法合成高纯 Cr_2AlC 粉末，过筛后得到的粉末颗粒尺寸不超过 38 μm。用 X 射线

收稿日期：2017-05-11；**录用日期**：2017-07-21；**网络出版时间**：2017-09-14 09:47
网络出版地址：kns.cnki.net/kcms/detail/11.2625.V.20170914.0947.002.html
基金项目：国家自然科学基金（51571008）
*通信作者. E-mail：nhmiao@buaa.edu.cn

引用格式：李志，赵文月，缪奶华. Zn-Cr_2AlC 复合材料的制备与摩擦磨损性能[J]. 北京航空航天大学学报，2018，44(4)：874-878.
LI Z，ZHAO W Y，MIAO N H. Preparation and tribological properties of Zn-Cr_2AlC composites[J]. Journal of Beijing University of Aeronautics and Astronautics，2018，44(4)：874-878 (in Chinese).

衍射仪（XRD）进行物相检测并进行定量分析，得到的 Cr_2AlC 粉末纯度大于 98%。购买得到的 Zn 粉颗粒尺寸小于 38 μm，纯度大于 99.9%。

在 Zn 粉中加入不同比例的 Cr_2AlC 粉末，增强相质量分数在 5% ~30% 之间变化（分别记为 Zn-5%Cr_2AlC、Zn-10%Cr_2AlC、Zn-20%Cr_2AlC、Zn-30%Cr_2AlC），用球磨机充分混合 12 h 后得到混合均匀的 Zn-Cr_2AlC 复合粉末。将粉末分别装进石墨模具中，在氩气气氛保护下，以 10℃/min 的升温速率升至 400℃，施加 50 MPa 压力，保温3 h，卸载压力，停止加热，随炉冷却至室温。为了对比材料性能的变化，在相同实验条件下制备了纯 Zn 样品。

制备的纯 Zn 和复合材料块状样品，用阿基米德排水法测量其密度，用型号为 D/MAX-2500 的 XRD 分析材料的物相，用 LEICA DM400 光学显微镜观察材料的金相组织，用 FM800 显微维氏硬度计测量硬度。用 UMT-2 摩擦磨损试验机进行摩擦磨损实验。选用 45# 钢钢球为对磨体，摩擦方式为球-块式往复运动摩擦。在室温下进行干摩擦试验，滑块运动频率 5 Hz，振幅 4 mm，载荷 1 N，滑动时间 20 min。相同条件下的样品重复 3 次实验。磨损率计算式为

$$WR = \frac{m_i - m_f}{\rho N d} \tag{1}$$

式中：WR 为材料的磨损率；m_i 为样品的初始质量；m_f 为样品磨损后的质量；ρ 为样品的实际密度；N 为实验加载载荷；d 为实验过程中的滑动距离。研究 Cr_2AlC 含量对复合材料摩擦磨损性能的影响，用 JSM7500 扫描电子显微镜（SEM）观察磨损后的材料表面形貌。

2　结果与分析

2.1　物相与金相

图 1 为纯 Zn 和复合材料样品的 XRD 图谱，可以发现，除了基体 Zn 和增强相 Cr_2AlC 的衍射峰之外，没有检测到任何其他物质的衍射峰，表明复合材料在制备过程中未发生氧化，并且基体 Zn 与增强相 Cr_2AlC 之间没有相互反应。

纯 Zn 和复合材料的金相组织照片如图 2 所示。从图 2(a)可以看出纯 Zn 样品中分布有一些小气孔。当 Cr_2AlC 质量分数为 5% 和 10% 时（见图 2(b)和 2(c)），基体与增强相界面接合良好，两相分布也较为均匀，增强相 Cr_2AlC 颗粒大小约为 30 μm，仍有少量气孔。而当 Cr_2AlC 质量分数为 20% 时，如图 2(d)所示，Cr_2AlC 颗粒大量团聚，分布不均匀，同时在基体与增强颗粒界面处有孔洞存在。Cr_2AlC 含量继续增长到 30%，如图 2(e)所示，Cr_2AlC 颗粒团聚严重，两相界面处存在大量的孔洞，使材料呈现为多孔的蓬松结构。

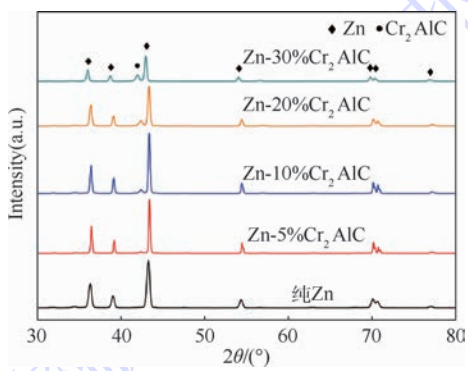
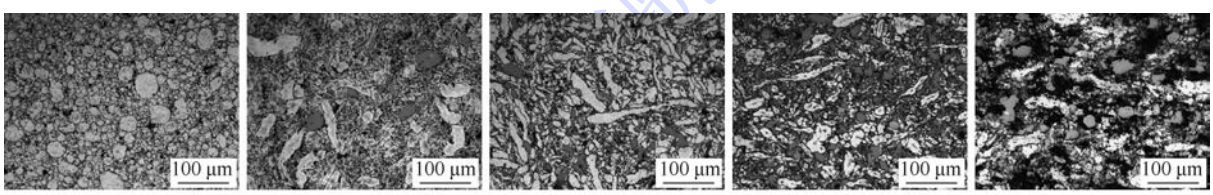


图 1　纯 Zn 及复合材料的 XRD 图谱

Fig. 1　XRD patterns of pure Zn and composites



(a) 纯Zn　(b) Zn-5%Cr_2AlC　(c) Zn-10%Cr_2AlC　(d) Zn-20%Cr_2AlC　(e) Zn-30%Cr_2AlC

图 2　纯 Zn 及复合材料的金相组织

Fig. 2　Microstructures of pure Zn and composites

2.2　维氏硬度与相对密度

图 3 为 Zn-Cr_2AlC 复合材料的维氏硬度和相对密度随 Cr_2AlC 的质量分数变化的曲线。由图 3可见，当 Cr_2AlC 的质量分数从 5% 增长到 20%时，材料维氏硬度逐渐提高，最高可达到纯 Zn 硬度的 1.52 倍；Cr_2AlC 质量分数为 30% 时，维氏硬度反而显著降低。原因在于增强颗粒 Cr_2AlC 的硬度大大高于基体 Zn，由于增强效应，适量的 Cr_2AlC 颗粒可使复合材料的维氏硬度明显提高，而过量的 Cr_2AlC 颗粒倾向于团聚，使两

北航学报 赠阅

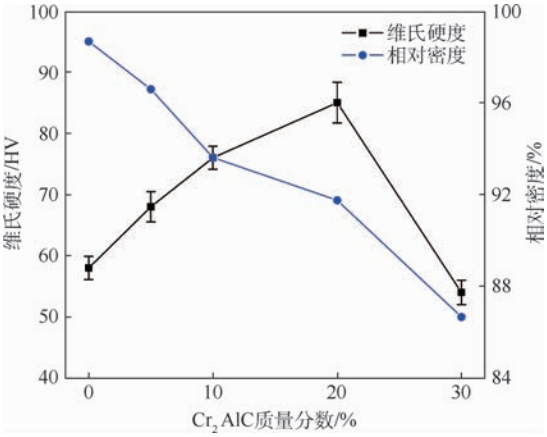


图 3 纯 Zn 及复合材料的硬度和相对密度

Fig. 3 Vickers-hardness and relative density of pure Zn and composites

相分布不均匀,影响了两相间的结合,孔隙增多,故而材料的硬度值不升反降。此外,材料的相对密度随增强相质量分数的增加而减小。一方面,Zn 的熔点为 419℃,在 400℃下为固相烧结,此时材料的致密化方式主要为固态扩散[16];另一方面,增强相质量分数越高,越容易团聚,会阻碍基体晶粒的晶界流动性,使材料致密化困难,导致复合材料的气孔逐渐增多,相对密度逐渐下降。

2.3 摩擦磨损性能

图 4(a)~(e)给出了纯 Zn 及复合材料的摩擦系数随滑动时间的变化曲线。由图 4(a)可见,纯 Zn 的摩擦系数约为 0.75,在整个滑动摩擦过程中摩擦系数都比较稳定。由图 4(b)可见,复合材料 Zn-5% Cr_2AlC 在摩擦运开始的前 5~10 min内,其摩擦系数随时间的变化幅度较大,为磨合阶段;10 min 后,过渡到稳定阶段,材料的摩擦系数随时间变化的曲线变得平滑,基本稳定在 0.85 左右。图 4(c)和 4(d)显示了与图 4(b)一样的摩擦系数变化趋势,最终 Zn-10% Cr_2AlC 和 Zn-20% Cr_2AlC 复合材料的摩擦系数分别稳定在 0.70 和 0.65 左右。由图 4(e)可见,当复合材料中 Cr_2AlC 的质量分数达到 30% 时,其摩擦系数异常稳定,波动性极小,为 0.15 ± 0.02,比纯 Zn 低 80%。图 4(f)为各复合材料在稳定状态下的摩擦系数。由图 4(f)可见,当 Cr_2AlC 质量分数为 5% ~30% 时,随着 Cr_2AlC 的质量分数的上升,复合材料的摩擦系数呈下降趋势。Zn 基体中添加 Cr_2AlC 颗粒对复合材料的摩擦磨损性能是有利的。特别是对于 Zn-30% Cr_2AlC 复合材料,其摩擦系数降到了 0.15。这是由于 Cr_2AlC 的高含量导致复合材料难以致密化,材料表面气孔较多,而复合材料与磨球之间的接触形式是点接触,样品的多孔结构使材料的摩擦系数急剧下降。因此,增强相 Cr_2AlC 颗粒的引入,降低了复合材料的摩擦系数,增强了材料的耐磨性。

复合材料的磨损率随 Cr_2AlC 质量分数变化关系如图 5 所示。纯 Zn 的磨损率为 9.61×10^{-4} $mm^3\cdot N^{-1}\cdot m^{-1}$,随着 Cr_2AlC 质量分数的增加,复合材料的磨损率逐渐降低,Zn-30% Cr_2AlC

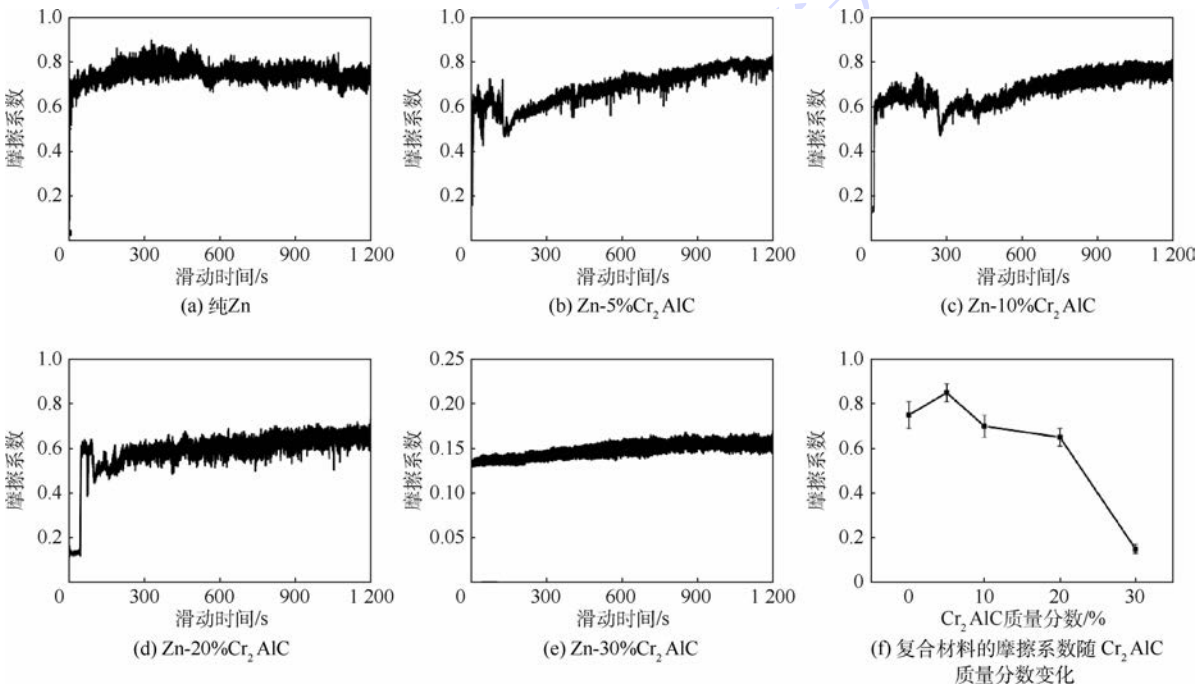


图 4 摩擦系数随滑动时间的变化

Fig. 4 Friction coefficient versus sliding time

的磨损率最小，为 $1.87\times10^{-4}\ mm^3\cdot N^{-1}\cdot m^{-1}$，比纯 Zn 降低了 80.54%。本实验所测试的复合材料试样磨损率的范围为 $(2\sim9)\times10^{-4}\ mm^3\cdot N^{-1}\cdot m^{-1}$。该结果与 Gupta 等[12]研究的 Ti_3SiC_2 颗粒增强 Zn 基复合材料得到的结果类似。

图 6(a)～(e)为复合材料的磨损表面形貌照片。由图 6(a)可见，纯 Zn 样品磨损表面有粉末碎屑和剥离凹坑。由图 6(b)可见，复合材料 Zn-5% Cr_2AlC 中，Cr_2AlC 含量较低，磨损表面存在松散的磨料碎片和大片脱层，可认为磨损类型是磨料磨损和剥层磨损。由图 6(c)可见，Zn-10% Cr_2AlC 复合材料磨损表面存在浅的犁沟和薄的脱层而没有剥离凹坑。由图 6(d)可见，Zn-20% Cr_2AlC 复合材料磨损表面相对平滑，分布着浅而直的犁沟，磨损程度比其他样品轻。这种磨损表面形貌的演变现象，是引入了具有自润滑性质的 Cr_2AlC 颗粒导致的。由图 6(e)可见，Zn-30% Cr_2AlC 复合材料，由于多孔特征，在磨损过程中脱落的磨料填充于孔隙中，使得样品的磨损表面非常光滑，这种磨损形貌在 Al-Ti_3SiC_2 复合材料相对于 Al_2O_3 的干摩擦实验中也可观察到[17]。结合 Gupta 和 Barsoum[18]研究基于 MAX 相复合材料的摩擦学结论，分析本实验中样品磨损表面的特性，可以得出复合材料样品的磨损机制是磨粒磨损和剥层磨损。

