Observability, Identifiability and Sensitivity of Vision-Aided Inertial Navigation

Joshua Hernandez[1] Konstantine Tsotsos[1] Stefano Soatto[1]

***Abstract*— We analyze the observability of 3-D pose from the fusion of visual and inertial sensors. Because the model contains unknown parameters, such as sensor biases, the problem is usually cast as a mixed filtering/identification, 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 patently 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. In other words, the resulting model is not observable. We therefore 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.**

I. Introduction

We present a novel approach to the analysis of observability/identifiability of three-dimensional (3-D) pose in visually-assisted navigation, whereby inertial sensors (accelerometers and gyrometers) are used in conjunction with optical sensors (vision) to yield an estimate of the 3-D position and orientation 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 [1], [2], [3], [4], [5].

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 [6], [7], for affine systems [8], and non-linear systems in [9], [10], [11]. 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 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.

A. Notation

We adopt the notation of [12], where a reference frame is represented by an orthogonal 3×3 positive-determinant (rotation) matrix $R \in \mathrm{SO}(3) \doteq \{R \in \mathbb{R}^{3\times 3} \mid R^T R = R R^T = I,\ \det(R) = +1\}$ and a translation vector $T \in \mathbb{R}^3$. They are collectively

[1]The authors are with the Computer Science Department, University of California, Los Angeles, USA. Email: {jheez,ktsotsos,soatto}@ucla.edu.

indicated by $g = (R, T) \in \mathrm{SE}(3)$. 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 $p_b = g_{ba}p_a$. In coordinates, 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 pose is indicated with $g(t) = (R(t), T(t))$ or $g_t = (R_t, T_t)$, and the entire trajectory from an initial time t_i and a final time t_f $\{g(t)\}_{t=t_i}^{t_f}$ is indicated in short-hand notation with $g_{t_i}^{t_f}$; when the initial time is $t_0 = 0$, we omit the subscript and call g^t the trajectory "up to time t". The time-index is sometimes omitted for simplicity of notation when it is clear from the context.

We indicate with $\widehat{V} = (\widehat{\omega}, v) \in \mathfrak{se}(3)$ the (generalized) velocity or "twist", where $\widehat{\omega}$ is a skew-symmetric matrix $\widehat{\omega} \in \mathfrak{so}(3) \doteq \{S \in \mathbb{R}^{3\times 3} \mid S^T = -S\}$ corresponding to the cross product with the vector $\omega \in \mathbb{R}^3$, so that $\widehat{\omega}v = \omega \times v$ for any vector $v \in \mathbb{R}^3$. We indicate the generalized velocity with $V = (\omega, v)$. We indicate the group composition $g_1 \circ g_2$ simply as $g_1 g_2$. In homogeneous coordinates, $\bar{X}_b = G_{ba}\bar{X}_a$ where $\bar{X}^T = [X^T \ 1]$ and

$$G \doteq \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \in \mathbb{R}^{4\times 4} \qquad \hat{V} \doteq \begin{bmatrix} \widehat{\omega} & v \\ 0 & 0 \end{bmatrix}.$$

Composition of rigid motions is then represented by matrix product.

