Abstract

We analyze the observability of motion estimates 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 identification/filtering, and the resulting observability analysis provides a necessary condition for any algorithm to converge to a unique point estimate. Unfortunately, most models treat sensor bias rates as “noise,” independent of other states including biases themselves, an assumption that is patently violated in practice. When this assumption is lifted, the resulting model is not observable, and therefore past 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 indistinguishable set is a singleton, which is not the case, we derive bounds on its volume, as a function of characteristics of the input and its sufficient excitation. This provides an explicit characterization of the indistinguishable set that can be used for analysis and validation purposes.

1 Introduction

This manuscript describes a novel approach to the analysis of observability/identifiability of visually-assisted navigation, whereby inertial sensors (accelerometers and gyroimeters) are used in conjunction with optical sensors (vision) to yield an estimate of the three-dimensional 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.

The unknown parameters in the model are typically treated as unknown constants (e.g., calibration parameters) or as random walks (e.g., accelerator and gyro biases), and treated as states in the model, driven by some kind of uninformative input (noise). Because noise does not affect the observability of a model, for the purpose of observability analysis, they are set to zero. This is because, by assumption, noise is “uninformative.” It is typically modeled as a realization of a white zero-mean, homoscedastic process, independent of the state of the model. However, the driving input to the random walk model of accelerometer and gyro bias is typically small but not independent of the state. In fact, far from being uninformative, it is strongly correlated with it, as it is its temporal derivative. Thus, it should be treated as an unknown input, rather than a “noise.”

Our first contribution is to show that while (a prototypical model of) assisted navigation and auto-calibration is observable in the absence of unknown input, 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), but it is...
nevertheless bounded to a set. We explicitly characterize this set and show that, interestingly, it may not contain the “true” state trajectory. Finally, we provide bounds on the volume of this subset as a function of the characteristics of the unknown inputs.

Rather than study observability of linearized system, or algebraically checking the rank conditions, that offers no insight on the structure of the indistinguishable states, we characterize observability directly in terms of indistinguishable sets.

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 [1, 3, 4], for affine systems [5], and non-linear systems in [3, 9, 2]. 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.

1.1 Notation
A reference frame is represented by an orthogonal $3 \times 3$ positive-determinant (rotation) matrix $R \in \text{SO}(3) \doteq \{ R \in \mathbb{R}^{3 \times 3} \mid R^T R = R R^T = I, \det(R) = +1 \}$ and a translation vector $T \in \mathbb{R}^3$. They are collectively indicated by $g = (R, T) \in \text{SE}(3)$. 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_i = t_f} is indicated in short-hand notation with $g_{i}^{f}$; 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 $\hat{V} = (\hat{\omega}, v) \in \mathfrak{s}(3)$ the (generalized) velocity or “twist”, where $\hat{\omega}$ is a skew-symmetric matrix $\hat{\omega} \in \mathfrak{s}(3) \doteq \{ S \in \mathbb{R}^{3 \times 3} \mid S^T = -S \}$ corresponding to the cross product with the vector $\omega \in \mathbb{R}^3$, so that $\hat{\omega} v = \omega \times v$ 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, $\hat{X}_b = G_{ba} \hat{X}_a$ where $\hat{X}_b = [X_b \ t_1]$ and

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

Composition of rigid motions is then represented by matrix product.

1.2 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 $V^b_{sb}$, defined by $\dot{g}_{sb}(t) = g_{sb}(t) \dot{V}^b_{sb}(t)$, which can be broken down into the rotational and translational components $\dot{R}_{sb}(t) = R_{sb}(t) \hat{\omega}^b_{sb}(t)$ and $\dot{T}_{sb}(t) = R_{sb}(t) v^b_{sb}(t)$. An ideal gyro (gyro) would measure $\hat{\omega}^b_{sb}$. The translational component of body velocity, $v^b_{sb}$, can be obtained from the last column of the matrix $\frac{d}{dt} \hat{V}^b_{sb}(t)$. That is, $v^b_{sb} = \dot{R}^T_{sb} \dot{T}_{sb} + R^T_{sb} \ddot{T}_{sb} = -\hat{\omega}^b_{sb} v^b_{sb} + R^T_{sb} \ddot{T}_{sb}$, which serves to define $\alpha^b_{sb} = \hat{\omega}^b_{sb} v^b_{sb} + \dot{\alpha}^b_{sb}$, which serves to define $\alpha^b_{sb} = \hat{\omega}^b_{sb} v^b_{sb} + \dot{\alpha}^b_{sb}$. These equations can be simplified by defining a new linear velocity, $\nu_{sb}$, which is neither the body velocity $v^b_{sb}$ nor the spatial velocity $v^s_{sb}$, but instead $\nu_{sb} = R_{sb} v^b_{sb}$. Consequently, we have that $\dot{\nu}_{sb}(t) = v_{sb}(t)$ and $\dot{\nu}_{sb}(t) = R_{sb} \dot{v}^b_{sb} + R_{sb} \dot{v}^s_{sb} = \ddot{T}_{sb} \hat{\omega}^b_{sb} + \dot{\alpha}^b_{sb}$ (where the last equation serves to define the new linear acceleration $\alpha_{sb}$; as one can easily verify we have that $\alpha_{sb} = R_{sb} \alpha^b_{sb}$). An ideal accelerometer (accel) would then measure $\alpha^b_{imu} = R^T_{sb}(t)(\alpha_{sb}(t) - \gamma)$. 


There are several reference frames to be considered in an aided navigation scenario. The spatial frame $s$, typically attached to Earth and oriented so that gravity $\gamma$ takes the form $\gamma^T = [0 \ 0 \ 1]^T \|\gamma\|$ 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\footnote{In practice, the IMU has several different frames due to the fact that the gyro and accel are not co-located and aligned, and even each sensor (gyro or accel) is composed of multiple sensors, each of which can have its own reference frame. Here we will assume that the IMU has been pre-calibrated so that accel and gyro yield measurements relative to a common reference frame, the body frame. In reality, it may be necessary to calibrate the alignment between the multiple-axes sensors (non-orthogonality), as well as the gains (scale factors) of each axis.} 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. Finally, the radar frame, or range frame $r$, is that of the antenna relative to which range measurements are provided.

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 $sb$ everywhere it appears. This yields

\[
\begin{aligned}
\dot{T} &= V \\
\dot{R} &= R \hat{\omega} \\
\dot{V} &= \alpha \\
\dot{\omega} &= \xi \\
\dot{\alpha} &= \xi
\end{aligned}
\tag{1}
\]

where $w \in \mathbb{R}^3$ is the rotational acceleration, and $\xi \in \mathbb{R}^3$ the translational jerk (derivative of acceleration). Although $\alpha$ corresponds to neither body nor spatial acceleration, it can be easily related to accel measurements:

\[
\alpha_{imu}(t) = R^T(t)(\alpha(t) - \gamma) + \alpha_b(t) + n_\alpha(t)
\tag{2}
\]

where the measurement error in bracket includes a slowly-varying mean (“bias”) $\alpha_b(t)$ and a residual term $n_\alpha(t)$ 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. Here $\gamma$ is the gravity vector expressed in the spatial frame.\footnote{The orientation of the body frame relative to gravity, $R_0$, is unknown, but can be approximated by keeping the IMU still (so $R^T(t) = R_0$) and averaging the accel measurements, so that $\frac{1}{T} \sum_{t=0}^{T-1} \alpha_{imu}(t) \approx -R_0^T \gamma + \alpha_b$. Assuming the bias to be small (zero), this equation defines $R_0$ up to a rotation around gravity, which is arbitrary. Note that if $\alpha_b \neq 0$, the initial bias will affect the initial orientation estimate.} Measurements from a gyro can be similarly modeled as

\[
\omega_{imu}(t) = \omega(t) + \omega_b(t) + n_\omega(t)
\tag{3}
\]