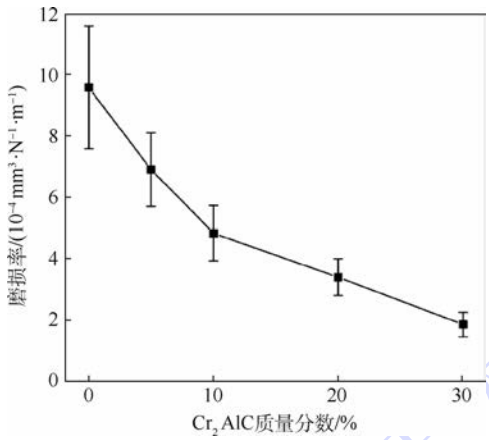


图 5　纯 Zn 及复合材料的磨损率

Fig. 5　Wear rates of pure Zn and composites

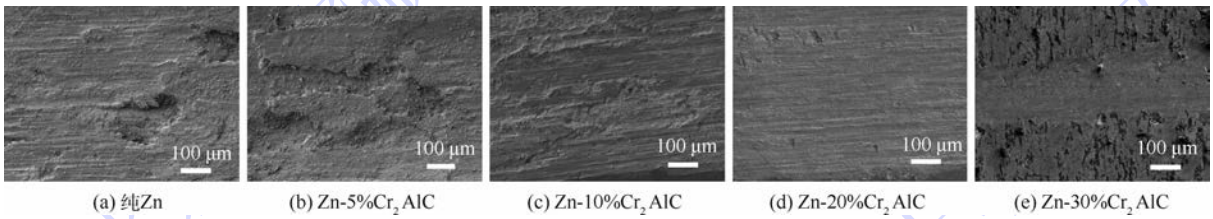


图 6　纯 Zn 及复合材料的磨损表面形貌

Fig. 6　Worn surface morphology of pure Zn and composites

3　结　论

本文采用无压烧结合成了高纯 Cr_2AlC 粉末，采用热压烧结法制备了 Zn-Cr_2AlC 复合材料，研究了 Cr_2AlC 颗粒的加入量对复合材料的金相组织、维氏硬度及摩擦磨损性能等的影响，可得出以下结论：

1）Cr_2AlC 质量分数在 5%～20% 范围内，复合材料的硬度随 Cr_2AlC 质量分数增加而提高，当 Cr_2AlC 质量分数过高(30%)时，复合材料难以致密化，硬度值不升反降。

2）在 Zn 基体中添加 Cr_2AlC 颗粒，摩擦系数由纯 Zn 的 0.75 降低到 Zn-20% Cr_2AlC 复合材料的 0.65；磨损率由纯 Zn 的 $9.61\times10^{-4}\ mm^3\cdot N^{-1}\cdot m^{-1}$ 降低到 Zn-30% Cr_2AlC 复合材料的 $1.87\times10^{-4}\ mm^3\cdot N^{-1}\cdot m^{-1}$，复合材料的摩擦磨损性能明显提升。

3）复合材料在干摩擦实验条件下的磨损机制以磨粒磨损为主，并有轻微的剥层磨损。

参考文献（References）

[1] EL-KHAIR M T A, LOTFY A, DAOUD A, et al. Microstructure, thermal behavior and mechanical properties of squeeze cast SiC, ZrO_2 or C reinforced ZA27 composites[J]. Materials Science and Engineering: A, 2011, 528(6): 2353-2362.

[2] BABIC M, SLOBODAN M, DZUNIC D, et al. Tribological behavior of composites based on ZA-27 alloy reinforced with g-raphite particles [J]. Tribology Letters, 2009, 37(2): 401-410.

[3] RANGANATH G, SHARMA S C, KRISHNA M. Dry sliding wear of garnet reinforced zinc/aluminium metal matrix composites[J]. Wear, 2001, 250(1-12): 1408-1413.

[4] LI Z G. Fabrication of in situ TiB_2 particulates reinforced zinc alloy matrix composite[J]. Materials Letters, 2014, 121: 1-4.

[5] MIROSLAV B, MITROVIC S, ZIVIC F, et al. Wear behavior of composites based on ZA-27 alloy reinforced by Al_2O_3 particles under dry sliding condition [J]. Tribology Letters, 2010, 38(3): 337-346.

[6] DEHSORKHI R N, QODS F, TAJALLY M. Investigation on microstructure and mechanical properties of Al-Zn composite

北航学报 赠阅

during accumulative roll bonding (ARB) process[J]. Materials Science and Engineering:A,2011,530(1):63-72.

[7] PRASAD B K. Sliding wear response of a zinc-based alloy and its composite and comparison with a gray cast iron:Influence of external lubrication and microstructural features[J]. Materials Science and Engineering:A,2005,392(1-2):427-439.

[8] TAVOOSI M, KARIMZADEH F, ENAYATI M H, et al. Bulk Al-Zn/Al_2O_3 nanocomposite prepared by reactive milling and hot pressing methods[J]. Journal of Alloys and Compounds, 2009,475(1-2):198-201.

[9] BARSOUM M W, RADOVIC M. Elastic and mechanical properties of the MAX phases[J]. Annual Review of Materials Research,2011,41(1):195-227.

[10] SUN Z M. Progress in research and development on MAX phases:A family of layered ternary compounds[J]. International Materials Reviews,2013,56(3):143-166.

[11] BARSOUM M W. The $M_{n+1}AX_n$ phases:A new class of solids: Thermodynamically stable nanolaminates[J]. Progress in Solid State Chemistry,2000,28(1):201-281.

[12] GUPTA S, HABIB M A, DUNNIGAN R, et al. Synthesis and characterization of Ti_3SiC_2 particulate-reinforced novel Zn matrix composites[J]. Journal of Materials Engineering and Performance,2015,24(10):4071-4076.

[13] LI H Y, ZHOU Y, CHEN C, et al. Microstructure and mechanical properties of Zn based composites reinforced by Ti_3AlC_2 [J]. Advances in Applied Ceramics,2015,114(6):315-320.

[14] LIN Z, ZHUO M, ZHOU Y, et al. Atomic scale characterization of layered ternary Cr_2AlC ceramic[J]. Journal of Applied Physics,2006,99(7):076109.

[15] TIAN W, WANG P, ZHANG G, et al. Synthesis and thermal and electrical properties of bulk Cr_2AlC[J]. Scripta Materialia, 2006,54(5):841-846.

[16] 雷宇,刘锦云,王敏,等. Cr_2AlC 颗粒增强 Cu 基复合材料的制备及力学性能研究[J]. 粉末冶金技术,2013,31(5):340-343.
LEI Y, LIU J Y, WANG M, et al. Preparation and mechanical properties of Cr_2AlC particulate reinforced Cu matrix composites[J]. Powder Metallurgy Technology,2013,31(5):340-343 (in Chinese).

[17] GUPTA S, HAMMANN T, JOHNSON R, et al. Synthesis and characterization of novel Al-matrix composites reinforced with Ti_3SiC_2 particulates[J]. Journal of Materials Engineering and Performance,2015,24(2):1-7.

[18] GUPTA S, BARSOUM M W. On the tribology of the MAX phases and their composites during dry sliding: A review[J]. Wear,2011,271(9-10):1878-1894.

作者简介:

李志 男,硕士研究生。主要研究方向:金属基复合材料。

缪奶华 男,博士,副教授,硕士生导师。主要研究方向:高熔点结构材料、计算材料学等。

Preparation and tribological properties of Zn-Cr_2AlC composites

LI Zhi, ZHAO Wenyue, MIAO Naihua*

(School of Materials Science and Engineering, Beijing University of Aeronautics and Astronautics, Beijing 100083, China)

Abstract: In order to improve the friction and wear properties of Zn, Cr_2AlC ceramic particle reinforced Zn matrix composites were prepared by hot pressing method. The effects of Cr_2AlC content on the metallographic structure, Vickers-hardness, relative density and tribological properties of the Zn-based composites were investigated. The results show that the hardness of the composites is improved obviously with the proper increase of Cr_2AlC content. When the mass fraction of Cr_2AlC reaches 20%, the hardness of the composite is 1.52 times higher than that of pure Zn. The introduced Cr_2AlC particles can significantly improve the tribological properties of the composites. The friction coefficient decreases from 0.75 in pure Zn to 0.65 in Zn-20% Cr_2AlC, and the wear rate of Zn-30% Cr_2AlC is reduced by 80.54% compared to pure Zn. Analysis on the worn surface morphology indicates that the wear mechanism is abrasive wear and delamination wear.

Keywords: Cr_2AlC; Zn matrix composites; Vickers-hardness; microstructure and morphology; tribological properties

Received: 2017-05-11; **Accepted**: 2017-07-21; **Published online**: 2017-09-14 09:47

URL: kns.cnki.net/kcms/detail/11.2625.V.20170914.0947.002.html

Foundation item: National Natural Science Foundation of China (51571008)

* **Corresponding author**. E-mail: nhmiao@buaa.edu.cn

北航学报 赠阅

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2017.0230

可吸收胶原膜的体内免疫反应评价

全下文载

张林，孙磊，徐梦洽，牛旭锋*
（北京航空航天大学 生物与医学工程学院，北京 100083）

摘 要：免疫应答反应一直是限制可植入生物材料应用的关键因素之一。实验评估了2种用于骨修复的胶原膜在体内的免疫反应，希望为其临床试验提供依据。在将2种膜皮下植入到BALB/c小鼠后的第14天，与阴性对照（NC，未植入材料）相比，脾和淋巴结没有明显肿大，淋巴结细胞数几乎没差异，而脾细胞数大约是NC的2倍。流式细胞术分析显示植入胶原膜1导致脾中T细胞比例减少了约13%，但是没有影响T细胞亚群，而植入胶原膜2对小鼠的脾细胞组成没有明显影响；2种胶原膜都激活了一定的B细胞，激活率大约为NC小鼠的2倍。淋巴细胞体外增殖实验显示与NC组无显著差异。酶联免疫吸附测试表明胶原膜1导致了第14天血清中的抗体浓度升高至NC小鼠的2倍。局部H&E染色显示2种材料都引起了轻微的细胞浸润。这2种胶原膜引起的免疫反应很微弱，可以被应用于临床试验。

关 键 词：胶原膜；骨修复；免疫应答；流式细胞术；炎症
中图分类号：R318.06
文献标识码：A **文章编号**：1001-5965(2018)04-0879-08

引导组织再生术（Guided Tissue Regeneration，GTR）常用于牙周修复，主要是在伤口处使用一种闭塞膜形成一种物理屏障，阻止牙结缔组织向牙根面生长，为牙周膜、牙骨质和牙槽骨的修复提供空间和时间。GTR已经被提出作为根管治疗的辅助手段来促进骨再生[1]，并且使用GTR以促进牙周再生已经取得了满意的结果[2]，尤其在较严重的根尖周病变和贯穿病变的情况下。目前的研究发现，使用可吸收的屏障材料比不可吸收的材料，或者不使用屏障材料的效果要好[1]。理想的屏障材料需满足以下几个条件：生物相容的，能够作为屏障防止其他细胞类型向根面生长，允许营养物质和空气通过，允许组织整合进材料却又不影响其防止内皮细胞向下生长的功能，提供并维持一个靠近根面的空间，具有便于加工的结构等[1]。

胶原已经被用于骨修复，不仅因为它是骨基质的主要成分，具有良好的生物相容性和生物降解性，无细胞毒性，较低的免疫原性，易被加工成为不同的形式[3]。此外，胶原还是一种牙周膜成纤维细胞的趋化因子[4]，能够阻挡内皮细胞的迁移，并且具有止血效应[5]。在胶原作为屏障材料的研究中都发现了牙再生的现象，并且5个月之后胶原全部降解[6]。可见，胶原作为引导膜治疗牙周病具有一定的优势，是一种较理想的材料，所以用于牙修复的胶原膜被广泛研究，但是其能否成为广泛应用的产品还得经过另一项检测，也就是免疫学评价，因为免疫原性是产生免疫毒性的主要原因。

众所周知，大多数胶原都来源于动物，因而提取的胶原中残留动物蛋白、DNA、细胞是难以避免的。有研究表明，胶原的结构、残留的非胶

收稿日期：2017-04-13；**录用日期**：2017-05-19；**网络出版时间**：2017-06-16 16:10
网络出版地址：kns.cnki.net/kcms/detail/11.2625.V.20170616.1610.002.html
基金项目：国家自然科学基金（31470915）；中央高校基本科研业务费专项资金（YWF-17-BJ-Y-45）
*通信作者. E-mail：nxf@buaa.edu.cn

引用格式：张林，孙磊，徐梦洽，等. 可吸收胶原膜的体内免疫反应评价[J]. 北京航空航天大学学报，2018，44(4)：879-886.
ZHANG L，SUN L，XU M H，et al. Immunological response evaluation of absorbable collagen membrane in vivo[J]. Journal of Beijing University of Aeronautics and Astronautics，2018，44(4)：879-886 (in Chinese).

