



北航学报
赠阅

ISSN 1001-5965
CODEN BHHDE8

北京航空航天大学 学报

JOURNAL OF BEIJING UNIVERSITY OF
AERONAUTICS AND ASTRONAUTICS



2021-07

Vol.47 No.7

目 次

纤维增强复合材料疲劳寿命预测及损伤分析模型研究进展 程小全, 杜晓渊 (1311)

基于 IMU 与 UKF 的船舶升沉运动信息测量方法 ... 卢道华, 付怀达, 王佳, 蔡雅轩, 宋世磊 (1323)

载人航天器体装太阳能电池阵有效发电面积计算方法 杨胜, 王曦 (1332)

基于导重法的自重载荷下悬臂梁结构拓扑优化 任毅如, 向剑辉, 何杰, 宁克焱, 杨玲玲 (1338)

基于改进遗传算法的航空集装箱装载问题研究 张长勇, 翟一鸣 (1345)

航天员出舱可伸缩安全系绳设计与验证 ... 孙启臣, 姜坤, 王海亮, 张子岚, 秦俊杰, 王露斯 (1353)

涂覆吸波材料圆柱的 RCS 预估方法 赵京城, 杨宗凯, 王健, 张扬 (1362)

抬升角对食蚜蝇飞行动稳定性的影响 许娜, 周帅至, 牟晓蕾 (1372)

弹性欧拉梁大变形分析的椭圆积分统一形式 安朝, 谢长川, 杨超 (1379)

低信噪比下分组交织器识别 吴昭军, 刘凯, 钟兆根, 但波, 周磊珂 (1387)

基于深度残差收缩网络的滚动轴承故障诊断 车畅畅, 王华伟, 倪晓梅, 蔺瑞管 (1399)

高精度在轨实时轨道机动决策 解树聪, 董云峰 (1407)

基于中心-对数半长的区间数据主成分分析 赵青, 王惠文, 王珊珊 (1414)

压电周期板中耦合禁带影响规律分析 姜周, 李琳, 范雨, 王文君 (1422)

S 弯进气道出口旋流对轴流压气机性能的影响 王加乐, 程邦勤, 费晓文, 尉洋, 冯路宁 (1438)

脉冲星方位误差估计的两步卡尔曼滤波算法 肖永强, 王宏力, 冯磊, 由四海, 许强 (1446)

面向 BDSBAS 电文认证的 OTAR 播发策略设计 穆盛林, 陈颖, 刘婷, 刘成, 陈潇 (1453)

一种新的最速下降算法在自适应噪声对消中的应用 王帅, 向建军, 彭芳, 肖冰松 (1462)

改进的深度神经网络下遥感机场区域目标检测
..... 韩永赛, 马时平, 何林远, 李承昊, 朱明明, 许悦雷 (1470)

混沌多精英鲸鱼优化算法 汤安迪, 韩统, 徐登武, 谢磊 (1481)

带扰动块的细长旋成体背部绕流数值模拟 刘克奇, 李国辉, 张浩南, 张宴嘉 (1495)

焊接缺陷对异种铝合金 TIG 对接接头疲劳行为的影响
..... 王池权, 石亮, 张祥春, 刘志毅, 邵成伟 (1505)

期刊基本参数: CN 11-2625/V * 1956 * m * A4 * 204 * zh * P * ¥ 50.00 * 900 * 22 * 2021-07

CONTENTS

Research development of fatigue life prediction and damage analysis model of fiber-reinforced composite
..... CHENG Xiaoquan, DU Xiaoyuan (1311)

Measurement method of ship's heave motion information based on IMU and UKF algorithm
..... LU Daohua, FU Huaida, WANG Jia, CAI Yaxuan, SONG Shilei (1323)

A calculation method for effective power supply area of manned spacecraft with body mounted solar array
..... YANG Sheng, WANG Xi (1332)

Topology optimization of cantilever structure with self-weight load based on guide-weight method
..... REN Yiru, XIANG Jianhui, HE Jie, NING Keyan, YANG Lingling (1338)

Air container loading based on improved genetic algorithm
..... ZHANG Changyong, ZHAI Yiming (1345)

Design and verification of an adjustable length safety tether for astronaut extravehicular activities
..... SUN Qichen, JIANG Kun, WANG Hailiang, ZHANG Zilan, QIN Junjie, WANG Lusi (1353)

RCS prediction method for cylinder containing microwave absorbing material
..... ZHAO Jingcheng, YANG Zongkai, WANG Jian, ZHANG Yang (1362)

Effects of stroke deviation on flight stability of true hoverfly
..... XU Na, ZHOU Shuaizhi, MOU Xiaolei (1372)

Unified form of large deflection analysis of elastic Euler beam based on elliptic integral
..... AN Chao, XIE Changchuan, YANG Chao (1379)

Recognition of packet interleaver at low SNR
..... WU Zhaojun, LIU Kai, ZHONG Zhaogen, DAN Bo, ZHOU Leike (1387)

Fault diagnosis of rolling bearing based on deep residual shrinkage network
..... CHE Changchang, WANG Huawei, NI Xiaomei, LIN Ruiguan (1399)

High-precision on-orbit real-time orbital maneuver decision
..... XIE Shucong, DONG Yunfeng (1407)

A principal component analysis of interval data based on center and log-radius
..... ZHAO Qing, WANG Huiwen, WANG Shanshan (1414)

Influence analysis of coupled band gap in piezoelectric periodic plate
..... JIANG Zhou, LI Lin, FAN Yu, WANG Wenjun (1422)

Effects of swirl at outlet of S-shaped inlet on axial flow compressor performance
..... WANG Jiale, CHENG Bangqin, FEI Xiaowen, YU Yang, FENG Luning (1438)

Two-step Kalman filter algorithm for pulsar position error estimation
..... XIAO Yongqiang, WANG Hongli, FENG Lei, YOU Sihai, XU Qiang (1446)

Design of message authentication and OTAR broadcast strategy for BDSBAS
..... MU Shenglin, CHEN Ying, LIU Ting, LIU Cheng, CHEN Xiao (1453)

Application of a new steepest descent algorithm in adaptive noise cancellation
..... WANG Shuai, XIANG Jianjun, PENG Fang, XIAO Bingsong (1462)

Regional object detection of remote sensing airport based on improved deep neural network
..... HAN Yongsai, MA Shiping, HE Linyuan, LI Chenghao, ZHU Mingming, XU Yuelei (1470)

Chaotic multi-leader whale optimization algorithm
..... TANG Andi, HAN Tong, XU Dengwu, XIE Lei (1481)

Numerical simulation of flow around slender body with disturbing block
..... LIU Keqi, LI Guohui, ZHANG Haonan, ZHANG Yanjia (1495)

Influence of welding defects on fatigue behavior of dissimilar aluminum alloy TIG butt joint
..... WANG Chiquan, SHI Liang, ZHANG Xiangchun, LIU Zhiyi, SHAO Chengwei (1505)

http://bhxb.buaa.edu.cn jbuua@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2020.0229

纤维增强复合材料疲劳寿命预测及 损伤分析模型研究进展



程小全*, 杜晓渊

(北京航空航天大学 航空科学与工程学院, 北京 100083)

摘 要: 随着纤维增强复合材料主承力结构在多领域内的广泛应用, 疲劳成为复合材料结构设计必须面对的问题, 出现了大量用于复合材料结构疲劳寿命预测及损伤演化的分析模型与方法。目前的复合材料疲劳性能分析模型可以分为疲劳寿命模型、唯象模型和渐进损伤模型。对这3类模型的发展情况进行了综述和对比分析。其中, 疲劳寿命模型原理相对简单, 适用于工程结构的寿命估算; 唯象模型建立了材料剩余强度/刚度与循环数的数学关系, 可以预测结构的寿命与材料剩余力学性能; 渐进损伤模型不仅可以预测结构寿命和材料剩余力学性能, 还能分析结构疲劳损伤机理。对各类疲劳性能分析模型的发展趋势进行了讨论。指出了降低实施成本和提高通用型是各类疲劳性能分析模型有待解决的问题。

关键词: 复合材料; 疲劳; 寿命预测; 唯象模型; 渐进损伤

中图分类号: TB332

文献标志码: A

文章编号: 1001-5965(2021)07-1311-12

纤维增强复合材料在航空航天等领域已经得到了越来越广泛的应用。大多数航空飞行器都有着一定的寿命要求, 这意味着其结构需要承受较长时间循环载荷的作用。早期复合材料在飞机上的使用大都为非主承力结构, 工作应力不高, 复合材料结构的疲劳问题并不突出。随着复合材料逐渐应用到主承力结构, 结构减重要求越来越高, 使得复合材料结构的疲劳问题不断地显现了出来^[1]。传统金属材料的疲劳断裂行为表现为裂纹的形成及其扩展, 而复合材料的疲劳行为往往以损伤场的形式进行演化, 其中包含有纤维断裂、基体开裂、纤维屈曲、纤维-基体脱黏、分层等多种损伤模式, 且彼此相互耦合^[2]。同时复合材料的疲劳性能受到材料类型、铺层顺序、制造工艺、环境条件和载荷形式等多种因素的影响^[3], 因此, 其疲劳与损伤行为实际上非常复杂。

多年来, 人们对复合材料的疲劳性能开展了大量的研究工作, 建立了众多不同的模型, 试图对复合材料的疲劳行为进行模拟, 对疲劳寿命进行预测。目前, 复合材料疲劳性能分析模型可以分为3类: 疲劳寿命模型、唯象模型和渐进损伤模型^[4]。大部分疲劳性能分析模型是建立在特定材料、特定铺层顺序和特定载荷形式基础上的, 并且模型的性能强烈依赖于大量的试验数据。由于现有的理论和分析模型仍不够成熟, 很多情况下复合材料结构设计中往往要采用较大的安全系数来保证结构安全, 并配合成本高昂的试验加以验证, 导致复合材料的减重优势难以充分发挥出来。因此, 完善疲劳寿命预测和损伤分析方法对复合材料结构的实际应用意义重大。

目前, 存在少量关于复合材料疲劳性能分析模型的综述。Vassilopoulos^[5-6]在综述复合材料疲

收稿日期: 2020-06-01; 录用日期: 2020-08-30; 网络出版时间: 2020-09-08 14:58

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20200907.1335.002.html

* 通信作者. E-mail: xiaoquan_cheng@buaa.edu.cn

引用格式: 程小全, 杜晓渊. 纤维增强复合材料疲劳寿命预测及损伤分析模型研究进展[J]. 北京航空航天大学学报, 2021, 47(7): 1311-1322. CHENG X Q, DU X Y. Research development of fatigue life prediction and damage analysis model of fiber-reinforced composite[J]. Journal of Beijing University of Aeronautics and Astronautics, 2021, 47(7): 1311-1322 (in Chinese).

劳理论发展情况时,单方面总结了疲劳寿命模型的发展历程。Bak^[7]和Pascoe^[8]等在综述复合材料疲劳分层理论时,只对疲劳分层预测模型进行了汇总。国内学者翟洪军和姚卫星^[9]综述了复合材料的剩余刚度唯象模型。这些综述都只考虑了单一类型的疲劳性能分析模型,缺乏不同类模型间的对比分析。

总结并对比分析各类复合材料疲劳性能分析模型,有助于全面了解复合材料的疲劳寿命预测和损伤分析方法,以及不同尺度下的疲劳损伤机理,同时也有助于明确未来的发展方向,并进一步完善各类模型。基于此,本文综述了具有代表性的复合材料疲劳性能分析模型,通过对各类复合材料疲劳性能分析模型研究现状的总结,分析了现存疲劳性能分析模型的不足,并讨论了疲劳性能分析模型的发展趋势,旨在为本领域内的研究工作和工程应用提供一定的参考。

1 疲劳寿命模型

疲劳寿命模型是从已有结构的S-N曲线或等寿命曲线(CLD)中提取出结构的疲劳寿命信息,所给出的结果是在某种载荷情况下试件发生疲劳破坏的循环数。

1.1 疲劳寿命模型发展

最初Hashin和Rotem^[10]基于Hashin静强度准则,提出了可以用于判定纤维失效和基体失效2种模式的疲劳判据。不同于静力下的失效判据,在疲劳寿命模型中极限强度不再是固定值,而是与加载应力水平、应力比和循环数有函数关系的变量。该模型中,极限强度的表达式需要通过各单向板在相应载荷情况下的S-N曲线中获得,这也意味着确定下来的一套准则方程只适用于相应载荷情况下的单层板结构。Reifsnider和Gao^[11]在文献[10]的基础上给出了应用于微观力学水平下的疲劳失效准则,其准则方程中考虑

了不同组分(纤维和基体)的材料强度参数。Philippidis和Vassilopoulos^[12]基于Tsai-Hill准则提出了相应的疲劳失效准则,同时基于Tsai-Wu张量准则提出了类似的疲劳准则,用于计算张量分量的材料静强度值是取之于相应的S-N曲线上的应力值。Fawaz和Ellyin^[13]提出了循环应力水平 S 和循环数 N 的半对数线性关系: $S = a \lg N + b$, a 和 b 为与材料相关的参数。

除了使用S-N曲线建立疲劳寿命模型外,CLD也被广泛地用于预测结构的疲劳寿命。典型的CLD如图1所示^[6]。把疲劳试验数据点画在以疲劳应力平均值 σ_m 为横轴、疲劳应力幅值 σ_a 为纵轴的坐标中,将具有相同循环寿命的点用曲线连起来即得到CLD曲线。在CLD中,同一条S-N曲线中的数据点具有相同的应力比 R ,这些点连接起来会成为一条从原点出发的直线。不同寿命对应的曲线互不相交,但是形式相同。不少学者提出了不同的CLD构建方法及相应的疲劳寿命模型。

类似于金属疲劳CLD,Brøndsted等^[14]认为复合材料的线性CLD可以通过应力比 $R = -1$ 时的应力幅值 σ_{a0} 、静力拉伸强度 U_{TS} 和静力压缩强度 U_{CS} 三个数据点连接而成,如图2(a)所示^[15]。线性CLD模型形式简单,但通过使用线性CLD模型所预测的疲劳寿命与试验结果比对存在较大误差^[15]。在线性CLD基础上,分段线性CLD被提出^[16]。除了线性CLD所需要的数据点,分段线性CLD还需要几条试验测得的S-N曲线来提供数据点,一般选择3个特殊应力比(0.1、-1和10)下的S-N曲线来分别提供拉-拉、拉-压和压-压载荷下的数据点,如图2(b)所示^[15]。图2中:从外而内分别为 $N = 10^3$ 到 10^7 的曲线。

由于复合材料拉伸性能和压缩性能并不相同,Harris^[17]认为复合材料的CLD不应关于应力幅值 σ_a 的纵轴对称。当拉伸性能高于压缩性能

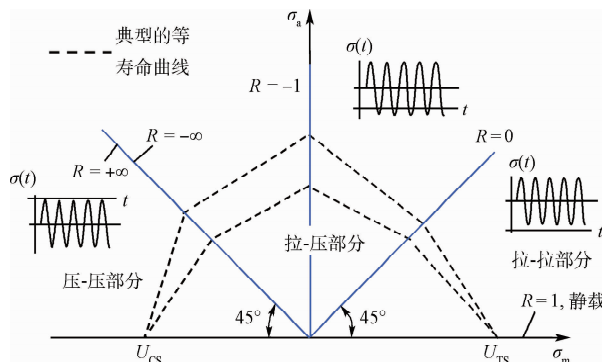
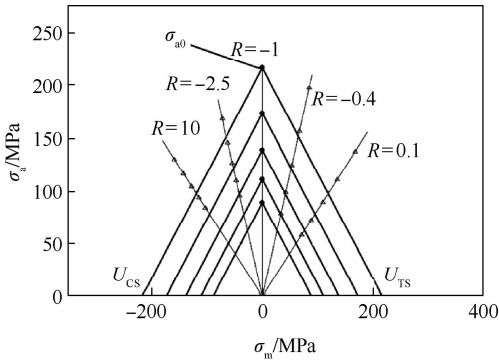
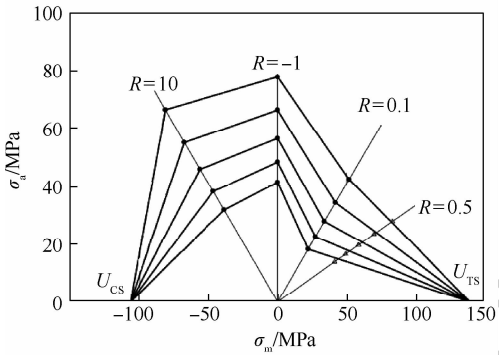


图1 典型等寿命曲线^[6]

Fig. 1 Typical constant fatigue life curve^[6]



(a) 线性模型



(b) 分段线性模型

● 试验数据 ▲ 待验证的试验数据 —— CLD 曲线

图 2 不同 CLD 模型^[15]

Fig. 2 Different CLD models^[15]

时,CLD 曲线应该向右倾斜;当压缩性能高于拉伸性能时,CLD 曲线应向左倾斜。基于以上特征,一种钟摆形 CLD 模型被提出。Kawai 和 Koizumi^[18]也提出了一种不对称的 CLD 模型,该模型只需要使用一条试验 S-N 曲线,即临界 S-N 曲线即可构造出 CLD。该临界 S-N 曲线的应力比 χ 被定义为材料的压缩极限强度与拉伸极限强度之比,称为临界应力比($\chi = U_{CS}/U_{TS}$)。

Boerstra^[19]提出的 CLD 模型中的应力比不再是一个定值参数,模型可以用于描述疲劳均值和幅值连续变化下的材料寿命,但该模型中的一些参数无法直接给出,而是需要进行多目标优化的过程来对参数进行估计。相比较之下,Kassapoglou^[20]则提出了一个简单的模型,该模型假设材料在一个循环中的失效概率是恒定的,与当前状态或到此为止的循环数无关。模型认为可以用材料的静力损伤统计数据分布来描述疲劳载荷下的损伤,从而该模型不需要进行疲劳试验,也不需要经验确定参数。但是这样的假设过于简化,且忽略了疲劳与静载下材料损伤机理的不同。Park 等^[21]认为拉伸疲劳载荷和压缩疲劳载荷对 CLD 曲线的影响不同,并且随着循环数的增加曲线的形状也会变化。基于此,考虑分段非线性

CLD 模型被提出,表示为

$$\sigma_a = \begin{cases} \sigma_{a0} [1 - (\sigma_m/U_{TS})^{\alpha_T}] & \sigma_m > 0 \\ \sigma_{a0} [1 - (\sigma_m/U_{CS})^{\alpha_C}] & \sigma_m < 0 \end{cases} \quad (1)$$

式中: α_T 和 α_C 分别为 CLD 曲线拉伸和压缩部分的形状参数。

近年来,随着计算机技术和人工智能(AI)理论的发展,使用基因遗传算法(GP)^[22-24]、神经网络算法(ANN)^[25-27]和模糊神经网络^[28-29]等理论来建立的复合材料疲劳寿命模型被陆续提出。Vassilopoulos 等^[22-23]把 GP 作为一种随机非线性回归分析工具,通过单一的数据输入对应单一的数据输出对算法进行训练,以拟合材料的疲劳寿命。由于 GP 算法与材料力学特性无关,只要有足够的数据进行训练,模型就可以用于预测任何材料的疲劳寿命。由于算法具有随机性,输出的结果可以不断扩充算法训练的数据库。ANN 模型^[25]使用误差反向传播算法进行训练。从试验结果的数据集中随机选取 50% 的数据用于构造训练集,剩下的 50% 用于模型预测结果的验证。模型的目的在于建立循环数 N 与输入参数 (R 和 σ_a 等)的函数关系。与传统的 CLD 构建方法相比,在相同的载荷形式下使用 ANN 算法可以节省大约 50% 的试验数据。Vassilopoulos 和 Bedi^[28]在 ANN 模型的基础上发展了自适应神经模糊推理系统(ANFIS)疲劳寿命预测模型。与线性回归、威布尔统计等传统复合材料疲劳寿命建模方法相比,ANFIS 模型只使用 50% ~ 60% 的试验数据就足以确定出可靠的 S-N 曲线,然而该模型不能在数据库之外做出任何预测。

1.2 疲劳寿命模型总结

目前,已有的疲劳寿命模型大多是使用传统的数学模型或者数值方法来构建公式或绘制 CLD 曲线。有的模型试验成本低,计算简单,但是预测结果与试验数据误差较大;另外一些预测结果相对准确的模型则需要大量的试验 S-N 曲线作为支撑,并且其模型中含有大量的参数需要优化或者经验性的估计。

与传统模型相比,AI 模型可以使用较少的试验数据给出复合材料在多轴疲劳载荷下的寿命预测结果,并且结果具有一定的可靠性。然而以 AI 方法构建的疲劳寿命模型也存在缺陷。例如,GP 对数据出现过拟合或欠拟合等缺陷时,预测结果会受到负面影响,并且算法程序容易过早收敛于局部最优解。所有的这些 AI 方法提供的数据输出是随机的,不能代表真正的试验结果,因此模型输出的随机数据不能用于进一步的疲劳分析。然

而, AI方法构建的复合材料疲劳寿命模型已被证实是高效而成本低廉的。随着机器学习和数据挖掘技术的不断发展, AI方法将逐渐成为复合材料疲劳寿命预测的主要工具之一。

2 唯象模型

复合材料的疲劳唯象模型是通过某一特定的材料或结构参数的衰减来描述材料因疲劳引起的损伤。剩余强度模型和剩余刚度模型是2种主流的疲劳唯象模型, 通过引入强度退化或刚度退化来考虑材料的实际损伤状态, 从而对复合材料的疲劳寿命进行预测。

2.1 剩余强度模型发展

试验表明, 复合材料结构的剩余强度变化如图3所示(标记点表示在给定的循环数 N' 下的剩余强度数据), 在一开始剩余强度下降平缓, 然后在临近失效时快速下降, 即所谓的“突然死亡现象”^[30]。基于这样的变化特性, 学者们提出了大量剩余强度与循环数的关系表达式以建立疲劳寿命预测模型。

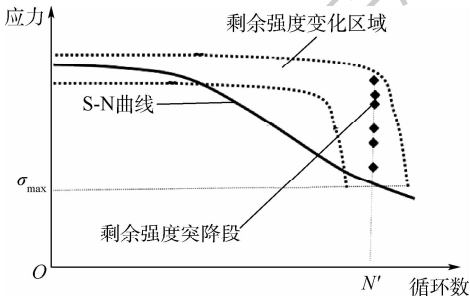


图3 最大循环应力 σ_{\max} 下剩余强度边界变化示意图^[31]

Fig.3 Schematic of residual strength boundaries at maximum cyclic stress σ_{\max} ^[31]

最初, Broutman 和 Sahu^[32] 认为复合材料剩余强度的衰减和循环数成线性关系, 而 Halpin 等^[33] 认为剩余强度的衰减与循环数呈幂律形式, 即

$$\frac{d\sigma_r(N)}{dN} = \frac{-A(\sigma_{\max})}{m[\sigma_r(N)]^{m-1}} \quad (2)$$

式中: $\sigma_r(N)$ 为循环数为 N 时的剩余强度; $A(\sigma_{\max})$ 为最大循环应力 σ_{\max} 的函数; m 为常数。

Halpin 的模型启发了大量的学者, 后人对该模型进行了深入研究与改进。Yang 等^[34-35] 基于幂律形式提出了另一种剩余强度表达式, 并分别推导了疲劳寿命对剩余强度的三参数和双参数威布尔分布表达式。Yang 的三参数模型被 Dick 等^[36] 加以修改, 提高了对疲劳寿命的预测精度。但是, 进一步研究发现, 这种三参数模型中的参数并非是与材料性能相关的常数, 使得模型的使用

异常困难。D'Amore 等^[37] 也提出了一种双参数模型, 该模型假定当剩余强度等于最大循环应力时发生失效, 其中明确地考虑了应力比对疲劳寿命的影响。Diao 等^[38] 根据材料静强度的分布, 导出了剩余强度分布的概率密度分布函数, 从而建立了可以预测剩余强度和疲劳寿命的统计模型, 该模型虽然与试验吻合较好, 但是需要大量的试验数据作为输入条件。

Yao 和 Himmel^[39] 通过试验观测, 发现大多数纤维增强复合材料在受拉情况下的剩余强度与损伤演化服从这样的变化过程, 即在疲劳寿命初期, 剩余强度的轻微下降与材料内的基体开裂相关; 在疲劳寿命中期, 剩余强度适度而均匀的下降主要与分层扩展和随机的纤维断裂相关; 在接近失效点阶段, 剩余强度剧烈下降。根据这样的变化过程, Yao 和 Himmel^[39] 提出了一种双参数剩余强度模型:

$$\sigma_r = \sigma_b - (\sigma_b - \sigma_{\max}) \frac{\sin(\beta x) \cos(\beta - \alpha)}{\sin \beta \cos(\beta x - \alpha)} \quad (3)$$

式中: σ_r 和 σ_b 分别为剩余强度和静强度; x 为疲劳寿命比 N/N_f , 即当前循环数与失效循环数之比; α 和 β 为由试验获得的常数。

Yao 和 Himmel^[39] 的模型可以用于变幅疲劳载荷的情况, 且不断地被后人改进并证明其有效性^[40], 不过改进的模型需要更多的试验数据来对参数 α 和 β 进行更明确的定义。D'Amore 等^[31, 41] 重点研究了剩余强度在失效前突然衰减阶段的描述方式, 提出了类似的双参数剩余强度模型, 结果表明, 疲劳寿命和剩余强度均与静强度的统计分布情况有关。

根据 FAA 的公开报告^[42], 在飞机复合材料结构使用比较广泛的是 Sendeckyj 等效静强度模型^[43] 和 Kassapoglou 模型^[44]。使用 Sendeckyj 模型可以把每条 S-N 曲线中的疲劳寿命和剩余强度数据转化为等效强度数据的点集, 利用 Sendeckyj 提出的方法将这些点集拟合为威布尔分布形式^[45]。Kassapoglou 模型是一种预测常幅疲劳寿命的概率型方法, 其假设在任何循环数下疲劳失效概率都是一个常值, 并且疲劳失效概率只与静强度测试结果的离散度相关。Kassapoglou 模型仅用静态数据就能对 S-N 曲线进行预测, 然而在负应力比情况下预测得出的 S-N 保守性较差, 寿命误差至少有 1~2 个数量级。

2.2 剩余刚度模型发展

疲劳试验中刚度有不同的定义方法, 如图4^[46]所示的初始切线刚度 $T(N)$ 和割线刚度

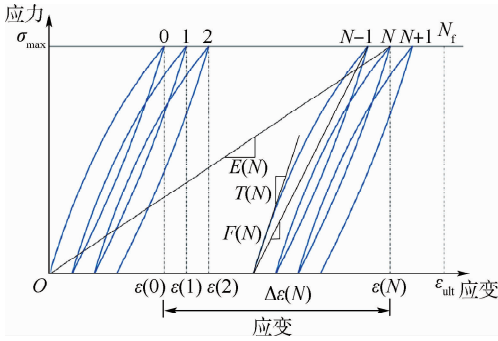


图4 疲劳载荷下的刚度^[46]

Fig. 4 Definition of stiffness under fatigue load^[46]

$F(N)$,但广泛用于预测疲劳寿命剩余刚度的则是由 Hwang 和 Han^[46] 提出的疲劳模量 $E(N)$, 并且指出 $E(N)$ 的衰减率满足幂律公式。而 Echtermeyer 等^[47] 根据试验结果认为, 剩余刚度与对数循环数呈线性关系。

Sidoroff 和 Subagio^[48] 在描述三点弯曲疲劳试验刚度衰减过程中, 率先引入了损伤量 $D = 1 - E(N)/E(0)$, $E(N)/E(0)$ 为某一循环下剩余刚度与初始刚度之比, 并建立了损伤量变化率 dD/dN 与应变幅值的数学关系。由于疲劳试验加载一般是用载荷控制, van Paepegem 和 Degrieck^[49] 将 Sidoroff 模型中的应变幅替换为应力幅。Vieilleigne 等^[50] 更严格地定义了疲劳载荷下的损伤变化增长率 dD/dN , 但 dD/dN 在压缩应力下被认为是 0。同样基于连续损伤变量的概念, Kawai 等^[51-53] 建立了可以用于预测复合材料偏轴疲劳寿命的模型。模型引入了一个基于 Tsai-Hill 准则定义的无量纲有效应力, 并且模型把单层板纤维主导的疲劳损伤和基体主导的疲劳损伤分开考虑。

Yang 等^[54] 提出了可以描述剩余刚度 $E(N)$ 退化的公式:

$$\frac{dE(N)}{dN} = -E(0)Q\nu N^{\nu-1} \quad (4)$$

式中: $E(0)$ 为初始刚度; Q 和 ν 为取决于载荷条件的参数, 可由试验结果拟合得到。基于式(4), 文献[55-56]使用不同的数学方法对该演化规律进行了研究, 得到了偏保守但令人满意的疲劳寿命预测结果。王丹勇和温卫东^[57] 改进了 Yang 的剩余刚度模型, 推导出了与试验曲线更加吻合的剩余刚度公式。模型中的剩余刚度除了与循环数相关, 还是最大循环应力和应力比的函数。该模型可以用于描述复合材料单层板各主方向上的刚度退化, 区分开了不同方向上剩余刚度的不同退化情况。

考虑到复合材料疲劳过程中剩余刚度衰减有如图 5 所示的 3 个阶段: 初始剧烈下降段(阶段 I)、

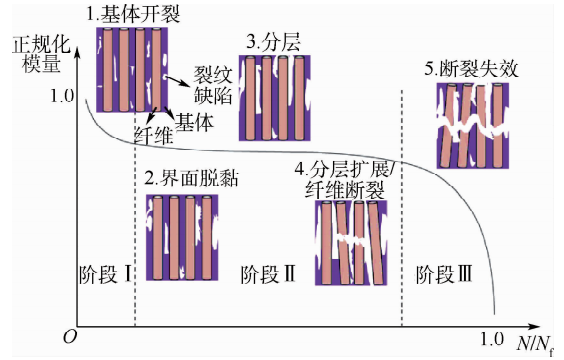


图5 复合材料疲劳过程中典型的归一化模量下降曲线与特征损伤状态^[58]

Fig. 5 Typical normalized modulus degradation curve and characteristic damage state during the fatigue life of composite laminate^[58]

中期平缓下降段(阶段 II)和后期剧烈下降段(阶段 III)^[58]。早期的大多数剩余刚度模型并非在 3 个衰减阶段都是有效的, 尤其在试件临近失效的阶段。Whitworth^[59] 提出的剩余刚度退化规律就须要使用到失效时的模量 $E(N_f)$ 。只要试件不发生失效, 该模型就无法得到 $E(N_f)$ 的确切值。

如何完整地描述剩余刚度的衰减规律成为近年来的一个研究热点。van Paepegem 和 Degrieck^[60] 分区描述了不同阶段的刚度衰减情况, 在考虑最终失效时使用了基于 Tsai-Wu 失效准则的应力度量, 使得模型参数可以与静力测试结果相关联。Wu 和 Yao^[61] 的双参数模型则可以完整地描述剩余刚度的衰减规律:

$$D(N) = \frac{E(0) - E(N)}{E(0) - E(N_f)} = 1 - \left[1 - \left(\frac{N}{N_f} \right)^B \right]^A \quad (5)$$

式中: $D(N)$ 为疲劳损伤变量, 初始时 $D(0) = 0$, 失效时 $D(N_f) = 1$; A 和 B 为试验确定的材料参数。

类似于 Wu 的模型, Shiri 等^[62] 使用三角函数关系式更好地逼近剩余刚度随循环数的衰减全过程。

2.3 唯象模型总结

由于每一根复合材料试件只能得到一个剩余强度数据, 不同循环数下试件的剩余强度测试是一个需要耗费大量时间和金钱的工作, 而剩余刚度可以在疲劳加载的过程中不间断地通过无损方式测得, 剩余强度模型的使用成本要比剩余刚度模型高很多。

目前的唯象模型存在的不足主要表现为: 不能完整而准确地描述整个强度/刚度衰减的过程; 只适用于特定的材料和有限范围的载荷水平; 需要试验确定的参数多; 仅适用于单轴载荷的情况。

将来的研究中, 除了建立更加准确的模型以

描述结构的剩余强度/刚度衰减外,还需要结合不同常规的测试技术及更多的试验投入,把该类模型应用于复杂载荷和环境条件下。根据一些疲劳试验的观测结果,其他材料性能,如泊松比^[63]、横向模量^[64]、面内剪切模量^[65]等参数的衰减也是不容忽视的。如何在唯象模型中考虑这些材料性能的衰减也有待进一步的研究。

3 渐进损伤模型

比起其他2类模型,渐进损伤模型不仅能够预测疲劳寿命,还能使用适当的失效准则和损伤变量对材料性能进行退化,进而实现预测损伤扩展和预测材料剩余力学性能的功能。根据实际需求,渐进损伤模型分别在分析宏观损伤、微观损伤和分层损伤方面得到了发展。

3.1 渐进损伤模型发展

自20世纪80年代末以来,可预测疲劳损伤扩展的模型被相继提出^[66-68]。碍于没有强大的计算机技术支撑,早期的渐进损伤模型仅是用于描述材料疲劳损伤演化过程的经验公式。随着计算机及有限元(FEM)技术的发展,功能更强大、适用范围更广的渐进损伤模型被陆续提出。

Shokrieh和Lessard^[69-70]结合应力分析、失效分析和材料性能衰减,提出了适用于碳纤维复合材料单向带的疲劳模型。该模型可以通过FEM预测疲劳损伤演化过程,是最受欢迎的渐进损伤模型之一,其运行原理如图6所示,其中使用了类似于静力Hashin的失效判据,以考虑不同的损伤模式。失效判据中材料性能参数不再是常数,而是循环数、应力水平和应力比的函数。FEM模型的单元彻底失效时,使用突降准则把某些相关的

材料性能衰减至0。对于没有彻底失效的单元,则使用渐进衰减准则来改变其材料参数。在应用于钉孔挤压这样拥有复杂应力状态的结构时,模型预测的疲劳寿命与试验吻合较好;其缺点是每种材料都需要一定数量的剩余强度/刚度试验数据以获取其中的参数值。

Papanikos等^[71]在Shokrieh模型^[69-70]的基础上做了改进。其模型对于纤维拉压失效改用了最大应力准则,同时还使用了Ye分层判据^[72]来判定法向拉压分层失效。在材料性能衰减方面,模型的突降准则不再把材料性能降至0,而是根据试验实测值进行衰减。模型对剩余刚度和剩余强度的渐进衰减准则也进行了简化。

Passipoularidis等^[73]建立了有相似运行原理的模型。与其他模型不同的是,该模型基于经典层合板理论对层合板单层进行面内应力分析,未考虑面外剪切和厚度方向的正应力,因此无法给出分层损伤的扩展情况;但模型采用了Puck失效判据^[74],包含了与分层相关的损伤。Eliopoulos和Philippidis^[75-76]进一步发展了Passipoularidis模型,通过疲劳试验和剩余性能试验验证了模型的有效性,并将其用于模拟不同铺层层合板的疲劳损伤。该模型忽视了应力、应变场的三维特性,导致分层损伤情况无法被模拟;基体材料的黏弹性没有明确考虑,因此应变率、加载频率和湿热等对材料性能的影响会被忽略,存在一定的局限性。

Zhao等^[77]改进并简化了渐进损伤模型,其中材料失效判据使用了修正过的最大应力准则,对材料性能突降准则也作了调整,使其更符合所研究的材料,并将材料性能渐进衰减准则简化为只有一个需要试验确定的参数。该模型用于简化载荷谱下的三钉双搭接结构疲劳性能分析,给出了损伤扩展的过程。

在文献[77]的基础上,Shan等^[78]把模型原有的材料性能衰减表达式改为了分段函数的形式,以更好地拟合试验测出的剩余性能衰减曲线,同时还引入了无量纲温度 $T^* = (T_g - T)/(T_g^0 - T_0)$ 对材料性能进行折减,实现了对湿热环境条件下复合材料层合板的疲劳损伤及寿命预测。其中, T_g 和 T_g^0 分别为湿态和干态下的材料玻璃化温度, T 和 T_0 分别为试验温度和室温。

由于复合材料的宏观性能与各组分的细观特性密切相关,不少学者拟通过复合材料的细观力学来揭示疲劳损伤的特征,促进了细观-宏观多尺度渐进损伤疲劳模型的发展。

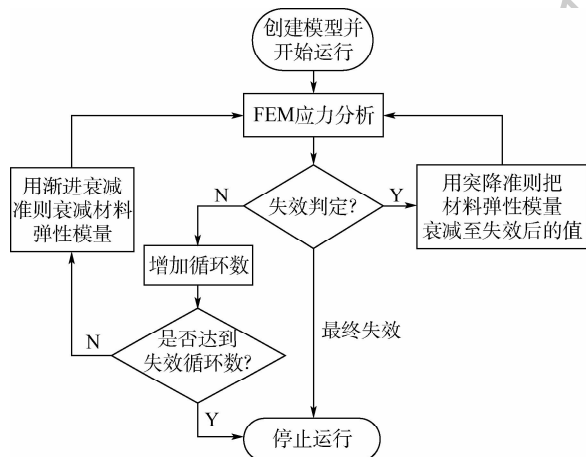


图6 Shokrieh 渐进损伤模型运行流程

Fig. 6 Operation flowchart of Shokrieh's progressive damage model

Sayyidmousavi 等^[79]把简化单胞细观力学模型(SUCM)结合到 FEM 中。SUCM 假设层合板的材料截面可以表示为矩形的纤维和基体亚胞。模型在每一次循环中的运行原理为:在宏观层面,层合板的应力场由 FEM 计算得到并输入到 SUCM 中;在细观层面,根据细观力学的失效准则,失效单元的某些相关材料性能将会衰减,更新后的材料参数又输入回宏观层面重新计算应力。模型在模拟层合板的疲劳过程中识别出了纤维拉伸、横向基体开裂和剪切失效 3 种失效模式。由于没有考虑分层损伤,模型只适用于分层不是主要失效模式的薄板结构。Li 等^[80]建立了与其具有相同运行原理的模型,但相比之下该细观单胞模型更加贴适于实际的组分形状。如图 7 所示^[80],圆柱纤维以面心阵列分布于单胞中,在单胞中还确定了用于提取细观应力的关键点。应用该模型,复合材料螺栓连接结构孔边的疲劳失效机理得到了具体分析,但损伤中仍然不包含分层。

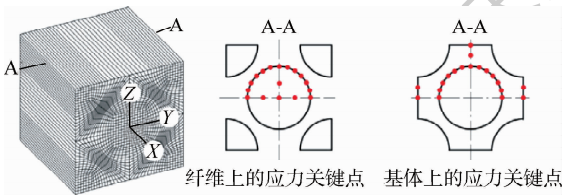


图 7 面心阵列分布的复合材料细观单胞模型^[80]

Fig. 7 Mesoscopic unit cell model with face-centered array distribution of composite^[80]

Hosseini Kordkheili 等^[81]建立的多尺度疲劳损伤预测模型把断裂动力学理论用于评估纤维和基体的损伤,并通过带有 2 个可变损伤参数的桥联模型把宏观应力与细观应力联系起来。该模型的优势在于只需要通过 0° 和 90° 单向带的疲劳试验即可分别获得纤维和基体断裂动力学公式中的参数,但该模型的缺陷在于除了纤维和基体损伤外无法考虑其他类型的损伤。

协同损伤力学(SDM)模型作为跨尺度的损伤分析方法,因其具有明确的物理意义并密切地与材料损伤机理相关,近年来越来越受到重视。其中,最具代表性的模型是 Talreja 等^[82-83]提出的一种建立在连续损伤力学(CDM)架构上的 SDM 模型。该模型利用 CDM 方法可以求解出复合材料在任一含损伤区域内的损伤张量。有了损伤张量,就可以使结构的刚度衰减与损伤建立数学关系。该模型需要通过试验得到 8 个宏观材料常数(4 个材料弹性参数和 4 个唯象损伤常数),并由微观尺度下 FEM 得到与裂纹张开位移相关的一

个约束参数。SDM 模型在 Shen 等^[84]的改进下被应用于描述复合材料疲劳初期由于散布损伤引起的结构刚度衰减。该模型中,试验件的损伤参数是通过无损检测横向基体裂纹密度得到的。由于无损检测的局限,模型忽略了基体裂纹尖端的局部分层损伤,导致预测结果与试验结果存在一定的误差。

在所有损伤模式中,导致分层所需的载荷是最低的^[85],而分层的进一步扩展会导致结构刚度下降甚至提前失效,因此,结构疲劳分层损伤是设计分析人员重点考虑的对象。目前,在大多数渐进损伤疲劳模型中,为了简化模型或者碍于模型原理的限制,分层损伤模式常常被忽略。若要准确预测复合材料疲劳分层扩展,往往需要建立专门的模型来进行分析。近年来,许多学者基于断裂力学概念对分层损伤进行描述,如使用 Paris 公式及其相应的变体来描述分层扩展行为^[86-87]。这些方法往往都离不开使用应变能释放率、应力集中因子及相应试验拟合参数来与裂纹扩展速率挂钩。

目前众多的疲劳分层模型中,使用 cohesive 界面单元模拟分层损伤是最近兴起的一种高效的渐进损伤方法。该单元异于普通单元之处在于其是一种双线性的本构关系,其应力-位移(σ - δ)曲线如图 8 所示。

Harper 和 Hallett^[88]建立了使用 cohesive 单元模拟疲劳分层的 FEM。为了描述单元损伤状态,模型引入了静力损伤变量 d_s 和疲劳损伤变量 d_f ,当单元的合损伤变量(d_s 与 d_f 之和)达到 1,则认为单元发生失效。该模型的局限是:需要在计算区域具备较高的应力集中现象,并明确初始的裂纹长度。Kawashita 和 Hallett^[89]在 Harper 模型的基础上提出了一种可以不需要预先明确 cohesive

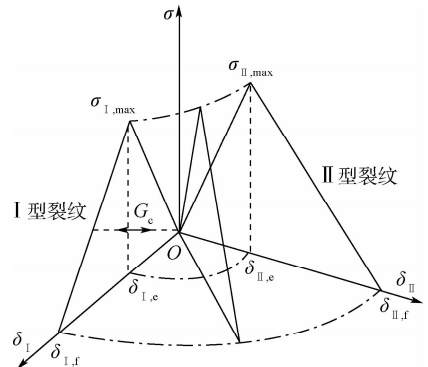


图 8 cohesive 单元 I 型、II 型和混合型裂纹加载对应的应力位移曲线

Fig. 8 Stress displacement curves for cohesive elements under I, II and hybrid crack loading modes

区域初始裂纹长度的模型,这一特点使得其可应用于分层形状及载荷条件较复杂的情况,同时也适用于三维几何模型。模型预测的分层扩展速率表现出较弱的网格敏感性,并且在较粗的网格下预测结果偏保守。

Tao 等^[90]在 cohesive 模型中引入了一个虚拟的损伤变量 v_{df} 以连续地追踪裂纹尖端的单元。该模型不需对界面单元进行性能衰减,从而可以使应变能释放率在计算过程中保持常值,基于该特点,还采用应变能释放率校正方法,减少了模型的网格敏感性。

3.2 渐进损伤模型总结

比起疲劳寿命模型和唯象模型,渐进损伤模型以其功能的全面和适用范围的宽泛倍受复合材料结构分析人员的青睐。然而,该类模型不可避免地也存在一些缺点。在静力分析中,复合材料的失效判据本身就还是一个不够完善的问题,而渐进损伤模型引入了一系列损伤起始、扩展判据和材料性能衰减准则,导致了模型复杂性增加。同时,模型中存在大量须要通过试验得到的拟合参数,使得模型的实施成本也更高。此外,由于复合材料疲劳损伤机理的研究还不够成熟,大部分渐进损伤模型仍然停留在通过建立数学公式去拟合试验数据或模拟材料行为的阶段。

在未来的发展中,渐进损伤模型的研究重点不应该是往上添加更多的依赖实验数据的参数,而是在保证一定精度的情况下尽可能简化模型,以适应工程人员对计算效率的要求。同时,从能量耗散、微观损伤和断裂力学等角度出发,明确复合材料疲劳损伤的物理机制并与宏观材料性能建立联系,将会是此类模型的另一大发展趋势。

4 结束语

随着复合材料疲劳理论的发展,大量用于预测寿命和损伤分析的疲劳性能分析模型被相继提出。目前为止,典型的疲劳性能分析模型可以分为疲劳寿命模型、唯象模型和渐进损伤模型三大类。

疲劳寿命模型是最早发展起来的疲劳性能分析模型。由于不需要理解损伤的物理机制,疲劳寿命模型是3类模型中最简单的一种,被广泛地应用于工程的结构寿命估算。然而,这些模型通常需要针对特定的案例进行校准,故其适用范围比较有限。相比之下,唯象模型借助剩余强度/刚度更加明确地表征了结构疲劳损伤累积特征,并可以通过简单的数学方法建立,因此该类模型在

工程实际中被大量应用于疲劳寿命和剩余力学性能的预测中。渐进损伤模型除了可以进行结构疲劳寿命与剩余力学性能的预测外,还可对疲劳损伤进行分类定量的描述,这是疲劳寿命模型和唯象模型所不能实现的。然而无论是宏观还是细观尺度下的模型,渐进损伤模型的复杂性和高成本往往不太能够被工程实际所接受,更多的用于对结构损伤机理分析的科研领域。

国内外学者尝试了不同的方法来提高疲劳寿命模型的性能。其中 AI 技术正在使疲劳寿命模型突破其适用范围狭窄的局限。AI 技术是一种没有材料和载荷限制的高效方法,因此结合 AI 技术的疲劳寿命预测将成为该类模型的发展趋势。制约唯象模型应用的一大重要因素在于其只能预测简单载荷作用下复合材料的疲劳寿命,无法考虑真实结构的复杂应力状态,因此未来会朝着解决这一问题的方向发展,当然,模拟真实结构复杂载荷情况的试验条件也须同步发展。渐进损伤模型有待进一步工程化,使用尽可能简单的模型和尽可能少量的试验数据就能分析不同材料及不同工况下的疲劳性能,将会是渐进损伤模型未来的发展目标。渐进损伤模型具有损伤分析的独有优势,需要进一步考虑损伤的物理机制,并从多尺度、多学科层面对模型进行完善。

复合材料疲劳性能分析模型的发展离不开试验技术和基础理论的进步。现有的大部分复合材料疲劳性能分析模型已经被试验证实具有较理想的准确性,而未来的疲劳性能分析模型将把降低实施成本和提高通用性作为发展重点。把疲劳性能分析模型应用于复杂工况(如随机载荷谱、湿热环境条件等)和复杂结构的情况,是复合材料疲劳性能分析模型研究的最主要目标。

参考文献 (References)

- [1] ROUCHON J, BOS M. Fatigue and damage tolerance evaluation of structures: The composite materials response; NLR-TP-2009-221[R]. Amsterdam: NLR, 2009.
- [2] BATHIAS C. An engineering point of view about fatigue of polymer matrix composite materials[J]. International Journal of Fatigue, 2006, 28(10): 1094-1099.
- [3] ANSARI M T A, SINGH K K, AZAM M S. Fatigue damage analysis of fiber-reinforced polymer composites—A review[J]. Journal of Reinforced Plastics and Composites, 2018, 37(9): 636-654.
- [4] BROD M, JUST G, DEAN A, et al. Numerical modelling and simulation of fatigue damage in carbon fibre reinforced plastics at different stress ratios[J]. Thin Walled Structures, 2019, 139: 219-231.

- [5] VASSILOPOULOS A P. Fatigue life modeling and prediction methods for composite materials and structures—Past, present, and future prospects[M] // VASSILOPOULOS A P. Fatigue life prediction of composites and composite structures. Amsterdam: Elsevier, 2020: 1-43.
- [6] VASSILOPOULOS A P. The history of fiber-reinforced polymer composite laminate fatigue[J]. International Journal of Fatigue, 2020, 134: 105512.
- [7] BAK B L, SARRADO C, TURON A, et al. Delamination under fatigue loads in composite laminates: A review on the observed phenomenology and computational methods[J]. Applied Mechanics Reviews, 2014, 66(6): 060803.
- [8] PASCOE J A, ALDERLIESTEN R C, BENEDICTUS R. Methods for the prediction of fatigue delamination growth in composites and adhesive bonds—A critical review[J]. Engineering Fracture Mechanics, 2013, 112-113: 72-96.
- [9] 翟洪军, 姚卫星. 纤维增强树脂基复合材料的疲劳剩余寿命研究进展[J]. 力学进展, 2002, 32(1): 69-80.
- ZHAI H J, YAO W X. A survey on stiffness reduction models of fiber reinforced plastics under cyclic loading[J]. Advances in Mechanics, 2002, 32(1): 69-80 (in Chinese).
- [10] HASHIN Z, ROTE M A. A fatigue failure criterion for fiber reinforced materials[J]. Journal of Composite Materials, 1973, 7(4): 448-464.
- [11] REIFSNIDER K, GAO Z. A micromechanics model for composites under fatigue loading[J]. International Journal of Fatigue, 1991, 13(2): 149-156.
- [12] PHILIPPIDIS T P, VASSILOPOULOS A P. Fatigue strength prediction under multiaxial stress[J]. Journal of Composite Materials, 1999, 33(17): 1578-1599.
- [13] FAWAZ Z, ELLYIN F. Fatigue failure model for fibre-reinforced materials under general loading conditions[J]. Journal of Composite Materials, 1994, 28(15): 1432-1451.
- [14] BRØNDSTED P, ANDERSEN S, LILHOLT H. Fatigue damage accumulation and lifetime prediction of GFRP materials under block loading and stochastic loading[C] // Proceedings of the 18th Risø International Symposium on Materials Science, 1997: 269-278.
- [15] VASSILOPOULOS A P, MANSHADI B D, KELLER T. Influence of the constant life diagram formulation on the fatigue life prediction of composite materials[J]. International Journal of Fatigue, 2010, 32(4): 659-669.
- [16] PHILIPPIDIS T P, VASSILOPOULOS A P. Life prediction methodology for GFRP laminates under spectrum loading[J]. Composites Part A: Applied Science and Manufacturing, 2004, 35(6): 657-666.
- [17] HARRIS B. A parametric constant-life model for prediction of the fatigue lives of fibre-reinforced plastics[M] // HARRIS B. Fatigue in composites. Amsterdam: Elsevier, 2003: 546-568.
- [18] KAWAI M, KOIZUMI M. Nonlinear constant fatigue life diagrams for carbon/epoxy laminates at room temperature[J]. Composites Part A: Applied Science and Manufacturing, 2007, 38(11): 2342-2353.
- [19] BOERSTRA G. The multislope model: A new description for the fatigue strength of glass fibre reinforced plastic[J]. International Journal of Fatigue, 2007, 29(8): 1571-1576.
- [20] KASSAPOGLOU C. Fatigue life prediction of composite structures under constant amplitude loading[J]. Journal of Composite Materials, 2007, 41(22): 2737-2754.
- [21] PARK T, KIM M, JANG B, et al. A nonlinear constant life model for the fatigue life prediction of composite structures[J]. Advanced Composite Materials, 2014, 23(4): 337-350.
- [22] VASSILOPOULOS A P, GEORGOPOULOS E F, KELLER T. Comparison of genetic programming with conventional methods for fatigue life modeling of FRP composite materials[J]. International Journal of Fatigue, 2008, 30(9): 1634-1645.
- [23] VASSILOPOULOS A P, KELLER T. Modeling of the fatigue life of adhesively-bonded FRP joints with genetic programming[C] // Proceedings of the 17th International Conference on Composite Materials, 2009: 27-31.
- [24] ZHOU C, ZHOU S, PENG Y, et al. Study on the fatigue modeling of FRP composite materials[C] // Proceedings of the 5th International Conference on Energy and Environmental Protection, 2016: 763-767.
- [25] VASSILOPOULOS A P, GEORGOPOULOS E F, DIONYSOPOULOS V. Artificial neural networks in spectrum fatigue life prediction of composite materials[J]. International Journal of Fatigue, 2007, 29(1): 20-29.
- [26] AL-ASSADI M, EL KADI H A, DEIAB I M. Using artificial neural networks to predict the fatigue life of different composite materials including the stress ratio effect[J]. Applied Composite Materials, 2011, 18(4): 297-309.
- [27] AL-ASSADI M, EL KADI H A, DEIAB I M. Predicting the fatigue life of different composite materials using artificial neural networks[J]. Applied Composite Materials, 2010, 17(1): 1-14.
- [28] VASSILOPOULOS A P, BEDI R. Adaptive neuro-fuzzy inference system in modelling fatigue life of multidirectional composite laminates[J]. Computational Materials Science, 2008, 43(4): 1086-1093.
- [29] JARRAH M, AL-ASSAF Y, KADI H E. Neuro-fuzzy modeling of fatigue life prediction of unidirectional glass fiber/epoxy composite laminates[J]. Journal of Composite Materials, 2002, 36(6): 685-700.
- [30] VASSILOPOULOS A P. Residual strength fatigue theories for composite materials[M] // VASSILOPOULOS A P. Fatigue life prediction of composites and composite structures. Amsterdam: Elsevier, 2020: 79-101.
- [31] D'AMORE A, GIORGIO M, GRASSIA L. Modeling the residual strength of carbon fiber reinforced composites subjected to cyclic loading[J]. International Journal of Fatigue, 2015, 78: 31-37.
- [32] BROUTMAN L, SAHU S. A new theory to predict cumulative fatigue damage in fiberglass reinforced plastics[C] // Composite materials: Testing and Design (Second Conference), 1972: 170-188.
- [33] HALPIN J C, JERINA K L, JOHNSON T A. Characterization of composites for the purpose of reliability evaluation; STP36479S [R]. West Conshohocken: ASTM International, 1973.
- [34] YANG J N, LIU M. Residual strength degradation model and theory of periodic proof tests for graphite/epoxy laminates[J].

- Journal of Composite Materials, 1977, 11(2): 176-203.
- [35] YANG J, JONES D. Fatigue of graphite/epoxy [0/90/45/-45]_s laminates under dual stress levels[J]. Journal of Composites, Technology and Research, 1982, 4(3): 63-70.
- [36] DICK T, JAR P Y, CHENG J J. Prediction of fatigue resistance of short-fibre-reinforced polymers[J]. International Journal of Fatigue, 2009, 31(2): 284-291.
- [37] D'AMORE A, CAPRINO G, STUPAK P, et al. Effect of stress ratio on the flexural fatigue behaviour of continuous strand mat reinforced plastics[J]. Science and Engineering of Composite Materials, 1996, 5(1): 1-8.
- [38] DIAO X, YE L, MAI Y W. A statistical model of residual strength and fatigue life of composite laminates[J]. Composites Science and Technology, 1995, 54(3): 329-336.
- [39] YAO W, HIMMEL N. A new cumulative fatigue damage model for fibre-reinforced plastics[J]. Composites Science and Technology, 2000, 60(1): 59-64.
- [40] HOSOI A, KAWADA H, YOSHINO H. Fatigue characteristics of quasi-isotropic CFRP laminates subjected to variable amplitude cyclic two-stage loading[J]. International Journal of Fatigue, 2006, 28(10): 1284-1289.
- [41] D'AMORE A, GRASSIA L. Constitutive law describing the strength degradation kinetics of fibre-reinforced composites subjected to constant amplitude cyclic loading[J]. Mechanics of Time-Dependent Materials, 2016, 20(1): 1-12.
- [42] JOHN T, WARUNA S. Determining the fatigue life of composites aircraft structures using life and load-enhancement factors: DOT/FAA/AR-10/6[R]. Washington, D. C.: FAA, 2011.
- [43] SENDECKYJ G P. Life prediction for resin-matrix composite materials[M]. Amsterdam: Elsevier, 1991: 431-483.
- [44] KASSAPOGLOU C. Fatigue model for composites based on the cycle-by-cycle probability of failure: Implications and applications[J]. Journal of Composite Materials, 2011, 45(3): 261-277.
- [45] SENDECKYJ G P. Fitting models to composite materials fatigue data; STP29314S[R]. West Conshohocken: ASTM International, 1981.
- [46] HWANG W, HAN K S. Fatigue of composites-fatigue modulus concept and life prediction[J]. Journal of Composite Materials, 1986, 20(2): 154-165.
- [47] ECHTERMEYER A T, ENGH B, BUENE L. Lifetime and Young's modulus changes of glass/phenolic and glass/polyester composites under fatigue[J]. Composites, 1995, 26(1): 10-16.
- [48] SIDOROFF F, SUBAGIO B. Fatigue damage modelling of composite materials from bending tests[C]//6th International Conference on Composite Materials (ICCM-VI) & Second European Conference on Composite Materials (ECCM-II), 1987, 4: 32-39.
- [49] VAN PAEPEGEM W, DEGRIECK J. Simulating damage and permanent strain in composites under in-plane fatigue loading[J]. Computers & Structures, 2005, 83(23-24): 1930-1942.
- [50] VIEILLEVIGNE S, JEULIN D, RENARD J, et al. Modelling of the fatigue behaviour of a unidirectional glass epoxy composite submitted to fatigue loadings[C]//Proceedings of the International Conference on Fatigue of Composites, 1997: 424-430.
- [51] KAWAI M, YAJIMA S, HACHINOHE A, et al. High-temperature off-axis fatigue behaviour of unidirectional carbon-fibre-reinforced composites with different resin matrices[J]. Composites Science and Technology, 2001, 61(9): 1285-1302.
- [52] KAWAI M, YAJIMA S, HACHINOHE A, et al. Off-axis fatigue behavior of unidirectional carbon fiber-reinforced composites at room and high temperatures[J]. Journal of Composite Materials, 2001, 35(7): 545-576.
- [53] KAWAI M, TANIGUCHI T. Off-axis fatigue behavior of plain weave carbon/epoxy fabric laminates at room and high temperatures and its mechanical modeling[J]. Composites Part A: Applied Science and Manufacturing, 2006, 37(2): 243-256.
- [54] YANG J, JONES D, YANG S, et al. A stiffness degradation model for graphite/epoxy laminates[J]. Journal of Composite Materials, 1990, 24(7): 753-769.
- [55] WU W F, LEE L, CHOI S T. A study of fatigue damage and fatigue life of composite laminates[J]. Journal of Composite Materials, 1996, 30(1): 123-137.
- [56] 程小全, 邹健, 杨琨, 等. 缝合复合材料层合板疲劳寿命预测[J]. 失效分析与预防, 2008, 3(4): 8-11.
CHENG X Q, ZOU J, YANG K, et al. Fatigue life prediction model for stitched composite laminates[J]. Failure Analysis and Prevention, 2008, 3(4): 8-11 (in Chinese).
- [57] 王丹勇, 温卫东. 复合材料单钉接头疲劳累积损伤破坏分析[J]. 复合材料学报, 2008, 25(1): 173-179.
WANG D Y, WEN W D. Fatigue progressive damage analysis of single fastener joints in composite laminates[J]. Acta Materiae Compositae Sinica, 2008, 25(1): 173-179 (in Chinese).
- [58] DORMOHAMMDI S, GODINES C, ABDI F, et al. Damage-tolerant composite design principles for aircraft components under fatigue service loading using multi-scale progressive failure analysis[J]. Journal of Composite Materials, 2017, 51(15): 2181-2202.
- [59] WHITWORTH H. A stiffness degradation model for composite laminates under fatigue loading[J]. Composite Structures, 1997, 40(2): 95-101.
- [60] VAN PAEPEGEM W, DEGRIECK J. A new coupled approach of residual stiffness and strength for fatigue of fibre-reinforced composites[J]. International Journal of Fatigue, 2002, 24(7): 747-762.
- [61] WU F, YAO W. A fatigue damage model of composite materials[J]. International Journal of Fatigue, 2010, 32(1): 134-138.
- [62] SHIRI S, YAZDANI M, POURGOL-MOHAMMAD M. A fatigue damage accumulation model based on stiffness degradation of composite materials[J]. Materials & Design, 2015, 88: 1290-1295.
- [63] PIDAPARTI R M, VOGT A. Experimental investigation of Poisson's ratio as a damage parameter for bone fatigue[J]. Journal of Biomedical Materials Research, 2002, 59(2): 282-287.
- [64] LECOMPTE D, SMITS A, SOL H, et al. Mixed numerical-experimental technique for orthotropic parameter identification using biaxial tensile tests on cruciform specimens[J]. International Journal of Solids and Structures, 2007, 44(5): 1643-1656.
- [65] VAN PAEPEGEM W, DE BAERE I, DEGRIECK J. Modelling

- the nonlinear shear stress-strain response of glass fibre-reinforced composites. Part I: Experimental results [J]. *Composites Science and Technology*, 2006, 66(10): 1455-1464.
- [66] BERGMANN H, PRINZ R. Fatigue life estimation of graphite/epoxy laminates under consideration of delamination growth [J]. *International Journal for Numerical Methods in Engineering*, 1989, 27(2): 323-341.
- [67] DAHLEN C, SPRINGER G S. Delamination growth in composites under cyclic loads [J]. *Journal of Composite Materials*, 1994, 28(8): 732-781.
- [68] FENG X, GILCHRIST M, KINLOCH A, et al. Development of a method for predicting the fatigue life of CFRP components [C] // *Proceedings of the International Conference on Fatigue of Composites*, 1997: 407-414.
- [69] SHOKRIEH M M, LESSARD L B. Progressive fatigue damage modeling of composite materials, Part I: Modeling [J]. *Journal of Composite Materials*, 2000, 34(13): 1056-1080.
- [70] SHOKRIEH M M, LESSARD L B. Progressive fatigue damage modeling of composite materials, Part II: Material characterization and model verification [J]. *Journal of Composite Materials*, 2000, 34(13): 1081-1116.
- [71] PAPANIKOS P, TSERPES K, PANTELAKIS S. Modelling of fatigue damage progression and life of CFRP laminates [J]. *Fatigue & Fracture of Engineering Materials & Structures*, 2003, 26(1): 37-47.
- [72] YE L. Role of matrix resin in delamination onset and growth in composite laminates [J]. *Composites Science and Technology*, 1988, 33(4): 257-277.
- [73] PASSIPOULARIDIS V, PHILIPPIDIS T, BRONSTED P. Fatigue life prediction in composites using progressive damage modelling under block and spectrum loading [J]. *International Journal of Fatigue*, 2011, 33(2): 132-144.
- [74] PUCK A, SCHURMANN H. Failure analysis of FRP laminates by means of physically based phenomenological models [J]. *Composites Science and Technology*, 2002, 62(12-13): 1633-1662.
- [75] ELIOPOULOS E N, PHILIPPIDIS T P. A progressive damage simulation algorithm for GFRP composites under cyclic loading. Part I: Material constitutive model [J]. *Composites Science and Technology*, 2011, 71(5): 742-749.
- [76] ELIOPOULOS E N, PHILIPPIDIS T P. A progressive damage simulation algorithm for GFRP composites under cyclic loading. Part II: FE implementation and model validation [J]. *Composites Science and Technology*, 2011, 71(5): 750-757.
- [77] ZHAO L, SHAN M, HONG H, et al. A residual strain model for progressive fatigue damage analysis of composite structures [J]. *Composite Structures*, 2017, 169: 69-78.
- [78] SHAN M, ZHAO L, HONG H, et al. A progressive fatigue damage model for composite structures in hygrothermal environments [J]. *International Journal of Fatigue*, 2018, 111: 299-307.
- [79] SAYYIDMOUSAVI A, BOUGHERARA H, FAWAZ Z. A multi-scale approach for fatigue life prediction of polymer matrix composite laminates [J]. *Journal of Reinforced Plastics and Composites*, 2015, 34(13): 1099-1109.
- [80] LI W, CAI H, LI C, et al. Micro-mechanics of failure for fatigue strength prediction of bolted joint structures of carbon fiber reinforced polymer composite [J]. *Composite Structures*, 2015, 124: 345-356.
- [81] HOSSEINI KORDKHEILI S A, TOOZANDEHJANI H, SOLTANI Z. A progressive multi-scale fatigue model for life prediction of laminated composites [J]. *Journal of Composite Materials*, 2017, 51(20): 2949-2960.
- [82] TALREJA R. Damage and fatigue in composites—A personal account [J]. *Composites Science and Technology*, 2008, 68(13): 2585-2591.
- [83] SINGH C V, TALREJA R. A synergistic damage mechanics approach for composite laminates with matrix cracks in multiple orientations [J]. *Mechanics of Materials*, 2009, 41(8): 954-968.
- [84] SHEN H, YAO W, WU Y. Synergistic damage mechanic model for stiffness properties of early fatigue damage in composite laminates [C] // *17th International Colloquium on Mechanical Fatigue of Metals*, 2014, 74: 199-209.
- [85] CHEN X M, SUN X S, CHEN P H, et al. A delamination failure criterion considering the effects of through-thickness compression on the interlaminar shear failure of composite laminates [J]. *Composite Structures*, 2020, 241: 112121.
- [86] NIXON-PEARSON O, HALLETT S, HARPER P, et al. Damage development in open-hole composite specimens in fatigue. Part 2: Numerical modelling [J]. *Composite Structures*, 2013, 106: 890-898.
- [87] SIMÓN I, BANKS-SILLS L, FOURMAN V. Mode I delamination propagation and *R*-ratio effects in woven composite DCB specimens for a multi-directional layup [J]. *International Journal of Fatigue*, 2017, 96: 237-251.
- [88] HARPER P W, HALLETT S R. A fatigue degradation law for cohesive interface elements—Development and application to composite materials [J]. *International Journal of Fatigue*, 2010, 32(11): 1774-1787.
- [89] KAWASHITA L F, HALLETT S R. A crack tip tracking algorithm for cohesive interface element analysis of fatigue delamination propagation in composite materials [J]. *International Journal of Solids and Structures*, 2012, 49(21): 2898-2913.
- [90] TAO C, QIU J, YAO W, et al. A novel method for fatigue delamination simulation in composite laminates [J]. *Composites Science and Technology*, 2016, 128: 104-115.

Research development of fatigue life prediction and damage analysis model of fiber-reinforced composite

CHENG Xiaoquan^{*}, DU Xiaoyuan

(School of Aeronautic Science and Engineering, Beihang University, Beijing 100083, China)

Abstract: With the wide application of fiber-reinforced composite primary structures in many fields, fatigue has become a critical problem in composite structural design and analysis. Based on the development of fatigue theory of composite materials, many theoretical methods and numerical models for life prediction and damage analysis have been proposed by scholars. The current composite fatigue performance analysis models can be classified as fatigue life model, phenomenological model and progressive damage model. The development of these three kinds of models is reviewed and compared and their advantages and disadvantages are analyzed. The theory of fatigue life model is relatively simple, and the model is suitable for the life estimation of engineering structures. The mathematical relationship between residual strength/stiffness and fatigue cycle number is established in phenomenological model, which can predict the structure fatigue life and material residual mechanical properties. The progressive damage model can not only predict the structure fatigue life and material residual mechanical properties, but also analyze the fatigue damage mechanism of structures. Finally, the development trends of these fatigue performance analysis models are discussed. It is pointed out that reducing the implementation cost and improving the generality are important problems of these models.

Keywords: composites; fatigue; life prediction; phenomenological models; progressive damage

http://bhxb.buaa.edu.cn jbuua@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2020.0223

基于 IMU 与 UKF 的船舶升沉运动信息测量方法



卢道华^{1,2,*}, 付怀达¹, 王佳¹, 蔡雅轩¹, 宋世磊¹

(1. 江苏科技大学 机械工程学院, 镇江 212003; 2. 江苏科技大学 海洋装备研究院, 镇江 212003)

摘 要: 为获取实时、精确的船舶升沉运动信息, 根据船舶升沉运动模型和频谱分析方法, 建立描述惯性测量单元(IMU)的加速度测量信息与船舶升沉运动状态量关系的解析模型。基于无迹卡尔曼滤波(UKF)算法非线性滤波的特点, 进行升沉运动滤波解算。通过仿真分析证明了所提方法在船舶升沉运动测量中的有效性。利用三自由度平台升沉运动测量实验验证, 结果表明, 同一模型下, 相比于扩展卡尔曼滤波(EKF)算法的解算结果, 所提方法具有更快的收敛速度和更高的测量精度; 对船舶升沉位移的估计精度达到最大升沉幅值的 5%, 可以得到精确、无时延的船舶升沉运动信息。

关键词: 升沉测量; 惯性测量单元(IMU); 无迹卡尔曼滤波(UKF); 频谱分析; 平台实验

中图分类号: U666.1

文献标志码: A

文章编号: 1001-5965(2021)07-1323-09

许多海上作业, 如航行补给、舰载机的起降、海底测绘和海上起重机工作等, 都需要对由海浪、海风和洋流等复杂海洋环境因素引起的船舶升沉运动进行补偿, 这就需要实时、精确的船舶升沉运动信息^[1-4]。

卡尔曼滤波器(KF)广泛应用于最优估计领域中。经典的卡尔曼滤波针对线性高斯系统对目标状态做出最优估计, 在一定条件下得到准确的估计状态^[5]。扩展卡尔曼滤波(EKF)针对非线性高斯系统, 对其非线性部分通过一阶泰勒展开使其模型线性化, 并且仅在特定范围的初始条件下收敛^[5], 而无迹卡尔曼滤波(UKF)通过 UT(Unscented Transformation)的非线性变换方法直接进行非线性函数的状态及其方差阵传播, 避免了非线性函数线性化近似过程中复杂的 Jacobian 矩阵的求解, 适用于强非线性系统, 改善了状态估计的准确性和收敛性^[6-7]。利用惯性测量单元

(IMU)可以测量船舶的升沉加速度, 通过积分计算得到升沉速度和升沉位移。由于惯性器件误差和惯导系统高度通道不稳定等原因, 加速度积分误差会随时间不断增长。故使用惯性传感器测量船舶升沉运动信息通常需要外部传感器的辅助或者通过滤波估计的方式实现。Godhavn^[8]提出使用 GPS 实时动态差分技术(RTK)辅助加速度计进行升沉运动测量, 得到了较高的升沉运动精度, 但该方法需要基站辅助, 而在海洋上很难获得高精度的参考信号。升沉运动测量中, 由于所需要的升沉信号频率与船舶主动航行运动、积分漂移的频率不同, 可以对积分信号进行滤波或估计来得到有效的升沉运动信息。通过设计高通滤波器的方法进行船舶升沉运动测量, 往往需要解决滤波器相位超前问题。Hu 和 Tao^[9]提出通过设计一对级联高通和全通 IIR 滤波器以解决相位误差。黄卫权等^[2]提出基于带限傅里叶线性组合

收稿日期: 2020-05-27; 录用日期: 2020-08-21; 网络出版时间: 2020-09-16 09:00

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20200915.1501.006.html

基金项目: 国家重点研发计划(2018YFC0309100)

* 通信作者。E-mail: ludaohua_just@126.com

引用格式: 卢道华, 付怀达, 王佳, 等. 基于 IMU 与 UKF 的船舶升沉运动信息测量方法[J]. 北京航空航天大学学报, 2021, 47(7): 1323-1331. LU D H, FU H D, WANG J, et al. Measurement method of ship's heave motion information based on IMU and UKF algorithm[J]. Journal of Beijing University of Aeronautics and Astronautics, 2021, 47(7): 1323-1331 (in Chinese).

(BMFLC)算法对船舶升沉运动频率实时估计来调整滤波器结构。严恭敏等^[10]通过先设计低通滤波器,再采用互补的方法将其转换为高通滤波器,改善了高频段的相频特性,但在阻通过渡带的幅频特性不理想。袁书明等^[11]提出了自适应延时校正算法来解决高通滤波器相位超前问题。相比于通过设计高通滤波器来进行船舶升沉测量,在时域内设计的卡尔曼滤波器提供了另一种解决方案,该方法不存在相位问题且可以实时进行船舶升沉测量。Küchler等^[12]采用EKF对船舶升沉位移进行滤波估计,但由于忽略了船舶横摇和纵摇角对船舶升沉加速度计算的影响,并且相比于非线性滤波存在收敛慢和线性化误差的缺点,会造成测量误差。

针对上述问题,本文提出了基于IMU器件与UKF算法测量船舶升沉运动信息的方法。首先,分析了船舶升沉运动模型,根据姿态更新算法和比力方程计算得到船舶升沉加速度;其次,通过对升沉加速度进行频谱分析和频域峰值检测,得到升沉运动测量所需的频率信息;然后,结合升沉运动模型,建立船舶升沉运动测量的状态方程和量测方程,并基于UKF理论进行滤波解算;最后,通过仿真分析和平台实验对本文方法进行了验证。

1 船舶升沉运动测量原理

1.1 船舶升沉运动模型

船舶的升沉运动从短时间来看,就是船舶沿地理坐标系天向轴方向所做的周期运动。根据信号叠加原理,可以通过多个振幅不同、频率不同、初始相位不同的余弦波的叠加得到^[13-14]。将其表示为关于时间 t 的函数:

$$z(t) = \sum_{j=1}^{N_m} A_j \cos(\omega_j t + \varphi_j) = \sum_{j=1}^{N_m} z_j(t) \quad (1)$$

式中: A_j 、 ω_j 和 φ_j 分别为第 j 个余弦波分量的幅值、特征频率和初始相位; N_m 为拟合海浪运动主要频率分量的余弦波个数。

船舶升沉运动测量依赖于对式(1)中各参数的准确估计。船舶升沉速度和升沉加速度可以分别通过升沉位移的一阶导和二阶导得到。

升沉速度表示为

$$\dot{z}(t) = - \sum_{j=1}^{N_m} A_j \omega_j \sin(\omega_j t + \varphi_j) = \sum_{j=1}^{N_m} \dot{z}_j(t) \quad (2)$$

升沉加速度表示为

$$\ddot{z}(t) = - \sum_{j=1}^{N_m} A_j \omega_j^2 \cos(\omega_j t + \varphi_j) = \sum_{j=1}^{N_m} \ddot{z}_j(t) \quad (3)$$

由式(1)和式(3)可以得到,船舶升沉位移各

余弦分量幅值与升沉加速度各余弦分量幅值的关系为

$$A(\omega_j) = - \frac{\ddot{A}(\omega_j)}{\omega_j^2} \quad j = 1, 2, \dots, N_m \quad (4)$$

式中: $A(\omega_j)$ 为船舶升沉位移各余弦分量幅值; $\ddot{A}(\omega_j)$ 为船舶升沉加速度各余弦分量幅值。

同样可以得到,船舶升沉位移和升沉速度的各余弦分量初始相位与升沉加速度各余弦分量初始相位之间的关系为

$$\varphi(\omega_j) = \ddot{\varphi}(\omega_j) - \pi \quad j = 1, 2, \dots, N_m \quad (5)$$

$$\dot{\varphi}(\omega_j) = \ddot{\varphi}(\omega_j) - \frac{\pi}{2} \quad j = 1, 2, \dots, N_m \quad (6)$$

式中: $\varphi(\omega_j)$ 、 $\dot{\varphi}(\omega_j)$ 和 $\ddot{\varphi}(\omega_j)$ 分别为船舶升沉位移、升沉速度和升沉加速度的各余弦分量的相位。

1.2 基于IMU的船舶升沉信息测量

船舶升沉运动测量如图1所示。以船舶中心为船体坐标系的原点, y_b 轴沿船体纵轴指向船体前向, x_b 轴与 y_b 轴垂直沿船体横轴指向船体右舷, z_b 轴与前两轴垂直沿船体竖直向上。取船体坐标系 $ox_b y_b z_b$ 为载体坐标系(b 系)、地理坐标系 $ox_n y_n z_n$ 为导航坐标系(n 系),这里取“东北天”坐标系,原点与船体坐标系原点重合。本文假设IMU安装在船体中心位置,所包含的三轴加速度计和三轴陀螺仪可以分别测量船舶在船体坐标系下沿3个坐标轴运动的加速度信息和绕3个坐标轴旋转的角速度信息,通过对陀螺仪和加速度计信息处理得到船舶的升沉信息。

船舶升沉运动信息测量依赖于准确的水平姿态测量,考虑平台失准角误差,捷联惯导姿态矩阵更新表达式为

$$\mathbf{C}_{b(k)}^n = \mathbf{C}_{b(k-1)}^n [\mathbf{I} + (\Delta\theta_k + \boldsymbol{\phi}^b) \times] \quad (7)$$

式中: $\mathbf{C}_{b(k)}^n$ 为 k 时刻的姿态矩阵; $\mathbf{C}_{b(k-1)}^n$ 为 $k-1$ 时刻的姿态矩阵; \mathbf{I} 为单位矩阵; $\Delta\theta_k$ 为 $k-1$ 时刻到 k 时刻陀螺仪的等效旋转矢量输出; $\boldsymbol{\phi}^b$ 为平台失准角在载体坐标系下的投影; $(\cdot) \times$ 为反对称矩阵。

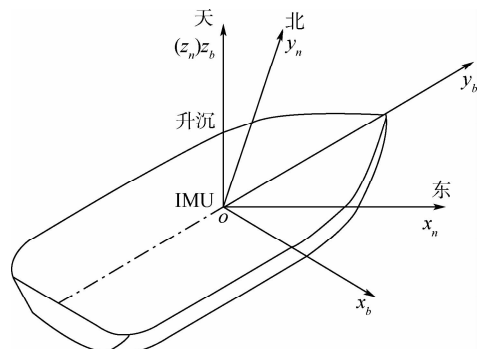


图1 船舶升沉运动测量原理

Fig.1 Schematic diagram of ship's heave motion measurement

船舶处于低加速度运动状态时,可以利用加速度计测量进行水平姿态对准得到初始姿态矩阵 $C_{b(0)}^n$ 。另外,平台水平失准角在载体坐标系的投影可以表示为^[15]

$$\phi^b = \frac{f_{sf}^b}{|f_{sf}^b|} \times C^T \quad (8)$$

式中: f_{sf}^b 为载体坐标系下的加速度计测量的比力信息; C 为姿态矩阵 C_b^n 的第 3 行向量。

捷联惯导系统的比力方程为

$$\dot{v}^n = C_b^n f_{sf}^b - (2\omega_{ie}^n + \omega_{en}^n) \times v^n + g^n \quad (9)$$

式中: v^n 为载体在地理坐标系下的速度; g^n 为重力加速度在地理坐标系下的投影; $(2\omega_{ie}^n + \omega_{en}^n) \times v^n$ 为地球自转和哥氏补偿项,可忽略不计。

式(9)简化为

$$\dot{v}^n = C_b^n f_{sf}^b + g^n \quad (10)$$

船舶升沉加速度可以表示为

$$a_z = \dot{v}^n(3) \quad (11)$$

式中: $\dot{v}^n(3)$ 为船舶加速度 \dot{v}^n 的天向分量,包含了加速度零偏和噪声等误差项。

对带有误差的升沉加速度信号 a_z 二次积分得到船舶的升沉位移估算,由于加速度信号存在误差,该积分值随着时间的推移会迅速地偏离船舶升沉位移的真实值。如图 2 所示,利用 IMU 采集波浪模拟平台升沉运动信息,通过计算得到导航坐标系下升沉加速度,经二次积分运算得到升沉位移曲线。从图 2 中可以看出,由于未经滤波处理,捷联解算得到的船舶升沉位移估算误差迅速增大。

为了抑制噪声并消除加速度零偏对船舶升沉运动信息测量的影响,通过对 IMU 输出信息进行捷联解算和滤波处理得到拟合升沉运动的主要余弦分量,从而得到船舶升沉位移信息。

船舶升沉运动估算算法流程如图 3 所示。首

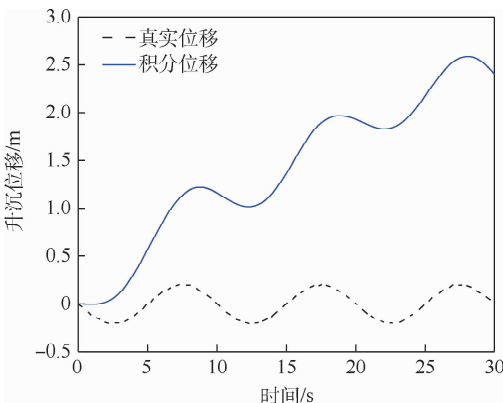


图 2 捷联解算船舶升沉位移

Fig. 2 Strapdown calculation of ship heave displacement

先,对 IMU 进行标定并做相应的误差补偿^[16],根据式(7)和式(8)对陀螺仪和加速度计数据处理得到系统的实时姿态矩阵,通过对载体坐标系下加速度信息进行坐标变换得到导航坐标系下的船舶升沉加速度 a_z 。然后,选择合适的窗口长度对升沉加速度信号 a_z 作快速傅里叶变换(FFT),保证可以检测到升沉运动波形主要余弦分量的最高频率,由此可以得到升沉加速度信号的幅度谱 $\tilde{A}(\omega)$ 和相位谱 $\tilde{\phi}(\omega)$,从而根据式(5)和式(6)得到升沉位移和升沉速度信息,使用峰值检测算法来检测升沉位移和升沉速度的幅度谱和相位谱。为了准确估计船舶幅度谱的局部最大值以确定模型参数 N_m ,同时计算得到 N_m 个频率分量余弦波所对应的幅值 A_j 、特征频率 ω_j 和相位 ϕ_j 。由于 FFT 得到的信息是在一段时间范围内各参数的均值、幅值和相位都不具有实时性,同时各余弦波的特征频率也会随时间发生变化,不能直接使用这些参数,但这些参数可以用于初始化卡尔曼滤波器。经过卡尔曼滤波算法多步滤波迭代,使误差逐渐缩小,才能获得最优参数,从而得到精确的船舶升沉运动信息。

通过仿真生成船舶升沉位移信息,并对其进行 FFT。图 4 为船舶升沉运动的幅度谱,峰值检测算法可以检测出幅度谱中的局部最大值。可以看出,波形中有 2 个主要的局部峰值,即 $N_m = 2$ 。峰值检测算法输出每个余弦波的幅值 A_j 和特征频

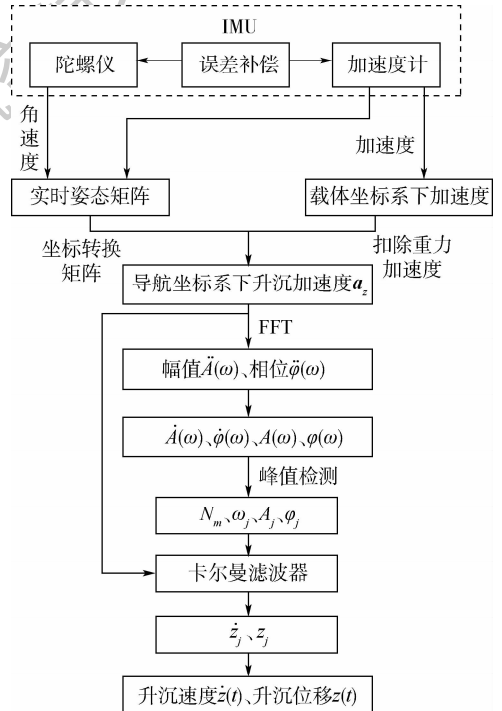


图 3 船舶升沉运动估算算法流程

Fig. 3 Ship heave motion estimation algorithm flowchart

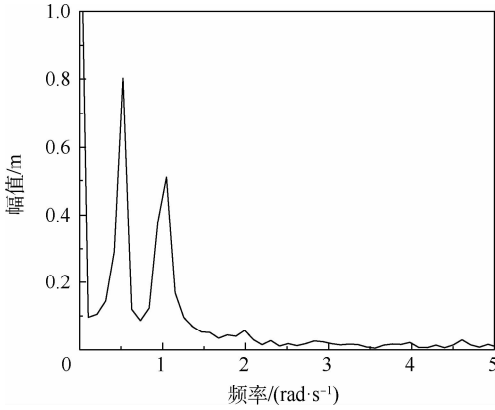


图4 模拟船舶升沉运动幅频曲线

Fig.4 Amplitude frequency curve of simulated ship heave motion

率 ω_j , 根据各波形的特征频率 ω_j 和 FFT 得到的相频谱得到对应的相位 φ_j 。

为了得到准确的船舶升沉运动模型, 必须去掉特征频率为零和特征频率接近的余弦波分量, 同时为了应对海浪时刻的变化, 必须在固定的时间间隔重新进行峰值检测。

2 基于 UKF 的船舶升沉运动测量

2.1 船舶升沉运动状态空间模型

式(1)为船舶升沉运动模型, 组成升沉运动的每个频率分量的余弦波都可以写成微分方程(12)的解:

$$\ddot{z}_j(t) + \omega_j^2 z_j(t) = 0 \quad t > t_0, j = 1, 2, \dots, N_m \quad (12)$$

初始时刻为 $z_j(t_0)$, $\dot{z}_j(t_0)$ 和 $\ddot{z}_j(t_0)$, 分别对应第 j 个频率分量余弦波的位移、速度和加速度的初始值。

由于船舶升沉运动模型中的频率信息需要通过固定时间间隔的 FFT 得到, 即组成船舶升沉运动周期分量的增加或减少都需要重新初始化系统, 系统模型需要根据海况变化进行调整。因为 FFT 得到的余弦波特征频率信息是在一段时间范围内的均值, 只能被用来初始化系统, 所以将特征频率 ω_j 作为系统变量进行估计, 建模为随机常值模型。

$$\dot{\omega}_j = 0 \quad (13)$$

建立系统的状态向量为

$$\mathbf{X}_j = [x_{1,j} \quad x_{2,j} \quad x_{3,j}]^T = [z_j \quad \dot{z}_j \quad \omega_j]^T \quad (14)$$

式中: \mathbf{X}_j 为第 j 个频率分量的系统状态量, 其成分量为该特征频率余弦波的位移、速度和特征频率, 初值由 FFT 得到的参数计算得到。

根据式(4)和式(10), 得到系统的状态空间模型为

$$\begin{cases} \dot{\mathbf{X}}_j = [x_{2,j} & -x_{3,j}^2 x_{1,j} & 0]^T \\ \mathbf{Y}_j = -x_{3,j}^2 x_{1,j} \end{cases} \quad j = 1, 2, \dots, N_m \quad (15)$$

式中: \mathbf{Y}_j 对应第 j 个频率分量余弦波的升沉加速度。

2.2 船舶升沉运动测量的状态方程

根据 IMU 输出计算得到的升沉加速度为各分量余弦波加速度的和, 同时考虑加速度测量的零偏和噪声, 得到

$$\mathbf{a}_z = \sum_{j=1}^{N_m} \ddot{z}_j(t) + b_z + \zeta_z \quad (16)$$

式中: b_z 为升沉加速度零偏, 通常由加速度计零偏造成; ζ_z 为升沉加速度测量噪声。

为了消除升沉加速度零偏对升沉运动信息测量的影响, 需要将零偏 b_z 作为系统变量进行准确的估计。结合式(14)得到最终的系统状态向量为

$$\mathbf{X}_k = [\mathbf{X}_{1,k} \quad \mathbf{X}_{2,k} \quad \dots \quad \mathbf{X}_{N_m,k} \quad b_{z,k}]^T \quad (17)$$

式中:

$$\mathbf{X}_{j,k} = [z_{j,k} \quad \dot{z}_{j,k} \quad \omega_{j,k}]^T \quad j = 1, 2, \dots, N_m \quad (18)$$

经离散化处理, 得到系统的状态方程为

$$\mathbf{X}_k = \begin{bmatrix} \mathbf{A}_1 & 0 & 0 & 0 & 0 \\ 0 & \mathbf{A}_2 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \mathbf{A}_{N_m} & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \mathbf{X}_{k-1} + \mathbf{w}_k \quad (19)$$

式中:

$$\mathbf{A}_j = \begin{bmatrix} 1 & T & 0 \\ -\omega_{j,k}^2 T & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad j = 1, 2, \dots, N_m \quad (20)$$

其中: T 为状态更新时间间隔; \mathbf{w}_k 为系统的过程噪声, 系统协方差矩阵为 $\mathbf{Q}_k = E[\mathbf{w}_k \mathbf{w}_k^T]$, 由各频率组成部分的系统变量的动力学模型确定。

2.3 船舶升沉运动测量的观测方程

由式(15)得到系统的观测方程为

$$\mathbf{Y}_k = \mathbf{a}_{z,k} \quad (21)$$

式中: $\mathbf{a}_{z,k}$ 为 k 时刻船舶在地理坐标系下的升沉加速度。

根据式(15)和式(16), 式(21)经离散化处理, 系统的观测方程为

$$\mathbf{Y}_k = [1 \quad 1 \quad \dots \quad 1 \quad 1] \begin{bmatrix} \mathbf{C}_1 \mathbf{X}_{1,k} \\ \mathbf{C}_2 \mathbf{X}_{2,k} \\ \vdots \\ \mathbf{C}_{N_m} \mathbf{X}_{N_m,k} \\ b_{z,k} \end{bmatrix} + \zeta_{z,k} \quad (22)$$

式中:

$$C_j = [-\omega_{j,k}^2 \quad 0 \quad 0] \quad (23)$$

量测噪声方差矩阵为 $R_k = \zeta_{z,k}^2$, 由升沉加速度测量噪声确定。

2.4 无迹卡尔曼滤波

2.4.1 无迹卡尔曼滤波模型

$$\begin{cases} X_k = f(X_{k-1}) + w_k \\ Y_k = h(X_k) + v_k \end{cases} \quad (24)$$

式中: $X_k \in \mathbf{R}^n$ 为 k 时刻系统的状态向量; $Y_k \in \mathbf{R}^m$ 为 k 时刻系统的观测向量; $f(\cdot)$ 为实的 n 维向量函数; $h(\cdot)$ 为实的 m 维向量函数; $f(\cdot)$ 和 $h(\cdot)$ 对其自变量而言都是非线性的; $w_k \in \mathbf{R}^n$ 为系统的过程噪声, 常设为协方差 $Q_k \geq 0$ 的零均值高斯白噪声; $v_k \in \mathbf{R}^m$ 为系统的量测噪声, 常设为方差 $R_k > 0$ 的零均值高斯白噪声。

2.4.2 无迹卡尔曼滤波器设计

船舶升沉运动模型属于非线性模型, 拟合船舶升沉运动的主要余弦分量是时刻变化的, 而 FFT 得到的各余弦运动的频率、幅值和相位都不具有实时性, 直接使用这些参数进行拟合是无意义的。UKF 适用于非线性系统, 相比于 EKF, 避免了非线性方程线性化过程产生的误差, 对模型变量估计的快速性和准确性是拟合船舶升沉运动的重要前提。

对加速度计测量得到的升沉运动加速度数据作 FFT, 计算得到滤波器的状态初始值 X_0 , 设定过程噪声协方差 Q_k 、量测噪声方差 R_k 和初始协方差矩阵 P_0 。式(19)和式(22)分别为系统的状态方程和量测方程, UKF 推导过程如下^[17]:

1) 给定状态估计 \hat{X}_{k-1} 及协方差矩阵 \hat{P}_{k-1} , 选取 sigma 点 χ_k 为

$$\chi_{j,k-1} = \begin{cases} \hat{X}_{k-1} & j = 0 \\ \hat{X}_{k-1} + \alpha \left(\sqrt{n\hat{P}_{k-1}} \right)_j & j = 1, 2, \dots, n \\ \hat{X}_{k-1} - \alpha \left(\sqrt{n\hat{P}_{k-1}} \right)_j & j = n+1, n+2, \dots, 2n \end{cases} \quad (25)$$

式中: n 为状态向量的维数; α 用于调节 sigma 点在均值 \hat{X}_{k-1} 附近的分布, 值越小则分布范围越小, 以避免方程非线性严重时采样点的非局部性影响, 取值范围为 $10^{-4} \sim 1$; $\left(\sqrt{n\hat{P}_{k-1}} \right)_j$ 为 $n\hat{P}_{k-1}$ 均方根矩阵的第 j 列。

2) 时间更新。

$$\begin{cases} \chi_{j,k/k-1}^* = f(\chi_{j,k-1}^*) \\ \hat{X}_{k/k-1} = \sum_{j=0}^{2n} W_j^m \chi_{j,k/k-1}^* \\ \hat{P}_{x,k/k-1} = \sum_{j=0}^{2n} W_j^c (\chi_{j,k/k-1}^* - \hat{X}_{k/k-1}) \cdot (\chi_{j,k/k-1}^* - \hat{X}_{k/k-1})^T + Q_k \\ \eta_{j,k/k-1} = h(\chi_{j,k/k-1}^*) \\ \hat{Y}_{k/k-1} = \sum_{j=0}^{2n} W_j^m \eta_{j,k/k-1} \end{cases} \quad (26)$$

式中:

$$\begin{cases} W_j^m = 1 - \frac{1}{\alpha^2}, W_j^c = 2 - \frac{1}{\alpha^2} - \alpha^2 + \beta, j = 0 \\ W_j^m = W_j^c = \frac{1}{2n\alpha^2}, j = 1, 2, \dots, 2n \end{cases} \quad (27)$$

其中: β 为先验知识常数, 高斯分布时最优值取 2。

$$\chi_{j,k/k-1} = \begin{cases} \hat{X}_{k/k-1} & j = 0 \\ \hat{X}_{k/k-1} + \alpha \left(\sqrt{(n + \kappa)\hat{P}_{k/k-1}} \right)_j & j = 1, 2, \dots, n \\ \hat{X}_{k/k-1} - \alpha \left(\sqrt{(n + \kappa)\hat{P}_{k/k-1}} \right)_j & j = n+1, n+2, \dots, 2n \end{cases} \quad (28)$$

式中: κ 为比例因子, 状态估计时通常设为 0;

$\left(\sqrt{n\hat{P}_{k/k-1}} \right)_j$ 为 $n\hat{P}_{k/k-1}$ 均方根矩阵的第 j 列。

3) 量测更新。

$$\begin{cases} P_{y,k/k-1} = \sum_{j=0}^{2n} W_j^c (\eta_{j,k/k-1} - \hat{Y}_{k/k-1}) \cdot (\eta_{j,k/k-1} - \hat{Y}_{k/k-1})^T + R_k \\ \hat{P}_{xy,k/k-1} = \sum_{j=0}^{2n} \hat{W}_j^c (\chi_{j,k/k-1} - \hat{X}_{k/k-1}) \cdot (\eta_{j,k/k-1} - \hat{Y}_{k/k-1})^T \\ K_k = \hat{P}_{xy,k/k-1} P_{y,k/k-1}^{-1} \\ \hat{X}_k = \hat{X}_{k/k-1} + K_k (Y_k - \hat{Y}_{k/k-1}) \\ P_k = P_{k/k-1} - K_k P_{y,k/k-1} K_k^T \end{cases} \quad (29)$$

式中: \hat{X}_k 和 P_k 分别为系统的状态滤波估计和对应的协方差矩阵。

3 仿真实验

基于 IMU 和 UKF 算法测量船舶升沉运动信

息,通过滤波估计可以得到组成升沉运动的主要余弦分量,叠加后得到准确的升沉速度和升沉位移信息。为了验证本文方法的有效性,利用仿真数据来进行验证。仿真条件如下:船舶升沉运动拟合频段为 0.05 ~ 0.2 Hz,模型参数 $N_m = 4$,升沉加速度由升沉位移真值的二阶导数与加速度零偏和噪声叠加而成,数据采样频率为 50 Hz,升沉加速度零偏为 0.05 m/s^2 。图 5 为升沉加速度仿真实验数据。初始滤波参数如下:各余弦分量的升沉位移滤波标准差为 0.2 m,升沉速度滤波标准差为 0.2 m/s,特征频率滤波标准差为 0.3 rad/s,加速度零偏滤波标准差为 0.5 m/s^2 。特征频率滤波噪声方差为 $4 \times 10^{-6} (rad/s)^2$,升沉加速度零偏滤波噪声方差为 $1 \times 10^{-8} (m/s^2)^2$,升沉加速度滤波噪声方差为 $4 \times 10^{-8} (m/s^2)^2$ 。升沉位移初值为 $[0\ m\ 0\ m\ 0\ m\ 0\ m]^T$,升沉速度初值为 $[0.03\ m/s\ 0.06\ m/s\ 0.04\ m/s\ 0.05\ m/s]^T$,特征频率初值为 $[0.31\ rad/s\ 0.42\ rad/s\ 0.54\ rad/s\ 0.6\ rad/s]^T$,加速度零偏初值为 $0\ m/s^2$ 。

图 6 ~ 图 8 分别为上述仿真条件下的船舶升沉加速度零偏估计、升沉速度和升沉位移的滤波结果。

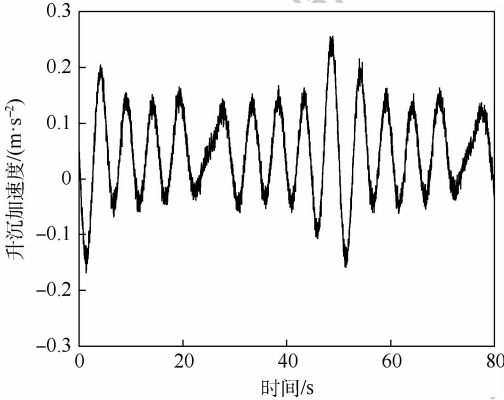


图 5 升沉加速度仿真数据

Fig. 5 Simulation data of heave acceleration

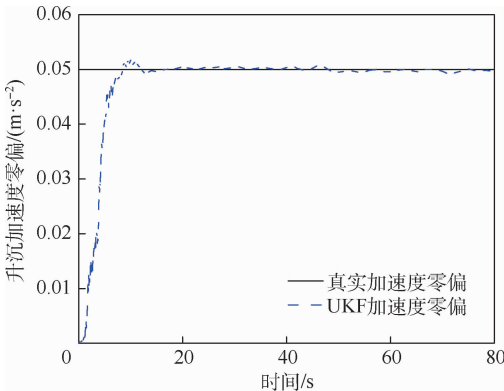


图 6 升沉加速度零偏估计仿真结果

Fig. 6 Simulation results of zero bias estimation of heave acceleration

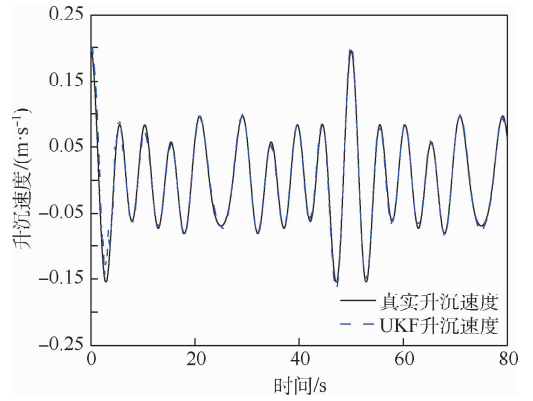


图 7 升沉速度测量仿真结果

Fig. 7 Simulation results of heave speed measurement

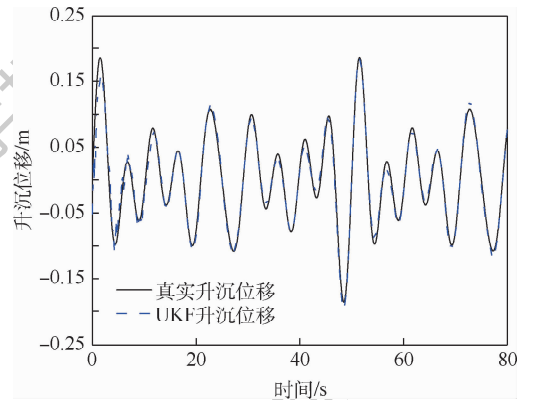


图 8 升沉位移测量仿真结果

Fig. 8 Simulation results of heave displacement measurement

通过图 6 的仿真结果可以看出,升沉加速度零偏估计在滤波开始后 8 s 左右收敛到 0.05 m/s^2 附近,与仿真条件给出的真值吻合。从图 7 和图 8 的仿真结果可以看出,升沉速度与升沉位移滤波估计值与真值吻合。同样可以看出,相比于升沉位移估计,升沉速度估计具有更高的精度,这是由于升沉速度为加速度信号的一次积分,而升沉位移是加速度信号的两次积分。因此,升沉加速度误差对位移的影响大于对速度的影响。仿真结果表明,升沉速度、位移误差保持在可接受的范围内,证明了本文方法的可行性。

4 平台实验

为了验证本文方法对船舶升沉运动速度、位移的估计效果,通过设计相关的实验进行验证。采用实验室现有的三自由度运动平台、Crossbow 公司的 IMU440 惯性测量单元和激光测距仪搭建实验环境。同时,在相同升沉运动测量模型和初始条件下与 EKF 算法结果相比较。如图 9 所示,三自由度运动平台用来模拟

船舶升沉运动,IMU440 惯性测量单元和激光测距仪安装在三自由度运动平台的中心位置,其中惯性测量单元用作升沉运动信息测量,激光测距仪测量安装平台与地面固定平台之间的相对运动,经过转换得到升沉运动的速度和位移基准。

IMU 采集、计算得到平台升沉加速度信息如图 10 所示。

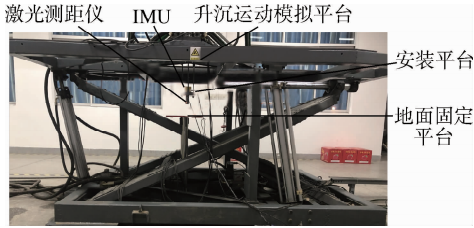


图9 升沉运动信息测量实验装置

Fig. 9 Heave motion information measuring experimental device

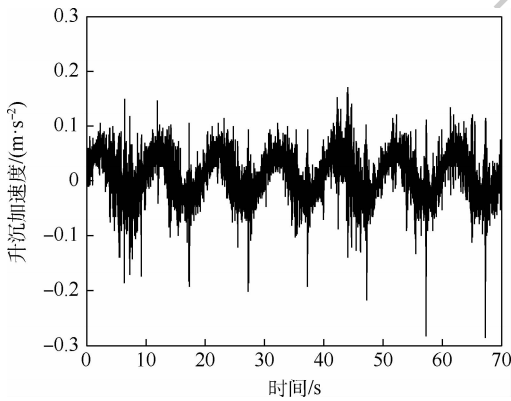


图10 升沉加速度

Fig. 10 Heave acceleration

根据升沉运动的频率分辨率要求,窗口长度取 20 ~ 30 s 数据长度,对升沉加速度数据作 FFT,通过峰值检测算法得到船舶升沉运动模型参数,计算得到系统的滤波初值 $X_0 = [-0.01 \text{ m} \ 0.01 \text{ m/s} \ 0.61 \text{ rad/s} \ 0 \text{ m/s}^2]^T$,根据实验调整确定系统状态滤波初始协方差矩阵 $P_0 = \text{diag}([0.4 \text{ m} \ 0.5 \text{ m/s} \ 0.5 \text{ rad/s} \ 0.8 \text{ m/s}^2]^T)^2$ 。系统的特征频率噪声方差通常由升沉运动的实际工况确定,而由于加速度计零偏是造成天向加速度偏移的主要原因,其噪声特性可以先根据 IMU440 惯性测量单元的加速度计性能参数给出,再根据实际的滤波结果做出调整,设定系统的过程噪声协方差矩阵和量测噪声方差: $Q_k = \text{diag}([0 \text{ m} \ 0 \text{ m/s} \ 2 \times 10^{-3} \text{ rad/s} \ 2 \times 10^{-4} \text{ m/s}^2]^T)^2$, $R_k = 9 \times 10^{-8} (\text{m/s}^2)^2$ 。

图 11 和图 12 为升沉运动滤波结果。图中:

蓝色虚线为本文方法滤波结果,红色实线为相同模型下 EKF 算法滤波结果。滤波估计状态标准差曲线显示,基于 UKF 的方法具有更快的收敛速度和更高的测量精度。

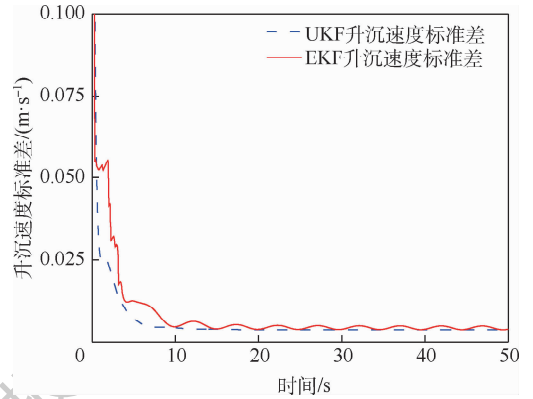


图11 升沉速度滤波标准差

Fig. 11 Standard deviation of heave motion speed filter

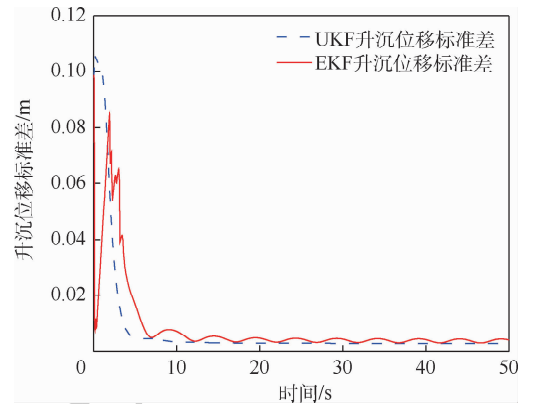


图12 升沉位移滤波标准差

Fig. 12 Standard deviation of heave motion displacement filter

由于在实际应用中,需要在固定时间间隔对海浪的主要频率分量进行检测,当模型中新增或减少频率分量时,需要对系统重新初始化,在面临高动态的海浪运动变化时,为得到高精度的升沉运动信息,往往需要滤波器具有快速收敛的特性。

图 13 和图 14 给出了升沉运动测量实验的速度和位移曲线。图中:黑色实线为激光测距仪测量得到的升沉速度、位移参考值,其中升沉位移幅值最大约 0.1 m,蓝色虚线为采用 UKF 算法得到的结果,红色点划线为采用 EKF 算法得到的结果。

图 15 给出了船舶升沉位移估计误差曲线。在滤波稳定后,本文方法升沉位移测量误差小于 0.5 cm,即小于最大升沉幅值的 5%,能够得到高精度、无时延的升沉运动信息。

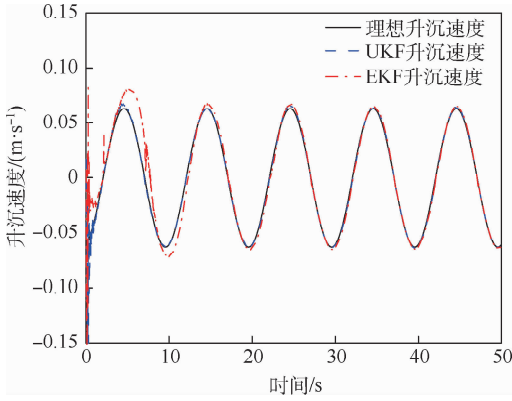


图 13 升沉速度测量实验结果

Fig. 13 Experimental results of heave speed measurement

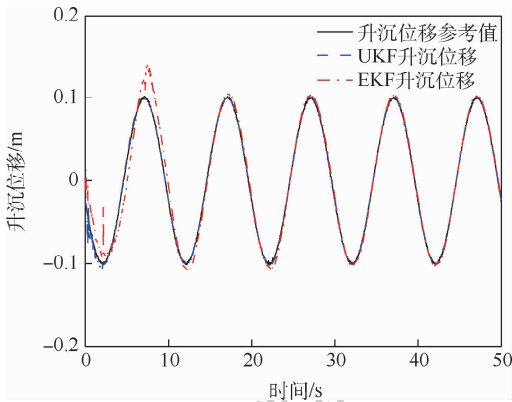


图 14 升沉位移测量实验结果

Fig. 14 Experimental results of heave displacement measurement

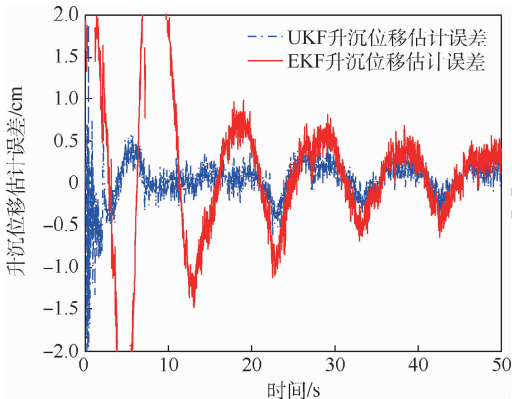


图 15 升沉位移估计误差

Fig. 15 Heave displacement estimation error

5 结论

1) 根据姿态更新算法和实时计算平台水平失准角得到船舶水平姿态信息,进而根据比力方程计算得到升沉加速度信息。通过对升沉加速度作 FFT 并结合升沉运动学模型得到系统滤波初值。

2) 算法实现了准确估计升沉加速度零偏和组成升沉运动的各余弦分量,进而通过叠加得到船舶升沉速度和升沉位移信息。

3) 仿真分析和平台实验结果表明,滤波估计的快速收敛和无时延特性适合船舶升沉运动组成频率分量变化、系统重新初始化的情况;对船舶升沉位移的估计精度达到最大升沉幅值的 5%。对满足船舶海上作业具有一定的工程借鉴价值。

参考文献 (References)

- [1] RICHTER M, SCHNEIDER K, WALSER D, et al. Real-time heave motion estimation using adaptive filtering techniques[J]. IFAC Proceedings Volumes, 2014, 47(3): 10119-10125.
- [2] 黄卫权, 李智超, 卢曼曼. 基于 BMFLC 算法的船舶升沉测量方法[J]. 系统工程与电子技术, 2017, 39(12): 159-164.
HUANG W Q, LI Z C, LU M M. Method of measurement for ship heave motion based on BMFLC algorithm[J]. Systems Engineering and Electronics, 2017, 39(12): 159-164 (in Chinese).
- [3] 孙伟, 孙枫. 基于惯导解算的舰船升沉测量技术[J]. 仪器仪表学报, 2012, 33(1): 162-172.
SUN W, SUN F. Measurement technology of ship heave movement based on SINS resolving[J]. Chinese Journal of Scientific Instrument, 2012, 33(1): 162-172 (in Chinese).
- [4] 刘星伯. 基于捷联航姿系统的舰船瞬时线运动信息测量研究[D]. 哈尔滨: 哈尔滨工程大学, 2009.
LIU X B. Study on the measuring of the instantaneous movements of the ships based on SINS [D]. Harbin: Harbin Engineering University, 2009 (in Chinese).
- [5] 秦永元. 卡尔曼滤波与组合导航原理[M]. 3版. 西安: 西北工业大学出版社, 2015: 199-220.
QIN Y Y. Kalman filtering and integrated navigation principle [M]. 3rd ed. Xi'an: Northwestern Polytechnical University Press, 2015: 199-220 (in Chinese).
- [6] JULIER S J, UHLMANN J K, DURRANT-WHYTE H F. A new approach for filtering nonlinear systems[C]// American Control Conference, 1995: 1628-1632.
- [7] JULIER S J. Unscented filtering and nonlinear estimation[J]. Proceedings of the IEEE, 2004, 92(3): 401-422.
- [8] GODHAVN J M. High quality heave measurements based on GPS RTK and accelerometer technology[C]// OCEANS 2000 MTS/IEEE Conference and Exhibition. Piscataway: IEEE Press, 2000: 309-314.
- [9] HU Y P, TAO L M. Real-time zero phase filtering for heave measurement[C]// The 11th IEEE International Conference on Electronic Measurement & Instruments. Piscataway: IEEE Press, 2013: 321-326.
- [10] 严恭敏, 苏幸君, 翁浚, 等. 基于惯导和零时延滤波器的船舶升沉测量[J]. 导航定位学报, 2016, 4(2): 91-93.
YAN G M, SU X J, WENG J, et al. Measurement of ship's heave motion based on INS and zero-phase-delay digital filter[J]. Journal of Navigation and Positioning, 2016, 4(2): 91-93 (in Chinese).
- [11] 袁书明, 程建华, 马斌. 基于自适应频率估计的船舶瞬时线运

- 动测量方法[J]. 中国惯性技术学报, 2016, 24(5): 565-570.
- YUAN S M, CHENG J H, MA B. Measurement method for ship instantaneous linear movement based on adaptive frequency estimation[J]. Journal of Chinese Inertial Technology, 2016, 24(5): 565-570 (in Chinese).
- [12] KÜCHLER S, EBERHARTER J K, LANGER K, et al. Heave motion estimation of a vessel using acceleration measurements [J]. IFAC Proceedings Volumes, 2011, 44(1): 14742-14747.
- [13] 刘婉婷. 船舶升沉运动预报研究[D]. 大连: 大连海事大学, 2016.
- LIU W T. Study on heave motion prediction of ships [D]. Dalian: Dalian Maritime University, 2016 (in Chinese).
- [14] 李智超. 基于惯导系统的舰船升沉测量技术研究[D]. 哈尔滨: 哈尔滨工程大学, 2018.
- LI Z C. Research on ship heave measurement technology based on inertial navigation system [D]. Harbin: Harbin Engineering University, 2018 (in Chinese).
- [15] 严恭敏, 翁浚. 捷联惯导算法与组合导航原理[M]. 西安: 西北工业大学出版社, 2019: 222-224.
- YAN G M, WENG J. Strapdown inertial navigation algorithm and integrated navigation principle [M]. Xi'an: Northwestern Polytechnical University Press, 2019: 222-224 (in Chinese).
- [16] 刘萍. 基于MEMS的高精度姿态检测系统设计[D]. 哈尔滨: 哈尔滨工程大学, 2016: 11-21.
- LIU P. Design of high precision attitude detection system based on MEMS [D]. Harbin: Harbin Engineering University, 2016: 11-21 (in Chinese).
- [17] 袁赣南, 张涛. 四元数UKF超紧密组合导航滤波方法[J]. 北京航空航天大学学报, 2010, 36(7): 762-766.
- YUAN G N, ZHANG T. Quaternion unscented Kalman filtering for ultra-tight integration [J]. Journal of Beijing University of Aeronautics and Astronautics, 2010, 36(7): 762-766 (in Chinese).

Measurement method of ship's heave motion information based on IMU and UKF algorithm

LU Daohua^{1,2,*}, FU Huaida¹, WANG Jia¹, CAI Yaxuan¹, SONG Shilei¹

(1. School of Mechanical Engineering, Jiangsu University of Science and Technology, Zhenjiang 212003, China;

2. Marine Equipment and Technology Institute, Jiangsu University of Science and Technology, Zhenjiang 212003, China)

Abstract: To obtain accurate estimation of ship heave motion information in real time, according to the ship heave kinematics model and spectral analysis method, an analytical model describing the relationship between the acceleration measurement information of the Inertial Measurement Unit (IMU) and the ship heave motion state quantity was established. Based on the characteristics of Unscented Kalman Filter (UKF) nonlinear filtering algorithm, the heave motion filtering solution is performed. Simulation analysis proves the effectiveness of the proposed algorithm in ship heave motion measurement. The experimental verification of the three-degree-of-freedom platform heave motion measurement is performed. The experimental results show that, compared with the solution of the Extended Kalman Filter (EKF) method, the algorithm in this paper has faster convergence speed and higher measurement accuracy; the estimation accuracy of the ship's heave displacement reaches 5% of the maximum heave amplitude, and accurate and no-delay ship heave motion information can be obtained.

Keywords: heave measurement; Inertial Measurement Unit (IMU); Unscented Kalman Filter (UKF); spectral analysis; platform experiment

Received: 2020-05-27; Accepted: 2020-08-21; Published online: 2020-09-16 09:00

URL: kns.cnki.net/kcms/detail/11.2625.V.20200915.1501.006.html

Foundation item: National Key R & D Program of China (2018YFC0309100)

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

http://bhxb.buaa.edu.cn jbuua@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2020.0216

载人航天器体装太阳能电池阵有效发电面积计算方法



杨胜*, 王曦

(北京空间飞行器总体设计部, 北京 100094)

摘 要: 航天器的发电能力与太阳能电池阵有效发电面积成正比。针对圆柱形载人航天器,提出了一种体装太阳能电池阵有效发电面积计算方法。首先,将太阳能电池阵沿圆周方向划分为多个子阵,通过坐标变换,计算太阳矢量与每个子阵法线的夹角(即太阳入射角);然后,将每个子阵面积与对应太阳入射角的余弦相乘并求和,得到体装太阳能电池阵的有效发电面积。对不同轨道日照角、不同飞行姿态下体装太阳能电池阵有效发电面积进行仿真分析,仿真结果表明:三轴对地姿态下平均有效发电面积为安装面积的 25%~32%,通过固定角度偏航可将有效发电面积提高至安装面积的 30%~44%。

关键词: 载人航天器;体装太阳能电池阵;有效发电面积;轨道日照角;飞行姿态

中图分类号: V529

文献标志码: A

文章编号: 1001-5965(2021)07-1332-06

传统航天器一般采用太阳能电池翼和蓄电池组合电源系统,入轨后太阳能电池翼展开,阳照区通过驱动太阳能电池翼对日定向为航天器供电、给蓄电池充电,阴影区由蓄电池为航天器供电^[1]。

航天器发电能力与轨道日照角、飞行姿态、太阳能电池翼安装方式、驱动方式等密切相关,发电能力分析是航天器总体设计工作的一项重要内容^[2-3]。航天器发电能力与太阳能电池翼有效发电面积成正比(太阳能电池翼安装面积和太阳入射角 θ 的余弦的乘积)。文献[4]研究了太阳能电池翼不对日定向、单轴对日定向模式下太阳入射角计算方法,文献[5]研究了斜装太阳能电池翼对日定向太阳入射角计算方法,用于航天器太阳能电池翼发电能力分析。

不同于传统航天器,日本 HTV 货运飞船、美国“载人龙”飞船使用了体装太阳能电池阵^[6-7]。其中,“载人龙”飞船在非密封舱外表面的 1/2 安装了新设计的太阳能电池阵,任何角度的阳光照射均

可发电。由于不再设置太阳能电池翼,体装太阳能电池阵式载人航天器无需进行帆板展开和驱动,提升了系统的可靠性。体装太阳能电池阵的安装方式比较特殊,一般应用于长方体构型的微小卫星^[8],对于圆柱体构型的载人航天器体装太阳能电池阵发电能力分析相关文献较为少见。

本文提出了一种基于坐标变换的体装太阳能电池阵有效发电面积计算方法,编制了仿真程序并通过算例验证了方法的有效性,对不同轨道日照角、飞行模式下有效发电面积进行对比分析,给出了体装太阳能电池阵载人航天器在轨飞行姿态建议,为体装太阳能电池阵载人航天器能源系统总体设计提供参考。

1 体装太阳能电池阵安装方式

载人航天器一般为圆柱形,沿纵轴从后向前看,按顺时针方向分为 I、II、III、IV 四个象限。各

收稿日期: 2020-05-27; 录用日期: 2020-08-07; 网络出版时间: 2020-08-24 17:06

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20200824.1422.001.html

* 通信作者。E-mail: cast_ys@tom.com

引用格式: 杨胜, 王曦. 载人航天器体装太阳能电池阵有效发电面积计算方法[J]. 北京航空航天大学学报, 2021, 47(7): 1332-1337.
YANG S, WANG X. A calculation method for effective power supply area of manned spacecraft with body mounted solar array[J].
Journal of Beijing University of Aeronautics and Astronautics, 2021, 47(7): 1332-1337 (in Chinese).