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,

\[
\dot{\omega}_b(t) = v_b(t), \quad \dot{\alpha}_b(t) = v_\alpha(t)
\tag{4}
\]

for some unknown input $v_b, v_\alpha$. While it is safe to assume that $v_b, v_\alpha$ are small, they certainly are not (white, zero-mean and, most importantly) independent of the biases (which are now part of the state); hence it cannot be assumed that they are uninformative. Nevertheless, it is common to consider $v_b, v_\alpha$, 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 repercussion on the analysis of the observability and identifiability of the resulting model (Sect. 2).
1.3 Standard and reduced models

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

\[
\begin{align*}
\dot{T} &= V & T(0) &= 0 \\
\dot{R} &= R\bar{\omega} & R(0) &= R_0 \\
\dot{V} &= \alpha \\
\dot{w} &= w \\
\dot{\alpha} &= \xi \\
\dot{\omega}_b &= n_{\omega_b} \\
\dot{\alpha}_b &= n_{\alpha_b} \\
\gamma &= 0 \\
\omega_{imu}(t) &= \omega(t) + \omega_b(t) + n_{\omega}(t) \\
\alpha_{imu}(t) &= R^T(t)(\alpha(t) - \gamma) + \alpha_b(t) + n_{\alpha}(t)
\end{align*}
\]

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 \(\omega_{imu}\) and \(\alpha\) in terms of \(\alpha_{imu}\):

\[
\begin{align*}
\omega &= \omega_{imu} - \omega_b + \frac{n_{R\omega}}{-n_{\omega}} \\
\alpha &= R(\alpha_{imu} - \alpha_b) + \gamma + \frac{n_V}{-R_{\alpha\omega}}
\end{align*}
\]

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

\[
\begin{align*}
\dot{T} &= V & T(0) &= 0 \\
\dot{R} &= R(\bar{\omega}_{imu} - \bar{\omega}_b) + n_R & R(0) &= R_0 \\
\dot{V} &= R(\alpha_{imu} - \alpha_b) + \gamma + n_V \\
\dot{\omega}_b &= n_{\omega_b} \\
\dot{\alpha}_b &= n_{\alpha_b}.
\end{align*}
\]

Next we will consider augmenting the models above with measurement equations coming either from range or bearing measurements for a finite set \(N\) of isolated points with coordinates \(X^i \in \mathbb{R}^3, \ i = 1, \ldots, N\).

1.4 Bearing augmentation (vision)

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 \(\pi : \mathbb{R}^3 \to \mathbb{R}^2; X \mapsto [X_1/X_3, \ X_2/X_3]\) is a canonical central (perspective) projection, assuming that the camera is calibrated\(^3\) and that the spatial frame coincides with the body frame at time 0, we have

\[
y(t) = \frac{R_{12}^t(t)(X^i - T_{12}(t))}{R_{3}^t(t)(X^i - T_{3}(t))} = \pi(g^{-1}(t)X^i) + n^i(t), \quad t \geq 0.
\]

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 \(y(t) = \pi(X^i) + n^i(0)\), and therefore

\[
X^i = \bar{y}(0)Z^i + \bar{n}^i
\]

\(^3\)Intrinsic calibration parameters are known and compensated for.

\(^4\)The pose of the camera relative to the IMU is known and compensated for.
where \( \tilde{y} \) is the homogeneous coordinate of \( y \), \( \tilde{y} = [y^T \ 1]^T \), and \( \tilde{n}^i = [n_i^T(0) Z^i \ 0]^T \). Here \( Z^i \) is the (unknown, scalar) depth of the point at time \( t = 0 \). With an abuse of notation, we write the map that collectively projects all points to their corresponding locations on the image plane as:

\[
y(t) = \begin{bmatrix}
y_1 \\
y_2 \\
\vdots \\
y_N \\
\end{bmatrix}(t) = \begin{bmatrix}
\pi(R^T(X^1 - T)) \\
\pi(R^T(X^2 - T)) \\
\vdots \\
\pi(R^T(X^N - T)) \\
\end{bmatrix} + \begin{bmatrix}
n_1^i(t) \\
n_2^i(t) \\
\vdots \\
n_N^i(t) \\
\end{bmatrix}
\]

(10)

### 1.5 Alignment (calibration)

Consider the model \( \mathbf{4} \) with measurements \( y^i(t) \) can representing either the range of a number of sparse reflectors or the position on the image plane of a sparse collection of point features. In the former case, the range is measured in the reference frame of the radar, and therefore we have

\[
y^i(t) = \pi \left( g_{rb} g^{-1}(t) X_s^i \right) + n^i(t) \in \mathbb{R}
\]

where \( \pi(X) = \|X\| \) and \( g_{rb} \) is the transformation from the body frame to the radar. In the latter we have

\[
y^i(t) = \pi \left( g_{rb} g^{-1}(t) X_b^i \right) + n^i(t) \in \mathbb{R}^2
\]

(12)

where \( \pi(X) = [X_1/X_3, X_2/X_3]^T \), and \( g_{rb} \) is the transformation from the body frame to the camera. The "alignment" transformations \( g_{cb}, g_{rb} \) are typically not known and should be inferred. We can then, as done for the points \( X^i \), add them to the state with trivial dynamics \( g_{cb} = g_{rb} = 0 \).

### 1.6 Groups (occlusions)

It may convenient in some cases to represent the points \( X^i_s \) in the reference frame where they first appear, 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. Consider \( X^i(t_i) = \tilde{y}^i(t_i) Z^i(t_i) \), then \( y^i(t) \) 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 \( g(t_i) \):

\[
X^i_s = X^i_s(g_{t_i}, y_{t_i}, Z_{t_i}) = g_{t_i} \tilde{y}_{t_i} Z_{t_i}
\]

(13)

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 \( j = 1, \ldots, K(t) \) groups, each of which has a number \( i = 1, \ldots, N_j(t) \) 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

\[
\begin{cases}
\dot{\tilde{y}}_{t_i}^j = 0, & i = 1, \ldots, N(j) \\
\tilde{Z}_{t_i}^j = 0 \\
\dot{\hat{y}}_j = 0, & j = 1, \ldots, K(t).
\end{cases}
\]

(14)

For the case of range, this is not relevant as there is no reference frame that offers a preferential treatment of uncertainty.
1.7 Compact notation