北航学报 赠阅

原蛋白和细胞、胶原的交联方式等都会对胶原的免疫原性产生影响[7]。因此,为了病人的安全和产品的广泛使用,进行免疫原性评价是十分必要的。

以往也有类似的研究,但是都不够深入全面,大多都只是停留在炎症的研究,而组织学检验和抗体检测是使用最广泛的评价方法[8-10]。在本实验中对猪来源的 2 种骨导膜进行了免疫学评价,希望为其广泛的应用于临床奠定基础。本实验制定了一套较完整的免疫原性检测方案,通过将其植入到 BALB/c 小鼠皮下,定期取样检测来分析胶原膜引起的免疫应答反应,检测指标涵盖了细胞免疫和体液免疫,局部免疫和系统免疫,免疫器官、免疫细胞和免疫分子,对胶原膜的临床应用具有重要意义。

1 实 验

1.1 动物与材料

动物实验已经获得了北京大学生物医学伦理委员会的批准,并严格按照其规定来饲养动物。本实验共需要 20 只 6 ~ 8 周,体重 20 g 左右的雌性 BALB/c 小鼠。实验共分为 4 组,每组 5 只,包括 2 个实验组(S1 和 S2,分别植入胶原膜 1 和胶原膜 2),阳性对照组(PC,植入猪皮),阴性对照组(NC,只做手术,不植入)。材料由深圳兰度生物材料有限公司提供。

1.2 皮下植入

以 50 mg/kg 的比例腹腔注射 w/v 分数 0.5% 的戊巴比妥钠生理盐水溶液麻醉小鼠,待小鼠昏迷之后,选取一块背部脊柱两侧的皮肤,脱毛并用体积分数 75% 酒精消毒,用无菌手术剪剪开一个约 1 cm 的伤口并撑开皮肤,放入材料,缝合伤口(NC 组不植入直接缝合)。总共在每只老鼠背部 3 个不同位置皮下植入各 1 次,一周一次。

1.3 取 样

1.3.1 取 血

分别在 3 次植入后第 3 天和第 7 天,分别剪去约 1 cm 长度的尾尖取血,用 EP 管收集,待血液分层之后,10 000 r/min 离心 5 min,收集上清液于另一干净的 EP 管,标记,置于 -20℃ 冰箱供检测抗体。到第 14 天时,直接摘除一只眼球取血,后面的处理同上。

1.3.2 脏器处理

第 14 天取血之后,拉颈处死小鼠,无菌环境下摘除脾和腋下淋巴结,浸泡于无菌磷酸盐缓冲液(Phosphate Buffer Solution, PBS)中。将脾脏研磨成单细胞悬液,800 r/min 离心 5 min,去上清,加入 1 mL 红细胞裂解液,室温 1 min,再加 PBS 至 5 mL以终止反应,离心去上清,加入一定量 PBS 悬浮细胞并混匀。取出少量稀释到适当倍数,进行细胞计数。淋巴结也经过类似的处理,只是不需要加红细胞裂解液,直接稀释计数。

1.3.3 局部组织处理

剪下最后一次植入部位的皮肤和材料,浸泡于 w/v 分数 4% 多聚甲醛中固定,供组织学分析。

1.4 细胞表型和流式细胞术

用于脾细胞表型识别和筛选的 5 种荧光单克隆抗体分别是 anti-CD19 PE-Cy7、anti-TCR APC、anti-CD4 FITC、anti-CD8 PerCP、anti-Ki67 PE。将单细胞悬液加入 2 mL 尖底离心管中,1 500 r/min 离心 5 min,弃上清液。加冷 PBS 1 mL,离心洗涤,弃上清液。加入用 PBS 稀释的荧光素标记的抗体 200 μL。用微量移液器轻轻吹打混匀,4℃ 或置冰上孵育 0.5 ~ 1 h。离心弃上清液,加入冷 PBS 1 mL,离心洗涤 2 次,以除去未结合的多余抗体成分。向细胞中加入冷 PBS 500 μL,吹打混匀,置流式管中,4℃ 避光保存,待测[11-12]。

1.5 细胞增殖

按每孔 2×10^5 的细胞数量取得细胞,离心去上清,加入 1640 培养基(含体积分数 10% 胎牛血清,体积分数 1% 双抗)200 μL,并按比例加入佛波酯(Phorbol-12-Myristate-13-Acetate, PMA)和离子霉素(Ionomycin, Iono)使之终浓度为 100 ng/mL 和 1 000 ng/mL,混匀后加入 96 孔板,放入二氧化碳培养箱培养。48 h 后用 MTS 试剂盒检测细胞增殖情况,具体来说,取出孔板,小心吸出上层培养液,每孔加 20 μL MTS 和 80 μL 1640 培养液混合液,培养箱孵育 1 h,用酶标仪测 490 nm 处的吸光度。

1.6 酶联免疫吸附测试

用包被液以 1∶1 000 稀释包被抗体,每个孔 100 μL,4℃ 过夜孵育。倒去包被液,每孔加入 200 μL PBST(体积分数 0.05% 吐温-20 的 PBS 溶液)洗涤 1 min,洗 3 次,用纱布包住拍干水分。加入 200 μL PBSB(w/v 分数 1% 牛血清白蛋白的 PBS 溶液)封闭液,室温放置 2 h 后,倾去液体,拍干水分。加入稀释好的样品(IgG 1∶100 000,IgM 1∶10 000)以及标准品(IgG 1∶25 000,IgM 1∶5 000)100 μL,并设置空白对照,室温2 h。洗涤6 次,拍干水分。辣根过氧化物酶标记的一抗用 PBSB 稀释 1 000 倍,每孔加 100 μL,室温2 h,洗涤 6 次,拍干。加 100 μL 3,3′,5,5′-四甲基联苯

北航学报 赠阅

胺，黑暗孵育 10 min（IgM）/20 min（IgG）后，加 50 μL 终止液（2 mol/L 硫酸溶液），450 nm 测吸光度[13]。

1.7　H&E

如文献[14]所述，皮肤组织经过 w/v 分数 4% 多聚甲醛固定之后，经过水洗，酒精梯度脱水，透明，石蜡包埋。石蜡块被切成 3 ~ 5 μm 的薄片，置于载玻片上，脱蜡之后再进行 H&E 染色，按标准步骤进行。

1.8　数据处理

所有数据都是由至少 3 个平行样本得出的，数据显示为 mean ± SD（均值 ± 标准差），作图是采用 GraphPad Prism 软件，数据分析采用单因素方差分析，当 $^{*}p<0.05$，$^{**}p<0.01$，$^{***}p<0.005$ 时，差异被视为具有统计学意义，文中所有比较都是与 NC 组比。

2　结果与讨论

2.1　免疫器官及细胞计数

脾和淋巴结是淋巴细胞的主要栖息地，也是免疫反应发生的主要场所，因此观察脾和淋巴结的大小以及其中的免疫细胞数目能从一定程度上反映免疫反应的强弱。图 1 所示为 3 次免疫后第 14 天时小鼠的脾和淋巴结的形态，可以看到 S1 和 S2 组的脾和淋巴结与正常的脾和淋巴结对比，都没有表现出明显的大小变化，而 PC 组的脾出现了剧烈的膨大，比 NC 组的 2 倍还大，淋巴结也明显较大。脾细胞计数结果见图 2（a），2 个实验组的脾细胞数约为 2×10^{8}，稍高于 NC 组的 1×10^{8}，但差异无统计学意义。而 PC 组细胞数目也正如看到的急剧膨大的脾一样，远远高于 NC 组，约 4.7×10^{8}。淋巴结细胞数目见图 2（b），无论是 PC 组还是 2 个试验组，都与 NC 组的差异无统计意义，但是也能看到 PC 的淋巴结中的细胞数目（约8.5×10^{7}）高于其他组，约是NC组（约2.8×10^{7}）的 3 倍。无论是免疫器官还是其中的免疫细胞数目，结果都基本一致，即 2 种胶原膜对小鼠的免疫刺激都较小。

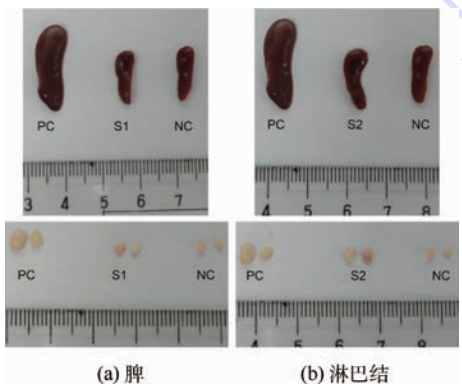


图 1　脾和淋巴结形态

Fig. 1　Images of spleen and lymph nodes

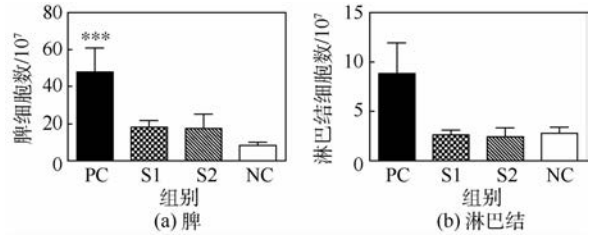


图 2　脾和淋巴结的免疫细胞计数结果（$^{***}p<0.005$）

Fig. 2　Immune cell population of spleen and lymph nodes（$^{***}p<0.005$）

2.2　脾淋巴细胞表型

正常脾淋巴细胞中 T 细胞约占 40%，B 细胞占 60%，CD4/CD8 = 2∶1。CD4 + T 细胞与 CD8 + T 细胞的比值变化能从一定程度上反应机体的免疫功能状态。本实验分析了脾淋巴细胞的表型，每个组有 3 个平行样本，规律类似，代表性流式图如图 3 所示，图中比例表示对应类别细胞占总测试细胞的比例。由图 3 可得到各组脾淋巴细胞中 CD4 + T细胞、CD8 + T 细胞和 B 细胞的含量百分比。对比试验组和对照组结果可以得出，S2 组的脾细胞组成比例正常，与 NC 组无明显差异；S1 组中，脾内 T 细胞比例减少了约 13%，B 细胞比例轻微增多，但整体变化较 PC 组（T 细胞比例与正常组比较下降了约 22%，T、B 细胞比例明显失衡）明显更小。且对比正常 T 细胞亚群 CD4/CD8 比例（约 2.3）可知 S1 组和 S2 组 CD4/CD8 比例（分别为 2.1 和 2.5）正常；而 PC 组 CD4/CD8 比偏低（约 1.5）。由此实验结果可初步得出结论，胶原膜 1 的植入会引起小鼠轻微的免疫应答反应，但未影响其免疫系统的正常机能；而胶原膜 2 对小鼠的影响极小。

2.3　脾淋巴细胞活化

T 细胞的活化通过 2 个亚群 CD4 + 细胞和 CD8 + 细胞反映，B 细胞的活化则通过 CD19 + 细胞反映。本实验对 CD4 + 、CD8 + 和 CD19 + 细胞表面的 Ki67 表达进行了检测，因为 Ki67 是与细胞活化和增殖相关的分子。每个组有 3 个平行样本，规律类似，代表性的流式分析图如图 4 所示。通过图 4 可以看出，各组别的 CD4 + 和 CD8 + T 细胞的活化状态无太大区别，但还是能看出 PC 组和 S1 组的 CD4 + T 细胞活化相对较多，活化比例比 NC 组高出约 5%；S1 组和 S2 组的B细胞活化稍有增加，约为NC组（5.14%）的

北航学报 赠阅

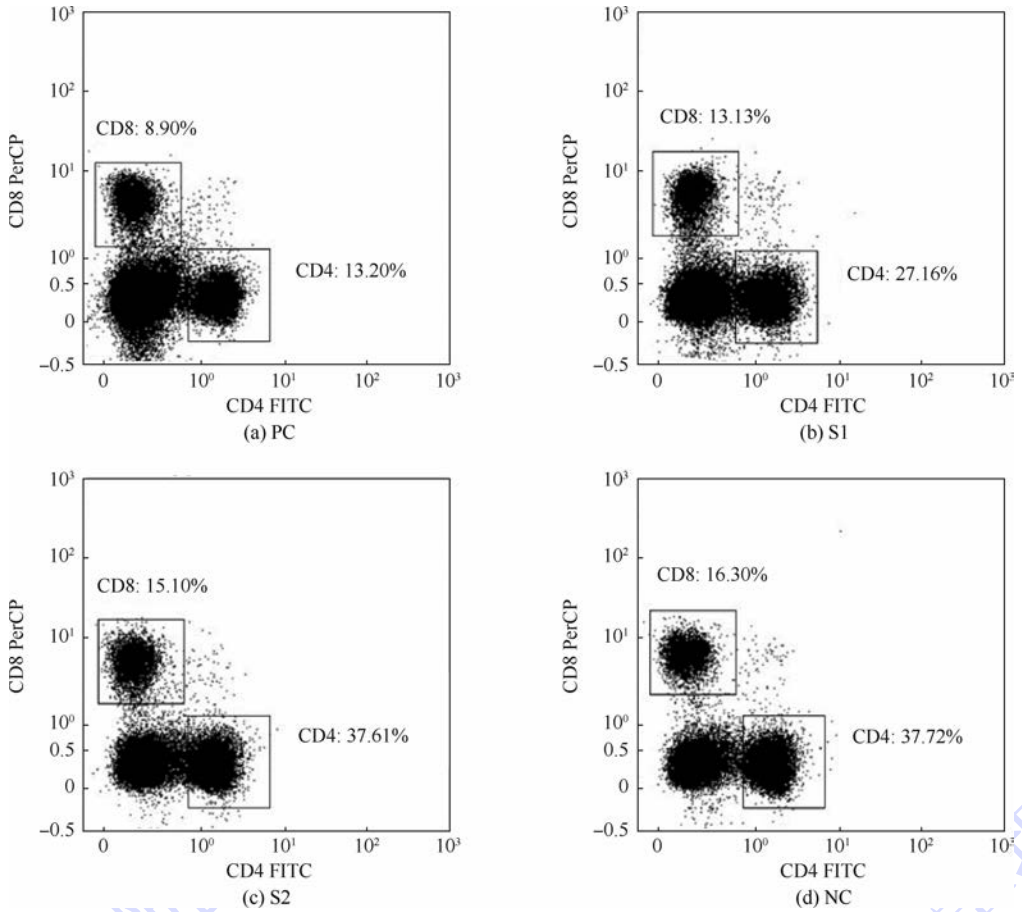


图 3 流式细胞术分析脾淋巴细胞表型

Fig. 3 Splenic cells' phenotype analysis by flow cytometry

2 倍,但 S1(13.29%)多于 S2(9.61%),PC 组 B 细胞明显处于过度活跃状态中,活化率高达 19.64%,大约是 NC 组的 4 倍。这些结果表明植入胶原膜引起了适当的免疫细胞激活,即小鼠对胶原膜作出了轻微免疫反应,对比 2 种胶原膜,胶原膜 1 引起的反应更明显。

2.4 淋巴细胞增殖

水溶性甲臜化合物(MTS)能被细胞线粒体中的脱氢酶还原为具有高度水溶性的有色甲臜产物,生成的甲臜物的数量与活细胞的数量成正比。用酶标仪在 490 nm 波长处测定其光吸收值,可间接反映活细胞数量。淋巴细胞的增殖能力代表了其向淋巴母细胞的转化情况,而淋巴母细胞的转化率可以反映免疫反应水平。本实验采用 PMA 和 Iono 体外刺激细胞增殖来检测细胞的活化情况,结果如图 5 所示,图中"**"表示 PC 组和 NC 组的吸光度差异具有高度的统计学意义。植入胶原膜第 14 天之后的老鼠的脾细胞和淋巴结细胞,均没有像 PC 组细胞那样表现出显著高于正常小鼠细胞的增殖能力,吸光度都在 0.5 左右,与正常小鼠的差异无统计学意义,也就是说明材料没有刺激大量的免疫细胞活化,即免疫反应微弱。