象限线间隔 90° , 正常在轨三轴对地飞行时 I 象限线对地^[9]。

体装太阳能电池阵一般均布舱体上半外表面, 从 II 象限线开始, 经 III 象限线, 至 IV 象限线结束。可以视为由 n 个长条形子阵构成, 如图 1 所示。

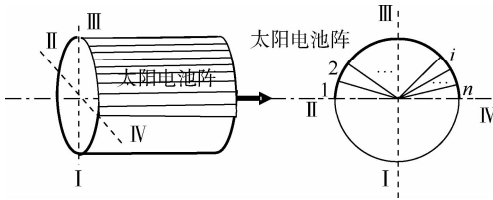


图1 体装太阳能电池阵安装示意图

Fig.1 Schematic of installation of body mounted solar array

2 计算模型

载人航天器体装太阳能电池阵某一时刻有效发电面积 S 等于 n 个子阵有效发电面积之和。第 i 个太阳能电池阵有效发电面积 S_i 等于其安装面积 ΔS 乘以对应太阳入射角 θ_i 的余弦。

$$S = \sum_{i=1}^n S_i = \sum_{i=1}^n (\Delta S \cdot \cos \theta_i) \quad (1)$$

式中: 太阳入射角 θ_i 为第 i 个太阳能电池阵法线矢量与太阳矢量的夹角。

计算载人航天器体装太阳能电池阵有效发电面积, 主要步骤如下:

步骤 1 在载人航天器本体几何坐标系下将体装太阳能电池阵划分为 n 个子阵。

步骤 2 在第 i 个太阳能电池阵安装坐标系下计算某一时刻太阳入射角 θ_i 。需要根据该时刻航天器轨道和飞行姿态, 通过一系列坐标变换将太阳矢量从地心赤道惯性坐标系投影到第 i 个太阳能电池阵安装坐标系下。

步骤 3 根据式(1)计算第 i 个太阳能电池阵有效发电面积 S_i 并求和, 得到该时刻体装太阳能电池阵有效发电面积 S 。

步骤 4 计算一个轨道周期内载人航天器在阳照区有效发电面积的平均值 S_{avg} , 用来表征体装太阳能电池阵式载人航天器的发电能力。

因此, 载人航天器体装太阳能电池阵有效发电面积计算模型主要包括太阳入射角计算、航天器轨道计算、航天器飞行姿态计算、阳照区计算等模块。

2.1 坐标系定义

下面介绍本文所用到的坐标系^[9]。

1) 地心赤道惯性坐标系 $O_1X_1Y_1Z_1$ 。原点在地球心, O_1X_1 在赤道面内指向平春分点; O_1Z_1 轴垂

直于赤道面, 与地球自转角速度矢量一致; O_1Y_1 满足右手法则。该坐标系用于太阳矢量计算。

2) 航天器轨道坐标系 $Ox_o y_o z_o$ 。原点在航天器质心, Oz_o 指向地心; Ox_o 在轨道平面内垂直于 Oz_o 指向航天器飞行方向; Oy_o 满足右手法则。根据航天器升交点赤经 (Ω)、纬度幅角 (u)、轨道倾角 (i), 可将太阳矢量从地心赤道惯性坐标系转换到航天器轨道坐标系, 如图 2 所示。

3) 本体控制坐标系 $Ox_c y_c z_c$ 。原点在航天器质心, Ox_c 沿航天器纵轴, 由航天器底部指向顶部方向; Oy_c 在与 Ox_c 垂直的平面内指向 IV 象限线方向; Oz_c 满足右手法则。本体控制坐标系相对轨道坐标系的转角即为航天器的姿态角, 按照 3-1-2 转序, 分别为偏航角 ψ 、滚动角 γ 和俯仰角 φ 。根据航天器姿态角可将太阳矢量从航天器轨道坐标系转换到本体控制坐标系, 如图 3 所示。

4) 本体质心坐标系 $Ox_b y_b z_b$ 。原点在航天器质心, Ox_b 沿航天器纵轴, 由航天器底部指向顶部方向; Oy_b 在与 Ox_b 垂直的平面内指向 III 象限线方向; Oz_b 满足右手法则。本体控制坐标系绕 Ox_b 旋转 -90° 即得到本体质心坐标系。

5) 本体几何坐标系 $O_d x_b y_b z_b$ 。原点在航天器与运载火箭对接面几何中心, $O_d x_b$ 、 $O_d y_b$ 、 $O_d z_b$ 与本体质心坐标系 Ox_b 、 Oy_b 、 Oz_b 平行, 且指向一致, 如图 4 所示, 该坐标系用于体装太阳能电池阵的安装。

6) 第 i 个太阳能电池阵安装坐标系 $O_{si} x_i y_i z_i$ 。原点 O_{si} 为第 i 个太阳能电池阵的安装中心, $O_{si} y_i$ 和 $O_{si} z_i$ 由本体几何坐标系 $O_d y_b$ 和 $O_d z_b$ 旋转一定角度 α 得到; $O_{si} x_i$ 满足右手法则, 如图 5 所示。其中, $\alpha = 3\pi/2 + (2i - 1)\pi/(2n)$; 若 $\alpha \geq 2\pi$, 则 $\alpha = \alpha - 2\pi$; n 为太阳子阵个数。

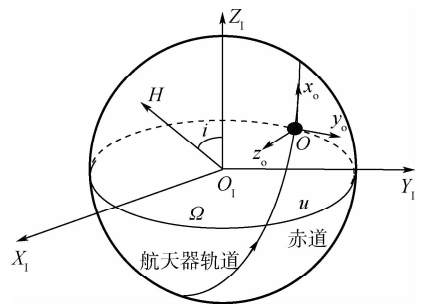


图2 地心赤道惯性坐标系与航天器轨道坐标系

Fig.2 Coordinate systems of geocentric equator inertia and spacecraft orbit

2.2 太阳入射角计算

在第 i 个太阳能电池阵安装坐标系下计算太阳入射角 θ_i , 如图 6 所示。

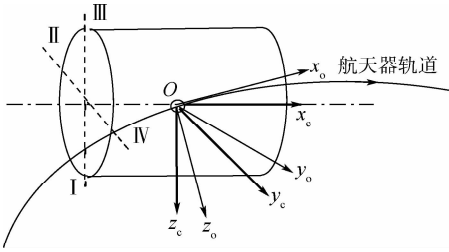


图3 航天器轨道坐标系与本体控制坐标系

Fig.3 Coordinate systems of spacecraft orbit and body control

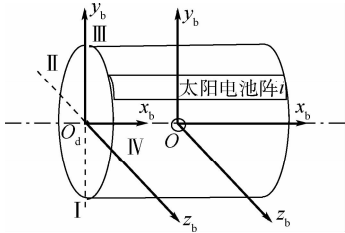


图4 航天器本体质心坐标系与本体几何坐标系

Fig.4 Coordinate system of spacecraft body centroid and geometric coordinate system of spacecraft body

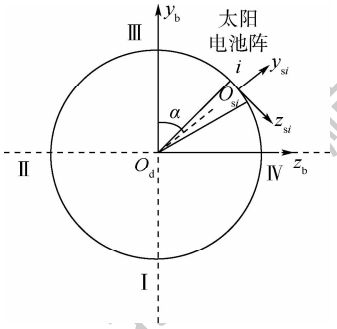


图5 航天器本体几何坐标系与太阳电池阵安装坐标系

Fig.5 Geometric coordinate system of spacecraft body and coordinate system of body mounted solar array

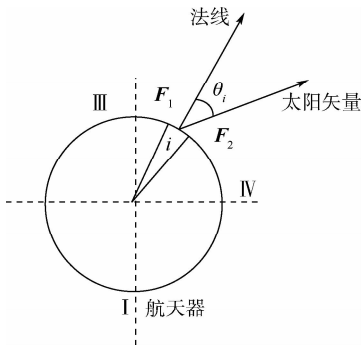


图6 太阳入射角计算模型示意图

Fig.6 Schematic of calculation model of solar incidence angle

其中, F_1 为第 i 个太阳电池阵的法线矢量 $[0, 1, 0]$, F_2 为太阳矢量在第 i 个太阳电池阵安装坐标系下的投影, 则

$$\theta_i = \arccos\left(\frac{F_1 \cdot F_2}{|F_1| \cdot |F_2|}\right) \quad (2)$$

当 $\theta_i < 90^\circ$ 时, 该太阳电池阵对太阳可见。在工程上, 还应考虑大入射角效应, 即大入射角情况下太阳电池阵的发电能力与 θ 的余弦不再成正

比, 而是快速衰减^[10-11]。因此, 本文在有效发电面积计算时, 当 θ_i 大于 65° 则直接置为 90° , 认为该太阳电池阵不发电。

式(2)中, F_2 的计算方法如下:

首先, 计算太阳在地心赤道惯性坐标系中的位置:

$$r_s = \begin{bmatrix} x_s \\ y_s \\ z_s \end{bmatrix} = \begin{bmatrix} \cos \delta_s \cos \alpha_s \\ \cos \delta_s \sin \alpha_s \\ \sin \delta_s \end{bmatrix} \quad (3)$$

式中: α_s 为太阳赤经; δ_s 为太阳赤纬。

然后, 通过 4 次坐标旋转变换(地心赤道惯性坐标系 \rightarrow 航天器轨道坐标系 \rightarrow 航天器本体控制坐标系 \rightarrow 航天器本体质心坐标系 \rightarrow 航天器本体几何坐标系 \rightarrow 第 i 个太阳电池阵安装坐标系。其中, 航天器本体质心坐标系 \rightarrow 本体几何坐标系, 只平移, 不旋转, 将太阳矢量 r_s 投影到第 i 个太阳电池阵安装坐标系。

$$F_2 = R_4 \cdot R_3 \cdot R_2 \cdot R_1 \cdot r_s \quad (4)$$

式中:

$$R_1 = R_x\left(-\frac{\pi}{2}\right) \cdot R_z\left(\frac{\pi}{2}\right) \cdot R_z(u) \cdot R_x(ii) \cdot R_z(\Omega)$$

$$R_2 = R_y(\varphi) \cdot R_x(\gamma) \cdot R_z(\psi)$$

$$R_3 = R_x\left(-\frac{\pi}{2}\right)$$

$$R_4 = R_x(\alpha)$$

其中: α 为第 i 个太阳电池阵的安装角度。

2.3 航天器轨道计算

太阳入射角 θ_i 与航天器在轨道上的位置有关。轨道计算可采用经典的 RKF7(8) 积分器, 积分过程中考虑的力学模型是目前常用的典型模型, 包括地球中心引力、地球非球形摄动、大气阻力摄动、太阳光压摄动、N 体摄动(日月引力)^[12]。

2.4 航天器飞行姿态

太阳入射角 θ_i 与航天器的飞行姿态有关。载人航天器在轨正常飞行姿态主要包括三轴对地、连续偏航、固定角度偏航等。

1) 三轴对地。航天器本体控制坐标系与轨道坐标系重合, 即偏航角 $\psi = 0^\circ$ 。

2) 固定角度偏航。航天器本体控制坐标系相对轨道坐标系绕 Oz_c 轴旋转一个固定角度, 一般取 90° 或 270° , 即偏航角 $\psi = 90^\circ$ 或 $\psi = 270^\circ$ 。

3) 连续偏航。航天器本体控制坐标系相对轨道坐标系绕 Oz_c 轴连续旋转, 使得 Ox_c 轴在 $x_o y_o$ 平面内始终跟踪太阳, 即

$$\psi = \arctan\left(\frac{\cos b}{\cos a}\right) \quad (5)$$

式中: a 为太阳矢量与 Ox_o 轴夹角; b 为太阳矢量

与 Oy_0 轴夹角。

2.5 阳照区计算

阳照区计算采用柱形地影模型^[5],如图7所示,则航天器在阳照区的条件为

$$|\mathbf{r}| \sin \delta > R_e \quad (6)$$

式中: \mathbf{r} 为航天器地心矢量; δ 为太阳矢量与航天器地心矢量的夹角; R_e 为地球半径。若航天器在阴影区,有效发电面积直接置为0。

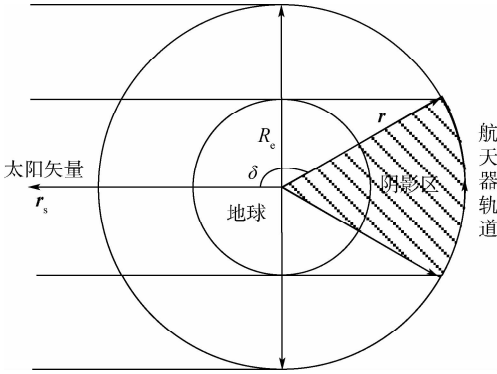


图7 地影模型

Fig.7 Model of earth eclipse

3 仿真验证

3.1 输入条件

航天器本体为直径3.35 m、长度2 m的圆柱体;航天器上半表面间隔 3° 均布60个太阳电池阵子阵,总安装面积约 9.6 m^2 。

航天器轨道高度400 km,偏心率0,倾角 40° 。选取不同的轨道日照角 β (太阳矢量与轨道面的夹角)进行仿真,每次仿真时间5 400 s(一个轨道周期),仿真步长60 s。

3.2 计算流程

步骤1 根据初始历元时刻计算轨道参数。

步骤2 计算飞行姿态。

步骤3 计算是否在阳照区。

1) 如果是,从 $i=1$ 开始,计算第 i 个太阳电池阵太阳入射角 θ_i 及有效发电面积 S_i 。当 $i=n$ 时,计算该时刻航天器有效发电面积 S ,转步骤4。

2) 如果不是,该时刻航天器有效发电面积置0,转步骤4。

步骤4 计算是否结束,如果是,转步骤5;如果不是,转步骤1。

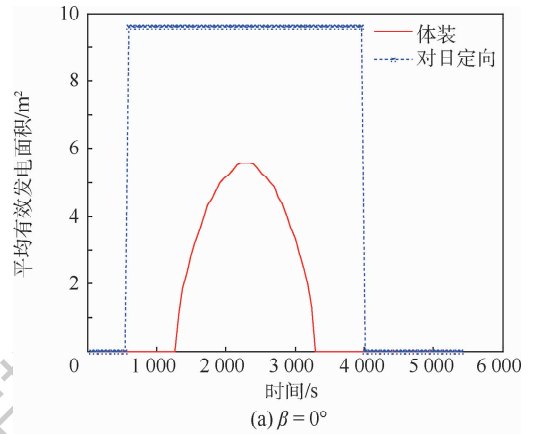
步骤5 计算一个轨道周期阳照区有效发电面积的平均值 S_{avg} (平均等效发电面积),仿真结束。

3.3 仿真结果

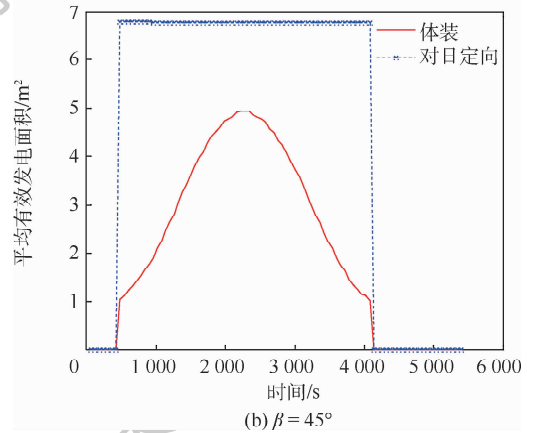
分别针对不同轨道日照角、不同飞行姿态开展了载人航天器体装太阳电池阵平均有效发电面积仿真。

1) 三轴对地飞行姿态仿真

首先,在三轴对地飞行姿态下,针对不同轨道日照角,分别计算对日定向太阳电池阵和体装太阳电池阵平均有效发电面积,仿真结果如图8和表1所示。



(a) $\beta = 0^\circ$



(b) $\beta = 45^\circ$

图8 三轴对地飞行姿态下平均有效发电面积

Fig.8 Average effective power supply area under flight attitude of three-axis orientating the earth

表1 三轴对地飞行姿态下不同安装方式平均有效发电面积

Table 1 Average effective area for different installation ways under earth-oriented three-axis stabilized attitude

安装方式	平均有效发电面积/ m^2			
	$\beta = 0^\circ$	$\beta = 30^\circ$	$\beta = 45^\circ$	$\beta = 63^\circ$
对日定向	9.6	8.3	6.8	4.3
体装	2.4(25%)	3.0(31%)	3.1(32%)	2.9(30%)

注:括号内为与太阳电池阵安装面积(9.6 m^2)的比值。

仿真结果表明,三轴对地飞行姿态下,对日定向太阳电池阵平均有效发电面积随轨道日照角增大而下降。体装太阳电池阵平均有效发电面积随轨道日照角增大先增大再减小,约为安装面积的25%~32%(考虑大入射角效应),轨道日照角为 45° 时达到最大值。其中,小轨道日照角情况下主要受进出阳照区时舱体自身遮挡影响^[13-14],大轨道日照角情况下主要受太阳电池阵安装区域(只覆盖

上半表面)影响,导致平均有效发电面积有所减小。

2) 不同飞行姿态仿真

对日定向太阳能电池阵载人航天器在大入射角情况下通常通过连续偏航,控制太阳能电池翼对日定向,提高发电能力^[15]。对体装太阳能电池阵载人

表 2 不同飞行姿态下体装太阳能电池阵平均有效发电面积

Table 2 Average effective power supply area of body mounted solar array under different flight attitudes

飞行姿态	平均有效发电面积/m ²			
	$\beta = 0^\circ$	$\beta = 30^\circ$	$\beta = 45^\circ$	$\beta = 63^\circ$
三轴对地	2.4	3.0	3.1	2.9
连续偏航	2.4(0%)	1.8(-40%)	1.2(-61%)	0.12(-96%)
固定角度偏航(90°)	4.2(+75%)	3.5(+17%)	2.6(-16%)	0.63(-78%)

注:括号内为相对三轴对地姿态下平均有效发电面积的变化量。

为了分析其他飞行姿态对发电能力的影响,进行90°固定角度偏航飞行姿态下平均有效发电面积仿真,如表2所示。仿真结果表明,体装太阳能电池阵在小轨道日照角情况下可以通过固定角度偏航提高有效发电面积(最大可提升至安装面积的44%),随着轨道日照角增大受进出阳照区时舱体自身遮挡的影响有效发电面积反而会下降。

不同飞行姿态下,体装太阳能电池阵平均有效发电面积对比如图9所示。

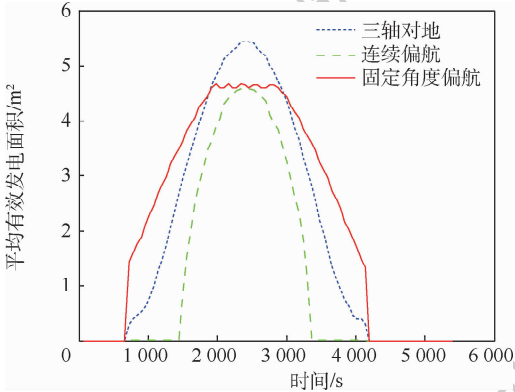


图9 不同飞行姿态下平均有效发电面积($\beta = 30^\circ$)

Fig.9 Average effective area under different flight attitudes ($\beta = 30^\circ$)

3) 不同轨道倾角仿真

针对载人航天器近地倾斜圆轨道的特点,进行了不同轨道倾角情况下平均有效发电面积的仿真分析,也得到类似的结论,限于篇幅不再赘述。

4 结 论

本文提出了一种载人航天器体装太阳能电池阵有效发电面积计算方法,针对近地飞行的载人航天器开展了不同飞行姿态下体装太阳能电池阵有效发电面积仿真分析,主要结论如下:

1) 三轴对地飞行姿态下,体装太阳能电池阵平

均有效发电面积约为安装面积的25%~32%(考虑大入射角效应),随轨道日照角增大先增大再减小,轨道日照角为45°时达到最大值。

2) 连续偏航飞行姿态下,无法提高体装太阳能电池阵有效发电面积,反而会损失一定的有效发电面积。

3) 固定角度偏航(90°/270°)飞行姿态下,在小轨道日照角情况下可以提高体装太阳能电池阵有效发电面积,但随着轨道日照角增大有效发电面积反而会下降。

4) 通过固定角度偏航(小轨道日照角)和三轴对地(大轨道日照角)飞行姿态结合,体装太阳能电池阵平均有效发电面积可由安装面积的25%~32%提高至30%~44%。

本文方法可推广应用于其他不规则构型航天器体装太阳能电池阵有效发电面积的计算。

参考文献 (References)

- [1] 刘志全,杨淑利,濮海玲.空间太阳能电池阵的发展现状及趋势[J].航天器工程,2012,21(6):112-118.
LIU Z Q, YANG S L, PU H L. Development and trend of space solar array technology [J]. Spacecraft Engineering, 2012, 21 (6):112-118 (in Chinese).
- [2] 李涛,李伟,杨雷.一种航天器太阳能电池阵供电能力计算方法[J].北京航空航天大学学报,2017,43(7):1355-1363.
LI T, LI W, YANG L. A calculation method for power supply capability of spacecraft solar array [J]. Journal of Beijing University of Aeronautics and Astronautics, 2017, 43 (7): 1355-1363 (in Chinese).
- [3] 任筱强,王鹏鹏.航天器组合体能量平衡分析系统设计与应用[J].航天器工程,2013,22(2):60-64.
REN X Q, WANG P P. Design and implementation of energy balance analysis system for spacecraft combination [J]. Spacecraft Engineering, 2013, 22 (2): 60-64 (in Chinese).
- [4] 陈忠贵,张志,廖瑛.航天器太阳翼在轨光照角度建模及仿真分析[J].航天器工程,2012,21(1):37-42.

- CHEN Z G, ZHANG Z, LIAO Y. Modeling and simulation analysis of solar illumination angle on spacecraft solar wing in-orbit [J]. *Spacecraft Engineering*, 2012, 21(1): 37-42 (in Chinese).
- [5] 万向成, 陆晴, 刘佩东. 斜装匀速对日驱动的太阳电池阵入射角计算方法[J]. *航天器工程*, 2017, 26(2): 38-44.
- WAN X C, LU Q, LIU P D. Method for calculating sunlight incident angle of solar array slanted mounted and driving with uniformly angular velocity [J]. *Spacecraft Engineering*, 2017, 26(2): 38-44 (in Chinese).
- [6] TEPPEI O, YASUSHI O, KIYOKAZU K, et al. Charging of the H-II transfer vehicle at rendezvous and docking phase [J]. *Journal of Spacecraft and Rockets*, 2018, 55(4): 971-983.
- [7] JEREMY H. The international space station is expecting two visitors this month; Starliner and Crew Dragon [J]. *IEEE Spectrum*, 2018, 55(8): 6-8.
- [8] 林君毅, 尹兴月. 微小卫星无姿控下太阳能电池阵输出功率研究[J]. *电源技术*, 2012, 36(10): 1478-1480.
- LIN J Y, YIN X Y. Research of solar array output for the mini-satellite without attitude control [J]. *Chinese Journal of Power Sources*, 2012, 36(10): 1478-1480 (in Chinese).
- [9] 杨宏, 等. 载人航天器技术 [M]. 北京: 北京理工大学出版社, 2017.
- YANG H, et al. *Human spacecraft technology* [M]. Beijing: Beijing Institute of Technology Press, 2017 (in Chinese).
- [10] 崔鹏, 郗毅鹏, 尹兴月, 等. 大光照角入射对太阳能电池阵的影响分析与验证[J]. *电源技术*, 2014, 38(9): 1637-1639.
- CUI P, QIE Y P, YIN X Y, et al. Impact of large light angle on solar array and its verification [J]. *Chinese Journal of Power Sources*, 2014, 38(9): 1637-1639 (in Chinese).
- [11] 程保义. 空间三结砷化镓太阳能电池辐照衰减模型研究 [D]. 天津: 南开大学, 2010.
- CHENG B Y. Research on the radiation model of space triple-junction solar cell [D]. Tianjin: Nankai University, 2010 (in Chinese).
- [12] 刘林. 航天器轨道理论 [M]. 北京: 国防工业出版社, 2000.
- LIU L. Orbit theory of spacecraft [M]. Beijing: Defense Industry Press, 2000 (in Chinese).
- [13] 李安寿, 张东来, 杨扬, 等. 一种精确计算航天器本体对太阳能电池阵遮挡的方法 [J]. *宇航学报*, 2013, 34(10): 1403-1409.
- LI A S, ZHANG D L, YANG Y, et al. A method for accurately calculating the shadow of solar array shaded by spacecraft itself [J]. *Journal of Astronautics*, 2013, 34(10): 1403-1409 (in Chinese).
- [14] RAMAPRABHA R, MATHUR B L. A comprehensive review and analysis of solar photovoltaic array configurations under partial shaded conditions [J]. *International Journal of Photoenergy*, 2012, 12(6): 1-16.
- [15] 王颖, 顾荃莹. 倾斜轨道航天器太阳翼对日跟踪方法探讨 [J]. *航天器工程*, 2009, 18(3): 36-40.
- WANG Y, GU Q Y. Method of solar array tracking for inclined-orbit satellite [J]. *Spacecraft Engineering*, 2009, 18(3): 36-40 (in Chinese).

A calculation method for effective power supply area of manned spacecraft with body mounted solar array

YANG Sheng*, WANG Xi

(Institute of Spacecraft System Engineering, Beijing 100094, China)

Abstract: The power supply capability is proportional to the effective area of spacecraft solar array. For cylindrical manned spacecraft, a calculation method for effective area of body mounted solar array is proposed. First, the mounted solar array was divided into n parts along the circumference. The angle between the sun vector and the normal of each part (incidence angle) was calculated by coordinate transform. Then, the effective area of body mounted solar array was obtained by summing up the products of the area of each part and the cosine of incidence angle. Finally, the proposed method was verified by effective area simulation under different orbit sunlight angles and different flight attitudes for manned spacecraft with body mounted solar array. The simulation results show that the average effective area can reach 25% - 32% of the mounted area under earth-oriented three-axis stabilized attitude the flight attitude of orientating the earth. By yawing the manned spacecraft with fixed angle, the ratio of effective area to mounted area can be expanded to 30% - 44%.

Keywords: manned spacecraft; body mounted solar array; effective power supply area; orbit sunlight angle; flight attitude

Received: 2020-05-27; Accepted: 2020-08-07; Published online: 2020-08-24 17:06

URL: kns.cnki.net/kcms/detail/11.2625.V.20200824.1422.001.html

* Corresponding author. E-mail: cast_ys@tom.com

http://bhxb.buaa.edu.cn jbuua@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2020.0191

基于导重法的自重载荷下悬臂梁结构拓扑优化



任毅如^{1,*}, 向剑辉¹, 何杰¹, 宁克焱², 杨玲玲²

(1. 湖南大学 机械与运载工程学院, 长沙 410082; 2. 中国北方车辆研究所, 北京 100072)

摘 要: 针对自重载荷悬臂梁结构拓扑优化末端区域材料分布不收敛的问题, 提出了一种拓扑优化方法。根据虚功相等原理, 利用四节点矩形单元的形函数、单元体积密度和质量的关系建立了载荷等效方法。根据优化模型的 Kuhn-Tucker 条件推导出了导重准则, 以及目标函数的灵敏度公式和考虑自重载荷拓扑优化的迭代算法。针对自重载荷作用下悬臂梁结构拓扑优化存在的末端区域材料分布模糊问题, 研究了变密度法和非结构质量相结合的求解策略, 揭示了典型因素对拓扑结构的影响规律。结果表明: 所提方法能够解决自重载荷下悬臂梁末端区域材料分布模糊问题。

关键词: 拓扑优化; 导重法; 变密度法; 自重载荷; 材料属性有理近似函数插值 (RAMP) 模型

中图分类号: TP314; TH111; O343

文献标志码: A

文章编号: 1001-5965(2021)07-1338-07

结构拓扑优化是指在一定的设计区域内, 在满足特定的约束条件和边界条件情况下, 寻求材料最优分配的过程。拓扑优化的问题自其被提出以来就受到了广泛的关注和研究, 包括载荷不确定问题^[1-2]、传热学问题^[3-4]、非线性问题^[5]及工程应用问题^[6-7]等。重力作为工程应用中无法避免的载荷, 在很多结构设计中是不可忽略的存在, 特别是在一些体积力在总载荷中占比较大的情况, 不考虑体积力可能导致完全不同的结果, 偏离实际的工况。

为了解决体积力情况下的拓扑优化设计问题, 大量的学者对此开展了研究。1984年, 陈树勋和叶尚辉^[8]为了解决天线设计问题, 采用了一种考虑自重的结构优化方法, 称为导重准则法。1989年, Bendsoe^[9]提出了基于固体各向同性材料惩罚 (Solid Isotropic Material with Penalization, SIMP) 模型的变密度法, 用于解决连续体结构的拓扑优化问题。Pedersen^[10]通过对采用 SIMP 模

型的变密度法的研究, 发现在考虑自重载荷时, 会出现结构无法支撑自身质量的情况, 该情况会导致优化结果中材料分布模糊, 称为低密度区域的“附属”效应。Stolpe 和 Svanberg^[11]提出了材料属性有理近似函数插值 (Rational Approximation of Material Properties, RAMP) 模型, 该模型增加了拓扑优化过程中获得松弛问题 0-1 解的可能性, 在一些情况下能够获得更加明确的材料分布。Bruyneel 和 Duysinx^[12]通过对 SIMP 模型进行修正, 选择基于梯度的运动渐进线逼近方法, 提出了一种具有密度依赖性的拓扑优化方法, 证明了体积力对结构拓扑优化的巨大影响。Lopes 等^[13]在 SIMP 模型的基础上, 创立了一种刚度和密度特性参数化的材料模型, 提出了一种适用于惯性力的三维结构拓扑优化设计公式。Liu 和 Li 等^[14-15]将陈树勋等提出的导重准则法引入到了连续体的拓扑优化问题中, 并且对其做了改进, 使其兼具优化准则法和数学规划法的特点。随后, 陈祥和刘辛

收稿日期: 2020-05-18; 录用日期: 2020-06-05; 网络出版时间: 2020-08-05 16:50

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20200805.1331.006.html

基金项目: 工业和信息化部基础产品创新科研项目 (23709900000170008)

* 通信作者。E-mail: renyiru@hnu.edu.cn

引用格式: 任毅如, 向剑辉, 何杰, 等. 基于导重法的自重载荷下悬臂梁结构拓扑优化[J]. 北京航空航天大学学报, 2021, 47(7): 1338-1344. REN Y R, XIANG J H, HE J, et al. Topology optimization of cantilever structure with self-weight load based on guide-weight method [J]. Journal of Beijing University of Aeronautics and Astronautics, 2021, 47(7): 1338-1344 (in Chinese).

军^[16]比较了 SIMP 模型和 RAMP 模型各自的特点,基于 RAMP 模型结构导重法对拓扑优化问题进行了求解。张晖等^[17]针对自重载荷作用下的连续体拓扑优化问题提出了采用 RAMP 模型和平均敏感度过滤技术相结合的求解策略。许华旻等^[18-19]继续基于导重法对考虑自重载荷作用的拓扑优化做了相关研究。Jain 和 Saxena^[20]比较自重和不同比值关系的外载荷作用下的优化结果,说明了自重对结构拓扑优化的影响及在优化设计中考虑结构自重的必要性。

自重载荷下结构拓扑优化会出现中间密度等问题,为此大量学者开展了研究。Bruyneel 和 Duysinx^[12]针对低密度区域的“附属”效应问题,采用了在合适位置添加非结构质量的方式解决问题。许华旻等^[19]同样也通过添加质量点的方式获得明确的材料分布。Holmberg 等^[21]比较了附加非结构质量情况下结构受各种最糟糕加速度时得到的不同优化结果,强调了结构自重的影响。虽然非结构质量在考虑体积力等问题的拓扑优化研究中已有大量的研究,但是鲜有非结构质量对拓扑优化的影响研究尤其是悬臂梁结构拓扑优化的研究。鉴于此,本文对结合导重准则法和 RAMP 模型对自重载荷作用下的结构拓扑优化进行了研究,在该研究的基础上针对悬臂梁结构在自重载荷作用下末端容易出现中间密度单元模糊区域的情况,采用在梁末端添加非结构质量的求解策略,讨论了非结构质量对优化结果的影响规律。

1 理论和方法

1.1 变密度法

变密度法是结构拓扑优化应用最为广泛的方法,其核心思想是:对材料的伪密度进行惩罚使其归于 0 和 1,从而完成材料的去除和保留。对于变密度法的惩罚模型,目前较为常用的有 SIMP 模型和 RAMP 模型。基于 SIMP 模型的变密度法又叫 SIMP 法,由 Bendsoe^[9]首次引入变密度法中,其表达式为

$$E_i = E_{\min} + x_i^p (E_0 - E_{\min}) \quad x_i \in [0,1] \quad (1)$$

式中: E_i 为单元的弹性模量; E_0 和 E_{\min} 分别为满材料单元和孔洞材料单元的弹性模量; x_i 为单元的伪密度; p 为 SIMP 模型的惩罚因子,一般取值为 3。

RAMP 模型由 Stolpe 和 Svanberg^[11]提出,该模型可以让施加有设计相关载荷时的优化设计获得更加稳定的优化结果,其表达式为

$$E_i = E_{\min} + \frac{x_i}{1 + q(1 - x_i)} (E_0 - E_{\min}) \quad x_i \in [0,1] \quad (2)$$

式中: q 为 RAMP 模型的惩罚因子,一般取值为 5~120。在结构拓扑优化中,自重载荷为设计相关载荷,故而采用 RAMP 模型。

变密度法拓扑优化常用的求解方法有数学规划法、优化准则法及导重准则法,导重准则法结合了数学规划法和优化准则法的优势,具有更高的求解效率。

1.2 体积约束下最大刚度的自重载荷结构拓扑优化

1.2.1 结构等效节点重力的计算

重力载荷作为一种设计相关载荷,其大小会随着设计变量的更新而变化。拓扑优化依赖于有限元计算,因此需要将单元的重力等效到单元节点上。通常,将每个单元重力按照虚功相等的原则分配到单元节点之上。

有限元分析采用四节点矩形单元,根据四节点矩形单元的形函数及单元体积密度和质量的关系,可以得到四节点矩形单元的第 i 个节点在重力加速度方向的等效节点力为^[18]

$$f_{vi}^e = \iint_A N_i (-\rho g) t dx dy = -\rho g t \iint_A N_i dx dy = -ab\rho g t = -\frac{1}{4}V_e \rho g \quad (3)$$

式中: a 和 b 分别为单元 1/2 长和 1/2 宽; N_i 表示四节点矩形单元第 i 个节点的形函数, $1 \leq i \leq 4$; ρ 为材料的物理密度; g 为重力加速度; A 为单元面积; t 为单元的厚度; V_e 为一个单元的体积。

在四节点单元有限元分析中,节点力为二维矢量,与重力垂直方向的力为 0,可以写出节点的体积力矢量为

$$f_{vi}^e = \begin{bmatrix} f_{vix}^e \\ f_{viy}^e \end{bmatrix} = \begin{bmatrix} 0 \\ -ab\rho g t \end{bmatrix} = \begin{bmatrix} 0 \\ -\frac{1}{4}V_e \rho g \end{bmatrix} \quad (4)$$

1.2.2 迭代公式和灵敏度

对结构进行拓扑优化设计时,以结构质量/体积为约束、最大刚度为目标是一种非常常用的优化方案。在结构的拓扑优化中,为了方便计算,通常以柔度最小来代替刚度最大的优化目标。

以体积为约束、柔度最小为目标的拓扑优化数学模型,其表达式为

$$\begin{cases} \text{find} & \mathbf{X} = [x_1, x_2, \dots, x_N]^T \in \mathbf{R}^N \\ \text{min} & C(\mathbf{X}) = \mathbf{U}^T \mathbf{K} \mathbf{U} = \sum_{i=1}^N \mathbf{u}_i^T \mathbf{k}_i \mathbf{u}_i \\ \text{s. t.} & g(\mathbf{X}) = M - M_0 \leq 0 \\ & x_i^L \leq x_i \leq x_i^U \quad i = 1, 2, \dots, N \end{cases} \quad (5)$$

式中: \mathbf{X} 为 N 维的设计变量,即伪密度; $C(\mathbf{X})$ 为结构的柔度; \mathbf{u}_i 为单元 i 的位移矢量; \mathbf{k}_i 为单元 i

的刚度矩阵; $g(\mathbf{X})$ 为结构的质量/体积约束; M 为结构在迭代优化过程中的实际质量/体积; M_0 为约束的质量/体积; x_i^L 和 x_i^U 分别为 x_i 的下限和上限。

根据 Kuhn-Tucker 条件对式(5)进行求解, 得到导重准则法求解的迭代准则^[14-15]为

$$x_i = \frac{G_i}{\lambda H_i} \quad i = 1, 2, \dots, N \quad (6)$$

式中: λ 为拉格朗日乘子; $H_i = \frac{\partial M}{\partial x_i} = v_i$ 为单元的容重, v_i 为单元 i 的体积, $M_i = x_i H_i = x_i v_i$ 为单元的等效质量; $G_i = x_i \frac{\partial C}{\partial x_i}$ 为单元的导重, $G = \sum_{i=1}^N G_i$ 为模型的总导重。

迭代公式可写为

$$x_i^{(k+1)} = \left(\frac{G_i}{\lambda H_i} \right)^{(k)} = \left(\frac{x_i \frac{\partial C}{\partial x_i}}{\lambda v_i} \right)^{(k)} \quad i = 1, 2, \dots, N \quad (7)$$

$$x_i^{(k+1)} = \begin{cases} x_i^U & x_i \geq x_i^U \\ \alpha \left(\frac{x_i \left(\frac{1}{2} v_i \rho g \cdot \mathbf{P}_0 \cdot \mathbf{u}_i^T - f'(x_i) \mathbf{u}_i^T \mathbf{k}_i \mathbf{u}_i \right)}{\lambda v_i} \right)^{(k)} + (1 - \alpha) x_i^{(k)} & x_i^L < x_i < x_i^U \\ x_i^L & x_i \leq x_i^L \end{cases} \quad i = 1, 2, \dots, N \quad (9)$$

$$\lambda = \frac{G}{M_0} = \frac{\sum_{i=1}^N \left(x_i \left(\frac{1}{2} v_i \rho g \cdot \mathbf{P}_0 \cdot \mathbf{u}_i^T - f'(x_i) \mathbf{u}_i^T \mathbf{k}_i \mathbf{u}_i \right) \right)}{M_0} \quad (10)$$

式中: α 为步长因子, 合适的取值可以保证迭代的收敛。关于 α 取值, 陈树勋^[22-23]进行了详细的数学推导和证明。在取值范围内, α 值越小, 越能精确地搜索最优解, 但相应的计算量也越大。考虑到计算量问题, 取 $\alpha = 0.5$ 能够适应大多数情况。

利用式(9)和式(10)计算拉格朗日乘子的值会产生很大的计算量, 并且不能保证式(9)和式(10)同时取得最优值, 解决这个问题的方法是: 利用二分法来求拉格朗日乘子, 在求解迭代过程中, 需要设定一个判断迭代是否停止的收敛条件, 如下:

$$\max(|\mathbf{X}^{(k+1)} - \mathbf{X}^{(k)}|) < \varepsilon \quad (11)$$

取 $\varepsilon = 0.01$, 即当设计变量矢量迭代更新值与更新前的值的差量的最大值小于 ε 时, 迭代终止。

2 算例和分析

2.1 自重载荷拓扑优化经典案例

图1 为一个二维的简支梁, 高 40 mm, 长

式中: k 为迭代的次数。

对目标函数关于设计变量求导数, 得到其灵敏度^[18]为

$$\frac{\partial C}{\partial x_i} = 2 \frac{\partial \mathbf{F}_i \mathbf{u}_i^T}{\partial x_i} - f'(x_i) \mathbf{u}_i^T \mathbf{k}_i \mathbf{u}_i = \frac{1}{2} v_i \rho g \cdot \mathbf{P}_0 \cdot \mathbf{u}_i^T - f'(x_i) \mathbf{u}_i^T \mathbf{k}_i \mathbf{u}_i \quad (8)$$

式中: $\mathbf{P}_0 = [0, -1, 0, -1, 0, -1, 0, -1]$ 为单元四节点载荷的位移方向矢量; \mathbf{F}_i 为第 i 个单元四节点等效载荷矢量; 案例中采用 RAMP 模型,

$$f'(x_i) = \frac{1 + q}{[1 + q(1 - x_i)]^2}$$

分析式(8)可知, 当区域应变能过小时, 由于数值化整误差的影响, 该区域的灵敏度会非常接近甚至相同, 从而导致不收敛的情况。

综上所述, 可以得到该案例的迭代公式及拉格朗日乘子的计算式^[18]为

120 mm, 两端的下方均铰链固定, 仅受该梁结构自身的重力作用。图2 为梁结构的有限元模型示意图, 单元类型为四节点矩形单元, 数量为 4 800, 厚度为 1。

对结构进行拓扑优化, 体积约束设置为 30%,

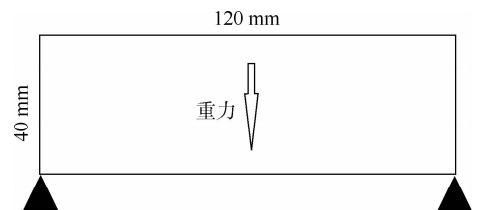


图1 自重载荷作用下的二维简支梁受力结构
Fig.1 Structural diagram of 2D simply supported beam under self-weight load

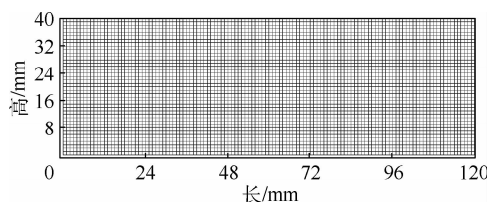


图2 待优化结构的有限元模型
Fig.2 Finite element model of structure to be optimized

经过 45 步迭代可以得到最终的优化结果。在迭代过程中,在第 1 次迭代到第 9 次迭代,结构发生非常大的改变;迭代到第 26 次时,优化的结构基本成型,趋于稳定,结构周围分布的中间密度单元逐渐收敛;从第 26 次迭代到第 45 次迭代,结构基本稳定,第 45 次迭代后得到了最终的优化结果,为拱形结构。图 3 为二维简支梁结构在自重载荷作用下的拓扑优化过程。

对优化出来的结构进行有限元分析,如图 4 所示,拱形结构的最大应力为 9.230×10^{-2} MPa。图 5 为优化前的简支梁结构,该结构的最大应力为 29.48×10^{-2} MPa。在仅受自重载荷作用下,拱形结构最大应力相对该结构减少了约 68.7%。

将简支梁重力等效为集中力施加于结构重心位置得到的拓扑优化结构,如图 6 所示,图中应力云图为对该结构进行实际自重作用情况下的结构应力分析所得。这个结构在自重载荷作用下最大应力为 10.72×10^{-2} MPa,相对于拱形结构增加了约 16%。



图 6 自重载荷作用下结构应力分析

Fig. 6 Structural stress analysis under self-weight load

2.2 二维悬臂梁在重力作用下的结构拓扑优化

图 7 为一个高 40 mm、长 120 mm 的悬臂梁,固定梁左边的端部,对自重载荷作用下的悬臂梁进行结构拓扑优化。

图 8 为自重载荷作用下的悬臂梁结构经过拓扑优化得到的优化结果。在结果中,距离固定端较远的那一端相对较模糊,单元处于中间密度,没有形成清晰明确的材料分布。

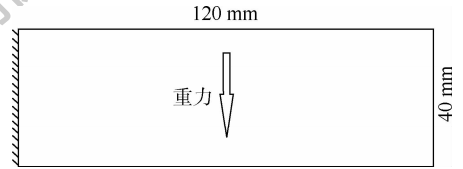


图 7 自重载荷作用下的二维悬臂梁结构示意图

Fig. 7 Schematic diagram of 2D cantilever beam under self-weight load

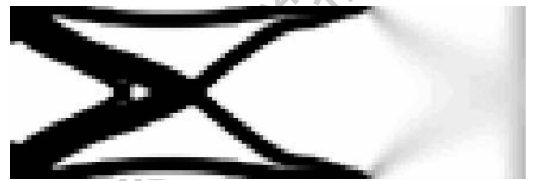


图 8 自重载荷作用下的二维悬臂梁拓扑优化结果

Fig. 8 Topology optimization results of 2D cantilever beam under self-weight load

图 9 为不同长高比下的悬臂梁进行结构拓扑优化的结果。结果表明,在长高比大于 120:60 时,优化结果没有形成完整的结构,在末端约 1/3 的区域存在中间密度单元,材料分布模糊;在长高比小于 120:60 时,优化结果呈现出一个完整的悬臂梁结构,但是材料仍然没有趋于完全收敛,在末端靠近边线的地方分布有一定的中间密度单元。这是因为悬臂梁在长高比较大时,其左右两端的应变能差距会很大,并且右端的应变能会很小,根据式(8)的分析,右端区域的灵敏度相对于左端的小很多,并且大小会非常接近甚至相同,导致材料分布不收敛,出现中间密度单元。另外,仅仅受自重载荷下的悬臂梁结构,优化结果中材料分布应该集中于固定端附近更为合理,这是由于拓扑优化算法中采用了网格过滤方法来控制棋盘格现象和网格依赖性问题,导致优化结果中材料并没有集中分布于固定端附近。根据图 9 所示的优化

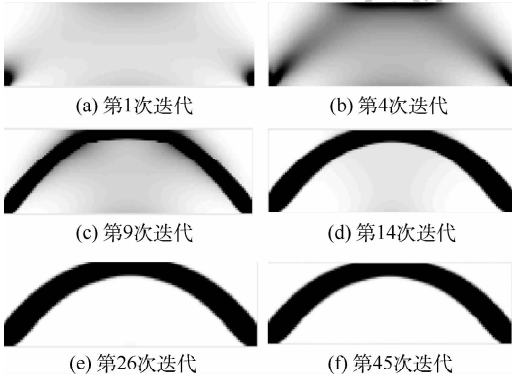


图 3 二维简支梁结构拓扑优化过程

Fig. 3 Topology optimization process of 2D simply supported beam structure

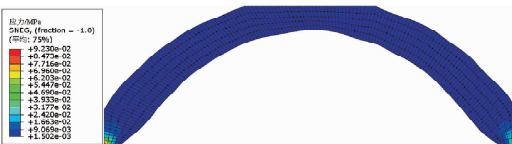


图 4 二维简支梁拓扑优化结果应力分析

Fig. 4 Stress analysis of 2D simply supported beam topology optimization results



图 5 优化前二维简支梁结构应力分析

Fig. 5 Stress analysis of 2D simply supported beam structure before optimization

结果,对自重载荷下悬臂梁的拓扑优化设计,优化出来的悬臂梁结构的长度是无法保证的。在对悬臂梁长度有要求的情况下,优化设计得到的结果并不能满足设计需求。

综合考虑上面的问题,采取在结构末端施加非结构质量的策略。末端非结构质量的添加保证了悬臂梁优化所得到结构的长度,同时可以使悬臂梁末端材料分布收敛,得到清晰的优化结果。施加非结构质量的悬臂梁结构模型如图 10 所示。

非结构质量的添加实际上改变了实际的工况,对拓扑优化得到的结果必然产生影响。但是通过研究非结构质量的合适大小,可以一定程度减少非结构质量对优化结果的影响。分别对图 10 所示悬臂梁进行拓扑优化设计,图 11 为在不同大小质量点情况下的拓扑优化结果, r 为质量点与设计前悬臂梁结构质量的比值。

上面的优化结果表明,在非结构质量大于等

于原悬臂梁结构质量 4% 的时候,悬臂梁结构的拓扑优化可以获得清晰明确的优化结果,收敛的情况有了很大的改善。分别对 $r=1$ 、 $r=0.1$ 和 $r=0.04$ 时优化得到的结果做自重载荷下的有限元分析,其应力云图分别如图 12 ~ 图 14 所示。如表 1 中的 3 种结构的应力峰值所示, $S_{r=1} > S_{r=0.1} > S_{r=0.04}$, 即非结构质量取到原悬臂梁质量的 4% 时优化能够获得相对更好的结构。

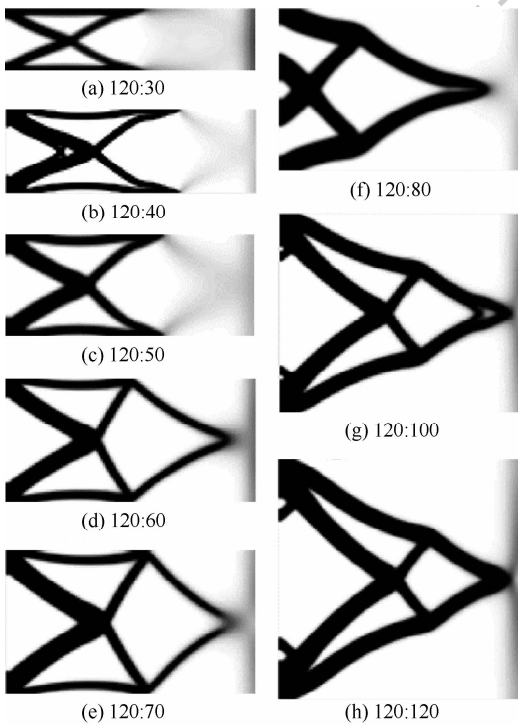


图 9 不同长高比下的悬臂梁拓扑优化结果
Fig.9 Topology optimization results of cantilever beams with different aspect ratios

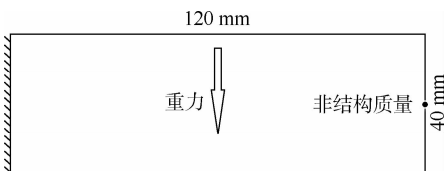


图 10 施加非结构质量的悬臂梁结构示意图
Fig.10 Schematic diagram of a cantilever beam with non-structural mass

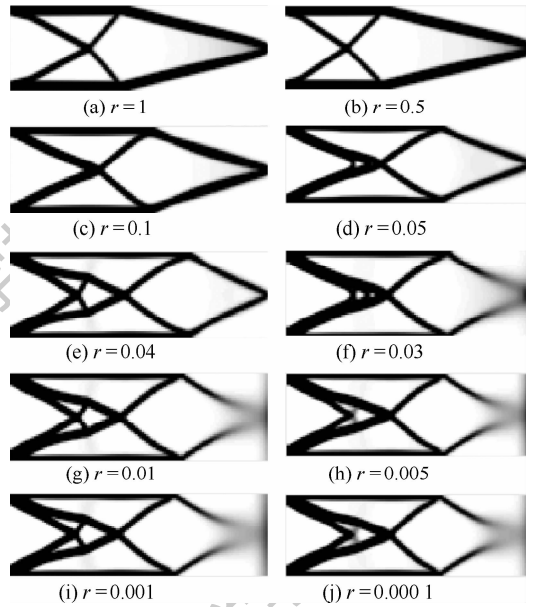


图 11 不同非结构质量情况下悬臂梁的拓扑优化结果
Fig.11 Topology optimization results of cantilever beams with different non-structural masses

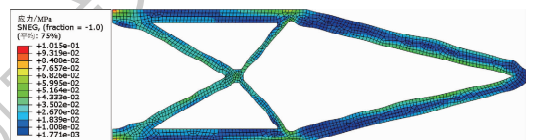


图 12 $r=1$ 悬臂梁拓扑优化结果应力分析
Fig.12 Stress analysis of $r=1$ cantilever beam topology optimization results

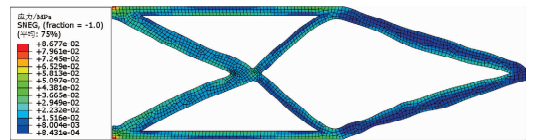


图 13 $r=0.1$ 悬臂梁拓扑优化结果应力分析
Fig.13 Stress analysis of $r=0.1$ cantilever beam topology optimization results

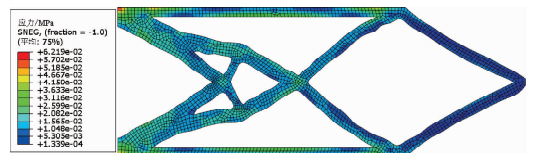


图 14 $r=0.04$ 悬臂梁拓扑优化结果应力分析
Fig.14 Stress analysis of $r=0.04$ cantilever beam topology optimization results

表 1 悬臂梁结构自重载荷下应力峰值

Table 1 Peak stress of cantilever beam structure under self-weight load

非结构质量比值	峰值应力/MPa
1	10.15×10^{-2}
0.1	8.677×10^{-2}
0.04	6.219×10^{-2}

3 结 论

采用基于导重准则法和 RAMP 模型的变密度法对自重载荷作用下的结构进行了拓扑优化研究,讨论了出现的问题及采取的解决措施。结合理论推导和算例的分析,可以得到以下结论:

1) 通过对灵敏度公式的分析,由于数值化整误差的存在,在结构的低应变能区域容易出现不收敛的情况,导致材料分布模糊,结合算例进行了说明。

2) 自重载荷根据材料分布均匀分配情况下进行拓扑优化相对于将自重等效集中施加于重心位置的情况会获得不同的优化结果,并且得到的结构在自重下性能更佳。

3) 在悬臂梁末端添加非结构质量可以有效解决悬臂梁末端材料分布模糊的情况,保证了悬臂梁的有效工作长度,但同时也改变了工况,选择合适大小的非结构质量可以在尽可能减少工况改变的影响的同时获得稳定收敛的优化结果。

4) 通过对比分析,在选择合适大小的非结构质量的情况下优化出来的结构相对于其他大小非结构质量能够获得更好的性能,表明了方法的有效性。

参考文献 (References)

- [1] 赵军鹏,王春洁. 载荷不确定条件下的结构拓扑优化算法[J]. 北京航空航天大学学报,2014,40(7):959-964.
ZHAO J P, WANG C J. Algorithm of structural topology optimization under loading uncertainty[J]. Journal of Beijing University of Aeronautics and Astronautics,2014,40(7):959-964 (in Chinese).
- [2] 付志方,王春洁. 载荷不确定的周期性结构稳健拓扑优化[J]. 北京航空航天大学学报,2017,43(4):747-753.
FU Z F, WANG C J. Robust topology optimization of periodic structures under uncertain loading[J]. Journal of Beijing University of Aeronautics and Astronautics,2017,43(4):747-753 (in Chinese).
- [3] 贾娇,程伟,龙凯. 基于 SIMP 法的周期性传热材料拓扑优化[J]. 北京航空航天大学学报,2015,41(6):1042-1048.
JIA J, CHENG W, LONG K. Topology optimization for periodic thermal conductive material using SIMP method[J]. Journal of Beijing University of Aeronautics and Astronautics, 2015, 41

(6):1042-1048 (in Chinese).

- [4] 李明高,李明,韩璐,等. 基于多目标的高速列车隔热结构拓扑优化[J]. 北京航空航天大学学报,2016,42(5):878-884.
LI M G, LI M, HAN L, et al. Multi-objective topological optimization of heat insulation structures used in high-speed train [J]. Journal of Beijing University of Aeronautics and Astronautics,2016,42(5):878-884 (in Chinese).
- [5] 郭连水,BANDI P,RENAUD J E. 一种大变形多空间域连续体结构拓扑优化方法[J]. 北京航空航天大学学报,2009,35(2):227-230.
GUO L S, BANDI P, RENAUD J E. Method of multi-domain topology optimization for continuum structures[J]. Journal of Beijing University of Aeronautics and Astronautics,2009,35(2):227-230 (in Chinese).
- [6] 刘丰睿,赵丽滨,韩邦成,等. 磁悬浮控制力矩陀螺框架结构的拓扑优化设计[J]. 北京航空航天大学学报,2010,36(4):455-458.
LIU F R, ZHAO L B, HAN B C, et al. Topology optimization design of frame structure for magnetic suspension control moment gyroscope[J]. Journal of Beijing University of Aeronautics and Astronautics,2010,36(4):455-458 (in Chinese).
- [7] 陶扬,王春洁,付志方. 卫星天线复合材料框架的铺层优化设计[J]. 北京航空航天大学学报,2017,43(5):1031-1037.
TAO Y, WANG C J, FU Z F. Layup design optimization of composite frame for satellite antenna[J]. Journal of Beijing University of Aeronautics and Astronautics,2017,43(5):1031-1037 (in Chinese).
- [8] 陈树勋,叶尚辉. 天线结构优化设计的准则法[J]. 固体力学学报,1984(4):482-497.
CHEN S X, YE S H. Guide-weight method for structural topology optimization[J]. Chinese Journal of Computational Mechanics,1984(4):482-497 (in Chinese).
- [9] BENDSOE M P. Optimal shape design as a material distribution problem[J]. Structural Optimization,1989,1(4):193-202.
- [10] PEDERSEN N L. Maximization of eigenvalues using topology optimization[J]. Structural and Multidisciplinary Optimization, 2000,20(1):2-11.
- [11] STOLPE M, SVANBERG K. An alternative interpolation scheme for minimum compliance topology optimization [J]. Structural and Multidisciplinary Optimization, 2001, 22(2):116-124.
- [12] BRUYNEEL M, DUYSINX P. Note on topology optimization of continuum structures including self-weight [J]. Structural and Multidisciplinary Optimization,2005,29(4):245-256.
- [13] LOPES R A, STUMP F V, SILVA E C N. Topology optimization of three dimensional structures under self-weight and inertial forces [C] //6th World Congress of Structural and Multidisciplinary Optimization,2005:5-6.
- [14] LIU X J, LI Z D, WANG L P, et al. Solving topology optimization problems by the guide-weight method[J]. Frontiers of Mechanical Engineering,2011,6(1):136-150.
- [15] LIU X J, LI Z D, CHEN X. A new solution for topology optimization problems with multiple loads: The guide-weight method [J]. Science China Technological Sciences, 2011, 54(6):

- 1505.
- [16] 陈祥,刘辛军. 基于 RAMP 插值模型结合导重法求解拓扑优化问题[J]. 机械工程学报,2012,48(1):135-140.
CHEN X, LIU X J. Solving topology optimization problems based on RAMP method combined with guide-weight method [J]. Journal of Mechanical Engineering,2012,48(1):135-140 (in Chinese).
- [17] 张晖,刘书田,张雄. 考虑自重载荷作用的连续体结构拓扑优化[J]. 力学学报,2009,41(1):98-104.
ZHANG H, LIU S T, ZHANG X. Topology optimization of continuum structures subjected to self weight loads [J]. Chinese Journal of Theoretical and Applied Mechanics, 2009, 41 (1): 98-104 (in Chinese).
- [18] XU H Y, GUAN L W, CHEN X, et al. Guide-weight method for topology optimization of continuum structures including body forces [J]. Finite Elements in Analysis and Design, 2013, 75: 38-49.
- [19] 许华扬,关立文,王立平,等. 惯性载荷下飞行模拟器大臂结构的拓扑优化[J]. 机械工程学报,2014,50(9):14-23.
XU H Y, GUAN L W, WANG L P, et al. Topology optimization for the arm of flight simulator under inertial loads [J]. Journal of Mechanical Engineering, 2014, 50 (9): 14-23 (in Chinese).
- [20] JAIN N, SAXENA R. Effect of self-weight on topological optimization of static loading structures [J]. Alexandria Engineering Journal, 2018, 57 (2): 527-535.
- [21] HOLMBERG E, THORE C J, KLARBRING A. Worst-case topology optimization of self-weight loaded structures using semi-definite programming [J]. Structural and Multidisciplinary Optimization, 2015, 52 (5): 915-928.
- [22] 陈树勋. 精密复杂结构的几种现代设计方法 [M]. 北京:北京航空航天大学出版社,1992:226-244.
CHEN X S. Several modern design methods for precise and complex structures [M]. Beijing: Beihang University Press, 1992:226-244 (in Chinese).
- [23] 陈树勋. 工程结构系统的分析,综合与优化设计 [M]. 北京:中国科学文化出版社,2008:99-108.
CHEN S X. Analysis, synthesis and optimization of engineering structural system [M]. Beijing: China Science Culture Publishing House, 2008:99-108 (in Chinese).

Topology optimization of cantilever structure with self-weight load based on guide-weight method

REN Yiru^{1,*}, XIANG Jianhui¹, HE Jie¹, NING Keyan², YANG Lingling²

(1. College of Mechanical and Vehicle Engineering, Hunan University, Changsha 410082, China;
2. China North Vehicle Research Institute, Beijing 100072, China)

Abstract: Aiming at the problem of non-convergence of the material distribution at the end of the cantilever beam structure under self-weight load, a topology optimization method for the cantilever beam structure under self-weight load is proposed to solve the problem. According to the principle of virtual work equal, the load equivalent method was established by the relationship among the shape function of the four-node rectangular element, the unit volume density and the mass. According to the Kuhn-Tucker condition of the optimization model, the guide-weight criterion was derived, the sensitivity formula of the objective function was obtained, and the iterative formula considering the topology optimization of the self-weight load was derived. Aimed at the problem of ambiguity of material distribution in the end region of the topological optimization of a cantilever beam structure under self-weight load, a solution strategy combining variable density method and non-structural mass was studied, and the influence of typical factors on the topological structure was revealed. The results show that this method can solve the problem of fuzzy material distribution at the end of a cantilever beam under self-weight.

Keywords: topology optimization; guide-weight method; variable density method; self-weight load; Rational Approximation of Material Properties (RAMP) model

Received: 2020-05-18; **Accepted:** 2020-06-05; **Published online:** 2020-08-05 16:50

URL: kns.cnki.net/kcms/detail/11.2625.V.20200805.1331.006.html

Foundation item: Innovation Research Project of Basic Product of Ministry of Industry and Information Technology of China (237099000000170008)

* **Corresponding author.** E-mail: renyiru@hnu.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2020.0197

基于改进遗传算法的航空集装箱装载问题研究



张长勇*, 翟一鸣

(中国民航大学 电子信息与自动化学院, 天津 300300)

摘 要: 针对标准遗传算法求解装载方案时存在收敛速度慢、易早熟、寻优结果欠佳的问题, 基于拟人装载策略, 提出了一种以集装箱空间利用率最大为目标, 考虑货物装载顺序、体积、质量、重心、不重叠等多种实际约束的改进遗传算法。首先, 采用与货物放置状态相结合的实数编码, 随机产生初始种群; 然后, 在常规选择操作中加入最优解保存策略, 并将稳定性、支撑限制、重心约束考虑到进行线性尺度变换后的适应度函数中, 以此来计算每种装载方案的评估值; 最后, 输出评估值最高的方案作为最优装载方案。实验采用异构性不同的测试算例进行性能测试, 结合3组具体货物装载数据证明算法的普适性与实用性。结果表明: 所提算法在求解强异构货物装载过程中具有较好的优化效果, 适用于求解集装箱装载问题。与标准遗传算法相比, 收敛性与搜索速度有所提高, 2种不同箱型的集装箱空间利用率分别提高了3.82%和3.66%, 运行时间分别缩短了7.9 s和5.58 s, 能快速找到最优装载方案, 可有效解决规则、不规则集装箱的货物装箱问题。基于MATLAB软件实现装载方案的可视化, 为集装箱的实时装载决策提供了理论基础。

关键词: 改进遗传算法; 拟人装载策略; 实际约束; 不同箱型集装箱; 可视化

中图分类号: V219; TP311

文献标志码: A **文章编号:** 1001-5965(2021)07-1345-08

目前, 在航空运输过程中, 集装箱装载大多依靠工人的经验, 因此存在一定的劳动效率低、装载效果差等问题, 装载和规划过程具有很大的优化空间。集装箱装载问题的研究, 对降低物流配送成本、提高经济效益等都有着很积极的意义^[1], 其本质属于三维装箱问题, 为改进装箱算法的实用性, 许多学者对装箱优化问题提出了有建设性的意见。按照常用算法类型, 可以分为启发式算法与遗传算法。

启发式算法是在直观、经验的基础上抽象化后得到的算法, 适合解决不是很复杂的装箱问题。根据实际约束及实现目标的不同, 学者们提出了多种启发式的求解方法。Maxence 和 Manuel^[2] 基于改进的最佳拟合启发式算法, 用于生成将初始

可行方案装载到箱中; 刘胜等^[3-4] 将树搜索的思想用于装箱, 使用箱子、片、条、层、实体的顺序进行装载; Aline 等^[5] 通过构建数学模型, 将一组不规则和规则的块分配给较大的矩形或不规则容器, 最大程度地减少了材料或空间的浪费; Mauro 等^[6] 研究了经典装箱问题的扩展, 提出一种综合不同方法的整体算法以找到最小数量的箱来装载所有物品; Yu 和 Amarnath^[7] 提出曲折排序的启发式算法, 构造了可行性保留规则以确保每个盒子的分配; 何琨等^[8-10] 利用占角策略, 通过占角动作提出了一种占角空间最小的穴度算法。

遗传算法作为鲁棒搜索算法, 不依赖问题的本质, 更适合解决复杂问题得到近似最优解。崔

收稿日期: 2020-05-20; 录用日期: 2020-09-04; 网络出版时间: 2020-09-22 15:52

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20200922.1507.003.html

基金项目: 国家自然科学基金(51707195); 中央高校基本科研业务费专项资金(3122016A009)

* 通信作者. E-mail: cyzhang@cauc.edu.cn

引用格式: 张长勇, 翟一鸣. 基于改进遗传算法的航空集装箱装载问题研究[J]. 北京航空航天大学学报, 2021, 47(7): 1345-1352.
ZHANG C Y, ZHAI Y M. Air container loading based on improved genetic algorithm [J]. Journal of Beijing University of Aeronautics and Astronautics, 2021, 47(7): 1345-1352 (in Chinese).

会芬等^[11]在遗传运算中使用排序选择方法和部分交叉来对遗传进行改进,在考虑部分约束下提高了集装箱的装载效率,但未考虑货物承重性、装箱顺序等因素;张钧和贺可太^[12]利用混合遗传、模拟退火2种算法结合求解模型,提出两段式的编码方式充分优化空间利用率,但是采用模拟退火法确定适应度,运行时间较长;Menghani和Guhani^[13]在考虑多种实际约束的情况下将货物合理装进多个容器中,但其搜索速度与收敛性很容易受到初始化条件的影响;代爱民^[14]考虑了货物装箱的及时性和顺序,但并未考虑货物的承载能力、重心分布等约束;于明正等^[15]提出基于2个搜索方向的遗传层,分别搜索可行方案的广度和深度以提高优化效率,但其约束条件考虑较少,不能满足实际应用需要。

综上所述,目前大多数算法求解装箱问题时,未考虑货物装载顺序约束的离线装载,以及垛形稳定性等实际装载过程中的现实约束,无法保证货物运输过程中的安全性。另外,对于装箱问题的研究较多集中在规则集装箱的装载上,对不规则集装箱的求解问题研究较少。

本文以航空货运集装箱装载为研究背景,考虑实际装载过程中的诸多现实约束,在此基础上搭建数学模型,提出一种改进遗传算法,以集装箱空间利用率为优化目标,结合稳定性、重心及支撑限制约束构建合理的适应度函数,并加入适应度尺度变换、最优解保存策略以进一步提高优化效果。采用异构性由弱到强的测试算例及3组具体货物装载数据进行实验,验证了算法的有效性与实用性。根据算法输出结果开发可视化程序,为工程应用创造了条件。

1 问题描述

1.1 符号说明及数学模型

以集装箱左后下点为原点,底面为XY平面,垂直底面向上为Z轴建立坐标系。 η_i 为0/1变量,若货物*i*已装入则 $\eta_i = 1$,若未装入则 $\eta_i = 0$; w_i, h_i, d_i, v_i, m_i 和 $[gx_i, gy_i, gz_i]$ 分别为货物*i*的长、宽、高、体积、质量和重心坐标; E_i 和 F_i 分别为货物*i*所承受的重量和最大承载力; B_i 和 C_i 分别为货物*i*的顺序号和码放序号; $(x_i, y_i, z_i), (x'_i, y'_i, z'_i)$ 分别为货物*i*左后下、右前上角坐标; $[conx_1, conx_2], [cony_1, cony_2], [0, conz]$ 分别为*x, y, z*方向的重心安全区间; V 和 M 分别为集装箱的体积和载重量。

给定航空集装箱和一批待装载货物 $B = \{b_1, b_2, \dots, b_n\}$, 装载要求为:在满足实际约束的情况下,得到使集装箱空间利用率最大的布局方案。

据此,目标函数如下:

$$U = \frac{\sum_{i=1}^n \eta_i v_i}{V} \quad i = \{1, 2, \dots, n\} \quad (1)$$

在实际装载过程中存在很多现实问题,假设条件可以简化优化模型,降低装箱问题的复杂度,因此对于假设条件的减少是非常必要的。具体讨论的假设条件围绕以下几点:①货物形状均为长方体;②货物在最大承重范围之内可接受多层装载;③货物重心为其几何中心;④不考虑货物在装载过程中出现的挤压变形等。

1.2 约束条件处理

1) 质量与体积约束。装载货物的总质量、总体积不能超过实际所允许的最大装载质量与体积,即

$$\sum_{i=1}^n m_i \eta_i \leq M \quad (2)$$

$$\sum_{i=1}^n v_i \eta_i \leq V \quad (3)$$

2) 货物装载顺序约束。考虑货物装箱过程中的顺序装载,货物按当前到来的顺序进行装载,即

$$B_i = C_i \quad (4)$$

3) 不重叠约束。货物之间不能重叠,即

$$\begin{cases} x_i + \frac{1}{2}w_i - x_j - \frac{1}{2}w_j \geq \frac{1}{2}w_i + \frac{1}{2}w_j \\ y_i + \frac{1}{2}h_i - y_j - \frac{1}{2}h_j \geq \frac{1}{2}h_i + \frac{1}{2}h_j \\ z_i + \frac{1}{2}d_i - z_j - \frac{1}{2}d_j \geq \frac{1}{2}d_i + \frac{1}{2}d_j \end{cases} \quad (5)$$

4) 稳定性约束。保证垛形稳定,货物不可悬空,其下需有一半以上的面积被支撑,即货物*i*的接触率 T_i 满足:

$$\frac{\{h_i + h_j - [\max(y_i + h_i, y_j + h_j) - \min(y_i, y_j)]\} \cdot \{w_i + w_j - [\max(x_i + w_i, x_j + w_j) - \min(x_i, x_j)]\}}{(w_i h_i)} > \frac{1}{2} \quad (6)$$

5) 支撑限制约束。在实际装载过程中,要考虑到货物的承载能力,避免出现货物毁坏的情况,即

$$E_i = \sum_q m_q + \sum_p m_p \frac{S_{pi}}{w_p h_p} < F_i \quad (7)$$

式中:

$$S_{pi} = [(x'_i - x_i + x'_p - x_p) - (\max\{x'_p, x'_i\} -$$

$$\min\{x_p, x_i\})[(y'_i - y_i + y'_p - y_p) - (\max\{y'_p, y'_i\} - \min\{y_p, y_i\})]$$

$$q \in \{q \mid \eta_q \eta_i = 1, z_q \geq z'_i, x_q \geq x_i, y_q \geq y_i, x'_q \leq x'_i, y'_q \leq y'_i\}$$

$$p \in \{p \mid \eta_p \eta_i = 1, z_p \geq z'_i, \max\{x'_p, x'_i\} - \min\{x_p, x_i\} < x'_i - x_i + x'_p - x_p, \max\{y'_p, y'_i\} - \min\{y_p, y_i\} < y'_i - y_i + y'_p - y_p\}$$

6) 重心约束。为了保证运输货物的安全性,要求其重心必须在一定的范围内,即

$$\begin{cases} \text{con}x_1 \leq \frac{\sum_{i=1}^n \eta_i m_i g x_i}{\sum_{i=1}^n \eta_i m_i} \leq \text{con}x_2 \\ \text{con}y_1 \leq \frac{\sum_{i=1}^n \eta_i m_i g y_i}{\sum_{i=1}^n \eta_i m_i} \leq \text{con}y_2 \\ \frac{\sum_{i=1}^n \eta_i m_i g z_i}{\sum_{i=1}^n \eta_i m_i} \leq \text{con}z \end{cases} \quad (8)$$

1.3 拟人装载策略

采用不同的装载策略^[16]会生成不同的装载方案,从而对集装箱的空间利用率产生影响,因此要选用合适的装载方法。采用可放置点构建、排序、首个匹配、参考线引入的拟人装载策略^[17]对货物进行装载。在码放过程中,水平与垂直方向上引入2条参考线以引导货物放置过程,利用更新可放置点的方法来确定货物的放置位置。拟人装载策略流程如图1所示。

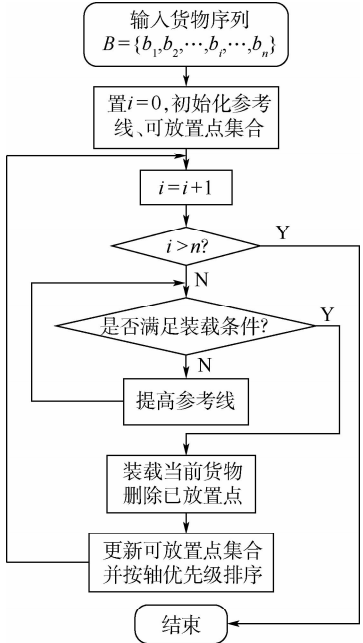


图1 拟人装载策略流程

Fig. 1 Anthropomorphic loading strategy flowchart

在该策略中,首先,将初始可放置点设为(0, 0, 0),并将水平、垂直参考线初始化为0。然后,按照货物到来顺序依次装载货物,对每个货物判断可放置点是否满足装载条件,装载条件为该货物不能与集装箱及其他货物重叠,且满足 $x + w_i \leq L_x, z + d_i \leq L_z$, 即不超过2条参考线,若不满足,通过水平参考线 L_x 执行不同的操作。最后,若当前货物找到了可以放置的位置,将其放在该位置,同步更新待装载货物与可放置点序列。算法结束时,输出货物装箱方案及空间利用率。

2 改进的遗传算法

遗传算法具有快速随机的全局搜索力,但收敛速度和局部搜索效果欠佳。为解决标准遗传算法过早收敛的问题,加入罚函数、适应度尺度变换、最优解保存策略来对其进行改进。

2.1 编码与解码

遗传算法的编码大多使用二进制编码,而应用在货物装箱这类问题上显然是不可行的。原因:①货物坐标使用二进制表示,会出现复杂度随着货物数量的增加而呈指数增长,降低了算法的性能;②坐标的变化描述了待装载货物的位置变化,有可能会货物重叠的现象。

货物在集装箱中的装载包含3个要素:装箱顺序、放置状态与放置位置。本文研究的是将货物按照其码放顺序装入到集装箱中,故编码时仅考虑第2个因素,采用实数编码,货物的每种布局方案对应 $S = (S_1, S_2, \dots, S_i, \dots, S_n), S_1 \sim S_n$ 为货物放置状态,分别用整数1,2,3,4,5,6代表货物的6种放置状态(见图2)。第3个因素在解码时完成,根据拟人装载策略实现解码操作。

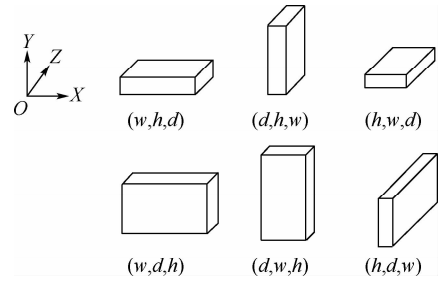


图2 货物放置状态

Fig. 2 Cargo orientations

2.2 解的不可行性与罚函数

惩罚是遗传算法解决受约束的优化问题的最常用办法,由于遗传操作可能会得到不可行的后代,加入惩罚项对违反约束的个体给予相应的惩罚,其本质是通过惩罚不可行的方案,把带约束的

问题转换成无约束问题。综合考虑稳定性约束、支撑限制约束、重心约束,惩罚项如下。

保证垛形稳定,即满足稳定性约束:

$$p_1(S) = \begin{cases} 0 & T_i > \frac{1}{2} \\ 1 & \text{否则} \end{cases} \quad (9)$$

货物所载全部上层货物总重量小于其最大载重量,即满足支撑限制约束:

$$p_2(S) = \begin{cases} 0 & E_i < F_i \\ 1 & \text{否则} \end{cases} \quad (10)$$

垛形重心在合理的范围以内,即满足重心约束:

$$p_3(S) = \begin{cases} 0 & \text{con}x_1 \leq \frac{\sum_{i=1}^n \eta_i m_i g x_i}{\sum_{i=1}^n \eta_i m_i} \leq \text{con}x_2 \\ 1 & \text{否则} \end{cases} \quad (11)$$

$$p_4(S) = \begin{cases} 0 & \text{con}y_1 \leq \frac{\sum_{i=1}^n \eta_i m_i g y_i}{\sum_{i=1}^n \eta_i m_i} \leq \text{con}y_2 \\ 1 & \text{否则} \end{cases} \quad (12)$$

$$p_5(S) = \begin{cases} 0 & \frac{\sum_{i=1}^n \eta_i m_i g z_i}{\sum_{i=1}^n \eta_i m_i} \leq \text{con}z \\ 1 & \text{否则} \end{cases} \quad (13)$$

记

$$\text{Penalties}(S) = \sum_{k=1}^5 p_k(S) \quad (14)$$

则式(9)~式(13)任何值为1,皆为不可行方案。

2.3 适应度尺度变换

研究目标是使集装箱空间利用率最大,即最小化集装箱的浪费空间,将适应度定义成空间利用率函数,如式(15)所示。为提高个体间竞争性,防止算法早熟,对其进行线性尺度变换,如式(16)所示。加入惩罚项后的评估函数如式(17)所示。

$$F(S) = \begin{cases} \frac{\sum_{i=1}^n \eta_i v_i}{V} & \sum_{i=1}^n \eta_i v_i \leq V \\ 0 & \text{否则} \end{cases} \quad (15)$$

$$\text{Fitness}(S) =$$

$$\begin{cases} \frac{F_{\text{avg}}}{F_{\text{avg}} - F_{\text{min}}} F + \frac{F_{\text{min}} F_{\text{avg}}}{F_{\text{avg}} - F_{\text{min}}} & \sum_{i=1}^n \eta_i v_i \leq V \\ 0 & \text{否则} \end{cases} \quad (16)$$

$$\text{Eval}(S) = \text{Fitness}(S) \cdot (5 - \text{Penalties}(S)) \quad (17)$$

式中: F_{min} 为 F 的最小值; F_{avg} 为 F 的平均值。

2.4 最优解保存策略

由于选择、交叉与变异具有随机性,可能会破坏评估值好的个体。为使最好的个体能够尽可能多地遗传到下一代,确保算法的全局收敛,采用最优解保存策略来实现优胜劣汰。通过设置参数 $\delta = 10\%$,确定被保留不需进行交叉、变异的个体。若上一代种群数为50,则通过计算评估值,将评估值前5的个体直接复制到下一代,减少了破坏最佳方案的可能性。

3 基于拟人装载策略的改进遗传算法

根据上述拟人装载策略和改进的遗传算法,设计了基于拟人装载策略的改进遗传算法,算法流程如图3所示,步骤如下:

步骤1 将集装箱的型号、尺寸与货物的基本信息输入程序,设置改进遗传算法的相关参数。

步骤2 考虑到根据特定的排序规则定义初始种群的生成会导致优化结果受到限制,采用随机生成初始种群的方式对装载结果进行寻优。

步骤3 根据拟人装载策略实现解码操作,以实现货物的装箱。

步骤4 根据式(2)~式(5)判断是否满足质量、体积、货物装载顺序与不重叠约束,若满足,转步骤5;否则,转步骤10。

步骤5 根据式(16),得到线性尺度变换后的适应度函数。

步骤6 根据式(9)~式(14)、式(17)计算加入稳定性约束、支撑限制约束、重心约束3项惩罚的评估函数。

步骤7 轮盘赌实现个体选择。

步骤8 判断个体是否直接成为下一代。若满足,直接复制到下一代中;若不满足,利用双点交叉、顺序逆转实现交叉与变异。

步骤9 将最优解保存策略与改进遗传算法得到的解结合起来,产生新一代种群。

步骤10 进行下一次循环,判断是否满足终止条件。若满足,输出最佳装载方案;若不满足,返回步骤8。

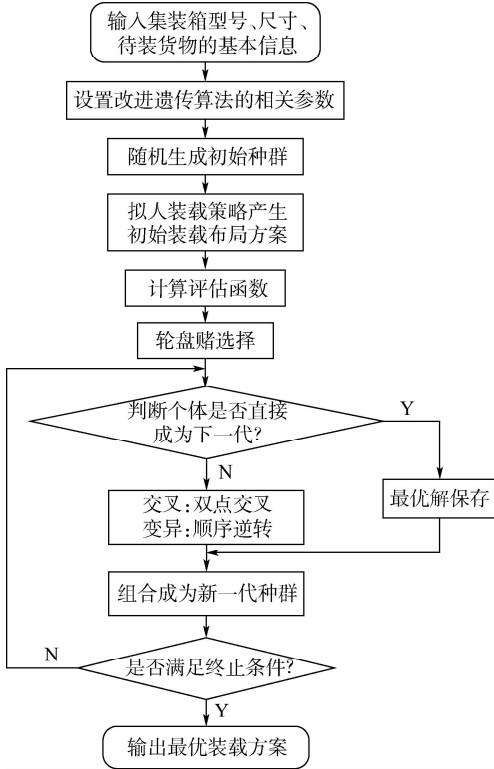


图 3 改进遗传算法流程

Fig. 3 Improved genetic algorithm flowchart

4 实验方案

采用同类三维装箱研究下的测试算例对本文算法进行性能测试,结合 3 组具体货物装载数据验证算法的实用性,并基于 MATLAB 实现装载效

果的可视化。

在 Visual Studio 2010 环境下,计算机型号为 Intel(R) Core(TM) i5-7200U CPU 2.50 GHz,操作系统为 Windows 10。遗传算法参数设为: {种群规模,交叉,变异,终止条件} = {50, 0.6, 0.1, 200}。

4.1 算例对比分析

与同类三维装箱研究用相同的测试算例进行比较验证,以证明本文算法的有效性和优越性。现有文献中大多是研究离线装箱问题,未考虑货物装载顺序约束,而本文研究的集装箱实时装载问题考虑了货物装箱的及时性和顺序,属于半在线装箱问题的研究范畴,两者不具有可比性。目前,仅发现文献[14]和本文属于同类装箱问题研究,故选取该文献中异构性由弱到强的算例进行实验。

表 1 为文献[14]的混合免疫遗传算法与改进遗传算法的对比结果,对每组数据分别做 5 次测试,得到平均空间利用率。可以发现,当货物异构性逐渐增强时,算法的优化效果总体呈上升趋势。经过改进遗传算法进行装载优化后,异构性较强的无尺寸相同的货物装载算例也可以获得较高的空间利用率,分别为 80.11% 和 84.47%,优于文献[14]应用的混合免疫遗传算法求解装载问题的结果。结果表明,本文算法在求解强异构货物装载问题时有明显的优势,有较好的优化效果。

表 1 算例测试结果对比

Table 1 Comparison of example test results

货物特征	编号	长	宽	高	平均空间利用率/%	
					混合免疫遗传算法	改进遗传算法
2 个尺寸相同	1	200	160	80 ~ 160	96.34	95.89
	2	200	100 ~ 200	100	91.92	92.43
	3	160 ~ 320	160	100	87.71	88.26
1 个尺寸相同	4	200	100 ~ 200	50 ~ 100	86.20	89.13
	5	140 ~ 280	140	50 ~ 100	83.20	86.56
	6	120 ~ 240	80 ~ 160	80	82.94	85.54
无尺寸相同	7	200 ~ 300	160 ~ 240	80 ~ 120	82.16	84.47
	8	200 ~ 400	160 ~ 320	80 ~ 160	72.75	80.11

4.2 实例验证

通过 4.1 节算例对比分析可知,本文算法在求解集装箱实时装载问题中有一定的优越性。为验证改进遗传算法可以解决规则、不规则集装箱的货物装载问题,将算法应用于实际集装箱装载。选用在航空运输领域应用广泛的 AMA、AKE 集装

箱作为待装载集装箱,其参数如表 2 所示。从某航空公司获取实际货物数据进行验证,采集了 3 组不同批次的货物数据,先后输入算法中进行实验。其中,第 1 组为较大型货物,第 2 组为混合型货物,第 3 组为较小型货物。3 个实验组组内货物基本信息如表 3 所示。

表2 集装箱参数

Table 2 Container parameters

箱型	尺寸/cm	可载重量/kg	适载机型
AMA 集装箱	$\{w, h, d\} = \{318, 214, 214\}$	6 444	747F
AKE 集装箱	$\{w_1, w_2, h, d\} = \{201, 156, 154, 163\}$	1 488	747、777、747F

表3 货物基本参数

Table 3 Basic cargo parameters

参数	第1组	第2组	第3组
平均长度/cm	65.32	51.56	45.74
平均宽度/cm	50.21	44.47	36.98
平均高度/cm	57.96	52.53	46.32
平均质量/kg	46.56	37.18	23.27
货物件数	50	100	150

基于算法的寻优过程,所考虑的约束条件较为全面,与实际工作环境相符合。将标准遗传算法与改进遗传算法进行对比,其中标准遗传算法

由拟人装载策略结合改进前的遗传算法得到。2种算法独立运行20次,对比结果如表4所示。结果表明,算法适用于解决规则、不规则集装箱的货物装箱问题,实用性较好,能满足实际工程需要。改进遗传算法所产生的方案有着较高的空间利用率,避免了不必要的空间浪费,能以较快的速度得到较为理想的装箱方案。究其原因,改进遗传算法加入罚函数、线性尺度变换、最优解保存策略后,算法的局部、全局寻优能力得到提高,能够更快速地找到最佳装载方案。

表4 不同箱型2种算法的对比结果

Table 4 Comparison results of two algorithms for different container types

箱型	货物实验组	标准遗传算法			改进遗传算法		
		平均空间利用率/%	平均装载货物件数	平均运行时间/s	平均空间利用率/%	平均装载货物件数	平均运行时间/s
AMA 集装箱	第1组	85.21	37	18.54	88.45	40	10.43
	第2组	85.89	75	29.37	89.89	80	19.65
	第3组	86.75	104	33.41	90.97	110	27.56
	平均	85.95	72	27.11	89.77	77	19.21
AKE 集装箱	第1组	84.66	22	12.53	87.67	25	6.28
	第2组	84.98	36	19.12	88.84	40	15.79
	第3组	85.29	49	25.63	89.42	55	18.45
	平均	84.98	36	19.09	88.64	40	13.51

遗传算法的收敛性是有效判断当前解是否达到最优解的恰当准则。文献[18]中 Rudolph 用马尔可夫链证明,在标准遗传算法中采用优秀个体保护策略,能保证算法最终收敛到全局最优解。本文算法加入了最优解保存策略,使优秀个体能直接进入下一代种群,故可以收敛于全局最优解。为对比改进前后遗传算法得到最优方案的性能,对 AMA 集装箱中的第3组货物数据得到的仿真结果进行统计。图4为评估值随迭代次数发生变化的对比效果。从收敛性角度看,随着优化过程的进展,评估函数最优值逐渐收敛。改进遗传算法在进化80代左右时,评估值趋于稳定,而标准遗传算法在53代开始出现“上升缓慢”现象,迭代到145代时才能趋于稳定,且性能较差。从搜索速度角度看,改进遗传算法能在较快的时间里搜索到最优解,搜索速度明显优于标准遗传算法。

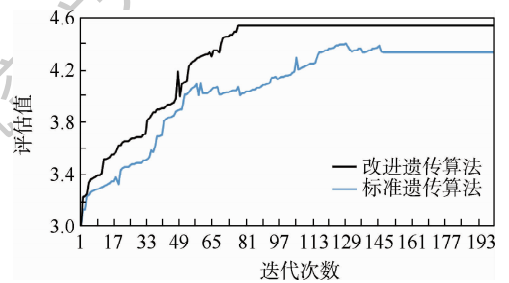


图4 优化性能比较

Fig. 4 Optimized performance comparison

先定义并整理输入数据与绘图空间,调用 meshgrid 函数根据货物放置的左后下、右前上角坐标生成三维网络采样点,再调用 slice 函数绘制每个采样点在垂直于 X、Y、Z 轴平面上的切片图,最终用不同颜色填充每个切片的颜色。

可视化货物的装箱方案,能直观地反映货物在箱内的码放位置、评估方案可行性,有必要根据算法输出结果开发可视化程序。在 MATLAB 中,

图5和图6分别为改进遗传算法得到的2种集装箱箱型的装载效果。可见,算法对于不同种类的强异构货物有着较好的适应性,垛形形状较为规则,能够满足规则、不规则集装箱的空间约束。货物摆放紧凑,满足了装载过程中稳定性的要求。

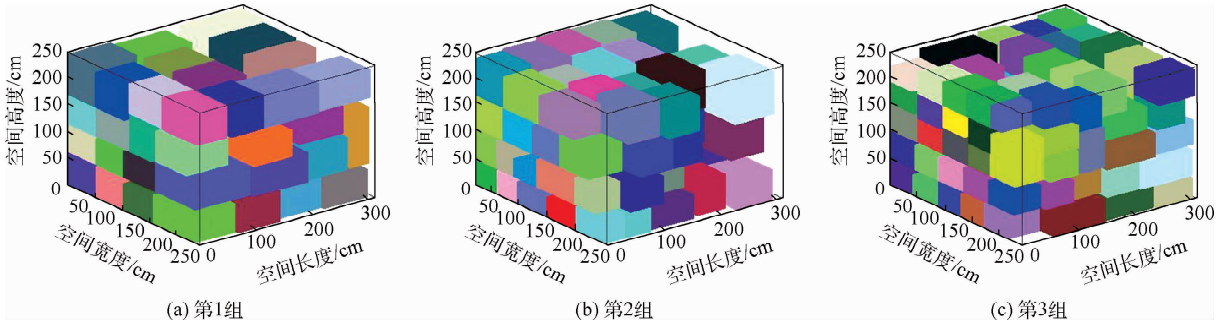


图5 规则集装箱装箱效果

Fig. 5 Regular packing of container

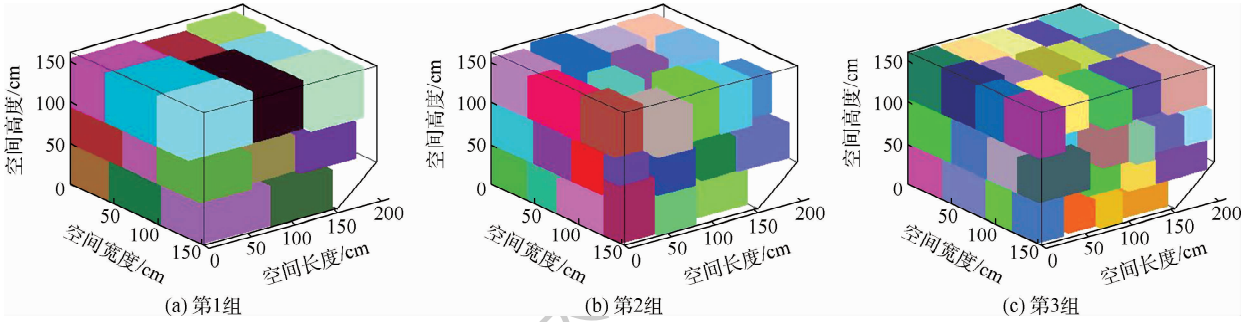


图6 不规则集装箱装箱效果

Fig. 6 Irregular packing of container

5 结论

1) 在考虑实际装载过程中的7个现实约束下,提出了一种适用于不同箱型的改进遗传算法,通过优化货物放置状态来最大化地利用集装箱装载空间,为复杂集装箱的实时装载问题提供了有效的解决办法。

2) 采用异构性不同的测试算例进行性能测试,证明了本文算法在求解集装箱实时装载问题中有一定的优越性。结合3组具体货物装载数据进行实验,证明了本文算法收敛性好,搜索速度快,能在获得最佳装载方案的同时缩短运行时间,对强异构货物有着较好的适应性与有效性。

3) 通过MTALB实现了集装箱布局方案的可视化,将货物按照其装载顺序合理的装载到集装箱中,得到空间利用率最大的布局方案,为实时装载决策提供了理论基础。

参考文献 (References)

[1] RAMOS A G, SILVA E, OLIVEIRA J F. A new load balance methodology for container loading problem in road transportation [J]. European Journal of Operational Research, 2018, 266 (3):1140-1152.

[2] MAXENCE D, MANUEL I. Enhanced pseudo-polynomial formulations for bin packing and cutting stock problems [J]. Informatics Journal on Computing, 2020, 32 (1):101-119.

[3] 刘胜,沈大勇,商秀芹,等. 求解三维装箱问题的多层树搜索算法[J]. 自动化学报, 2020, 46 (6):1178-1187.

LIU S, SHEN D Y, SHANG X Q, et al. Multi-layer tree search

algorithm for solving three-dimensional packing problem [J]. Journal of Automation, 2020, 46 (6):1178-1187 (in Chinese).

[4] LIU S, TAN W, XU Z, et al. A tree search algorithm for the container loading problem [J]. Computers & Industrial Engineering, 2014, 75:20-30.

[5] ALINE A S, FRANKLINA M B, TOLEDO J, et al. Irregular packing problems: A review of mathematical models [J]. European Journal of Operational Research, 2020, 282 (3):803-822.

[6] MAURO D, FABIO F, MANUEL I. A branch-and-price algorithm for the temporal bin packing problem [J]. Computers and Operations Research, 2020, 114:1-16.

[7] YU F, AMARNATH B. Heuristic/meta-heuristic methods for restricted bin packing problem [J]. Journal of Heuristics, 2020, 26:637-662.

[8] 何琨,黄文奇. 基于动作空间的三维装箱问题的确定性高效率求解算法[J]. 计算机学报, 2014, 37 (8):1786-1793.

HE K, HUANG W Q. A deterministic and efficient solution algorithm for the three-dimensional packing problem based on action space [J]. Journal of Computer Science, 2014, 37 (8):1786-1793 (in Chinese).

[9] 何琨,黄文奇,胡骞. 基于动作空间的求解三维矩形装箱问题的穴度算法[J]. 计算机科学, 2010, 37 (10):181-183.

HE K, HUANG W Q, HU Q. Acuity algorithm for solving three-dimensional rectangular boxing problem based on action space [J]. Computer Science, 2010, 37 (10):181-183 (in Chinese).

[10] 何琨,黄文奇. 三维矩形 Packing 问题的拟人求解算法[J]. 中国科学:信息科学, 2010, 40 (12):1586-1595.

HE K, HUANG W Q. Anthropomorphic algorithm for solving three-dimensional rectangular Packing problems [J]. Science in China: Information Science, 2010, 40 (12):1586-1595 (in Chinese).

[11] 崔会芬,许佳瑜,朱鸿国,等. 基于改进遗传算法的三维单箱装箱问题研究[J]. 工业工程与管理, 2018, 23 (1):86-89.

- CUI H F, XU J Y, ZHU H G, et al. Research on three-dimensional single box packing based on improved genetic algorithm [J]. *Industrial Engineering and Management*, 2018, 23(1): 86-89 (in Chinese).
- [12] 张钧, 贺可太. 求解三维装箱问题的混合遗传模拟退火算法[J]. *计算机工程与应用*, 2019, 55(14): 32-39.
- ZHANG J, HE K T. Hybrid genetic simulated annealing algorithm for solving the three-dimensional packing problem [J]. *Computer Engineering and Applications*, 2019, 55(14): 32-39 (in Chinese).
- [13] MENGHANI D, GUHA A. Packing boxes into multiple containers using genetic algorithm [J]. *Journal of the Institution of Engineers (India): Series C*, 2016, 97(3): 441-450.
- [14] 代爱民. 基于混合免疫遗传算法的半在线三维装箱问题研究[D]. 重庆: 重庆大学, 2018.
- DAI A M. Research on semi-online 3D packing problem based on hybrid immune genetic algorithm [D]. Chongqing: Chongqing University, 2018 (in Chinese).
- [15] 于明正, 徐斌, 陈佳. 基于双层启发式遗传算法的三维装箱问题[J]. *科学技术与工程*, 2020, 20(5): 2042-2047.
- YU M Z, XU B, CHEN J. Three-dimensional packing problem based on double-layer heuristic genetic algorithm [J]. *Science Technology and Engineering*, 2020, 20(5): 2042-2047 (in Chinese).
- [16] 李鹏, 汤勇. 三维货物装箱问题的研究进展[J]. *铁道科学与工程学报*, 2015, 12(5): 1232-1242.
- LI P, TANG Y. Research progress of three-dimensional cargo packing [J]. *Journal of Railway Science and Engineering*, 2015, 12(5): 1232-1242 (in Chinese).
- [17] 张德富, 魏丽军, 陈青山, 等. 三维装箱问题的组合启发式算法[J]. *软件学报*, 2007, 18(9): 2083-2089.
- ZHANG D F, WEI L J, CHEN Q S, et al. Combined heuristic algorithm for three-dimensional packing problem [J]. *Journal of Software*, 2007, 18(9): 2083-2089 (in Chinese).
- [18] RUDOLPH G. Convergence analysis of canonical GA [J]. *IEEE Trans on Neural Networks*, 1994, 5(1): 96-101.

Air container loading based on improved genetic algorithm

ZHANG Changyong*, ZHAI Yiming

(College of Electronic Information and Automation, Civil Aviation University of China, Tianjin 300300, China)

Abstract: Aimed at the problems of slow convergence speed, premature maturity, and poor optimization results when the standard genetic algorithm solves the loading plan, based on the anthropomorphic loading strategy, an improved genetic algorithm is proposed to maximize the utilization of container space, considering the loading sequence, volume, and quality of the goods, center of gravity, non-overlapping and other practical constraints. First, the real number code combined with the placement state of the goods is used to randomly generate the initial population. Second, the optimal solution preservation strategy is added to the routine selection operation, and the stability, support constraints, and center of gravity constraints are taken into account after linear scale transformation. In the fitness function, the evaluation value of each loading scheme is calculated by this. Finally, the scheme with the highest evaluation value is output as the optimal loading scheme. In the experimental part, the performance test was performed using test cases with different heterogeneity, and then three sets of specific cargo loading data were combined to prove the universality and practicability of the algorithm. The results show that the proposed algorithm has better optimization effect in solving the process of strong heterogeneous cargo loading, and is suitable for solving the container loading problem. Compared with the standard genetic algorithm, the convergence and search speed have been improved. The space utilization of the two different container types has increased by 3.82% and 3.66%, and the running time has been shortened by 7.9 s and 5.58 s. The optimal loading can be found quickly. The solution can effectively solve the problem of cargo packing in regular and irregular containers. At the same time, the visualization of the loading plan is realized based on MATLAB software, which provides a theoretical basis for the real-time loading decision of the container.

Keywords: improved genetic algorithm; anthropomorphic loading strategy; actual constraints; different container types; visualization

Received: 2020-05-20; **Accepted:** 2020-09-04; **Published online:** 2020-09-22 15:52

URL: kns.cnki.net/kcms/detail/11.2625.V.20200922.1507.003.html

Foundation items: National Natural Science Foundation of China (51707195); the Fundamental Research Funds for the Central Universities (3122016A009)

* **Corresponding author.** E-mail: cyzhang@cauc.edu.cn

http://bhxb.buaa.edu.cn jbuua@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2020.0184

航天员出舱可伸缩安全系绳设计与验证



孙启臣¹, 姜坤¹, 王海亮², 张子岚¹, 秦俊杰¹, 王露斯^{1,*}

(1. 北京卫星制造厂有限公司, 北京 100093; 2. 中国航天员科研训练中心, 北京 100093)

摘 要: 随着中国神舟飞船系列任务和空间站建设任务的稳步进行, 中国航天员将面临频繁的在轨出舱作业任务, 航天员出舱活动范围将逐渐增大, 作业位置也将频繁变换, 传统安全绳已无法满足使用要求。提出了一种基于恒力矩弹簧驱动的航天员出舱可伸缩安全系绳设计方案, 并优化设计方法。对可伸缩安全系绳的核心部分——驱动单元进行设计, 研制航天员出舱可伸缩安全系绳样机, 并完成功能性能、环境适应性和驱动单元寿命试验验证。结果表明: 所提方案可顺利完成钢丝绳收放, 且实测收绳力、放绳力与理论值基本吻合。方案具有体积小、质量轻、可耐大量级振动和高低温环境等优点, 其驱动单元满足 10 000 次的长寿命使用要求, 可用于后续工程产品中。

关键词: 航天员; 安全系绳; 可伸缩; 恒力矩弹簧; 长寿命

中图分类号: V476.1

文献标志码: A

文章编号: 1001-5965(2021)07-1353-09

航天员出舱安全系绳是航天员在轨出舱活动的重要装备^[1], 其作用是保证航天员与航天器之间的可靠连接, 以保障航天员人身安全。航天员出舱安全系绳可分为固定长度安全系绳和可伸缩安全系绳(长度可调)2 种类型。

早在 20 世纪 60—70 年代, 国外航天员就已经开始在出舱活动中利用安全系绳, 当时采用的安全系绳大部分为固定长度安全系绳。随着航天员出舱活动半径的日益增加, 固定长度安全系绳的长度逐步增加至数米甚至更长, 且绳径较大。航天员反映“又粗又复杂的系绳很难控制”, 冗余的系绳甚至有缠绕住航天员颈部的危险^[2-3]。

自 20 世纪 90 年代以来, 可伸缩安全系绳逐渐成为航天员出舱活动的标准配置^[4]。可伸缩安全系绳不但体积小, 而且可在航天员舱外作业过程中保持一定的拉力, 确保钢丝绳始终处于拉直状态, 有效地避免了航天员作业过程中发生缠绕、钩挂等危险情况^[5-6]。相比于固定长度安

全系绳, 可伸缩安全系绳具有较为明显的优势^[7-8]。

随着中国神舟飞船系列任务及空间站建设任务的稳步进行, 中国航天员将面临频繁的在轨出舱作业任务, 如出舱进行设备维修、检查等。目前, 中国航天员出舱作业时已配备了固定长度的安全系绳。在舱外作业时, 需将其挂在附近的扶栏上, 以保证自身安全^[9-10]。但由于目前航天员舱外作业模式及范围的限制(均为近距离作业), 在可伸缩安全系绳方面尚无需求, 国内尚无成熟产品。所以, 中国航天员对长度可调的安全系绳的需求越来越强烈, 亟需根据中国航天任务的进度研制出可伸缩安全系绳产品。同时由于出舱作业的特殊性, 该产品需具有可靠性高、寿命高、耐高低温性能好、作业过程平稳及满足人机工效学要求等特性, 研制难度极大。

针对中国航天员对可伸缩安全系绳的迫切需求, 本文设计了一种基于恒力矩弹簧驱动的航天

收稿日期: 2020-05-14; 录用日期: 2020-07-18; 网络出版时间: 2020-08-31 16:17

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20200831.1502.002.html

* 通信作者。E-mail: 335808408@qq.com

引用格式: 孙启臣, 姜坤, 王海亮, 等. 航天员出舱可伸缩安全系绳设计与验证[J]. 北京航空航天大学学报, 2021, 47(7): 1353-1361. SUN Q C, JIANG K, WANG H L, et al. Design and verification of an adjustable length safety tether for astronaut extravehicular activities[J]. Journal of Beijing University of Aeronautics and Astronautics, 2021, 47(7): 1353-1361 (in Chinese).

员出舱可伸缩安全系绳。该产品可通过钢丝绳的收放,保证航天员与航天器之间的连接一直处于有张力的拉直状态,且具有体积小、质量轻、可耐大量级振动和高低温环境、驱动单元满足 10 000 次的长寿命使用要求等特点,可用于后续出舱作业中。同时,本文所提方案采用恒力矩弹簧作为驱动源,该弹簧具有随着变形的增大,输出力矩基本保持恒定的特点,现已成功应用于月球 1/6 重力模拟及仪器仪表制造等领域^[11]。本文驱动方式可使收绳力、拉绳力平稳,满足航天员用产品对人工工效学要求。

1 结构设计

本文提出的可伸缩安全系绳构型如图 1 所示。在其连接环处安装连接带后,通过连接带与航天服连接;在安全挂钩接口处安装安全挂钩,安全挂钩用于与航天器连接;抓握带可辅助航天员在失重环境中快速抓牢机构主体。

可伸缩安全系绳工作原理如图 2 所示。其工作原理如下:以恒力矩弹簧 2 为动力源,拉出钢丝绳 8 时,钢丝绳 8 带动卷绳筒 7 旋转,同时将恒力矩弹簧 2 由输入卷筒 1 缠绕至输出卷筒 9 而储能;回收钢丝绳时,恒力矩弹簧 2 由输出卷筒 9 缠绕至输入卷筒 1 而释放弹性势能,带动卷绳筒 7 对钢丝绳 8 进行卷绕复位。滑轮 5 可对钢丝绳 8 卷绕起到一定的导向作用,避免钢丝绳 8 卷绕过于紊乱。

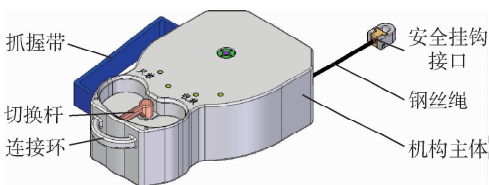


图 1 可伸缩安全系绳构型

Fig. 1 Prototype of adjustable length safety tether configuration

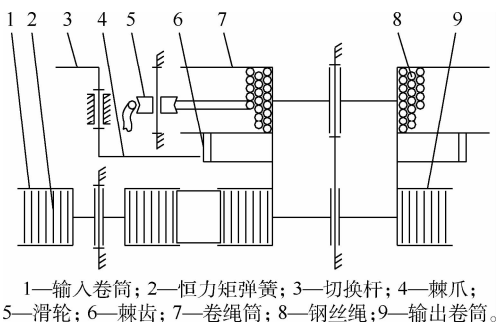


图 2 可伸缩安全系绳工作原理

Fig. 2 Operating principle of adjustable length safety tether

可伸缩安全系绳有 2 种工作模式:“收放”模式和“只放”模式。“收放”模式下,可实时收绳和放绳,拉出的钢丝绳长度可随航天员的位置变化而实时自适应调节,并使钢丝绳一直处于张紧状态,适用于航天员频繁变换作业位置的工况;“只放”模式下,钢丝绳只能被拉出,而不能收回,拉出钢丝绳后钢丝绳可保持在所需长度,适用于航天员长时间在某一固定位置作业的工况。2 种工作模式之间通过拨动切换杆实现切换,切换杆控制棘爪与棘齿开合,当棘爪和棘齿相互作用使卷绳筒只能单向旋转时即为“只放”模式,棘爪和棘齿分开而不相互作用时则为“收放”模式。

本文方案采用恒力矩弹簧作为储能驱动元件,可保证收绳、拉绳过程中力的平稳性,最大程度地减少了系绳对航天员移动的干扰;其内部结构较为简单,使用寿命长,耐真空、振动和高低温性能较好,无需对其施加额外热控措施;机构构型小巧,具有便于携带、操作简便等特点,且其外形、抓握部位、工作模式切换部位均具有良好的人机工效特性。

2 驱动单元设计

可伸缩安全系绳中的驱动单元由恒力矩弹簧、输出卷筒和输入卷筒组成,驱动单元为可伸缩安全系绳卷绕钢丝绳提供动力来源。由图 2 所示的工作原理可看出,驱动单元的设计将直接关系到可伸缩安全系绳的功能实现。因此,驱动单元设计在本文所提方案中至关重要。

2.1 设计原理与方法

大多弹簧特性为变量与输出力呈正比,而恒力矩弹簧的特性为不同变量下输出力相同^[12]。这是由于其成形后,曲率半径相同,进而实现恒定输出^[13]。恒力矩弹簧工作原理为:储能阶段弹簧由输入卷筒向输出卷筒缠绕;释放阶段在自身应力下,弹簧自动缠绕至输入卷筒,并通过输出卷筒输出动力。恒力矩弹簧安装状态如图 3 所示。

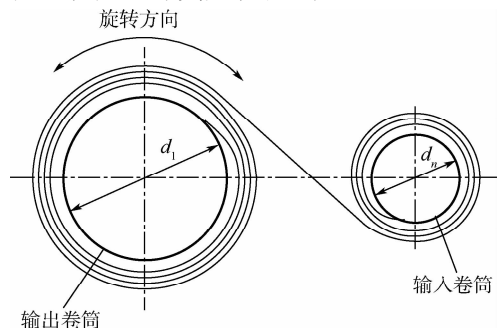


图 3 恒力矩弹簧原理

Fig. 3 Principle of constant torque spring

影响恒力矩弹簧输出性能的参数包括材料的弹性模量、簧片宽度、簧片厚度、弹簧自由状态下

的内圈直径、弹簧的芯轴直径等。

恒力矩弹簧完整设计流程如图 4 所示。

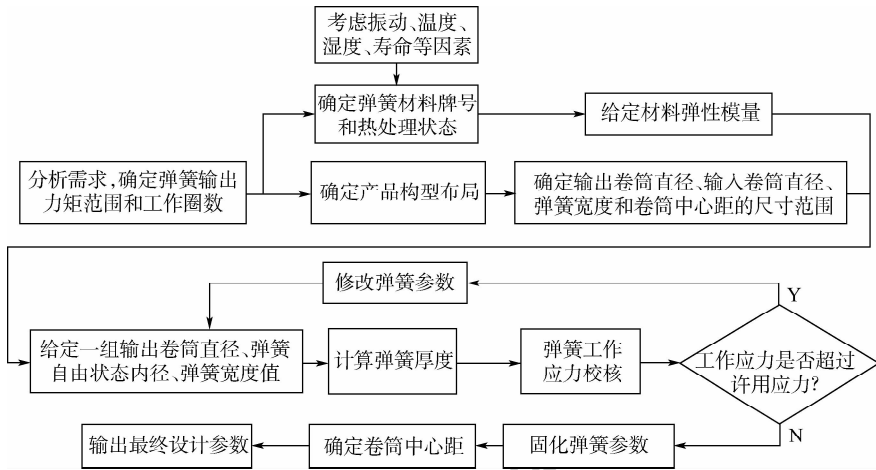


图 4 恒力矩弹簧设计流程

Fig. 4 Constant torque spring design flow

2.2 驱动单元参数计算

恒力矩弹簧输出力矩和工作应力的传统计算公式如下。

恒力矩弹簧输出力矩为^[12]

$$T = \frac{Ebh^3d_1}{12} \left(\frac{1}{d_0} + \frac{1}{d_1} \right)^2 \quad (1)$$

恒力矩弹簧工作应力为^[12]

$$\sigma = Eh \left(\frac{1}{d_0} + \frac{1}{d_1} \right) \times 10^3 \quad (2)$$

式中: T 为输出力矩, $N \cdot m$; E 为材料的弹性模量, GPa ; b 为弹簧宽度, mm ; h 为弹簧厚度, mm ; d_0 为弹簧自由状态下的内径, mm ; d_1 为输出卷筒直径, mm ; σ 为弹簧工作应力, MPa 。

由恒力矩弹簧工作原理和输出特性可知:

1) 恒力矩弹簧输出力矩和工作应力计算时,应以最外圈弹簧厚度方向的中心位置作为实际输出卷筒直径。

2) 当输出卷筒上缠绕一圈以上弹簧时,弹簧的实际输出卷筒直径须叠加已绕弹簧厚度。

因此,恒力矩弹簧在输出卷筒上缠绕 n 圈时的实际输出卷筒直径应为 $d_1 + h + 2h(n - 1)$,即随着弹簧卷绕圈数的增加,弹簧的实际输出卷筒直径会逐渐增大,弹簧输出力矩也随之增大。因此,对恒力矩弹簧输出力矩和工作应力计算公式进行如下修正。

恒力矩弹簧第 n 圈输出力矩为

$$T_n = \frac{Ebh^3 [d_1 + h + 2h(n - 1)]}{12} \left[\frac{1}{d_0} + \frac{1}{d_1 + h + 2h(n - 1)} \right]^2 \quad (3)$$

恒力矩弹簧第 n 圈工作应力为

$$\sigma_n = Eh \left[\frac{1}{d_0} + \frac{1}{d_1 + h + 2h(n - 1)} \right] \times 10^3 \quad (4)$$

式中: T_n 为第 n 圈输出力矩, $N \cdot m$; n 为输出卷筒上已绕弹簧圈数($n \geq 1$)。

由上述公式可看出,当输出卷筒上卷绕 1 圈弹簧时,实际输出卷筒直径为 $d_1 + h$;当输出卷筒上卷绕 2 圈弹簧时,实际输出卷筒直径为 $d_1 + 3h$;以此类推。

按照 2.1 节设计流程,并结合修正后的输出力矩计算公式(式(3)、式(4)),得到恒力矩弹簧相关设计参数,如表 1 所示。

表 1 恒力矩弹簧设计参数

Table 1 Design parameters of constant torque spring

参数	数值
输出卷筒直径/mm	80
弹簧自由状态下的内径/mm	44
弹簧厚度/mm	0.25
弹簧宽度/mm	16
第 1 圈输出力矩/($N \cdot m$)	0.393
第 45 圈输出力矩/($N \cdot m$)	0.428
输入卷筒直径/mm	46
卷筒中心距/mm	116
弹簧工作应力/MPa	1 544 ~ 1 672

2.3 收绳力、拉绳力理论计算

考虑到航天员在轨出舱操作的能力及太空作业的安全性,可伸缩安全系绳的拉绳力要求不大于 25 N,防止对航天员行动和操作带来过大阻力;要求收绳力不大于 10 N,防止收绳过程对航

航天员产生过大的加速度,造成航天员与飞船舱壁高速碰撞的风险。由可伸缩安全系绳工作原理(见图2)可知:

1)“收放”模式下,棘爪与棘齿分开,机构自动回收钢丝绳的“收绳力”为恒力矩弹簧在卷绳筒上作用对钢丝绳产生的拉力减去滑轮与壳体出口处阻力之和;手动拉出钢丝绳的“拉绳力”为恒力矩弹簧在卷绳筒上作用对钢丝绳产生的拉力、滑轮阻力和壳体出口处阻力三者之和;各轴系回转阻力极小,可忽略。

2)“只放”模式下,棘爪与棘齿啮合,手动拉出钢丝绳的“拉绳力”为恒力矩弹簧在卷绳筒上作用对钢丝绳产生的拉力、滑轮阻力、壳体出口处阻力和棘轮处阻力四者之和;各轴系回转阻力极小,可忽略。

$$F_{SF-SS} = T_n \frac{D}{2 \times 10^3} - F_{HL} - F_{SK} \quad (5)$$

$$F_{SF-FS} = T_n \frac{D}{2 \times 10^3} + F_{HL} + F_{SK} \quad (6)$$

$$F_{ZF-FS} = T_n \frac{D}{2 \times 10^3} + F_{HL} + F_{SK} + F_{JL} \quad (7)$$

式中: F_{SF-SS} 为“收放”模式下收绳力; F_{SF-FS} 为“收放”模式下拉绳力; F_{ZF-FS} 为“只放”模式下拉绳力; D 为卷绳筒卷绕直径; F_{HL} 为滑轮平均阻力; F_{SK} 为绳口平均阻力; F_{JL} 为棘轮平均阻力。

经样机实测可得部分参数值,如表2所示。

由式(5)~式(7)计算可得,各工作模式下钢丝绳拉出不同长度时,“收绳力”和“拉绳力”理论值如表3所示。由表3可知,本文设计方案满足指标要求。

表2 部分参数实测值

Table 2 Measured values of some parameters

参数	数值
第1~3圈	75
第4~9圈	80
第10~15圈	85
卷绳筒卷绕直径/mm	
第16~21圈	90
第22~27圈	96
第28~33圈	101
第34~38圈	106
滑轮平均阻力/N	1.32
绳口平均阻力/N	0.85
棘轮平均阻力/N	1.13

表3 收绳力、拉绳力理论值

Table 3 Theoretical values of rope pulling force and rope pulling force

拉出绳长/m	弹簧工作圈数	卷绳筒卷绕直径/mm	“收放”模式收绳力/N	“收放”模式拉绳力/N	“只放”模式拉绳力/N
1.1	5	0.106	5.3	9.7	10.8
1.9	8	0.101	5.7	10.1	11.2
3.0	12	0.096	6.2	10.5	11.7
4.1	16	0.096	6.3	10.6	11.7
5.1	20	0.09	6.9	11.2	12.4
5.9	23	0.09	6.9	11.3	12.4
7.0	27	0.085	7.6	11.9	13.0
8.1	31	0.08	8.2	12.6	13.7
8.9	34	0.08	8.3	12.6	13.8
10.0	38	0.075	9.1	13.4	14.6

注:随拉出钢丝绳长度增加,弹簧工作圈数增加,卷绳筒卷绕直径减小,因此收绳力与拉绳力会发生变化。

3 力学仿真分析

可伸缩安全系绳为航天产品,上行过程会经历力学环境。同时,可伸缩安全系绳在轨工作时,钢丝绳受拉也可能对承载部位造成破坏。因此,对本文方案开展了模态分析和关键部件的应力应变分析。

3.1 力学特性仿真分析

各阶模态频率、模态振型等模态参数可用来衡量结构的动态特性。通过模态分析,可以对结构的动态特性进行评价^[14]。因此,对可伸缩安全

系绳方案开展了模态分析,分析结果如图5所示。

由分析结果可知,可伸缩安全系绳基频为518.86 Hz,前六阶频率在518~867 Hz之间,可认为可伸缩安全系绳上行阶段不会产生共振。

3.2 关键部位应力分析

航天员在失重状态下出舱作业时,由于外界因素的不确定性,特殊情况下航天员可能出现运动轨迹不受控的情况,从而对可伸缩安全系绳钢丝绳端及机构本体产生冲击过载拉力,该冲击过载拉力也是可伸缩安全系绳所有工况中受力最苛刻的工况。

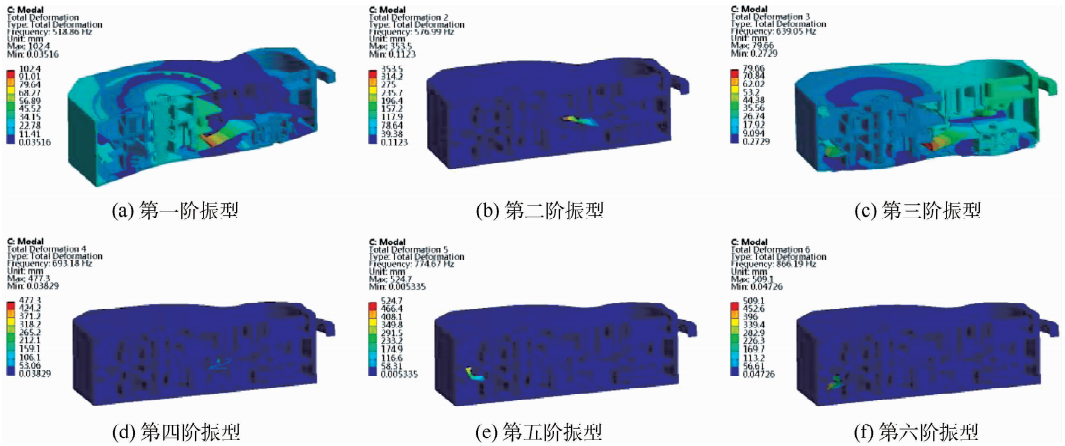


图 5 前六阶振型分析结果

Fig. 5 Analysis results of the first six modes

结合可伸缩安全系绳的工作原理和传力路径,开展了最苛刻工况下机构关键部位的应力分析,主要承力零部件(壳体、滑轮轴、卷绳筒)静力分析结果如图 6 所示。

由分析结果可知,在最苛刻工况下,各主要零部件应力水平较其材料许用应力均有较大余量。由此可认为,可伸缩安全系绳各零部件均满足各工况下强度要求。

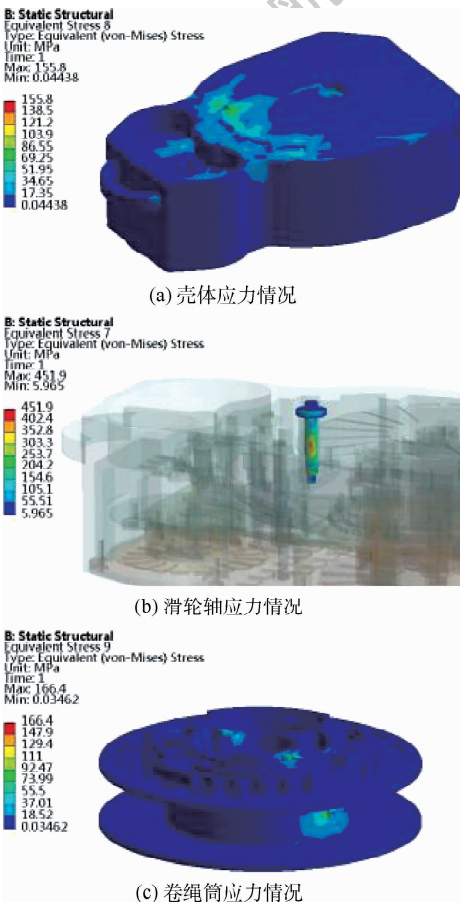


图 6 各零部件应力分布

Fig. 6 Stress distribution of each component

4 试验验证

可伸缩安全系绳样机实物如图 7 所示。为验证可伸缩安全系绳收放绳功能性能、环境适应性及驱动单元寿命,对可伸缩安全系绳开展了相应的试验验证。

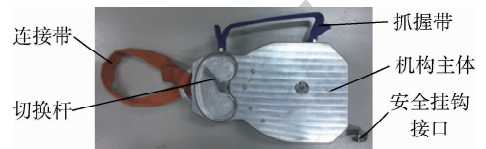


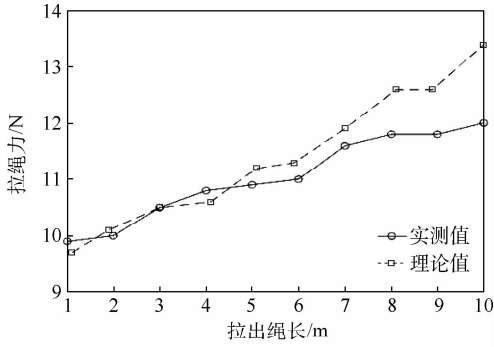
图 7 可伸缩安全系绳原理样机实物图

Fig. 7 Schematic diagram of principle prototype of adjustable length safety tether

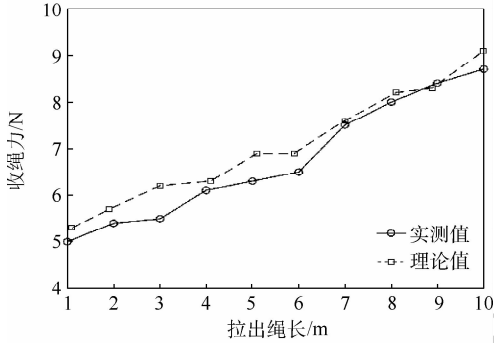
4.1 功能性能试验验证

根据功能需求,可伸缩安全系绳需具备手动拉绳、自动收绳、绳长保持功能,且其收绳力不大于 10 N,拉绳力不大于 25 N。通过对可伸缩安全系绳测试可知,收放绳功能可实现;切换杆可实现“收放”和“只放”2 种工作模式的切换。分别对“只放”模式下拉出钢丝绳所需拉力(只放模式拉绳力)、“收放”模式下拉出钢丝绳所需拉力(收放模式拉绳力)、“收放”模式下机构对钢丝绳的回收力(收放模式收绳力)进行了测试,测试结果如图 8 所示。

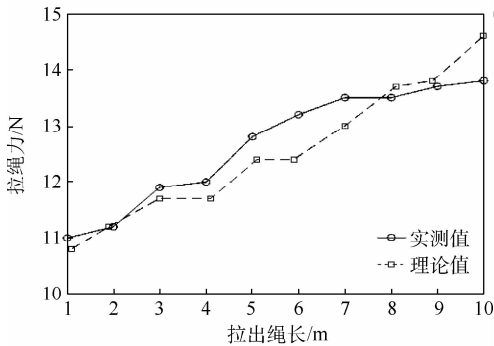
试验结果表明,可伸缩安全系绳在各工作模式下,收绳力、拉绳力平稳,“收放”模式下收绳力范围在 5~9 N 之间,“只放”模式下拉绳力在 11~14 N 之间,“收放”模式下拉绳力在 10~13 N 之间,收绳力和拉绳力均满足设计指标要求。相对于“收放”模式,“只放”模式下拉出钢丝绳的同时还需克服棘爪与棘齿间作用力,因此,“只放”模式下拉绳力相对较大。



(a) “收放”模式拉绳力实测值与理论值对比



(b) “收放”模式收绳力实测值与理论值对比



(c) “只放”模式拉绳力实测值与理论值对比

图8 各模式下性能曲线

Fig. 8 Performance curves in various modes

4.2 力学环境适应性试验验证

可伸缩安全系绳在发射过程中会承受由火箭诱发的各类力学环境^[15-16]。因此,为验证可伸缩安全系绳的力学环境适应性,对其进行了正弦、随机振动试验,试验状态如图9所示,其试验条件如表4、表5所示。

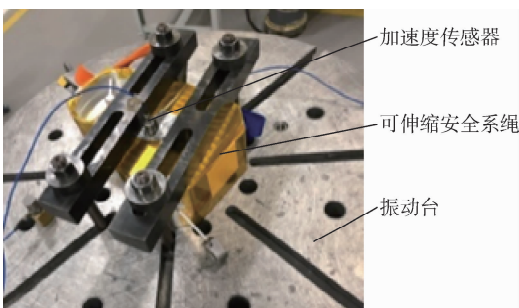


图9 正弦、随机振动试验

Fig. 9 Sinusoidal, random vibration test

正弦、随机振动试验前后,对各模式下的收绳力、拉绳力进行测试,并将测试结果进行了对比,如图10所示。

试验结果表明,可伸缩安全系绳在经历正弦、随机振动试验后,收绳力值、拉绳力与试验前一致性较好,性能未发生变化,满足力学环境的要求。

表4 正弦振动试验条件

Table 4 Sinusoidal vibration test conditions

频率范围/Hz	幅值	
	验收级	鉴定级
4 ~ 10	14.7 mm	22 mm
10 ~ 17	3.6 g	5.4 g
17 ~ 60	8 g	12 g
60 ~ 100	3.8 g	5.6 g
加载扫描率	4 oct/min	2 oct/min
加载方向	3个轴向	

注:oct为倍频程。

表5 随机振动试验条件

Table 5 Random vibration test conditions

频率范围/Hz	鉴定级	验收级
10 ~ 50	3 dB/oct	3 dB/oct
50 ~ 300	0.25 g ² /Hz	0.1 g ² /Hz
300 ~ 2 000	-12 dB/oct	
总均方根加速度 (grms)	9.68	6.11
加载时间 (s · 轴 ⁻¹)	120	60
加载方向	3个轴向	

注:g为重力加速度,rms为均方根。

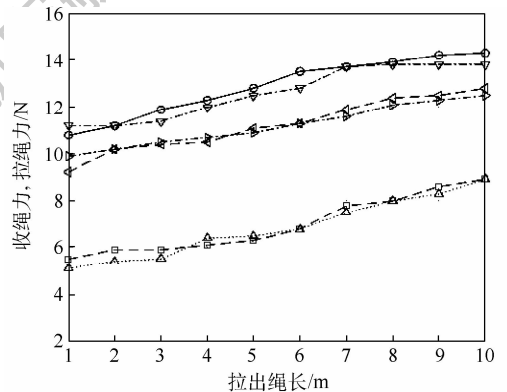


图10 力学试验前后收绳力和拉绳力测试结果

Fig. 10 Test results of rope pulling force and rope pulling force before and after mechanical test

4.3 热环境适应性试验验证

可伸缩安全系绳在舱外使用过程中,由于作业位置、作业时间、光照变化等因素的影响,壳体及内部温度会发生变化,从而影响工作性能。

因此,为验证可伸缩安全系绳的热环境适应性,对其进行了热循环试验,试验状态如图 11 所示。试验温度采用可能出现的最极限温度范围,即 $-100 \sim +100^{\circ}\text{C}$,循环次数为 10.5 次。

热循环试验中,对各模式下的收绳力、拉绳力进行了测试,高温工况下测试结果如图 12 所示,低温工况下测试结果如图 13 所示。

由试验结果可知,高温工况下,收绳力、拉绳

力与常温下相比几乎未发生变化;低温工况下,由于恒力矩弹簧刚度增加,收绳力、拉绳力略有增大,但仍满足使用要求。热循环试验结果表明,可伸缩安全系绳满足热环境的要求。

4.4 驱动单元寿命试验验证

由于航天员频繁的在轨出舱作业任务,可伸缩安全系绳也将长期在轨工作,且工作时会频繁收放钢丝绳,其寿命要求非常高。通过对可伸缩安全系绳设计方案进行分析,得出影响其寿命的主要环节为驱动单元。针对其驱动元件——恒力矩弹簧,开展了专项寿命验证,搭建了由动力源、测试工装、控制系统及恒力矩弹簧试验件组成的寿命试验系统,如图 14 所示。

寿命试验的方法为:用测试工装真实模拟恒力矩弹簧使用状态,先用电机将恒力矩弹簧卷绕至测试工装的输出卷筒模拟件上,进行循环往复测试。

试验过程中,定期对卷绕和复位过程中的恒力矩弹簧驱动力矩进行测试,部分测试结果如图 15 所示。

由试验结果可知,经过 10 000 次寿命试验后,恒力矩弹簧力矩值和力矩变化趋势与初始状态基本一致。由试验结果可知,经过 10 000 次寿命试验,恒力矩弹簧驱动性能未下降。



图 11 热循环试验

Fig. 11 Thermal cycle test

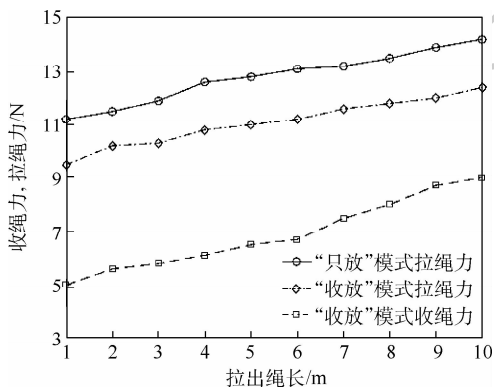


图 12 热循环试验高温工况下收绳力和拉绳力测试结果
Fig. 12 Test results of rope pulling force and rope pulling force under high temperature condition in thermal cycle test

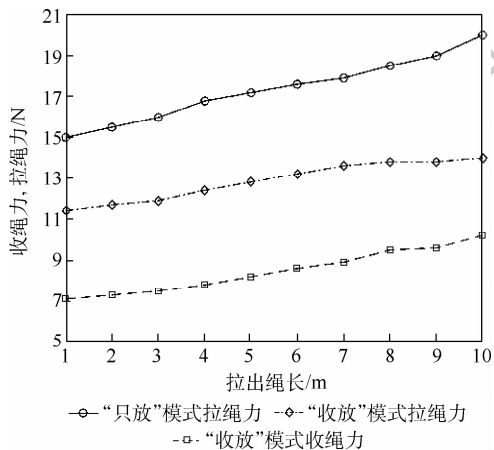


图 13 热循环试验低温工况下收绳力和拉绳力测试结果

Fig. 13 Test results of retracting force and retracting force under low temperature conditions in thermal cycle test

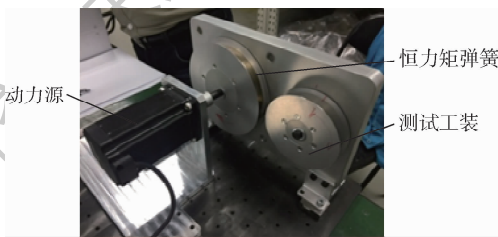


图 14 寿命试验系统组成

Fig. 14 Composition of life test system

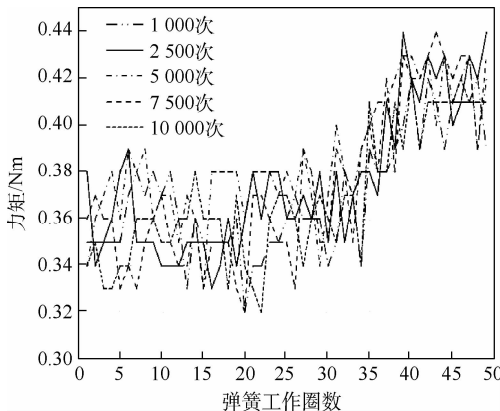


图 15 恒力矩弹簧寿命试验曲线

Fig. 15 Life test curves of constant torque spring

5 结 论

1) 通过理论、仿真与试验研究等手段,对航天员出舱可伸缩安全系绳进行了设计与验证。结果表明,本文方案合理可行,其功能性能、环境适应、寿命等满足使用要求。

2) 本文方案结构简单可靠,轻量化水平较高,便携且易操作,人机工效性能较好,可用于后续工程产品中。

后续将继续开展优化设计,进一步提高产品工作性能与可靠性,提高轻量化与人机工效水平。

参考文献 (References)

- [1] 朱仁璋,王鸿芳,王晓光. 舱外活动系统述评[J]. 航天器工程,2008,17(6):7-32.
ZHU R Z, WANG H F, WANG X G. A review of EVA system [J]. Spacecraft Engineering, 2008, 17(6): 7-32 (in Chinese).
- [2] MINOR M A, HIRSCHI C R. Automated tether management system for extravehicular activities[J]. Journal of Field Robotics, 2007, 24(4): 311-335.
- [3] ATWELL W. Extra vehicular activity (EVA) operations in the geostationary environment: AIAA 2012-3646 [R]. Reston: AIAA, 2012.
- [4] VOGT G L. Suited for spacewalking. Teacher's guide with activities, mathematics and science: EG-1998-03-112 [R]. Washington D. C.: NASA, 1998.
- [5] BORTHWICK D E, CRONCH D F, NIXON G R. Design of a reusable kinetic energy absorber for an astronaut safety tether to be used during extravehicular activities on the space station: NASA-CR-192015 [R]. Washington, D. C.: NASA, 1991.
- [6] MINOR M A, HIRSCHI C R, AMBROSE R O. An automated tether management system for microgravity extravehicular activities [C] // Proceedings of 2002 IEEE International Conference on Robotics and Automation. Piscataway: IEEE Press, 2002: 7379378.
- [7] HIRSCHI C R, MINOR M A. Testing and evaluation of an automated tether management system for microgravity extravehicular activities [C] // Proceedings of 2004 IEEE International Conference on Robotics and Automation. Piscataway: IEEE Press, 2004: 8330186.
- [8] VAN PELT M. Space tethers and space elevators [M]. Berlin: Springer, 2009.
- [9] 陈学文,黄鹏,张炎,等. 航天员舱外活动安全绳虚拟现实操作仿真方法研究[J]. 载人航天, 2018, 24(6): 724-739.
CHEN X W, HUANG P, ZHANG Y, et al. Study on simulation methods of safety rope operation in virtual reality during astronaut extravehicular activity [J]. Manned Spaceflight, 2018, 24(6): 724-739 (in Chinese).
- [10] 杨锋,丁立,杨春信,等. 舱外活动中航天员利用扶手移动的路径规划[J]. 北京航空航天大学学报, 2005, 31(3): 326-330.
YANG F, DING L, YANG C X, et al. Path planning for astronaut translation by means of continuous handrails during extravehicular activity [J]. Journal of Beijing University of Aeronautics and Astronautics, 2005, 31(3): 326-330 (in Chinese).
- [11] 付铁,陈杰,丁洪生. 恒力弹簧的载荷-变形特性分析[J]. 北京理工大学学报, 2011, 3(8): 905-908.
FU T, CHEN J, DING H S. Analysis on load-deflection behaviors of constant force springs [J]. Transactions of Beijing Institute of Technology, 2011, 3(8): 905-908 (in Chinese).
- [12] 张英会,刘辉航,王德成. 弹簧手册 [M]. 北京:机械工业出版社, 2008.
ZHANG Y H, LIU H H, WANG D C. Spring handbook [M]. Beijing: China Machine Press, 2008 (in Chinese).
- [13] 李占华,韩静涛,于春宇,等. 拉压复合连续弯曲成形恒力弹簧工艺研究[J]. 锻压技术, 2016, 41(4): 49-53.
LI Z H, HAN J T, YU C Y, et al. Research on the constant force spring process in tension-compression composite progressive bending [J]. Forging & Stamping Technology, 2016, 41(4): 49-53 (in Chinese).
- [14] 曹树谦,张文德,萧龙翔. 振动结构模态分析: 理论与应用 [M]. 2版. 天津:天津大学出版社, 2014.
CAO S Q, ZHANG W D, XIAO L X. Modal analysis of vibration structure: Theory, experiment and application [M]. 2nd ed. Tianjin: Tianjin University Press, 2014 (in Chinese).
- [15] 马兴瑞,韩增尧,邹元杰,等. 航天器力学环境分析与条件设计研究进展[J]. 宇航学报, 2012, 33(1): 1-12.
MA X R, HAN Z Y, ZOU Y J, et al. Review and assessment of spacecraft mechanical environment analysis and specification determination [J]. Journal of Astronautics, 2012, 33(1): 1-12 (in Chinese).
- [16] 孟光,周徐斌,苗军. 航天重大工程中的力学问题[J]. 力学进展, 2016, 46(1): 267-322.
MENG G, ZHOU X B, MIAO J. Mechanical problems in momentous projects of aerospace engineering [J]. Progress in Mechanics, 2016, 46(1): 267-322 (in Chinese).

Design and verification of an adjustable length safety tether for astronaut extravehicular activities

SUN Qichen¹, JIANG Kun¹, WANG Hailiang², ZHANG Zilan¹, QIN Junjie¹, WANG Lusi^{1,*}

(1. Beijing Spacecraft, Beijing 100093, China;

2. China Astronaut Research and Training Center, Beijing 100093, China)

Abstract: With the steady progress of Shenzhou spaceship series tasks and space station construction tasks in China, Chinese astronauts are facing frequent extravehicular activities, the range of astronauts' extravehicular activities will gradually increase, and the operation position of astronauts will change frequently. The traditional safety tether can not operation the requirements. In this paper, a design scheme of an adjustable length safety tether for astronaut extravehicular activities driven by constant force spring is put forward, the design method is optimized, and the detailed design of the driving unit which is the core part of the safety tether is performed. Finally, the prototype of safety tether is developed, and the functional performance test, environmental adaptability test and driving unit life test are completed. The results show that the scheme of the safety tether can successfully complete the retraction and release of steel wire rope, and the measured retracting force and pulling force are basically consistent with the theoretical values. It has the advantages of small volume, light weight, and resistance to a large number of levels of vibration and high and low temperature environment. The driving unit can meet the requirements of 10 000 times of long-life use. The scheme of the safety tether can be used in subsequent engineering products.

