Autonomous Navigation of Unmanned Aircraft Using Space Target LOS Measurements and QLEKF

Kai Xiong *, Peng Zhou and Chunling Wei

Abstract: An autonomous navigation method based on the fusion of INS (inertial navigation system) measurements with the line-of-sight (LOS) observations of space targets is presented for unmanned aircrafts. INS/GNSS (global navigation satellite system) integration is the conventional approach to achieving the long-term and high-precision navigation of unmanned aircrafts. However, the performance of INS/GNSS integrated navigation may be degraded gradually in a GNSS-denied environment. INS/CNS (celestial navigation system) integrated navigation has been developed as a supplement to the GNSS. A limitation of traditional INS/CNS integrated navigation is that the CNS is not efficient in suppressing the position error of the INS. To solve the abovementioned problems, we studied a novel integrated navigation method, where the position, velocity and attitude errors of the INS were corrected using a star camera mounted on the aircraft in order to observe the space targets whose absolute positions were available. Additionally, a QLEKF (Q-learning extended Kalman filter) is designed for the performance enhancement of the integrated navigation system. The effectiveness of the presented autonomous navigation method based on the star camera and the IMU (inertial measurement unit) is demonstrated via CRLB (Cramer–Rao lower bounds) analysis and numerical simulations.

Keywords: unmanned aircraft; autonomous navigation; LOS measurement; Q-learning; extended Kalman filter

1. Introduction

Recently, unmanned aircrafts flying at altitudes ranging from 20–100 km above sea level have received increasing interest in both the aeronautic and astronautic fields with respect to various applications [1–3]. As one of the key technologies of unmanned aircrafts, the precise navigation technique is essential for ensuring the feasibility and efficiency of the onboard guidance and control system [4–6].

Currently, the INS/GNSS integrated navigation system is widely used to provide information about the position, velocity and attitude of unmanned aircrafts, such as X-43A, X-51A and HTV-2. The INS/GNSS integration takes advantage of the consistently high accuracy of the GNSS and the short-term stability of the INS, which is promising for the achievement of precise navigation parameters [7–9]. However, the performance of INS/GNSS integrated navigation is easily affected by the disturbance of the GNSS signals. The use of INS alone is limited, as the navigation error increases with time, and it is sensitive to the initial errors [10,11].

To enhance the navigation performance, many researchers have focused on other autonomous navigation approaches [12–16]. A typical method is INS/CNS integration, where the accumulated errors of the gyroscopes in the IMU are compensated by the measurements of star sensors mounted on the aircraft [17–19]. Compared with the GNSS, one of the main characteristics of INS/CNS integrated navigation is that it relies on the optical signals from celestial bodies, such that the high-autonomous and anti-disturbance
abilities of the navigation system are retained. The main limitation of traditional INS/CNS integrated navigation is that the position error accumulation caused by the bias of the accelerators in the IMU cannot be suppressed efficiently. Thus, it also suffers from an increasing number of errors over long flights. Achieving precise navigation information for unmanned aircrafts with a large flight envelope in the GNSS-denied environment remains a challenging problem.

In order to address this problem, this paper presents an INS/Vision integrated navigation method based on the star camera, which is mounted on the aircraft in order to measure the LOS vectors of space targets with the available position-related information, such as low-orbit satellites with a known ephemeris [20]. Different from the traditional star sensor, the star camera captures images of both the low-orbit satellites and the background stars in the field of view (FOV), and the LOS vectors of the satellites in the inertial frame can be extracted directly using the star catalogue and the least square (LS) algorithm. The current star cameras have the ability to observe low-orbit satellites and stars simultaneously during daytime when attached to unmanned aircrafts in near space. The measurements of the IMU and the star camera can be fused with a navigation filter in order to estimate the kinematic state of the aircraft.

The benefits of the presented navigation method are rooted in three aspects. Firstly, for unmanned aircraft in near space, the observation of the space targets, which are widely distributed within the celestial sphere, is less sensitive to the effect of clouds in comparison with landmark-based navigation [21,22]. Secondly, as it is easier to accurately measure the LOS vector of the space target than the direction vector of the Earth center using current technology, the presented method is promising for the achievement of a high rate of accuracy compared to navigation approaches based on the starlight and horizon reference [23,24]. Thirdly, both the position and the attitude information of the unmanned aircrafts can be measured with the star camera, which is able to observe the space target and the starlight simultaneously.

The main contributions of this work are as follows. Firstly, a novel autonomous navigation method is presented to determine the position, velocity and attitude of the unmanned aircrafts. Secondly, the theoretical accuracy of the presented method is analyzed through the calculation of the Cramer-Rao lower bounds (CRLB) based on the visibility analysis of the space targets. Thirdly, a Q-learning extended Kalman filter (QLEKF), which was designed to improve the navigation accuracy of space-target-based INS/Vision integration, is discussed.

This paper is divided into five sections. Following this introduction, the state and observation models of space-target-based INS/Vision integrated navigation are established in Section 2. Then, the method of the visibility analysis is outlined in Section 3. In addition, the QLEKF structure and an explanation of the learning process are given in Section 4. To demonstrate the performance of the presented method, a CRLB analysis and simulation study are described in Section 5. Finally, a brief conclusion is offered in Section 6.

2. System Model
2.1. Main Idea

The presented navigation method is based on the observation of a space target LOS vector, as shown in Figure 1. The LOS vector of the space target in the inertial frame is acquired through the processing of the image captured by the star camera, which contains both the space targets with a known ephemeris and the background stars in the star catalog. As the observation of the space target LOS vector can be implemented by the star camera alone, it does not depend on the misalignments between the different instruments, such that the possible sources of error are reduced. The observations of the space target LOS vector and the starlight vector are used to determine the position and attitude of the aircraft. It should be mentioned that the starlight vectors of the stars in the inertial frame can be calculated with an accuracy of less than 1 mas. The number and magnitude of the observed stars depends on the FOV and the sensitivity of the star camera. Generally, a star
camera with an FOV on the order of $10^\circ \times 10^\circ$ and a sensitivity of about 8 Mv is feasible for navigation.

![Figure 1. The measurement of the space target LOS vector.](image1)

The space-target-based INS/Vision integrated navigation system is mainly composed of three parts: the star camera, the IMU and the navigation filtering algorithm, as shown in Figure 2. The IMU is composed of three orthogonal accelerometers and three orthogonal gyroscopes, which are used to measure the angular rate and the specific force of the unmanned aircraft.

![Figure 2. Space-target-based INS/Vision integrated navigation method.](image2)

The prediction of the aircraft’s position, velocity and attitude are obtained from the measurements taken by the gyroscopes and the accelerometers. The navigation filter, which is designed based on the state model and the observation model of the navigation system, is implemented in order to update the prediction of the position, velocity and attitude based on the measurements by the star camera. As the position, velocity and attitude
of the unmanned aircraft are predicted using the IMU, with a high data update rate, the real-time accuracy can be guaranteed in the case when the exposure time of the star camera is recorded accurately. For the implementation of the navigation method, multiple space targets should be tracked by the ground stations, and the ephemeris of the space targets should be updated before the unmanned aircraft takes off. Then, the positions of the space targets can be obtained through orbital propagation during the flight of the aircraft.

### 2.2. Model for the Attitude Determination

The star camera and the gyroscopes are integrated in order to generate the attitude solution of unmanned aircrafts via the observation of the starlight so as to correct the attitude error and the gyroscope drift. The quaternion \( q_k = [q_{1k} \ q_{2k} \ q_{3k} \ q_{4k}]^T \) is used to describe the attitude of the aircraft [25,26]. For the attitude determination system composed of the star camera and the gyroscopes, the state vector \( x_{A,k} \) is considered as:

\[
x_{A,k} = [\delta \rho_k^T \ \delta b_k^T]^T
\]

where \( \delta b_k = b_k - \hat{b}_k \) is the estimation error of the gyroscope drift, \( b_k \) and \( \hat{b}_k \) are the gyroscope drift and its estimate, respectively, and the subscript \( k \) is the time label. \( \delta \rho_k = [\delta q_{1k} \ \delta q_{2k} \ \delta q_{3k}]^T \) is a part of the error quaternion \( \delta q_k = [\delta \rho_k^T \ \delta q_{4k}]^T \), where \( \delta q_{4k} = (1 - ||\delta \rho_k||^2)^{0.5} \). The operator \( || \cdot || \) represent the Euclidean norm. The error quaternion \( \delta q_k \) represents the difference between the estimated quaternion \( \hat{q}_k = [\hat{q}_{1k} \ \hat{q}_{2k} \ \hat{q}_{3k} \ \hat{q}_{4k}]^T \) and the true quaternion \( q_k \), which is expressed as:

\[
\delta q_k = q_k \otimes \hat{q}_k^{-1}
\]

or

\[
q_k = \delta q_k \otimes \hat{q}_k
\]