2.5 血清中抗体水平

抗体是由浆细胞分泌的用来鉴别和中和外来抗原的物质,在体液免疫中起着重要的作用。本实验检测了血清中的 IgM 和 IgG,二者分别在初次免疫应答和再次免疫应答中起作用,因此 IgG 产生较晚。酶联免疫吸附测试结果显示各组中 IgM 浓度没有显著差异(本文未给出)。对此分析这可能是因为在 3 次植入免疫中,第 1 次植入时机体免疫主要为初次免疫应答,IgM 水平较高;随着后续增强免疫的进行,机体免疫应答逐渐以再次免疫应答为主,初次免疫应答的水平逐渐减弱,故在3 次免疫后第 3 天检测时机体初次应答水平已经处于较弱阶段,因此 IgM 浓度整体水平较低,故4 组小鼠组间无明显差异。IgG 浓度的检测结果如图 6 所示,S1 组的 IgG 浓度一直处于上升趋势,但曲线斜率明显小于 PC 组,S1 组第 14 天血清中的 IgG 浓度(约 12 mg/mL)约是 NC 组(约 5.5 mg/mL)的 2 倍;而 S2 组的 IgG 浓度在第 7 天出现了轻微地上升,从 4 mg/mL 到 8 mg/mL,第 14 天时与第7天差别不大。总体来说,2个实验

北航学报 赠阅

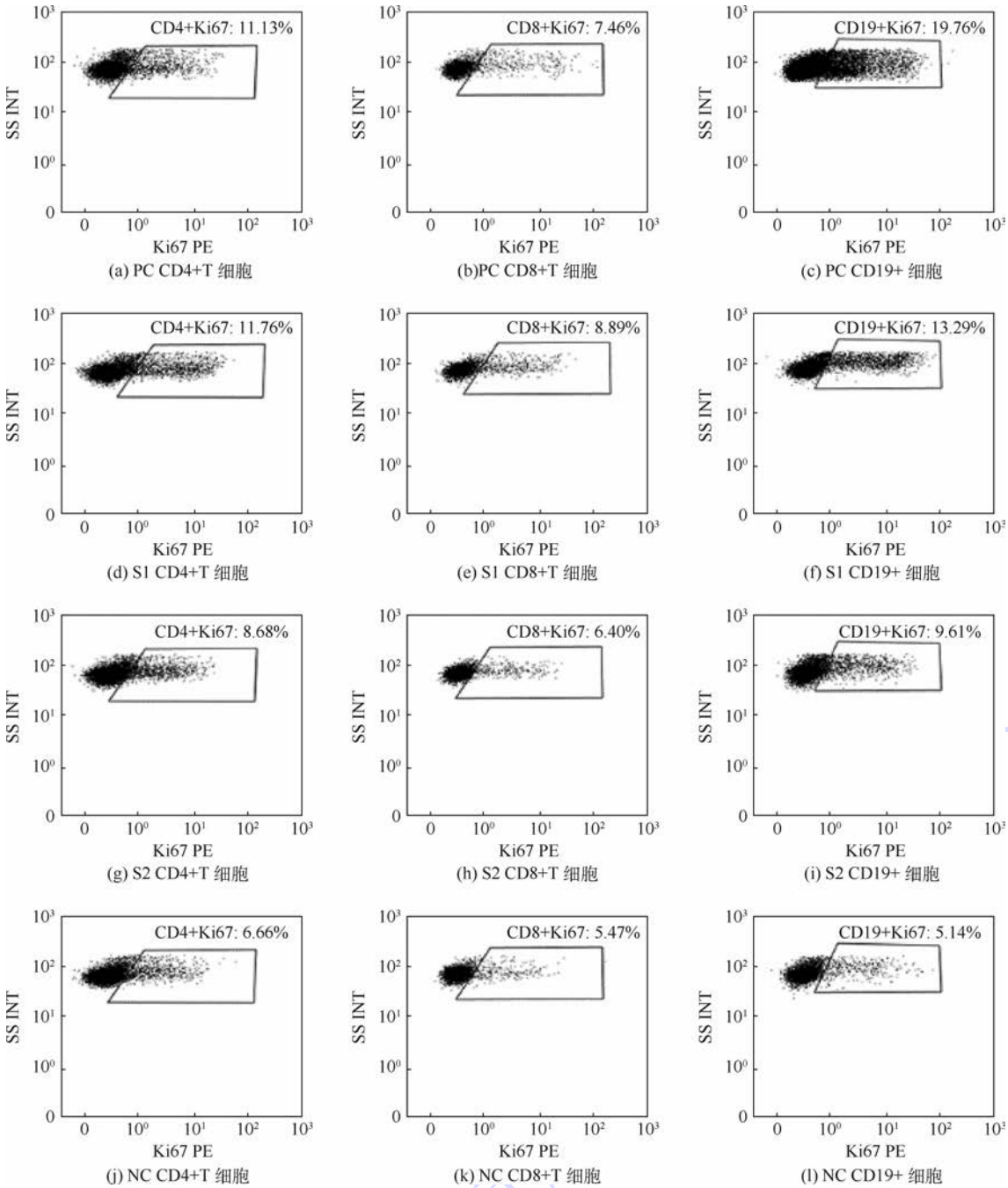


图 4　CD4 + T 细胞、CD8 + T 细胞和 CD19 + 细胞的 Ki67 表达情况

Fig. 4　Ki67 expression of CD4 + T lymphocytes, CD8 + T lymphocytes and CD19 + lymphocytes

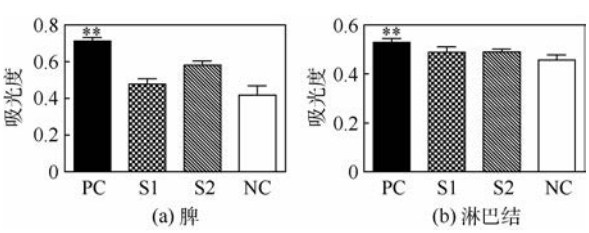


图 5　脾和淋巴结细胞体外增殖情况（$^{**}p<0.01$）

Fig. 5　Proliferation in vitro of splenic cells and lymph nodes cells ($^{**}p<0.01$)

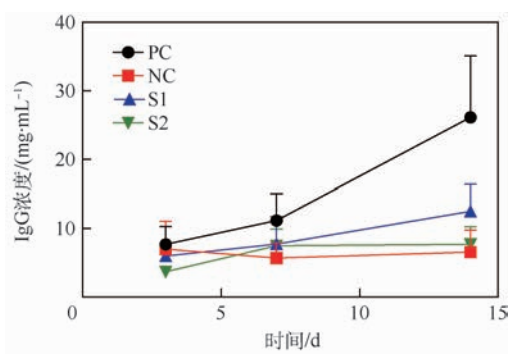


图 6　血清中 IgG 浓度随时间的变化

Fig. 6　Variation of concentration of IgG in serum with time

北航学报 赠阅

组与 NC 组的 IgG 浓度变化差异不大，尤其是 S2 组。Liu 等[15]也有类似的发现，将猪源Ⅰ型胶原膜作为角膜原位移植到 BALB/c 小鼠后，在第 14 天的血清中也检测到了 IgG，并且浓度随时间升高。而 Patino 等[16]的研究表明在 Wistar 大鼠的血清中没有检测到针对猪源胶原膜的抗体，这种差异可能是由于动物的种类不同而导致的。由此可见，本实验中的猪来源的胶原膜 1 有可能对小鼠产生了一定的刺激作用，以至于小鼠体内产生了适当的 IgG 抗体来中和这种外来抗原，而胶原膜 2 对小鼠的刺激是不明显的。

2.6 细胞浸润

以上的实验都是针对系统性的免疫反应，细胞浸润测试则是用来评价局部免疫反应。局部炎症反应的一个特征是炎性细胞浸润，因此在本实验中笔者采用 H&E 染色来检查植入局部的炎症发生情况。代表性的 H&E 染色如图 7 所示，M 为植入的材料，右图是左图的局部放大。植入之后第14 天，在 PC 组的组织中，大量的细胞聚集在整个材料之中，也就是说炎症反应剧烈；与未植入的皮肤相比，在胶原膜1 和胶原膜2 的周围有稍多的细胞浸润，并且胶原膜 1 周围的细胞数量多于胶原膜 2，但都只是围绕在材料周围。这种现象与以往的研究结果一致，Liu 等[15]的研究在植入后第 12 天也发现细胞浸润现象只发生在胶原材料周围，而没有侵入其中。猜测可能是由于胶原本身的免疫原性比较低，以至于浸润细胞少且细胞浸润的速度慢。由此可见这 2 种材料会引起轻微的局部炎症反应，并且胶原膜 1 引起的炎症稍强于胶原膜 2。

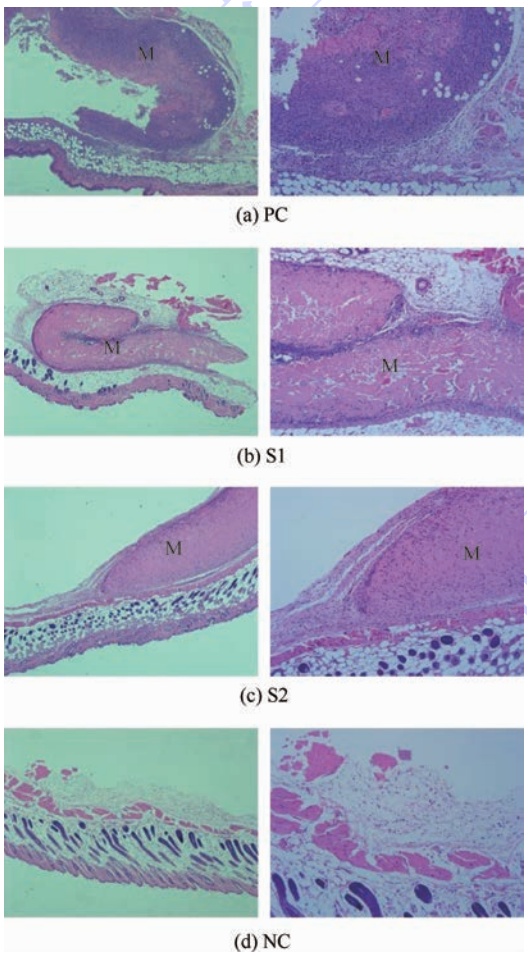


图 7 植入部位 H&E 染色

Fig. 7 H&E staining of implantation sites

本文实验中的 2 种胶原膜唯一的差别在于胶原膜 1 是经过了过氧化氢灭菌处理，而胶原膜 2 没有。过氧化氢是一种强氧化剂，其灭活微生物的原理是通过解离高活性的羟基，攻击破坏微生物的细胞壁、DNA、脂类以及蛋白质等，对包膜病毒作用尤为明显，生成产物只有水和氧气，广泛应用于生物医药领域[17-18]。本实验中几乎所有的结果都表明胶原膜 1 在 BALB/c 小鼠中引起的免疫反应明显强于胶原膜 2，推测主要原因是过氧化氢破坏了胶原的结构，使之抗原决定簇更多地暴露，从而增强了其免疫原性。有研究结果表明[19]，当过氧化氢浓度低于 2%（体积分数）时，主要以氧化基团与胶原的交联为主，而当过氧化氢浓度达到 3% 时，氧化能力变强，以至于破坏了胶原的结构，此时的作用以降解为主。

3 结 论

1）本文提出了一套免疫原性检测方案，可以用于检测所有可植入生物材料的免疫原性，能为生物医用材料的生物安全性检测提供参考。

2）胶原膜 1 在 BALB/c 小鼠中引起了轻微的系统性免疫反应，包括脾细胞数目和活化以及 IgG 抗体分泌增多，但是与未植入材料的老鼠无具有统计学意义的差异。另外，胶原膜 1 在植入局部引起了一定的细胞浸润，即炎症。

3）胶原膜 2 在 BALB/c 小鼠中也引起了轻微的系统性免疫反应，主要是脾细胞数目增多，和 B 细胞活化增多，但是程度比胶原膜 1 要轻，还发生了反应程度比胶原膜 1 稍轻的局部炎症。

4）2 种胶原膜在 BALB/c 小鼠中引起的免疫反应较弱，比 NC 组稍强，远远不及 PC 组，因此，这 2 种胶原膜，尤其是胶原膜 2，可被认为是安全可靠的骨修复材料，能被应用于临床试验。

致谢 感谢北京大学医学部王巍老师在实验中给予的指导与帮助，感谢谷歌生物科技有限公司给予技术支持。

参考文献（References）

[1] TSESIS I, ROSEN E, TAMSE A, et al. Effect of guided tissue regeneration on the outcome of surgical endodontic treatment: A systematic review and meta-analysis[J]. Journal of Endodontics, 2011, 37(8): 1039-1045.

[2] LEKOVIC V, CAMARGO P M, WEINLAENDER M, et al. Effectiveness of a combination of platelet-rich plasma, bovine porous bone mineral and guided tissue regeneration in the treatment of mandibular grade Ⅱ molar furcations in humans[J]. Journal of Clinical Periodontology, 2003, 30(8): 746-751.

[3] SUMITA Y, HONDA M J, OHARA T, et al. Performance of collagen sponge as a 3-D scaffold for tooth-tissue engineering[J]. Biomaterials, 2006, 27(17): 3238-3248.

[4] LU H K, LEE S Y, LIN F P. Elastic modulus, permeation time and swelling ratio of a new porcine dermal collagen membrane[J]. Journal of Periodontal Research, 1998, 33(5): 243-248.

[5] PITARU S, TAL H, SOLDINGER M, et al. Collagen membranes prevent apical migration of epithelium and support new connective tissue attachment during periodontal wound healing in dogs[J]. Journal of Periodontal Research, 1989, 24(4): 247-253.

[6] SCULEAN A, NIKOLIDAKIS D, NIKOU G, et al. Biomaterials for promoting periodontal regeneration in human intrabony defects: A systematic review[J]. Periodontol 2000, 2015, 68(1): 182-216.

[7] LYNN A K, YANNAS I V, BONFIELD W. Antigenicity and immunogenicity of collagen[J]. Journal of Biomedical Materials Research Part B: Applied Biomaterials, 2004, 71(2): 343-354.