B. Mechanization Equations

The motion of a sensor platform is represented as the time-varying pose g_{sb} of the body relative to the spatial frame. To relate this to measurements of an inertial measurement unit (IMU) we compute the temporal derivatives of g_{sb}, which yield the (generalized) body velocity V_{sb}^b, defined by $\dot{g}_{sb}(t) = g_{sb}(t)\widehat{V}_{sb}^b(t)$, which can be broken down into the rotational and translational components $\dot{R}_{sb}(t) = R_{sb}(t)\widehat{\omega}_{sb}^b(t)$ and $\dot{T}_{sb}(t) = R_{sb}(t)v_{sb}^b(t)$. An ideal gyrometer (gyro) would measure $\omega_{\mathrm{imu}} = \omega_{sb}^b$. The translational component of body velocity, v_{sb}^b, can be obtained from the last column of the matrix $\frac{d}{dt}\widehat{V}_{sb}^b(t)$. That is, $\dot{v}_{sb}^b = \dot{R}_{sb}^T\dot{T}_{sb} + R_{sb}^T\ddot{T}_{sb} = -\widehat{\omega}_{sb}^b v_{sb}^b + R_{sb}^T\ddot{T}_{sb} \doteq -\widehat{\omega}_{sb}^b v_{sb}^b + \alpha_{sb}^b$, which serves to define $\alpha_{sb}^b \doteq R_{sb}^T\ddot{T}_{sb}$. These equations can be simplified by defining a new linear velocity, v_{sb}, which is neither the body velocity v_{sb}^b nor the spatial velocity v_{sb}^s, but instead $v_{sb} \doteq R_{sb}v_{sb}^b$. Consequently, we have that $\dot{T}_{sb}(t) = v_{sb}(t)$ and $\dot{v}_{sb}(t) = \dot{R}_{sb}v_{sb}^b + R_{sb}\dot{v}_{sb}^b = \ddot{T}_{sb} \doteq \alpha_{sb}(t)$ where the last equation serves to define the new linear acceleration α_{sb}; as one can easily verify, we have that $\alpha_{sb} = R_{sb}\alpha_{sb}^b$. An ideal accelerometer (accel) would then measure $\alpha_{\mathrm{imu}} = R_{sb}^T(t)(\alpha_{sb}(t) - \gamma)$ where $\gamma \in \mathbb{R}^3$ is the gravity vector.

There are several reference frames to be considered in a navigation scenario. The *spatial frame* s, typically attached to Earth and oriented so that gravity γ takes the form $\gamma^T = [0\ 0\ 1]^T\|\gamma\|$ where $\|\gamma\|$ can be read from tabulates based on location and is typically around $9.8m/s^2$. The *body frame* b is attached to the IMU. The *camera frame* c, relative to which image measurements are captured, is also unknown, although we will assume that *intrinsic calibration* has been performed, so that measurements on the image plane are provided in metric units.

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 sb wherever it appears. This yields $\dot{T} = v,\ \dot{R} = R\widehat{\omega},\ \dot{v} = \alpha, \dot{\omega} = w,\ \dot{\alpha} = \xi$ where $w \in \mathbb{R}^3$ is the rotational acceleration, and $\xi \in \mathbb{R}^3$ the translational jerk (derivative of acceleration).

C. Sensor model

Although the acceleration α defined above corresponds to neither body nor spatial acceleration, it is conveniently related to accelerometer measurements α_{imu}:

$$\alpha_{\mathrm{imu}}(t) = R^T(t)(\alpha(t) - \gamma) + \underbrace{\alpha_b(t) + n_\alpha(t)} \qquad (1)$$

where the measurement error in bracket includes a slowly-varying mean ("bias") $\alpha_b(t)$ and a residual term n_α 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_α is independent of α, hence uninformative. Here γ is the gravity vector expressed in the spatial frame. Measurements from a gyro, ω_{imu}, can be similarly modeled as

$$\omega_{\mathrm{imu}}(t) = \omega(t) + \underbrace{\omega_b(t) + n_\omega(t)} \qquad (2)$$

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

Other than the fact that the biases α_b, ω_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,

$$\dot{\omega}_b(t) = w_b(t), \quad \dot{\alpha}_b(t) = \xi_b(t) \tag{3}$$

for some unknown inputs w_b, ξ_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 ω_b, α_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.

D. Model reduction

The mechanization equations above define a dynamical model having as output the IMU measurements. Including the initial conditions and biases, we have

$$\begin{cases} \dot{T} = v & T(0) = 0 \\ \dot{R} = R\widehat{\omega} & R(0) = R_0 \\ \dot{v} = \alpha \\ \dot{\omega} = w \\ \dot{\alpha} = \xi \\ \dot{\omega}_b = w_b \\ \dot{\alpha}_b = \xi_b \\ \dot{\gamma} = 0 \\ \omega_{\text{imu}}(t) = \omega(t) + \omega_b(t) + n_\omega(t) \\ \alpha_{\text{imu}}(t) = R^T(t)(\alpha(t) - \gamma) + \alpha_b(t) + n_\alpha(t) \end{cases} \tag{4}$$

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 ω in terms of ω_{imu} and α in terms of α_{imu}:

$$\omega = \omega_{\text{imu}} - \omega_b + \underbrace{n_R}_{-n_\omega} \qquad \alpha = R(\alpha_{\text{imu}} - \alpha_b) + \gamma + \underbrace{n_v}_{-Rn_\alpha} \tag{5}$$