The operator \( \otimes \) represents the quaternion product, which is defined as:

\[
q_k \otimes \hat{q}_k^{-1} = [\Psi(q_k) \ q_k]q_k^{-1}
\]

with

\[
\hat{q}_k^{-1} = [-\hat{q}_{1k} \ \hat{q}_{2k} \ -\hat{q}_{3k} \ \hat{q}_{4k}]^T
\]

\[
\Psi(q_k) = \begin{bmatrix} q_{4k}I_{3} \times 3 - [\rho_k \times] \\ -\rho_k^T \end{bmatrix}
\]

where \( \rho_k = [q_{1k} \ q_{2k} \ q_{3k}]^T \), \( [\rho_k \times] \) is a cross-product matrix defined by

\[
[\rho_k \times] = \begin{bmatrix} 0 & -q_{3k} & q_{2k} \\ q_{3k} & 0 & -q_{1k} \\ -q_{2k} & q_{1k} & 0 \end{bmatrix}
\]

The discrete-time state model of the integrated navigation system for the attitude determination is expressed as:

\[
x_{A,k} = F_{A,k}x_{A,k-1} + w_{A,k}
\]

with

\[
F_{A,k} = \begin{bmatrix} I_{3} \times 3 - \tau_A [\omega_k \times] & -\tau_A^2 I_{3} \times 3 \\ 0_{3 \times 3} \\ I_{3} \times 3 \end{bmatrix}
\]

where \( \tau_A \) is the time step for the state prediction and \( \omega_k = [\omega_{xk} \ \omega_{yk} \ \omega_{zk}]^T \) is the estimated angular velocity, which can be obtained from the gyroscopes measurements after the compensation for the drift. The denotation \( [\omega_k \times] \) is formulated as:
\[
[\mathbf{\omega}_k \times] = \begin{bmatrix}
0 & -\omega_{zk} & \omega_{yk} \\
\omega_{zk} & 0 & -\omega_{xk} \\
-\omega_{yk} & \omega_{xk} & 0
\end{bmatrix}
\] (10)

The process noise \( w_{A,k} \) is modeled to obtain zero-mean Gaussian white noise with the covariance matrix \( Q_{A,k} \).

For simplicity, as the star sensor is mounted on the aircraft, the aircraft body frame and the star camera frame are not distinguished in this paper. For the attitude determination, the starlight vector on the aircraft body frame is observed using the star camera. The measurement is considered as:

\[
y_{A,k} = \begin{bmatrix}
\mathbf{u}^{(1)}_{Bk} - \mathbf{\hat{u}}^{(1)}_{Bk} \\
\mathbf{u}^{(2)}_{Bk} - \mathbf{\hat{u}}^{(2)}_{Bk} \\
\vdots \\
\mathbf{u}^{(N)}_{Bk} - \mathbf{\hat{u}}^{(N)}_{Bk}
\end{bmatrix}
\] (11)

where \( \mathbf{u}^{(j)}_{Bk} \) and \( \mathbf{\hat{u}}^{(j)}_{Bk} = [\mathbf{\hat{u}}^{(j)}_{Bxk} \mathbf{\hat{u}}^{(j)}_{Byk} \mathbf{\hat{u}}^{(j)}_{Bzk}]^T \) \( (j = 1, 2, \ldots, N) \) are the starlight vectors of the \( j \)-th star in the aircraft body frame and its estimate, while \( N \) is an integer, which denotes the number of visible stars in the FOV of the star camera.

To construct an observation model which relates the state vector to the measurements, the following function of the quaternion is introduced:

\[
\mathbf{u}^{(j)}_{Bk} = A(q_k)\mathbf{u}^{(j)}_{Ik}
\] (12)

where \( \mathbf{u}^{(j)}_{Ik} \) is the starlight vector of the \( j \)-th star in the inertial frame, which can be calculated according to the star catalog, while \( A(q_k) \) is the attitude transformation matrix from the inertial frame to the aircraft body frame, which is formulated as:

\[
A(q_k) = \left( q_k^2 - \rho_k^T \rho_k \right) I_{3 \times 3} + 2 \rho_k \rho_k^T - 2 q_k [\rho_k \times]
\] (13)

Accordingly, \( \mathbf{\hat{u}}^{(j)}_{Bk} \) is calculated as:

\[
\mathbf{\hat{u}}^{(j)}_{Bk} = A(\mathbf{\hat{q}}_k)\mathbf{u}^{(j)}_{Ik}
\] (14)

Substituting (3) into (12), we have:

\[
\mathbf{u}^{(j)}_{Bk} = A(\delta q_k \otimes \mathbf{\hat{q}}_k)\mathbf{u}^{(j)}_{Ik} = A(\delta q_k)A(\mathbf{\hat{q}}_k)\mathbf{u}^{(j)}_{Ik}
\] (15)

Inserting (14) into (15) yields:

\[
\mathbf{u}^{(j)}_{Bk} = A(\delta q_k)\mathbf{\hat{u}}^{(j)}_{Bk}
\] (16)

As the measurement accuracy of the star camera is on the order of a few arcsec, neglecting the high-order terms, the matrix \( A(\delta q_k) \) is simplified as:

\[
A(\delta q_k) = \begin{bmatrix}
1 & 2\delta q_{3k} & -2\delta q_{2k} \\
-2\delta q_{3k} & 1 & 2\delta q_{1k} \\
2\delta q_{2k} & -2\delta q_{1k} & 1
\end{bmatrix}
\] (17)

From (16) and (17), we obtain:

\[
\mathbf{u}^{(j)}_{Bk} - \mathbf{\hat{u}}^{(j)}_{Bk} = 2 [\mathbf{\hat{u}}^{(j)}_{Bk} \times] \delta q_k
\] (18)
where

\[
\mathbf{A}_{Bk}^{(i)} = \begin{bmatrix}
0 & -\dot{\theta}_{Bxk}^{(i)} & \dot{\theta}_{Byk}^{(i)} \\
\dot{\theta}_{Bxk}^{(i)} & 0 & -\dot{\theta}_{Bzk}^{(i)} \\
-\dot{\theta}_{Byk}^{(i)} & \dot{\theta}_{Bzk}^{(i)} & 0 \\
\end{bmatrix}
\] (19)

According to (11) and (18), for synchronously observed multiple stars, the measurement model is expressed as:

\[
y_{A,k} = H_{A,k}x_{A,k} + v_{A,k}
\] (20)

where

\[
H_{A,k} = \begin{bmatrix}
2 [\mathbf{A}_{Bk}^{(1)} \times] & 0_{3 \times 3} \\
2 [\mathbf{A}_{Bk}^{(2)} \times] & 0_{3 \times 3} \\
\vdots & \vdots \\
2 [\mathbf{A}_{Bk}^{(N)} \times] & 0_{3 \times 3} \\
\end{bmatrix}
\] (21)

The measurement noise \(v_{A,k}\) is modeled as Gaussian white noise with the zero mean and covariance matrix \(R_{A,k}\).

Using the system model shown in (8) and measurement model shown in (20), it is convenient to design a Kalman filter (KF) in order to estimate the state vector \(x_{A,k}\) by means of the measurement \(y_{A,k}\). Then, the estimates of \(\delta p_k\) and \(\delta b_k\) are used to correct the attitude quaternion \(\hat{q}_k\) and gyroscope drift \(\dot{b}_k\). The procedure of the attitude determination using the Kalman filtering algorithm is similar to that in [27], and it is omitted here for simplicity.