[8] PENG Y Y, GLATTAUER V, RAMSHAW J A, et al. Evaluation of the immunogenicity and cell compatibility of avian collagen for biomedical applications[J]. Journal of Biomedical Materials Research Part A, 2010, 93(4): 1235-1244.

[9] ZIV O, AVTALION R R, MARGEL S. Immunogenicity of bioactive magnetic nanoparticles: Natural and acquired antibodies[J]. Journal of Biomedical Materials Research Part A, 2008, 85(4): 1011-1021.

[10] DILGIMEN A S, MUSTAFAEVA Z, DEMCHENKO M, et al. Water-soluble covalent conjugates of bovine serum albumin with anionic poly(n-isopropyl-acrylamide) and their immunogenicity[J]. Biomaterials, 2001, 22(17): 2383-2392.

[11] KREISER S, ECKHARDT J, KUHNT C, et al. Murine cd83-positive T cells mediate suppressor functions in vitro and in vivo[J]. Immunobiology, 2015, 220(2): 270-279.

[12] O'DONNELL H, PHAM O H, LI L X, et al. Toll-like receptor and inflammasome signals converge to amplify the innate bactericidal capacity of t helper 1 cells[J]. Immunity, 2014, 40(2): 213-224.

[13] WELCH R J, LITWIN C M. A comparison of brucella igg and igm elisa assays with agglutination methodology[J]. Journal of Clinical Laboratory Analysis, 2010, 24(3): 160-162.

[14] LIU H, WISE S G, RNJAK-KOVACINA J, et al. Biocompatibility of silk-tropoelastin protein polymers[J]. Biomaterials, 2014, 35(19): 5138-5147.

[15] LIU L, KUFFOVA L, GRIFFITH M, et al. Immunological responses in mice to full-thickness corneal grafts engineered from porcine collagen[J]. Biomaterials, 2007, 28(26): 3807-3814.

[16] PATINO M G, NEIDERS M E, ANDREANA S, et al. Cellular inflammatory response to porcine collagen membranes[J]. Journal of Periodontal Research, 2003, 38(5): 458-464.

[17] 方哲翔,王建华,袁平,等. 医用胶原修复膜病毒灭活/去除工艺的验证和评价[J]. 中国生物制品学杂志, 2016, 29(12): 1341-1345.
FANG Z X, WANG J H, YUAN P, et al. Validation and evaluation of inactivation/removal of medical collagen repair membrane[J]. Chinese Journal of Biologicals, 2016, 29(12): 1341-1345(in Chinese).

[18] 王朋. 过氧化氢消毒灭菌技术介绍[J]. 化工与医药工程, 2012, 33(6): 51-53.
WANG P. Introduction of sterilization technology with vaporized hydroge peroxide[J]. Chemical and Pharmaceutical Engineering, 2012, 33(6): 51-53(in Chinese).

[19] 龚居霞,刘新华,但卫华,等. 过氧化氢处理对Ⅰ型胶原结构和性能的影响[J]. 中国皮革, 2015(1): 10-13.
GONG J X, LIU X H, DAN W H, et al. Effects of hydrogen peroxide treatment on structure and performance of type Ⅰ collagen[J]. China Leather, 2015(1): 10-13(in Chinese).

作者简介:

张林　女,硕士研究生。主要研究方向:生物医用材料。

牛旭锋　男,博士,副教授,博士生导师。主要研究方向:生物医用材料。

北京航空航天大学学报

北航学报 赠阅

Immunological response evaluation of absorbable collagen membrane in vivo

ZHANG Lin, SUN Lei, XU Menghan, NIU Xufeng*

(School of Biological Science and Medical Engineering, Beijing University of Aeronautics and Astronautics, Beijing 100083, China)

Abstract: Immune response is one of the critical factors to limit implantable biomaterials' application. In this study, two kinds of collagen membranes for bone regeneration were evaluated in terms of immunological response in vivo, which is hoped to lay foundations for clinical trials. At day 14 following subcutaneous implantation of the two membranes in BALB/c mice, compared with negative control (NC, no implant), spleen and lymph nodes showed no obvious swelling, and normal lymph nodes' cell population as well as one-fold more splenic cell population was observed. Flow cytometry analysis demonstrated that in mice with collagen membrane 1, splenic T lymphocytes percentage decreased by about 13%, but no significant change was discovered in T lymphocyte subsets; collagen membrane 2 did not trigger apparent splenic cells composition alteration; the two membranes activated one-fold more B lymphocytes than NC. Enzyme linked immunosorbent assay indicated that one-fold higher IgG concentration than NC was detected in mice with collagen membrane 1 in day 14. Lymphocytes proliferation assay in vitro did not show significant difference from NC. H&E staining of local tissues exhibited slight cell infiltration around the two membranes. The two collagen membranes have engendered mild immune response and can be applied to clinical trials.

Keywords: collagen membrane; bone regeneration; immune response; flow cytometry; inflammation

北京航空航天大学学报

Received: 2017-04-13; **Accepted**: 2017-05-19; **Published online**: 2017-06-16 16:10

URL: kns.cnki.net/kcms/detail/11.2625.V.20170616.1610.002.html

Foundation items: National Natural Science Foundation of China (31470915); the Fundamental Research Funds for the Central Universities (YWF-17-BJ-Y-45)

* **Corresponding author**. E-mail: nxf@buaa.edu.cn

北航学报 赠阅

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2017.0367

基于竞争博弈的多目标可靠性优化设计方法

全下文载

冯嘉珍[1,2], 张建国[1,2,*], 邱继伟[1,2]

(1. 北京航空航天大学 可靠性与系统工程学院, 北京 100083;
2. 北京航空航天大学 可靠性与环境工程技术重点实验室, 北京 100083)

摘 要: 针对目标权重选取的主观性问题,提出基于竞争博弈进行多目标可靠性优化设计的方法。首先,将各设计目标视为不同的博弈方,通过随机设计变量集映射(RDVSM)技术,将优化模型中的随机设计变量集分解为各博弈方所拥有的策略集;然后,各博弈方以自身收益为目标,结合性能测量方法在各自的策略集中进行单目标可靠性优化设计,并由所有的优化设计结果形成一轮博弈的策略组合;经过多轮博弈之后得博弈均衡解。压力容器和齿轮减速器2个案例的分析表明,所提方法有效避免了目标权重的选择,设计结果具有较高客观性。

关 键 词: 多目标; 可靠性优化设计; 竞争博弈; 随机设计变量集映射 (RDVSM); 博弈均衡解

中图分类号: TB114.3

文献标识码: A **文章编号**: 1001-5965(2018)04-0887-08

在实际工程问题当中,广泛存在着不确定性因素,例如,材料特性、载荷、结构参数、边界条件以及测量误差等的不确定性[1-2],使得设计会因为这些不确定性而发生改变,影响产品性能。对产品的优化设计已不再是单纯的追求性能最佳或费用最小,而是需要在性能、可靠性以及经济性等多种设计要求之间取得平衡。所以,产品的优化设计是一个多目标可靠性优化设计(Multi-objective Reliability Design Optimization, MRDO)。

通常情况下,多个设计目标之间是相互冲突的,解决MRDO问题需要在各目标之间进行权衡和折中,使各目标尽可能达到最优[3]。当前,MRDO的传统方法是按照重要程度赋予各目标权重,转化为单目标可靠性优化设计(Single-objective Reliability Design Optimization, SRDO)问题之后,再结合一次二阶矩(First Order Second Moment, FOSM)法等可靠性分析方法进行求解。Kogiso等利用加权法对汽车车身进行了基于可靠性的多目标优化设计[4];张瑞军等基于灰色理论确定各目标的权重,再采用加权法将机械产品的可靠性稳健优化设计问题转换为单目标优化设计问题[5];王若冰等将分层序列法应用于某型航空飞行器气动、质量与性能的三目标可靠性优化设计[6];于淼等采用物理规划法构造各目标的偏好函数,在此基础上通过非线性加权将MRDO问题转化为SRDO问题,并应用于湿式多盘制动器的设计[7]。上述研究中,各目标的权重系数或偏好函数的选取体现了设计者的主观意识,人为经验性较大,优化结果客观性较低。针对这一问题,后续研究产生了新的MRDO方法,例如,基于多目标粒子群优化(Multi-Objective Particle Swarm Optimization, MOPSO)算法或多目标遗传算法(Multi-objective Genetic Algorithm, MOGA)进行寻优,并利用FOSM等方法判断寻优点是否满足可靠性约束的要求。在岩土工程设计领域,Juang和Wang利用MOGA开展了扩展基础的多目标可靠

收稿日期: 2017-05-31; **录用日期**: 2017-08-01; **网络出版时间**: 2017-09-22 15:07
网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20170922.1507.009.html
基金项目: 国家重点研发计划(2013CB733000); 国家自然科学基金(51675026,71671009)
* **通信作者**. E-mail: zjg@buaa.edu.cn

引用格式: *冯嘉珍, 张建国, 邱继伟. 基于竞争博弈的多目标可靠性优化设计方法[J]. 北京航空航天大学学报, 2018, 44(4): 887-894. FENG J Z, ZHANG J G, QIU J W. Multi-objective reliability design optimization approach based on competition game[J]. Journal of Beijing University of Aeronautics and Astronautics, 2018, 44(4): 887-894 (in Chinese).*

北航学报 赠阅

性优化设计[8]；张干清等则采用 MOPSO 算法对盾构机行星减速器进行了多目标可靠性优化设计[9]。但是，MOPSO 等算法在处理 MRDO 问题时，需要求解大量 Pareto 非劣解，计算量巨大，收敛速度缓慢，这一问题随着设计目标与随机设计变量的增多会更加突出[10]。鉴于多目标优化设计与博弈之间的相似性，近年来，国内外不断有研究将博弈论应用于求解工程中的多目标优化设计问题。Song 等将竞争博弈模型应用于汽车悬架参数的多目标优化设计[11]；Desideri 利用竞争博弈模型对某型机翼开展了外形与空气动力学的优化设计[12]；Xiao 等利用竞争博弈模型对小水线面双体船进行了多目标优化设计[13]。基于博弈的方法无需人为设置目标权重，相较于 MOPSO 等算法显著降低了计算量（无需求解 Pareto 非劣解集），在求解多目标优化设计问题时有着独特优势。当前，该方法多针对确定性的多目标优化设计问题，将其应用于 MRDO 鲜有研究成果发表。

为解决权重法所带来的优化结果客观性不足的问题，本文提出基于竞争博弈求解 MRDO 问题的方法。首先，将各设计目标视为不同的博弈方；其次，通过随机设计变量集映射（Random Design Variables Set Mapping，RDVSM）技术，将随机设计变量集分解为各博弈方所拥有的策略集；最后，各博弈方以自身收益最佳为目标，结合可靠性分析的性能测量方法（Performance Measurement Approach，PMA）在各自策略集中进行 SRDO，由所有优化结果形成一轮博弈的策略组合，经过多轮博弈获得均衡解（即 MRDO 问题的优化解）。通过 2 个案例验证方法的有效性。

1 基础理论

1.1 可靠性分析的性能测量方法

考虑设计变量 $\boldsymbol{X}=\{x_1,x_2,\cdots,x_n\}$ 的随机不确定性，MRDO 问题的数学表达式可以描述为

$$\begin{cases}\min\{\mu_{f_1}(\boldsymbol{X}),\mu_{f_2}(\boldsymbol{X}),\cdots,\mu_{f_m}(\boldsymbol{X})\}\\ \text{s.t. } \Pr\{g_j(\boldsymbol{X})\geqslant 0\}\geqslant R_{jt}\quad j=1,2,\cdots,q\\ \boldsymbol{\mu}_{\boldsymbol{X}}^{\mathrm{L}}\leqslant\boldsymbol{\mu}_{\boldsymbol{X}}\leqslant\boldsymbol{\mu}_{\boldsymbol{X}}^{\mathrm{U}}\end{cases}\tag{1}$$

式中：m 为目标函数的个数；q 为可靠性约束个数；$\mu_{f_i}(\cdot)$ 为第 i 个目标函数 $f_i(\cdot)$ 的期望；$g_j(\cdot)$ 为第 j 个功能函数；R_{jt} 为 $g_j(\cdot)$ 对应的可靠度目标值；$\boldsymbol{\mu}_{\boldsymbol{X}}^{\mathrm{L}}$ 和 $\boldsymbol{\mu}_{\boldsymbol{X}}^{\mathrm{U}}$ 分别为随机变量 $\boldsymbol{X}$ 均值 $\boldsymbol{\mu}_{\boldsymbol{X}}$ 的下限和上限。

式(1)中可靠性约束的分析评估本身是一个优化迭代过程。因此，MRDO 问题最直接的求解方法可以表示为一个双层优化模型，外层为优化循环，里层为可靠性分析循环。高效、稳健的可靠性分析方法是求解 MRDO 问题的重要环节。针对这一要求，采用 PMA 进行可靠性分析[14]。该方法表述为功能函数的功能测量应不小于 0，否则可靠性约束不满足。功能测量是指在规定的可靠度目标值下的最小功能函数值，其求解可以表示为一个优化问题：

$$\begin{cases}\min\quad g_j(\boldsymbol{u})\\ \text{s.t. } \|\boldsymbol{u}\|=\Phi^{-1}(R_{jt})\end{cases}\tag{2}$$

式中：$\boldsymbol{u}$ 为标准正态分布随机变量；$\|\boldsymbol{u}\|$ 为变量 $\boldsymbol{u}$ 的模；$\Phi^{-1}(\cdot)$ 为标准正态分布逆函数。式(2)的最优解 $\boldsymbol{u}_{R_{jt}}^{\mathrm{IMPP}}$ 称为逆最大可能点（Inverse Most Probable Point，IMPP），由此可以得到 $g_j(\cdot)$ 的功能度量 $g_j(\boldsymbol{u}_{R_{jt}}^{\mathrm{IMPP}})$。求解 IMPP 的迭代步骤如下：

步骤 1 选择 **0** 点作为初始迭代点，求解 $g_j(\cdot)$ 的梯度 $\nabla g_j(\boldsymbol{0})$。

步骤 2 求解第 1 次迭代计算之后得到的点 $\boldsymbol{u}^{(1)}=-(\nabla g_j(\boldsymbol{0})/\|\nabla g_j(\boldsymbol{0})\|)\beta_{jt}$，$\beta_{jt}$ 为对应于 R_{jt} 的可靠度指标目标值。

步骤 3 当 $k\geqslant 1$ 时，求解第 $k+1$ 个迭代点 $\boldsymbol{u}^{(k+1)}=-(\nabla g_j(\boldsymbol{u}^{(k)})/\|\nabla g_j(\boldsymbol{u}^{(k)})\|)\beta_{jt}$。

