

# QLEKF的木星探测环绕段自主导航方法

戴志雯<sup>1</sup>, 侯博文<sup>1</sup>, 张艺捷<sup>1</sup>, 何章鸣<sup>1</sup>, 陆丹丹<sup>2</sup>

(1. 国防科技大学 理学院, 长沙 410073; 2. 深空探测实验室, 北京 100195)

**摘要:** 针对木星的复杂环境导致噪声不确定性对滤波性能的影响, 提出了基于多个木星卫星相对视线方向信息建立光学自主导航方案, 采用单滤波器简化的Q学习扩展卡尔曼滤波(Q Learning Extend Kalman Filter, QLEKF)算法对探测器的位置和速度进行估计。单滤波器Q学习扩展卡尔曼滤波(QLEKF-single)基于单个扩展卡尔曼滤波器(Extend Kalman Filter, EKF)新息设计奖励函数, 采用Q学习算法自适应地选择噪声协方差阵的取值, 利用SoftMax策略进行动作选择, 最终通过EKF实现系统状态的迭代估计。仿真过程中经过随机生成的初始状态估计和测量噪声, 对木星真实的轨道动力学模型进行简化, 验证了在不确定噪声的影响条件下, QLEKF-single算法相对于传统滤波方法有效地提高了导航精度; 与QLEKF算法相比, 在精度变化不大的情况下, 运行时间减少了10%以上。

**关键词:** 自主导航; QLEKF; 贪婪算法; SoftMax算法; 木星环绕

**中图分类号:** V11

**文献标识码:** A

**文章编号:** 2096-9287(2025)02-0162-10

**DOI:** 10.15982/j.issn.2096-9287.2025.20240048

**引用格式:** 戴志雯, 侯博文, 张艺捷, 等. QLEKF的木星探测环绕段自主导航方法[J]. 深空探测学报(中英文), 2025, 12(2): 162-171.

**Reference format:** DAI Z W, HOU B W, ZHANG Y J, et al. Research on an autonomous navigation method for the orbital phase of jupiter probe based on QLEKF[J]. Journal of Deep Space Exploration, 2025, 12(2): 162-171.

## 引言

近些年来, 木星探测已成为继月球和火星探测后深空探测的又一重要领域<sup>[1]</sup>。目前, 全球范围内已启动或正在计划多项木星探测项目。美国通过“伽利略号”(Galileo)和“朱诺号”(Juno)探测器搜集到关于木星大气和磁层结构的关键数据<sup>[2-4]</sup>; 2023年4月14日, 欧洲航天局(European Space Agency, ESA)发射了“木星冰月探测”(JUper ICy moons Explorer, JUICE)任务, 通过环绕轨道多次飞越的方式深入探测木星卫星<sup>[5]</sup>; 中国计划于2030年左右开展木星探测任务<sup>[6]</sup>, 将重点关注木星及其卫星的环绕探测, 旨在进一步探索木星系统和行星际空间环境, 详细研究木星大气、磁场、离子体和高能粒子的分布。

木星与地球距离遥远, 存在约34 min的通信延时<sup>[2]</sup>, 地面难以对探测器实时测控, 因此要求探测器具备极强的自主管理能力, 包括自主导航与控制、自主任务规划、自主故障诊断及处理等多个方面<sup>[7]</sup>。其中: 自主导航技术可实时提供探测器的状态信息, 是确保探测任务成功实施的关键, 要求探测器不依赖外部支持,

仅利用自身携带的测量设备实时确定自身相对于某一坐标系的位置、速度、姿态等信息来引导航行<sup>[8]</sup>。

现有的自主导航算法, 多建立在火星、月球探测背景, 分别基于导航星的天体光学信息、陀螺仪和加速计测量信息、脉冲星信号和基于星间测量信标等测量信息<sup>[9]</sup>利用滤波算法解算。由于计算资源有限、木星周边磁场和引力场复杂, 探测器无法连续、精确地接收远处星系的测量数据, 而木星周边丰富的天然卫星运行稳定、星历信息已知, 其相对于探测器的视线矢量是理想的测量数据来源<sup>[7]</sup>。

现阶段木星环绕探测的导航精度需求在千米级<sup>[8]</sup>, 然而木星周边区域内复杂的空间扰动和干扰源导致过程噪声和测量噪声难以精准建模, 显著地降低了导航精度, 仅使用卡尔曼滤波(Extend Kalman Filter, EKF)进行状态估计无法达到精度需求。为抑制噪声的不确定性, 学者提出了多种自适应滤波方法。鲁平等<sup>[10]</sup>基于协方差匹配技术改进了Sage-Husa自适应滤波, 对测量噪声进行实时估计; Wang等<sup>[11]</sup>提出了一种基于递归方差估计的增广型扩展卡尔曼滤波(Augmented Kalman Filter, AKF)算法以解决过程噪声协方差不确定的问

题; Ge等<sup>[12]</sup>构造了一种基于模型参数比 (Model Parameter Ratio, MPR) 的AKF方法, 将自适应估计问题转化为约束优化问题, 更准确地估计过程噪声协方差阵和测量噪声协方差阵。基于上述自适应算法, 为进一步提高运算速度和估计精度, Xiong等<sup>[13-15]</sup>提出了一种结合Q学习扩展卡尔曼滤波 (Q Learning Extend Kalman Filter, QLEKF), 将EKF算法与强化学习相结合, 通过与环境交互得到反馈, 自适应地调整噪声协方差的取值。QLEKF可解决多种情景下高精度自主导航的问题, 降低噪声不确定性对EKF的负面影响。Dai等<sup>[16]</sup>利用QLEKF算法, 基于无人机真实飞行数据实现了磁强计、加速度计与陀螺仪的组合 (Magnetic Angular Rate and Gravity, MARG) 传感器对姿态的精确估计; Tao等<sup>[17]</sup>围绕太阳系边界巡航背景, 开展了利用QLEKF算法进行高精度导航的研究。

本文利用探测器在轨环绕期间获得的多个天然木星卫星的视距 (Line of Sight, LOS) 矢量, 建立了基于木星引力及 $J_2$ 摄动效应的自主导航模型。为进一步降低模型的求解时间, 本文提出了一种单滤波器奖励反馈的QLEKF-single算法。现有的QLEKF滤波器设计, 大多基于并行滤波器之间的新息序列设计强化学习的奖励值, 根据累积奖励, 使用贪婪策略选择最佳的噪声矩阵取值。QLEKF-single算法仅将搜索EKF新息序列模的对数函数作为即时奖励, 使用SoftMax策略作为智能体在不同噪声协方差对之间转移的依据。在收敛后估计精度基本不变的前提下, 与QLEKF算法相比, QLEKF-single提高了计算效率。

## 1 木星环绕动力学与测量模型

木星绕飞任务是木星探测的关键环节。为确定探测器在选定参考系的位置和速度, 便于模型的建立和解算, 本文描述了离散时间条件下坐标系、探测器状态变量及测量变量的定义<sup>[18]</sup>。

### 1) 坐标系

(1) 木星惯性坐标系 (i系) 原点定义为木星质心 $O$ ,  $OX$ 位于赤道平面内指向春分点;  $OZ$ 指向木星北极;  $OY$ 位于赤道平面, 与 $OX$ 、 $OZ$ 构成右手坐标系。

(2) 本体坐标系 (b系) 原点定义为探测器质心 $O$ ,  $OX$ 、 $OY$ 、 $OZ$ 分别沿探测器3个惯性主轴方向构成右手坐标系。

(3) 相机坐标系 (c系) 原点定义为相机光心 $O$ ,  $OY$ 平行于像平面,  $OZ$ 、 $OX$ 、 $OY$ 构成右手坐标系。

### 2) 探测器状态变量

木星惯性系定义 $k$ 时刻探测器的状态变量为

$$\mathbf{x}_k = \begin{bmatrix} \mathbf{r}_k \\ \mathbf{v}_k \end{bmatrix} \quad (1)$$

$$\mathbf{r}_k = [x_k \ y_k \ z_k]^T, \mathbf{v}_k = [\dot{x}_k \ \dot{y}_k \ \dot{z}_k]^T \quad (2)$$