2.3. Model for the Position Estimation

In space-target-based INS/Vision integration, the space target LOS vectors observed using the star camera are used to correct the INS position and velocity errors in the inertial frame. The state vector for the position estimation consists of position vector \(r_k = [r_{xk} \ r_{yk} \ r_{zk}]^T\), velocity vector \(v_k\) and accelerometer bias \(\nabla_k\), which is expressed as:

\[
x_{P,k} = \begin{bmatrix}
r_k^T \\
v_k^T \\
\nabla_k^T \\
\end{bmatrix}^T
\] (22)

The state model is established to describe the dynamics of the state vector, which is represented as:

\[
x_{P,k} = f(x_{P,k-1}) + w_{P,k}
\] (23)

where \(w_{P,k}\) is the process noise with the covariance matrix \(Q_{P,k}\). The discretized nonlinear function \(f(\cdot)\) is formulated as:

\[
f(x_{P,k}) = \begin{bmatrix}
r_k \\
v_k \\
\nabla_k \\
\end{bmatrix} + \begin{bmatrix}
A^T(q_k)(f_{Bk} + \nabla_k) + g(x_{P,k}) \\
0 \\
\end{bmatrix}\tau_p
\] (24)

where \(\tau_p\) is the time step for the state prediction, \(q_k\) can be obtained through the method shown in the previous sub-section, \(f_{Bk}\) is the output of the accelerometers in the aircraft body frame, and \(g(x_{P,k})\) is a function used to describe the gravitational acceleration, which can be written as:

\[
g(x_{P,k}) = -\frac{\mu_E r_k}{||r_k||^3} + p(r_k)
\] (25)

where \(\mu_E\) is the gravitational constant of Earth and the function \(p(r_k)\) encapsulates the perturbation accelerations of the aircraft other than those that are cause by the two-body gravitational acceleration. The Jacobian matrix of the state model is derived as:

\[
F_{P,k} = \frac{\partial f(x_{P,k})}{\partial x_{P,k}} = I_{9 \times 9} + \Phi_{P,k}\tau_p
\] (26)
with
\[
\Phi_{P,k} = \begin{bmatrix}
0_{3 \times 3} & I_{3 \times 3} & 0_{3 \times 3} \\
\frac{\partial g(x_{P,k})}{\partial x_{P,k}} & 0_{3 \times 3} & A^T(q_k) \\
0_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 3}
\end{bmatrix}
\]

Let
\[
\frac{\partial g(x_{P,k})}{\partial x_{P,k}} = \begin{bmatrix}
\Phi_{P,k}(4,1) & \Phi_{P,k}(4,2) & \Phi_{P,k}(4,3) \\
\Phi_{P,k}(5,1) & \Phi_{P,k}(5,2) & \Phi_{P,k}(5,3) \\
\Phi_{P,k}(6,1) & \Phi_{P,k}(6,2) & \Phi_{P,k}(6,3)
\end{bmatrix}
\]

For simplicity, only the two-body gravitational acceleration is taken into consideration. The elements in the matrix shown in (28) are written as:

\[
\Phi_{P,k}(4,1) = -\frac{\mu_v}{\|r_k\|^3} \left(1 - \frac{3r_{yk}^2}{\|r_k\|^2}\right)
\]

(29)

\[
\Phi_{P,k}(4,2) = \frac{3\mu_v r_{yk} r_{yk}}{\|r_k\|^5}
\]

(30)

\[
\Phi_{P,k}(4,3) = \frac{3\mu_v r_{yk} r_{zk}}{\|r_k\|^5}
\]

(31)

\[
\Phi_{P,k}(5,1) = \Phi_{P,k}(4,2)
\]

(32)

\[
\Phi_{P,k}(5,2) = -\frac{\mu_v}{\|r_k\|^3} \left(1 - \frac{3r_{zk}^2}{\|r_k\|^2}\right)
\]

(33)

\[
\Phi_{P,k}(5,3) = \frac{3\mu_v r_{yk} r_{zk}}{\|r_k\|^5}
\]

(34)

\[
\Phi_{P,k}(6,1) = \Phi_{P,k}(4,3)
\]

(35)

\[
\Phi_{P,k}(6,2) = \Phi_{P,k}(5,3)
\]

(36)

\[
\Phi_{P,k}(6,3) = -\frac{\mu_v}{\|r_k\|^3} \left(1 - \frac{3r_{zk}^2}{\|r_k\|^2}\right)
\]

(37)

The space target LOS vector in the inertial frame is used as the measurement for the position estimation. Accordingly, the measurement model is expressed as:

\[
y_{P,k}^{(i)} = h^{(i)}(x_{P,k}) + v_{P,k}^{(i)}
\]

(38)

with

\[
h^{(i)}(x_{P,k}) = \frac{r_{T_k}^{(i)} - r_k}{\|r_{T_k}^{(i)} - r_k\|}
\]

(39)

where \(y_{P,k}^{(i)}\) is the measurement of the \(i\)-th space target LOS vector, \(r_{T_k}^{(i)}\) is the position vector of the \(i\)-th space target, which can be calculated according to the known ephemeris and orbit propagation, and \(v_{P,k}^{(i)}\) is the measurement noise with the covariance matrix \(R_{P,k}^{(i)}\). The corresponding Jacobian matrix is:

\[
H_{P,k}^{(i)} = \frac{\partial h^{(i)}(x_{P,k})}{\partial x_{P,k}} = -\frac{1}{\|r_{T_k}^{(i)} - r_k\|^3} I_{3 \times 3} + \frac{(r_{T_k}^{(i)} - r_k)(r_{T_k}^{(i)} - r_k)^T}{\|r_{T_k}^{(i)} - r_k\|^2} 0_{3 \times 3}
\]

(40)

According to the state model shown in (23) and the measurement model shown in (38), the extended Kalman filter (EKF) can be designed to fuse the accelerometer and star
camera data. Furthermore, an advanced filtering algorithm, which is presented in Section 4, can be used to enhance the navigation performance.

3. Principle of Visibility Analysis

The visibility analysis of the space targets is the basis for the performance evaluation of the integrated navigation method. In practice, only the visible space targets can be observed with the star camera in order to provide valuable navigation-related information. The user should confirm whether certain space targets are visible to the aircraft before the implementation of the navigation scheme. To facilitate the application, certain principles used to analyze the visibility of the space targets are described in this section.

First, for a space target to be visible, its apparent magnitude should be sufficient so as to be detected with the star camera, i.e.,

$$m_T^{(i)} < m_0$$  \hspace*{1cm} (41)

where $m_T^{(i)}$ is the apparent magnitude of the $i$-th space target and $m_0$ is a magnitude bound specified by the sensitivity of the star camera. To simplify the notation, the time label $k$ is omitted in this section. The apparent magnitude $m_T^{(i)}$ can be calculated as:

$$m_T^{(i)} = M_T^{(i)} + 2.5 \log_{10} \left[ \frac{r_{Sun} - r_T^{(i)}}{d_0} \| r - r_T^{(i)} \| \right]^2$$ \hspace*{1cm} (42)

where $M_T^{(i)}$ is the absolute magnitude of the space target, which is approximated by:

$$M_T^{(i)} = m_{Sun} - 5 \log_{10} \left( \frac{r_d^{(i)}}{d_0} \sqrt{a_l^{(i)}} \right)$$ \hspace*{1cm} (43)

where $m_{Sun} \approx -27$ is the apparent magnitude of Sun, $r_d^{(i)}$ is the observed body radius, $a_l^{(i)}$ is the body albedo of the space target and $d_0 \approx 1.4959787 \times 10^8$ km is an astronomical unit. $r_{Sun}$ is the position vector of Sun. $p(\xi^{(i)})$ is a function to describe the integration of the reflected light, which is approximated by:

$$p(\xi^{(i)}) = \frac{2}{3\pi} \left[ (\pi - \xi^{(i)}) \cos(\xi^{(i)}) + \sin(\xi^{(i)}) \right]$$ \hspace*{1cm} (44)

where $\xi^{(i)}$ is the sun–target–aircraft angle of the $i$-th space target, which is expressed as:

$$\xi^{(i)} = \cos^{-1} \left[ \frac{r_{Sun} - r_T^{(i)}}{\| r_{Sun} - r_T^{(i)} \| \| r - r_T^{(i)} \|} \right]$$ \hspace*{1cm} (45)

It describes the relative positions of the sun, the space target and the unmanned aircraft, as shown in Figure 3.

![Figure 3. Relative positions of the Sun, space target and unmanned aircraft.](image-url)
Secondly, a necessary condition for the visibility of the space target is that it is not hidden by the shadow of Earth, which can be judged with the following condition:

$$\psi^{(i)} < 90^\circ + \beta^{(i)}$$

(46)

where $$\psi^{(i)}$$ is the Sun–Earth–target angle of the $$i$$-th target, which is expressed as:

$$\psi^{(i)} = \cos^{-1}\left(\frac{r_T^{(i)} \cdot r_{Sun}}{\|r_T^{(i)}\| \cdot \|r_{Sun}\|}\right)$$

(47)

From Figure 4, the angle bound $$\beta^{(i)}$$ is calculated as:

$$\beta^{(i)} = \sin^{-1}\left[\left(\frac{\|r_T^{(i)}\|^2 - R_e^2}{\|r_T^{(i)}\|}\right)^{0.5}\right]$$

(48)

where $$R_e$$ is the Earth’s radius. The space target is illuminated by the Sun and can be seen as a candidate observation object when it is not within the shadow of the Earth.

