

Observability, Identifiability, Sensitivity and Model Reduction For Vision-Assisted Inertial Navigation

Stefano Soatto

Technical Report UCLA CSD13022
Aug. 20, 2013; Revised Nov. 12, 2013

Abstract

Observability analysis of vision-aided navigation typically assumes input noise to be absent. This leads to conclusions that are not valid when such inputs are possibly small, but not uninformative. We frame the analysis of observability of three-dimensional trajectory (position and orientation) given inertial and visual sensors as the determination of the volume of the set of indistinguishable state trajectories, and show that this volume is not zero, unlike suggested by previous observability analysis. Moreover, the indistinguishable set may even fail to contain the “true” state trajectory. More constructively, we provide bounds on the volume of the indistinguishable set as a function of the characteristics of the unknown input. This yields conditions under which the state trajectory can be estimated within a given accuracy, despite being unobservable. Our analysis also includes alignment (relative position and orientation between visual and inertial sensors) and biases.

1 Introduction

This manuscript describes a novel approach to the analysis of observability/identifiability of visually-assisted navigation, whereby inertial sensors (accelerometers and gyroscopes) 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.*, accelerometer and gyro biases), and treated as states in the model, driven by some kind of uninformative input (“noise”).

Observability of the model is a necessary condition for *any* filter/observer to operate, hence the problem has received some attention in the literature [5, 21, 13, 2, 14, 19, 15, 18, 27, 10, 16, 11, 7, 20, 4]. Unfortunately, *existing analysis* of visual-inertial observability is of limited value because it *relies on a critical assumption that is not satisfied in practice and that, if removed, renders the conclusions invalid*.

Observability is a structural property of a model, which is not affected by noise. As a result, noise is usually set to zero for the purpose of observability analysis, which is reasonable because, by assumption, noise is uninformative. However, the driving input to the random walk model of accelerometer and gyro bias is typically small but *not* independent of the state. Far from being uninformative, it is strongly correlated with it, as it is its temporal derivative. While modeling it as noise is acceptable for the purpose of inference, it is not for the purpose of analysis, where it should be treated as *unknown input*, rather than “noise.” As such, it should be included in the observability/identifiability analysis. In this manuscript, we offer the following contributions to the analysis of observability/identifiability of visual-inertial navigation

1. We show that while (a prototypical model of) assisted navigation and auto-calibration is *observable* in the absence of unknown inputs, it is *not* observable when unknown inputs are taken into account.

2. We therefore 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.
3. We explicitly characterize this indistinguishable set and show that it may not contain the “true” state trajectory.
4. Finally, we provide bounds on the volume of the indistinguishable set as a function of the characteristics of the unknown input.

Unlike other studies of observability, we bypass rank conditions or other algebraic tests that do not offer insight on the structure of the indistinguishable set. Instead, we characterize observability directly in terms of the volume indistinguishable sets.

1.1 Notation

A reference frame is represented by an orthogonal 3×3 positive-determinant (rotation) matrix $R \in SO(3) \doteq \{R \in \mathbb{R}^{3 \times 3} \mid R^T R = RR^T = I, \det(R) = +1\}$ and a translation vector $T \in \mathbb{R}^3$. They are collectively indicated by $g = (R, T) \in SE(3)$. 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 $\hat{V} = (\hat{\omega}, v) \in se(3)$ the (generalized) velocity or “twist”, where $\hat{\omega}$ is a skew-symmetric matrix $\hat{\omega} \in so(3) \doteq \{S \in \mathbb{R}^{3 \times 3} \mid S^T = -S\}$ corresponding to the cross product with the vector $\omega \in \mathbb{R}^3$, so that $\hat{\omega}v = \omega \times v$ 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, $\bar{X}_b = G_{ba}\bar{X}_a$ where $\bar{X}^T = [X^T \ 1]$ and