其中:  $\mathbf{r}_k$ 和 $\mathbf{v}_k$ 分别为 $k$ 时刻探测器到木星质心的相对位置矢量和速度矢量;  $x_k$ 、 $y_k$ 、 $z_k$ 为 $k$ 时刻探测器的位置参数;  $\dot{x}_k$ 、 $\dot{y}_k$ 、 $\dot{z}_k$ 为探测器的速度参数。

### 3) 探测器观测变量

为便于分析测量方程, 假设探测器姿态可通过星敏感器进行估计<sup>[19]</sup>, 即在环绕木星的过程中, 探测器的姿态是已知的。探测器利用搭载相机对3颗木星卫星进行成像识别, 捕捉其平面二维像素数据并提取观测值。在相机坐标系下, 记 $k$ 时刻探测器指向木星卫星 $i$ 的LOS矢量为<sup>[20]</sup>

$$\mathbf{l}_{ik} = \frac{1}{\sqrt{x_{ik,c}^2 + y_{ik,c}^2 + f^2}} \begin{bmatrix} -x_{ik,c} \\ -y_{ik,c} \\ f \end{bmatrix} + \boldsymbol{\eta}_k = C_b^c C_i^b \frac{\mathbf{r}_{ik} - \mathbf{r}_k}{\|\mathbf{r}_{ik} - \mathbf{r}_k\|} + \boldsymbol{\eta}_k, \quad i = 1, 2, 3 \quad (3)$$

其中:  $x_{ik,c}$ 、 $y_{ik,c}$ 分别为根据相机拍摄图像提取的卫星 $i$ 质心特征点所在成像平面坐标;  $f$ 为相机焦距;  $\mathbf{r}_{ik} = [x_{ik} \ y_{ik} \ z_{ik}]^T, i = 1, 2, 3$ , 代表木星惯性系3颗卫星到木星的相对位置矢量;  $C_i^b$ 为木星惯性坐标系 (i系) 到探测器本体坐标系 (b系) 的旋转矩阵;  $C_b^c$ 为本体坐标系 (b系) 到相机坐标系 (c系) 的旋转矩阵。

### 1.1 动力模型

假设绕飞探测器主要受木星引力的影响, 仅考虑木星形状摄动的 $J_2$ 项, 采用过程噪声 $\mathbf{w}_k$ 描述忽略其它摄动项造成的误差。在惯性坐标系建立的探测器动力学绕飞模型<sup>[19]</sup>为

$$\mathbf{x}_k = f(\mathbf{x}_{k-1}) + \mathbf{w}_k \quad (4)$$

状态方程的一阶泰勒展开式为

$$f(\mathbf{x}_{k-1}) = \mathbf{x}_{k-1} + \boldsymbol{\phi}(\mathbf{x}_{k-1})\tau \quad (5)$$

$$\boldsymbol{\phi}(\mathbf{x}_k) = \begin{bmatrix} \dot{x}_k \\ \dot{y}_k \\ \dot{z}_k \\ -\mu \frac{x_k}{r_k^3} \left[ 1 - J_2 \left( \frac{R}{r_k} \right)^2 \left( 7.5 \frac{z_k^2}{r_k^2} - 1.5 \right) \right] \\ -\mu \frac{y_k}{r_k^3} \left[ 1 - J_2 \left( \frac{R}{r_k} \right)^2 \left( 7.5 \frac{z_k^2}{r_k^2} - 1.5 \right) \right] \\ -\mu \frac{z_k}{r_k^3} \left[ 1 - J_2 \left( \frac{R}{r_k} \right)^2 \left( 7.5 \frac{z_k^2}{r_k^2} - 4.5 \right) \right] \end{bmatrix} \quad (6)$$

其中:  $r_k = \|\mathbf{r}_k\|$ ;  $\mathbf{w}_k$ 为过程噪声。

其统计特性为

$$E(\mathbf{w}_k) = \mathbf{0}, \text{Cov}(\mathbf{w}_k, \mathbf{w}_j^T) = \mathbf{Q}_k \delta_{kj} \quad (7)$$

其中:  $\delta_{kj}$  为狄利克雷函数, 满足条件

$$\delta_{kj} = \begin{cases} 1, & k = j \\ 0, & k \neq j \end{cases} \quad (8)$$

根据木星轨道动力学<sup>[21]</sup>, 其中各参数取值及含义见表1。

表1 动力模型参数

Table 1 Dynamic model parameters

参数	取值
万有引力常数 $G/(\text{N}\cdot\text{m}^2\cdot\text{kg}^{-2})$	$6.67 \times 10^{-11}$
木星质量 $m/\text{kg}$	$1.90 \times 10^{27}$
木星引力常数 $\mu/(\text{km}^3\cdot\text{s}^{-2})$	$1.2673 \times 10^8$
木星赤道半径 $R/\text{m}$	$7.1398 \times 10^7$
木星球谐系数 $J_2$	$1.469657 \times 10^{-2}$
时间步长 $\tau/\text{s}$	60

## 1.2 视觉相机测量模型

在探测器绕飞过程中, 天然卫星无法返回测距数据, 固将LOS矢量作为观测变量。且探测器姿态信息已知, 因此不考虑坐标系的转换对观测变量的大小的影响, 即认为  $\mathbf{C}_i^b$ 、 $\mathbf{C}_i^s$  是单位阵, 模型可写作为

$$\mathbf{y}_k = \begin{pmatrix} \frac{\mathbf{r}_{1k} - \mathbf{r}_k}{\|\mathbf{r}_{1k} - \mathbf{r}_k\|} \\ \frac{\mathbf{r}_{2k} - \mathbf{r}_k}{\|\mathbf{r}_{2k} - \mathbf{r}_k\|} \\ \frac{\mathbf{r}_{3k} - \mathbf{r}_k}{\|\mathbf{r}_{3k} - \mathbf{r}_k\|} \end{pmatrix} + \boldsymbol{\eta}_k = \mathbf{h}(\mathbf{x}_k) + \boldsymbol{\eta}_k \quad (9)$$

其中:  $\boldsymbol{\eta}_k$  为测量噪声, 统计特性为

$$E(\boldsymbol{\eta}_k) = \mathbf{0}, \text{Cov}(\boldsymbol{\eta}_k, \boldsymbol{\eta}_j^T) = \mathbf{R}_k \delta_{kj} \quad (10)$$

$$\delta_{kj} = \begin{cases} 1, & k = j \\ 0, & k \neq j \end{cases}$$

且过程噪声与测量噪声相独立, 即对任意  $k, j$  有

$$\text{Cov}(\mathbf{w}_k, \boldsymbol{\eta}_j^T) = \mathbf{0} \quad (11)$$

## 1.3 扩展卡尔曼滤波算法

结合动力模型式(4)和视觉相机测量模型式(9)建立的木星环绕自主导航模型为

$$\begin{cases} \mathbf{x}_k = \mathbf{f}(\mathbf{x}_{k-1}) + \mathbf{w}_k \\ \mathbf{y}_k = \mathbf{h}(\mathbf{x}_k) + \boldsymbol{\eta}_k \end{cases} \quad (12)$$

EKF最早由Stanley Schmidt提出, 是QLEKF滤波器设计和实现的基础, 也是最常用的非线性滤波器。利用EKF解算木星的环绕自主导航模型, 首先计算线

性化状态矩阵<sup>[19]</sup>为

$$\mathbf{F}_k = \left. \frac{\partial \mathbf{f}}{\partial \mathbf{x}} \right|_{\mathbf{x}_k} = \mathbf{I}_6 + \boldsymbol{\Phi}_k \boldsymbol{\tau} \quad (13)$$

