A Nonlinear Observer for 6 DOF Pose estimation from Inertial and Bearing Measurements

Grant Baldwin, Robert Mahony, and Jochen Trumpf

Abstract—This paper considers the problem of estimating pose from inertial and bearing-only vision measurements. We present a non-linear observer that evolves directly on the special Euclidean group SE(3) from inertial measurements and bearing measurements, such as provided by a visual system tracking known landmarks. Local asymptotic convergence of the observer is proved. The observer is computationally simple and its gains are easy to tune. Simulation results demonstrate robustness to measurement noise and initial conditions.

Inertial Vision, Non-linear Observer, Bearing-Only, Pose Estimation.

I. INTRODUCTION

The availability of cheap inertial measurement units (IMUs) has spurred the development of new and innovative applications in low-cost, small-scale robotics with a requirement for robust six degree of freedom pose estimation. As these cheap IMUs tend to suffer from significant non-Gaussian noise and time varying biases, most work has investigated sensor fusion of the IMU data with data from a complementary sensor that is free of bias, such as a global positioning system (GPS), e.g. [1]–[3], or a vision system, e.g. [4]–[8]. Vision from a monocular camera offers several advantages in its low unit cost and ability to operate in a wide variety of environments, including indoors or in urban environments where GPS signals may be unavailable or severely degraded. Image features from vision may be used directly as bearing measurements or reconstructed into a pose measurement. Pose reconstruction is a well understood problem in computer vision, e.g. [9], but requires significant computational resources not always available on embedded systems.

Extended and unscented Kalman filters have been applied to the pose estimation problem [10], [11] and the non-linear sub-problem of attitude estimation [12], with difficulties experienced due to the non-linear equations [7] and non-Gaussian measurement noise [13]. Particle filters have also been considered [14], [15] as a means of handling non-linearities. Estimation from bearing-only measurements has been considered in the SLAM community using Kalman and particle filters, e.g. [16].

Non-linear observers, exploiting the natural geometry of the problem, promise to handle the problem of non-linear state and measurement equations and can lead to strong (almost global) stability results. Salcudean designed a non-linear observer for angular velocity of rigid body motion [17]. Ghosh and subsequent authors consider non-linear observers on the class of perspective systems [18]–[20]. More recent work expands on the estimation of rigid body motion to consider attitude, position and linear velocity. Mahony et. al. consider attitude estimation as an observer problem on the special orthogonal group, SO(3), complementing rate gyros with measurement of gravitational direction [21]. Thielen and Sanner derive an attitude and bias observer of which bias converges exponentially fast [22]. Bonnabel, Martin and Rouchon investigate invariant observers on SO(3) × R^3, estimating attitude, linear velocity and biases from inertial measurements [23], [24]. Vik and Fossen develop a non-linear observer on SO(3) × R^3 for pose from GPS and inertial measurements. [25]. In [26], the authors present an observer on the special Euclidean group SE(3) using a pose measurement reconstructed from each vision frame. Other work uses features present in an image directly. Vasconcelos et. al. develop a filter on SE(3) from inertial measurements and vector measurements of landmarks [27]. Rehbinder and Ghosh investigated pose estimation using point and line image features [8].

In this paper, we present an observer evolving on SE(3) from measurements of angular and linear velocities and bearing-only measurements of landmarks. A Lyapunov function is defined using the chordal bearing error and we present a local non-linear Lyapunov stability argument using Barbalat’s lemma. The use of the geometric structure of the problem in observer design leads to an observer that is computationally simple and has gains that are easy to tune. Simulation results show the observer is highly robust to measurement noise and initial conditions.

The remainder of this paper is organised as follows; Section II reviews and introduces the notation used in this paper. In section III the dynamical system to be estimated is presented. Measurement models are given in Section III-A. The non-linear observer is detailed in Section IV. Simulation results are presented in Section V.