Keywords: astronaut; safety tether; adjustable length; constant force spring; long service life

http://bhxb.buaa.edu.cn jbuua@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2020.0214



涂覆吸波材料圆柱的 RCS 预估方法

赵京城^{1,*}, 杨宗凯¹, 王健², 张扬²

(1. 北京航空航天大学 电子信息工程学院, 北京 100083; 2. 航空工业第一飞机设计研究院, 西安 710089)

摘 要: 用缩比模型测量结果预估是研制阶段获得大尺寸目标雷达散射截面(RCS)的常用方法,但根据经典电磁相似理论,严格满足缩比条件的涂覆吸波材料缩比目标测量难以实现。针对涂覆吸波材料缩比目标的 RCS 预估问题,提出了采用多元对数线性回归模型的预估方法。设计了2组圆柱模型,在微波暗室中对缩比因子分别为1、2、4、8的2组模型进行了测试。在完成角度矫正等数据预处理基础上,将缩比模型 RCS 数据作为训练集代入模型当中求得参数,对原模型的 RCS 进行预估并与实际实测数据进行对比分析。结果表明:所提方法预估数据与实测数据曲线拟合度较好,相较于传统平方率模型,误差下降了3~5 dB,在回归模型中加入吸波材料因子后误差进一步下降了0.3~0.8 dB。

关键词: 雷达散射截面(RCS); 缩比模型; 多元线性回归; 吸波材料; 相似原理

中图分类号: TN951

文献标志码: A

文章编号: 1001-5965(2021)07-1362-10

雷达散射截面(RCS)是武器系统隐身性能的重要指标^[1],相比于全尺寸测量,缩比模型测量具有成本低廉、测试方便、受外部环境影响较小等优点^[2]。但由于经典电磁相似理论的限制,目前 RCS 缩比测量多见于目标为金属或无耗、低损耗介质的情况^[3-6]。而在实际工程当中,在金属表面涂覆吸波材料又是最常使用的一种缩减目标 RCS 的手段,因此,研究含介质及吸波材料等有耗目标的缩比规律具有重要意义。

对有耗目标的 RCS 缩比研究,时振栋等^[7-10]对其进行了理论推导,利用 Leontovich 阻抗边界条件,提出了有耗散射体缩比因子的概念,对理想缩比条件下的公式进行调整,但仅限于缩比模型表面与原模型表面具有相同的表面阻抗。胡艳等^[11]分析了只有工作波长变化或者只有模型尺寸变化时简单散射体的 RCS 变化,并介绍了其在缩比测量中的应用。陈晓洁^[12]采用物理光学法研究了当测试波长不能满足大比例缩短时电大目

标与缩比模型的缩比关系。文献[13]总结了缩比测量所用的紧缩场、背景降噪、辐射源等技术,并介绍了为满足理想缩比定律而特定制作材料成本及难度较高。Susetio 等^[14]对单层介质不同比例的缩比模型进行了仿真,在仿真过程中根据电磁相似原理改变介质的参数,验证了不同波段的理想电磁学相似原理。对于涂覆吸波材料目标的 RCS 缩比预估问题,实用性较强的成果还有待进一步的研究。

本文针对涂覆吸波材料的大尺寸目标 RCS 预估问题进行研究,使用缩比模型的 RCS 预估原模型的 RCS。提出了2种多元对数线性回归的预估模型,数据来源是模型在紧缩场暗室环境下进行的实验测量,得到测试数据后对数据进行了角度矫正等预处理工作。将缩比模型数据作为训练集,回归出了不同极化条件下的缩比预估公式,并将预估的曲线与实测曲线进行对比,验证了模型的有效性。

收稿日期: 2020-05-25; 录用日期: 2020-08-01; 网络出版时间: 2020-08-06 08:33

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20200805.1840.010.html

* 通信作者。E-mail: zjccool@126.com

引用格式: 赵京城, 杨宗凯, 王健, 等. 涂覆吸波材料圆柱的 RCS 预估方法[J]. 北京航空航天大学学报, 2021, 47(7): 1362-1371.
ZHAO J C, YANG Z K, WANG J, et al. RCS prediction method for cylinder containing microwave absorbing material[J]. Journal of Beijing University of Aeronautics and Astronautics, 2021, 47(7): 1362-1371 (in Chinese).

1 预估模型及原理

1.1 相似原理

缩比模型的预估方法是基于 Sinclair^[15] 提出的电磁相似原理,相似条件如表 1 所示。表中: n 为缩比因子,即缩比模型的尺寸缩小倍数,当模型缩小 n 倍、电导率变大 n 倍、测试频率增大 n 倍而其他电磁量不变时,缩比模型的 RCS 为原模型 RCS 的 $1/n^2$ 。对 RCS 取对数处理后,其单位为 dBsm,当严格满足电磁相似原理时,原模型的 RCS 等于缩比模型 RCS 加上 $20\lg n$ 。

表 1 电磁相似原理条件

Table 1 Conditions of electromagnetic similarity principle

物理量	原模型	缩比模型
长度	l	$l' = l/n$
频率	f	$f' = nf$
波长	λ	$\lambda' = \lambda/n$
电导率	σ	$\sigma' = n\sigma$
磁导率	μ	$\mu' = \mu$
介电常数	ε	$\varepsilon' = \varepsilon$
表面阻抗	η	$\eta' = \eta$

1.2 多元对数线性回归模型

多元对数线性回归模型可以将原本为相乘关系的因变量转化为求和的形式,若因变量 y 与多个自变量 x_1, x_2, \dots, x_n 之间的关系为

$$y = x_1^{a_1} x_2^{a_2} \dots x_n^{a_n} \quad (1)$$

式中: a_1, a_2, \dots, a_n 为所求待定系数。

对式(1)等号左右两边同时取对数,可得

$$\lg y = a_1 \lg x_1 + a_2 \lg x_2 + \dots + a_n \lg x_n \quad (2)$$

本文以使用缩比模型 RCS 来预估原模型 RCS 为目标,而 RCS 在测试计算过程当中进行了取对数的处理,由经典相似原理可得,以原模型 RCS 为因变量,缩比模型 RCS 与 $\lg n$ 为自变量,符合多元线性模型。

当模型的材料为 $\sigma \rightarrow \infty$ 的理想导体或 $\sigma = 0$ 的理想介质时,很容易满足理想的相似原理所要求的 $\sigma' = n\sigma$ 条件,但在实际的工程当中,所研究的模型多为表面涂覆吸波材料的复合模型,此时很难满足理想状态下的缩比原理,理想缩比关系为多元对数线性的关系。在此基础上假设不满足理想缩比关系的模型仍满足多元对数线性模型,只是参数发生了一定的变化。设原模型 RCS 为 y (单位为 dBsm),缩比模型 RCS 为 x_1 (单位为 dBsm), $\lg n$ 为 x_2 ,故本文所研究的模型 1 为

$$y = a_1 x_1 + a_2 x_2 + \delta \quad (3)$$

式中: a_1, a_2 为所求待定系数; δ 为符合正态分布的

随机误差。

模型 1 没有考虑吸波材料参数的作用,在不同的测试频率下,吸波材料的参数会相应地发生变化,模型的 RCS 也会受到影响,本文所选用的吸波材料反射率随频率变化的曲线如图 1 所示。

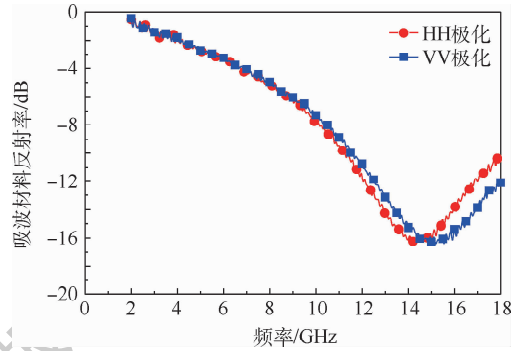


图 1 吸波材料反射率随频率变化

Fig. 1 Reflectance of microwave absorbing materials varies with frequency

在模型中添加吸波材料反射率作为新的因子,设原模型测试频率下的吸波材料反射率为 x_3 (单位为 dB),缩比模型测试频率表下的吸波材料反射率为 x_4 (单位为 dB),可得本文所研究的模型 2 为

$$y = a_1 x_1 + a_2 x_2 + a_3 x_3 + a_4 x_4 + \delta \quad (4)$$

式中: a_3, a_4 为所求待定系数。

1.3 模型的求解及检验方法

多元对数线性回归模型的求解多用最小二乘法^[16-18]进行估计,记

$$\left\{ \begin{aligned} \mathbf{Y} &= \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \\ \mathbf{X} &= \begin{bmatrix} 1 & x_{11} & \dots & x_{n1} \\ 1 & x_{12} & \dots & x_{n2} \\ \vdots & \vdots & & \vdots \\ 1 & x_{1k} & \dots & x_{nk} \end{bmatrix} \\ \mathbf{A} &= \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \\ \boldsymbol{\varepsilon} &= \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_k \end{bmatrix} \end{aligned} \right. \quad (5)$$

式中: $(x_{1i}, x_{2i}, \dots, x_{ni} : y_i)$ 为 k 组观测值; \mathbf{Y} 为观测

向量; X 为设计矩阵; A 为回归系数的估计值向量; ϵ 为误差向量。

则模型可用矩阵表示为

$$Y = XA + \epsilon \quad (6)$$

通过推导, 可得回归系数的估计值向量为

$$A = (X^T X)^{-1} X^T Y \quad (7)$$

多元对数线性回归模型的检验方法^[16-19]有拟合度判定(R 检验)、方差总体线性的显著性检验(F 检验)、变量的显著性检验(t 检验), 分别如下。

1) 拟合度判定(R 检验)

拟合度 R 的计算公式为

$$R = \frac{\sqrt{\frac{\sum_{i=1}^k (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^k (y_i - \bar{y})^2}}}{\sqrt{1 - \frac{\sum_{i=1}^k (y_i - \hat{y}_i)^2}{\sum_{i=1}^k (y_i - \bar{y})^2}}} \quad (8)$$

式中: \bar{y} 为观测值 y_i 的平均值; \hat{y}_i 为预测值。当 R^2 接近于 1 时, y_i 与 \hat{y}_i 更接近, 回归模型的拟合度更高, 效果更好。

2) 方差总体线性的显著性检验(F 检验)

构造原假设与备选假设:

$$H_0: a_1 = 0, a_2 = 0, \dots, a_n = 0。$$

$$H_1: a_j (j=1, 2, \dots, n) \text{ 不全为 } 0。$$

构造统计量:

$$F = \frac{\sum_{i=1}^k (y_i - \hat{y}_i)^2}{\sum_{i=1}^k (y_i - \bar{y})^2} \cdot \frac{k - n - 1}{n} \quad (9)$$

给定显著性水平 α , 比较 F_α 与 F 值大小:

若 $F > F_\alpha(n, k - n - \alpha)$, 拒绝接受原假设 H_0 , 回归效果显著, 存在线性关系。

若 $F \leq F_\alpha(n, k - n - \alpha)$, 接受原假设 H_0 , 回归效果不显著, 不存在线性关系。

3) 变量的显著性检验(t 检验)

多元对数线性回归模型的方程总体线性关系式显著不能说明每个自变量对因变量都是显著的, 为了判断各个变量能否作为自变量保留在回归模型中, 需要对每个自变量进行显著性检验。

构造原假设与备选假设:

$$H_0: a_j = 0。$$

$$H_1: a_j \neq 0, j=1, 2, \dots, n。$$

在给定的显著水平 α , 则有:

若 $t_j > t_{\alpha/2}(k - n - 1)$, 则拒绝 H_0 , 说明 x_j 是影响自变量 y 的主要因素。

若 $t_j \leq t_{\alpha/2}(k - n - 1)$, x_j 的作用不显著, 应删除该元素。

2 实验测试及数据预处理

2.1 实验设计

本文的实验环境为紧缩场实验室, 实验的主要仪器设施为: 2 m 静区紧缩场、E8363B 天线/RCS 测试系统、目标定位转台、RCS 测试及成像系统软件。测试场地示意图如图 2 所示。

本文使用缩比模型的 RCS 来预估原模型的 RCS, 预估的目标是电大尺寸目标, 电大尺寸目标一般指模型的尺寸超过测试波长的 10 倍以上。从散射机理来看, 飞机是复杂目标, 存在多种散射类型。从研究步骤来看, 从比较简单的镜面散射、边缘散射入手, 获得较好结果之后再行其他类型散射研究。本文选择以圆柱为研究对象, 选用了 2 种尺寸的模型, 模型的电尺寸分别为 13.3 和 3.2, 模型的实际照片如图 3 所示。

圆柱 1、圆柱 2 的缩比因子、模型尺寸、测试频率、测试组数分别如表 2、表 3 所示。由于圆柱 1 原模型尺寸较大, 若采用机加工方式完成会使得质量很大, 不适合作为 RCS 测试的模型, 该模型是选择工业标准型材二次加工完成的, 但工业型材只有几种有限的尺寸可选。在兼顾 4 种缩比因子的情况下, 圆柱 1 选择了缩比因子为 0.96 的型材。

将圆柱涂敷吸波材料后, 实际测试场景如图 4 所示。

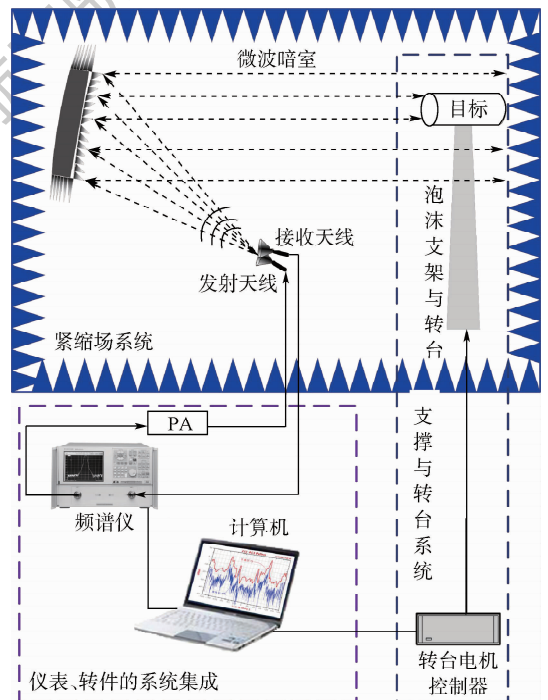


图 2 测试场地示意图

Fig. 2 Sketch map of test site

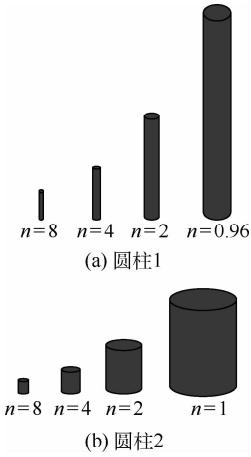


图 3 两种圆柱模型的设计图

Fig. 3 Design drawing of two kinds of cylinder model

表 2 圆柱 1 的尺寸及测试频率

Table 2 Dimension and test frequency of cylinder 1

n	长度/mm	直径/mm	测试频率/GHz	组数
0.96	2 092	159	1.5 ~ 2	101
2	1 000	76	3.5 ~ 4.5	201
4	500	38	7 ~ 9	401
8	250	19	14 ~ 18	801

表 3 圆柱 2 的尺寸及测试频率

Table 3 Dimension and test frequency of cylinder 2

n	长度/mm	直径/mm	测试频率/GHz	组数
1	480.0	240	2 ~ 2.5	101
2	240	120	3.5 ~ 4.5	201
4	120	60	7 ~ 9	401
8	60	30	14 ~ 18	801



图 4 实际测试示意图

Fig. 4 Sketch map of actual test

将待测目标放置在转台中央,通过计算机控制转台以 0.2° 为角度间隔转动 360° ,共 1 801 个角度位置,每转动一次按照测试频率间隔为 5 MHz 进行对应频率的扫频测试,在实验过程当中,网络分析仪提供扫频激励信号,并利用高性能接收机对反射信号进行接收,使用 MATLAB 与北航 RCS 测试及成像系统软件进行原始数据整合。

测试流程及数据处理流程如图 5 所示。

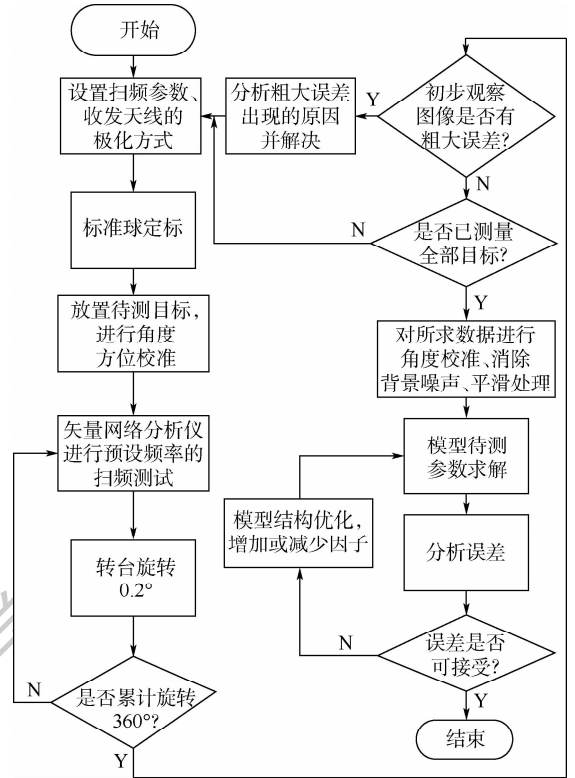


图 5 测试及数据处理流程

Fig. 5 Flowchart of test and data processing

2.2 实验误差分析

本次 RCS 测试依据 GJB 5022—2001^[20]。在本实验室条件下,影响 RCS 测试精度的误差主要有:人工摆放误差、测试系统误差、背景电平误差及目标支架误差。

在测试过程当中,需要将各个模型人工摆放到转台上,对模型的垂直方向与水平方向进行校准,垂直方向可使用水平尺校准,故垂直方向精度较高,而水平方向需要人工校准,可能会出现 1° 以内的误差,故在数据处理的过程当中需要先对数据的角度进行校准。

对于系统误差方面,实验结果表明,对于同一目标、同一转角,一天内多次测得的 RCS 最大误差小于 0.25 dB。另外,目标支架引起的误差是由于泡沫支架在方位角上的不均匀造成的,其对测试结果的影响可以与背景电平合在一起讨论。测量误差与目标和背景电平的差值是指数下降关系,表 4 列出了几个典型数据。

经过测试,测试环境的背景噪声在 -50 dBsm 左右。由表 4 可知,当测试数据在 -40 dBsm 以下时,目标除以背景的比值小于 10 dB,测试的误差可能接近 3 dB,而在实际的测试过程当中有部分数据小于 -50 dBsm,这些数据受背景噪声的影响较大,故在数据处理过程中去除 -45 dBsm 以下误差较大的数据。

表4 测量误差与目标和背景电平的差值典型数据

Table 4 Typical data of measurement error according to difference of target and background ratio

目标电平 背景电平	测量误差/dB
31	±0.25
25	±0.5
19	±1
16	±1.5
10.69	±3

对于实验的随机误差,本文采用5°的窗口进行平滑处理,从而降低随机误差,提高信噪比。

2.3 角度校准

观察测试数据的 RCS 特征分布,随机抽取到测试频率为 1.835 GHz 下的圆柱 1 数据,绘制单站 RCS 对角度 φ 的曲线如图 6 所示。

由图 6 可得,圆柱的 RCS 在 2 种极化情况下均会在 90°与 270°处出现峰值,故以 90°与 270°两点附近的峰值角度与 90°与 270°的差值作为偏差

角度进行处理。

以圆柱 2 在 HH 极化下为例对测试角度误差进行统计,观察其分布如图 7 所示,其中偏差角度正为顺时针偏差,偏差角度负为逆时针偏差。

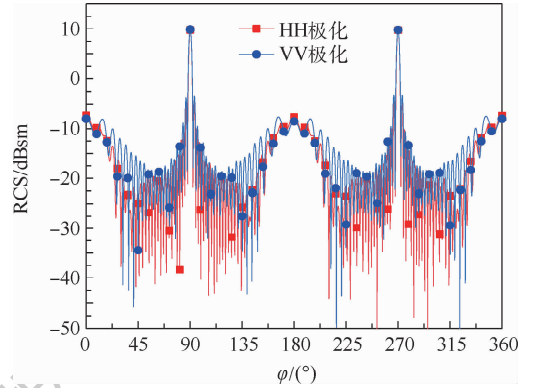
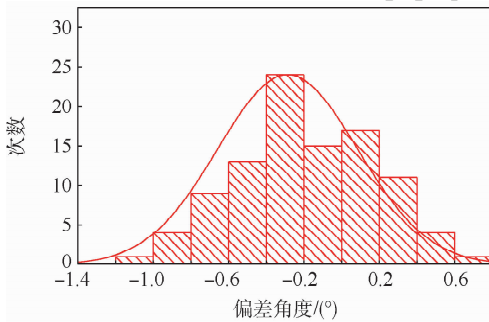
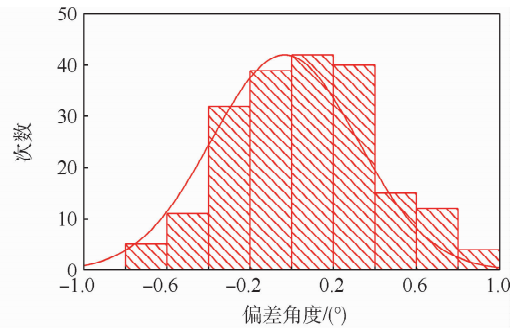


图6 测试频率 1.835 GHz 下圆柱 1 单站 RCS 对角度曲线

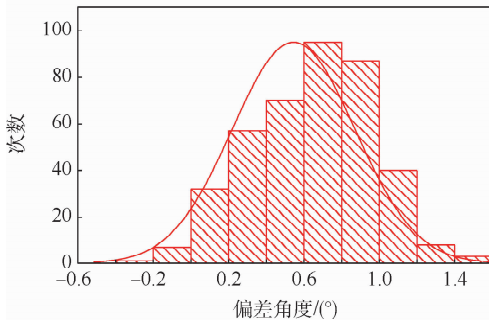
Fig. 6 Influence of angle on cylinder 1's monostatic RCS with test frequency of 1.835 GHz



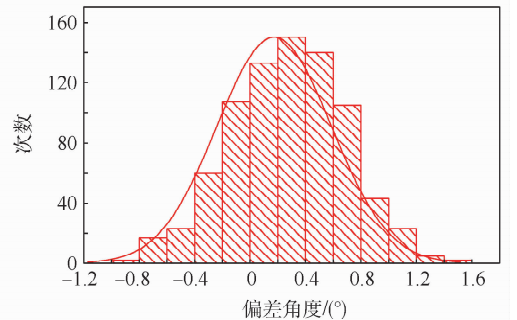
(a) 缩比因子为1时偏差角度出现次数统计



(b) 缩比因子为2时偏差角度出现次数统计



(c) 缩比因子为4时偏差角度出现次数统计



(d) 缩比因子为8时偏差角度出现次数统计

图7 不同缩比因子模型的角度偏差统计对比

Fig. 7 Comparison of angle deviation of models with different scaling factors

由图 7 可得,角度偏差整体上成正态分布,若误差是指服从均值为 0 的正态分布,则误差之间相互独立,可认为其为随机误差,随机误差是由于信号测量环境当中的高斯白噪声、电力供应波动、传输线干扰等因素引起的,具有不可重复性^[21],故在进行角度校准时,将偏差角度校准为均值为 0 的正态分布即可。

涂覆吸波材料后的圆柱 1 在不同的缩比因子与不同的极化方式下的角度校准结果如表 5 所示。

涂覆吸波材料后的圆柱 2 在不同的缩比因子与不同的极化方式下的角度校准结果如表 6 所示。

表5 圆柱 1 数据角度校准结果

Table 5 Calibration results of angle of cylinder 1 (°)

极化方式	n = 0.96	n = 2	n = 4	n = 8
HH	-0.6	0.2	-0.2	-0.2
VV	-0.6	0	-0.2	0.4

表6 圆柱2数据角度校准结果

Table 6 Calibration results of angle of cylinder 2 (°)

极化方式	$n=1$	$n=2$	$n=4$	$n=8$
HH	0.2	0.2	0.2	0.8
VV	0	0.2	0.4	-0.2

3 预估结果分析

3.1 预估过程说明

在数据处理过程当中,将缩比因子为2、4、8的数据作为训练集,以缩比因子为2的数据作为新的原模型,组成缩比因子分别为1、2、4的数据代入模型当中对回归模型的参数进行训练。以缩比因子为1、4的数据作为测试集,使用缩比因子为4的数据代入到训练好的模型当中,对原模型的数据进行回归预测,并与原模型的实测数据进行对比分析。

以圆柱1数据(HH极化)在模型1的回归结果为例,经过计算,回归模型的详细结果如表7和表8所示。

由表7可得,模型的相关系数为0.980,线性相关性较强,经过查表并与表8对比可得在显著

表7 圆柱1的模型1摘要(HH极化)

Table 7 Model one's summary of cylinder 1 (HH polarization)

R	R^2	调整后的 R^2	标准估算的误差
0.980	0.961	0.961	1.602 808

表8 圆柱1的模型1回归方程系数(HH极化)

Table 8 Model one's regression equation coefficients of cylinder 1 (HH polarization)

因子	系数	标准错误	t	显著性
x_1	0.943	0	2 830.873	0
x_2	27.175	0.041	967.286	0

表9 圆柱1不同模型回归公式与平均绝对误差对比

Table 9 Comparison of regression formula and absolute error of cylinder 1 in different models

极化方式	模型	回归公式	相关系数	拟合度	平均绝对误差/dB
HH极化	传统平方率模型	$y = x_1 + 20x_2$	0.927	0.859	4.93
HH极化	模型1	$y = 0.943x_1 + 27.175x_2$	0.980	0.961	1.94
HH极化	模型2	$y = 0.996x_1 + 21.279x_2 - 0.08x_3 - 0.708x_4$	0.998	0.995	1.11
VV极化	传统平方率模型	$y = x_1 + 20x_2$	0.934	0.872	5.26
VV极化	模型1	$y = 0.991x_1 + 30.630x_2$	0.992	0.984	2.14
VV极化	模型2	$y = 1.039x_1 + 19.476x_2 - 1.728x_3 - 0.378x_4$	0.994	0.988	1.29

性水平 $\alpha = 0.05$ 时,有

$$t_{\alpha/2}(k - n - 1) \approx 1.964 \ll t_i \quad (10)$$

且显著性 $p_i \ll 0.05$,故通过 t 检验可得各个自变量与因变量成强相关。

经过查表与计算,可得

$$F = 3\,913\,057.5 \gg F_{\alpha}(n, k - n - \alpha) \approx 3.011 \quad (11)$$

故通过 F 检验也可得到模型各个自变量与因变量成强相关,其他情况下均可得到类似的结果,在接下来的模型分析当中不再进行过多赘述。

由表8可得,圆柱1在HH极化模型1的情况下得到的回归方程为

$$y = 0.943x_1 + 27.175x_2 \quad (12)$$

3.2 圆柱1结果

采用3.1节相同的方法对圆柱1两种极化下的数据进行处理,模型1、模型2、传统平方率模型的回归方程与平均绝对误差对比如表9所示。

以原模型在2 GHz的测试频率为例,将圆柱1缩比因子为4的数据代入到各个回归方程当中,可得不同模型的预估结果与实测结果对比如图8所示。

由表9可得,在HH极化的情况下,模型1预估平均绝对误差相较于传统平方率模型的预估误差下降了约3 dB,而模型2的预估误差则在模型1的基础上下降了约0.8 dB。故对于圆柱1来说,在模型的回归方程当中添加吸波材料反射率因子取得了较好的效果。本文所提出的2种模型在预估效果上均好于传统平方率模型。

由图8(a)、(b)可得,模型1、模型2预估的RCS对角度曲线的拟合度比传统平方率模型所预估的曲线拟合度高,整体曲线预估效果良好,并且在 0° 、 90° 、 180° 、 270° 、 360° 的峰值附近,模型1

与模型2预估的数据与实测数据几乎重合,预估效果良好,由图8(c)、(d)中的红色曲线分布可知,模型2的误差整体较小,除个别点外均小于传统平方率模型误差与模型1预估的误差。

3.3 圆柱2结果

采用3.1节相同的方法对圆柱2两种极化下的数据进行处理,模型1、模型2、传统平方率模型的回归方程与平均绝对误差对比如表10所示。

以原模型在2 GHz的测试频率为例,将圆柱1缩比因子为4的数据代入到各个回归方程当中,可得不同模型的预估结果与实测结果对比如

图9所示。

由图9(c)、(d)可得,模型2的预估误差整体小于模型1和传统平方率模型,通过图9(a)、(b)可得,模型1与模型2的拟合度良好,模型2预估的数据在峰值处相较于模型1效果更好。

由表10可得,圆柱2在HH极化的情况下,模型1的预估平均绝对误差较传统平方率模型的平均预估误差下降了4.88 dB,优化效果明显,模型2相较于模型1的平均绝对误差下降了0.68 dB。在VV极化的情况下,模型1的预估平均绝对误差较传统平方率模型的平均预估误差下降了约3.5 dB。

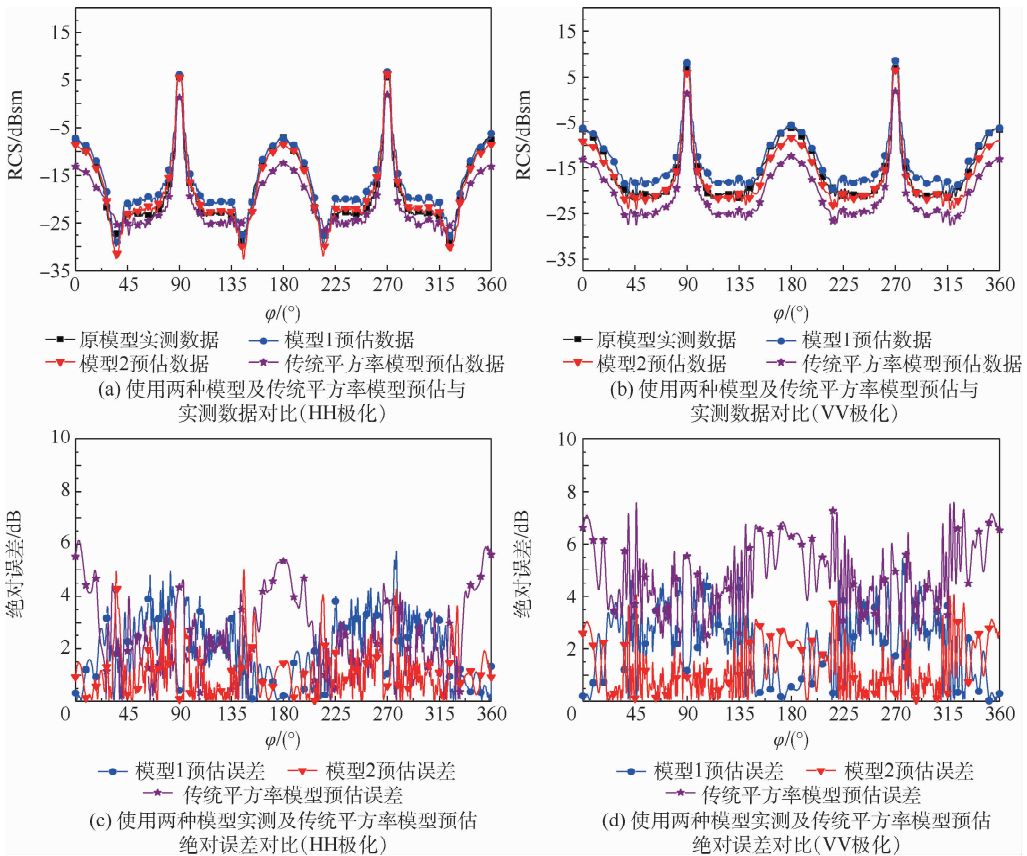


图8 圆柱1不同模型下单站RCS预估与实测对比

Fig. 8 Comparison of prediction and measurement of cylinder 1's monostatic RCS in different models

表10 圆柱2不同模型回归公式与平均绝对误差对比

Table 10 Comparison of regression formula and absolute error of cylinder 2 in different models

极化方式	模型	回归公式	相关系数	拟合度	平均绝对误差/dB
HH极化	传统平方率模型	$y = x_1 + 20x_2$	0.934	0.872	7.18
HH极化	模型1	$y = 0.969x_1 + 33.130x_2$	0.996	0.992	2.30
HH极化	模型2	$y = 0.964x_1 + 26.455x_2 + 0.355x_3 - 0.483x_4$	0.997	0.995	1.62
VV极化	传统平方率模型	$y = x_1 + 20x_2$	0.946	0.895	5.92
VV极化	模型1	$y = 1.005x_1 + 30.119x_2$	0.989	0.979	2.41
VV极化	模型2	$y = 0.969x_1 + 19.469x_2 + 0.046x_3 - 0.894x_4$	0.991	0.982	2.18

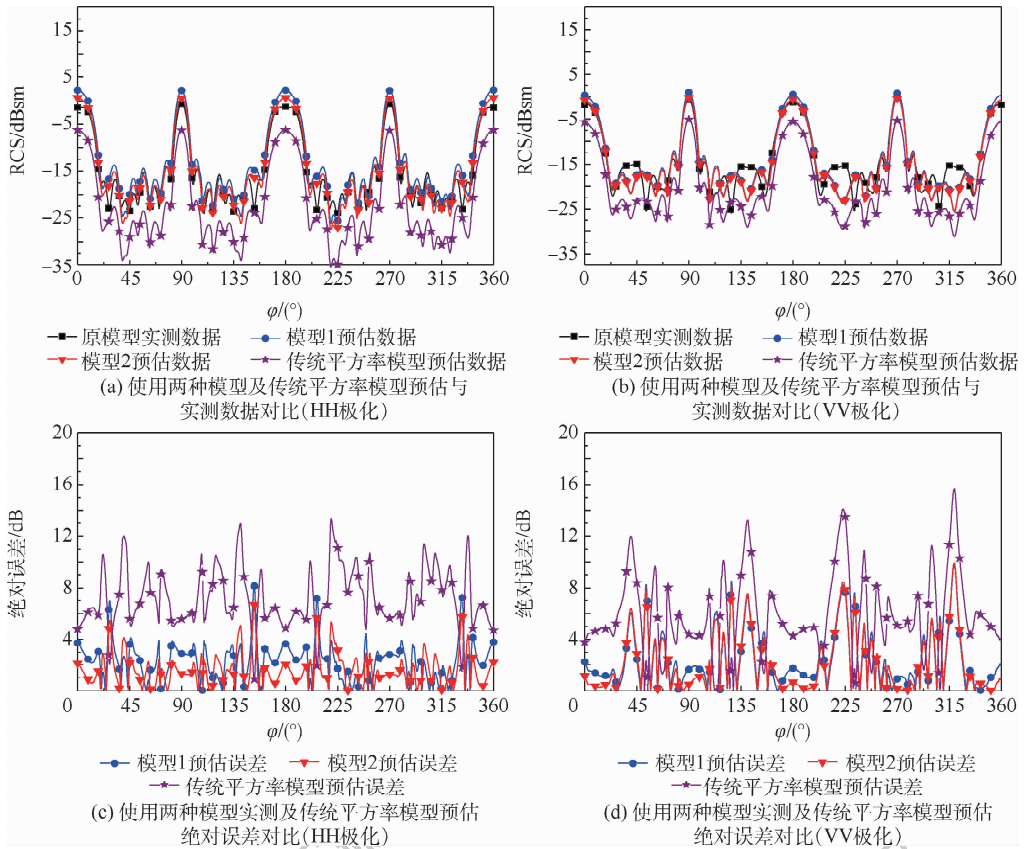


图 9 圆柱 2 不同模型下单站 RCS 预估与实测对比

Fig. 9 Comparison of prediction and measurement of cylinder 2's monostatic RCS in different models

4 结 论

1) 分析了多元对数线性回归模型应用于涂覆吸波材料目标的 RCS 缩比预估当中的理论可行性,并提出了 2 种回归预估模型。

2) 对不同尺寸的表面附有吸波材料的圆柱进行了实际 RCS 测试,分析了实验当中可能出现的误差并提出了角度矫正等数据预处理的方法。

3) 将 2 种回归模型得到的预估数据与传统平方率模型所预估的数据进行了对比,本文提出的 2 种预估模型的效果明显好于传统平方率模型,可以将平均绝对误差降低 3 ~ 5 dB。

4) 模型 2 的预估效果比模型 1 的预估效果好,模型 2 在 RCS 的峰值处拟合效果明显优于模型 1,模型 2 所预估数据的平均绝对误差相较于模型 1 下降了 0.3 ~ 0.8 dB,验证了增加反射率因子的可行性。

进一步研究翼面前缘等更多的模型效果、在模型当中合理添加新的影响因子将是今后的研究方向。

参考文献 (References)

[1] VAN SCHALKWYK R F, SMIT J C. Dynamic radar cross sec-

tion measurements of a full-scale aircraft for RCS modelling validation [C] // International Conference on Radar Systems (Radar 2017), 2017: 1-6.

[2] 李江海,孙秦. 导体与非导体介质翼型的缩比模型 RCS 数值计算分析 [J]. 航空计算技术, 2008, 38(1): 23-25.

LI J H, SUN Q. RCS numerical analysis of scaling model for conductor and nonconductor dielectric airfoil [J]. Aeronautical Computing Technology, 2008, 38(1): 23-25 (in Chinese).

[3] 夏应清,杨河林,鲁述,等. 超大电复杂目标 RCS 缩比模型预估方法 [J]. 微波学报, 2003, 19(1): 8-11.

XIA Y Q, YANG H L, LU S, et al. Prediction of RCS for extremely large electric size target by using scaling model [J]. Chinese Journal of Radio Science, 2003, 19(1): 8-11 (in Chinese).

[4] 陈晓洁,李磊,史小卫,等. 半空间理想导体目标 RCS 缩比关系的研究 [C] // 全国微波毫米波会议, 2006: 96-99.

CHEN X J, LI L, SHI X W, et al. Study on RCS scaling relationship of perfect conducting objects in half space [J]. National Conference on Microwave and Millimeter Waves (NCM-MW), 2006: 96-99 (in Chinese).

[5] 赵佼. 导体目标与海面复合电磁散射的缩比诱导关系 [D]. 西安: 西安电子科技大学, 2014: 16-18.

ZHAO J. The scale relation of the composite electromagnetic scattering from a conductor target above sea surface [D]. Xi'an: Xi'an University of Electronic Science and Technology, 2014: 16-18 (in Chinese).

[6] 李良超,吴振森,邓蓉. 复杂目标后向激光雷达散射截面计算与缩比模型测量比较 [J]. 中国激光, 2005, 32(6): 44-48.

- LI L C, WU Z S, DENG R. Comparison of the back laser radar cross-section calculation of complex object with the measurement of scale-model[J]. *China Laser*, 2005, 32(6): 44-48 (in Chinese).
- [7] SHI Z D, DING C S, CHEN J Y. Dimensional analysis and physical similarity of lossy electromagnetic systems[J]. *Chinese Physics Letters*, 1993, 10(6): 347-350.
- [8] 时振栋, 刘宏伟. 隐身目标雷达截面的缩比测量及反演计算[J]. *电子科技大学学报*, 1995(S1): 13-17.
SHI Z D, LIU H W. Scale measurement and inversion calculation of radar cross section of invisible target[J]. *Journal of University of Electronic Science and Technology*, 1995(S1): 13-17 (in Chinese).
- [9] 时振栋, 杨仕文, 丁春生. 有耗雷达目标散射截面的缩比计算——再论有耗电磁系统的相似性[J]. *应用科学学报*, 1993, 11(2): 109-114.
SHI Z D, YANG S W, DING C S. Computation of the scattering cross section of lossy radar targets—On the similarity of lossy electromagnetic systems [J]. *Journal of Applied Sciences*, 1993, 11(2): 109-114 (in Chinese).
- [10] 刘宏伟, 时振栋, 唐璞. 有耗雷达目标的 RCS 模型测量方法[J]. *电子科学学刊*, 1996, 18(1): 95-98.
LIU H W, SHI Z D, TANG P. RCS model measurement method for lossy radar targets[J]. *Journal of Electronic Science*, 1996, 18(1): 95-98 (in Chinese).
- [11] 胡艳, 时振栋, 唐璞, 等. 电磁相似性在计算基本散射体 RCS 方面的应用[J]. *应用科学学报*, 2004, 22(1): 46-49.
HU Y, SHI Z D, TANG P, et al. The application of electromagnetic similarity in the calculation of RCS of the basic scatterer [J]. *Journal of Applied Sciences*, 2004, 22(1): 46-49 (in Chinese).
- [12] 陈晓洁. 电大目标雷达散射截面的研究[D]. 西安: 西安电子科技大学, 2006: 14-16.
CHEN X J. Research on radar cross section of electrically large target[D]. Xi'an: Xi'an University of Electronic Science and Technology, 2006: 14-16 (in Chinese).
- [13] 蒋文亭, 吴德伟, 何晶. 目标雷达散射截面缩比测量关键技术概述[J]. *信息工程大学学报*, 2013, 14(2): 184-188.
JANG W T, WU D W, HE J. Summary of key technologies of scale model measurement of RCS [J]. *Journal of Information Engineering University*, 2013, 14(2): 184-188 (in Chinese).
- [14] SUSETIO A, OKI T, MORISHITA H. Monostatic radar cross section evaluations of scale model single layer dielectric radar absorbing material at S and X bands [C] // IEEE 4th Asia-Pacific Conference on Antennas and Propagation (APCAP). Piscataway: IEEE Press, 2015: 568-569.
- [15] SINCLAIR G. Theory of models of electromagnetic systems[J]. *Proceedings of the IRE*, 1948, 36(11): 1364-1370.
- [16] 刘闯, 金仁瀚, 巩二磊, 等. 多元线性回归方法预测燃气轮机发电机组性能[J]. *中国电机工程学报*, 2017, 37(16): 4731-4738.
LIU C, JIN R H, GONG E L, et al. Prediction for the performance of gas turbine units using multiple linear regression [J]. *Proceedings of the CSEE*, 2017, 37(16): 4731-4738 (in Chinese).
- [17] 王振友, 陈莉娥. 多元线性回归统计预测模型的应用[J]. *统计与决策*, 2008(5): 46-47.
WANG Z Y, CHEN L E. Application of multivariate linear regression statistical prediction model [J]. *Statistics and Decision*, 2008(5): 46-47 (in Chinese).
- [18] 黄兵, 胥云, 廖映华, 等. 基于 BP 神经网络和多元线性回归的航发叶片铣削力预测研究[J]. *机电工程*, 2019, 36(8): 824-829.
HUANG B, XU Y, LIAO Y H, et al. Milling force forecasting of aero-engine blade based on BP neural network and multivariate linear regression [J]. *Journal of Mechanical & Electrical Engineering*, 2019, 36(8): 824-829 (in Chinese).
- [19] PERRON P, YAMAMOTO Y. Pitfalls of two-step testing for changes in the error variance and coefficients of a linear regression model [J]. *Econometrics*, 2019, 7(2): 1-11.
- [20] 中国航天标准化研究所. 室内场目标雷达散射截面测试方法: GJB 5022—2001 [S]. 北京: 国防科学技术工业委员会, 2001: 1-14.
China Aerospace Standardization Institute. Method for measurement of radar cross section of scale target indoor range: GJB 5022—2001 [S]. Beijing: Commission on Science, Technology and Industry for National Defense, 2001: 1-14 (in Chinese).
- [21] 李贵晓. 化工过程动态数据校正与过失误差侦破的研究 [D]. 青岛: 青岛科技大学, 2014: 5-8.
LI G X. Research on dynamic data correction and error detection of chemical process [D]. Qingdao: Qingdao University of Science and Technology, 2014: 5-8 (in Chinese).

RCS prediction method for cylinder containing microwave absorbing material

ZHAO Jingcheng^{1,*}, YANG Zongkai¹, WANG Jian², ZHANG Yang²

(1. School of Electronic and Information Engineering, Beihang University, Beijing 100083, China;

2. AVIC The First Aircraft Institute, Xi'an 710089, China)

Abstract: Large target's Radar Cross Section (RCS) estimation using scaling model is a common method to obtain RCS at the development phase. However, according to classical electromagnetic similarity theory, measurement of scaled target coated with microwave absorbing material is difficult to meet the scaling condition strictly. A multivariate logarithmic linear regression model is proposed to estimate RCS for the scaled target coated with microwave absorbing materials. Two sets of cylindrical models were designed and tested in the microwave anechoic chamber with scaling factors of 1:1, 1:2, 1:4 and 1:8, respectively. On the basis of data preprocessing such as angle correction, RCS data of scaling model is substituted into the model as training set to obtain parameters, and RCS of the original model is estimated and compared with the actual measured data. The results show that the curves of the predicted data and the measured data fit well. Compared with the traditional square rate formula, the error of the proposed method decreases by 3 – 5 dB, and the error decreases by 0.3 – 0.8 dB after adding the microwave absorbing material factor to the regression model.

Keywords: Radar Cross Section (RCS); scaling model; multiple linear regression; microwave absorbing material; similarity principle

http://bhxb.buaa.edu.cn jbuua@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2020.0719

抬升角对食蚜蝇飞行动稳定性的影响



许娜¹, 周帅至¹, 牟晓蕾^{2,*}

(1. 烟台大学 机电汽车工程学院, 烟台 264005; 2. 烟台大学 土木工程学院, 烟台 264005)

摘 要: 抬升角的存在对昆虫飞行时的气动力会产生不同程度的影响, 其对昆虫飞行动稳定性的影响也非常值得探索。首先, 求解 Navier-Stokes 方程得到了有抬升角时食蚜蝇的纵向与横向稳定性导数; 然后, 用特征模态分析法研究其动稳定性。结果表明: 有抬升角时, 在侧向来流作用下食蚜蝇的滚转力矩对应的导数比无抬升角时明显减小, 而其余导数无明显变化, 导数减小是由于抬升角的存在使得有侧向来流时因左右翅举力不同产生的正向滚转力矩数值明显减小, 而由侧向力产生的负向滚转力矩数值略有增大, 从而使得其总的负向滚转力矩数值增大; 但有侧向来流时滚转力矩所对应导数的减小并未引起食蚜蝇飞行动稳定性的改变, 其纵向和横向的特征模态仍与无抬升角时相同。

关键词: 抬升角; 食蚜蝇; 飞行动稳定性; Navier-Stokes 方程; 运动模态

中图分类号: V211.3

文献标志码: A

文章编号: 1001-5965(2021)07-1372-07

昆虫在飞行过程中, 其翅膀通常会不同程度背离拍动平面^[1], 形成如“8”字形^[1-2]、椭圆形、“U”形^[2-3]及这些形状组合而成的扑翼轨迹^[4-7]。扑翼轨迹的多样性与不确定性给人们带来了很大的困惑, 为了弄清楚扑翼轨迹对其气动性能的影响, 人们进行了一系列的研究^[8-12], 结果表明, 不同扑翼轨迹对昆虫的飞行气动力和能耗不尽相同但影响不大, 基本在 10% 左右。除了气动力和能耗以外, 飞行动稳定性对昆虫的飞行至关重要。Jankauski 等^[13]指出抬升角能使昆虫在稳定飞行状态下产生转矩, 从而调整其飞行姿态, 对昆虫的飞行控制和机动飞行产生影响。人们针对昆虫飞行动稳定性开展的大量研究工作, 都将昆虫翅膀的运动简化为只有拍动运动和翻转运动, 忽略了抬升角^[14-19]。

本文在此前工作^[17,19]的基础上, 计算了有抬升角时(即食蚜蝇以“U”形扑翼轨迹飞行时)食

蚜蝇的纵向与横向稳定性导数, 求解了稳定性矩阵的特征值, 并将这些结果与无抬升角时的动稳定性结果^[17,19]进行比较, 研究了抬升角对食蚜蝇飞行动稳定性的影响。

1 计算模型及方法

1.1 运动方程与分析方法

为方便描述食蚜蝇的运动, 分别引入固定在食蚜蝇身体上随其一起运动的随体坐标系 $O_b x_b y_b z_b$ 和在地面上的固定坐标系 $OXYZ$, 如图 1 所示。 O_b 位于身体的质心, 当食蚜蝇平衡飞行时, x_b 和 y_b 轴在水平面内, x_b 轴水平向前, y_b 轴指向食蚜蝇身体右侧, z_b 轴垂直向下。 u 、 v 和 w 分别为昆虫质心在 x_b 、 y_b 和 z_b 轴方向上的平动速度, 而 p 、 q 和 r 分别为昆虫身体绕 x_b 、 y_b 和 z_b 轴的转动角速度, θ_b 和 γ 分别为食蚜蝇身体上俯和向右滚转时 x_b 和 y_b 轴与水平面的夹角。

收稿日期: 2020-12-28; 录用日期: 2021-01-30; 网络出版时间: 2021-02-07 14:45

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20210207.1404.001.html

基金项目: 国家自然科学基金(11802262, 11502228)

* 通信作者。E-mail: xiaoleimou@126.com

引用格式: 许娜, 周帅至, 牟晓蕾. 抬升角对食蚜蝇飞行动稳定性的影响[J]. 北京航空航天大学学报, 2021, 47(7): 1372-1378.

XU N, ZHOU S Z, MOU X L. Effects of stroke deviation on flight stability of true hoverfly [J]. Journal of Beijing University of Aeronautics and Astronautics, 2021, 47(7): 1372-1378 (in Chinese).

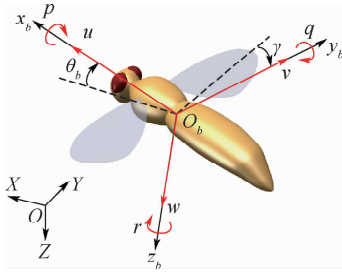


图1 食蚜蝇的身体坐标系和地面上的固定坐标系及各变量示意图

Fig.1 Schematic diagram of true hoverfly body coordinate system and fixed ground coordinate system with variables

与此前研究昆虫飞行稳定性方法^[17,19-21]相同,本文仍采用平均模型和线性理论分析方法。 X, Y 和 Z 分别为气动力一个拍动周期内的平均值在 x_b, y_b 和 z_b 轴方向的分量, L, M 和 N 分别为绕 x_b, y_b 和 z_b 轴的平均气动力矩。食蚜蝇在悬停飞行时, u, v, w, p, q 和 r 及其导数均为零,且 X, Y, L, M 和 N 也均等于零,而 $Z = -mg, m$ 为食蚜蝇的总质量, g 为重力加速度。运动方程线性化之后,

便可将其解耦为纵向和横向2种运动方程^[20](见式(1)和式(2)),分别进行研究。

$$\begin{bmatrix} \delta \dot{u}^+ \\ \delta \dot{w}^+ \\ \delta \dot{q}^+ \\ \delta \dot{\theta}_b^+ \end{bmatrix} = \mathbf{A}_1^+ \begin{bmatrix} \delta u^+ \\ \delta w^+ \\ \delta q^+ \\ \delta \theta_b^+ \end{bmatrix} \quad (1)$$

$$\begin{bmatrix} \delta \dot{v}^+ \\ \delta \dot{p}^+ \\ \delta \dot{r}^+ \\ \delta \dot{\gamma}^+ \end{bmatrix} = \mathbf{A}_2^+ \begin{bmatrix} \delta v^+ \\ \delta p^+ \\ \delta r^+ \\ \delta \gamma^+ \end{bmatrix} \quad (2)$$

式中:“ \cdot ”表示各运动量关于时间 t 的导数; \mathbf{A}_1^+ 为纵向稳定性矩阵; \mathbf{A}_2^+ 为横向稳定性矩阵。

$$\mathbf{A}_1^+ = \begin{bmatrix} X_u^+/m^+ & X_w^+/m^+ & X_q^+/m^+ & -g^+ \\ Z_u^+/m^+ & Z_w^+/m^+ & Z_q^+/m^+ & 0 \\ M_u^+/I_y^+ & M_w^+/I_y^+ & M_q^+/I_y^+ & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad (3)$$

$$\mathbf{A}_2^+ = \begin{bmatrix} Y_v^+/m^+ & Y_p^+/m^+ & Y_r^+/m^+ & g^+ \\ (I_z^+ L_v^+ + I_{xz}^+ N_v^+)/ (I_x^+ I_z^+ - I_{xz}^{+2}) & (I_z^+ L_p^+ + I_{xz}^+ N_p^+)/ (I_x^+ I_z^+ - I_{xz}^{+2}) & (I_z^+ L_r^+ + I_{xz}^+ N_r^+)/ (I_x^+ I_z^+ - I_{xz}^{+2}) & 0 \\ (I_{xz}^+ L_v^+ + I_x^+ N_v^+)/ (I_x^+ I_z^+ - I_{xz}^{+2}) & (I_{xz}^+ L_p^+ + I_x^+ N_p^+)/ (I_x^+ I_z^+ - I_{xz}^{+2}) & (I_{xz}^+ L_r^+ + I_x^+ N_r^+)/ (I_x^+ I_z^+ - I_{xz}^{+2}) & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \quad (4)$$

式中: $X_u^+, Z_u^+, M_u^+, X_w^+, Z_w^+, M_w^+, X_q^+, Z_q^+$ 和 M_q^+ 为食蚜蝇在扰动运动时的纵向稳定性导数; $Y_v^+, L_v^+, N_v^+, Y_p^+, L_p^+, N_p^+, Y_r^+, L_r^+$ 和 N_r^+ 为食蚜蝇在扰动运动时的横向稳定性导数; I_x^+, I_y^+, I_z^+ 和 I_{xz}^+ 分别为身体坐标系 $O_b x_b y_b z_b$ 下的转动惯量和惯性积;“+”的变量和参数均用参考速度 $U (U = 2\Phi nr_2)$ 为平均拍动速度, Φ 为拍翅幅值, n 为拍翅频率, r_2 为翅膀面积二阶矩折合半径、参考长度 $c (c$ 为翅膀平均弦长)和参考时间 $T (T = 1/n$ 为拍翅周期)进行了无量纲化处理,各无量纲表达式如下:
 $\delta u^+ = \delta u/U, \delta v^+ = \delta v/U, \delta w^+ = \delta w/U, \delta p^+ = \delta pT, \delta q^+ = \delta qT, \delta r^+ = \delta rT, m^+ = m/(0.5\rho U S_i T), I_x^+ = I_x/(0.5\rho U^2 S_i c T^2), I_y^+ = I_y/(0.5\rho U^2 S_i c T^2), I_z^+ = I_z/(0.5\rho U^2 S_i c T^2), I_{xz}^+ = I_{xz}/(0.5\rho U^2 S_i c T^2), g^+ = gT/U, X^+ = X/(0.5\rho U^2 S_i), Y^+ = Y/(0.5\rho U^2 S_i), Z^+ = Z/(0.5\rho U^2 S_i), L^+ = L/(0.5\rho U^2 S_i c), M^+ = M/(0.5\rho U^2 S_i c), N^+ = N/(0.5\rho U^2 S_i c), t^+ = t/T, \rho$ 为空气密度, S_i 为双翅面积。

本文所选用的食蚜蝇的形态学参数(m, I_x

等)在文献^[17,19]中已经给出,因此只需计算得到有抬升角时的纵向和横向的气动导数即可确定稳定性矩阵 \mathbf{A}_1^+ 和 \mathbf{A}_2^+ 。基于稳定性矩阵可求出特征值^[17],相应的特征值代表了相应的简单运动,该运动即为系统的特征模态;处于平衡飞行状态的食蚜蝇,受到小扰动后,其自由运动即为这些特征模态的组合。特征值为正数时($\lambda > 0$)表示不稳定发散模态,初始扰动会随时间增长,扰动振幅的倍幅期为 $0.693/|\lambda|$ 。特征值为负数时($\lambda < 0$)表示衰减模态,初始扰动会随时间减小最终消失,扰动振幅的半衰期为 $0.693/|\lambda|$ 。特征值为复数时,如 $\lambda_{1,2} = \hat{n} \pm \hat{\omega}i$,扰动会随时间振荡,其振荡周期取决于特征值虚部 $\hat{\omega}$,扰动的增长或衰减取决于特征值的实部 \hat{n} 。当 $\hat{n} > 0$ 时,扰动会增长,表示不稳定振荡模态;反之表示扰动会衰减,表示稳定的振荡模态,振荡模态的倍幅期(或半衰期)和周期分别由 $0.693/|\hat{n}|$ 和 $2\pi/\hat{\omega}$ 给出。

1.2 流动方程求解与稳定性导数

1.2.1 翅膀运动

本文中食蚜蝇的模型翅膀与文献^{[3,12,17,}

19]中用的模型翅膀相同,其计算网格沿展向和弦向的剖面如图2所示。

翅膀的各运动学参数定义如图3所示。图中: β 为拍动平面的倾角(见图3(a)),而翅膀运动的欧拉角 ϕ (拍动角)、 ψ (翻转角)和 θ (抬升角)定义在拍动平面坐标系($O_w X_w Y_w Z_w$)内如图3(b)所示,其中 $O_w x_w y_w z_w$ 为有抬升角 θ 时固定在食蚜蝇翅膀上的动坐标系, $O_w x'_w y'_w z'_w$ 为无抬升角 θ 时固定在其翅膀上的动坐标系。

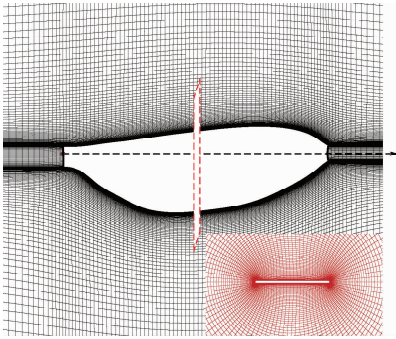


图2 食蚜蝇翅膀网格的展向和弦向剖面
Fig. 2 True hoverfly wing planform and portion of computational grid

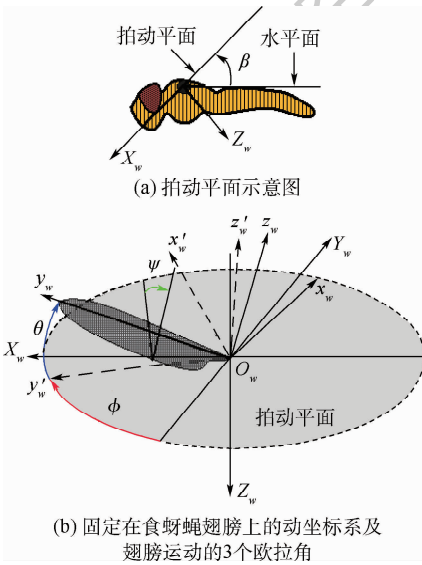


图3 拍动翅的运动学参数定义

Fig. 3 Definition of kinematic parameters of flapping wing

翅膀转动的3个欧拉角随时间的变化可以用不同的谐波函数表示,拍动角 ϕ 可用下面的余弦函数表示:

$$\phi = \bar{\phi} + 0.5\Phi\cos(2\pi nt) \tag{5}$$

式中: $\bar{\phi}$ 为平均拍动角, $\bar{\phi} = (\phi_{\max} + \phi_{\min})/2$, $\Phi = \phi_{\max} - \phi_{\min}$, ϕ_{\max} 和 ϕ_{\min} 分别为拍动角的最大值和最小值。

对于翻转角,其在翅膀上/下拍动的中部不随时间改变近似为常数,而在翻转过程中随时间变

化,如在第 N 个拍动周期翅膀由下拍变为上拍时 ψ 可表示为

$$\psi = \alpha_d + a\{(t - t_1) - [\Delta t_s / (2\pi)]\sin(2\pi(t - t_1) / \Delta t_s)\} \tag{6}$$
$$t_1 \leq t \leq t_1 + \Delta t_s$$

式中: a 为常数, Δt_s 为翅膀上仰时间, $a = (180^\circ - \alpha_u - \alpha_d) / \Delta t_s$, α_d 为翅膀下拍中部攻角, α_u 为上拍中部攻角; t_1 为翅膀开始上仰的时刻, $t_1 = NT - 0.5T - \Delta t_s / 2$ 。翅膀下俯运动时与以上定义类似。

实验观测结果^[3]显示,抬升角周期变化的频率约为拍翅频率的2倍,一般用如下谐波函数^[2-3]来拟合曲线:

$$\theta = \theta_0 + \theta_1 \sin(2\pi t + \eta_1) + \theta_2 \sin(4\pi t + \eta_2) \tag{7}$$

式中: θ_0 为偏离拍动平面的平均值; θ_1 、 θ_2 和 η_1 、 η_2 分别为第一、二阶项的幅度和相位角。

根据实验测量结果^[3], $\Phi = 65.6^\circ$, $n = 162 \text{ Hz}$, $\Delta t_s = 0.35$, $\theta_0 = 6^\circ$ 。加上抬升角后,本文通过力和力矩的平衡条件在无抬升角的基础^[19]上对上下拍攻角 α_d 、 α_u 和平均拍动角 $\bar{\phi}$ 进行了微调,新的平衡状态下 $\alpha_d = 18.9^\circ$, $\alpha_u = 49^\circ$, $\bar{\phi} = 16.9^\circ$ 与无抬升角^[19]时差别不大。

1.2.2 流动方程求解

稳定性导数需要通过求解 Navier-Stokes 方程来获得,Navier-Stokes 方程中的无量纲参数为雷诺数 Re ($Re = Uc/\nu$, ν 为运动黏性系数),食蚜蝇悬停飞行时 $Re \approx 323$ ^[3,19]。本文求解 Navier-Stokes 方程的算法基于 Rogers 等^[22]发展的拟压缩性算法,计算程序与之前的研究所采用的相同^[16-17,19,23],能够保证计算结果的可靠性和精度。

本文所使用的翅膀网格沿展向和弦向的剖面形状如图2所示,网格近壁面法向第1层间距为平均弦长 c 的1/1000,翅膀壁面沿法向到远场边界的距离约为 $20c$,沿展向方向到远场边界的距离为 $6c$ 。文献[3]中已经对网格密度、计算域、时间步长等进行了验证,进一步证明了本文所采用的计算参数和计算方法的可靠性。

1.2.3 稳定性导数计算

本文中,纵向和横向稳定性导数的计算方法与文献[15,17,19,24]中对昆虫纵向和横向稳定性问题研究时所采用的方法相同。以横向稳定性导数 Y_v 、 L_v 和 N_v 为例,简要介绍导数计算。食蚜蝇悬停飞行,其各向力与力矩处于平衡状态时,保持 p^+ 、 r^+ 、 γ 均为0,对 v^+ 在0附近进行微调,计算得到一系列气动力和力矩相对悬停时的变化量(ΔY^+ 、 ΔL^+ 和 ΔN^+)随 v^+ 的变化曲线,拟合该系

列曲线,各曲线的斜率即为相应的稳定性导数。同理,可以计算得到 u^+ 、 w^+ 和 q^+ 系列等纵向稳定性导数与 p^+ 和 r^+ 系列等横向稳定性导数。计算得到稳定性导数后即可确定稳定性矩阵 A_1^+ 和 A_2^+ ,进而可以对有抬升角时食蚜蝇的飞行稳定性进行分析。

2 计算结果分析

2.1 稳定性导数比较

如 1.2.1 节所述,通过计算确定了有抬升角时食蚜蝇的平衡飞行状态,并在此平衡状态下对 u^+ 、 v^+ 、 w^+ 、 p^+ 、 q^+ 和 r^+ 等状态变量进行微调,分别得到了相应的气动力 (X^+ 、 Y^+ 、 Z^+) 与气动力矩 (L^+ 、 M^+ 、 N^+)。这里将有抬升角和无抬升角时的各气动力与气动力矩随各状态变量变化曲线进行对比,结果如图 4 和图 5 所示。可以发现,有抬升角时各气动力和力矩系数随状态变量的变化线性较好,变化趋势也与无抬升角时大致相同。但值得注意的是,如图 5 所示,有抬升角时 ΔL^+

随横向速度 v^+ 的变化率比无抬升角时要大。为进行更深入的比较,根据图 4 和图 5 的变化曲线,计算出了纵向和横向稳定性导数(见表 1 和表 2)。通过比较可以发现,有无抬升角 2 种情形下,所有纵向稳定性导数的差异都很小,且绝大多数的横向稳定性导数也无明显差异,但是有抬升角时的 L_v^+ 导数比无抬升角时更小,数值上约为无抬升角时的 2 倍。有抬升角时 L_v^+ 为何会有如此明显的变化,其对横向稳定性是否会产生影响值得进行深入的探讨。

首先分析抬升角的存在为何使得 L_v^+ 导数进一步减小,在此之前需要明确有侧向来流时滚转力矩是如何产生的。侧向来流的存在引起了食蚜蝇左右翅气动力的不对称,这里将计算得到的两翅的气动力系数分别沿 Y 和 Z 轴投影,可得其侧向力系数 Y^+ 和举力系数 Z^+ ,通过分析比较左右翅 Y^+ 和 Z^+ 的差异可以得出滚转力矩系数 ΔL^+ 产生的原因。以横向扰动运动速度 $v^+ = 0.15$ 为例,将有抬升角和无抬升角 2 种情形下左右两翅的

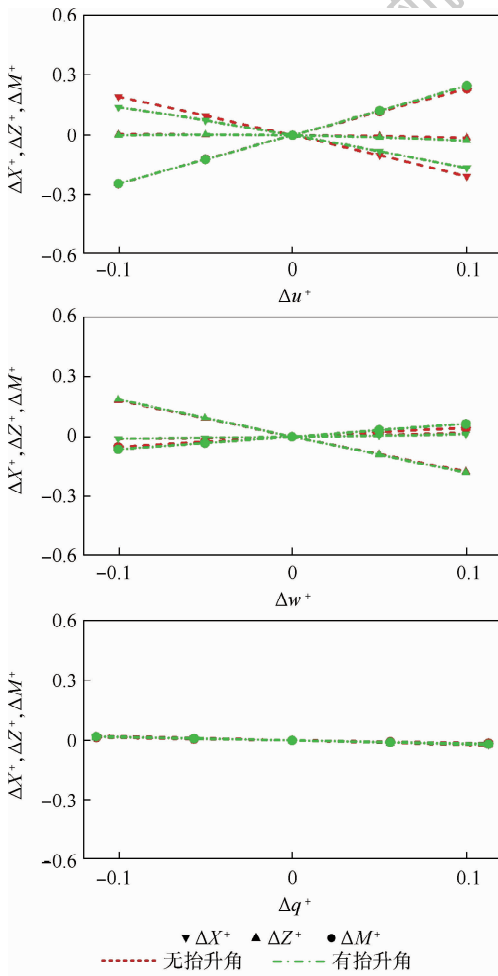


图 4 纵向 u^+ 、 w^+ 和 q^+ 系列气动力和力矩系数
Fig. 4 Longitudinal u^+ 、 w^+ and q^+ series aerodynamic force and moment coefficients

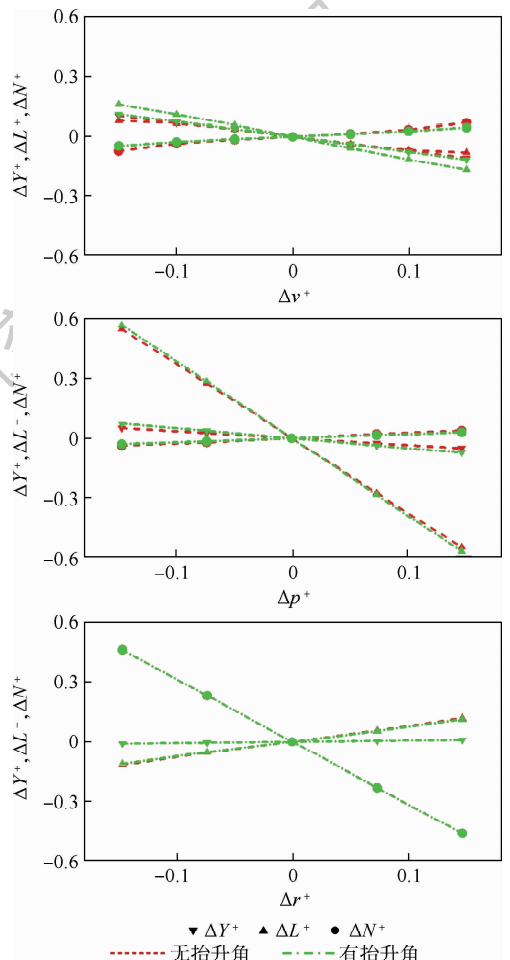


图 5 横向 v^+ 、 p^+ 和 r^+ 系列气动力和力矩系数
Fig. 5 Lateral v^+ 、 p^+ and r^+ series aerodynamic force and moment coefficients

Y^+ 、 Z^+ 和绕 X 轴的滚转力矩系数 ΔL^+ 进行了对比,如表3所示。通过比较可以发现,2种情形下,右翅产生的沿 Y 轴负向的侧向力系数数值均大于左翅沿 Y 轴正向的侧向力系数,左右两翅侧向力系数的差值 ΔY^+ 分别为 -0.210 和 -0.228 ,而食蚜蝇的形态学数据测量结果^[3]表明翅根位于身体质心的上方,因此 ΔY^+ 都会使食蚜蝇产生绕 X 轴负向的滚转力矩,且有抬升角时的左右两翅侧向力系数的差值 ΔY^+ 比无抬升角时略大,即有抬升角时绕 X 轴负向的滚转力矩比无抬升角时大。再者2种情形下,左翅沿 Z 轴的举力系数

均大于右翅,故 Z^+ 会使食蚜蝇产生绕 X 轴正向的滚转力矩,但有抬升角时左右两翅举力系数的差值 ΔZ^+ 仅为无抬升角时的约 $1/3$,这说明抬升角的存在使得左右两翅举力系数的差值减小,从而削弱了食蚜蝇正向的滚转力矩。综上,抬升角的存在使得左右两翅的侧向力差值 ΔY^+ 增大,增大了食蚜蝇的负向滚转力矩;却减小了左右两翅举力系数的差值 ΔZ^+ ,即减小了相应的正向滚转力矩,从而导致有抬升角时绕 X 轴负向的滚转力矩 ΔL^+ 是无抬升角时的2倍左右,最终使得有抬升角时的横向稳定性导数 L_v^+ 的数值约为无抬升角时的2倍。

表1 纵向无量纲稳定性导数

Table 1 Longitudinal non-dimensional stability derivatives

有/无抬升角	X_u^+	Z_u^+	M_u^+	X_w^+	Z_w^+	M_w^+	X_q^+	Z_q^+	M_q^+
无抬升角	-1.998	-0.087	2.378	0.146	-1.777	0.475	-0.206	-0.220	-0.128
有抬升角	-1.519	-0.136	2.449	0.112	-1.829	0.630	-0.155	-0.188	-0.174

表2 横向无量纲稳定性导数

Table 2 Lateral non-dimensional stability derivatives

有/无抬升角	Y_v^+	L_v^+	N_v^+	Y_p^+	L_p^+	N_p^+	Y_r^+	L_r^+	N_r^+
无抬升角	-0.700	-0.530	0.473	-0.341	-3.727	0.249	0.066	0.805	-3.156
有抬升角	-0.759	-1.100	0.303	-0.495	-3.891	0.204	0.066	0.768	-3.155

表3 $v^+ = 0.15$ 时有无抬升角2种情形下左右两翅的侧向力系数、举力系数和滚转力矩系数

Table 3 Coefficients of lateral force, vertical force and rolling moment of left and right wing at

$v^+ = 0.15$ with and without stroke deviation

有/无抬升角	Y_L^+	Y_R^+	ΔY^+	Z_L^+	Z_R^+	ΔZ^+	L_L^+	L_R^+	ΔL^+
无抬升角	0.093	-0.303	-0.210	1.773	1.690	0.083	2.843	-2.922	-0.079
有抬升角	0.326	-0.554	-0.228	1.780	1.749	0.031	2.961	-3.125	-0.164

2.2 稳定性分析

通过稳定性导数的比较可以发现,有无抬升角2种情形下除了 L_v^+ 导数的显著差异外其余稳定性导数差别不大,因此可以预见抬升角存在与否对食蚜蝇的纵向稳定性可能不会产生影响,但是有抬升角时的 L_v^+ 比无抬升角时小,此前的研究^[19]发现随着 L_v^+ 的变小,昆虫的横向不稳定性会变弱, L_v^+ 的不同可能会对其横向稳定性产生影响。因此,为了进一步分析抬升角对食蚜蝇稳定性的影响,将得到的纵向和横向稳定性导数代入式(3)和式(4)中的纵向稳定性矩阵 A_1^+ 和横向稳定性矩阵 A_2^+ 中,并用MATLAB软件求解得到了 A_1^+ 和 A_2^+ 的特征值,这里将有抬升角和无抬升角时特征值的对比结果分别列于表4和表5中。

表4 纵向稳定性矩阵 A_1^+ 的特征值

Table 4 Eigenvalues of longitudinal system matrix A_1^+

有/无抬升角	模态1	模态2	模态3
	λ_1	λ_2	$\lambda_{3,4}$
无抬升角	-0.169	-0.028	$0.065 \pm 0.136i$
有抬升角	-0.169	-0.028	$0.067 \pm 0.138i$

通过比较表4中的数据可以发现,有抬升角时食蚜蝇的纵向运动的特征值(特征模态)与无抬升角^[17]时相似,也分别为一个数值较大的负实根 λ_1 (代表快衰减模态)、一个数值较小的负实根 λ_2 (代表慢衰减模态)和一对具有正实部的共轭复根 $\lambda_{3,4}$ (代表不稳定振荡模态)。这与文献[14-15, 25]中对蝇类、蜂类、蛾类等昆虫的纵向稳定性研究结果一致。而由表5则可以发现,有抬升角时食蚜蝇的横向运动的特征值(特征模态)与无抬升角^[19]时也相似,分别为一个数值较大的负实根 λ_1 (代表快衰减模态)、一个数值较小的负实根 λ_2 (代表慢衰减模态)和一对具有极小负实部的共轭复根 $\lambda_{3,4}$ (代表弱稳定振荡模态)。综上可以发现,抬升角的存在与否除了对食蚜蝇悬停时

表5 横向稳定性矩阵 A_2^+ 的特征值

Table 5 Eigenvalues of lateral system matrix A_2^+

有/无抬升角	模态1	模态2	模态3
	λ_1	λ_2	$\lambda_{3,4}$
无抬升角	-1.269	-0.153	$-0.006 \pm 0.058i$
有抬升角	-1.330	-0.157	$-0.003 \pm 0.089i$

的气动力有微小影响外,其对食蚜蝇的纵向与横向扰动运动特征模态没有影响,这也说明此前在研究食蚜蝇悬停飞行的气动力与动稳定性问题时只考虑了翅膀的拍动角和攻角而忽略了抬升角这一做法是合理的。

3 结论

1) 有抬升角时,食蚜蝇的各气动力和气动力矩系数随各状态变量的变化规律线性较好,变化率与无抬升角时基本相同。

2) 除 L_v^+ 导数外,其余稳定性导数与无抬升角时的差异很小,甚至相同;有抬升角时的 L_v^+ 比无抬升角时更小,数值上约为无抬升角时的2倍;该值变小是由于抬升角的存在使得有侧向来流时左右翅的举力系数差别减小,即减小了左右翅举力不同产生的正滚转力矩;而左右翅侧向力的差值比无抬升角时略有增大,即有抬升角时侧向力产生的负滚转力矩数值略有增大,从而使得总的负滚转力矩变大。

3) 抬升角的存在并未改变食蚜蝇的纵向和横向动稳定性,其特征模态与无抬升角时相同,只是特征值的数值略有差异。因此,研究昆虫悬停飞行的动稳定性问题时,当抬升角在一定范围内只考虑翅膀的拍动角和攻角而忽略了抬升角这一做法是合理的。

参考文献 (References)

- [1] ELLINGTON C. The aerodynamics of hovering insect flight. III. Kinematics[J]. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 1984, 305(1122): 41-78.
- [2] LIU Y, SUN M. Wing kinematics measurement and aerodynamics of hovering droneflies[J]. Journal of Experimental Biology, 2008, 211(13): 2014-2025.
- [3] MOU X L, LIU Y P, SUN M. Wing motion measurement and aerodynamics of hovering true hoverflies[J]. Journal of Experimental Biology, 2011, 214(17): 2832-2844.
- [4] ENNOS A R. The kinematics and aerodynamics of the free flight of some diptera[J]. Journal of Experimental Biology, 1989, 142(1): 49-85.
- [5] WILLMOTT A P, ELLINGTON C P. The mechanics of flight in the hawkmoth *Manduca sexta*. I. Kinematics of hovering and forward flight[J]. Journal of Experimental Biology, 1997, 200(21): 2705-2722.
- [6] FRY S N, SAYAMAN R, DICKINSON M H. The aerodynamics of free-flight maneuvers in *Drosophila*[J]. Science, 2003, 300(5618): 495-498.
- [7] FRY S N, SAYAMAN R, DICKINSON M H. The aerodynamics of hovering flight in *Drosophila*[J]. Journal of Experimental Biology, 2005, 208(12): 2303-2318.
- [8] WANG C, ZHOU C Y, XIE P. Numerical investigation into the effects of stroke trajectory on the aerodynamic performance of insect hovering flight[J]. Journal of Mechanical Science and Technology, 2016, 30(4): 1659-1669.
- [9] LUO G Y, DU G, SUN M. Effects of stroke deviation on aerodynamic force production of a flapping wing[J]. AIAA Journal, 2018, 56(1): 25-35.
- [10] HU F J, LIU X M. Effects of stroke deviation on hovering aerodynamic performance of flapping wings[J]. Physics of Fluids, 2019, 31(11): 111901.
- [11] LEI M L, LI C Y. The aerodynamic performance of passive wing pitch in hovering flight[J]. Physics of Fluids, 2020, 32(5): 051902.
- [12] 牟晓蕾, 许娜. 翅尖轨迹对食蚜蝇悬停时气动特性的影响[J]. 北京航空航天大学学报, 2016, 42(12): 2603-2609.
MOU X L, XU N. Effect of wing-tip trajectory on aerodynamics of hovering true hoverfly[J]. Journal of Beijing University of Aeronautics and Astronautics, 2016, 42(12): 2603-2609 (in Chinese).
- [13] JANKAUSKI M, DANIEL T L, SHEN I Y. Asymmetries in wing inertial and aerodynamic torques contribute to steering in flying insects[J]. Bioinspiration & Biomimetics, 2017, 12(4): 046001.
- [14] SUN M, XIONG Y. Dynamic flight stability of a hovering bumblebee[J]. Journal of Experimental Biology, 2005, 208(3): 447-59.
- [15] SUN M, WANG J K, XIONG Y. Dynamic flight stability of hovering insects[J]. Acta Mechanica Sinica, 2007, 23(3): 231-246.
- [16] ZHANG Y, SUN M. Dynamic flight stability of a hovering model insect; Lateral motion[J]. Acta Mechanica Sinica, 2010, 26(2): 175-190.
- [17] MOU X L, SUN M. Dynamic flight stability of a model hoverfly in inclined-stroke-plane hovering[J]. Journal of Bionic Engineering, 2012, 9(3): 294-303.
- [18] XU N, SUN M. Lateral dynamic flight stability of a model bumblebee in hovering and forward flight[J]. Journal of Theoretical Biology, 2013, 319: 102-115.
- [19] XU N, SUN M. Lateral dynamic flight stability of a model hoverfly in normal and inclined stroke-plane hovering[J]. Bioinspiration & Biomimetics, 2014, 9(3): 036019.
- [20] ETKIN B, REID L D. Dynamics of flight: Stability and control[M]. New York: John Wiley and Sons, Inc, 1996.
- [21] TAYLOR G K, THOMAS A L R. Animal flight dynamics. II. Longitudinal stability in flapping flight[J]. Journal of Theoretical Biology, 2002, 214(3): 351-370.
- [22] ROGERS S E, KWAK D, KIRIS C. Numerical solution of the incompressible Navier-Stokes equations for steady-state and time-dependent problems[C]//27th Aerospace Sciences Meeting, 1989: 0463.
- [23] SUN M, TANG J. Lift and power requirements of hovering flight in *Drosophila virilis*[J]. Journal of Experimental Biology, 2002, 205(16): 2413-2427.
- [24] ZHU H J, MENG X G, SUN M. Forward flight stability in a drone-fly[J]. Scientific Reports, 2020, 10(1): 12.
- [25] CHENG B, DENG X. Translational and rotational damping of flapping flight and its dynamics and stability at hovering[J]. IEEE Transactions on Robotics, 2011, 27(5): 849-864.

Effects of stroke deviation on flight stability of true hoverfly

XU Na¹, ZHOU Shuaizhi¹, MOU Xiaolei^{2,*}

(1. School of Electromechanical and Automotive Engineering, Yantai University, Yantai 264005, China;

2. School of Civil Engineering, Yantai University, Yantai 264005, China)

Abstract: The aerodynamic performance of insect can be affected by wing stroke deviation, which may also influence the flight stability. The longitudinal and lateral non-dimensional stability derivatives of hoverfly with stroke deviation are obtained by solving Navier-Stokes equation, and then the natural modes of motion analysis method is used to analyze the flight stability. The results show that, when stroke deviation exists, the stability derivative of the rolling moment induced by the side motion velocity decreases significantly, while the other derivatives have little difference by comparing the situation with no stroke deviation. The reason why the existence of stroke deviation causes the decrease of the derivative is that the positive rolling moment of the left and right wings declines obviously under lateral wind circumstance, while the negative rolling moment caused by the lateral force increases slightly, which makes the total negative rolling moment increase. However, the decrease of the derivative of the rolling moment caused by the lateral flow does not alter the flight stability of hoverfly, and the longitudinal and lateral characteristic modes of motion are still the same as those without stroke deviation.

Keywords: stroke deviation; true hoverfly; flight stability; Navier-Stokes equation; natural mode of motion

Received: 2020-12-28; **Accepted:** 2021-01-30; **Published online:** 2021-02-07 14:45

URL: kns.cnki.net/kcms/detail/11.2625.V.20210207.1404.001.html

Foundation items: National Natural Science Foundation of China (11802262,11502228)

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

http://bhxb.buaa.edu.cn jbuua@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2020.0186

弹性欧拉梁大变形分析的椭圆积分统一形式



安朝, 谢长川*, 杨超

(北京航空航天大学 航空科学与工程学院, 北京 100083)

摘 要: 一端固支、一端受集中载荷的欧拉梁受载问题是一种基础的力学模型, 具有重要的理论研究意义。针对传统的线性求解方法在大变形分析中不适用及无法计算中心定向受压杆在载荷系数超过临界值后屈曲变形的问题, 提出一种非线性精确解来进行受集中载荷梁的大变形计算方法。通过椭圆积分形式来推导受集中载荷梁的变形表达, 考虑在固支梁自由端加载任意角度下的定向载荷及随动载荷, 给出形式统一的梁大变形方程, 求解一定载荷因子系数及载荷方位角组合下的变形结果; 同时利用此形式对定向受压杆的平衡分支解问题进行了分析。所提方法计算结果准确, 可以应用于弹性欧拉梁受定向及随动载荷的大变形分析。

关键词: 大变形梁; 椭圆积分; 随动载荷; 平衡分支解; 压杆稳定性

中图分类号: O341

文献标志码: A

文章编号: 1001-5965(2021)07-1379-08

大变形欧拉梁作为一种基础的力学模型, 其理论研究对于分析结构几何非线性特性具有重要意义。随着工程技术及材料科学的发展, 欧拉梁大变形分析需求也日益迫切, 一些典型的工程结构如大展弦比柔性机翼、细长导弹等都可以简化为该模型进行分析研究。Dado 与 Al-Sadder^[1]总结了梁大变形问题的主要分析方法, 包括椭圆积分分析方法^[2-3]、打靶法^[4-5]、有限差分方法^[6]及有限元方法^[7-8]。其中, 椭圆积分分析方法可以给出半解析形式, 具备结果准确度高、形式简洁及计算效率高等优势。

考虑几何非线性因素, Bisshop 和 Drucker^[2]最早给出了受定向载荷梁弯曲大变形的椭圆积分计算形式。Midha^[9]、Kimball^[10]、Lau^[11]等将这一形式扩展到受力矩载荷分析及受混合载荷分析。在处理变几何参数梁问题时, 这一算法需要复杂的迭代流程。Shinohara 和 Hara^[12]将这一算法扩展至曲梁变形分析。王恩惠^[13]和李银山等^[14]利用椭圆积分方法研究压杆稳定性及变形分析问

题, 能够准确计算失稳后的屈曲力学特征。当考虑弹性梁受随动载荷时, 复杂性提高, 利用椭圆积分方法分析的研究不多^[15]。Srpacic^[6]和 Argyris^[8]等分别利用有限差分方法及有限元方法给出梁端部受随动载荷的变形解。Rao 等^[4,16]利用改进打靶法分析了这一问题, 提高了计算效率。Navae^[17]及 Shavartsman^[18]等进一步改进分析流程, 通过改变积分顺序将边值问题转化为初值问题, 进行 Runge-Kutta 数值积分求解, 降低分析难度。多数现有研究局限于载荷与梁轴线垂直的情况。

本文利用椭圆积分, 给出考虑弹性欧拉梁受端部定向载荷, 即所谓“死载”(Dead Force)及随动载荷(Follower Force)下大变形的分析求解方法, 分析定向受压杆平衡分支解问题。利用不同的变量变换关系使这两类问题控制方程具有统一形式。不限制载荷与梁轴线的角度关系, 提供适用程度更高的梁大变形分析方法。本文计算结果与非线性有限元方法计算结果进行了对比。

收稿日期: 2020-05-14; 录用日期: 2020-06-30; 网络出版时间: 2020-07-15 09:12

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20200715.0904.001.html

* 通信作者. E-mail: xiechangc@buaa.edu.cn

引用格式: 安朝, 谢长川, 杨超. 弹性欧拉梁大变形分析的椭圆积分统一形式[J]. 北京航空航天大学学报, 2021, 47(7): 1379-1386. AN C, XIE C C, YANG C. Unified form of large deflection analysis of elastic Euler beam based on elliptic integral[J]. Journal of Beijing University of Aeronautics and Astronautics, 2021, 47(7): 1379-1386 (in Chinese).

1 梁大变形方程

考虑根部固支悬臂梁,长度 L , 弯曲刚度 $EI(s)$, 端部受载 P 。悬臂梁采用各向同性的线弹性材料组成。梁单元平衡方程满足:

$$\kappa = \frac{d\varphi}{ds} = \frac{M(s)}{EI} \quad (1)$$

式中: κ 为梁上某点变形曲率; φ 为梁变形角; $M(s)$ 为梁弯矩内力; EI 为梁弯曲刚度; s 为弧长坐标。

1.1 端部受定向载荷

考虑在悬臂梁端部给定一方向不变、大小不变的载荷。受载下梁几何关系如图 1 所示。图中: α 为载荷方位角, φ 为梁变形角。

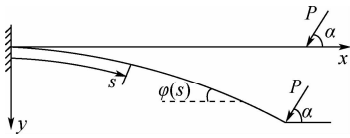


图 1 定向载荷作用下梁几何关系

Fig. 1 Geometric relationship of beam with dead force

将式(1)微分可得控制方程:

$$EI \frac{d^2\varphi}{ds^2} + P \sin(\alpha + \varphi) = 0 \quad (2)$$

边界条件为

$$\begin{cases} \varphi(0) = 0 \\ \frac{d\varphi}{ds}(L) = 0 \end{cases} \quad (3)$$

给出变换关系:

$$z(s) = \alpha + \varphi(s) \quad (4)$$

代入式(2)、式(3)可得新的控制方程及边界条件为

$$EI \frac{d^2z}{ds^2} + P \sin z = 0 \quad (5)$$

$$\begin{cases} z(0) = \alpha \\ \frac{dz}{ds}(L) = 0 \end{cases} \quad (6)$$

积分式(5), 可得

$$\frac{1}{2} \left[\left(\frac{dz}{ds} \right)^2 \right]_s - \frac{1}{2} \left[\left(\frac{dz}{ds} \right)^2 \right]_0 = \frac{P}{EI} [\cos z|_s - \cos z|_0] \quad (7)$$

代入边界条件整理可得

$$\left(\frac{dz}{ds} \right)^2 \Big|_s = \frac{4P}{EI} \left[\sin^2 \left(\frac{z_L}{2} \right) - \sin^2 \left(\frac{z}{2} \right) \right] \quad (8)$$

式中: $z_L = \alpha + \varphi_L$, $\varphi_L = \varphi(L)$ 。令 $k^2 = \sin^2 \left(\frac{z_L}{2} \right)$, 整理可得

$$ds = \frac{dz}{2\sqrt{\frac{P}{EI} \left(k^2 - \sin^2 \left(\frac{z}{2} \right) \right)}} \quad (9)$$

引入变量 θ , 令 $\sin \left(\frac{z}{2} \right) = k \sin \theta = \sin \left(\frac{z_L}{2} \right) \sin \theta$:

$$ds = \frac{d\theta}{\sqrt{\frac{P}{EI} (1 - k^2 \sin^2 \theta)}} \quad (10)$$

积分式(10)可得

$$s = \int_0^s ds = \frac{1}{\sqrt{P/EI}} \int_{\theta_1}^{\theta_2} \frac{d\theta}{\sqrt{1 - k^2 \sin^2 \theta}} \quad (11)$$

$$\begin{cases} z(0) = \alpha, \sin \theta_1 = \frac{1}{k} \sin \left(\frac{\alpha}{2} \right) \\ z(s) = z, \sin \theta_2 = \frac{1}{k} \sin \left(\frac{z}{2} \right) \end{cases} \quad (12)$$

令相对弧长坐标 $\bar{s} = \frac{s}{L}$, 载荷因子 $\beta = \frac{PL^2}{2EI}$, 即有变形与载荷关系:

$$\bar{s} = \frac{1}{\sqrt{2\beta}} [F(k, \theta_2) - F(k, \theta_1)] \quad (13)$$

$$\begin{cases} F(k, \theta) = \int_0^\theta \frac{d\theta}{\sqrt{1 - k^2 \sin^2 \theta}} \\ k = \sin \left(\frac{z_L}{2} \right) = \sin \left(\frac{\alpha + \varphi_L}{2} \right) \\ \sin \theta_1 = \frac{1}{k} \sin \left(\frac{\alpha}{2} \right) \\ \sin \theta_2 = \frac{1}{k} \sin \left(\frac{z}{2} \right) \end{cases} \quad (14)$$

$F(k, \theta)$ 为第一类椭圆积分^[19]。当 $\bar{s} = 1$ 时, 梁端部变形与载荷关系有

$$\sqrt{2\beta} = \left[F \left(k, \frac{\pi}{2} \right) - F(k, \theta_1) \right] \quad (15)$$

1.2 端部受随动载荷

考虑在悬臂梁端部给定一方向与梁轴向夹角不变、大小不变的载荷, 即随动载荷 (Follower Force)。受载下梁几何关系如图 2 所示。

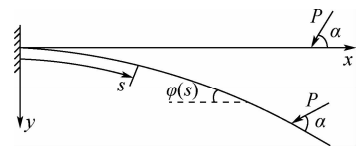


图 2 随动载荷作用下梁几何关系

Fig. 2 Geometric relationship of beam with follower force

将式(1)微分可得控制方程:

$$EI \frac{d^2\varphi}{ds^2} + P \sin(\alpha + \varphi - \varphi_L) = 0 \quad (16)$$

边界条件为

$$\begin{cases} \varphi(0) = 0 \\ \frac{d\varphi}{ds}(L) = 0 \end{cases}$$

给出变换关系:

$$z(s) = \alpha + \varphi(s) - \varphi_L \quad (17)$$

代入式(16)、式(3)可得新的控制方程为

$$EI \frac{d^2 z}{ds^2} + P \sin(z) = 0$$

边界条件为

$$\begin{cases} z(0) = \alpha - \varphi_L \\ \frac{dz}{ds}(L) = 0 \end{cases} \quad (18)$$

积分式(5),可得

$$\frac{1}{2} \left[\left(\frac{dz}{ds} \right)^2 \right]_s - \frac{1}{2} \left[\left(\frac{dz}{ds} \right)^2 \right]_0 = \frac{P}{EI} [\cos z|_s - \cos z|_0]$$

代入边界条件整理可得

$$\left(\frac{dz}{ds} \right)^2 \Big|_s = \frac{4P}{EI} \left[\sin^2 \left(\frac{z_L}{2} \right) - \sin^2 \left(\frac{z}{2} \right) \right]$$

式中: $z_L = \alpha$ 。令 $k^2 = \sin^2 \left(\frac{z_L}{2} \right)$, 整理可得

$$ds = \frac{dz}{2\sqrt{\frac{P}{EI} \left(k^2 - \sin^2 \left(\frac{z}{2} \right) \right)}}$$

引入变量 θ , 令 $\sin \left(\frac{z}{2} \right) = k \sin \theta = \sin \left(\frac{z_L}{2} \right) \sin \theta$:

$$ds = \frac{d\theta}{\sqrt{\frac{P}{EI} (1 - k^2 \sin^2 \theta)}}$$

积分式(10)可得

$$s = \int_0^s ds = \frac{1}{\sqrt{\frac{P}{EI}}} \int_{\theta_1}^{\theta_2} \frac{d\theta}{\sqrt{1 - k^2 \sin^2 \theta}}$$

$$\begin{cases} z(0) = \alpha - \varphi_L, \sin \theta_1 = \frac{1}{k} \sin \left(\frac{\alpha - \varphi_L}{2} \right) \\ z(s) = z, \sin \theta_2 = \frac{1}{k} \sin \left(\frac{z}{2} \right) \end{cases} \quad (19)$$

即有变形与载荷关系:

$$\bar{s} = \frac{1}{\sqrt{2\beta}} [F(k, \theta_2) - F(k, \theta_1)]$$

$$\begin{cases} F(k, \theta) = \int_0^\theta \frac{d\theta}{\sqrt{1 - k^2 \sin^2 \theta}} \\ k = \sin \left(\frac{z_L}{2} \right) = \sin \left(\frac{\alpha}{2} \right) \\ \sin \theta_1 = \frac{1}{k} \sin \left(\frac{\alpha - \varphi_L}{2} \right) \\ \sin \theta_2 = \frac{1}{k} \sin \left(\frac{z}{2} \right) \end{cases} \quad (20)$$

同样地, 梁端部变形与载荷关系有

$$\sqrt{2\beta} = \left[F \left(k, \frac{\pi}{2} \right) - F(k, \theta_1) \right]$$

对比可见, 定向载荷与随动载荷加载后的梁控制方程形式一致, 其差别表现在求解方程的边

界条件差异。

2 定向受压杆问题的平衡分支解

定向受压杆的屈曲问题是梁非线性变形研究中的经典问题, 超临界状态下对应一组载荷存在多种平衡分支现象。在端部受定向载荷解中, 给定载荷方位角 $\alpha = 0^\circ$, 变换关系 $z(s) = \varphi(s)$, 变形与载荷关系为

$$\begin{aligned} \bar{s} &= \frac{1}{\sqrt{2\beta}} [F(k, \theta_2) - F(k, \theta_1)] \\ \begin{cases} F(k, \theta) = \int_0^\theta \frac{d\theta}{\sqrt{1 - k^2 \sin^2 \theta}} \\ k = \sin \left(\frac{z_L}{2} \right) = \sin \left(\frac{\varphi_L}{2} \right) \\ \sin \theta_1 = 0 \\ \sin \theta_2 = \frac{1}{k} \sin \left(\frac{z}{2} \right) \end{cases} \end{aligned} \quad (21)$$

$$\sqrt{2\beta} = \left[F \left(k, \frac{\pi}{2} \right) - F(k, \theta_1) \right]$$

式中: $\theta_1 = m\pi, m = 0, -1, -2, \dots$, 方程均成立, 即存在多平衡解分支。需要说明的是, 由边界条件及变换关系可知, $\theta_1 < \theta_2 \leq \frac{\pi}{2}$, 因此 $m \leq 0$ 。

当 $m = 0$ 时, 称为平衡 I 态, 其临界载荷因子 $\beta_{cr-1} = \min \left(\frac{1}{2} \left[F \left(k, \frac{\pi}{2} \right) - F(k, 0) \right]^2 \right)$ 。载荷超过该临界载荷时, 平衡解为平衡 I 态解曲线, 由式(11)及式(15)可知其解曲线由积分区间 $0 \leq \theta \leq \frac{\pi}{2}$ 单独确定。

当 $m = -1$ 时, 称为平衡 II 态, 其临界载荷因子 $\beta_{cr-2} = \min \left(\frac{1}{2} \left[F \left(k, \frac{\pi}{2} \right) - F(k, -\pi) \right]^2 \right)$ 。载荷超过该临界载荷时, 平衡解出现新的分支解即平衡 II 态解曲线, 由式(11)、式(15)及引入变量变换关系 $\sin \left(\frac{z}{2} \right) = k \sin \theta$ 可知, 解曲线由积分区间 $-\pi \leq \theta \leq \frac{\pi}{2}$ 确定。

当 $m = -2$ 时, 称为平衡 III 态, 其临界载荷因子 $\beta_{cr-3} = \min \left(\frac{1}{2} \left[F \left(k, \frac{\pi}{2} \right) - F(k, -2\pi) \right]^2 \right)$ 。载荷超过该临界载荷时, 平衡解在平衡 I 态/II 态解曲线基础上出现新的分支解即平衡 III 态解曲线, 由引入变换关系可知, 解曲线由积分区间 $-2\pi \leq \theta \leq \frac{\pi}{2}$ 确定。以此类推, 当载荷因子 β 介于平衡 n 态及平衡 $n+1$ 态的临界载荷因子之间时, 梁变形挠曲线共有 n 种形式, 即 n 个平衡分支解。

值得说明的是,当梁端部随动载荷受压时,由式(20)可知 $k=0$,解曲线为0,即随动载荷压杆不存在静不稳定解。

3 计算流程

梁端部受定向载荷时,求解式(13)~式(15)。由于椭圆积分参数 k 与梁末端变形角 φ_L 相关,无法直接给出。首先在式(15)中可以得到不同 k 值对应的载荷因子 β 。找到与工况指定的载荷因子对应的 k 即可确定椭圆积分参数及梁末端变形角。由式(13)即可得到弧长坐标 \bar{s} 与变量 z 的关系。根据变换关系式(4)可恢复梁实际变形角。挠曲线结果可通过沿弧长坐标积分计算求解:

$$\bar{x} = \int_0^{\bar{s}} \sin \varphi ds \tag{22}$$

$$\bar{y} = \int_0^{\bar{s}} \cos \varphi ds \tag{23}$$

当 $\bar{s} = 1$ 即 $s = L$ 时,梁端部 y 方向变形量 $\delta = \bar{y}(1)$, x 方向变形量 $\Delta = L - \bar{x}(1)$,梁挠曲线几何关系如图3所示。计算流程如图4所示。

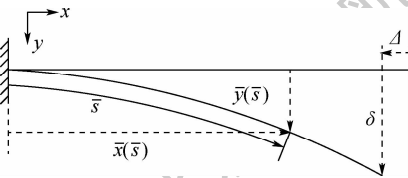


图3 梁变形挠曲线

Fig. 3 Deflection curves of beam

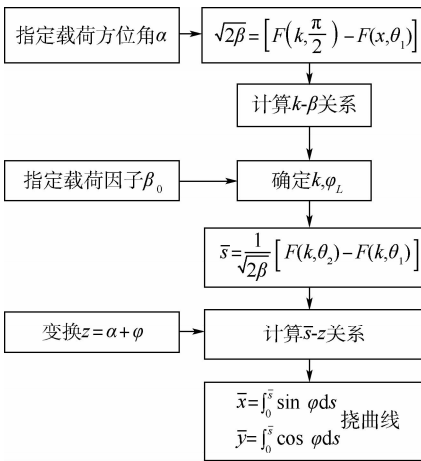


图4 梁端部受定向载荷计算流程

Fig. 4 Flowchart of calculation with dead force at the end of beam

梁端部受随动载荷时,求解式(13)、式(15)、式(20)。椭圆积分参数仅与载荷方位角 α 相关,可以直接给出。在式(15)中可以得到不同 φ_L 值对应的载荷因子。找到与工况指定的载荷因子对应的 φ_L 即可确定梁末端变形角及中间变量 θ_1 。由式(13)即可得到弧长坐标与变量的关系。根

据变换关系式(17)可恢复梁实际变形角。挠曲线结果同样可通过沿弧长坐标积分式(13)、式(21)计算求解。计算流程如图5所示。

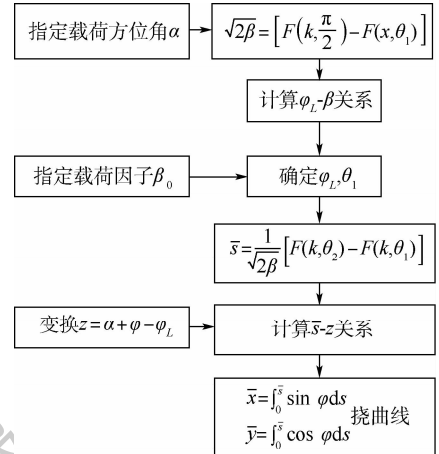


图5 梁端部受随动载荷计算流程

Fig. 5 Flowchart of calculation with follower force at the end of beam

4 数值算例及结果

4.1 大变形梁模型

给定一柔性梁模型,梁截面圆形,不考虑沿轴向伸长/缩短,模型参数如表1所示。

表1 梁模型参数

Table 1 Parameters of beam model

参数	数值
长度 L/m	1
截面半径 r/m	0.001 7
弹性模量 $E/(N \cdot m^{-2})$	7×10^{10}
泊松比	0.29

4.2 梁变形计算结果

给定不同的载荷方位角 α 后,指定不同的载荷因子 $\beta = PL^2/(2EI)$ 后,计算梁挠曲线变形。为验证计算结果的准确性,变形结果与非线性有限元计算结果进行对比。图6给出了梁端部受定向载荷时不同 α/β 组合下的梁挠曲线计算结果对比。随着载荷因子增大,梁变形逐渐增大,本文方法计算结果与非线性有限元方法计算结果保持高度一致,且计算精度没有随载荷因子的增大及载荷方位角的改变而降低。

图7给出了梁端部受随动载荷时不同 α/β 组合下的梁挠曲线计算结果对比。同样的,随着载荷因子增大,梁变形逐渐增大。计算结果与非线性有限元方法计算结果保持高度一致。非线性有限元计算通过 MSC. Nastran 软件实现,计算精度没有随载荷因子的增大及载荷方位角的改变而降低,计算方式适用性高。

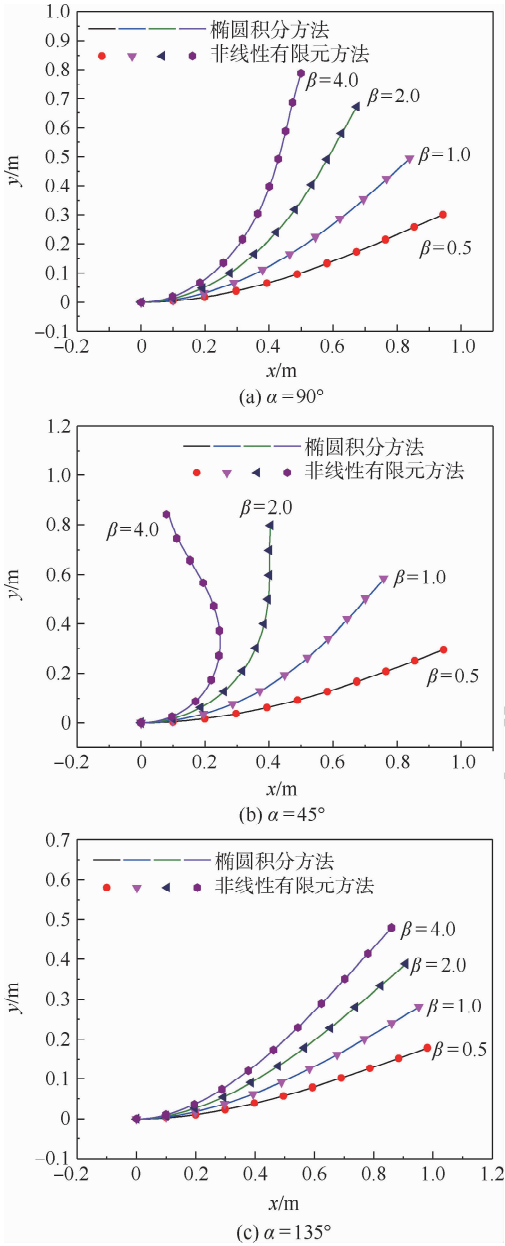


图6 梁端部受定向载荷挠曲线

Fig. 6 Deflection curves with dead force at the end of beam

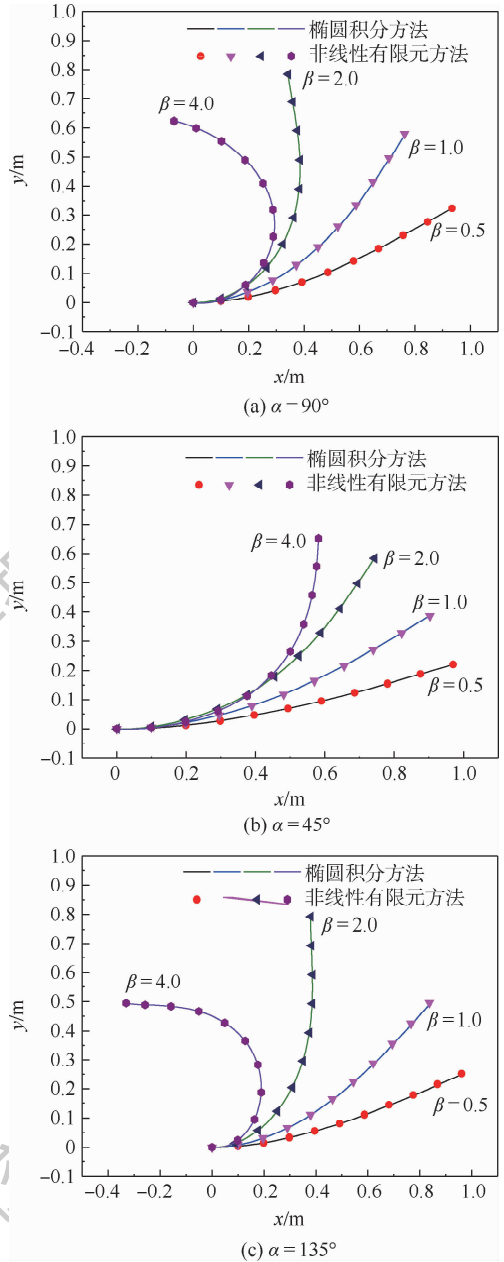


图7 梁端部受随动载荷挠曲线

Fig. 7 Deflection curves with follower force at the end of beam

图8给出了梁端部变形在载荷方位角 $\alpha = 90^\circ$ 时受载荷因子影响的变化曲线,同时给出非线性有限元方法计算结果及文献[2]计算结果,三者保持高度一致。与相关文献结果^[2]的对比进一步验证了本文方法对于梁变形计算的准确性。表2和表3给出了定向载荷及随动载荷下,不同 α/β 组合下梁端部变形 δ 的具体数值计算结果对比。可以发现,其与非线性有限元方法计算结果偏差值很小,已有状态下,最大偏差值不超过0.75%。

图9展示了受定向载荷时,梁端部变形在不同载荷方位角下,受载荷因子影响的变化曲线。梁端挠度随载荷增大呈现逐渐增大趋势,最终趋于平缓。图10展示了受随动载荷时,梁端部变形在不同载荷方位角下,受载荷因子影响的变化曲

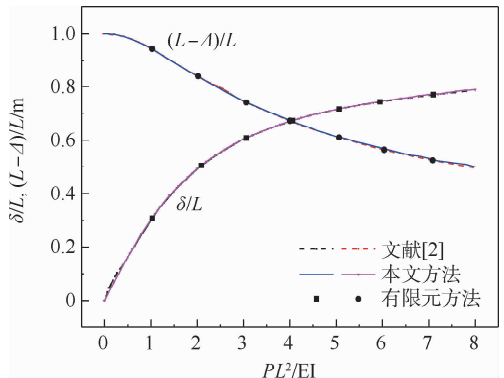


图8 定向载荷下梁变形对比($\alpha = 90^\circ$)

Fig. 8 Contrast of beam deflection with dead force ($\alpha = 90^\circ$)
线。与受定向载荷情况不同,梁端部挠度随载荷

表2 定向载荷下梁端部变形 δ 对比
Table 2 Contrast of beam end deflection
 δ with dead force

载荷方位角/ $^\circ$	载荷因子	椭圆积分解/m	有限元解/m	偏差/%
90	0.5	0.301 63	0.301 81	0.06
90	1.0	0.493 5	0.494 96	0.3
90	2.0	0.670 65	0.671 49	0.12
90	4.0	0.787 85	0.788 11	0.03
45	0.5	0.295 45	0.295 58	0.04
45	1.0	0.582 65	0.584 92	0.39
45	2.0	0.797 67	0.798 19	0.65
45	4.0	0.847 02	0.843 12	0.46
135	0.5	0.178 43	0.178 44	0.01
135	1.0	0.280 25	0.281 05	0.28
135	2.0	0.388 11	0.388 32	0.05
135	4.0	0.480 46	0.480 58	0.02

表3 随动载荷下梁端部变形 δ 对比
Table 3 Contrast of beam end deflection
 δ with follower force

载荷方位角/ $^\circ$	载荷因子	椭圆积分解/m	有限元解/m	偏差/%
90	0.5	0.324 8	0.324 64	0.05
90	1.0	0.576 22	0.579 23	0.52
90	2.0	0.785 65	0.784 63	0.13
90	4.0	0.623 69	0.623 01	0.11
45	0.5	0.221 02	0.221 0	0.01
45	1.0	0.386 93	0.385 97	0.25
45	2.0	0.584 41	0.583 68	0.12
45	4.0	0.652 39	0.651 96	0.15
135	0.5	0.252 09	0.252 06	0.01
135	1.0	0.492 37	0.496 02	0.74
135	2.0	0.793 19	0.792 17	0.13
135	4.0	0.495 42	0.495 06	0.07

的增大呈现先增大后减小的趋势。这与两类载荷加载中梁局部剪力变化趋势不同密切相关。

4.3 定向受压杆平衡分支解计算结果

给定载荷方位角 $\alpha = 0^\circ$ 计算该算例模型前3项临界载荷因子及平衡态解曲线。由椭圆积分性质可得 $\min(F(k)) = \frac{\pi}{2}$, 有

$$\beta_{cr-1} = \min\left(\frac{1}{2}[F(k)]^2\right) = \frac{\pi^2}{8} = 1.233 7 \quad (24)$$

结合载荷因子定义,临界载荷为

$$P_{cr-1} = \frac{EI\pi^2}{4L^2} \quad (25)$$

这与材料力学中心压杆失稳欧拉解一致。类似的,通过计算给定参数椭圆积分极值,可以得到第2及第3项临界载荷因子:

$$\beta_{cr-2} = \min\left(\frac{1}{2}[F(k) - F(k, -\pi)]^2\right) = 11.1 \quad (26)$$

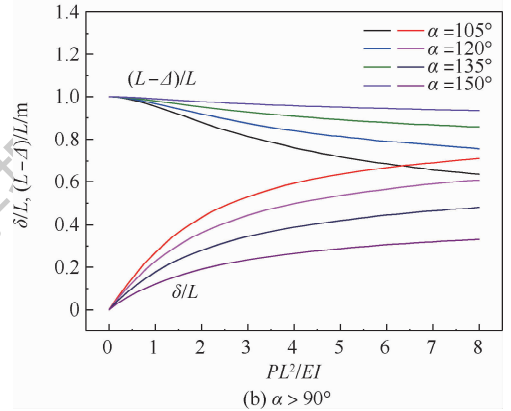
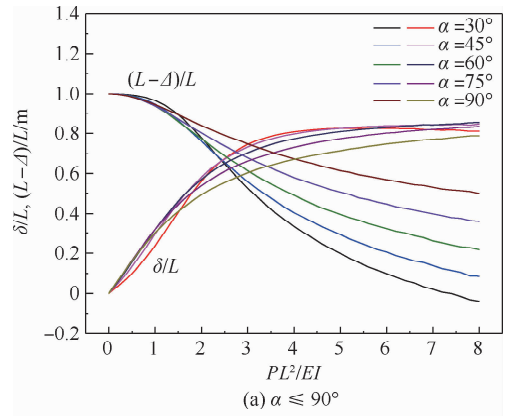


图9 定向载荷下梁变形

Fig. 9 Deflection of beam with dead force

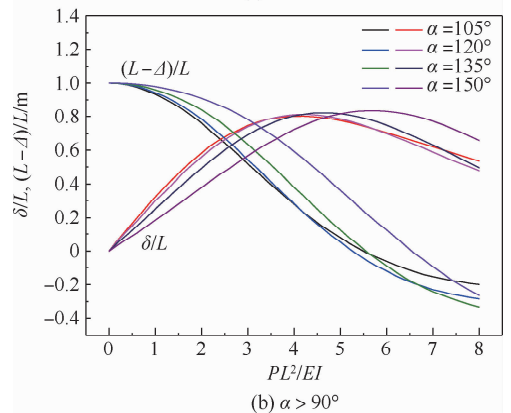
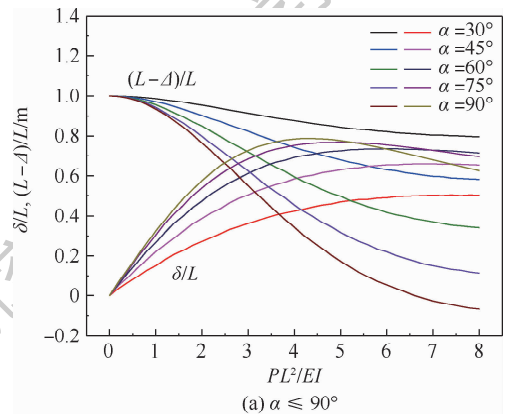


图10 随动载荷下梁变形

Fig. 10 Deflection of beam with follower force

$$\beta_{cr-3} = \min\left(\frac{1}{2}[F(k) - F(k, -2\pi)]^2\right) = 30.84 \quad (27)$$

图 11 ~ 图 13 给出了在不同载荷因子下,前 3 个平衡态解曲线计算结果,可以发现越靠后出现的平衡态其梁曲线越复杂。结合前述理论分析,平衡 I 态解曲线由 $0 \leq \theta \leq \pi/2$ 积分区间确定,其解曲线包括 1 个驻点、0 个反弯点;平衡 II 态解曲线由 $-\pi \leq \theta \leq \pi/2$ 积分区间确定,其解曲线包括 2 个驻点、1 个反弯点;平衡 III 态解曲线由 $-2\pi \leq \theta \leq \pi/2$ 积分区间确定,其解曲线包括 3 个驻点、2 个反弯点;对于平衡 n 态,其解曲线将由 $-(n-1)\pi \leq \theta \leq \pi/2$ 积分区间确定,其解曲线包括 n 个驻点、 $n-1$ 个反弯点。本文方法适用于计算定向压杆各级平衡态临界载荷及给定载荷下的解曲线。

5 结 论

1) 利用椭圆积分推导了悬臂梁端部受集中定向载荷及随动载荷的大变形分析方程,通过变量代换将两者形式进行了统一,使两者拥有一致的控制方程,仅在中间变量计算中存在不同。

2) 计算了给定载荷因子及载荷方位角组合下的梁大变形,展示了大变形的变化规律,并将结果与非线性有限元分析结果进行对比,两者一致性很高。当载荷逐渐增大时,非线性有限元方法收敛性逐渐变差,而椭圆积分不会出现该问题。

3) 对定向受压杆平衡分支解问题进行了分析计算,给出了各级平衡态临界载荷及解曲线的计算方法,并简要分析了各级平衡态解曲线性质。后续可对不同载荷方位角下的平衡分支解问题进行深入研究。

参考文献 (References)

[1] DADO M, AL-SADDER S. A new technique for large deflection analysis of non-prismatic cantilever beams [J]. Mechanics Research Communications, 2005, 32 (6): 692-703.

[2] BISSHOP K E, DRUCKER D C. Large deflection cantilever beams [J]. Quarterly of Applied Mathematics, 1945, 3 (3): 272-275.

[3] TIMOSHENKO S P, GERE J M. Theory of elastic stability [M]. New York: McGraw-Hill, 1963: 125-127.

[4] RAO B N, RAO G V. Applicability of static and dynamic criterion for the stability of a cantilever column under a tip-concentrated subtangential follower force [J]. Journal of Sound and Vibration, 1987, 120 (1): 197-200.

[5] WANG C M, KITIPORNCHAI S. Shooting optimization technique for large deflection analysis of structural members [J]. Engineering Structure, 1992, 14 (4): 231-240.

[6] SRPCIC S, SAJE M. Large deflections of thin curved plane beam of constant initial curvature [J]. International Journal of Mechanical Science, 1986, 28 (5): 275-287.

[7] WANG G, SHAHINPOOR M. Design prototyping and computer simulations of a novel large bending actuator made with a shape memory alloy contractile wire [J]. Smart Materials Structure, 1996, 6 (2): 214-221.

[8] ARGYRIS J H, SYMEONIDS S P. Non-linear finite element analysis of elastic systems under nonconservative loading-natural formulation. Part 1. Quasi-static problems [J]. Compute Methods of Applied Mechanical Engineering, 1981, 26 (1): 75-123.

[9] HOWELL L L, MIDHA A. Parametric deflection approximations for end-loaded large deflection beams in compliant mechanic [J]. Journal of Mechanical Design, 1995, 117 (1): 156-165.

[10] KIMBALL C, TSAI L W. Modeling of flexural beams subjected to arbitrary end loads [J]. Journal of Mechanical Design, 2002,

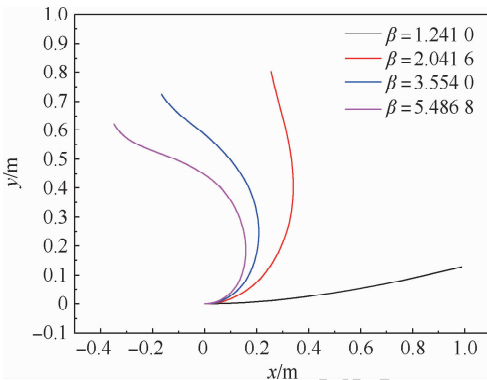


图 11 平衡 I 态解曲线

Fig. 11 Solution curves of balance branch I

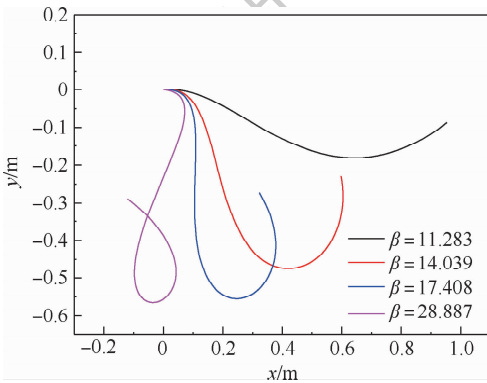


图 12 平衡 II 态解曲线

Fig. 12 Solution curves of balance branch II

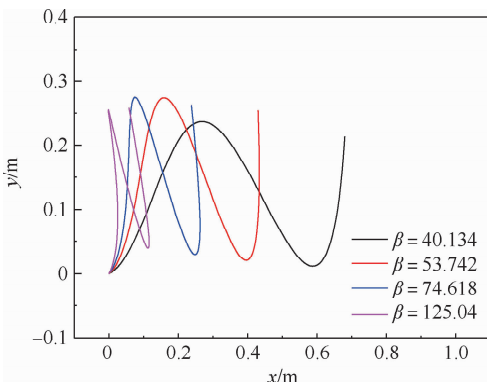


图 13 平衡 III 态解曲线

Fig. 13 Solution curves of balance branch III

- 124(2):223-234.
- [11] LAU J H. Closed-form solutions for the large deflections of curved optical glass fibers under combined loads[J]. *Journal of Electron Package*, 1993, 115(3):337-339.
- [12] SHINOHARA A, HARA M. Large deflection of a circular c-shaped spring[J]. *International Journal of Mechanical Science*, 1979, 21(3):179-186.
- [13] 王恩惠. 弹性悬臂杆的平衡分支[J]. *固体力学学报*, 1983, 4:143-151.
WANG E H. Balance branch of elastic cantilever rod[J]. *Acta Mechanica Solida Sinica*, 1983, 4:143-151 (in Chinese).
- [14] 李银山, 刘波, 潘文波, 等. 弹性压杆的大变形分析[J]. *河北工业大学学报*, 2011, 40(5):31-35.
LI Y S, LIU B, PAN W B, et al. Analysis of large deflection of flexible compression bars[J]. *Journal of Hebei University of Technology*, 2011, 40(5):31-35 (in Chinese).
- [15] FRISCH-FAY R. *Flexible bars* [M]. London: Butterworths, 1962:119-151.
- [16] RAO B N, RAO G V. Applicability of static or dynamic criterion for the stability of a non-uniform cantilever column subjected to a tip-concentrated subtangential follower force[J]. *Journal of Sound and Vibration*, 1988, 122(1):188-191.
- [17] NAVAEE S, ELLING R. Equilibrium configurations of cantilever beams subjected to inclined end loads[J]. *Journal of Applied Mechanics*, 1992, 59(3):572-579.
- [18] SHAVARTSMAN B S. Large deflections of cantilever beam subjected to a follower force[J]. *Journal of Sound and Vibration*, 2007, 304(3-5):969-973.
- [19] 王竹溪, 郭敦仁. *特殊函数概论* [M]. 北京: 北京大学出版社, 2000:520-556.
WANG Z X, GUO D R. *Special functions* [M]. Beijing: Peking University Press, 2000:520-556 (in Chinese).

Unified form of large deflection analysis of elastic Euler beam based on elliptic integral

AN Chao, XIE Changchuan*, YANG Chao

(School of Aeronautic Science and Engineering, Beihang University, Beijing 100083, China)

Abstract: The loading problem of Euler beam with one end clamped and one end having concentrated load is a basic mechanical model, which has important theoretical significance. In order to solve the problem that the traditional linear method is not applicable to large deflection analysis and cannot calculate the post buckling deflection of the central direction compressed column when the load coefficient exceeds the critical value, a nonlinear exact solution is proposed to calculate the large deformation of the beam under concentrated load. In this paper, the elliptic integral form is used to derive the deflection expression of the beam under concentrated load. Considering the dead force and follower force at any angle on the free end of the fixed-supported beam, a large deflection equation of the beam in a unified form is given. The deflection results under the combination of load factor and load angle are calculated. At the same time, the balance branch solution of the compressed column with dead force is analyzed using this form. The calculation results of this method are accurate and can be applied to the large deflection analysis of elastic beam under dead force and follower force.

Keywords: large deflection beam; elliptic integral; follower force; balance branch solution; stability of compressed column

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2020.0193



低信噪比下分组交织器识别

吴昭军^{1,2}, 刘凯^{1,*}, 钟兆根³, 但波⁴, 周磊²

(1. 海军航空大学 信息融合研究所, 烟台 264001; 2. 西南电子电信研究所, 成都 610041;
3. 海军航空大学 航空基础学院, 烟台 264001; 4. 海军航空大学 岸防兵学院, 烟台 264001)

摘 要: 针对现有的分组交织器识别算法计算复杂高且容错性差缺点, 从分组交织后的同步码分布规律出发, 提出了一种新的识别算法。首先, 利用数据矩阵统计特性, 给出了在任意矩阵列数下, 同步码和随机业务数据位置上的概率密度分布函数, 基于最小错误判决准则, 设定了同步码检测门限, 同时基于 3 倍标准差准则, 求解出稳健的交织周期识别门限; 其次, 分析了数据矩阵中每一行与每一列累积量之间的对应关系, 提出了一种快速交织周期遍历方法, 使得数据矩阵的构建次数大大减少; 最后, 总结了 4 个分组交织后同步码分布规律, 通过遍历同步码序列, 利用同步码之间的位置关系, 实现交织同步位置、分组交织列与交织行参数快速识别。仿真结果表明: 所提算法具有较强的低信噪比容错性, 在信噪比为 -6 dB 条件下, 参数识别率能够达到 98% 以上, 同时与现有的算法相比, 其性能提升 4~10 dB 且计算效率明显提高。

关键词: 认知无线电; 信道编码; 分组交织器; 同步码; 帧同步特征; 识别

中图分类号: V243.1; TN911.7

文献标志码: A **文章编号:** 1001-5965(2021)07-1387-12

在突发信道, 如天电噪声、宇宙强射线干扰下, 对均匀误码具有较强纠错能力的编码方案如 BCH 码、卷积码等往往会失效。为了克服这种缺陷, 将突发误码转化为均匀误码, 交织器通常被广泛应用于现代数字通信系统中。对于非合作通信方而言, 现有的信道编码盲识别算法, 如线性分组码^[1-2]、卷积码^[3-4]以及 Turbo 码^[5-6]识别等都是将交织关系作为前提条件, 因此在低信噪比下完成交织器的识别对于整个信道编码链路的分析具有非常重要的意义。

从现有的研究来看, 对于交织器的识别主要集中于分组交织器与卷积交织器。针对卷积交织器而言, Burel 和 Gautier^[7]最早针对该问题进行研究, 提出了一种基于矩阵分析的识别方法, 该方法将截获的序列排入分析矩阵中, 当分析矩阵的

列数等于交织周期的整数倍时, 矩阵将出现秩亏现象, 这种方法虽能够实现参数识别, 但仅能识别出交织周期与交织起点, 同时不具有容错性。Sicot 和 Houcke^[8]基于矩阵分析方法, 将误码对消元的影响考虑在内, 利用高斯消元后, 线性相关列与独立列中 0、1 概率分布不同, 从而判断矩阵是否出现秩亏, 该方法在文献[7]的基础上, 容错性能得到较大的提高, 但也仅仅能够识别出部分交织参数。为实现对卷积交织器的完整识别, Gan^[9]和 Xu^[10]等深入分析了卷积交织器的结构特点, 首先利用高斯消元法识别出交织周期, 然后通过三重遍历搜索方式识别出交织深度、交织宽度以及交织起点, 虽然该方法能够完整识别出卷积交织器, 但是计算复杂度较高, 同时容错性能较差。为了克服高斯消元的缺点, 解辉^[11]和于沛东^[12]等

收稿日期: 2020-05-18; 录用日期: 2020-08-07; 网络出版时间: 2020-09-15 16:08

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20200915.1401.005.html

基金项目: 国家自然科学基金(91538201); “泰山学者”建设工程专项(ts201511020)

* 通信作者. E-mail: wendao_2008@163.com

引用格式: 吴昭军, 刘凯, 钟兆根, 等. 低信噪比下分组交织器识别[J]. 北京航空航天大学学报, 2021, 47(7): 1387-1398.

WU Z J, LIU K, ZHONG Z G, et al. Recognition of packet interleaver at low SNR [J]. Journal of Beijing University of Aeronautics and Astronautics, 2021, 47(7): 1387-1398 (in Chinese).