![Figure 4](image_url)  
Figure 4. Earth’s shadow in the orbit of space target.

Thirdly, for the space target to be visible, it must not be occulted by celestial bodies, such as the Earth, Sun and Moon. A space target is believed to be occulted when its LOS vector points to a celestial body. Generally, the LOS vector of a space target cannot be measured effectively by the star camera in this case, as the images of the space target and the celestial body overlap. For example, the relative positions of the space target, the aircraft and Earth are shown in Figure 5.

![Figure 5](image_url)  
Figure 5. Relative positions of the Earth, aircraft and target.
From the figure, to judge whether the space target is occulted by the Earth, the following condition is given:

\[ \theta^{(i)} > \theta_0 \]  

(49)

where \( \theta^{(i)} \) is the Earth–aircraft–target angle, which is expressed as:

\[ \theta^{(i)} = \cos^{-1}\left( \frac{r - r_T^{(i)}}{\|r - r_T^{(i)}\| \cdot \|r\|} \right) \]  

(50)

The angle bound \( \theta_0 \) is calculated as:

\[ \theta_0 = \sin^{-1}\left( \frac{R_e}{\|r\|} \right) \]  

(51)

Similar conditions can be derived to judge whether the space target is occulted by the Sun or Moon.

4. Q-Learning Extended Kalman Filter
4.1. Q-Learning for the Filter Design

In this section, we describe an advanced navigation filter that was designed to estimate the position and velocity of the unmanned aircraft using the IMU and the star camera mounted on the aircraft in cases where the GNSS is unavailable. The EKF is a classic algorithm used to fuse the measurements of different sensors. An important procedure for the optimization of the EKF is the fine-tuning of the filtering parameters, such as the noise covariance matrices. However, the pre-determined noise covariance matrices based on the prior information may be not appropriate for aircrafts during flight, which will deteriorate the state estimation. In order to address this problem, we attempted to tune the process noise covariance with the measurement data online using the Q-learning approach so as to achieve a higher accuracy of the navigation system.

The Q-learning is a kind of reinforcement learning (RL) approach for solving sequential decision problems, where the agent interacts with an initially unknown environment and optimizes its behavior in order to maximize its cumulative reward [28–31]. Q-learning has received considerable attention, with many successful applications in various fields, such as gaming, robots, industrial process control and network management. Figure 6 illustrates the main components of Q-learning. An agent takes action \( a \) in the environment described by the state \( s \), which may cause the change in the environment’s state (from \( s \) to \( s' \)), and receives an immediate reward \( R(s, a) \) from the environment as a feedback on the action. The learning objective is to modify the action selection policy such that the cumulative reward, which is represented by the Q function \( Q(s, a) \), is maximized. Q-learning has been studied as an important data-based approach for adaptive control and optimal filters and is preferred for practical systems with uncertain models [32–34].

Figure 6. Main components of the iterative learning procedure.
In this study, to deal with the continuous state and action spaces, a Q-learning method with function approximation is adopted, where the Q function is expressed as:

\[ Q_i(s, a) = \Phi^T(s, a) \Theta_i \]  

(52)

with

\[
\Phi(s, a) = \begin{bmatrix} \phi_1(s, a) \\ \phi_2(s, a) \\ \vdots \\ \phi_D(s, a) \end{bmatrix}, \quad \Theta_i = \begin{bmatrix} \theta_{1i} \\ \theta_{2i} \\ \vdots \\ \theta_{Di} \end{bmatrix}
\]

(53)

where \( \phi_d(s, a) (d = 1, 2, \ldots, D) \) are basis functions, \( \theta_{di} (d = 1, 2, \ldots, D) \) are weights and the subscript \( i \) is the number of iterations. In this paper, the basis function is chosen as:

\[
\phi_d(s, a) = \exp \left( -[s - \mu_{sd} \ a - \mu_{ad}] \left[ \frac{\sigma_{sd}^2}{\sigma_{ad}^2} \ s - \mu_{ad} \right] \right) \sum_{i=1}^{D} \exp \left( -2[s - \mu_{sr} \ a - \mu_{ar}] \left[ \frac{\sigma_{sr}^2}{\sigma_{ar}^2} \ s - \mu_{ar} \right] \right) \frac{\mu_{sr}}{\sigma_{sr}}
\]

(54)

where \( \mu_{sd}, \mu_{ad}, \sigma_{sd} \) and \( \sigma_{ad} \) are the parameters of the basis function. It should be mentioned that the choice of the basis function is determined by the user, and a poor choice may lead to a poor filtering performance. In this study, the basis function is designed with the structure similar to that in [35], such that the appropriate action can be activated. The incremental weight update rule of Q-learning is given by:

\[
\Theta_{i+1} = \Theta_i + \alpha \left[ R(s, a) + \gamma \max_{a'} \left( \Phi^T(s', a') \Theta_i \right) - \Phi^T(s, a) \Theta_i \right] \Phi(s, a)
\]

(55)

where \( 0 \leq \alpha < 1 \) is the learning rate, \( 0 < \gamma \leq 1 \) is the discount factor and \( a \) and \( \gamma \) are the tuning parameters. It is easy to observe from the previous equations that:

\[
\Phi^T(s, a) \Theta_{i+1} = \Phi^T(s, a) \Theta_i + \alpha \left[ R(s, a) + \gamma \max_{a'} \left( \Phi^T(s', a') \Theta_i \right) - \Phi^T(s, a) \Theta_i \right] \Phi^T(s, a) \Phi(s, a)
\]

(56)

or

\[
Q_{i+1}(s, a) = Q_i(s, a) + \alpha \left[ R(s, a) + \gamma \max_{a'} Q_i(s', a') - Q_i(s, a) \right]
\]

(57)

The convergence of the iterative Q function is studied in the following theorem, which can help readers to comprehend the key idea of the learning procedure.

**Theorem 1:** Considering the iterative Q function sequence \( \{Q_i(s, a), i = 1, 2, \ldots\} \), if the initial Q function \( Q_0(s, a) \) satisfies the following condition:

\[
0 \leq Q_0(s, a) \leq R(s, a) + \gamma \max_{a'} Q_0(s', a')
\]

(58)

and there exist the constants \( \delta \geq 1 \geq \delta \geq 0 \) and \( \lambda \geq 0 \), such that the following inequalities hold:

\[
\bar{Q}_0(s, a) \leq Q_0(s, a) \leq \bar{Q}_0(s, a)
\]

(59)

\[
\lambda R(s, a) \geq \max_{a'} \bar{Q}_0(s', a')
\]

(60)

where

\[
\bar{Q}_0(s, a) = R(s, a) + \max_{a'} \bar{Q}_0(s', a')
\]

(61)

then the iterative Q function \( Q_i(s, a) \) converges to the optimal Q function \( Q^\ast(s, a) \) as \( i \to \infty \), i.e.,

\[
\lim_{i \to \infty} Q_i(s, a) = Q^\ast(s, a)
\]

(62)
where the function $Q^*(s, a)$ is formulated as:

$$Q^*(s, a) = R(s, a) + \gamma \max_{a'} Q^*(s', a')$$  \hspace{1cm} (63)$$

The proof of the theorem is provided in Appendix A. It shows that, under certain conditions, the iterative Q function $Q_i(s, a)$ is convergent to the optimal Q function $Q^*(s, a)$, which represents the maximum cumulative reward. Generally, if the constants $\delta$ and $\lambda$ are sufficiently large, and $\delta$ is sufficiently small, the conditions shown in (59) and (60) can be satisfied.

4.2. Filter Algorithm

Based on the system model for space-target-based INS/Vision integrated navigation, we designed the QLEKF algorithm, where the Q-learning is coupled with the EKF for the fine-tuning of the filter parameters with the objective of improving the estimation accuracy. The process noise covariance matrix was chosen for tuning, as it is crucial for calculating the filtering gain for the information contained in the system model and for the measurement to be utilized reasonably. In the Q-learning approach, the state $s$ is related to the current value of the process noise covariance matrix. For example, it can be assigned as 0 for the nominal noise covariance matrix. The action $a$ is related to the magnitude so as to enlarge or reduce the current process noise covariance matrix. It can be assigned as different numbers, such as 0, 1, 2, ..., for different degrees of tuning. The reward $R(s, a)$ is constructed based on the innovation of the EKF, which is a well-known indication of the estimation accuracy of the filter. Generally, a smaller innovation indicates a superior filtering performance, and vice versa.