II. NOTATION

Define an inner product on the set of R^{n×n} matrices as

\[ \langle A, B \rangle = \text{tr}(A^\top B). \]  (1)
The Frobenius norm $\| \cdot \|_F$ is given by $\|A\|_F^2 = \langle A, A \rangle = \text{tr}(A^T A)$. Let $\pi_X$ and $\pi_X^\perp$ denote the projection matrices onto the subspace spanned by $X \in \mathbb{R}^3$ and the subspace perpendicular to the span of $X$, respectively, such that $I = \pi_X + \pi_X^\perp$. One has
\begin{align}
\pi_X &= \frac{1}{\|X\|^2} XX^T, \\
\pi_X^\perp &= \frac{1}{\|X\|^2} (I - XX^T).
\end{align}

III. Problem Description

Let $A$ denote an inertial frame attached to the earth such that $e_3$ points vertically down. Let $B$ denote a body-fixed frame attached to a vehicle. The origin of $B$ expressed in $A$ is given by the vector $p$, and the attitude of $B$ expressed in $A$ is given by the rotation matrix $R$. The element of $\text{SE}(3)$
\begin{equation}
T = \begin{pmatrix} R & p \\ 0 & 1 \end{pmatrix}
\end{equation}
has the triple interpretation as the pose of $B$ expressed in $A$: the coordinate frame transformation mapping objects expressed in $B$ to objects expressed in $A$; and the rigid body transformation moving an object from the pose given by the attitude and position of $A$ to the attitude and position of $B$, expressed in $A$. We henceforth adopt the convention that positions and vectors expressed in the inertial frame are denoted by lower case letters while quantities expressed in other frames are denoted by the majuscule.

The kinematics of $B$ are given by
\begin{align}
\dot{R} &= R \Omega_x, \\
\dot{p} &= v
\end{align}
where $\Omega \in B$ and $v \in A$ denote the angular and linear velocities, respectively, of the body-fixed frame, and $(\cdot)_x$ is an operator taking the vector $\Omega \in \mathbb{R}^3$ to a skew-symmetric matrix
\begin{equation}
\Omega_x = \begin{pmatrix} 0 & -\Omega_3 & \Omega_2 \\ \Omega_3 & 0 & -\Omega_1 \\ -\Omega_2 & \Omega_1 & 0 \end{pmatrix}.
\end{equation}

Let $P = -R^T p$, the position of the origin of $A$ expressed in $B$, then $\dot{P} = -\Omega_x P - v$ where $V = R^T v$.

Represented in $\text{SE}(3)$, the kinematics of $B$, (5), are
\begin{equation}
\dot{T} = T \Xi, \quad \Xi = (\Omega, V)^\wedge
\end{equation}
where
\begin{equation}
(\Omega, V)^\wedge = \begin{pmatrix} \Omega_x \\ V \\ 0 \end{pmatrix}.
\end{equation}
\(\Xi \in \text{se}(3)\) is a representation of the velocity of $B$ expressed in $B$. The wedge operator, $(\cdot, \cdot)^\wedge$, maps between the $\mathbb{R}^6$ and $\text{SE}(3)$ interpretations of the system velocity.

Consider an ensemble of landmark points $\{z_i\} \in A$ that are stationary in the inertial frame and distributed in space such that they are non-collinear. Let $Y_i^h = T^{-1} z_i^h$, the position of the $i$-th landmark expressed in $B$, where $Y_i^h$ and $z_i^h$ denotes the homogeneous coordinate representation of $Y_i$ and $z_i$.

Since the $z_i$ are stationary in the inertial frame then $\dot{z}_i = 0$ and one has
\begin{equation}
\dot{Y}_i = -\Omega_x Y_i - V = -\Xi Y_i^h.
\end{equation}

In this paper, we will use measurements of the bearing of the $i$-th landmark from the origin of $B$ as the visual measurements. A bearing can be expressed as a direction in 3D, that is as a direction on the sphere $X_i \in \mathbb{S}^2$. Based on this representation of bearing measurements one has
\begin{align}
X_i &= \frac{Y_i}{\|Y_i\|}, \\
\dot{X}_i &= -\Omega_x X_i - \frac{1}{\|Y_i\|} \pi_{X_i}^\perp(V)
\end{align}
where $X_i \in B$ is expressed in the body-fixed frame.