If we call the “state” \( x = \{T, R, \alpha_0, \omega_b, X\} = \{x_1, x_2, x_3, x_4, x_5, x_6\} \) the “known input” \( u = \{\omega_{imu}, \alpha_{imu}\} = \{u_1, u_2\} \), the unknown input \( v = \{n_{\omega_b}, n_{\alpha_b}\} = \{v_1, v_2\} \), we can write the mechanization equations \( \text{(7)} \) as

\[
\dot{x} = f(x) + c(x)u + Dv
\]

where

\[
f(x) = \begin{bmatrix}
x_3 \\
x_2x_4 \\
x_2x_5 + \gamma \\
0 \\
0 \\
0
\end{bmatrix}, \quad c(x) = \begin{bmatrix}
0 \\
R \\
R \\
0 \\
0 \\
0
\end{bmatrix}, \quad D = \begin{bmatrix}
0 & 0 \\
0 & 0 \\
0 & 0 \\
0 & 0 \\
0 & 0
\end{bmatrix}
\]

and the measurement equation \( \text{(10)} \) as

\[
y = h(x) + n
\]

where

\[
h(x) = \begin{bmatrix}
\pi(x_2^T(x_6^2 - x_1)) \\
\vdots \\
\end{bmatrix}
\]

Putting together \( \text{(7)} - \text{(10)} \) we have a model of the form

\[
\begin{cases}
\dot{x} = f(x) + c(x)u + Dv \\
y = h(x) + n.
\end{cases}
\]

1.8 Definitions

We call \( y^t = \{y(\tau)\}_{\tau=0}^t \), a collection of output measurements, and \( x^t = \{x(\tau)\}_{\tau=0}^t \) a state trajectory. In the absence of unknown inputs, \( v = 0 \), given output measurements \( y^t \) and known inputs \( u^t \), we call

\[
\mathcal{I}(y^t|u^t; \tilde{x}_0) = \{\tilde{x}^t \mid y^t = h(\tilde{x}^t) \text{ s. t. } \tilde{x}(t) = f(\tilde{x}) + c(\tilde{x})u(t), \ \tilde{x}(0) = \tilde{x}_0 \ \forall \ t\}
\]

the indistinguishable set, or set of indistinguishable trajectories, for a given input \( u^t \). If the initial condition \( \tilde{x}_0 = x_0 \) equals the “true” one, the indistinguishable set contains at least one element, the “true” trajectory \( x^t \). However, if \( \tilde{x}_0 \neq x_0 \), the true trajectory may not even be part of this set.

If the indistinguishable set is a singleton (it contains only one element, \( \tilde{x}^t \), which is a function of the initial condition \( \tilde{x}_0 \)), we say that the model is observable up to the initial condition, or simply observable.\(^5\) If \( \{x^t\} \) is further independent of the initial condition, we say that the model is strongly observable: \( \mathcal{I}(y^t|u^t; \tilde{x}_0) = \{x^t\} \ \forall \ \tilde{x}_0, \ u^t \).

If the state includes unknown parameters with a trivial dynamic, and there is no unknown input, \( v = 0 \), then observability of the resulting model implies that the parameters are identifiable. That usually requires the input \( u^t \) to be sufficiently exciting (SE), in order to enable disambiguating the indistinguishable states\(^6\) that will disambiguate them. as the definition does not require that every input disambiguates any state trajectories.

\(^5\) We overload the notation to indicate with \( x_2 \) both the actual pose \( x_2 = R \in SO(3) \) and its local (exponential) coordinate \( x_2 = \Omega \mid \exp(\Omega) = R \), as it is clear from the context which we refer to. With this notation, we will assume that the state-space is linear.

\(^6\) We will assume that the solution of the differential equation \( \dot{x} = f(x) + c(x)u \) is unique and continuously dependent on the initial condition, so if we impose \( \tilde{x}_0 = x_0 \), then \( \tilde{x}^t = x^t \).

\(^7\) Sufficient excitation means that the input is generic, and does not lie on a thin set. That is, even if we could find a particular input \( u^t \) that yields indistinguishable states, there will be another input that is infinitesimally close to it.
In the presence of unknown inputs \( v \neq 0 \), consider the following definition

\[
\mathcal{I}_v(y^t|u^t; \tilde{x}_0) \doteq \{ \tilde{x}^t \mid \exists \, v^t \text{ s. t. } y^t = h(\tilde{x}^t), \ \dot{\tilde{x}}(t) = f(\tilde{x}) + c(\tilde{x})u(t) + Dv(t) \forall \, t; \ \tilde{x}(0) = \tilde{x}_0 \}
\]  

which is the set of unknown-input indistinguishable states. The model \( \{f, c, D\} \) is said to be unknown-input observable (up to initial conditions) if the unknown-input indistinguishable set is a singleton. If such a singleton is further independent of the initial conditions, the model is strongly observable. The two definitions coincide once the only admissible unknown input is \( v^t = 0 \) for all \( t \).

It is possible for a model to be observable (the indistinguishable set is a singleton), but not unknown-input observable (the unknown-input indistinguishable set is dense). In that case, the notion of sensitivity arises naturally, as one would want to measure the “size” of the unknown-input indistinguishable set as a function of the “size” of the unknown input. For instance, it is possible that if the set of unknown inputs is small in some sense, the resulting set of indistinguishable states is also small. If \( v \in V \) and for any \( \epsilon > 0 \) there exists a \( \delta > 0 \) such that \( \text{vol}(V) \leq \epsilon \) for some measure of volume implies \( \text{vol}(\mathcal{I}_v(y^t|u^t; \tilde{x}_0)) < \delta \) for any \( u^t, \tilde{x}_0 \), then we say that the model is bounded-unknown-input/unknown-output observable (up to the initial condition). If the latter volume is independent of \( \tilde{x}_0 \) we say that model is strongly bounded-unknown-input/unknown-output observable.

2 Analysis of Bearing-Augmented Navigation

2.1 Preliminary claims

**Lemma 1.** Given \( S \in \text{SO}(3) \) and \( \tilde{S} \in T_{\text{SO}(3)}(S) \), and \( a \in \mathbb{R} \), the matrix \( (aS + \tilde{S}) \) is nonsingular unless \( a = 0 \), in which case it has rank 2 or 0.

**Proof.** The tangent \( \tilde{S} \) has the form \( SM \), where \( M \) is some skew-symmetric matrix. As such, \( Mx \perp x \) for any \( x \in \mathbb{R}^3 \), so

\[
\|(aS + \tilde{S})x\|^2_2 = \|S(aI + M)x\|^2_2 = \|ax\|^2_2 + \|Mx\|^2_2.
\]

The above is zero only if \( ax = 0 \), so \( (aS + \tilde{S}) \) is nonsingular. For the remaining cases, observe that a \( 3 \times 3 \) skew-symmetric matrix has rank 2 or 0. \( \square \)

**Lemma 2.** Let \( (R(t), T(t)) \) and \( (\tilde{R}(t), \tilde{T}(t)) \) be differentiable trajectories in \( \text{SE}(3) \). For each time \( t' \in [0, T] \), there exists an open, full-measure subset \( \mathcal{A}_{t'} \subset \mathbb{R}^3 \) such that:

For any two static point-clouds \( \{X^i\}_{i=1}^N \subset \mathcal{A}_{t'} \) and \( \{\tilde{X}^i\}_{i=1}^N \subset \mathbb{R}^3 \) that satisfy

\[
\pi(R^{-1}(t)(X^i - T(t))) = \pi(\tilde{R}^{-1}(t)(\tilde{X}^i - \tilde{T}(t))) \quad \text{for all } i \text{ and } t
\]

there exist constant scalings \( \sigma_{i,t'} > 0 \) and a constant rotation \( S_{t'} = \tilde{R}(t')R^{-1}(t') \) such that

\[
\sigma_{i,t'}S_{t'}(X^i - T(t)) = (\tilde{X}^i - \tilde{T}(t)) + O((t-t')^2) \quad \text{for all } i \text{ and } t.
\]

Furthermore, if \( T(t') \neq 0 \), then \( \sigma_{i,t'} = \sigma_{i} \) for all \( i \).

