Improving the Real-Time Efficiency of Inertial SLAM and Understanding its Observability

Jonghyuk Kim  
ARC Centre of Excellence in Autonomous Systems  
The University of Sydney  
NSW 2006, Australia  
Email: jhkim@acfr.usyd.edu.au

Salah Sukkarieh  
ARC Centre of Excellence in Autonomous Systems  
The University of Sydney  
NSW 2006, Australia  
Email: salah@acfr.usyd.edu.au

Abstract—This paper addresses the computational complexity associated with 6DoF inertial SLAM by recasting the filter into an indirect or complimentary structure. By doing so the standard non-linear inertial SLAM algorithm is piecewise linearised, providing two significant breakthroughs. Firstly, through the process of linearisation, the filter estimates the errors in the inertial states (errors in position, velocity and attitude) as opposed to the inertial states themselves, thus significantly reducing the amount of prediction updates required and hence the computational burden generally associated with inertial SLAM. Secondly, due to the piecewise linear nature of the underlying model, a new level of understanding of the observability of SLAM can be carried out through the use of the Stripped Observability Matrix. This paper focuses on the 6DoF implementation of SLAM in an airborne vehicle. Results illustrating that the new filter structure can efficiently estimate the errors and provide a navigation solution which is comparable to the standard implementation with significantly less computational cost is provided, along with theoretical results on the observability of 6DoF SLAM.

I. INTRODUCTION

Most of the research conducted in Simultaneous Localisation and Mapping (SLAM) use estimation algorithms where there are two cycles, the prediction cycle and the estimation procedure (which handles the feature extraction and data association as well as the statistical estimation). One of the main goals undertaken by the global research community is to improve the computational efficiency of the algorithm and hence the focus is on map management techniques. The basic assumption is that the prediction cycle is not the computationally expensive component of the algorithm, but instead the goal is to limit the computation required by ensuring that the local map or update region is small enough to guarantee real-time computation. Given such conditions, the “direct state” formulation where the SLAM filter directly estimates the vehicle’s state and landmarks simultaneously is effective, especially given the situation where the vehicle navigates in a 2D environment and dead-reckoning sensors such as encoders and steering sensors can be implemented.

If however a situation is called for where SLAM is to work in a 3D environment and the vehicle undergoes 6DoF motion (almost all outdoor field robotic applications!), then the use of encoders and steering sensors to drive the vehicle model is not sufficient. Instead one is limited to the application of inertial navigation. A further benefit of implementing inertial navigation techniques is that the SLAM solution is independent of the type of vehicle since one is not required to develop a kinematic model and is simply left with the dynamic equations of a 6DoF system. In fact, inertial navigation techniques can, and have been applied, to the 2D case as well thus making it a versatile sensor suite.

In [3] the authors present a 6DoF SLAM implementation using an Inertial Navigation System (INS) and provide results for an Uninhabited Air Vehicle (UAV). The implementation was a direct SLAM structure using an Extended Kalman Filter (EKF), where the vehicle and landmark states were predicted and estimated together.

In this paper, the SLAM structure is recast into the indirect form. In this form the state transition matrix which drives the filter is an INS error model, which is developed from perturbation techniques on the non-linear INS equations. Since the error model represents the low dynamics of the INS error, the filter is not required to run at the same rate as the INS sampling rate which can reach up to several kHz. This form is not just relevant for inertial systems, but any kinematic model which can be cast into a linear error form (through perturbation techniques) will be able to implement and benefit from such a structure.

Figure 1 compares the direct SLAM structure with its indirect form. In both cases an Inertial Measurement Unit (IMU) provides the acceleration and rotation rate of the vehicle. The observation sensor provides the range, bearing and elevation to the landmarks. In direct SLAM the filter accepts the raw data from the IMU and passes this into a non-linear 6DoF model, and the EKF proceeds through the process of predicting the states of the vehicle and landmark locations. When the sensor observes a landmark, the estimation cycle proceeds [3]. In indirect SLAM however, the INS loop is separated from the filter, and so the inertial navigation equations transform the raw inertial data to position, velocity and attitude outside of the filter with sufficiently high rates. The estimation algorithm is a linear Kalman Filter (KF), and the state transition model is an error model of both the INS and the landmarks. When an observation occurs, a predicted observation is generated which is based on the current location of the vehicle and location of the landmark as dictated by the map. The difference between the predicted and actual observations
underpinning of this paper. Section VI provides the conclusion to the paper.