This equality is valid for *samples* (realizations) of the stochastic processes involved, but it can be misleading as, if considered as stochastic processes, the noises above are *not* independent of the states. Such a dependency is nevertheless typically neglected. The resulting mechanization model is

$$\begin{cases} \dot{T} = v & T(0) = 0 \\ \dot{R} = R(\widehat{\omega}_{\text{imu}} - \widehat{\omega}_b) + n_R & R(0) = R_0 \\ \dot{v} = R(\alpha_{\text{imu}} - \alpha_b) + \gamma + n_v \\ \dot{\omega}_b = w_b \\ \dot{\alpha}_b = \xi_b. \end{cases} \tag{6}$$

E. Imaging model and alignment

Initially we assume there is a collection of points $X^i,\ i = 1, \dots, N$, visible from time $t = 0$ to the current time t. If $\pi : \mathbb{R}^3 \to \mathbb{R}^2; X \mapsto [X_1/X_3,\ X_2/X_3]$ 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

$$y^i(t) = \frac{R^T_{1:2}(t)(X^i - T_{1:2}(t))}{R^T_3(t)(X^i - T_3(t))} \doteq \pi(g^{-1}(t)X^i) + n^i(t) \tag{7}$$

If the feature first appears at time $t = 0$ *and if the camera reference frame is chosen to be the origin the world reference frame* so that $T(0) = 0; R(0) = I$, then we have that $y^i(0) = \pi(X^i) + n^i(0)$, and therefore

$$X^i = \bar{y}^i(0)Z^i + \tilde{n}^i \tag{8}$$

where $\bar{y}$ is the homogeneous coordinate of y, $\bar{y} = [y^T\ 1]^T$, and $\tilde{n}^i = [{n^i}^T(0)Z^i \quad 0]^T$. Here Z^i is the (unknown, scalar) depth of the point at time $t = 0$, and again the dependency of the noise on the state is neglected. With an abuse of notation, we write the map that collectively projects all points to their corresponding locations on the image plane as $y(t) = \pi(g^{-1}(t)\mathbf{X}) + n(t)$, or:

$$y(t) \doteq \begin{bmatrix} y^1 \\ y^2 \\ \vdots \\ y^N \end{bmatrix}(t) = \begin{bmatrix} \pi(R^T(X^1 - T)) \\ \pi(R^T(X^2 - T)) \\ \vdots \\ \pi(R^T(X^N - T)) \end{bmatrix} + \begin{bmatrix} n^1(t) \\ n^2(t) \\ \vdots \\ n^N(t) \end{bmatrix} \tag{9}$$

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":

$$y^i(t) = \pi\left(g_{cb}g^{-1}(t)X^i_s\right) + n^i(t) \in \mathbb{R}^2 \tag{10}$$

We can then, as done for the points X^i, add it to the state with trivial dynamics $\dot{g}_{cb} = 0$.

It may be convenient in some cases to represent the points X^i_s in the reference frame where they first appear,

[1] Intrinsic calibration parameters are known and compensated for.

[2] The pose of the camera relative to the IMU is known and compensated for.

say at time t_i, rather than in the spatial frame. This is because the uncertainty is highly structured in the frame where they first appear: if $X^i(t_i) = \bar{y}^i(t_i)Z^i(t_i)$, then $y^i(t_i)$ has the same uncertainty of the feature detector (small and isotropic on the image plane) and Z^i has a large uncertainty, but it is constrained to be positive.

However, to relate $X^i(t_i)$ to the state, we must bring it to the spatial frame, via $g(t_i)$, which is unknown. Although we may have a good approximation of it, the current estimate of the state $\hat{g}(t_i)$, the pose when the point first appears should be estimated along with the coordinates of the points. Therefore, we can represent X^i using $y^i(t_i)$, $Z^i(t_i)$ *and* $g(t_i)$:

$$X^i_s = X^i_s(g_{t_i}, y_{t_i}, Z_{t_i}) = g_{t_i}\bar{y}_{t_i}Z_{t_i} \tag{11}$$

