Observability, Identifiability and Sensitivity of Vision-Aided Inertial Navigation

Joshua Hernandez and Konstantine Tsotsos and Stefano Soatto
University of California, Los Angeles
{jheez,ktrotsos,soatto}@ucla.edu

Abstract
We analyze the observability of 3-D position and orientation from the fusion of visual and inertial sensors. The model contains unknown parameters, such as sensor biases, and so the problem is usually cast as a mixed filtering/identification problem, with the resulting observability analysis providing necessary conditions for convergence to a unique point estimate. Most models treat sensor bias rates as “noise,” independent of other states, including biases themselves, an assumption that is violated in practice. We show that, when this assumption is lifted, the resulting model is not observable, and therefore existing analyses cannot be used to conclude that the set of states that are indistinguishable from the measurements is a singleton. We re-cast the analysis as one of sensitivity: Rather than attempting to prove that the set of indistinguishable trajectories is a singleton, we derive bounds on its volume, as a function of characteristics of the sensor and other sufficient excitation conditions. This provides an explicit characterization of the indistinguishable set that can be used for analysis and validation purposes.

1 Introduction
We present a novel approach to the analysis of observability/identifiability of three-dimensional (3-D) position and orientation (termed pose) in visually-assisted navigation, whereby inertial sensors (accelerometers and gyrometers, jointly referred to as an Inertial Measurement Unit, or IMU) are used in conjunction with optical sensors (vision) to yield an estimate of the 3-D pose of the sensor platform. It is customary to frame this as a filtering problem, where the time-series of positions and orientations of the sensor platform is modeled as the state trajectory of a dynamical system, that produces sensor measurements as outputs, up to some uncertainty. Observability/identifiability analysis refers to the characterization of the set of possible state trajectories that produce the same measurements, and therefore are indistinguishable given the outputs [Soatto, 1994; Kelly and Sukhatme, 2009; Mourikis and Roumeliotis, 2007; Jones et al., 2007; Martinelli and others, 2014].

The parameters in the model are either treated as unknown constants (e.g., calibration parameters) or as random processes (e.g., accelerometer and gyro biases) and included in the state of the model, which is then driven by some kind of uninformative (“noise”) input. Because noise does not affect the observability of a model, for the purpose of analysis it is usually set to zero. However, the input to the model of accelerometer and gyro bias is typically small but not independent of the state. Thus, it should be treated as an unknown input, which is known to be “small” in some sense, rather than “noise.”

Our first contribution is to show that while (a prototypical model of) assisted navigation is observable in the absence of unknown inputs, it is not observable when unknown inputs are taken into account. Our second contribution is to reframe observability as a sensitivity analysis, and to show that while the set of indistinguishable trajectories is not a singleton (as it would be if the model was observable), it is nevertheless bounded. We explicitly characterize this set and bound its volume as a function of the characteristics of the inputs, which include sensor characteristics (bias rates) and the motion undergone by the platform (sufficient excitation).

Related work
In addition to the above-referenced work on visual-inertial observability, our work relates to general unknown-input observability of linear time-invariant systems addressed in [Basile and Marro, 1969; Hamano and Basile, 1983], for affine systems [Hammouri and Tmar, 2010], and non-linear systems in [Dimassi et al., 2010; Tanwani, 2011; Bezzaoucha et al., 2011]. The literature on robust filtering and robust identification is relevant, if the unknown input is treated as a disturbance. However, the form of the models involved in vision-aided navigation does not fit in the classes treated in the literature above, which motivates our analysis. The model we employ includes alignment parameters for the (unknown) pose of the inertial sensor relative to the camera.

1.1 Notation
We adopt the notation of [Murray et al., 1994], where a reference frame is represented by an orthogonal $3 \times 3$ positive-determinant (rotation) matrix $R \in SO(3) = \{ R \in \mathbb{R}^{3 \times 3} \mid R^T R = RR^T = I, \det(R) = +1 \}$ (the special orthogonal group) and a translation vector $T \in \mathbb{R}^3$. They are
collectively indicated by \( g = (R, T) \in SE(3) \) (the special Euclidean group). When \( g \) represents the change of coordinates from a reference frame \( "a" \) to another ("b"), it is indicated by \( g_{ba} \). Then the columns of \( R_{ba} \) are the coordinate axes of \( a \) relative to the reference frame \( b \), and \( T_{ba} \) is the origin of \( a \) in the reference frame \( b \). If \( p_a \) is a point relative to the reference frame \( a \), then its representation relative to \( b \) is \( p_b = g_{ba} p_a \). If \( X_a \) are the coordinates of \( p_a \), then \( X_b = R_{ba} X_a + T_{ba} \) are the coordinates of \( p_b \). A time-varying transformation (or pose) is indicated with \( g(t) = (R(t), T(t)) \).