A. Measurement Model

We consider the vehicle of interest equipped with an inertial-vision sensor package; an IMU and a monocular camera, both affixed to the craft and taking measurements in the body-fixed frame, $B$. Such an inertial-vision sensor arrangement is detailed in e.g. [6]. Commonly in such a system the inertial measurement rate is faster, by as much as an order of magnitude, than the vision measurement, but the inertial measurement may be corrupted by low frequency noise including time-varying biases. Conversely, the slower vision measurements are stable and free of bias at low frequency. Hence, the two measurements are fused to take advantage of their complementary characteristics.

For this paper, we consider IMU measurements of angular and linear velocity, $\Xi = (\Omega, V)^\wedge$ that are free of bias and corrupted by a zero mean noise process $n_\Xi$. Such a case may occur where raw inertial measurements of angular velocity and linear acceleration are pre-filtered for bias correction and to obtain linear velocity as in [26]. We consider complementary vision measurements of the bearings of image features, $\{X_i\}$ corrupted by zero-mean noise processes on the tangent plane of the unit sphere.

Let $\Xi_y$ and $X_{yi}$, respectively, denote measurements of velocity and the bearing of the $i$-th image feature. The measurements are given by
\begin{align}
\Xi_y &= \Xi + n_\Xi, \\
X_{yi} &= \frac{X_i + n_{X_i}}{\|X_i + n_{X_i}\|}
\end{align}
where $n_\Xi = (n_\Omega, n_V)^\wedge$ denotes a zero-mean noise process composed from the vectorial zero-mean noise processes $n_\Omega$ and $n_V$, and $n_{X_i}$ denotes a zero-mean random perturbation in a direction tangential to $X_i$.

IV. $\text{SE}(3)$ Filter

In this section we consider the problem of estimating the pose $T$ of a moving vehicle given measurements of its angular and linear velocity, $\Xi$, and the bearing of the known landmarks, $X_i$, together with a model of the landmark point.
ensemble in the inertial frame, \{z_i\}. We assume that the landmarks are arranged such that there exists a separating hyperplane between the region containing the landmarks and the region in which the vehicle operates (e.g., the vehicle is always above landmarks distributed across the ground). We define an estimate and suitable errors, then define a non-linear observer and prove local asymptotic convergence by non-linear Lyapunov analysis.

Let \( E \) denote a new frame representing a current estimate of \( B \). The pose of \( E \) expressed in \( A \) is given by the SE(3) element

\[
\hat{T} = \begin{pmatrix} \hat{R} & \hat{p} \\ 0 & 1 \end{pmatrix}
\] (12)

where \( \hat{p} \) is the origin of \( E \) expressed in \( A \) and \( \hat{R} \) is the rotation matrix giving the attitude of \( E \) expressed in \( A \). Let

\[
\hat{Y}_i = \hat{T}^{-1}z_i
\] (13)

and

\[
\hat{X}_i = \hat{Y}_i / \|\hat{Y}_i\|
\] (14)

denote the position and bearing of the \( i \)-th landmark expressed in \( E \).

**Lemma 1:** Consider the system (7) and constellation of landmarks \( \{z_i\} \). For a set of three or more non-collinear landmarks \( \{z_i\} \) then \( \hat{X}_i = X_i \) for all \( i \) if and only if \( \hat{T}T^{-1} = I \).

**Proof:** [Proof of Lemma 1]

Consider the stabiliser group for a single landmark bearing measurement, \( X_i \). It consists of translation along and rotation about the axis in direction \( X_i \).

For a set \( \{X_i\} \) derived from a set \( \{z_i\} \) of three or more non-collinear elements, the intersection of the stabiliser groups contains only the identity element and the result follows.

**Theorem 2 (Local Observer Stability):** Consider the system (7) and constellation of landmarks \( z_i \) such that at least three landmarks are non-collinear, together with a bounded piecewise continuous driving term \( \Xi \) and trajectory \( T \) such that, for all \( i \) and for all \( t \), \( \|Y_i(t)\| > \epsilon > 0 \).