Clearly this is an over-parametrization, since each point is now represented by $3 + 6$ parameters instead of 3. However, the pose g_{t_i} can be pooled among all points that appear at time t_i, considered therefore as a *group*. At each time, there may be a number $j = 1, \dots, K(t)$ groups, each of which has a number $i = 1, \dots, N_j(t)$ points. We indicate the group index with j and the point index with $i = i(j)$, omitting the dependency on j for simplicity. The representation of X^i_s then evolves according to

$$\begin{cases} \dot{y}^i_{t_i} = 0, \quad i = 1, \dots, N(j) \\ \dot{Z}^i_{t_i} = 0 \\ \dot{g}_j = 0, \quad j = 1, \dots, K(t). \end{cases} \tag{12}$$

II. 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 $\omega_{\text{imu}}, \alpha_{\text{imu}}$ and y should constrain the set of trajectories that satisfy (6)-(12) to be unique. If there are different trajectories that satisfy (4) with the same outputs, 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 (4) or its equivalent reduced realization, 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 [13].

Claim 1 (Indistinguishable Trajectories): Let $g(t) = (R(t), T(t)) \in \text{SE}(3)$ satisfy (6)-(12) for some known constant γ and functions $\alpha_{\text{imu}}(t)$, $\omega_{\text{imu}}(t)$ and for some unknown functions $\alpha_b(t), \omega_b(t)$ that are constrained to have $\|\dot{\alpha}_b(t)\| \leq \epsilon$, $\|\dot{\omega}_b(t)\| \leq \epsilon$, and $\|\ddot{\omega}_b(t)\| \leq \epsilon$ at all t, for some $\epsilon < 1$.

Suppose $\tilde{g}(t) \doteq \sigma(g_B g(t) g_A)$ 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[3] $\|T_A\| \leq M_A$ and $0 < m_\sigma \leq |\sigma| \leq M_\sigma$. Then, under sufficient excitation conditions, $\tilde{g}(t)$ satisfies (6)-(12) if and only if

$$\|I - R_A\| \leq \frac{2\epsilon}{m(\dot{\omega}_{\text{imu}} : \mathbb{R}^+)} \tag{13}$$

$$|\sigma - 1| \leq \frac{k_1\epsilon + M_\sigma \|I - R_A\|}{m(\dot{\alpha}_{\text{imu}} : \mathcal{I}_1)} \tag{14}$$

$$\|T_A\| \leq \frac{\epsilon(k_2 + (2M_\sigma + 1)M_A)}{m_\sigma \, m(\ddot{\omega}_{\text{imu}} : \mathcal{I}_2)} \tag{15}$$

$$\begin{aligned} \|(1 - R_B^T)\gamma\| \leq \frac{\epsilon(k_3 + M_\sigma M_A)}{m_\sigma \, m(\omega_{\text{imu}} - \omega_b : \mathcal{I}_3)} + \\ + \frac{(|\sigma - 1| + \epsilon)M(\omega_{\text{imu}} - \omega_b : \mathcal{I}_3)\|\gamma\|}{m_\sigma \, m(\omega_{\text{imu}} - \omega_b : \mathcal{I}_3)} \end{aligned} \tag{16}$$

for $\mathcal{I}_i$ and k_i determined by the sufficient excitation conditions.

The set of indistinguishable trajectories in the limit where $\epsilon \to 0$ is parametrized by an arbitrary $T_B \in \mathbb{R}^3$ and $\theta \in \mathbb{R}$,

$$\begin{cases} \tilde{T} = \exp(\widehat{\gamma}\theta)T + T_B \\ \tilde{R} = \exp(\widehat{\gamma}\theta)R \\ \tilde{T}_{t_i} = \exp(\widehat{\gamma}\theta)\bar{T}_{t_i} + T_B \\ \tilde{R}_{t_i} = \exp(\widehat{\gamma}\theta)\bar{R}_{t_i} \qquad \text{up to } \mathcal{O}\left(\frac{\|\dot{\omega}_b\|}{\|\dot{\omega}_{\text{imu}}\|}, \frac{\|\dot{\alpha}_b\|}{\|\dot{\alpha}_{\text{imu}}\|}, \frac{1}{\|\gamma\|}\right) \\ \tilde{T}_{cb} = T_{cb} \\ \tilde{R}_{cb} = R_{cb} \end{cases} \tag{17}$$

If we impose that $T(0) = \tilde{T}(0) = 0$, then $T_B = 0$ is determined; similarly, if we impose the initial pose to be aligned with gravity (so gravity is in the form $[0 \;\; 0 \;\; \|\gamma\|]^T$), then $\theta = 0$. But while we can impose this condition, we cannot *enforce* it, since the initial condition is not a part of the state of the filter, so we cannot relate the measurements at each time t directly to it.