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

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_{sb}^b , defined by $\dot{g}_{sb}(t) = g_{sb}(t)\hat{V}_{sb}^b(t)$, which can be broken down into the rotational and translational components $\dot{R}_{sb}(t) = R_{sb}(t)\hat{\omega}_{sb}^b(t)$ and $\dot{T}_{sb}(t) = R_{sb}(t)v_{sb}^b(t)$. An ideal gyrometer (gyro) would measure $\omega_{imu} = \omega_{sb}^b$. The translational component of body velocity, v_{sb}^b , can be obtained from the last column of the matrix $\frac{d}{dt}\hat{V}_{sb}^b(t)$. That is, $\dot{v}_{sb}^b = \dot{R}_{sb}^T \dot{T}_{sb} + R_{sb}^T \ddot{T}_{sb} = -\hat{\omega}_{sb}^b v_{sb}^b + R_{sb}^T \ddot{T}_{sb} \doteq -\hat{\omega}_{sb}^b v_{sb}^b + \alpha_{sb}^b$, 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 $\alpha_{imu} = R_{sb}^T(t)(\alpha_{sb}(t) - \gamma)$.

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 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.¹ 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

$$\begin{cases} \dot{T} = V \\ \dot{R} = R\hat{\omega} \\ \dot{V} = \alpha \\ \dot{\omega} = w \\ \dot{\alpha} = \xi \end{cases} \quad (2)$$

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

$$\alpha_{imu}(t) = R^T(t)(\alpha(t) - \gamma) + \underbrace{\alpha_b(t)}_{\text{bias}} + \underbrace{n_\alpha(t)}_{\text{noise}} \quad (3)$$

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.² Measurements from a gyro can be similarly modeled as

$$\omega_{imu}(t) = \omega(t) + \underbrace{\omega_b(t)}_{\text{bias}} + \underbrace{n_\omega(t)}_{\text{noise}} \quad (4)$$

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 instance

$$\dot{\omega}_b(t) = v_b(t), \quad \dot{\alpha}_b(t) = v_\alpha(t) \quad (5)$$

for some unknown input v_b, v_α . While it is safe to assume that v_b, v_α are *small*, they certainly are not (white, zero-mean and, most importantly) uninformative. Nevertheless, it is common to consider v_b, v_α , 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 repercussion on the analysis of the observability and identifiability of the resulting model (Sect. 2).

¹In 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.

²The 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 $\frac{1}{T} \sum_{t=0}^T \alpha_{imu}(t) \simeq -R_0^T \gamma + \alpha_b$. 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.

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

$$\left\{ \begin{array}{l} \dot{T} = V \quad T(0) = 0 \\ \dot{R} = R\hat{\omega} \quad R(0) = R_0 \\ \dot{V} = \alpha \\ \dot{\omega} = w \\ \dot{\alpha} = \xi \\ \dot{\omega}_b = n_{\omega_b} \\ \dot{\alpha}_b = n_{\alpha_b} \\ \dot{\gamma} = 0 \\ \omega_{imu}(t) = \omega(t) + \omega_b(t) + n_{\omega}(t) \\ \alpha_{imu}(t) = R^T(t)(\alpha(t) - \gamma) + \alpha_b(t) + n_{\alpha}(t) \end{array} \right. \quad (6)$$

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} :

$$\boxed{\omega = \omega_{imu} - \omega_b + \underbrace{n_R}_{-n_{\omega}}} \quad \boxed{\alpha = R(\alpha_{imu} - \alpha_b) + \gamma + \underbrace{n_V}_{-Rn_{\alpha}}} \quad (7)$$

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

$$\boxed{\left\{ \begin{array}{l} \dot{T} = V \quad T(0) = 0 \\ \dot{R} = R(\hat{\omega}_{imu} - \hat{\omega}_b) + n_R \quad R(0) = R_0 \\ \dot{V} = R(\alpha_{imu} - \alpha_b) + \gamma + n_V \\ \dot{\omega}_b = n_{\omega_b} \\ \dot{\alpha}_b = n_{\alpha_b}. \end{array} \right.} \quad (8)$$

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 $X^i \in \mathbb{R}^3$, $i = 1, \dots, N$.

1.4 Bearing augmentation (vision)

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 $\pi : \mathbb{R}^3 \rightarrow \mathbb{R}^2; X \mapsto [X_1/X_3, X_2/X_3]$ is a canonical central (perspective) projection, assuming that the camera is *calibrated*,³ *aligned*,⁴ and that the spatial frame coincides with the body frame at time 0, we have

$$\boxed{y^i(t) = \frac{R_{1:2}^T(t)(X^i - T_{1:2}(t))}{R_3^T(t)(X^i - T_3(t))} \doteq \pi(g^{-1}(t)X^i) + n^i(t), \quad t \geq 0.} \quad (9)$$

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

$$\boxed{X^i = \bar{y}^i(0)Z^i + \tilde{n}^i} \quad (10)$$

³Intrinsic calibration parameters are known and compensated for.