其中

$$\boldsymbol{\Phi}_k = \left. \frac{\partial \phi}{\partial \mathbf{x}} \right|_{\mathbf{x}_k} = \begin{bmatrix} \mathbf{O}_{3 \times 3} & \mathbf{I}_{3 \times 3} \\ \mathbf{A}_{3 \times 3} & \mathbf{O}_{3 \times 3} \end{bmatrix} \quad (14)$$

$$\mathbf{A}_{3 \times 3} = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix} \quad (15)$$

代入矩阵  $\mathbf{A}_{3 \times 3}$  各元素的取值, 即

$$\begin{cases} A_{11} = \frac{\mu R^2 J_2}{r_k^5} \left[ \left( 7.5 \frac{z_k^2}{r_k^2} - 1.5 \right) - \frac{x_k^2}{r_k^2} \left( 52.5 \frac{z_k^2}{r_k^2} - 7.5 \right) \right] - \frac{\mu}{r_k^3} \left( 1 - \frac{3x_k^2}{r_k^2} \right) \\ A_{12} = A_{21} = \mu \frac{x_k y_k}{r_k^5} \left[ 3 - \frac{R^2 J_2}{r_k^2} \left( 52.5 \frac{z_k^2}{r_k^2} - 7.5 \right) \right] \\ A_{13} = A_{31} = \mu \frac{x_k z_k}{r_k^5} \left[ 3 - \frac{R^2 J_2}{r_k^2} \left( 52.5 \frac{z_k^2}{r_k^2} - 22.5 \right) \right] \\ A_{22} = \frac{\mu R^2 J_2}{r_k^5} \left[ \left( 7.5 \frac{z_k^2}{r_k^2} - 1.5 \right) - \frac{y_k^2}{r_k^2} \left( 52.5 \frac{z_k^2}{r_k^2} - 7.5 \right) \right] - \frac{\mu}{r_k^3} \left( 1 - \frac{3y_k^2}{r_k^2} \right) \\ A_{23} = A_{32} = \mu \frac{y_k z_k}{r_k^5} \left[ 3 - \frac{R^2 J_2}{r_k^2} \left( 52.5 \frac{z_k^2}{r_k^2} - 22.5 \right) \right] \\ A_{33} = -\frac{\mu R^2 J_2}{r_k^5} \left( 52.5 \frac{z_k^4}{r_k^4} - 45 \frac{z_k^2}{r_k^2} + 4.5 \right) - \frac{\mu}{r_k^3} \left( 1 - \frac{3z_k^2}{r_k^2} \right) \end{cases}$$

进行状态估计和误差协方差估计的时间更新为

$$\begin{cases} \hat{\mathbf{x}}_{k|k-1} = \mathbf{f}(\hat{\mathbf{x}}_{k-1}) \\ \hat{\mathbf{x}}_{k-1} = [\hat{x}_{k-1}, \hat{y}_{k-1}, \hat{z}_{k-1}, \hat{x}_{k-1}, \hat{y}_{k-1}, \hat{z}_{k-1}] \end{cases} \quad (16)$$

探测器实际运行过程中, 通过实时得到状态估计, 有助于探测器避免偏离预定轨道, 根据状态估计规划探测任务, 确保优良的探测条件和数据采集。

$$\mathbf{P}_{k|k-1} = \mathbf{F}_k \mathbf{P}_{k-1} \mathbf{F}_k^T + \mathbf{Q}_k \quad (17)$$

根据式(9)获得线性化测量矩阵, 其中:  $i = 1, 2, 3$ 。

$$\mathbf{H}_k = \left. \frac{\partial \mathbf{h}}{\partial \mathbf{x}} \right|_{\mathbf{x}_k} = \begin{bmatrix} h_{11} & h_{12} & h_{13} & 0 & 0 & 0 \\ h_{14} & h_{15} & h_{16} & 0 & 0 & 0 \\ h_{17} & h_{18} & h_{19} & 0 & 0 & 0 \\ h_{21} & h_{22} & h_{23} & 0 & 0 & 0 \\ h_{24} & h_{25} & h_{26} & 0 & 0 & 0 \\ h_{27} & h_{28} & h_{29} & 0 & 0 & 0 \\ h_{31} & h_{32} & h_{33} & 0 & 0 & 0 \\ h_{34} & h_{35} & h_{36} & 0 & 0 & 0 \\ h_{37} & h_{38} & h_{39} & 0 & 0 & 0 \end{bmatrix} \quad (18)$$

$$\left\{ \begin{array}{l} h_{i1} = \frac{(x_{ik} - x_k)^2 - r_{ik}^2}{r_{ik}^3}, h_{i2} = h_{i4} = \frac{(x_{ik} - x_k)(y_{ik} - y_k)}{r_{ik}^3} \\ h_{i3} = h_{i7} = \frac{(x_{ik} - x_k)(z_{ik} - z_k)}{r_{ik}^3}; h_{i5} = \frac{(y_{ik} - y_k)^2 - r_{ik}^2}{r_{ik}^3} \\ h_{i6} = h_{i8} = \frac{(y_{ik} - y_k)(z_{ik} - z_k)}{r_{ik}^3}, h_{i9} = \frac{(z_{ik} - z_k)^2 - r_{ik}^2}{r_{ik}^3} \end{array} \right. \quad (19)$$

得到的状态估计和估计误差协方差表达式<sup>[19]</sup>为

$$\mathbf{K}_k = \mathbf{P}_{k|k-1} \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^T + \mathbf{R}_k)^{-1} \quad (20)$$

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k (\mathbf{y}_k - h(\hat{\mathbf{x}}_{k|k-1})) \quad (21)$$

$$\mathbf{P}_k = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_{k|k-1} (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k)^T + \mathbf{K}_k \mathbf{R}_k \mathbf{K}_k^T \quad (22)$$

设 $\hat{\mathbf{Q}}_k$ 、 $\hat{\mathbf{R}}_k$ 为正定的对称矩阵, 在EKF算法作为真实过程噪声和测量噪声协方差阵 $\mathbf{Q}_k$ 、 $\mathbf{R}_k$ 的估计值。EKF算法流程见表2。

表2 EKF算法流程  
Table 2 EKF algorithm process

Extended Kalman filter	
1: 输入:	$(\hat{\mathbf{x}}_{k-1}, \mathbf{P}_{k-1}, \mathbf{y}_k, \hat{\mathbf{Q}}_k, \hat{\mathbf{R}}_k)$
2: 输出:	$\hat{\mathbf{x}}_k, \mathbf{P}_k, \tilde{\mathbf{y}}_k$
3: 状态一步预测值:	$\hat{\mathbf{x}}_{k k-1} = f(\hat{\mathbf{x}}_{k-1})$
4: 预测方差:	$\mathbf{P}_{k k-1} = \mathbf{F}_k \mathbf{P}_{k-1} \mathbf{F}_k^T + \hat{\mathbf{Q}}_k$
5: 增益:	$\mathbf{K}_k = \mathbf{P}_{k k-1} \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_{k k-1} \mathbf{H}_k^T + \hat{\mathbf{R}}_k)^{-1}$
6: 新息:	$\tilde{\mathbf{y}}_k = \mathbf{y}_k - h(\hat{\mathbf{x}}_{k k-1})$
7: 滤波更新:	$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_{k k-1} + \mathbf{K}_k \tilde{\mathbf{y}}_k$
8: 滤波方差:	$\mathbf{P}_k = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_{k k-1} (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k)^T + \mathbf{K}_k \hat{\mathbf{R}}_k \mathbf{K}_k^T$