However, if the reference can be associated to *constant parameters* that are part of the state of the model, it can be enforced in a consistent manner. For instance, the ambiguous set of points is

$$\tilde{X}^j = g_a \bar{g}_i^{-1} g_i g_a^{-1} X^j, \tag{18}$$

[3]Here $\sigma(g)$ is a scaled rigid motion: if $g = (R, T)$, then $\sigma(g) = (R, \sigma T)$.

if each group i contains at least 3 non-coplanar points, it is possible to fix $\bar{g}_i$ by parameterizing $X^j \doteq \bar{y}^j_{t_i} Z^j$ and imposing three directions $y^j_{t_i} = \tilde{y}^j_{t_i} = y^j(t_i), j = 1, \dots, 3$, the measurement of these directions at time t_i when they first appear. This yields $\bar{g}_i = g_i$ and $\tilde{X}^j = X^j$ for that group. Note that it is necessary to impose this constraint in *each group*.

The residual set of indistinguishable trajectories is parameterized by *constants* θ, T_B, that determine a Gauge transformation for the groups, that can be fixed by always fixing the pose of *one* of the groups. This can be done in a number of ways. For instance, if for a certain group of points indexed by i we impose

$$R_{t_i} = \tilde{R}_{t_i} = \hat{R}(t_i) \text{ and } T_{t_i} = \tilde{T}_{t_i} = \hat{T}(t_i) \tag{19}$$

by assigning their value to the current best estimate of pose and not including the corresponding variables in the state of the model, then we have that

$$\hat{R}(t_i) = \exp(\widehat{\gamma}\theta)\hat{R}(t_i) \tag{20}$$

and therefore $\theta = 0$; similarly,

$$T_B = (I - \exp(\widehat{\gamma}\theta))T(t_i) = 0. \tag{21}$$

Therefore, the gauge transformation is enforced explicitly at each instant of time, as each measurement provides a constraint on the states. After the Gauge Transformation has been fixed, the model is observable in the limit $\epsilon \to 0$, and otherwise the state of an observer is related to the true one as follows:

$$\begin{aligned}
\tilde{X}^{\text{ref}} &= (1+\tilde{\sigma})\tilde{R}_{cb}e^{\omega_B}e^{\widehat{\gamma}\theta}e^{\omega_A}\tilde{R}^T_{cb}(X^{\text{ref}} - T_A) + \\
&\quad +(1+\tilde{\sigma})(\tilde{R}_{cb}e^{\omega_A}T_B + \tilde{R}_{cb}T_A + \tilde{T}_{cb}) && (22)\\
\tilde{X}^j &= (1+\tilde{\sigma})\tilde{R}_{cb}\bar{R}_i\tilde{R}_{t_i}\tilde{R}^T_{cb}(X^j - T_A) + \\
&\quad +(1+\tilde{\sigma})(\tilde{R}_{cb}\bar{R}_i\tilde{T}_{t_i} + \tilde{R}_{cb}\bar{T}_i + \tilde{T}_{cb}) && (23)\\
\tilde{T} &= e^{\widehat{\gamma}\theta}T + T_B(1+\tilde{\sigma}) + \\
&\quad +\omega_B e^{\widehat{\gamma}\theta}T + e^{\omega_B}e^{\widehat{\gamma}\theta}RT_A(1+\tilde{\sigma}) && (24)\\
\tilde{R} &= e^{\omega_B}e^{\widehat{\gamma}\theta}Re^{\omega_A} && (25)\\
\tilde{T}_{t_i} &= e^{\widehat{\gamma}\theta}\bar{T}_i + T_B(1+\tilde{\sigma}) + \\
&\quad +\omega_B e^{\widehat{\gamma}\theta}\bar{T}_i + e^{\omega_B}e^{\widehat{\gamma}\theta}\bar{R}_iT_A(1+\tilde{\sigma}) && (26)\\
\tilde{R}_{t_i} &= e^{\omega_B}e^{\widehat{\gamma}\theta}\bar{R}_ie^{\omega_A} && (27)\\
\tilde{T}_{cb} &= T_{cb} + \tilde{\sigma}T_{cb} + R_{cb}T_A(1+\tilde{\sigma}) && (28)\\
\tilde{R}_{cb} &= R_{cb}\exp(\omega_A) && (29)
\end{aligned}$$