利用数据帧中同步码经过卷积交织后位置分布规律,实现高误码下卷积交织器的快速识别,但是该方法仅仅针对卷积交织结构,并不适用于分组交织器的识别。针对分组交织器的识别,郑鹏鹏等^[13]利用分组码交织后码重分布不均衡的特点,实现交织周期的识别,该方法虽然具有较好的容错性,但是仅仅适用于低码率的分组码,同时不能有效完成交织起点、分组交织块行与列识别。陆凯等^[14]将卷积码下的分组交织器作为研究对象,首先利用高斯消元方法确定出交织周期,然后通过遍历分组交织器的行与列对原始数据进行采样,构造分析矩阵,完成参数识别,该方法虽然能够完整识别分组交织参数,但是容错性太差。陈泽亮等^[15]为克服文献[14]中方法容错性差的缺点,采用软解调序列进行校验关系验证,该方法虽然有效地提升了原有方法的容错性能,但是需要遍历校验关系,使得计算复杂度较高。由上述的分析可知,针对分组交织器参数识别算法还存在着计算复杂度高和容错性差的缺点。

基于此,本文算法从同步码序列经过分组交织后的位置规律出发,直接利用软解调序列构建分析矩阵进行参数识别。首先,基于同步码与随机业务数据位置在分析矩阵列上的统计特性不同,实现同步码的快速检测,从而完成交织周期的快速识别;然后,分析帧同步后,同步码位置的分布特性,利用该特性实现了分组交织起点、交织块行与列的快速解算。与传统算法相比,本文算法计算效率与低信噪比下的识别性能都有明显提高。

1 分组交织器原理

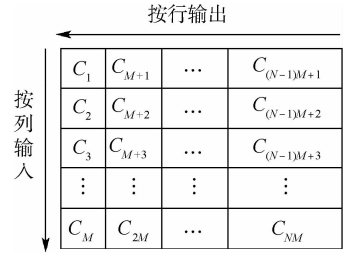
分组交织器先接收来自于信道编码器的输出码元,再对帧数据中码元位置进行置换,最终通过信道发送出去。一般情况下,分组交织的过程是:首先,将数据序列按照列填充到 $M \times N$ 的寄存器阵列中;然后,码元按行输出,送入调制器,在接收端,解交织的过程与此相反,此时码元按行填充到 $M \times N$ 的矩阵中;最后,按列输出。图 1 给出了分组交织与解交织的过程。

从图 1 分组交织原理可知,其分组交织的位置置换关系为

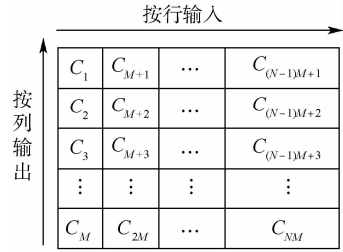
$$I(i) = \text{mod}((i-1)N, MN) + \left\lfloor \frac{i-1}{M} \right\rfloor + 1 \quad (1)$$

式中:函数 $\text{mod}(a, b)$ 表示对 a 的模 b 运算; $\lfloor \cdot \rfloor$ 表示向下取整; i 为交织前位置。

对于任意小于 N 个连续突发的错误,通过解



(a) 分组交织过程



(b) 解交织过程

图 1 分组交织与解交织过程

Fig. 1 Process of packet interleaving and deinterleaving

交织过程,错误能被转化为随机独立错误,且错误之间至少被 M 个码元隔开。由于收发两端都需要 $M \times N$ 个码元数据缓存,故其时间延时近似为 $2M \times N$ 个码元时间。

由于通信系统中数据总是按帧发送,为了使收发端数据延时不至于太长,分组交织规模一般要远远小于数据帧长,同时为了最大可能利用上分组交织器性能,数据帧长 L 一般取值为 MN 的整数倍,即

$$\text{mod}(L, MN) = 0 \quad (2)$$

在非协作通信中,非合作方需要识别的内容包括分组交织周期、交织起点以及分组交织器矩阵行 M 与列 N 。本文重点讨论的内容为在低信噪比下,利用截获的软解调序列以及经分组交织后的同步码位置规律完成上述参数识别。

2 基于同步码位置特征参数识别

2.1 分组交织周期识别

在实际通信系统中,为了实现数据帧同步,通常会在每一帧数据前添加固定的同步码序列,交织器不仅会将业务数据进行位置置换,对于同步码序列同样也会进行交织打乱,如 ATSC、DVB-S 以及 DVB-T 等协议^[11]。由于每帧数据中同步码序列是固定的,经过交织后,在交织周期中的位置分布一定是固定的。设每一帧中固定的同步码序列为 (a_1, a_2, \dots, a_s) , 经过交织置换后,其位置变为 $(a_{I(1)}, a_{I(2)}, \dots, a_{I(s)})$, 若将截获的软解调序列按行排列成分析矩阵,当分析矩阵的列数等于交

织周期整数倍时,则在矩阵 $I(1), I(2), \dots, I(s)$ 列上元素一阶累积量将会出现峰值;反之,当不为交织周期的整数倍时,数据矩阵中每一列元素随机出现,其累积量会在 0 附近波动。由此可见,分组交织周期的识别问题可以归结为数据矩阵中同步码位置检测问题。

为了使算法具有较强的低信噪比适应能力,本文将直接采用软解调序列。设调制方式为 BPSK,码元 1 被映射为 1, 0 被映射为 -1, 设信道噪声为均值为 0、方差为 σ^2 的高斯白噪声。若待发送的码元为 c , 则经信道后,得到的软解调码元为

$$x = 2c - 1 + r' \quad (3)$$

式中: $c \in \{0, 1\}$; r' 服从均值为 0、方差为 σ^2 的高斯分布,即 $r' \sim \mathcal{N}(0, \sigma^2)$ 。

由前面分析可知,当构造的分析矩阵列数不同时,矩阵中每一列的同步码分布将会有很大的差异,因此进一步研究在不同列数下分析矩阵中列的一阶累积量概率分布,对于同步码的检测具有重要意义。设截获的软解调序列长度为 l , 构建的分析矩阵为 X , 其列数为 l' , 即

$$X = [v_1, v_2, \dots, v_{l'}] \quad (4)$$

式中: $v_i = (v_{1,i}, v_{2,i}, \dots, v_{w,i})^T$, $W = \lfloor l/l' \rfloor$ 。

X 中第 i 列的一阶累积量定义为

$$\text{cum}(i) = \frac{1}{W} \sum_{j=0}^{W-1} x_{j,i} \quad (5)$$

将截获的序列按式(4)进行构造分析矩阵,当同步码的位置分布再次重复第 1 行中的分布规律时,则已经排入分析矩阵中的数据量 l'' 定是帧长 L 的整数倍,由于分析矩阵的列数为 l' , 已排入分析矩阵中的数据量也一定是 l' 的整数倍,由此可知, l'' 一定满足如下条件:

$$l'' = \text{lcm}(l', L) \quad (6)$$

式中:函数 $\text{lcm}(\cdot)$ 表示取最小公倍数。此时,分析矩阵中每一行的同步码分布特征将以 lcm/l' 为周期重复出现,故分析矩阵行中出现相同同步码分布规律的概率为 l'/lcm ,不妨记 $\eta = l'/\text{lcm}$, 由于 lcm 可整除 l' , $\eta \in \{1/n \mid n = 1, 2, \dots\}$ 。由分析矩阵中列上一阶累积量的定义可知,当分析矩阵列上存在同步码时,其一阶累积量的均值为 $\eta(2c - 1)$, 方差为 $(1 - \eta^2 + \sigma^2)/W$;反之,当分析矩阵列上仅仅存在随机业务数据时,由均值与方差的定义,同理可得,其均值为 0, 方差为 σ^2/W 。当 W 较大时,由中心极限定理可知,二者近似服从高斯分布,由此可得到,在分析矩阵为任意列 l' 情况下,矩阵中第 i 列累积量服从以下概率分布:

$\text{cum}(i)$:

$$\begin{cases} \mathcal{N}(\eta(2c - 1), (1 - \eta^2 + \sigma^2)/W) & \text{存在同步码} \\ \mathcal{N}(0, (\sigma^2 + 1)/W) & \text{无同步码} \end{cases} \quad (7)$$

由式(7)可知,当 η 趋近于 0 时,存在同步码位置列的统计特性也将趋近于仅由随机业务数据所构成的列;当 l' 等于 L 时, $\eta = 1$, 与其统计特性最为接近的是 $\eta = 1/2$ 情况。图 2 给出了在噪声方差 $\sigma^2 = 3$, $W = 100$ 条件下,不同 η , $\eta \in \{1/n \mid n = 1, 2, \dots\}$ 值下分析矩阵列的一阶累积量概率密度函数曲线。

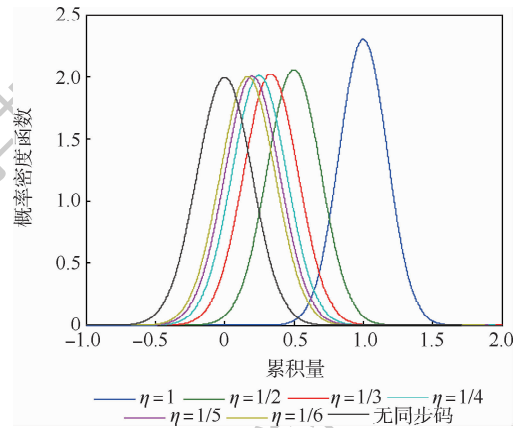


图 2 不同 η 下的概率密度函数

Fig. 2 Probability density function under different η

从图 2 结果来看,随着 η 的减小,概率密度函数曲线逐渐趋近于无同步码的情况,同时与 $\eta = 1$ 统计特性最为接近的是 $\eta = 1/2$ 的曲线,由此可知,上述分析是正确的,并且设定的判决门限只要能够有效区分出 $\mathcal{N}(2c - 1, \sigma^2/W)$ 与 $\mathcal{N}(c - 1/2, (3/4 + \sigma^2)/W)$ 两种分布,即可检测出同步码位置,从而识别交织周期。考虑在 $c = 1$ 情况 ($c = 0$ 同理可得)下,设判决门限为 Λ , 为方便说明,令 $\sigma_1^2 = \sigma^2/W$, $\sigma_0^2 = (\sigma^2 + 3/4)/W$, 则虚警概率 P_a 与漏警概率 P_f 可表示为

$$P_a = \int_{\Lambda}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_0} e^{-(y-0.5)^2/(2\sigma_0^2)} dy \quad (8)$$

$$P_f = \int_{-\infty}^{\Lambda} \frac{1}{\sqrt{2\pi}\sigma_1} e^{-(y-1)^2/(2\sigma_1^2)} dy \quad (9)$$

综合这 2 类错误,得到平均错误判决错误概率 $P_e = 0.5(P_a + P_f)$, 将 P_e 对门限 Λ 求导数,并令其等于 0, 得到

$$\frac{1}{\sqrt{2\pi}\sigma_0} e^{-(\Lambda-0.5)^2/(2\sigma_0^2)} = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-(\Lambda-1)^2/(2\sigma_1^2)} \quad (10)$$

两边取对数,化简后,得到一元二次方程,即 $b_2\Lambda^2 + b_1\Lambda + b_0 = 0$ (11) 式中:

$$b_2 = \sigma_0^2 - \sigma_1^2 \quad (12)$$

$$b_1 = \sigma_1^2 - 2\sigma_0^2 \quad (13)$$

$$b_0 = \sigma_0^2 - \sigma_1^2/4 + \sigma_0^2\sigma_1^2(\ln \sigma_1 - \ln \sigma_0) \quad (14)$$

求解式(14),得到最小错误判决门限 Λ_{opt} 为

$$\Lambda_{opt} = \frac{\sigma_0^2 - 0.5\sigma_1^2 - \sigma_0\sigma_1\sqrt{1/4 + (\sigma_1^2 - \sigma_0^2)\ln(\sigma_1/\sigma_0)}}{\sigma_0^2 - \sigma_1^2} \quad (15)$$

当 $c=0$ 时, c 被映射为 -1 , 此时最小错误判决门限与式(15)正好互为相反数, 在进行同步码检测时, 为了方便计算, 可以将求取的一阶累积量取绝对值, 将判决门限统一于 $c=1$ 的情况。当信噪比较低时, 即便是分析矩阵的列数不等于交织周期, 矩阵列中的一阶累积量也有极大的可能性出现大于门限 Λ_{opt} 的情况, 如果将是否出现一阶累积量大于设定判决作为交织周期识别的标准, 那么算法将无法在低信噪比下完成参数的正确识别, 由此可知, 还需要进一步研究在最小错误判决门限下, 同步码检测数目的统计规律。

首先, 考虑当 $\eta=1/2$ 情况, 当分析矩阵列上存在同步码时, 其一阶累积量服从 $\mathcal{N}(1/2, \sigma_0^2)$, 反之, 当矩阵列仅由随机业务数据构成时, 其一阶累积量服从 $\mathcal{N}(0, (1 + \sigma^2)/W)$, 为方便描述, 记 $\sigma_2^2 = (1 + \sigma^2)/W$, 则在门限 Λ_{opt} 下, 存在同步码与完全由随机业务数据构成的列, 其虚警概率 P_{sa0} 与 P_{sa2} 分别为

$$P_{sa0} = \int_{\Lambda_{opt}}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_0} e^{-(y-0.5)^2/(2\sigma_0^2)} dy = 0.5 \operatorname{erfc}\left(\frac{\Lambda_{opt} - 0.5}{\sqrt{2}\sigma_0}\right) \quad (16)$$

$$P_{sa2} = \int_{\Lambda_{opt}}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_2} e^{-y^2/(2\sigma_2^2)} dy = 0.5 \operatorname{erfc}\left(\frac{\Lambda_{opt}}{\sqrt{2}\sigma_2}\right) \quad (17)$$

式中: $\operatorname{erfc}(y) = \frac{2}{\sqrt{\pi}} \int_y^{\infty} e^{-t^2} dt$ 。

设每一帧数据中, 同步码序列长度为 s_0 , 则在 $l=L/2$ 时, 如果检测出同步码, 那么同步码是由 s_0 个存在同步码的矩阵列以及 $l-s_0$ 个随机业务数据列造成的虚警构成。由二项式分布定义可知, 存在同步码的 s_0 列检测出同步码的个数服从均值为 $s_0 P_{sa0}$ 、方差为 $s_0 P_{sa0}(1 - P_{sa0})$ 的二项分布, 而仅有随机业务数据构成的 $l-s_0$ 个矩阵列, 检测出同步码的个数服从均值为 $(l-s_0) P_{sa2}$ 、方差为 $(l-s_0) P_{sa2}(1 - P_{sa2})$ 的二项分布, 将这 2 种概率分布综合考虑, 得到在 $\eta=1/2$ 条件下, 检测出同

步码个数的均值与方差为

$$\mu_{syn} = s_0 P_{sa0} + (L/2 - s_0) P_{sa2} \quad (18)$$

$$\sigma_{syn}^2 = s_0 P_{sa0}(1 - P_{sa0}) + P_{sa2}(L/2 - s_0)(1 - P_{sa2}) \quad (19)$$

由统计学中 3 倍标准差准则, 可知在 $\eta=1/2$ 条件下, 检测出同步码个数大于 $\mu_{syn} + 3\sigma_{syn}$ 事件为小概率事件, 故可以得到在低信噪比下交织周期识别门限 Λ_{s_0} :

$$\Lambda_{s_0} = \mu_{syn} + 3\sigma_{syn} \quad (20)$$

其次, 当信噪比较高时, 则主要考虑在 $\eta=1$ 下, 真实的同步码位置漏警对算法的影响。当分析矩阵列等于交织周期时, 在门限 Λ_{opt} 下, 漏警概率 P_{sf} 为

$$P_{sf} = \int_{-\infty}^{\Lambda_{opt}} \frac{1}{\sqrt{2\pi}\sigma_1} e^{-(y-1)^2/(2\sigma_1^2)} dy = 0.5 \operatorname{erfc}\left(\frac{1 - \Lambda_{opt}}{\sqrt{2}\sigma_1}\right) \quad (21)$$

同理得到, 同步码的漏警个数服从均值为 $s_0 P_{sf}$ 、方差为 $s_0 P_{sf}(1 - P_{sf})$ 的二项分布, 同样由 3 倍标准差准则可知, 在 $\eta=1$ 情况下, 检测出同步码数目门限 Λ_{s_1} 为

$$\Lambda_{s_1} = s_0(1 - P_{sf}) - 3\sqrt{s_0 P_{sf}(1 - P_{sf})} \quad (22)$$

综合上述 2 种门限, 得到交织周期识别门限为

$$\Lambda_s = \max(\Lambda_{s_0}, \Lambda_{s_1}) \quad (23)$$

由式(23)可知, 在交织周期判决门限解算过程中, 需要事先预知帧同步码长度 s_0 , 虽然在一般情况下较难获得, 但是从实际通信系统来看, 同步码位数一般以字节出现(8 位), 故在初始设定 s_0 时, 可以按照 8 的整数倍从小到大设定。

在对交织周期识别过程中, 需要对潜在的周期进行遍历, 不妨设遍历的周期最小值与最大值分别为 T_{min} 和 T_{max} , 一般的方法是从 $T_{min} \sim T_{max}$ 遍历构建分析矩阵, 然后求解一阶累积量, 这种方法的效率较低, 特别是当截获的数据量很大时, 缺点尤为明显。本文提出一种快速遍历的方法, 利用分析矩阵结构特点, 仅仅构造一次分析矩阵, 就可完成多个不同矩阵列数下的一阶累积量计算。

设在遍历的交织周期为 T 、截获的序列为 $\{x_1, x_2, \dots, x_l\}$ 下构建的分析矩阵为 \mathbf{A} , 即

$$\mathbf{A} = \begin{bmatrix} x_1 & x_2 & \cdots & x_T \\ x_{T+1} & x_{T+2} & \cdots & x_{2T} \\ \vdots & \vdots & & \vdots \\ x_{(w-1)T+1} & x_{(w-1)T+2} & \cdots & x_{wT} \end{bmatrix} \quad (24)$$

式中: $w = \lfloor L/T \rfloor$ 。

当构建的分析矩阵列数为 kT 时, 构造的新分

析矩阵每一列可由式(24)中 \mathbf{A} 的行间隔 k 抽样构成,由此可知,当分析矩阵的列数由 T 扩展为 kT 时,对应于其每列的一阶累积量计算方法为

$$\mathbf{cum}_{kT}(jT+i) = \frac{1}{w'} \sum_{t=1}^{w'} x_{k(t-1)T+i} \quad (25)$$

式中: $0 \leq j \leq k-1; 1 \leq i \leq T; w' = \lfloor w/k \rfloor$ 。

由式(25)可知,仅仅通过构造一次式(24)的分析矩阵,即可求解出在列数为 $T, 2T, \dots, kT$ 下的一阶累积量值,其计算效率相比较于依次遍历的方法,最大可提升近 $\lfloor T_{\max}/T_{\min} \rfloor$ 倍,当交织周期遍历的范围很大时,本文提出的算法优势将会非常明显。

基于上述分析,可以得到分组交织周期快速识别具体步骤,如算法 1 所示。

算法 1 分组交织周期快速识别。

输入: T_{\min}, T_{\max} 、估计的同步码数目 SL 、软解调序列(序列长为 l)。

输出: EsT 。

1. 初始化遍历集合 $set = \{T_{\min}, T_{\min} + 1, \dots, T_{\max}\}$, $EsT = 0$, 标志位 $flag = 0$ 。
2. while $set \neq \emptyset$ do
3. $l' = set\{1\}$, $w = \lfloor l/l' \rfloor$, 构造数据矩阵 $\mathbf{A}_{w \times l'}$ 。
4. for $k = 1$ to $\lfloor set\{end\}/l' \rfloor$ do
5. 计算门限 Λ_{opt} 与 Λ_s 。
6. 计算 $\mathbf{A}_{w \times l'}$ 中每列累积量,并取绝对值。
7. 将步骤 6 中结果与 Λ_{opt} 比较,检测出同步码个数 $NUMS$ 。
8. if $NUMS \geq \Lambda_s$ do
9. $EsT = kl'$, $flag = 1$
10. break;
11. end if
12. end for
13. if $flag = 1$ do
14. break;
15. end if
16. end while

当实际的交织周期较小时,按照算法 1 识别出的参数 EsT 有可能为实际交织周期的整数倍,此时还需要遍历一次集合 $\{2, 3, \dots, \lfloor EsT/T_{\min} \rfloor\}$ 中元素,构造列数为 EsT/k 的分析矩阵(k 属于集合中元素且能被 EsT 整除),若检测出同步码数目大于门限且 k 为最大值,则实际交织周期为 EsT/k 。

2.2 分组交织起点与行列识别

当分组交织周期完成识别后,进一步对可恢复数据的交织起点、分组交织行与列进行识别。

传统的方法是利用三重循环遍历交织起点、交织行与列对截获的数据进行重新排列,再采用高斯消元或是遍历校验关系进行识别,这种方法计算复杂度高,同时容错性较差,而且目前还仅仅应用于短码(如卷积码)下的分组交织识别。本文针对该问题,利用帧同步后,同步码经分组交织后位置分布规律,直接完成交织同步识别,同时利用位置分布规律解算分组交织行与列。下面重点探讨在帧同步后,同步码经分组交织后的位置分布规律。

不妨设交织前每一帧中同步码数目为 s , 同步码为 $\mathbf{syn} = (a_1, a_2, \dots, a_s)$, 每一帧数据由同步码以及业务数据构成,即 $(\mathbf{syn}, \mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_K)$, 其中业务数据由 K 个码字构成。按照图 1 分组交织的规则,由同步码长度,分以下 3 种情况讨论。

1) 同步码数目 s 小于交织行 M 。

在情况 1 下,将帧数据进行分组交织,按照分组交织规则,得到数据排入分组交织块中的形式,如图 3 所示。

当按行输出后,同步码与随机业务数据分布特征如图 4 所示。从图 4 结果来看,当 $s < M$ 时,同步码被随机业务数据打散,并且相邻同步码之间间隔正好为分组交织列数。

2) 同步码长度 s 为 M 的整数倍,即 $s = kM$ 。

在情况 2 下,按照分组交织规则,可得到在分组交织矩阵中数据位置分布,如图 5 所示。

按行输出后,得到分组交织后同步码的分布规律如图 6 所示。

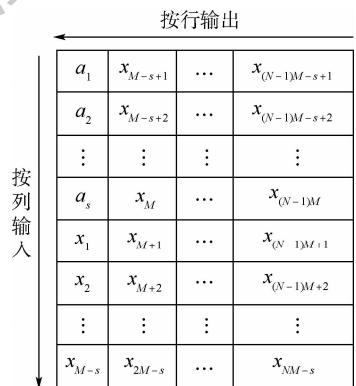


图 3 $s < M$ 时,数据排列情况

Fig. 3 Data arrangement at $s < M$

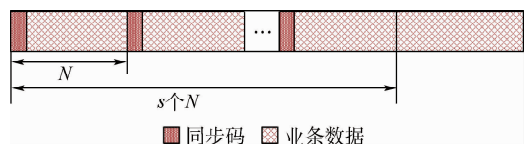


图 4 $s < M$ 时,同步码交织后分布规律

Fig. 4 Distribution rule of synchronization codes

after interleaving at $s < M$

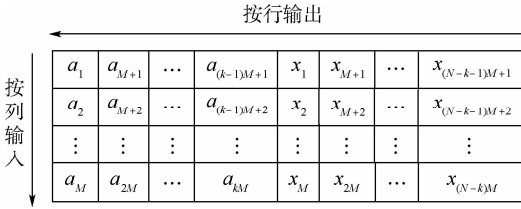


图5 $s = kM$ 时,数据排列情况

Fig. 5 Data arrangement at $s = kM$

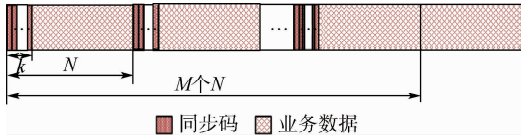


图6 $s = kM$ 时,同步码交织后分布规律

Fig. 6 Distribution rule of synchronization codes after interleaving at $s = kM$

从图6分组交织后同步码分布来看,当 $s = kM$ 时,每帧前 k bit 为同步码,紧跟后续 $(N - k)$ bit 为随机业务数据,这种结构周期性重复 M 次后,以全随机业务数据结尾。在该种情况下,只要正确检测出同步码位置,即可有效完成交织同步、分组交织列与行识别。

3) 同步码长度 s 大于 M , 且不为其整数倍, 即 $s = kM + r$ 。

情况3综合了情况1与情况2两种情况,其同步码的分布规律更为复杂,按照分组交织定义,首先给出在 $s = kM + r$ 情况下,帧数据在交织矩阵中的排列关系,如图7所示。

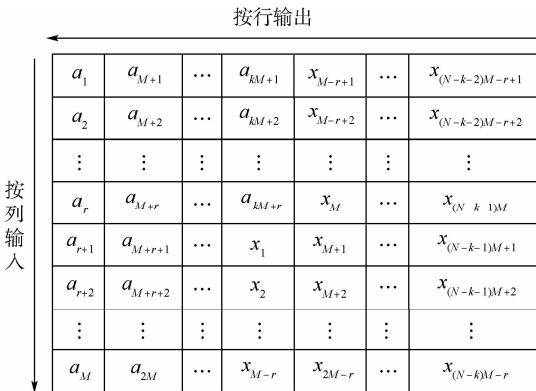


图7 $s = kM + r$ 时,数据排列情况

Fig. 7 Data arrangement at $s = kM + r$

图7中比特流按行输出交织后的序列,在输出的交织序列中,交织起点为 $k + 1$ 个同步码,紧接着为 $N - k - 1$ 个业务数据,相同的数据段重复 r 个周期,随后是 k 个同步码,同步码之后为 $N - k$ 个随机业务数据,该数据段特征重复 $M - r$ 个周期,具体同步码分布规律如图8所示。

从上述讨论的3种情况,可知在分组交织后,帧同步码的位置分布特征如下:

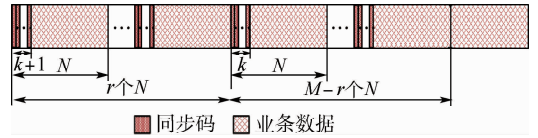


图8 $s = kM + r$ 时,同步码交织后分布规律

Fig. 8 Distribution rule of synchronization codes after interleaving at $s = kM + r$

- 1) 分组交织起点位置一定是同步码位置。
- 2) 由于随机业务数据量要远远大于同步码位数,交织起点同步后最后一个同步码位置要远远小于交织周期。
- 3) 交织起点同步后,若检测出的同步码位置不连续(情况1),则相邻两同步码位置之差等于分组交织列 N 且交织行 $M > s$ 。
- 4) 交织起点同步后,若检测出的同步码位置存在连续的情况(情况2与情况3),则同步码连续簇个数等于交织行 M 且相邻2个连续簇之间相距为交织列 N 。

利用上述分组交织下帧同步的特征,可以快速完成交织起点、分组交织列 N 识别。对于分组交织行 M 的识别,在 $s > M$ 情况下,利用连续同步码簇个数可以实现快速识别,而当 $s \leq M$ 时,由于 s 较小,无法精确获得 M 的值,此时只能利用参数之间的关系,缩小 M 的取值范围,不妨设 M 参数的范围为 $M_{\min} \sim M_{\max}$ 。由于交织周期一般为 MN 的整数倍,故 M 与 N 以及 EsT 必然满足以下关系:

$$\text{mod}(EsT/N, M) = 0 \tag{26}$$

同时, M 还满足取值范围: $\max(s, M_{\min}) \leq M \leq M_{\max}$, 利用式(26)关系约束,能够使 M 的取值范围大大减少,最后通过遍历 M 可能值,对数据帧解交织,当遍历的 M 为正确时,经过解交织过程,具有约束关系的比特恢复到了一个码组,此时码元之间的校验关系成立,可能性将达到最大。

设解交织后的软解调序列为 \mathbf{v}_l , 即

$$\mathbf{v}_l = [v_{l(1)}, v_{l(2)}, \dots, v_{l(l)}] \tag{27}$$

式中: $l'(\cdot)$ 为解交织映射函数。

设码字的校验向量为 $\mathbf{h} = [h_0, h_2, \dots, h_m]$, 为衡量在软解调序列下的校验关系成立可能性大小,引入文献[16]中余弦符合度概念,在 M 的集合中,使得序列 \mathbf{v}_l 校验符合度值最大的 M 即为识别出来的参数,即

$$\hat{M} = \max_{M \in \Phi} \frac{1}{l - m} \sum_{i=1}^{l-m} \prod_{i=0}^m \cos\left(\pi \frac{e^{2v_{l'(i+i)}/\sigma^2}}{e^{2v_{l'(i+i)}/\sigma^2} + 1} h_{m-i}\right) \tag{28}$$

式中: m 为编码约束长度;集合 $\Phi = \{M | \text{mod}(EsT/N, M) = 0\}$, 同时 M 满足: $\max(s, M_{\min}) \leq M \leq$

M_{\max} , M_{\min} 为设定的最小 M , M_{\max} 为设定的最大 M 。

由上述分析,可以得到分组交织起点、分组交织行与分组交织列参数识别流程,如算法2所示。

算法2 分组交织起点、分组交织列与行识别。

输入: 交织周期 EsT , 软解调序列, M_{\min} 、 M_{\max} 、 ΔM 。

输出: $Esstart$ 、 \hat{N} 、 \hat{M} 。

1. 初始化标志位 $flag1 = 0$, $flag2 = 0$ 。
2. 利用软解调序列构造列数为 EsT 的分析矩阵。
3. 求解分析矩阵中每一列的一阶累积量。
4. 计算同步码检测门限。
5. 检测同步码位置,将其保存于集合 Set_0 中,设同步码个数为 s 。
6. for $i = 1$ to s do。
7. 将 $Set_0(i)$ 作为帧头,构造列数为 EsT 的分析矩阵。
8. 重复步骤2与步骤3,记录末尾同步码位置 P_{end} 。
9. if $EsT - P_{end} > \Delta M$ do
10. $Esstart = Set_0(i)$
11. if 存在连续同步码簇 do
12. 确定分组交织行 \hat{M} 与分组交织列 \hat{N} 。
13. else
14. $flag2 = 1$, 确定分组交织列 \hat{N} , 确定分组交织行 \hat{M} 集合 $Set_1 = \{M | \text{mod}(EsT/N, M) = 0\}$ 。
15. end if
16. $flag1 = 1$, break
17. end if
18. if $flag1 = 1$ do
19. break
20. end if
21. end for
22. if $flag2 = 1$ do
23. 遍历集合 Set_1 , 计算每一 M 下的余弦符合度值。

24. 求取最大余弦符合度对应的 M , 即为识别的 \hat{M} 。

25. end if

从算法1与算法2的参数识别过程可知,首先当同步码数据大于 M 时,仅仅利用同步码的分布规律,即可识别出分组交织参数,与传统方法相比,具有绝对优势;其次是在同步码数目小于 M 时,不可避免需要遍历 M 值,但是与传统方法相

比,本文利用 M 、 N 以及交织周期之间的数量关系,使得遍历的 M 大大减少,同时引入了余弦符合度作为 M 的判定依据,这使得算法具有较强的低信噪比下容错性。综合2个方面来看,本文算法的性能具有明显优势。

2.3 计算复杂度分析

设截获的数据量为 l , 周期遍历范围为 $T_{\min} \sim T_{\max}$ 。在识别过程中,本文算法需要利用遍历的参数构建分析矩阵,然后计算矩阵每列的一阶累积量,故在最不利情况,周期识别总共需要进行 $(T_{\max} - T_{\min} + 1)l$ 次加法,按照本文算法中一阶累积量的求解方法,构建矩阵次数约等于 $2(T_{\max} - T_{\min} + 1) / (\lfloor T_{\max}/T_{\min} \rfloor + 1)$, 故本文算法最大的计算量近似为 $O(T_{\max}^2 l)$ 。对基于高斯消元的识别算法而言,首先需要遍历识别出交织周期,由于高斯消元法计算复杂度近似与矩阵列数的3次方成正比,交织周期识别的计算复杂度近似为 $\sum_{i=T_{\min}}^{T_{\max}} i^3$ 。然后采用3重循环遍历识别出交织参数,设遍历的分组交织行范围为 $M_{\min} \sim M_{\max}$, 列范围为 $N_{\min} \sim N_{\max}$, 其计算复杂度近似为 $(N_{\max} - N_{\min} + 1)(M_{\max} - M_{\min} + 1)T_{\max}^3$, 故其总的计算复杂度近似为 $O(T_{\max}^4 + N_{\max}M_{\max}T_{\max}^3)$ 。由此可知,相比较于高斯消元的方法,本文算法的计算复杂度大大减少。

3 仿真分析

本节主要从3个部分对算法进行验证,首先验证算法的有效性,并从同步码数目 $s < M$ 、 $s = kM$ 以及 $s = kM + r$ 三种情况验证帧同步后同步码的分布规律 r 为 s 除以 M 后的余数;其次验证存在估计同步码偏差,不同同步码数目以及截获的帧数目对算法性能的影响;最后与现有的算法性能进行对比。

3.1 算法有效性验证

仿真设定分组交织器行数 $M = 18$, 列数为 $N = 14$, 每帧同步码长度设定为 8 bit、24 bit 以及 36 bit 3种情况,将卫星通信中常用的(2,1,7)卷积码作为随机业务数据的编码方案,其编码的生成多项式为 $(1 + D^2 + D^3 + D^5 + D^6, 1 + D + D^2 + D^3 + D^6)$, D 表示寄存器延时。每帧数据总长度设定为 3 024, 截获的帧块数目为 100 帧,将数据前任意比特舍去(仿真设定为 23 bit, 下一帧同步位置为 3 002 bit), 为验证帧同步识别准备条件。设定的信道环境为 -3 dB。首先验证在3种同步码长度下,对交织周期识别有效,设定遍历的交织周

期 $T_{\min} = 1\ 000, T_{\max} = 10\ 000$, 当遍历到交织周期等于 3 024 时, 出现检测出的同步码个数大于门限 A_s , 由于在 1 000 和 10 000 之间能被 3 024 整除数目为 1 512 以及 1 008, 故还需要进一步考察在分析矩阵列数为 1 512 与 1 008 情况, 表 1 进一步给出了在 3 种同步码长度以及 3 种交织周期下, 同步码检测门限 A_{opt} 、交织周期识别门限 A_s 、同步码检测数目结果。

表 1 交织周期识别结果

Table 1 Recognition results of interleaving period

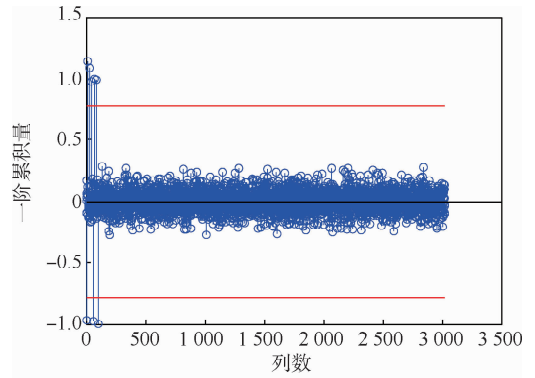
s/bit	EsT	A_s	A_{opt}	检测数目	结果判断
8	3 024	7.219 3	0.781 71	8	√
	1 512	7.839 8	0.783 26	0	×
	1 008	7.963 7	0.783 78	0	×
24	3 024	22.578 9	0.781 73	23	√
	1 512	23.725 5	0.783 27	0	×
	1 008	23.939 2	0.783 78	0	×
36	3 024	34.203 6	0.781 73	36	√
	1 512	35.661 3	0.783 27	0	×
	1 008	35.925 5	0.783 78	0	×

从交织周期识别过程来看, 在 3 种同步码长度下, 当遍历交织周期 EsT 等于 3 024 时, 唯一出现检测到的同步码数目大于门限 A_s , 这与设定的参数相同, 这说明算法对交织周期识别有效。然后进一步遍历检测出的同步码位置, 进行交织起点同步, 在 3 种同步码长度下同步后, 分析矩阵每一列的一阶累积量分布结果如图 9 所示。

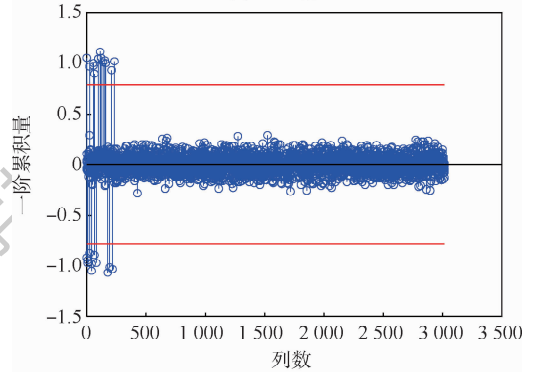
通过分析同步后同步码位置之间的数量关系, 进一步得到交织起点、分组交织行与列参数, 记录如表 2 所示。

从表 2 结果来看, 算法通过遍历同步码位置, 最终得到了相同的交织起点位置, 这与仿真设定正好一致, 当 $s > M$ 时, 利用同步码位置关系以及连续同步码簇, 解算出的分组交织列 N 与行 M 与实际设定的结果也完全一致, 这说明总结的同步码经交织后的位置规律是和实际相符的。当 $SL < M$ 时, 由于不存在连续的同步簇, 此时只能通过遍历可能的 M 值, 先对截获数据进行解交织操作, 然后利用 $(2, 1, 7)$ 卷积码校验向量进行校验符合度统计, 由于 $EsT = 3\ 024$, 故满足条件的 M 集合为 $\{8, 9, 12, 18, 24, 27, 36, 54, 72\}$, 遍历 M 集合, 解交织后余弦符合度结果如图 10 所示。

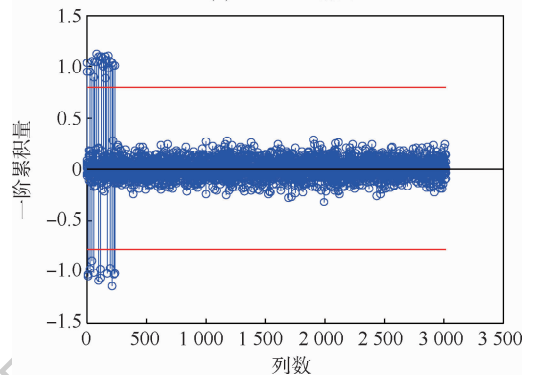
从图 10 结果来看, 在 $M = 18$ 时, 余弦符合度出现最大值, 故识别出分组交织行为 18, 这与设定条件是一致的。综合上述实验可知, 本文算法能够在低信噪比下有效完成分组交织参数识别。



(a) $s < M$ 情况



(b) $s = kM + r$ 情况



(c) $s = kM$ 情况

—●— 累积量 ——— 门限

图 9 帧同步后一阶累积量分布

Fig. 9 First-order cumulant distribution after frame synchronization

表 2 交织起点、分组交织列与行识别结果

Table 2 Results for recognition of interleaving starting point, packet interleaving column and row

s/bit	遍历同步码位置	起点位置	连续同步码簇数	\hat{N}	\hat{M}
8	7	3 002	无	14	×
24	21	3 002	14	14	18
36	31	3 002	14	14	18

3.2 算法容错性能验证

本节主要是对算法的性能进行验证, 仿真中设定业务数据为 $(2, 1, 7)$ 卷积码, 每帧数据由 6 300 比特构成, 设定分组交织器交织行 $M = 18$, 交织列 $N = 14$, 仿真中将前 23 bit 数据剔除, 用以进行交织起点识别, 此时下一帧的交织起点为 6 278。考察

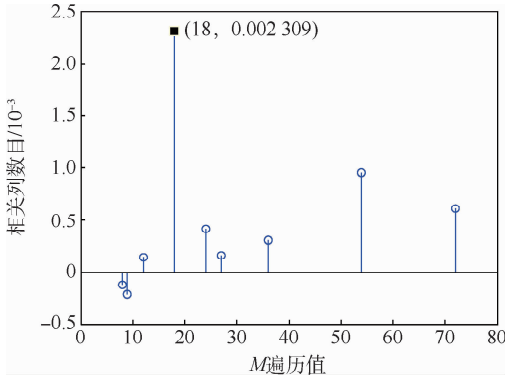


图 10 校验统计结果

Fig. 10 Results of verification statistics

存在估计的同步码偏差、实际的同步码数目以及截获的帧数目 3 个因素对算法性能的影响。

1) 仿真 1:同步码偏差数目影响。

仿真设定同步码数目为 32 bit,截获的帧数目为 100,由于在实际通信系统中同步码总是以字节形式出现(8 位),在算法中初始设定估计的同步码位数时按照 8 的倍数设定,考察 2 种情况:①估计的同步码数目小于等于实际的数目,估计的同步码位数设定为 8,16,24,32 bit;②估计的同步码数目大于实际的数目,估计的同步码位数设定为 40,48,56,64 bit。仿真设定信噪比范围为 -12 ~ -5 dB,蒙特卡罗实验次数为 1 000,统计的参数正确识别率结果如图 11 所示。

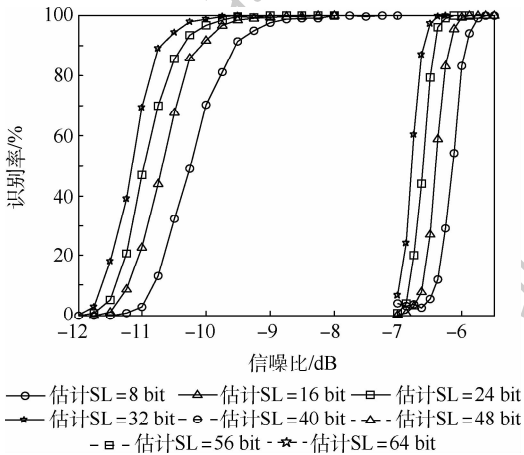


图 11 同步码偏差数目对算法性能影响

Fig. 11 Influence of deviation number of synchronization codes on algorithm performance

首先从图 11 结果来看,本文算法具有较好的容错性,在 -6 dB 条件下,参数识别率能达到 98% 以上;其次还可看出,估计的同步码数目小于实际同步码数目下的性能要明显好于实际数目下的性能,同时估计的同步码数目越接近实际同步码数目的整数倍,识别的正确率越高。主要原因在于:当估计的同步码数目大于实际数目时,计算

所得到的交织周期判决门限要大于最小正周期下同步码数目,无法完成最小正周期的识别,故仅能识别出为 2 倍的最小正周期,此时实际的同步码数目由 32 变成 64,由于构建的矩阵列数变成 2 倍,参与一阶累计的行数变为原来的 1/2,性能会明显下降。由此可知,当估计同步码时,采取的策略是从小往大进行遍历,这样更容易识别出交织周期。

2) 仿真 2:同步码数目影响。

仿真设定截获的帧数为 100,每帧同步码设定长度为 40,32,24,16,8 bit 等 5 种,为了单纯考察同步码数目对算法影响,识别过程中令估计的同步码数目等于实际同步码数目,信噪比范围设定为 -12 ~ -7 dB,间隔 0.25 dB 取值,统计在不同同步码长度以及信噪比下参数正确识别率,结果如图 12 所示。

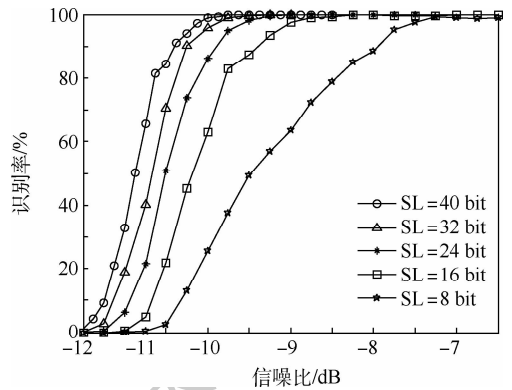


图 12 同步码数目对算法性能影响

Fig. 12 Influence of the number of synchronization codes on algorithm

从图 12 结果来看,随着每帧中同步码数目的增加,参数的正确识别率也在逐渐提升,主要原因在于:同步码数目增加后,实际检测到的同步码数目也会增加,而算法正好是以检测到的同步码数目作为交织周期识别的统计量,同时同步码数目越多,就越能够反映出分组交织参数与同步码位置之间的数量关系。从最终结果来看,在最不利情况下,参数识别率能够在 -8 dB 条件下达到 95% 以上的识别率,这说明本文算法具有较强的容错性能。

3) 仿真 3:截获帧块数目影响。

仿真设定每帧同步码数目为 32 bit,识别过程中设定估计的同步码数目为 16,设定截获的帧块数据分别为 100,200,300,400 和 500,噪声信道环境中,信噪比设定范围为 -20 ~ -8 dB,间隔 0.25 dB 取值,蒙特卡罗仿真次数为 1 000 次,分别统计在不同截获帧块数目以及信噪比条件下参数的正确识别率,结果如图 13 所示。

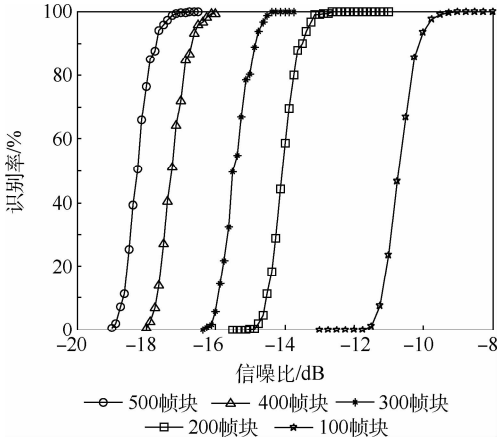


图 13 截获帧块数目对算法性能影响
Fig. 13 Influence of intercepted frame blocks on algorithm performance

从图 13 结果来看,通过增加截获帧块数目,可以有效提升分组交织参数识别的正确率。当帧块数目增加时,构建的分析矩阵行数相应也会增加,此时概率分布中方差值会越趋近于 0,故在设定的判决门限下,检测出同步码的虚警与漏警概率会减少,从而使参数的正确识别率得到提升。由此可知,当实际的信道环境极端恶劣时,可以通过增加截获的数据帧块,来有效克服噪声的影响。

3.3 与其他算法对比

与本文算法进行对比的是基于软信息的识别方法^[15]与硬判决下的高斯消元方法^[14]。设定分组交织 $M = 19$, 交织列 $N = 26$; 随机业务数据采用编码系数为 (15, 17) 的 (2, 1, 4) 卷积码以及编码系数为 (133, 171) 的 (2, 1, 7) 卷积码, 每一帧数据长度设定为 2 470, 其同步码位置设定为前 16 bit 以及前 32 bit 2 种, 截获的帧数目设定为 200 帧, 分别记录 3 种算法对参数的正确识别率, 结果如图 14 所示。

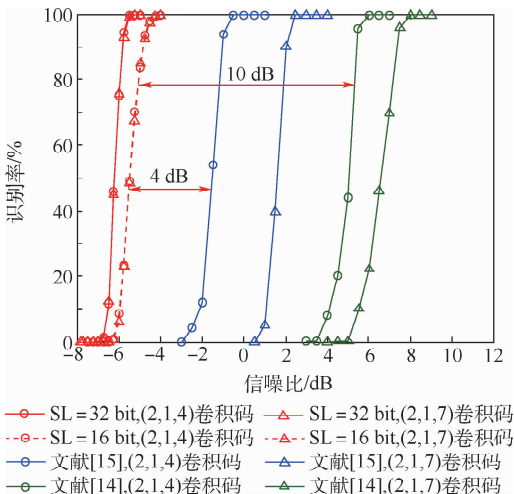


图 14 三种算法的性能对比

Fig. 14 Performance comparison among three algorithms

从图 14 中 3 种算法的识别性能对比来看,首先本文算法性能要远远好于其他 2 种算法,即使是在同步码为 16 bit 的条件下,与文献 [15] 基于软信息的算法相比,性能至少提升 4 dB; 与文献 [14] 中基于硬判决的算法相比,性能提升近 10 dB; 其次由于文献 [14-15] 主要利用了码元之间的约束关系, 编码约束长度对其算法的性能会具有较大的影响, 从图 14 可以明显看出文献 [14-15] 在 (2, 1, 7) 卷积码下的性能要差于在 (2, 1, 4) 卷积码下的性能, 而对本文算法而言, 其性能仅仅随着同步码长度的增加而提升, 与采用的何种编码方式无关。

由此可知,与文献 [14-15] 的方法相比,本文算法不光具有较强的低信噪比适应性,同时还具有很强的通用性。

4 结论

- 1) 分析了在数据矩阵中,同步码与随机业务数据所在列的一阶累积量统计特性,基于最小错误判决准则,实现同步码的快速检测。
- 2) 充分考虑了在低信噪比下,随机业务数据列的虚警概率与同步码位置列的漏警概率,基于 3 倍标准差准则,设定了稳健的分组交织周期识别门限。
- 3) 利用了数据矩阵结构特点,提出了一种仅构造一次分析矩阵就能实现多个交织周期遍历的方法,使得算法的计算效率得到较大的提升。
- 4) 最后总结出分组交织帧同步后,同步码位置分布的特征,通过遍历同步码位置,与帧同步后的特征进行比对,利用交织后同步码位置关系,最终完成交织同步、分组交织列与交织行的快速识别。

参考文献 (References)

[1] 刘杰,张立民,占超. 基于矩阵分析的线性分组码盲识别 [J]. 系统工程与电子技术,2017,39(2):404-409.
LIU J,ZHANG L M,ZHAN C. Blind recognition of linear block codes based on matrix analysis [J]. Systems Engineering and Electronics,2017,39(2):404-409 (in Chinese).
[2] 刘杰,张立民,钟兆根. 基于二元域等效的 RS 码编码参数盲识别 [J]. 电子学报,2018,46(12):2888-2895.
LIU J,ZHANG L M,ZHONG Z G. Blind parameter identification of RS code based on binary field equivalence [J]. Acta Electronica Sinica,2018,46(12):2888-2895 (in Chinese).
[3] 于沛东,彭华,巩克现,等. 基于最小二乘代价函数的卷积码盲识别方法 [J]. 电子学报,2018,46(7):1545-1552.
YU P D,PENG H,GONG K X, et al. Blind recognition of convolutional codes based on least-square cost-function [J]. Acta

- Electronica Sinica, 2018, 46(7): 1545-1552 (in Chinese).
- [4] 姚智刚, 解辉, 韩壮志, 等. 基于分段 Walsh-hadamard 变换的卷积码盲重建算法[J]. 电子与信息学报, 2019, 41(9): 2047-2054.
- YAO Z G, XIE H, HAN Z Z, et al. Blind recognition of convolutional code based on partitioned Walsh-Hadamard transform [J]. Journal of Electronics & Information Technology, 2019, 41(9): 2047-2054 (in Chinese).
- [5] 吴昭军, 张立民, 钟兆根. 基于最大序列相关性的 Turbo 码交织器识别[J]. 航空学报, 2019, 40(6): 257-268.
- WU Z J, ZHANG L M, ZHONG Z G. Blind recognition of interleaver for Turbo codes based on maximum sequence correlation [J]. Acta Aeronautica et Astronautica Sinica, 2019, 40(6): 257-268 (in Chinese).
- [6] 陈泽亮, 李静, 彭华, 等. 利用 Gibbs 采样进行优化的 Turbo 码交织器识别[J]. 电子学报, 2018, 46(1): 15-23.
- CHEN Z L, LI J, PENG H, et al. An optimization method using Gibbs sampler for turbo code interleaver identification [J]. Acta Electronica Sinica, 2018, 46(1): 15-23 (in Chinese).
- [7] BUREL G, GAUTIER R. Blind estimation of encoder and interleaver characteristics in non-cooperative context [C] // IASTED International Conference on Communications, Internet and Information Technology. Piscataway: IEEE Press, 2003: 17-19.
- [8] SICOT G, HOUCKE S. Blind detection of interleaver parameters [J]. Signal Processing, 2009, 89(4): 450-462.
- [9] GAN L, LI D, LIU Z H, et al. A low complexity algorithm of blind estimation of convolutional interleaver parameters [J]. Science in China Series F: Information Sciences, 2012, 55(10): 1-9.
- [10] XU Y Y, ZHONG Y, HUANG Z P. An improved blind recognition method of the convolutional interleaver parameters in a noisy channel [J]. IEEE Access, 2019, 6(7): 101775-101784.
- [11] 解辉, 王丰华, 黄知涛. 卷积交织器盲识别方法[J]. 电子与信息学报, 2013, 35(8): 1952-1957.
- XIE H, WANG F H, HUANG Z T. A method for blind recognition of convolutional interleaver [J]. Journal of Electronics & Information Technology, 2013, 35(8): 1952-1957 (in Chinese).
- [12] 于沛东, 彭华, 巩克家, 等. 利用帧同步码的卷积交织器快速盲识别方法[J]. 电子学报, 2018, 46(6): 1530-1536.
- YU P D, PENG H, GONG K J, et al. Fast blind recognition of convolutional interleavers based on existence of frame sync codes [J]. Acta Electronica Sinica, 2018, 46(6): 1530-1536 (in Chinese).
- [13] 郑鹏鹏, 张玉, 杨晓静. 基于低码率线性分组码的分组交织长度识别[J]. 电子信息对抗技术, 2012, 27(4): 9-12.
- ZHENG P P, ZHANG Y, YANG X J. Identification of the length of packet interleaver based on low code-rate binary linear block codes [J]. Electronic Information Warfare Technology, 2012, 27(4): 9-12 (in Chinese).
- [14] 陆凯, 张旻, 李敬昊. 基于矩阵秩统计的卷积码分组交织盲识别[J]. 探测与控制学报, 2015, 37(3): 40-44.
- LU K, ZHANG M, LI X H. Blind identifying of convolutional code and packet interleaver based on matrix rank statistic [J]. Journal of Detection & Control, 2015, 37(3): 40-44 (in Chinese).
- [15] 陈泽亮, 巩克现, 彭华, 等. 基于软信息的分组交织和卷积码联合识别[J]. 电子学报, 2018, 46(6): 1454-1460.
- CHEN Z L, GONG K X, PENG H, et al. Joint blind recognition of packet interleaver and convolution code based on soft information [J]. Acta Electronica Sinica, 2018, 46(6): 1454-1460 (in Chinese).
- [16] WU Z J, ZHANG L M, ZHONG Z G. A maximum cosinoidal cost function method for parameter estimation of RSC Turbo codes [J]. IEEE Communications Letters, 2019, 23(3): 390-393.

Recognition of packet interleaver at low SNR

WU Zhaojun^{1,2}, LIU Kai^{1,*}, ZHONG Zhaogen³, DAN Bo⁴, ZHOU Leike²

(1. Institute of Information Fusion, Naval Aviation University, Yantai 264001, China;

2. Southwest Institute of Electronics and Telecommunications, Chengdu 610041, China;

3. School of Aviation Basis, Naval Aviation University, Yantai 264001, China;

4. School of Coastal Defense, Naval Aviation University, Yantai 264001, China)

Abstract: In view of the shortcomings of the existing algorithms for blind recognition of packet interleaver, which are high computational complexity and poor fault tolerance, a new recognition algorithm based on the distribution of synchronization codes after packet interleaving is proposed in this paper. Firstly, the proposed algorithm based on the statistical characteristics of data matrix gives the function of probability density distribution for synchronous code and random traffic data in any number of matrix columns, and based on the minimum error decision criterion, the detection threshold of synchronous code is set. At the same time, the detection threshold of robust interleaving period is set based on the criterion of 3 times standard deviation. Secondly, the corresponding relationship between each row and column in the data matrix is analyzed, and a fast interleaving period traversal method is proposed, which greatly reduces the number of times of data matrix construction. Finally, the four rules of distribution of synchronization codes are summarized, and by traversing the synchronization codes and utilizing the relationship of positions between synchronization codes, the parameters of synchronization positions, interleaving column and row can be identified efficiently. The simulation results show that the algorithm has a strong error tolerance at low SNR and the correct rate of parameter recognition can reach more than 98% at the SNR of -6 dB. At the same time, compared with the existing methods, its performance is improved by $4-10$ dB and the calculation efficiency is significantly improved.

Keywords: cognitive radio; channel coding; packet interleaver; synchronization code; feature of frame synchronization; recognition

Received: 2020-05-18; **Accepted:** 2020-08-07; **Published online:** 2020-09-15 16:08

URL: kns.cnki.net/kcms/detail/11.2625.V.20200915.1401.005.html

Foundation items: National Natural Science Foundation of China (91538201); Taishan Scholar Special Foundation (ts201511020)

* **Corresponding author.** E-mail: wendao_2008@163.com

http://bhxb.buaa.edu.cn jbuua@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2020.0194

基于深度残差收缩网络的滚动轴承故障诊断



车畅畅, 王华伟*, 倪晓梅, 蔺瑞管

(南京航空航天大学 民航学院, 南京 210016)

摘 要: 滚动轴承的准确故障诊断是确保机械设备安全可靠运行的必要手段。针对多故障、长时间序列的滚动轴承振动信号,提出了一种基于深度残差收缩网络(DRSN)模型的故障诊断方法。首先,根据采集到的滚动轴承数据构造故障样本,针对多种故障类型下的长时间序列的振动信号,按照一定尺寸将长时间序列矩阵化,构成多故障类型的灰度图故障样本。从正常到故障的滚动轴承性能退化过程,通过多个采样点的随机采样,构造全寿命周期的故障样本用于故障诊断。其次,在多层深度学习模型基础上,将残差收缩网络模块加入到卷积神经网络(CNN)中构建深度残差收缩网络模型用于故障诊断,其中通过将残差项加入到网络中训练解决了多层网络模型的模型退化问题,利用软阈值化实现了样本降噪。最后,为了验证所提方法的有效性,采集了滚动轴承的多故障时间序列样本和全寿命周期故障样本用于故障诊断。实例验证的结果表明:所提深度残差收缩网络模型在处理含噪声样本时仍具有良好的鲁棒性,多层网络模型下没有明显的网络退化,能够保持较高的故障诊断正确率。在处理2种轴承故障数据集时,与其他模型相比,所提方法训练误差更低,平均故障诊断正确率提高1%~6%。

关键词: 滚动轴承; 故障诊断; 深度残差收缩网络(DRSN); 卷积神经网络(CNN); 软阈值化

中图分类号: TH17

文献标志码: A

文章编号: 1001-5965(2021)07-1399-08

滚动轴承在机械设备中应用广泛,由于其长时间处在高速、高负荷、变工况、强噪声干扰运转环境下,故障率很高。及时发现滚动轴承潜在故障并判断故障类型能够指导维修工作、提高维修效率以及机械设备的可靠性和安全性。目前,主流的滚动轴承故障诊断方法大多都基于对振动、光信号和电信号的研究,其中基于滚动轴承振动信号的故障诊断是当前研究的热点。多种故障的振动信号在长时间序列下包含着丰富的故障特征信息。通过数据驱动方法发掘振动信号与故障规律之间的映射关系,能够提取出通用特征用于故障诊断。常用的方法包括统计分析方法、传统机器学习方法、深度学习方法^[1]。

统计分析方法能够在数学模型的基础上通过公式推导得到相关指标用于衡量故障发生的可能性。常用的统计分析方法包括滤波算法、奇异值分解、模态分解、小波变换等^[2-5]。统计分析方法针对特定研究对象和样本集能够得到综合故障指标,然而对于多工况下数据集的故障诊断需要重复建模,因此通用性有限。传统机器学习方法包括支持向量机、隐马尔可夫方法、BP神经网络、径向基神经网络、集成学习等^[6-7]。传统机器学习方法大多具有浅层结构,对复杂数据集的非线性特征提取效果有限,因此常常被用作深度学习模型输出层。

深度学习模型是机器学习模型的重要分支,

收稿日期: 2020-05-18; 录用日期: 2020-07-10; 网络出版时间: 2020-07-21 13:57

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20200721.0931.001.html

基金项目: 国家自然科学基金(U1833110)

* 通信作者. E-mail: wang_hw66@163.com

引用格式: 车畅畅, 王华伟, 倪晓梅, 等. 基于深度残差收缩网络的滚动轴承故障诊断[J]. 北京航空航天大学学报, 2021, 47(7): 1399-1406. CHE C C, WANG H W, NI X M, et al. Fault diagnosis of rolling bearing based on deep residual shrinkage network [J]. Journal of Beijing University of Aeronautics and Astronautics, 2021, 47(7): 1399-1406 (in Chinese).

通过贪婪逐层预训练和反向微调以及各种非线性激活函数的组合应用实现复杂特征提取。常用的深度学习模型特征包括自编码器、卷积神经网络(Convolutional Neural Networks, CNN)、深度信念网络、循环神经网络等^[8]。深度学习模型因为其出色的隐含特征提取和大数据处理能力在工程领域也有广泛的应用。

宫文峰等^[9]提出了一种改进 CNN 的新方法。使用全局均值池化技术、数据增强、Dropout 等深度学习训练技巧实现滚动轴承故障的快速智能诊断;朱丹宸等^[10]将不同故障下多个传感器测得的一维振动信号转化为二维灰度图像作为 CNN 的输入,用于滚动轴承故障诊断;Qiu 等^[11]在利用时域分析、频域分析、小波包分解等特征提取方法获得相关指标的基础上,利用 CNN 对齿轮箱进行智能故障诊断;Sun 等^[12]针对实际故障诊断分布的多样性和可变性,提出了一种求解域自适应的稀疏叠加上噪声自编码器算法,用于提高目标区域的滚动轴承故障诊断精度;Che 等^[13]针对滚动轴承在多种工况下缺乏样本标签的问题,提出了一种适用于滚动轴承故障诊断的区域自适应深度信念网络。

随着深度学习模型层数的提高,会出现模型退化和过拟合的问题。在实际运行环境中存在的较强噪声干扰也会影响滚动轴承故障诊断的精度。为了解决这些问题,本文提出了基于深度残差收缩网络(DRSN)的滚动轴承故障诊断模型。首先,对滚动轴承数据集进行预处理,针对多故障对应的长时间序列的振动信号数据,将其矩阵化为多维度的灰度图故障样本,并将对应的故障类型热编码化,从而得到深度学习模型易于处理的样本形式。针对轴承从正常到故障采集到的全寿命周期故障样本,通过多个采样点的随机采样,构造故障样本用于进一步的深度故障特征提取。其次,在多层 CNN 模型的基础上,将残差项加入到模型中进行训练,降低了多层训练过程中的样本特征损失,再将软阈值化和注意力机制用于网络模型的特征激活,从而针对不同等级的含噪声样本设定不同的阈值,实现样本的降噪处理。最后,通过基于不同轴承故障数据集构造深度学习模型,并通过训练和优化实现精准化故障诊断。

1 深度残差收缩网络模型

1.1 深度学习算法理论

深度学习模型具有多层网络结构和非线性激活函数,能够实现大数据样本的复杂特征提取,在

工程领域有很好的应用。本文选取了常用的 CNN 模型作为研究对象。CNN 包括输入层、卷积层、池化层、全连接层、输出层。输入层能够收集相关数据样本将其矩阵化,从而转化为模型易于处理的图片样本,同时对图片样本进行归一化、降维、去均值、白化等预处理。卷积层中将每个神经元都作为滤波器,通过将一组固定的权重作为卷积核和不同窗口内的数据做内积运算,通过滑动窗口完成整个样本的卷积运算,其中通过多个卷积核对样本进行处理可以得到包含多通道的特征图,从而将低维度特征映射到高维度空间。池化层用于压缩数据和参数的量,减小过拟合,常用的池化操作包括 Max Pooling 和 Average Pooling。全连接层在卷积神经网络尾部,由多个通道的不同卷积核的隐层一维展开而组成。通过激活函数将全连接层与输出层相连接,从而输出模型分类结果。一个基础的 CNN 模型的结构如图 1 所示。

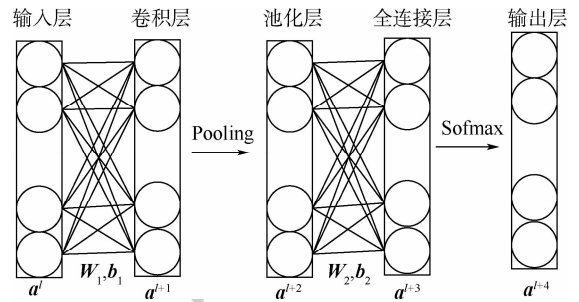


图 1 CNN 模型

Fig. 1 Model of CNN

CNN 模型前向传播各层的公式如下^[14]:

$$a^{l+1} = \text{ReLU}(W_1 a^l + b_1) \quad (1)$$

$$a^{l+2} = \text{Pooling}(a^{l+1}) \quad (2)$$

$$a^{l+3} = W_2 \cdot a^{l+2} + b_2 \quad (3)$$

$$a^{l+4} = S(a^{l+3}) \quad (4)$$

式中: W 和 b 分别为权重向量和偏置向量;输入 a^l 经过卷积层 a^{l+1} 、池化层 a^{l+2} 、全连接层 a^{l+3} ,通过激活函数 $S(x) = 1/(1 + e^{-x})$ 输出分类结果 a^{l+4} 。

假设实际样本标签为 y ,定义损失函数为

$$J(W, b, a, y) = \frac{1}{2} \|a^{l+4} - y\|^2 \quad (5)$$

在求解梯度时假设各层的中间依赖部分 δ 为^[15]

$$\delta^{l+4} = \frac{\partial J(W, b, a, y)}{\partial a^{l+3}} = (a^{l+4} - y) \odot S''(a^{l+3}) \quad (6)$$

$$\delta^{l+3} = (W_2)^T \delta^{l+4} \odot S'(a^{l+3}) \quad (7)$$

$$\delta^{l+2} = \text{Unsample}(\delta^{l+3}) \odot S'(a^{l+3}) \quad (8)$$

$$\delta^{l+1} = \delta^{l+2} \cdot \text{rot180}(W_1) \odot S'(a^{l+3}) \quad (9)$$

式中:Upsample 函数表示把 δ^{l+3} 中所有子矩阵的

各个池化局域的值平均后放在还原后的子矩阵位置的过程;⊙为 Hadamard 积,表示 2 个矩阵的对应元素相乘; $S(x) = 1/(1 + e^{-x})$ 为 Sigmoid 激活函数;函数 $\text{rot}180(\cdot)$ 表示将卷积核反转 180° 。

综上所述可以得到反向传播过程中模型参数的更新公式为

$$W_2 = W_2 - \lambda \delta^{l+3} (a^{l+2})^T \quad (10)$$

$$b_2 = b_2 - \lambda \delta^{l+3} \quad (11)$$

$$W_1 = W_1 - \lambda \delta^{l+1} \cdot a^l \quad (12)$$

$$b_1 = b_1 - \lambda \sum_{u,v} (\delta^{l+1})_{u,v} \quad (13)$$

式中: λ 为学习率; $\sum_{u,v} (\delta^{l+1})_{u,v}$ 表示将 δ^{l+1} 的各个子矩阵的项分别求和,得到一个误差向量。

经过前向传播和反向传播微调可以得到训练好的 CNN 网络模型。

1.2 深度残差收缩网络

对于传统的深度学习模型,其网络层数越多非线性的表达能力就越强,该模型就能学习到更多的特征。然而随着网络的增加,传统的多层网络结构的非线性表达很难去表示恒等映射,因此模型会出现网络退化问题。振动信号数据中普遍存在噪声干扰,影响故障诊断准确率。为了解决上述问题,本文提出了深度残差收缩网络用于振动信号样本的故障诊断。深度残差收缩网络是一种面向含噪声数据的深度学习方法,在深度残差网络(DRN)进行基于反向传播的模型训练时,其损失不仅能够通过卷积层等进行逐层的反向传播,还能够通过残差项的恒等映射进行更为方便的反向传播。再利用软阈值化对振动信号进行降噪处理,能够得到更优的模型^[16]。

假设要求解的映射为 $H(a^l)$,将这个问题转换为求解网络的残差映射函数 $F(a^l)$, $F(a^l) = H(a^l) - a^l$ 。相较于 ReLU 函数,软阈值化能够更加灵活地设置特征取值区间。在残差收缩网络中,阈值是可以通过注意力机制根据样本自身情况自动调整的。深度残差收缩网络模型的一部分如图 2 所示,其中输入 a^l 的尺寸为 $C \times N$,在经过隐层 1 后通过 ReLU 函数得到 a^{l+1} 作为隐层 2 的输入。在隐层 2 通过构建一个小性子网络,学习得到一组取值区间在 $0 \sim 1$ 之间的阈值 α ,再进行特征的软阈值化并加上残差项 $F(a^l)$ 从而得到输出 a^{l+2} 。其中各层的输出如下^[17]:

$$a^{l+1} = \text{ReLU}(W^{l+1} a^l + b^{l+1}) \quad (14)$$

$$a^l = W^{l+2} \text{ReLU}(W^{l+1} a^l + b^{l+1}) + b^{l+2} \quad (15)$$

式(16)是对每个维度下的 a^l 和对应的阈值

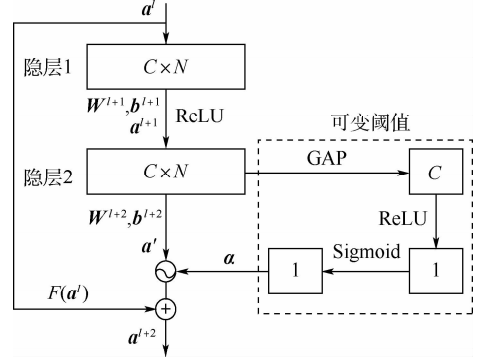


图 2 深度残差收缩网络模型

Fig. 2 Model of deep residual shrinkage network

α 进行比较的基础上得到的软阈值化结果。

$$a^s = \begin{cases} a^l - \alpha & a^l > \alpha \\ 0 & -\alpha \leq a^l \leq \alpha \\ a^l + \alpha & a^l < -\alpha \end{cases} \quad (16)$$

$$a^{l+2} = a^s + F(a^l) \quad (17)$$

2 滚动轴承故障诊断模型

本文构建了基于深度残差收缩网络的滚动轴承故障诊断模型,该模型在多层 CNN 的基础上添加了残差项,并进行了软阈值化。其中恒等映射是深度残差网络的核心贡献,通过将残差项加入网络进行训练,极大程度地降低了深度神经网络训练的难度。软阈值化和注意力机制能够让深度残差网络根据数据集的噪声含量设定不同的阈值,从而使模型具有一定的降噪能力。

基于深度残差收缩网络的滚动轴承故障诊断模型的流程如图 3 所示。首先,采集滚动轴承多故障、长时间序列、全寿命周期的振动信号构成故

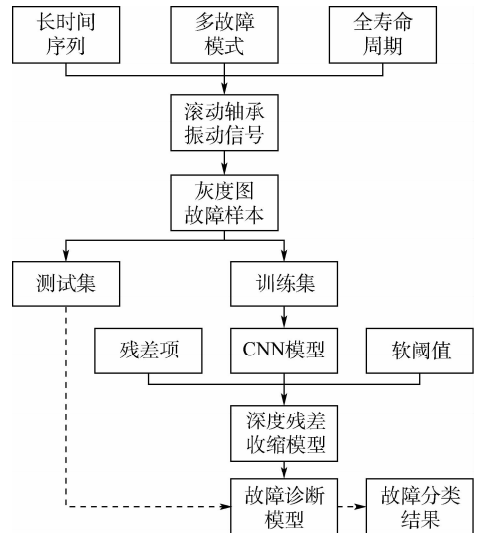


图 3 基于深度残差收缩网络的滚动轴承故障诊断流程

Fig. 3 Fault diagnosis flowchart of rolling bearing based on deep residual shrinkage network

障样本。然后,将样本矩阵化从而构成灰度图故障样本用于深度学习训练,再将灰度图故障样本分为训练集和测试集,其中将训练集代入到由残差项、软阈值和多层 CNN 而组成的深度残差收缩网络中,经过训练和优化得到故障诊断模型。最后,将测试集代入到故障诊断模型中得到故障诊断结果,通过故障分类正确率来衡量模型的有效性。

3 实例分析

为了验证本文方法的有效性,分别采用了凯斯西储大学(Case Western Reserve University)的滚动轴承数据集和西南交通大学的轴承数据集作为实验对象^[18-19],分别构成多故障长时间序列故障样本和全寿命周期故障样本。

3.1 基于多故障长时间序列样本的滚动轴承故障诊断

选用凯斯西储大学的滚动轴承数据集中采样频率为 48 kHz 的故障尺寸为 0.007 in(1 in(英尺) = 0.0254 m)的故障样本集,该数据集的 6 种故障以及对应的长时间序列数如表 1 所示。

表 1 滚动轴承故障样本集

Table 1 Failure dataset of rolling bearing

故障代号	故障描述	时间序列数
F1	内圈故障	1×10^4
F2	滚珠故障	1×10^4
F3	外圈承压端故障	1×10^4
F4	外圈侧面故障	1×10^4
F5	外圈承压端对面故障	1×10^4
F6	正常状态	1×10^4

3.1.1 故障样本预处理

将收集到的长时间序列的振动信号样本经过 Z-score 标准化后可以得到如图 4 所示的结果。其中 Z-score 标准化是将数据按比例缩放,使之落入一个特定区间。假设样本 $\mathbf{x} = (x_1, x_2, \dots, x_k, \dots)$ 的均值为 μ , 方差为 σ , 可以得到 Z-score 标准化后的样本 \mathbf{z} 的求解公式为

$$z_k = \frac{x_k - \mu}{\sigma} \quad (18)$$

为了将振动信号样本转化为深度学习模型易于处理的样本形式,本文将一维的长时间序列样本矩阵化为 20×20 的灰度图故障样本,选用 F1 的一个故障样本如图 5 所示。

为了模拟滚动轴承在实际运行环境中的噪声干扰,本文在原有样本上添加了具有不同方差的高斯白噪声。定义 $\sigma = 0.01$ 为噪声等级 1, $\sigma =$

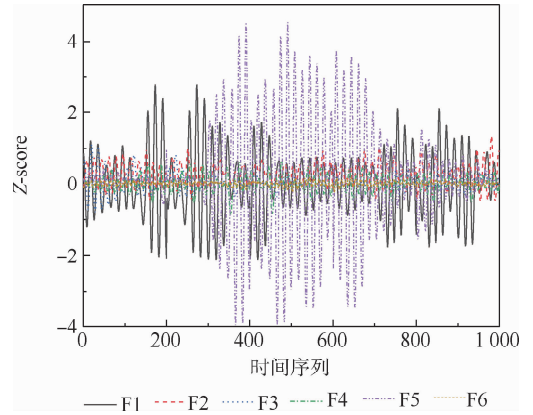


图 4 标准化后的振动信号样本

Fig. 4 Sample of standardized vibration signal

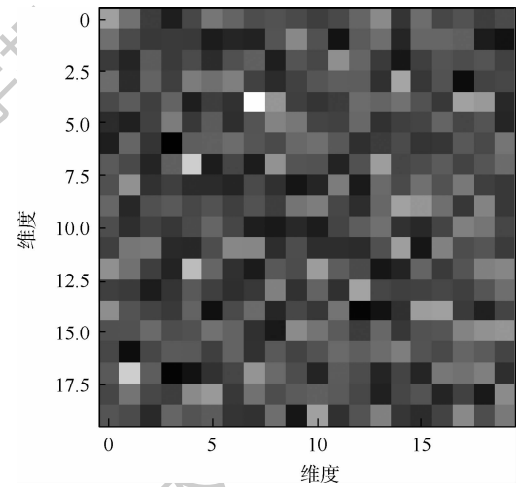


图 5 灰度图故障样本

Fig. 5 Gray image fault sample

0.02 为噪声等级 2。

通过重复间隔采样从每种类别中分别抽取 1 200 个样本构成总数为 7 200 的样本集。

3.1.2 模型的构建与训练

定义深度残差收缩网络的模型参数如图 6 所示。本文构建的深度残差收缩网络包含了 10 个隐层,其中每隔 2 个隐层添加残差项,并进行软阈值操作。其中输入的参数 (20×20 , Conv, 10) 表示输入样本的尺寸是 20×20 , 模型是 CNN, 通道数是 10。将输入层进行 Pooling 操作和 ReLU 函数处理得到尺寸为 10×10 的特征输入到隐层中,其中前 4 个隐层的参数相同,在第 4 个隐层后通过 Pooling 将特征尺寸压缩到 5×5 ,同时通道数增加到 20。在第 10 个隐层,经过 Pooling 操作和 ReLU 函数处理后,将多个通道的特征图展开并输入到全连接层,然后通过 Softmax 函数得到故障分类结果。

定义每个批次包含 32 个样本,训练 50 个周期(Epoch)可以得到训练均方误差(Mean Square

Error, MSE) 和训练过程中训练集正确率的变化过程,如图 7 所示。经过 50 个 Epoch,训练误差从 0.58 降低到 0.019,分类正确率从 79.4% 增加到 99.5%。

将噪声等级 1 和噪声等级 2 的样本集作为测试集代入到训练好的模型中进行故障分类可以得到测试结果,如图 8 所示。其中噪声等级 1(N1)和噪声等级 2(N2)的样本集故障诊断正确率分别为 98.6% 和 97.1%。可以看出,在添加噪声后,只有 F3 和 F4 的故障诊断出现一些偏差。考虑到 F3 和 F4 都属于外圈故障,其故障机理有高度相似性,因此可以认为该模型在噪声干扰下得到了较高的故障诊断正确率,能够用于指导实际工况下的滚动轴承故障诊断。

将深度残差收缩网络模型与深度残差网络和不含残差收缩功能的 CNN 模型相比较,其训练过程中的误差变化如图 9 所示。DRSN、DRN 和 CNN

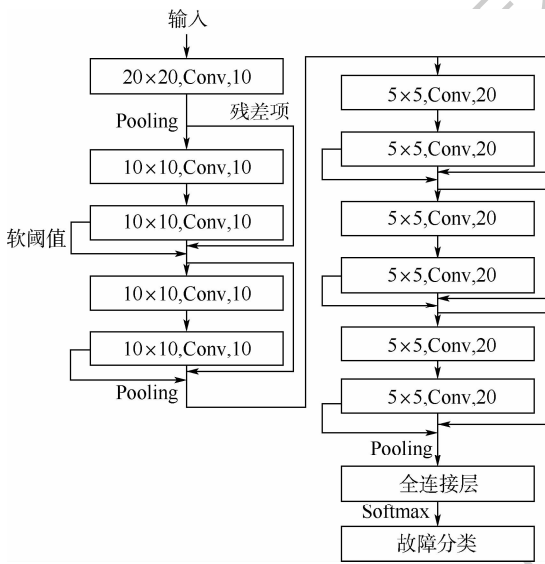


图 6 深度残差收缩网络结构

Fig. 6 Structure of deep residual shrinkage network

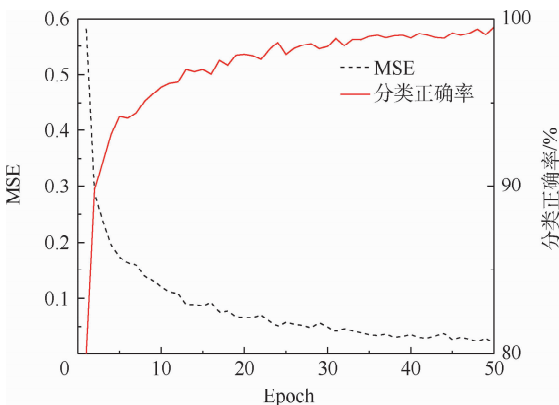
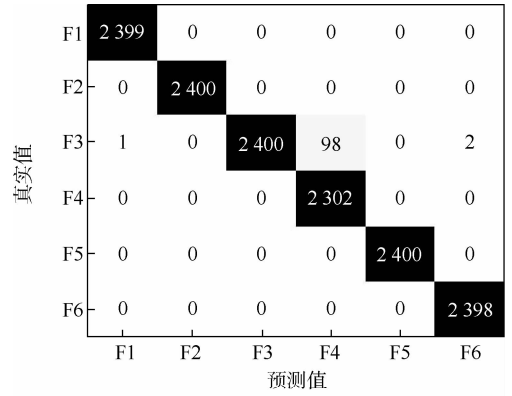
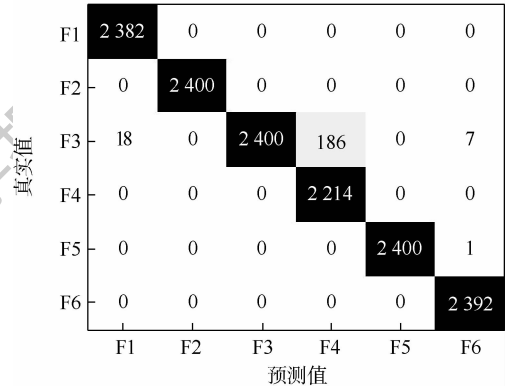


图 7 训练误差与分类正确率变化

Fig. 7 Change of training error and classification accuracy



(a) 噪声等级 1



(b) 噪声等级 2

图 8 样本故障诊断结果

Fig. 8 Fault diagnosis results of samples

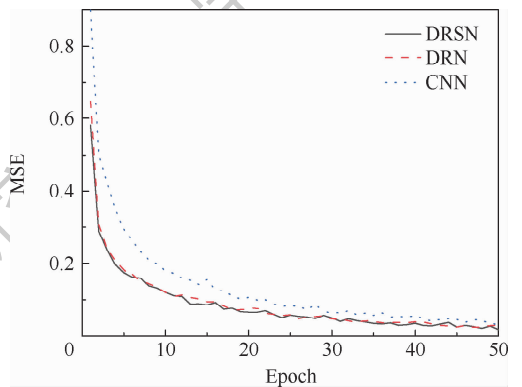


图 9 训练误差对比

Fig. 9 Comparison of training error

在第 50 个 Epoch 的训练误差分别为 0.018 7、0.029 8 和 0.031 2。可以看出,DRSN 模型和 DRN 模型比 CNN 模型的误差下降更快,DRSN 有最低的训练误差。

3.2 基于全寿命周期样本的滚动轴承故障诊断

本节从西南交通大学的轴承数据集中选取了在 37.5 Hz/11 kN 的运行条件下的外圈故障和保持架故障作为研究对象^[20]。其中外圈和保持架分别从正常到故障运行了 5 小时 39 分钟和 8 小时 53 分钟,并在 339 个和 533 个采样点采集到了

水平振动信号和垂直信号数据集。为了方便模型处理,选取2种故障发生前的200个采样点以内的振动信号数据,将每个采样点对应的长时间序列振动信号按照一定长度分为多个小样本,并从所有采样点随机选取小样本,将其对应振动信号的振幅作为指标构造包含200维的水平振动信号振幅和200维的垂直信号振幅的全寿命周期故障样本。选取2种故障的全寿命周期样本如图10所示。

在3.1节的样本预处理和模型构造的基础上,将输出节点改为2,可以构造深度残差收缩网络用于全寿命周期样本的故障诊断。在经过50个Epoch后,可以得到其训练误差降低到0.005,故障诊断正确率也达到了99.5%。该实验说明在经过一定的数据集处理后,通过深度残差收缩网络能够准确地判断特征指标变化曲线对应的故障类别,从而实现基于滚动轴承振动信号的精准化故障诊断。

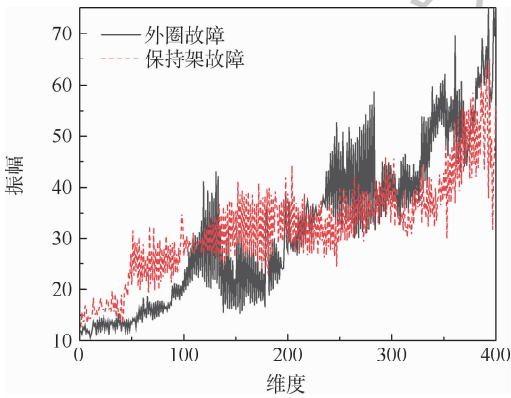


图10 滚动轴承全寿命周期故障样本

Fig. 10 Life cycle failure sample of rolling bearing

3.3 对比实验

为了全面验证本文方法的有效性,构建了包含10个隐层的深度信念网络(Deep Belief Network, DBN)和人工神经网络(Artificial Neural Network, ANN),其隐层节点为[360, 320, 280, 240, 200, 160, 120, 80, 40, 20]。同时,还构建了传统的支持向量机(Support Vector Machine, SVM)实现从400维的样本到输出节点为6的数据集1和输出节点为2的数据集2的故障分类。对比模型在2种故障样本集的诊断正确率如表2所示。可以看出,DRSN在不同数据集和噪声等级下的诊断正确率都最高,与其他深度学习模型和传统模型,其平均正确率提高了1%~6%。

表2 故障诊断正确率对比

Table 2 Comparison of fault diagnosis accuracy

模型	数据集1			数据集2			平均 值
	原始	N1	N2	原始	N1	N2	
	%						
DRSN	99.2	98.6	97.1	99.5	99.1	98.7	98.7
DRN	98.2	97.8	96.1	98.8	98.2	97.5	97.8
CNN	97.4	96.2	94.3	98.1	97.7	97.3	96.8
DBN	96.5	94.7	93.8	97.9	96.7	95.8	95.9
SVM	90.8	88.2	85.4	95.6	93.4	91.5	90.8
ANN	91.3	89.4	87.6	96.9	95.1	94.7	92.5

4 结论

1) 针对多故障下的长时间序列滚动轴承振动数据,将滚动轴承的一维振动信号数据矩阵化为多维度的灰度图故障样本,从而便于深度学习模型训练。

2) 针对从正常状态到故障的滚动轴承性能退化过程,通过多个采样点的随机采样,构造全寿命周期的故障样本用于故障诊断。

3) 使用多层卷积网络模型框架用于滚动轴承故障诊断,能够充分挖掘长时间序列多故障类型的振动信号数据的深层次特征。

4) 将残差项加入到模型中进行训练,降低了多层训练过程中的样本特征损失,从而解决了多层深度学习模型的网络退化问题。

5) 将软阈值化和注意力机制用于网络模型的特征激活,从而针对不同等级的含噪声样本设定不同的阈值,实现样本的降噪处理。

为了提高模型的鲁棒性,得到更准确的故障诊断结果,需要选取更多工况下的样本集用于模型训练,并探究多种深度学习模型框架添加残差收缩模块的可能性。

参考文献 (References)

- [1] LEI Y G, YANG B, JIANG X W, et al. Applications of machine learning to machine fault diagnosis: A review and roadmap[J]. Mechanical Systems and Signal Processing, 2020, 138: 106587.
- [2] 余建波, 吕靖香, 程辉, 等. 基于ITD和改进形态滤波的滚动轴承故障诊断[J]. 北京航空航天大学学报, 2018, 44(2): 241-249.
YU J B, LV J X, CHENG H, et al. Fault diagnosis for rolling bearing based on ITD and improved morphological filter[J]. Journal of Beijing University of Aeronautics and Astronautics, 2018, 44(2): 241-249 (in Chinese).
- [3] 丁显, 徐进, 滕伟, 等. 基于无参数经验小波变换的风电齿轮箱故障特征提取[J]. 振动与冲击, 2020, 39(8): 99-105.
DING X, XU J, TENG W, et al. Fault feature extraction of a wind turbine gearbox using adaptive parameterless empirical

- wavelet transform[J]. *Journal of Vibration and Shock*, 2020, 39(8):99-105 (in Chinese).
- [4] WANG L M, SHAO Y M. Fault feature extraction of rotating machinery using a reweighted complete ensemble empirical mode decomposition with adaptive noise and demodulation analysis[J]. *Mechanical Systems and Signal Processing*, 2020, 138: 106545.
- [5] 李政, 张伟, 明安波, 等. 基于 IEWT 和 MCKD 的滚动轴承故障诊断方法[J]. *机械工程学报*, 2019, 55(23):136-146.
LI Z, ZHANG W, MING A B, et al. A novel fault diagnosis method based on improved empirical wavelet transform and maximum correlated kurtosis deconvolution for rolling element bearing[J]. *Journal of Mechanical Engineering*, 2019, 55(23): 136-146 (in Chinese).
- [6] 曹惠玲, 高升, 薛鹏. 基于多分类 AdaBoost 的航空发动机故障诊断[J]. *北京航空航天大学学报*, 2018, 44(9):1818-1825.
CAO H L, GAO S, XUE P. Aeroengine fault diagnosis based on multi-classification AdaBoost[J]. *Journal of Beijing University of Aeronautics and Astronautics*, 2018, 44(9):1818-1825 (in Chinese).
- [7] 崔路瑶. 基于机器学习的滚动轴承智能故障诊断方法研究[D]. 南昌: 华东交通大学, 2019:1-77.
CUI L Y. Research on fault diagnosis method of rolling bearing based on machine learning[D]. Nanchang: East China Jiaotong University, 2019:1-77 (in Chinese).
- [8] 车畅畅, 王华伟, 倪晓梅, 等. 基于深度学习的航空发动机故障融合诊断[J]. *北京航空航天大学学报*, 2018, 44(3):621-628.
CHE C C, WANG H W, NI X M, et al. Fault fusion diagnosis of aero-engine based on deep learning[J]. *Journal of Beijing University of Aeronautics and Astronautics*, 2018, 44(3):621-628 (in Chinese).
- [9] 宫文峰, 陈辉, 张泽辉, 等. 基于改进卷积神经网络的滚动轴承智能故障诊断研究[J]. *振动工程学报*, 2020, 33(2):400-413.
GONG W F, CHEN H, ZHANG Z H, et al. Intelligent fault diagnosis for rolling bearing based on improved convolutional neural network[J]. *Journal of Vibration Engineering*, 2020, 33(2):400-413 (in Chinese).
- [10] 朱丹宸, 张永祥, 潘洋洋, 等. 基于多传感器信号和卷积神经网络滚动轴承故障诊断[J]. *振动与冲击*, 2020, 39(4):172-178.
ZHU D C, ZHANG Y X, PAN Y Y, et al. Fault diagnosis for rolling element bearings based on multi-sensor signals and CNN[J]. *Journal of Vibration and Shock*, 2020, 39(4):172-178 (in Chinese).
- [11] QIU G Q, GU Y K, CAI Q. A deep convolutional neural networks model for intelligent fault diagnosis of a gearbox under different operational conditions[J]. *Measurement*, 2019, 145: 94-107.
- [12] SUN M D, WANG H, LIU P, et al. A sparse stacked denoising autoencoder with optimized transfer learning applied to the fault diagnosis of rolling bearings[J]. *Measurement*, 2019, 146: 305-314.
- [13] CHE C C, WANG H W, NI X M, et al. Domain adaptive deep belief network for rolling bearing fault diagnosis[J]. *Computers & Industrial Engineering*, 2020, 143:106427.
- [14] ZHANG Y, XING K S, BAI R X, et al. An enhanced convolutional neural network for bearing fault diagnosis based on time-frequency image[J]. *Measurement*, 2020, 157:107667.
- [15] LIANG P F, DENG C, WU J, et al. Compound fault diagnosis of gearboxes via multi-label convolutional neural network and wavelet transform[J]. *Computers in Industry*, 2019, 113: 103132.
- [16] ZHAO M, ZHONG S, FU X, et al. Deep residual shrinkage networks for fault diagnosis[J]. *IEEE Transactions on Industrial Informatics*, 2020, 16(7):4681-4690.
- [17] QIU W Y, LI D X, JIN X Y, et al. Deep neural network inspired by iterative shrinkage-thresholding algorithm with data consistency (NISTAD) for fast undersampled MRI reconstruction[J]. *Magnetic Resonance Imaging*, 2020, 70:134-144.
- [18] OCAK H, LOPARO K A. HMM-based fault detection and diagnosis scheme for rolling element bearings[J]. *Journal of Vibration and Acoustics*, 2005, 127:299-306.
- [19] SMITH W A, RANDALL R B. Rolling element bearing diagnostics using the Case Western Reserve University data; A benchmark study[J]. *Mechanical Systems and Signal Processing*, 2015, 64-65:100-131.
- [20] WANG B, LEI Y, LI N, et al. A hybrid prognostics approach for estimating remaining useful life of rolling element bearings[J]. *IEEE Transactions on Reliability*, 2020, 69:401-412.

Fault diagnosis of rolling bearing based on deep residual shrinkage network

CHE Changchang, WANG Huawei*, NI Xiaomei, LIN Ruiguan

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

Abstract: Accurate fault diagnosis of rolling bearing is a necessary means to ensure the safe and reliable operation of mechanical equipment. In this paper, a fault diagnosis method based on Deep Residual Shrinkage Network (DRSN) is proposed for the vibration signal of rolling bearing with multiple faults and long time series. Firstly, fault samples are constructed according to the collected rolling bearing data. For the vibration signals of long time series under various fault types, the long time series are matrixed according to a certain size, so as to form the gray image fault samples of multiple fault types. Aimed at the performance degradation process of rolling bearings from normal to fault, the whole life cycle fault samples are constructed for fault diagnosis through random sampling of multiple sampling points. Secondly, based on the multi-layer deep learning model, the residual shrinkage network module is added to the Convolutional Neural Network (CNN) to build the deep residual shrinkage network model, in which the model degradation problem of the multi-layer network model is solved by adding the residual term to the network training, and the sample noise reduction is realized by using soft thresholding. Finally, in order to verify the effectiveness of the proposed method, multi-fault time series samples and life cycle fault samples of rolling bearing are collected for fault diagnosis. The result of the example shows that the proposed model has good robustness under the noise interference, there is no obvious network degradation under the multi-layer network model, and it can maintain a high accuracy of fault diagnosis. When dealing with two kinds of bearing fault datasets, compared with other models, this method has lower training error, and the average accuracy of fault classification is increased by 1% - 6%.

Keywords: rolling bearing; fault diagnosis; Deep Residual Shrinkage Network (DRSN); Convolutional Neural Network (CNN); soft thresholding

Received: 2020-05-18; **Accepted:** 2020-07-10; **Published online:** 2020-07-21 13:57

URL: kns.cnki.net/kcms/detail/11.2625.V.20200721.0931.001.html

Foundation item: National Natural Science Foundation of China (U1833110)

* **Corresponding author.** E-mail: wang_hw66@163.com

http://bhxb.buaa.edu.cn jbuua@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2020.0195

高精度在轨实时轨道机动决策



解树聪, 董云峰*

(北京航空航天大学 宇航学院, 北京 100083)

摘 要: 为保证在轨机动实时性和高精度的要求,提出了一种基于机器学习的在轨实时机动决策方法。通过优化算法离线获得摄动下的精确解,减去二体解得到速度增量差,将其投影到轨道坐标系获得速度增量摄动修正项,以此作为神经网络输出,设计网络参数并训练得到摄动修正网络、组合应用摄动修正网络和二体解实现高精度的在轨实时轨道机动决策。仿真结果表明:卫星按照该决策机动完成后的终端位置偏差与按照优化算法给出的决策机动完成后终端位置偏差精度一致,且前者决策耗时仅为后者决策耗时的0.01%左右。所提轨道机动决策方法兼顾了精度与实时性,适用于星上决策。

关键词: 轨道机动; 神经网络; 机器学习; Lambert 机动; 摄动修正

中图分类号: V448.23

文献标志码: A

文章编号: 1001-5965(2021)07-1407-07

高精度的轨道机动决策是空间交会、在轨监视与捕获等空间任务的关键技术^[1-4]。空间中一点到另一点固定时间的机动是轨道机动的基础问题。在二体模型条件下,该问题可通过 Gauss 法^[5]、Bate 法^[5]等快速求解,实际在轨机动要考虑摄动因素,直接应用二体解会产生较大的末端位置偏差,通过迭代优化的方式可以获得考虑摄动影响的高精度解^[6-10]。然而,迭代优化运算量大,不利于在轨自主实时解算。

近年来,机器学习不断发展,在实时控制方面展现出优势^[11-13]。Sánchez-Sánchez 和 Izzo^[14]的研究表明深度学习技术可以应用于航天器导航和控制的星上决策。在行星际任务^[15]和月球着陆^[16]方面,神经网络也得到应用,通过优化等间接方法获得精确解,训练网络用于实时决策。

对于高精度在轨机动问题,直接对问题的精确解进行训练得到的神经网络决策解精度不够。考虑将二体运动规律滤除,只对摄动引起的复杂规律进行训练学习。据此,本文提出了一种基于

机器学习的在轨实时轨道机动决策方法,使得决策精度与优化算法精度一致,且耗时大大缩短。

1 轨道机动问题

本文的轨道机动问题可描述为:给定卫星初始时刻的位置 \mathbf{r}_1 和速度 \mathbf{v}_1 ,要求经过 Δt 飞行时间,卫星达到终端时刻的位置 \mathbf{r}_2 和速度 \mathbf{v}_2 ,确定出卫星在初始时刻所需的速度增量 $\Delta \mathbf{v}_1$ 。

在航天工程应用中,往往会根据复杂约束确定标称工况,对于上述问题的标称工况,考虑摄动的影响,上述问题通常采用迭代优化的方法求解^[6]。

已知卫星的运动满足如下微分方程^[17]:

$$\begin{aligned} d\mathbf{r} &= \dot{\mathbf{r}} \\ d\dot{\mathbf{r}} &= -\frac{\mu}{|\mathbf{r}|^3}\mathbf{r} + \frac{\mathbf{F}_p}{m} \end{aligned} \quad (1)$$

式中: μ 为地心引力常数; \mathbf{F}_p 为各项摄动力,包括地球非球型引力、日月引力、太阳光压、大气阻力等^[17]; m 为卫星质量。

在初始时刻,对卫星施加速度增量 $\Delta \mathbf{v}_1$,则上

收稿日期: 2020-05-20; 录用日期: 2020-08-14; 网络出版时间: 2020-09-15 11:27

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20200915.0959.002.html

* 通信作者。E-mail: sinosat@buaa.edu.cn

引用格式: 解树聪, 董云峰. 高精度在轨实时轨道机动决策[J]. 北京航空航天大学学报, 2021, 47(7): 1407-1413.

XIE S C, DONG Y F. High-precision on-orbit real-time orbital maneuver decision[J]. Journal of Beijing University of Aeronautics and Astronautics, 2021, 47(7): 1407-1413 (in Chinese).

述问题的边界条件为

$$\begin{cases} \mathbf{r}(0) = \mathbf{r}_1 \\ d\mathbf{r}(0) = \mathbf{v}_1 + \Delta\mathbf{v}_1 \end{cases} \quad (2)$$

卫星飞行 Δt 时间后到达位置 $\mathbf{r}(\Delta t)$, $\mathbf{r}(\Delta t)$ 可通过求解微分方程式(1)得到,则终端位置偏差为

$$d_i = |\mathbf{r}(\Delta t) - \mathbf{r}_2| \quad (3)$$

于是, $\Delta\mathbf{v}_1$ 可通过求解下述优化问题得到:

Find: $\Delta\mathbf{v}_1$

$$\min d_i = |\mathbf{r}(\Delta t) - \mathbf{r}_2|$$

$$\text{s. t. } \begin{cases} d\mathbf{r} = \dot{\mathbf{r}}, d\dot{\mathbf{r}} = -\frac{\mu}{|\mathbf{r}|^3}\mathbf{r} + \frac{\mathbf{F}_p}{m} \\ \mathbf{r}(0) = \mathbf{r}_1, d\mathbf{r}(0) = \mathbf{v}_1 + \Delta\mathbf{v}_1 \end{cases}$$

2 机器学习方法

在实际工程应用中,由于不确定性因素的存在,会在标称工况周围形成一个可行解域,即实际初始位置速度和终端位置速度会在一定范围内变动。在该范围内采样并经优化算法得到每个样本点的解,训练神经网络,实现该范围内任一点的轨道机动问题的实时高精度求解。

2.1 初始采样

给定样本中心点的输入向量为

$$\mathbf{X}_0 = (\mathbf{r}_1^T, \mathbf{v}_1^T, \mathbf{r}_2^T, \mathbf{v}_2^T)^T = (r_{1x0}, r_{1y0}, r_{1z0}, v_{1x0}, v_{1y0}, v_{1z0}, r_{2x0}, r_{2y0}, r_{2z0}, v_{2x0}, v_{2y0}, v_{2z0})^T$$

给定样本点输入向量到样本中心点输入向量各维度的最远距离 $\delta r_{1x}, \delta r_{1y}, \delta r_{1z}, \delta v_{1x}, \delta v_{1y}, \delta v_{1z}, \delta r_{2x}, \delta r_{2y}, \delta r_{2z}, \delta v_{2x}, \delta v_{2y}, \delta v_{2z}$, 可以得到初始位置速度和终端位置速度各分量的变动范围,即样本输入向量的取值范围 $\mathbf{X}_{\min} \sim \mathbf{X}_{\max}$ 。

$$\mathbf{X}_{\min} = (r_{1x0} - \delta r_{1x}, r_{1y0} - \delta r_{1y}, r_{1z0} - \delta r_{1z}, v_{1x0} - \delta v_{1x}, v_{1y0} - \delta v_{1y}, v_{1z0} - \delta v_{1z}, r_{2x0} - \delta r_{2x}, r_{2y0} - \delta r_{2y}, r_{2z0} - \delta r_{2z}, v_{2x0} - \delta v_{2x}, v_{2y0} - \delta v_{2y}, v_{2z0} - \delta v_{2z})^T$$

$$\mathbf{X}_{\max} = (r_{1x0} + \delta r_{1x}, r_{1y0} + \delta r_{1y}, r_{1z0} + \delta r_{1z}, v_{1x0} + \delta v_{1x}, v_{1y0} + \delta v_{1y}, v_{1z0} + \delta v_{1z}, r_{2x0} + \delta r_{2x}, r_{2y0} + \delta r_{2y}, r_{2z0} + \delta r_{2z}, v_{2x0} + \delta v_{2x}, v_{2y0} + \delta v_{2y}, v_{2z0} + \delta v_{2z})^T$$

在上述范围中,对于每个维度,按照均匀分布采样一次得到一个样本输入向量 $\mathbf{X} = (X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9, X_{10}, X_{11}, X_{12})^T$ 。其中, $X_i \sim U((\mathbf{X}_{\min})_i, (\mathbf{X}_{\max})_i)$ ($i=1, 2, \dots, 12$), X_i 表示 \mathbf{X} 第 i 维度取值, $(\mathbf{X}_{\min})_i$ 表示 \mathbf{X}_{\min} 第 i 维度取值, $(\mathbf{X}_{\max})_i$ 表示 \mathbf{X}_{\max} 第 i 维度取值。

给定样本个数,按照上述方式完成采样。给定飞行时间 Δt ,对每一个样本点输入向量通过优化算法得到样本输出向量,即摄动精确解 $\Delta\mathbf{v}_1$ 。

2.2 摄动解与二体解作差

在二体模型条件下,通过 Bate 法可快速求得二体解 $\Delta\mathbf{v}_{1TB}$ 。

首先,计算转移角度 $\Delta\theta$:

$$\Delta\theta = \arccos\left(\frac{\mathbf{r}_1 \mathbf{r}_2}{|\mathbf{r}_1| |\mathbf{r}_2|}\right) \quad (4)$$

然后,计算常数 A :

$$A = \frac{\sqrt{|\mathbf{r}_1| |\mathbf{r}_2|} \sin \Delta\theta}{\sqrt{1 - \cos \Delta\theta}} \quad (5)$$

之后,通过牛顿迭代法求解飞行时间 Δt 的普适方程:

$$\Delta t = \frac{1}{\sqrt{\mu}} \left[\left(\sqrt{\frac{y(z)}{C(z)}} \right)^3 S(z) + A \sqrt{y(z)} \right] \quad (6)$$

$$\text{式中: } S(z) = \frac{\sqrt{z} - \sin \sqrt{z}}{\sqrt{z^3}}; C(z) = \frac{1 - \cos \sqrt{z}}{z};$$

$$y(z) = |\mathbf{r}_1| + |\mathbf{r}_2| - A \frac{zS(z) - 1}{\sqrt{C(z)}}。$$

解得 z 之后,计算拉格朗日系数 f 和 g :

$$\begin{cases} f = 1 - \frac{y(z)}{|\mathbf{r}_1|} \\ g = A \sqrt{\frac{y(z)}{\mu}} \end{cases} \quad (7)$$

于是,得到二体模型下所需的速度增量 $\Delta\mathbf{v}_{1TB}$:

$$\Delta\mathbf{v}_{1TB} = \frac{\mathbf{r}_2 - f\mathbf{r}_1}{g} - \mathbf{v}_1 = f_{\text{Bate}}(\mathbf{r}_1, \mathbf{v}_1, \mathbf{r}_2, \mathbf{v}_2, \Delta t) \quad (8)$$

式中: $f_{\text{Bate}}(\cdot)$ 为将 Bate 法求解速度增量的过程函数。

摄动解与二体解作差得到摄动项引起的速度增量差为

$$\Delta\Delta\mathbf{v}_1 = \Delta\mathbf{v}_1 - \Delta\mathbf{v}_{1TB} \quad (9)$$

2.3 坐标变换

已知卫星初始时刻位置 \mathbf{r}_1 和速度 \mathbf{v}_1 , 则

$$(\mathbf{k}_o)_i = -\frac{\mathbf{r}_1}{|\mathbf{r}_1|} \quad (10)$$

$$\mathbf{H} = \mathbf{r}_1 \times \mathbf{v}_1 \quad (11)$$

$$(\mathbf{j}_o)_i = -\frac{\mathbf{H}}{|\mathbf{H}|} \quad (12)$$

$$(\mathbf{i}_o)_i = (\mathbf{j}_o)_i \times (\mathbf{k}_o)_i \quad (13)$$

式中: \mathbf{H} 为动量矩矢量; \mathbf{i}_o 、 \mathbf{j}_o 和 \mathbf{k}_o 分别为轨道坐标系 x 、 y 和 z 的单位矢量。

于是,地心赤道惯性坐标系到质心轨道坐标系的坐标变换矩阵为

$$\mathbf{R}_{oi} = \begin{bmatrix} (\mathbf{i}_o)_i^T \\ (\mathbf{j}_o)_i^T \\ (\mathbf{k}_o)_i^T \end{bmatrix} \quad (14)$$

进而得到速度增量差在轨道坐标系下的表示

形式:

$$(\Delta\Delta\mathbf{v}_1)_o = \mathbf{R}_{oi} \cdot \Delta\Delta\mathbf{v}_1 \quad (15)$$

至此,得到最终的样本输出速度增量摄动修正项 $(\Delta\Delta\mathbf{v}_1)_o$ 。

2.4 神经网络设计

设计摄动修正网络 $N_{\text{HpopCorrect}}$,其输入为 $\mathbf{r}_1, \mathbf{v}_1, \mathbf{r}_2, \mathbf{v}_2$ 在惯性坐标系的三轴分量形式,即 $\mathbf{X} = (r_{1x}, r_{1y}, r_{1z}, v_{1x}, v_{1y}, v_{1z}, r_{2x}, r_{2y}, r_{2z}, v_{2x}, v_{2y}, v_{2z})^T$,输出为速度增量差在轨道坐标系的三轴分量形式,即 $\mathbf{Y} = ((\Delta\Delta v_{1x})_o, (\Delta\Delta v_{1y})_o, (\Delta\Delta v_{1z})_o)^T$ 。

对网络包含的隐层层数、层内神经元节点数、每层的激活函数进行遍历,假设隐层层数有 L_n 个取值,层内神经元节点数有 N_n 个取值,激活函数有 A_n 个取值,则获得 $L_n \times N_n \times A_n$ 个网络,对比获得性能最优的网络。网络性能指标采用均方误差(Mean Squared Error, MSE)。

$$\text{MSE} = \frac{1}{MD_{\text{OUT}}} \sum_{j=1}^M \sum_{k=1}^{D_{\text{OUT}}} (y_{j,k} - \hat{y}_{j,k})^2 \quad (16)$$

式中: M 为测试样本个数; D_{OUT} 为样本输出维度,此处为3; $y_{j,k}$ 为第 j 个测试样本第 k 个维度的输出真值; $\hat{y}_{j,k}$ 为第 j 个测试样本第 k 个维度的网络预测输出值。

2.5 应用网络进行轨道机动决策

如图1所示,轨道机动决策流程包括了如下步骤:

- 步骤1 $\forall \mathbf{X} = (\mathbf{r}_1^T, \mathbf{v}_1^T, \mathbf{r}_2^T, \mathbf{v}_2^T)^T \in \mathbf{X}_{\min} \sim \mathbf{X}_{\max}$,
 $(\Delta\Delta\mathbf{v}_1)_o = N_{\text{HpopCorrect}}(\mathbf{r}_1, \mathbf{v}_1, \mathbf{r}_2, \mathbf{v}_2)$ 。
- 步骤2 $\Delta\Delta\mathbf{v}_1 = \mathbf{R}_{oi}^{-1} \cdot (\Delta\Delta\mathbf{v}_1)_o$ 。
- 步骤3 $\Delta\mathbf{v}_{\text{TB}} = f_{\text{Bate}}(\mathbf{r}_1, \mathbf{v}_1, \mathbf{r}_2, \mathbf{v}_2, \Delta t)$ 。
- 步骤4 $\Delta\mathbf{v}_1 = \Delta\Delta\mathbf{v}_1 + \Delta\mathbf{v}_{\text{TB}}$ 。

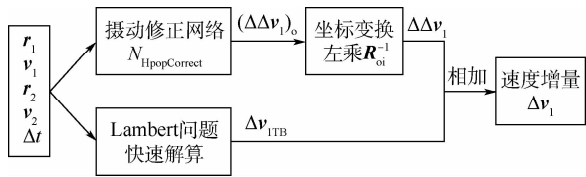


图1 轨道机动决策流程

Fig. 1 Decision-making process for orbital maneuvers

3 仿真验证

选定样本中心点为 $\mathbf{X}_0 = (5\ 846\ 591.286\ 6, -1\ 761\ 003.240\ 6, 3\ 479\ 801.993\ 5, 3\ 176.387\ 9, -2\ 216.637\ 1, -6\ 458.570\ 7, 5\ 432\ 239.472\ 1, -2\ 636\ 976.590\ 5, -3\ 602\ 112.081\ 7, -3\ 952.612\ 7, 604.677\ 3, -6\ 351.835\ 6)^T$,选定飞行时间 Δt 为1 000 s。

选定样本点到样本中心点各维度的最远距离如表1所示。

表1 样本点到样本中心点最远距离

Table 1 The farthest distance from sample point to sample center point

参数	数值
$\delta r_{1x}/\text{m}$	1 000
$\delta r_{1y}/\text{m}$	1 000
$\delta r_{1z}/\text{m}$	1 000
$\delta v_{1x}/(\text{m} \cdot \text{s}^{-1})$	1
$\delta v_{1y}/(\text{m} \cdot \text{s}^{-1})$	1
$\delta v_{1z}/(\text{m} \cdot \text{s}^{-1})$	1
$\delta r_{2x}/\text{m}$	1 000
$\delta r_{2y}/\text{m}$	1 000
$\delta r_{2z}/\text{m}$	1 000
$\delta v_{2x}/(\text{m} \cdot \text{s}^{-1})$	1
$\delta v_{2y}/(\text{m} \cdot \text{s}^{-1})$	1
$\delta v_{2z}/(\text{m} \cdot \text{s}^{-1})$	1

于是,样本取值范围为 $\mathbf{X}_{\min} \sim \mathbf{X}_{\max}$ 。

$$\mathbf{X}_{\min} = (5\ 845\ 591.286\ 6, -1\ 762\ 003.240\ 6, 3\ 478\ 801.993\ 5, 3\ 175.387\ 9, -2\ 217.637\ 1, -6\ 459.570\ 7, 5\ 431\ 239.472\ 1, -2\ 637\ 976.590\ 5, -3\ 603\ 112.081\ 7, -3\ 953.612\ 7, 603.677\ 3, -6\ 352.835\ 6)^T$$

$$\mathbf{X}_{\max} = (5\ 847\ 591.286\ 6, -1\ 760\ 003.240\ 6, 3\ 480\ 801.993\ 5, 3\ 177.387\ 9, -2\ 215.637\ 1, -6\ 457.570\ 7, 5\ 433\ 239.472\ 1, -2\ 635\ 976.590\ 5, -3\ 601\ 112.081\ 7, -3\ 951.612\ 7, 605.677\ 3, -6\ 350.835\ 6)^T$$

在上述范围 $\mathbf{X}_{\min} \sim \mathbf{X}_{\max}$ 内按均匀分布($X_i \sim U((\mathbf{X}_{\min})_i, (\mathbf{X}_{\max})_i) i = 1, 2, \dots, 12$)随机采取6 000个样本点,每个样本点经优化获得摄动下的精确解 $\Delta\mathbf{v}_1$,作为样本的初始输出。摄动精确解与二体解作差,再经坐标变换得到最终输出。

为比较不同输出网络的性能,设计3种神经网络,分别为速度增量网络 N_{Dir} 、摄动偏差网络 N_{DeTB} 、摄动修正网络 $N_{\text{HpopCorrect}}$ 。其中,速度增量网络以速度增量 $\Delta\mathbf{v}_1$ 为网络输出,摄动偏差网络以速度增量差 $\Delta\Delta\mathbf{v}_1$ 为网络输出,摄动修正网络以速度增量摄动修正项 $(\Delta\Delta\mathbf{v}_1)_o$ 作为网络输出。

对3种网络的结构参数分别进行优化设计。隐层层数和层内神经元节点数取值如表2所示。

表2 网络结构参数取值

Table 2 Parameter value of network structure

神经网络参数	数值
网络隐层数	1,2,3,4
隐层节点数	8,16,32,64,128

隐层激活函数选取最常见的几种,分别是logsig[0, 1], softmax[0, 1], poslin[0, 1], purelin[-∞, ∞], tansig[-1, 1]。输出层激活函数取purelin[-∞, ∞]。网络训练最大迭代次数

取5 000次,梯度下降下限取 1.0×10^{-5} ,误差最大连续不下降次数取 50 次。

在 6 000 个样本数据中,取前 5 400 个样本进行网络训练,后 600 个样本进行测试。

1) 直接对速度增量进行训练。统计不同参数网络的 MSE 记录在表 3 中。由表 3 可以看出,

直接对速度增量进行训练,最优的速度增量网络包含 3 层隐层,隐层有 16 个神经元节点,隐层激活函数为 purelin。将 600 个测试样本输入到最优速度增量网络,得到对应的速度增量 Δv_{1NDir} 。

2) 对速度增量的摄动因素进行训练。统计不同参数网络的 MSE 记录在表 4 中。由表 4 可以

表 3 不同层数、节点数和激活函数的速度增量网络的性能

Table 3 Performance of speed increment networks with different layers, units and activation functions

隐层层数/节点数	MSE				
	logsig	softmax	poslin	purelin	tansig
1/8	9.95×10^{-6}	8.38×10^{-5}	8.23×10^{-9}	2.06×10^{-9}	5.76×10^{-6}
1/16	3.05×10^{-6}	1.81×10^{-5}	5.42×10^{-6}	2.02×10^{-9}	1.45×10^{-6}
1/32	6.40×10^{-6}	2.20×10^{-6}	2.17×10^{-9}	2.02×10^{-9}	3.83×10^{-6}
1/64	6.41×10^{-6}	7.47×10^{-6}	2.73×10^{-7}	2.01×10^{-9}	4.27×10^{-6}
1/128	2.46×10^{-4}	2.62×10^{-6}	7.91×10^{-6}	2.02×10^{-9}	2.97×10^{-6}
2/8	2.26×10^{-5}	5.98×10^{-5}	2.86×10^{-5}	2.02×10^{-9}	3.10×10^{-5}
2/16	5.66×10^{-6}	2.69×10^{-5}	3.86×10^{-6}	2.02×10^{-9}	3.70×10^{-6}
2/32	2.60×10^{-6}	4.37×10^{-6}	6.06×10^{-5}	2.02×10^{-9}	8.53×10^{-6}
2/64	3.33×10^{-4}	1.54×10^{-6}	5.09×10^{-4}	2.02×10^{-9}	6.34×10^{-5}
2/128	1.89×10^{-3}	2.91×10^{-6}	5.29×10^{-3}	2.02×10^{-9}	9.91×10^{-4}
3/8	2.94×10^{-5}	1.18×10^{-4}	2.96×10^{-1}	2.02×10^{-9}	8.34×10^{-6}
3/16	3.23×10^{-6}	1.60×10^{-5}	4.88×10^{-5}	2.01×10^{-9}	4.08×10^{-6}
3/32	5.15×10^{-6}	1.64×10^{-5}	2.43×10^{-3}	2.03×10^{-9}	3.28×10^{-6}
3/64	4.24×10^{-4}	9.36×10^{-6}	5.23×10^{-3}	2.02×10^{-9}	1.25×10^{-4}
3/128	2.90×10^{-3}	2.93×10^{-6}	9.91×10^{-2}	2.02×10^{-9}	3.13×10^{-3}
4/8	4.02×10^{-5}	2.98×10^{-1}	2.96×10^{-1}	2.02×10^{-9}	2.57×10^{-5}
4/16	2.45×10^{-6}	3.03×10^{-1}	1.05×10^{-2}	2.02×10^{-9}	8.95×10^{-6}
4/32	5.82×10^{-6}	1.31×10^{-5}	7.84×10^{-3}	2.03×10^{-9}	9.21×10^{-6}
4/64	4.66×10^{-4}	1.13×10^{-5}	1.35×10^{-1}	2.02×10^{-9}	2.91×10^{-4}
4/128	2.65×10^{-3}	2.98×10^{-1}	1.50	2.02×10^{-9}	2.94×10^{-3}

表 4 不同层数、节点数和激活函数的摄动偏差网络的性能

Table 4 Performance of perturbation deviation networks with different layers, units and activation functions

隐层层数/节点数	MSE				
	logsig	softmax	poslin	purelin	tansig
1/8	2.80×10^{-6}	1.04×10^{-6}	1.91×10^{-6}	6.11×10^{-6}	5.50×10^{-6}
1/16	1.62×10^{-6}	1.04×10^{-6}	2.24×10^{-6}	5.87×10^{-6}	4.11×10^{-6}
1/32	3.60×10^{-6}	1.34×10^{-6}	8.77×10^{-6}	1.87×10^{-6}	4.86×10^{-6}
1/64	3.27×10^{-6}	1.19×10^{-6}	7.61×10^{-6}	2.58×10^{-6}	3.04×10^{-6}
1/128	8.65×10^{-6}	2.26×10^{-6}	8.98×10^{-6}	7.81×10^{-7}	7.07×10^{-6}
2/8	2.08×10^{-6}	1.30×10^{-6}	3.33×10^{-6}	3.96×10^{-6}	4.73×10^{-6}
2/16	3.13×10^{-6}	2.07×10^{-6}	4.83×10^{-6}	8.44×10^{-7}	4.59×10^{-6}
2/32	3.20×10^{-6}	1.78×10^{-6}	4.35×10^{-6}	3.59×10^{-7}	3.19×10^{-6}
2/64	1.86×10^{-6}	8.88×10^{-7}	5.48×10^{-6}	1.09×10^{-7}	2.16×10^{-6}
2/128	5.23×10^{-6}	5.71×10^{-7}	3.04×10^{-6}	1.52×10^{-8}	2.02×10^{-6}
3/8	6.12×10^{-6}	5.85×10^{-7}	1.61×10^{-6}	9.80×10^{-7}	2.06×10^{-6}
3/16	5.45×10^{-6}	8.64×10^{-7}	2.22×10^{-6}	4.71×10^{-7}	2.32×10^{-6}
3/32	2.44×10^{-6}	7.58×10^{-7}	3.87×10^{-6}	8.09×10^{-8}	2.74×10^{-6}
3/64	2.49×10^{-6}	2.02×10^{-6}	3.55×10^{-6}	2.98×10^{-9}	2.52×10^{-6}
3/128	2.43×10^{-6}	7.33×10^{-7}	9.31×10^{-7}	4.11×10^{-10}	2.98×10^{-6}
4/8	7.63×10^{-7}	1.52×10^{-6}	9.29×10^{-7}	1.03×10^{-6}	9.74×10^{-7}
4/16	9.66×10^{-7}	8.19×10^{-7}	3.84×10^{-6}	4.16×10^{-7}	2.02×10^{-6}
4/32	1.99×10^{-6}	1.24×10^{-6}	3.53×10^{-6}	6.47×10^{-9}	1.12×10^{-6}
4/64	1.84×10^{-6}	1.51×10^{-6}	3.57×10^{-7}	1.06×10^{-10}	1.57×10^{-6}
4/128	1.87×10^{-6}	3.99×10^{-7}	6.99×10^{-7}	8.84×10^{-12}	1.18×10^{-6}

看出,对速度增量的摄动因素进行训练,最优摄动偏差网络包含 4 层隐层,每层隐层有 128 个神经元节点,隐层激活函数为 purelin。将 600 个测试样本输入到最优摄动偏差网络,得到对应的 $\Delta\Delta\mathbf{v}_{1N_{DeTB}}$,加上对应的二体解 $\Delta\mathbf{v}_{1TB}$,得到对应的速度增量 $\Delta\mathbf{v}_{1N_{DeTB}}$ 。

3) 对速度增量摄动因素坐标变换后进行训练。统计不同参数网络的 MSE 记录在表 5 中。由表 5 知,对速度增量摄动因素坐标变换后进行训练,最优的摄动修正网络包含 4 层隐层,每层隐

层有 128 个神经元节点,隐层激活函数为 purelin。将 600 个测试样本输入到最优摄动修正网络,得到每个样本的对应输出 $(\Delta\Delta\mathbf{v}_1)_o$,经坐标变换得到对应的 $\Delta\Delta\mathbf{v}_{1N_{HpspCorrect}}$,加上对应的二体解 $\Delta\mathbf{v}_{1TB}$,得到对应的速度增量 $\Delta\mathbf{v}_{1N_{HpspCorrect}}$ 。

将 3 种最优网络得到的速度增量施加到初始时刻进行仿真,计算 Δt 时刻后卫星的终端位置偏差 d_i 。统计 600 个样本点对应 d_i 的最大值、最小值、平均值,分别记为 $d_{max}, d_{min}, d_{mean}$,结果记录在表 6 中。

表 5 不同层数,节点数和激活函数的摄动修正网络的性能

Table 5 Performance of perturbation correction networks with different layers, units and activation functions

隐层层数/节点数	MSE				
	logsig	softmax	poslin	purelin	tansig
1/8	9.44×10^{-7}	3.87×10^{-7}	3.16×10^{-6}	2.70×10^{-6}	5.14×10^{-6}
1/16	2.29×10^{-6}	9.71×10^{-7}	6.09×10^{-6}	3.01×10^{-6}	1.83×10^{-6}
1/32	1.64×10^{-6}	1.36×10^{-6}	7.63×10^{-6}	3.61×10^{-6}	4.02×10^{-6}
1/64	2.87×10^{-6}	7.55×10^{-7}	4.58×10^{-6}	3.67×10^{-6}	9.22×10^{-6}
1/128	7.44×10^{-6}	2.81×10^{-7}	6.37×10^{-6}	5.56×10^{-7}	7.38×10^{-6}
2/8	1.24×10^{-6}	2.06×10^{-6}	2.71×10^{-6}	2.29×10^{-6}	1.21×10^{-6}
2/16	9.09×10^{-7}	6.05×10^{-7}	4.78×10^{-6}	2.10×10^{-6}	1.84×10^{-6}
2/32	2.73×10^{-6}	1.26×10^{-6}	1.99×10^{-6}	1.02×10^{-6}	4.42×10^{-6}
2/64	1.89×10^{-6}	1.08×10^{-6}	5.50×10^{-6}	6.50×10^{-8}	2.95×10^{-6}
2/128	4.18×10^{-6}	9.30×10^{-7}	5.67×10^{-6}	2.20×10^{-8}	3.87×10^{-6}
3/8	3.00×10^{-7}	2.51×10^{-6}	1.52×10^{-6}	1.51×10^{-6}	1.16×10^{-6}
3/16	2.18×10^{-6}	1.81×10^{-6}	4.38×10^{-6}	1.00×10^{-6}	1.05×10^{-6}
3/32	1.65×10^{-6}	1.10×10^{-6}	5.48×10^{-6}	1.01×10^{-7}	2.51×10^{-6}
3/64	2.15×10^{-6}	1.40×10^{-6}	4.15×10^{-6}	4.33×10^{-9}	1.95×10^{-6}
3/128	2.95×10^{-6}	9.17×10^{-7}	2.87×10^{-6}	5.72×10^{-10}	2.33×10^{-6}
4/8	4.20×10^{-6}	7.73×10^{-7}	1.46×10^{-6}	1.55×10^{-6}	1.26×10^{-6}
4/16	7.32×10^{-7}	1.25×10^{-6}	2.00×10^{-6}	3.41×10^{-7}	1.65×10^{-6}
4/32	8.72×10^{-7}	1.00×10^{-6}	1.85×10^{-6}	5.47×10^{-9}	2.16×10^{-6}
4/64	2.11×10^{-6}	1.54×10^{-6}	1.93×10^{-6}	1.14×10^{-10}	1.54×10^{-6}
4/128	2.63×10^{-6}	1.00×10^{-6}	3.29×10^{-7}	7.28×10^{-12}	2.03×10^{-6}

表 6 终端位置偏差统计

Table 6 Statistics of terminal position deviation

方法	d_{max}/m	d_{min}/m	d_{mean}/m
优化算法(样本)	$3.864 0 \times 10^{-3}$	$2.340 0 \times 10^{-4}$	$2.008 0 \times 10^{-3}$
最优速度增量网络	$2.892 5 \times 10^{-1}$	$3.076 0 \times 10^{-3}$	$7.396 2 \times 10^{-2}$
最优摄动偏差网络	$1.192 3 \times 10^{-2}$	$6.240 0 \times 10^{-4}$	$4.751 9 \times 10^{-3}$
最优摄动修正网络	$1.117 8 \times 10^{-2}$	$3.150 0 \times 10^{-4}$	$4.175 8 \times 10^{-3}$

分析表 6 数据,可以得出:

1) 对比最优速度增量网络和最优摄动偏差网络的结果,不管是终端位置偏差的最大值、最小值还是平均值,最优摄动偏差网络的效果都要明显优于最优速度增量网络的效果(1 个数量级),即对速度增量摄动项因素进行训练的效果要明显优于直接对速度增量训练的效果。

2) 对比最优摄动偏差网络和最优摄动修正

网络结果,不管是终端位置偏差的最大值、最小值还是平均值,最优摄动修正网络的效果都要略微优于最优摄动偏差网络的效果,即将速度增量的摄动因素进行坐标变换,投影到轨道坐标系下,训练效果更好一点。

3) 最优摄动修正网络与优化算法的结果很接近,即对速度增量摄动因素坐标变换后进行训练出的神经网络在精度方面几乎与优化算法精度

一致。

仿真所用计算机性能为:内存 8 GB;处理器 Inter Core(TM) i7-2600@2.70 GHz;GPU NVIDIA GeForce GTX1050。优化算法给出一次决策的时间大约为 200 s,综合应用最优摄动修正网络和二体解给出一次决策的时间为 0.02 s 左右,相比优化算法,后者耗时大大缩短,实现了精度与耗时的统一,适于星上在轨实时决策。

4 结束语

对于固定两点间固定时间的轨道机动问题,在初始位置速度和终端位置速度一定误差范围内,本文得到的最优摄动修正网络决策精度与优化算法的精度一致,但其决策耗时仅为优化算法决策耗时的 0.01%,实时性高。

本文为在轨机动、行星际转移等复杂问题的快速高精度求解提供了一种思路,将精确解分解为简化模型解与残差的和,对残差进行训练得到神经网络,再与简化模型解组合获得原问题的高精度解。

参考文献 (References)

- [1] 夏红伟,李莉,曲耀斌,等.卫星编队构型设计与轨道机动算法优化[J].中国惯性技术学报,2013,21(2):186-191.
XIA H W, LI L, QU Y B, et al. Satellites formation configuration design and orbit maneuver algorithm optimization[J]. Journal of Chinese Inertial Technology, 2013, 21(2): 186-191 (in Chinese).
- [2] 于瀚.航天器轨道机动可达区域研究[D].哈尔滨:哈尔滨工业大学,2017.
YU H. The study of reachable domain for spacecraft maneuver [D]. Harbin: Harbin Institute of Technology, 2017 (in Chinese).
- [3] 谭丽芬.赤道椭圆交会轨道规划与制导方法[D].长沙:国防科技大学,2011.
TAN L F. Rendezvous trajectory planning and guidance approach for equatorial elliptical orbit [D]. Changsha: National University of Defense Technology, 2011 (in Chinese).
- [4] 张守玉,姜振东.基于 STK 的卫星轨道机动模型设计与仿真[J].计算机仿真,2004,21(10):25-27.
ZHANG S Y, JIANG Z D. Design and simulation of satellite orbital maneuver model on STK[J]. Computer Integrated Manufacturing Systems, 2004, 21(10): 25-27 (in Chinese).
- [5] BATE R R. 航天动力学基础[M].吴鹤鸣,李肇杰,译.北京:北京航空航天大学出版社,1990.
BATE R R. Fundamentals of astrodynamics [M]. WU H M, LI Z J, translated. Beijing: Beihang University Press, 1990 (in Chinese).
- [6] 李栋林,黄福铭.基于 Lambert 问题的精确拦截与交会策略研究[J].飞行力学,2008,26(2):57-59.
LI D L, HUANG F M. Research into accurate interception and rendezvous scheme based on Lambert problem[J]. Flight Dynamics, 2008, 26(2): 57-59 (in Chinese).
- [7] 桑艳,周进.基于 Lambert 算法的脉冲精确变轨策略[J].国防科技大学学报,2009,31(3):29-32.
SANG Y, ZHOU J. An approach of accurate impulse transfer based on Lambert algorithm[J]. Journal of National University of Defense Technology, 2009, 31(3): 29-32 (in Chinese).
- [8] JEZEWSKI D J. Optimal rendezvous trajectories subject to arbitrary perturbations and constraints [C] // AIAA/AAS Astrodynamics Specialist Conference. Reston: AIAA Press, 1992:4507.
- [9] CHANG Y, ZHOU J. Orbital correction method for two-impulse rendezvous between non-coplanar elliptic orbits considering the J2 perturbation[J]. Journal of Astronautics, 2008, 29(4): 1172-1176.
- [10] 周须峰,唐硕.固定时间拦截变轨段制导的摄动修正方法[J].飞行力学,2006,24(4):46-49.
ZHOU X F, TANG S. Disturbed modify method of fixed-time interception's guidance in orbit-change stage[J]. Flight Dynamics, 2006, 24(4): 46-49 (in Chinese).
- [11] 裴忠才,尹丽,王占林.基于神经网络的仿真转台控制系统[J].北京航空航天大学学报,2005,31(9):1045-1048.
PEI Z C, YIN L, WANG Z L. Simulating turntable control system with neural network [J]. Journal of Beijing University of Aeronautics and Astronautics, 2005, 31(9): 1045-1048 (in Chinese).
- [12] CHENG L, WANG Z, JIANG F, et al. Real-time optimal control for spacecraft orbit transfer via multiscale deep neural networks [J]. IEEE Transactions on Aerospace and Electronic Systems, 2018, 55(5): 2436-2450.
- [13] ZHONG R, XU S. Neural-network-based terminal sliding-mode control for thrust regulation of a tethered space-tug [J]. Astrodynamics, 2018, 2(10): 175-185.
- [14] SÁNCHEZ-SÁNCHEZ C, IZZO D. Real-time optimal control via deep neural networks: Study on landing problems [J]. Journal of Guidance, Control, and Dynamics, 2018, 41(5): 1122-1135.
- [15] IZZO D, SPRAGUE C I, TAILOR D V. Machine learning and evolutionary techniques in interplanetary trajectory design [M] // FASANO G, PINIER J. Modeling and optimization in space engineering. Berlin: Springer, 2019: 191-210.
- [16] FURFARO R, BLOISE I, ORLANDELLI M, et al. Deep learning for autonomous lunar landing [C] // AIAA/AAS Astrodynamics Specialist Conference. Reston: AIAA, 2018: 18-363.
- [17] 董云峰,陈士明,苏建敏,等.卫星姿态控制动态模拟技术 [M].北京:科学出版社,2010:300-301.
DONG Y F, CHEN S M, SU J M, et al. Dynamic simulation technology of satellite attitude control [M]. Beijing: Science Press, 2010: 300-301 (in Chinese).

High-precision on-orbit real-time orbital maneuver decision

XIE Shucong, DONG Yunfeng*

(School of Astronautics, Beihang University, Beijing 100083, China)

Abstract: In order to ensure the real-time maneuverability and high-precision requirements of orbital maneuver, a real-time maneuver decision-making method based on machine learning is proposed. The optimal solution under perturbation is obtained offline through the optimization algorithm. The two-body solution is subtracted to obtain the speed increment difference, which is projected onto the orbital system to obtain the speed increment perturbation correction term, which is used as the output of the neural network. The network parameters are designed and trained to obtain perturbation correction network. The combination of perturbation correction network and two-body solution is used to achieve high-precision real-time orbital maneuver decision. The simulation results show that the terminal position deviation after the completion of the maneuver according to the decision is consistent with the accuracy of the terminal position deviation after the completion of the decision maneuver according to the optimization algorithm, and the former decision time is only about 0.01% of the latter decision time. The orbital maneuver decision-making method proposed in this paper takes into account both accuracy and real-time performance, and is suitable for on-board decision-making.