For the system model shown in (23) and (38), the QLEKF algorithm is formulated in Algorithm 1. For simplicity, the subscript $i$ of the weight vector $\Theta_i$, representing the number of iterations, is omitted here.

---

**Algorithm 1:** Q-learning extended Kalman filter

1. $\hat{x}_{P,0}^B = \hat{x}_{P,0}^E = \hat{x}_{P,0}, P_{P,0}^B = P_{P,0}^E \triangleright$ Initialization
2. $k \leftarrow 0$
3. $\Theta \leftarrow 0$
4. for each period, do
5. for all $a \in A$, do
6. $C^B \leftarrow 0, C^E \leftarrow 0$
7. for $t = 1, 2, \ldots, T$, do
8. $k \leftarrow k + 1$
9. $[\hat{x}_{P,k}^B, P_{P,k}^B, y_{P,k}^B] \leftarrow$ EKF$(\hat{x}_{P,k-1}^B, P_{P,k-1}^B, y_{P,k}, Q_{P,k}, R_{P,k}) \triangleright$ Benchmark filter
10. $[\hat{x}_{P,k}^E, P_{P,k}^E, y_{P,k}^E] \leftarrow$ EKF$(\hat{x}_{P,k-1}^E, P_{P,k-1}^E, y_{P,k}, Q_{P,k}^{(s,a)}, R_{P,k}) \triangleright$ Exploring filter
11. $[\hat{x}_{P,k}, P_{P,k}, y_{P,k}] \leftarrow$ EKF$(\hat{x}_{P,k-1}, P_{P,k-1}, y_{P,k}, Q_{P,k}^{(s,a)}), R_{P,k}) \triangleright$ Main filter
12. $C^B \leftarrow C^B + \frac{1}{T-1} \left[ (y_{k}^B)^T R_k^{-1} y_{k}^B - C^B \right]$  
13. $C^E \leftarrow C^E + \frac{1}{T-1} \left[ (y_{k}^E)^T R_k^{-1} y_{k}^E - C^E \right]$  
14. end for
15. $R(s,a) \leftarrow C^B - C^E \triangleright$ Calculation of Reward
16. $\Theta \leftarrow \Theta + \alpha \begin{bmatrix} R(s,a) + \gamma \max_{a'} (\Phi^T(s',a')\Theta) - \Phi^T(s,a)\Theta \end{bmatrix} \Phi(s,a) \triangleright$ Update of weight
17. $\hat{x}_{P,k} \leftarrow \hat{x}_{P,k}^B, P_{P,k} \leftarrow P_{P,k}^B \triangleright$ Resetting of exploring filter
18. end for
19. $a_{max} \leftarrow \arg \max_{a} \left( \Phi^T(s,a)\Theta \right) \triangleright$ Selection of the best action
20. end for
21. return $\{\hat{x}_{P,k}\}$ and $\{P_{P,k}\} \triangleright$ Result of state estimation
In Algorithm 1, \( A \) is the pre-determined action space to be explored, \( T \) is a positive number representing the time interval for the calculation of the immediate reward \( R(s, a) \), \( \hat{x}_{P,k} \) and \( P_{P,k} \) are the state estimate and its corresponding error covariance matrix of the main filter, and \( \hat{y}_{P,k} \) is the innovation. Note that, in regard to the main filter, the process noise covariance matrix \( \hat{Q}_P^{s,a_{\text{max}}} \) tuned through Q-learning is used instead of \( Q_{P,k} \) in the system model.

It can be seen from the algorithm that there are another two parallel filters in addition to the main filter, where the benchmark filter and the exploring filter are distinguished by the superscripts B and S. The difference between the benchmark filter and the exploring filter is that the process noise covariance matrix of the benchmark filter \( Q_{P,k}^{s,a} \) is obtained from the prior knowledge of the system, while that of the exploring filter \( Q_{P,k}^{s,a_{\text{max}}} \) is altered by enlarging or reducing \( Q_{P,k} \).

The statistical values of the innovations obtained from the benchmark filter and the exploring filter are compared in order to construct the immediate reward \( R(s, a) \). The output of the QLEKF is the estimation result of the main filter. For clarity, the subroutine of the EKF is shown in Algorithm 2.

**Algorithm 2: Extended Kalman filter**

```plaintext
1: function EKF(\( \hat{x}_{P,k-1}, P_{P,k-1}, y_{P,k}, Q_{P,k}, R_{P,k} \))
2: \( \hat{x}_{P,k|k-1} \leftarrow f(\hat{x}_{P,k-1}) \) \( \triangleright \) Prediction
3: \( P_{P,k|k-1} \leftarrow F_{P,k}P_{P,k-1}F_{P,k}^T + Q_{P,k} \)
4: \( K_{P,k} \leftarrow P_{P,k|k-1}H_{P,k}^T \left( H_{P,k}P_{P,k|k-1}H_{P,k}^T + R_{P,k} \right)^{-1} \)
5: \( \hat{y}_{P,k} \leftarrow y_{P,k} - h(\hat{x}_{P,k|k-1}) \)
6: \( \hat{x}_{P,k} \leftarrow \hat{x}_{P,k|k-1} + K_{P,k}\hat{y}_{P,k} \) \( \triangleright \) Update
7: \( P_{P,k} \leftarrow (I - K_{P,k}H_{P,k})P_{P,k|k-1}(I - K_{P,k}H_{P,k})^T + K_{P,k}R_{P,k}K_{P,k}^T \)
8: return \( \hat{x}_{P,k}, P_{P,k}, \hat{y}_{P,k} \)
9: end function
```

In Algorithm 2, \( K_{P,k} \) denotes the Kalman gain. It is easy to see from the algorithm that the EKF is designed based on the system model shown in Section 2.3. The innovation \( \hat{y}_{P,k} \) as the output of Algorithm 2, is not used in the next iteration, but it is used to construct the immediate reward \( R(s, a) \) in Algorithm 1.

5. Simulations

5.1. Simulation Conditions

Simulations were performed to evaluate the performance of the space-target-based INS/Vision integrated navigation. An ideal trajectory was produced for the unmanned aircraft with an altitude of about 30 km and velocity of about 680 m/s, including the true position, velocity and attitude. The noises of the IMU and the star camera were added to the ideal values to obtain the simulated measurement data. For the data generation of the gyroscope, the angle random walk, the rate random walk and the bias were taken into consideration. The LEO satellites (low Earth orbit) with a semi-major axis of 6938 km and inclination of 55° were seen as the observed space targets. The construction of the satellite constellation is shown in Figure 7.
The parameters for the integrated navigation simulation are listed in Table 1.

**Table 1. Basic simulation configuration.**

| Parameter                                      | Value                      |
|-----------------------------------------------|----------------------------|
| Initial condition                             |                            |
| Initial position error (200 m, 200 m, 200 m)  |                            |
| Initial velocity error (0.02 m/s, 0.02 m/s, 0.02 m/s) |                            |
| Initial attitude error (0.01°, 0.01°, 0.01°)   |                            |
| Duration of navigation process                | 2 h                        |
| Measurement performance                       |                            |
| Accuracy of gyroscope                         | 0.02°/hour                 |
| Accuracy of accelerometer                    | 100 µg                     |
| Accuracy of star camera                       | 5"                         |
| Sensitivity of star camera                    | 8 Mv                       |
| Measurement update frequency                  | 1 Hz                       |
| Observed target in each period                | 1                          |
| Total number of space target                  | 180                        |
| QLEKF parameter                               |                            |
| Learning rate                                 | 0.1                        |
| Discount factor                               | 0.9                        |
| Parameter of basis function μ_{ad}            | (−2, −1, 0, 1, 2)          |
| Parameter of basis function σ_{ad}            | 0.3                        |
| Tuning of filtering parameter Q^{P,ad}_{P,k}  | 10^{2α} Q^{P,k}           |

It should be pointed out that the estimation performance and computational effort of the algorithm depend on the QLEKF parameters. In particular, it can be seen from the simulation results that the QLEKF parameters shown in Table 1 yield an improved performance, and the convenience of calculation is retained. Further works are planned in which we will perform a rigorous comparison analysis for the estimation performance using different QLEKF parameters. The absolute position, velocity and attitude of the aircraft are estimated with the method shown in Sections 2 and 4 in cases where the positions of the space targets are known. The comparison between the state estimation and the ideal trajectory yields the navigation results.

5.2. CRLB Based on the Visibility Analysis

The visibility of the space targets for the star camera on the aircraft is analyzed using the principles shown in Section 3. The curves of the apparent magnitude of some space targets over time are shown in Figure 8. It can be seen from the figure that, as the specified
sensitivity of the star camera in Algorithm 1 is 8 Mv, some space targets are bright enough to be observed in certain areas of the orbit.