Define the observer

\[
\hat{T} = \hat{T}(\Xi_g + \xi) \quad (15a)
\]

\[
\xi = (\xi_\Omega, \xi_V) \quad (15b)
\]

\[
\xi_\Omega = -k_\Omega \sum_{i=1}^{n} \hat{X}_i \times X_{y_i} \quad (15c)
\]

\[
\xi_V = -k_V \sum_{i=1}^{n} \frac{\pi_{\hat{X}_i} X_{y_i}}{\|Y_i\|}, \quad (15d)
\]

and the error \( \hat{T} \)

\[
\hat{T} = \hat{T}T^{-1} \quad (16)
\]

Then, for exact measurements \( X_{y_i} \) and \( \Xi_g \), there exists \( k_V > 0 \) such that for all choices of \( k_\Omega > 0 \), \( \hat{T} \) is locally asymptotically stable about \( I \).

**Proof:** [Proof of Theorem 2]

Recall the definition of \( \hat{X}_i \) from (14). It is straightforward to verify that

\[
\dot{\hat{X}} = -\left(\Omega + \xi_\Omega\right) \times \hat{X}_i - \frac{\pi_{\hat{X}_i} (V + \xi_V)}{\|Y_i\|} \quad (17)
\]

Consider the per-feature error \( X_i^\Delta := \hat{X}_i - X_i \). One has

\[
\dot{X}_i^\Delta = -\left[\Omega \times X_i^\Delta - \xi_\Omega \times \hat{X}_i \right.
\]

\[
\left. - \frac{\pi_{\hat{X}_i} (V - \frac{\pi_{\hat{X}_i}}{\|Y_i\|} \xi_V)}{\|Y_i\|} \right]. \quad (18)
\]

Define the positive definite cost function

\[
V : (\mathbb{R}^3)^n \rightarrow \mathbb{R} \\
V = \sum_{i=1}^{n} \frac{1}{2} \|X_i^\Delta\|^2. \quad (19)
\]

Then

\[
\dot{\hat{Y}} = -\sum_{i=1}^{n} \left(\hat{X}_i \times X_i^\Delta, \xi_\Omega\right) - \sum_{i=1}^{n} \left(X_i^\Delta, \frac{\pi_{\hat{X}_i}}{\|Y_i\|} \xi_V\right) \quad (20a)
\]

\[
- \sum_{i=1}^{n} \left(X_i^\Delta, \frac{\pi_{\hat{X}_i}}{\|Y_i\|} \frac{\pi_{\hat{X}_i}}{\|Y_i\|} \xi_V\right) \|Y_i\| \right) \|V\|
\]

\[
= \left\langle \sum_{i=1}^{n} \hat{X}_i \times X_i, \xi_\Omega \right\rangle + \sum_{i=1}^{n} \frac{\pi_{\hat{X}_i} X_i}{\|Y_i\|}, \xi_V \right\rangle \quad (20b)
\]

\[
+ \sum_{i=1}^{n} \frac{\pi_{\hat{X}_i} X_i}{\|Y_i\|} + \frac{\pi_{\hat{X}_i} X_i}{\|Y_i\|}, V \right\rangle.
\]

Substituting \( \xi_\Omega \) and \( \xi_V \) from (15c) and (15d) and setting \( \Omega_y \equiv \Omega, V_y \equiv V \), and \( X_{y_i} \equiv X_i \). One has

\[
\dot{\hat{Y}} = -k_\Omega \sum_{i=1}^{n} \hat{X}_i \times X_i^\Delta - k_V \sum_{i=1}^{n} \frac{\pi_{\hat{X}_i} X_i}{\|Y_i\|}^2 \quad (21)
\]