## 2 基于智能滤波的木星环绕自主导航

### 2.1 Q学习扩展卡尔曼滤波器

Q学习是强化学习 (Reinforcement Learning, RL) 的一种有效实现方法, 它指导智能体通过探索和利用, 在具有随机性质的环境学习最优策略。在此过程中, 智能体执行某一动作并接收环境返回的反馈奖励 $R$ , 作为评估所采取动作的效益指标。Q值表示在给定状态下采取特定动作所能获得的预期累积奖励。智能体通过一定的行动选择策略, 使长期的累积奖励最大化。其基本工作原理<sup>[22]</sup>如图1所示。

将Q学习方法与EKF相结合, 得到Q学习扩展卡尔曼滤波器。QLEKF可在不确定的环境通过与其交互学习, 根据奖励值的反馈自适应地选择噪声协方差阵的取值, 达到比传统EKF更高的估计精度<sup>[14]</sup>。

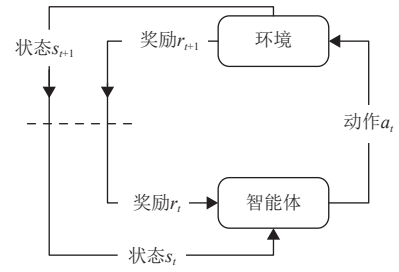


图1 强化学习原理图  
Fig. 1 Reinforcement learning principle diagram

#### 2.1.1 状态集和动作空间设计

将状态集 $S$ 和各状态下可采取的动作空间 $A$ 均定义为离散集合。构造状态集 $S$ 为 $M \times N$ 的网格, 包含 $M$ 个过程噪声协方差矩阵的设计值 $\{\hat{\mathbf{Q}}_1^i, \hat{\mathbf{Q}}_2^i, \dots, \hat{\mathbf{Q}}_M^i\}$ 和 $N$ 个测量噪声协方差矩阵的设计值 $\{\hat{\mathbf{R}}_1^j, \hat{\mathbf{R}}_2^j, \dots, \hat{\mathbf{R}}_N^j\}$ 。将状态集 $S$ 在一个具有 $M$ 行和 $N$ 列的平面上分成3个区域。

1) 角区域:

$$(S_{11}, S_{1N}, S_{M1}, S_{MN})$$

2) 边界区域:

$$(S_{1j}, j = 2, \dots, N-1; S_{i1}, i = 2, \dots, M-1) \\ (S_{Mj}, j = 2, \dots, N-1; S_{iN}, i = 2, \dots, M-1)$$

3) 中心区域:

$$(S_{ij}, i = 2, \dots, M-1, j = 2, \dots, N-1)$$

动作空间定义了智能体可以执行的状态转移, 即智能体可以无障碍地转移到相邻的状态或停留在当前状态。与状态集相对应的动作空间如图2所示。

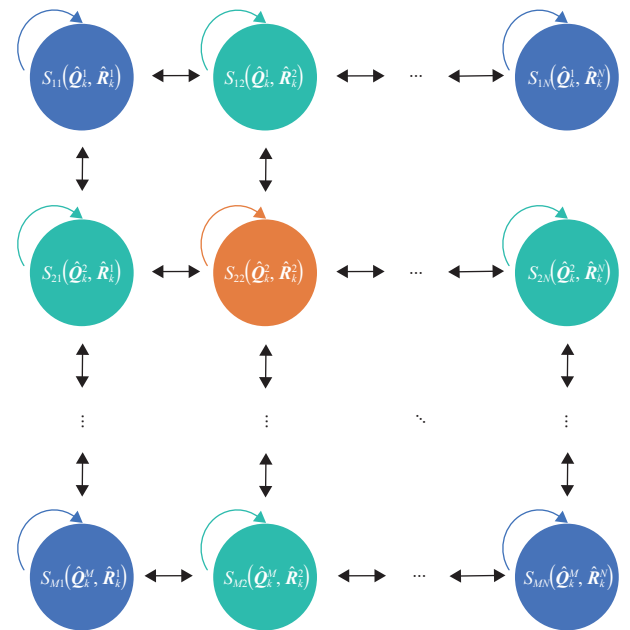


图2 状态集和动作转移  
Fig. 2 State set and action transition

### 2.1.2 动作选择策略

SoftMax策略根据Q值函数的大小来计算动作选择概率, 其数学表示为

$$P(a_i|s) = \frac{\exp\left\{\frac{Q(s, a_i)}{\text{temp}}\right\}}{\sum_{j=1}^m \exp\left\{\frac{Q(s, a_j)}{\text{temp}}\right\}} \quad (23)$$

其中:  $\{a_1, a_2, \dots, a_m\}$  为状态  $s$  下所有可采取的动作集合; 温度参数  $\text{temp}$  衡量智能体对于探索和利用的偏好,  $\text{temp} \rightarrow \infty$  时, SoftMax算法趋于仅探索,  $\text{temp} \rightarrow 0$  时趋于仅利用。

SoftMax与贪婪算法的不同之处在于, SoftMax策略可根据当前状态  $S(Mr, Nr)$  与周围状态的相对取值大小变更利用或探索的偏好, 如表3所示。即如果当前状态的Q值远大于其它相邻的状态, 则智能体以较大的概率停留在当前状态, 以小概率访问其它状态, 智能体倾向于利用; 而如果当前状态  $S(Mr, Nr)$  的Q取值与相邻的一个或多个状态相近, 智能体会以相对均等的概率进行动作选择, 此时智能体倾向于探索, 避免陷入局部最优的情况。但需要注意的是SoftMax策略中的指数函数会放大Q值, 因此在设计奖励函数和初始化时Q表格应注意Q值取值范围。

表3 SoftMax策略  
Table 3 SoftMax strategy

SoftMax策略:	
1:	function $\text{SoftMax}(S, A, Q(s, a), \text{temp})$
2:	for all $a_i \in A(s)$
3:	$P(a_i s) = \frac{\exp\left\{\frac{Q(s, a_i)}{\text{temp}}\right\}}{\sum_{j=1}^m \exp\left\{\frac{Q(s, a_j)}{\text{temp}}\right\}}$
4:	结束循环
5:	依概率 $P(a^* s)$ 选择动作 $a^*$
7:	结束
8:	返回 $a^*$
9:	结束函数

### 2.1.3 奖励函数设计

智能体执行动作后, 环境会对此做出响应, 并以奖励  $R$  的形式向智能体反映动作在短期内的好坏程度。本文基于Xiong等<sup>[14]</sup>的方法进行简化, 构造搜索EKF滤波器和估计EKF滤波器, 以实现QLEKF-single算法, 其原理如图3所示。

为评估给定状态-动作的性能, 将新息序列的对数函数  $R = -1.25 - 0.1 \times \lg(\|\tilde{y}_k^{sc}\|)$  作为奖励函数。新息序

列的模越大, 奖励函数越小, 当新息序列的模  $\|\tilde{y}_k^{sc}\| > 3.7267 \times 10^{-6}$  时, 奖励函数为负值, 说明执行该动作的收益较差。

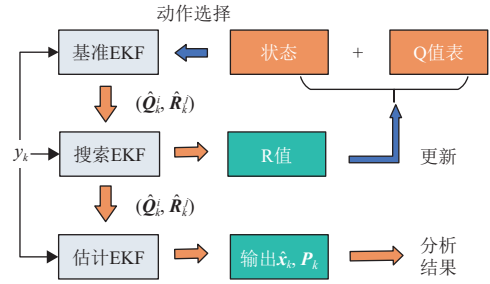


图3 QLEKF-single滤波器设计  
Fig. 3 QLEKF-single filter design

反之, 环境将通过奖励函数给予当前状态动作对正反馈。且  $\|\tilde{y}_k^{sc}\|$  越接近0值, 奖励函数  $R$  增加的幅度越大; 奖励函数越大, 说明选择  $\hat{Q}_k^i, \hat{R}_k^j$  作为噪声协方差阵的估计精度越高, QLEKF相较于EKF的滤波结果也更精确。