We indicate with \( \hat{\omega} \) a skew-symmetric matrix \( \omega \in so(3) \cong \{ S \in \mathbb{R}^{3 \times 3} | S^T = -S \} \) corresponding to the cross product with the vector \( \omega \in \mathbb{R}^3 \), so that \( \hat{\omega} v = \omega \times v \) for any vector \( v \in \mathbb{R}^3 \). In homogeneous coordinates, we write \( \bar{X}_b = g_{ba} \bar{X}_a \) where \( X^T = [X^T 1] \) and \( \bar{X}_b^T = [(R_{ba} X_a + T_{ba})^T 1] \).

### 1.2 Motion Model

There are several reference frames to be considered in a navigation scenario. The spatial frame \( s \) is typically attached to the Earth and oriented so that gravity \( g \) takes the form \( \gamma^T = [0 0 1]^T | \gamma | \) where \( | \gamma | \) can be read from tabulates based on location and is typically around 9.8\( m/s^2 \). The body frame \( b \) is attached to the IMU. The camera frame \( c \), relative to which image measurements are measured, is also known, although we will assume that intrinsic calibration has been performed, so that measurements on the image plane are provided in metric units [Ma et al., 2003]. The motion of a sensor platform is represented as the time-varying pose \( g_{sb} \) of the body relative to the spatial frame.

The equations of motion (known as mechanization equations) are usually described in terms of the body frame at time \( t \) relative to the spatial frame \( g_{sb}(t) \). Since the spatial frame is arbitrary (other than for being aligned to gravity), it is often chosen to be co-located with the body frame at time \( t = 0 \). To simplify the notation, we indicate this time-varying frame \( g_{sb}(t) \) simply as \( g \), and so for \( R_{sb}, T_{sb}, \omega_{sb}, v_{sb} \), thus effectively omitting the subscript \( s \) wherever it appears. This yields \( \bar{T} = v, \bar{R} = R \bar{\omega}, \bar{v} = \dot{\omega}, \bar{\omega} = \bar{\omega}, \bar{\alpha} = \xi \) where \( \omega, \alpha \) and \( \xi \) are the translational and rotational jerk (of acceleration). For further details, see [Jones and Soatto, 2011].

### 1.3 Sensor Model

Although the acceleration \( \alpha \) defined above corresponds to neither body nor spatial acceleration, it is conveniently related to accelerometer measurements \( \alpha_{imu} \):

\[
\alpha_{imu}(t) = R_T^T(t)(\alpha(t) - \gamma) + \alpha_b(t) + n_{\alpha}(t)
\]

where the measurement error in bracket includes a slowly-varying mean ("bias") \( \alpha_b(t) \) and a residual term \( n_{\alpha} \) that is commonly modeled as a zero-mean (its mean is captured by the bias), white, homoscedastic and Gaussian noise process. In other words, it is assumed that \( n_{\alpha} \) is independent of \( \alpha \), hence uninformative. Measurements from a gyro, \( \omega_{imu} \), can be similarly modeled as

\[
\omega_{imu}(t) = \omega(t) + \omega_b(t) + n_{\omega}(t)
\]

where the measurement error in bracket includes a slowly-varying bias \( \omega_b(t) \) and a residual "noise" \( n_{\omega} \) also assumed zero-mean, white, homoscedastic and Gaussian, independent of \( \omega \).

Other than the fact that the biases \( \alpha_b, \omega_b \) change slowly, they can change arbitrarily. One can therefore consider them an unknown input to the model, or a state in the model, in which case one has to hypothesize a dynamical model for them. For instance,

\[
\alpha_b(t) = \xi_b(t) \quad \omega_b(t) = w_b(t), \quad \dot{\alpha}_b(t) = \xi_b(t)
\]

for some unknown inputs \( \omega_b, \xi_b \) that can be safely assumed to be small, but not (white, zero-mean and, most importantly) independent of the biases. Nevertheless, it is common to consider them to be realizations of a Brownian motion that is independent of \( \omega_b, \alpha_b \). This is done for convenience as one can then consider all unknown inputs as "noise." Unfortunately, however, this has implications on the analysis of the observability and identifiability of the resulting model.