步骤 4 当不等式 $|\|\boldsymbol{u}^{(k+1)}\|-\|\boldsymbol{u}^{(k)}\||\leqslant\delta$ 成立时（δ 为收敛判据），可得到 IMPP 的值 $\boldsymbol{u}_{R_{jt}}^{\mathrm{IMPP}}=\boldsymbol{u}^{(k+1)}$。

1.2 博弈理论

博弈论是一种应用于存在利益冲突场合的数学工具，为分析各博弈方彼此间决策会相互影响的问题提供了有效的解决方案[13]。用 G 代表一个博弈，如果 G 有 m 个博弈方 $p_1,p_2,\cdots,p_m$，每个博弈方可选择的全部策略的集合称为策略集，分别用 $\boldsymbol{S}_1,\boldsymbol{S}_2,\cdots,\boldsymbol{S}_m$ 表示，每个博弈方做出策略决策之后的收益分别用 $u_1,u_2,\cdots,u_m$ 表示，则博弈 G 可以表示为

$$G=\{p_1,p_2,\cdots,p_m;\boldsymbol{S}_1,\boldsymbol{S}_2,\cdots,\boldsymbol{S}_m;u_1,u_2,\cdots,u_m\}\tag{3}$$

博弈分为竞争博弈与合作博弈 2 类[11]。竞争博弈是指各博弈方以竞争的方式追求自身收益最佳为目标；合作博弈是指各博弈方以合作的方式追求整体收益最佳。对于 MRDO 问题，各目标之间相互冲突与竞争，某个目标的改善可能引起其他目标的恶化，与竞争博弈高度类似。将 MRDO 模型中的设计目标对应于博弈方；随机设计变量集 $\boldsymbol{X}$ 对应于所有博弈方所拥有的策略集，并进一步将 $\boldsymbol{X}$ 分解为各博弈方所拥有的策略集 $\boldsymbol{S}_1,\boldsymbol{S}_2,\cdots,\boldsymbol{S}_m$；目标函数的响应对应于博弈方的收益

值；可靠性约束条件用于限制博弈方的策略取值。MRDO 问题可以用博弈模型来描述，其中，$\boldsymbol{S}_1=\{x_i,x_{i+1},\cdots,x_j\}$，$\boldsymbol{S}_2=\{x_k,x_{k+1},\cdots,x_l\}$，$\cdots$，$\boldsymbol{S}_m=\{x_v,x_{v+1},\cdots,x_w\}$，满足 $\boldsymbol{S}_1\cup\boldsymbol{S}_2\cup\cdots\cup\boldsymbol{S}_m=\boldsymbol{X}$ 且 $\boldsymbol{S}_c\cap\boldsymbol{S}_d=\varnothing(c,d=1,2,\cdots,m$ 且 $c\neq d)$；收益函数与目标函数之间的映射关系为 $u_i(\boldsymbol{S}_i)=\mu_{f_i}(\boldsymbol{S}_i)$，$i=1,2,\cdots,m$。

2　MRDO 的竞争博弈方法

基于竞争博弈求解 MRDO 问题，包含 2 项关键技术：各博弈方策略集的分解技术，竞争博弈的求解算法。

2.1　基于 RDVSM 的博弈方策略集分解

通过 RDVSM，随机设计变量集 $\boldsymbol{X}$ 可以分解为各博弈方所拥有的策略集。RDVSM 由影响因子集合构建、基于模糊聚类的策略集分解等 2 个部分组成。

1）影响因子集合构建

构建影响因子集合用于形成分类样本，然后通过模糊聚类方法对样本进行分类处理以便获得各博弈方的策略集。影响因子集合的构建步骤如下：

步骤 1　结合 PMA，对式(1)中的 m 个目标函数分别进行 SRDO，得到的目标函数最优值分别为 $\mu_{f_1}(\boldsymbol{X}_1^*)$，$\mu_{f_2}(\boldsymbol{X}_2^*)$，$\cdots$，$\mu_{f_m}(\boldsymbol{X}_m^*)$，其中：$\boldsymbol{X}_i^*=[x_{i1}^*,x_{i2}^*,\cdots,x_{ij}^*,\cdots,x_{in}^*]$，$i=1,2,\cdots,m$；$j=1,2,\cdots,n$。

步骤 2　将与随机设计变量 x_j 相关的变量 $\boldsymbol{X}_i^j=[x_{i1}^*,x_{i2}^*,\cdots,x_{i(j-1)}^*,x_j,x_{i(j+1)}^*,\cdots,x_{in}^*]$ 代入目标函数 $\mu_{f_i}(\cdot)$ 中，并对 x_j 的均值 μ_{x_j} 求偏导数得到灵敏度函数 $h_i(x_j)=\partial\mu_{f_i}/\partial\mu_{x_j}$。

步骤 3　设 λ_j 为一常数，其由关于 x_j 的所有灵敏度函数所决定，其数学表达式为

$$\lambda_j=0.5(\min|h_i(x_j)|/|\mu_{f_i}(\boldsymbol{X}_i^*)|+\max|h_i(x_j)|/|\mu_{f_i}(\boldsymbol{X}_i^*)|)\quad i=1,2,\cdots,m \tag{4}$$

步骤 4　求解不等式 $|h_i(x_j)|/|\mu_{f_i}(\boldsymbol{X}_i^*)|\geqslant\lambda_j$，得到 μ_{x_j} 占其可行域 $[\mu_{x_j}^{\mathrm{L}},\mu_{x_j}^{\mathrm{U}}]$ 的比例 η_{ji}，η_{ji} 为 x_j 对目标函数 $\mu_{f_i}(\cdot)$ 的影响因子。x_j 对所有目标函数的影响因子集为 $\boldsymbol{\eta}_j=\{\eta_{j1},\eta_{j2},\cdots,\eta_{jm}\}$。

步骤 5　随机设计变量集 $\boldsymbol{X}$ 对所有目标函数的影响因子集为 $\boldsymbol{\eta}=\{\eta_1,\eta_2,\cdots,\eta_n\}$。

2）基于模糊聚类的策略集分解

对影响因子集 $\boldsymbol{\eta}$ 进行聚类，将关联性和相似性较强的样本划分为同一类。由于 x_j 与 $\boldsymbol{\eta}_j$ 之间的一一对应关系，$\boldsymbol{\eta}$ 的聚类结果就代表了 $\boldsymbol{X}$ 的聚类结果。将待分类的对象严格地划分到某个类中，存在不合理之处，为此采用模糊方法处理 $\boldsymbol{\eta}$ 的聚类。

① 构建模糊相似矩阵 $\boldsymbol{R}$。

在模糊聚类之前，首先需要建立 $\boldsymbol{\eta}$ 上的模糊相似关系，可以表示为一个模糊相似矩阵 $\boldsymbol{R}=(\gamma_{ij})_{n\times n}$，其中 γ_{ij} 为任意 2 个影响因子 $\boldsymbol{\eta}_i$ 与 $\boldsymbol{\eta}_j$ 之间的相似度，且 $0\leqslant\gamma_{ij}\leqslant1$，数学表达式为[15]

$$\gamma_{ij}=\begin{cases}1 & i=j\\ 1-c_0\sum\limits_{k=1}^{m}|\eta_{ik}-\eta_{jk}| & i\neq j\end{cases}\quad i,j=1,2,\cdots,n \tag{5}$$

式中：c_0 为常数。

② 构建模糊等价矩阵 $\tilde{\boldsymbol{R}}$。

模糊相似矩阵 $\boldsymbol{R}$ 不满足传递性，无法直接对其进行聚类。采用平方法求解 $\boldsymbol{R}$ 的模糊等价矩阵 $\tilde{\boldsymbol{R}}$[15]：依次计算 $\boldsymbol{R}\rightarrow\boldsymbol{R}^2\rightarrow\boldsymbol{R}^4\rightarrow\cdots$，直至首次出现 $\boldsymbol{R}^k=\boldsymbol{R}^{2k}$，$k=1,2,\cdots$。$\boldsymbol{R}^k$ 即为 $\boldsymbol{R}$ 的模糊等价矩阵 $\tilde{\boldsymbol{R}}$。由于 $\tilde{\boldsymbol{R}}$ 满足传递性，可以对其进行聚类。

③ 模糊聚类。

记置信水平为 θ，且 $\theta\in[0,1]$。选取适当的 θ 值，按照如下方式对 $\tilde{\boldsymbol{R}}$ 进行截取：在矩阵 $\tilde{\boldsymbol{R}}$ 中，将大于或等于 θ 的元素变为 1，小于 θ 的元素变为 0，则 $\tilde{\boldsymbol{R}}$ 变为只含有 2 种元素 0 和 1 的矩阵，观察矩阵的各行元素，把元素相等的行分为同一类，由此得到影响因子集 $\boldsymbol{\eta}$ 的分类结果，亦即随机设计变量集 $\boldsymbol{X}$ 的分类结果。

2.2　竞争博弈的求解算法

在一个竞争博弈模型 G 中，各博弈方的策略组合 $\boldsymbol{s}^*=\{\boldsymbol{s}_1^*,\boldsymbol{s}_2^*,\cdots,\boldsymbol{s}_i^*,\cdots,\boldsymbol{s}_m^*\}$ 称为一个博弈均衡：如果对任意博弈方 i 来说，策略 $\boldsymbol{s}_i^*$ 是在给定其余博弈方的策略组合 $\overline{\boldsymbol{s}_i^*}=\{\boldsymbol{s}_1^*,\boldsymbol{s}_2^*,\cdots,\boldsymbol{s}_{i-1}^*,\boldsymbol{s}_{i+1}^*,\cdots,\boldsymbol{s}_m^*\}$ 情况下该博弈方所做出的最优策略，即 $u_i(\boldsymbol{s}_i^*,\overline{\boldsymbol{s}_i^*})\leqslant u_i(\boldsymbol{s}_i,\overline{\boldsymbol{s}_i^*})$ 对任意 $\boldsymbol{s}_i\in\boldsymbol{S}_i$ 都成立。求解博弈均衡解的算法步骤具体如下：

步骤 1　通过 RDVSM，得到隶属于各博弈方的策略集 $\boldsymbol{S}_1,\boldsymbol{S}_2,\cdots,\boldsymbol{S}_m$。

步骤 2　在所有博弈方的策略集中形成一个初始策略组合 $\boldsymbol{s}^{(0)}=\{\boldsymbol{s}_1^{(0)},\boldsymbol{s}_2^{(0)},\cdots,\boldsymbol{s}_i^{(0)},\cdots,\boldsymbol{s}_m^{(0)}\}$。

步骤 3　假设 $\overline{\boldsymbol{s}_1^{(0)}},\overline{\boldsymbol{s}_2^{(0)}},\cdots,\overline{\boldsymbol{s}_i^{(0)}},\cdots,\overline{\boldsymbol{s}_m^{(0)}}$ 分别为 $\boldsymbol{s}_1^{(0)},\boldsymbol{s}_2^{(0)},\cdots,\boldsymbol{s}_i^{(0)},\cdots,\boldsymbol{s}_m^{(0)}$ 在 $\boldsymbol{s}^{(0)}$ 中的补集，结合 PMA 对任意博弈方 i 求解其最优策略 $\boldsymbol{s}_i^*$，使得自身收益最佳化：

$$\begin{cases}\min & u_i(\boldsymbol{s}_i^*,\overline{\boldsymbol{s}_i^{(0)}})\\ \text{s.t.} & \Pr\{g_j(\boldsymbol{s}_i^*,\overline{\boldsymbol{s}_i^{(0)}})\geqslant0\}\geqslant R_j\\ & j=1,2,\cdots,q\end{cases} \tag{6}$$

北京航空航天大学学报

北航学报 赠阅

步骤 4 将新的策略组合表示为 $\boldsymbol{s}^{(1)}=\{\boldsymbol{s}_1^*,\boldsymbol{s}_2^*,\cdots,\boldsymbol{s}_i^*,\cdots,\boldsymbol{s}_m^*\}$，计算相邻 2 个策略组合 $\boldsymbol{s}^{(0)}$ 与 $\boldsymbol{s}^{(1)}$ 之间的距离是否满足不等式 $\|\boldsymbol{s}^{(1)}-\boldsymbol{s}^{(0)}\|\leqslant\xi$，其中常数 ξ 为博弈收敛判据，若满足，则竞争博弈过程结束；否则以 $\boldsymbol{s}^{(1)}$ 为基础转至步骤 3 继续执行，直至所有博弈方的最优策略不再随迭代过程而发生变化，得到博弈均衡解。

3 案例分析

3.1 压力容器 MRDO 案例

1）压力容器 MRDO 数学模型

图 1 为某型压力容器的示意图[16]。案例包含：3 个正态分布的随机设计变量，概率分布参数如表 1 所示；4 个可靠性约束条件；2 个设计目标，分别使得质量 $w(\cdot)$ 的期望最小，容积 $v(\cdot)$ 的期望最大。

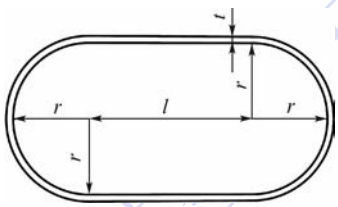


图 1 压力容器示意图[16]

Fig. 1 Schematic diagram of pressure vessel[16]

表 1 压力容器随机变量的概率分布

Table 1 Probability distribution of random variables of pressure vessel

变 量	均 值	变异系数	均值范围
r/mm	595.861 1	0.05	[2.54, 914.4]
l/mm	999.774 5	0.05	[2.54, 3 556]
t/mm	62.451 0	0.05	[12.7, 152.4]

压力容器的 MRDO 数学模型如式(7)所示，$w(\cdot)$ 和 $v(\cdot)$ 的表达式如式(8)所示：

$$\begin{cases}\min & \mu_w(r,l,t)\\ \max & \mu_v(r,l,t)\\ \text{s.t.} & \Pr\{\sigma_c=pr/t\leqslant\sigma_R\}\geqslant 0.95\\ & \Pr\{5t-r\leqslant 0\}\geqslant 0.95\\ & \Pr\{r+t-40\leqslant 0\}\geqslant 0.95\\ & \Pr\{l+2r+2t-150\leqslant 0\}\geqslant 0.95\end{cases}\tag{7}$$

$$\begin{cases}w(r,l,t)=\rho\pi\left[\dfrac{4}{3}(r+t)^3+(r+t)^2l-\dfrac{4}{3}r^3-r^2l\right]\\ v(r,l)=\dfrac{4}{3}\pi r^3+\pi r^2l\end{cases}\tag{8}$$