### 2.1.4 Q值函数和价值函数

Q值衡量累积奖励的大小, 该值每周期更新一次, 一步更新规则为

$$Q_{\text{new}}(s, a) \leftarrow (1 - \alpha) Q_{\text{old}}(s, a) + \alpha [R + \gamma V(s')] \quad (24)$$

执行动作  $a$  后到达状态  $s'$ , 基于状态  $s'$  所能实现的最大Q值记作价值函数, 即

$$V(s') = \text{Max}_a Q(s', a) \quad (25)$$

式(24)中,  $Q(s, a)$  为在  $s$  状态下执行动作  $a$  的Q值,  $R + \gamma V(s')$  表示本次行动  $a$  所能带来的奖励, 包括短期奖励  $R$  和长期奖励  $V(s')$ ;  $\alpha \in (0, 1)$  为学习率, 衡量Q函数更新奖励和继承旧值的比重;  $\gamma \in (0, 1)$  为折扣因子, 衡量长期奖励的重要性。

### 2.2 算法设计

根据状态和动作空间设计, 将探测器看作智能体, 结合EKF算法和Q学习原理构建QLEKFsingle-soft算法<sup>[14]</sup>。该算法流程见表4, QLEKFsingle-greedy算法与之相似, 只需更改动作选择策略, 将Step2中的SoftMax策略换成贪婪策略。QLEKF-single算法通过周期性地更新状态和动作, 搜索合适的过程噪声和观测噪声矩阵, 最终得到精确度较高的状态估计。

将非线性时变系统式(12)进行线性定常简化, 根据文献<sup>[10]</sup>中结论可知当系统连续两个时刻以上获得测量信息时, 可观性矩阵满秩, 非线性系统具有局部可观测性。

表 4 QLEKFsingle-soft算法流程  
Table 4 QLEKFsingle-soft algorithm process

**Algorithm QLEKFsingle-soft算法**

**输入:** 测量值 $y_k$ , 学习率 $\alpha$ , 折扣因子 $\gamma$ , 状态集 $S$ , 动作空间 $A$ , 周期长度 $T$ , 最大迭代次数 $k_{max}$

**输出:**  $\{\hat{x}_k\}, \{P_k\}, \{\bar{y}_k\}$  序列

1. 初始化估计值和误差方差矩阵:  $\hat{x}_0^{Se} = \hat{x}_0, P_0^{Se} = P_0$
2. 初始化Q值函数和价值函数:  $Q(s, a) = 0, V(s) = 0$
3. 初始化时间步 $k \leftarrow 0$ , 奖励值 $R \leftarrow 0$ , 生成随机数 $Mr, Nr$ , 则当前状态为 $s = S(Mr, Nr)$
4. **for 每个周期. do:**
5. 通过 $SoftMax(A, Q(s, a), s)$ 策略选择一个动作 $a$
6. 执行动作 $a$ , 到达新状态 $s'$
7. **for  $t = 1, 2, \dots, T$  (每个周期内的时间步), do:**
8.  $k \leftarrow k + 1$ , 根据状态 $s'$ 确定 $\hat{Q}_k^i$ 和 $\hat{R}_k^j$
9. 根据上一个时间步的状态估计、误差协方差矩阵、测量值和噪声协方差矩阵, 更新搜索EKF的状态估计、误差协方差矩阵和新息序列:  
 $[\hat{x}_k^{Se}, P_k^{Se}, \bar{y}_k^{Se}] \leftarrow EKF(\hat{x}_{k-1}^{Se}, P_{k-1}^{Se}, y_k, \hat{Q}_k^i, \hat{R}_k^j)$
10. 计算量化的奖励 $R$ :  $R \leftarrow R - 1.25 - 10^{-1} \log(\|\bar{y}_k^{Se}\|)$
11. 更新估计EKF的状态估计、误差协方差矩阵和新息序列:  
 $[\hat{x}_k, P_k, \bar{y}_k] \leftarrow EKF(\hat{x}_{k-1}, P_{k-1}, y_k, \hat{Q}_k^i, \hat{R}_k^j)$
12. **end for**
13. 更新Q值函数和价值函数:  
 $Q(s, a) \leftarrow Q(s, a) + \alpha [R + \gamma V(s') - Q(s, a)],$   
 $V(s') \leftarrow \max_a Q(s', a).$
14. 重置:  $s \leftarrow s', R \leftarrow 0.$
15. **end for**

表 5 探测器及卫星轨迹参数  
Table 5 Parameters of probe and satellite trajectories

木卫一	
初始位置/m	(4.214 3 × 10 <sup>8</sup> , 0, 0)
初始速度/(m·s <sup>-1</sup> )	(0, 1.734 6 × 10 <sup>4</sup> , 15.137 6)
半长轴/km	421 700
轨道倾角/(°)	0.05
离心率	0.0041
木卫四	
初始位置/m	(1.881 4 × 10 <sup>9</sup> , 0, 0)
初始速度/(m·s <sup>-1</sup> )	(0, 8.209 7 × 10 <sup>3</sup> , 29.373 7)
半长轴/km	1 882 709
轨道倾角/(°)	0.205
离心率	0.007 4
木卫十四	
初始位置/m	(2.216 8 × 10 <sup>8</sup> , 0, 0)
初始速度/(m·s <sup>-1</sup> )	(0, 2.391 3 × 10 <sup>4</sup> , 449.129 4)
半长轴/km	221 889
轨道倾角/(°)	1.076
离心率	0.0175
探测器	
初始位置/m	(6.705 6 × 10 <sup>8</sup> , 0, 0)
初始速度/(m·s <sup>-1</sup> )	(0, 1.375 1 × 10 <sup>4</sup> , 113.044 4)
半长轴/km	671 034
轨道倾角/(°)	0.471
离心率	0.009 4

### 3 仿真验证

为验证提出的算法, 本文选择木卫一、木卫四和木卫十四3颗天然卫星作为观测源, 探测器及卫星轨迹仿真参数如表5所示, 通过数值仿真进行验证, 探测器及木星卫星运行轨迹如图4所示, 仿真流程如图5所示。

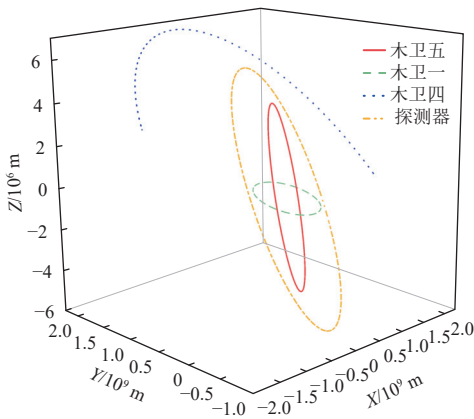


图 4 木星卫星与探测器轨迹图  
Fig. 4 Jupiter's moons and probe trajectory diagram

在木星环绕探测场景中, 对于 $\hat{Q}_k, \hat{R}_k$ 的估计可通过取定先验值 $Q_0, R_0$ 确定状态集的上、下界, 建立寻优网

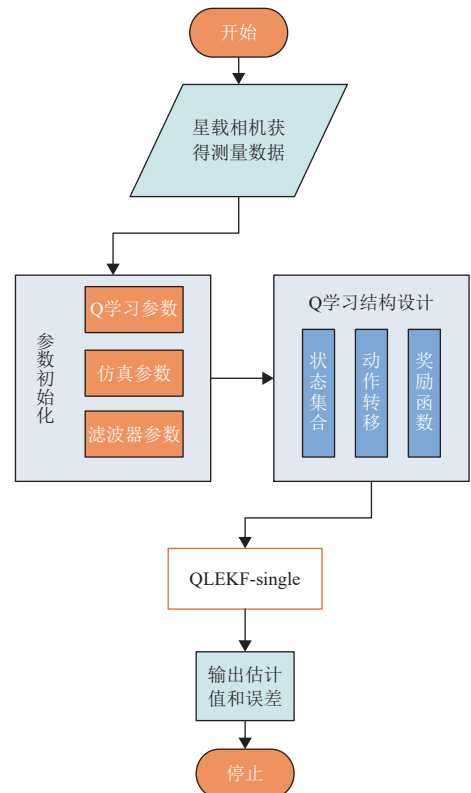


图 5 仿真流程图  
Fig. 5 Simulation flowchart

格。探测器通过周期性更新奖励函数和Q值,在网格中进行动作转移,训练出最优的 $\hat{Q}_k, \hat{R}_k$ 。

对状态集的上下界取对数,即

$$\begin{aligned} -\lg Q_0 &\leq \lg \hat{Q}_k^i \leq \lg Q_0, \\ -2\lg R_0 &\leq \lg \hat{R}_k^j \leq 2\lg R_0, \end{aligned}$$