II. THE INERTIAL NAVIGATION ALGORITHM AND LANDMARK OBSERVATION MODEL

The INS algorithms are formulated in the Earth-Fixed Local-Tangent frame [3]. The acceleration in the body frame \( \mathbf{f}^b \) is read from the IMU. Its is converted to acceleration in the navigation frame \( \mathbf{f}^n \) via a Direction Cosine Matrix (DCM), \( \mathbf{C}_{ib}^n \). The DCM is evaluated from a quaternion vector \( \mathbf{q} \) which makes use of a four component vector to represent the attitude of the vehicle, and is determined from the rotation rate, \( \omega^b \), from the IMU. The acceleration in the navigation frame is then integrated to give position and velocity in the navigation frame (once acceleration due to gravity has been removed). The INS equations in continuous time are,

\[
\begin{bmatrix}
\dot{\mathbf{p}}^n \\
\dot{\mathbf{v}}^n \\
\dot{\mathbf{q}}^n
\end{bmatrix} = \begin{bmatrix}
\mathbf{C}_{ib}^n \mathbf{f}^b + \mathbf{g}^n \\
\frac{1}{2} \mathbf{q}^\top \otimes \dot{\mathbf{q}}^b
\end{bmatrix},
\]

where \( \otimes \) represents the quaternion product, and \( \dot{\omega}^b \) is the quaternion form of the rotation rate measured from the gyros.

The on-board range, bearing and elevation sensor provides relative observations between the vehicle and landmarks. The relationship between range, bearing, and elevation in the sensor frame and the location of the landmark in the navigation frame can be found in [3]. The \( i^{th} \) landmark position \( \mathbf{m}_i^n \) in the navigation frame is a function of the vehicle position \( \mathbf{P}^n \), the sensor lever-arm offset \( \mathbf{P}_{sb}^b \), and the relative distance \( \mathbf{P}_{ms}^s \) measured from the sensor in the sensor frame [3] and is

\[
\mathbf{m}_i^n = \mathbf{g}_1(\mathbf{P}^n, \psi^i, \mathbf{z}^i) = \mathbf{p}^n + \mathbf{C}_{ib}^n \mathbf{p}_{sb}^b + \mathbf{C}_{ib}^n \mathbf{C}_{sb}^n \mathbf{P}_{ms}^s(\mathbf{z}^i),
\]

where \( \psi \) is the Euler angle representation of the vehicle, \( \mathbf{z}^i \) is the range \((r)\), bearing \((\phi)\) and elevation \((\theta)\) measured from the on-board sensor, \( \mathbf{C}_{ib}^n \) is a DCM which transforms the vector in the sensor frame to the body frame, and \( \mathbf{P}_{ms}^s \) is the relative distance of the map from the sensor. \( \mathbf{P}_{ms}^s = [x \ y \ z]^T \) is computed from the range, bearing and elevation observations using a polar to cartesian transformation,

\[
\mathbf{p}_{ms}^s(\mathbf{z}^i) = \mathbf{g}_2(r, \phi, \theta) = \begin{bmatrix}
\rho \cos(\phi) \cos(\theta) \\
\rho \sin(\phi) \cos(\theta) \\
\rho \sin(\theta)
\end{bmatrix}.
\]

III. THE INDIRECT 6DoF SLAM

In the indirect SLAM structure, the filter estimates the errors in the INS and map position. By perturbing the INS, a linearised error model can be obtained. The resulting error equation in continuous time can be found in [4] and [6] and is

\[
\begin{bmatrix}
\delta \mathbf{p}^n \\
\delta \mathbf{v}^n \\
\delta \dot{\mathbf{q}}^n
\end{bmatrix} = \begin{bmatrix}
\mathbf{f}^n \times \delta \dot{\psi}^n + C_{ib}^n \delta \mathbf{f}^b \\
-C_{ib}^n \delta \dot{\psi}^b
\end{bmatrix}.
\]
In SLAM a stationary landmark dynamic model is used, hence its error model also becomes a constant model,

$$\delta m_{i-\ldots-N}^n = 0_{1\ldots-N},$$  

where $N$ is the number of landmarks.