**Proof.** Write \( S(t) = \tilde{R}(t)R^{-1}(t) \). Equality under the projection \( \pi \) implies that there exists a scaling \( \sigma_i(t) \) (possibly varying with \( X^i \) and \( t \)) such that

\[
\sigma_i S(X^i - T) = \tilde{X}^i - \tilde{T}.
\]

For a given time \( t' \), we wish to find a suitably large set \( \mathcal{A}_{t'} \) such that \( \dot{\sigma}_i(t') = \tilde{S}(t') = 0 \) and \( \sigma_i(t') \) is independent of \( X^i \), when \( X^i \in \mathcal{A}_{t'} \). Taking time derivatives,

\[
(\dot{\sigma}_i S + \sigma_i \dot{S})(X^i - T) - \sigma_i ST = -\dot{\tilde{T}}
\]
or, dividing by $\sigma_i$,
\[
\left(\frac{\partial}{\partial \sigma_i} S + \dot{S}\right)(X^i - T) - S \dot{T} = -\frac{1}{\sigma_i} \dot{T}.
\] (24)
Differentiating both sides with respect to $X^i$,
\[
\left(\frac{\partial}{\partial \sigma_i} S + \dot{S}\right)\delta X^i + \left(\frac{d}{dX^i} \left(\frac{\partial}{\partial \sigma_i}\right)\delta X^i\right) S(X^i - T) = -\left(\frac{d}{dX^i} \left(\frac{1}{\sigma_i}\right)\delta X^i\right) \dot{T}.
\] (25)
Observe that $\frac{d}{dX^i} \left(\frac{\partial}{\partial \sigma_i}\right)\delta X^i$ and $\frac{d}{dX^i} \left(\frac{1}{\sigma_i}\right)\delta X^i$ are scalars. By Lemma 1, the LHS has rank 2 or greater (as a linear map on $\delta X^i$), unless $\dot{\sigma}_i(t') = 0$. The RHS, however, has rank at most 1. Thus, (24) is invalid for almost all $X^i$, unless $\dot{\sigma}_i(t') = 0$ (two maps of different ranks can only agree on a submanifold). Plugging $\dot{\sigma}_i = 0$ into (25), we are left with
\[
\dot{S} \delta X^i = -\left(\frac{d}{dX^i} \left(\frac{1}{\sigma_i}\right)\delta X^i\right) \dot{T}.
\] (26)
Now, the LHS has rank 2 or 0, while the RHS has rank 1 or 0. Again, (24) is invalid for almost all $X^i$, unless $\dot{S}(t') = 0$. Let $A \subset \mathbb{R}^3$ be the open, full-measure subset (being the complement of two submanifolds) on which the latter must hold. If, in addition, $T(t') \neq 0$, then $\dot{T}(t') \neq 0$ and $\frac{d}{dX^i}(t') = 0$, we can finally write
\[
\sigma_i' S(t) (X^i - T) = \dot{X}^i - \dot{T} + O((t-t')^2).
\]
\[\Box\]

**Claim 1** (Indistinguishable Trajectories from Bearing Data Sequences). Let $g(t)$ and $\tilde{g}(t)$ be differentiable trajectories in $\text{SO}(3)$. There exists an open, full-measure subset $A \subset \mathbb{R}^3$ such that

Given two static, generic (non-coplanar) point clouds $\{X^i\}_{i=1}^N \subset A$ and $\{\tilde{X}^i\}_{i=1}^N \subset \mathbb{R}^3$, satisfying
\[
\pi(g^{-1}(t) X^i) = \pi(\tilde{g}^{-1}(t) \tilde{X}^i) \quad \text{for all } i \text{ and } t,
\]
there exist constant scalings $\sigma_i > 0$ and a constant transformation $\tilde{g} \in \text{SE}(3)$ such that
\[
\begin{cases}
\tilde{X}^i = \sigma_i (\tilde{g} X^i) \\
\tilde{g}(t) = \sigma_i (\tilde{g} g(t))
\end{cases} \quad \text{for all } i \text{ and } t. \quad (27)
\]

Furthermore, if $g(t)$ has a non-constant translational component, then $\sigma_i = \sigma$ for all $i$.

**Proof.** Write $g(t) = (R(t), T(t))$ and $\tilde{g}(t) = (\tilde{R}(t), \tilde{T}(t))$. Let $A = \{X \in \mathbb{R}^3 : X \in A \}$, with $A \subset \mathbb{R}^3$ defined as in Lemma 1. By Fubini’s theorem, this has full measure in $\mathbb{R}^3$. If $\{X^i\} \subset A$, then the conditions for Lemma 1 are satisfied for almost all $t$, and thus there exists constant (being stationary for almost all $t$) scalings $\sigma_i$ and rotation $S = \tilde{R}(t)R(t)^{-1} \in \text{SO}(3)$ such that $\tilde{X}^i = \sigma_i S(X^i - T_i) + \tilde{T}_i$.

Define $\tilde{g}(t) = (\sigma_i^{-1} \tilde{g}(t)) g(t)^{-1}$, and observe that
\[
\tilde{X}^i = \sigma_i S(X^i - T_i) + \tilde{T}_i = \sigma_i(\tilde{R}_i(\tilde{g}^{-1}(t) X^i) + \sigma_i^{-1} \tilde{T}_i) = \sigma_i((\sigma_i^{-1} \tilde{g}(t)) g(t)^{-1} X^i) = (\sigma_i(g(t)) X^i).
\]
If this affine relation holds for the generic set $\{X^i\}$, then $\tilde{g}(t)$ must be constant. Next,
\[
\sigma_i(\tilde{g} g(t)) = \sigma_i((\sigma_i^{-1} \tilde{g}(t)) g(t)^{-1} g(t)) = \sigma_i(\sigma_i^{-1} \tilde{g}(t)) = \tilde{g}(t).
\]
Finally, if $T(t') = 0$ for some $t'$, then $\sigma_i = \sigma_i(t') = \sigma(t') = \sigma$ for all $i$. $\Box$
Definition 1 (Sufficiently Exciting Motion). A trajectory $g(t)$ is **sufficiently exciting** relative to a point-cloud $\{X^i\}_{i=1}^N \subset \mathbb{R}^3$ if, for all $\{X^i\}_{i=1}^N \subset \mathbb{R}^3$ and $\tilde{g}(t)$ in SE(3),

\[
\pi(g(t)^{-1}(t)X^i) = \pi(\tilde{g}(t)^{-1}\tilde{X}^i) \quad \text{for all } i \text{ and } t \iff \begin{cases} 
\tilde{X}^i = \sigma(\tilde{g}X^i) \\
\tilde{g}(t) = \sigma(\tilde{g}g(t))
\end{cases} \text{for all } i \text{ and } t \text{ for some constant } \sigma > 0 \text{ and } \tilde{g} \in \text{SE}(3).
\] (28)

That is, if the projection map $\pi(g(t)X^i)$ defines $g(t)$ and $\{X^i\}$ up to a constant rotation and mapping.