并等距分为 $M \times N$ 的网格。如 $M=3$ 时,有 $[\hat{Q}_k^1, \hat{Q}_k^2, \hat{Q}_k^3] = (0.1Q_0, Q_0, 10Q_0)$ ,通过对数等距分割保证 $\hat{Q}_k^i, \hat{R}_k^j$ 取值搜索可高效进行,同时尽可能地考察较大的取值范围。

仿真中,假设每1min获取一次测量信息,需要考虑的仿真参数、模型误差和Q学习参数如表6所示。在表6的仿真条件下,在RAM 16.0 GB, Intel (R) Core (TM) i7-13620H CPU@ 2.40 GHz的计算机进行实验,得到状态估计结果如表7及图6~10所示。

表6 仿真参数表  
Table 6 Simulation parameter table

仿真设置	
仿真时间/s	10000
预测时间步长/s	60
循环周期	$T = 100$
探测器初始位置误差	$N(\mathbf{0}, 10\,000\text{ m}^2 \mathbf{I}_3)$
探测器初始速度误差	$N(\mathbf{0}, (1\text{ m/s})^2 \mathbf{I}_3)$
探测器测量误差	$N(\mathbf{0}, \sigma^2 \mathbf{I}_6)$ $\sigma = 1 \times 10^{-6}$
滤波器参数	
系统噪声方差阵:	
$\mathbf{Q}_0 = \begin{bmatrix} 160\,000\text{ m}^2 \mathbf{I}_2 & & & \\ & 10\,000\text{ m}^2 & & \\ & & 0.0064\text{ (m/s)}^2 \mathbf{I}_2 & \\ & & & 0.000\,025\text{ (m/s)}^2 \end{bmatrix}$	
测量噪声方差阵: $\mathbf{R}_0 = \sigma_R^2 \mathbf{I}_6, \sigma_R = 5 \times 10^{-6}$	
初始估计误差方差阵:	
$\mathbf{P}_0 = \begin{bmatrix} 10\,000\text{ m}^2 \mathbf{I}_3 & & \\ & (1\text{ m/s})^2 \mathbf{I}_3 & \end{bmatrix}$	
Q学习参数	
学习速率	$\alpha = 0.1$
折扣因子	$\gamma = 0.9$
状态集大小	$M = N = 9$
状态集边界	$10^{-1} Q_0 \leq \hat{Q}_k^i \leq 10^1 Q_0,$ $10^{-2} R_0 \leq \hat{R}_k^j \leq 10^2 R_0.$
随机动作选择概率	$\varepsilon = 0.1$
SoftMax温度参数	temp = 0.8

表7 程序运行时长  
Table 7 Runtime of the program

EKF/s	QLEKF/s	QLEKFsingle-greedy/s	QLEKFsingle-soft/s
4.765	6.059	5.155	5.153

本文采用位置估计的均方根误差作为导航精度的评价指标为

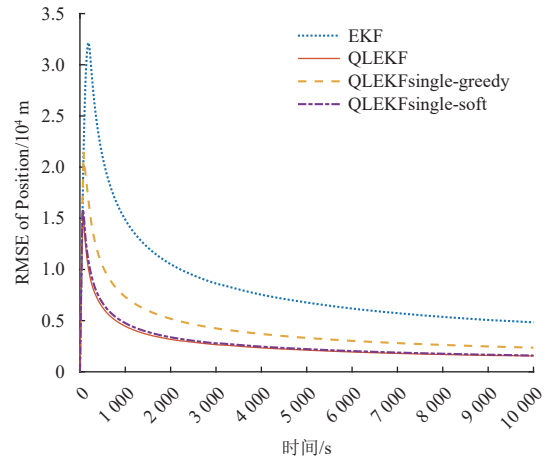


图6 滤波估计位置结果

Fig. 6 Filtering estimated position results

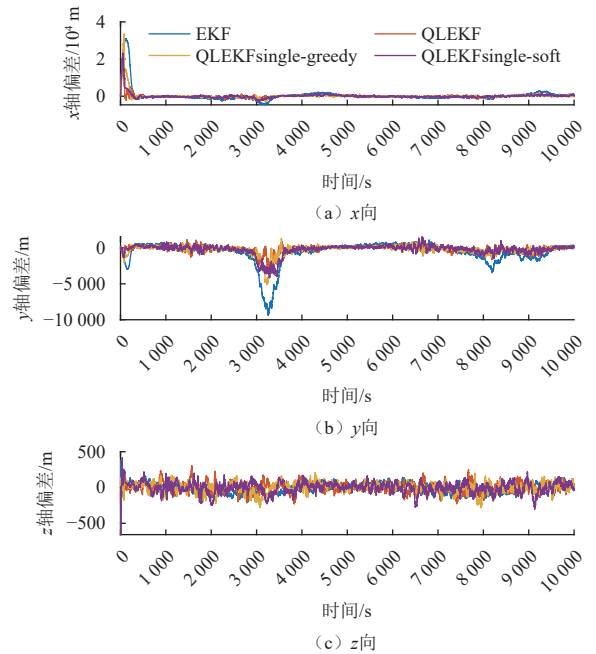


图7 3个方向状态估计位置误差

Fig. 7 Position error in three-dimensional state estimation

$$RMSE = \sqrt{\frac{1}{k} \sum_{i=1}^k (\mathbf{r}_i - \hat{\mathbf{r}}_i)^T (\mathbf{r}_i - \hat{\mathbf{r}}_i)} \quad (26)$$

4种不同滤波算法的性能如图6和表7所示,其中EKF运算效率最高、精度最低;QLEKF算法精度最高但程序运行时间最长,QLEKF-single算法相比QLEKF算法,运算时间减少约14%,且此时QLEKF-soft算法的精度与QLEKF相比变化不大。

沿3个坐标方向的位置和速度的估计偏差如图7和图8所示,X轴、Y轴和Z轴的位置偏差分别从 $10^4$ 、 $10^3$ 、 $10^2$  m数量级收敛到0值附近;速度误差均稳定在0值周围,震荡幅度不超过10 m/s。QLEKFsingle-soft算法位置估计和速度估计的误差曲线如图9和图10