A. System Dynamic and Observation Model

The system model in state space can now be constructed using equations 4 and 5.

$$\delta x(k+1) = F(k)\delta x(k) + G(k)w(k),$$  

where $\delta x$ is the error state vector, $F(k)$ is the linearised system transition matrix, $G(k)$ is the system noise input matrix and $w(k)$ is the system noise vector with noise covariance $Q(k)$. The error state vector is defined as the error in position, velocity and attitude of the INS and the error in map position:

$$\delta x(k) = [\delta p^n(k), \delta \psi^n(k), \delta m^n_{i-\ldots-n}(k)]^T.$$  

The observation model can be obtained in terms of the observation residual, or measurement differences, $\delta z(k)$ and the error states $\delta x(k)$,

$$\delta z(k) = H(k)\delta x(k) + v(k),$$  

where, $H(k)$ is the linearised observation matrix and $v(k)$ is the observation noise vector with noise covariance $R(k)$. The observation residuals are generated at each observation update by subtracting the measured range, bearing and elevation, $\tilde{z}(k)$, from the SLAM indicated range, bearing and elevation value, $\tilde{z}(k)$:

$$\delta z(k) = \tilde{z}(k) - z(k) = [\delta \rho \delta \varphi \delta \theta]^T.$$  

The predicted range, bearing and elevation are computed from the vehicle state and the associated landmark position by,

$$z^n(k) = h(p^n, \psi^n, m^n) = g^{-1}_1(p^n, \psi^n, m^n),$$  

where $g^{-1}_1$ and $g^{-2}_2$ are obtained from equations 2 and 3.

B. Prediction and Estimation

With the state and observation models defined in equations 6 and 8, the KF estimation procedure can proceed. The state and its covariance are predicted using the process noise input. The state covariance is propagated using the state transition model and process noise matrix by,

$$\delta x(k|k) = F(k)\delta x(k|k-1) = 0$$  

$$P(k|k-1) = F(k)P(k-1|k-1)F^T + G(k)Q(k)G^T(k).$$  

Not only is the linear prediction much simpler and computational more efficient then in the direct SLAM approach, but furthermore the predicted error $\delta x(k|k-1)$ is zero. This is because if one assumes that the only error in the vehicle and map states is zero mean Gaussian noise, then there is no error to propagate in the state prediction cycle, and the uncertainty in this assumption is provided in the covariance matrix propagation.

When an observation occurs, the state vector and its covariance are updated according to

$$\delta x(k|k) = \delta x(k|k-1) + W(k)v(k)$$  

$$P(k|k) = P(k|k-1) - W(k)S(k)W^T(k),$$  

where the innovation vector ($v(k)$), Kalman gain ($W(k)$), and innovation covariance ($S(k)$) are computed as,

$$v(k) = \delta z(k) - H\delta x(k|k-1) = \delta z(k)$$  

$$W(k) = P(k|k-1)H^T(k)S^{-1}(k)$$  

$$S(k) = H(k)P(k|k-1)H^T(k) + R(k).$$

For the same reason as in the prediction cycle, $H\delta x(k|k-1)$ is zero and hence is not computed.

Once the observation update has been processed successfully, the estimated errors of the INS and map are propagated to correct the actual vehicle and landmark states and the error state is set to zero, that is, $\delta x(k|k) = 0$.

C. Data Association and New landmark Augmentation

Whenever a landmark is observed a data association process is conducted which checks to see if the landmark has been previously observed. If the landmark has been previously registered in the external map, the observation is used to update the state and covariance, and if the landmark is a new one then a new landmark error state is augmented to the filter state.

The direct state SLAM performs this task by using the current estimates of the vehicle and landmarks within the filter. However, the state vector in indirect SLAM only maintains the error estimates it requires for aiding the INS and landmarks, both of which reside outside the filter. Given the current prediction of the INS and map errors along with their covariances, and the observation, a statistical data association can be performed using the Normalised Innovation Square (NIS). The NIS ($\gamma$) is computed by

$$\gamma = \delta z(k)^T S^{-1}(k)\delta z(k).$$

Given an innovation and its covariance with the assumption of Gaussian distribution, $\gamma$ forms a $\chi^2$ (chi-square) distribution. If $\gamma$ is less than a predefined threshold, then the observation and the landmark that were used to form the innovation are associated and the innovation is now used to update the state and covariance.