Observe that the right-to-left implication is always true: if the RHS holds, then

\[
\pi(\tilde{g}(t)^{-1}\tilde{X}^i) = \pi((\sigma\tilde{g}g(t))^{-1}\sigma(gX^i))\pi(g(t)^{-1}\tilde{g}^{-1}\sigma^{-1}\sigma\tilde{g}X^i) = \pi(g(t)^{-1}X^i).
\]

We will see that the sufficient excitation condition is very easily satisfied.

Claim 2. Given trajectories $g(t)$ and $\tilde{g}(t)$ in SE(3) with non-constant translation, and a set $\{X^i\}_{i=1}^N$ of $N \geq 4$ points sampled i.i.d. from a non-singular distribution over $\mathbb{R}^3$, the trajectory $g(t)$ is a.s. sufficiently exciting relative to $\{X^i\}$.

Proof. Fix $g(t)$. By Claim 1 there exists a full-measure $\mathcal{A} \subset \mathbb{R}^3$ such that (28) holds for any static, generic point clouds $\{X^i\}_{i=1}^N \subset \mathcal{A}$ and $\{\tilde{X}^i\}_{i=1}^N \subset \mathbb{R}^3$. If $\{X^i\}$ is sampled i.i.d. from a non-singular distribution over $\mathbb{R}^3$, then $\{X^i\} \subset \mathcal{A}$ almost surely.

Equation (27) establishes the fact that the indistinguishable trajectories are an equivalence class parameterized by a group $\sigma(\tilde{g})$, called a gauge transformation. We now include a constant reference frame $g_a$. We then have the following claim.

Claim 3 (Indistinguishable Alignments). For a point cloud $\{X^i\}_{i=1}^{N(t)}$, $N(t) > 3$, in general position (non-coplanar), and sufficiently exciting motion,

\[
\pi(g_ag^{-1}(t)X^i) = \pi(\tilde{g}_a\tilde{g}^{-1}(t)\tilde{X}^i)
\] (29)

if and only if there exist constants $\sigma > 0$, $g_A$ and $g_B \in \text{SE}(3)$ such that

\[
\begin{cases}
\tilde{X}^i = \sigma(gBX^i) \\
\tilde{g}(t) = \sigma(gBg(t)g_A) \\
\tilde{g}_a = \sigma(g_ag_A)
\end{cases}
\] (30)

Proof. From Claim 1 we get constant $g_B \in \text{SE}(3)$ and $\sigma > 0$ such that $\tilde{X}^i = \sigma(gBX^i)$ and $\tilde{g}(t)\tilde{g}_a^{-1} = \sigma(gBg(t)g_a^{-1})$ (31)

Let $g_A = g_a^{-1}\sigma^{-1}(\tilde{g}_a)$. Then $\tilde{g}_a = \sigma(g_ag_A)$ and

$\tilde{g}(t) = \sigma(gBg(t)g_A)$.

We now include groups of points, each with its own reference frame.

Claim 4 (Indistinguishable Groups). For a number $i = 1, \ldots, K$ of groups each with a number $j = 1, \ldots, N_i \geq 3$ of points in general position (non-coplanar), and sufficiently exciting motion,

\[
\pi(g_ag^{-1}(t)g_jg_a^{-1}X^j) = \pi(\tilde{g}_a\tilde{g}^{-1}(t)\tilde{g}_j\tilde{g}_a^{-1}\tilde{X}^j)
\] (32)
Applying the definitions of $\bar{g}$.

Then, applying the definition of $\bar{g}$, we are interested under the usual assumptions it should converge to a state trajectory that is related to the true one by an

Finally, rearrange the definition of $\bar{g}$ from the measurements, we impose that the centroid of the points in that one group (the “reference group”) be one, which fixes

Proof. From Claim 1, we get constant $g_C \in SE(3)$ and $\sigma > 0$ such that

Define

Then, applying the definition of $\tilde{g}$ to (34),

Applying the definitions of $g_A$ and $g_B$ to (35),

Rearranging the definitions of $g_A$, $g_B$ and $\tilde{g}$,

Finally, rearrange the definition of $g_A$ to get

Eq. (33) describes the ambiguous state trajectories if only bearing measurement time series are given. In that case, there is no alignment to other sensor, so we can assume without loss of generality that $g_a = \text{Id}$ and so for $g_a$, which in turn implies $g_A = \text{Id}$. The resulting ambiguity is well-known and shows that scale $\sigma$ is constant but arbitrary, that the global reference frame is arbitrary (since $g_B$ is), and that the reference frame of each group is also arbitrary (since $\bar{g_i}$ is). To lock these ambiguities, we can fix three directions for each group (thus fixing $\tilde{g_i}$) and, in addition, for one of the groups fix the pose (thus fixing $g_B$); finally, we can impose that the centroid of the points in that one group (the “reference group”) be one, which fixes $\sigma$. Thus, an observer designed based on the standard model, where 3 directions within each group are saturated, and where the pose of one group is fixed, and the centroid of the group is at distance one, is observable, and under the usual assumptions it should converge to a state trajectory that is related to the true one by an arbitrary unknown scaling, and global reference frame.

Now, when inertial measurements are present, of all the possible trajectories that are indistinguishable from the measurements, we are interested only in those that are compatible with the dynamical model driven by IMU measurements. Since the fact that $X^j$ and $g_a$ are constant has already been enforced, the model will impose no constraints on $\tilde{X}^j$, $\tilde{g}_i$ and $\tilde{g}_a$. However, it will offer constraints on $\tilde{g}(t)$, that depends on the arbitrary constants $\sigma, g_A, g_B$. 


2.2 Indistinguishable trajectories in bearing augmentation

Definition 2. For an \( \mathbb{R}^3 \)-valued trajectory \( f : \mathbb{R} \to \mathbb{R}^3 \) and interval \( \mathcal{I} \subset \mathbb{R}^+ \), define

\[
m(f; \mathcal{I}) := \inf_{\| x \| = 1} \left( \sup_{t \in \mathcal{I}} | f(t) \cdot x | \right) = \inf_{\| x \| = 1} \left( \sup_{t \in \mathcal{I}} \| f(t) \times x \| \right),
\]

\[
M(f; \mathcal{I}) := \sup_{\| x \| = 1} \left( \sup_{t \in \mathcal{I}} | f(t) \cdot x | \right) = \sup_{t \in \mathcal{I}} \| f(t) \|, \quad \text{and} \quad \bar{m}(f; \mathcal{I}) := \sqrt{\max(0, 2m(f; \mathcal{I})^2 - M(f; \mathcal{I})^2)}.
\]

Observe that \( M(f; \mathcal{I}) \geq m(f; \mathcal{I}) \geq \bar{m}(f; \mathcal{I}) \), and that the inequalities are strict unless \( \{ \pm f(t) | t \in \mathcal{I} \} \) is dense on the sphere of radius \( M(f; \mathcal{I}) \). We use these “minimum-excitation” bounds in order to prove a partial converse of the Cauchy-Schwarz inequality:

Lemma 3. Let \( A = c_1 I + c_2 R \), for some rotation \( R \in \text{SO}(3) \) and scalars \( c_1 \) and \( c_2 \). Then, for any trajectory \( f : \mathbb{R}^+ \to \mathbb{R}^3 \) and set of times \( \mathcal{I} \subset \mathbb{R}^+ \),

\[
\sup_{t \in \mathcal{I}} \| Af(t) \| \geq \| A \| \bar{m}(f; \mathcal{I}).
\]