The proof continues by considering a given trajectory \( T(t) \) and considering a local analysis around the associated constant error trajectory \( \hat{T}(t) \equiv I \). Consider the first order local approximation \( \hat{X}_i = X_i + \alpha, \|\hat{Y}_i\| = \|Y_i\| + \beta_i \), where \( \alpha_i \in T_X, S^2 \) and \( \beta_i \in \mathbb{R} \) are small perturbations induced by small variations in \( \hat{T} \). One has that \( \alpha_i \times X_i = 0 \). Hence

\[
\pi_{\hat{X}_i} X_i = -\alpha_i \quad \pi_{\hat{X}_i} \hat{X}_i = \alpha_i 
\]

\[
\frac{\pi_{\hat{X}_i} X_i}{\|Y_i\|} + \frac{\pi_{\hat{X}_i} \hat{X}_i}{\|Y_i\|} = \frac{\beta_i \alpha_i}{\|Y_i\|} (\|Y_i\| + \beta_i)
\]
The derivatives of (24) around differentials of the mappings $M \in \mathbb{R}^3$ is negative definite. Applying Theorem 8.4 of [28] the local system linearisation is stable and $X_i - \hat{X}_i \to 0 \forall i$ locally. Further, by Lemma 1, $\hat{T} \to I$ locally.

\[ \mathcal{V} = \sum_{i=1}^{n} \frac{1}{2} \| \alpha_i \|^2 \]  
\[ (23a) \]

\[ \dot{\mathcal{V}} = -k_{\Omega} \left( \sum_{i=1}^{n} \alpha_i \times X_i \right)^2 - k_{V} \left( \sum_{i=1}^{n} \frac{\alpha_i}{\| Y_i \| + \beta_i} \right)^2 \]  
\[ (23b) \]

\[ + \left\langle \sum_{i=1}^{n} \frac{\beta_i \alpha_i}{\| Y_i \| (\| Y_i \| + \beta_i)} \| V \right\rangle \]

\[ \leq -k_{\Omega} \left( \sum_{i=1}^{n} \alpha_i \times X_i \right)^2 - k_{V} \left( \sum_{i=1}^{n} \frac{\alpha_i}{\| Y_i \| + \beta_i} \right)^2 \]  
\[ + \left\langle \sum_{i=1}^{n} \frac{\beta_i \alpha_i}{\| Y_i \| (\| Y_i \| + \beta_i)} \| V \right\rangle \]

Remark 3: On Theorem 2, we note:

1) The per-feature error $X_i^\Delta$ is a projective analogue to the vectorial per-landmark error

\[ Y_i^\Delta := \hat{Y}_i - Y_i \]

\[ = \hat{T}^{-1} z_i - T^{-1} z_i \]

\[ = T^{-1} (\hat{T} - I) z_i \]

where $\hat{T} = \hat{T}^-1$. Baldwin et al. [26] provide an interpretation for $\hat{T}$ as a rigid body transformation taking $\mathcal{B}$ to $\mathcal{E}$ represented in the inertial frame, and derive a filter based on the cost function $\| \hat{T} - I \|$.

2) The requirement $\forall i \| Y_i \| > \epsilon > 0$ is not a constraint in a practical situation as the set of landmarks $\{ z_i \}$ may be modified online to dynamically remove any landmarks the vehicle moves close to and return them when the vehicle moves away.

Similarly, occluded landmarks are not expected to significantly degrade the scheme provided there is always a suitable non-colinear set of landmarks visible.

3) The cross term, containing linear velocity, in $\dot{\mathcal{V}}$ (20a) is unavoidable when working with projective measurements. Each summand corresponds to the difference between components of linear velocity perpendicular to the actual and estimated i-th bearing, scaled by the inverse of the actual or estimated distance to the landmark. As the actual landmark distance, $\| Y_i \|$ is not a measured quantity, the resulting error term can not be cancelled by clever choice of innovation term. In the analogous system with non-projective measurements $Y_i$, these terms cancel and the observer admits an almost-global stability proof.

We believe that understanding the bounds of this cross term is critical to proving global observer properties. In particular, the existence of a uniform bound between this cross term and the translation error innovation, i.e.

\[ \left\| \sum_{i=1}^{n} \frac{\pi \hat{X}_i}{\| Y_i \|} \right\| \leq c \left\| \sum_{i=1}^{n} \frac{\pi \hat{X}_i}{\| Y_i \|} \right\| \]

would permit a simple almost-global stability proof for this system.