所示, 估计偏差大部分时间都保持在 $2\sigma$ 范围内, 可以认为QLEKFsingle-soft算法适用当前木星环绕导航系统。

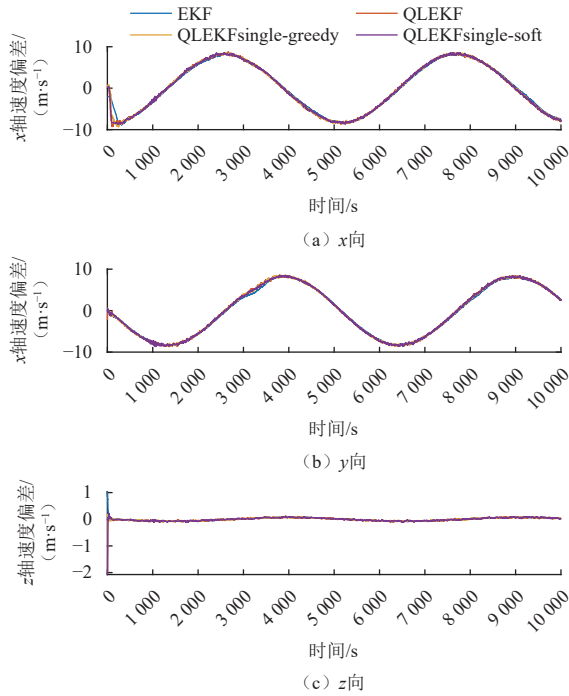


图8 三个方向状态估计速度误差

Fig. 8 Velocity error in three-directional state estimation

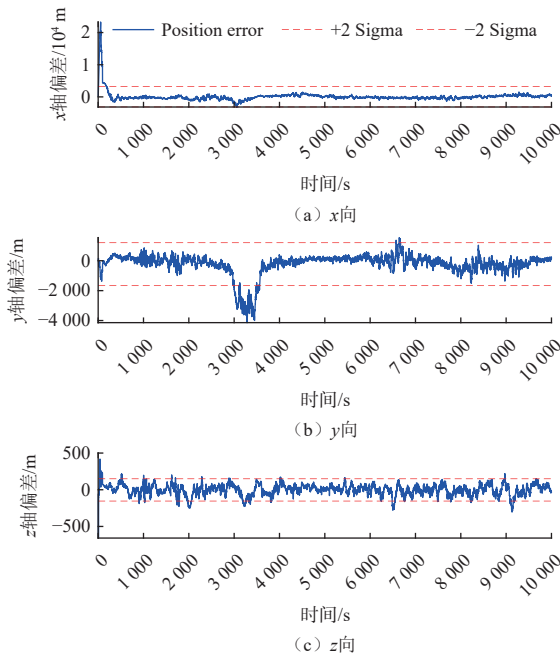


图9 QLEKFsingle-soft算法位置偏差

Fig. 9 Position bias of the QLEKFsingle-soft algorithm

在仿真过程中, 状态集大小的变化显著影响最后收敛的结果, 假设 $M = N$ , 分别计算20次实验3种滤波算法最后100次迭代的平均RMSE值, 结果见表8。

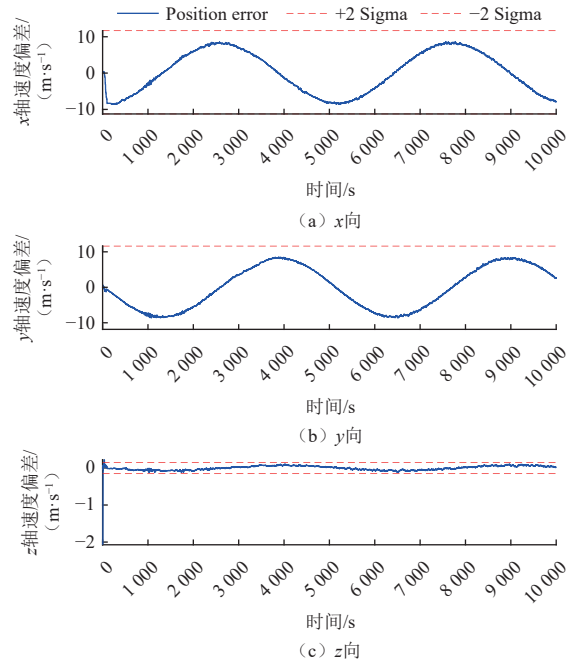


图10 QLEKFsingle-soft算法速度偏差

Fig. 10 Velocity bias of the QLEKFsingle-soft algorithm

表8 状态集大小变动时的状态估计位置RMSE

Table 8 RMSE of position estimation amidst variations in the state vector dimensionality

$N$	QLEKFsingle-soft/10 km	QLEKFsingle-greedy/10 km	EKF/10 km
3	0.222 8	0.222 8	0.563 5
6	0.275 6	0.296 5	0.438 8
9	0.306 6	0.370 0	0.513 6
12	0.390 4	0.413 9	0.470 3
15	0.395 3	0.485 1	0.551 3
18	0.391 3	0.499 4	0.519 8
21	0.592 9	0.556 5	0.593 5
...	...	...	...

状态集规模对QLEKFsingle-soft和QLEKFsingle-greedy算法性能, 总体而言QLEKFsingle算法优于EKF算法; 当状态集规模适中( $9 \leq N \leq 18$ )时, 因为Q值表得到充分更新, QLEKF-soft算法优于QLEKF-greedy算法, 状态 $s$ 下SoftMax策略以 $P^* = e^{Q(s,a^*)} / \sum_{a \in A} e^{Q(s,a^*)}$ 概率选择次优动作; 其中 $A$ 为状态 $s$ 下所有可采取的动作集合; 贪婪算法以概率 $(1 - \epsilon)/A(s)$ 选择最优动作, 次优动作则以较小概率被选中, SoftMax策略能够更稳健地得到全局最优解。当状态集较小时, 智能体可遍历每个状态且Q值表得到充分更新, 此时选用SoftMax策略和贪婪策略得到的结果精度相近; 状态集规模过大( $N \geq 21$ )时, Q值表为稀疏矩阵, SoftMax策略会放大当前时刻高价值行动的选择概率<sup>[23]</sup>, 并在Q值周期更新的过程中不断累积奖励, 使得智能体无法探索其他行动而造成较大误差, 采用贪婪策略得到的结果误差更小。

## 4 结 论

本文引入了QLEKF-single算法,作为木星环绕自主导航过程中模型噪声协方差矩阵的调整方案,并验证了可观测性和有效性,对比了不同的动作策略对于估计结果的影响。通过探测器绕木星飞行仿真实验,验证了SoftMax和贪婪算法都可以作为稳定有效的动作选择策略,且SoftMax算法在适中状态集大小下结果优于贪婪算法:它能在噪声协方差矩阵难以准确取定的情况下,在一定搜索范围内找到合适的噪声协方差矩阵值以提升QLEKF-single算法的性能,从而实现比EKF算法更精确的估计效果,在收敛后的位置和速度估计方面有了较大的改进。

在木星环绕任务场景中,由于星载计算能力有限,为减少并行滤波器对计算资源的占用,本文采用双滤波器并行运行的QLEKF-single算法。然而当探测器与木卫的几何关系、光照条件等动态变化时,采用贪婪策略可能陷入局部最优,进而导致噪声参数调整滞后。为此,引入了SoftMax动作选择策略,该策略通过概率化选择噪声协方差对(动作),平衡探索与利用的关系,增强算法在动态环境中的自适应能力。

在测量信息存在延迟、计算速度受限的情况下,如何在复杂空间环境中利用有限计算资源,实时估计探测器位置、速度和姿态,是下一步亟待解决的关键问题。未来,可以通过进一步优化并行滤波器设计和状态模型构建,以适应更加复杂多变的深空环境、完成更富有挑战性的自主导航任务。