Proof. First, observe that \( A \) is normal:

\[
AA^T = (c_1 I + c_2 R)(c_1 I + c_2 R^T) = 2c_1 c_2 I + c_1 c_2 (R + R^T) = A^T A.
\]

Let \( \{ (\lambda_i, v_i) \}_{i=1}^3 \) be orthonormal eigenvalue/eigenvector pairs of \( A \), with \( \lambda_1 \geq \lambda_2 \geq \lambda_3 \).

\[
\| Af(t) \|^2 = \lambda_1^2 (v_1 \cdot f(t))^2 + \lambda_2^2 (v_2 \cdot f(t))^2 + \lambda_3^2 (v_3 \cdot f(t))^2 \\
\geq \lambda_1^2 ((v_1 \cdot f(t))^2 - (v_2 \cdot f(t))^2 - (v_3 \cdot f(t))^2) \\
= \| A \|^2 (2(v_1 \cdot f(t))^2 - \| f(t) \|^2).
\]

Taking the supremum over \( \mathcal{I} \),

\[
\sup_{t \in \mathcal{I}} \| Af(t) \|^2 \geq \| A \|^2 \sup_{t \in \mathcal{I}} (2(v_1 \cdot f(t))^2 - \| f(t) \|^2) \\
\geq \| A \|^2 (2 \sup_{t \in \mathcal{I}} (v_1 \cdot f(t))^2 - \sup_{t \in \mathcal{I}} \| f(t) \|^2) \\
\geq \| A \|^2 (2m(f; \mathcal{I})^2 - M(f; \mathcal{I})^2)
\]

Lemma 4. Let \( A = I - R \), for some rotation \( R \in \text{SO}(3) \). Then, for trajectory \( f : \mathbb{R}^+ \to \mathbb{R}^3 \) and \( \mathcal{I} \subset \mathbb{R}^+ \),

\[
\sup_{t \in \mathcal{I}} \| Af(t) \| \geq \| A \| m(f; \mathcal{I}).
\]

Proof. Let \( \{ (\lambda, v_1), (\lambda, v_2), (1, 0) \} \) be the orthonormal eigenvalue/eigenvector pairs of \( R \). Since \( R \) and \( I \) commute, \( \{ (\lambda - 1, v_1), (\lambda - 1, v_2), (0, u) \} \) are the eigenpairs of \( A \), and \( \| A \| = |\lambda - 1| = |\lambda - 1| \). Then,

\[
\| Af(t) \|^2 = |\lambda - 1|^2 (v_1 \cdot f(t))^2 + |\lambda - 1|^2 (v_2 \cdot f(t))^2 + 0 = \| A \|^2 (w \cdot f(t))^2,
\]

where

\[
w := \frac{(v_1 \cdot f(t))v_1 + (v_2 \cdot f(t))v_2}{\|(v_1 \cdot f(t))v_1 + (v_2 \cdot f(t))v_2\|} = \frac{(v_1 \cdot f(t))v_1 + (v_2 \cdot f(t))v_2}{\sqrt{(v_1 \cdot f(t))^2 + (v_2 \cdot f(t))^2}}.
\]

Taking the supremum over \( \mathcal{I} \),

\[
\sup_{t \in \mathcal{I}} \| Af(t) \|^2 = \| A \|^2 \sup_{t \in \mathcal{I}} \| w \cdot f(t) \|^2 \geq \| A \|^2 m(f; \mathcal{I})^2.
\]
Claim 5 (Indistinguishable Trajectories from IMU Data). Let \( g(t) = (R(t), T(t)) \in \text{SE}(3) \) be such that

\[
\begin{align*}
\dot{R} &= R(\tilde{\omega}_{\text{imu}} - \tilde{\omega}_b) \\
\dot{T} &= V \\
\dot{V} &= R(\alpha_{\text{imu}} - \alpha_b) + \gamma
\end{align*}
\] (36)

for some known constant \( \gamma \) 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 \( \|\dot{\omega}_b(t)\| \leq \epsilon \) 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 \( \|T_A\| \leq M_A \) and \( 0 < M_\sigma \leq |\sigma| \leq M_\gamma \). Then, under sufficient excitation conditions (described in this proof), \( \tilde{g}(t) \) satisfies (36) if and only if

\[
\begin{align*}
\|I - R_A\| &\leq \frac{2\epsilon}{m(\omega_{\text{imu}}: \mathbb{R}^+)} \\
|\sigma - 1| &\leq k_1 \epsilon + M_\sigma \|I - R_A\| \\
\|T_A\| &\leq \epsilon(k_2 + (2M_\sigma + 1)M_A) \\
\|(1 - R_A^T)\| &\leq \epsilon(k_3 + M_\sigma M_A + (|\sigma - 1| + \epsilon)M(\omega_{\text{imu}} - \omega_b: \mathbb{R}^3)) \|\gamma\| \\
&\leq m_\sigma m(\omega_{\text{imu}} - \omega_b: \mathbb{R}^3)
\end{align*}
\] (37) (38) (39) (40)

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

Proof.

(37) The ambiguous rotation \( \tilde{R} \) must satisfy \( \dot{\tilde{R}} = \tilde{R}(\tilde{\omega}_{\text{imu}} - \tilde{\omega}_b) \) for some \( \tilde{\omega}_b \):

\[
\dot{\tilde{R}} = \tilde{R} B R(\tilde{\omega}_{\text{imu}} - \tilde{\omega}_b) = \tilde{R} R_A^T(\tilde{\omega}_{\text{imu}} - \tilde{\omega}_b) R_A = \tilde{R} (R_A^T \tilde{\omega}_{\text{imu}} - \tilde{R}_A^T \tilde{\omega}_b)
\]

where the quantity in brackets is \( -\tilde{\omega}_b \), which defines

\[
\tilde{\omega}_b := R_A^T \omega_b + (I - R_A^T) \omega_{\text{imu}}.
\] (41)

Taking derivatives and rearranging,

\[
2\epsilon \geq \|\dot{\tilde{\omega}}_b - R_A^T \tilde{\omega}_b\| = \|(I - R_A^T)\tilde{\omega}_{\text{imu}}\|
\]

Since this is true for all \( t \in \mathbb{R} \), we can write

\[
2\epsilon \geq \sup_{t \in \mathbb{R}} \|(I - R_A^T)\tilde{\omega}_{\text{imu}}(t)\| \geq \|I - R_A^T\| m(\omega_{\text{imu}}: \mathbb{R}^+).
\]

This rearranges to give (37).