Keywords: orbital maneuver; neural networks; machine learning; Lambert maneuver; perturbation correction

http://bhxb.buaa.edu.cn jbuua@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2020.0227

基于中心-对数半长的区间数据主成分分析



赵青^{1,2}, 王惠文^{1,3}, 王珊珊^{1,2,*}

1. 北京航空航天大学 经济管理学院, 北京 100083; 2. 城市运行应急保障模拟技术北京市重点实验室, 北京 100083;

3. 北京航空航天大学 大数据科学与脑机智能高精尖创新中心, 北京 100083)

摘 要: 为研究多变量区间数据的降维和可视化,采用包含中心点和半长对数值的二维数组表征区间数据,建立了区间数据的代数运算法则,并在此基础上提出了一种新的区间数据主成分分析(PCA)方法。对区间半长取对数的处理保证了最终得到的区间主成分半长非负的合理性,计算过程简单、复杂度较低,并且使得降维前后样本集中中点之间相对位置的改变尽可能小。通过对高维空间进行变量降维,从而多种经典的统计分析方法能够得到运用,同时能够在低维空间中描绘原始高维空间中的样本点,使得多变量区间数据的可视化成为可能。仿真实验结果表明了所提方法的有效性。

关键词: 区间数据; 主成分分析(PCA); 中心-对数半长; 降维; 协方差矩阵

中图分类号: O212.4

文献标志码: A

文章编号: 1001-5965(2021)07-1414-08

主成分分析(Principal Component Analysis, PCA)是一种对包含多个变量的平面数据表进行最佳综合简化的多元分析方法^[1]。其主要目的是在保证数据信息损失最小的前提下,对多元数据进行降维处理,基本原理是通过正交变换将 p 个相互关联的变量转换为 p 个相互无关的“主成分”,并省却数据变异不大的方向信息,提取前 m ($m \leq p$) 个成分来概括数据系统^[2]。主成分分析可以将难以可视化的高维空间样本点投影到较为直观的低维空间(如二维平面),从而实现数据的可视化,有助于提取有效信息得以辅助相关决策^[3]。然而,传统的主成分分析方法在处理大规模数据时,存在较大的局限性^[4]:一方面,海量数据的运算会增加计算机的内存负荷,从而大大降低了主成分分析的计算效率;另一方面,较大的样本数量会对降维后的可视化效果造成影响。

不同于传统的单值数据,区间数据通过“数

据打包”的思想将原数据集合处理为若干个高维超矩形^[5],区间数据表中的每个单元是一个区间而不是一个数值。区间数据的应用,一方面可以有效地降低样本容量,同时能够有效解决数据中的可变性和不确定性,反映样本群点的集中趋势和离散程度等整体特征^[6]。近年来,一系列区间数据主成分分析方法被提出,例如,顶点主成分分析法(Vertices Principal Component Analysis, V-PCA)和中心主成分分析法(Centers Principal Component Analysis, C-PCA)是最经典的2种区间数据主成分分析方法。

具体地,对包含 n 个样本点、 p 个变量的 $n \times p$ 维区间数据表 $\mathbf{X}_{n \times p}$, V-PCA 把每个区间样本点视为具有 2^p 个顶点的超矩形^[7-8],将 $\mathbf{X}_{n \times p}$ 展开成 $(n \cdot 2^p) \times p$ 维的普通数值型数据表,然后对其进行传统的主成分分析并投影到低维空间中,最后重新组合还原为区间数据。但 V-PCA 的计算过

收稿日期: 2020-05-29; 录用日期: 2020-08-22; 网络出版时间: 2020-09-21 17:08

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20200921.1523.001.html

基金项目: 国家自然科学基金(71420107025, 11701023)

* 通信作者. E-mail: sswang@buaa.edu.cn

引用格式: 赵青, 王惠文, 王珊珊. 基于中心-对数半长的区间数据主成分分析[J]. 北京航空航天大学学报, 2021, 47(7): 1414-1421. ZHAO Q, WANG H W, WANG S S. A principal component analysis of interval data based on center and log-radius [J]. Journal of Beijing University of Aeronautics and Astronautics, 2021, 47(7): 1414-1421 (in Chinese).

程随着变量维数 p 的增大而更加复杂,导致出现“维数灾难”。Cazes等^[7]给出一种简化算法,直接计算 V-PCA 展开后矩阵的相关系数矩阵,可以消除大量的冗余计算。王惠文等^[9]比较了 V-PCA 和简化的 V-PCA 的计算过程和结果,验证了二者计算结果的等价性以及 V-PCA 简化方法的高效性。而 C-PCA 用每个区间的中点代替原始区间^[10-11],计算过程与 V-PCA 类似,对用中点表示的 $n \times p$ 维普通数值型数据表进行传统的主成分分析,相比 V-PCA 方法,C-PCA 的计算量得到有效降低,但同时丢失了更多的数据信息。

此外,郭均鹏和李汶华^[12]研究了服从均匀分布的区间数据的经验描述统计量,基于经验相关矩阵对区间数据进行主成分分析,结果与 V-PCA 和 C-PCA 类似,但降低了计算复杂度。Palumbo 和 Lauro^[13]提出了中心半长法,将区间样本矩阵分为中心矩阵和半长矩阵,分别进行主成分分析,然后加以综合得到最终的区间主成分,但该方法假设 2 个矩阵主成分相同,存在一定的理论缺陷。

为解决上述研究方法存在的一些问题,郭均鹏和李汶华^[14]提出了基于误差传递公式的区间主成分分析方法,首先得到测量值的主成分,通过误差传递公式得到极限误差,然后组合二者得到最终的区间主成分。Wang 等^[15]提出了基于全信息的区间数据主成分分析方法,基于区间数据的基本代数算子和运算规则,在计算过程中保留区间样本内部的全部信息,得到能够准确反映区间内部结构特征的主轴。侯自盼和李生刚^[16]依据区间矩阵的运算方法得到区间数据样本的协方差矩阵和样本相关矩阵,计算所得对应的特征值和特征向量的分量均为区间数,并且重新定义贡献率,由此给出一种新的区间数据主成分分析方法。刘清贤^[17]在服从正态分布区间数的研究基础上,提出了 2 种改进的算法。

不同于上述方法,本文考虑对区间的半长值取对数,基于中心和对数半长的表达,提出一种新的区间数据主成分分析方法,实现多变量区间数据的降维与可视化。与已有方法相比,本文方法计算复杂度较低,且保证了降维前后样本集中点点之间相对位置的改变尽可能小。仿真实验也验证了该区间数据主成分分析方法的有效性。

1 基本概念与理论

1.1 区间数据的基本概念

区间数据的取值可表示为一个集合:

$$x = [\underline{x}, \bar{x}] = \{s \mid \underline{x} < s < \bar{x}\} \quad (1)$$

式中: $\underline{x} \leq \bar{x}$ 且 \underline{x} 和 $\bar{x} \in \mathbf{R}$, 称 \underline{x} 和 \bar{x} 分别为区间 x 的下界和上界; $s \in \mathbf{R}$ 表示实数。注意到区间数据的这种表达形式存在一个约束条件,即区间下界必须小于等于上界。

在中心半长法^[13]中,区间数据表达为一个由区间中心和半长构成的数组:

$$x = (x^c, x^r) \quad (2)$$

式中: $x^c = (\underline{x} + \bar{x})/2$ 和 $x^r = (\bar{x} - \underline{x})/2$ 分别表示区间的中心和半长。此种表达形式直接体现了区间数据的位置信息和大小信息,但仍然存在一个约束条件,即区间的半长必须为非负数。

带约束的表达形式极大地限制了多种经典的统计分析方法在区间数据领域的应用。为解决上述问题,考虑对区间的半长进行取自然对数的处理(以下简称为“对数半长”),将区间数据重新表达为由中心和对数半长组成的二维数组:

$$x = (x^c, x^{\ln r}) \quad (3)$$

式中: $x^{\ln r} = \ln(x^r) = \ln[(\bar{x} - \underline{x})/2]$ 表示区间的对数半长; $x^c = (\underline{x} + \bar{x})/2$ 仍表示区间中心。由对数函数的性质可得,对区间半长 x^r 取自然对数的处理可以解除其必须为非负数的约束条件,并且 x^r 与其自然对数值 $x^{\ln r} = \ln(x^r)$ 是一一对应。需要说明的是,对半长进行取对数的处理适用于半长恒为正数的区间,故本文不考虑区间数据退化为普通数值型数据的情形。

易知,上述区间数据的 3 种表达形式是互相对应的,且能唯一确定一个区间。若无特殊说明,下文关于区间数据的论述均采用式(3)基于中心-对数半长的表达形式。

若 n 维向量 $\mathbf{X} = [x_1, x_2, \dots, x_n]'$ 的每个分量均为区间数据,即 $x_i = (x_i^c, x_i^{\ln r}) (1 \leq i \leq n)$, 则称 \mathbf{X} 为区间向量。类似的,若 $n \times p$ 维矩阵 $\mathbf{X}_{n \times p}$ 的每个元素均为区间数据 $x_{ij} = (x_{ij}^c, x_{ij}^{\ln r})$, 则称 $\mathbf{X}_{n \times p}$ 为区间矩阵,记为

$$\mathbf{X}_{n \times p} = [\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_p] = \begin{bmatrix} \mathbf{e}'_1 \\ \mathbf{e}'_2 \\ \dots \\ \mathbf{e}'_n \end{bmatrix} = [(x_{ij}^c, x_{ij}^{\ln r})]_{n \times p} \quad (4)$$

式中:每一行 \mathbf{e}'_i 为一个区间样本,由 p 个变量刻画;每一列 \mathbf{X}_j 表示一个区间变量,包含 n 个独立样本观测;第 i 行第 j 列元素 $x_{ij} = (x_{ij}^c, x_{ij}^{\ln r})$ 表示样本点 \mathbf{e}'_i 对应区间变量 \mathbf{X}_j 的区间观测值。

本文关于区间数据主成分分析的对象正是形

如式(4)包含 n 个独立样本并由 p 个区间变量描述的区间数据表。

1.2 区间数据的代数运算

基于中心-对数半长的表达形式,对于任意 2 个区间数据 $a = (a^c, a^{lnr})$ 和 $b = (b^c, b^{lnr})$, 及任意实数 $\lambda \in \mathbf{R}$, 定义区间数据加法 \oplus 和数乘 \odot 的运算法则遵循二维向量的对应运算, 即

$$a \oplus b = (a^c + b^c, a^{lnr} + b^{lnr}) \quad (5)$$

$$\lambda \odot a = \lambda \odot (a^c, a^{lnr}) = (\lambda a^c, \lambda a^{lnr}) \quad (6)$$

可以看到, 2 个区间数据相加即 2 个区间的中心和对数半长分别相加, 实数与区间数据的数乘即中心和对数半长分别与实数相乘, 加法和数乘运算的结果仍为区间数据。进而可推导减法运算 \ominus 为

$$a \ominus b = (a^c - b^c, a^{lnr} - b^{lnr}) \quad (7)$$

类似地, 根据欧式空间二维向量内积运算规则, 给出区间数据内积 $\langle \cdot, \cdot \rangle_{c-lnr}$ 的定义如下:

$$\langle a, b \rangle_{c-lnr} = a^c b^c + a^{lnr} b^{lnr} \quad (8)$$

易证上述内积的定义满足正定性、对称性和线性。进一步, 可以得到由内积(8)导出的区间数据的范数 $\| \cdot \|_{c-lnr}$ 和距离 $d_{c-lnr}(\cdot, \cdot)$ 定义为

$$\| a \|_{c-lnr} = \sqrt{\langle a, a \rangle_{c-lnr}} = \sqrt{(a^c)^2 + (a^{lnr})^2} \quad (9)$$

$$d_{c-lnr}(a, b) = \| a \ominus b \|_{c-lnr} = \sqrt{(a^c - b^c)^2 + (a^{lnr} - b^{lnr})^2} \quad (10)$$

易证式(9)给出的范数定义满足正定性、齐次性和三角不等式。式(10)在考虑 2 个区间数据的距离时, 同时考虑了区间中心(即位置信息)和对数半长(即长度信息) 2 个维度的差异, 并且满足正定性、对称性和三角不等式。

尽管已有文献提出了一些基于 Moore 算术的区间数据代数运算, 却很少提及区间数据代数体系的零元素。基于中心-对数半长的表达和式(5)~式(10)的运算法则, 可定义此框架下区间数据的零元素为 $\eta_e = (\eta_e^c, \eta_e^{lnr}) = (0, 0)$, 还原为区间上下界的表达形式为 $[-1, 1]$, 满足对任意区间数据 $a = (a^c, a^{lnr})$ 有:

$$a \oplus \eta_e = (a^c, a^{lnr}) \oplus (0, 0) = a \quad (11)$$

1.3 区间向量的代数运算

进而可将上述运算法则推广至区间向量空间。对区间向量 $\mathbf{A} = [a_1, a_2, \dots, a_n]'$ 和 $\mathbf{B} = [b_1, b_2, \dots, b_n]'$, 其中对 $\forall i = 1, 2, \dots, n, a_i = (a_i^c, a_i^{lnr}), b_i = (b_i^c, b_i^{lnr})$ 均为区间数据。则区间向量之间的加法 \oplus 、数乘 \odot 、内积 $\langle \cdot, \cdot \rangle_{c-lnr}$ 、范数 $\| \cdot \|_{c-lnr}$ 和距离 $d_{c-lnr}(\cdot, \cdot)$ 运算可分别定义为

$$\mathbf{A} \oplus \mathbf{B} = (a_1 \oplus b_1, a_2 \oplus b_2, \dots, a_n \oplus b_n)' \quad (12)$$

$$\lambda \odot \mathbf{A} = (\lambda \odot a_1, \lambda \odot a_2, \dots, \lambda \odot a_n)' \quad (13)$$

$$\langle \mathbf{A}, \mathbf{B} \rangle_{c-lnr} = \sum_{i=1}^n \langle a_i, b_i \rangle_{c-lnr} = \sum_{i=1}^n a_i^c b_i^c + \sum_{i=1}^n a_i^{lnr} b_i^{lnr} \quad (14)$$

$$\| \mathbf{A} \|_{c-lnr} = \sqrt{\sum_{i=1}^n \| a_i \|_{c-lnr}^2} = \sqrt{\sum_{i=1}^n [(a_i^c)^2 + (a_i^{lnr})^2]} \quad (15)$$

$$d_{c-lnr}(\mathbf{A}, \mathbf{B}) = \| \mathbf{A} \ominus \mathbf{B} \|_{c-lnr} = \sqrt{\sum_{i=1}^n \| a_i \ominus b_i \|_{c-lnr}^2} \quad (16)$$

根据式(16)所示区间向量的距离运算, 可定义 2 个 $n \times p$ 维区间矩阵 $\mathbf{X}_{n \times p}$ 和 $\mathbf{Y}_{n \times p}$ 的距离为

$$d_{c-lnr}(\mathbf{X}_{n \times p}, \mathbf{Y}_{n \times p}) = \sum_{i=1}^n d_{c-lnr}(\mathbf{e}_i, \boldsymbol{\xi}_i) = \sum_{i=1}^n \sqrt{\sum_{j=1}^p \| x_{ij} \ominus y_{ij} \|_{c-lnr}^2} \quad (17)$$

式中: $\mathbf{e}_i = [x_{i1}, x_{i2}, \dots, x_{ip}]'$ 和 $\boldsymbol{\xi}_i = [y_{i1}, y_{i2}, \dots, y_{ip}]'$ 分别为区间矩阵 $\mathbf{X}_{n \times p}$ 和 $\mathbf{Y}_{n \times p}$ 的第 i 行区间向量转置。

此外, 定义 $n \times p$ 维区间矩阵 $\mathbf{X}_{n \times p} = (x_{ij})_{n \times p}$ 与 $p \times m$ 维实数矩阵 $\mathbf{U}_{p \times m} = (u_{jk})_{p \times m} (u_{jk} \in \mathbf{R})$ 相乘仍为一个区间矩阵, 记为 $\mathbf{W}_{n \times m} = (w_{ik})_{n \times m} = \mathbf{X} \odot \mathbf{U}$, $\mathbf{W}_{n \times m}$ 的第 i 行第 k 列元素为

$$w_{ik} = \bigoplus_{j=1}^p u_{jk} \odot x_{ij} \quad (18)$$

式中: x_{ij} 为区间矩阵 \mathbf{X} 的第 i 行第 j 列元素, 为一个区间; u_{jk} 为实数矩阵 \mathbf{U} 的第 j 行第 k 列元素, 为一个实数。

2 区间数据主成分分析的理论推导

2.1 经典的主成分分析

首先简单回顾经典的主成分分析。对于给定的 p 个变量 $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_p$, 记其协方差矩阵为 \mathbf{S} , 则第 $h (1 \leq h \leq p)$ 个主成分是 p 个变量的线性组合, 即 $\mathbf{Y}_h = \sum_{j=1}^p u_{jh} \mathbf{X}_j$ 。已知数据所携带信息量的多少可以由变量的方差表示, 方差越大, 则所含信息越多。因此, 为了在信息损失最少的前提下实现对变量的降维, 主成分分析要求前 $m (m \leq p)$ 个主成分的方差之和尽可能大, 并满足约束条件: $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m$ 是标准正交的, $\mathbf{u}_h = [u_{1h}, u_{2h}, \dots, u_{ph}]' \in \mathbf{R}^p, h = 1, 2, \dots, m$ 。上述优化问题的解恰是协方差矩阵 \mathbf{S} 的前 m 个最大特征值所对应的 m 个特征向量, 并且第 h 个最大特征值 λ_h 的取值恰好是第 h 个主成分的方差。综上所述, 对变量 $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_p$ 进行主成分分析的过程实际上是对

其协方差矩阵 S 的特征分解过程。

2.2 区间变量的数字特征

在式(4)所示的 $n \times p$ 维区间数据表 $\mathbf{X}_{n \times p}$ 中, 每一行 \mathbf{e}_i' 为一个由 p 个区间变量刻画的区间样本, 每一列表示所有 n 个样本在第 j 个区间变量 X_j 上的观测值, 由此可得到区间变量 X_j 的一些统计特征。

区间变量 X_j 的样本均值为

$$E(X_j) = \bar{X}_j = \left(\frac{1}{n} \sum_{i=1}^n x_{ij}^c, \frac{1}{n} \sum_{i=1}^n x_{ij}^{\text{lnr}} \right) \quad (19)$$

记 $\bar{X}_j^c = \frac{1}{n} \sum_{i=1}^n x_{ij}^c$ 和 $\bar{X}_j^{\text{lnr}} = \frac{1}{n} \sum_{i=1}^n x_{ij}^{\text{lnr}}$ 分别表示

区间变量 X_j 中心和和对数半长的样本均值。则 X_j 的样本均值 $E(X_j) = (\bar{X}_j^c, \bar{X}_j^{\text{lnr}})$ 仍为一个二维数组表示的区间数据。

区间变量 X_j 的样本方差为

$$\text{Var}(X_j) = S_j^2 = \frac{1}{2n} \sum_{i=1}^n \langle x_{ij} - \bar{X}_j, x_{ij} - \bar{X}_j \rangle_{\text{c-lnr}} \quad (20)$$

记 $(S_j^c)^2 = \frac{1}{n} \sum_{i=1}^n (x_{ij}^c - \bar{X}_j^c)^2$ 表示区间变量 X_j

中心的方差, $(S_j^{\text{lnr}})^2 = \frac{1}{n} \sum_{i=1}^n (x_{ij}^{\text{lnr}} - \bar{X}_j^{\text{lnr}})^2$ 表示 X_j

对数半长的方差, 易证 $S_j^2 = \frac{1}{2} [(S_j^c)^2 + (S_j^{\text{lnr}})^2]$ 。

区间变量 X_j 和 X_k 的样本协方差为

$$\text{Cov}(X_j, X_k) = S_{jk} = \frac{1}{2n} \sum_{i=1}^n \langle x_{ij} - \bar{X}_j, x_{ik} - \bar{X}_k \rangle_{\text{c-lnr}} \quad (21)$$

区间变量 X_j 和 X_k 的样本相关系数为

$$\text{corr}(X_j, X_k) = r_{jk} = \frac{S_{jk}}{\sqrt{S_j^2 S_k^2}} \quad (22)$$

从而可以得到区间数据表 $\mathbf{X}_{n \times p}$ 的样本协方差矩阵 \mathbf{V} 和样本相关系数矩阵 \mathbf{R} , 分别表示为

$$\left\{ \begin{aligned} \mathbf{V} &= \begin{bmatrix} S_1^2 & S_{12} & \cdots & S_{1p} \\ S_{21} & S_2^2 & \cdots & S_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ S_{p1} & S_{p2} & \cdots & S_p^2 \end{bmatrix} \\ \mathbf{R} &= \begin{bmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{21} & 1 & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & 1 \end{bmatrix} \end{aligned} \right. \quad (23)$$

2.3 区间数据主成分分析理论推导

考虑对形如式(4)的 $n \times p$ 维区间数据表 $\mathbf{X}_{n \times p}$ 进行区间主成分分析, 为便于推导, 这里假

设区间变量 X_1, X_2, \dots, X_p 的样本均值均为零元素。类似于经典的主成分分析, 第 h ($1 \leq h \leq p$) 个区间主成分 Y_h 是 p 个区间变量的线性组合, 即

$$\mathbf{Y}_h = \mathbf{X} \odot \mathbf{u}_h = \bigoplus_{j=1}^p u_{jh} \odot \mathbf{X}_j \quad (24)$$

式中: $\mathbf{u}_h = [u_{1h}, u_{2h}, \dots, u_{ph}]' \in \mathbf{R}^p$, 并满足约束条件 $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_p$ 标准正交。此外, 对 $\forall j = 1, 2, \dots, p$, 有 $E(X_j) = (0, 0)$, 则 $E(\mathbf{Y}_h) = (0, 0)$, 结合式(20) ~ 式(23)可得 \mathbf{Y}_h 的方差为

$$\text{Var}(\mathbf{Y}_h) = \frac{1}{2n} \langle \mathbf{Y}_h, \mathbf{Y}_h \rangle_{\text{c-lnr}} = \mathbf{u}_h' \mathbf{V} \mathbf{u}_h \quad (25)$$

区间数据主成分分析同样要求前 m ($m \leq p$) 个区间主成分携带尽可能多的信息, 即前 m 个区间主成分方差之和要达到最大, 从而根据式(25)可以转化为一个二次型的极值问题:

$$\begin{aligned} \max_{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m} \sum_{h=1}^m \mathbf{u}_h' \mathbf{V} \mathbf{u}_h \\ \text{s. t. } \mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m \text{ 标准正交} \end{aligned} \quad (26)$$

与经典主成分分析类似, 由二次型极大值定理可得, 极值问题(26)的最优解是 $\mathbf{X}_{n \times p}$ 的样本协方差矩阵 \mathbf{V} 的前 m 个最大特征值 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$ 对应的特征向量 $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m$, 且最优值是 $\sum_{h=1}^m \lambda_h$ 。

上述基于中心-对数半长的区间数据代数运算框架下的主成分分析方法, 以下简称 C-lnR PCA 方法。通过 C-lnR PCA 方法, 原始的 p 维区间数据表投影到 m ($m \leq p$) 维空间, 从而能够在低维空间上运用多种经典的统计分析方法, 进一步提取更多有效信息, 同时能够在低维空间(如二维平面)上描绘和分析原始高维空间中的样本点, 使得多变量区间数据表的可视化成为可能。

3 区间数据 C-lnR PCA 的建模过程

3.1 区间变量的相关处理方法

1) 区间变量 X_j 的中心化处理

将 $x_{ij} = (x_{ij}^c, x_{ij}^{\text{lnr}})$ 变换为

$$\hat{x}_{ij} = x_{ij} - \bar{X}_j = \left(x_{ij}^c - \frac{1}{n} \sum_{i=1}^n x_{ij}^c, x_{ij}^{\text{lnr}} - \frac{1}{n} \sum_{i=1}^n x_{ij}^{\text{lnr}} \right) \quad (27)$$

中心化处理对所有观测值进行平移处理, 不会改变样本之间的相对位置关系和变量之间的相关关系。记 \hat{X}_j 表示区间变量 X_j 中心化处理后的区间变量, 则 \hat{X}_j 满足

$$E(\hat{X}_j) = (0, 0) \quad (28)$$

$$\text{Var}(\widehat{X}_j) = \frac{1}{2n} \langle \widehat{X}_j, \widehat{X}_j \rangle_{c-\ln r} \quad (29)$$

记 \widehat{X}_k 表示区间变量 X_k 中心化处理后的区间变量, 则 $\widehat{X}_j, \widehat{X}_k$ 满足

$$\text{Cov}(\widehat{X}_j, \widehat{X}_k) = \frac{1}{2n} \langle \widehat{X}_j, \widehat{X}_k \rangle_{c-\ln r} \quad (30)$$

2) 区间变量 X_j 的标准化处理

将 $x_{ij} = (x_{ij}^c, x_{ij}^{\ln r})$ 变换为

$$\widehat{x}_{ij} = \left(\frac{x_{ij}^c - \bar{X}_j^c}{S_j^c}, \frac{x_{ij}^{\ln r} - \bar{X}_j^{\ln r}}{S_j^{\ln r}} \right) \quad (31)$$

标准化处理使区间变量方差变为 1, 可以更真实地体现样本的离散程度。设 \widehat{X}_j 表示区间变量 X_j 经过标准化处理后的区间变量, 则 \widehat{X}_j 满足

$$E(\widehat{X}_j) = (0, 0) \quad (32)$$

$$\text{Var}(\widehat{X}_j) = 1 \quad (33)$$

记 \widehat{X}_k 表示区间变量 X_k 经过标准化处理后的区间变量, 则 $\widehat{X}_j, \widehat{X}_k$ 满足

$$\text{corr}(\widehat{X}_j, \widehat{X}_k) = \text{Cov}(\widehat{X}_j, \widehat{X}_k) = \frac{1}{2n} \langle \widehat{X}_j, \widehat{X}_k \rangle_{c-\ln r} \quad (34)$$

3.2 C-lnR PCA 方法建模步骤

综上, 对式(4)所示区间数据表 $X_{n \times p}$ 进行 C-lnR PCA 方法的建模步骤如下:

步骤 1 对区间数据表 $X_{n \times p}$ 进行中心化或标准化处理。标准化处理目的是使样本点集合的中心与坐标原点重合, 并消除由量纲不同引起的虚假变异信息, 得到更加合理的分析结果。为方便起见, 仍记中心化或标准化处理后的数据表为 $X_{n \times p}$ 。

步骤 2 计算区间数据表 $X_{n \times p}$ 的协方差矩阵 V 。

步骤 3 通过特征分解求解协方差矩阵 V 的特征值 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$, 以及特征值对应的特征向量 u_1, u_2, \dots, u_p (要求是标准正交的)。

步骤 4 根据确定主成分个数的一些准则, 如 Kaiser-Guttman 准则或需要的累积贡献率来选取最终保留的主成分的个数 $m (m \leq p)$ 。

步骤 5 对应的 u_1, u_2, \dots, u_m 称为主轴, 可计算样本点 $e_i = [x_{i1}, x_{i2}, \dots, x_{ip}]'$ 在主轴 $u_h = [u_{1h}, u_{2h}, \dots, u_{ph}]'$ 上的投影 $F_{ih} (h = 1, 2, \dots, m)$ 为

$$F_{ih} = e_i' \odot u_h = \bigoplus_{j=1}^p u_{jh} x_{ij}^c = \left(\sum_{j=1}^p u_{jh} x_{ij}^c, \sum_{j=1}^p u_{jh} x_{ij}^{\ln r} \right) \quad (35)$$

记 $F_{ih}^c = \sum_{j=1}^p u_{jh} x_{ij}^c, F_{ih}^{\ln r} = \sum_{j=1}^p u_{jh} x_{ij}^{\ln r}$, 可将投影

F_{ih} 还原为样本点 e_i 在主轴 u_h 上的区间下界和上界分别为

$$\begin{cases} \underline{F}_{ih} = F_{ih}^c - \exp(F_{ih}^{\ln r}) \\ \overline{F}_{ih} = F_{ih}^c + \exp(F_{ih}^{\ln r}) \end{cases} \quad (36)$$

定理 1 对区间数据表 $X_{n \times p}$ 实施 C-lnR PCA, 用 \hat{e}_i 表示原 p 维空间中样本点 e_i 在区间主成分分析后的 m 维空间上的投影, 易证

$$\frac{1}{n^2} \sum_{i=1}^n \sum_{l=1}^n \|e_i \ominus e_l\|^2 - \frac{1}{n^2} \sum_{i=1}^n \sum_{l=1}^n \|\hat{e}_i \ominus \hat{e}_l\|^2 = 4[\text{Trace}(V) - \sum_{h=1}^m u_h' V u_h] \quad (37)$$

式中: Trace 表示迹函数, 即矩阵对角线上的所有元素之和。

对一个确定的区间数据表而言, 样本协方差矩阵 V 是确定的, 因此 $\text{Trace}(V)$ 也是确定的, 由最优化问题(26)和式(37)可得

$$\max \sum_{h=1}^m u_h' V u_h \Leftrightarrow \min \left(\sum_{i=1}^n \sum_{l=1}^n \|e_i \ominus e_l\|^2 - \sum_{i=1}^n \sum_{l=1}^n \|\hat{e}_i \ominus \hat{e}_l\|^2 \right) \quad (38)$$

上述定理表明, C-lnR PCA 方法将原 p 维空间降维到一个 m 维空间后, 样本集中点点之间相对位置的改变是尽可能小的, 即尽可能地保持降维前后点点之间位置关系的一致性。

4 区间 PCA 方法有效性比较

4.1 不同区间 PCA 方法复杂度比较

考虑对区间数据表 $X_{n \times p}$ 进行区间主成分分析, 将经典的 V-PCA、C-PCA 与本文提出的 C-lnR PCA 方法进行比较。

关于空间复杂度, 经典的 V-PCA 方法需要首先将 $X_{n \times p}$ 完全展开为 $(n \cdot 2^p) \times p$ 维普通数值型数据表, C-PCA 方法的处理对象则是 $n \times p$ 维数值型中心矩阵, 而本文提出的 C-lnR PCA 方法等价于处理 $2n \times p$ 维的中心-对数半长数值型矩阵。显然 C-PCA 方法和 C-lnR PCA 方法所需要的存储空间远小于 V-PCA 方法。

关于时间复杂度, 3 种区间数据主成分分析方法的比较结果如表 1 所示, V-PCA 方法的计算量要远大于 C-PCA 方法和 C-lnR PCA 方法, 后 2 种方法计算过程较为简单, 因此时间复杂度显著较低。

综上所述, C-PCA 和 C-lnR PCA 2 种方法在空间复杂度和时间复杂度上都优于 V-PCA 方法。

表1 V-PCA、C-PCA 和 C-lnR PCA 时间复杂度的比较

Table 1 Comparison of time complexity among V-PCA, C-PCA and C-lnR PCA methods

计算步骤	复杂度		
	V-PCA	C-PCA	C-lnR PCA
计算 X_j 均值	$n \cdot 2^p$ 个值参与运算	n 个值参与运算	$2n$ 个值参与运算
计算 X_j 方差	$n \cdot 2^p$ 个值参与运算	n 个值参与运算	$2n$ 个值参与运算
对 X_j 标准化	$n \cdot 2^p$ 个值参与运算	n 个值参与运算	$2n$ 个值参与运算
计算协方差矩阵	$C_p^2 \times (n \cdot 2^p)$ 次乘法	$C_p^2 \times n$ 次乘法	$C_p^2 \times (2n)$ 次乘法
计算每个样本的第 h 区间主成分	$p \cdot 2^p$ 次乘法	$2p$ 次乘法	$(2p + 1)$ 次乘法

4.2 区间数据主成分分析的有效性指标

为能够定量地衡量不同区间数据主成分分析方法的有效性,本文参考文献[18]给出一种区间数据 PCA 的效度指标。对原始区间数据表 $X_{n \times p}$ 进行某种区间数据主成分分析方法,假设最终选取前 m ($m \leq p$) 个主成分,可以得到由对应前 m ($m \leq p$) 个最大特征值的特征向量所构成的 $p \times m$ 维正交矩阵 $U_{p \times m}$,由 $Y = X \odot U$ 可得 $n \times m$ 维区间矩阵 $Y_{n \times m}$,再根据数据重构的思想由 $X^* = Y \odot U'$ 计算出 $X_{n \times p}$ 的拟合区间数据表 $X_{n \times p}^*$ 。由于主成分分析旨在保留原始变量尽可能多的信息,因此拟合区间数据表 $X_{n \times p}^*$ 与原始数据表 $X_{n \times p}$ 越接近,则区间数据主成分分析的方法越为有效。

定义1 区间数据表 $X_{n \times p}$ 经过某种区间数据主成分分析方法后得到拟合区间数据表 $X_{n \times p}^*$,则定义此种区间数据主成分分析方法的有效性指标为

$$r = 1 - \frac{d_{c-\ln r}(X, X^*)}{d_{c-\ln r}(X, \mathbf{0}) + d_{c-\ln r}(X^*, \mathbf{0})} \quad (39)$$

式中: $\mathbf{0}$ 表示每个单元均为 $\eta_e = (\eta_e^c, \eta_e^{\ln r}) = (0, 0)$ 的 $n \times p$ 维区间数据表。由 $d_{c-\ln r}$ 距离的非负性和三角不等式易知有效性指标 r 满足 $0 \leq r \leq 1$ 。当 r 越接近 1,则 $X_{n \times p}$ 与 $X_{n \times p}^*$ 越近似,即区间主成分分析方法越有效;反之,当 r 越接近 0,则区间主成分分析方法有效性越低。

4.3 不同区间 PCA 方法有效性比较的仿真实验

首先要构造样本区间数据表 $X_{n \times p}$,其中每个元素 $x_{ij} = (x_{ij}^c, x_{ij}^{\ln r})$ 通过产生随机数的方法分别得到服从均匀分布的中心和对数-半长,从而可得对应的区间上下界。其次分别用 C-lnR PCA、C-PCA 和 V-PCA 3 种方法对 $X_{n \times p}$ 进行区间数据主成分分析得到对应的拟合区间数据表 $X_{n \times p}^*$,继而由式(39)可计算出每种方法的有效性指标。

为更加全面分析区间数据主成分分析方法的有效性,对每种区间数据主成分分析方法,分别取样本个数为 $n = 6, 12, 24, 48$, 变量个数为 $p = 4, 8, 12$ 。这里对第 j ($j = 1, 2, \dots, p$) 个区间变量 X_j ,假设其中心和对数半长分别在区间 $[c, d]$ 和 $[f, g]$

上服从均匀分布。在每次实验中,首先对区间变量 X_j 的中心和对数半长分别随机选择 n 个单值数据样本,从而构成样本规模为 n , 变量个数为 p 的区间数据表;然后用 3 种方法对区间数据表进行主成分分析,采用 Kaiser-Guttman 准则决定最终保留的主成分个数,即选取那些特征值大于 1 的主成分,而舍弃那些比原变量所包含信息更少(特征值小于 1)的主成分。

对样本个数和变量个数的每种组合方式分别进行 200 次重复实验,并计算出 200 次重复实验的有效性指标的平均值,得到如表 2 所示的结果。

表2 C-lnR PCA、C-PCA 和 V-PCA 有效性指标的平均值

Table 2 Average values of validity index of C-lnR PCA, C-PCA and V-PCA

方法	$p = 4$			
	$n = 6$	$n = 12$	$n = 24$	$n = 48$
C-lnR PCA	0.717 7	0.664 1	0.657 9	0.627 8
C-PCA	0.668 1	0.644 2	0.638 0	0.623 1
V-PCA	0.626 7	0.623 3	0.616 5	0.603 9
方法	$p = 8$			
	$n = 6$	$n = 12$	$n = 24$	$n = 48$
C-lnR PCA	0.749 6	0.698 5	0.665 8	0.640 5
C-PCA	0.665 1	0.641 4	0.633 2	0.620 6
V-PCA	0.653 4	0.632 2	0.622 8	0.620 0
方法	$p = 12$			
	$n = 6$	$n = 12$	$n = 24$	$n = 48$
C-lnR PCA	0.792 9	0.723 0	0.681 9	0.651 8
C-PCA	0.667 6	0.655 4	0.636 6	0.619 9
V-PCA	0.661 3	0.641 8	0.627 0	0.612 5

从表 2 中可以得出以下结论:

1) 当区间变量个数不变时,随着样本数目的增加,C-lnR PCA、C-PCA 和 V-PCA 这 3 种区间数据主成分分析方法的有效性均有所降低。这是由于随着样本量增大,数据集的多样性增加,往往需要提取更多的主成分才能较好地涵盖所有样本的特征,根据 Kaiser-Guttman 准则,包含信息较少的主成分被舍弃,从而使得数据集中部分信息有所损失,导致 3 种方法的有效性随着样本量的增加

有所下降。

2) 当样本较少时, C-lnR PCA 相比于 C-PCA 和 V-PCA 的有效性优势较为明显, 随着样本量的增加, 3 种方法的差异逐渐缩小。C-lnR PCA 通过中心和对数半长 2 个特征来构造协方差矩阵, 更能充分刻画区间数据表包含的信息, 尤其是在样本量较小时有着更好的表现。随着样本量的增大, 3 种方法提取主成分后损失的信息都随之增加, C-lnR PCA 方法仍然优于 C-PCA 和 V-PCA, 但优势不再明显。

3) 当区间样本数量不变时, 随着区间变量个数的增加, C-lnR PCA、C-PCA 和 V-PCA 这 3 种方法的有效性指标整体上均呈现上升趋势。

4) 在任何一种样本数量与变量个数组合的情形下, C-lnR PCA 的有效性指标都要高于 C-PCA 和 V-PCA 方法, 即采用 C-lnR PCA 对区间数据表进行主成分分析的有效性更高。

5 结 论

基于中心-对数半长的表达形式, 本文将区间数据表征为由中心和对数半长组成的二维数组, 建立了区间数据和区间向量的运算法则, 并在此框架下提出了一种新的区间数据主成分分析方法, 具有以下优点:

1) 对区间半长取对数的处理可以避免出现最终得到的区间主成分半长为负数的不合理情形。

2) 尽可能地保持降维前后点点之间位置关系的一致性。

3) 所需存储空间小, 空间复杂度较低; 计算过程简单, 时间复杂度较低。

4) 与经典的 V-PCA 和 C-PCA 相比, 对区间数据表进行主成分分析的有效性更好。

5) 原始高维空间被投影至低维空间, 多种经典的统计分析方法可以被运用从而进一步提取有效信息, 同时能够在低维空间中描绘原始高维空间中的样本点, 使得多变量区间数据的可视化成为可能。

参考文献 (References)

- [1] WOLD S, ESBENSEN K, GELADI, P. Principal component analysis[J]. *Chemometrics and Intelligent Laboratory Systems*, 1987, 2(1-3): 37-52.
- [2] 任若恩, 王惠文. 多元统计数据分析: 理论、方法、实例[M]. 北京: 国防工业出版社, 1997: 92-95.
REN R E, WANG H W. Multivariate statistical data analysis: Theory, method and examples[M]. Beijing: National Defense Industry Press, 1997: 92-95 (in Chinese).
- [3] SPETSIERIS P G, MA Y, DHAWAN V, et al. Differential diagnosis of parkinsonian syndromes using PCA-based functional imaging features[J]. *NeuroImage*, 2009, 45(4): 1241-1252.
- [4] 胡艳, 王惠文. 一种海量数据的分析技术——符号数据分析及应用[J]. *北京航空航天大学学报*, 2002, 17(2): 40-44.
HU Y, WANG H W. A new data mining method based on huge data and its application[J]. *Journal of Beijing University of Aeronautics and Astronautics*, 2002, 17(2): 40-44 (in Chinese).
- [5] DIDAY E. Thinking by classes in data science: The symbolic data analysis paradigm: Symbolic data analysis[J]. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2016, 8(5): 172-205.
- [6] 张寅, 王岩, 王惠文. 重点学术期刊专项基金管理中的期刊评价——基于简化的区间数据主成分分析方法[J]. *管理科学学报*, 2010, 13(7): 92-98.
ZHANG Y, WANG Y, WANG H W. Evaluating of academic journals in management of key academic journal fund: An application of simplified principal component analysis based on interval data[J]. *Journal of Management Sciences in China*, 2010, 13(7): 92-98 (in Chinese).
- [7] CAZES P, CHOUAKRIA A, DIDAY E, et al. Extension de l'analyse en composantes principales à des données de type intervalle[J]. *Revue de Statistique Appliquée*, 1997(3): 5-24.
CAZES P, CHOUAKRIA A, DIDAY E, et al. Extending principal component analysis to interval data[J]. *Applied Statistics Review*, 1997(3): 5-24 (in France).
- [8] DIDAY E, BOCK H H. Analysis of symbolic data: Exploratory methods for extracting statistical information from complex data[J]. *Journal of Classification*, 2000, 18(2): 291-294.
- [9] 王惠文, 李岩, 关蓉. 两种区间数据主成分分析方法的比较研究[J]. *北京航空航天大学学报*, 2010, 24(4): 86-89.
WANG H W, LI Y, GUAN R. A comparison study of two methods for principal component analysis of interval data[J]. *Journal of Beijing University of Aeronautics and Astronautics*, 2010, 24(4): 86-89 (in Chinese).
- [10] CHOUAKRIA A, DIDAY E, CAZES P. Vertices principal components analysis with an improved factorial representation[C]// *Proceedings of the 6th Conference of the International Federation of Classification Societies (IFCS-98)*. Berlin: Springer, 1998: 397-402.
- [11] LAURO C N, PALUMBO F. Principal components analysis of interval data: A symbolic data analysis approach[J]. *Computational Statistics*, 2000, 15(1): 73-87.
- [12] 郭均鹏, 李文华. 基于经验相关矩阵的区间主成分分析[J]. *管理科学学报*, 2008, 11(3): 49-52.
GUO J P, LI W H. Interval PCA based on empirical correlation matrix[J]. *Journal of Management Sciences in China*, 2008, 11(3): 49-52 (in Chinese).
- [13] PALUMBO F, LAURO C N. A PCA for interval-valued data based on midpoints and radii[C]// *Proceedings of the International Meeting of the Psychometric Society IMPS2001*. Berlin: Springer, 2003: 641-648.
- [14] 郭均鹏, 李文华. 基于误差理论的区间主成分分析及其应

- 用[J].数理统计与管理,2007,26(4):636-640.
- GUO J P, LI W H. Principle component analysis based on error theory and its application [J]. Application of Statistics and Management, 2007, 26(4): 636-640 (in Chinese).
- [15] WANG H H, GUAN R, WU J J. CIPCA: Complete-information-based principal component analysis for interval-valued data [J]. Neurocomputing, 2012, 86(5): 158-169.
- [16] 侯自盼, 李生刚. 一种针对区间型数据的新主成分分析法[J]. 纺织高校基础科学学报, 2016, 29(2): 184-189.
- HOU Z P, LI S G. A new principal component analysis method for interval data [J]. Basic Sciences Journal of Textile Universities, 2016, 29(2): 184-189 (in Chinese).
- [17] 刘清贤. 区间型符号数据主成分分析及有效性研究[D]. 西安: 西安科技大学, 2019: 19-24.
- LIU Q X. Principal component analysis of interval symbol data and validity study [D]. Xi'an: Xi'an University of Science and Technology, 2019: 19-24 (in Chinese).
- [18] 郭均鹏, 李汶华. 一种区间 PCA 的效度分析方法[J]. 系统工程学报, 2009, 24(2): 226-230.
- GUO J P, LI W H. Analysis of validity of the PCA for interval data [J]. Journal of Systems Engineering, 2009, 24(2): 226-230 (in Chinese).

A principal component analysis of interval data based on center and log-radius

ZHAO Qing^{1,2}, WANG Huiwen^{1,3}, WANG Shanshan^{1,2,*}

(1. School of Economics and Management, Beihang University, Beijing 100083, China;

2. Beijing Key Laboratory of Emergency Support Simulation Technologies for City Operations, Beijing 100083, China;

3. Beijing Advanced Innovation Center for Big Data and Brain Computing, Beihang University, Beijing 100083, China)

Abstract: In order to study the dimension reduction and visualization of multivariate interval data, a two-dimensional array including center and log-radius is used as the expression of interval data. Then the algebraic algorithm of interval data is given, and a new Principal Component Analysis (PCA) method of interval data is proposed on this basis. The processing of the logarithm of interval radius ensures the rationality that the range of the final interval principal components are non-negative. The calculation of this new method is simple, and the complexity is low. Furthermore, the change of the relative position between the points in the sample group before and after the dimension reduction is as small as possible. By reducing the dimension of variables in the high-dimensional space, various classical statistical analysis methods can be used. Besides, the sample points in the original high-dimensional space can be depicted in the low-dimensional space, which makes it possible to visualize multivariate interval data. The results of simulation experiment verify the effectiveness of the proposed method.

Keywords: interval data; Principal Component Analysis (PCA); center and log-radius; dimension reduction; covariance matrix

http://bhxb.buaa.edu.cn jbuua@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2020.0230

压电周期板中耦合禁带影响规律分析



姜周¹, 李琳^{1,2}, 范雨^{1,2,*}, 王文君¹

(1. 北京航空航天大学 能源与动力工程学院, 北京 100083;

2. 北京航空航天大学 先进航空发动机协同创新中心, 北京 100083)

摘 要:近年来,利用周期结构中弹性波禁带特性减振的研究思路受到了广泛关注。针对以往周期结构难以实现宽频且可调禁带的问题,设计了一种含电路网络的压电周期结构。该结构中的弯曲波和电路网络中的电波通过压电效应可以产生一个较宽的耦合禁带,且通过调整电路参数就可以达到调节禁带位置的目的。首先,为了高效计算该结构的波动特性,开发了适用于压电周期结构的减缩波有限元算法,该算法可以在保证结果准确性的基础上节约90%以上的计算时间。利用该算法研究了压电材料尺寸和形状对耦合禁带性能的影响。结果表明:相同电学参数下,随着压电片尺寸的增大,耦合禁带向低频移动,且禁带带宽增加;相同面积下含圆形和方形压电片的机电系统内耦合禁带差异较小,即形状对耦合禁带的分布影响不大。其次,针对不同尺寸和形状的机电系统,设计了电学参数使得在同一频率附近产生耦合禁带,并分析了其性能差异。最后,为了证明耦合禁带的减振效果,设计了一种有限压电周期板模型,其强迫响应的结果证明了耦合禁带对结构内弹性波可以进行有效调控。

关键词:压电材料;周期结构;耦合禁带;波有限元算法;减缩算法;参数分析

中图分类号: V241.8

文献标志码: A **文章编号:** 1001-5965(2021)07-1422-16

弹性波在周期结构中传播时,由于受到内部周期机械阻抗的作用,形成了特殊的色散关系(波长和频率的关系,又称能带结构),使得在某些频率内弹性波无法传播,即为禁带^[1-2]。这一特性在振动控制^[3-6]、能量收集^[7]以及结构健康检测^[8]等方面都有广阔的应用前景,因此受到了广泛关注。其中如何在特定位置构造低频且宽频的禁带一直是研究的热点问题^[9-11]。

常见的禁带构造机理有3种,分别可以构造布拉格禁带、局域共振禁带和耦合禁带。布拉格禁带是周期性变化的材料或几何特性与弹性波相互作用形成的,结构中各组分的弹性常数和密度越大越容易产生禁带。这一禁带由于其宽频特性

在减振方面有一些应用^[12-14],但其受制于产生的布拉格条件,导致该禁带往往位于高频,因此难以实现低频减振^[15-18]。为了突破这一限制,Liu等^[19]提出了局域共振禁带的构造机理。该禁带主要由结构中谐振子的共振产生,导致禁带波长可以随其共振频率变化,从而实现低频隔振。然而其禁带带宽很窄,虽然有研究对其进行拓宽。如朱兴一等^[20]就研究了谐振子的排列方式、材料性质等对禁带调节的影响,以获得宽频禁带。但这种定性研究大多是为了研究禁带的演变规律,而非为了满足工程减振的应用需求^[3,10-21]。耦合禁带是结构中2种波由于耦合效应产生的^[22-23],调节一种波的传播特性产生的耦合禁带可以影响

收稿日期: 2020-06-01; 录用日期: 2020-09-25; 网络出版时间: 2020-09-30 11:12

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20200929.1830.004.html

基金项目: 国家自然科学基金(51675022,11702011); 航空科学基金(2019ZB051002)

*通信作者: E-mail: fanyu04@buaa.edu.cn

引用格式: 姜周, 李琳, 范雨, 等. 压电周期板中耦合禁带影响规律分析[J]. 北京航空航天大学学报, 2021, 47(7): 1422-1437.
JIANG Z, LI L, FAN Y, et al. Influence analysis of coupled band gap in piezoelectric periodic plate[J]. Journal of Beijing University of Aeronautics and Astronautics, 2021, 47(7): 1422-1437 (in Chinese).

另一种波,因此具有丰富的设计空间和调节裕度,为突破布拉格禁带和局域共振禁带的局限提供可能。但机械结构内耦合禁带^[24-25]随着结构确定而确定,因此无法针对特定频带进行设计。近年来,压电材料的出现为实现可调耦合禁带提供了可能。

压电材料作为一种成熟的商用智能材料,结合电路后可以对结构中的能量进行收集或耗散^[8,26-29]。因此,Li等^[30]在压电周期板中施加了电感电路网络,形成压电网络板,利用其中的电波和结构中的弯曲波通过压电效应进行耦合,从而产生了可调且频带较宽的耦合禁带。此处的“电波”实际上是一种电磁波,只是由于存在电感等电学元件,导致其波长增加到可以与弯曲波在弹性波传导的低频段产生耦合禁带的程度。故分析压电结构时,一般认为该电磁波的磁场分量与电场分量相比可以忽略^[31]。用术语“电波”来代替一般意义上的“电磁波”,并只用电场量(电势)对其进行描述。

考虑到文献[30]是针对可调耦合禁带生成的可行性研究,并未分析不同压电片形状对耦合禁带的影响。实际上,大部分基于压电周期结构的研究都集中在通过设计主动或被动电路的方式,以调整禁带的位置和带宽^[32-33],较少有研究探索压电片形状和尺寸对禁带特性的影响。然而为了在低频得到宽频禁带,压电片几何尺寸的设计也是一个十分关键的因素。同时在航空航天等领域,轻质化一直是设计的核心需求,对于附加质量和形状有着严格的限制。因此为了在最小质量比的条件下得到最佳性能的耦合禁带,需要从压电网络板中耦合禁带的影响规律入手,从而为选取最佳的压电片尺寸提供理论依据。

考虑到研究尺寸变化需要计算多种模型的频散曲线,为了提高计算效率,采用了减缩的波有限元算法^[34]。波有限元算法的主要优势在于可以借助商业有限元软件建立模型,导出动力学矩阵后施加周期性边界条件便可求解系统的频散特性,确定禁带分布^[35]。在减缩模型方面,Zhou等^[36-37]将固定界面模态综合法^[38]与波有限元算法结合,将元胞(结构最小可重复单元)内部自由度用模态坐标进行代替,从而进一步提高了计算效率。但这种针对机械系统的减缩波有限元算法无法直接用于机电耦合系统。因此,对该算法进行了改进,推导了适用于压电周期板的减缩波有限元算法。

考虑到结构中主要依赖弯曲波和电波的耦合,因此对保留模态阶数和仅保留弯曲模态的情况进行了误差分析。在此基础上,针对市场上常

见的圆形和正方形压电片,研究了其尺寸对耦合禁带性能的影响。同时设计电学参数,对比了不同压电模型在同一位置产生耦合禁带时的性能差异。为工程应用中如何在特定频带设计合适的耦合禁带提供指导。最后利用强迫响应对耦合禁带的减振效果进行了验证。为压电周期结构的合理设计提供了计算工具和方法指导。

1 压电结构的减缩波有限元算法

用波有限元算法分析周期结构的波传导特性包含2个主要步骤。首先,通过商用软件或有限元子程序获得一个元胞的动力学方程组;其次,在元胞的边界自由度(与相邻元胞共有)上施加周期边界条件,构造求解自由波动特性的特征值问题。对于压电结构的波传导特性分析也是如此,只需在建立元胞动力学模型时考虑压电材料及其电路的影响即可。本节将按照上述步骤阐述具有网络电路的压电板的波动特性分析流程,并着重介绍提出的元胞动力学模型减缩算法。

1.1 压电网络板的元胞动力学模型

本文研究的压电网络板如图1(a)所示。

图1(a)中压电片周期性贴在基板上,相邻压电片通过相同的电感相连,以构成电路网络。该周期结构可以由虚线方框内的元胞沿 x 方向和 y 方向延拓得到。考虑到机电耦合系统内含有电场和机械场的自由度,在后续处理时需要分别处理,故画出元胞所有机械场自由度向量 \mathbf{a} 分布如图1(b)所示。由图1(a)可知,施加电路网络后,会在元胞边界处增加4个电学节点,其分布如图1(c)所示。

考虑到结构中不只存在一种机械波,需要确保其他波(如剪切波等)不会对耦合禁带的减振效果产生影响^[11]。因此如图1(c)所示,将压电片对称贴在基板两面的相同位置,且极化方向相同,从而保证弯曲波上下表面刚度一致,结构内的剪切波不会导致元胞发生弯曲变形。图中, v 表示单节点的自由度。另外设置压电片与基板相连接面接地,另一面并联后与电感电路相连,可以使拉压波产生的电荷相互抵消,从而拉压波也可以与电波解耦^[10]。

由于难以获得图1(a)中压电网络板波动特性的理论解,故通过波有限元算法对图1(c)中的元胞进行建模和求解,以获得系统内的禁带分布。由周期结构理论^[39]可知,系统波动特性求解时需要给元胞施加复数形式的边界条件,而大部分有限元软件难以完成这一设置。同时利用有限

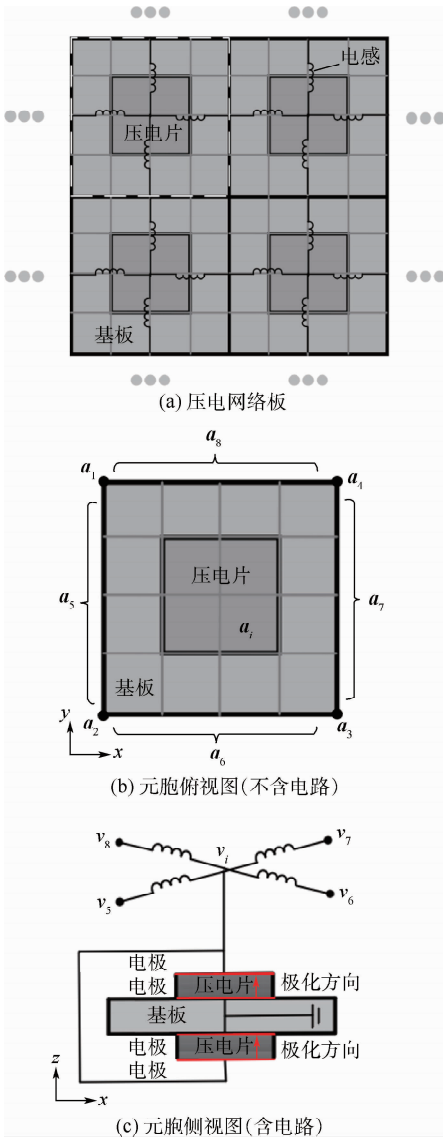


图1 压电网络板及元胞示意图

Fig.1 Illustration of plate with piezoelectric network and its unit cell

元软件建立元胞模型时且不支持导出含电路网络的动力学矩阵,故难以直接建立这一元胞的动力学方程。因此此处利用软件 ANSYS 建立该压电系统的机械场模型,导出其动力学矩阵后,编程实现与电场动力学矩阵的组集,以得到压电网络板元胞的动力学方程。对其施加边界条件后,可求解其波动特性。基板选取 SHELL181 单元,压电片由 SOLID5 单元构成。基板材料为环氧树脂,密度为 1180 kg/m^3 ,弹性模量为 $4.35 \times 10^9 \text{ N/m}^2$,泊松比为 0.37。压电片材料为 PZT-5H,其材料常数如表 1 所示,表中: ρ_p 为压电片密度, $s_{11}^E, s_{12}^E, s_{55}^E$ 为弹性柔顺常数, d_{31} 为压电应变常数, ϵ_{33}^T 为厚度方向的恒应力介电常数。固定基板尺寸为 $80 \text{ mm} \times 80 \text{ mm} \times 5 \text{ mm}$,选取最常用的正方形和圆形 2 种形状的压电片,分别研究其长度和直径对耦合禁带的影响。

表 1 压电片材料参数

Table 1 Material parameters of piezoelectric patches

参数	数值
$\rho_p / (\text{kg} \cdot \text{m}^{-3})$	7 500
$s_{11}^E / (\text{m}^2 \cdot \text{N}^{-1})$	13×10^{-12}
$s_{12}^E / (\text{m}^2 \cdot \text{N}^{-1})$	-4.29×10^{-12}
$s_{55}^E / (\text{m}^2 \cdot \text{N}^{-1})$	22×10^{-12}
$d_{31} / (\text{C} \cdot \text{N}^{-1})$	-1.86×10^{-10}
$\epsilon_{33}^T / (\text{F} \cdot \text{m}^{-1})$	3.009×10^{-8}

同时为了纵向对比形状对禁带的影响,设置每一组正方形和圆形压电片的面积相等,即 $L_p^2 = \pi D_p^2 / 4$, L_p 为正方形压电片边长, D_p 为圆形压电片直径。

建立不含电路的压电网络板元胞有限元模型后,可以得到如下的动力学方程:

$$\begin{cases} m\ddot{\mathbf{a}} + \mathbf{k}\mathbf{a} + \mathbf{k}_p v_i = \mathbf{f} \\ \mathbf{k}_p^T \mathbf{a} - C_p v_i = Q_i \end{cases} \quad (1)$$

式中: \mathbf{a} 为节点位移向量; \mathbf{f} 为节点力向量; Q_i 为压电片表面电荷量; v_i 为压电片上表面电势; \mathbf{m} 为元胞的质量矩阵; \mathbf{k} 为元胞的刚度矩阵; \mathbf{k}_p 为耦合矩阵; C_p 为常应变下压电片的电容值,下标 p 表示压电片相关参数;下标 i 为内部节点;上标 T 为转置符号。

为了方便建立该电感电路网络的动力学方程,利用有限元算法的思路,建立单个电感的单元动力学矩阵后,依据节点位置进行组集,可得该电路网络的动力学方程如下:

$$\begin{bmatrix} 4 & -1 & -1 & -1 & -1 \\ -1 & 1 & 0 & 0 & 0 \\ \frac{1}{\omega^2 L_e} & -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} v_i \\ v_5 \\ v_6 \\ v_7 \\ v_8 \end{bmatrix} = \begin{bmatrix} Q_i \\ Q_5 \\ Q_6 \\ Q_7 \\ Q_8 \end{bmatrix} \quad (2)$$

式中: L_e 为网络中的电感值; ω 为谐响应下的频率; $\mathbf{v} = [v_i \ v_5 \ v_6 \ v_7 \ v_8]^T$ 为电学节点的电势; $\mathbf{Q} = [Q_i \ Q_5 \ Q_6 \ Q_7 \ Q_8]^T$ 为电学节点对应的电荷分布;下标 5、6、7、8 分别表示左边节点、底边节点、右边节点和上端节点,其位置如图 1(c) 所示。

由于施加电路网络后,压电片上表面电势 v_i 的电荷满足 $Q_i = 0$ 。将方程和按照这一关系进行组集,得到压电网络板元胞的动力学方程如下:

$$\begin{bmatrix} \mathbf{k} - \omega^2 \mathbf{m} & \mathbf{k}_p & \mathbf{0} \\ \mathbf{k}_p^T & C_p - \frac{4}{\omega^2 L_e} & -\mathbf{E}_{ib} \\ \mathbf{0} & -\mathbf{E}_{ib}^T & -\mathbf{E}_{bb} \end{bmatrix} \begin{bmatrix} \mathbf{a} \\ v_i \\ \mathbf{v}_b \end{bmatrix} = \begin{bmatrix} \mathbf{f} \\ 0 \\ \mathbf{Q}_b \end{bmatrix} \quad (3)$$

式中: \mathbf{E} 为电场的动刚度矩阵,其中的分块矩阵为

$$\mathbf{E}_{bb} = -\frac{1}{\omega^2 L_e} \mathbf{I}_4, \mathbf{I}_4 \text{ 为 } 4 \text{ 阶单位矩阵; } \mathbf{E}_{ib} = -\frac{1}{\omega^2 L_e} \cdot$$

$$[-1 \quad -1 \quad -1 \quad -1]; \mathbf{v}_b = [v_5 \quad v_6 \quad v_7 \quad v_8]^T$$
 为电路网络边界节点的电势自由度; $\mathbf{Q}_b = [Q_5 \quad Q_6 \quad Q_7 \quad Q_8]^T$ 为电学边界节点的电势。

对方程施加周期性边界条件后即可求解未减缩模型的波动特性。为了确保计算结果的正确性,每个波长内至少需要有6~12个网格,导致模型自由度较大,计算费时。同时考虑到需要计算不同压电片尺寸和形状下的频散曲线,为了节约计算时间,在求解前对方程进行减缩处理。

1.2 压电网络板元胞减缩模型及波动特性求解

该减缩主要基于Craig-Bampton方法,通过借鉴固定界面模态综合法的思路^[35],将自由度分为保留自由度和减缩自由度,将保留自由度固定后,对减缩自由度进行模态分析,获得将减缩自由度转到模态坐标的转换矩阵,即可完成减缩。由于电场的动刚度矩阵 \mathbf{E} 与频率有关,将电场自由度归为减缩自由度求解得到的固定界面主模态无法进行对角化,导致不同模态间存在耦合,从而截断高阶模态会导致计算出错。故只对机械场内部自由度 \mathbf{a}_i 进行减缩,保留自由度为 $\mathbf{q}_R = [\mathbf{a}_b^T \quad \mathbf{v}_i^T]^T$,下标 R 意为保留节点。包括机械场边界节点 \mathbf{a}_b ,电学内部节点 \mathbf{v}_i ,电学边界节点 \mathbf{v}_b 。按照这一关系对方程中的矩阵重新划分为

$$\left(\mathbf{K} - \omega^2 \mathbf{M} - \frac{1}{\omega^2} \mathbf{H} \right) \begin{bmatrix} \mathbf{a}_i \\ \mathbf{q}_R \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{f}_R \end{bmatrix} \quad (4)$$

式中: $\mathbf{f}_R = [\mathbf{f}_b^T \quad \mathbf{0} \quad \mathbf{Q}_b^T]^T$ 为机械场边界自由度、电学内部自由度以及电学边界自由度对应的边界力,此时只有一个电学内部自由度。根据区分的机械场内部自由度 \mathbf{a}_i 和保留自由度 \mathbf{q}_R ,将式(4)的刚度矩阵 \mathbf{K} 、质量矩阵 \mathbf{M} 和电学矩阵 \mathbf{H} 重新分块,有

$$\mathbf{K} = \begin{bmatrix} \mathbf{K}_{ii} & \mathbf{K}_{iR} \\ \mathbf{K}_{iR}^T & \mathbf{K}_{RR} \end{bmatrix} \quad (5)$$

$$\mathbf{M} = \begin{bmatrix} \mathbf{M}_{ii} & \mathbf{M}_{iR} \\ \mathbf{M}_{iR}^T & \mathbf{M}_{RR} \end{bmatrix} \quad (6)$$

$$\mathbf{H} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{E} \end{bmatrix} \quad (7)$$

由固定界面模态综合法可知,将内部自由度 \mathbf{a}_i 转化到模态坐标 \mathbf{y} 的转换矩阵 \mathbf{B} 为

$$\begin{bmatrix} \mathbf{a}_i \\ \mathbf{q}_R \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Psi} & -\mathbf{K}_{ii}^{-1} \mathbf{K}_{iR} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{y} \\ \mathbf{q}_R \end{bmatrix} = \mathbf{B} \begin{bmatrix} \mathbf{y} \\ \mathbf{q}_R \end{bmatrix} \quad (8)$$

式中: \mathbf{I} 为单位矩阵; $\boldsymbol{\Psi}$ 为固定界面主模态,通过固定保留自由度 $\mathbf{q}_R = \mathbf{0}$,求解如下特征值方程获得

$$\mathbf{K}_{ii} \mathbf{a}_i - \mathbf{M}_{ii} \ddot{\mathbf{a}}_i = \mathbf{0} \quad (9)$$

考虑到高模态对描述内部自由度的贡献较小,故仅保留前 n 阶模态,最终得到的固定界面主模态为 $\boldsymbol{\Psi} = [\boldsymbol{\psi}_1^T \quad \boldsymbol{\psi}_2^T \quad \cdots \quad \boldsymbol{\psi}_n^T]^T$ 。保留模态阶数选取的原则为

$$\omega_i < \alpha_f \omega_N \quad i \in [1, 2, \dots, n] \quad (10)$$

即保留模态的固有频率 ω_i 不超过分析频带的上界 ω_N 的若干倍。通过选取不同的系数 α_f 来控制保留的模态阶数。将式(8)代入方程(4),并左乘转换矩阵 \mathbf{B} 的转置,即可得到减缩后的动力学矩阵:

$$\tilde{\mathbf{D}} = \tilde{\mathbf{K}} - \omega^2 \tilde{\mathbf{M}} - \frac{1}{\omega^2} \tilde{\mathbf{H}} \quad (11)$$

式中: $\tilde{\mathbf{K}} = \mathbf{B}^T \mathbf{K} \mathbf{B}$; $\tilde{\mathbf{M}} = \mathbf{B}^T \mathbf{M} \mathbf{B}$; $\tilde{\mathbf{H}} = \mathbf{B}^T \mathbf{H} \mathbf{B}$ 。此时再将减缩后的动力学矩阵 $\tilde{\mathbf{D}}$ 按内部自由度和边界自由度进行分块,得到减缩后的动力学方程可写为

$$\begin{bmatrix} \tilde{\mathbf{D}}_{ii} & \tilde{\mathbf{D}}_{ib} \\ \tilde{\mathbf{D}}_{ib}^T & \tilde{\mathbf{D}}_{bb} \end{bmatrix} \begin{bmatrix} \mathbf{P}_C \\ \mathbf{q}_b \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{F}_b \end{bmatrix} \quad (12)$$

式中: $\mathbf{P}_C = [\mathbf{y} \quad \mathbf{v}_i]^T$,包括模态坐标 \mathbf{y} 和电学内部自由度 \mathbf{v}_i ; $\mathbf{q}_b = [\mathbf{a}_b^T \quad \mathbf{v}_b^T]^T$ 为广义坐标的边界自由度,包括边界的位移自由度 \mathbf{a}_b 和电势自由度 \mathbf{v}_b ; $\mathbf{F}_b = [\mathbf{f}_b^T \quad \mathbf{Q}_b^T]^T$ 为广义坐标边界节点的力向量。考虑到求解前需要对边界节点施加周期性边界条件,故需要按照节点位置进行划分。另外由于内部节点不受外力,因此可进一步减缩求解矩阵的规模:

$$\hat{\mathbf{D}}_s \mathbf{q}_b = \mathbf{F}_b \quad (13)$$

式中: $\hat{\mathbf{D}}_s = \tilde{\mathbf{D}}_{bb} - \tilde{\mathbf{D}}_{ib}^T \tilde{\mathbf{D}}_{ii}^{-1} \tilde{\mathbf{D}}_{ib}$ 。对比方程(4)可知,此时矩阵 $\tilde{\mathbf{D}}_{ii}^{-1}$ 的规模已大大降低,同时左乘和右乘转换矩阵 \mathbf{B} 后,该矩阵为对角矩阵,求逆时可以节约大量时间。对方程(13)施加边界条件后即可进行波动特性求解。

由Bloch边界理论可知^[39],节点自由度应当满足如下关系:

$$\mathbf{q}_b = \mathbf{T}(\lambda_x, \lambda_y) \hat{\mathbf{q}} \quad (14)$$

式中: $\hat{\mathbf{q}} = [\mathbf{q}_5^T \quad \mathbf{q}_6^T \quad \mathbf{q}_1^T]^T$,下标5、6、1分别代表包含电学自由度和机械自由度的左边节点、底边节点和左下角节点。矩阵 \mathbf{T} 为

$$T(\lambda_x, \lambda_y) = \begin{bmatrix} I & 0 & 0 \\ I\lambda_y & 0 & 0 \\ 0 & I & 0 \\ 0 & I\lambda_x & 0 \\ 0 & 0 & I \\ 0 & 0 & I\lambda_x \\ 0 & 0 & I\lambda_y \\ 0 & 0 & I\lambda_x\lambda_y \end{bmatrix} \quad (15)$$

同样,根据节点力平衡可知:

$$T^T(\lambda_x^{-1}, \lambda_y^{-1})F_b = 0 \quad (16)$$

式中: $\lambda_x = e^{\mu_x}$; $\lambda_y = e^{\mu_y}$; $\mu_x = -jk_x L_x$; $\mu_y = -jk_y L_y$ 分别为沿着 x 和 y 方向的传播常数,其幅值和相位分别反映波在经过一个周期单元的传播后产生的幅值衰减和相位变化。 k_x 和 k_y 为沿着 x 和 y 方向的传播常数的波数,其实部为传播常数,虚部为衰减常数。 L_x 和 L_y 为元胞沿着 x 和 y 方向的尺寸,此处由于选取了正方形元胞,有 $L_x = L_y = 80 \text{ mm}$ 。

将式(14)和式(16)代入方程(13)可得

$$T^T(\lambda_x^{-1}, \lambda_y^{-1})\hat{D}_s(\omega)T(\lambda_x, \lambda_y)\hat{q} = 0 \quad (17)$$

方程中包含 3 个未知量 ω , λ_x 和 λ_y 。其中 (λ_x, λ_y) 表征了板结构内波的方向性分布,可以将其转化为波矢 (k_x, k_y) 。由周期结构理论可知,只需考虑分布在布里渊区 $([-\pi/L_x, \pi/L_x], [-\pi/L_y, \pi/L_y])$ 中波矢 (k_x, k_y) 的分布,就可以刻画系统的传播特性。然而根据结构的对称性,往往只需计算得到位于简约布里渊区 $([0, \pi/L_x], [0, \pi/L_y])$ 的分布。因此在不同频率 ω 下,固定波矢中的一个值以求解另一个。从而可以得到该频率下波空间内所有传播波的分布,如果存在某些角度范围内不存在传播波,则为禁带。

固定频率 ω 和 (λ_x, λ_y) 中的一个(譬如 λ_y),就可以将其转化为二次特征值问题进行 λ_x 的求解^[11],从而获得减缩后压电网络板沿某一方向的波动特性。另外对于减缩模型求解出的特征向量 $\hat{\Phi}$ 与波模态 Φ 的关系为

$$\Phi = BT(\lambda_x, \lambda_y)\hat{\Phi} \quad (18)$$

对方程(4)直接施加周期性边界条件求解,则可得到未减缩模型的频散曲线。

1.3 保留模态对计算精度的影响分析

选取尺寸为 $L_p = 40 \text{ mm}$ 的正方形压电片,设定电感 $L_e = 0.08 \text{ H}$ 。选取 $0 \sim 2000 \text{ Hz}$ 为目标频带,保留模态系数 $\alpha_f = 5$ 。以验证减缩效果。模型 1 有限元模型如图 2 所示。

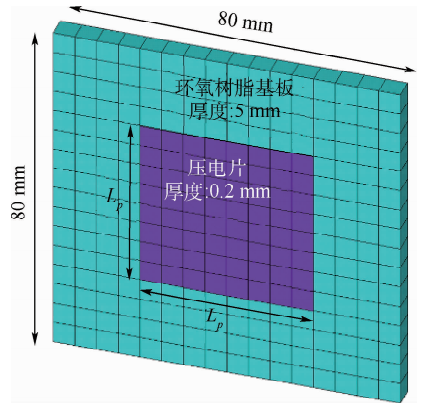


图 2 含正方形压电片元胞有限元模型

Fig. 2 Finite element model of unit cell with square lead zirconate titanate patch

压电片贴于基板中央。计算其 $\lambda_y = 1$ 且未加电路的波动特性,得到系统沿 x 方向的频散曲线,结果如图 3 所示,图中横坐标为频率,纵坐标为传播常数,表明弹性波相位的变化。可见其中存在 3 种波,其中由于压电片在基板中的周期分布,在 $755 \sim 951 \text{ Hz}$ 内产生了波 1 的布拉格禁带。

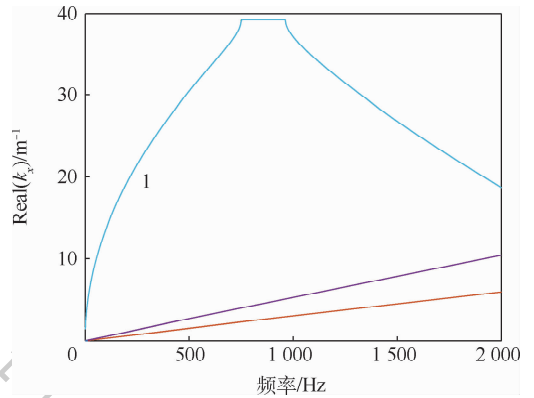


图 3 未加电路时系统传播常数分布

Fig. 3 Distribution of system propagation constants without circuits

施加电路网络后依然令 $\lambda_y = 1$,即求解系统沿 x 方向的频散曲线,结果如图 4 所示,图中横坐标都为频率,图 4(a)中纵坐标为传播常数,图 4(b)中纵坐标为衰减常数。图中虚线为未减缩计算结果,实线为减缩计算结果。

由图 4 可知,减缩前后频散曲线吻合。其中图 4 中波 2 为施加电路网络后产生的电波,该网络形式下,波数随频率增加而增大,经过耦合禁带后进入永久禁带。波 1 和波 2 在目标频带各存在一个布拉格禁带,即满足 $\text{Real}(k_x) = \pi/L_x$ 。同时波 1 和波 2 在 1300 Hz 附近,由于耦合作用产生了耦合禁带,该禁带的衰减常数和带宽都优于波 1 的布拉格禁带。由于压电效应只影响耦合区附近的频散曲线^[22],因此随着耦合结束,波 1 的

波数依然随频率增加而减小,波 2 的波数随频率增加而增大。画出 700 Hz 下各个波的波形如图 5 所示。

可知波 1 和波 2 有弯曲变形,故为机电耦合波,波 3 为纵波,波 4 为剪切波。

减缩矩阵规模和效率提升如表 2 所示。

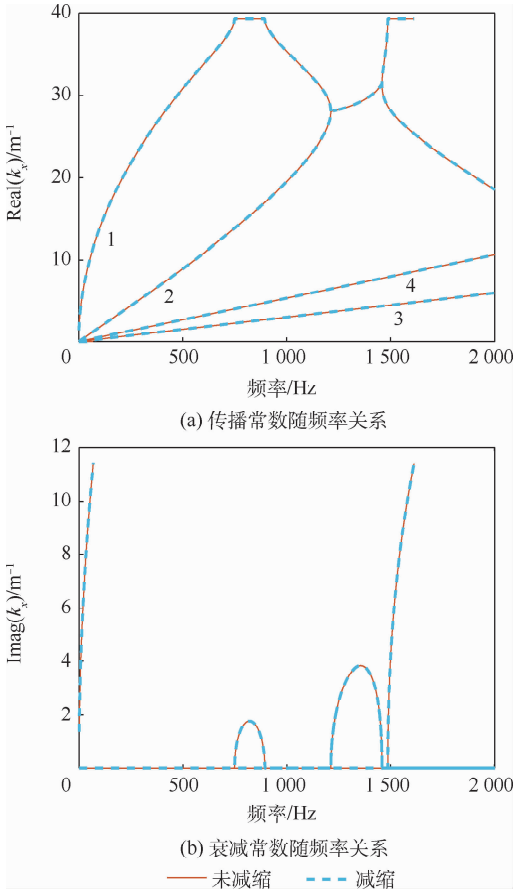


图 4 $\alpha_f = 5$ 时减缩前后频散曲线对比

Fig. 4 Comparison of frequency dispersion curves with/without reduction when $\alpha_f = 5$

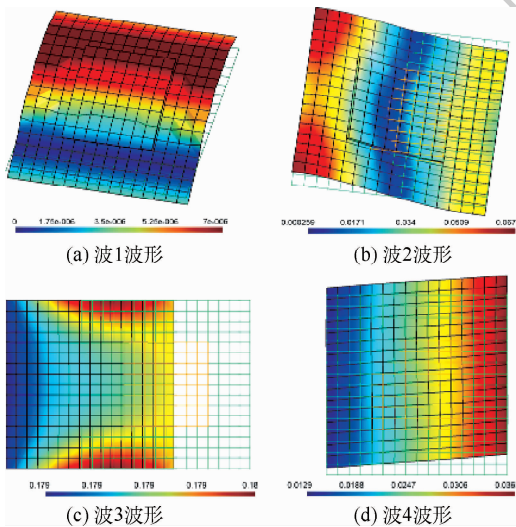


图 5 700 Hz 下各行波波形

Fig. 5 Wave shapes of propagation waves at 700 Hz

表 2 减缩模型对比

Table 2 Comparison between full model and reduced model

模型对比	计算时间/s	矩阵规模	自由度数
完整模型	2 843.17	8 365 × 8 365	1 175
减缩模型	227.98	313 × 313	297
增益/%	91.98	96.25	74.72

由表 2 可知,该减缩算法可以有效提升计算效率。另外选取弯机电耦合波 1 的传播常数做误差分析,结果如图 6 所示。由图 6 可知,其误差基本在 0.01% 附近,可见该算法在提升效率的同时还能保证结果的准确性。

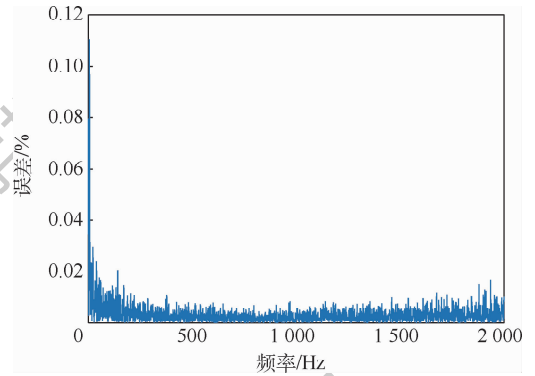


图 6 各频率下模型 1 内波 1 减缩前后误差

Fig. 6 Errors of wave 1 with/without reduction in model 1 at different frequency

值得注意的是,之所以选取保留模态系数 $\alpha_f = 5$,是因为求解特征值方程时发现其特征频率较高,如果 α_f 过低则会导致保留模态数过少而导致计算结果误差较大。图 7 给出了 $\alpha_f = 3$ 时传播常数的对比,图中虚线为减缩计算结果,实线为未减缩计算结果。可见虽然减缩后纵波 3 和剪切波 4 的吻合较好,波 1 远离耦合区的频散曲线也在一定程度上吻合。但波 1 布拉格禁带后的波数以及波 2 的波数计算的误差都较大,因此在进行减缩时需要酌情选取 α_f 的值。

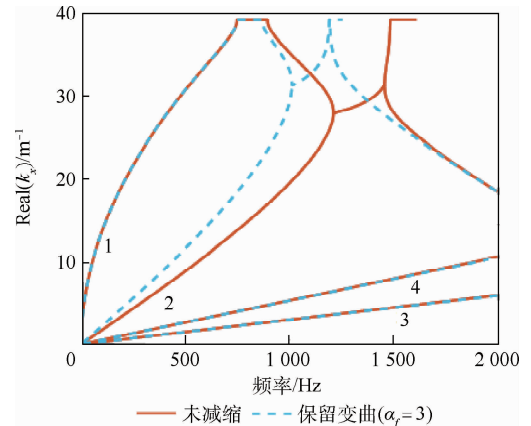


图 7 $\alpha_f = 3$ 时减缩前后传播常数对比

Fig. 7 Comparison of propagation constants with/without reduction when $\alpha_f = 3$

考虑到系数 α_f 对剪切波和纵波影响较小,同时耦合禁带主要是弯曲波和电波为主。因此为了进一步降低保留的模态阶数,在 $\alpha_f = 5$ 时只考虑其中的弯曲模态。其对比如图 8 所示。

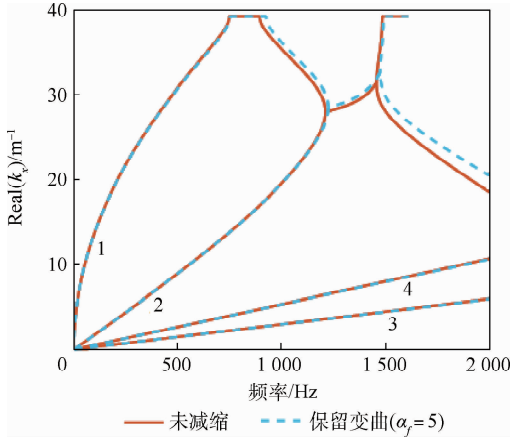


图 8 $\alpha_f = 5$ 时只考虑弯曲模态减缩前后传播常数对比
Fig. 8 Comparison of propagation constants with/without reduction when $\alpha_f = 5$ with only bending modes considered

此时虽然波 1 和波 2 布拉格禁带前的波数计算结果都较好,但布拉格禁带后的波数依然存在可见误差。因此无法对保留的模态阶数进行进一步的选择。

2 结构尺寸对耦合禁带影响分析

由图 4 可知,当电波和弯曲波在耦合禁带内,剪切波和纵波依然可以传播,因此只考虑结构中的传播波无法确定禁带的分布。故采用单波形占比的指标 RSW 以消除纵波和剪切波的影响。这一指标的详细定义参见文献[11]。

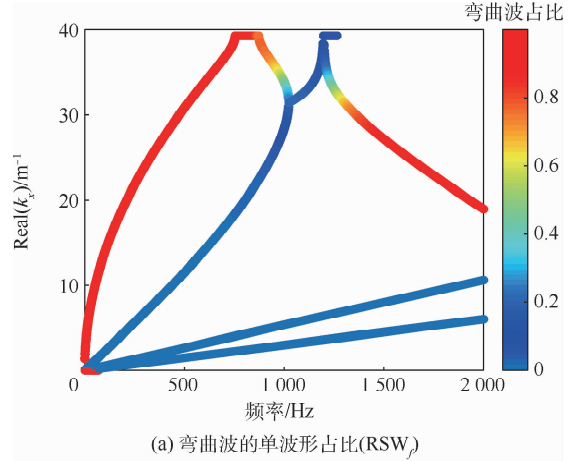
该指标主要通过设定某种特定变形的参考波模态 Φ_{ref} , 计算其与完全波模态 Φ_{full} 的模态置信因子(MAC):

$$RSW = MAC(\Phi_{ref}, \Phi_{full}) = \frac{|\Phi_{ref}^H \cdot \Phi_{full}|^2}{(\Phi_{ref}^H \cdot \Phi_{ref})(\Phi_{full}^H \cdot \Phi_{full})} \quad (19)$$

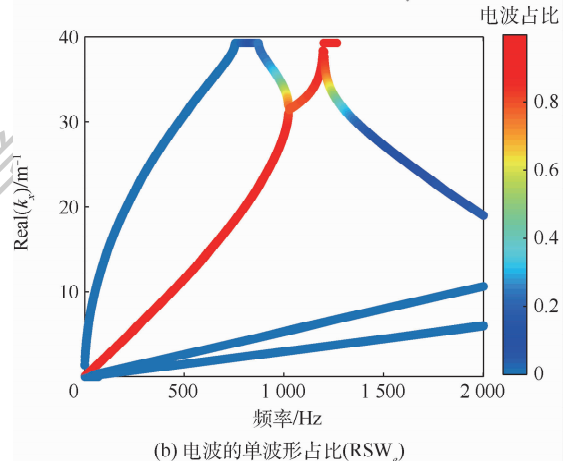
式中:上标 H 为共轭转置。RSW 是位于 $[0, 1]$ 的实数,0 表示完全波模态 Φ_{full} 中不含有变形 Φ_{ref} ,1 则为完全波模态 Φ_{full} 与参考波模态 Φ_{ref} 变形完全一致。每种波的波模态包含 7 类自由度, $\Phi_{full} = [u_x^T, u_y^T, u_z^T, \theta_x^T, \theta_y^T, \theta_z^T, \lambda_e^T]^T$, 即 3 个方向的平动自由度(u_x, u_y, u_z)和 3 个方向的转动自由度($\theta_x, \theta_y, \theta_z$)以及电学自由度 λ_e (磁通量)。对于波模态中弯曲变形的参考波模态有

$$\Phi_{ref} = \Phi_f = [0, 0, u_z^T, \theta_x^T, \theta_y^T, 0, 0]^T \quad (20)$$

下标 f 代表弯曲波(Flexural Wave)。画出电波和弯曲波的占比如图 9 所示。电波对应的参考



(a) 弯曲波的单波形占比(RSW_f)



(b) 电波的单波形占比(RSW_e)

图 9 电波和弯曲波占比分布

Fig. 9 Distribution for RSW_f and RSW_e

波模态为只保留 λ_e 。由图 9(a)可知,弯曲波的 RSW_f 除了耦合禁带附近都接近于 1,表明该波完全由弯曲变形主导;图 9(b)中电波的 RSW_e 也有类似的结论。在耦合禁带附近,弯曲波和电波的单波形占比变化剧烈(由接近 1 下降为 0.5 左右),说明二者存在强烈的能量交换,同时验证了元胞压电片布置形式和极化方向的有效性。设置弯曲波的占比 $RSW_f > 0.2$, 以排除剪切波和纵波的影响。

同时考虑到板为二维结构,需要考虑不同方向上耦合禁带性能的差异。所以计算不同 λ_y 下 λ_x 的值,以得到禁带的方向性分布。故使用减缩算法可以节约大量计算时间。

2.1 正方形压电片尺寸变化

参考表 1,计算 5 种边长不同的正方形压电片的机电系统,具体尺寸为 $L_p = \{20, 35, 40, 45, 50\}$ mm,设定电感 $L_e = 0.1$ H,使得在 $0 \sim 2000$ Hz 产生耦合禁带和布拉格禁带,利用减缩算法研究其波动特性。

当 $L_p = 20$ mm 时,得到 x 方向频散曲线如图 10 所示。其中布拉格禁带范围为 736 ~ 774 Hz,

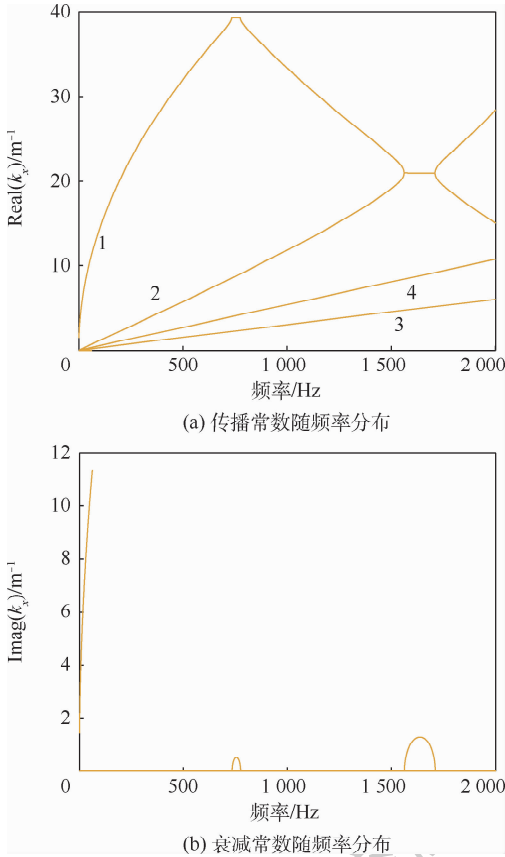


图 10 $L_p = 20$ mm 时频散曲线

Fig. 10 Frequency dispersion curves when $L_p = 20$ mm

耦合禁带范围为 1 562 ~ 1 710 Hz。可知当压电片面积较小时,基板和散射体刚度相差较小,导致布拉格禁带较小,同时将机械能转化为电能的能力有限,因此耦合禁带效果也有限。但耦合禁带依然比布拉格禁带宽 110 Hz,衰减常数是布拉格禁带的 3 倍。

调整 λ_y 的值,反复计算可以得到该结构参数下禁带的方向性分布,如图 11 所示,图中横坐标为频率,纵坐标为禁带存在的角度。图中左侧为布拉格禁带分布,右侧区域为耦合禁带分布,考虑

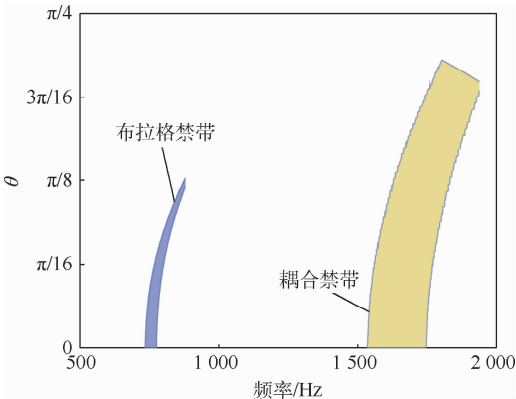


图 11 $L_p = 20$ mm 时禁带方向性分布

Fig. 11 Directional distribution of band gaps when $L_p = 20$ mm

到选取的元胞为正方形,因此根据对称性只用画出 $[0, \pi/4]$ 的区域即可表示其方向性。

由图 11 可知,随着角度增加,布拉格禁带往高频移动,且带宽迅速降低。而耦合禁带虽然也向高频移动,但带宽近似不变。所以当压电片尺寸过小时,难以利用结构中的布拉格禁带,而耦合禁带依然有较好的性能。考虑到研究对象为二维板,因此主要研究其尺寸对禁带方向性的影响, x 方向的波动特性可以从角度分布图中看出,因此后续仅展示角度分布图。

保持电感 $L_e = 0.1$ H 不变,进一步增大其尺寸到 35 mm,可得禁带的方向性分布如图 12 所示。增大压电片尺寸可以明显拓宽布拉格禁带和耦合禁带。在相同电感值的情况下,耦合禁带向低频移动,原因在于增大压电片尺寸使得其本征电容增大,导致电波向低频移动从而耦合禁带出现在较低频率范围。

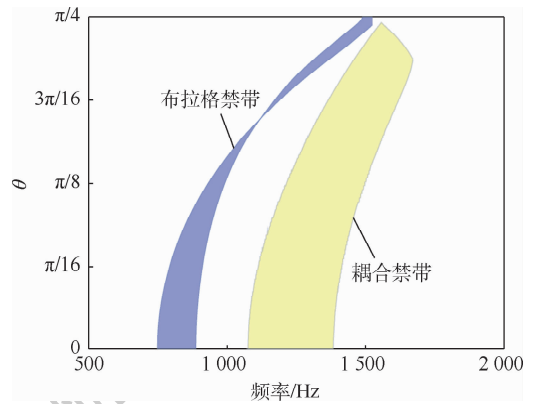


图 12 $L_p = 35$ mm 时禁带方向性分布

Fig. 12 Directional distribution of band gaps when $L_p = 35$ mm

当增大压电片尺寸后,布拉格禁带虽然得到了拓宽,但其带宽随着传播角度增加依然明显衰减。相比于 20 mm 的情况,耦合禁带虽然整体在向低频移动,同时其带宽也随着角度增加而略微降低,即沿着 x 方向带宽最宽。

当 $L_p = 40$ mm 时,画出其方向性分布如图 13 所示。可见当尺寸增大到一定程度后,对布拉格禁带影响近似不变,然而压电片增大导致其电容增加,使得耦合禁带不断向低频移动,与布拉格禁带靠近后,二者相互作用,使得在较高的传播角度上具有较宽的禁带。因此在利用耦合禁带进行方向性隔振时,考虑到最佳角度附近禁带带宽变化剧烈,需要有针对性地设计角度范围,以达到最佳的禁带宽度。

当压电片尺寸增大到 $L_p = 45$ mm 时,在 x 方向上耦合禁带和布拉格禁带也开始相互影响。其

方向性分布如图 14 所示。与 40 mm 的情况相比,虽然其带宽较宽且频率较低,但随着角度增加其带宽迅速降低。此时耦合禁带更适合用于沿 x 方向的隔振或减振,以达到最佳减振效果。

同时,需要配合设计电学参数,使得布拉格禁带和耦合禁带都具有最佳性能。

2.2 圆形压电片尺寸变化

选取圆形的压电片,计算 5 种直径的机电系统,有限元模型如图 16 所示。

设定电感不变 $L_e = 0.1 \text{ H}$,选取圆形尺寸为 $D_p = \{22.57, 39.49, 45.14, 50.78, 56.42\} \text{ mm}$,计算 $0 \sim 2\,000 \text{ Hz}$ 内禁带的波动特性和方向性分布。当 $D_p = 22.57 \text{ mm}$ 时,得到 x 方向频散曲线如图 17 所示。

与同一面积的方形压电片相比,参考图 10 可知,耦合禁带和布拉格禁带都相差不多,对比耦合

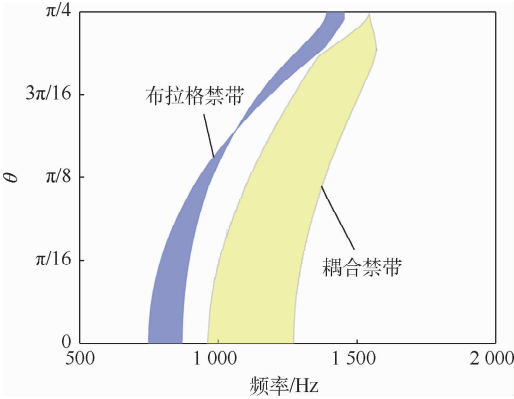


图 13 $L_p = 40 \text{ mm}$ 时禁带方向性分布

Fig. 13 Directional distribution of band gaps when $L_p = 40 \text{ mm}$

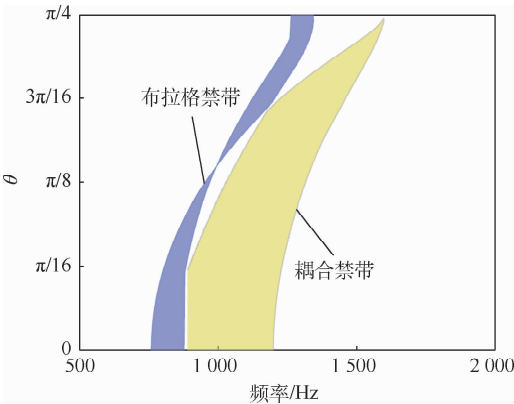


图 14 $L_p = 45 \text{ mm}$ 时禁带方向性分布

Fig. 14 Directional distribution of band gaps when $L_p = 45 \text{ mm}$

当压电片尺寸增大到 $L_p = 50 \text{ mm}$ 时,布拉格禁带和耦合禁带的相互作用越发强烈,此时其方向性分布如图 15 所示。可见二者几乎合二为一,耦合禁带具有了布拉格禁带的特征,即随着角度增加带宽迅速降低,然而原布拉格禁带在角度较高时带宽却近似不变。可见在调整压电片尺寸的

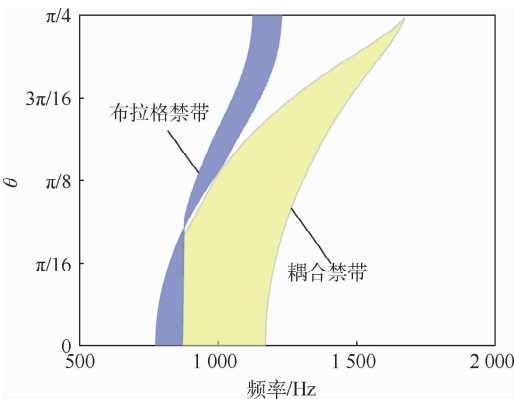


图 15 $L_p = 50 \text{ mm}$ 时禁带方向性分布

Fig. 15 Directional distribution of band gaps with $L_p = 50 \text{ mm}$

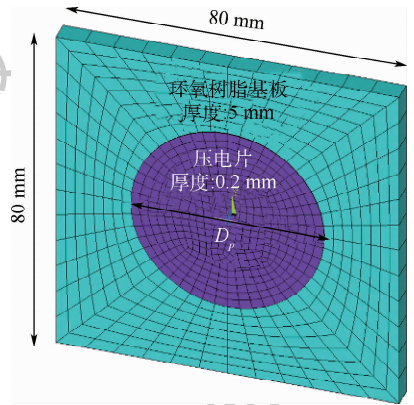
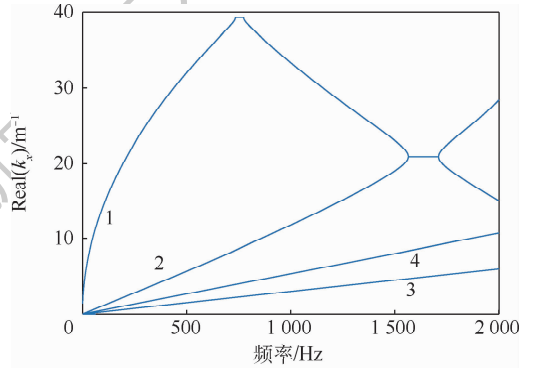
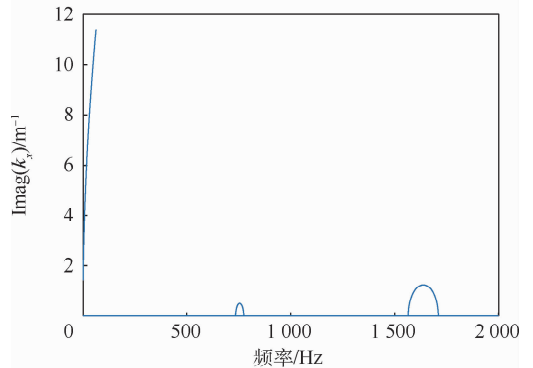


图 16 含圆形压电片元胞有限元模型

Fig. 16 Finite element model of unit cell with circular lead zirconate titanate patch



(a) 传播常数随频率分布



(b) 衰减常数随频率分布

图 17 $D_p = 22.57 \text{ mm}$ 时频散曲线

Fig. 17 Frequency dispersion curves with $D_p = 22.57 \text{ mm}$

禁带边界频率发现圆形压电片的耦合禁带略窄,但二者的中心频率不变。原因在于二者压电片电容近似相等,导致未耦合前电波具有一致的频散特性,因此中心频率不变。然而正方形压电片将机械能转化为电能的能力略优,导致耦合禁带略宽。

对比同一面积的 2 种压电片,得到其方向性分布如图 18 所示。由图可知,同一面积下,不同性状对耦合禁带影响有限,正方形情况下耦合禁带带宽略宽。由于布拉格禁带变化不大,因此并未在图中列出。可见在不同传播方向上,正方形情况依然略优于圆形情况,当传播角度高到一定程度时,耦合禁带带宽才趋于一致。

进一步增加直径至 $D_p = 39.49$ mm,计算其方向性分布与同等面积的正方形压电片对比如图 19 所示。二者依然具有相似的方向性分布,只是随着面积的增大,在 x 方向的带宽差异扩大,这一差异随着角度增加而减小。

进一步增大压电片直径至 45.14 mm(见图 20),由于此时耦合禁带和布拉格禁带相互作用,导致

圆形和方形的耦合禁带在高频时发生微小差异。但二者的中心频率一直保持不变。类似的现象在 $D_p = 50.78$ mm 时都可以看到,如图 21 所示。

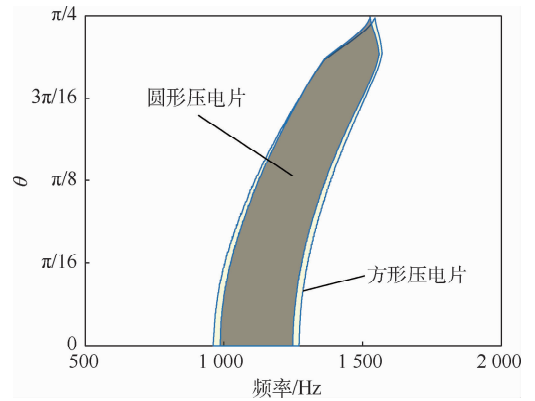


图 20 $D_p = 45.14$ mm 时耦合禁带方向性对比
Fig. 20 Comparison of directional distribution for coupled band gaps with $D_p = 45.14$ mm

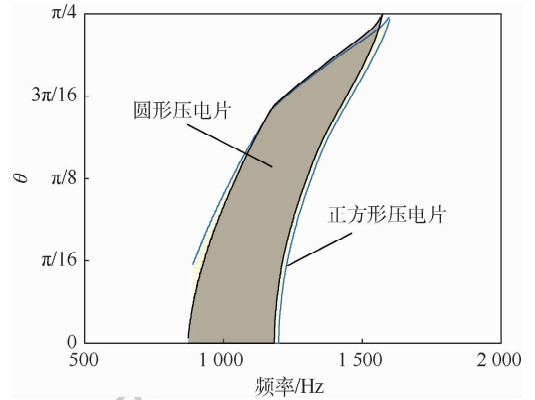


图 21 $D_p = 50.78$ mm 时耦合禁带方向性对比
Fig. 21 Comparison of directional distribution for coupled band gaps with $D_p = 50.78$ mm

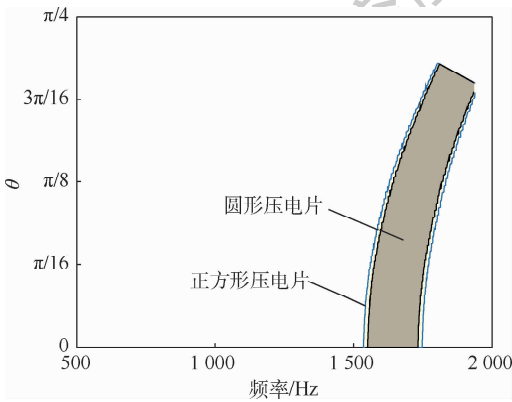


图 18 $D_p = 22.57$ mm 时耦合禁带方向性对比
Fig. 18 Comparison of directional distribution for coupled band gaps with $D_p = 22.57$ mm

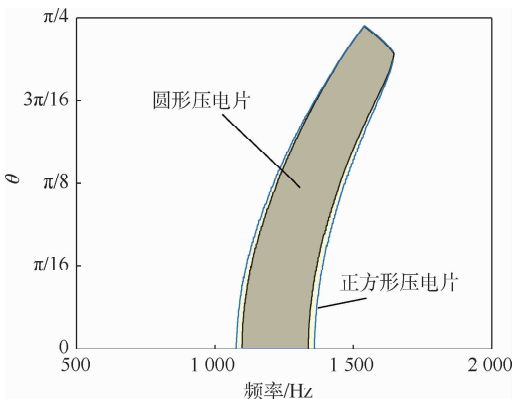


图 19 $D_p = 39.49$ mm 时禁带方向性分布
Fig. 19 Directional distribution of band gaps with $D_p = 39.49$ mm

2.3 不同尺寸压电片电学参数设计

由 2.2 节分析可知,随着压电片面积的增大,不改变电感时,耦合禁带会向低频移动。当与布拉格禁带发生耦合时会导致禁带带宽随着传播角度增加而迅速降低。因此为了避免这一情况需要针对压电片对电学参数进行设计。

根据耦合禁带形成机理,耦合禁带产生于未耦合 2 种波相交的频率附近^[23]。因此设计原则为:通过选取电学参数,使得未耦合电波具有相同的频散曲线。

考虑到压电片相当于电容,与纯电感电路相连后,可以得到耦合前电波的理论波动曲线:

$$\omega = 2 \sin\left(k \frac{L_{\text{all}}}{2}\right) \sqrt{\frac{1}{L_c \tilde{C}_p}} \quad (21)$$

式中: ω 为系统角频率; L_{all} 为基板长度; k 为波数。值得注意的是,此时应当选取压电片开路时的电

容 \tilde{C}_p 与压电片本征电容关系为^[11]

$$\tilde{C}_p = C_p + \mathbf{k}_p^T \mathbf{k}^{-1} \mathbf{k}_p \quad (22)$$

由 2.2 节可知, $L_p = 35 \text{ mm}$, $L_e = 0.1 \text{ H}$ 的参数下耦合禁带效果较好, 因此选为基准。同时由于正方形和圆形压电片对耦合禁带中心频率影响较小, 故同一面积的压电片的电学参数保持一致。由图 4 可知, 电波具有永久布拉格禁带, 该禁带满足波数 $\text{Real}(k) = \pi/L_{\text{all}}$ 。代入式(22), 可得禁带初始频率有 $\omega_0 = 2\sqrt{1/L_e \tilde{C}_p}$, 保证不同电学参数下该禁带初始频率不变, 就可获得不同开路电容下的电感值。

计算不同面积的开路电容后, 得到不同尺寸对应的电感值如表 3 所示。

表 3 压电片结构参数

Table 3 Structural parameters of piezoelectric patches

情况编号	长度/mm	直径/mm	电感/H	覆盖率/%
1	20	22.567 6	0.319	6.25
2	35	39.493 3	0.1	19.14
3	40	45.135 2	0.078	25.00
4	45	50.777 1	0.060 6	31.64
5	50	56.419 0	0.051 2	39.06

注: 表中覆盖率为压电片面积与基板总面积之比。

由表 3 可知, 当压电片面积增大时, 在同一位置产生耦合禁带所需的电感值迅速下降, 对比情况 1 和情况 3 可知, 当面积增大为原来的 4 倍时, 电感变为 24.45%。因此在难以实现大电感的情况时, 可以通过增加压电片的面积使得构造的耦合禁带满足预期效果。

画出情况 1 的方向性分布如图 22 所示。由图 22 可知, 方形压电片性能依然略优于圆形压电片, 但这 2 种形状对耦合禁带的方向性影响较小。

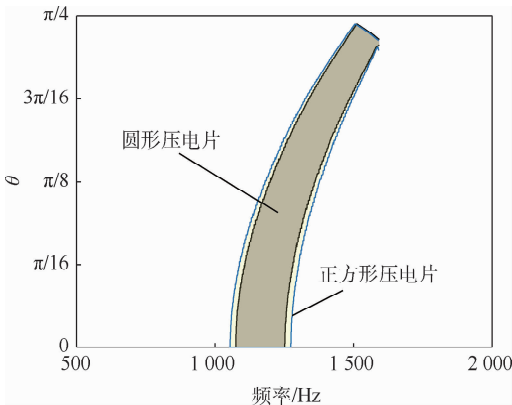


图 22 情况 1 时耦合禁带方向性对比

Fig. 22 Comparison of directional distribution for coupled band gaps with configuration 1

同时对比图 18 发现, 在同一尺寸下, 降低电感会使得耦合禁带向低频移动, 且带宽随角度增加而降低, 在靠近 $\pi/4$ 时, 原来不存在耦合禁带的角度下出现了禁带分布, 即该电学参数下禁带的方向性更好。

由于选取了图 19 的结构和电学参数作为对照, 且表 3 中情况 3 ~ 情况 5 耦合禁带方向性分布类似, 因此为了节约篇幅, 仅展示情况 4 的结果, 如图 23 所示。

由图 23 可知, 虽然设计的电学参数可以使得未耦合电波频散曲线一致, 但由于各尺寸耦合刚度不同, 使得随尺寸增加, 禁带起始频率近似不变, 中心频率往高频移动, 使得耦合禁带整体变宽。

针对方形压电片, 选取 3 个角度, 可以画出禁带和终止频率随尺寸增大的影响, 如图 24 所示, 图中横坐标为情况编号, 纵坐标为频率。

对比情况 4 和情况 5 可知, 当压电片覆盖率从 25% 提升到 39.06% 时, 耦合禁带起始频率近似不变, 终止频率略有上升, 可见其性能增加并不

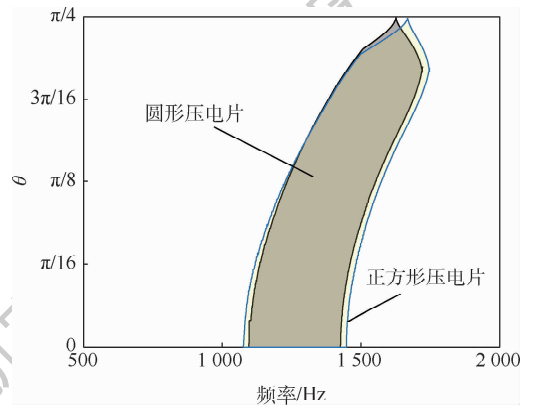


图 23 情况 4 时耦合禁带方向性对比

Fig. 23 Comparison of directional distribution for coupled band gaps with configuration 4

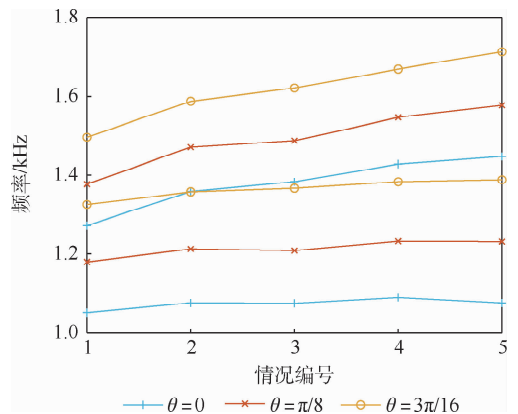


图 24 耦合禁带边界频率对比

Fig. 24 Comparison of boundary frequencies for coupled band gaps

明显。从而在根据实际情况设计压电片时,并不是覆盖率越高越好,需要从实际情况出发,综合考虑覆盖面积和目标性能的影响。

3 耦合禁带强迫响应仿真验证

为了验证耦合禁带的减振效果,选用耦合禁带带宽较窄,衰减常数较低的 20 mm 正方形压电片建立机电系统,设计验证模型如图 25 所示。该模型参考了文献[40-41]中的禁带强迫响应验证的思路,建立含有均质部分和周期部分的模型,相互对照以体现禁带对弹性波的控制作用。其中模型右半部分为压电网络板,相邻压电片施加电感 $L_e = 0.319$ H 以构成电路网络,左半部分为环氧树脂基板作为对照。考虑到其方向性分布如图 22 所示,因此得到特定频率下该结构的动力学响应后,与该频率下耦合禁带的方向性分布对比,即可验证耦合禁带的方向性分布。同时与均质板部分对比,可以研究有限周期结构内耦合禁带的减振性能。

模型采用了四边自由边界条件,在中心施加沿板厚度方向的瞬时力,即图中 F 点。为了降低频散效应,激励信号选为施加汉宁窗的正弦信号。其中心频率设定为 1 140 Hz,信号长度为 3 个周期,最大振幅为 0.5 N,采样频率为中心频率的 30 倍,以保证波包的信号质量。初始位移、速度和加速度都是 0。结构中不含任何形式的阻尼,强迫响应利用 ANSYS17 中的完全法进行求解,从而保证结果的准确性。

为了监测各个方向的传播特性,选取距离激励点 6 个元胞长度(即 0.48 m)圆弧上的单元能量作为指标,考虑到结构的对称性,只选取了上半部分的单元,如图 25 所示。对于压电网络板部分,

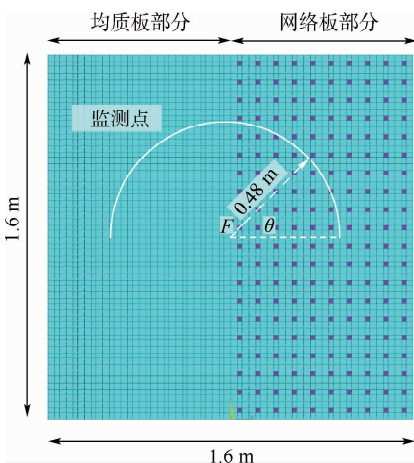


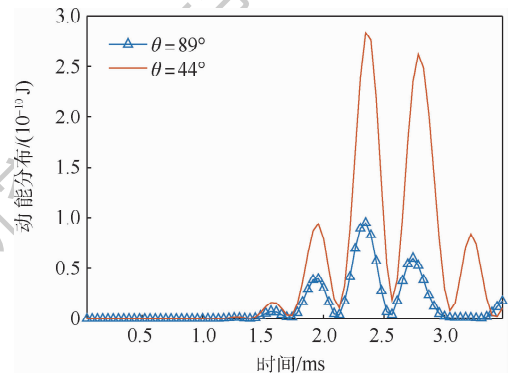
图 25 强迫响应计算模型

Fig. 25 Calculation model for forced response

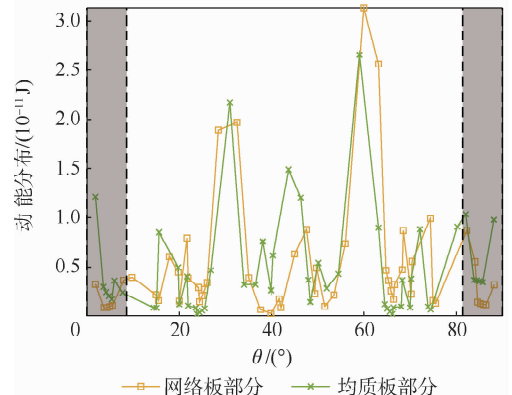
传播角 $\theta = 0$ 即为沿着 x 轴正方向传播,对于均质板部分,传播角 $\theta = 0$ 即为沿着 x 轴负方向传播。计算结果如图 26 所示,图 26(a)为弹性波传播角度为 44° 和 89° 时,动能随时间变化的分布。由于 1 140 Hz 位于耦合禁带内,导致压电网络板内不同角度动能随时间变化分布区分明显,1 140 Hz 时位于 $\theta = 89^\circ$ 的耦合禁带内,位于 $\theta = 44^\circ$ 的通带内。因此可知禁带内的能量得到了明显控制。选取图 26(a)中虚线时刻($t = 1.87 \times 10^{-3}$ s),研究监测点不同方向的能量分布可得图 26(b),图中横坐标为弹性波传播的角度,纵坐标为动能分布。

图 26(b)阴影部分为存在耦合禁带的角度,可见在禁带内,相比于均质板部分,动能大幅下降,通带内二者相差不大,这验证了禁带对弹性波传播的控制能力。

为了进一步比较二者的差异,计算了均质板部分和压电板部分所有动能之和随时间变化的关系曲线,如图 27 所示,图中横坐标为时间,纵坐标为动能分布。图 27(a)为情况 1 对应正方形压电片尺寸下模型的能量分布,可见虽然刚开始时二者相差不大,但随着时间增加,弹性波每经过一次元胞总动能都有所衰减。



(a) 不同角度单元动能随时间分布图



(b) $t = 1.87 \times 10^{-3}$ s 时两部分动能分布对比

图 26 1 140 Hz 下强迫响应计算结果

Fig. 26 Forced response calculation results at 1 140 Hz

图 27(b)为情况 3 对应正方形压电片尺寸下模型的能量分布,可见在耦合禁带频率内(1 300 Hz)依然计算强迫响应,得到能量对比如图 27 所示,可见随着压电片面积增大,能量转化效率提高,对振动控制效果越明显。

图 28 进一步对比了这 2 种情况下,均质板部分和网络板部分同一时刻的能量场分布,图中横坐标为结构 x 方向的坐标,纵坐标为结构 y 方向的坐标。考虑到结构的对称性,仅画出上半部分的能量分布。图 28 中由原点出发的 2 条直线间的角度范围为 2.3 节计算的耦合禁带分布,在 $0^\circ \sim 45^\circ$ 内,情况 1 耦合禁带的涵盖角度为 $0^\circ \sim 9.4^\circ$,情况 3 耦合禁带的涵盖角度为 $7.01^\circ \sim 26.7^\circ$ 。

可见禁带内都没有能量分布。另外可以看出,在均质板部分,由于没有禁带,因此能量沿四周均匀传播。在网络板部分,由于在特定角度范围存在禁带,导致能量主要集中在 45° 方向。

虽然结构内的电学参数并非最优值,导致情况 1 的能量对比差异并不十分明显,但也足以验证耦合禁带对有限周期结构的振动控制能力突出,在低频宽频振动控制领域具有广阔的应用前景。

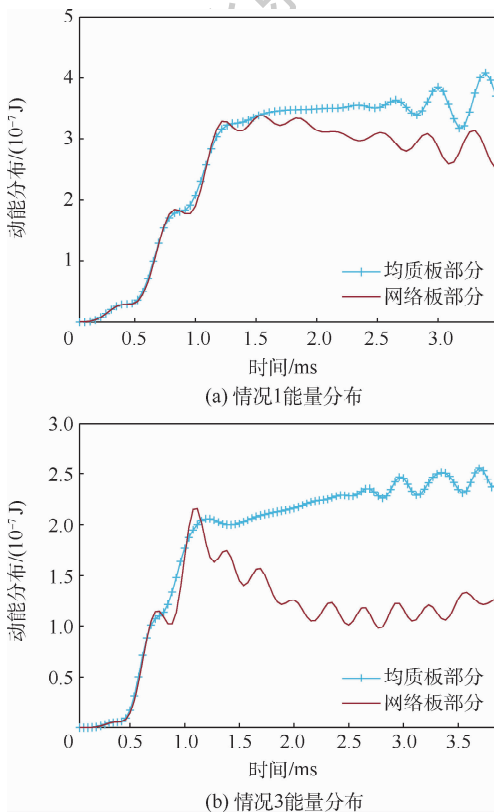


图 27 不同情况下强迫响应能量结果对比

Fig. 27 Comparison of forced response energy results in different situations

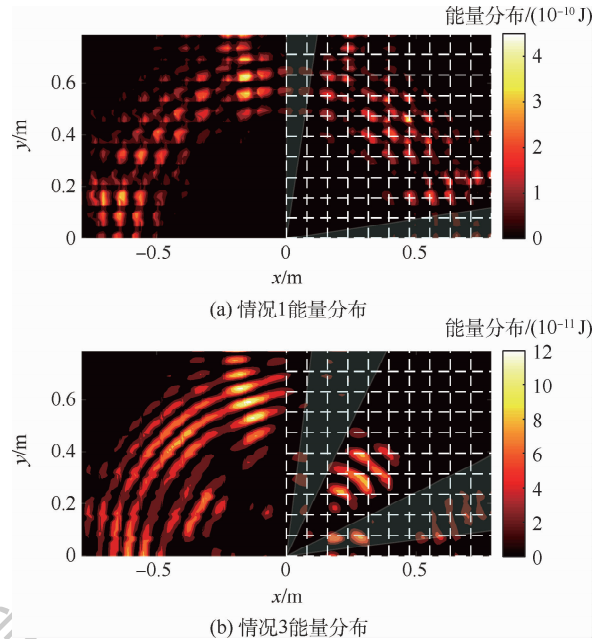


图 28 不同情况下能量云图对比

Fig. 28 Distribution of kinetic energy in different situations

4 结 论

在压电周期结构的设计中,需要尽可能提高禁带的性能,此时结构参数的设计至关重要。开发了适用于压电周期网络系统的减缩波有限元算法,在此基础上,结合单波形占比的概念研究了二维结构内耦合禁带的方向性分布。同时分析了不同压电片形状和尺寸对耦合禁带的影响,并提出了如何针对特定频率设计结构的电学参数,分析结果由强迫响应验证。得到如下结论:

1) 利用该减缩波有限元算法可以节约 90% 以上的计算时间。但为了保证计算精度,至少需要保留 5 倍计算目标频带内的所有模态阶数,仅选择特定变形的模态会使得高频的结果出现较大误差。

2) 圆形压电片和方形压电片随着尺寸增大都使得耦合禁带向低频移动。相同面积下,二者具有相似的禁带中心频率和方向性分布,方形压电片禁带带宽略宽于圆形压电片,因此在后续设计中,可以把压电片面积作为主要设计指标。

3) 提出了针对不同尺寸和形状的压电片如何在同一频率附近设计耦合禁带的准则。随着压电片面积的增大,所需的电学参数降低,耦合禁带宽度增强,然而压电覆盖率超过一定程度后,对禁带的宽度提升并不明显。因此在无法实现大电感等电学元件的情况下,可以通过设计压电片的尺寸,使得耦合禁带涵盖目标频带。

4) 对于带宽较弱的耦合禁带,利用强迫响应

分析可知,即使在不含任何阻尼的情况下,对有限结构内的振动依然有明显的控制效果。

参考文献 (References)

- [1] HUSSEIN M I, LEAMY M J, RUZZENE M. Dynamics of phononic materials and structures; Historical origins, recent progress, and future outlook [J]. *Applied Mechanics Reviews*, 2014, 66(4):1-38.
- [2] 张荣英,姜根山,王璋奇,等. 声子晶体的研究进展及应用前景[J]. *声学技术*, 2006, 25(1):35-42.
ZHANG R Y, JIANG G S, WANG Z Q, et al. Progress in researches of phononic crystal and the application perspectives [J]. *Technical Acoustics*, 2006, 25(1):35-42 (in Chinese).
- [3] ROCA D, YAGO D, CANTE J, et al. Computational design of locally resonant acoustic metamaterials [J]. *Computer Methods in Applied Mechanics and Engineering*, 2019, 345(1):161-182.
- [4] NING L, WANG Y Z, WANG Y S, et al. Active control of elastic metamaterials consisting of symmetric double Helmholtz resonator cavities [J]. *International Journal of Mechanical Sciences*, 2019, 153-154:287-298.
- [5] PELAT A, GALLOT T, GAUTIER F. On the control of the first Bragg band gap in periodic continuously corrugated beam for flexural vibration [J]. *Journal of Sound and Vibration*, 2019, 446:249-262.
- [6] FAN Y, COLLET M, ICHCHOU M, et al. Energy flow prediction in built-up structures through a hybrid finite element/wave and finite element approach [J]. *Mechanical Systems and Signal Processing*, 2016, 66-67(1):137-158.
- [7] QI S, OUDICH M, LI Y, et al. Acoustic energy harvesting based on a planar acoustic metamaterial [J]. *Applied Physics Letters*, 2016, 108(26):1-4.
- [8] WU Y G, LI L, FAN Y, et al. Design of dry friction and piezoelectric hybrid ring dampers for integrally bladed disks based on complex nonlinear modes [J]. *Computers and Structures*, 2020, 233:1-19.
- [9] WEN S R, XIONG Y H, HAO S M, et al. Enhanced band-gap properties of an acoustic metamaterial beam with periodically variable cross-sections [J]. *International Journal of Mechanical Sciences*, 2020, 166:1-10.
- [10] MOSCATELLI M, ARDITO R, DRIEMEIER L, et al. Band-gap structure in two- and three-dimensional cellular locally resonant materials [J]. *Journal of Sound and Vibration*, Academic Press, 2019, 454(18):73-84.
- [11] LI L, JIANG Z, FAN Y, et al. Creating the coupled band gaps in piezoelectric composite plates by interconnected electric impedance [J]. *Materials*, 2018, 11(9):1656.
- [12] 舒海生,张法,刘少刚,等. 一种特殊的布拉格型声子晶体杆振动带隙研究 [J]. *振动与冲击*, 2014, 33(19):147-151.
SHU H S, ZHANG F, LIU S G, et al. Vibration band gap of a special rod of phononic crystals [J]. *Journal of Vibration and Shock*, 2014, 33(19):147-151 (in Chinese).
- [13] 舒海生,董立强,李世丹,等. 布拉格型声子晶体弦的横向振动带隙研究 [J]. *人工晶体学报*, 2013, 42(9):1918-1923.
SHU H S, DONG L Q, LI S D, et al. Study on the lateral band gap of string of phononic crystals [J]. *Journal of Synthetic Crystals*, 2013, 42(9):1918-1923 (in Chinese).
- [14] 张若军,肖勇,温激鸿,等. 四边固支局域共振型板的低频隔声特性研究 [J]. *振动工程学报*, 2016, 29(5):905-912.
ZHANG R J, XIAO Y, WEN J H, et al. Analysis of sound transmission through clamped locally resonant plate in low frequency [J]. *Journal of Vibration Engineering*, 2016, 29(5):905-912 (in Chinese).
- [15] LIU L, HUSSEIN M I. Wave motion in periodic flexural beams and characterization of the transition between bragg scattering and local resonance [J]. *Journal of Applied Mechanics*, 2012, 79(1):1-17.
- [16] XIAO Y, WEN J, WANG G, et al. Theoretical and experimental study of locally resonant and bragg band gaps in flexural beams carrying periodic arrays of beam-like resonators [J]. *Journal of Vibration and Acoustics*, 2013, 135(4):1-17.
- [17] DONG Y, YAO H, DU J, et al. Research on local resonance and Bragg scattering coexistence in phononic crystal [J]. *Modern Physics Letters B*, 2017, 31(11):1-10.
- [18] CAI C, WANG Z, CHU Y, et al. The phononic band gaps of Bragg scattering and locally resonant pentamode metamaterials [J]. *Journal of Physics D: Applied Physics*, 2017, 50(41):1-7.
- [19] LIU Z, ZHANG X, MAO Y, et al. Locally resonant sonic materials [J]. *Science*, 2000, 289(5485):1734-1736.
- [20] 朱兴一,钟盛,叶安珂,等. 声子晶体禁带特性及局域共振现象的试验研究 [J]. *人工晶体学报*, 2014(11):2852-2859.
ZHU X Y, ZHONG S, YE A K, et al. Experimental study on band gap properties and local resonance phenomenon of phononic crystals [J]. *Journal of Synthetic Crystals*, 2014(11):2852-2859 (in Chinese).
- [21] LEE T, HZUKA H. Bragg scattering based acoustic topological transition controlled by local resonance [J]. *Physical Review B*, 2019, 99(6):1-11.
- [22] MANCONI E, MACE B. Veering and strong coupling effects in structural dynamics [J]. *Journal of Vibration and Acoustics*, 2017, 139(2):1-10.
- [23] MACE B R, MANCONI E. Wave motion and dispersion phenomena: Veering, locking and strong coupling effects [J]. *The Journal of the Acoustical Society of America*, 2012, 131(2):1015-1028.
- [24] THOMPSON D J, FERGUSON N S, YOO J W, et al. Structural waveguide behaviour of a beam-plate system [J]. *Journal of Sound and Vibration*, 2008, 318(1):206-226.
- [25] YU D, WEN J, ZHAO H, et al. Vibration reduction by using the idea of phononic crystals in a pipe-conveying fluid [J]. *Journal of Sound and Vibration*, 2008, 318(1):193-205.
- [26] LIU J, LI L, FAN Y. A comparison between the friction and piezoelectric synchronized switch dampers for blisks [J]. *Journal of Intelligent Material Systems and Structures*, 2018, 29(12):2693-2705.
- [27] LIU J, LI L, HUANG X, et al. Dynamic characteristics of the blisk with synchronized switch damping based on negative capacitor [J]. *Mechanical Systems and Signal Processing*, 2017,

- 95:425-445.
- [28] YE G, YAN J, WONG Z J, et al. Optimisation of a piezoelectric system for energy harvesting from traffic vibrations [C] // Proceedings of the IEEE Ultrasonics Symposium. Piscataway: IEEE Press, 2009: 759-762.
- [29] WANG W, LI L, FAN Y, et al. Piezoelectric transducers for structural health monitoring of joint structures in cylinders: A wave-based design approach [J]. *Sensors*, 2020, 20(3): 1-25.
- [30] LI L, JIANG Z, FAN Y, et al. Coupled band gaps in the piezoelectric composite plate with interconnected electric impedance [C] // ASME 2018 Conference on Smart Materials, Adaptive Structures and Intelligent Systems, 2018: 1-27.
- [31] MEITZLER A H, BERLINCOURT D, WELSH F S, et al. IEEE standard on piezoelectricity: ANSI/IEEE Std 176-1987 [S]. Piscataway: IEEE Press, 1988.
- [32] CHEN S, WANG G, SONG Y. Low-frequency vibration isolation in sandwich plates by piezoelectric shunting arrays [J]. *Smart Materials and Structures*, 2016, 25(12): 1-8.
- [33] WANG G, CHEN S. Large low-frequency vibration attenuation induced by arrays of piezoelectric patches shunted with amplifier-resonator feedback circuits [J]. *Smart Materials and Structures*, 2015, 25(1): 1-15.
- [34] DUHAMEL D, MACE B R, BRENNAN M J. Finite element analysis of the vibrations of waveguides and periodic structures [J]. *Journal of Sound and Vibration*, 2006, 294(1): 205-220.
- [35] FAN Y, COLLET M, ICHCHOU M, et al. A wave-based design of semi-active piezoelectric composites for broadband vibration control [J]. *Smart Materials and Structures*, 2016, 25(5): 1-15.
- [36] ZHOU C W, LAINÉ J P, ICHCHOU M N, et al. Numerical and experimental investigation on broadband wave propagation features in perforated plates [J]. *Mechanical Systems and Signal Processing*, 2016, 75: 556-575.
- [37] ZHOU C W, LAINÉ J P, ICHCHOU M N, et al. Multi-scale modelling for two-dimensional periodic structures using a combined mode/wave based approach [J]. *Computers and Structures*, 2015, 154: 145-162.
- [38] BAMPTON M C C, CRAIG J R R. Coupling of substructures for dynamic analyses [J]. *AIAA Journal*, 1968, 6(7): 1313-1319.
- [39] WILCOX C H. Theory of Bloch waves [J]. *Journal d'Analyse Mathématique*, 1978, 33(1): 146-167.
- [40] MING-HUI L U, ZHANG C, FENG L, et al. Negative birefracton of acoustic waves in a sonic crystal [J]. *Nature Materials*, 2007, 6(10): 744-748.
- [41] YANG W P, WU L Y, CHEN L W. Refractive and focusing behaviours of tunable sonic crystals with dielectric elastomer cylindrical actuators [J]. *Journal of Physics D Applied Physics*, 2008, 41(13): 135408.