If the landmark is reobserved then the estimation cycle proceeds, otherwise it is a new landmark and must be added into the external map database. In the indirect SLAM filter, the error state is augmented with the new landmark error and the covariance matrix is augmented:

$$\delta x_{aug}(k|k) = G(k)\begin{bmatrix} \delta x(k|k) \\ \delta z(k) \end{bmatrix} = 0$$  

$$P_{aug}(k|k) = G(k)\begin{bmatrix} P(k|k) & 0 \\ 0 & R(k) \end{bmatrix} G(k)^T.$$
where $G(k)$ is

$$G(k) = \begin{bmatrix} I & 0 \\ \nabla g_{1,x}(k) & \nabla g_{1,z}(k) \end{bmatrix}, \quad (21)$$

with $\nabla g_{1,x}(k)$ and $\nabla g_{1,z}(k)$ are the Jacobians for the current state and observation respectively in equation 10.

IV. INDIRECT SLAM RESULTS

In [3] the SLAM algorithm using an INS in a direct filter structure is demonstrated and validated in a real-time UAV implementation, however the computational expense associated with that implementation was excessive. The results shown here provide a successful alternative means through the indirect or complementary SLAM implementation.

A vision sensor is simulated as part of the observation sensor. The vision sensor has good bearing accuracy but it can provide poor range data if the size of the landmark is known. It also has a limited Field-Of-View (FOV) below the vehicle. To overcome this poor range accuracy, a laser sensor is attached which can provide $0.5\,\text{m}$ accuracy. The sensor parameters are listed in table I and the specifications are from the actual flight test. The flight vehicle undergoes a figure of eight trajectory at approximately 100m above the ground. The flight time is 300 seconds and the average flight speed is $40\,\text{m/s}$. The vision sensor detects landmarks below the flight path and registers 27 landmarks from the total of 30 landmarks on the ground, and it extracts information at 10Hz.

Figure 2 shows the estimated vehicle trajectory from the indirect SLAM filter with 400Hz prediction rate. The standalone INS result is also plotted for comparison. The vehicle’s position from the indirect SLAM filter was maintained within approximately 10m and the attitude was maintained within approximately $0.5^{\circ}$. The map accuracy was within $5.25\,\text{m}$ with the initial vehicle uncertainty of $5\,\text{m}$. Figure 3 compares the vehicle position errors in the north direction from the direct and indirect SLAM implementation at 400Hz, and Figure 4 compares the heading errors. For the purpose of comparison, the same set of filter parameters were used and both forms show comparable results. In fact although both filter structures are theoretically consistent the difference lay in the ability to tune the indirect form more effectively given its linear structure. The low-dynamic characteristics of the INS errors allows for the reduction in the filter prediction rate. Three prediction rates were applied with the same filter tuning parameters: 400, 100 and 10Hz. Figure 5 compares the vehicle position errors in the north direction and Figure 6 compares the heading errors. Despite the dramatic reduction in the prediction rate from 400Hz to 10Hz, it can be observed that the estimated values are almost identical. Hence the indirect SLAM structure can be implemented in a much lower rate than in direct SLAM which is the direct consequence of the low-dynamic characteristics of the inertial errors. The degree of reduction in SLAM prediction rate can be determined from the accuracy of the IMU implemented and the performance required. The validation of the indirect approach illustrating that it can be used as an alternate means for SLAM navigation provides us with the opportunity now to discuss the observability of SLAM in its linear form.

V. SLAM OBSERVABILITY ANALYSIS

The observability analysis for nonlinear SLAM is quite cumbersome to evaluate and requires numerical computation of the observability Grammian rather than analytical. However, for a time-varying linear system which can be modelled as a piecewise constant system, such as the indirect SLAM filter, then the observability analysis can be performed analytically by introducing the Stripped Observability Matrix (SOM) [1].