(38) The ambiguous translation \( \tilde{T} \) must satisfy the dynamics in (36):

\[
\dot{\tilde{T}} = \dot{V} = \tilde{R}(\alpha_{\text{imu}} - \tilde{\alpha}_b) + \gamma = R_B R_A (\alpha_{\text{imu}} - \tilde{\alpha}_b) + \gamma
\]

Alternatively, working with \( \tilde{T} = \sigma R_B (RT_A + T) \) and applying the dynamics to \( T \),

\[
\ddot{T} = \sigma R_B (\ddot{RT}_A + \ddot{T}) = \sigma R_B (\ddot{RT}_A + R(\alpha_{\text{imu}} - \alpha_b) + \gamma).
\]

\[\text{12}\]
Taking the difference between these two expressions,

$$0 = \sigma R_B \ddot{R} T_A + R_B R (R_A \dot{\alpha}_b - \sigma \alpha_b) + R_B R (\sigma \alpha_{imu} - R_A \alpha_{imu}) + (\sigma R_B - I) \gamma,$$

and multiplying by $R_T R_B^T$,

$$0 = \sigma (R_T \ddot{R}) T_A + (R_A \dot{\alpha}_b - \sigma \alpha_b) + (\sigma \alpha_{imu} - R_A \alpha_{imu}) + R_T^T (\sigma - R_B^T) \gamma$$

$$= \sigma ((\dot{\omega}_{imu} - \dot{\omega}_b)^2 + (\ddot{\omega}_{imu} - \ddot{\omega}_b)) T_A + (R_A \dot{\alpha}_b - \sigma \alpha_b) + (\sigma \alpha_{imu} - R_A \alpha_{imu}) + R_T^T (\sigma - R_B^T) \gamma.$$

Differentiating again,

$$0 = \sigma (R_T \ddot{R} + R_T \dddot{R}) T_A + \sigma (I - R_A) \sigma + (\sigma - 1) R_A \dot{\alpha}_{imu}$$

$$+ \dot{R}_T ((I - R_B^T) \sigma + (\sigma - 1) R_B) \gamma$$

$$+ (R_A \dot{\alpha}_b - \sigma \alpha_b).$$

As a sufficient excitation condition, assume that $\|\ddot{R}(t)\| \leq \epsilon$, $\|\dddot{R}(t)\| \leq \epsilon$, and $\|\ddot{R}(t) - \gamma\| \leq \epsilon$, for $t \in \mathcal{T}_1$. Under these constraints, $\|\dot{\alpha}_{imu}\| \leq 2\epsilon$, and (42) is bounded by $k_1 \epsilon$, where, e.g. $k_1 := 2M_{\sigma} M_A + (2M_{\sigma} + 1)(\|\gamma\| + 1)$. In that case,

$$k_1 \epsilon \geq \max_{t \in \mathcal{T}_1} \|((I - R_A) \sigma + (\sigma - 1) R_A) \dot{\alpha}_{imu}(t)\|$$

$$\geq |\sigma - 1| m(\dot{\alpha}_{imu}: \mathcal{T}_1) - M_{\sigma} \|I - R_A\|.$$

This rearranges to give (38).

Now, assume that $\|\dot{R}(t)\| \leq \epsilon$, $\|\ddot{R}(t)\| \leq \epsilon$, and $\|\ddot{R}(t) - \gamma\| \leq \epsilon$, for $t \in \mathcal{T}_2$. Under these constraints, $\|\dot{\alpha}_{imu}\| \leq 2\epsilon$, and (42) is bounded by $k_2 \epsilon$, where, e.g. $k_2 := (2M_{\sigma} + 1)(\|\gamma\| + 3)$. In that case,

$$k_2 \epsilon \geq \max_{t \in \mathcal{T}_2} \|\sigma ((\dot{\omega}_{imu} - \dot{\omega}_b)(\ddot{\omega}_{imu} - \ddot{\omega}_b) + (\dddot{\omega}_{imu} - \dddot{\omega}_b)) T_A\|$$

$$= \max_{t \in \mathcal{T}_2} \|\sigma ((R_T \ddot{R} - (R_T \dddot{R})^2) + (\dddot{\omega}_{imu} - \dddot{\omega}_b)) T_A\|$$

$$\geq m_{\sigma} \max_{t \in \mathcal{T}_2} \|\dot{\omega}_{imu}(t) \times T_A\| - (2M_{\sigma} + 1) M_A \epsilon$$

$$\geq m_{\sigma} \|T_A\| m(\dot{\omega}_{imu}: \mathcal{T}_2) - (2M_{\sigma} + 1) M_A \epsilon.$$

This rearranges to give (39).

Finally, assume that $\|\ddot{R}(t)\| \leq \epsilon$, $\|\dddot{R}(t)\| \leq \epsilon$, and $\|\ddot{R}(t) - \gamma\| \leq \epsilon$ for $t \in \mathcal{T}_3$. As before, $\|\dot{\alpha}_{imu}\| \leq 2\epsilon$. Then, (42) + (43) is bounded by $k_3 \epsilon$, where, e.g. $k_3 := 3M_{\sigma} + 3$. In that case,

$$k_3 \epsilon \geq \|\sigma (R_T \ddot{R} + R_T \dddot{R}) T_A + \dddot{R} ((I - R_B^T) \sigma + (\sigma - 1) R_B^T) \gamma\|$$

$$\geq \|\sigma \ddot{R} ((I - R_B^T) \gamma - M_{\sigma} M_A \epsilon - (\sigma - 1) \|\dddot{R}^T\| \|\gamma\|$$

$$\geq m_{\sigma} \|\dddot{R} (I - R_B^T) \gamma\| - M_{\sigma} M_A \epsilon - (\sigma - 1 + \epsilon) \|\dddot{R}^T\| \|\gamma\|$$

$$\geq m_{\sigma} m(\dddot{R}: \mathcal{T}_3)(1 - R_B^T) \gamma - \epsilon(k_3 + M_{\sigma} M_A) - (\sigma - 1 + \epsilon) M(\dddot{R}: \mathcal{T}_3) \|\gamma\|$$

This rearranges to give (40).
2.3 Gauge transformations

The set of indistinguishable trajectories \( \mathcal{I} \) is an equivalence class, and when the model is observable \textit{up to the initial condition}, it is parametrized by \( \hat{x}_0 \). Choosing the “true” initial condition \( \hat{x}_0 = x_0 \) produces an indistinguishable set consisting of the sole “true” trajectory, otherwise it is a singleton other than the true trajectory. In some cases, the initial condition corresponds to an arbitrary choice of reference frame, and therefore the equivalence class of indistinguishable trajectories are related by a gauge transformation (a change of coordinates). As the equivalence class can be represented by any element, enforcing a particular reference for the gauge transformation yields strong observability (although the singleton may not correspond to the true trajectory).

Formally, an arbitrary choice of initial condition is sufficient to fix the gauge reference. For instance, the set of indistinguishable trajectories in the limit where \( \epsilon \to 0 \) is parametrized by an arbitrary \( T_B \in \mathbb{R}^3 \) and \( \theta \in \mathbb{R} \),

\[
\begin{align*}
\hat{T} &= \exp(\hat{\gamma}\theta)T + T_B \\
\hat{R} &= \exp(\hat{\gamma}\theta)R \\
\hat{T}_{ti} &= \exp(\hat{\gamma}\theta)\hat{T}_{ti} + T_B \\
\hat{R}_{ti} &= \exp(\hat{\gamma}\theta)\hat{R}_{ti} \\
\hat{T}_{cb} &= T_{cb} \\
\hat{R}_{cb} &= R_{cb}
\end{align*}
\]

If we impose that \( T(0) = \hat{T}(0) = 0 \), then \( T_B = 0 \) is determined; similarly, if we impose the initial pose to be aligned with gravity (so gravity is in the form \([0 \ 0 \ \| \omega_b \|] \), 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 gauge reference can be associated to \textit{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

\[
\hat{X}^j = g_a \hat{g}_i^{-1} g_i a_i^{-1} X^j.
\]