Influence analysis of coupled band gap in piezoelectric periodic plate

JIANG Zhou¹, LI Lin^{1,2}, FAN Yu^{1,2,*}, WANG Wenjun¹

(1. School of Energy and Power Engineering, Beihang University, Beijing 100083, China;

2. Collaborative Innovation Center for Advanced Aero-Engine, Beihang University, Beijing 100083, China)

Abstract: In recent years, researches on the elastic band gaps in periodic structures to reduce vibration have attracted widespread attention. However, it is difficult to design a band gap with wide bandwidth and good tunability. Aiming at this problem, we designed a periodic structure with piezoelectric network. By bonding piezoelectric patches periodically into structure, a coupled band gap can be created between the elastic waves and electric waves thanks to the piezoelectric effect. This band gap can be tailored with the help of external circuits. In order to calculate the propagation characteristics of the structure efficiently, a reduced wave finite element method suitable for piezoelectric periodic structures was developed to improve the calculation efficiency. It was found that more than 90% of the calculation time can be saved with great accuracy. Using this method, the influence of the size and shape of the piezoelectric material on the performance of the coupled band gap was studied. The results show that when fixing the electrical parameters, as the size of the piezoelectric patches increases, the coupled band gap moves to lower frequency range and its bandwidth increases. Moreover, the bandwidth in the system with square patches is slightly wider than that with circular patches. However, these two shapes have little impact on the directional distribution of coupled band gap. Then, the guideline is proposed for designing electrical parameters to make sure that coupled band gaps are generated around the desired frequency for electromechanical systems with different sizes and shapes. Finally, in order to prove the vibration reduction effect of the coupled band gap, a finite periodic piezoelectric plate was employed. The results show that the coupled band gap can effectively control the elastic wave in structure.

Keywords: piezoelectric material; periodic structure; coupled band gap; wave finite element method; reduction algorithm; parametric analysis

Received: 2020-06-01; **Accepted:** 2020-09-25; **Published online:** 2020-09-30 11:12

URL: kns.cnki.net/kcms/detail/11.2625.V.20200929.1830.004.html

Foundation items: National Natural Science Foundation of China (51675022,11702011); Aeronautical Science Foundation of China (2019ZB051002)

* **Corresponding author.** E-mail: fanyu04@buaa.edu.cn

http://bhxb.buaa.edu.cn jbuua@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2020.0190

S弯进气道出口旋流对轴流压气机性能的影响



王加乐, 程邦勤*, 费晓文, 尉洋, 冯路宁

(空军工程大学 航空工程学院, 西安 710038)

摘 要: 为探究 S 弯进气道出口旋流对轴流压气机性能的影响, 优化设计了旋流畸变网以模拟旋流, 利用数值模拟的方法探究了单级轴流压气机在 S 弯进气道出口旋流作用下的气动响应, 获得均匀进气条件和旋流进气条件下的压气机特性线和流场分布。结果表明: 优化后的旋流畸变网总体旋流角误差降低了。S 弯进气道出口旋流对增压能力影响不大, 但会导致压气机效率下降, 稳定工作范围减小。在 100% 和 80% 换算转速, 压气机的压比最大降幅分别为 0.12% 和 0.28%, 在峰值效率点附近的效率最大降幅为 3.2% 和 14.4%。S 弯进气道出口旋流中的反向旋流区增大了转子叶片进气攻角, 导致气流叶背分离、叶片通道堵塞, 最终导致压气机失稳。

关键词: S 弯进气道; 旋流畸变; 压气机; 性能; 稳定性

中图分类号: V231.3

文献标志码: A

文章编号: 1001-5965(2021)07-1438-08

现代战机对隐身性和机动性的要求日益提高, 第五代战机提出的 4S^[1] 中就包含了隐身性 (Stealth) 和超机动性 (Super Maneuverability)。为达到相应的指标要求, 先进战机广泛采用了 S 弯进气道, 然而 S 弯进气道被运用到战机上的历程却并非一帆风顺, 20 世纪 70 年代, 欧洲的 Tornado 战机采用了 S 弯进气道, 在地面试验时, 进气道/发动机匹配正常, 然而试飞过程中, 在特定的飞行工况下, 发动机发生了严重的喘振现象^[2], 事后调查组得出结论, 事故是由 S 弯进气道所产生的旋流引起。类似的, 在战斧巡航导弹的研制过程中, 也发生了因旋流而引发的事故^[3]。这些事故, 使得研究人员逐步重视起对旋流畸变的研究。

在早期, 研究人员主要结合具体的装备和事故开展相应的研究。后来发现, 在现实中出现最多, 而且对发动机稳定性影响最严重的是整体涡旋流, 但人们对于对涡旋流也十分关注^[4], 因此,

便针对整体涡和对涡旋流开展了大量的研究。Sheoran 等^[5-6] 探究了整体涡、对涡等旋流畸变对压气机性能和稳定性的影响。Govardhan 和 Viswanath^[7] 探究了轴流风扇在不同旋流进气条件下的响应。Naseri 等^[8-9] 开展了进气旋流畸变的模拟, 并开展了旋流对压气机性能的影响。在整体涡和对涡的研究基础之上, 开始针对 S 弯进气道产生的旋流问题展开研究, 美国 CLENN 中心^[10] 以及国内徐诸霖等^[11-12] 基于体积力模型开展了 S 弯进气道与压气机的耦合计算研究。Pardo 等^[13] 探究了 Rotor67 在进气旋流畸变作用下的气动响应; 王加乐等^[14] 探究了 S 弯进气道出口旋流对跨声速压气机 Stage67 性能的影响; 刘雷等^[15] 也探究了 S 弯进气道对风扇性能的影响。

本文基于课题组此前设计的旋流畸变网^[16] (以下简称旋流网) 进行优化设计, 获得了模拟精度更高的旋流网, 并结合单级低速轴流压气机开展联合数值仿真研究, 旨在探究 S 弯进气道出口

收稿日期: 2020-05-15; 录用日期: 2020-07-10; 网络出版时间: 2020-07-27 14:27

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20200727.1116.001.html

* 通信作者. E-mail: 1695438620@qq.com

引用格式: 王加乐, 程邦勤, 费晓文, 等. S 弯进气道出口旋流对轴流压气机性能的影响[J]. 北京航空航天大学学报, 2021, 47(7): 1438-1445. WANG J L, CHENG B Q, FEI X W, et al. Effects of swirl at outlet of S-shaped inlet on axial flow compressor performance [J]. Journal of Beijing University of Aeronautics and Astronautics, 2021, 47(7): 1438-1445 (in Chinese).

旋流(以下简称 S 弯旋流)对压气机性能和稳定性的影响,为抑制旋流带来的不利影响,改善压气机性能奠定基础。

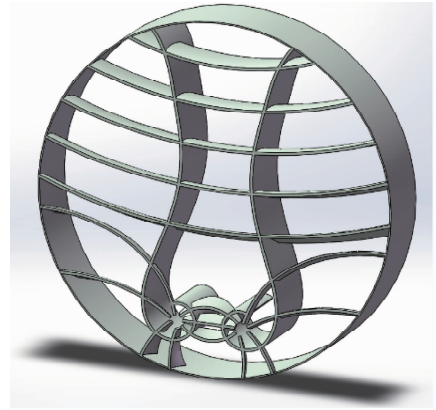
1 研究对象与计算方法

本文研究对象为旋流网与单级低速轴流压气机,图 1 中展示了速度矢量场及旋流角云图,是美国国家航空航天局(NASA)设计的某型 S 弯进气道出口流场^[16],该型进气道进口为 D 形,出口为圆形。上部旋流角较小,在底部位置存在一对强度较大的涡,通过设计旋流畸变网来模拟此目标流场。压气机为等离子体重点实验室的单级低速轴流式压气机。

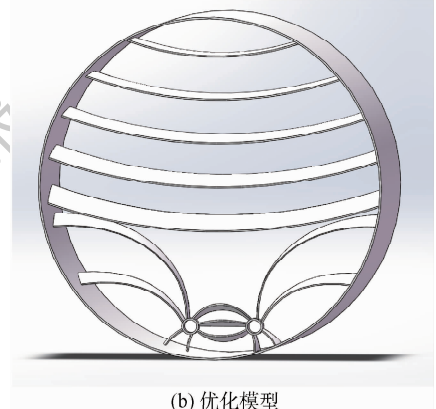
1.1 几何模型

图 2(a)^[11]为课题组此前所设计的旋流网,在此基础上,进行了进一步的优化设计,微调了叶片布局位置和叶片偏转角度,最终获得如图 2(b)所示的模拟精度更高的旋流网。

图 3 为低速轴流压气机实验台结构示意图^[17],压气机采用等径设计,轮毂直径 400 mm,机匣内壁直径 600 mm,其几何参数和设计参数如表 1 和表 2 所示。



(a) 未优化模型^[11]



(b) 优化模型

图 2 旋流网

Fig. 2 Swirl network

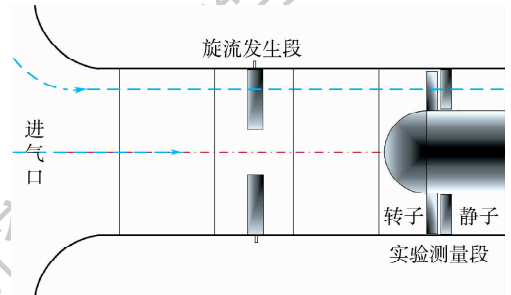


图 3 压气机实验系统^[17]

Fig. 3 Compressor experiment system^[17]

表 1 低速轴流压气机几何参数

Table 1 Geometric parameters of low-speed axial flow compressor

参数	转子	静子
叶片数目	45	60
弦长/mm	53.6	40
稠度	1.28	1.27
展弦比	1.87	2.5
叶顶间隙/mm	1.2	1.2

表 2 低速轴流压气机设计点性能参数

Table 2 Designed point performance parameters of low-speed axial flow compressor

参数	数值
流量/($\text{kg} \cdot \text{s}^{-1}$)	6.5
转速/($\text{r} \cdot \text{min}^{-1}$)	3 000
压比	1.025

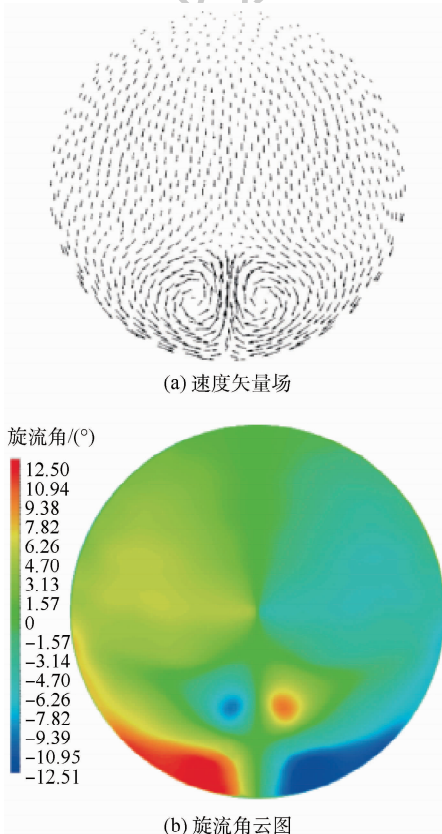


图 1 目标旋流场^[16]

Fig. 1 Target swirl flow field^[16]

1.2 求解方法

如图4所示,本文计算域分为2部分:一方面是因为计算域较大,网格数量巨大,在同一个计算域计算对计算机硬件要求较高;另一方面,不同的仿真软件有各自的专长,旋流网结构复杂且不规则,因此选用了ANSYS CFX进行求解,在叶轮机械的求解问题上,NUMECA IGG则具有更大的优势,因此选用NUMECA IGG对压气机进行求解。计算时,通过流量平衡连接两计算域。

前置管道、旋流网、后置管道及整流罩部分采用ANSYS ICEM划分网格,因旋流网结构复杂,故网格类型为四面体非结构网格,采用ANSYS CFX进行计算求解;压气机部分采用NUMECA IGG划分网格,网格类型为全六面体结构网格,叶片通道采用O4H网格拓扑,其中H型网格用于模拟进出口等部位,O型拓扑结构模拟叶片边界,在叶顶间隙部位采用了“蝶形”网格,对近壁面网格进行加密,第1层网格厚度为0.01 mm,确保 Y^+ 值小于8。图5为50%叶高截面的网格,计算求解采用NUMECA Fine turbo。

ANSYS CFX求解计算采用Navier-Stokes控制方程,SST湍流模型,残差目标设定为 1×10^{-5} ,进口设定总压101 325 Pa,总温288.15 K,出口设定质量流量,通过监控残差曲线及进出口流量是否平衡来判断收敛情况。

NUMECA采用Navier-Stokes方程开展计算,计算求解使用S-A模型,叶片、轮毂、畸变网和机匣壁面均设置为绝热无滑移壁面边界条件,其中与转子连接的轮毂设为转动无滑移壁面,其余轮

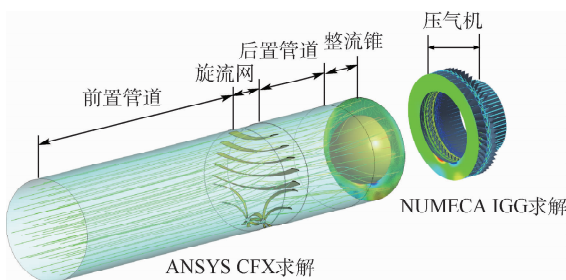


图4 计算域

Fig. 4 Computational domain

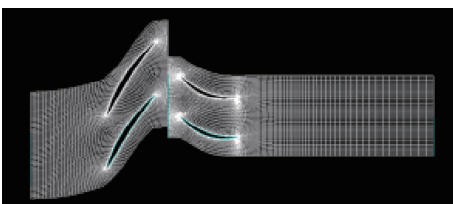


图5 50%叶高处网格

Fig. 5 Mesh at 50% span of compressor

毂和机匣设为静止无滑移壁面。边界条件设置如下:进口通过Profile文件给定畸变流场,出口设定为质量流量。监控均方根残差、进出口流量平衡、压比和效率波动情况来判断计算结果是否收敛。

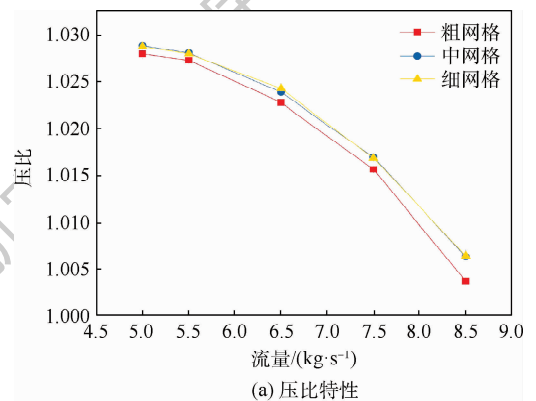
1.3 网格无依赖性验证

研究过程中,共划分了3套不同密度的网格,分别为粗网格、中网格和细网格,单通道网格数量如表3所示,3套网格计算的设计转速下压气机压比和效率特性曲线如图6所示。可知,网格数量达到 60.3×10^4 时,计算结果精确度满足要求,继续增加网格数量对计算结果的影响不大,故综合考虑求解精度和计算速度选取满足计算准确性的中网格进行计算求解。压气机全通道网格节点总数为 $3\,236.9 \times 10^4$,全通道网格如图7所示。

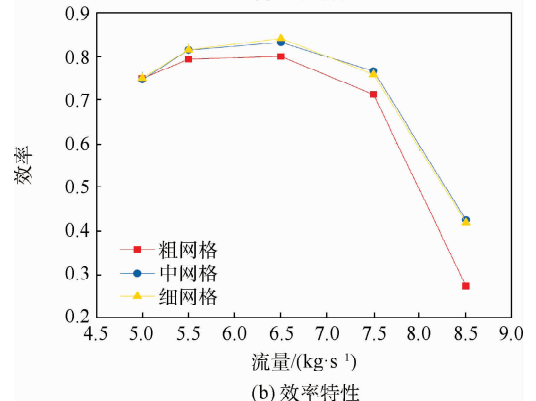
表3 不同网格类型的网格数量

Table 3 Number of grids for different grid types

网格类型	单通道网格数量/ 10^4
粗网格	33.9
中网格	60.3
细网格	91.8



(a) 压比特性



(b) 效率特性

图6 不同网格计算的压气机特性对比

Fig. 6 Comparison of compressor characteristics calculated by different grids



图7 全通道网格
Fig.7 Whole-passage mesh

2 结果分析

2.1 旋流场流场特性

如图8所示,气动交界面(AIP)选取在旋流网出口截面向下游400 mm处,随着气流向下发展,受整流罩的影响,涡核位置发生变化,最终在压气机进口截面的流场如图9所示,可见2个旋流严重的区域将作用在压气机叶根处。

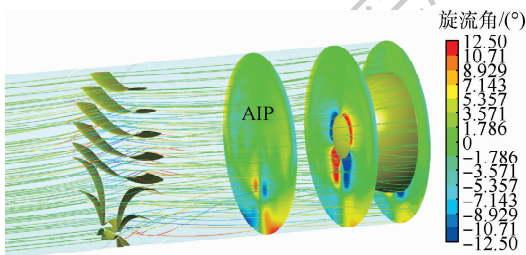


图8 旋流网下游各截面旋流角云图

Fig.8 Swirl angle contour on each section downstream of swirl network

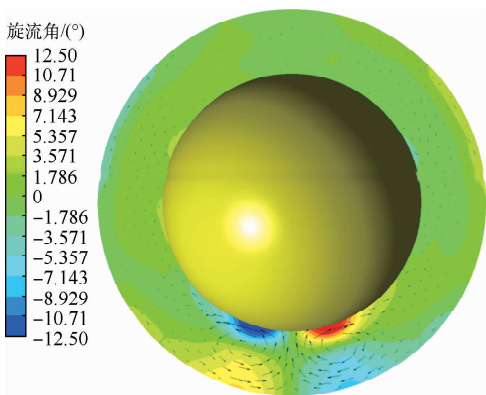


图9 压气机进口截面速度矢量场与旋流角云图

Fig.9 Velocity vector field and swirl angle contour of compressor inlet section

旋流角是评定旋流场最基本的指标^[18],其定义为

$$\alpha = \arctan\left(\frac{U_{\theta}}{U_x}\right) \quad (1)$$

式中: U_{θ} 为周向速度分量; U_x 为轴向速度分量。

在AIP设置5个测环,目标流场旋流网模拟的流场在5个测环上(见图10(a))的旋流角如图10(b)~(f)所示。在5个测环上的平均误差分别为 1.13° 、 0.97° 、 1.09° 、 0.82° 和 1.33° ,总体平均误差为 1.068° ,此前课题组设计的旋流网模拟误差为 1.56° ,模拟精度有了进一步提高。

2.2 压气机总体性能变化

图11为均匀进气和旋流进气条件下压气机性能对比图,可见在S弯旋流的作用下,压气机增压能力降低幅度较小,但喘振流量增加,稳定工作范围减小,效率降低,在100%和80%换算转速,压气机的压比最大降幅分别为0.12%和0.28%、在峰值效率点附近的效率最大降幅为3.2%和14.4%,压比特性线和效率特性线均向下方移动。

2.3 叶片流场变化

本节选取设计转速、设计流量点在进气旋流畸变的工况点进行分析,旨在探究S弯旋流对不同叶高位置的影响。

为便于展示进气旋流畸变对叶片内气流的影响,如图12所示,进口截面展示了旋流角云图,在5%、50%以及99%叶高位置展示了相对马赫数云图。由图12(a)可知,当叶根位置的反向旋流区作用于压气机时,气流在5%叶高截面出现了明显的低速区(椭圆区域),后经过同向旋流区后,相对马赫数随之增加。在50%叶高位置,进口截面无强度较大的旋流区,故该截面的流动相对稳定,未受较大扰动。在99%叶高处,与5%叶高截面类似,当反向旋流作用于压气机时,在2段虚线中间形成了低速流动区域,可见,对压气机性能和稳定性的主要影响区域,是叶根位置和叶顶位置的2对旋流区域。

2.4 压气机失速机理分析

S弯旋流与标准对涡旋流相似,根据此前研究发现,对涡旋流可看作同向整体涡与反向整体涡旋流的叠加,S弯旋流也可看成对涡结构,但其涡核位置位于截面底部(见图9),因此,相较于标准对涡旋流,S弯旋流影响的周向范围相对较小,但旋流变化梯度较大。

在进气流量为 4.8 kg/s 时,均匀进气时压气机未失速,而在S弯旋流的作用下,压气机已进入失速状态,因此,本节选取流量为 4.8 kg/s 的S弯旋流进气工况进行分析,探究压气机失速的原因。

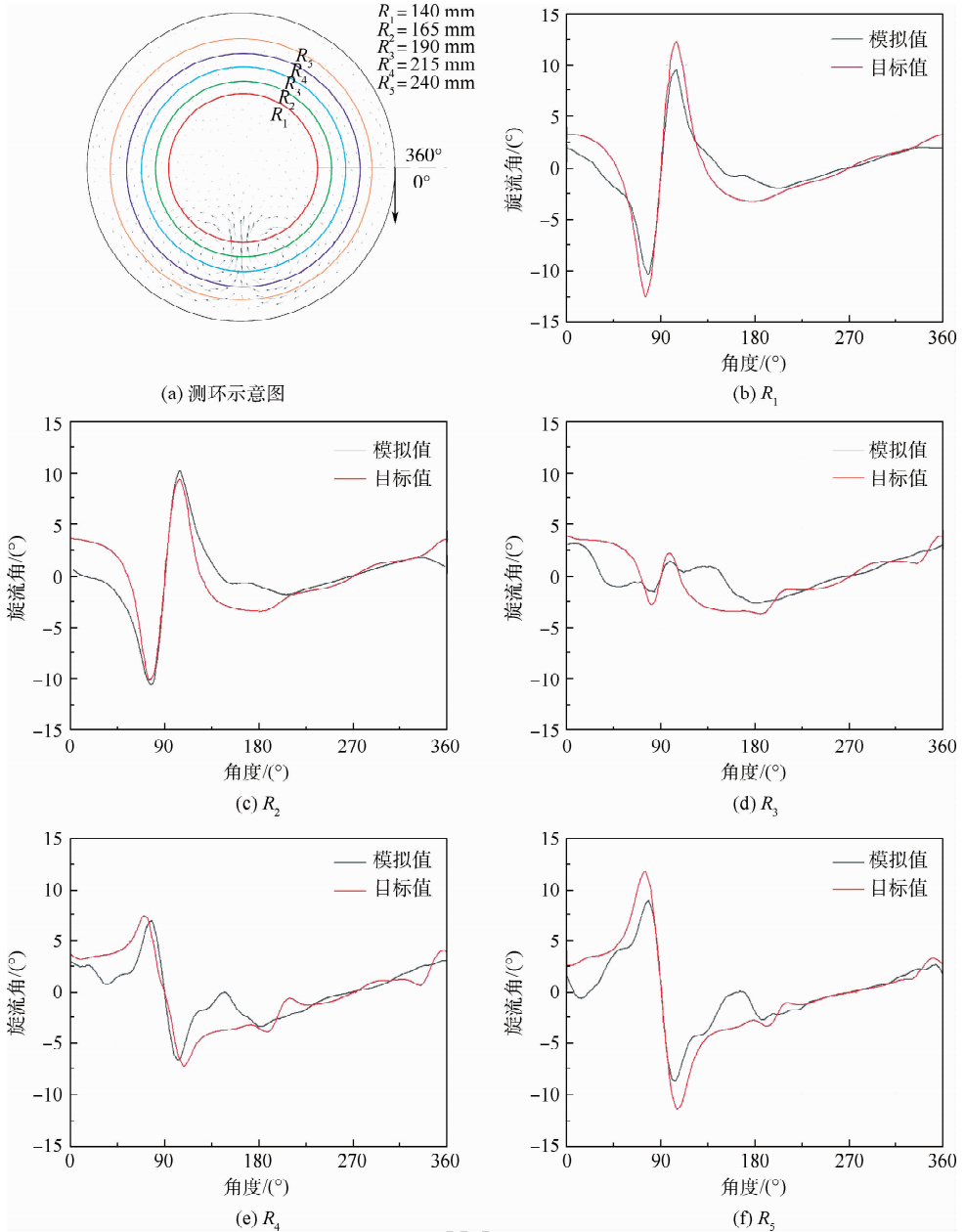


图 10 测环示意图及测环上旋流角分布

Fig. 10 Schematic of measuring ring and distribution of swirl angle on measuring ring

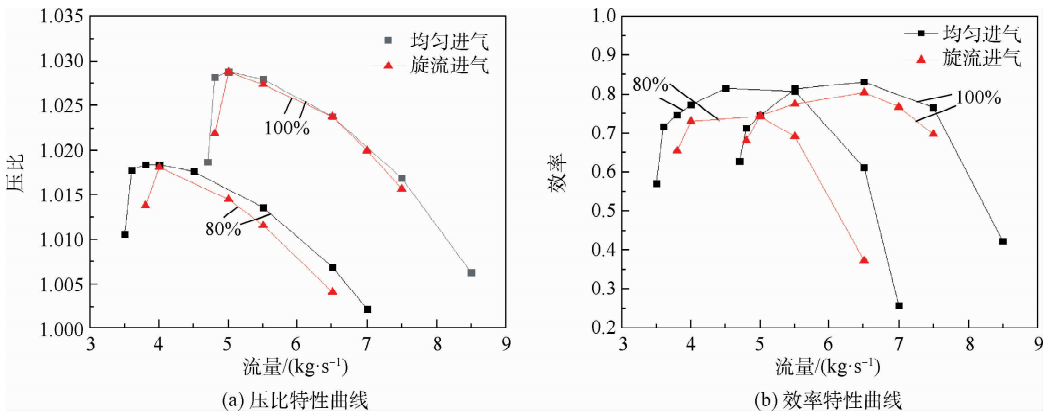


图 11 S弯旋流对压气机性能的影响

Fig. 11 Effect of S-shaped swirl on compressor performance

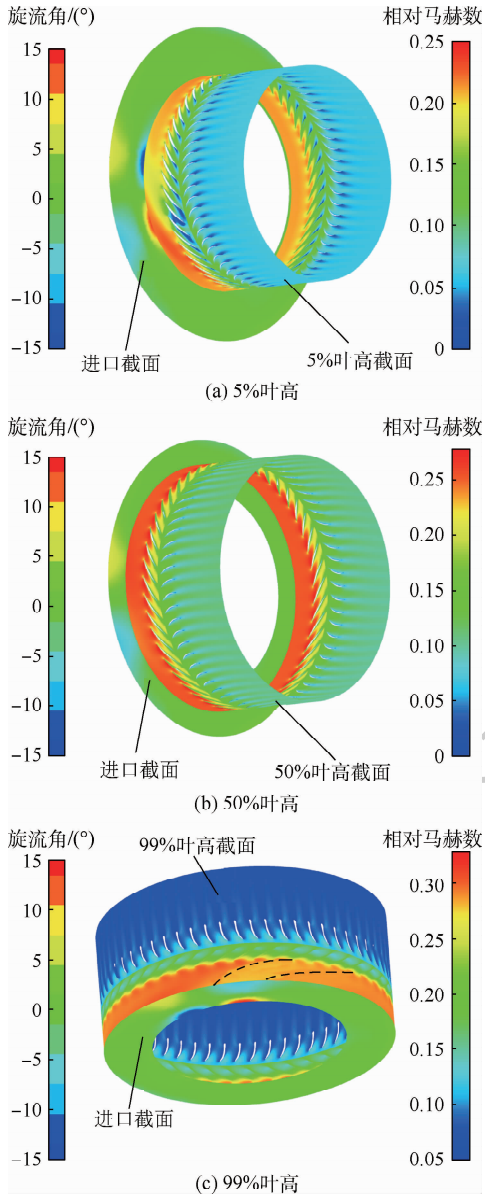


图 12 S弯进气道各叶高处相对马赫数云图
Fig. 12 Relative Mach number contour on each section with S-shaped inlet

如图 13 所示,观察到在叶尖区域存在一个反向旋流区,而且随着气流向下游流动,反向旋流区的位置有所变化,且因与压气机的耦合作用,旋流强度增加,反向旋流区范围扩大,到转子前缘时,反向旋流使得进气攻角增大,进而造成气流在叶背分离,造成叶片通道内气流堵塞,经过几个流道的发展,最终导致气流在叶片前缘溢出,溢出的气流沿周向迅速发展,最终导致压气机失稳。

经 2.3 节分析,对压气机稳定性造成影响还有叶根位置的一对旋流,如图 14 所示,在 5% 叶高截面,当负向旋流区作用于压气机时,在红色椭圆区域形成了明显的低速区域,转子通道叶根区域堵塞逐渐严重,上一通道堵塞的气流将转子前缘的气流向上游方向“推出”,使其无法顺利流

入下一通道中,如此形成恶性循环,这也是压气机失稳的一个原因。由此可知,反向旋流对压气机的稳定工作威胁性最大。

由此前的研究可知^[19-21],若进气旋流全部为同向整体涡时,会导致转子进气攻角减小,使得叶片负荷变小,进而降低叶片的增压能力,但增强了气流在叶片通道的流通能力,因此会增强压气机的稳定性;反之若进气旋流全为反向整体涡时,会使得转子进气攻角变大,增大叶片负荷,压气机增压能力提高,但因通道内气流堵塞原因,稳定工作裕度会变小,压气机更易失稳。

在本文研究的 S 弯旋流中发现,S 弯旋流使得压比有较小幅度的下降,也使得压气机提前失速,稳定工作裕度降低。这是因为压气机失稳是由最不利于其稳定工作的进气区域决定。

结合此前的研究分析可知,只有在反向整体涡或反向整体涡占主导地位时,进气旋流才会使得压气机增压能力增加,对于标准对涡旋流或 S 弯旋流而言,同向旋流和反向旋流的强度和范围大致相等,因此对压气机增压能力影响不大,但与均匀进气相比,仍有较小的负面影响。而受 S 弯

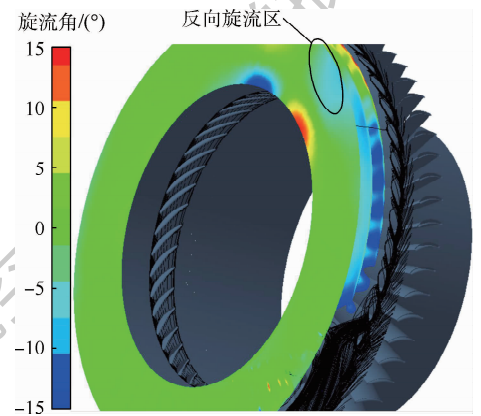


图 13 旋流发展趋势及叶顶区域流线
Fig. 13 Development trend of swirl and streamlines of blade tip area

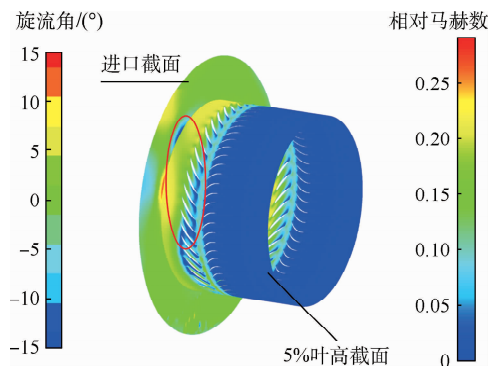


图 14 失速点 5% 叶高截面相对马赫数云图
Fig. 14 Relative Mach number contour on span 5% near stall point

旋流场中反向旋流的作用,压气机的稳定工作性能恶化,因此对涡类旋流对压气机只会产生不利影响。

3 结论

本文基于此前课题组设计的旋流畸变网进行了优化,并探究了S弯旋流对单级低速轴流式压气机性能与稳定性的影响,得到的结论如下:

1) S弯旋流对压气机压比影响较小,但对效率产生较大的消极影响。在100%和80%换算转速,压气机的压比最大降幅分别为0.12%和0.28%,在峰值效率点附近的效率最大降幅为3.2%和14.4%,压比特性线和效率特性线均向下方移动。

2) S弯旋流降低了压气机工作的稳定性,S弯旋流的2对高强度旋流区域共同导致了压气机失稳,其中反向旋流区域是造成压气机失稳的根本原因,反向旋流区会导致叶片攻角增大,使得气流在叶背分离,导致叶片通道内气流堵塞溢出,最终导致压气机失稳。

3) S弯旋流对压气机压比影响较小,这是因为S弯旋流场呈对称分布,在同向旋流区会降低叶片的增压能力,在反向旋流区会增加叶片的增压能力,同向旋流区域反向旋流区域的强度和面积大致相等,因此,对全通道压气机叶片的增压能力影响较小。

参考文献 (References)

- [1] 王如根,高坤华.航空发动机新技术[M].北京:航空工业出版社,2003.
WANG R G,GAO K H. New technology of aeronautical engine [M]. Beijing: Aviation Industry Press,2003 (in Chinese).
- [2] STOCKS C P,BISSINGER N C. The design and development of the tornado engine air intake: CP-301-1981 [R]. [S. l.]: AGARD Aerodynamics of Power Plant Installation,1981.
- [3] LUDWIG G. Tomahawk engine/inlet compatibility study for f107-wr-400/402 engines[C]//Williams International Report, 1989.
- [4] 姜健,于芳芳,赵海刚,等.进气道/发动机相容性评价体系的完善与发展[J].科学技术与工程,2009,9(21):6474-6483.
JIANG J,YU F F,ZHAO H G,et al. Perfection and development of engine/intake compatibility evaluate criterion[J]. Science Technology and Engineering, 2009, 9 (21): 6474-6483 (in Chinese).
- [5] SHEORAN Y, BRUCE B P, KRISHNAN M. Compressor performance and operability in swirl distortion[J]. Journal of Turbomachinery,2012,134(4):1-13.
- [6] SHEORAN Y, BRUCE B P, KRISHNAN M. Advancements in the design of an adaptable swirl distortion generator for testing gas turbine engines[C]//ASME Turbo Expo 2009: Power for Land, Sea, and Air. Orlando: ASME,2010:23-32.
- [7] GOVARDHAN M,VISWANATH K. Investigations on an Axial flow fan stage subjected to circumferential inlet flow distortion and swirl[J]. Journal of Thermal Science,1997,6(4):241-250.
- [8] NASERI A M, BORROOMAND M, TOUSI A M. The effect of inlet flow distortion on performance of a MicroJet engine: Part I—Development of an inlet simulator [C]//Proceedings of the ASME 2012 International Mechanical Engineering Congress and Exposition. Orlando: ASME,2012:317-324.
- [9] NASERI A M, BORROOMAND A M, TOUSI A M, et al. The effect of inlet flow distortion on performance of a micro-jet engine: Part 2—Engine tests [C]//Proceedings of the ASME 2012 International Mechanical Engineering Congress and Exposition. Orlando: ASME,2012:311-316.
- [10] AREND D J, CASTNER R S, FRATE F C. Low cost, compact and versatile rig for integrated inlet and propulsion systems research [C]//44th AIAA Aerospace Science Meeting and Exhibit. Reston: AIAA,2006:1313.
- [11] 徐诸霖,达兴亚,吴军强.基于体积力模型的大S弯进气道与风扇耦合计算研究[J].推进技术,2019,40(7):1441-1448.
XU Z L, DA X Y, WU J Q. Computational study on S-shaped inlet and fan coupling based on body force model [J]. Journal of Propulsion Technology,2019,40(7):1441-1448 (in Chinese).
- [12] 徐诸霖,达兴亚,范兆林.基于五孔探针的大S弯进气道旋流畸变评估[J].航空学报,2017,38(12):48-57.
XU Z L, DA X Y, FAN Z L. Assessment of swirl distortion of serpentine inlet based on five-hole probe [J]. Acta Aeronautica et Astronautica Sinica,2017,38(12):48-57 (in Chinese).
- [13] PARDO A C, MEHDI A, PACHIDIS V, et al. Numerical study of the effect of multiple tightly-wound vortices on a transonic fan stage performance [C]//ASME Turbo Expo 2014: Turbine Technical Conference and Exposition. Orlando: ASME,2014.
- [14] 王加乐,程邦勤,张磊,等.特定涡旋流畸变对跨声速压气机性能的影响[J].航空动力学报,2020,35(3):540-551.
WANG J L, CHENG B Q, ZHANG L, et al. Effects of specific swirl distortion on performance of transonic compressor [J]. Journal of Aerospace Power,2020,35(3):540-551 (in Chinese).
- [15] 刘雷,宋彦萍,陈焕龙,等.S弯进气道优化对其内流场及性能影响研究[J].工程热物理学报,2015,138(2):50-54.
LIU L, SONG Y P, CHEN H L, et al. Investigation on effect of S-shaped inlet optimization to internal flow characteristic and aerodynamic performance [J]. Journal of Engineering Thermophysics,2015,138(2):50-54 (in Chinese).
- [16] 张磊,程邦勤,王加乐,等.新型旋流畸变网的设计与仿真研究[J].推进技术,2018,39(9):2110-2120.
ZHANG L, CHENG B Q, WANG J L, et al. Design and numerical simulation of a new swirl distortion screen [J]. Journal of Propulsion Technology,2018,39(9):2110-2120 (in Chinese).
- [17] 宋国兴,李军,周游天,等.轴流压气机进气旋流畸变实验与仿真研究[J].风机技术,2017,59(5):1-12.
SONG G X, LI J, ZHOU Y T, et al. The experiment and simulation of inlet swirl distortion for axial compressor [J]. Chinese Journal of Turbomachinery,2017,59(5):1-12 (in Chinese).

- [18] Society of Automotive Engineers. A method for assessing inlet swirl distortion; AIR 5686[R]. Warrendale: SAE, 2010.
- [19] 胡伟波,程邦勤,陈志敏,等. 整体涡旋流畸变对压气机性能影响的研究[J]. 推进技术, 2015, 36(9): 1324-1330.
HU W B, CHENG B Q, CHEN Z M, et al. Investigation on effects of bulk swirl distortion on compressor performance[J]. Journal of Propulsion Technology, 2015, 36(9): 1324-1330 (in Chinese).
- [20] 程邦勤,纪振伟,胡伟波,等. 对涡旋流畸变对单级跨声速压气机性能影响的数值研究[J]. 推进技术, 2017, 38(8): 1776-1786.
CHENG B Q, JI Z W, HU W B, et al. Effects of paired swirl distortion on single stage transonic compressor performance using numerical simulation[J]. Journal of Propulsion Technology, 2017, 38(8): 1776-1786 (in Chinese).
- [21] 胡伟波,程邦勤,陈志敏,等. 应用改进的平行压气机模型预测旋流畸变对压气机性能的影响[J]. 推进技术, 2016, 37(8): 1485-1489.
HU W B, CHENG B Q, CHEN Z M, et al. Prognostication for effects of swirl distortion on compressor performance with modified parallel compressor model[J]. Journal of Propulsion Technology, 2016, 37(8): 1485-1489 (in Chinese).

Effects of swirl at outlet of S-shaped inlet on axial flow compressor performance

WANG Jiale, CHENG Bangqin*, FEI Xiaowen, YU Yang, FENG Luning

(Aeronautics Engineering College, Air Force Engineering University, Xi'an 710038, China)

Abstract: In order to estimate the effects of swirl at outlet of S-shaped inlet on the performance of axial flow compressor, a swirl distortion network was optimized and designed to simulate the swirl flow field. The aerodynamic response of the single-stage axial flow compressor under the effect of the swirl at the outlet of the S-shaped inlet was investigated by numerical simulation. The characteristic map and flow field distributions of the compressor under the condition of uniform intake and swirling intake were obtained. The results show that the overall swirl angle error of the optimized swirl distortion network is reduced. The swirl flow at the outlet of S-shaped inlet has little influence on the pressurization ability, but it will lead to the decrease of compressor efficiency and decrease of stability. At the corrected speed of 100% and 80%, the maximum reduction of pressure ratio is 0.12% and 0.28% respectively, and the maximum reduction of efficiency near the peak efficiency point is 3.2% and 14.4% respectively. The reverse swirl zone in the swirl at the outlet of the S-shaped inlet increases the inlet attack angle of the rotor blade, resulting in the separation of airflow at the suction side and the blockage of the blade passage, which eventually leads to the instability of the compressor.

Keywords: S-shaped inlet; swirl distortion; compressor; performance; stability

http://bhxb.buaa.edu.cn jbuua@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2020.0201

脉冲星方位误差估计的两步卡尔曼滤波算法



肖永强¹, 王宏力^{1,*}, 冯磊¹, 由四海¹, 许强²

(1. 火箭军工程大学 制导教研室, 西安 710025; 2. 青州高新技术研究所 测试控制系, 青州 262500)

摘 要: 为了克服钟差和卫星位置误差对脉冲星方位误差估计的影响, 设计了两步卡尔曼滤波(TSKF)算法。首先, 介绍了脉冲星方位误差估计的传统模型, 并通过分析和仿真验证了钟差、卫星位置误差以及2种误差同时存在时会使脉冲星方位误差估计结果产生较大偏差。其次, 在传统的估计模型中加入了钟差和卫星位置误差, 并将钟差和钟差变化率增广为新的状态量, 从而推导出包含2种误差的新模型, 并证明了该模型的完全可观测性; 根据该模型并按照两步卡尔曼滤波原理, 得到了TSKF算法的步骤。最后, 通过仿真表明: 在钟差和卫星位置误差同时影响下, 传统脉冲星方位误差估计算法偏差较大且发散; TSKF算法则能够有效隔离2种误差的影响, 使赤经和赤纬误差估计达到0.2 mas之内的精度。

关键词: 方位误差; 时钟钟差; 卫星位置误差; 增广状态; 两步卡尔曼滤波(TSKF)

中图分类号: V324.2⁺1; P129

文献标志码: A

文章编号: 1001-5965(2021)07-1446-07

脉冲星导航具有较强的自主性和较高的可靠性, 在民用和军事领域有着巨大的发展前景^[1-2]。然而在导航过程中, 0.001"的方位误差就会引起几百米的导航误差, 从而严重影响导航精度^[3]。当前主要依靠甚长基线干涉测量技术对脉冲星进行观测, 然而由于脉冲星信号微弱以及测量设备、精度等因素的影响, 该技术不能达到脉冲星导航精度的需求。为此, 国内学者采用了基于信标卫星的估计^[4]、鲁棒滤波估计^[5-6]、组合导航^[7]等方法来提高其精度。许强等在利用信标卫星估计方位误差时, 对卫星位置误差进行了修正^[8], 但是该方法假设卫星位置误差在3个坐标轴的投影相同且都为正, 但是卫星实际运行复杂, 其位置误差在不同方向的投影不尽相同, 因此该方法应用范围较低。而且上述方法都没有考虑导航过程中的时钟钟差问题。虽然孙守明、王璐等修正了脉冲星导航与惯性、多普勒导航等组合导航中钟差的

影响^[9-10], 但是目前脉冲星方位误差估计过程中的钟差问题研究较少。

为此, 本文以利用信标卫星进行脉冲星方位误差估计为背景, 设计了两步卡尔曼滤波(Two-step Kalman filter, TSKF)算法解决了同时存在钟差和卫星位置误差影响的问题, 从而提高了脉冲星方位误差估计的精度。

1 脉冲星方位误差估计的传统模型

利用卫星估计脉冲星方位误差是将卫星上X射线探测器测量的脉冲到达时间转换为到达太阳系质心坐标系(SSB)原点的时间 t'_{SSB} , 并用该时间减去脉冲星相位时间模型计算的真实到达SSB原点的时间 t_{SSB} , 而脉冲星方位误差就反映在两者之差中^[3], 如图1所示。

转换过程如下^[11]:

收稿日期: 2020-05-22; 录用日期: 2020-08-14; 网络出版时间: 2020-09-08 10:12

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20200905.1337.001.html

基金项目: 国家自然科学基金(61503391); 中国博士后科学基金(2017M613327)

* 通信作者. E-mail: wanghongli19650526@163.com

引用格式: 肖永强, 王宏力, 冯磊, 等. 脉冲星方位误差估计的两步卡尔曼滤波算法[J]. 北京航空航天大学学报, 2021, 47(7): 1446-1452. XIAO Y Q, WANG H L, FENG L, et al. Two-step Kalman filter algorithm for pulsar position error estimation[J]. Journal of Beijing University of Aeronautics and Astronautics, 2021, 47(7): 1446-1452 (in Chinese).

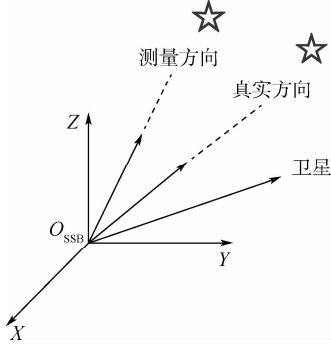


图1 脉冲星方位误差估计原理

Fig.1 Principle of pulsar position error estimation

$$t'_{SSB} = t'_{sat} + \frac{\mathbf{n}' \cdot \mathbf{r}_{sat}}{c} + o(t) \quad (1)$$

式中: t'_{sat} 为信标卫星时钟测得的脉冲到达时间; \mathbf{n}' 为测量得到的带误差的单位方向矢量; \mathbf{r}_{sat} 为卫星在SSB中的位置矢量; c 为光速; $o(t)$ 为周年视差效应等引起的高阶项。由相位时间模型得到脉冲到达SSB的时间 t_{SSB} 为

$$t_{SSB} = t_{sat} + \frac{\mathbf{n} \cdot \mathbf{r}_{sat}}{c} + o(t) \quad (2)$$

式中: t_{sat} 为脉冲到达卫星的真实时间; \mathbf{n} 为脉冲星真实的单位方向矢量。设脉冲星的赤经为 α , 赤纬为 β , 则满足:

$$\mathbf{n} = \begin{bmatrix} \cos \beta \cos \alpha \\ \cos \beta \sin \alpha \\ \sin \beta \end{bmatrix} \quad (3)$$

设 $(\Delta\alpha, \Delta\beta)$ 为脉冲星方位误差, 则脉冲星方位信息的真实值与带误差的值之间满足:

$$\begin{cases} \alpha = \alpha' + \Delta\alpha \\ \beta = \beta' + \Delta\beta \end{cases} \quad (4)$$

将式(4)代入式(3), 忽略二阶小项得

$$\mathbf{H}_k = \begin{bmatrix} -\cos \beta' \sin \alpha' \cdot r_{sat/x} + \cos \beta' \cos \alpha' \cdot r_{sat/y} \\ \cos \beta' \cdot r_{sat/z} - \sin \beta' \cos \alpha' \cdot r_{sat/x} - \sin \beta' \sin \alpha' \cdot r_{sat/y} \end{bmatrix}^T$$

$r_{sat/x}, r_{sat/y}, r_{sat/z}$ 为在3个坐标轴方向的投影。

2 误差影响分析

2.1 钟差的影响

卫星上的时钟频率和相位会发生漂移, 从而会造成脉冲到达卫星的真实时间和卫星时钟测得的脉冲到达时间存在偏差, 这种偏差就是时钟钟差, 设其为 δt , 则满足:

$$t_{sat} - \delta t = t'_{sat} \quad (12)$$

$$c(t_{SSB} - t'_{SSB}) = \Delta\mathbf{n} \cdot \mathbf{r}_{sat} + c \cdot \delta t \quad (13)$$

$$\mathbf{n} = \begin{bmatrix} \cos \beta' \cos \alpha' \\ \cos \beta' \sin \alpha' \\ \sin \beta' \end{bmatrix} + \begin{bmatrix} -\cos \beta' \sin \alpha' \cdot \Delta\alpha - \sin \beta' \cos \alpha' \cdot \Delta\beta \\ \cos \beta' \cos \alpha' \cdot \Delta\alpha - \sin \beta' \sin \alpha' \cdot \Delta\beta \\ \cos \alpha' \cdot \Delta\beta \end{bmatrix} \quad (5)$$

脉冲星带误差的单位方向矢量为

$$\mathbf{n}' = \begin{bmatrix} \cos \beta' \cos \alpha' \\ \cos \beta' \sin \alpha' \\ \sin \beta' \end{bmatrix} \quad (6)$$

记单位方向矢量误差 $\Delta\mathbf{n}$ 为

$$\Delta\mathbf{n} = \begin{bmatrix} -\cos \beta' \sin \alpha' \cdot \Delta\alpha - \sin \beta' \cos \alpha' \cdot \Delta\beta \\ \cos \beta' \cos \alpha' \cdot \Delta\alpha - \sin \beta' \sin \alpha' \cdot \Delta\beta \\ \cos \alpha' \cdot \Delta\beta \end{bmatrix} \quad (7)$$

则可以将式(5)表示为

$$\mathbf{n} - \Delta\mathbf{n} = \mathbf{n}' \quad (8)$$

若不考虑时钟钟差, 则 $t'_{sat} \approx t_{sat}$, 可得

$$c(t_{SSB} - t'_{SSB}) = \Delta\mathbf{n} \cdot \mathbf{r}_{sat} \quad (9)$$

取状态变量为 $\mathbf{X} = [\Delta\alpha \ \Delta\beta]^T$, 则脉冲星方位误差估计的传统模型为

$$\mathbf{X}_{k+1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \mathbf{X}_k + \mathbf{W}_k = \mathbf{A}_k \mathbf{X}_k + \mathbf{W}_k \quad (10)$$

$$\mathbf{Z}_k = c(t_{SSB} - t'_{SSB}) = \mathbf{H}_k \mathbf{X}_k + \boldsymbol{\eta}_k \quad (11)$$

式中: \mathbf{W}_k 和 $\boldsymbol{\eta}_k$ 分别为系统噪声和量测噪声, k 表示第 k 时刻; 状态转移矩阵 $\mathbf{A}_k = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$; 观测矩阵 \mathbf{H}_k 为

卫星钟差模型为^[12]

$$x_1(t_{k+1}) = x_1(t_k) + \tau \cdot x_2(t_k) + \boldsymbol{\omega}(k) \quad (14)$$

式中: x_1, x_2 分别为钟差、钟差漂移率; τ 为步长; $\boldsymbol{\omega}$ 为白噪声, 其方差为

$$\mathbf{Q}(\tau) = E[\boldsymbol{\omega}(k)\boldsymbol{\omega}(k)^T] = \begin{bmatrix} q_1\tau + \frac{1}{3}q_2\tau^3 & \frac{1}{2}q_2\tau^2 \\ \frac{1}{2}q_2\tau^2 & q_2\tau \end{bmatrix} \quad (15)$$

式中: q_1, q_2 为噪声功率谱密度。其离散过程模型为

$$\begin{bmatrix} x_1(t_{k+1}) \\ x_2(t_{k+1}) \end{bmatrix} = \boldsymbol{\Phi}(\tau) \begin{bmatrix} x_1(t_k) \\ x_2(t_k) \end{bmatrix} + \boldsymbol{\omega}(k) \quad (16)$$

式中:状态转移矩阵为 $\Phi(\tau) = \begin{bmatrix} 1 & \tau \\ 0 & 1 \end{bmatrix}$ 。

取时钟误差漂移率为 $3.637\ 979 \times 10^{-12}$,根据铷原子钟模型,取时钟噪声功率谱密度分别为 $q_1 = 1.11 \times 10^{-22}$ s 和 $q_2 = 2.22 \times 10^{-32}/s^{[9]}$ 。给定时钟初始时刻的钟差为 0 s,取步长为 1 s,则可得钟差随时间的变化如图 2 所示。

由图 2 可知,随着时间的推移,钟差达到 5×10^{-6} s,将其与光速相乘理论上将造成 1 500 m 的误差,将会给系统造成不容忽略的影响。因此为验证其影响,仿真实验以脉冲星 B0531 + 21 作为观测脉冲星,其参数如表 1 所示。

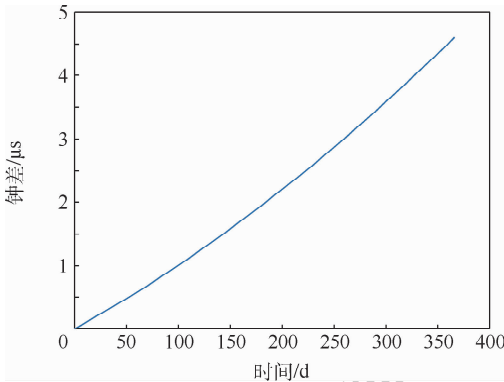


图 2 钟差随时间变化

Fig. 2 Clock error changes with time

表 1 脉冲星 B0531 + 21 参数

Table 1 Parameters of pulsar B0531 + 21

参数	数值
赤经/(°)	83.63
赤纬/(°)	22.01
距离/kpc	2.0
P/ms	33.4
W/ms	1.7
$F_x/(ph \cdot cm^{-2} \cdot s^{-1})$	1.54
$p_f/\%$	70

注:kpc 为秒差距,是天文学中使用的距离单位; $ph \cdot cm^{-2} \cdot s^{-1}$ 为宇宙背景噪声单位。

其中,脉冲周期、宽度、辐射光子流量分别为 P 、 W 、 F_x ; p_f 为脉冲周期中辐射流量与平均值之比。脉冲星的观测噪声方差^[13]可由式(17)计算得到:

$$\sigma = \frac{cW\sqrt{[B_x + F_x(1 - p_f)](At_{obs}d) + F_xAp_f t_{obs}}}{2F_xAp_f t_{obs}} \quad (17)$$

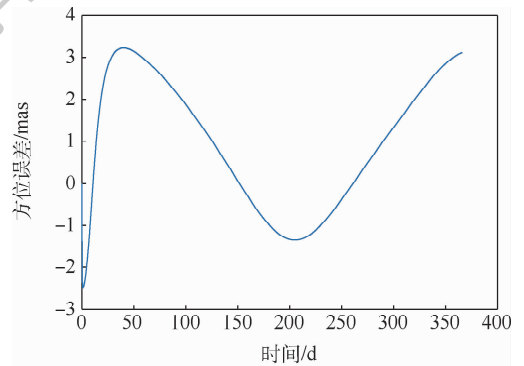
式中:设探测器有效面积 A 为 $1\ m^2$; $B_x = 0.005\ ph \cdot cm^{-2} \cdot s^{-1}$ 为宇宙背景噪声; d 为 W 与 P 之比; 设观测时间 t_{obs} 为 1 000 s,则可计算得到^[8] $\sigma = (77.69\ m)^2$ 。设方位误差初始值为 (2, 2) mas (毫角秒),并取初始状态为 0。

表 2 和图 3 为卫星轨道参数和仿真结果。由图 3 可知,在钟差的影响下赤经和赤纬误差估计结果均出现了较大的偏差,尤其是后者,甚至超过了 40 mas,且估计结果都不能收敛到一个固定值。因此,为了使赤经和赤纬误差估计更加准确,有必要考虑修正钟差的影响。

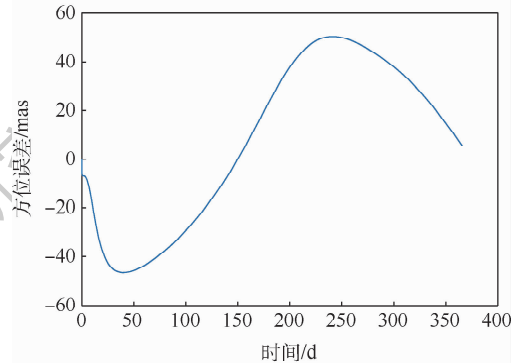
表 2 卫星轨道参数

Table 2 Parameters of satellite orbit

参数	数值
半长轴/km	7 460
离心率	4.55×10^{-16}
轨道倾角/(°)	25
近地点幅角/(°)	45
升交点赤经/(°)	0
初始真近地点/(°)	30
起始时间	2015-07-01T12:00:00



(a) 赤经误差估计结果



(b) 赤纬误差估计结果

图 3 未修正钟差的估计结果

Fig. 3 Estimation result with uncorrected clock error

2.2 卫星位置误差的影响

使用卫星估计方位误差的前提是卫星位置是精确已知的^[3],然而实际情况中,地面站得到的卫星位置会不可避免地存在偏差。因此当卫星位置存在误差 Δr 时,设 r_{sat} 与偏差位置 r'_{sat} 之间满足:

$$r_{sat} - \Delta r = r'_{sat} \quad (18)$$

忽略二阶小项,则观测模型变为

$$c(t_{SSB} - t'_{SSB}) = n \cdot r_{sat} - n' \cdot r'_{sat} = n' \cdot \Delta r + \Delta n \cdot r'_{sat} + \Delta n \cdot \Delta r \approx n' \cdot \Delta r + \Delta n \cdot r'_{sat} \quad (19)$$

此时若仍以式(9)作为观测模型,则会给系统引入一定的偏差。仿真时卫星轨道和脉冲星数据不变,设卫星位置误差在3个坐标方向的投影为100 m,估计结果如图4所示。

分析以上仿真结果可得,当信标卫星存在位置误差时,估计结果误差较大,且不收敛。可见,信标卫星位置误差也是不可忽略、需要被修正的。

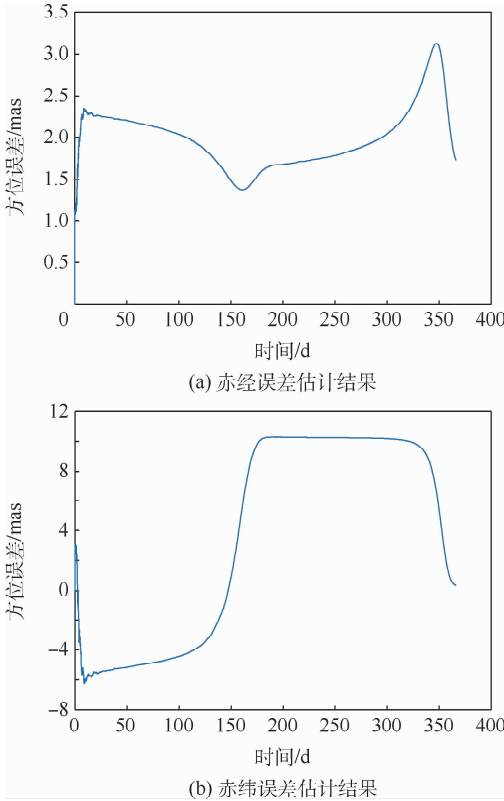


图4 未修正卫星位置误差的估计结果

Fig.4 Estimation results with uncorrected satellite position errors

2.3 钟差和卫星位置误差同时存在的影响

2.1节和2.2节是假设一种误差存在而另一种误差不存在而进行分析的,但是当利用信标卫星进行脉冲星方位误差估计时,时钟会发生漂移,同时卫星位置也不可避免地存在偏差。因此,当钟差和卫星位置误差同时存在时,可得系统新的观测模型为

$$c(t_{SSB} - t'_{SSB}) = c \cdot \delta t + (\mathbf{n}' + \Delta \mathbf{n}) \cdot (\mathbf{r}'_{sat} + \Delta \mathbf{r}) - \mathbf{n}' \cdot \mathbf{r}'_{sat} \approx c \cdot \delta t + \mathbf{n}' \cdot \Delta \mathbf{r} + \Delta \mathbf{n} \cdot \mathbf{r}'_{sat} \quad (20)$$

仿真条件不变,当2种误差都未修正时的仿真结果如图5所示。

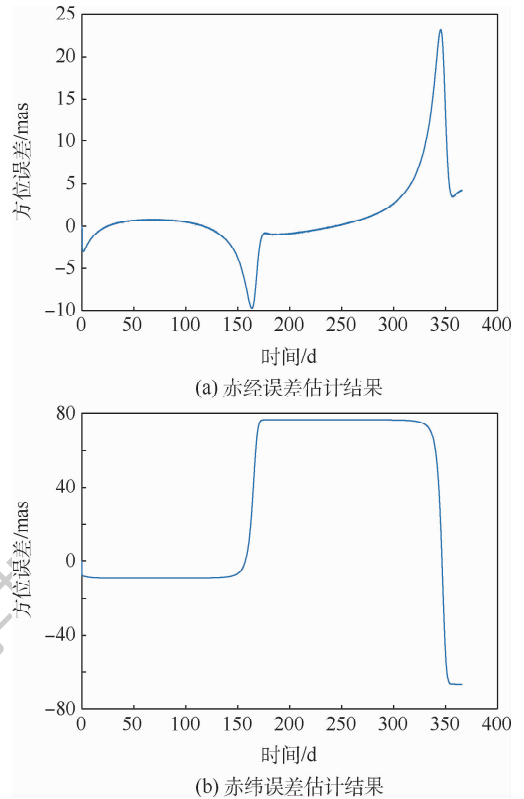


图5 2种误差都未修正的估计结果

Fig.5 Estimation results with two types of error uncorrected

可见,当同时存在钟差和卫星位置误差时,赤经和赤纬估计结果偏差较大,存在较大的发散。因此,为了得到更加精确的估计结果,要同时修正钟差和卫星位置误差的影响。

3 修正钟差和卫星位置误差的算法

3.1 增广状态模型

为了修正钟差的影响,对观测模型(20),在传统模型状态变量为 $[\Delta \alpha \quad \Delta \beta]^T$ 的基础上,将钟差和钟差漂移率扩展为新的状态量。扩展后的状态量为 $\mathbf{X} = [\Delta \alpha \quad \Delta \beta \quad x_1 \quad x_2]^T$,进而得到新的状态和观测方程为

$$\mathbf{X}_{k+1} = \mathbf{A}_k \mathbf{X}_k + \mathbf{W}_k \quad (21)$$

$$\mathbf{Z}_k = c(t_{SSB} - t'_{SSB}) = \mathbf{H}_k \mathbf{X}_k + \mathbf{n}' \cdot \Delta \mathbf{r} + \boldsymbol{\eta}_k \quad (22)$$

状态转移矩阵 $\mathbf{A}_k = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \tau \\ 0 & 0 & 0 & 1 \end{bmatrix}$, 观测矩阵

\mathbf{H}_k 为

$$\mathbf{H}_k = \begin{bmatrix} -\cos \beta' \sin \alpha' \cdot r_{sat/x} + \cos \beta' \cos \alpha' \cdot r_{sat/y} \\ \cos \beta' \cdot r_{sat/z} - \sin \beta' \cos \alpha' \cdot r_{sat/x} - \sin \beta' \sin \alpha' \cdot r_{sat/y} \\ 1 \\ 0 \end{bmatrix}^T$$

由于是线性时变系统,采用 PWCS 判据分析其可观性,将系统分为 j 个时间段,在各个时间段都将其视为线性定常系统^[3]。

系统总的可观性矩阵为

$$\mathbf{Q}(q) = \begin{bmatrix} \mathbf{Q}_1 \\ \mathbf{Q}_2 \mathbf{A}_1^{n-1} \\ \vdots \\ \mathbf{Q}_q \mathbf{A}_{q-1}^{n-1} \dots \mathbf{A}_1^{n-1} \end{bmatrix} \quad (23)$$

第 j 个时间段的可观矩阵为

$$\mathbf{Q}_j = \begin{bmatrix} \mathbf{H}_j \\ \mathbf{H}_j \mathbf{A} \\ \mathbf{H}_j \mathbf{A}^2 \\ \mathbf{H}_j \mathbf{A}^3 \end{bmatrix} = \begin{bmatrix} a^j & b^j & 1 & 0 \\ a^j & b^j & 1 & \tau \\ a^j & b^j & 1 & 2\tau \\ a^j & b^j & 1 & 3\tau \end{bmatrix} \quad j = 1, 2, \dots, q \quad (24)$$

式中:

$$\begin{cases} a_j = -\cos \beta' \sin \alpha' \cdot r_{\text{sat}/x}^j + \cos \beta' \cos \alpha' \cdot r_{\text{sat}/y}^j \\ b_j = \cos \beta' \cdot r_{\text{sat}/z}^j - \sin \beta' \cos \alpha' \cdot r_{\text{sat}/x}^j - \sin \beta' \sin \alpha' \cdot r_{\text{sat}/y}^j \end{cases} \quad (25)$$

其中任意相邻的 3 个时间段满足:

$$\text{rank} \begin{bmatrix} \mathbf{Q}_j \\ \mathbf{Q}_{j+1} \\ \mathbf{Q}_{j+2} \end{bmatrix} \leq \text{rank}(\mathbf{Q}(q)) \quad (26)$$

对其进行初等行变换可得

$$\text{rank} \begin{bmatrix} \mathbf{Q}_j \\ \mathbf{Q}_{j+1} \mathbf{A}^{n-1} \\ \mathbf{Q}_{j+2} \mathbf{A}^{n-1} \mathbf{A}^{n-1} \end{bmatrix} = \text{rank} \begin{bmatrix} a^{j+1} - a^j & b^{j+1} - b^j & 0 & 0 \\ a^{j+2} - a^j & b^{j+2} - b^j & 0 & 0 \\ a^j & b^j & 1 & 0 \\ 0 & 0 & 0 & \tau \end{bmatrix} \quad (27)$$

由于 $\tau > 0$,且卫星运行时相邻 3 个位置一般不共面,可得

$$\text{rank} \begin{bmatrix} 1 & 0 \\ 0 & \tau \end{bmatrix} = 2 \quad (28)$$

$$\text{rank} \begin{bmatrix} a^{j+1} - a^j & b^{j+1} - b^j \\ a^{j+2} - a^j & b^{j+2} - b^j \end{bmatrix} = 2 \quad (29)$$

所以

$$\text{rank} \begin{bmatrix} \mathbf{Q}_j \\ \mathbf{Q}_{j+1} \mathbf{A}^{n-1} \\ \mathbf{Q}_{j+2} \mathbf{A}^{n-1} \mathbf{A}^{n-1} \end{bmatrix} = 4 \quad (30)$$

即系统满足可观性条件。

3.2 TSKF 算法

对于观测模型(22),为修正卫星位置误差的影响,若仍采用增广状态法将卫星位置误差在 3 个坐标轴的分量增广为状态变量,则此时状态变量由 4 个增加到 7 个,不仅会增加计算量,还容易出现数值不稳定的问题^[11]。因此为减小计算量并确保算法的可观测性,采用 TSKF 算法进行解算。

Friedland 等^[14-15]最先提出 TSKF 算法并将其用在处理系统常值偏差问题上,Ignagni^[16]在此基础上将其应用到处理缓变偏差上。因此本文在新的状态和观测方程(21)和(22)基础上,根据两步卡尔曼滤波原理取第一步滤波的状态量为 $\mathbf{X} = [\Delta\alpha \ \Delta\beta \ x_1 \ x_2]^T$,第二步滤波状态量 \mathbf{b} 为卫星位置误差 $\Delta\mathbf{r}$,可将该算法的更新方程写为^[17-18]

一步滤波方程

$$\begin{cases} \mathbf{X}_{k/k-1} = \mathbf{A}_k \mathbf{X}_{k-1} \\ \mathbf{P}_{k/k-1} = \mathbf{A}_k \mathbf{P}_{k-1} \mathbf{A}_k^T + \mathbf{Q}_k \\ \mathbf{K}_k = \mathbf{P}_{k/k-1} \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_{k/k-1} \mathbf{H}_k^T + \mathbf{R}_k)^{-1} \\ \mathbf{X}_k = \mathbf{X}_{k/k-1} + \mathbf{K}_k (\mathbf{Z}_k - \mathbf{H}_k \mathbf{X}_{k/k-1}) \\ \mathbf{P}_k = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_{k/k-1} \end{cases} \quad (31)$$

二步滤波方程

$$\begin{cases} \mathbf{U}_k = \mathbf{A}_{k-1} \mathbf{V}_{k-1} + \mathbf{B}_{k-1} \\ \mathbf{S}_k = \mathbf{H}_k \mathbf{U}_k + \mathbf{C}_k \\ \mathbf{M}_k = \mathbf{M}_{k-1} - \mathbf{M}_{k-1} \mathbf{S}_k^T (\mathbf{H}_k \mathbf{P}_{k-1} \mathbf{H}_k^T + \mathbf{R}_k + \mathbf{S}_k \mathbf{M}_{k-1} \mathbf{S}_k^T)^{-1} \mathbf{S}_k \mathbf{M}_{k-1} \\ \mathbf{V}_k = \mathbf{U}_k - \mathbf{K}_k \mathbf{S}_k \\ \bar{\mathbf{K}}_k = \mathbf{M}_k (\mathbf{V}_k^T \mathbf{H}_k^T + \mathbf{C}_k^T) \mathbf{R}_k^{-1} \\ \mathbf{b}_k = (\mathbf{I} - \bar{\mathbf{K}}_k \mathbf{S}_k) \mathbf{b}_{k-1} + \bar{\mathbf{K}}_k (\mathbf{Z}_k - \mathbf{H}_k \mathbf{X}_{k/k-1}) \end{cases} \quad (32)$$

最终估计结果可写为

$$\tilde{\mathbf{X}}_k = \mathbf{X}_k + \mathbf{V}_k \mathbf{b}_k \quad (33)$$

式中: \mathbf{Q}_k 为 \mathbf{W}_k 的方差; \mathbf{P}_k 为状态量协方差; \mathbf{V}_{k-1} 为第二步状态变量对第一步状态变量的纠正矩阵; \mathbf{R}_k 为 $\boldsymbol{\eta}_k$ 的方差; \mathbf{B}_k 为 $\Delta\mathbf{r}$ 在式(21)中的驱动方程; \mathbf{C}_k 为 $\Delta\mathbf{r}$ 在式(22)中的驱动方程。由上述分析可得: $\mathbf{B}_k = \mathbf{0}, \mathbf{C}_k = \mathbf{n}'$ 。

仿真时,仍取脉冲星方位误差为(2,2) mas,时钟初始时刻的钟差为 0 s,卫星轨道参数、脉冲星参数、钟差模型参数及其他参数不变,可得同时修正钟差和卫星位置误差的 TSKF 算法估计结果如图 6 所示。

为进一步验证 TSKF 算法的有效性,其他条

件不变,设置不同的钟差和卫星位置误差,设置的条件和仿真结果分别如表3和表4所示。

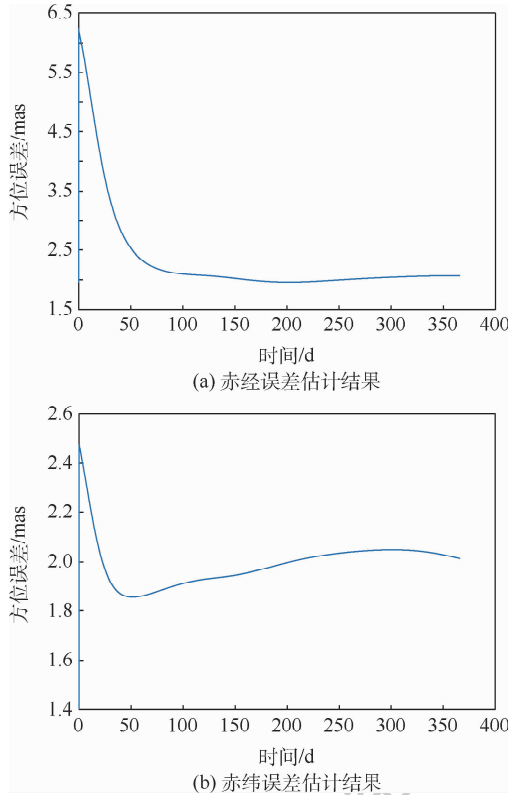


图6 TSKF算法估计结果

Fig. 6 Estimation results of TSKF algorithm

表3 条件设置

Table 3 Condition setup

钟差/ μs	卫星位置误差/km
0	(0.1,0.1,0.1)
1	(0,0,0)
1	(0.1,-0.1,0.1)
2	(0.1,0.1,0.1)
2	(0.2,-0.2,0.2)
2	(0.5,0.5,0.5)
5	(0.1,0.1,0.1)
5	(0.2,-0.2,0.2)

表4 仿真结果

Table 4 Simulation results

赤经估计偏差/mas	赤纬估计偏差/mas
0.050	0.043
0.061	0.047
0.150	0.065
0.150	0.072
0.172	0.110
0.221	0.150
0.182	0.143

由图6可得,赤经误差估计结果收敛到2.2 mas以内,赤纬误差估计结果也在1.8 mas和2.1 mas之间,可见TSKF算法能使方位误差估计

结果实现较好的收敛,使赤经和赤纬估计结果保持在约0.2 mas以内的精度。且由表4可得,在不同的钟差和卫星位置误差下,TSKF算法均能使估计结果保持较高的精度。因此可得,TSKF算法能有效隔离钟差和卫星位置误差的影响,使赤经和赤纬误差估计结果保持在没有误差影响的水平下。

4 结论

1) 在利用信标卫星进行脉冲星方位误差估计时,时钟钟差和卫星位置误差是需要被修正的,否则会严重影响估计精度。

2) 在重新推得的包含钟差和卫星位置误差模型基础上,采用TSKF算法能有效克服钟差和卫星位置误差的影响,使脉冲星方位误差估计精度显著提高。

参考文献 (References)

[1] SHEIKH S I. The use of variable celestial X-ray sources for spacecraft navigation [D]. Maryland: University of Maryland, 2005.

[2] 刘劲. 基于X射线脉冲星的航天器自主导航方法研究[D]. 武汉:华中科技大学,2011.
LIU J. X-ray pulsar-based spacecraft autonomous navigation [D]. Wuhan: Huazhong University of Science and Technology, 2011 (in Chinese).

[3] 孙守明. 基于X射线脉冲星的航天器自主导航方法研究[D]. 长沙:国防科学技术大学,2011.
SUN S M. Study on autonomous navigation method of spacecraft based on X-ray pulsars [D]. Changsha: National University of Defense Technology, 2011 (in Chinese).

[4] 孙守明,郑伟,汤国建. X射线脉冲星表方位误差估计算法研究[J]. 飞行器测控学报,2010,29(2):57-60.
SUN S M, ZHENG W, TANG G J. A new estimation algorithm of the X-ray pulsar position error [J]. Journal of Spacecraft TT&C Technology, 2010, 29(2): 57-60 (in Chinese).

[5] 熊凯,魏春岭,刘良栋. 鲁棒滤波技术在脉冲星导航中的应用[J]. 空间控制技术与应用,2008,34(6):8-11.
XIONG K, WEI C L, LIU L D. Application of robust filtering in pulsars-based navigation [J]. Aerospace Control and Application, 2008, 34(6): 8-11 (in Chinese).

[6] XU Q, WANG H L, FENG L, et al. An improved augmented X-ray pulsar navigation algorithm based on the norm of pulsar direction error [J]. Advances in Space Research, 2018, 62(11): 3187-3198.

[7] NING X L, GUI M Z, FANG J C, et al. Differential X-ray pulsar aided celestial navigation for Mars exploration [J]. Aerospace Science and Technology, 2017, 62: 36-45.

[8] 王宏力,许强,由四海,等. 考虑卫星位置误差的增广脉冲星方位误差估计算法[J]. 国防科技大学学报,2018,40(5):177-182.
WANG H L, XU Q, YOU S H, et al. Augmented estimation al-

- gorithm for pulsar position error with satellite position error[J]. Journal of National University of Defense Technology, 2018, 40(5):177-182 (in Chinese).
- [9] 孙守明, 郑伟, 汤国建. X射线脉冲星/SINS组合导航中的钟差修正方法研究[J]. 国防科技大学学报, 2010, 32(6): 82-86.
SUN S M, ZHENG W, TANG G J. A new clock error control algorithm of X-ray pulsars/SINS integrated navigation[J]. Journal of National University of Defense Technology, 2010, 32(6): 82-86 (in Chinese).
- [10] 王璐, 史晨曦, 李建勋, 等. 考虑钟差修正的脉冲星与多普勒差分组合导航[J]. 电波科学学报, 2018, 33(2): 217-224.
WANG L, SHI C X, LI J X, et al. Pulsar and Doppler difference integrated navigation considering clock error control[J]. Chinese Journal of Radio Science, 2018, 33(2): 217-224 (in Chinese).
- [11] EMADZADEH A A, SPEYER J L. X-ray pulsar-based relative navigation using epoch folding[J]. IEEE Transactions on Aerospace and Electronic Systems, 2011, 47(4): 2317-2328.
- [12] 李孝辉, 吴海涛, 高海军, 等. 用 Kalman 滤波器对原子钟进行控制[J]. 控制理论与应用, 2003, 20(4): 551-554.
LI X H, WU H T, GAO H J, et al. Clock disciplined method by using Kalman filter[J]. Control Theory & Applications, 2003, 20(4): 551-554 (in Chinese).
- [13] 帅平, 陈绍龙, 吴一帆, 等. X射线脉冲星导航技术研究进展[J]. 空间科学学报, 2007, 27(2): 169-176.
SHUAI P, CHEN S L, WU Y F, et al. Advance in X-ray pulsar navigation technology[J]. Chinese Journal of Space Science, 2007, 27(2): 169-176 (in Chinese).
- [14] FRIEDLAND B. Treatment of bias in recursive filtering[J]. IEEE Transactions on Automatic Control, 1969, 14(4): 359-367.
- [15] HAESSIG D, FRIEDLAND B. Separate-bias estimation with reduced-order Kalman filters[J]. IEEE Transactions on Automatic Control, 1998, 43(7): 983-987.
- [16] IGNAGNI M B. Separate bias Kalman estimator with bias state noise[J]. IEEE Transactions on Automatic Control, 1990, 35(3): 338-341.
- [17] 王奕迪, 郑伟, 孙守明, 等. 考虑系统偏差的脉冲星守时算法研究[J]. 国防科技大学学报, 2013, 35(2): 12-16.
WANG Y D, ZHENG W, SUN S M, et al. Algorithm for the pulsar timing system with the system bias[J]. Journal of National University of Defense Technology, 2013, 35(2): 12-16 (in Chinese).
- [18] 许强, 范小虎, 徐利国, 等. 脉冲星方位误差估计的 TSKF 算法[J]. 北京航空航天大学学报, 2020, 46(4): 761-768.
XU Q, FAN X H, XU L G, et al. TSKF algorithm for pulsar position error estimation[J]. Journal of Beijing University of Aeronautics and Astronautics, 2020, 46(4): 761-768 (in Chinese).

Two-step Kalman filter algorithm for pulsar position error estimation

XIAO Yongqiang¹, WANG Hongli^{1,*}, FENG Lei¹, YOU Sihai¹, XU Qiang²

(1. Guidance Office, Rocket Force University of Engineering, Xi'an 710025, China;

2. Department of Test and Control, Qingzhou Research Institute of High-technology, Qingzhou 262500, China)

Abstract: A Two-Step Kalman Filter (TSKF) algorithm is designed to overcome the influence of clock error and satellite position error on pulsar position error estimation. First, the traditional model of pulsar position error estimation is introduced, and it is confirmed by analysis that the clock error, satellite position error, and both errors will have serious impact on the estimation. Second, the clock error and satellite position error are added to the traditional estimation model, and the clock error and its rate of change are expanded to a new state quantity, thereby deducing a new model of pulsar position error estimation containing these two errors. And its observability is proved through theoretical analysis. Then the update equation of the TSKF algorithm is written combined with the new model and based on the two-step Kalman filter principle. Finally, simulations show that the TSKF algorithm can effectively isolate the influences of the two errors and make the estimation accuracy kept within 0.2 mas while the traditional pulsar position error estimation algorithm has a large deviation and divergence of the right ascension and declination errors under the influence of the two errors.

Keywords: position error; clock error; satellite position error; augmented state; Two-Step Kalman Filtering (TSKF)

Received: 2020-05-22; Accepted: 2020-08-14; Published online: 2020-09-08 10:12

URL: kns.cnki.net/kcms/detail/11.2625.V.20200905.1337.001.html

Foundation items: National Natural Science Foundation of China (61503391); China Postdoctoral Science Foundation (2017M613327)

* Corresponding author. E-mail: wanghongli19650526@163.com

http://bhxb.buaa.edu.cn jbuua@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2020.0222

面向 BDSBAS 电文认证的 OTAR 播发策略设计



穆盛林^{1,2}, 陈颖³, 刘婷¹, 刘成³, 陈潇^{1,2,*}

(1. 中国科学院大学 光电学院, 北京 100049; 2. 中国科学院 空天信息创新研究院, 北京 100094;

3. 北京跟踪与通信技术研究所, 北京 100094)

摘 要: 星基增强系统(SBAS)是向用户提供完好性信息的广域增强系统,涉及生命安全领域。为了提高 SBAS 系统的抗欺骗能力,基于椭圆曲线数字签名算法(ECDSA)提出面向北斗星基增强系统(BDSBAS)的电文认证设计方案。根据国内商用密码标准椭圆曲线(SM2)算法进行相应参数设计和电文排布方案,并描述了空中密钥更新(OTAR)的电文播发策略。为了进一步对试验方案进行验证,采用蒙特卡罗 OTAR 仿真器,基于3种播发调度算法进行仿真试验,通过调整权重对不同情况下 OTAR 电文接收时间进行分析。仿真结果为具体情况下 OTAR 电文方案播发调度的选择提供了重要参考依据。

关键词: 北斗星基增强系统(BDSBAS); 电文认证; 椭圆曲线数字签名算法(ECDSA); 空中密钥更新(OTAR); 播发策略

中图分类号: V19; X949

文献标志码: A

文章编号: 1001-5965(2021)07-1453-09

星基增强系统(Satellite-based Augmentation System, SBAS)是利用地球轨道同步(GEO)卫星作为通信媒介,向用户提供差分校正参数和完好性信息的广域增强系统。SBAS 涉及生命安全领域,满足民航从航路飞行阶段到垂向引导精密进近阶段的导航需求。随着近年来频发的 GNSS 欺骗攻击事件,SBAS 系统的信号格式公开以及电文可预测等脆弱性逐步凸显,SBAS 系统面临遭受欺骗攻击的安全风险。

SBAS 系统为用户提供 GNSS 差分参数和完好性增强电文,而针对 SBAS 用户的欺骗模式是产生与真实 SBAS 信号高度相似的虚假信号,并篡改其中的 SBAS 电文。面对 SBAS 电文篡改的欺骗方式,欧洲和美国均开展相应研究,并提出了系统端解决方案——SBAS 电文认证方法及服务策略,并计划将电文认证服务加入未来双频多星座(Dual Frequency Multi Constellation, DFMC)标

准。2016年,欧洲启动了 EAST(EGNOS Authentication Security Test-bed)项目^[1],研究适用于 EGNOS 的认证服务方案并评估其服务性能,提出了基于椭圆曲线数字签名算法(Elliptic Curve Digital Signature Algorithm, ECDSA)和时间效应流丢失容错认证方案(Time Efficient Stream Loss-tolerant Authentication, TESLA)^[2]。美国斯坦福大学同时开展了基于美国密码标准的2种认证方案设计与仿真,并提出了基于空中密钥更新(Over the Air Rekeying, OTAR)的密钥播发策略^[3]。随着 ICAO NSP DS2 工作组积极推进 SBAS 认证服务标准化以及 BDSBAS 系统建设与发展,迫切需要面向 BDSBAS 的认证服务开展适用于中国密码标准的认证服务方案,并开展 OTAR 密钥播发策略设计。

本文面向北斗星基增强系统(BDSBAS)设计了一种基于中国密码标准椭圆曲线(SM2)的电文认证方法,并开展了基于该方法的 OTAR 播发策

收稿日期: 2020-05-27; 录用日期: 2020-07-28; 网络出版时间: 2020-08-05 13:02

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20200805.1000.001.html

* 通信作者。E-mail: chenxiao@aircas.ac.cn

引用格式: 穆盛林, 陈颖, 刘婷, 等. 面向 BDSBAS 电文认证的 OTAR 播发策略设计[J]. 北京航空航天大学学报, 2021, 47(7): 1453-1461. MU S L, CHEN Y, LIU T, et al. Design of message authentication and OTAR broadcast strategy for BDSBAS[J]. Journal of Beijing University of Aeronautics and Astronautics, 2021, 47(7): 1453-1461 (in Chinese).

略仿真试验。首先,简要概述了 ECDSA 的原理,然后,提出了基于 ECDSA 算法的 SBAS 认证服务方案和 OTAR 播发策略,最后,针对 OTAR 播发策略进行了相关的仿真试验,评估了 3 种播发策略的性能。

1 SBAS 电文认证

1.1 电文认证原理

SBAS 电文认证是在现有 SBAS 广播电文中加入特殊的认证电文,使用户确认 SBAS 信号是否来自真实的 GEO 卫星以及 SBAS 电文是否被伪造及篡改^[4]。电文认证重点保障 SBAS 电文的完整性,在不影响用户正常获得基本的 SBAS 服务的情况下,通过增加电文完整性核验和信号源身份认证等方法为用户提供更加安全的导航服务,以提升抗欺骗攻击能力^[5-6]。

如图 1 所示,地面站观测 GNSS 导航信号,获得观测数据和导航电文,数据处理中心对地面站观测数据进行处理得到增强电文,并接收来自密钥管理中心的密钥对,利用私钥使用数字签名的方式为增强电文生成对应的认证电文,并通过上注站将调制 SBAS 电文(含认证电文和公钥)的 SBAS 信号上注至 GEO 卫星,GEO 卫星播发 SBAS 信号包含认证电文,用户端利用接收的 SBAS 电文和认证电文实现发射源身份认证;而欺骗攻击设备产生 SBAS 欺骗信号难以伪造认证电文(主要是生成式欺骗攻击),用户端可以通过认证电文辨别真实 SBAS 信号,提高抗欺骗的安全性^[7]。

认证算法包括 2 种:结合消息认证码和密钥延迟发布的 TESLA 算法,以及 ECDSA 算法^[8]。相对于 TESLA 算法,ECDSA 算法的优势是具有国际标准,安全性和通用性有所保障,且不需要满足时间同步要求,实现简单。因此,后续章节将重点介绍以 ECDSA 算法为基础的电文认证方法。

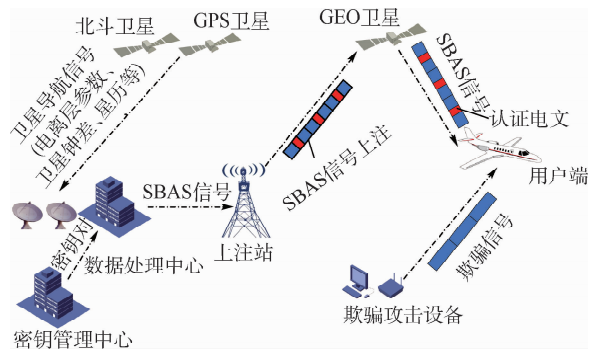


图 1 SBAS 电文认证原理

Fig. 1 Principle of SBAS message authentication

1.2 ECDSA 算法

ECDSA 算法是利用椭圆曲线密码体制(ECC)对数字签名算法的模拟。与其他公钥密码体制相比,椭圆曲线密码体制的优点为:密钥长度较短,参数选择灵活,具有数学难题保证的安全性,实现速度较快^[9-10]。

ECDSA 算法的原理是“私钥签名,公钥验证”,具体包括 4 个步骤:

步骤 1 系统建立。输入安全参数 k ,输入系统公开参数 $\text{params} = \{p, q, E, G\}$, p, q 表示有限域的规模,为大素数或为 2 的幂次, E 表示定义在有限域 F_p 上的椭圆曲线, G 表示椭圆曲线 E 上阶为 Q 的生成元点。

步骤 2 密钥产生。输入系统公开参数 params ,用户随机选取秘密值 $d \in [1, q - 1]$,计算 $P = d[*]G$,将 P 作为公钥公开, d 作为私钥秘密保存。

步骤 3 签名生成。输入系统公开参数 params ,私钥 d 和待签名消息 M ,签名者按如下步骤生成签名:

- 1) 令 $\overline{M} = Z \| M$, Z 表示签名者身份标识与系统参数的杂凑值。
- 2) 计算 $e = \text{hash}(\overline{M})$,将 e 转换为整数, $\text{hash}()$ 为单向哈希函数。
- 3) 产生安全参数 $k \in [1, q - 1]$,计算 $k[*] \cdot G = (x_1, y_1)$ 。
- 4) 计算 $r = (x_1 + e) \bmod q$,若 $r = 0$ 或 $r + k = q$,则重新选取安全参数 k 。
- 5) 计算 $s = (1 + d)^{-1}(k - rd) \bmod q$,若 $s = 0$,则重新选取安全参数 k ,否则将 (r, s) 作为签名结果输出。

步骤 4 签名验证。输入系统公开参数 params ,公钥 P 以及收到的消息 M' 的签名 (r', s') ,验证者按照如下步骤验证签名:

- 1) 检验 $s' \in [1, q - 1]$ 是否成立,若不成立,则验证不通过。
- 2) 检验 $r' \in [1, q - 1]$ 是否成立,若不成立,则验证不通过。
- 3) 令 $M' = Z \| M$ 。
- 4) 计算 $e' = \text{hash}(M')$ 。
- 5) 计算 $t = (r' + s') \bmod q$,若 $t = 0$,则验证不通过。
- 6) 计算椭圆曲线点 $(x'_1, y'_1) = s'G + tP$ 。
- 7) 计算 $r'' = (x'_1 + e') \bmod q$,检验 $r'' = r'$ 是否成立,若成立,则验证通过,否则验证不通过。

1.2.1 ECDSA 认证电文格式

随着 DFMC 标准的推进,SBAS 认证服务将考虑在 L5 频点播发。L5 频点采用 QPSK 调制方式,包括 I 支路和 Q 支路^[11]。

I 支路:播发多星座增强电文,播发速率 250 bit/s,包括 4 bits 前导码,24 bits CRC 以及 216 bits 增强电文数据位。

Q 支路:仅作为导频通道,未调制电文。

认证电文考虑 2 种播发方式: I 支路和 Q 支路。

I 支路认证电文播发框架如图 2 所示。认证电文可根据规定的认证时间间隔插入增强电文中,每个认证电文包含对前一个认证电文之后的所有增强电文的签名。这种方法存在延迟认证的问题,每一个签名需要保护的增强电文数量较多。

Q 支路认证电文播发框架如图 3 所示。增强电文与认证电文一一对应,但并不意味着一个签名仅保护一个增强电文,因为一般情况下,签名长度会长于增强电文。这种方式可减小认证延迟,但需要考虑 I/Q 支路的功率分配问题。

ECDSA 算法的数字签名长度超过 216 bits,若将认证电文在 I 支路播发,则至少占用 2 s 数据位,会大大增加认证的延迟;若在 Q 支路播发,则不需要考虑认证延迟问题,至多可以每 2 s 生成一次签名。因此,本文选择 L5Q-ECDSA 的方式播发认证电文。



图 2 L5I 认证电文播发框架

Fig.2 L5I authentication message broadcasting framework

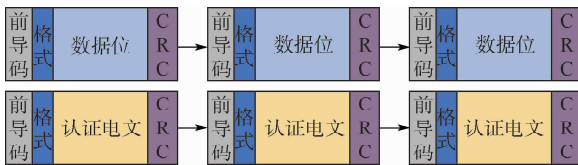


图 3 L5Q 认证电文播发框架

Fig.3 L5Q authentication message broadcasting framework

1.2.2 ECDSA 安全等级

安全等级用于衡量密码算法被暴力破解的难度。美国国家标准与技术研究院(NIST)发布了密码算法安全等级和未来使用标准的建议,如表 1 所示。依据美国密码安全标准,在 2030 年之前,安全等级在 112 位以上的密码算法均处于安全范围内,可以使用^[12]。

表 1 NIST 密码算法安全等级

Table 1 NIST cryptographic algorithm security level

安全等级/位	应用层级	2030 年之前	2031 年以后
< 112	应用	不可用	不可用
	进程	存在风险	不可用
112	应用	可用	不可用
	进程	可用	存在风险
128	应用/进程	可用	可用
192	应用/进程	可用	可用
256	应用/进程	可用	可用

仅依靠密码算法自身的安全等级并不能满足 SBAS 对认证安全性的要求。由于公钥是公开的,不能保证公钥来源的真实性,欺骗攻击方可以伪造一对密钥,用伪造的私钥生成签名,将签名和伪造的公钥发送至用户端,冒充真实签名和公钥,从而完成欺骗攻击。

应对 BDSBAS 的密钥安全问题,可利用公钥基础设施(Public Key Infrastructure, PKI)实现公钥证书管理。PKI 通过为公钥及相关的 SBAS 服务提供身份信息签发公钥证书,绑定证书持有者的身份和相关的密钥对,为终端用户提供方便的证书申请、证书作废、证书获取、证书状态查询的途径,并利用公钥证书及相关的各种服务实现 SBAS 服务的身份认证、完整性、抗抵赖性和保密性。

1.3 关键性能指标

SBAS 电文认证关键性能指标(Key Performance Indicator, KPI)包括认证时间间隔、首次认证时间、最大认证延迟、认证告警时间^[13]。

认证时间间隔(Time Between Authentication, TBA):表示 2 次认证事件时间间隔,TBA 需要权衡 SBAS 系统服务性能。例如,当 TBA 较小则需要频繁传输的认证电文,会消耗大量的带宽从而降低 SBAS 性能;当 TBA 较大,导致认证电文传输时间过长(几分钟甚至更长),用户接收机会在较长的时间使用未经认证的 SBAS 信号,将增加用户受到欺骗攻击的风险。

最大认证延迟(Max Authentication Latency, MAL):表示用户从接收 SBAS 信号到完成 SBAS 认证所需最大时间延迟。认证延迟与完好性告警时间(Time to Alert, TTA)直接相关。理想的认证延迟应该是 6 s 或更少,因为 SBAS 完好性告警时间是 6 s。TBA 和 MAL 的关系如图 4 所示。

首次认证时间(Time to First Authentication, TTFA):表示用户在完成第一次信号采集后检测电文是否真实或伪造所需的时间;指检索被测身

份验证协议所需信息的时间,需要在 SBAS 完好性告警时间内完成。

认证告警时间 (Alert Time to Authentication, ATTA):表示用户正确停止使用未认证信号的时间,或者消除未认证信号对接收机输出影响的时间,即对于给定的攻击事件,一个或多个接收机的输出受到欺骗信号影响的时间与接收机收到认证失败通知或移除欺骗信号对接收机输出影响的时间的差值。理论条件下认证失败会马上发出告警。

现有 SBAS 系统完好性认证告警时间是 6 s, SBAS 认证服务不能影响完好性服务,因此 SBAS 信号认证服务需要满足 6 s 最大认证延迟。

表 2 展示了对本文方案设计产生影响的关 键性能指标及理论结果,是设计方案的 主要依据。

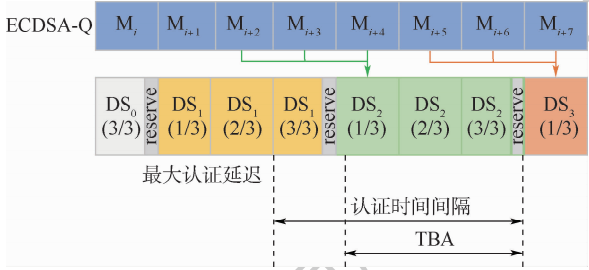


图 4 最大认证延迟和认证时间间隔

Fig. 4 Authentication latency and time between authentication

表 2 关键性能指标

Table 2 Key performance indicators

Table with 3 columns: 核心指标 (Core Indicator), 定义 (Definition), 指标数值/s (Indicator Value/s). Rows include TBA, TTFA, MAL, ATTA.

1.4 设计方案

电文发送端产生及加入认证电文的方式如图 5 所示。SBAS 电文由 I 支路承载,SBAS 电文与存储在发送端的 ECDSA 私钥计算生成对应 SBAS 电文的数字签名,并将数字签名及 ECDSA

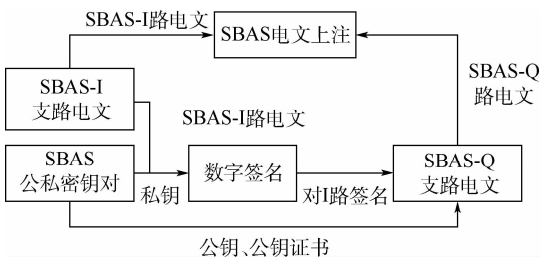


图 5 电文发送端运行流程

Fig. 5 Operation flowchart of message sending end

公钥、数字证书等电文整合,由 Q 支路承载,将 I 支路和 Q 支路电文整合后生成调制信号,通过上注站上注至 GEO 卫星。

中国 ECDSA 标准采用中国商用密码标准 SM2 算法,SM2 算法拥有固定的 128 位安全等级,完全满足 NIST 对密码算法安全等级的标准。在 128 位安全等级下,私钥长度为 256 bits,公钥长度为 512 bits,签名长度为 512 bits。由于 Q 支路播发速率为 250 bit/s,播发一条完整的签名需要占用 3 s 的数据位,剩余数据位可用于密钥管理。图 6 为 L5Q-ECDSA 方案的电文序列,其中 M 表示增强电文,S 为对应电文生成的签名,OTAR 为密钥更新信息。

结合 SBAS 电文信息类型和目前的增强电文播发策略,本文给出了 L5Q-ECDSA 认证方案的 I 支路和 Q 支路电文播发时序安排,如图 7 所示。由于完好性信息更新周期为 6 s,以 6 s 为 1 个周期,表征了 2 个周期的电文编排。其中,MT₀ 为测试电文,表示该编排格式为测试样例,MT₃₄、MT₃₅、MT₃₆ 为传递卫星完好性信息的电文,MT₃₉、MT₄₀ 为传递卫星伪距及星历等信息的电文,自定义类型 MT₅₁ 用于传输认证电文,具体包括数字签名和 OTAR。

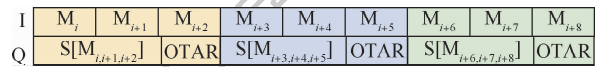


图 6 L5Q-ECDSA 电文序列

Fig. 6 L5Q-ECDSA message sequence

Table showing ECDSA message sequence of time with columns for time slots (时序1-12) and rows for I and Q branches across two 6-second intervals.

图 7 ECDSA 电文时序

Fig. 7 ECDSA message sequence of time

SM2 标准公钥长度为 512 bits,签名长度为 512 bits。L5Q 电文设计如图 8 所示。为播发全部签名,L5Q 支路按照 250 bit/s 速率,至少需要 3 s 的数据位进行播发,共 750 bits,其他数据位包括 4 bits 前导码,72 bits CRC (24 bit/s) 和 162 bits OTAR 位,在 OTAR 中,前 4 位是 OTAR 头,用于声明包含在 OTAR 位中的信息,根据要传输的 OTAR 信息的长短,可能要将 OTAR 信息分为几段进行传输,x 位序列号用于标注 OTAR 信息的顺序,占用的位数由 OTAR 信息长度决定,其余位将用来传输相关的 OTAR 数据信息。

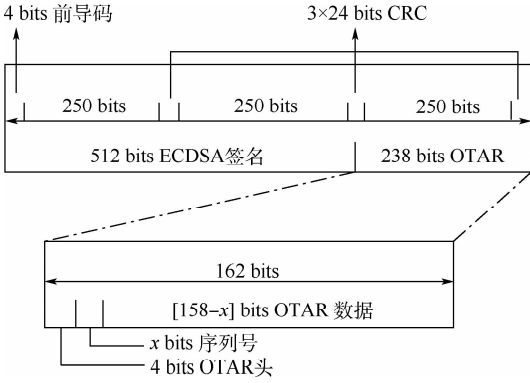


图 8 L5Q 路电文结构

Fig. 8 L5Q-channel message structure

2 OTAR 播发策略

SBAS 认证服务使用的密钥需要定期更换,防止被恶意攻击者截获利用,当密钥出现问题时,也需要及时更新密钥,防止影响用户端接收机的认证工作。密钥的更新工作需要通过 OTAR 位进行。OTAR 空中密钥更新指利用 SBAS 广播电文进行密钥的更新和管理。

2.1 播发策略

OTAR 位中传递的具体内容主要包括公钥、公钥证书以及公钥到期声明。公钥指当前签名对应的公钥,用于验证签名;公钥证书指用于保护公钥安全性的签名,其内容包括公钥签名、服务方身份标示、证书有效期等内容;公钥到期声明指在当前公钥到达有效期时的声明,提醒用户方公钥以及接收新的公钥。其中公钥和公钥证书为播发周期相对较短的信息,为了保证短暂关闭的接收机可以获取当前使用的公钥,公钥到期声明仅在更换公钥期间保证广播频率。表 3 展示了本文定义的 OTAR 播发内容。其中系统公钥指验证 SBAS 认证服务中签名的公钥,公钥证书公钥指验证公钥证书的公钥。

表 3 OTAR 播发内容

Table 3 OTAR broadcast content

OTAR 电文格式 (OMT)	定义	长度/bit
OMT1	当前系统公钥	512
OMT2	OMT1 的公钥证书	512
OMT3	下一个系统公钥	512
OMT4	OMT3 的公钥证书	512
OMT5	当前系统公钥/公钥证书公钥到期声明	162
OMT6	OMT5 的公钥证书	512
OMT7	当前公钥证书公钥	512
OMT8	OMT7 的公钥证书	512
OMT9	下一个公钥证书公钥	512
OMT10	OMT9 的公钥证书	512
OMT11-13	保留位置	

OTAR 的播发基于 3 种调度算法^[14-15]:

1) 标准分组公平队列 (Packet Fair Queuing, PFQ) 算法,该算法将 OTAR 电文分为多个组,每个组都被分配相同的带宽。使用标准分组公平队列算法进行播发时,会将 OTAR 电文分成若干段,这些分段电文会按照原始顺序进行播发,如 OMT_i 电文会分为 OMT_{i1}、OMT_{i2}、OMT_{i3},播发时这些序列也同时按此顺序播发,而不会被其他电文中断。PFQ 算法允许为每个组添加权重,根据权重比可以为权重高的组分配更高的带宽。

2) 半刚性分组公平队列 (Semi Rigid Packet Fair Queuing, PFQ-Semi-Rigid) 算法,是 PFQ 算法的演进算法,该算法允许设置其中重要子队列的最大传输时间间隔。采用半刚性分组公平队列算法传输 OTAR 电文,可以设置 OMT1 和 OMT2 为重要电文,通过设置 2 种电文间的最大时间间隔,可有效确保按时发送重要电文的时间上限。

3) 分裂型分组公平队列 (Split Packet Fair Queuing, PFQ-Split) 算法,是 PFQ 算法的演进算法,该算法消除了分段电文必须按照顺序播发的限制,例如 OMT_i 电文会分为 OMT_{i1}、OMT_{i2}、OMT_{i3},播发时这些电文可随机播发。

2.2 评价标准

评估 OTAR 设计方案的主要指标为接收机在冷启动后接收当前系统公钥所需要的时间,以及接收全部 OTAR 电文所需要的时间。表 4 展示了本文主要讨论的评估指标。由于当前系统公钥和公钥证书播发频率最高,接收当前系统公钥和公钥证书所需要的时间为主要指标,包括平均时间、最小时间、最大时间和时间范围。

表 4 评估指标

Table 4 Evaluation indicators

评估指标	当前系统公钥/公钥证书	全部 OTAR
平均时间	表征接收 OMT1 和 OMT2 所需平均时间	表征接收全部 OTAR 电文所需平均时间
最大时间	表征接收 OMT1 和 OMT2 所需最大时间	表征接收全部 OTAR 电文所需最大时间
最小时间	表征接收 OMT1 和 OMT2 所需最小时间	表征接收全部 OTAR 电文所需最小时间
时间范围	表征接收 OMT1 和 OMT2 所需时间区间	表征接收全部 OTAR 电文所需时间区间

3 仿真结果与分析

3.1 仿真场景

本文使用蒙特卡罗 OTAR 仿真器验证试验方

案。蒙特卡罗仿真模拟器是一款 MATLAB 中的工具包,主要功能是仿真从电文播发到用户接收的过程。仿真包括 3 个阶段:

1) 定义 OTAR 电文格式。指定义 OTAR 消息类型以及长度的分配。

2) 确定电文播发策略。指电文在进行播发时存在多种播发调度算法,调度算法会对 OTAR 电文进行分配,决定每个 OTAR 电文会在何时播发。

3) 用户端接收模拟。指对用户接收电文过程的模拟,仿真过程中模拟的用户接收机均为冷启动,即第一次接收系统公钥及公钥证书,因此可以接收所有需要的 OTAR 电文。

3.2 结果与分析

影响仿真结果的因素有 3 个,即权重分配、调度算法和解调错误率。为播发频率较高的 OTAR 内容赋予较高的权重,会使这些内容的接收时间更短;采用不同的调度算法,接收时间的结果也会发生变化;接收机接收 SBAS 电文时有一定概率发生解调错误,使当前电文无法进行认证。本章分别考虑上述各因素影响下的信号收发情况的仿真。每次仿真时仅考虑一项因素的影响,同时保持其他两项因素的一致性。

3.2.1 权重分析

OMT1 和 OMT2 为播发频率最高的电文,因此为这组电文赋予更高的权重 W_1 ,其他 OTAR 电文赋予较低的权重 W_r ,使接收端有更大的可能在最短的时间内接收 OMT1 和 OMT2。通过调整 W_1/W_r 的比值,分析权重比的变化对 OTAR 电文接收时间的影响。仿真试验在解调错误率 $PER = 0$,调度算法为 PFQ 算法的条件下进行,分别对权重比 $W_1/W_r = [10^0, 10^1, 10^2, 10^3, 10^4]$ 时,每种权重比进行 10 000 次播发模拟,并绘制出变化趋势图,结果如图 9 所示。图中下方曲线范围表示接收 OMT1 和 OMT2 所需要的时间范围,上方曲线范围表示接收全部 OTAR 电文所需要的时间范围。

表 5 展示了图 9 中几个关键点位的详细时间结果。权重比 W_1/W_r 从 10^0 到 10^4 增长过程中,接收 OMT1 和 OMT2 的最小时间始终为 24 s,最大时间从 123 s 降至 45 s,平均时间从 58 s 降至 24.5 s, $W_1/W_r = 10^4$ 时,平均时间和最小时间几乎相同;接收全部 OTAR 电文的时间持续增长,平均时间从 187 s 增至 5 896 s。

结果显示,权重比 W_1/W_r 越高,接收权重高的系统公钥所需的时间越短,但接收全部 OTAR

电文的总时间更长,且随着权重比的增高,接收系统公钥所需平均时间越接近最小时间。

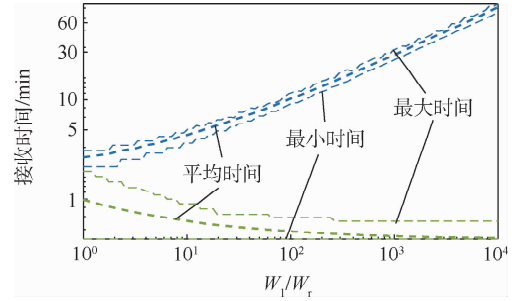


图 9 权重比影响分析

Fig. 9 Impact analysis of weight ratio

表 5 权重比对接收时间影响对比结果

Table 5 Comparison results of weight ratio on reception time

权重比 W_1/W_r	系统公钥 平均时间/s	系统公钥 最大时间/s	系统公钥 最小时间/s	OTAR 电文 平均时间/s
10^0	58	123	24	187
10^1	36	57	24	317
10^2	28	48	24	725
10^3	25	45	24	2 025
10^4	24.5	45	24	5 896

3.2.2 播发调度算法分析

基于 3 种调度算法分别进行仿真试验,试验条件为解调错误率 $PER = 0$, $W_1/W_r = 10^2$,通过 10 000 次播发试验,将每次接收机接收的时间汇总绘制成直方图,结果如图 10 ~ 图 12 所示。图中纵坐标表示在 10 000 次模拟中各种结果占的比例。

图 10 表示在 PFQ 算法下系统公钥的接收时间结果。接收系统公钥平均时间 28 s,最大时间 39 s,最小时间 24 s,10 000 次模拟中最小时间占比最大,为 0.49;接收全部 OTAR 电文平均时间 822 s,最小时间 657 s,最大时间 1 803 s,时间分布区间较小。

图 11 表示在 PFQ-Semi-Rigid 算法下系统公钥的接收时间结果。接收公钥平均时间 31 s,最大时间 90 s,最小时间 24 s,10 000 次模拟中最小时间占比最大,为 0.76;接收全部 OTAR 电文平均时间 745 s,最小时间 195 s,最大时间 2 076 s,时间分布区间较大。

图 12 表示在 PFQ-Split 算法下系统公钥的接收时间结果。接收公钥平均时间 28 s,最大时间 36 s,最小时间 24 s,10 000 次模拟中接近平均时间占比最高,为 0.45;接收全部 OTAR 电文平均时间 863 s,最小时间 723 s,最大时间 1 851 s,时间分布区间较小。

表 6 为上述仿真结果的汇总。3 种调度算法接收系统公钥平均时间相差不大,最小时间一致,其中 PFQ-Semi-Rigid 算法的系统公钥接收时间大都集中于最小时间处,方差较大,分布极不均匀;PFQ 算法系统公钥接收时间同样偏向最小时间,方差较小;PFQ-Split 算法系统公钥接收时间均匀,方差最小,但整体时间偏大。

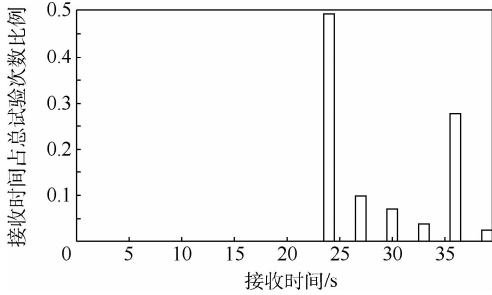


图 10 PFQ 系统公钥接收时间直方图

Fig. 10 Histogram of system public key reception time of PFQ

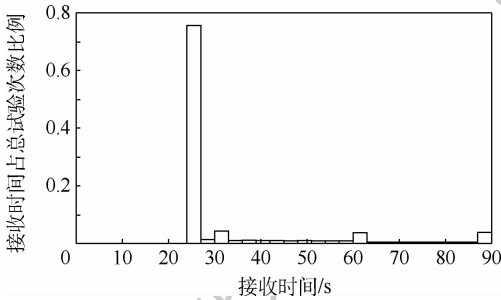


图 11 PFQ-Semi-Rigid 系统公钥接收时间直方图

Fig. 11 Histogram of system public key reception time of PFQ-Semi-Rigid

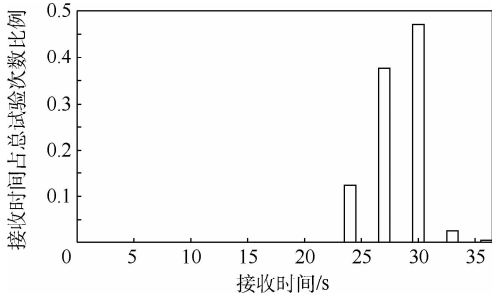


图 12 PFQ-Split 系统公钥接收时间直方图

Fig. 12 Histogram of system public key reception time of PFQ-Split

表 6 调度算法对接收时间影响对比结果

Table 6 Comparison results of scheduling algorithms on reception time

调度算法	系统公钥	系统公钥	系统公钥	OTAR 电文
	平均时间/s	最大时间/s	最小时间/s	平均时间/s
PFQ	28	39	24	822
PFQ-Semi-Rigid	31	90	24	745
PFQ-Split	28	36	24	863

3.2.3 解调错误率分析

上述仿真试验均在解调错误率 $PER = 0$ 的完美条件下进行,但实际情况中一定有解调错误的事件发生,本节为接收机设置一定概率的解调错误,错误率范围从 $PER = 10^{-5}$ 至 $PER = 10^{-1}$,每种情况分别进行 10 000 次播发模拟,权重比 $W_1/W_r = 10^2$,并将接收机每次接收时间汇总为图 13。图中下方曲线范围表示接收 OMT1 和 OMT2 所需要的时间范围。

表 7 详细展示了图 13 中 5 个关键点位的详细时间结果。解调错误率从 $PER = 10^{-5}$ 到 $PER = 10^{-3}$ 增长过程中,接收当前系统公钥和全部 OTAR 电文所需平均时间基本未变,接收当前系统公钥平均时间 29 s,最小时间 24 s,最大时间有所变化,从 54 s 增至 75 s,从 $PER = 10^{-2}$ 至 $PER = 10^{-1}$ 增长过程中,接收时间大幅增长,接收当前系统公钥平均时间从 30 s 增至 41 s,最大时间从 111 s 增至 177 s,最小时间未变;接收全部 OTAR 电文平均时间从 815 s 增至 1 317 s。

结果表明,解调错误率主要对最大接收时间造成影响,在 $PER = 10^{-2}$ 之前,解调错误率对接收平均时间几乎没有影响,在 $PER > 10^{-2}$ 后,接收平均时间开始随解调错误率的增大而变长。

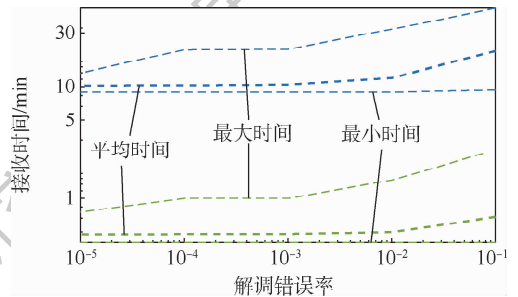


图 13 解调错误率影响分析

Fig. 13 Impact analysis of demodulation error rate

表 7 解调错误率对接收时间影响对比结果

Table 7 Comparison results of demodulation error rate on reception time

解调错误率 PER	系统公钥 平均时间/s	系统公钥 最大时间/s	系统公钥 最小时间/s	OTAR 电文 平均时间/s
10^{-1}	41	177	24	1 317
10^{-2}	30	111	24	815
10^{-3}	29	75	24	722
10^{-4}	29	66	24	714
10^{-5}	29	54	24	712

3.3 与现有方案对比

美国斯坦福大学也进行了有关 OTAR 电文接收时间的研究^[16-17]。斯坦福大学采用的 ECDSA 标准为 P-224 标准,安全等级 112 位,签名长度

448 bits, 数字证书采用 P-384 标准, 安全等级 192 位, 签名长度 768 bits。本文方案系统签名和数字证书签名采用同一标准, 安全等级相同, 斯坦福大学的方案系统签名和数字证书签名选择了不同的 SM2 模型, 安全等级递增, 安全性相对更高, 但带来的问题是 OTAR 需要承载的电文内容更多, 接收时间相对更长, 对接收机端的存储压力也更大。表 8 展示了斯坦福大学采用美国 ECDSA 标准的认证服务方案和本文采用 SM2 标准的认证服务方案的条件下, 分别使用 3 种调度算法的仿真结果对比, 其中解调错误率 $PER = 0$, 权重比 $W_1/W_r = 10^2$ 。

斯坦福大学设计的方案采用多级数字签名, 且安全等级更高, 然而安全性的提高使得 OTAR 位需要播发的电文更多, 本文采用的方案仿真的结果相较于斯坦福大学的结果, 平均接收时间相对更短。

表 8 国内外方案结果对比

Table 8 Comparison of results of domestic and foreign programs

调度算法	系统公钥平均时间/s		OTAR 电文平均时间/s	
	本文	斯坦福	本文	斯坦福
PFQ	28	63.6	822	1 386
PFQ-Semi-Rigid	31	50.3	745	2 394
PFQ-Split	28	62.6	863	1 464

4 结 论

1) ECDSA 算法操作简单, 且在国际和国内都具有通用的密码标准, 国内为商用密码标准 SM2 算法, 安全性和通用型都有所保障, 数字签名为非对称认证, 不需要维持时间同步。由于数字签名长度较长, SBAS-L5 频段中 I 支路用于播发增强信息, 剩余数据位较少, Q 支路没有特别的需求, 本文选择在 L5Q 支路播发认证电文。

2) 本文仿真试验主要针对用户接收机在冷启动时接收 OTAR 电文的时间展开, 分析得出对接收时间产生影响的因素有 3 个: 权重比、播发策略和解调错误率。重要 OTAR 电文和次要 OTAR 电文权重比越高, 接收重要电文所需时间越短, 但总体时间越长; 3 种调度算法各有优劣, 可根据具体应用环境进行选择; 当解调错误率 $PER > 10^{-2}$ 时对接收时间影响开始增大。

3) SBAS 信号认证是导航领域的热点问题, 欧美双方研究起步较早, 本文提出一种面向 BDS-BAS 的电文认证方案, 为 BDSBAS 认证服务的搭建提供一定的理论支持。

参考文献 (References)

- [1] CHIARA A D, BROI G D, POZZOBON O, et al. Authentication concepts for satellite-based augmentation systems [C] // Proceedings of the 29th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GNSS + 2016), 2016: 3208-3221.
- [2] CHIARA A D, BROI G D, POZZOBON O, et al. SBAS authentication proposals and performance assessment [C] // Proceedings of the 30th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GNSS + 2017), 2017: 2106-2116.
- [3] NEISH A, WALTER T, ENGE P. Parameter selection for the TESLA keychain [C] // Proceedings of the 31st International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GNSS + 2018), 2018: 2155-2171.
- [4] NEISH A, WALTER T, ENGE P. Quantum-resistant authentication algorithms for satellite-based augmentation systems [J]. Navigation, 2019, 66(1): 199-209.
- [5] WU Z J, LIU R S, CAO H J. ECDSA-based message authentication scheme for BeiDou-II navigation satellite system [J]. IEEE Transactions on Aerospace and Electronic Systems, 2019, 55(4): 1666-1682.
- [6] NEISH A, WALTER T, POWELL J D. Design and analysis of a public key infrastructure for SBAS data authentication [J]. Navigation, 2019, 66(4): 831-844.
- [7] WU Z J, ZHANG Y, LIU R S. BD-II NMA&SSI: An scheme of anti-spoofing and open BeiDou II D2 navigation message authentication [J]. IEEE Access, 2020, 8: 23759-23775.
- [8] NEISH A, WALTER T, POWELL D, et al. SBAS data authentication: A concept of operations [C] // Proceedings of the 32nd International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GNSS + 2019), 2019: 1812-1823.
- [9] 中华人民共和国国家质量监督检验检疫总局, 中国国家标准化管理委员会. 信息安全技术 SM2 椭圆曲线公钥密码算法 第 1 部分: 总则: GB/T 32918.1—2016 [S]. 北京: 中国标准出版社, 2017.
General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China, Standardization Administration of the People's Republic of China. Information security technology—Public key cryptographic algorithm SM2 based on elliptic curves—Part 1: General: GB/T 32918.1—2016 [S]. Beijing: Standards Press of China, 2017 (in Chinese).
- [10] 中华人民共和国国家质量监督检验检疫总局, 中国国家标准化管理委员会. 信息安全技术 SM2 椭圆曲线公钥密码算法 第 2 部分: 数字签名算法: GB/T 32918.2—2016 [S]. 北京: 中国标准出版社, 2017.
General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China, Standardization Administration of the People's Republic of China. Information security technology—Public key cryptographic algorithm SM2 based on elliptic curves—Part 2: Digital signature algorithm: GB/T 32918.2—2016 [S]. Beijing: Standards Press of China, 2017

- (in Chinese).
- [11] 孙荣燕,蔡昌曙,周洲,等. 国密 SM2 数字签名算法与 ECDSA 算法对比分析研究[J]. 网络安全技术与应用, 2013(2):60-62.
- SUN R Y, CAI C S, ZHOU Z, et al. The comparison between digital signature based on SM2 and ECDSA[J]. Network Security Technology & Application, 2013(2):60-62 (in Chinese).
- [12] 黄双临,辛洁,王冬霞,等. 星基增强系统电文及播发特性研究[J]. 数字通信世界, 2019(2):4-6.
- HUANG S L, XIN J, WANG D X, et al. Research on propagating message and strategy of satellite-based augmentation system [J]. Digital Communication World, 2019(2):4-6 (in Chinese).
- [13] 申成良,郭承军. 卫星导航信号电文加密技术研究与评估[J]. 全球定位系统, 2018, 43(3):7-12.
- SHEN C L, GUO C J. Study and evaluation of GNSS signal cryptographic authentication defenses[J]. GNSS World of China, 2018, 43(3):7-12 (in Chinese).
- [14] 梁曦,陶晓霞,周昀,等. 星基增强系统导航电文及完好性信息研究[J]. 空间电子技术, 2016, 13(5):39-42.
- LIANG X, TAO X X, ZHOU Y, et al. Research of SBAS navigation message and integrity message[J]. Space Electronic Technology, 2016, 13(5):39-42 (in Chinese).
- [15] 汪朝晖,张振峰. SM2 椭圆曲线公钥密码算法综述[J]. 信息安全研究, 2016, 2(11):972-982.
- WANG Z H, ZHANG Z F. Overview on public key cryptographic algorithm SM2 based on elliptic curves[J]. Journal of Information Security Research, 2016, 2(11):972-982 (in Chinese).
- [16] FERNÁNDEZ-HERNÁNDEZ I, CHÂTRE E, DALLA CHIARA A, et al. Impact analysis of SBAS authentication[J]. Navigation, 2018, 65(4):517-532.
- [17] BONELLI N, GIORDANO S, PROCISSI G. Network traffic processing with PFQ[J]. IEEE Journal on Selected Areas in Communications, 2016, 34(6):1819-1833.

Design of message authentication and OTAR broadcast strategy for BDSBAS

MU Shenglin^{1,2}, CHEN Ying³, LIU Ting¹, LIU Cheng³, CHEN Xiao^{1,2,*}

(1. School of Optoelectronics, University of Chinese Academy of Sciences, Beijing 100049, China;

2. Airspace Information Innovation Institute, Chinese Academy of Sciences (CAS), Beijing 100094, China;

3. Beijing Institute of Tracking and Communication Technology, Beijing 100094, China)

Abstract: Satellite-based Augmentation System (SBAS) relates to the field of life safety through a wide-area augmentation system that provides integrity information to users. To improve SBAS's ability to resist spoofing, firstly, based on the Elliptic Curve Digital Signature Algorithm (ECDSA), this paper proposes a message authentication design scheme for BeiDou Satellite-based Augmentation System (BDSBAS). According to the domestic commercial cryptographic standard elliptic curve (SM2) algorithm, the corresponding parameter design and message distribution scheme are carried out, and the message distribution strategy of Over the Air Rekeying (OTAR) is described. In order to further verify the test scheme, this paper uses Monte Carlo OTAR simulator to carry out simulation tests based on three broadcast scheduling algorithms. By adjusting the weights, the receiving time of OTAR messages in different situations is analyzed. The simulation results provide an important reference for the selection of broadcast scheduling of OTAR message schemes in specific situations.

Keywords: BeiDou Satellite-based Augmentation System (BDSBAS); message authentication; Elliptic Curve Digital Signature Algorithm (ECDSA); Over the Air Rekeying (OTAR); broadcast strategy

http://bhxb.buaa.edu.cn jbuua@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2020.0218

一种新的最速下降算法在自适应噪声对消中的应用



王帅, 向建军*, 彭芳, 肖冰松

(空军工程大学 航空工程学院, 西安 710038)

摘 要: 自适应噪声对消在很多领域有着重要应用。为了进行自适应噪声对消, 提出了一种新的最速下降算法。该算法主要原理是对多元二次函数进行降维处理, 使其变成一元二次函数, 再应用抛物线的性质分别循环迭代地求解每一个维度上的极值。在自适应噪声对消应用中, 所提算法与传统的自适应算法进行对比, 具有收敛速度快, 滤波效果好, 滤波效果可调节, 抗恶劣信噪比以及急剧变化信噪比, 不需选择迭代步长, 适合计算机和可编程硬件实现等优点。

关键词: 自适应算法; 最速下降算法; 噪声对消; 最小均方误差; 自适应滤波器

中图分类号: TN911.7

文献标志码: A

文章编号: 1001-5965(2021)07-1462-08

在飞机机舱内部、医学信号检测以及一些特殊工业环境中, 受噪声的影响语音信号的质量会急剧下降。为了滤除信号中的噪声提高语音信号的质量, 自适应噪声对消技术得到了广泛的应用。

自适应噪声对消技术本质是一种将噪声作为期望信号的自适应滤波器。对噪声消除能力的好坏主要由滤波器中自适应算法的性能决定, 常见的自适应算法有最小均方误差(Least Mean Square, LMS)算法、递归最小二乘(Recursive Least-Square, RLS)算法等。虽然传统的自适应算法在自适应对消领域展现出了很好的效果, 但却都存在着一定的不足, 对此本文针对传统自适应算法所存在的缺点提出基于一种新的最速下降算法的自适应算法。

1 传统的自适应算法

自适应滤波算法最早可以追溯到1960年, 由Widrow和Hoffman最先提出了LMS算法。自适应滤波可以根据信号自动调整滤波器的独特优势

被广泛应用于系统辨识、自适应均衡、噪声对消和波束形成等领域^[1-3]。

自适应滤波的原理是通过引入反馈和期望信号, 根据自适应算法来自动地调节滤波器的权系数。自适应滤波器的核心就是自适应算法, 经典的自适应算法有LMS算法和RLS算法^[4-6]。

LMS算法是自适应算法中最基础的算法, 具有运算量小的优点, 但缺点是收敛速度慢, 而且需要考虑迭代步长的选择问题, 迭代步长选择的过大则会造成无法收敛, 选择过小则会减小滤波器的收敛速度^[7-9]。LMS算法的迭代方程为

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu \mathbf{x}^T e(n) \quad (1)$$

式中: μ 为迭代步长; \mathbf{x} 为输入信号; $e(n)$ 为期望信号与输出信号之差所得到的误差信号; $\mathbf{w}(n)$ 为当前时刻的滤波器权系数; $\mathbf{w}(n+1)$ 为下一时刻的滤波器权系数。

经典的LMS算法性能存在着很大的局限性, 如收敛速度慢、步长固定等。对此文献[10]提出

收稿日期: 2020-05-27; 录用日期: 2020-09-04; 网络出版时间: 2020-10-10 13:15

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20201010.1121.001.html

基金项目: 航空科学基金(20175596020)

*通信作者: E-mail: xiang_jj@sohu.com

引用格式: 王帅, 向建军, 彭芳, 等. 一种新的最速下降算法在自适应噪声对消中的应用[J]. 北京航空航天大学学报, 2021, 47(7): 1462-1469. WANG S, XIANG J J, PENG F, et al. Application of a new steepest descent algorithm in adaptive noise cancellation [J]. Journal of Beijing University of Aeronautics and Astronautics, 2021, 47(7): 1462-1469 (in Chinese).