### 1.4 Model Reduction

The equations above define a dynamical model having as output the IMU measurements. In this standard model, data from the IMU are considered as (output) measurements. However, it is customary to treat them as (known) input to the system, by writing \( \omega \) in terms of \( \hat{\omega}_{imu} \) and \( \alpha \) in terms of \( \alpha_{imu} \). Including the initial conditions and biases, the resulting mechanization model is

\[
\begin{align*}
\dot{T} &= v \\
\dot{R} &= R(\hat{\omega}_{imu} - \hat{\omega}) + n_R \\
\dot{v} &= R(\alpha_{imu} - \alpha_b) + \gamma + \nu \\
\dot{\omega}_b &= w_b \\
\dot{\alpha}_b &= \xi_b
\end{align*}
\]

with \( n_R = -n_{\alpha} \) and \( \nu = -R_{\alpha} \), both typically considered independent of the state.

### 1.5 Imaging Model and Alignment

Initially we assume there is a collection of points \( X_i, i = 1, \ldots, N \), visible from time \( t = 0 \) to the current time \( t \). If \( \pi : \mathbb{R}^3 \rightarrow \mathbb{R}^2; X \rightarrow [X_1/X_3, X_2/X_3] \) is a canonical central (perspective) projection, assuming that the camera is calibrated,\(^1\) aligned,\(^2\) and that the spatial frame coincides with the body frame at time \( 0 \), we have

\[
y_i(t) = R_{T12}^1(t)(X_i - T_{12}(t)) \quad \pi(g^{-1}(t)X_i) + n_i(t)
\]

In practice, the measurements \( y(t) \) are known only up to a transformation \( g_{cb} \) mapping the body frame to the camera, often referred to as "alignment":

\[
y_i(t) = \pi(g_{cb}g^{-1}(t)X_i) + n_i(t) \in \mathbb{R}^2
\]

Which we can add, along with the points \( X_i \), to the state with trivial dynamics \( \dot{g}_{cb} = 0 \):

\[
\begin{align*}
\dot{X}_i &= 0, & i = 1, \ldots, N \\
\dot{g}_{cb} &= 0
\end{align*}
\]

\(^1\)Intrinsic calibration parameters are known.

\(^2\)The pose of the camera relative to the IMU is known.
2 Analysis of the Model

The goal here is to exploit imaging and inertial measurements to infer the sensor platform trajectory. For this problem to be well-posed, a (sufficiently exciting) realization of \(\omega_{imu}, \alpha_{imu}\) and \(y\) should constrain the set of trajectories that satisfy (4)-(7) to be unique. If there are different trajectories that satisfy (4) with the same outputs and inputs, they are indistinguishable. If the set of indistinguishable trajectories is a singleton (contains only one element, presumably the “true” trajectory), the model (4) is observable, and one may be able to retrieve a unique point-estimate of the state using a filter, or observer.

While it is commonly accepted that the model (specifically, its equivalent reduced realization) (4), is observable, this is the case only when biases are exactly constant. But if biases are allowed to change, however slowly, the observability analysis conducted thus far cannot be used to conclude that the indistinguishable set is a singleton. Indeed, we show that this is the not the case, by computing the indistinguishable set explicitly. The following claim is proven in [Hernandez et al., 2013].

Claim 1 (Indistinguishable Trajectories) Let \(g(t) = (R(t), T(t)) \in SE(3)\) satisfy (4)-(7) for some known constant \(\gamma\) and functions \(\alpha_{imu}(t), \omega_{imu}(t)\) and for some unknown functions \(\alpha_0(t), \omega_0(t)\) that are constrained to have \(|\alpha_0(t)| \leq \epsilon, |\dot{\omega}_0(t)| \leq \epsilon, \) and \(|\ddot{\omega}_0(t)| \leq \epsilon\) at all t, for some \(\epsilon < 1\).

Let \(\hat{g}(t) = \sigma(g_B g(t)) g_A\) for some constant \(g_A = (R_A, T_A), g_B = (R_B, T_B), \sigma > 0,\) with bounds on the configuration space such that \(\|T_A\| \leq M_A\) and \(0 < m_\sigma \leq |\sigma| \leq M_\sigma.\) Then, under sufficient excitation conditions, \(\hat{g}(t)\) satisfies (4)-(7) if and only if

\[
\|I - R_A\| \leq \frac{2\epsilon}{m(\omega_{imu} : \mathbb{R}^+)}
\] (8)

\[
|\sigma - 1| \leq \frac{k_1 \epsilon + M_\sigma \|I - R_A\|}{m(\omega_{imu} : \mathbb{I}_1)}
\] (9)

\[
\|T_A\| \leq \frac{\epsilon (k_2 + (2M_\sigma + 1)M_A)}{m_\sigma m(\omega_{imu} : \mathbb{I}_2)}
\] (10)

\[
\|\gamma\| \leq \frac{\epsilon (k_3 + M_\sigma M_A)}{m_\sigma m(\omega_{imu} - \omega_0 : \mathbb{I}_3)} + \frac{(|\sigma - 1| + \epsilon) M(\omega_{imu} - \omega_0 : \mathbb{I}_3)}{m_\sigma m(\omega_{imu} - \omega_0 : \mathbb{I}_3)}
\] (11)

for \(\mathbb{I}_1\) and \(k_i\) determined by the sufficient excitation conditions.