![Figure 8. Apparent magnitudes of the space targets.](image)

The curves of the Sun-Earth-target angles of some space targets over time are shown in Figure 9. The angle bound $90^\circ + \beta^{(i)}$, representing Earth’s shadow, is plotted as the shadow in the figure. From this result, we can see that the space targets are not in Earth’s shadow for most of the time.

![Figure 9. Sun–Earth–target angles and the corresponding bounds.](image)

The curves of the Earth-aircraft-target angles, Sun-aircraft-target angles of some space targets over time and the corresponding angle bounds are shown in Figures 10 and 11, respectively. The figures illustrate that, sometimes, the space targets are not occulted by the Earth or Sun.
Based on the previous analysis, the potential performance of the presented navigation method was evaluated through the CRLB, which was derived from the system model. The CRLB provides a theoretical bound on the achievable accuracy of the navigation system. Thus, it is useful to study the feasibility of a novel navigation scheme. The calculation procedure of the CRLB can be found in the literature [36,37].

Figures 12 and 13 describe the theoretical bounds of the three-axis position estimation errors of the integrated navigation system with different noise levels of the accelerometer and the star camera in cases where only the visible space targets are observed. It indicates that a high navigation performance is achievable using the state-of-the-art sensors by applying the space-target-based INS/Vision integrated navigation method.
5.3. Simulation Results of the Navigation Filter

The navigation results of the space-target-based INS/Vision integration analysis obtained from the EKF are shown in Figures 14–16, including the curves of the position, velocity and attitude errors, where the solid lines represent the estimation error and the dashed lines represent the $3\sigma$ bound, calculated from the error covariance matrix of the EKF.

As expected, the position, velocity and attitude can all be estimated correctly by the measurement of the LOS vectors of the space targets and starlight vectors. The mean estimation errors of the position, velocity and attitude of the space-target-based INS/Vision integration were 12.96 m, 0.32 m/s and 0.94", respectively. This shows that the presented method has the potential to realize the absolute navigation of unmanned aircrafts with a high accuracy. For the implementation of the navigation method, the position error of the space target should be sufficiently less than the accuracy specification of the unmanned aircraft. In the simulation, only the measurements of the visible space targets are used to update the state vector. In the case where there are no visible space targets, the state vector...
is predicted by the IMU. The effect of the visibility is not distinct, as there are multiple space targets, as shown in Figure 7.

Figure 14. Position error of space-target-based INS/Vision integrated navigation.

Figure 15. Velocity error of space-target-based INS/Vision integrated navigation.

Figure 16. Attitude error of space-target-based INS/Vision integrated navigation.
Next, the ability of the QLEKF to improve the navigation accuracy for aircraft navigation through space-target-based INS/Vision integration was assessed. Monte Carlo simulations were conducted to evaluate the performance of the QLEKF in comparison with the traditional EKF and adaptive extended Kalman filter (AEKF) [38]. The QLEKF aims to fine-tune the noise covariance matrix obtained from the prior knowledge, while the AEKF aims to estimate the noise covariance matrix together with the state vector. Figures 17 and 18 show the norms of the root mean squared errors (RMSE) of the position and attitude estimates obtained from the EKF, the AEKF and the QLEKF.

![Figure 17. RMSE of the position estimation for EKF, AEKF and QLEKF.](image)

![Figure 18. RMSE of the attitude estimation for EKF, AEKF and QLEKF.](image)

According to the simulation results, the EKF provides similar results to the AEKF, while the QLEKF outperforms both the EKF and the AEKF with higher precision on average. These results demonstrate that the Q-learning approach is valuable for fine-tuning the critical parameters of the navigation filter.

6. Conclusions

This paper presents a novel navigation method that can be used to estimate the position, along with velocity and attitude, of unmanned aircrafts, using the star camera...
to observe the LOS vectors of the space targets with known positions. This method is especially suitable for aircrafts with long-term and high-accuracy autonomous navigation requirement. Based on the mathematical model and the visibility analysis, the CRLB of the integrated navigation system was calculated in order to validate the effectiveness of the presented method. In addition, the QLEKF was designed to enhance the navigation performance. The simulation results demonstrate the high performance of the space-target-based INS/Vision integrated navigation method. Further research will focus on the improvement of the navigation scheme in order to address the unfavorable effects of the space target position error and optical measurement error caused by thermal environment on the performance of the system.

Author Contributions: Conceptualization, K.X. and C.W.; methodology, K.X.; software, P.Z.; validation, K.X., P.Z. and C.W.; formal analysis, K.X.; writing—original draft preparation, K.X.; writing—review and editing, P.Z.; supervision, C.W.; project administration, K.X.; funding acquisition, C.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Civil Aerospace Advance Research Project, grant number D020403, and the National Natural Science Foundation of China, grant number U21B6001.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Proof of Theorem 1. First, to facilitate the analysis of the learning property, consider a specific case. In the case, where \( \gamma = 1 \), with the initial value \( Q_0(s,a) = Q_0(s,a) \), the iterative Q function shown in (57) can be written as:

\[
Q_{i+1}(s,a) = Q_i(s,a) + \alpha[R(s,a) + \max_{a'} Q_i(s',a') - Q_i(s,a)]
\]

(A1)

The mathematical induction is adopted to prove the following statement:

\[
[1 + (\prod_{j=0}^{i-1} \left(1 - \frac{\alpha}{\lambda + 1}\right))]Q_\infty(s,a) \leq Q_i(s,a) \leq [1 + (\prod_{j=0}^{i-1} \left(1 - \frac{\alpha}{\lambda + 1}\right))]Q_\infty(s,a)
\]

(A2)

For \( i = 0 \), from (A1) and the condition shown in (59), we have:

\[
Q_1(s,a) = Q_0(s,a) + \alpha[R(s,a) + \max_{a'} Q_0(s',a') - Q_0(s,a)] \leq (1 - \alpha)\delta Q_\infty(s,a) + \alpha R(s,a) + \delta_{\max_{a'} Q_\infty(s',a')}
\]

(A3)

From (60), it can be verified that:

\[
R(s,a) + \delta_{\max_{a'} Q_\infty(s',a')} \leq [1 + \frac{\lambda(\delta - 1)}{\lambda + 1}] [R(s,a) + \max_{a'} Q_\infty(s',a')]
\]

(A4)

Substituting (61) into (A4) yields:

\[
R(s,a) + \delta_{\max_{a'} Q_\infty(s',a')} \leq [1 + \frac{\lambda(\delta - 1)}{\lambda + 1}] Q_\infty(s,a)
\]

(A5)

It is derived from (A3) and (A5) that:

\[
Q_1(s,a) \leq [1 + (\delta - 1)(1 - \frac{\alpha}{\lambda + 1})] Q_\infty(s,a)
\]

(A6)
Similarly, we can obtain:

\[
\bar{Q}_1(s, a) \geq [1 + (\delta - 1)(1 - \frac{a}{\lambda + 1})] \bar{Q}_\infty(s, a) \tag{A7}
\]

According to (A6) and (A7), the inequality (A2) holds for \( i = 1 \). Assume that the inequality (A2) holds for \( i = l \), i.e.,

\[
\left[ 1 + \left( \prod_{j=0}^{l-1} (1 - \frac{a}{\lambda + 1}) \right) (\delta - 1) \right] \bar{Q}_\infty(s, a) \leq \bar{Q}_l(s, a) \leq \left[ 1 + \left( \prod_{j=0}^{l-1} (1 - \frac{a}{\lambda + 1}) \right) (\delta - 1) \right] \bar{Q}_\infty(s, a) \tag{A8}
\]

From (A1) and the condition shown in (59), we obtain:

\[
\bar{Q}_{l+1}(s, a) \leq (1 - \alpha) [1 + (\prod_{j=0}^{l-1} (1 - \frac{a}{\lambda + 1})) (\delta - 1)] \bar{Q}_\infty(s, a) + \alpha \{ R(s, a) + [1 + (\prod_{j=0}^{l-1} (1 - \frac{a}{\lambda + 1})) (\delta - 1)] \bar{Q}_\infty(s, a) \} \tag{A9}
\]

From (60), it can be verified that:

\[
R(s, a) + [1 + (\prod_{j=0}^{l-1} (1 - \frac{a}{\lambda + 1})) (\delta - 1)] \max_{a'} \bar{Q}_\infty(s', a') \leq [1 + \frac{\lambda}{\lambda + 1} (\prod_{j=0}^{l-1} (1 - \frac{a}{\lambda + 1})) (\delta - 1)] \bar{Q}_\infty(s, a) \tag{A10}
\]