## 参考文献

- [1] 邹昕,彭兢,缪远明. 木星系科学探测研究与展望[J]. *中国空间科学技术*, 2023, 43(6): 1-10.  
ZOU X, PENG J, MIU Y M. Research and prospect of scientific exploration for Jovian system[J]. *Chinese Space Science and Technology*, 2023, 43(6): 1-10
- [2] 高博宇,陈忠贵,周文艳. 国外木星探测任务进展与分析[J]. *航天器工程*, 2021, 30(5): 107-114.  
GAO B Y, CHEN Z G, ZHOU W Y. Progress and analysis of foreign Jovian exploration missions[J]. *Spacecraft Engineering*, 2021, 30(5): 107-114.
- [3] MELTZER M. Mission to Jupiter: a history of the Galileo project: NASA/SP-2007-4231[R]. USA: NASA, 2007.
- [4] STEPHENS S K. Juno at Jupiter: the mission and its path to unveiling secrets of the history of the solar system[C]//Proceedings of 2018 IEEE Aerospace Conference. Big Sky, MT, USA: IEEE, 2018.
- [5] HUSSMANN H, PALUMBO P, JAUMANN R, et al. JUICE JUpter ICy moons Explorer: exploring the emergence of habitable worlds around gas giants[EB/OL]. (2014)[2024-7-4]. <https://research.tudelft.nl/en/publications/juice-jupiter-icy-moons-explorer-exploring-the-emergence-of-habit>.
- [6] 吴季. 2016-2030年空间科学规划 研究报告[M]. 北京: 科学出版社, 2016.
- [7] 李佳宁,李杨,林海森,等. 基于光学/脉冲星的木星探测转移段自主组合导航[J]. *上海航天*, 2020, 37(4): 1-9.  
LI J N, LI Y, LIN H M, et al. Autonomous integrated navigation for jupiter probe transfer phase based on optical/pulsar[J]. *Aerospace Shanghai*, 2020, 37(4): 1-9.
- [8] 唐鹏. 测速测角组合自主导航性能评估研究[D]. 成都: 电子科技大学, 2016.  
TANG P. A research on the performance evaluation of autonomous navigation system[D]. Chengdu: University of Electronic Science and Technology of China, 2016
- [9] 王巍,那朝洋,冯文帅. 自主导航技术发展现状与趋势[J]. *航空学报*, 2021, 42(11): 525049.  
WANG W, XING C Y, FENG W S. State of the art and perspectives of autonomous navigation technology[J]. *Acta Aeronautica et Astronautica Sinica*, 2021, 42(11): 525049.
- [10] 鲁平,赵龙,陈哲. 改进的Sage-Husa自适应滤波及其应用[J]. *系统仿真学报*, 2007, 19(15): 3503-3505.  
LU P, ZHAO L, CHEN Z. Improved Sage-Husa adaptive filtering and its application[J]. *Journal of System Simulation*, 2007, 19(15): 3503-3505.
- [11] WANG H R, DENG Z H, FENG B, et al. An adaptive Kalman filter estimating process noise covariance[J]. *Neurocomputing*, 2017, 223: 12-17.
- [12] GE Q B, LI Y Y, WANG Y L, et al. Adaptive Kalman filtering based on model parameter ratios[J]. *IEEE Transactions on Automatic Control*, 2024, 69(9): 6230-6237.
- [13] XIONG K, WEI C L, ZHANG H Y. Q-learning for noise covariance adaptation in extended KALMAN filter[J]. *Asian Journal of Control*, 2021, 23(4): 1803-1816.
- [14] XIONG K, WEI C L. Q-learning-based target selection for bearings-only autonomous navigation[J]. *Journal of Systems Science and Complexity*, 2021, 34(4): 1401-1425.
- [15] 熊凯,魏春岭,李连升,等. 基于扩维QLEKF的脉冲星/星间定向组合导航[J]. *航空学报*, 2023, 44(3): 193-205.  
XIONG K, WEI C L, LI L S, et al. Pulsar/linter-satellite LOS integrated navigation based on augmented QLEKF[J]. *Acta Aeronautica et Astronautica Sinica*, 2023, 44(3): 526232.
- [16] DAI X, NATEGHI V, FOURATI H, et al. Q-learning-based noise covariance adaptation in Kalman filter for MARG sensors attitude estimation[C]//Proceedings of 2022 IEEE International Symposium on Inertial Sensors and Systems (INERTIAL). Avignon, France: IEEE, 2022.
- [17] TAO W J, ZHANG J X, HU H, et al. Intelligent navigation for the cruise phase of solar system boundary exploration based on Q-learning EKF[J]. *Complex & Intelligent Systems*, 2024, 10(2): 2653-2672.
- [18] 熊凯,魏春岭,郭建新. 航天器导航滤波器设计方法[M]. 北京: 中国宇航出版有限责任公司, 2022.
- [19] 王大轶,李茂登,黄翔宇,等. 航天器多源信息融合自主导航技术[M]. 北京: 北京理工大学出版社, 2018.  
WANG D Y, LI M D, HUANG X Y, et al. Spacecraft autonomous navigation technology based on multi-sources fusion[M]. Beijing: Beijing Institute of Technology, 2018.

- [20] 周海银, 王炯琦, 潘晓刚, 等. 卫星状态融合估计理论与方法[M]. 北京: 科学出版社, 2013.  
ZHOU H Y, WANG J Q, PAN X G, et al. Fusion theory and methods for satellite estimation[M]. Beijing: Beijing Science Press, 2013.
- [21] 曹建峰, 黄勇, 段建锋, 等. 木星探测器轨道计算中的动力学[J]. 宇航学报, 2020, 41(10): 1251-1258.  
CAO J F, HUANG Y, DUAN J F, et al. Dynamics in orbital determination of Jupiter spacecraft[J]. *Journal of Astronautics*, 2020, 41(10): 1251-1258.
- [22] 王琦, 杨毅远, 江季. Easy RL 强化学习教程[M]. 北京: 人民邮电出版社, 2022.
- [23] MIAHI E, MACQUEEN R, AYOUB A, et al. Resmax: an alternative soft-greedy operator for reinforcement learning[EB/OL]. (2016-12-16) [2024-7-4]. <https://arxiv.org/abs/1612.05628>.

作者简介:

**戴志雯**(2000-), 女, 硕士研究生, 主要研究方向: 航天器自主导航智能滤波算法。

通信地址: 湖南省长沙市开福区三一大道万煦园B9(410003)

**何章鸣**(1985-), 男, 副教授, 主要研究方向: 智能导航滤波算法、信息融合理论及其应用。**本文通信作者**。

通信地址: 湖南省长沙市开福区三一大道万煦园B9(410003)

E-mail: [hzmudt@163.com](mailto:hzmudt@163.com)

## Research on an Autonomous Navigation Method for the Orbital Phase of Jupiter Probe Based on QLEKF

DAI Zhiwen<sup>1</sup>, HOU Bowen<sup>1</sup>, ZHANG Yijie<sup>1</sup>, HE Zhangming<sup>1</sup>, LU Dandan<sup>2</sup>

(1. College of Science, National University of Defense Technology, Changsha 410073, China;

2. Deep Space Exploration Laboratory, Beijing 100195, China)

**Abstract:** To address the challenge of noise uncertainty affecting filtering performance in Jupiter's complex environment, an optical autonomous navigation scheme was established based on the relative line-of-sight information from multiple Jovian moons, using a simplified QLEKF (Q-Learning Extended Kalman Filter) algorithm with a single filter to estimate the position and velocity of the probe. The QLEKF-single (Single Filter Q-learning Extended Kalman Filter) designs a reward function based on the innovation of a single EKF filter. The Q-learning algorithm adaptively selected the values of the noise covariance matrix, while the SoftMax strategy was employed for action selection, ultimately achieving iterative system state estimation by EKF filter. Through simulation by randomly generating initial state estimates and measurement noise, the simplified model of Jupiter's real orbital dynamics was verified. It demonstrated that in scenarios with noise uncertainty, the QLEKF-single algorithm effectively improved navigation accuracy compared to traditional filtering methods. Moreover, compared to the QLEKF algorithm, the run time was reduced by more than 10% with little change in accuracy.

**Keywords:** autonomous navigation; QLEKF; greedy algorithm; SoftMax algorithm; Jupiter orbit

### Highlights:

- Adopted the SoftMax strategy and transfer actions based on probabilities.
- Designed a logarithmic function of innovation as the reward function to simplify the parallel filters in QLEKF.
- Analyzed the impact of state set size on estimation results under different action selection strategies.

[责任编辑: 杨晓燕, 英文审校: 宋利辉]