where ω_A, R_A, σ, T_A, ω_B, R_B satisfy (13)-(16), and θ, T_B are arbitrary. The groups will be defined up to an arbitrary reference frame $(\bar{R}_i, \bar{T}_i)$, except for the reference group where that transformation is fixed. Note that, as the reference group "switches" (when points in the reference group become occluded or otherwise disappear due to failure in the data association mechanism), a small error in pose is accumulated. This error affects the gauge transformation, not the *state* of the system, and therefore is not reflected in the innovation, nor in the covariance of the state estimate, that remains bounded. This is unlike [3], where the covariance of the translation state T_B and the rotation about gravity θ grows unbounded over time, possibly affecting the numerical aspects of the implementation. Notice that in the limit where $\dot{\omega}_b = \dot{\alpha}_b = 0$, we obtain back Eq. (17). Otherwise, the equations above immediately imply the following

Claim 2 (unknown-input observability): The model (6)-(12) 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 [13] for 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.

III. 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 and the connection is stable to high precision). If alignment parameters were identifiable (or the augmented state observable), we would expect convergence to the same parameters across all trials. Instead, Fig. 1 shows that the filter reaches steady-state, with the estimates of the parameters stabilizing, but to different values at each run. Nevertheless, the parameter values are in a set, whose volume can be bounded based on the analysis above and the characteristics of the sensor. In particular, less stable biases, and less exciting motions, result in a larger indistinguishable set: Fig. 2 shows the same experiments with more gentle (hence less exciting) motions. Fig. 3 shows the same where the accel and gyro biases have been artificially inflated by adding a slowly time-varying offset to the IMU measurements. To further support the conclusions of the analysis, Monte-Carlo experiments were conducted on the model in simulation

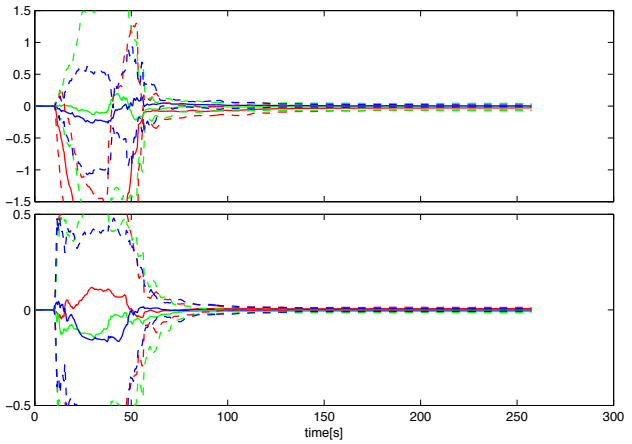

Fig. 1. *Convergence of alignment parameters (top translational, bottom rotational) to a set, rather than a unique point estimate, due to the lack of unknown-input observability in the presence of non-constant biases. The mean (solid line) and twice the standard deviation (dashed lines) of the change in estimated parameters relative to their initial nominal values across multiple trials on real data collected with our experimental framework, show that different trials converge to different parameter values, but to within a bounded set. The standard deviations of the converged translational parameters (in centimeters) are [1.76 2.8 0.77] and [0.0032 0.0029 0.0033] for the rotational parameters (in radians).*