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

The residual set of indistinguishable trajectories is parameterized by \textit{constants} \( \theta, T_B \), 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

\[
R_{ti} = \hat{R}_{ti} = \hat{R}(t_i) \quad \text{and} \quad T_{ti} = \hat{T}_{ti} = \hat{T}(t_i)
\]

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

\[
\hat{R}(t_i) = \exp(\hat{\gamma}\theta)\hat{R}(t_i)
\]

and therefore \( \theta = 0 \); similarly,

\[
T_B = (I - \exp(\hat{\gamma}\theta))T(t_i) = 0.
\]

Therefore, the gauge transformation is enforced explicitly at each instant of time, as each measurement provides a constraint on the states. This suggests the following modeling procedure in the design of a filter/observer for bearing-assisted navigation:

1. Set \( T(0) = 0 \) with zero model error covariance, and zero initial covariance.
2. Set \( R(0) = R_0 \) such that \([I_{2 \times 2}]R_0 \alpha_{imu} = 0\), with zero model error and non-zero initial covariance.
3. Fix gravity to \([0, 0, \|\gamma\|]^{T}\) from tabulates.

4. Initialize at rest, then perform some fast motions before groups of features are added.

5. Add \(K\) groups, each with \(2N + N\) states, plus their pose for each group but one.

6. Fix 2 directions per group (\([6]\) fixes all directions; this results in a non-zero mean component of the innovation, that in turn results in a small bias in all other states, that have to account for/absorb the mean)

7. Fix the pose of one group (remove its pose from the state)

8. Triage groups before adding them to the state.

After the Gauge Transformation has been fixed, the model is observable, and therefore a properly designed observer will converge to a solution \(\hat{x}\) that is related to the true one \(x\) as follows:

\[
\begin{align*}
\dot{X}^{\text{ref}} &= (1 + \bar{\sigma})\bar{R}_{cb} e^{\omega_{B}^{A} \bar{T}^{A}}(X^{\text{ref}} - T) + (1 + \bar{\sigma})\bar{R}_{cb} e^{\omega_{A}^{A} T_{B} + \bar{R}_{cb} T_{A} + \bar{T}_{cb}} \\
\dot{X}^{j} &= (1 + \bar{\sigma})\bar{R}_{cb} \bar{R}_{i} \bar{R}_{cb}(X^{j} - T) + (1 + \bar{\sigma})(\bar{R}_{cb} \bar{R}_{i} \bar{T}_{i} + \bar{R}_{cb} \bar{T}_{i} + \bar{T}_{cb}) \\
\dot{T} &= e^{\gamma_{\theta}^{[\bar{T}]}} + T_{B}(1 + \bar{\sigma}) + \omega_{B} e^{\gamma_{\theta}^{[\bar{T}]}} + e^{\omega_{B}^{A} e^{\gamma_{\theta}^{[\bar{T}]} T_{A}}(1 + \bar{\sigma})} \\
\bar{R} &= e^{\omega_{B}^{A} e^{\gamma_{\theta}^{[\bar{T}]} R^{A}}} \\
\bar{T}_{i} &= e^{\gamma_{\theta}^{[\bar{T}]} \bar{T}_{i}} + T_{B}(1 + \bar{\sigma}) + \omega_{B} e^{\gamma_{\theta}^{[\bar{T}]} \bar{T}_{i}} + e^{\omega_{B}^{A} e^{\gamma_{\theta}^{[\bar{T}]} T_{A}}(1 + \bar{\sigma})} \\
\bar{R}_{i} &= e^{\omega_{B}^{A} e^{\gamma_{\theta}^{[\bar{T]]} R^{A}}} \\
\bar{T}_{cb} &= T_{cb} + \bar{\sigma} T_{cb} + R_{cb} T_{A}(1 + \bar{\sigma}) \\
\bar{R}_{cb} &= R_{cb} \exp(\omega_{A})
\end{align*}
\]

where

\[
\begin{align*}
\|\omega_{A}\| &\leq 2\|I - R_{A}\| < \frac{2\epsilon}{m(\dot{\omega}_{imu} : \bar{I}^{+})} \\
|\sigma - 1| &\leq \frac{k_{1}\epsilon + M_{\sigma} \|I - R_{A}\|}{m(\dot{\omega}_{imu} : \bar{I}_{1})} \\
\|T_{A}\| &\leq \frac{\epsilon(k_{2} + (2M_{\sigma} + 1)M_{\sigma})}{m_{\sigma} m(\dot{\omega}_{imu} : \bar{I}_{2})} \\
\|\omega_{B}\| &\leq 2\|I - R_{B}^{A}\|\gamma\| + \left(\sigma - 1\right) |\omega_{imu} - \omega_{b} : \bar{I}_{3}| \|\gamma\|
\end{align*}
\]

and arbitrary \(\theta, T_{B}\) and suitable constants \(k_{i}\). 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 \([11]\), where the covariance of the translation state \(T_{B}\) and the rotation about gravity \(\theta\) 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. \([40]\).

\section{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. That is, the set of indistinguishable states is not a singleton, as one would be led to believe if assuming...
that bias rates are “white noise” that is independent of all other states (including the biases themselves). While the treatment of bias rates as (possibly small but otherwise) unknown inputs, as opposed to noise, is a matter of modeling, empirical evidence dating back to [6] shows that indeed convergence is not to the “ground truth” but to a steady-state error that can be bounded as a function of the input characteristics.

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 of certain matrices or distributions, is that we characterize the indistinguishable set explicitly. In addition to being opaque, rank conditions are “fragile” in the sense that the model can be nominally observable, and yet the condition number of the observability matrix (or co-distribution) 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.

References

[1] G. Basile and G. Marro. On the observability of linear, time-invariant systems with unknown inputs. Journal of Optimization theory and applications, 3(6):410–415, 1969.

[2] 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.

[3] 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, pages 2602–2607. IEEE, 2010.

[4] F Hamano and G Basile. Unknown-input present-state observability of discrete-time linear systems. Journal of Optimization Theory and Applications, 40(2):293–307, 1983.

[5] H. Hammouri and Z. Tmar. Unknown input observer for state affine systems: A necessary and sufficient condition. Automatica, 46(2):271–278, 2010.

[6] E. Jones and S. Soatto. Visual-inertial navigation, localization and mapping: A scalable real-time large-scale approach. Intl. J. of Robotics Res., Apr. 2011.

[7] E. S. Jones, A. Vedaldi, and S. Soatto. Inertial structure from motion and autocalibration. In Workshop on Dynamical Vision, October 2007.

[8] J. Kelly and G. Sukhatme. Fast Relative Pose Calibration for Visual and Inertial Sensors. In Experimental Robotics, pages 515–524, 2009.

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

[10] Agostino Martinelli et al. Visual-inertial structure from motion: observability and resolvability. Foundations and Trends® in Computer Graphics and Vision, 1(1), 2014.

[11] 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, pages 3565–3572. IEEE, 2007.
[12] S. Soatto. 3-d structure from visual motion: modeling, representation and observability. *Automatica*, 33:1287–1312, 1997.

[13] Stefano Soatto. Observability/identifiability of rigid motion under perspective projection. In *Decision and Control, 1994.*, Proceedings of the 33rd IEEE Conference on, volume 4, pages 3235–3240. IEEE, 1994.