Inserting (61) into (A10), we have:

\[
R(s, a) + [1 + (\prod_{j=0}^{l-1} (1 - \frac{a}{\lambda + 1})) (\delta - 1)] \max_{a'} \bar{Q}_\infty(s', a') \leq [1 + \frac{\lambda}{\lambda + 1} (\prod_{j=0}^{l-1} (1 - \frac{a}{\lambda + 1})) (\delta - 1)] \bar{Q}_\infty(s, a) \tag{A11}
\]

It follows from (A9) and (A11) that:

\[
\bar{Q}_{l+1}(s, a) \leq \left[ 1 + \left( \prod_{j=0}^{l} (1 - \frac{a}{\lambda + 1}) \right) (\delta - 1) \right] \bar{Q}_\infty(s, a) \tag{A12}
\]

Similarly, we can obtain:

\[
\bar{Q}_{l+1}(s, a) \geq \left[ 1 + \left( \prod_{j=0}^{l} (1 - \frac{a}{\lambda + 1}) \right) (\delta - 1) \right] \bar{Q}_\infty(s, a) \tag{A13}
\]

From (A12) and (A13), the inequality (A2) holds for \( i = l + 1 \). Hence, for \( i = 0, 1, 2, \ldots \), the statement (A2) holds. The mathematical induction is complete.

As \( 0 \leq \alpha < 1 \) and \( \lambda \geq 0 \), we have:

\[
\lim_{i \to \infty} \left( \prod_{j=0}^{l} \left( 1 - \frac{\alpha}{\lambda + 1} \right) \right) = 0 \tag{A14}
\]

This leads to:

\[
\lim_{i \to \infty} \left[ 1 + \left( \prod_{j=0}^{l-1} \left( 1 - \frac{\alpha}{\lambda + 1} \right) \right) (\delta - 1) \right] = \lim_{i \to \infty} \left[ 1 + \left( \prod_{j=0}^{l-1} \left( 1 - \frac{\alpha}{\lambda + 1} \right) \right) (\delta - 1) \right] = 1 \tag{A15}
\]

It is derived from (A2) and (A15) that:

\[
\lim_{i \to \infty} \bar{Q}_i(s, a) = \bar{Q}_\infty(s, a) \tag{A16}
\]

Secondly, the following statement is proven by mathematical induction:

\[
\bar{Q}_{i+1}(s, a) \geq \bar{Q}_i(s, a) \tag{A17}
\]
For $i = 0$, from (57) and the condition shown in (58), we get:

$$Q_1(s,a) = (1 - \alpha)Q_0(s,a) + \alpha[R(s,a) + \gamma\max_{a'}Q_0(s',a')] \geq (1 - \alpha)Q_0(s,a) + \alpha Q_0(s,a) = Q_0(s,a)$$  \hspace{1cm} (A18)

Thus, the statement (A17) holds for $i = 0$. Assume that the inequality (A17) holds for $i = l - 1$, i.e.,

$$Q_i(s,a) \geq Q_{l-1}(s,a)$$  \hspace{1cm} (A19)

It is derived from (57) and (A19) that:

$$Q_{l+1}(s,a) = (1 - \alpha)Q_l(s,a) + \alpha[R(s,a) + \gamma\max_{a'}Q_l(s',a')] \geq (1 - \alpha)Q_{l-1}(s,a) +$$

$$a[R(s,a) + \gamma\max_{a'}Q_{l-1}(s',a')] = Q_l(s,a)$$  \hspace{1cm} (A20)

It is seen from (A20) that the statement (A17) holds for $i = l$. Hence, for $i = 0, 1, 2, \ldots$, the statement (A17) holds. The mathematical induction is complete.

Finally, we prove the following statement by contradiction:

$$Q_i(s,a) \leq \overline{Q}_i(s,a)$$  \hspace{1cm} (A21)

For $i = 0$, from (57), (A1) and the value range of the parameter $0 < \gamma \leq 1$, we have:

$$Q_1(s,a) = (1 - \alpha)Q_0(s,a) + \alpha[R(s,a) + \gamma\max_{a'}Q_0(s',a')] \leq (1 - \alpha)Q_0(s,a) + \alpha[R(s,a) +$$

$$\max_{a'}Q_0(s',a')] = Q_0(s,a)$$  \hspace{1cm} (A22)

Thus, the statement (A21) holds for $i = 0$. Assume that the inequality (A21) holds for $i = l - 1$, i.e.,

$$Q_i(s,a) \leq \overline{Q}_i(s,a)$$  \hspace{1cm} (A23)

It follows from (57) and (A23) that:

$$Q_{l+1}(s,a) = (1 - \alpha)Q_l(s,a) + \alpha[R(s,a) + \gamma\max_{a'}Q_l(s',a')] \leq (1 - \alpha)Q_{l-1}(s,a) + \alpha[R(s,a) +$$

$$\max_{a'}Q_l(s',a')] \leq (1 - \alpha)\overline{Q}_l(s,a) + \alpha[R(s,a) + \max_{a'}\overline{Q}_l(s',a')] = \overline{Q}_{l+1}(s,a)$$  \hspace{1cm} (A24)

Thus, the statement (A21) holds for $i = l$. According to (A22) and (A24), for $i = 0, 1, 2, \ldots$, the statement (A21) holds. The mathematical induction is complete.

From (A17), (A21) and (A16), there exists a limit of $Q_i(s,a)$. Let

$$\lim_{i \to \infty} Q_i(s,a) = Q_{\infty}(s,a)$$  \hspace{1cm} (A25)

Finally, we prove the following statement by contradiction:

$$Q_{\infty}(s,a) = R(s,a) + \gamma\max_{a'}Q_{\infty}(s',a')$$  \hspace{1cm} (A26)

Assume that (A26) is false. Define the function

$$\varepsilon(s,a) = R(s,a) + \gamma\max_{a'}Q_{\infty}(s',a') - Q_{\infty}(s,a)$$  \hspace{1cm} (A27)

Let $\varepsilon$ be a small positive constant, such that the signs of $\varepsilon(s,a) + \varepsilon$ and $\varepsilon(s,a) - \varepsilon$ are the same. Considering (A25), there exists a positive integer $N$, such that the following inequality holds for $i > N$:

$$\varepsilon(s,a) + \varepsilon \geq R(s,a) + \gamma\max_{a'}Q_{\infty}(s',a') - Q_{\infty}(s,a) \geq \varepsilon(s,a) - \varepsilon$$  \hspace{1cm} (A28)
According to (57), the iterative Q function is written as:

\[
Q_1(s, a) = Q_0(s, a) + a[R(s, a) + \gamma \max_{a'} Q_0(s', a') - Q_0(s, a)]
\]

\[
Q_2(s, a) = Q_0(s, a) + a \sum_{j=0}^{1} [R(s, a) + \gamma \max_{a'} Q_j(s', a') - Q_j(s, a)]
\]

\[
Q_{i+1}(s, a) = Q_0(s, a) + a \sum_{j=0}^{i} [R(s, a) + \gamma \max_{a'} Q_j(s', a') - Q_j(s, a)]
\]

Similarly, \( Q_\infty(s, a) \) can be expressed as:

\[
Q_\infty(s, a) = Q_0(s, a) + a \sum_{j=0}^{N-1} [R(s, a) + \gamma \max_{a'} Q_j(s', a') - Q_j(s, a)] + a \sum_{j=N}^{\infty} [R(s, a) + \gamma \max_{a'} Q_j(s', a') - Q_j(s, a)]
\]

From (A28) and (A32), we obtain:

\[
Q_0(s, a) + a \sum_{j=0}^{N-1} [R(s, a) + \gamma \max_{a'} Q_j(s', a') - Q_j(s, a)] + a \sum_{j=N}^{\infty} [\epsilon(s, a) + \epsilon] \geq Q_\infty(s, a) \geq Q_0(s, a) + a \sum_{j=0}^{N-1} [R(s, a) + \gamma \max_{a'} Q_j(s', a') - Q_j(s, a)] + a \sum_{j=N}^{\infty} [\epsilon(s, a) - \epsilon]
\]

According to (A33), if \( \epsilon(s, a) > 0 \), then \( Q_\infty(s, a) \geq \infty \). On the contrary, if \( \epsilon(s, a) < 0 \), then \( Q_\infty(s, a) \leq -\infty \). This is in contradiction with the fact that there exists a limit of \( Q_i(s, a) \), as \( i \to \infty \). Thus, the assumption is false, and the equation (A26) holds. Note that the formulation of (A26) is essentially the same as that of (63). □