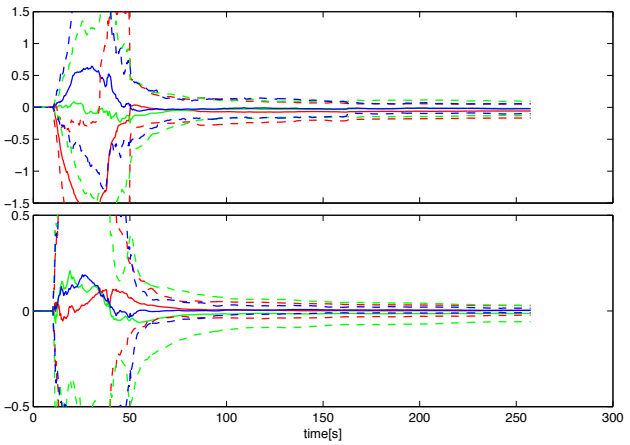

Fig. 2. *The indistinguishable set is bounded depending on the characteristic of the motion, that has to be sufficiently exciting. Gentler motion produces multiple trials that converge (top translational, bottom rotational) to a set of larger volume compared to Fig. 1. The standard deviations of the converged translational parameters (in centimeters) are [4.43 5.98 3.57] and [0.0069 0.0079 0.0062] for the rotational parameters (in radians).*

using stationary and time-varying biases while undergoing sufficiently exciting motion. For each trial, the platform views a consistent set of randomly generated points (no occlusions) while circling the point set on randomly generated trajectories. Figures 4 and 5 show the resulting estimation errors of the alignment states for 20 trials each using a constant and white-noise driven bias respectively. As seen in the experiments with real data, estimates in the time-varying bias scenario do not converge to a singleton.

The experiments thus confirm the analysis.

IV. Discussion

We have shown that when inertial sensor biases are included as model parameters in the state of a filter used for navigation estimates, with bias rates treated as unknown inputs, the resulting model is *not observable*.

Consequently, we have re-formulated the problem of analyzing the convergence characteristics of (any) filters for 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. Furthermore, rank conditions are "fragile" in the sense that the model can be nominally observable, and yet the condition number of the observability matrix be so small as to render the model effectively unobservable. We quantify the "degree of unobservability" as the sensitivity of the solution set to the input; provided that sufficient-excitation conditions are satisfied, the unobservable set can be bounded and effectively be treated as a singleton. More in general, however, the analysis provides an estimate of the uncertainty surrounding 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).

References

[1] S. Soatto, "Observability/identifiability of rigid motion under perspective projection," in *Decision and Control, 1994., Proceedings of the 33rd IEEE Conference on*, vol. 4. IEEE, 1994, pp. 3235–3240.

[2] J. Kelly and G. Sukhatme, "Fast Relative Pose Calibration for Visual and Inertial Sensors," in *Experimental Robotics*, 2009, pp. 515–524.

[3] A. Mourikis and S. Roumeliotis, "A multi-state constraint kalman filter for vision-aided inertial navigation," in *Robotics and Automation, 2007 IEEE International Conference on*. IEEE, 2007, pp. 3565–3572.

[4] E. S. Jones, A. Vedaldi, and S. Soatto, "Inertial structure from motion and autocalibration," in *Workshop on Dynamical Vision*, October 2007.

[5] A. Martinelli *et al.*, "Visual-inertial structure from motion: observability and resolvability," *Foundations and Trends® in Computer Graphics and Vision*, vol. 1, no. 1, 2014.

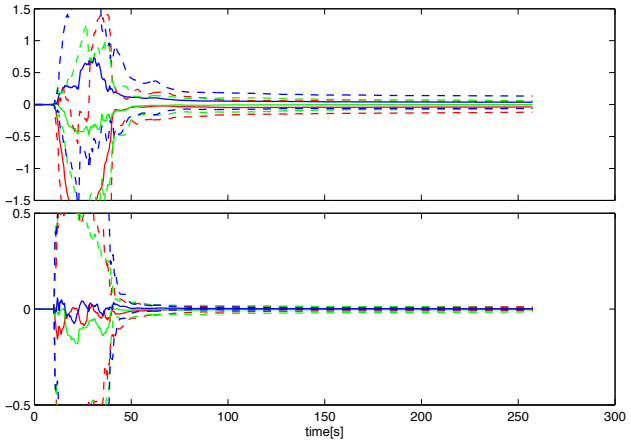

Fig. 3. *The indistinguishable set also depends on the characteristics of the sensor, and its volume is directly proportional to the sensor bias rate. Here artificial bias is added to the measurements, resulting in a larger indistinguishable set (top translational alignment, bottom rotational alignment) compared to Fig. 1. The standard deviations of the converged translational parameters (in centimeters) are [4.09 3.1 4.88] and [0.0053 0.0061 0.002] for the rotational parameters (in radians).*