了一种变步长的 LMS 改进算法。该算法的步长基于对数函数进行变化,改善了传统 LMS 算法步长固定的缺点,使 LMS 算法的收敛速度加快,滤波性能得到了很大的改进。

改进后的 LMS 算法,迭代步长与当前时刻的误差以及上一时刻误差的乘积成对数关系,步长的更新公式为

$$\mu(n) = \text{blg}(a | e(n)e(n-1) |) \quad (2)$$

式中:参数 a 和参数 b 控制曲线的形状和整体数值的大小。将式(2)代入式(1)得到改进 LMS 算法迭代公式如下:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \text{blg}(a | e(n)e(n-1) |) \mathbf{x}^T e(n) \quad (3)$$

RLS 算法是利用上一时刻的滤波器权系数,根据代价函数自动调节得到当前时刻的滤波器权系数。这种算法克服了 LMS 收敛速度慢的缺点,但是缺点是涉及矩阵的运算,致使程序编写比较复杂而且计算量较大^[11-14]。RLS 算法的迭代方程为

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mathbf{k}(n)e(n) \quad (4)$$

式中: $\mathbf{k}(n)$ 为增益向量,表达式为

$$\mathbf{k}(n) = \frac{\mathbf{R}^{-1}(n-1)\mathbf{x}(n)}{\lambda + \mathbf{x}^T \mathbf{R}^{-1}(n-1)\mathbf{x}(n)} \quad (5)$$

式中: λ 为遗忘因子; $\mathbf{R}(n) = \sum_{i=0}^n \lambda^{n-i} \mathbf{x}(i)\mathbf{x}^T$ 。

上述 2 种经典的自适应算法一经提出便得到了广泛的应用,针对其存在的缺点,后期也有很多改进的方法,但是基本的求解极值的原理没有太大变化。本文提出了一种新的求极值的最速下降算法,并将其应用到自适应滤波上。该算法主要原理是对多元二次函数进行降维处理,使其变成一元二次函数,再应用抛物线的性质分别循环迭代地求解每一个维度上的极值,使所求得的值渐渐逼近函数的极值。

抛物线的一般式为 $f(z) = b_0 z^2 + b_1 z + b_2$; 抛物线是轴对称图形;对称轴为直线 $z = -b_1/2b_0$; 对称轴与抛物线唯一的交点为抛物线的极值点。

2 改进算法原理

2.1 数学推导

本文算法主要利用二次函数越接近极值函数曲面越集中,来加快求极值过程中的收敛速度。为了简单解释该算法,可以应用二元二次函数来进行解释。假设二元二次函数方程为 $u = v^2 + w^2$, 是一个抛物面凸函数,任取一个 v_0 作为 v 的初始值,对抛物面进行平行于 uow 平面的纵切得到一条

以 w 为自变量、 u 为因变量的抛物线,求解这条抛物线的极值 w_0 ,再以 w_0 作为 w 的初始值,对抛物面进行平行于 uov 平面的纵切将会得到第二条抛物线,可以想象第二条抛物线的极值与第一条抛物线相比更接近于抛物面的极值点,依照这种方式循环迭代地求解 v 和 w 纵切面上的极值,得到抛物线的极值将越来越逼近于抛物面凸函数的极值点。

本文算法详细的推导过程如下:将多元二次函数降维到二维平面上;多元二次函数降维到二维平面上的形状是一条抛物线;应用抛物线的性质根据其他系数求出抛物线的极值。

1) 假设多元二次函数的一般形式为 $f_1^2 = (a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n + c_1)^2$,将多元二次函数的一个自变量 x_1 当作未知数,其他自变量 x_2, x_3, \dots, x_n 代入初始值当作系数,此时多元二次函数变成了一个以 x_1 为自变量、 f_1^2 为因变量的二次抛物线,根据抛物线的性质可以很容易地求出此抛物线的极值。

2) 以第 1) 步求得的极值作为 x_1 的初始值,将 x_2 作为自变量,其他自变量代入初始值作为系数,求解以 x_2 为自变量、 f_1^2 为因变量的二次抛物线,用抛物线的性质求此抛物线的极值。

3) 将第 2) 步求得的极值作为 x_2 的初始值,应用同样的方法依次求解其他自变量的取值。

4) 当求完 x_n 的取值之后,再返回来从 x_1 开始求解,按照这种方式循环迭代的运算。最终求得的价值将会接近多元二次函数极值处的各个自变量的取值。

这种二次函数求极值的方法需要约束方程的个数大于等于自变量的个数,即多元二次函数存在极值点。具体解释如下:

$$F(x_1, x_2, \dots, x_n) = f_1^2 + f_2^2 + \dots + f_m^2 \quad (6)$$

式(6)中:

$$\begin{cases} f_1 = a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n + c_1 \\ f_2 = a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n + c_2 \\ \vdots \\ f_m = a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n + c_m \end{cases} \quad (7)$$

式中: m 为约束方程的个数; n 为自变量的个数。若应用这种方法求二次函数极值,就要求 $n \leq m$, 这里要求 f_1, f_2, \dots, f_m 是线性无关的,若是存在线性相关组则要求 f_1, f_2, \dots, f_m 的最大线性无关组的个数大于等于自变量的个数,即二次函数存在极值点。

一元二次函数一般形式可以设为 $f(z) = c_0 z^2 + c_1 z + c_2$,根据抛物线的性质可以求得抛物线的极

值点位于 $z = -c_1/(2c_0)$ 处,为了求解函数中的系数 c_1, c_0 ,可以采用对函数求导数的方法,即 $c_1 = y'_z|_{z=0}, 2c_0 = y''_{zz}$,因此多元二次函数的极值点位于 $x = -f'_x|_{x=0}/f''_{xx}$ 处。

式(6)所示的是 n 元二次函数,应用这种方

法求解式(6)的极值,步骤如下:

首先,给 n 个自变量随机设定初始值 $x_1^{(0)}, x_2^{(0)}, \dots, x_n^{(0)}$;应用此种方法多次循环迭代使 n 个自变量的取值向着极值点收敛,迭代方程如下:

$$\left\{ \begin{aligned} x_1^{(k+1)} &= \frac{-F'_{x_1}(x_1=0, x_2=x_2^{(k)}, \dots, x_n=x_n^{(k)})}{F''_{x_1x_1}} = -\frac{2a_{11}f_1 + 2a_{21}f_2 + \dots + 2a_{m1}f_m |_{x_1=0, x_2=x_2^{(k)}, \dots, x_n=x_n^{(k)}}}{2a_{11}^2 + 2a_{21}^2 + \dots + 2a_{m1}^2} = \\ &x_1^{(k)} - \frac{a_{11}f_1 + a_{21}f_2 + \dots + a_{m1}f_m |_{x_1=x_1^{(k)}, x_2=x_2^{(k)}, \dots, x_n=x_n^{(k)}}}{a_{11}^2 + a_{21}^2 + \dots + a_{m1}^2} \\ x_2^{(k+1)} &= \frac{-F'_{x_2}(x_1=x_1^{(k+1)}, x_2=0, \dots, x_n=x_n^{(k)})}{F''_{x_2x_2}} = -\frac{2a_{12}f_1 + 2a_{22}f_2 + \dots + 2a_{m2}f_m |_{x_1=x_1^{(k+1)}, x_2=0, \dots, x_n=x_n^{(k)}}}{2a_{12}^2 + 2a_{22}^2 + \dots + 2a_{m2}^2} = \\ &x_2^{(k)} - \frac{a_{12}f_1 + a_{22}f_2 + \dots + a_{m2}f_m |_{x_1=x_1^{(k+1)}, x_2=x_2^{(k)}, \dots, x_n=x_n^{(k)}}}{a_{12}^2 + a_{22}^2 + \dots + a_{m2}^2} \\ &\vdots \\ x_n^{(k+1)} &= \frac{-F'_{x_n}(x_1=x_1^{(k+1)}, x_2=x_2^{(k+1)}, \dots, x_n=0)}{F''_{x_nx_n}} = -\frac{2a_{1n}f_1 + 2a_{2n}f_2 + \dots + 2a_{mn}f_m |_{x_1=x_1^{(k+1)}, x_2=x_2^{(k+1)}, \dots, x_n=0}}{2a_{1n}^2 + 2a_{2n}^2 + \dots + 2a_{mn}^2} = \\ &x_n^{(k)} - \frac{a_{1n}f_1 + a_{2n}f_2 + \dots + a_{mn}f_m |_{x_1=x_1^{(k+1)}, x_2=x_2^{(k+1)}, \dots, x_n=x_n^{(k)}}}{a_{1n}^2 + a_{2n}^2 + \dots + a_{mn}^2} \end{aligned} \right. \quad (8)$$

2.2 基于新的最速下降算法的自适应滤波器原理

自适应滤波器一般遵从的是最小均方误差准则,使经过滤波器的输出信号和期望信号之间的均方误差最小^[15-18],自适应滤波器的示意图如图1所示。

图1中 $x(n)$ 是输入信号, $y(n)$ 是经滤波器进行滤波之后的输出信号, $d(n)$ 是期望信号, $e(n)$ 是输出信号和期望信号做差所得到的误差信号,通过误差信号模的大小,根据相应的自适应滤波算法自动地调整滤波器的权值,最终使误差信号的模达到最小。自适应噪声对消就是将有用信号和噪声的混合信号作为输入信号,将噪声信号作为期望信号,当误差信号最小时最接近有用信号^[19-20]。

自适应滤波器所用到的是有限长 FIR 滤波器,因此信号处理的方式是输入信号与滤波器权

值做卷积。假设 N 阶滤波器的权值分别为 h_1, h_2, \dots, h_N ,则输出信号为

$$y(n) = h_1x(n) + h_2x(n-1) + \dots + h_Nx(n-N+1) \quad (9)$$

$$e(n) = y(n) - d(n) = h_1x(n) + h_2x(n-1) + \dots + h_Nx(n-N+1) - d(n) \quad (10)$$

误差信号的模最小时,误差信号的平方也最小,因此可以计算误差信号平方的最小值。

$$e^2(n) = [h_1x(n) + h_2x(n-1) + \dots + h_Nx(n-N+1) - d(n)]^2 \quad (11)$$

为了使用本文所提出的方法,这里求解均方误差的最小值,均方误差如下:

$$\left\{ \begin{aligned} F(h_1, h_2, \dots, h_N) &= \frac{1}{M} [e^2(n) + e^2(n-1) + \dots + e^2(n-M+1)] \\ e(n) &= h_1x(n) + h_2x(n-1) + \dots + h_Nx(n-N+1) - d(n) \\ e(n-1) &= h_1x(n-1) + h_2x(n-2) + \dots + h_Nx(n-N) - d(n-1) \\ &\vdots \\ e(n-M+1) &= h_1x(n-M+1) + h_2x(n-M) + \dots + h_Nx(n-N-M+2) - d(n-M+1) \end{aligned} \right. \quad (12)$$

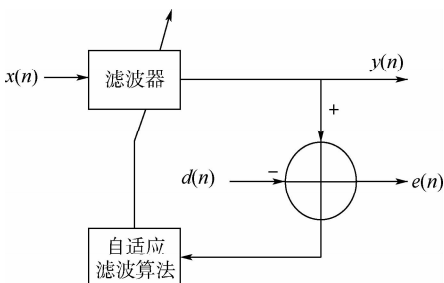


图1 自适应滤波器示意图

Fig. 1 Schematic diagram of adaptive filter

用式(8)求解式(12)均方误差的最小值得到自适应滤波算法的迭代方程如下:

$$\left\{ \begin{aligned} h_1^{(k+1)} &= \frac{-F'_{h_1}(h_1 = 0, h_2 = h_2^{(k)}, \dots, h_n = h_n^{(k)})}{F''_{h_1 h_1}} = \\ & \frac{h_1^{(k)} - [x(n)e(n) + x(n-1)e(n-1) + \dots + x(n-M+1)e(n-M+1)]_{h_1=h_1^{(k)}, h_2=h_2^{(k)}, \dots, h_N=h_N^{(k)}}}{[x^2(n) + x^2(n-1) + \dots + x^2(n-M+1)]} \\ h_2^{(k+1)} &= \frac{-F'_{h_2}(h_1 = h_1^{(k+1)}, h_2 = 0, \dots, h_n = h_n^{(k)})}{F''_{h_2 h_2}} = \\ & \frac{h_2^{(k)} - [x(n-1)e(n) + x(n-2)e(n-1) + \dots + x(n-M)e(n-M+1)]_{h_1=h_1^{(k+1)}, h_2=h_2^{(k)}, \dots, h_N=h_N^{(k)}}}{[x^2(n-1) + x^2(n-2) + \dots + x^2(n-M)]} \\ & \vdots \\ h_N^{(k+1)} &= \frac{-F'_{h_N}(h_1 = h_1^{(k+1)}, h_2 = h_2^{(k+1)}, \dots, h_N = 0)}{F''_{h_N h_N}} = \\ & \frac{h_N^{(k)} - [x(n-N+1)e(n) + x(n-N)e(n-1) + \dots + x(n-M)e(n-N-M+2)]_{h_1=h_1^{(k+1)}, h_2=h_2^{(k+1)}, \dots, h_N=h_N^{(k)}}}{[x^2(n-N+1) + x^2(n-N) + \dots + x^2(n-N-M+2)]} \end{aligned} \right. \quad (13)$$

式中:要求约束方程的个数 M 大于等于滤波器的阶数 N 。

根据以上对基于新的最速下降算法的自适应滤波器原理的推导,将这种新的自适应算法的应用步骤总结如下:

- 1) 设置滤波器的阶数 N 和约束方程的个数 M , 注意在设置这 2 个常数时应该满足 $N \leq M$ 的条件。
- 2) 对滤波器的权值 h_1, h_2, \dots, h_N 进行初始化赋值分别为 $h_1^{(0)}, h_2^{(0)}, \dots, h_N^{(0)}$ 。
- 3) 将输入信号和误差信号代入式(13)中, 按照从上到下的顺序求解新的滤波器权值 $h_1^{(1)}, h_2^{(1)}, \dots, h_N^{(1)}$ 。
- 4) 输入下一时刻的输入信号 $x(n+1)$, 根据新的输入信号 $x(n+1)$ 和新的滤波器权值 $h_1^{(1)}, h_2^{(1)}, \dots, h_N^{(1)}$ 对误差信号 $e(n)$ 进行更新。
- 5) 重新调用式(13)求解出新的滤波器权值 $h_1^{(2)}, h_2^{(2)}, \dots, h_N^{(2)}$, 重复第 3) 步和第 4) 步使自适应滤波器的权值向着使误差信号模最小的极值点收敛。
- 6) 当误差信号达到允许的范围时停止循环。

3 仿真分析

使用 MATLAB 软件对本文算法在自适应噪

声对消中的应用进行仿真分析。仿真的理想信号是正弦波和声音信号,噪声是服从高斯分布的白噪声信号,滤波器的输入信号是白噪声信号,期望信号是混入噪声的理想信号。仿真环境的信噪比设置为 -10 dB,滤波器阶数为 5,约束方程个数为 150。

3.1 时域波形图仿真

图 2 为采用本文算法进行噪声对消得到的时域仿真波形图。图中显示的信号分别是理想信号加上信噪比为 -10 dB 的高斯白噪声之后的时域波形图,理想信号的时域波形图,混入噪声之后的理想信号经由本文算法进行自适应噪声对消处理后的时域波形图。

从仿真的时域波形图可以看出,高斯白噪声对信号的干扰是十分严重的,经由本文算法进行自适应噪声对消之后的波形图基本上接近理想信号,滤波效果很好。可以看到,经处理之后的时域波形图一开始有一小段失真,这是由于算法一开始处于收敛阶段造成的,当算法收敛到最小之后,滤波效果达到最佳,恢复出了和理想信号最接近的时域波形。

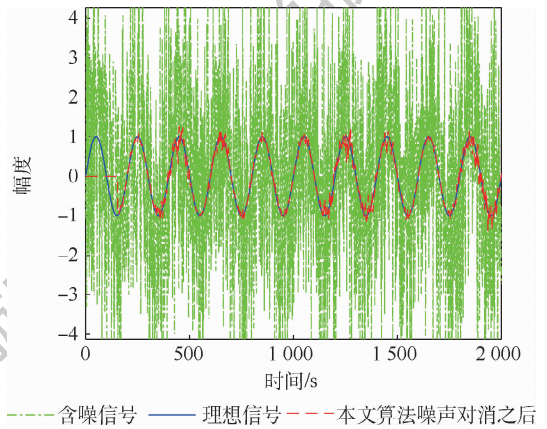


图 2 时域波形图

Fig. 2 Time domain waveform

3.2 收敛速度仿真分析

对本文算法、LMS 算法、文献[19]中的改进 LMS 算法以及 RLS 算法 4 种算法的收敛速度进行仿真分析,得到图 3。图 3 为用 MATLAB 软件在相同的仿真环境下进行蒙特卡罗仿真 500 次得到的数据。图中横轴代表滤波器的迭代次数,纵轴代表经滤波器输出的信号与理想信号之间的均方误差。

图 3 中:LMS 算法的迭代步长是经过多次测试之后选取的最优步长,即可以使算法收敛的最大步长 μ 为 0.001,改进 LMS 算法的参数 a 和参数 b 也是经过多次测试之后选取的最优参数 $a =$

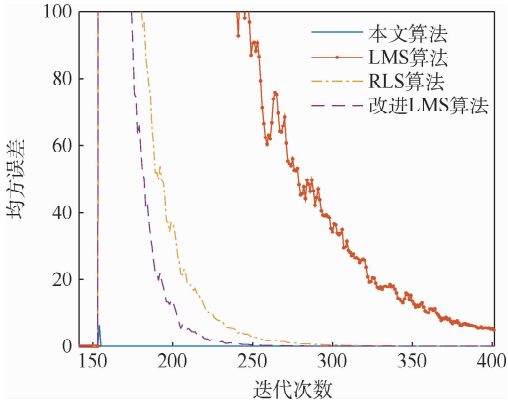


图3 4种算法收敛速度对比

Fig. 3 Comparison of convergence rates among four algorithms

1 200, $b = 0.001$ 。从图3中可以看出,由本文算法进行滤波之后的均方误差最先降到最小,改进LMS算法其次,RLS算法第3,LMS算法最慢。可见本文算法的收敛速度是十分快的,明显优于其他算法。

3.3 滤波效果仿真分析

对本文算法、LMS算法、RLS算法和改进LMS算法的滤波效果进行仿真分析得到图4。如图4所示,在误差信号趋于平稳之后RLS的最大均方误差在0.3左右,LMS算法的最大均方误差出现在0.2左右,改进LMS算法出现在0.5左右,本文算法的最大均方误差出现在0.1左右,可以很直观地看出本文算法的滤波效果优于LMS算法,LMS算法的滤波效果优于RLS算法,改进LMS算法的效果最差,这是由于改进LMS算法在高信噪比条件下效果较好,而仿真条件信噪比为 -10 dB,改进LMS算法的滤波效果较差。因此,本文算法的滤波效果要优于其他算法,但是这一结论也不绝对,因为新算法的性能与约束方程的数量有很大关系,将会在下面对本文算法性能进行定量分析。

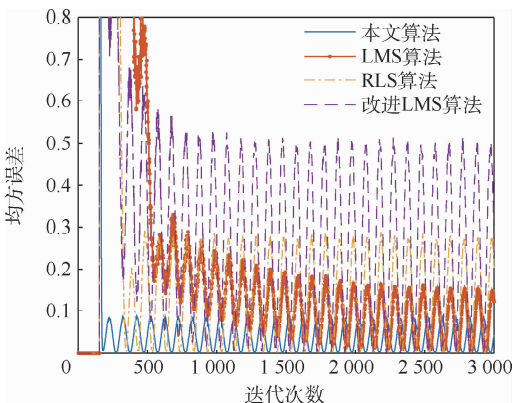


图4 4种算法滤波效果对比

Fig. 4 Comparison of filtering results among four algorithms

4 在实际信号噪声对消中的应用

引入一段语音信号作为输入信号来测试本文算法对实际信号的噪声对消性能。测试流程为:①使用MATLAB软件读取一段已经录制好的语音信号;②给语音信号加上功率可以随意调节的高斯白噪声;③使信号输入基于本文算法的自适应滤波器得到输出信号;④对输出信号进行时域、频域等分析来测试本文算法在实际信号噪声对消中的性能。

4.1 噪声对消后的时频输出

给语音信号加入信噪比为 -20 dB的噪声,此时语音信号已经完全淹没在噪声中无法分辨,采用本文算法进行自适应滤波之后,可以很好地恢复出原语音信号,里面仅掺杂着少量噪声。

实际信号处理的时域波形图如图5所示。图5中:绿色点划线代表加上噪声信号的语音信号,可以看到由于信噪比极低语音信号已经完全淹没于噪声信号中无法辨认,蓝色实线代表的是不含噪声的声音信号,红色虚线代表经新算法滤波之后输出的声音信号,可以看到输出信号与原始语音信号几乎重合。

图6为图5中3个信号的频谱。图中:绿色的波形是含噪信号的频谱,可以看出噪声频谱完全淹没了声音信号的频谱。蓝色的波形显示的是原始语音信号的频谱,红色波形显示的是经新算法滤波后输出信号的频谱,可以看出原始信号与输出信号的频谱基本重合。

对信号经过时域和频域分析之后,可以看到本文算法具有很强的滤波能力,即使在信噪比很低的情况下也能使输出信号的时域和频域与原始信号高度相似。

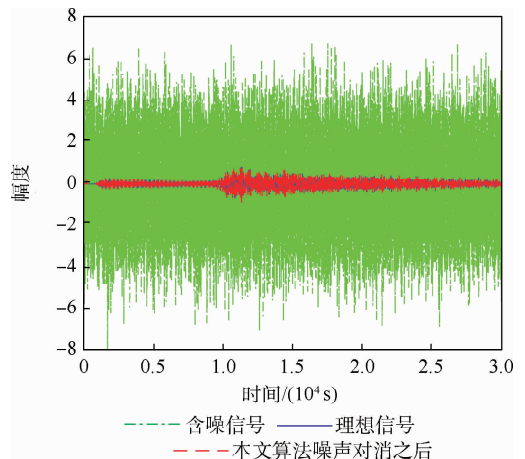


图5 实际信号处理的时域波形图

Fig. 5 Time domain waveform of actual signal processing

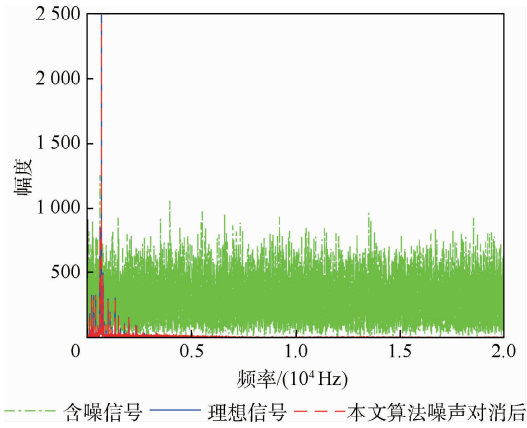


图6 实际信号处理的频域波形图

Fig.6 Frequency domain waveform of actual signal processing

4.2 算法性能的定量分析

在本文算法对实际信号进行处理的环境下,对本文算法的性能进行定量分析。主要分析约束方程的数量对算法性能的影响和输入信噪比的变化对算法性能的影响。

首先,对算法性能与约束方程的关系做定量分析。以输出信噪比作为滤波算法性能的衡量指标,如图7所示,当输入信噪比为 -20 dB,约束方程的个数 M 的取值范围为 $5 \sim 200$ 个,输出信噪比的值。

从图7可以看出,经过滤波器进行滤波之后,声音信号的信噪比整体都得到了改善。随着约束方程的个数 M 取值的不断增大,输出信噪比整体呈现出上升趋势,对声音信号信噪比的改善程度越来越好。可以得到结论:约束方程的个数越大,算法的滤波效果越好,去噪能力越强。在实际应用中,可以通过调节约束方程个数的方式来达到希望的滤波效果,这一特点克服了一般自适应滤波算法滤波性能固定不变的缺点。

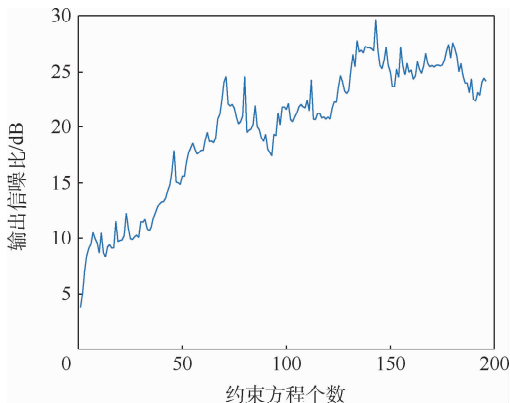


图7 约束方程个数对滤波效果的影响

Fig.7 Influence of the number of constraint equations on filtering result

然后,对输入信噪比与算法性能的关系做定量分析。设定约束方程的个数为 150 个,声音信号的输入信噪比变化范围为 $-50 \sim 10$ dB,观察输出信噪比的变化情况,得到输入信噪比与输出信噪比的变化如图8所示。

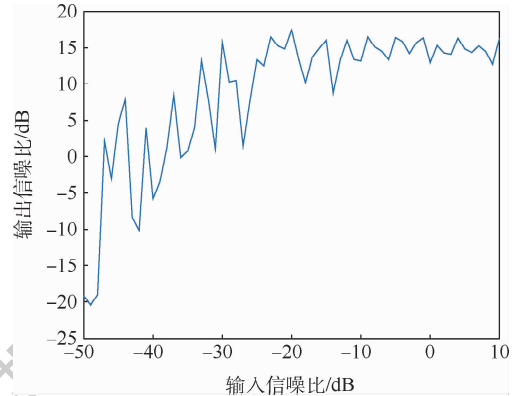


图8 信噪比对滤波效果影响

Fig.8 Influence of SNR on filtering result

如图8所示,输出信噪比整体都大于输入信噪比,说明滤波器在输入信噪比为 $-50 \sim 10$ dB的范围内对信号有很好的滤波作用,即使输入信噪比很恶劣或者输入信噪比急剧变化的情况下,本文算法都能展现出很好的效果。从图8的整体趋势可以看出,随着输入信噪比的增大,输出信噪比一开始也呈现出上升趋势,随着输入信噪比增大到一定值时,滤波器的滤波性能达到极限使输出信噪比不会有过大的提升。

5 结束语

通过对新的自适应滤波算法进行仿真以及实际应用,可以看到新的最速下降算法可以很好地完成自适应噪声对消的功能。与传统自适应算法相比,新的最速下降算法具有如下优点:收敛速度快;滤波效果好;滤波效果可以进行调节;在信噪比急剧变化以及信噪比十分恶劣的情况下,具有显著优势;不需要考虑迭代步长的选择问题;主要使用循环运算的方式,不涉及矩阵运算,更适合使用计算机或可编程硬件实现算法。

参考文献 (References)

- [1] 蔡卫菊. 自适应 RLS 的噪声对消系统 simulink 仿真[J]. 科技经济导刊, 2019, 27(36): 31.
CAI W J. Simulink simulation of adaptive RLS noise cancellation system [J]. Technology and Economic Guide, 2019, 27(36): 31 (in Chinese).
- [2] 邢怀玺, 吴华, 陈游. LMS 预处理的相位差机载单站无源定位方法[J]. 空军工程大学学报(自然科学版), 2019, 20(6): 9-14.

- XING H X, WU H, CHEN Y. A phase difference airborne single-station passive location method based on LMS preprocessing [J]. *Journal of Air Force Engineering University (Natural Science Edition)*, 2019, 20(6): 9-14 (in Chinese).
- [3] 景源, 汪洪涛. 一种新的变步长 LMS 算法及在自适应波束形成中的应用[J]. *辽宁大学学报(自然科学版)*, 2018, 45(4): 326-330.
- JING Y, WANG H T. A new variable step size LMS algorithm and its application in adaptive beamforming[J]. *Journal of Liaoning University (Natural Sciences Edition)*, 2018, 45(4): 326-330 (in Chinese).
- [4] GOEL N, KAUR H, SAXENA J. Modified decision based unsymmetric adaptive neighborhood trimmed mean filter for removal of very high density salt and pepper noise[J]. *Multimedia Tools and Applications*, 2020, 79(1): 1-30.
- [5] 李轲, 张国栋, 冀启东. 基于 FPGA 的自适应滤波技术在雷达信号处理中的应用研究[J]. *舰船电子工程*, 2018, 38(12): 83-87.
- LI K, ZHANG G D, JI Q D. Application of adaptive filtering in radar signal processing based on FPGA[J]. *Ship Electronic Engineering*, 2018, 38(12): 83-87 (in Chinese).
- [6] 霍陆陆. 基于自适应噪声相消的语音增强系统研究[D]. 西安: 西安电子科技大学, 2019: 1-20.
- HUO L L. Speech enhancement system based on adaptive noise cancellation[D]. Xi'an: Xidian University, 2019: 1-20 (in Chinese).
- [7] KWONG R H, JOHNSTON E W. A variable step size LMS algorithm[J]. *IEEE Transactions on Signal Processing*, 1992, 40(7): 1633-1642.
- [8] 靳翼, 邵怀宗. 一种新的变步长 LMS 自适应滤波算法及其仿真[J]. *信号处理*, 2010, 26(9): 1385-1388.
- JIN Y, SHAO H Z. A novel variable step size LMS adaptive filtering algorithm and its simulation[J]. *Signal Processing*, 2010, 26(9): 1385-1388 (in Chinese).
- [9] MENG X S, LIN Y, YAN L, et al. Airborne LiDAR point cloud filtering by a multilevel adaptive filter based on morphological reconstruction and thin plate spline interpolation[J]. *Electronics*, 2019, 8(10): 1153.
- [11] SHRIVASTAVA P C, KUMAR P, TIWARI M, et al. Efficient architecture for the realization of 2-D adaptive FIR filter using distributed arithmetic[J]. *Circuits, Systems, and Signal Processing*, 2021, 40(5): 1458-1478.
- [12] 白征东, 李帅, 陈波波, 等. 带约束的自适应抗差整体最小二乘滤波[J]. *测绘科学*, 2019, 44(1): 5-9.
- BAI Z D, LI S, CHEN B B, et al. Adaptive robust global least square filtering with constraints[J]. *Science of Surveying and Mapping*, 2019, 44(1): 5-9 (in Chinese).
- [13] 覃景繁, 欧阳景正. 一种新的变步长 LMS 自适应滤波算法[J]. *数据采集与处理*, 1997, 12(3): 171-174.
- QIN J F, OUYANG J Z. A new variable step size LMS adaptive filtering algorithm[J]. *Journal of Data Acquisition and Processing*, 1997, 12(3): 171-174 (in Chinese).
- [14] 高鹰, 谢胜利. 一种变步长 LMS 自适应滤波算法及分析[J]. *电子学报*, 2001, 29(8): 1094-1097.
- GAO Y, XIE S L. A variable step size LMS adaptive filtering algorithm and its analysis[J]. *Acta Electronica Sinica*, 2001, 29(8): 1094-1097 (in Chinese).
- [15] 兰瑞明, 唐普英. 一种新的变步长 LMS 自适应算法[J]. *系统工程与电子技术*, 2005, 27(7): 1307-1310.
- LAN R M, TANG P Y. New variable step-size LMS adaptive algorithm[J]. *Systems Engineering and Electronics*, 2005, 27(7): 1307-1310 (in Chinese).
- [16] CHENG X Z, XU J D, WEI A Y, et al. Application of RLS adaptive filtering in signal de-noising[J]. *Journal of Measurement Science and Instrumentation*, 2014, 5(1): 32-36.
- [17] 韩伟, 王大志, 刘震. 基于可变遗忘因子 RLS 算法的谐波电流检测方法[J]. *电工技术学报*, 2013, 28(12): 70-74.
- HAN W, WANG D Z, LIU Z. A harmonic current detection method based on variable forgetting factor RLS algorithm[J]. *Transactions of China Electrotechnical Society*, 2013, 28(12): 70-74 (in Chinese).
- [18] SABZI M, KAMAREI M, RAZBAN T, et al. New noise cancellation topology in common-gate LNAs[J]. *Microelectronics Journal*, 2020, 100: 104800.
- [19] HAN F L, ZHANG H B, RUI J W, et al. Multiple point geostatistical simulation with adaptive filter derived from neural network for sedimentary facies classification[J]. *Marine and Petroleum Geology*, 2020, 118: 104406.
- [20] 茹国宝, 黄燕, 郭英杰, 等. 基于对数函数的新变步长 LMS 算法[J]. *武汉大学学报(理学版)*, 2015, 61(3): 295-298.
- RU G B, HUANG Y, GUO Y J, et al. New variable step size LMS algorithm based on logarithmic function[J]. *Journal of Wuhan University (Natural Science Edition)*, 2015, 61(3): 295-298 (in Chinese).

Application of a new steepest descent algorithm in adaptive noise cancellation

WANG Shuai, XIANG Jianjun^{*}, PENG Fang, XIAO Bingsong

(College of Aeronautical Engineering, Air Force Engineering University, Xi'an 710038, China)

Abstract: Adaptive noise cancellation has important applications in many fields. In order to carry out adaptive noise cancellation, a new steepest descent algorithm is proposed. The main principle of the algorithm is to reduce the dimension of multivariate quadratic function to make it become a univariate quadratic function, and then apply the properties of parabola to iteratively solve the extremum of each dimension. Compared with the traditional adaptive algorithm, the new steepest descent algorithm shows the advantages of fast convergence, good filtering effect, adjustable filtering effect, resisting bad SNR and sharp change SNR, no need to choose iteration step, and suitable for computer and programmable hardware implementation.

Keywords: adaptive algorithm; steepest descent algorithm; noise cancellation; minimum mean square error; adaptive filter

http://bhxb.buaa.edu.cn jbuua@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2020.0225

改进的深度神经网络下遥感机场区域目标检测



韩永赛¹, 马时平^{1,*}, 何林远¹, 李承昊¹, 朱明明¹, 许悦雷²

(1. 空军工程大学 航空工程学院, 西安 710038; 2. 西北工业大学 无人系统技术研究院, 西安 710072)

摘 要: 卫星遥感监测器下的机场区域多类目标检测在实际生活中有着重大的军用和民用意义。为了有效提升机场区域遥感图片的检测精确率,以主流目标检测方法中更快的区域卷积神经网络(Faster R-CNN)为基础框架,针对数据侧提出了 ReMD 数据增强算法。同时使用更具深度的残差神经网络(ResNet)以及特征融合部件-特征金字塔网络(FPN)来提取机场区域目标更鲁棒的深层区分性特征。在末端检测网络添加新的全连接层并根据目标的类间关联性组合 softmax 分类器以及 4 个 logistic regression 分类器进行机场区域多类目标的精确分类。实验结果表明:相比原网络改进后的网络带来了 11.6% 的多类平均检测精确率的提升,达到了 80.5% 的 mAP,与其他主流网络进行对比也有更好的精确率;同时通过适当减小建议区域的输入量,可以在降低 3.2% 精确率的前提下将 0.512 s 的检测时间提速 3 倍,至 0.173 s,根据具体任务可以合理权衡精确率和检测速度,体现了该网络的有效性以及实用性。

关键词: 目标检测; 图像处理; 遥感; 机场区域; 神经网络

中图分类号: TP183; TP751.1

文献标志码: A **文章编号:** 1001-5965(2021)07-1470-11

目标检测目前为计算机视觉和数字图像处理的一个热门方向,其通过机器自主识别人们需要检测目标的方式,可以大大减少人力资源的消耗。而机场对于军用和民用领域都有重大的意义,对于民用来说是商业活动的中心同时也是城市的资源配置中心;对于军用来说,是空中力量的关键基础,也是夺取制空权的坚强力量。而随着卫星遥感图像的分辨率越来越高,其蕴含的信息量也越来越大,为人们从高分遥感图像中获取准确的有效信息提供了可能^[1],也为实现机器自动化地对机场区域实施大范围监控,以及精确搜索区域内目标,提供了关键的源保障。

针对机场区域的目标检测问题,国内外已经有了众多研究成果。其中针对机场检测,传统方法一般为先进行边缘检测然后结合相关算法提取

出机场跑道的直线特征来检测跑道区域,进而定位出机场,其中提取跑道边缘直线特征的算法包括 Hough 变换^[2]以及线段检测器(Line Segment Detector, LSD)等,这种方法检测速度快、计算量小但同时也存在一定的问题,比如形似物的干扰,如同样具有长直线特征的海岸线、河流、公路等。另一类方法是结合机场的显著性结构特征^[3]的检测,即根据机场的显著性特征提取图像中的感兴趣区域,通常使用的方法为图像分割^[4],该方法相比提取直线特征的方法在一定程度上提高了检测精确度,但是过多的滑动窗口冗余检测计算成本也是其一个缺点,使得检测效率很难得到保证,因此很难应用于实际检测任务中。众多学者基于以上 2 类方法提出了相应的改进并取得了较好的效果,但基于传统的方法依旧存在人工设

收稿日期: 2020-05-28; 录用日期: 2020-08-21; 网络出版时间: 2020-09-22 15:34

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20200922.1435.001.html

基金项目: 国家自然科学基金(61701524,61773397); 航空科学基金(20175896022)

* 通信作者. E-mail: 1013765061@qq.com

引用格式: 韩永赛, 马时平, 何林远, 等. 改进的深度神经网络下遥感机场区域目标检测[J]. 北京航空航天大学学报, 2021, 47(7): 1470-1480. HAN Y S, MA S P, HE L Y, et al. Regional object detection of remote sensing airport based on improved deep neural network [J]. Journal of Beijing University of Aeronautics and Astronautics, 2021, 47(7): 1470-1480 (in Chinese).

计特征的鲁棒性不足以及滑动窗口设计带来过多冗余计算使得检测效率较慢的问题。针对民航飞机的检测,传统的方法有基于特征学习以及分类学习的方法。Li等^[5]结合视觉显著性和对称性特征的方法对民航飞机进行了检测。林煜东等^[6]结合稀疏表示与飞机类目标几何形状得到飞机的几何部件再结合显著图进行民航飞机的检测。仇建斌等^[7]结合角点以及边缘特征的方法进行了检测。An等^[8]结合圆周频率滤波器进行候选区域定位并结合方向梯度直方图以及Adaboost算法对其进行检测。但基于几何特征学习的检测,较易受到边缘噪声以及形似物的干扰,基于分类学习的方法较为依赖特征设计。

随着深度学习的发展,近年来深度学习技术成为了各大领域的热门研究,基于学习的目标检测也愈发成熟,深度学习俨然已经成为了目标检测领域的“标配”。基于学习的机场区域检测方法也得到了越来越多的应用,并相对传统的检测方法在检测精确度以及效率方面有了很大的提升。Zhang等^[9]用卷积神经网络(Convolutional Neural Network, CNN)提取图片特征并结合机场线性分布来进行较为有效的区域建议来对机场进行检测。Zhu等^[10]使用残差神经网络(Residual Net, ResNet)进行特征提取并使用机场的显著图进行区域建议来对机场进行检测。Chen等^[11]结合深度置信网络(Deep Belief Network, DBN)用于飞机检测。Wu等^[12]采取BING + CNN的方法对飞机进行目标检测。上述方法应用了深度神经网络强大的特征提取能力,相比人工设计特征拥有更好的鲁棒性,从而大大提高了检测准确率,但依旧存在候选框过多使得检测效率偏慢的问题。针对以上问题,国内外提出了众多有贡献的改进算法,其中Faster R-CNN^[13]是其中的一大代表检测框架,其区域建议网络(Regional Proposal Net, RPN)的提出大大加快了检测效率,其中包含的各方面改进也使得检测效果达到了新高,其对多类目标强大的分类、定位能力以及较好的检测效率使得人们对其的关注度越来越高,并在此框架基础上根据自己的数据集做出相应的改进并取得了很好的效果。戴陈卡和李毅^[14]将Faster R-CNN和多部件结合方法用于机场区域内飞机目标的检测,相较本文前述算法,较大地提高了检测性能且降低了时间成本,但由于数据集以及网络本身的限制,对小目标飞机检测不理想。朱明明等^[15]在区域卷积神经网络的基础上应用特征融合和软判决的方式对机场区域内飞机目标

的检测进行了改进,有效地提高了对小目标的检测率,并在一定程度上改善了遮挡问题。Chen等^[16]通过合理增加锚框的尺度范围以及进行候选框中跑道区域的面积占比来进行候选区域的约束进行机场的检测,在一定程度上提高了机场的检测精度。但是众实验多是在对机场区域内单类目标的研究,如对机场或飞机目标的检测研究。对机场区域内多类目标检测的研究不足,且由于网络本身结构以及所用数据集本身的局限性使得检测精确率还有较大的提升空间。

以机场区域目标为检测对象,包括机场、民航飞机、战斗机、运输机、直升机、油罐以及桥梁目标;以卫星遥感图片为检测源,以深度学习下Faster R-CNN为检测框架,通过数据集对其进行调试并改进来完成对机场区域的目标检测。主要工作以及创新点如下:

- 1) 针对数据侧,提出ReMD数据增强算法,在不损失图像信息的前提下丰富检测对象的背景并使得模型可以更多地从局部了解目标相关信息。
- 2) 在网络的搭建中,加入更深的基础网络以及特征融合检测部件以提取更深层次的区分性鲁棒特征。
- 3) 针对检测对象的类间关联特性组合更优的新型末端检测器。

1 机场区域目标检测原理与改进

1.1 检测框架的搭建

1.1.1 检测原理

以尽可能地提高卫星遥感监测任务中目标的检准性以及检全性为出发点,应用Faster R-CNN为检测的基础框架,具体检测原理如图1所示。

在遥感机场区域目标检测任务中,当待训练或检测的图片进入神经网络后,首先会进入特征提取网络(Feature Extraction Network)对其进行特征提取,特征提取网络由众多隐含层构成,主要以众多 $3 \times 3 \times n$ 卷积(Convolution)来对输入的图片进行特征提取,并借助池化(Pooling)操作来控制特征图的尺寸,方便后续计算,最后得到完整的全局特征图。然后,特征图进入区域建议网络将映射到原图,借助softmax分类器以及bounding box regression函数以及数据集中相应的标签信息进行区域建议(Regional Proposal),再经过非极大值抑制(Non-Maximum Suppression, NMS)算法减少可选候选框的数量并映射回特征图。其次,进入感兴趣池化(ROI Pooling)层进行感兴趣的建议区域池化,将其改变为固定尺寸,方便后续输入。

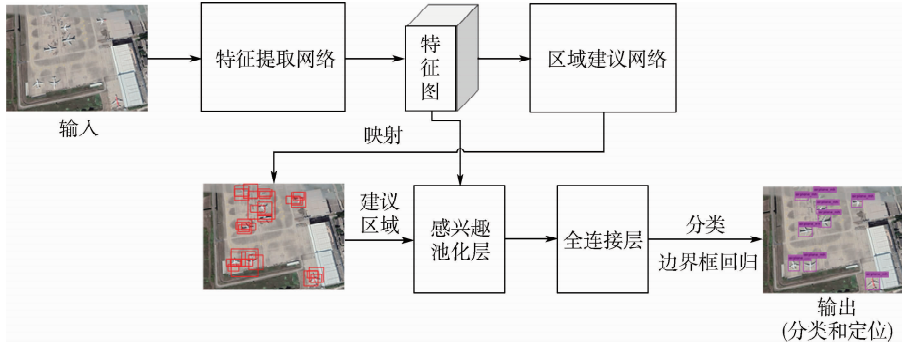


图1 机场区域目标检测示意图

Fig. 1 Schematic diagram of airport area target detection

最后进入全连接层(Full Connection)结合 softmax 分类器以及 bounding box regression 函数来实现目标的精确分类以及定位。最终生成遥感机场区域目标检测输出图,如图1最终的输出图所示。

1.1.2 更好的特征提取以及融合网络

神经网络的深度对于模型性能的提升有着至关重要的作用,在 ImageNet 数据集^[17]上获胜的网络大多都采用了更深的网络结构,比如 ZFNet^[18]、VGG-16^[19]、VGG-19、GoogLeNet^[20]等,以此也揭示了网络深度的重要性。然而得到性能提升的同时,也不可避免地带来了因单纯增加网络深度造成梯度消失/爆炸以及精确率饱和后下降的问题,在解决梯度问题时可以使用 L1 或 L2 正则化来得到部分程度的补偿,但是精确率饱和的问题却需要在单纯增加网络深度的基础上进行更具创新性的改进。针对此类问题,最具代表的解释为:图像具有局部相关性,神经网络越来越深的时候,反传回来的梯度之间的相关性会越来越差,最后接近白噪声。ResNet 网络^[21]根据这一问题采用了“跳层连接”的方式,即把输入直接叠加到输出上,在反向传播的时候,这一路梯度是原封不动的往回传的,相关性很强,如图2所示。

假设图2(a)经过2次卷积以及激活整形操作后的结果为 $F(x)$, 则 $H(x) = F(x)$, $H'(x) = F(x) + x$, x 为上一层的输出。由此可以看出 ResNet 在进行卷积层之间的运算时通过“跳层连接”

将上一层的输出可以直接作用在下一层的输出之中,从而保留了临近层之间提取的图像局部相关性。

同时在进行特征提取的过程中,随着网络深度的增加,提取到的特征会有越来越大的差异,通常可以将其区分为低层特征和高层特征,低层特征的语义信息很弱,但是有较强的纹理信息,因而更适合用来定位;而高层特征可以提取到更为抽象的特征且含有更强的语义信息,因此更适合用来分类。针对研究目标中飞机类的目标而言,其在一幅图片中可能拥有较少的像素信息,因此随着网络深度的增加,最后提取到的高层特征将很难对目标进行有效的定位以及良好的分类。在此可以使用特征融合的方式来进行一定程度的改进。在具体实施的过程中,为了充分应用各层提取的特征,借鉴 FPN^[22]的思想,进行“top-down path”预测,即在进行多层特征融合预测的同时进行分层单独预测。

在进行特征融合方式设计时,为了获得更好的特征融合质量,尽量使得融合后的特征图有尽可能大的特征尺寸以及维度,因此在特征融合过程中将高层特征做2倍上采样,同时用 1×1 卷积仅改变上一层特征的维度与高层特征维度相同,然后使用“add”的方式将2层特征进行融合,加权时二者权重相同。同理将低层特征与融合后的特征进行融合,得到最终的融合特征(Predict-3),同时3个特征(Predict-1 ~ Predict-3)都进入特征检测网络,如图3所示。

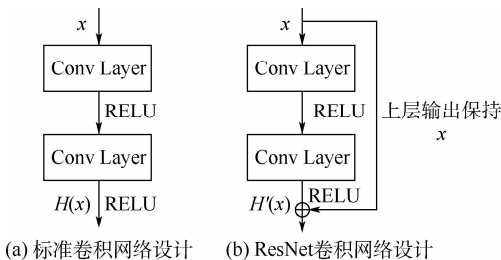


图2 “跳层连接”示意图

Fig. 2 Schematic diagram of “layer jump connection”

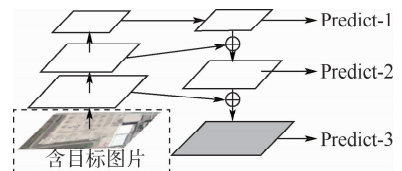


图3 FPN 特征融合检测示意图

Fig. 3 Schematic diagram of FPN feature fusion detection

1.2 数据集的准备

目前公用机场区域的卫星遥感图像数据集很少,这也是特定目标检测存在的一大难点,因为想要训练一个高效的检测网络,较为完备的数据集是其必要条件。已有文献[16]表明从高精度卫星检测软件获取的图片对目标检测具有很高的适应性以及泛化性,因此以 Google Earth 为自主获取包含目标信息图片的源,以 VOC2007^[23]为制作数据集的标准格式来进行数据集的制作。

制作的数据集共包含 200 多个不同的机场区域,共有 7 264 张原始图片,其中包含 1 982 张机场图片、1 838 张民航飞机图片、565 张战斗机图片、715 张直升机图片、813 张运输机图片、583 张桥梁图片以及 768 张油罐图片。数据集示意如图 4 所示。

图 4 所示。



图 4 数据集示意图

Fig. 4 Schematic diagram of data set

各目标对应标签信息如表 1 所示。为了方便类别间的重组,将所有飞机类的标签统一添加“airplane_”前缀。其中,airplane_mh 代表民航飞机,airplane_y 代表运输机,airplane_z 代表战斗机,airplane_zs 代表直升机。

表 1 目标-标签对应表

Table 1 Object-label correspondence table

目标	airport	civil airplane	transport plane	fighter	helicopter	bridge	oil tank
标签	airport	airplane_mh	airplane_y	airplane_z	airplane_zs	bridge	oil tank

1.3 网络的改进

1.3.1 ReMD 数据增强算法

数据增强,是指对有限的训练数据通过某种变换操作,从而生成新数据的过程。对于数据集较少时,数据增强技术可以起到非常重要的作用,其中对训练过程中的各类别数据不均衡问题也能起到一定的改善。

ReMD 图像增强算法,即指随机擦除、区域填充以及逆向重组(Random erase + Mix + Double)相关联的算法。原理如下所示:

假设训练样本集为: $X = \{x_1, x_2, \dots, x_m\}$, x_a 、 x_b 为其中的 2 个样本。

首先对 x_a 样本进行随机擦除:选取一个需要擦除区域的边界框 $A = (a_x, a_y, a_w, a_h)$, 其中 (a_x, a_y) 是边界框的中心点坐标; (a_w, a_h) 是边界框的几何特征,即其宽度和高度。

对擦除区域的边界框进行采样:

$$\begin{cases} a_x \sim \text{Unif}(0, W_a) \\ a_y \sim \text{Unif}(0, H_a) \end{cases} \quad (1)$$

式中: W_a 为选取样本的宽度; H_a 为选取样本的高度。

引入参数 $\lambda \sim U(0, 1)$, 使得

$$\begin{cases} a_w = W_a \sqrt{\lambda} \\ a_h = H_a \sqrt{\lambda} \end{cases} \quad (2)$$

确保擦除的区域与样本的关系为

$$a_w a_h = \lambda W_a H_a \quad (3)$$

使得随机擦除的区域边界框始终在样本有效像素点内。

选取 $R \in \{0, 1\}^{W_a \times H_a}$, 为随机擦除部分区域并

进行区域填充而设定的二进制掩码。将二进制掩码中的擦除区域 A 取值为 0, 其他区域取值为 1, 即

$$R_{ij} = \begin{cases} 0 & (i, j) \in A \\ 1 & \text{其他} \end{cases} \quad (4)$$

以完成二进制掩码的采样,然后生成所需的随机擦除结合区域填充后的对应新样本:

$$x'_a = R \otimes x_a \oplus (I - R) \otimes x_b \quad (5)$$

式中: \otimes 为逐像素相乘; \oplus 为逐像素相加; I 为所有元素为 1 的二进制掩码。

最后再进行逆向重组,即用相同的擦除区域边界框以及二进制掩码 R 来对样本 x_b 进行随机擦除以及区域填充。同时,在混合图像过程中,当切割线穿过目标时,在合成图像的目标位置按照比例分配分类(Label)的结果,比如在一个矩形分类框的 Label 中民航飞机为 0.6、运输机为 0.4。效果如图 5 所示。

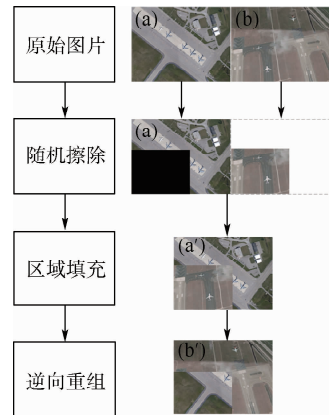


图 5 ReMD 效果示意图

Fig. 5 Schematic diagram of ReMD effect

应用该算法有如下优点:

1) 对随机擦除的区域中添加其他样本的信息,可以使得模型能够更多地从局部了解目标相关信息,提升模型定位的表现。

2) 丰富了检测物体的背景,因为原理类似图像组合,不会出现类似将2张图像“交叠”在一起时的混合后不自然的情形,同时保留了 regional dropout 的优势,能够提升模型分类的表现。

3) BN 计算时相当于一次计算多张图片,提

高单张图片的训练效率。经过逆向重组(见图5)可知,该算法可以充分利用样本的每一个像素信息,同时训练和推理代价保持不变。

1.3.2 新型末端检测器

为了进一步提高关联性较强的飞机类目标中民航飞机、直升机、战斗机以及运输机的平均检测精度,降低类间误检率。提出在 Faster R-CNN 检测框架的末端检测区域,根据目标类别间的关联特性构建一个结合 softmax 分类器以及 logistic regression 的末端检测器,如图6所示。

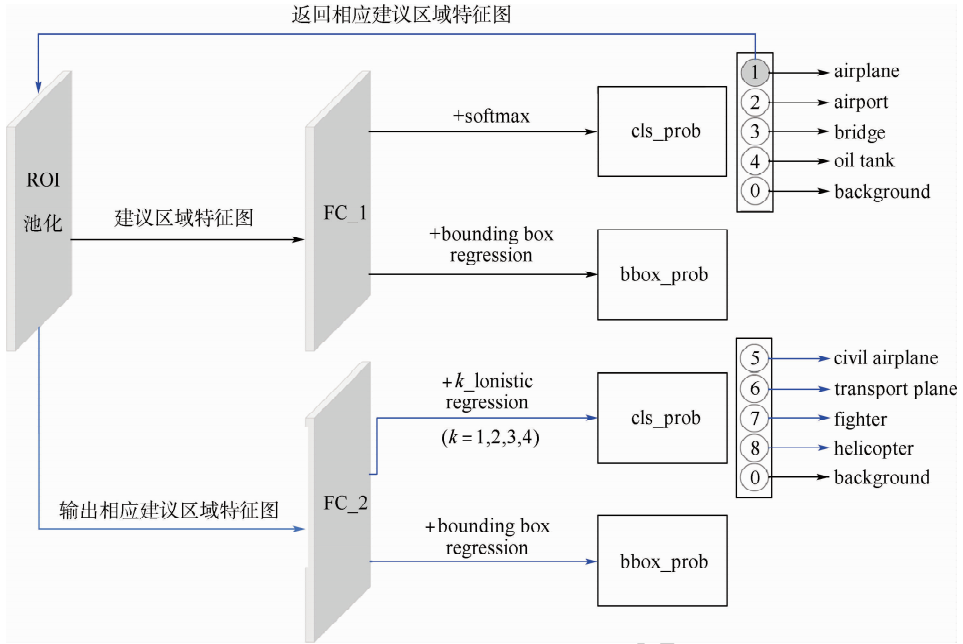


图6 新型末端检测器示意图

Fig.6 Schematic diagram of new end detector

在具体分类任务中,首先将各类待检测的目标分配类别号,如表2所示。

表2 目标-类别号对应表

Table 2 Object-number correspondence table

目标	airplane	airport	bridge	oil tank	civil airplane
标签	airplane	airport	bridge	oil tank	airplane_mh
类别号	1	2	3	4	5
目标	transport plane	fighter	helicopter	background	
标签	airplane_y	airplane_z	airplane_zs		
类别号	6	7	8	0	

表2中 airplane 是一个相对大类,包含 civil airplane、transport plane、fighter 以及 helicopter 4 个小类。在各类目标的层次特征中,机场、飞机类、桥梁、油罐的特征区分度较大,如颜色、形状、纹理等,而飞机类目标中的特征区分度则相对较小,比如部分飞机类之间具有相似的形状、尺寸等,因此理论上机场、飞机类、桥梁、油罐之间存在互斥关系,而飞机类目标间则有较强的关联性。

softmax 是多类目标分类器,在所有类别明显互斥的时候 softmax 分类器具有良好的分类性能,其分类机理如下:

如图6所示,当候选特征区域进入 FC1 后,会将输入信息转换为多个 $K \times 1$ 维变量输出到 softmax 层,然后由其输出最终各类别的置信度,如下所示:

$$a^{[L]} = \frac{e^{z^{[L]}}}{\sum_{i=1}^K e_i^{z^{[L]}}} \tag{6}$$

式中: $K=5$, 与设计中所需预测的类别数相同; $z^{[L]}$ 为 FC1 的输出变量; $a^{[L]}$ 为 softmax 层最终的输出变量。

logistic regression 可以看作 softmax 回归在 $K=2$ 时的特殊情况,已有研究^[24]表明多个 logistic regression 在关联性较强的数据集中相比单独的 softmax 分类器有更好的分类性能,因此在识别民航飞机、运输机、战斗机以及直升机飞机类目标

时,选择使用 4 个 logistic regression 来代替 softmax 分类器,在进行训练以及测试时,选择“一对其余”(One vs. Rest, OvR)的方式对分类任务进行拆分,即将识别 4 类飞机目标的任务拆分为 4 个二分类任务,在训练以及测试时,将每一类的样例作为正例,其余各类均为反例来训练,再对 4 个 logistic regression 分类器的预测结果进行集成输出。集成输出准则为:如有一个分类器预测为正,则输出该标签值,若有多个分类器预测为正,则考虑各个预测置信度,选择置信度最大的分类标签值作为输出结果。

2 实验配置

2.1 实验环境

CPU:Core i7-7700@3.60 GHz;RAM:16.0 GB; Graphics Card: GTX 1080Ti@6 GB; Operating System: Ubuntu; Frame: Caffe。

2.2 训练设置

使用模型迁移学习方法,用 imagenet_models 对网络进行预训练,但由于两者数据集相似度不高,且由于实际限制所构建的数据集较小,所以仅保留了每个预训练模型的前置层数参数,其余各层参数服从均值为 0,方差为 0.01 的高斯分布随机初始化,进行重新训练;初始学习率设置为 0.001,动量设置为 0.9,权重衰减设置为 0.0005;初始阈值设置为 0.7;训练迭代轮次设置为 8×10^4 次;用随机数法按照训练集:验证集:测试集 = 7:2:1 的比例进行数据集的随机分组。

3 针对搭建网络以及各改进部分的结果分析

3.1 ReMD 数据增强算法结果分析

各数据增强算法都可能造成数据量成倍数增

长,为了避免在数据增强过程中的爆炸性增长,训练时选择在每一个小批量 (Mini-Batch) 上执行相关数据增强算法,并使用 GPU 进行加速。实验结果如表 3 所示。

其中旋转 (Spin) 操作是 Faster R-CNN 在训练过程中本身具有的数据增强方法,镜像 (Mirror)、缩放 (Scaling)、平移 (Pan)、亮度变换 (Brightness Change)、裁剪 (Crop)、高斯加噪 (Gaussian Noise) 是常用且有效的已有数据增强算法,ReMD 为本文的数据增强算法。在进行对比训练时均使用 VGG-16 基础网络。

从上述对比实验结果可以得到,对数据量较小的数据集进行数据增强有助于提高模型识别的准确率。其中对于常规且有效的数据增强算法,如旋转、镜像、缩放等可以使得模型学到更多的信息,如旋转、位移等不变性,对多个目标类别平均检测精确率 (mAP) 提供了一定程度的提升;其中 ReMD (本文算法) 通过对随机分对样本进行随机擦除后区域填充以及逆向重组的方式,丰富了检测物体的背景并使得模型可以更多地从局部了解目标的相关信息,同时保留了 regional dropout 的优势,实验结果表明该算法可以为模型带来 1.3% 的 mAP 提升,证明了 ReMD 数据增强算法的有效性。最后通过在送入神经网络的每一个小批次样本 (Mini-Batch) 中应用以上多种数据增强技术,模型的 mAP 提高了 3.7%,有着较大的改善效果。

3.2 选用更好的基础网络以及进行特征融合后的检测结果分析

为了验证在 Faster R-CNN 中加入 ResNet 作为基础网络以及 FPN 作为检测部件的有效性,在高度一致的实验环境下,用自主标定好的数据集对其进行了对比实验测试,结果如表 4 所示。

表 3 各数据增强算法效果

Table 3 Effect of each data enhancement algorithm

数据增强算法	各算法使用情况							
Spin	√	√	√	√	√	√	√	√
Mirror		√	√	√	√	√	√	√
Scaling			√	√	√	√	√	√
Pan				√	√	√	√	√
Brightness Change					√	√	√	√
Crop						√	√	√
Gaussian Noise							√	√
ReMD (proposed)								√
mAP/%	68.9	69.4	69.8	69.9	70.4	71.1	71.3	72.6

表4 基础网络对比实验结果

Table 4 Comparison of experiment results of basic network

模型	mAP/%	Average IOU
ZFNet	58.4	0.392
VGG_CNN_M_1024	63.5	0.425
VGG-16	72.6	0.566
VGG-19	73.9	0.571
ResNet-50	74.1	0.572
ResNet-101	75.8	0.574
ResNet-50 + FPN	78.9	0.643
ResNet-101 + FPN	80.2	0.645

其中:

$$IOU = \frac{\text{Detection Result} \cap \text{Ground Truth}}{\text{Detection Result} \cup \text{Ground Truth}} \quad (7)$$

IOU 代表测试集中网络预测出来的检测框与原图像中对应目标的真实标记框的重合程度。Average IOU 即指测试集中所有图像的 IOU 的算术平均值。

实验结果表明:随着网络深度的增加,模型的 mAP 和 Average IOU 的提升将会越来越大,因为随着网络深度的增加,特征提取能力会越来越强,特征表达能力也会越来越好。其中 VGG-16 是深层网络的代表之一,相比仅有 5 层卷积层的 ZFNet 以及 VGG_CNN_M_1024 来看,对模型的性能已有了较大的提升,而 ResNet 网络因为其特有的“跳层连接”设计,使得深层网络模型性能的提升又迈出了一大步,ResNet-101 相比 VGG-16 有了 3.2% 的 mAP 提升。而加入 FPN 检测部件后,通

表5 加入新型末端检测器前后对比实验结果

Table 5 Comparison of experiment results before and after adding the new end detector

网络	AP/%							mAP/%
	airport	bridge	oil tank	Civil airplane	Transport plane	fighter	helicopter	
T1	89.7	75.8	76.9	89.6	78.5	72.0	78.9	80.2
T2	89.8	75.9	76.9	90.1	79.1	72.6	79.4	80.5

表6 各检测部件与所带来的时间成本

Table 6 Summary of each testing component and time cost

Faster R-CNN 网络及增加的部件	原网络	ReMD	ResNet-101	FPN	N2	N1
检测时间/s	0.215	0.215	0.483	0.510	0.512	0.217
$\Delta t/s$	0	0	0.268	0.027	0.002	0.002

通过表 6 可知,ReMD 数据增强算法并未带来额外的检测时间成本,因为该操作生效在训练时的线上增强过程中,而并未参与到检测框架之中;造成检测时间成本较大增加的为引入了更深的 ResNet-101 基础网络,该操作增加了 0.268 s 的检测耗时,而建立在其上的 FPN 特征融合部件以及新型末端检测器分别仅带来了 0.027 s 以及 0.002 s 的检测耗时,主要因为 ResNet-101 网络所

过有效的特征融合操作,在 ResNet-101 基础上有了 4.4% 的 mAP 以及 7.1% 的 Average IOU 的提升,证明了加入 FPN 特征融合的有效性。

3.3 新型末端检测器检测结果分析

用 T1 代表 Faster R-CNN 进行本文数据增强以及应用 ResNet-101 基础网络和 FPN 检测部件后的网络,T2 代表在 T1 的基础上应用该末端检测器后的网络。实验结果如表 5 所示。

从测试结果得到,所构建的新型检测框架应用于前述 Faster R-CNN 框架后,对于机场区域的飞机类目标的检测精确率提高了 0.5% 左右,对于机场区域各类目标的多类平均检测精度提高了 0.3% 左右,原因在于通过新增全连接层以及后续的多 logistic regression 配合原网络的 softmax 分类器,使得模型检测时可以结合具体所检测目标间的特性组合更优的分类器。对于该实验结果也证明了该方法对模型的检测产生了有益的影响。

3.4 针对各算法以及检测部件对网络带来的检测时间成本分析

通过控制变量法,在相同实验环境下进行多次实验,结果如表 6 所示。其中 N1 代表细分网络,则表格中“N1”表示仅在进行 ReMD 数据增强的 Faster R-CNN 上加入 N1,“N2”表示网络在进行 ReMD 数据增强以及加入 ResNet-101 和 FPN 的基础上加入 N1。检测时间增量为 Δt ,全部数据保留小数点后 3 位。

增加的用于深层次特征提取的网络层数以及参数大大多于 FPN 以及新型末端检测器所增加的额外参数。

4 可视化检测结果示例以及对比试验

4.1 检测结果示例

本文所改进后的模型检测示例如图 7 所示。

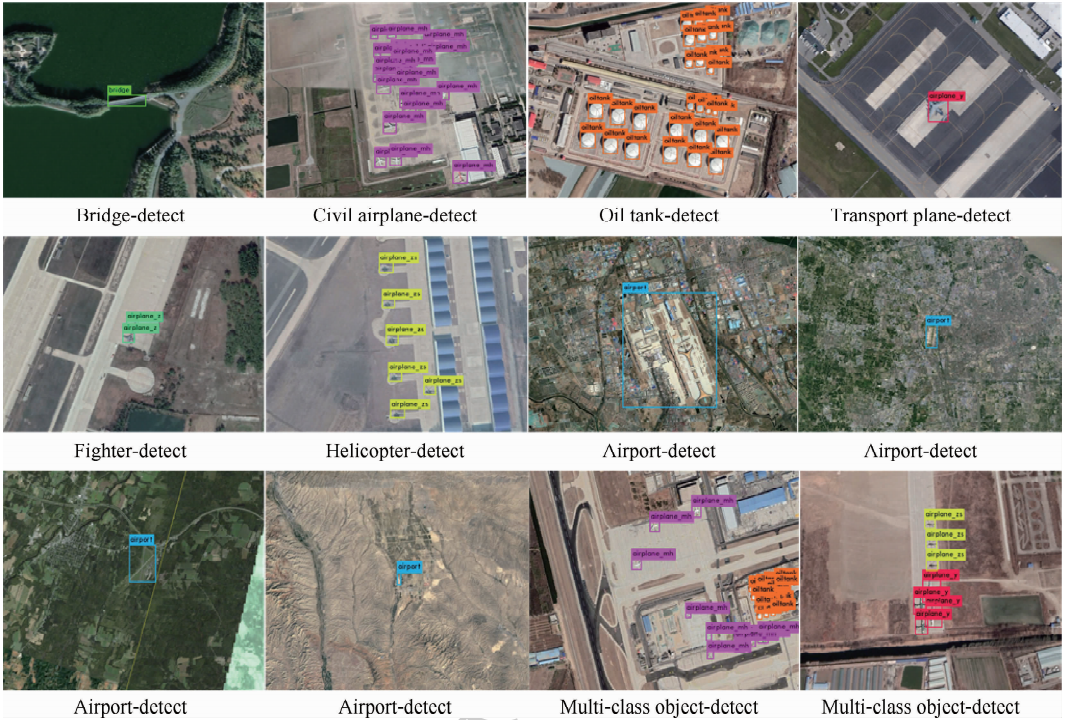


图 7 机场区域目标检测效果示意图

Fig. 7 Schematic diagram of object detection effect in airport area

4.2 对比实验分析

在相同的实验环境下,针对相同的数据集进行对比实验,实验结果如表 7 所示。相关图示如图 8 所示。

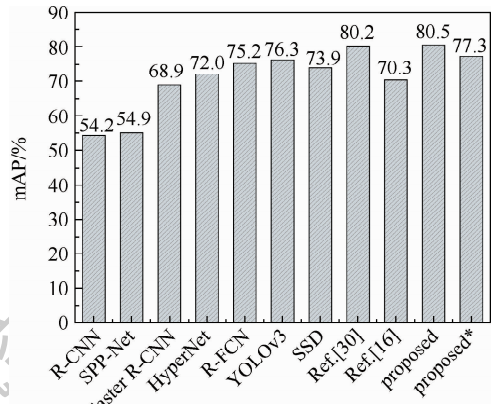
表 7 不同检测网络间的对比实验结果

Table 7 Comparison of experiment results between different detection networks

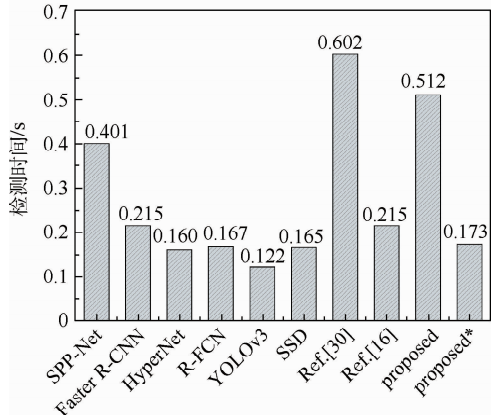
算法	mAP/%	检测时间/s
R-CNN	54.2	> 10.000
SPP-Net	54.9	0.401
Faster R-CNN	68.9	0.215
HyperNet	72.0	0.160
R-FCN	75.2	0.167
YOLOv3	76.3	0.122
SSD	73.9	0.165
Ref. [30]	80.2	0.602
Ref. [16]	70.3	0.215
proposed	80.5	0.512
proposed*	77.3	0.173

注:proposed*表示将 RPN 末端生成的 proposal 数量减少到 50 (原设为 300)。

其中图 8(a)展示了各算法在本文数据集下的 mAP 对比效果;图 8(b)展示了各算法的检测时间对比效果。通过对比实验,可以发现本文网络相比目前主流的目标检测算法 R-CNN^[25]、SPP-Net^[26]、Faster R-CNN、HyperNet^[27]、R-FCN^[28]、YOLOv3^[24]、SSD^[29],以及性能较好的复合网络 Faster R-CNN + InceptionResNetv2 + TDM^[30] 和



(a) 本文数据集下的 mAP



(b) 各算法的检测时间对比

图 8 对比实验结果示意图

Fig. 8 Schematic diagram of comparative experiment results Faster R-CNN++^[16],在 mAP 方面取得了更好的效果,mAP 达到了 80.5%,相比 Faster R-CNN 总体提高了 11.6%,比同样具有较高检测率的

YOLOv3 提高了 4.2%, 比文献[30]提高了 0.3%。但是, 由于应用了更深的特征提取网络 ResNet-101、FPN 检测部件以及在末端检测器加入新的全连接层, 使得相比 Faster R-CNN 增加了一倍多的时间成本, 检测时间达到 0.512 s, 虽依旧比文献[30]快 0.09 s, 检测速度却不到 YOLOv3 的 1/4, 这也是本文网络存在的一个缺陷。经过实验发现, 当 proposal 减小到 50 时(原设为 300), 网络降低了 3.2% 的 mAP, 但是检测时间却缩短至原本的 1/3 左右(0.173 s), 且此时依旧有较好的 mAP。在这里 proposal 指的是 RPN 末端生成的 region proposal 经过越界剔除以及 NMS 抑制后保留的前 N 个 proposal(下一步将进入 ROI 池化层, 如图 1 所示), 而在网络中进行更改时是比较容易进行的。因此在具体应用中可以根据检测精确率和检测速度的权重进行再设定, 寻优最适合任务的方法。

5 结论

1) ReMD 数据增强算法为模型带来了 1.3% 的 mAP 提升, 相比部分经典算法平均对 mAP 有 0.4% 的提升, 证明了 ReMD 数据增强算法的有效性, 同时综合多种数据增强技术, 实验结果表明, 对模型的 mAP 提高了 3.7%, 起到了一定的改善作用。

2) ResNet-101 网络以及特征融合检测部件 FPN 的加入, 使得模型可以提取到更鲁棒的特征, 相比原检测框架(Faster R-CNN + VGG-16)提高了 7.6% 的 mAP 以及 7.9% 的 Average IOU, 对模型的精确分类以及准确定位有了较大的提升。

3) 新型末端检测器通过对各类目标进行更优的组合检测分配, 使得模型的 mAP 提高了 0.3% 并使得飞机类目标的 mAP 提高了约 0.5%, 在目标精确分类方面起到了一定作用。

4) 与其他部分主流检测网络进行实验对比, 本文网络在检测精确率上均取得了较高的效果, 但是检测时间却相比原 Faster R-CNN 检测框架增加了 0.297 s, 同时通过减少 proposal 至 50 时可以在降低一定精确率的前提下提速 3 倍, 且此时检测精确率依旧高于其他部分主流检测网络。在具体任务中可以权衡精度以及速度来选择更合适的方法。

通过实验验证了本文方法的有效性, 对该类任务有较强的实用价值。在下一步工作中, 将着眼于扩充数据集的容量以及进一步提高网络的检测性能。

参考文献 (References)

- [1] RICHARDS J R. Remote sensing digital image analysis [M]. Berlin: Springer, 1999: 20-21.
- [2] 杨四海, 陈锐生, 谢维波. Hough 变换的特性分析: 一种全局观点[J]. 计算机辅助设计与图形学学报, 2006, 18(8): 1197-1204.
YANG S H, CHEN D S, XIE W B. Characteristics of hough transform: A global view[J]. Journal of Computer-Aided Design & Computer Graphics, 2006, 18(8): 1197-1204 (in Chinese).
- [3] 梁浩然. 自然图像的视觉显著性特征分析与检测方法及其应用研究[D]. 杭州: 浙江工业大学, 2016: 16-17.
LIANG H R. Research on saliency detection of natural image and its application [D]. Hangzhou: Zhejiang University of Technology, 2016: 16-17 (in Chinese).
- [4] CHEN L C, PAPANDEOU G, KOKKINOS I, et al. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(4): 834-848.
- [5] LI W, XIANG S M, WANG H B, et al. Robust airplane detection in satellite images [C] // 2011 18th IEEE International Conference on Image Processing. Piscataway: IEEE Press, 2011: 2821-2824.
- [6] 林煜东, 和红杰, 尹忠科, 等. 基于稀疏表示的可见光遥感图像飞机检测算法[J]. 光子学报, 2014, 43(9): 196-201.
LIN Y D, HE H J, YIN Z K, et al. Airplane detection in optical remote sensing image based on sparse-representation[J]. Acta Photonica Sinica, 2014, 43(9): 196-201 (in Chinese).
- [7] 仇建斌, 李士进, 王玮. 角点与边缘信息相结合的遥感图像飞机检测新方法[J]. 微电子学与计算机, 2011, 28(9): 214-216.
QIU J B, LI S J, WANG W. A new approach to detect aircrafts in remote sensing images based on corner and edge information fusion [J]. Microelectronics & Computer, 2011, 28(9): 214-216 (in Chinese).
- [8] AN Z Y, SHI Z W, TENG X C, et al. An automated airplane detection system for large panchromatic image with high spatial resolution [J]. Optik, 2014, 125(12): 2768-2775.
- [9] ZHANG P, NIU X, DOU Y, et al. Airport detection on optical satellite images using deep convolutional neural networks [J]. IEEE Geoscience and Remote Sensing Letters, 2017, 14(8): 1183-1187.
- [10] ZHU T H, LI Y H, YE Q K, et al. Integrating saliency and ResNet for airport detection in large-size remote sensing images [C] // 2017 2nd International Conference on Image, Vision and Computing (ICIVC). Piscataway: IEEE Press, 2017: 20-25.
- [11] CHEN X Y, XIANG S M, LIU C L, et al. Aircraft detection by deep belief nets [C] // 2013 2nd IAPR Asian Conference on Pattern Recognition. Piscataway: IEEE Press, 2013: 54-58.
- [12] WU H, ZHANG H, ZHANG J F, et al. Fast aircraft detection in satellite images based on convolutional neural networks [C] // 2015 IEEE International Conference on Image Processing (ICIP). Piscataway: IEEE Press, 2015: 4210-4214.

- [13] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(6): 1137-1149.
- [14] 戴陈卡, 李毅. 基于 Faster RCNN 以及多部件结合的机场场面静态飞机检测 [J]. *计算机应用*, 2017, 37(S2): 85-88.
DAI C K, LI Y. Aeroplane detection in static aerodrome based on Faster RCNN and multi-part model [J]. *Journal of Computer Applications*, 2017, 37(S2): 85-88 (in Chinese).
- [15] 朱明明, 许悦雷, 马时平, 等. 基于特征融合与软判决的遥感图像飞机检测 [J]. *光学学报*, 2019, 39(2): 71-77.
ZHU M M, XU Y L, MA S P, et al. Airplane detection based on feature fusion and soft decision in remote sensing images [J]. *Acta Optica Sinica*, 2019, 39(2): 71-77 (in Chinese).
- [16] CHEN F, REN R L, VAN DE VOORDE T, et al. Fast automatic airport detection in remote sensing images using convolutional neural networks [J]. *Remote Sensing*, 2018, 10(3): 443.
- [17] RUSSAKOVSKY O, DENG J, SU H, et al. ImageNet large scale visual recognition challenge [J]. *International Journal of Computer Vision*, 2015, 115(3): 211-252.
- [18] ZEILER M D, FERGUS R. Visualizing and understanding convolutional networks [M] // *Computer Vision-ECCV 2014*. Berlin: Springer, 2014: 818-833.
- [19] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [C] // *Computer Vision and Pattern Recognition*. Piscataway: IEEE Press, 2014: 121-124.
- [20] SZEGEDY C, LIU W, JIA Y Q, et al. Going deeper with convolutions [C] // *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Piscataway: IEEE Press, 2015: 1-9.
- [21] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition [C] // *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Piscataway: IEEE Press, 2016: 770-778.
- [22] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C] // *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Piscataway: IEEE Press, 2017: 936-944.
- [23] EVERINGHAM M, GOOL L, WILLIAMS C K I, et al. The pascal visual object classes (VOC) challenge [J]. *International Journal of Computer Vision*, 2010, 88(2): 303-338.
- [24] REDMON J, FARHADI A. YOLOv3: An incremental improvement [EB/OL]. (2018-04-08) [2019-07-18]. <https://arxiv.org/abs/1804.02767>.
- [25] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C] // *2014 IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE Press, 2014: 580-587.
- [26] HE K M, ZHANG X Y, REN S Q, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(9): 1904-1916.
- [27] KONG T, YAO A B, CHEN Y R, et al. HyperNet: Towards accurate region proposal generation and joint object detection [C] // *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Piscataway: IEEE Press, 2016: 845-853.
- [28] DAI J F, LI Y, HE K M, et al. R-FCN: Object detection via region-based fully convolutional networks [C] // *Neural Information Processing Systems*. Piscataway: IEEE Press, 2016: 379-387.
- [29] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot MultiBox detector [M] // *Computer Vision-ECCV 2016*. Berlin: Springer, 2016: 21-37.
- [30] SHRIVASTAVA A, SUKTHANKAR R, MALIK J, et al. Beyond skip connections: Top-down modulation for object detection [C] // *Computer Vision and Pattern Recognition*. Piscataway: IEEE Press, 2016: 256-266.

Regional object detection of remote sensing airport based on improved deep neural network

HAN Yongsai¹, MA Shiping^{1,*}, HE Linyuan¹, LI Chenghao¹, ZHU Mingming¹, XU Yuelei²

(1. School of Aeronautical Engineering, Air Force Engineering University, Xi'an 710038, China;

2. Institute of Unmanned Systems Technology, Northwestern Polytechnical University, Xi'an 710072, China)

Abstract: The detection of multiple types of targets in the airport area under the satellite remote sensing monitor is of great military and civilian significance in real life. In order to effectively improve the detection accuracy of remote sensing images in the airport area, based on the representative deep network Faster R-CNN in the mainstream target detection method, the ReMD data enhancement algorithm is proposed for the data side. The deep ResNet network and the feature fusion component-FPN are used to extract more robust deep distinguishing features of airport area target. Finally, a new fully connected layer is added to the end detection network, and the softmax classifier and 4 logistic regression classifiers are combined to accurately classify airport area multi-class targets according to the target class correlation. Experiments show that the improvement of the original network brings a 11.6% increase in the average detection accuracy rate of the original network, reaching 80.5% mAP. Compared with other mainstream networks, it also has a better accuracy rate. At the same time, by appropriately reducing the input amount of the recommended area, under the premise of 3.2% reduction of accuracy rate, the detection time of 0.512 s is improved by 3 times to 0.173 s. According to the specific task, the accuracy and detection speed can be reasonably weighed, which reflects the effectiveness and practicability of the network.

Keywords: object detection; image processing; remote sensing; airport area; neural network

Received: 2020-05-28; **Accepted:** 2020-08-21; **Published online:** 2020-09-22 15:34

URL: kns.cnki.net/kcms/detail/11.2625.V.20200922.1435.001.html

Foundation items: National Natural Science Foundation of China (61701524, 61773397); Aeronautical Science Foundation of China (20175896022)

* **Corresponding author.** E-mail: 1013765061@qq.com

http://bhxb.buaa.edu.cn jbuua@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2020.0585

混沌多精英鲸鱼优化算法

汤安迪¹, 韩统^{2,*}, 徐登武³, 谢磊¹

(1. 空军工程大学 研究生院, 西安 710038; 2. 空军工程大学 航空工程学院, 西安 710038;

3. 中国人民解放军 94855 部队, 衢州 324000)

摘 要: 针对无人机(UAV)的航迹规划问题,提出了一种基于混沌多精英鲸鱼优化算法(CML-WOA)的航迹规划方法。首先,在已知飞行环境下,建立3D飞行空间模型和航迹代价模型。通过引入罚函数,将有约束3D航迹规划问题转化为无约束多维函数优化问题,利用CML-WOA求解模型来获得最优航迹。其次,为克服WOA易陷入局部最优的缺陷,引入立方映射混沌算子改善初始种群,增强种群多样性,并通过自适应框架融入正余弦算法(SCA),利用多精英搜索策略有效地提高了算法开发能力和探索能力。最后,使用贪婪策略保证了收敛效率。通过20个基准函数测试和航迹规划仿真实验对提出的改进WOA进行验证。结果表明:所提算法相对其他算法,寻优性能明显提升,具有较强局部最优规避能力和更高的收敛精度与收敛速度;能够稳定快速地规划出代价最少、满足约束的安全可行的飞行航迹。

关键词: 航迹规划; 鲸鱼优化算法(WOA); 混沌算子; 多精英搜索; 正余弦算法(SCA); 全局优化

中图分类号: TP301.6

文献标志码: A 文章编号: 1001-5965(2021)07-1481-14

随着无人机(UAV)技术的不断发展,UAV逐渐在战争中发挥着越来越重要的作用。UAV航迹规划是UAV作战任务规划的重要组成部分,是一个典型的非确定性多项式问题。随着规划问题维度的不断增加,求解难度和计算量呈指数式增长,因此很难找到一种有效应对所有复杂环境的航迹规划方法。为解决不同条件下的航迹规划问题,学者们不断对航迹规划算法进行研究和改进。典型的方法有2类^[1],一类是传统优化方法,如Dijkstra算法^[2]、人工势场法^[3]等;另一类则是启发式智能优化算法,如粒子群算法^[4]、遗传算法^[5]、差分进化算法^[6]、灰狼优化算法^[7]、乌贼算法^[8]和蚁群算法^[9]等已被应用于求解航迹规划问题。

自1975年美国教授Holland根据达尔文进化

论以及自然界优胜劣汰机制提出了遗传算法以后,不断有学者通过对不同生物种群行为和物理现象进行分析,提出多种智能优化算法。其中澳大利亚学者Mirjalili于2016年提出了一种新的群智能优化算法鲸鱼优化算法^[10](Whale Optimization Algorithm, WOA)。WOA相比其他算法,具有参数少、原理简单和寻优能力强的优点,已被应用于许多工程领域^[11-17],但同其他智能优化算法类似,容易陷入局部最优,导致早熟。文献^[18]通过引入等级制度来扩大可行解的搜索范围,以此增强了算法的探索能力;文献^[19]通过和黏菌算法融合,增强了算法的局部开发能力;文献^[20]引入随机差分变异策略,增强了算法的开发能力;文献^[21]引入自适应调整权重,通过控

收稿日期: 2020-10-15; 录用日期: 2021-01-22; 网络出版时间: 2021-02-01 17:18

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20210201.1547.004.html

基金项目: 陕西省自然科学基金(2020JQ-481, 2021JM224); 航空科学基金(201951096002)

* 通信作者. E-mail: 418932433@qq.com

引用格式: 汤安迪, 韩统, 徐登武, 等. 混沌多精英鲸鱼优化算法[J]. 北京航空航天大学学报, 2021, 47(7): 1481-1494.

TANG A D, HAN T, XU D W, et al. Chaotic multi-leader whale optimization algorithm[J]. Journal of Beijing University of Aeronautics and Astronautics, 2021, 47(7): 1481-1494 (in Chinese).

制 WOA 的关键参数来增强算法的全局搜索能力。

针对 WOA 存在的不足,本文提出了一种混沌多精英鲸鱼优化算法(Chaotic Multi-Leader Whale Optimization Algorithm, CML-WOA)。首先,利用立方映射来产生均匀分布的初始种群,增强初始种群的多样性;其次,提出了一种自适应策略,将正余弦算法(Sine Cosine Algorithm, SCA)融入 WOA 中,每个个体根据自适应参数选择使用 WOA 策略还是 SCA 策略来更新自身位置,同时对每代的最优个体利用 SCA 策略进行更新,充分利用 SCA 的周期振荡能力,帮助最优个体改善全局搜索能力,从而跳出局部最优;当算法陷入停滞时,使用多精英搜索策略,采取 2 种更新方式,平衡算法的开发和探索能力,增强种群多样性,加快算法的收敛精度和收敛速度;最后,利用贪婪策略从父代种群和生成的子代个体中选择优势群体,保证算法的收敛效率。通过求解 20 个基准测试函数验证改进算法的有效性,将其应用于航迹规划问题求解,结果表明:CML-WOA 寻优性能明显提升;能够有效地解决航迹规划问题,高效稳定地规划出高质量飞行航迹。

1 鲸鱼优化算法

WOA 是澳大利亚学者 Mirjalili 等根据座头鲸的狩猎方式提出的一种新的群智能优化算法。鲸鱼在大海中随机游走寻找猎物,使用一种泡网觅食法的狩猎方式,先深入水底收缩包围猎物,而后螺旋上升在靠近海面的地方捕食猎物。WOA 数学模型主要包括 3 部分:游走搜索猎物、收缩包围猎物和螺旋捕食猎物。

1.1 游走搜索猎物

WOA 在寻优过程中,通过参数 A 的控制,来控制种群的搜索策略。当 $|A| > 1$ 时,WOA 进行全局搜索,随机选择一个个体作为参考个体,按式(5)生成新个体。

$$a = 2 - 2t/t_{\max} \quad (1)$$

$$A = 2ar - a \quad (2)$$

$$C = 2r \quad (3)$$

$$D = |C \cdot X'_{\text{rand}} - X^t| \quad (4)$$

$$X^{t+1} = X'_{\text{rand}} - A \cdot D \quad (5)$$

式中: t 为当前迭代次数; t_{\max} 为最大迭代次数; r 为 $[0, 1]$ 之间的随机数; X'_{rand} 为随机选择的搜索个体; X^t 为第 t 代种群个体; D 为包围步长; A 为生成个体离参考个体的远近程度。

1.2 收缩包围猎物

在收缩包围阶段,鲸鱼个体不再随机选择个

体进行位置更新,而是选择当前最优个体进行位置更新,实施包围策略,从而使种群获得最优解。收缩包围猎物的位置更新公式为

$$D_{\text{best}} = |C \cdot X^t_{\text{best}} - X^t| \quad (6)$$

$$X^{t+1} = X^t_{\text{best}} - A \cdot D_{\text{best}} \quad (7)$$

式中: X^t_{best} 为搜索至第 t 代种群时最好鲸鱼个体的位置,通过向鲸鱼种群最优个体包围,模拟收缩包围猎物的行为。

1.3 螺旋捕食猎物

鲸鱼通过一种螺旋式游走的方式形成气泡网进行捕食,具体游走模型如下:

$$X^{t+1} = X^t_{\text{best}} + D_{\text{best}} \cdot e^{bl} \cdot \cos(2\pi l) \quad (8)$$

式中: b 为控制螺旋线形状的常数; l 为 $[0, 1]$ 的随机变量。鲸鱼的捕猎行为和螺旋收缩是同时进行的,假设两者概率均为 50%,具体数学模型如下:

$$X^{t+1} = \begin{cases} X^t_{\text{best}} - A \cdot D_{\text{best}} & p \geq 0.5 \\ D_{\text{best}} \cdot e^{bl} \cdot \cos(2\pi l) + X^t_{\text{best}} & p < 0.5 \end{cases} \quad (9)$$

2 改进策略

2.1 立方混沌映射

当前大多数智能优化算法的初始种群都是在搜索空间中随机生成的,而初始化的种群质量对于优化算法的效率有很大影响,均匀分布的种群有利于扩大搜索范围,进而提高算法的收敛速度和精度。

WOA 和其他智能优化算法一样,在求解复杂优化问题时,存在迭代后期种群个体多样性减少,易陷入局部最优的缺陷。为了增强种群个体多样性,考虑到混沌算子具有随机性与规律性的特点,且能在一定范围内不重复遍历所有状态,于是采用混沌算子对种群初始化,如图 1 所示,立方映射比 Logistic 映射的均匀性好,因此采用立方映射混沌算子。

立方映射公式如下所示:

$$y(o+1) = 4y(o)^3 - 3y(o) \quad (10)$$

$$-1 < y(o) < 1, y(o) \neq 0 \quad o = 0, 1, 2, \dots \quad (11)$$

初始化鲸鱼种群由 M 个 d 维个体组成,采用立方映射初始化鲸鱼种群的具体过程为:随机产生一个上下限为 1 和 -1 的 d 维矩阵作为第 1 个个体,之后用式(10)对第 1 个个体的每一维进行迭代得到剩余 $M-1$ 个个体,最后使用式(12)将立方映射产生的变量值映射到鲸鱼个体上。

$$X = X_{lb} + (X_{ub} - X_{lb}) \times (y + 1)/2 \quad (12)$$

式中: X_{lb} 、 X_{ub} 为每个个体每个维度的上下限; X 为

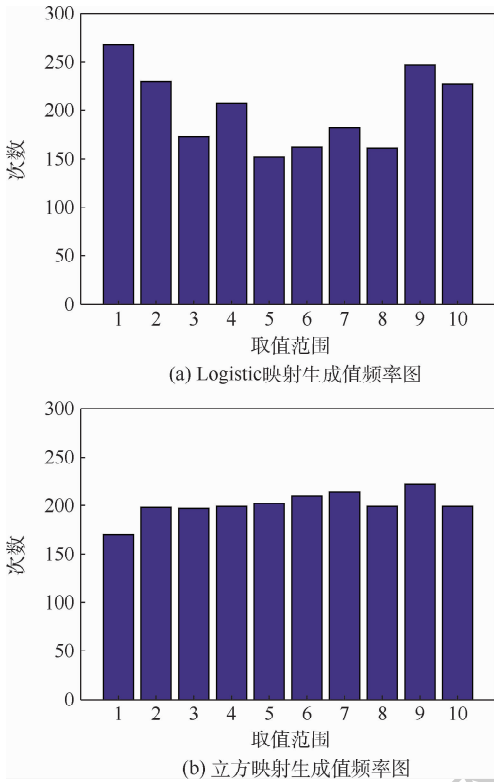


图1 映射对比图

Fig.1 Mapping contrast diagram

映射后个体。

2.2 正余弦优化策略

正余弦优化策略^[22]是利用正弦函数与余弦函数的振荡特性进行寻优,其优点是收敛性好,易于实现。

首先随机生成 N 个搜索个体,依据适应度函数计算每个个体的适应度值,将适应度值最好的个体记为最优个体 X^* ,寻优过程中个体位置更新公式为

$$X^{t+1} = \begin{cases} X^t + r_1 \cdot \sin r_2 \cdot |r_3 \cdot X^* - X^t| & r_4 < 0.5 \\ X^t + r_1 \cdot \cos r_2 \cdot |r_3 \cdot X^* - X^t| & r_4 \geq 0.5 \end{cases} \quad (13)$$

$$r_1 = c(1 - t/t_{\max}) \quad (14)$$

式中: c 为大于1的常数; r_2 为服从 $0 \sim 2\pi$ 均匀分布的随机数; r_3 为服从 $0 \sim 2$ 均匀分布的随机数; r_4 为服从 $0 \sim 2$ 均匀分布的随机数。

2.3 自适应策略

为了将 SCA 有效地嵌入到 WOA 中,受文献[23]选择坐标系的自适应框架启发,提出一种自适应策略,每个搜索个体根据自适应概率参数 pb 选择采用哪种策略,并根据迭代过程中表现优劣对 pb 进行更新,更新策略如下:

WOA 策略较好:使用 WOA 策略更新个体位置,子代个体优于父代个体,此时 pb 的更新模型为

$$pb_i^{t+1} = \alpha \cdot pb_i^{t+1} + (1 - \alpha)(1 - pb_i^{t+1})(t/t_{\max}) \quad (15)$$

WOA 策略较差:使用 WOA 策略更新个体位置,子代个体劣于父代个体,此时 pb 的更新模型为

$$pb_i^{t+1} = \alpha \cdot pb_i^{t+1} + (1 - \alpha)pb_i^{t+1}(1 - t/t_{\max}) \quad (16)$$

SCA 策略较好:使用 SCA 策略更新个体位置,子代个体优于父代个体,此时 pb 的更新模型为

$$pb_i^{t+1} = \alpha \cdot pb_i^{t+1} + (1 - \alpha)pb_i^{t+1}(1 - t/t_{\max})$$

SCA 策略较差:使用 SCA 策略更新个体位置,子代个体劣于父代个体,此时 pb 的更新模型为

$$pb_i^{t+1} = \alpha \cdot pb_i^{t+1} + (1 - \alpha)(1 - pb_i^{t+1})(t/t_{\max})$$

2.4 多精英选择策略

在算法迭代寻优过程中,采用前一半优势种群的平均适应值来判断算法是否陷入停滞,如果优势种群的平均适应值在连续2次迭代中没有发生变化,则认为算法搜索陷入停滞,此时使用多精英搜索(Multi-Leader Search, MLS)策略生成新个体,以帮助算法跳出局部最优,避免算法的早熟收敛。MLS 策略围绕多个精英个体进行局部搜索,设置一个保存当前精英个体且具有自适应规模的集合 L ,在初始状态时,只包含一个当前最优解,当算法每次陷入停滞时,集合 L 便增加一个最优解,储存到一个事先给定的最大值时,集合 L 数量不再增加。

MLS 策略对集合中的精英解执行2种搜索行为。MLS 策略将围绕集合 L 中的精英个体进行局部搜索来产生新个体,对于没有被选入集合 L 的个体,采用式(19)进行更新。

$$X^{t+1} = X_{\text{best}}^t + \sigma \cdot (X_{\text{best}}^t - X^t) + r_5 \cdot X_{\text{best}}^t - r_6 \cdot X^t \quad (17)$$

$$\sigma \sim N(0, I), r_5, r_6 \sim N(0, 1), X_{\text{best}}^t \in L$$

式中: X_{best}^t 为集合 L 中随机选择的精英个体。通过 r_5 和 r_6 两项添加小扰动,避免搜索陷入局部最优。

对被选入集合的个体,采用式(20)生成子代。

$$X^{t+1} = X^t + |z| \cdot (X_{\text{best}}^t - X^t) + D \cdot (1 - t/t_{\max}) \cdot y \quad (18)$$

$$z \sim N(0, 1), X_{\text{best}} \in L, X_{\text{best}} \neq X$$

此外,当前最优解 X^* 的更新和其他个体依赖相同的种群信息,进而导致 X^* 的全局搜索能力较差,为进一步增加算法的探索能力,在每一次迭代中,对 X^* 采用正余弦策略进行更新。

最后采用贪婪策略保证 CML-WOA 的全局收敛效率,根据适应度值从子代个体与父代个体中

选择较优个体,组成新种群,进而充分保留优势个体,提高算法收敛效率。

2.5 改进鲸鱼优化算法

针对基本 WOA 存在的不足,本文提出了 CML-WOA。使用立方映射混沌算子初始化种群,改进后的种群初始位置分布更加均匀,扩大了鲸鱼群在空间中的搜索范围,增加了群体位置的多样性;利用自适应框架将 SCA 嵌入到 WOA 中,种群中的每个个体能够根据自适应概率参数选择使

用 WOA 策略或 SCA 策略;对每一代最优个体使用 SCA 策略,利用 SCA 的振荡周期性帮助最优个体改善探索能力,进一步增强算法的全局搜索能力;当算法陷入停滞时,利用 MLS 策略,采取 2 种更新方式,平衡算法的开发和探索能力,增强种群多样性,加快算法的收敛精度和收敛速度;最后利用贪婪策略从父代种群和生成的新个体中选择优势群体,保证算法的收敛效率。CML-WOA 的流程图如图 2 所示。

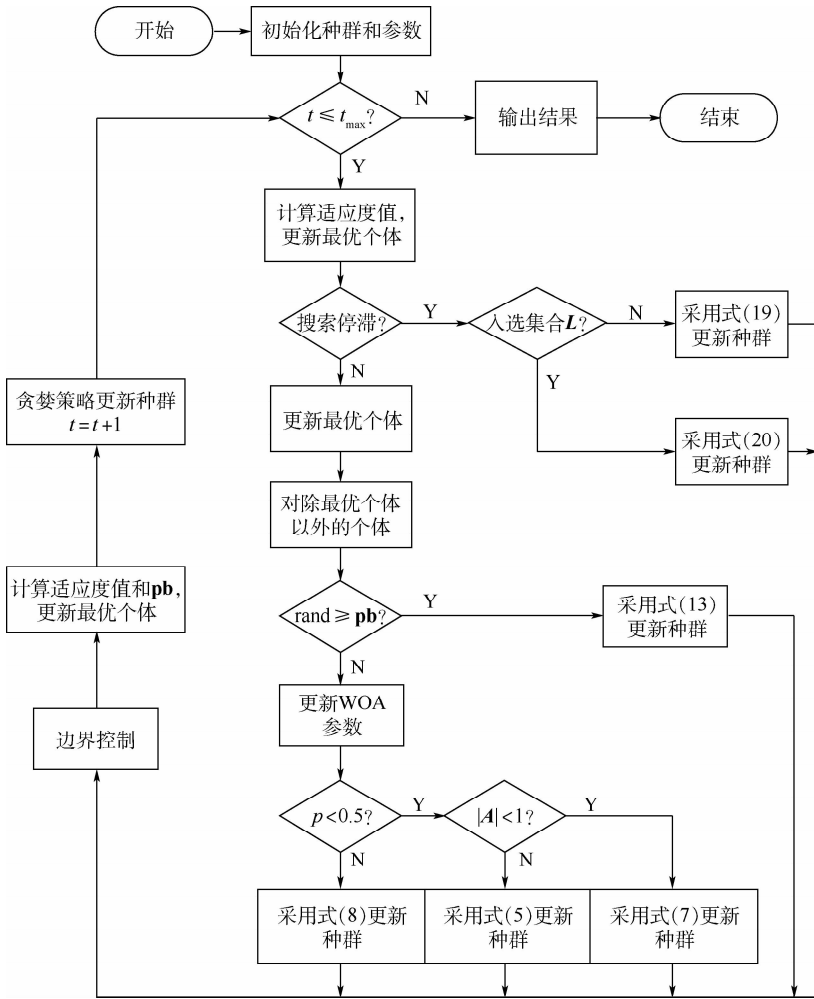


图 2 CML-WOA 流程图

Fig. 2 Flow chart of CML-WOA

3 航迹规划问题

3.1 航迹规划原理

航迹规划指在已知飞行区域内规划一条能够在规定时间段内从起飞点 (U) 到达目标点 (T)、且满足各类约束的最优航迹。本文采用离散的航迹点表示飞行航迹,通过将航迹点 ($U, R_1, R_2, R_3, \dots, R_N, T$) 进行连线得到一条满足约束的最优航迹,如图 3 所示,由于是点与点之间的连线段,需要进行航迹平滑处理,本文采用 B 样条法处理。

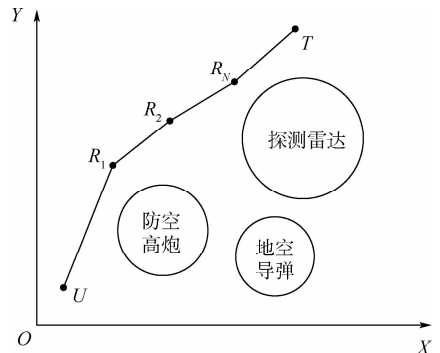


图 3 航迹规划原理

Fig. 3 Principles of path planning

3.2 坐标系转换

传统二维航迹规划考虑 2 个方向上的变量,为简化研究问题维数,提高计算效率,本文采用新建坐标系进行降维,以 UAV 起点 U 为坐标原点,将起点 U 和目标点 T 连线作为横轴,垂直于线段 UT 作纵轴,建立新的坐标系。新坐标系如图 4 所示。

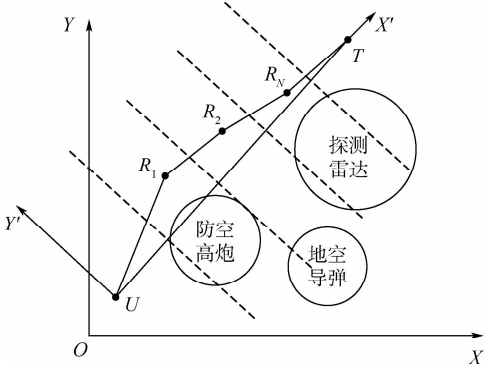


图 4 坐标系转换

Fig. 4 Coordinate system transformation

在图 4 中,将线段 UT 等分为 $N + 1$ 份,则转换后的航迹点在 $X'UY'$ 坐标系中的横坐标已知,只需求解出航迹点对应纵坐标,然后将得到的横纵坐标转换到 XOY 坐标系中,即可得到所规划的航迹点,最后通过 B 样条法处理得到可行飞行航迹。

3.3 B-spline 曲线

B-spline 曲线是贝塞尔曲线的一般化形式,公式如下:

$$B(x) = \sum_{i=0}^{m-n-2} P_i L_{i,n}(x) \quad x \in [x_n, x_{m-n-1}] \quad (19)$$

式中: m 为节点 x 的个数,且 x 的取值范围为 $\{x_0, x_1, \dots, x_{m-1} | x_0 \leq x_1 \leq \dots \leq x_{m-1}\}$; P_i 为控制节点,共有 $m - n - 1$ 个控制节点; $L_{i,n}$ 表示 n 阶的 B-spline 曲线基数,公式如下:

$$L_{i,1} = \begin{cases} 1 & x_i \leq x < x_{i+1} \\ 0 & \text{其他} \end{cases} \quad (20)$$

$$L_{i,n}(x) = \frac{x - x_i}{x_{i+1} - x_i} L_{i,n-1}(x) + \frac{x_{i+k} - x}{x_{i+k} - x_{i+1}} L_{i+1,k-1}(x) \quad (21)$$

B-spline 曲线的计算成本较低,能够控制样条形状,易于保证路径连续性。而 Dubins 曲线不具有曲率连续性,贝塞尔曲线稳定性不够,因此本文选用 B-spline 曲线对航迹段进行处理。航迹平滑效果如图 5 所示。

3.4 约束函数和适应度函数

为了更好的规划出满足约束的高质量飞行航迹,需要建立合适的适应度函数以及考虑各种约

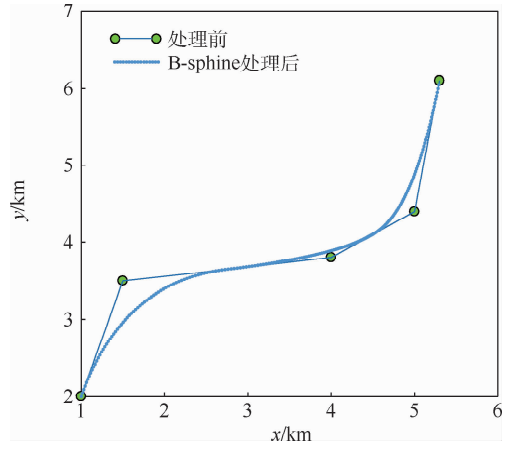


图 5 B 样条曲线航迹平滑效果

Fig. 5 Effect of B-spline trajectory smoothing

束,尽可能符合实际要求。静态三维全局航迹规划主要包括 2 部分,第一部分是适应度函数,需要考虑 UAV 的飞行油量代价、飞行高度代价以及威胁代价;另一部分则是各种约束条件,如最大转弯角、最大俯仰角。

3.4.1 飞行油量代价

在规划航迹时考虑飞行油量代价,是为了保证 UAV 执行任务后能够顺利返航。假定 UAV 在执行任务中保持恒定速率,则油量与航迹长度成正比,表示为如下代价函数:

$$\text{Cost}_g = \varepsilon \sum L_i \quad (22)$$

$$L_i = \sqrt{(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2 + (z_i - z_{i-1})^2} \quad (23)$$

式中: ε 为油量约束和航迹长度的比值; $\sum L_i$ 为相邻航迹点连线段之和,本文采用欧氏距离, L_i 为第 i 段航迹距离; (x_i, y_i, z_i) 为第 i 个航迹点的坐标。

3.4.2 高度代价

任务中 UAV 飞行高度也是影响作战效果的因素之一,低空飞行有利于 UAV 隐蔽,因此考虑 UAV 在飞行过程中的高度代价,模型如下:

$$\text{Cost}_h = \sum H_i \quad (24)$$

式中: H_i 为每个航迹点的高度。

3.4.3 威胁代价和禁飞区约束

UAV 在执行任务时会遇到敌方的防空系统,其中包括探测雷达、防空高炮、地空导弹等威胁,本文将上述威胁在二维平面上近似看作一个个圆形区域,探测范围或打击范围作为其半径,规定 UAV 不能通过圆形区域。

对于如何判断是否穿过威胁区,本文采用如下方法:

- 1) 分别计算两相邻航迹点与威胁源圆心距

离 D_1 、 D_2 。

2) 情况 1: D_1 、 D_2 有 1 个或 2 个都小于威胁源半径 r , 则穿过该威胁区, 如图 6(a) 所示。

3) D_1 、 D_2 均大于威胁源半径 r , 则计算威胁源心和两航迹点连线距离为 D_3 。

4) 情况 2: 若 $D_3 \geq r$, 则不穿过该威胁区, 如图 6(b) 所示。若 $D_3 < r$, 则计算威胁源圆心到相邻航迹点连线段中点距离为 D_4 。

5) 情况 3: $D_4 \geq r$, 则穿过威胁区, 如图 6(c) 所示。

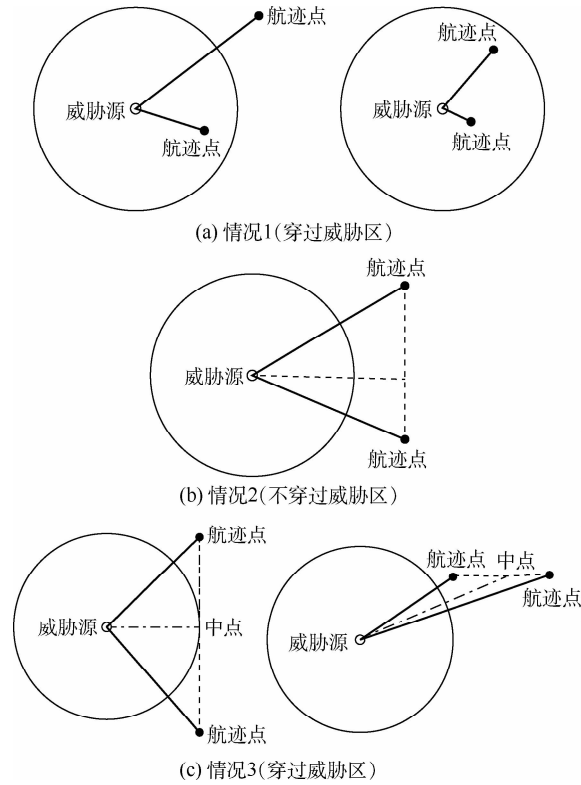


图 6 威胁区穿越判断

Fig. 6 Judgment of crossing the threat zone

3.4.4 最大转弯角约束

由于 UAV 性能限制, 实际飞行的最大转弯角不能无限制^[24], 同时在转弯过程中, UAV 会稍微偏离规划的航迹, 此时航迹曲率就要被考虑, 为确保规划出的航迹可以被 UAV 有效跟随, 所以要求生成的相邻航迹段间的夹角不能超过最大转弯角。在本文, 规定最大转弯角为 60° 。

3.4.5 最大俯仰角约束

UAV 在三维空间飞行时, 其爬升或下降的角度同最大转弯角类似, 不能超过最大俯仰角, 本文设定相邻航迹段间爬升角不能超过 60° 。

3.4.6 最大航程约束

假设 UAV 执行任务后需要原路安全返回基地, 那么 UAV 携带燃油要能够进行往返飞行。本文设定最大航程为 550 km。

综上, 本文使用代价函数为

$$\text{Fit} = \omega_1 \cdot \text{Cost}_g + \omega_2 \cdot \text{Cost}_h + \text{Constarints} \cdot \zeta \quad (25)$$

$$\text{Constraints} = \begin{cases} 0 & \text{结果满足约束条件} \\ 1 & \text{结果不满足约束条件} \end{cases} \quad (26)$$

式中: $\omega_1 + \omega_2 = 1$; ζ 为威胁惩罚系数, 通过引入惩罚函数系数将约束优化问题转化为无约束优化问题, 本文设 ζ 为 10^5 。

4 仿真与验证

4.1 算法测试

为验证本文所提改进算法的有效性与优越性, 选择 WOA 与改进算法 IWOA^[18]、PSO^[25]、BSO^[26]、SSA^[27]、VCS^[28] 和 GSA^[29] 进行对比分析。为保证公平, 在相同的实验平台基础上, 所有算法迭代次数相同, 均设置为 500 次, 种群规模设为 30, 各算法参数与原文献保持一致。所有算法均使用 MATLAB R2016b 编程, 计算机操作系统为 Windows10, 处理器为 AMD R7 4800U 16 GB RAM。对表 1 所示测试函数连续独立运行 30 次, 记录 30 次实验中各测试函数的最优值、平均值、标准差和时间, 测试结果如表 2 ~ 表 4 所示, 对其中最好的结果加粗显示。

其中 $F_1 \sim F_7$ 为单峰测试函数, 主要用于测试算法的开发能力, 根据表 2 可知, CML-WOA 在求解 $F_1 \sim F_4$ 和 F_7 时, 寻优效果最佳, 均能稳定得到全局最优解; 在求解 F_5 和 F_6 时, 虽然效果不是最好, 但寻优能力仍居于 2 个测试函数的第 3 位, 且明显优于 WOA 和 IWOA 算法。从整体上看, 对于单峰测试函数, CML-WOA 收敛效果优于其他算法, 且稳定性较好。

$F_8 \sim F_{20}$ 为多峰测试函数, 该类测试函数有多个局部最优解, 智能优化算法在求解时易于陷入局部最优。因此多峰测试函数主要用于测试算法的探索能力。由表 3 和表 4 可知, 在求解 F_8 和 F_{10} 时, CML-WOA 优于 4 种算法, 在性能相近的 3 种算法中, 耗时最短; 在求解 F_9 时, VCS 优于其他对比算法, 在耗时上优于 VCS 算法; 在求解 F_{11} 时, CML-WOA 效果不佳, 但仍明显优于 WOA 及其改进算法; 在求解 F_{12} 时, CML-WOA 明显优于其他对比算法, 且耗时较短。在求解 $F_{13} \sim F_{14}$ 和 F_{18} 时, CML-WOA 均优于 6 种对比算法; 在求解 $F_{15} \sim F_{17}$ 时, 除 WOA 和 IWOA 外, 另外 6 种算法性能相近, 但 CML-WOA 稳定性最强。在求解 F_{20} 时, CML-WOA 优于其他对比算法。因此, 提出的

CML-WOA 在求解 20 个基准测试函数时,在 14 均优于 WOA 和 IWOA,证明 CML-WOA 有较强的个测试函数中,收敛精度最佳,且在所有函数中,局部最优规避能力。

表 1 测试函数

Table 1 Test functions

函数分类	测试函数	维度	范围	最优值
单峰测试函数	$F_1(x) = \sum_{i=1}^n x_i^2$	30	$[-100, 100]$	0
	$F_2(x) = \sum_{i=1}^n x_i + \prod_{i=1}^n x_i $	30	$[-10, 10]$	0
	$F_3 = \sum_{i=1}^n \left(\sum_{i=1}^n x_i \right)^2$	30	$[-100, 100]$	0
	$F_4 = \max_i \{ x_i , -1 < i < n \}$	30	$[-100, 100]$	0
	$F_5(x) = \sum_{i=1}^n [100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2]$	30	$[-30, 30]$	0
	$F_6(x) = \sum_{i=1}^n ([x_i + 0.5])^2$	30	$[-100, 100]$	0
	$F_7(x) = \sum_{i=1}^n ix_i^4 + \text{random}[0, 1]$	30	$[-1.28, 1.28]$	0
多峰测试函数	$F_8(x) = \sum_{i=1}^n [x_i^2 - 10\cos(2\pi x) + 10]$	30	$[-5.12, 5.12]$	0
	$F_9(x) = -20\exp\left(-0.2\sqrt{\frac{1}{n}\sum_{i=1}^n x_i^2}\right) - \exp\left(\frac{1}{n}\sum_{i=1}^n \cos(2\pi x_i)\right) + 20 + e$	30	$[-32, 32]$	0
	$F_{10}(x) = \frac{1}{4000}\sum_{i=1}^n x_i^2 - \prod_{i=1}^n \cos\left(\frac{x_i}{\sqrt{i}}\right) + 1$	30	$[-600, 600]$	0
	$F_{11}(x) = \frac{\pi}{n} \left\{ 10\sin(\pi y_1) + \sum_{i=1}^{n-1} (y_i - 1)^2 [1 + 10\sin^2(\pi y_i + 1)] + (y_n + 1)^2 \right\} + \sum_{i=1}^n u(x_i, 10, 100, 4);$	30	$[-50, 50]$	0
	$y_i = 1 + \frac{x_i + 1}{4}, u(x_i, a, k, m) = \begin{cases} k(x_i - a)^m & x_i > a \\ 0 & -a < x_i < a \\ k(-x_i - a)^m & x_i < -a \end{cases}$			
	$F_{12}(x) = 0.1 \left\{ \sin^2(3\pi x_1) + \sum_{i=1}^n (x_i - 1)^2 [1 + \sin^2(3\pi x_i + 1)] + (x_n - 1)^2 [1 + \sin^2(2\pi x_n)] \right\} + \sum_{i=1}^n u(x_i, 5, 100, 4)$	30	$[-50, 50]$	0
固定维度测试函数	$F_{13}(x) = \left(\frac{1}{500} + \sum_{j=1}^{25} \left(j + \sum_{i=1}^2 (x_i - a_{ij})^6 \right)^{-1} \right)^{-1}$	2	$[-65, 65]$	0.998
	$F_{14}(x) = \sum_{i=1}^{11} \left[a_i - \frac{x_1(b^2 + b_i x_2)}{b_i^2 + b_i x_3 + x_4} \right]^2$	4	$[-5, 5]$	0.000 3
	$F_{15}(x) = 4x_1^2 - 2.1x_1^4 + \frac{1}{3}x_1^6 + x_1x_2 - 4x_2^2 + 4x_2^4$	2	$[-5, 5]$	-1.031 6
	$F_{16}(x) = [1 + (x_1 + x_2 + 1)2(19 - 14x_1 + 3x_2^2 - 14x_2 + 6x_1x_2 + 3x_2^2)] \times [30 + 2x_1 - 3x_2]^2 \times (18 - 32x_1 + 12x_1^2 + 48x_2 - 36x_1x_2 + 27x_2^2)$	2	$[-2, 2]$	3
	$F_{17}(x) = -\sum_{i=1}^4 c_i \exp\left(-\sum_{j=1}^3 a_{ij}(x_j - p_{ij})^2\right)$	3	$[1, 3]$	-3.86
	$F_{18}(x) = -\sum_{i=1}^4 c_i \exp\left(-\sum_{j=1}^6 a_{ij}(x_j - p_{ij})^2\right)$	6	$[0, 1]$	-3.32
固定维度测试函数	$F_{19}(x) = -\sum_{i=1}^5 [(X - a_i)(X - a_i)^T + c_i]^{-1}$	4	$[0, 10]$	-10.153 2
	$F_{20}(x) = -\sum_{i=1}^{10} [(X - a_i)(X - a_i)^T + c_i]^{-1}$	4	$[0, 10]$	-10.536 3

表2 $F_1 \sim F_7$ 算法测试结果比较

Table 2 Results comparison among test functions $F_1 \sim F_7$ of algorithms

测试函数	统计值	WOA	IWOA	VCS	BSO	PSO	SSA	GSA	CML-WOA
F_1	最优值	4.01×10^{-116}	1.30×10^{-118}	7.21×10^{-121}	1.65×10^{-53}	2.73×10^{-26}	0	2.57×10^{-19}	0
	平均值	2.56×10^{-99}	2.19×10^{-106}	2.93×10^{-106}	1.11×10^{-25}	1.60×10^{-39}	9.94×10^{-100}	1.31×10^{-18}	0
	标准差	1.37×10^{-98}	8.11×10^{-106}	1.37×10^{-105}	5.19×10^{-25}	3.17×10^{-39}	5.32×10^{-99}	6.82×10^{-19}	0
	耗时/s	0.788 1	0.779 3	3.336 5	3.837 1	2.662 3	4.377 1	8.332 5	3.013 5
F_2	最优值	1.38×10^{-72}	1.25×10^{-76}	3.22×10^{-59}	8.71×10^{-11}	2.07×10^{-22}	2.50×10^{-76}	1.16×10^{-9}	5.90×10^{-294}
	平均值	5.92×10^{-65}	1.50×10^{-69}	1.02×10^{-52}	7.58×10^{-3}	2.86×10^{-21}	6.25×10^{-53}	2.36×10^{-9}	2.60×10^{-160}
	标准差	1.84×10^{-64}	8.38×10^{-69}	5.12×10^{-52}	1.17×10^{-2}	3.43×10^{-21}	3.40×10^{-52}	7.60×10^{-10}	0
	耗时/s	1.063 2	1.061 5	3.645 4	4.700 3	2.868 0	4.773 7	8.296 9	3.305 5
F_3	最优值	1.04×10^{-44}	4.01×10^{-28}	4.42×10^{-111}	7.70×10^{-54}	3.98×10^{-33}	0	3.02×10^{-19}	0
	平均值	2.22×10^{-4}	1.33×10^{-4}	8.68×10^{-97}	2.98×10^{-8}	2.47×10^{-30}	2.99×10^{-58}	1.76×10^{-18}	3.55×10^{-297}
	标准差	8.10×10^{-4}	5.71×10^{-4}	3.67×10^{-96}	1.63×10^{-7}	3.92×10^{-30}	9.64×10^{-58}	1.31×10^{-18}	0
	耗时/s	1.898 1	1.980 2	4.532 4	7.302 0	3.707 5	6.210 0	9.224 7	4.354 4
F_4	最优值	6.60×10^{-2}	5.77×10^{-8}	1.09×10^{-57}	8.95×10^{-16}	3.07×10^{-20}	3.24×10^{-242}	3.81×10^{-10}	0
	平均值	2.13×10^1	8.78	1.10×10^{-51}	3.54×10^{-2}	8.17×10^{-19}	1.75×10^{-43}	8.45×10^{-10}	0
	标准差	1.31×10^1	9.07	5.31×10^{-51}	1.44×10^{-1}	6.12×10^{-19}	6.69×10^{-43}	2.59×10^{-10}	0
	耗时/s	0.784 8	0.772 6	3.434 0	3.774 7	2.624 0	4.318 3	8.101 3	3.204 7
F_5	最优值	9.96×10^{-1}	9.76×10^{-1}	1.94×10^{-13}	0	4.58×10^{-5}	1.70×10^{-11}	1.01	9.53×10^{-4}
	平均值	2.04×10^1	5.49×10^1	8.86×10^{-9}	1.91	2.22	1.51×10^{-5}	7.43	2.93×10^{-1}
	标准差	9.19×10^1	2.78×10^2	2.74×10^{-8}	2.02	4.91	3.90×10^{-5}	3.21×10^1	1.02
	耗时/s	1.176 1	1.170 9	3.809 7	5.174 9	2.982 4	4.910 2	8.335 2	5.534 0
F_6	最优值	2.26×10^{-3}	4.5×10^{-3}	3.89×10^{-25}	0	0	0	4.50×10^{19}	1.84×10^{-32}
	平均值	3.38×10^{-3}	1.16×10^{-1}	4.18×10^{-16}	8.32×10^{-21}	0	0	1.15×10^{-18}	8.60×10^{-3}
	标准差	4.00×10^{-2}	1.14×10^{-1}	2.13×10^{-15}	4.42×10^{-20}	0	0	6.32×10^{-19}	1.89×10^{-2}
	耗时/s	0.794 4	0.776 5	3.312 0	3.932 3	2.619 2	4.354 3	8.082 8	3.476 1
F_7	最优值	5.12×10^{-5}	1.09×10^{-5}	1.19×10^{-5}	1.98×10^{-4}	9.60×10^{-5}	3.50×10^{-5}	3.62×10^{-4}	4.63×10^{-6}
	平均值	2.42×10^{-3}	2.04×10^{-3}	3.42×10^{-4}	1.27×10^{-3}	7.21×10^{-4}	6.46×10^{-4}	5.38×10^{-3}	2.47×10^{-4}
	标准差	2.82×10^{-3}	3.27×10^{-3}	2.69×10^{-4}	1.11×10^{-3}	5.28×10^{-4}	3.57×10^{-4}	3.44×10^{-3}	1.93×10^{-4}
	耗时/s	1.130 0	1.180 0	3.691 2	4.844 1	2.927 3	4.906 7	8.551 8	3.478 1

表3 $F_8 \sim F_{12}$ 算法测试结果比较

Table 3 Results comparison among test functions $F_8 \sim F_{12}$ of algorithms

测试函数	统计值	WOA	IWOA	VCS	BSO	PSO	SSA	GSA	CML-WOA
F_8	最优值	0	0	0	0	0	0	0	0
	平均值	0	2.36×10^{-16}	0	5.67	4.64×10^{-1}	0	2.08	0
	标准差	0	1.29×10^{-15}	0	4.34	5.04×10^{-1}	0	1.59	0
	耗时/s	0.824 4	0.812 5	3.379 4	4.112 4	2.686 8	4.472 2	7.798 9	3.097 1
F_9	最优值	8.88×10^{-16}	8.88×10^{-16}	8.88×10^{-16}	4.44×10^{-15}	8.88×10^{-16}	8.88×10^{-16}	1.09×10^{-9}	8.88×10^{-16}
	平均值	3.84×10^{-15}	3.37×10^{-15}	8.88×10^{-16}	6.80×10^{-1}	3.01×10^{-15}	8.88×10^{-16}	2.29×10^{-9}	8.88×10^{-16}
	标准差	2.30×10^{-15}	2.31×10^{-15}	0	8.56×10^{-1}	1.77×10^{-15}	8.88×10^{-16}	6.82×10^{-10}	0
	耗时/s	1.139 8	1.118 9	3.660 8	5.039 3	2.990 2	4.954 5	8.690 0	3.366 7
F_{10}	最优值	0	0	0	2.71×10^{-2}	0	0	9.36×10^{-2}	0
	平均值	4.13×10^{-2}	4.47×10^{-2}	0	1.75×10^{-1}	3.88×10^{-2}	0	9.05×10^{-1}	0
	标准差	9.08×10^{-2}	1.12×10^{-1}	0	1.03×10^{-1}	2.67×10^{-2}	0	6.32×10^{-1}	0
	耗时/s	1.396 5	1.374 9	4.066 0	5.773 5	3.227 2	5.491 3	8.750 0	3.699 0
F_{11}	最优值	2.41×10^{-3}	5.57×10^{-3}	1.61×10^{-24}	9.42×10^{-32}	9.42×10^{-32}	9.42×10^{-32}	1.07×10^{-20}	9.42×10^{-32}
	平均值	1.61×10^{-1}	2.27×10^{-1}	4.91×10^{-18}	2.44×10^{-1}	9.42×10^{-32}	9.42×10^{-32}	5.29×10^{-20}	4.25×10^{-5}
	标准差	3.86×10^{-1}	4.19×10^{-1}	2.14×10^{-17}	1.02	3.34×10^{-47}	3.53×10^{-34}	3.56×10^{-20}	2.97×10^{-4}
	耗时/s	3.155 6	3.140 2	5.822 1	10.999 5	4.959 2	8.483 5	10.533 0	6.326 9
F_{12}	最优值	3.35×10^{-3}	2.29×10^{-2}	3.02×10^{-23}	1.34×10^{-32}	3.44×10^{-32}	1.34×10^{-32}	2.74×10^{-20}	1.34×10^{-32}
	平均值	9.82×10^{-2}	1.71×10^{-1}	1.48×10^{-17}	3.84×10^{-3}	2.21×10^{-3}	1.38×10^{-32}	1.57×10^{-19}	1.34×10^{-32}
	标准差	6.99×10^{-2}	1.11×10^{-1}	4.88×10^{-17}	1.04×10^{-2}	4.20×10^{-3}	1.25×10^{-33}	1.05×10^{-19}	5.56×10^{-48}
	耗时/s	3.149 3	3.145 4	5.768 7	11.117 3	4.993 7	8.603 5	10.547 6	6.234 5

表 4 $F_{13} \sim F_{20}$ 算法测试结果比较

Table 4 Results comparison among test functions $F_{13} \sim F_{20}$ of algorithms

测试函数	统计值	WOA	IWOA	VCS	BSO	PSO	SSA	GSA	CML-WOA
F_{13}	最优值	9.98×10^{-1}	9.98×10^{-1}	9.98×10^{-1}	9.98×10^{-1}	9.98×10^{-1}	9.98×10^{-1}	9.98×10^{-1}	9.98×10^{-1}
	平均值	5.89	5.09	1.42	3.35	9.98×10^{-1}	5.22	5.2	6.01
	标准差	4.93	3.95	2.13	3.45	5.82×10^{-17}	5.49	3.79	5.12
	耗时/s	15.793 1	15.603 5	19.221 3	48.949 5	17.548 9	30.023 6	22.608 2	16.261 7
F_{14}	最优值	3.49×10^{-4}	4.58×10^{-4}	3.07×10^{-4}	3.07×10^{-4}	3.07×10^{-4}	3.07×10^{-4}	1.88×10^{-3}	3.11×10^{-4}
	平均值	3.59×10^{-3}	5.08×10^{-3}	5.65×10^{-4}	7.71×10^{-3}	5.95×10^{-4}	3.37×10^{-4}	5.51×10^{-3}	4.08×10^{-4}
	标准差	4.70×10^{-3}	3.07×10^{-4}	3.45×10^{-4}	1.92×10^{-2}	3.88×10^{-4}	8.26×10^{-5}	3.97×10^{-3}	3.31×10^{-4}
	耗时/s	0.974 8	0.953 1	3.491 5	4.506 8	2.796 9	4.593 2	8.054 0	3.124 3
F_{15}	最优值	-1.03	-1.03	-1.03	-1.03	-1.03	-1.03	-1.03	-1.03
	平均值	-1.03	-1.03	-1.03	-1.03	-1.03	-1.03	-1.03	-1.03
	标准差	3.70×10^{-9}	7.35×10^{-10}	6.04×10^{-16}	5.92×10^{-16}	6.45×10^{-16}	5.04×10^{-16}	4.72×10^{-16}	3.06×10^{-16}
	耗时/s	0.651 0	0.573 5	3.118 7	3.270 6	2.408 0	4.057 0	6.912 4	3.552 3
F_{16}	最优值	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00
	平均值	1.21×10^1	1.28×10^1	3.00	3.00	3.00	3.00	3.00	3.00
	标准差	1.93×10^1	1.77×10^1	1.716×10^{-15}	2.28×10^{-15}	1.56×10^{-15}	1.37×10^{-15}	3.94×10^{-15}	1.06×10^{-15}
	耗时/s	0.443 3	0.428 9	2.929 4	2.821 0	2.297 8	3.679 1	6.846 8	2.344 8
F_{17}	最优值	-3.86	-3.86	-3.86	-3.71	-3.86	-3.86	-3.86	-3.86
	平均值	-3.75	-3.76	-3.86	-3.18	-3.86	-3.86	-3.86	-3.86
	标准差	1.02×10^{-1}	1.54×10^{-1}	2.69×10^{-15}	4.20×10^{-1}	2.68×10^{-15}	2.32×10^{-15}	2.31×10^{-15}	2.17×10^{-15}
	耗时/s	1.716 9	1.714 9	4.247 6	6.941 1	3.571 1	6.145 1	8.411 6	3.473 2
F_{18}	最优值	-3.27	-3.14	-3.32	-2.73	-3.32	-3.32	-3.32	-3.32
	平均值	-3.01	-2.70	-3.29	-1.35	-3.27	-3.27	-3.31	-3.30
	标准差	1.77×10^{-1}	4.83×10^{-1}	5.34×10^{-2}	7.33×10^{-1}	5.83×10^{-2}	5.92×10^{-2}	2.22×10^{-2}	5.79×10^{-2}
	耗时/s	1.773 1	1.170 8	4.322 7	6.934 4	3.630 3	6.176 7	9.296 6	4.463 4
F_{19}	最优值	-9.81	-10.12	-10.15	-1.72	-10.15	-10.15	-10.15	-10.15
	平均值	-5.86	-4.71	-10.15	-6.8×10^{-1}	-5.97	-8.18	-6.94	-10.15
	标准差	2.30	2.42	6.56×10^{-15}	3.38×10^{-1}	3.55	2.49	3.72	6.35×10^{-15}
	耗时/s	3.210 0	3.190 9	5.737 1	11.526 2	5.126 5	8.617 4	10.264 8	5.079 3
F_{20}	最优值	-10.11	-10.42	-10.53	-4.57	-10.53	-10.53	-10.53	-10.53
	平均值	-4.17	-3.68	-10.53	-1.00	-7.79	-8.88	-9.68	-10.53
	标准差	2.37	2.19	2.91×10^{-15}	8.32×10^{-1}	3.71	2.65	2.34	2.51×10^{-15}
	耗时/s	6.000 4	6.001 4	8.826 4	19.921 0	7.758 9	13.714 2	12.980 1	12.753 2

为进一步分析 8 种算法的寻优能力,根据表 2 ~ 表 4 的均值对算法在各个测试函数中的结果进行比较排序,若相等则比较标准差,表 5 展示了各算法的平均排序结果。CML-WOA 排序第 1, 寻优性能在 8 个算法中最强,且明显优于 WOA。其余算法按照性能由好到差排序为:SSA、VCS、PSO、GSA、BSO、IWOA 和 WOA。为更加直观显示 8 种算法在不同测试函数上的排序结果,采用雷达图将各算法进行绘图,如图 7 所示。图中算法性能曲线所围面积越小,表明算法的排序越靠前,其性能就越好。黑色加粗曲线即为 CML-WOA 算法排序结果曲线,可以直观地看到,CML-WOA 在

F_6 、 F_{11} 和 F_{12} 上性能相对较差,在其他测试函数中均表现较好,且其所围面积最小,说明 CML-WOA 整体上具有最好的寻优性能。

图 8 为 8 种算法独立求解 9 个具有代表性的基准测试函数 30 次所得结果的箱式图,从图中可

表 5 算法性能排序结果

Table 5 Results of sort of algorithm performance

算法	平均排序	算法	平均排序
WOA	5.9	PSO	4.35
IWOA	6.4	SSA	3.1
VCS	3.2	GSA	5
BSO	6.2	CML-WOA	1.85