V. SIMULATION

The filter (15) has been implemented in MATLAB 2007b as a discrete event simulation. Simulated velocity and feature bearing measurements have been generated with noise added as per the measurement model detailed in Section III-A. The simulation used velocity measurements at 100 Hz and pose measurements at 5 Hz with synchronized measurement clocks. Simulations used a feature constellation of $\{(1, 1, 0), (1, -1, 0), (-1, -1, 0), (-1, 1, 0)\}$ with features never occluded by camera position or orientation.

A multirate discrete integration scheme has been implemented where the low rate pose error information, the bearing measurements, are used to make impulsive updates...
to the high rate inertial information update. That is,

\[
\hat{T}_{k+1} = \begin{cases} 
\hat{T}_k \exp(\tau \Xi + s\tau \xi) & \text{vision available} \\
\hat{T}_k \exp(\tau \Xi) & \text{vision unavailable}
\end{cases}
\] (31)

where \(\tau\) is the integrations time step and \(s\) is number of time steps from the last available vision measurement. This process avoids error due to translating old pose error into a frame of reference. The resulting observer output displays a multirate character with small regular updates from the high frame of reference. The resulting observer output displays a pose error should not grow during periods where no pose information is available.

Figure 1 depicts the pose estimate for a body descending along a trim trajectory from above the observed feature constellation, described in (32). Observer gains used in simulations are given in Table I. Tuning observer gains was found to be easy, with changes in gains resulting in good performance for values \(0.001 \leq k_0 \leq 10\) and \(0.01 \leq k_v \leq 100\), with gain choice governed by the tradeoff between convergence rate and overshoot due to discrete time implementation. Other trajectories, of varying scales, have been simulated with similar convergence results, strengthening the case for global observer properties.

\[
\begin{align*}
\tau & = 0.1 \text{ m} \\
t_f & = 120 \text{ s} \\
x_0 & = -0.1 \text{ m} \\
y_0 & = 0 \text{ m} \\
z_0 & = -1.5 \text{ m} \\
\delta_z & = 0.5 \text{ m} \\
\omega & = \frac{2\pi}{t_f} \text{ rad s}^{-1} \\
x(t) & = \begin{pmatrix} \cos(\omega t) & -\sin(\omega t) \\ \sin(\omega t) & \cos(\omega t) \end{pmatrix} \begin{pmatrix} r \\
y \end{pmatrix} + \begin{pmatrix} x_0 - r \\ y_0 \end{pmatrix} \\
z(t) & = z_0 + \frac{\delta_z}{t_f} t \\
\phi(t) & = \frac{\pi}{2} + \omega t \\
\theta(t) & = \arctan\left(\frac{x(0)}{y(0)}\right) = \arctan\left(\frac{\delta_z}{t_f r}\right) \\
\psi(t) & = 2\theta(t) \\
T(t) & = \begin{pmatrix}
R_{\psi(t)} R_{\theta(t)} & x(t) \\
0 & 1
\end{pmatrix} \begin{pmatrix}
y(t) \\
z(t)
\end{pmatrix}
\end{align*}
\] (32)

VI. CONCLUSIONS

We proposed an observer that evolves on \(SE(3)\) with local stability from a combination of angular and linear velocity and bearing measurements, such as those obtained from a combination of inertial and vision sensors. Local asymptotic convergence of the error system was proven using a non-linear Lyapunov argument. Simulation results have shown that the observer is robust to measurement noise and initial conditions, with a wide range of gains yielding good performance. Although only local convergence is proven, simulations indicate almost-global convergence.

REFERENCES

Fig. 1: Simulation results of pose estimation using simulated measurements along a trim trajectory specified by (32), with filter gains as per Table I. Dashed lines represent the reference signal while the solid line represent the observer estimates. Simulation with image feature noise variance of 0.5 m on an tangent image plane at 1 m, angular velocity noise variance of 0.01 rad. s\(^{-1}\) and linear velocity noise variance of 0.1 m s\(^{-1}\). Simulation used a velocity measurement rate of 100 Hz and an image feature measurement rate of 5 Hz.