式中：σ_c 为圆周应力；$\sigma_R=241.3$ MPa 为材料的许用抗拉强度；$\rho=7\,833.4$ kg/m^3 为密度；$p=26.8$ MPa为压强。

2）基于 RDVSM 的策略集分解

首先，计算 3 个设计变量 r、l 和 t 对 2 个目标函数 $\mu_w(\cdot)$ 和 $\mu_v(\cdot)$ 的影响因子集分别为：$\boldsymbol{\eta}_1=\{0.505\,2,0\}$、$\boldsymbol{\eta}_2=\{1,0\}$、$\boldsymbol{\eta}_3=\{0.387\,1,0\}$。然后，对集合 $\boldsymbol{\eta}=\{\boldsymbol{\eta}_1,\boldsymbol{\eta}_2,\boldsymbol{\eta}_3\}$ 进行模糊聚类得：策略集 $\boldsymbol{S}_1=\{r,t\}$ 隶属于博弈方 $\mu_w(\cdot)$，$\boldsymbol{S}_2=\{l\}$ 隶属于博弈方 $\mu_v(\cdot)$。

3）竞争博弈求解

在所有的博弈方策略集中形成初始策略组合 $\boldsymbol{s}^{(0)}=\{\boldsymbol{s}_1^{(0)},\boldsymbol{s}_2^{(0)}\}$，其中，$\boldsymbol{s}_1^{(0)}=\{595.861\,1,62.451\,0\}$、$\boldsymbol{s}_2^{(0)}=\{999.774\,5\}$。依据竞争博弈算法的步骤，得到的博弈均衡解为 $\boldsymbol{s}_1^*=\{702.334\,6,106.996\,2\}$、$\boldsymbol{s}_2^*=\{1\,725.805\,5\}$，相应的目标函数值分别为 $\mu_w^*=12\,889.816\,6$，$\mu_v^*=4.123\,5$，实际的可靠度指标为 $\boldsymbol{R}_a=[0.999\,9,0.999\,9,0.999\,9,0.999\,9]$，满足约束要求。

4）优化结果对比分析

通过多学科优化设计软件 iSIGHT 中的带精英策略的快速非支配排序遗传算法(Non-dominated Sorting Genetic Algorithm Ⅱ，NSGA-Ⅱ)，结合 PMA 对压力容器进行 MRDO，NSGA-Ⅱ的参数设置为，种群规模 12，交叉概率 0.9，变异概率0.01，最大进化代数 100，得到部分 Pareto 非劣解及相应的目标函数值如表 2 所示，其中 μ_t^N、μ_l^N和 μ_r^N 为优化后的设计变量均值，μ_w^N 和 μ_v^N 为优化后的目标函数值。利用加权法将多目标问题转换为单目标问题，结合 PMA 与序列二次规划(Sequential Quadratic Programming，SQP)算法进行 SRDO，SQP 算法的参数设置为：最大迭代次数 40，终止精度 10^{-6}，步长 0.001，得到的优化结果如表 3 所示，其中 ω_1 和 ω_2 分别为 2 个目标函数的权重系数，μ_t^W、μ_l^W 和 μ_r^W 为优化后的设计变量均值，μ_w^W 和 μ_v^W 为优化后的目标函数值。

根据表 2，基于竞争博弈的目标函数值 $\mu_w^*=12\,889.816\,6$ 和 $\mu_v^*=4.123\,5$ 分别接近于非劣解 1 所对应的2个目标函数值$\mu_w^N=12\,393.860\,9$和

表 2 基于 NSGA-Ⅱ的压力容器部分优化结果

Table 2 Part optimization results of pressure vessel based on NSGA-Ⅱ

变 量	非劣解 1	非劣解 2	非劣解 3
μ_t^N/mm	107.554 0	108.746 8	106.403 1
μ_l^N/mm	1 059.278 0	1 509.354 4	1 601.394 3
μ_r^N/mm	795.133 5	815.710 1	827.923 2
μ_w^N/kg	12 393.860 9	15 135.898 9	15 523.401 2
μ_v^N/m^3	4.207 6	5.425 9	5.822 7

表 3　基于加权法的压力容器优化结果

Table 3　Optimization results of pressure vessel based on weighted method

变　量	权重组合 1 ($\omega_1=0.01,\omega_2=0.99$)	权重组合 2 ($\omega_1=0.1,\omega_2=0.9$)	权重组合 3 ($\omega_1=\omega_2=0.5$)	权重组合 4 ($\omega_1=0.9,\omega_2=0.1$)	权重组合 5 ($\omega_1=0.99,\omega_2=0.01$)
μ_t^{W}/mm	105.092 8	105.099 4	105.096 6	83.953 1	12.7
μ_l^{W}/mm	1 719.527 4	1 713.351 9	1 721.302 9	2 093.012 8	962.011 8
μ_r^{W}/mm	840.228 2	841.113 6	837.490 1	672.022 5	71.356 5
μ_w^{W}/kg	16 187.338 1	16 184.143 9	16 121.364 6	10 387.539 0	54.265 2
μ_v^{W}/m^3	6.295 3	6.297 5	6.250 2	4.238 7	0.016 9

$\mu_v^{\mathrm{N}}=4.207\,6$，表明利用竞争博弈求解 MRDO 问题的可行性、结果的有效性。竞争博弈方法的本质是多方决策，通过各博弈方的多轮博弈实现收益均衡、满足设计要求，期间不存在人为干预，博弈结果具有较高的客观性。

根据表 3 可得：①对比权重组合 3 ~ 权重组合 5，目标函数值 μ_w^{W} 和 μ_v^{W} 均随着权重的变化而产生显著的变化，随着权重的增大，容器质量 μ_w^{W} 呈现减小趋势，容积 μ_v^{W} 呈现减小趋势，体现出设计者的主观偏好；②对比权重组合 1 ~ 权重组合 3，权重的显著差异并未造成明显不同的优化结果，相反这 3 组权重组合所对应的目标函数值十分接近，并没有反映出设计者的主观偏好。产生上述结果的原因在于：权重法本质上属于单方决策，人为将多目标问题转化为单目标问题，掩盖了各目标间的冲突，使得优化结果缺乏客观性，甚至无法保证可信度。

3.2　减速器 MRDO 案例

1）减速器 MRDO 数学模型

图 2 为某小型航空发动机齿轮减速器传动原理的示意图[17]。案例包含 1 个确定性设计变量（齿轮 1 齿数），6 个正态分布随机设计变量，概率分布参数见表 4；6 个可靠性约束条件，其中 c1 为轮齿弯曲应力约束，c2 为轮齿接触应力约束，c3 和 c4 分别为轴 1 和轴 2 的横向变形约束，c5 和 c6 分别为轴 1 和轴 2 的应力约束；5 个由经验规定的确定性几何约束 c7 ~ c11；3 个设计目标分别使得减速器体积 $f_1(\cdot)$、齿轮轴1及齿轮轴2

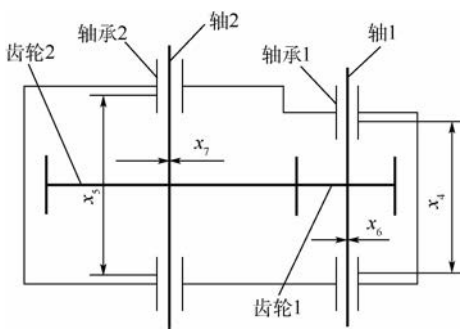


图 2　减速器传动原理示意图[17]

Fig. 2　Schematic diagram of drive principle of reducer[17]

表 4　减速器随机变量的概率分布

Table 4　Probability distribution of random variables of reducer　　cm

变　量	均　值	标准差	均值范围
齿面宽度 x_1	3.58	0.05	[2.6, 3.6]
齿轮模数 x_2	0.72	0.01	[0.7, 0.8]
轴 1 轴承间距 x_4	7.48	0.05	[7.3, 8.3]
轴 2 轴承间距 x_5	7.83	0.05	[7.3, 8.3]
轴 1 直径 x_6	3.37	0.05	[2.9, 3.9]
轴 2 直径 x_7	5.26	0.05	[5.0, 5.5]

注：齿轮 1 齿数 x_3（取整数）均值范围为[17,28]。

的应力 $f_2(\cdot)$ 和 $f_3(\cdot)$ 的期望最小。$f_1(\cdot)$、$f_2(\cdot)$ 与 $f_3(\cdot)$ 的表达式如式(9)所示，减速器的 MRDO 数学模型如式(10)所示：

$$\begin{cases} f_1(\boldsymbol{X}) = 0.785\,4x_1x_2^2(3.333\,3x_3^2+14.933\,4x_3-43.093\,4)-1.507\,9x_1(x_6^2+x_7^2)+7.477(x_6^3+x_7^3)+0.785\,4(x_4x_6^2+x_5x_7^2) \\ f_2(\boldsymbol{X}) = 10\sqrt{[745x_4/(x_2x_3)]^2+1.69\times10^7}/x_6^3 \\ f_3(\boldsymbol{X}) = 10\sqrt{[745x_5/(x_2x_3)]^2+1.575\times10^8}/x_7^3 \end{cases} \tag{9}$$

$$\begin{cases} \min\{\mu_{f_1}(\boldsymbol{X}),\mu_{f_2}(\boldsymbol{X}),\mu_{f_3}(\boldsymbol{X})\} \\ \text{s.t.} \\ \quad c1:\Pr\{g_1=27/(x_1x_2^2x_3)-1\geqslant0\}\geqslant0.95 \\ \quad c2:\Pr\{g_2=397.5/(x_1x_2^2x_3^2)-1\geqslant0\}\geqslant0.95 \\ \quad c3:\Pr\{g_3=1.93x_4^3/(x_2x_3x_6^4)-1\geqslant0\}\geqslant0.95 \\ \quad c4:\Pr\{g_4=1.93x_5^3/(x_2x_3x_7^4)-1\geqslant0\}\geqslant0.95 \\ \quad c5:\Pr\{g_5=\sqrt{[745x_4/(x_2x_3)]^2+1.69\times10^7}/x_6^3-110\geqslant0\}\geqslant0.95 \\ \quad c6:\Pr\{g_6=\sqrt{[745x_4/(x_2x_3)]^2+1.575\times10^8}/x_7^3-85\geqslant0\}\geqslant0.95 \\ \quad c7:x_2x_3-40\leqslant0 \\ \quad c8:5-x_1/x_2\leqslant0 \\ \quad c9:x_1/x_2-12\leqslant0 \\ \quad c10:(1.5x_6+1.9)/x_4-1\leqslant0 \\ \quad c11:(1.1x_7+1.9)/x_5-1\leqslant0 \end{cases} \tag{10}$$

北航学报 赠阅

2）减速器的博弈均衡解

根据竞争博弈求解算法的步骤，在对齿轮 1 的齿数 x_3 进行圆整处理之后，得到的博弈均衡解为 $\boldsymbol{s}_1^* = \{3.6, 0.7, 17, 8.1760, 8.3\}$，$\boldsymbol{s}_2^* = \{3.4892\}$，$\boldsymbol{s}_3^* = \{5.4307\}$；相应的目标函数值分别为 $\mu_{f_1}^* = 3186.0606$，$\mu_{f_2}^* = 97.5266$，$\mu_{f_3}^* = 78.4229$；$\boldsymbol{R}_a = [0.9994, 0.9999, 0.9999, 1.0, 0.9970, 0.9980]$ 满足可靠度要求。

3）优化结果对比分析

结合 NSGA-Ⅱ 和 PMA 对减速器进行 MRDO，NSGA-Ⅱ 的参数设置为：种群规模 12，交叉概率 0.92，变异概率 0.01，最大进化代数 120。所得 Pareto 非劣解集的一部分如表 5 所示，其中 $\mu_{x_1}^N \sim \mu_{x_7}^N$ 为优化后的设计变量均值，$\mu_{f_1}^N$、$\mu_{f_2}^N$ 和 $\mu_{f_3}^N$ 为优化后的目标函数值。利用加权法并结合 SQP 算法与 PMA 进行减速器的 SRDO，SQP 算法的参数设置为，最大迭代次数 60，终止精度 10^{-6}，步长 0.001，得到的优化结果如表 6 所示，其中 $\omega_1 \sim \omega_3$ 分别为 3 个目标函数的权重系数，$\mu_{x_1}^W \sim \mu_{x_7}^W$ 为优化后的设计变量均值，$\mu_{f_1}^W$、$\mu_{f_2}^W$ 和 $\mu_{f_3}^W$ 为优化后的目标函数值。

根据表 5 的最后 3 行数据，基于竞争博弈的目标函数值 $\mu_{f_1}^* = 3186.0606$、$\mu_{f_2}^* = 97.5266$ 和 $\mu_{f_3}^* = 78.4229$ 接近于 3 个非劣解对应的目标函数值，表明博弈均衡解的有效性和正确性。

根据表 6 可得：①分别对比权重组合 1 与权重组合 2（ω_1 存在显著差异）、权重组合 4 与权重组合 5（ω_2 和 ω_3 均存在显著差异），权重的差异没有引起目标函数值 $\mu_{f_1}^W$、$\mu_{f_2}^W$ 和 $\mu_{f_3}^W$ 明显的改变，未能体现出设计者的主观偏好；②分别对比权重组合 4 与权重组合 1、权重组合 4 与权重组合 2、权重组合 5 与权重组合 1、权重组合 5 与权重组合 2，目标函数值随着权重的变化而产生了明显的改变，但是权重的选取依赖于主观经验，设计结果个性化明显，客观性不足。

表 5 基于 NSGA-Ⅱ 的减速器部分优化结果

Table 5 Part optimization results of reducer based on NSGA-Ⅱ

变 量	非劣解 1	非劣解 2	非劣解 3
$\mu_{x_1}^N$/cm	3.5798	3.5884	3.5823
$\mu_{x_2}^N$/cm	0.7017	0.7009	0.7011
$\mu_{x_3}^N$	17	17	17
$\mu_{x_4}^N$/cm	8.0423	8.0438	8.0396
$\mu_{x_5}^N$/cm	7.9276	7.9623	8.0460
$\mu_{x_6}^N$/cm	3.5379	3.4561	3.5091
$\mu_{x_7}^N$/cm	5.3792	5.3905	5.3959
$\mu_{f_1}^N$/cm^3	3194.9760	3195.3554	3210.5762
$\mu_{f_2}^N$/MPa	93.5042	100.2957	95.8238
$\mu_{f_3}^N$/MPa	80.6881	80.1837	79.9444