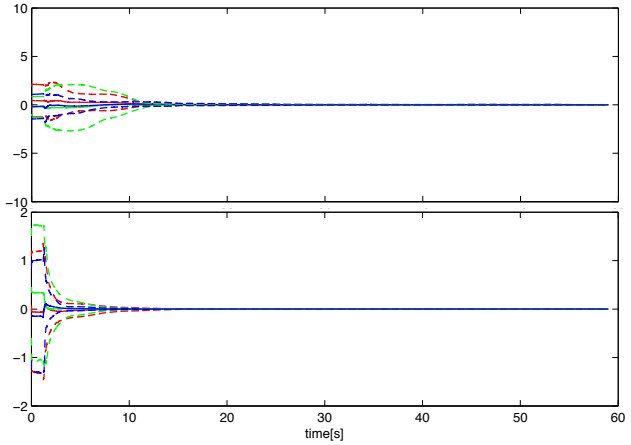

Fig. 4. *Mean (solid line) and twice the standard deviation (dashed lines) of squared estimation errors of alignment parameters (top translational, bottom rotational) aggregated over 50 Monte-Carlo trials with a constant bias.*

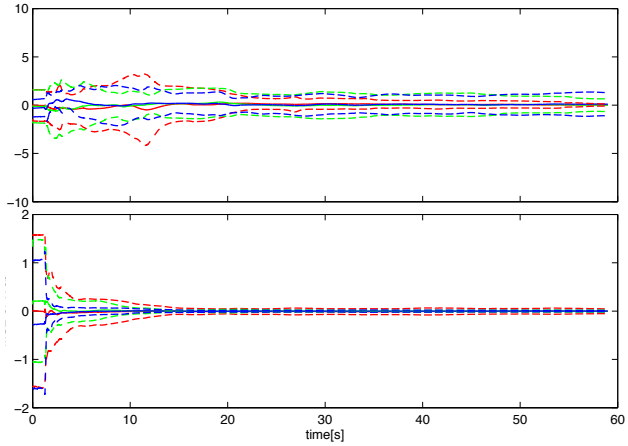

Fig. 5. *Mean (solid line) and twice the standard deviation (dashed lines) of squared estimation errors of alignment parameters (top translational, bottom rotational) aggregated over 50 Monte-Carlo trials with a time-varying bias with similar noise characteristics to the simulated sensors models.*

[6] G. Basile and G. Marro, "On the observability of linear, time-invariant systems with unknown inputs," *Journal of Optimization theory and applications*, vol. 3, no. 6, pp. 410–415, 1969.

[7] F. Hamano and G. Basile, "Unknown-input present-state observability of discrete-time linear systems," *Journal of Optimization Theory and Applications*, vol. 40, no. 2, pp. 293–307, 1983.

[8] H. Hammouri and Z. Tmar, "Unknown input observer for state affine systems: A necessary and sufficient condition," *Automatica*, vol. 46, no. 2, pp. 271–278, 2010.

[9] H. Dimassi, A. Loría, and S. Belghith, "A robust adaptive observer for nonlinear systems with unknown inputs and disturbances," in *Decision and Control (CDC), 2010 49th IEEE Conference on*. IEEE, 2010, pp. 2602–2607.

[10] D. Liberzon, P. R. Kumar, A. Dominguez-Garcia, and S. Mitra, "Invertibility and observability of switched systems with inputs and outputs," 2012.

[11] S. Bezzaoucha, B. Marx, D. Maquin, J. Ragot, *et al.*, "On the unknown input observer design: a decoupling class approach with application to sensor fault diagnosis," in *1st International Conference on Automation and Mechatronics, CIAM'2011*, 2011.

[12] R. M. Murray, Z. Li, and S. S. Sastry, *A Mathematical Introduction to Robotic Manipulation*. CRC Press, 1994.

[13] J. Hernandez and S. Soatto, "Observability, identifiability, sensitivity, and model reduction for vision-assisted inertial navigation," *UCLA CSD TR13022, http://arxiv.org/abs/1311.7434*, Aug. 20, 2013 (revised Nov. 12, 2013; Nov 29. 2013).