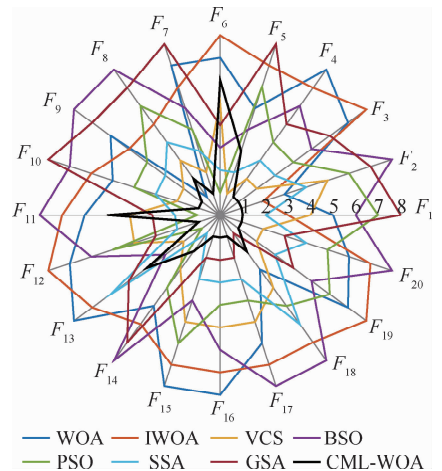


图 7 算法排序雷达图

Fig. 7 Radar chart of sort of algorithms

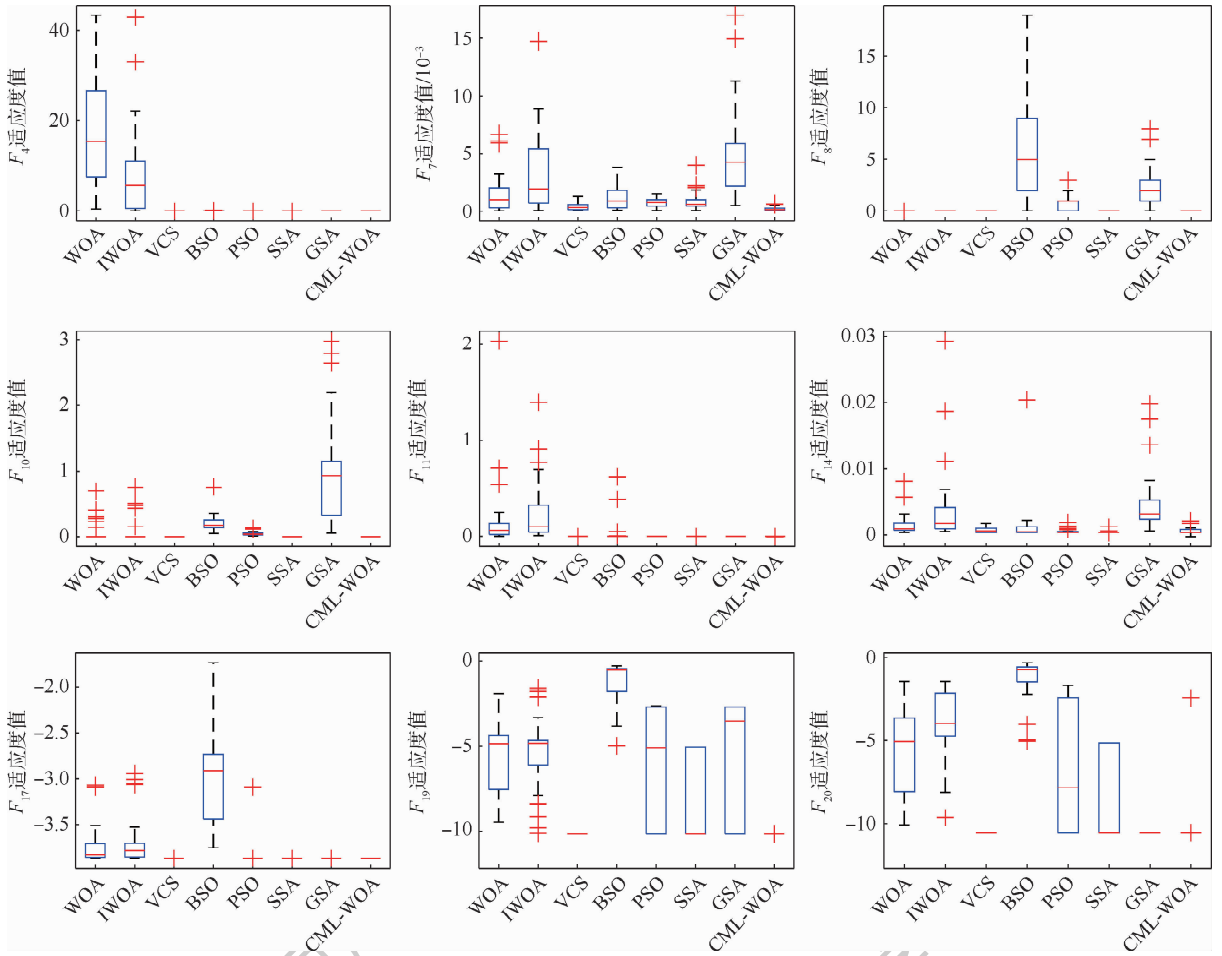


图8 8种算法在9个基准测试函数上的收敛箱式图

Fig.8 Convergence box diagram of eight algorithms on nine benchmark test functions

以得知,在求解 F_4 、 F_8 、 F_{10} 、 F_{11} 和 F_{17} 时,均无异常点出现,说明 CML-WOA 在求解这些函数时性能较稳定;在求解其余函数时,有个别异常点出现,但均少于对比算法,且在求解所有测试函数时,收敛值的分布相比其他对比算法整体上更为集中,明显优于其他对比算法,说明改进的 CML-WOA 算法具有较强的鲁棒性。

通过对 30 次独立运行求解测试函数结果的平均值和标准差进行分析比较,并不能精确分析每次运行的结果,且在运行过程中仍有一定概率出现偶然,致使算法在均值上具有较好表现。因此在统计学层面上来判断不同算法整体结果的显著性区别,采用 Wilcoxon 统计检验。将 8 种算法独立求解 20 个测试函数 30 次得到的结果作为样本,在置信度为 0.05 的条件下进行检验,判断 7 个对比算法所得结果与 CML-WOA 所得结果的显著性区别。Wilcoxon 统计检验结果如表 6 所示,最后一行为检验结果的统计数量。

Wilcoxon 统计检验结果表明:在 20 个基准测试函数中,CML-WOA 的性能在 16 个测试函数上优于 WOA、IWOA 和 GSA,在 10 个测试函数上优

表6 Wilcoxon 统计检验结果

Table 6 Results of statistical tests of Wilcoxon

算法	WOA	IWOA	VCS	BSO	PSO	SSA	GSA
F_1	-	-	-	-	-	-	-
F_2	-	-	-	-	-	-	-
F_3	-	-	-	-	-	-	-
F_4	-	-	-	-	-	-	-
F_5	-	-	+	-	=	=	-
F_6	-	-	=	+	+	+	=
F_7	-	-	-	-	-	-	-
F_8	=	=	=	-	-	-	-
F_9	-	-	=	-	=	-	-
F_{10}	=	=	-	-	-	=	-
F_{11}	-	-	-	-	+	+	-
F_{12}	=	=	=	=	=	-	=
F_{13}	=	=	=	=	+	=	=
F_{14}	-	-	=	=	=	=	-
F_{15}	-	-	-	=	-	-	-
F_{16}	-	-	-	-	-	-	-
F_{17}	-	-	=	-	=	-	-
F_{18}	=	=	+	=	+	+	+
F_{19}	-	-	=	-	-	-	-
F_{20}	-	-	=	-	-	-	-

(-/+)= 16/4/0 16/4/0 10/8/2 14/5/1 11/5/4 13/4/3 16/3/1

注:表中“+”表示该算法性能明显优于 CML-WOA;“=”表明该算法性能与 CML-WOA 无明显差别;“-”表示该算法性能明显劣于 CML-WOA。

于 VCS,在 14 个测试函数上优于 BSO,在 11 个测试函数上优于 PSO,在 13 个测试函数上优于 SSA。综上,CML-WOA 在求解 20 个测试函数时,至少在一半以上的函数优于另外 7 个对比算法,仅在个别测试函数中劣于个别函数。因此基于统计学理论分析,在收敛精度方面,CML-WOA 的寻优性能明显优于 7 种对比算法。

为了进一步阐述 CML-WOA 的收敛性能,8 种算法独立求解 F_1 、 F_3 、 F_4 、 F_8 、 F_{10} 、 F_{14} 、 F_{16} 、 F_{19} 和 F_{20} 各 30 次的平均收敛曲线如图 9 所示。

在求解前 6 个算法时,CML-WOA 有更快的收敛速度和更好的收敛精度,在求解 F_{16} 时,CML-WOA 在前期收敛速度慢于 GSA,但后期收敛速度

超过所有对比算法,先收敛到最优值;在求解 F_{19} 和 F_{20} 时,在前期收敛速度较慢,但后期收敛速度同样超过对比算法,且对比算法收敛精度更好。因此 CML-WOA 相比 WOA 和 IWOA,其寻优性能具有明显提升,具有较强局部最优规避能力和更高的收敛精度与收敛速度。

4.2 航迹规划问题实验

本次仿真实验环境为:AMD R7 4800U 16 GB RAM,Windows10 64 位。航迹规划模型参数设置如下:战场空间为 $x_{\max} = 100$ km, $y_{\max} = 100$ km,航迹节点数 Num 为 20,最大转弯角为 60° ,最大航程为 550 km,UAV 起点为 (10,10) km,目标点为 (95,90) km,威胁源参数如表 7 所示。

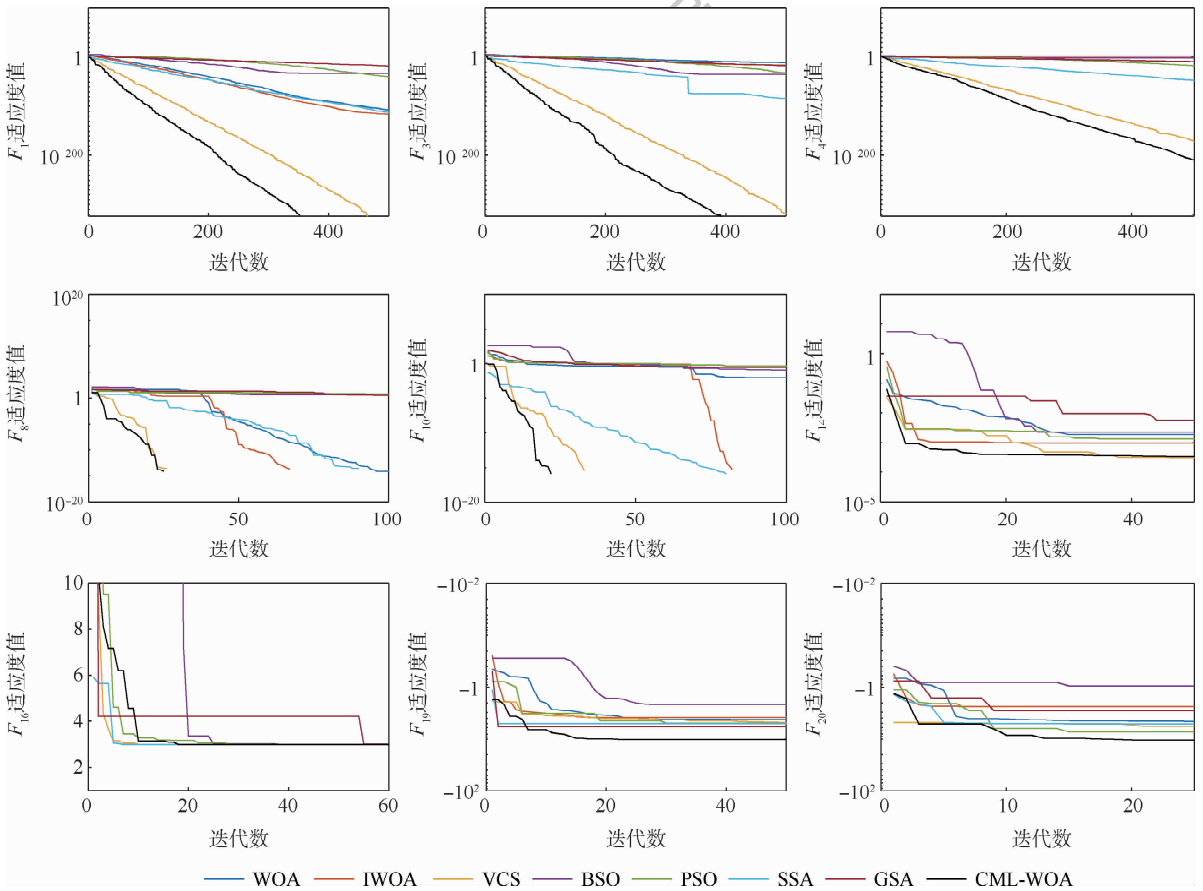


图9 算法收敛曲线

Fig.9 Algorithm convergence curves

表7 威胁源参数

Table 7 Parameters of threat source

威胁源	威胁类型	位置/km	作用半径/km	高度/km
Threat 1	探测雷达	(25,20)	9.8	2.8
Threat 2	地空导弹	(55,50)	11.7	2.9
Threat 3	防空高炮	(85,45)	8.9	3.0
Threat 4	地空导弹	(45,70)	8.0	3.0
Threat 5	探测雷达	(24.5,50)	10.0	2.8
Threat 6	防空高炮	(75,80)	12.0	3.1

分别使用 WOA、IWOA、PSO、GWO 以及 CML-WOA 对静态航迹规划模型进行求解,算法种群规模设置为 100,最大迭代数量为 200,进行 30 次仿真实验。结果如表 8 所示。图 10 ~ 图 14 为 4 种算法求解航迹节点数为 20 的问题结果。

从表 8 可以得知,在 20 维的情况下求解航迹规划模型,本文提出的 CML-WOA 结果中的平均适应值、最小适应值、最大适应值以及标准差均是 5 种算法中最小的,说明本文算法性能明显优于对

表 8 20 维下的规划代价值

Table 8 Planning cost in 20 dimensions

航迹节点数	结果	WOA	IWOA	PSO	GWO	CML-WOA
20	平均适应值	2.31×10^4	2.22×10^4	2.00×10^4	2.01×10^4	8.32×10^1
	标准差	8.84×10^3	7.35×10^3	6.35×10^3	8.21×10^3	2.92
	最大适应值	5.29×10^4	5.16×10^4	5.58×10^4	6.08×10^4	86.237 8
	最小适应值	80.709 8	79.061 2	76.668 6	75.486 7	73.773 7
	成功率	0.65	0.60	0.55	0.60	1.00
	耗时/s	10.477 7	11.153 1	10.606 1	11.082 9	13.256 1

比算法,虽然耗时不是最少,但通过改进策略增加了算法的收敛精度,使得该改进算法在求解航迹规划问题时,成功率达到了百分之百。如图 10 和图 11 所示,CML-WOA 生成航迹优于对比算法生成航迹,能够尽可能选择地形高度较低的区域前进,从而实现低空突防,而其他算法求得的飞行航迹都会通过一个小山坡,不利于 UAV 隐蔽飞行。如图 12 和图 13 所示,各算法在整个规划航迹中,转弯角和俯仰角均远离阈值,说明规划的航迹有利于 UAV 飞行,降低 UAV 飞行条件。图 14 是各算法求解航迹模型代价值的平均收敛曲线图,如图 14 所示,CML-WOA 能够迅速收敛至较优解,且收敛精度优于对比算法。在无人机航迹规划问题中,计算耗时是一个需要考虑的重要因素。表 8 显示 CML-WOA 的时间成本较高,但总的来说,CML-WOA 的运行时间和其他算法差异并不大,并且,多花费的时间既提高了求解精度,又保证

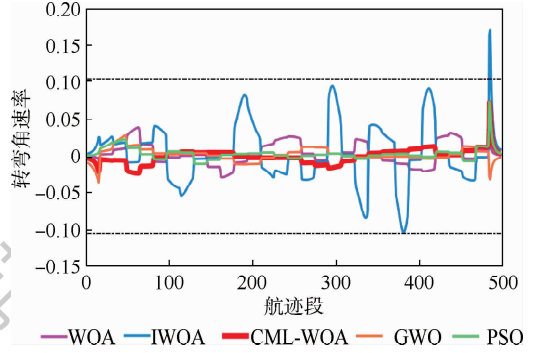


图 12 转弯角的变化率

Fig. 12 Change rate of turning angle

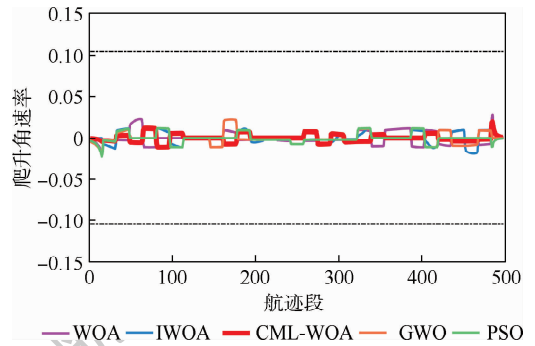


图 13 爬升角的变化率

Fig. 13 Change rate of climbing angle

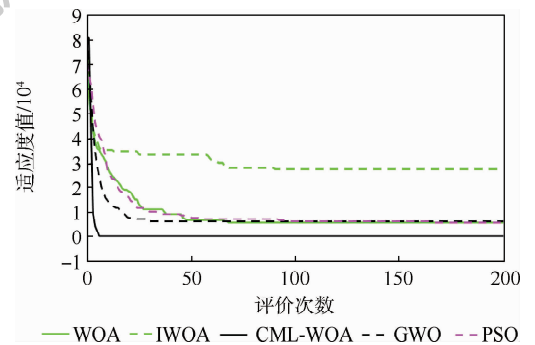


图 14 航迹代价收敛曲线

Fig. 14 Convergence curves of path cost

了求解质量,同时,本文中涉及的无人机航迹规划问题为静态航迹规划问题,对于实时性的要求没有动态航迹规划要求高,本文工程运用的主要目的是获得最优或近似最优的飞行航迹。因此,虽然 CML-WOA 的计算耗时有所增加,但其求解精度和成功率得到了明显提高,表明了本文提出改

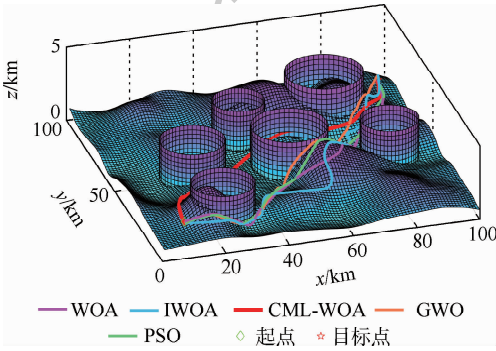


图 10 三维航迹图

Fig. 10 Three-dimensional path map

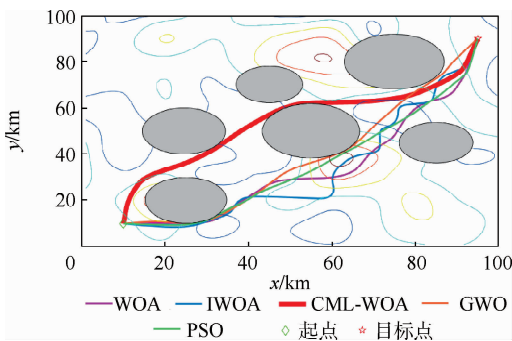


图 11 二维等高线航迹图

Fig. 11 Two-dimensional contour path map

进策略的有效性,证明了本文提出的基于 CML-WOA 的 UAV 航迹规划方法的有效性与优越性。

5 结 论

1) CML-WOA 寻优性能明显提升,在 20 个基准测试函数性能测试中,CML-WOA 的性能在 16 个测试函数上优于 WOA、IWOA 和 GSA,在 10 个测试函数上优于 VCS,在 14 个测试函数上优于 BSO,在 11 个测试函数上优于 PSO,在 13 个测试函数上优于 SSA。

2) CML-WOA 寻优精度高,引入立方映射和自适应正弦策略,丰富种群多样性,利用多精英搜索策略,有效跳出局部最优,增强算法全局搜索能力。

3) CML-WOA 具有较强的稳定性,由测试函数标准差和箱式图可知,收敛值的分布更为集中,寻优结果稳定。

4) CML-WOA 收敛速度较快,从各算法收敛曲线图可知,CML-WOA 的收敛效果优于其他对比算法。

5) CML-WOA 能够求解出安全可行的飞行航迹。

参考文献 (References)

- [1] 胡中华,赵敏,姚敏,等. 无人机航迹规划技术研究及发展趋势[J]. 航空电子技术,2009,40(2):24-29.
HU Z H,ZHAO M,YAO M, et al. Research and development trend of path planning for unmanned air vehicle[J]. Avionics Technology,2009,40(2):24-29 (in Chinese).
- [2] 符小卫,高晓光. 一种无人机组路规划算法研究[J]. 系统仿真学报,2004,16(1):20-21.
FU X W,GAO X G. Study on a kind of path planning algorithm for UAV[J]. Acta Simulata Systematica Sinica,2004,16(1):20-21 (in Chinese).
- [3] 湛海云,陈华胄,刘强. 基于改进人工势场法的多无人机三维编队路径规划[J]. 系统仿真学报,2020,32(3):414-420.
CHEN H Y,CHEN H Z,LIU Q. Multi-UAV 3D formation path planning based on improved artificial potential field[J]. Journal of System Simulation,2020,32(3):414-420 (in Chinese).
- [4] 单文昭,崔乃刚,黄蓓,等. 基于 PSO-HJ 算法的多无人机协同航迹规划方法[J]. 中国惯性技术学报,2020,28(1):122-128.
SHAN W Z,CUI N G,HUANG B, et al. Multiple UAV cooperative path planning based on PSO-HJ method[J]. Journal of Chinese Inertial Technology,2020,28(1):122-128 (in Chinese).
- [5] 李文广,胡永江,庞强伟,等. 基于改进遗传算法的多无人机协同侦察航迹规划[J]. 中国惯性技术学报,2020,28(2):248-255.
LI W G,HU Y J,PANG Q W, et al. Track planning of multi-UAV cooperative reconnaissance based on improved genetic al-

- gorithm[J]. Journal of Chinese Inertial Technology,2020,28(2):248-255 (in Chinese).
- [6] YU X B,LI C L,ZHOU J F. A constrained differential evolution algorithm to solve UAV path planning in disaster scenarios[J]. Knowledge-Based Systems,2020,204:106209.
- [7] WANG X F,ZHAO H,HAN T, et al. A grey wolf optimizer using Gaussian estimation of distribution and its application in the multi-UAV multi-target urban tracking problem[J]. Applied Soft Computing,2019,78:240-260.
- [8] 钱洲元,雷明. 面向无人机航迹规划的自适应乌贼算法[J]. 哈尔滨工业大学学报,2019,51(10):37-46.
QIAN Z Y,LEI M. Adaptive cuttlefish algorithm for UAV path planning[J]. Journal of Harbin Institute of Technology,2019,51(10):37-46 (in Chinese).
- [9] 魏江,王建军,王健,等. 基于改进蚁群算法的三维航迹规划[J]. 计算机工程与应用,2020,56(17):217-223.
WEI J,WANG J J,WANG J, et al. 3D path planning based on improved ant colony algorithm[J]. Computer Engineering and Applications,2020,56(17):217-223 (in Chinese).
- [10] MIRJALILI S,LEWIS A. The whale optimization algorithm[J]. Advances in Engineering Software,2016,95:51-67.
- [11] 崔东文. 鲸鱼优化算法在水库优化调度中的应用[J]. 水利水电科技进展,2017,37(3):72-76.
CUI D W. Application of whale optimization algorithm in reservoir optimal operation[J]. Advances in Science and Technology of Water Resources,2017,37(3):72-76 (in Chinese).
- [12] BHASKAR N,KUMAR P M. Optimal processing of nearest-neighbor user queries in crowdsourcing based on the whale optimization algorithm[J]. Soft Computing,2020,24(17):13037-13050.
- [13] AGRAWAL R K,KAUR B,SHARMA S. Quantum based whale optimization algorithm for wrapper feature selection[J]. Applied Soft Computing,2020,89:106092.
- [14] MOHANAKRISHNAN U,RAMAKRISHNAN B. MCTRP: An energy efficient tree routing protocol for vehicular ad hoc network using genetic whale optimization algorithm[J]. Wireless Personal Communications,2020,110(1):185-206.
- [15] LONG W,WU T B,JIAO J J, et al. Refraction-learning-based whale optimization algorithm for high-dimensional problems and parameter estimation of PV model[J]. Engineering Applications of Artificial Intelligence,2020,89:103457.
- [16] PHAM Q V,MIRJALILI S,KUMAR N, et al. Whale optimization algorithm with applications to resource allocation in wireless networks[J]. IEEE Transactions on Vehicular Technology,2020,69(4):4285-4297.
- [17] ZHANG H P,WANG H L,LI N, et al. Time-optimal memetic whale optimization algorithm for hypersonic vehicle reentry trajectory optimization with no-fly zones[J]. Neural Computing and Applications,2020,32(7):2735-2749.
- [18] 吴坤,谭劭昌. 基于改进鲸鱼优化算法的无人机航路规划[J]. 航空学报,2020,41(S2):724286.
WU K,TAN S C. Path planning of UAVs based on improved whale optimization algorithm[J]. Acta Aeronautica et Astronautica Sinica,2020,41(S2):724286 (in Chinese).
- [19] ABDEL-BASSET M,CHANG V,MOHAMED R. HSMA_WOA:

- A hybrid novel slime mould algorithm with whale optimization algorithm for tackling the image segmentation problem of chest X-ray images[J]. *Applied Soft Computing*, 2020, 95:106642.
- [20] 覃溪, 龙文. 基于随机差分变异的改进鲸鱼优化算法[J]. *中国科技论文*, 2018, 13(8):937-942.
- QIN X, LONG W. An improved whale optimization algorithm based on stochastic differential mutation[J]. *China Sciencepaper*, 2018, 13(8):937-942 (in Chinese).
- [21] 孔芝, 杨青峰, 赵杰, 等. 基于自适应调整权重和搜索策略的鲸鱼优化算法[J]. *东北大学学报(自然科学版)*, 2020, 41(1):35-43.
- KONG Z, YANG Q F, ZHAO J, et al. Adaptive adjustment of weights and search strategies-based whale optimization algorithm[J]. *Journal of Northeastern University (Natural Science)*, 2020, 41(1):35-43 (in Chinese).
- [22] MIRJALILI S. SCA: A sine cosine algorithm for solving optimization problems[J]. *Knowledge-Based Systems*, 2016, 96:120-133.
- [23] LIU Z Z, WANG Y, YANG S X, et al. An adaptive framework to tune the coordinate systems in nature-inspired optimization algorithms[J]. *IEEE Transactions on Cybernetics*, 2019, 49(4):1403-1416.
- [24] 郑昌文. 飞行器航迹规划方法研究[D]. 武汉: 华中科技大学, 2003:99-102.
- ZHENG C W. Research on route planning for air vehicles[D]. Wuhan: Huazhong University of Science and Technology, 2003:99-102 (in Chinese).
- [25] POLI R, KENNEDY J, BLACKWELL T. Particle swarm optimization[J]. *Swarm Intelligence*, 2007, 1(1):33-57.
- [26] WANG T T, YANG L, LIU Q. Beetle swarm optimization algorithm: theory and application[EB/OL]. (2018-08-01)[2020-10-14]. <https://arxiv.org/abs/1808.00206>.
- [27] XUE J K, SHEN B. A novel swarm intelligence optimization approach: sparrow search algorithm[J]. *Systems Science & Control Engineering*, 2020, 8(1):22-34.
- [28] LI M D, ZHAO H, WENG X W, et al. A novel nature-inspired algorithm for optimization: Virus colony search[J]. *Advances in Engineering Software*, 2016, 92:65-88.
- [29] RASHEDI E, NEZAMABADI-POUR H, SARYAZDI S. GSA: A gravitational search algorithm[J]. *Information Sciences*, 2009, 179(13):2232-2248.

Chaotic multi-leader whale optimization algorithm

TANG Andi¹, HAN Tong^{2,*}, XU Dengwu³, XIE Lei¹

(1. Graduate School, Air Force Engineering University, Xi'an 710038, China;

2. Aeronautics Engineering Institute, Air Force Engineering University, Xi'an 710038, China; 3. Team 94855 of PLA, Quzhou 324000, China)

Abstract: Aimed at the path planning problem of Unmanned Aerial Vehicle (UAV), a path planning method based on a Chaotic Multi-Leader Whale Optimization Algorithm (CML-WOA) is proposed. In the known flight environment, 3D model of flight area and a flight path cost model are established. By introducing penalty functions, the constrained 3D path planning problem is transformed into an unconstrained multi-dimensional function optimization problem, which is solved using CML-WOA to obtain the optimal flight path. To overcome the defect that the WOA is easy to fall into local optimum, this paper introduces the cubic mapping chaos operator to improve the initial population and enhance the population diversity. And the Sine Cosine Algorithm (SCA) is integrated through an adaptive framework. A multi-leader search strategy is used to effectively improve the algorithm exploitation and exploration capability. Finally, a greedy strategy is used to ensure the convergence efficiency. The proposed improved CML-WOA is tested and validated by 20 benchmark functions test and path planning simulation experiments. The results show that the algorithm has significantly improved performance compared to other algorithms, with strong local optimal avoidance capability, higher convergence accuracy and convergence speed. Also, it is able to provide stable and fast planning of safe and feasible flight path with minimum cost and satisfying constraints.

Keywords: path planning; Whale Optimization Algorithm (WOA); chaotic operator; multi-leader search; Sine Cosine Algorithm (SCA); global optimization

Received: 2020-10-15; **Accepted:** 2021-01-22; **Published online:** 2021-02-01 17:18

URL: kns.cnki.net/kcms/detail/11.2625.V.20210201.1547.004.html

Foundation items: Natural Science Foundation of Shaanxi Province (2020JQ-481, 2021JM224); Aeronautical Science Foundation of China (201951096002)

* **Corresponding author.** E-mail: 418932433@qq.com

http://bhxb.buaa.edu.cn jbuua@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2020.0676

带扰动块的细长旋成体背部绕流数值模拟



刘克奇, 李国辉*, 张浩南, 张宴嘉

(空军航空大学 航空作战勤务学院, 长春 130000)

摘 要: 为了更好地了解细长旋成体背部绕流非对称涡的形成机理, 研究了头部带扰动块对细长旋成体背部绕流结构的影响, 通过雷诺平均 Navier-Stokes (RANS) 法对细长旋成体模型在攻角 $5^\circ \sim 60^\circ$ 范围内进行仿真。在攻角分别为 20° 和 30° 时对是否添加扰动块模型进行对比, 分析了不同截面绕流沿轴向位置的发展, 提出了验证拓扑结构的一种方法, 找到了各流态下奇点的位置, 通过涡核位置对模型背部绕流的发展进行了分析。研究表明: 添加已知规则扰动块可以加快各绕流结构间的转换速度, 使非对称涡产生的攻角减小。

关键词: 绕流结构; 细长旋成体; 非对称流动; 鞍点; 拓扑结构

中图分类号: V221; TJ760

文献标志码: A **文章编号:** 1001-5965(2021)07-1495-10

细长旋成体作为未来先进飞机以及导弹弹体的主要外形之一, 在大攻角飞行时, 会产生绕流。随着攻角的变化, 细长旋成体背部绕流旋涡会经历对称流动和非对称流动现象。由于旋涡的非对称流动, 细长旋成体会产生一个侧向力, 这个侧向力大小达到和法向力同一量级甚至超过法向力, 进而产生较大的偏航力矩, 影响飞机以及导弹的飞行状态。Zilliac 等^[1]研究表明旋成体背部旋涡发展经历 4 种流动状态, 即稳定对称流动、不稳定对称流动、双稳态流动以及类卡门涡街(或随机尾迹)流态。侧向力的变化是由不同流态相互转换所引起的^[2], Rizzetta^[3]和 Fagley^[4]等的研究结果显示, 细长旋成体模型在大攻角状态下对模型头部微小几何扰动的敏感性非常高, 使得旋涡由对称流态转变为非对称流态, 但是在具体实验中模型头部的自然扰动如加工精细程度和粗糙度是不可避免的, 所以实验测得的背部旋涡和各项参数数据是不可重复的, 但是在模型头部加上固定的人工扰动, 自然扰动所产生的影响就可

以忽略, 实验结果重复性也较好。贺中等^[5-6]通过实验对旋成体头部微小扰动进行了一系列的研究; 除此之外 Wu 等^[7]采用数值模拟的方法对旋成体模型进行数值分析, 在数值模拟过程中加入扰动^[8]或几何缺陷^[9]是不可或缺的; 也有部分研究人员采用大涡模拟 (LES) 方法^[10]对添加扰动的模型进行研究。Moskovitz 等^[11]研究发现旋成体头部几何微小扰动对模型背部的非对称流态影响特别大; Zhu 等^[12]通过对头部缺陷模型进行实验对旋涡非对称性形成机理有了更进一步的理解。为了实现侧向力的有效控制以及对细长旋成体背部绕流结构形成机理有更深入的认识, 研究头部带微小扰动块对绕流结构的影响具有重要意义。

本文主要工作是对尖拱-圆柱形细长旋成体模型分别在不添加任何人工扰动和加入人工微小几何扰动块 2 种情况下采用雷诺平均 Navier-Stokes (RANS) 法, 在攻角 $5^\circ \leq \alpha \leq 60^\circ$ 范围内对旋成体模型背部绕流结构进行数值模拟。为研

收稿日期: 2020-12-02; 录用日期: 2021-01-30; 网络出版时间: 2021-02-03 11:39

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20210203.0835.001.html

* 通信作者。E-mail: ghleeaaf@sina.com

引用格式: 刘克奇, 李国辉, 张浩南, 等. 带扰动块的细长旋成体背部绕流数值模拟[J]. 北京航空航天大学学报, 2021, 47(7): 1495-1504. LIU K Q, LI G H, ZHANG H N, et al. Numerical simulation of flow around slender body with disturbing block[J]. Journal of Beijing University of Aeronautics and Astronautics, 2021, 47(7): 1495-1504 (in Chinese).

究人工微小几何扰动块对模型背部不同绕流结构的影响,选用攻角 $\alpha = 20^\circ$ 和 $\alpha = 30^\circ$ 的情况,对比分析了2种模型不同截面绕流沿轴向位置的发展,采用截面压力系数分析了侧向力在模型轴向位置上的发展,找到了不同流态的奇点,并通过截面流场中奇点的精确数量、位置和性质,验证了细长旋成体不同绕流结构下的截面流场的拓扑结构,对非对称涡流态的形成与发展进行了分析,结果证明,数值模拟结果与理论分析一致^[13]。

1 数值方法

本文所采用的三维黏性 Navier-Stokes 方程是通过有限体积法求解出来的,主控方程守恒形式为

$$\frac{\partial U}{\partial t} + \frac{\partial E}{\partial x} + \frac{\partial F}{\partial y} + \frac{\partial G}{\partial z} = \frac{\partial E_\mu}{\partial x} + \frac{\partial F_\mu}{\partial y} + \frac{\partial G_\mu}{\partial z} \quad (1)$$

式中: U 为非定常项; E 、 F 、 G 为无黏项; E_μ 、 F_μ 、 G_μ 为黏性项。

本文采用时间平均法对三维 Navier-Stokes 方程进行处理以模拟出湍流流动形式。采用雷诺平均法对控制方程(1)进行时均处理得到 RANS 方程,该方程为封闭 RANS 方程组。选用多种湍流模型进行求解,最终选用对该类流动适应性相对较好的 SST $k-\omega$ 湍流模型^[14],SST $k-\omega$ 湍流模型中湍流动能 k 和比耗散率 ω 的输运方程如下所示:

$$\frac{\partial}{\partial t}(\rho k) + \frac{\partial}{\partial x_i}(\rho k u_i) = \frac{\partial}{\partial x_j} \left(\Gamma_k \frac{\partial k}{\partial x_j} \right) + G_k - Y_k + S_k \quad (2)$$

$$\frac{\partial}{\partial t}(\rho \omega) + \frac{\partial}{\partial x_i}(\rho \omega u_i) = \frac{\partial}{\partial x_j} \left(\Gamma_\omega \frac{\partial \omega}{\partial x_j} \right) + G_\omega - Y_\omega + D_\omega + S_\omega \quad (3)$$

式中: ρ 为密度; u_i 为速度矢量; t 为时间; G_k 为湍流的动能; G_ω 为 ω 的生成项; Γ_k 和 Γ_ω 分别为 k 与 ω 的有效扩散项; Y_k 和 Y_ω 分别为 k 与 ω 的耗散项; D_ω 为正交发散项; S_k 和 S_ω 分别为 k 与 ω 的用户自定义广义源项。

SST $k-\omega$ 湍流模型在靠近物面区相当于 $k-\omega$ 湍流模型,在远离物面区域采用的是标准的 $k-\varepsilon$ 湍流模型。对于三维 Navier-Stokes 方程的空间离散方式,采用二阶中心差分格式对扩散项进行处理,对流项采用三阶 MUSCL 格式,用 PISO 算法对速度与压力进行耦合。

2 模型、边界条件及验证

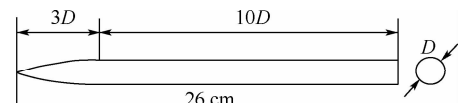
本文所选用的几何模型为尖拱-圆柱形细长旋成体,圆柱段直径 $D = 2 \text{ cm}$ 。细长旋成体模型总体长度为 26 cm ,模型尖拱段头部长细比为 3 ,圆柱段后体长细比为 10 (见图 1(a))。曲线方程为

$$\frac{y(X)}{D} = -0.00262 \left(\frac{X}{D} \right)^3 - 0.0389 \left(\frac{X}{D} \right)^2 + 0.30984 \left(\frac{X}{D} \right) \quad (4)$$

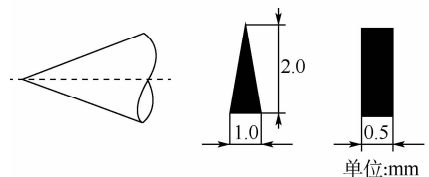
扰动块安装于头部尖拱处,其尺寸以及具体添加位置如图 1 所示,本文对于计算域的选择采用的是圆柱形计算域,其中具体尺寸为尖拱段前端延长至 $6D$ 处,圆柱段与径向均延长至 $15D$ 处(见图 2(a))。整体网格选用分块结构化网格,第 1 层网格高度设置为 $0.0001D$,周向 \times 轴向 \times 径向节点数为 $320 \times 120 \times 210$,总网格数达到 1.8×10^6 (见图 2(b))。将细长旋成体头部、背部以及头部扰动块的网格进行了加密与图 2(b) 保持一致,头部加密网格如图 2(c) 所示。

采用 Fluent 软件进行数值模拟,计算所用边界条件中入口为速度入口,出口为压力出口,外边界采用自由来流无反射条件,自由来流设置为 $0.2Ma$,壁面选用无滑移绝热壁面,此外,湍流度为 0.2% 。

为了验证所选的计算方法以及选用网格的可靠性,运用本文数值模拟方法,与邓学莹等^[15]在风洞试验中得到的试验结果进行对比。在攻角 $\alpha = 50^\circ$, $Ma = 0.3$ 的情况下对尖拱-圆柱形细长旋成体模型进行计算,图 3 为数值模拟和实验所获得的截面侧向力系数 C_y 沿轴向变化的对比,可以得到,侧向力峰值以及变化趋势与风洞试验吻合度较高,证明选用的计算方法可靠性较高。



(a) 整体模型



(b) 扰动块尺寸及安装位置

图 1 尖拱-圆柱形细长旋成体模型

Fig. 1 Pointed arch-cylindrical elongated slender body model

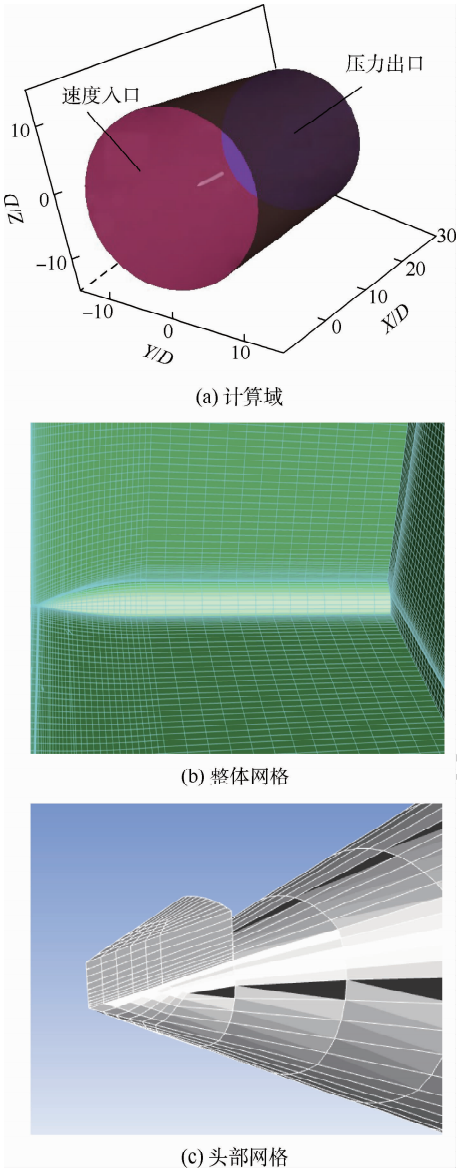


图 2 计算域及网格划分

Fig. 2 Computational domain and mesh division

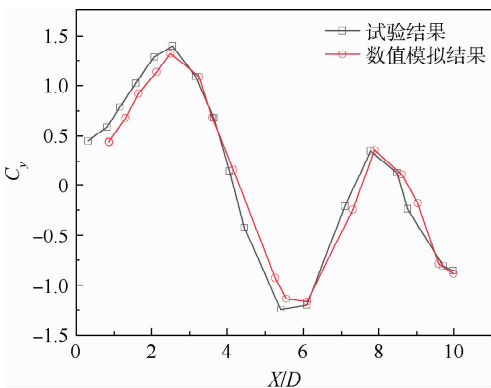


图 3 截面侧向力系数对比

Fig. 3 Comparison of lateral force coefficients

3 数值模拟结果

在本节中,通过对各种参数进行比较,如截面压力分布、截面侧向力、鞍点位置等,定性地

分析了旋涡的对称性;应用壁面切向速度曲线找到了所选截面壁上奇点的位置和数量,应用截面法向量与速度矢量夹角的零点确定空间鞍点的位置和数量,并根据其在空间上的数量、位置和性质,验证了其他学者所提出的细长旋成体的截面拓扑结构;此外,精确定位涡核位置,并绘制其在模型轴向上的发展曲线,分析了其变化规律。

3.1 绕流特性分析

图 4 为扰动模型在不同攻角下部分截面的 x 涡量等值线图。从图中可得出,背部旋涡的非对称性是沿轴向发展的,且随着攻角的增加其非对称性越来越明显。在 $5^\circ \leq \alpha \leq 20^\circ$ 范围内,背部旋涡保持为对称涡状态,非对称性在 $\alpha = 25^\circ$ 时开始出现,但是刚出现时非对称性较小,在模型头部尖拱段时没有出现分离现象,处于附着流态,随截面沿轴向发展开始出现对称涡,继而出现非对称涡,但是刚出现非对称性时两涡核高度相差较小,直到模型圆柱段后体才较为明显。在 $30^\circ \leq \alpha \leq 50^\circ$ 范围内,非对称涡得到了充分的发展,到 $\alpha = 45^\circ$ 时,旋涡表现为高度非对称涡,其非对称性在尖拱段已较为明显,在 $\alpha = 60^\circ$ 时,背部旋涡达到高度非对称,涡飘起,形成的脱落涡速度更快。对于无扰动模型参见文献[16],在此不做叙述。

绘制不同攻角下各截面侧向力系数 C_y 沿模型轴向位置变化曲线(见图 5)和各截面压力系数

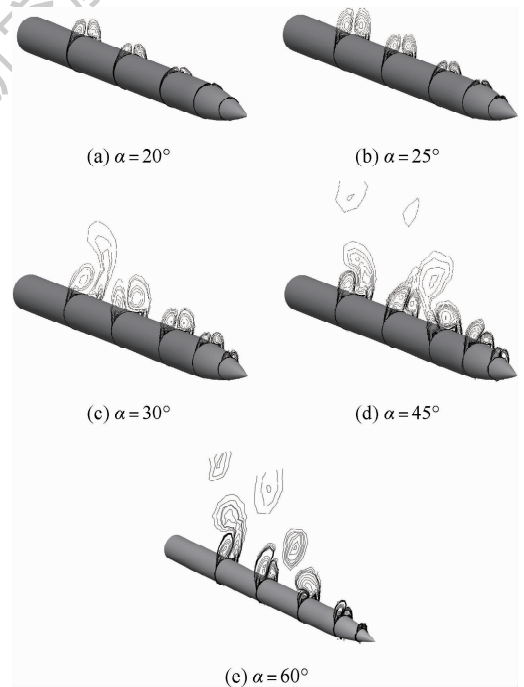


图 4 平均 x 涡量等值线

Fig. 4 Mean x vorticity contour

C_p 沿轴向位置分布曲线(见图6)。压力系数 C_p 和截面侧向力系数 C_y 定义为

$$C_y = \frac{1}{2} \int_0^{2\pi} C_p \sin \theta d\theta \quad (5)$$

$$C_p = (p - p_\infty) / \left(\frac{1}{2} \rho_\infty v_\infty^2 \right) \quad (6)$$

式中: p 为压力; p_∞ 为来流压力; ρ_∞ 为来流密度; v_∞ 为来流速度; θ 为周向角度。其中坐标系的定义与文献[17]保持一致。

图5为2种模型的截面侧向力系数曲线。在 $20^\circ \leq \alpha \leq 40^\circ$ 范围内无扰动模型 C_y 基本为0, 说明此时模型背部旋涡没有出现非对称性, 但是添加扰动的模型在 $\alpha = 25^\circ$ 时, C_y 在 $X/D = 5$ 附近, 即模型圆柱段开始发生改变, 此时背部旋涡表现为非对称涡, 并且随着截面轴向位置后移其非对称性会增加; 在 $30^\circ \leq \alpha \leq 45^\circ$ 范围内, 背部旋涡的非对称性得到充分发展, C_y 沿轴向位置的变化表现为正弦分布, 然而无扰动模型在 $\alpha = 50^\circ$ 时 C_y 才不为0, 在 $\alpha = 60^\circ$ 时其背部旋涡的非对称性得到充分发展。除此之外, 随着攻角的增加, $|C_y|$ 的最大值增大, 这是由于攻角的增加会导致模型背风侧的吸力变大。对于添加扰动块的模型, $\alpha = 25^\circ$ 时, 在 $X/D = 5$ 截面处 $|C_y|$ 发生改变, 但是 $|C_y|$ 在 $X/D = 7.5$ 截面附近达到最大值; $\alpha = 35^\circ$ 时, $|C_y|$ 在 $X/D = 5$ 截面附近达到最大值, 在 $\alpha = 45^\circ$ 时, $|C_y|$ 最大值出现在 $X/D = 3$ 截面处, $\alpha = 60^\circ$ 时, 涡量和体积增大, 壁面也会出现 Tertiary 涡, 其 $|C_y|$ 最大值是所有计算攻角中的最大值, $|C_y|$ 最大值相比 45° 增加了约 23%, 相比最开始出现侧向力增加了约 253%。以 $\alpha = 60^\circ$ 为例, 有扰动 $|C_y|$ 最大值约为无扰动的 2.1 倍, 对侧向力的影响范围约为无扰动的 1.5 倍。上述分析表明, 添加扰动块之后, 使得非对称涡出现的攻角 α 减小, $|C_y|$ 幅值增大及影响范围变大; 攻角的增加会导致 $|C_y|$ 最大值所处截面位置沿轴向头部移动, 意味着模型背部旋涡非对称性出现位置提前。

为更好观察不同流态下旋涡的压力变化情况, 在 $\alpha = 20^\circ$ 和 30° 时, 根据数值模拟结果, 绘制了 $X/D = 1, X/D = 3, X/D = 5$ 和 $X/D = 7$ 这4个截面下的物面压力轴向分布曲线(见图6), 由图6(a)可知, 在攻角 20° 工况下, 无论是否添加扰动块, 旋成体模型 C_p 变化曲线基本保持对称, 与图5中 C_y 基本为0保持一致; 在攻角 30° 工况下, 无扰动模型4个截面的 C_p 仍然保持对称, 但是相对于无扰动模型来说, 添加扰动块的模型在 $X/D = 1$ 和 $X/D = 3$ 两个截面的 C_p 基本保持对称, 直到 $X/D = 5$ 截面,

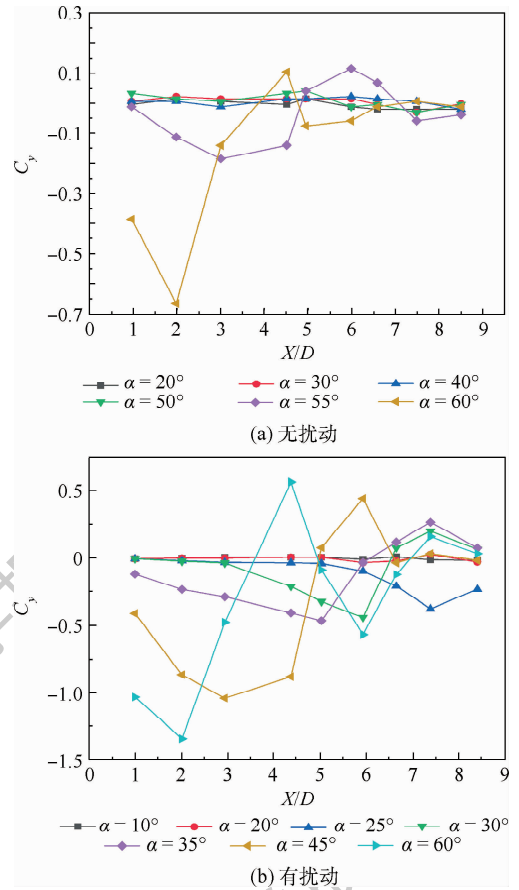


图5 截面侧向力系数曲线

Fig.5 Cross-section lateral force coefficient curves

C_p 呈现非对称分布, 这是因为添加扰动块之后促使两涡产生高度差, 距离模型表面较远的高位涡对模型表面的诱导作用相对于物面附近的低位涡来说较小, 使 $|C_p|$ 较小, 相反, 较低的涡所产生的诱导作用对模型的影响更大 $|C_p|$ 也较大, 此时平衡状态被打破, 进而导致截面侧向力的产生, 表现为非对称涡流态, 并且沿着模型轴向发展, 高位涡不断向上飘起, 形成高低有别的2个涡。综上所述, 添加扰动块之后, 使得非对称涡出现的攻角 α 减小; 当旋成体模型背部开始出现非对称涡后, 伴随着攻角 α 的增加, 非对称涡出现的位置会沿着模型轴向从后体圆柱段逐渐前移到前体尖拱段。

细长旋成体模型在有攻角下会在背部产生非常复杂的旋涡。故本文对比分析头部扰动块对细长旋成体模型背部绕流结构的影响, 根据数值模拟结果, 当 $\alpha = 20^\circ$ 时, 2个模型均处于对称涡状态, 当 $\alpha = 30^\circ$ 时, 带扰动的模型后体的非对称涡已较为明显, 故选用这2个角度进行分析。以有扰动模型在 30° 攻角下为例, 绘制 $X/D = 0.3, X/D = 1, X/D = 3, X/D = 5$ 截面的速度矢量图, 如图7所示。对于旋成体头部尖拱段, 此时附面层没有发生分离, 旋成体背部没有出现旋涡, 故该截面为附

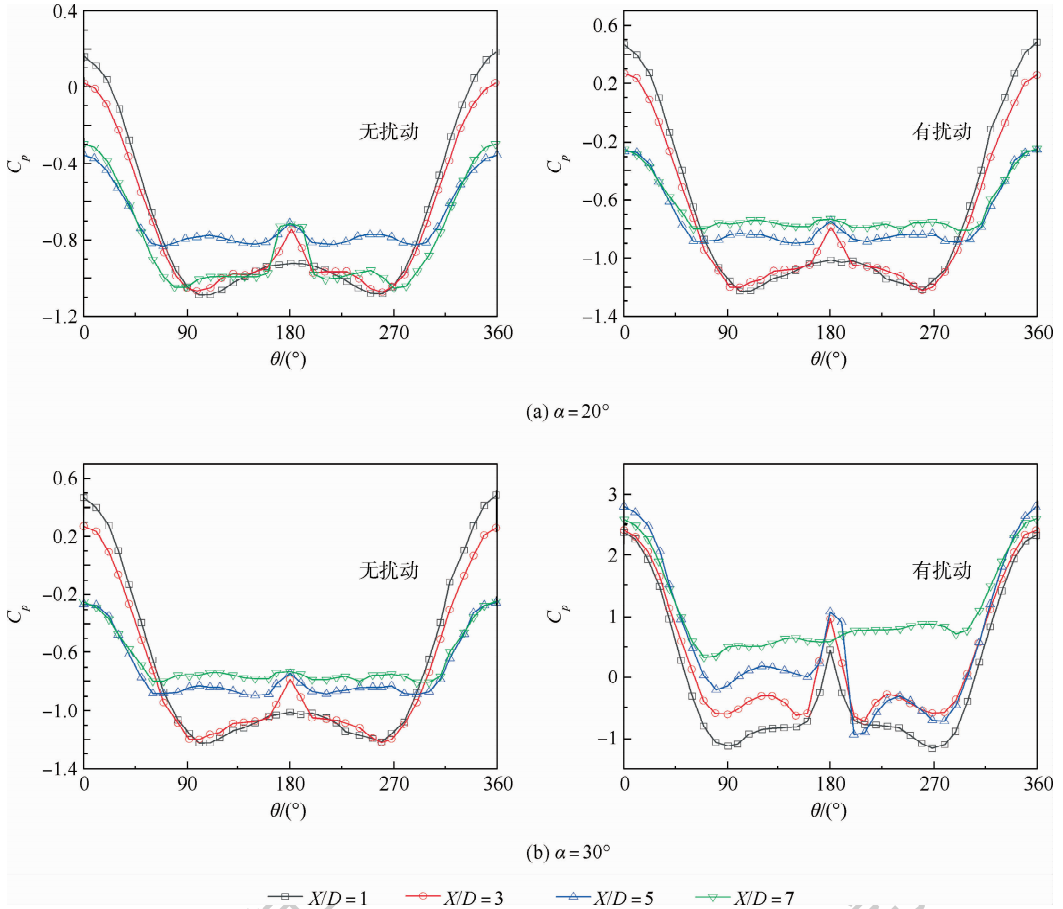


图 6 物面压力轴向分布曲线

Fig. 6 Axial distribution curves of surface pressure

着流态(见图 7(a));随着截面沿模型轴向后移动,模型背部出现逆压,模型背部边界层流动迅速发生分离,此时,由于主流对脱离物面的涡量产生影响,使附面层卷起,故尖拱段的附着流态发生改变,形成集中涡,细长旋成体模型背部刚开始出现对称涡如图 7(b)所示,此时的流动分离现象比较弱,使得该截面产生的对称涡不仅体积小而且强度相对较弱,是由于此时旋成体背部卷入对称涡中的涡量相对较少,因此,在离对称面相对较近的位置会出现再附现象;当模型截面处于一对对称涡状态时,对称涡之间的相互诱导作用处于一种平衡状态,如上文分析,此时侧向力几乎为 0,除此之外,由于两涡位置距离相对较远,它们之间的相互影响几乎可以忽略不计,从而使得这一对对称涡处于一种较为稳定的平衡状态;流动的分离程度随着截面位置继续沿模型轴向往后移动而增强,对称涡的涡量会随着剪切层的持续输入而增加,使其强度和体积变大,同时两涡涡核相互靠近,相互影响作用加大,但是两涡仍为一对对称涡,但随着两涡相互吸引,位置相互靠近,该截面下的对称涡处于不稳定状态(见图 7(c));随着截面沿轴向后移,到模型后部,之前的平衡状态被

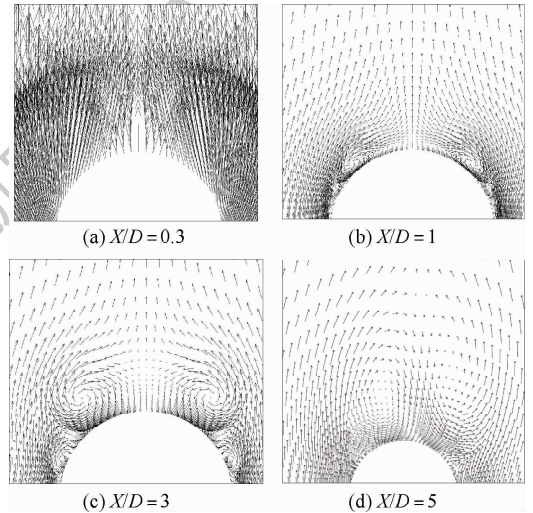


图 7 速度矢量分布

Fig. 7 Velocity vector distribution

打破,产生高度差,故背部旋涡由一对对称涡转变为 一对非对称涡(见图 7(d))。

3.2 壁面鞍点分析

为了进一步的研究细长旋成体背部旋涡拓扑结构,基于 Lowson 和 Ponton^[18]所提出的细长旋成体绕流截面拓扑结构(见图 8),对有/无扰动 2 种模型在攻角 $\alpha = 20^\circ$ 和 30° 下对 $X/D = 0.3$,

$X/D=1$, $X/D=3$ 和 $X/D=5$ 四个截面的流态进行分析。图 9 为壁面处的切向速度曲线, V_θ 为在壁面周向角 θ 处的切向速度。整体来看, 当攻角 $\alpha=30^\circ$ 时, 2 种模型都在 $X/D=0.3$ 截面处于附着流态, 在 $X/D=1$ 截面处于稳定对称涡流态, 在 $X/D=3$ 截面处于不稳定对称涡流态, 但是添加扰动后在 $\theta=180^\circ$ 处切向速度斜率的绝对值会变小, 在 $X/D=5$ 截面有扰动处于非对称涡流态但无扰动却没有, 当 $\alpha=20^\circ$ 时 $X/D=1$ 截面有扰动处于稳定对称涡流态而无扰动时仍处于附着流态, 在 $X/D=3$ 截面处有扰动为不稳定对称涡流态, 无扰动为稳定对称涡流态, 这说明添加扰动会加快结构转换, 下面对此进行详细分析。

对于壁面上奇点, 通过绘制各截面的切向速度变化曲线(见图 9), 找到了图 8 所示拓扑结构中物面上奇点较为精确的位置, 在 $X/D=0.3$ 截面处(见图 9(a)) 无论是否添加扰动, 此时都存在 2 个切向速度为 0 的点, 分别位于 $\theta=0^\circ$ 和 $\theta=180^\circ$ 处, 因为壁面是不可穿透的, 故这 2 点为奇点, 同时根据其流动方向可以确定这 2 点分别为再附型半鞍点 R_{h1} (速度斜率为负) 和分离型半鞍点 S_{h1} (速度斜率为正); 在 $X/D=1$ 截面(见图 9(b)), 当 $\alpha=20^\circ$ 时, 无扰动模型仍有 2 个奇点, 其位置和性质并未发生改变, 但是在 180° 处斜率变小, 表明此时鞍点性质有发生改变的趋势, 对于其他 3 种工况, 各有 6 个奇点, 以添加扰动模型 $\alpha=30^\circ$ 为例, 除上述 2 个奇点外, 还有再附型半鞍点 R_{h2} 和 R_{h3} ($\theta=165^\circ$ 与 $\theta=195^\circ$) 及分离型半鞍点 S_{h2} 和 S_{h3} ($\theta=85^\circ$ 与 $\theta=275^\circ$), 对比无扰动 $\alpha=30^\circ$, 2 个再附型半鞍点位置更靠近 180° , 说明添加扰动块会加快鞍点的转换, 无论是否添加扰动, 攻角的增加也会使鞍点转换速度变快; 在 $X/D=3$ 截面处(见图 9(c)), 只有无扰动模型 $\alpha=20^\circ$ 时有 6 个奇点, 其余 3 个工况均有 4 个奇点, 以有扰动 $\alpha=30^\circ$ 为例, 分别为再附型半鞍点 R_{h1} 和 R_{h2} ($\theta=0^\circ$ 和 $\theta=180^\circ$) 和分离型半鞍点 R_{h1} 和 R_{h2} ($\theta=80^\circ$ 和 $\theta=280^\circ$), 对比图 8(b) 与 8(c), 图 8(b) 中 R_{h2} 和 R_{h3} 由于 2 个对称涡的不断增强, 体积逐渐变大, 相互挤压, 促使这 2 个点不断靠近, 直至与 S_{h1} 重合, S_{h1} 由分离型半鞍点转化为再附型半鞍点, 衍生出空间的(不在壁面上的点)鞍点 S_1 , 其位置会由于两涡挤压作用上移, 故此时处于不稳定的对称涡流态; 在 $X/D=5$ 截面处(见图 9(d)), 此时 4 种情况均有 4 个奇点, 但是只有添加扰动模型在 $\alpha=30^\circ$ 时鞍点不关于 $\theta=180^\circ$ 对称, 故专门对此进行分析, 这 4 个奇点分别为再附

型半鞍点 R_{h1} 和 R_{h2} ($\theta=0^\circ$ 和 $\theta=180^\circ$) 和分离型半鞍点 R_{h1} 和 R_{h2} ($\theta=121^\circ$ 和 $\theta=246^\circ$), 此时可以发现, 相较于对称涡流态, 非对称涡流态的奇点个数和性质与其保持一致, 但是点 R_{h2} 所处位置由 $\theta=180^\circ$ 推迟至 $\theta=188^\circ$ 处, 是由于头部的微小扰动破坏了两涡的诱导平衡, 使两涡形成高度差, 进一步证明了此时背部旋涡处于非对称流态, 此时也有空间鞍点 S_1 , 其余 3 种情况仍为对称涡状态。

图 9 中各曲线所表示的各流态的鞍点数量与性质和图 8 中不同流态拓扑结构一一对应。分析图 7 各截面矢量图发现在 $X/D=1$, $X/D=3$ 和 $X/D=5$ 截面出现的 2 个空间螺旋点利用截面法向量与速度矢量夹角为 0 的方法确定其精确位置, 下面将对涡核和空间鞍点位置沿轴向变化规律进行分析。

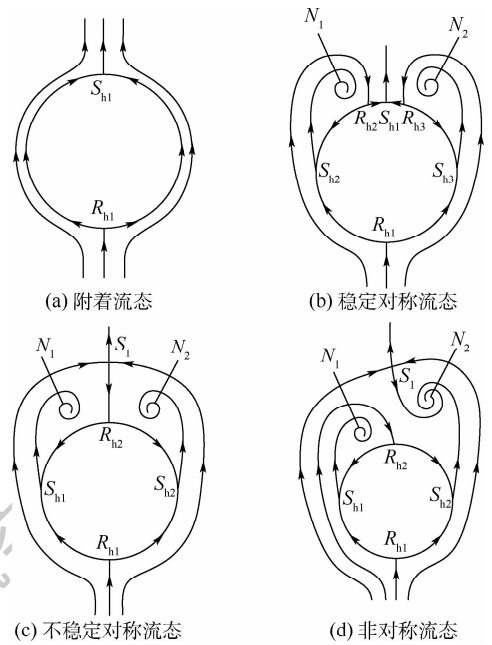


图 8 细长旋成体绕流拓扑结构

Fig. 8 Topological structure of flow around slender body

3.3 空间鞍点及涡核位置分析

空间奇点分为空间鞍点 (S_1) 和空间螺旋点 (即涡核), 其位置的确定采取了寻找截面法向量与速度矢量夹角为 0 的方法。图 10 为 2 种模型在 $\alpha=20^\circ$ 和 30° 时, 空间鞍点所在位置的纵向(高度)和横向(左右偏移)变化曲线。鞍点随截面沿旋成体后移位置不断升高, 但是添加扰动的模型点 S_1 位置变化较快, 对于 $\alpha=30^\circ$ 来讲, 点 S_1 出现的轴向截面位置比无扰动时提前约 0.8, 这是由于添加扰动块后会加快各个流态之间的转换; $\alpha=20^\circ$ 时 2 种模型均处于对称涡状态, 点 S_1 不会发生左右偏移, 但是有扰动模型鞍点横向位置产生了微小的波动, 有扰动模型点 S_1 纵向位置最大

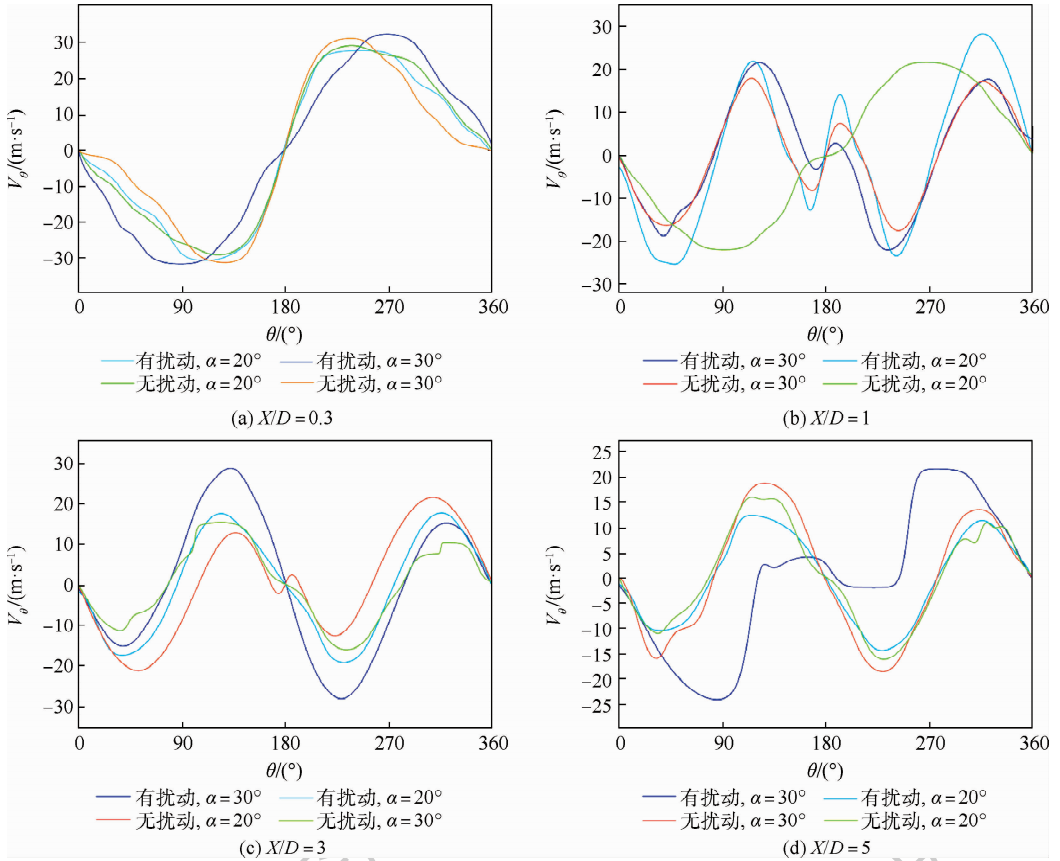


图 9 壁面切向速度曲线

Fig. 9 Tangential speed curves

约为无扰动的 5 倍,说明此时有由对称流态向非对称流态发展的趋势,也说明在该角度下扰动块对于点 S_1 的横向位置影响较小,但是在 $\alpha = 30^\circ$ 时添加扰动的模型刚开始处于对称涡状态,不会产生偏移,沿着轴向位置后移,右涡首先飘起,左涡在靠近物面的位置也会向右涡移动,出现非对称涡,点 S_1 发生横向偏移,在第 1 个峰值有扰动模型点 S_1 与无扰动的左右偏移量约为 0.04,第 2 个左右偏移峰值约为 0.15,这是由于随着轴向位置沿模型后移,两涡高度逐渐变大,物面会产生新的涡,迫使两涡向两侧和更高位置变化,点 S_1 偏移量也会随着轴向的发展而加大。

图 11(a) 为涡核的纵向(高度)位置, $\alpha = 20^\circ$ 时无论是否添加扰动,同一工况下两涡的涡核位置高度一样,并沿着轴向位置向上移,但是添加扰动后,涡核上移速度变快,但都处于对称涡状态下,所以 2 个模型的左右涡纵向曲线重合,但有扰动模型高度最大约为无扰动的 20 倍,说明在较小攻角下,添加扰动块不能改变绕流流态,但会加快各流态间的转换; $\alpha = 30^\circ$ 时无扰动模型涡核变化趋势与 $\alpha = 20^\circ$ 时一样,但是变化速度加快,对于有扰动模型,在其头部,两涡的涡核高度保持一致,不会形成高度差,故此时处于对称涡流态,随

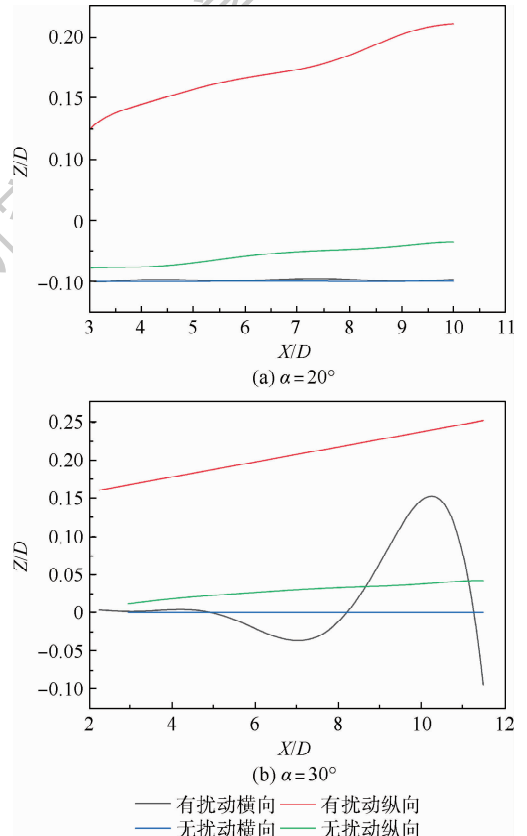
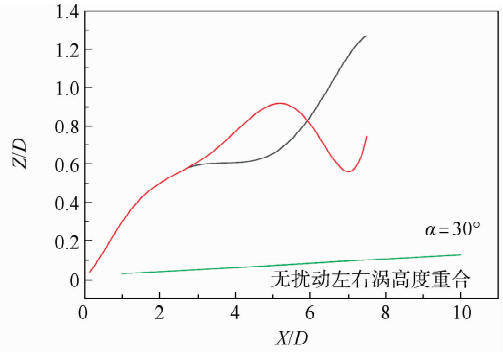
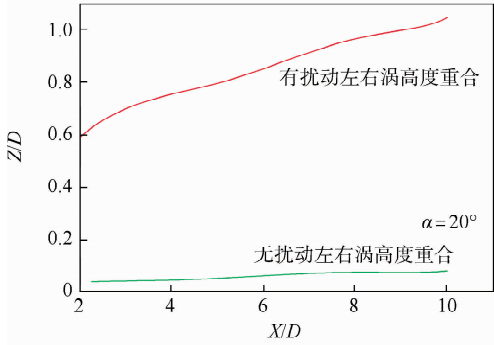


图 10 空间鞍点的纵向与横向位置

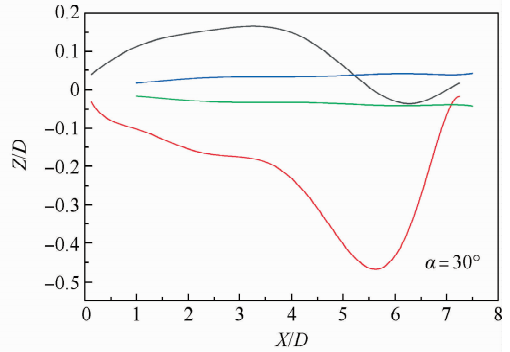
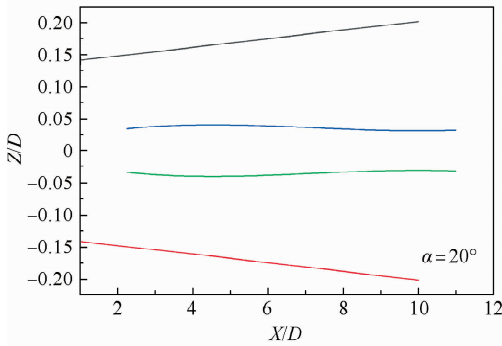
Fig. 10 Vertical and horizontal positions of saddle point in space

着轴向位置后移,右涡不断从剪切层获得涡量,致使其涡量增加,此时平衡状态遭到破坏,形成高度差,此时已处于非对称涡流态,同时2个主涡会交替飘起,甚至破裂,也会形成新的主涡,而且越靠近模型后部高位涡与低位涡的高度差会越来越大,此时已进入多涡流态,本文暂不研究;对于涡核的横向位置来说,当 $\alpha = 20^\circ$ 时,2种模型横向位置均基本关于 $y = 0$ 平面对称,此时处于对称涡流态;当 $\alpha = 30^\circ$ 时,无扰动模型涡核横向位置仍保持对称,但有扰动模型仅在模型头部保持对称,这

一对对称涡沿着轴向往后发展不断从剪切层吸收涡量,体积和涡量增加,平衡被破坏,导致右侧涡飘起(偏右),同时,左涡在靠近物面的位置靠右移动,直至右涡脱落,左涡不断飘起(偏左),右涡成为低位涡在靠近物面的位置靠左移动,整体呈现往复交替的趋势(若左涡为高位涡趋势同上),两涡横向(左右偏移)位置最大差值约为0.44。故添加扰动块后,在较小攻角时对于绕流结构的影响较小,就整体来看,扰动块加快了各绕流结构相互转换的速度。



(a) 纵向位置



(b) 横向位置

— 有扰动左涡 — 有扰动右涡 — 无扰动左涡 — 无扰动右涡

图 11 涡核位置

Fig. 11 Vortex core position

4 结论

本文对是否添加头部微小扰动块的细长旋成体模型在攻角 $\alpha = 5^\circ \sim 60^\circ$ 范围内,利用 SST $k-\omega$ 湍流模型进行数值模拟,得到了其绕流结构,并在 $\alpha = 20^\circ$ 和 30° 对绕流沿轴向发展进行研究,分析了扰动块对绕流结构的影响,得出以下结论:

1) 在添加已知规则扰动块的前提下,会使得非对称涡产生的攻角 α 减小,其非对称性会随着攻角的增加而增强,而且非对称涡出现的位置会随着攻角的增加沿轴向往头部移动。

2) 添加扰动块后,在较小攻角(如 $\alpha \leq 20^\circ$)时对绕流结构影响较小,但是会加快各绕流结构

间的相互转换。

3) 通过找到部分截面的壁面鞍点、涡核和空间鞍点的数量、精确位置和鞍点性质对相应的截面扰流拓扑结构进行验证,数值模拟结果与其保持一致。

4) 空间鞍点在对称涡阶段不会发生左右偏移;添加扰动后空间鞍点位置变化会加快。

5) 涡核高度会随攻角的增加而增加;添加扰动后涡核位置变化会加快。

参考文献 (References)

[1] ZILLIAC G G, DEGANI D, TOBAK M. Asymmetric vortices on a slender body of revolution [J]. AIAA Journal, 1991, 29 (5): 667-675.

- [2] BANKS D W, FISHER D F, HALL R M, et al. The F/A-18 high-angle-of-attack ground-to-flight correlation: Lessons learned: NASA-TM-4783[R]. Washington, D. C.: NASA, 1997.
- [3] RIZZETTA D P, VISBAL M R, STANEK M J. Numerical investigation of synthetic-jet flowfields[J]. AIAA Journal, 1999, 37(8):919-927.
- [4] FAGLEY C, PORTER C, MCLAUGHLIN T. Experimental closed-loop flow control of a von kármán ogive at high incidence[J]. AIAA Journal, 2014, 52(12):2891-2898.
- [5] 贺中, 吴军强, 蒋卫民, 等. 高速下模型头部扰动与非对称涡流动响应研究[J]. 实验流体力学, 2013, 27(6):6-13.
HE Z, WU J Q, JIANG W M, et al. Investigation on response between model tip perturbation and asymmetric vortex flow over slender body in high speed[J]. Journal of Experiments in Fluid Mechanics, 2013, 27(6):6-13(in Chinese).
- [6] 贺中, 吴军强, 蒋卫民, 等. 细长体大迎角非对称流动的高速 PIV 风洞试验研究[J]. 空气动力学学报, 2014, 32(3):295-299.
HE Z, WU J Q, JIANG W M, et al. Study on asymmetric flow over slender body at high angles of attack via particle image velocimetry test in high speed wind tunnel[J]. Acta Aerodynamica Sinica, 2014, 32(3):295-299(in Chinese).
- [7] WU M, YUE H H, CAO M J, et al. Research on forebody active disturbed flow characteristics of slender body in supersonic field[J]. Journal of the Brazilian Society of Mechanical Sciences and Engineering, 2018, 40(9):431.
- [8] MURMAN S. Geometric perturbations and asymmetric vortex shedding about slender pointed bodies[C]// Atmospheric Flight Mechanics Conference. Reston: AIAA, 2000:4103.
- [9] KUMAR P, PRASAD J K. Mechanism of side force generation and its alleviation over a slender body[J]. Journal of Spacecraft and Rockets, 2016, 53(1):195-208.
- [10] MA B F, HUANG Y, LIU T X. Low-frequency unsteadiness of vortex wakes over slender bodies at high angle of attack[J]. Chinese Journal of Aeronautics, 2014, 27(4):772-780.
- [11] MOSKOVITZ C, DEJARNETTE F, HALL R. Effects of surface perturbations on the asymmetric vortex flow over a slender body[C]//26th Aerospace Sciences Meeting. Reston: AIAA, 1988.
- [12] ZHU Y D, YUAN H J, LEE C. Experimental investigations of the initial growth of flow asymmetries over a slender body of revolution at high angles of attack[J]. Physics of Fluids, 2015, 27(8):084103.
- [13] 李国辉, 邓学莹. 细长体截面绕流中的一种临界状态[J]. 应用数学和力学, 2005, 26(1):25-31.
LI G H, DENG X Y. A critical pattern of crossflow around a slender[J]. Applied Mathematics and Mechanics, 2005, 26(1):25-31(in Chinese).
- [14] 刘跃, 管小荣, 徐诚. 不同 SST 模式在细长体绕流模拟中的应用比较[J]. 工程力学, 2016, 33(11):240-248.
LIU Y, GUAN X R, XU C. Comparison amongst different shear-stress transport models in simulating the flow field over a slender body[J]. Engineering Mechanics, 2016, 33(11):240-248(in Chinese).
- [15] 邓学莹, 王刚, 陈学锐, 等. 细长旋成体大迎角正则态非对称涡系结构的物理模型[J]. 中国科学 E 辑: 技术科学, 2004, 34(1):79-93.
DENG X Y, WANG G, CHEN X R, et al. Physical model of asymmetrical vortex structure of slender body at high angle of attack[J]. Science in China Ser. E: Technological Sciences, 2004, 34(1):79-93(in Chinese).
- [16] 张浩南, 李国辉, 徐宇. 尖拱细长旋成体低速大攻角绕流数值模拟[J/OL]. 兵器装备工程学报, 2020(2020-10-28)[2020-12-01]. <http://kns.cnki.net/kcms/detail/50.1213.TJ.20201028.1326.002.html>.
ZHANG H N, LI G H, XU Y. Numerical simulation of flow around a pointed arch slender body at low speed and high angle of attack[J/OL]. Journal of Ordnance Equipment Engineering, 2020(2020-10-28)[2020-12-01]. <http://kns.cnki.net/kcms/detail/50.1213.TJ.20201028.1326.002.html>. (in Chinese).
- [17] 孟轩, 陈志敏, 姚伟刚. 细长体非对称涡流动及侧向力分析研究[J]. 弹箭与制导学报, 2007, 27(2):219-221.
MENG X, CHEN Z M, YAO W G. Investigation of asymmetrical flow and side force around slender body[J]. Journal of Projectiles, Rockets, Missiles and Guidance, 2007, 27(2):219-221(in Chinese).
- [18] LOWSON M V, PONTON A J C. Symmetry breaking in vortex flows on conical bodies[J]. AIAA Journal, 1992, 30(6):1576-1583.

Numerical simulation of flow around slender body with disturbing block

LIU Keqi, LI Guohui*, ZHANG Haonan, ZHANG Yanjia

(Aviation Operations Service Academy, Aviation University Air Force, Changchun 130000, China)

Abstract: In this paper, the influence of the head disturbance block on the structure of the flow around the back of the slender body is studied. The slender body model is simulated at the angle of attack of $5^\circ - 60^\circ$ through numerical simulation. The numerical calculation method selected is the Reynolds Average Navier-Stokes (RANS) method. This paper mainly compares whether to add a disturbance block model at an angle of attack of $\alpha = 20^\circ, 30^\circ$, analyzes the development of the flow around different sections along with the axial position, proposes a method to verify the topological structure, and finds the position of the singularity in each flow state. The position of the vortex core is used to analyze the development of the flow around the back of the model. It is found that the premise of adding a known regular disturbance block can speed up the conversion speed between the surrounding flow structures, and will reduce the angle of attack generated by the asymmetric vortex.

Keywords: surrounding flow structure; slender body; asymmetric flow; saddle point; topological structure

http://bhxb.buaa.edu.cn jbuua@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2020.0370

焊接缺陷对异种铝合金 TIG 对接接头 疲劳行为的影响



王池权*, 石亮, 张祥春, 刘志毅, 邵成伟

(中国航空综合技术研究所, 北京 100028)

摘 要: 以 5A06-O/7A05-T6 异种铝合金钨极氩弧焊(TIG)对接接头为对象,通过疲劳试验数据和断口形貌分析,研究了气孔缺陷和未熔合缺陷对焊接接头疲劳性能的影响规律及机理。结果表明:气孔缺陷和未熔合缺陷对 5A06-O/7A05-T6 对接接头的疲劳性能均产生不利影响,且缺陷的大小、位置与载荷的交互作用是影响疲劳裂纹提前萌生的主要因素,同一应力水平下,疲劳裂纹更易萌生于尺寸较大且位置更接近于材料表面的焊接缺陷处,而随着应力水平的降低,焊接缺陷对焊接接头疲劳性能的不利影响更为显著;与气孔缺陷相比,未熔合缺陷边缘的应力集中效应更明显,更易导致疲劳裂纹萌生,且焊接接头组织相较于焊接母材组织更脆,疲劳裂纹以穿晶和沿晶形式交替扩展,使疲劳寿命进一步缩短。

关键词: 铝合金; 钨极氩弧焊(TIG); 焊接缺陷; 疲劳; SEM

中图分类号: V261.3; TG40

文献标志码: A

文章编号: 1001-5965(2021)07-1505-10

铝合金材料具有比强度高、耐腐蚀性好等优点,其焊接结构在汽车、高铁和航空航天领域得到了大量应用^[1]。由于铝合金具有较好的导电导热性,且与氧的亲合力大,在焊接母材和填充焊丝的重熔过程中容易形成气孔、未熔合、裂纹等缺陷^[2-5],使焊接接头成为焊接结构的薄弱部位,而铝合金焊接结构在服役过程中又常常受到交变载荷的作用^[6-8],为此,针对焊接缺陷对铝合金焊接接头疲劳性能的影响开展了大量研究。张红霞等^[9]对 5A06 铝合金及其焊接接头疲劳性能断裂行为进行了研究,结果发现,5A06 铝合金对接接头、横向十字接头等焊接接头的疲劳强度均低于 5A06 铝合金母材。何超等^[10]采用超声疲劳试验方法研究了气孔对铝合金 5052-H32 焊接接头超高周疲劳性能的影响,结果表明,气孔缺陷是诱发

疲劳裂纹萌生的主要原因,相同疲劳寿命下焊接接头疲劳强度较母材下降了 73.3%。赵旭和陈辉^[11]通过疲劳试验与断口形貌分析研究了 7N01 铝合金平滑焊接接头疲劳裂纹萌生的机理,结果发现,焊接接头疲劳裂纹起源于焊缝中近熔合线的液化微孔或微裂纹处,随循环载荷不断作用,裂纹逐步合并、扩展,直至断裂。宋哲^[12]和胡志力^[13]等分别研究了气孔缺陷对 7020 铝合金激光复合焊接接头和 2024 铝合金搅拌摩擦焊接接头疲劳行为的影响,结果发现,疲劳裂纹多萌生于表面和近表面较大气孔处或者近表面簇集的小气孔处,且焊接接头的疲劳性能随着气孔尺寸和外加应力的增大而不断降低。邓彩艳等^[14]采用超高周疲劳试验系统研究了 7050-T7451 铝合金搅拌摩擦焊接头的超高周疲劳性能,结果表明,裂纹萌

收稿日期: 2020-07-29; 录用日期: 2020-10-23; 网络出版时间: 2020-11-05 14:27

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20201104.1502.005.html

基金项目: 国家重点研发计划(2018YFB1106300)

*通信作者. E-mail: wangchiquan301@163.com

引用格式: 王池权, 石亮, 张祥春, 等. 焊接缺陷对异种铝合金 TIG 对接接头疲劳行为的影响[J]. 北京航空航天大学学报, 2021, 47(7): 1505-1514. WANG C Q, SHI L, ZHANG X C, et al. Influence of welding defects on fatigue behavior of dissimilar aluminum alloy TIG butt joint[J]. Journal of Beijing University of Aeronautics and Astronautics, 2021, 47(7): 1505-1514 (in Chinese).

生位置取决于表面起裂和内部起裂相互竞争的结果,应力范围较高时,试件的疲劳裂纹往往在表面萌生,随着应力范围的降低,裂纹有亚表面和内部萌生的倾向。吴圣川等^[15]通过高周疲劳试验研究了气孔缺陷对铝合金 7020-T651 和 7050-T7451 焊接接头疲劳极限的影响,结果发现,铝合金 7020-T651 和 7050-T7451 焊接接头疲劳极限均低于母材,其中 7020-T651 焊接接头疲劳极限约为母材的 63.1%,铝合金 7050-T7451 焊接接头疲劳极限约为母材的 43.5%。

异种金属材料焊接结构可以充分发挥不同材料的性能优势,而在常见的焊接方法中,钨极氩弧焊(TIG)由于具有焊缝金属致密,强度、韧性和塑性良好等优点,常用于异种金属材料之间的焊接^[16-19],但异种金属材料的熔点、热导率等物理特性和力学性能均存在较大差异,因此异种金属材料在焊接过程中更容易产生气孔、未熔合等缺陷^[20-22]。目前,国内外对于异种金属材料特别是异种铝合金焊接接头的疲劳性能研究较少,含缺陷异种铝合金焊接接头疲劳裂纹萌生及扩展的影响机制也尚待进一步认识。为此,本文以 5A06-O/7A05-T6 异种铝合金 TIG 对接接头作为研究对象,试验测定了 5A06-O/7A05-T6 异种铝合金 TIG 对接接头的疲劳性能,通过试验数据对比以及疲劳断口 SEM 分析,研究焊接缺陷对 5A06-O/7A05-T6 异种铝合金 TIG 对接接头(以下简称 5A06-O/

7A05-T6 对接接头)疲劳性能的影响规律及机理,为异种铝合金焊接结构的工程应用提供参考。

1 疲劳试验

为了试验研究焊接缺陷对 5A06-O/7A05-T6 对接接头疲劳性能的影响,采用 ER5356 焊丝通过 TIG 焊将母材 5A06-O 和 7A05-T6 焊接在一起,再根据焊接接头部位的超声 TOFT 及胶片射线检测结果,确定含气孔缺陷和未熔合缺陷的区域,分别制备了去除焊缝余高的 5A06-O/7A05-T6 平滑焊接接头疲劳测试试样(气孔缺陷、未熔合缺陷)。焊接母材及焊丝的力学性能和化学成分如表 1 和表 2 所示。5A06-O/7A05-T6 对接接头疲劳试样的几何形状与尺寸如图 1 所示,考虑到实际使用情况,对 5A06-O/7A05-T6 对接接头余高进行打磨,再进行表面抛光处理以降低残余应力的影响。

表 1 5A06-O/7A05-T6 焊接母材及焊丝的力学性能
Table 1 Mechanical properties of 5A06-O/7A05-T6 welding base metal and welding wire

材料	σ_b /MPa	σ_s /MPa	δ /%
5A06-O	356.4	161.8	21.14
7A05-T6	414.4	311.6	13.56
5A06-O/7A05-T6	281.08	157.33	11.43

注: σ_b 为强度极限, σ_s 为屈服极限, δ 为延伸率。

表 2 5A06-O/7A05-T6 焊接母材及焊丝的化学成分

Table 2 Chemical composition of 5A06-O/7A05-T6 welding base metal and welding wire

材料	元素含量							
	Mg	Mn	Zn	Fe	Si	Cu	Ti	Al
5A06-O	6.2	0.62	0.23	0.13	0.12	0.11	0.03	Bal.
	Zn	Mg	Mn	Cu	Fe	Si	Ti	Al
7A05-T6	4.58	1.37	0.22	0.16	0.15	0.08	0.04	Bal.
	Mg	Fe	Mn	Ti	Si	Zn	Cu	Al
ER5356	4.59	0.13	0.12	0.1	0.04	0.01	0.001	Bal.

注: Bal. 表示 Al 元素的含量为除了表中已经列出的元素含量外剩下的化学元素含量。

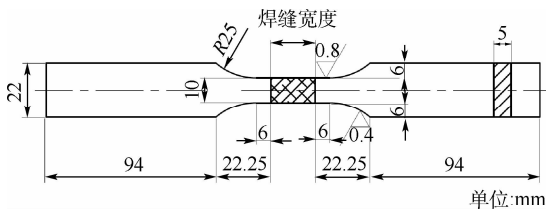


图 1 5A06-O/7A05-T6 对接接头疲劳测试试样

Fig. 1 Fatigue test specimen of 5A06-O/7A05-T6 butt joint

参照标准 ASTM E466 - 2015^[23],在 MTS-100 kN 液压伺服机上进行 5A06-O/7A05-T6 对接接头(气孔缺陷、未熔合缺陷)的恒载疲劳试验

(见图 2),试验环境及加载方式如表 3 所示,加载波形为正弦波,加载应力比 $R = 0.1$,加载频率 $f = 10$ Hz。

5A06-O/7A05-T6 对接接头(气孔缺陷、未熔合缺陷)在中、短疲劳寿命区的中值疲劳寿命($p = 50\%$)采用三级成组试验法测定^[24],每级试验的应力水平参照焊接母材的疲劳寿命试验数据进行选择,其中 5A06-O/7A05-T6 对接接头(气孔缺陷)所选应力水平分别为 200 MPa、160 MPa 和 120 MPa;5A06-O/7A05-T6 对接接头(未熔合缺



图 2 疲劳测试

Fig. 2 Fatigue test

表 3 试验环境

Table 3 Test environments

环境	温度/℃	试样	加载方式
干燥大气	25 ± 3	5A06-O/7A05-T6 对接接头(气孔缺陷)	T-T 恒幅加载
干燥大气	25 ± 3	5A06-O/7A05-T6 对接接头(未熔合缺陷)	T-T 恒幅加载

所应力水平分别为 120 MPa、90 MPa 和 60 MPa,且每级应力水平下的有效试样数不少于 3 个。

试验后利用 JSM-6010LA 型扫描电子显微镜对焊接接头(气孔缺陷、未熔合缺陷)试样的断口进行观察,分析其破坏机理,并结合图像处理软件对疲劳源处缺陷尺寸进行测量及统计。

2 试验结果分析与讨论

2.1 含缺陷焊接接头与焊接母材疲劳性能对比

通过观察 5A06-O/7A05-T6 对接接头(气孔缺陷、未熔合缺陷)试样的破坏位置,发现 5A06-O/7A05-T6 对接接头的破坏位置均位于靠近母材 7A05-T6 一侧的热影响区,这与同种铝合金焊接接头破坏位置发生在随机一侧热影响区不同^[25-26],这是由于 5A06-O/7A05-T6 对接接头采用的 ER5356 焊丝物理性能及化学成分与 7A05-T6 相差较大,易在 7A05-T6 铝合金一端形成气孔、未熔合等缺陷,从而导致疲劳裂纹提前萌生。

5A06-O/7A05-T6 对接接头(气孔缺陷、未熔合缺陷)和焊接母材^[27]的疲劳试验数据点如图 3 所示, N 为疲劳循环次数, S_{max} 为疲劳载荷最大应力水平值。从图中可以看出,5A06-O/7A05-T6 对接接头(气孔缺陷)同一组疲劳数据点与焊接母

材疲劳数据点的分散性基本一致,这是由于气孔缺陷的数量一般较多,且其大小及分布具有与材料初始缺陷类似的随机性^[12,15],而 5A06-O/7A05-T6 对接接头(未熔合缺陷)同一组疲劳数据点与焊接母材疲劳数据点的分散性相差很大,这是由于 5A06-O/7A05-T6 对接接头(未熔合缺陷)试样中通常只存在一个未熔合缺陷,且不同试样中的未熔合缺陷尺寸相差较大,导致其数据点的分散性增大。

由于疲劳性能曲线本身考虑了材料初始缺陷、表面状态等特性的随机性,可以将气孔缺陷作为材料的一种初始缺陷,采用最小二乘法拟合得到的焊接母材 5A06-O、7A05-T6 和 5A06-O/7A05-T6 对接接头(气孔缺陷)的中值疲劳 $S-N$ 曲线及其表达式分别如图 3 及表 4 所示。从图 3 中可以看出,5A06-O/7A05-T6 对接接头(气孔缺陷、未熔合缺陷)的疲劳寿命均随着应力水平的增大而缩短,而在同一应力水平下,5A06-O/7A05-T6 对接接头(气孔缺陷、未熔合缺陷)的疲劳寿命均低于焊接母材 5A06-O 和 7A05-T6 的疲劳寿命,且含未熔合缺陷的 5A06-O/7A05-T6 对接接头的疲劳寿命最低,这说明气孔缺陷和未熔合缺陷对 5A06-O/7A05-T6 对接接头的疲劳性能均产生不利影响,且未熔合缺陷对 5A06-O/7A05-T6 对接接头疲劳性能的不利影响最大。

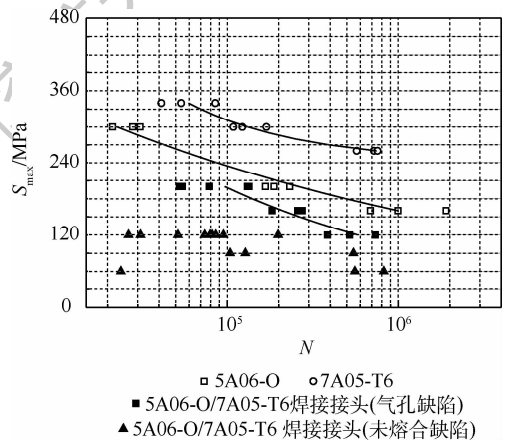


图 3 疲劳测试结果

Fig. 3 Fatigue test results

表 4 疲劳 S-N 曲线

Table 4 Fatigue S-N curves

试样	试验环境	S-N 曲线
5A06-O	干燥大气	$S_{max}^{6.02} N = 1.86 \times 10^{19}$
7A05-T6	干燥大气	$(S_{max} - 240.3)^{1.56} N = 7.8 \times 10^7$
5A06-O/7A05-T6 焊接接头(气孔缺陷)	干燥大气	$S_{max}^{3.50} N = 1.08 \times 10^{13}$

2.2 气孔缺陷对 5A06-O/7A05-T6 对接接头疲劳性能的影响

气孔缺陷作为焊接过程中难以避免出现的一类缺陷,其大小及位置对焊接接头疲劳裂纹的萌生均会产生影响,因此,首先对 5A06-O/7A05-T6 对接接头(气孔缺陷)试样疲劳裂纹萌生处的气孔缺陷尺寸及其距试样表面的最小距离进行测量。由于气孔缺陷近似为球形,其尺寸由直径来表征,测量过程如图 4 所示,测量结果如图 5 所示。从图 5 中可以看出,导致疲劳裂纹萌生的气孔缺陷直径大小集中在 85 ~ 140 μm 之间,且气孔缺陷距试样表面的最小距离位于 130 ~ 300 μm 范围内。

为了进一步研究气孔缺陷对 5A06-O/7A05-T6 对接接头疲劳性能的影响,定义疲劳寿命为 N 时,5A06-O/7A05-T6 对接接头的气孔缺陷影响系数 K_{pore} 为

$$K_{\text{pore}} = \frac{S_{\text{bm}}}{S_{\text{pore}}} \quad (1)$$

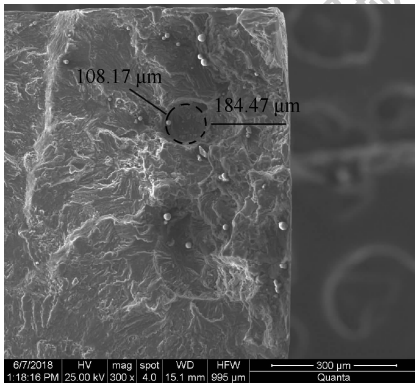


图 4 疲劳裂纹萌生处气孔缺陷尺寸及位置测量

Fig. 4 Measurement of size and location of pore defects at the position of fatigue crack initiation

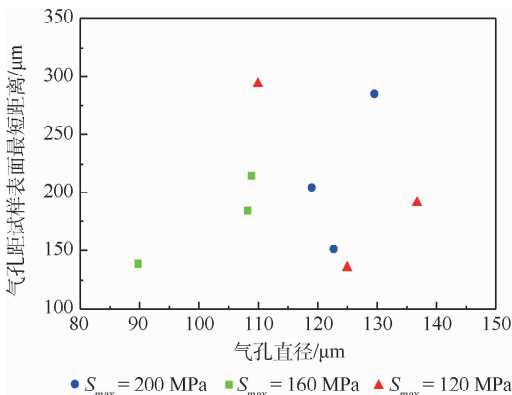


图 5 疲劳裂纹萌生处气孔缺陷尺寸及位置分布

Fig. 5 Size and location distribution of pore defects at the position of fatigue crack initiation

式中: S_{bm} 为焊接母材 5A06-O 疲劳寿命为 N 时对应的疲劳强度; S_{pore} 为 5A06-O/7A05-T6 对接接头(气孔缺陷)疲劳寿命为 N 时对应的疲劳强度。

根据拟合得到的焊接母材 5A06-O 和 5A06-O/7A05-T6 对接接头(气孔缺陷)的疲劳 $S-N$ 曲线表达式,可以得到 5A06-O/7A05-T6 对接接头的气孔缺陷影响系数 K_{pore} 曲线如图 6 所示。从图 6 中可以看出,5A06-O/7A05-T6 对接接头的气孔缺陷影响系数 K_{pore} 随着疲劳寿命 N 的增加(即应力水平的降低)而不断增大且均大于 1,这说明与高应力水平(对应低周疲劳寿命 N)相比,低应力水平(对应高周疲劳寿命 N)下的 5A06-O/7A05-T6 对接接头(气孔缺陷)的疲劳强度相较于焊接母材 5A06-O 下降更快,即应力水平越低,气孔缺陷对焊接接头疲劳性能的不利影响越大。

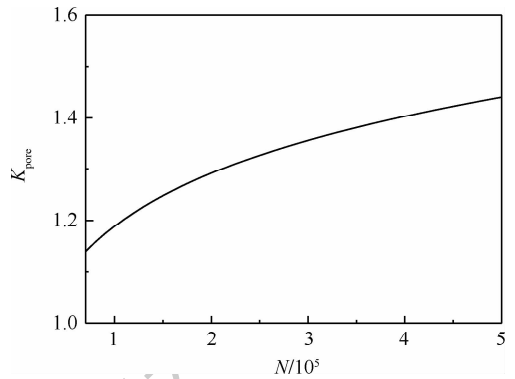


图 6 气孔缺陷影响系数曲线

Fig. 6 Influence coefficient curve of pore defect

2.3 未熔合缺陷对 5A06-O/7A05-T6 对接接头疲劳性能的影响

在现有焊接质量验收规范中,一般不允许焊缝处出现未熔合缺陷,但受限于结构形式、焊接工艺及经济性等因素影响,5A06-O/7A05-T6 异种铝合金焊接结构在焊接过程中容易在局部位置产生未熔合缺陷,这就需要进一步研究未熔合缺陷对 5A06-O/7A05-T6 对接接头疲劳性能的影响,为 5A06-O/7A05-T6 异种铝合金焊接结构剩余寿命的评定提供数据支撑。

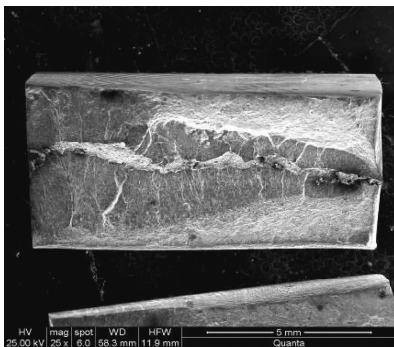
首先利用 Photoshop 图像处理软件提取 5A06-O/7A05-T6 对接接头(未熔合缺陷)断口照片中的未熔合边界,进而计算得到未熔合缺陷面积占断口截面面积的比例值,具体方法如下:首先利用 JSM-6010LA 型扫描电子显微镜对含未熔合缺陷的 5A06-O/7A05-T6 对接接头断口进行拍照(见图 7(a));然后利用 Photoshop 图像处理软件提取未熔合区域的边界,并将未熔合区域内部设定为白色,其他区域设定为黑色(见图 7(b));最

后根据颜色分布直方图中白色区域的百分比、断口以及照片的实际面积,即可计算得到未熔合缺陷面积占断口截面面积的比例值。采用上述方法计算得到的未熔合缺陷面积占断口截面面积的比例值分布如图 8 所示。从图 8 中统计可得,绝大多数未熔合缺陷的比例值小于 30%,占试样总数的 84.2%。

定义疲劳寿命为 N 时,5A06-O/7A05-T6 对接接头的未熔合缺陷影响系数 K_{if} 为

$$K_{if} = \frac{S_{bm}}{S_{if}} \quad (2)$$

式中: S_{if} 为 5A06-O/7A05-T6 对接接头(未熔合缺陷)疲劳寿命为 N 时对应的疲劳强度。



(a) 未熔合缺陷断口照片



(b) 未熔合缺陷图像处理后效果图

图 7 未熔合缺陷图像处理过程

Fig. 7 Image processing procedure of incomplete fusion defect

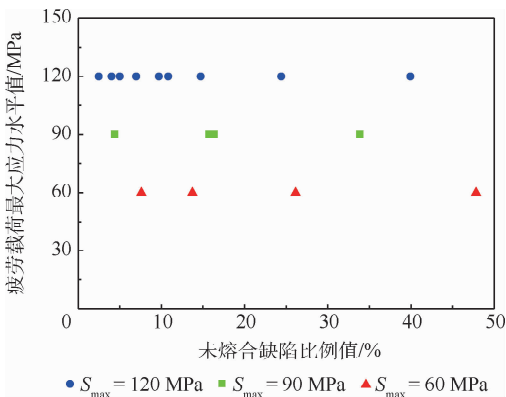


图 8 未熔合缺陷比例值分布

Fig. 8 Proportion distribution of incomplete fusion defects

根据拟合得到的焊接母材 5A06-O 的疲劳 $S-N$ 曲线表达式和含未熔合缺陷 5A06-O/7A05-T6 对接接头疲劳数据,可得到不同应力水平下的未熔合缺陷影响系数 K_{if} ,如图 9 所示。从图 9 中可以看出,不同应力水平下的焊接接头未熔合缺陷影响系数均大于 1,且同一应力水平下的未熔合缺陷影响系数随着未熔合缺陷比例的增加而不断增大,说明未熔合缺陷比例越大,5A06-O/7A05-T6 对接接头(未熔合缺陷)的疲劳性能相较于焊接母材的疲劳性能下降越快。

在图 9 数据点的基础上,采用最小二乘法进一步拟合得到不同应力水平下的未熔合缺陷影响系数 K_{if} 曲线如图 9 和表 5 所示。从图 9 中可以看出,同一应力水平下,未熔合缺陷影响系数与未熔合缺陷比例值近似呈线性关系,且不同应力水平下的未熔合缺陷影响系数 K_{if} 曲线近似平行。此外,应力水平低的未熔合缺陷影响系数曲线均位于应力水平高的未熔合缺陷影响系数曲线上方,说明未熔合缺陷比例相同的情况下,缺陷处引起的应力集中随着应力水平的增加而均匀增大,并且与高应力水平(对应低周疲劳寿命 N)相比,低应力水平(对应高周疲劳寿命 N)下的 5A06-O/7A05-T6 对接接头(未熔合缺陷)的疲劳强度相较于焊接母材下降更快,即应力水平越低,未熔合缺陷对焊接接头疲劳性能的不利影响越大。

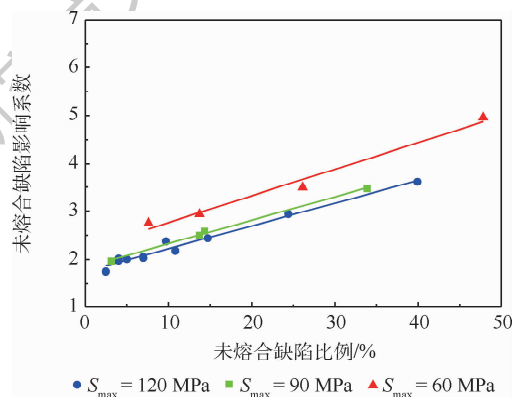


图 9 未熔合缺陷影响系数曲线

Fig. 9 Influence coefficient curves of incomplete fusion defect

表 5 未熔合缺陷影响系数拟合曲线

Table 5 Fitting curves of influence coefficient of incomplete fusion defect

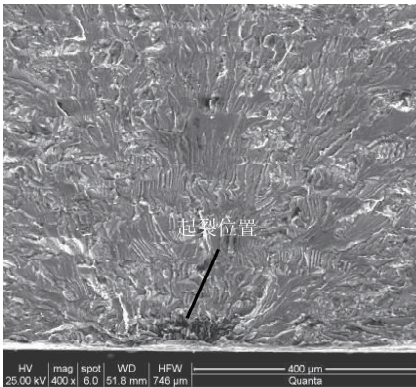
S_{max}/MPa	未熔合缺陷影响系数曲线	相关系数 R^2
120	$y = 0.047x + 1.75$	0.979
90	$y = 0.049x + 1.84$	0.997
60	$y = 0.056x + 2.21$	0.983

2.4 焊接缺陷对 5A06-O/7A05-T6 对接接头疲劳裂纹萌生及扩展的影响

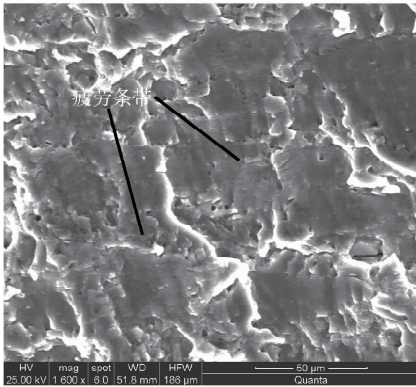
为了进一步探究焊接缺陷的大小、位置等对 5A06-O/7A05-T6 对接接头疲劳裂纹萌生及扩展的影响机理,利用 JSM-6010LA 型扫描电子显微镜对铝合金 5A06-O 和 5A06-O/7A05-T6 对接接头(气孔缺陷、未熔合缺陷)中的典型疲劳断口进行观测,如图 10 ~ 图 12 所示。图 10 中, $S_{max} = 160 \text{ MPa}$, $N = 6.87 \times 10^5$;图 11 中, $S_{max} = 120 \text{ MPa}$, $N = 5.20 \times 10^5$ 。

对比铝合金 5A06-O 和 5A06-O/7A05-T6 对接接头(气孔缺陷、未熔合缺陷)的疲劳源区,可以看出,铝合金 5A06-O 和 5A06-O/7A05-T6 对接接头(气孔缺陷)断口处的疲劳裂纹均在试样表

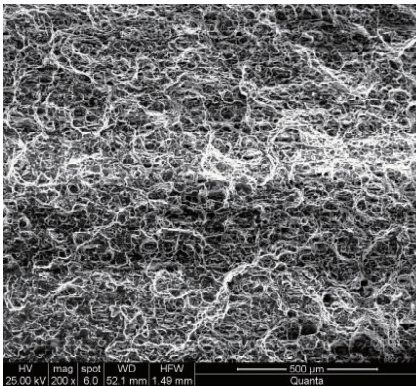
面或近表面初始缺陷处萌生并以点源形式逐渐扩展(见图 10(a)和图 11(a))。其中铝合金 5A06-O 起裂位置处存在塑性滑移特征,说明铝合金材料中的第二相粒子或夹杂等缺陷在位错滑移过程中引起的塑性变形是导致疲劳裂纹萌生的主要原因;而 5A06-O/7A05-T6 对接接头(气孔缺陷)起裂位置处存在明显的气孔缺陷(见图 11(a)),说明近表面气孔缺陷引起的应力集中是导致疲劳裂纹萌生的主要原因^[11-13]。对于 5A06-O/7A05-T6 对接接头(未熔合缺陷),根据未熔合缺陷的大小及分布,断口处的疲劳源区除了主疲劳源外还存在次疲劳源(见图 12(a)),且疲劳裂纹均在试样内部未熔合缺陷边缘处萌生,并以多个临近点源



(a) 疲劳源区



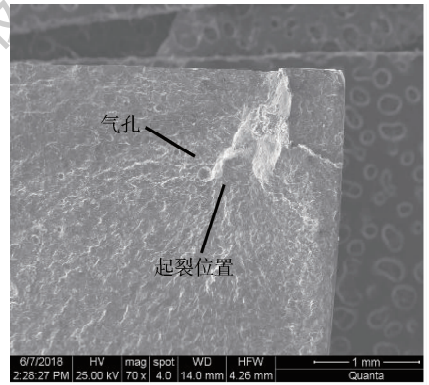
(b) 裂纹扩展区



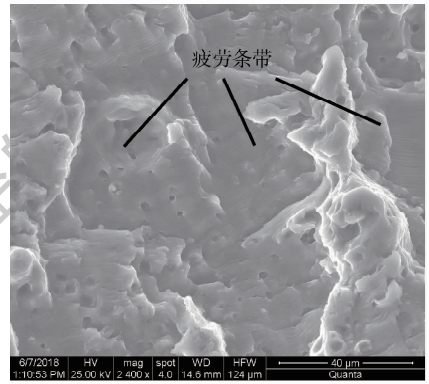
(c) 瞬断区

图 10 铝合金 5A06-O 疲劳断口

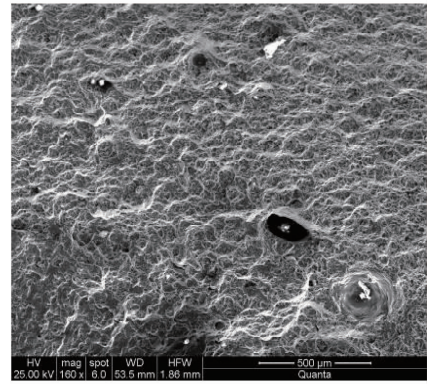
Fig. 10 Fatigue fracture of 5A06-O aluminum alloy



(a) 疲劳源区



(b) 裂纹扩展区



(c) 瞬断区

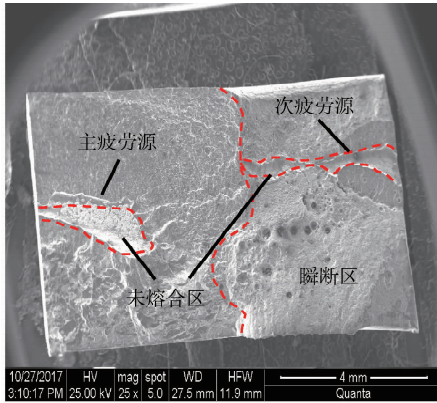
图 11 5A06-O/7A05-T6 对接接头疲劳断口(气孔缺陷)

Fig. 11 Fatigue fracture of 5A06-O/7A05-T6 butt joint (pore defect)

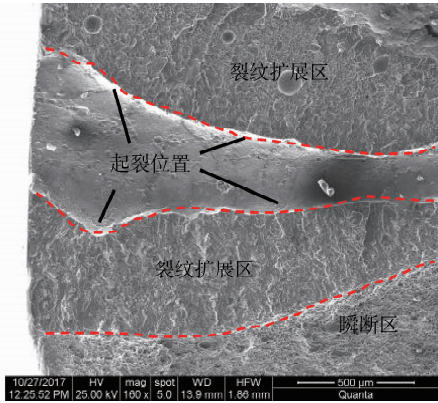
形成的线源形式逐渐扩展,不同点源处的裂纹不断扩展直至交汇形成台阶特征(见图 12(b)和(d)),此外,疲劳源区还可见焊接过程中形成的气孔缺陷,这说明与气孔缺陷相比,未熔合缺陷边缘处的应力集中效应更加明显,更容易导致疲劳裂纹萌生。

对比铝合金 5A06-O 和 5A06-O/7A05-T6 对接接头(气孔缺陷、未熔合缺陷)的疲劳裂纹扩展区,可以看出,铝合金 5A06-O 和 5A06-O/7A05-T6 对接接头(气孔缺陷、未熔合缺陷)的疲劳裂纹扩展区均可见致密、连续的疲劳条带,其中 5A06-O

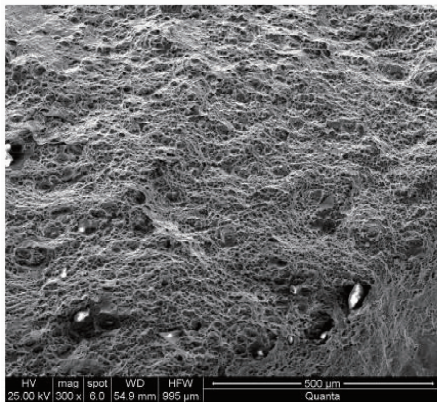
疲劳条带附近可见由二次相颗粒脱落形成的韧窝特征(见图 10(b)),而 5A06-O/7A05-T6 对接接头(气孔缺陷、未熔合缺陷)的疲劳条带附近还存在大量夹杂脱落和气孔形成的孔洞特征(见图 11(b)、图 12(e)和图 12(f))。与铝合金 5A06-O 的裂纹扩展区相比,5A06-O/7A05-T6 对接接头(气孔缺陷、未熔合缺陷)裂纹扩展区中的疲劳条带较不明显,且存在脆性开裂特征(见图 10(b)和图 12(e)),说明 5A06-O/7A05-T6 对接接头相较于铝合金 5A06-O 更脆,从而在裂纹扩展的过程中容易形成短促、不连续的二次裂纹。



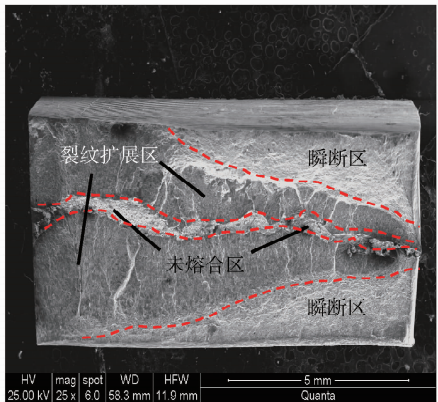
(a) 疲劳源区, $S_{max} = 120 \text{ MPa}$, $N = 8.57 \times 10^4$



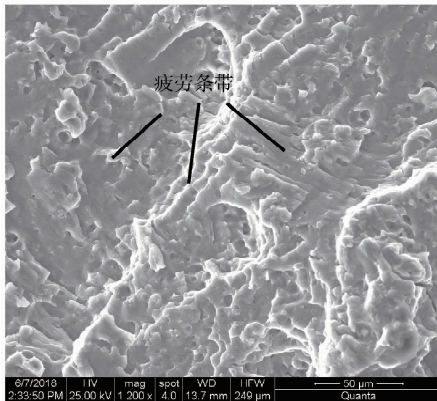
(b) 裂纹萌生处, $S_{max} = 120 \text{ MPa}$, $N = 8.57 \times 10^4$



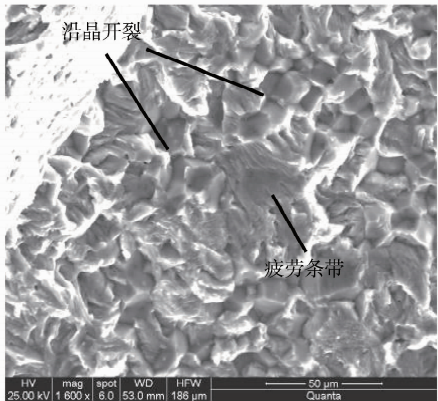
(c) 瞬断区, $S_{max} = 120 \text{ MPa}$, $N = 8.57 \times 10^4$



(d) 疲劳源区, $S_{max} = 120 \text{ MPa}$, $N = 7.94 \times 10^4$



(e) 裂纹稳定扩展区, $S_{max} = 120 \text{ MPa}$, $N = 7.94 \times 10^4$



(f) 裂纹快速扩展区, $S_{max} = 120 \text{ MPa}$, $N = 7.94 \times 10^4$

图 12 5A06-O/7A05-T6 对接接头疲劳断口(未熔合缺陷)

Fig. 12 Fatigue fracture of 5A06-O/7A05-T6 butt joint (incomplete fusion defect)

此外,在裂纹快速扩展区中的疲劳条带附近还存在大量沿晶开裂特征(见图 12(f)),说明焊接接头的疲劳裂纹在扩展中后期是以穿晶和沿晶形式交替扩展。对比铝合金 5A06-O 和 5A06-O/7A05-T6 对接接头(气孔缺陷、未熔合缺陷)瞬断区形貌,可以发现 5A06-O/7A05-T6 对接接头瞬断区的韧窝大小不一,同时存在大量大小不一的气孔、夹杂等焊接缺陷,且随机分布于整个断口截面处。

3 结 论

试验测定了 5A06-O/7A05-T6 异种铝合金 TIG 对接接头(气孔缺陷、未熔合缺陷)在干燥大气环境下的疲劳性能,并对试样断口进行 SEM 分析,研究了焊接缺陷与载荷联合作用对 5A06-O/7A05-T6 异种铝合金 TIG 对接接头疲劳性能的影响机理,得到以下结论:

1) 5A06-O/7A05-T6 异种铝合金 TIG 对接接头更易在 7A05-T6 一侧的热影响区处产生气孔、未熔合等缺陷,导致疲劳裂纹萌生,且气孔缺陷和未熔合缺陷影响系数均大于 1,说明气孔缺陷和未熔合缺陷对 5A06-O/7A05-T6 对接接头的疲劳性能均产生不利影响,并且应力水平越低,焊接缺陷与疲劳载荷的交互作用越充分,疲劳性能下降越明显。

2) 与含气孔缺陷的 5A06-O/7A05-T6 对接接头相比,含未熔合缺陷的 5A06-O/7A05-T6 对接接头的疲劳性能下降更为显著,且未熔合缺陷比例越大,焊接接头的疲劳性能下降越快。

3) 缺陷的大小、位置与载荷的交互作用是影响疲劳裂纹提前萌生的主要因素,同一应力水平下,疲劳裂纹更易萌生于尺寸较大且位置更接近材料表面的缺陷处,相较于气孔缺陷未熔合边缘更容易引起应力集中,导致疲劳裂纹从材料内部提前萌生,并以穿晶和沿晶形式交替扩展,使疲劳寿命进一步缩短。

参考文献 (References)

- [1] 马青娜,邵飞,高磊. 铝合金焊接接头疲劳研究进展综述[J]. 建筑结构,2018,48(S2):1037-1041.
MA Q N, SHAO F, GAO L. Overview of research progress of fatigue of aluminum alloy butt joints[J]. Building Structure, 2018,48(S2):1037-1041 (in Chinese).
- [2] 李占明,朱有利,杜晓坤,等. 铝合金手工 GTAW 焊接缺陷及其对疲劳性能的影响[J]. 热加工工艺,2012,41(15):133-135.
LI Z M, ZHU Y L, DU X K, et al. Effect of aluminum alloy welding defects in gas tungsten arc welding on fatigue properties [J]. Hot Working Technology, 2012,41(15):133-135 (in Chinese).
- [3] 马思群,谷理想,袁永文,等. 焊接缺陷对动车组铝合金车体疲劳寿命影响研究[J]. 铁道学报,2014,36(2):42-48.
MA S Q, GU L X, YUAN Y W, et al. Research on influence of welding defects on fatigue life of EMU aluminum-alloy car body [J]. Journal of the China Railway Society, 2014,36(2):42-48 (in Chinese).
- [4] ZHU C X, TANG X H, HE Y, et al. Effect of preheating on the defects and microstructure in NG-GMA welding of 5083 Al-alloy[J]. Journal of Materials Processing Technology, 2018,251:214-224.
- [5] 邹东利,陈向林,肖大武,等. 5A06 铝合金激光焊接接头显微组织及缺陷分析[J]. 稀有金属材料与工程,2019,48(9):2951-2956.
ZOU D L, CHEN X L, XIAO D W, et al. Microstructure and defects analysis of 5A06 aluminium alloy laser butt joint[J]. Rare Metal Materials and Engineering, 2019,48(9):2951-2956 (in Chinese).
- [6] 王池权,熊峻江,马少俊,等. 航空铝合金材料腐蚀裂纹扩展性能试验[J]. 北京航空航天大学学报,2017,43(5):935-941.
WANG C Q, XIONG J J, MA S J, et al. Tests for corrosion crack propagation behavior of aeronautical aluminum alloys[J]. Journal of Beijing University of Aeronautics and Astronautics, 2017,43(5):935-941 (in Chinese).
- [7] ZHANG B, CHEN W, POIRIER D R. Effect of solidification cooling rate on the fatigue life of A356.2-T6 cast aluminum alloy[J]. Fatigue and Fracture of Engineering Materials and Structures, 2000,23(5):417-423.
- [8] 王晨,王贝贝,薛鹏,等. SiCp/6092Al 复合材料搅拌摩擦焊接头的疲劳行为研究[J]. 金属学报,2019,55(1):149-159.
WANG C, WANG B B, XUE P, et al. Fatigue behavior of friction stir welded SiCp/6092Al composite[J]. Acta Metallurgica Sinica, 2019,55(1):149-159 (in Chinese).
- [9] 张红霞,吴广贺,闫志峰,等. 5A06 铝合金及其焊接接头的疲劳断裂行为[J]. 中国有色金属学报,2013(2):327-335.
ZHANG H X, WU G H, YAN Z F, et al. Fatigue fracture behavior of 5A06 aluminum alloy and its butt joint[J]. The Chinese Journal of Nonferrous Metals, 2013(2):327-335 (in Chinese).
- [10] 何超,崔仕明,刘永杰,等. 气孔对铝合金焊接接头超长疲劳寿命的影响[J]. 焊接学报,2014(11):18-22.
HE C, CUI S M, LIU Y J, et al. Arc behavior of A-MIG welding and microstructure of aluminum alloy butt joint[J]. Transactions of the China Welding Institution, 2014(11):18-22 (in Chinese).
- [11] 赵旭,陈辉. 7N01 铝合金平滑焊接接头疲劳性能研究[J]. 热加工工艺,2016,45(21):183-186.
ZHAO X, CHEN H. Research on fatigue property of smooth weld joints in 7N01 aluminum alloy[J]. Hot Working Technology, 2016,45(21):183-186 (in Chinese).
- [12] 宋哲,吴圣川,胡雅楠,等. 冶金型气孔对熔化焊接 7020 铝

- 合金疲劳行为的影响[J]. 金属学报, 2018, 54(8): 1131-1140.
- SONG Z, WU S C, HU Y N, et al. The influence of metallurgical pores on fatigue behaviors of fusion welded AA7020 joints[J]. Acta Metallurgica Sinica, 2018, 54(8): 1131-1140 (in Chinese).
- [13] 胡志力, 李锦, 万心勇, 等. 铝合金搅拌摩擦焊接头服役孔洞裂纹扩展规律[J]. 稀有金属材料与工程, 2019, 48(3): 892-897.
- HU Z L, LI J, WAN X Y, et al. Crack propagation law of holes in aluminum alloy friction stir welding joint[J]. Rare Metal Materials and Engineering, 2019, 48(3): 892-897 (in Chinese).
- [14] 邓彩艳, 高仁, 龚宝明, 等. 7050 铝合金搅拌摩擦焊接头超高周疲劳性能[J]. 焊接学报, 2018, 39(11): 114-118.
- DENG C Y, GAO R, GONG B M, et al. Research on ultra-high-cycle fatigue properties of 7050 aluminum alloy FSW butt joints[J]. Transactions of the China Welding Institution, 2018, 39(11): 114-118 (in Chinese).
- [15] 吴圣川, 谢成, 胡雅楠, 等. 中高强度铝合金熔焊焊接头的缺陷容忍评价方法[J]. 机械工程学报, 2020, 56(8): 46-59.
- WU S C, XIE C, HU Y N, et al. Defect tolerance assessment method of fusion welded medium and high strength Al alloy joints[J]. Journal of Mechanical Engineering, 2020, 56(8): 46-59 (in Chinese).
- [16] 孔令明, 杜双明, 韩赞, 等. 镁合金/不锈钢钎极氩弧熔焊钎焊接头结合界面微观组织分析[J]. 矿冶工程, 2017, 37(5): 118-121.
- KONG L M, DU S M, HAN Y, et al. Analysis of interface microstructure of TIG brazing joint of magnesium alloy and stainless steel[J]. Mining and Metallurgical Engineering, 2017, 37(5): 118-121 (in Chinese).
- [17] LIU F, LIU Y, WU Y. Effect of lattice matching degree and intermetallic compound on the properties of Mg/Al dissimilar material butt joints[J]. Science and Technology of Welding and Joining, 2017, 22(8): 719-725.
- [18] DU H G, LI Y J, XU X Y, et al. Analysis of microstructure and properties of Mg/Al dissimilar joints by GTAW with Al-Si filler wire[J]. Kovove Mater, 2015, 53: 113-118.
- [19] 饶文姬, 魏守征, 李志勇, 等. AZ31B/5A06 脉冲熔极氩弧焊焊缝组织特性[J]. 热加工工艺, 2019, 48(21): 26-30.
- RAO W J, WEI S Z, LI Z Y, et al. Microstructures of weld seam during pulsed metal inert-gas welding of AZ31B and 5A06 dissimilar alloys[J]. Hot Working Technology, 2019, 48(21): 26-30 (in Chinese).
- [20] 杜红燕, 李亚江. AZ31/7005 异种材料填丝 GTAW 焊接接头的组织与性能[J]. 材料工程, 2014(9): 14-19.
- DU H Y, LI Y J. Microstructures and properties of AZ31/7005 dissimilar joints by GTAW with filler wire[J]. Journal of Materials Engineering, 2014(9): 14-19 (in Chinese).
- [21] 廖传清, 宿国友, 高艳芳, 等. 7075/5A06 异种铝合金 TIG 焊接头的显微组织和力学性能[J]. 中国有色金属学报, 2015(1): 43-48.
- LIAO C Q, SU G Y, GAO Y F, et al. Microstructure and mechanical properties of 7075/5A06 dissimilar aluminum alloy joints made by TIG welding[J]. The Chinese Journal of Nonferrous Metals, 2015(1): 43-48 (in Chinese).
- [22] 吕堂祺, 丁凯, 施睿寰, 等. 20 钢-铝青铜焊接接头的组织和高周次疲劳性能研究[J]. 上海金属, 2020, 42(1): 13-16.
- LV T Q, DING K, SHI R Y, et al. Investigation on microstructure and high-cycle fatigue behavior of butt joint of 20 steel and aluminum[J]. Shanghai Metals, 2020, 42(1): 13-16 (in Chinese).
- [23] Standard Practice for conducting force controlled constant amplitude axial fatigue tests of materials: ASTM E466 - 2015 [S]. American Society for Testing and Materials International, 2015.
- [24] 高镇同, 蒋新桐, 熊峻江, 等. 疲劳性能试验设计和数据处理[M]. 北京: 北京航空航天大学出版社, 1999: 21-23.
- GAO Z T, JIANG X T, XIONG J J, et al. Fatigue experiment design and data processing [M]. Beijing: Beihang University Press, 1999: 21-23 (in Chinese).
- [25] 李宏伟, 姚为, 王志敏, 等. 激光焊接 ZL114A/5A06 异种铝合金接头组织及性能[J]. 宇航材料工艺, 2010, 40(6): 70-73.
- LI H W, YAO W, WANG Z M, et al. Microstructure and properties of dissimilar aluminum alloys ZL114A/5A06 joints by laser welding[J]. Aerospace Materials & Technology, 2010, 40(6): 70-73 (in Chinese).
- [26] 佟建华, 张坤, 林松, 等. 搅拌摩擦焊和熔极气体保护焊 6082 铝合金疲劳性能分析[J]. 焊接学报, 2015, 36(7): 105-108.
- TONG J H, ZHANG K, LIN S, et al. Comparison of fatigue property of 6082 aluminum alloy joint by friction stir welding and metal inert-gas welding [J]. Transactions of the China Welding Institution, 2015, 36(7): 105-108 (in Chinese).
- [27] WANG W K, CAO Z Q, SHI L, et al. Fatigue behavior investigation on dissimilar aluminum alloy ring butt joints in a new type sandwich high-pressure shell[J]. Materials Research Express, 2019, 6(6): 066534.

Influence of welding defects on fatigue behavior of dissimilar aluminum alloy TIG butt joint

WANG Chiquan^{*}, SHI Liang, ZHANG Xiangchun, LIU Zhiyi, SHAO Chengwei

(China Aero Polytechnology Establishment, Beijing 100028, China)

Abstract: Fatigue tests and fracture morphology analyses were performed to determine the influence law and mechanism of pore defect and incomplete fusion defect on the fatigue performance of 5A06-O/7A05-T6 dissimilar aluminum alloy TIG butt joint. The results show that both pore defect and incomplete fusion defect have adverse effects on the fatigue properties of 5A06-O/7A05-T6 butt joint, and the interaction between the size and location of defects and the load is the main factor affecting the early initiation of fatigue cracks. Fatigue cracks are more likely to initiate at the welding defects with larger size and closer to the material surface under the same stress level, and the interaction between welding defects and fatigue loading increases with the decrease of stress level, and ultimately decreases the fatigue strength. Compared with pore defect, the effect of stress concentration at the edge of incomplete fusion defect is more significant, which is more likely to cause fatigue crack initiation, the microstructure of welded joint is more brittle than that of welding base metal, and the fatigue crack propagates alternately in the transgranular and intergranular form, which further shortens the fatigue life.

Keywords: aluminum alloy; TIG; welding defects; fatigue; SEM

Received: 2020-07-29; **Accepted:** 2020-10-23; **Published online:** 2020-11-05 14:27

URL: kns.cnki.net/kcms/detail/11.2625.V.20201104.1502.005.html

Foundation item: National Key R & D Program of China (2018YFB1106300)

*** Corresponding author.** E-mail: wangchiquan301@163.com

《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人,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》
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47卷第7期 2021年7月

JOURNAL OF BEIJING UNIVERSITY OF
AERONAUTICS AND ASTRONAUTICS
(JBUEAA)

(Monthly, Started in 1956)

Vol.47 No.7 July 2021

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 jbuua@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 Qinqing

Edited and Published by Editorial Board of JBUEAA

Printed by Beijing Kexin Printing Co., Ltd.

Distributed by BUAA Culture Media Group Limited

Telephone (010) 82315594 82338922

82314839

E-mail jbuua@buaa.edu.cn

http://bhxb.buaa.edu.cn

中国标准连续出版物号: $\frac{\text{ISSN } 1001-5965}{\text{CN } 11-2625/V}$

国内定价: 50.00元/期

ISSN 1001-5965



9 771001 596212