References

1. Chen, K.; Zhou, J.; Shen, F.Q.; Sun, H.Y.; Fan, H. Hypersonic boost-glide vehicle strapdown inertial navigation system/global positioning system algorithm in a launch-centered earth-fixed frame. Aerosp. Sci. Technol. 2020, 98, 1–17. [CrossRef]
2. Hu, G.; Gao, B.; Zhong, Y.; Ni, L.; Gu, C. Robust unscented kalman filtering with measurement error detection for tightly coupled INS/GNSS integration in hypersonic vehicle navigation. IEEE Access 2019, 7, 151409–151421. [CrossRef]
3. Cheng, X.; Yang, Y.; Hao, Q. Analysis of the effects of thermal environment on optical systems for navigation guidance and control in supersonic aircraft based on empirical equations. Sensors 2016, 16, 1717. [CrossRef] [PubMed]
4. Kahn, A.D.; Edwards, D.J. Navigation, guidance, and control of a micro unmanned aerial glider. J. Guid. Control Dyn. 2019, 42, 2474–2484. [CrossRef]
5. Suna, J.; Bloma, H.A.P.; Ellerbroeka, J.; Hoekstraab, J.M. Particle filter for aircraft mass estimation and uncertainty modeling. Transp. Res. Part C 2019, 105, 145–162. [CrossRef]
6. Hiba, A.; Gati, A.; Manecy, A. Optical navigation sensor for runway relative positioning of aircraft during final approach. Sensors 2021, 21, 2203. [CrossRef]
7. Hu, G.; Ni, L.; Gao, B.; Zhu, X.; Wang, W.; Zhong, Y. Model predictive based unscented Kalman filter for hypersonic vehicle navigation with INS/GNSS integration. IEEE Access 2020, 8, 4814–4823. [CrossRef]
8. Pan, W.; Zhan, X.; Zhang, X. Fault exclusion method for ARAIM based on tight GNSS/INS integration to achieve CAT-I approach. IET Radar Sonar Navig. 2019, 13, 1909–1917. [CrossRef]
9. Krasuski, K.; Cwiklak, J.; Jafarin, H. Aircraft positioning using PPP method in GLONASS system. Aircr. Eng. Aerosp. Technol. 2018, 90, 1413–1420. [CrossRef]
10. Jurevicius, R.; Marcinkevicius, V.; Šeibokas, J. Robust GNSS-denied localization for UAV using particle filter and visual odometry. Mach. Vis. Appl. 2019, 30, 1181–1190. [CrossRef]
11. Chen, D.; Neusyip, K.;Selezneva, M.; Mu, Z. New algorithms for autonomous inertial navigation systems correction with precession angle sensors in aircrafts. Sensors 2019, 19, 5016. [CrossRef]
12. Lei, X.; Liu, X. A high performance altitude navigation system for small rotorcraft unmanned aircraft. Mechatronics 2019, 62, 102248. [CrossRef]
13. Gui, M.; Zhao, D.; Ning, X.; Zhang, C.; DAI, M. A time delay/star angle integrated navigation method based on the event-triggered implicit unscented Kalman filter. IEEE Trans. Instrum. Meas. 2021, 70, 1–10. [CrossRef]
14. Zhang, H.; Jiao, R.; Xu, L.; Xu, C.; Mi, P. Formation of a satellite navigation system using X-ray pulsars. Publ. Astron. Soc. Pac. 2019, 131, 045002.
15. Wang, T.; Somani, A.K. Aerial-DEM geolocalization for GPS-denied UAS navigation. *Mach. Vis. Appl.* 2020, 31, 3. [CrossRef]

16. Ye, L.; Yang, Y.; Ma, J.; Deng, L.; Li, H. Research on an LEO constellation multi-aircraft collaborative navigation algorithm based on a dual-way asynchronous precision communication-time service measurement system (DWAPC-TSM). *Sensors* 2022, 22, 3213. [CrossRef]

17. Wang, R.; Xiong, Z.; Liu, J.; Shi, L. A new tightly-coupled INS/CNS integrated navigation algorithm with weighted multi-stars observations. *Proc. Inst. Mech. Eng. Part G J. Aerosp. Eng.* 2016, 230, 698–712. [CrossRef]

18. Ning, X.; Yuan, W.; Liu, Y. A tightly coupled rotational SINS/CNS integrated navigation method for aircraft. *J. Syst. Eng. Electron.* 2019, 30, 770–782.

19. Chen, H.; Gao, H.; Zhang, H. Integrated navigation approaches of vehicle aided by the strapdown celestial angles. *Int. J. Adv. Robot. Syst.* 2020, 5, 1–12. [CrossRef]

20. Xiong, K.; Wei, C.; Liu, L. Autonomous navigation for a group of satellites with star sensors and inter-satellite links. *Acta Astronaut.* 2013, 86, 10–23.

21. Huang, L.; Song, J.; Zhang, C.; Cai, G. Observable modes and absolute navigation capability for landmark-based IMU/Vision navigation system of UAV. *Optik* 2020, 202, 1–19. [CrossRef]

22. Kim, Y.; Bang, H. Vision-based navigation for unmanned aircraft using ground feature points and terrain elevation data. *Proc. Inst. Mech. Eng. Part G J. Aerosp. Eng.* 2018, 232, 1334–1346. [CrossRef]

23. Christian, J.A. Optical navigation using planet’s centroid and apparent diameter in image. *J. Guid. Control Dyn.* 2015, 38, 192–204. [CrossRef]

24. Chang, J.; Geng, Y.; Guo, J.; Fan, W. Calibration of satellite autonomous navigation based on attitude sensor. *J. Guid. Control Dyn.* 2017, 40, 185–191. [CrossRef]

25. Lefert, E.J.; Markley, F.L.; Shuster, M.D. Kalman filtering for spacecraft attitude estimation. *J. Guid. Control Dyn.* 1982, 5, 417–292. [CrossRef]

26. Markley, F.L. Attitude error representations for Kalman filtering. *J. Guid. Control Dyn.* 2003, 5, 311–317. [CrossRef]

27. Xiong, K.; Wei, C. Integrated celestial navigation for spacecraft using interferometer and earth sensor. *Proc. Inst. Mech. Eng. Part G J. Aerosp. Eng.* 2020, 234, 2248–2262. [CrossRef]

28. Sutton, R.S.; Barto, A.G. *Reinforcement Learning: An Introduction*; The MIT Press: London, UK, 2018.

29. Gosavi, A. Reinforcement learning: A tutorial survey and recent advances. *INFORMS J. Comput.* 2009, 21, 178–192. [CrossRef]

30. Xu, X.; Zuo, L.; Huang, Z. Reinforcement learning algorithms with function approximation: Recent advance and applications. *Inf. Sci.* 2014, 261, 1–31. [CrossRef]

31. Watkins, C.J.; Dayan, P. Q-learning. *Mach. Learn.* 1992, 8, 279–292. [CrossRef]

32. Wei, Q.; Lewis, F.L.; Sun, Q.; Yan, P.; Song, R. Discrete-time deterministic Q-learning: A novel convergence analysis. *IEEE Trans. Cybern.* 2017, 47, 1224–1237. [CrossRef]

33. Luo, B.; Wu, H.N.; Huang, T. Optimal output regulation for model-free quanser helicopter with multistep Q-learning. *IEEE Trans. Ind. Electron.* 2018, 65, 4953–4961. [CrossRef]

34. Xiong, K.; Wei, C.; Zhang, H. Q-learning for noise covariance adaptation in extended Kalman filter. *Asian J. Control* 2021, 23, 1803–1816. [CrossRef]

35. Ding, L.; Yang, Q. Research on air combat maneuver decision of UAVs based on reinforcement learning. *Avion. Technol.* 2018, 49, 29–35.

36. Ristic, B.; Farina, A.; Benvenuti, D.; Arulampalam, M.S. Performance bounds and comparison of nonlinear filters for tracking a ballistic object on re-entry. *IEEE P-Radar Sonar Navig.* 2003, 150, 65–70. [CrossRef]

37. Lei, M.; Wyk, B.J.; Qi, Y. Online estimation of the approximate posterior cramer-rao lower bound for discrete-time nonlinear filtering. *IEEE Trans. Aerosp. Electron. Syst.* 2011, 47, 37–57. [CrossRef]

38. Xiong, K.; Liu, L. Design of parallel adaptive extended Kalman filter for online estimation of noise covariance. *Aircr. Eng. Aerosp. Technol.* 2019, 91, 112–123. [CrossRef]