For simplicity, a single landmark case is considered at first. In each $i^{th}$ segment, the discrete state transition and observation matrix of SLAM can be re-written as

$$F_i = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & F_{if} & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad (22)$$

$$H_i = \begin{bmatrix} H_{ip} & 0 & H_{ia} & -H_{ip} \end{bmatrix}, \quad (23)$$

where $F_{if}$ is the skew-symmetric matrix of the acceleration in the navigation frame at the $i^{th}$ segment,

$$F_{if} = \begin{bmatrix} 0 & -f_{ip}^n & f_{ip}^n & 0 \\ -f_{ip}^n & 0 & -f_{ia}^n & 0 \\ f_{ip}^n & -f_{ia}^n & 0 & 0 \end{bmatrix} \quad (24)$$

and $H_{ip}$ and $H_{ia}$ are the observation sub-matrices for the vehicle position and attitude [3].

By using equations 22 and 23, the Total Observability Matrix (TOM) of SLAM at the $j^{th}$ segment, $Q_{TOM}(j)$, can be constructed including the current and previous segments:

$$Q_{TOM}(j) = \begin{bmatrix} Q_1 \\ Q_2 e^{F_1 \Delta_1} \\ \vdots \\ Q_{j} e^{F_{j-1} \Delta_{j-1} \ldots e^{F_1 \Delta_1}} \end{bmatrix}, \quad (25)$$

where, $Q_i^T = [H_i^T (H_i F_i) (H_i F_i)^T \ldots (H_i F_i^{n-1})^T]^T$ for $1 \leq i \leq j$, and $\Delta_i$ is the time interval at $i^{th}$ segment. It can be shown [1] that if $null(Q_1) \subsetneq null(F_1)$, $\forall i$, then $\text{rank}(Q_{TOM}(j)) = \text{rank}(Q_{TOM}(j))$, where $Q_{TOM}(j)$ is

$$Q_{TOM}(j) = \begin{bmatrix} Q_1^T \\ Q_2^T \\ \vdots \\ Q_{j}^T \end{bmatrix}, \quad (26)$$

<table>
<thead>
<tr>
<th>Sensor</th>
<th>Type</th>
<th>Spec.</th>
</tr>
</thead>
<tbody>
<tr>
<td>IMU</td>
<td>Sampling rate</td>
<td>400 Hz</td>
</tr>
<tr>
<td></td>
<td>Accel noise</td>
<td>$0.5m/s^2/\sqrt{Hz}$</td>
</tr>
<tr>
<td></td>
<td>Gyro noise</td>
<td>$0.5^\circ/s/\sqrt{Hz}$</td>
</tr>
<tr>
<td>Vision/Laser</td>
<td>Frame rate</td>
<td>10Hz</td>
</tr>
<tr>
<td></td>
<td>FOV angle</td>
<td>$\pm 15^\circ$</td>
</tr>
<tr>
<td></td>
<td>Bearing/Elev noise strength</td>
<td>$0.1604^\circ/0.1206^\circ$</td>
</tr>
<tr>
<td></td>
<td>Range noise strength</td>
<td>0.2m</td>
</tr>
<tr>
<td>Initial Error</td>
<td>(Roll, Pitch)/Heading</td>
<td>$0.5^\circ/0.0^\circ$</td>
</tr>
</tbody>
</table>
Hence, the observability analysis is greatly simplified by evaluating equation 26, instead of equation 25. Firstly, the local observability matrix for the \(i^{th}\) segment is computed using \(F_i\) and \(H_i\):

\[
Q_i = \begin{bmatrix}
H_i^T & (H_i F_i)^T & (H_i F_i^2)^T
\end{bmatrix}^T
= \begin{bmatrix}
H_{ip} & 0 & H_{ia} - H_{ip}
0 & H_{ip} & 0
0 & 0 & H_{ip} F_{if}
\end{bmatrix},
\]

where \(F_i^3 = F_i^2 = \cdots = 0\) is used. Then, equation 26 can be computed as

\[
Q_{SOM}(j) = \begin{bmatrix}
H_{ip} & 0 & H_{ia} - H_{ip}
0 & H_{ip} & 0
0 & 0 & H_{ip} F_{if}
H_{ip} & 0 & H_{ja} - H_{ip}
0 & H_{ip} & 0
0 & 0 & H_{ip} F_{if}
\end{bmatrix},
\]

From the structure of \(Q_{SOM}(j)\) the following can be observed:

- The observability of SLAM at the \(j^{th}\) segment, \(Q_{SOM}(j)\), depends on all preceding segments including the current one.
- The order in which the piecewise constant segments are arranged has no effect on the final observability of the system.
- The first and second rows in \(Q_{SOM}(j)\) provide six linearly independent rows and repeating these at later segments does not provide any further linearly independent rows, thus does not enhance the observability of the system.
- \(F_{if}\) in each segment can provide only two linearly independent rows [5]. This is due to the property of the skew-symmetric matrix which is not invertible. It can be easily shown from equation 24 that the third row is the linear combination of the first and second rows. Accordingly \(H_{ip} F_{if}\) can only provide two linearly independent rows at each segments.

What is interesting is that a range, bearing and elevation observation in SLAM can only provide a rank of 8 (the dimension of 6DoF SLAM with one landmark is 12) or 9 when the vehicle undergoes sufficient maneuvers such as accelerating, decelerating or turning motion (under these conditions, \(F_{if}\) varies in each segment and can provide an additional linearly independent row). This result coincides with the fact that the relative distance observation between the vehicle and landmark provides information only about the difference of these two quantities not the individual one, which results in a rank deficiency of three. As an additional landmark is registered in SLAM, the column space of the SOM increases by three, which provides an additional three linearly independent rows to the SOM. That is, if there are \(N\) landmarks registered in SLAM, then the dimension of the system is \(3N+9\) and the rank of the SOM will either be \(3N+5\) or \(3N+6\) depending on the history of maneuvers.

Important question arises from these results: What makes 6DoF SLAM completely observable (assuming that the vehicle undergoes enough dynamics)? Since 6DoF SLAM always has a rank deficiency of three regardless of the number of landmarks within the filter, additional position observations that can provide three independent rows to the SOM can make SLAM fully observable. This can be achieved either by observing the vehicle position (for example by using the Global Positioning System (GPS), but this defeats the purpose of SLAM), or by observing a landmark whose position is known in the global coordinates (such as a man made or natural feature with known coordinates). Once this observation occurs, the vehicle and the all landmarks errors become fully observable and can be estimated from the filter, otherwise the rank deficiency will always occur.

If SLAM is fully observable, then the ability to estimate the state of SLAM is only dependent on 1) the statistical characteristics of the noise both in the process and observation model and 2) the relative geometry between the vehicle and landmarks. The former factor can be analysed through the probability density distribution of the noise. The latter one is related to the sensitivity of the observation matrix and can be studied by analysing the information contributions from the spatially distributed information sources (this issue is not tackled in this paper but can be seen in [6]).

VI. CONCLUSIONS

This paper has presented a novel approach to tackling the computational complexity of the 6DoF SLAM algorithm, which has in turn provided the capability of understanding the observability of the SLAM process. By recasting the SLAM algorithm to the indirect implementation, and by modelling the state transition matrix as an error model of both the dynamics of the vehicle and of the map, then a significant reduction in the filter prediction rate was achieved. The simulation results verified that the proposed SLAM structure can estimate the errors of the vehicle and landmarks without any significant degradation in accuracy even with a low filter prediction rate of 10Hz from 400Hz. More importantly, the observability analysis was also performed using the linearised SLAM model and the stripped observability matrix, and it showed that with \(N\)-landmarks a rank deficiency of at least three is encountered depending on the history of maneuvers, and that in order to cause the system to become fully observable requires an external piece of information about either the state of the vehicle or the state of a landmark.

ACKNOWLEDGEMENTS

This work is supported in part by the ARC Centre of Excellence programme, funded by the Australian Research Council (ARC) and the New South Wales State Government.
REFERENCES


Fig. 2. The estimated vehicle and 2D map position with estimated $1\sigma$ uncertainty using the indirect SLAM filter. The UAV starts at (0,0) and traverses in a figure eight.

Fig. 3. Comparison of the direct and the indirect SLAM position errors.

Fig. 4. Comparison of the direct and the indirect SLAM heading errors. Both show comparable results.

Fig. 5. Comparison of the indirect SLAM with reduced prediction rates for the position error with its $(1\sigma)$ uncertainty in the north direction.

Fig. 6. Comparison of the indirect SLAM with reduced prediction rates for the heading error with its $(1\sigma)$ uncertainty.