Here, sufficient excitation \((m, M)\) refers to the nature of the motion, with more dynamic motion leading to larger values, as described in [Hernandez et al., 2013]. From these bounds, we find that the set of indistinguishable trajectories in the limit where \(\epsilon \to 0\) is parametrized by an arbitrary \(T_B \in \mathbb{R}^3\) and rotation \(\theta \in \mathbb{R}\) about gravity, termed the Gauge ambiguity, which can be explicitly fixed [Hernandez et al., 2013]. This immediately implies the following

Claim 2 (unknown-input observability) The model (4)-(7) is not observable, even after fixing the Gauge ambiguity, as the indistinguishable set is not a singleton, unless biases are constant \((\epsilon = 0)\) or their derivative is known exactly.

We refer the reader to [Hernandez et al., 2013] for additional details and proofs, which are articulated into several steps. In practice, once the Gauge transformations are fixed, a properly designed filter can be designed to converge to a point estimate, but there is no guarantee that such an estimate coincides with the true trajectory. Instead, the estimate can deviate from the true trajectory depending on the biases. The analysis above quantifies how far from the true trajectory the estimated one can be, provided that the estimation algorithm uses bounds on the bias drift rates and the characteristics of the motion. Often these bounds are not strictly enforced but rather modeled through the driving noise covariance.

3 Empirical Validation

To validate the analysis, we run repeated trials to estimate the state of the platform under different motion but identical alignment (the camera is rigidly connected to the IMU) using our experimental platform [Tsotsos et al., 2015]. If alignment parameters \((T_{cb}, \Omega_{cb})\) were identifiable (or the augmented state observable), we would expect convergence to the same parameters across all trials. Instead, Fig. 1 shows that the estimates of the parameters stabilize, but to different values at each run. Nevertheless, the parameter values are in a set, whose vol-

![Figure 1: Convergence of alignment parameters to a set, rather than a unique point estimate, due to the lack of unknown-input observability in the presence of (realistic) non-constant biases. The mean (solid) and twice the std. dev. (dashed) of the change in estimated parameters relative to their initial nominal values across multiple trials on real data, show that different trials converge to different parameter values, but to within a bounded set.](image-url)

Here \(\sigma(g)\) is a scaled rigid motion: if \(g = (R, T)\), then \(\sigma(g) = (R, \sigma T)\).
analyzing the convergence characteristics of (any) filters for
is, the set of indistinguishable states is not a singleton.

as unknown inputs, the resulting model is
a filter used for navigation estimates, with bias rates treated
sensor biases are included as model parameters in the state of

This paper presents an overview of the analysis presented in
4 Discussion
sufficiently exciting motion. Figures 4 and 5 show the re-
cted Monte-Carlo experiments on the model in simulation
duced static and time-varying biases while undergoing
sulting estimation errors of alignment states for 50 trials each

the accel and gyro biases have been artificially inflated by
Adding, we con-

Figure 3: The indistinguishable set also depends on the char-
acteristics of the sensor, and its volume is directly propor-
tional to the sensor bias rate. Here artificial bias drift is added,
resulting in a larger indistinguishable set compared to Fig. 1.

Figure 4: Mean (solid line) and twice the standard deviation
(dashed lines) of estimation errors of alignment parameters
aggregated over 50 Monte-Carlo trials with a constant bias.

vision-aided inertial navigation not as one of observability or
identifiability, but one of sensitivity, by bounding the set of
indistinguishable trajectories to a set whose volume depends
on motion characteristics.

The advantage of this approach, compared to the standard
observability analysis based on rank conditions, is that we
characterize the indistinguishable set explicitly. We quantify
the “degree of unobservability” as the sensitivity of the solu-
tion set to the input; provided that sufficient-excitation condi-
tions are satisfied, the unobservable set can be bounded and
effectively be treated as a singleton. More generally, however,
the analysis provides an estimate of the uncertainty surround-
ning the solution set, as well as a guideline on how to limit it
by enforcing certain gauge transformations.

Acknowledgements
This work was supported by the Air Force Office of Scientific
Research (grant no. AFOSR FA9550-12-1-0364) and the
Office of Naval Research (grant no. ONR N00014-13-1-034).

4173
References