表 6 基于加权法的减速器优化结果

Table 6 Optimization results of reducer based on weighted method

变 量	权重组合 1（$\omega_1=0.02$，$\omega_2=\omega_3=0.49$）	权重组合 2（$\omega_1=0.2$，$\omega_2=\omega_3=0.4$）	权重组合 3（$\omega_1=\omega_2=\omega_3=1/3$）	权重组合 4（$\omega_1=0.8$，$\omega_2=\omega_3=0.1$）	权重组合 5（$\omega_1=0.98$，$\omega_2=\omega_3=0.01$）
$\mu_{x_1}^W$/cm	3.50	3.50	3.50	3.50	3.50
$\mu_{x_2}^W$/cm	0.7	0.7	0.7	0.7	0.7
$\mu_{x_3}^N$	17	17	17	17	17
$\mu_{x_4}^W$/cm	7.9868	7.7501	7.7506	7.4751	7.4764
$\mu_{x_5}^W$/cm	7.9501	7.9520	8.1129	7.9959	7.9953
$\mu_{x_6}^W$/cm	3.9	3.9	3.9	3.4342	3.4340
$\mu_{x_7}^W$/cm	5.5	5.5	5.3690	5.3692	5.3692
$\mu_{f_1}^W$/cm^3	3313.3453	3310.5163	3226.0713	3077.5333	3077.4632
$\mu_{f_2}^W$/MPa	69.8133	69.7836	69.7837	102.1553	102.1777
$\mu_{f_3}^W$/MPa	75.4907	75.4939	81.1534	81.1432	81.1432

4 结 论

1）基于竞争博弈提出了一种解决 MRDO 问题的方法。通过 RDVSM 将随机设计变量集转换成策略集，所有博弈方均以自身收益为目标在各自策略集中进行 SRDO，由优化结果构成一轮博弈的策略组合，然后经过多轮博弈获得均衡解。2 个案例设计结果表明了方法的可行性，所求博弈均衡解具有较高的客观性。

2）在利用 PMA 对 MRDO 模型中的可靠性约束进行分析评估时，只考虑了单一的随机不确定性，工程实际当中有多种类型不确定性共存，应考虑在此基础上对基于竞争博弈的 MRDO 求解方法做进一步的研究和改进。

参考文献（References）

[1] 许孟辉，邱志平. 结构模糊非概率混合可靠性分析方法[J]. 北京航空航天大学学报，2014，40(2)：222-228.
XU M H，QIU Z P. Reliability analysis of structures with fuzzy and non-probabilistic hybrid variables[J]. Journal of Beijing University of Aeronautics and Astronautics，2014，40(2)：222-228(in Chinese).

[2] 孟广伟，冯昕宇，李锋，等. 基于降维算法和 Edgeworth 级数的结构可靠性分析[J]. 北京航空航天大学学报，2016，42(3)：421-425.
MENG G W，FENG X Y，LI F，et al. Structural reliability analysis based on dimensionality reduction and Edgeworth series[J]. Journal of Beijing University of Aeronautics and Astronautics，2016，42(3)：421-425(in Chinese).

[3] GARG H，RANI M，SHARMA S P，et al. Intuitionistic fuzzy optimization technique for solving multi-objective reliability optimization problems in interval environment[J]. Expert Systems with Applications，2014，41(7)：3157-3167.

[4] KOGISO N，KODAMA R，TOYODA M. Reliability-based multi-objective optimization using the satisficing trade-off method[J]. Mechanical Engineering Journal，2014，1(6)：1-12.

[5] 张瑞军，邱继伟，贾庆轩. 灰色系统理论的多目标可靠性稳健设计[J]. 北京邮电大学学报，2014，37(3)：23-26.
ZHANG R J，QIU J W，JIA Q X. Multi-objective robust design for reliability based on grey system theory[J]. Journal of Beijing University of Posts and Telecommunications，2014，37(3)：23-26(in Chinese).

[6] 王若冰，谷良贤，龚春林. 随机-区间混合不确定性分层序列化多学科可靠性分析方法[J]. 西北工业大学学报，2016，34(1)：139-146.
WANG R B，GU L X，GONG C L. A stratified sequencing multi-disciplinary reliability analysis method under random and interval uncertainty[J]. Journal of Northwestern Polytechnical University，2016，34(1)：139-146(in Chinese).

[7] 于淼，石博强，姜勇. 基于物理规划的湿式多盘制动器不确定性优化设计[J]. 农业机械学报，2011，42(4)：22-26.
YU M，SHI B Q，JIANG Y. Application of physical programming in uncertainty optimization design of multi-disc wet brake[J]. Transactions of the Chinese Society of Agricultural Machinery，2011，42(4)：22-26(in Chinese).

[8] JUANG C H，WANG L. Reliability-based robust geotechnical design of spread foundations using multi-objective genetic algorithm[J]. Computers and Geotechnics，2013，48(4)：96-106.

[9] 张干清，龚宪生，王欢欢，等. 基于可靠灰色粒子群算法的盾构机行星减速器轮系的多目标优化设计[J]. 机械工程学报，2010，46(23)：135-145.
ZHANG G Q，GONG X S，WANG H H，et al. Multi-objective optimization design on gear train of planetary reducer in shield tunneling machine based on reliably grey particle swarm optimization[J]. Journal of Mechanical Engineering，2010，46(23)：135-145(in Chinese).

[10] 龙腾，李学亮，黄波，等. 基于自适应代理模型的翼型气动隐身多目标优化[J]. 机械工程学报，2016，52(22)：101-111.
LONG T，LI X L，HUANG B，et al. Aerodynamic and stealthy performance optimization of airfoil based on adaptive surrogate model[J]. Journal of Mechanical Engineering，2016，52(22)：101-111(in Chinese).

[11] SONG C Z，ZHAO Y Q，WANG L. Tri-objective co-evolutionary algorithm and application of suspension parameter design based on lizard behavior bionics[J]. Journal of Mechanical Science and Technology，2014，28(12)：4857-4867.

[12] DESIDERI J A. Cooperation and competition in multidisciplinary optimization application to the aero-structural aircraft wing shape optimization[J]. Computational Optimization and Applications，2012，52(1)：29-68.

[13] XIAO M，SHAO X Y，GAO L，et al. A new methodology for multi-objective multidisciplinary design optimization problems based on game theory[J]. Expert Systems with Applications，2015，42(3)：1602-1612.

[14] YAO W，CHEN X Q，OUYANG Q，et al. A reliability-based multidisciplinary design optimization procedure based on combined probability and evidence theory[J]. Structural and Multidisciplinary Optimization，2013，48(2)：339-354.

[15] 曲福恒，崔广才，李岩芳，等. 模糊聚类算法及应用[M]. 北京：国防工业出版社，2011：57-66.
QU F H，CUI G C，LI Y F，et al. Fuzzy clustering algorithm and its application[M]. Beijing：National Defense Industry Press，2011：57-66(in Chinese).

[16] RAO J R J，BADHRINATH K，PAKALA R，et al. A study of optimal design under conflict using models of multi-player games[J]. Engineering Optimization，1997，28(1-2)：63-94.

[17] GOLINSKI J. Optimal synthesis problems solved by means of nonlinear programming and random methods[J]. Journal of Mechanisms，1970，5(3)：287-309.

作者简介：

冯嘉珍　男，博士研究生。主要研究方向：机械可靠性。

张建国　男，博士，教授，博士生导师。主要研究方向：机械/机构/结构可靠性。

北航学报 赠阅

Multi-objective reliability design optimization approach based on competition game

FENG Jiazhen[1,2], ZHANG Jianguo[1,2,*], QIU Jiwei[1,2]

(1. School of Reliability and Systems Engineering, Beijing University of Aeronautics and Astronautics, Beijing 100083, China;
2. Science and Technology on Reliability and Environmental Engineering Laboratory, Beijing University of Aeronautics and Astronautics, Beijing 100083, China)

Abstract: Aimed at the subjectivity of the selection of target weights, multi-objective reliability design optimization approach based on competition game is proposed. In this approach, every design objective is treated as the corresponding game player, and the random design variable set is decomposed into multiple strategy sets that are allocated to the corresponding player through the random design variable set mapping (RDVSM) technology. Then, combined with the performance measurement analysis method, every player takes its payoff as the single objective function for the reliability design optimization in its own strategy set, and the optimal results of all players form a group of strategies in this game round. After multi-round games, the equilibrium solution of the game is acquired. The study of a pressure vessel case and a gear reducer case shows that the proposed approach avoids the selection of target weights, and the design results have high objectivity.

Keywords: multi-objective; reliability design optimization; competition game; random design variable set mapping (RDVSM); game equilibrium solution

北京航空航天大学学报

Received: 2017-05-31; **Accepted**: 2017-08-01; **Published online**: 2017-09-22 15:07

URL: kns.cnki.net/kcms/detail/11.2625.V.20170922.1507.009.html

Foundation items: National Key R&D Program of China (2013CB733000); National Natural Science Foundation of China (51675026, 71671009)

* **Corresponding author**. E-mail: zjg@buaa.edu.cn

北航学报
赠 阅

《北京航空航天大学学报》征稿简则

《北京航空航天大学学报》是北京航空航天大学主办的以航空航天科学技术为特色的综合性自然科学学术期刊(月刊)。本刊以反映航空航天领域研究成果与动态、促进学术交流、培养科技人才和推动科技成果向社会生产力转化为办刊宗旨。本刊为中国自然科学技术核心期刊,并被 Ei Compendex 等国内外权威文献检索系统收录。本刊向国内外公开发行,为进一步提高办刊质量和所刊出文章的学术水平,特制定本简则。

1 论文作者及内容

1.1 本刊面向海内外所有学者。

1.2 主要刊载与航空航天科学技术有关的材料科学及工程、飞行器设计与制造、宇航科学与工程、信息与电子技术、控制技术和自动化工程、流体力学和动力工程、计算机科学及应用技术、可靠性工程与失效分析等领域的研究文章。航空航天科学技术民用方面以及具有航空航天工程背景的应用数学、应用物理、应用力学和工程管理等方面的文章也在本刊优先考虑之列。

2 来稿要求

2.1 论文应具有创新性、科学性、学术性和可读性。

2.2 论文为原创作品,尚未公开发表过,并且不涉及泄密问题。若发生侵权或泄密问题,一切责任由作者承担。

2.3 主题明确,数据可靠,图表清晰,逻辑严谨,文字精练,标点符号正确。

2.4 文稿撰写顺序:中文题名(一般不超过 20 个汉字),作者中文姓名、单位、所在城市、邮政编码,中文摘要(包括目的、方法、结果及结论),中文关键词(5~8 个),中图分类号,英文题名,作者英文姓名、单位、所在城市、邮政编码、国别,英文摘要,英文关键词,引言,正文,参考文献。首页下角注明基金项目名称及编号,作者信息。

2.5 作者请登录本刊网页进行在线投稿。

3 稿件的审核、录用与版权

3.1 来稿须经专家两审和主编、编委讨论后决定刊用与否。

3.2 若来稿经审查后认定不宜在本刊发表,将及时告知作者。如果在投稿满 3 个月后仍未收到本刊任何通知,作者有权改投它刊。在此之前,请勿一稿多投,否则一切后果自负。

3.3 来稿一经刊登,即赠送单行本。

3.4 来稿一经作者签字并在本刊刊出,即表明所有作者都已经认可其版权转至本刊编辑部。本刊在与国内外文献数据库或检索系统进行交流及合作时,不再征询作者意见。

邮寄地址:100191 北京市海淀区学院路 37 号 北京航空航天大学学报编辑部
办公地点:北京航空航天大学办公楼 405,407,409 房间
电 话:(010)82315594,82338922,82314839,82315426
E-mail:jbuaa@ buaa. edu. cn
http://bhxb. buaa. edu. cn
http://www. buaa. edu. cn

北京航空航天大学学报

北航学报
赠 阅

《北京航空航天大学学报》
第五届编辑委员会

主　任（主　编）：赵沁平

（以下按姓氏笔划为序）

副主任（副主编）：丁希仑　王少萍　孙志梅　李秋实　李焕喜　杨嘉陵
苗俊刚　相　艳　徐立军　钱德沛　曹晋滨

编　委：马殿富　王　琪　王　聪　邓小燕　王青云　王荣明　刘　宇
刘　红　江　洁　刘　强　闫　鹏　朱天乐　刘铁钢　齐铂金
陈万春　邹正平　苏东林　杨世春　沈成平　邱志平　宋知人
杨树斌　张晓林　杨晓奕　杨继萍　李惠峰　吴新开　张瑞丰
杨照华　宋凝芳　周　锐　林宇震　林贵平　战　强　姚仰平
胡庆雷　赵秋红　段海滨　赵巍胜　席　平　郭　宏　徐　洁
徐世杰　郭洪波　康　锐　翟　锦　熊华钢

北京航空航天大学学报
Beijing Hangkong Hangtian Daxue Xuebao
（原《北京航空学院学报》）
（月刊　1956 年创刊）
第 44 卷 第 4 期　2018 年 4 月

JOURNAL OF BEIJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS (JBUAA)
(Monthly, Started in 1956)
Vol.44 No.4 April 2018

主管单位　中华人民共和国工业和信息化部
主办单位　北京航空航天大学
主　　编　赵沁平
编辑出版　《北京航空航天大学学报》编辑部
邮　　编　100083
地　　址　北京市海淀区学院路 37 号
印　　刷　北京科信印刷有限公司
发　　行　北航文化传媒集团
发行范围　国内外发行
联系电话　(010) 82315594 82338922 82314839
电子信箱　jbuaa@buaa.edu.cn

Administrated by Ministry of Industry and Information Technology of the People's Republic of China
Sponsored by Beijing University of Aeronautics and Astronautics (BUAA)
(Beijing 100083, P. R. China)
Chief Editor ZHAO Qinping
Edited and Published by Editorial Board of JBUAA
Printed by Beijing Kexin Printing Co., Ltd.
Distributed by BUAA Culture Media Group Limited
Telephone (010) 82315594 82338922 82314839
E-mail jbuaa@buaa.edu.cn
http://bhxb.buaa.edu.cn

刊　　号　ISSN 1001-5965 / CN 11-2625/V
国内定价　50.00 元／期

ISSN 1001-5965
9 771001 596182 04>