⁴The pose of the camera relative to the IMU is known and compensated for.

where \bar{y} is the homogeneous coordinate of y , $\bar{y} = [y^T \ 1]^T$, and $\tilde{n}^i = [n^{i^T}(0)Z^i \ 0]^T$. 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:

$$y(t) \doteq \begin{bmatrix} y^1 \\ y^2 \\ \vdots \\ y^N \end{bmatrix} (t) = \begin{bmatrix} \pi(R^T(X^1 - T)) \\ \pi(R^T(X^2 - T)) \\ \vdots \\ \pi(R^T(X^N - T)) \end{bmatrix} + \begin{bmatrix} n^1(t) \\ n^2(t) \\ \vdots \\ n^N(t) \end{bmatrix} \quad (11)$$

1.5 Alignment (calibration)

When the camera and the body frame are not coincident, we have

$$y^i(t) = \pi(g_{cb}g^{-1}(t)X_s^i) + n^i(t) \in \mathbb{R}^2 \quad (12)$$

where $\pi(X) = [X_1/X_3, X_2/X_3]^T$, and g_{cb} 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 $\dot{g}_{cb} = \dot{g}_{rb} = 0$.

1.6 Groups (occlusions)

It may convenient in some cases to represent the points X_s^i 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 $X^i(t_i) = \bar{y}^i(t_i)Z^i(t_i)$, 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 $g(t_i)$:

$$X_s^i = X_s^i(g_{t_i}, y_{t_i}, Z_{t_i}) = g_{t_i}\bar{y}_{t_i}Z_{t_i} \quad (13)$$

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 $j = 1, \dots, K(t)$ 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_s^i then evolves according to

$$\begin{cases} \dot{y}_{t_i}^i = 0, & i = 1, \dots, N(j) \\ \dot{Z}_{t_i}^i = 0 \\ \dot{g}_j = 0, & j = 1, \dots, K(t). \end{cases} \quad (14)$$

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” $x = \{T, R, V, \alpha_b, \omega_b, X\} = \{x_1, x_2, x_3, x_4, x_5, x_6\}$ the “known input” $u = \{\hat{\omega}_{imu}, \alpha_{imu}\} = \{u_1, u_2\}$, the *unknown input* $v = \{n_{\omega_b}, n_{\alpha_b}\} = \{v_1, v_2\}$, we can write the mechanization equations (8) as

$$\dot{x} = f(x) + c(x)u + Dv \quad (15)$$

where

$$f(x) \doteq \begin{bmatrix} x_3 \\ -x_2 x_4 \\ -x_2 x_5 + \gamma \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad c(x) \doteq \begin{bmatrix} 0 \\ R \\ R \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad D \doteq \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ I & 0 \\ 0 & I \\ 0 & 0 \end{bmatrix} \quad (16)$$

and the measurement equation (11) as

$$y = h(x) + n \quad (17)$$

where

$$h(x) \doteq \begin{bmatrix} \vdots \\ \pi(x'_2(x'_6 - x_1)) \\ \vdots \end{bmatrix} \quad (18)$$

Putting together (8)-(11) we have a model of the form

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

(19)

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*. Given output measurements y^t and known inputs u^t , we call

$$\mathcal{I}(y^t | u^t; \tilde{x}_0) \doteq \{\tilde{x}^t \mid y^t = h(\tilde{x}^t) \text{ s. t. } \dot{\tilde{x}}(t) = f(\tilde{x}) + c(\tilde{x})u(t), \tilde{x}(0) = \tilde{x}_0 \forall t\} \quad (20)$$

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 $\{\tilde{x}^t\}$ 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 as the definition does not require that every input disambiguates states.

In the presence of *unknown inputs* $v \neq 0$, consider the following definition

$$\mathcal{I}_v(y^t | u^t; \tilde{x}_0) \doteq \{\tilde{x}^t \mid \exists v^t \text{ s. t. } y^t = h(\tilde{x}^t), \dot{\tilde{x}}(t) = f(\tilde{x}) + c(\tilde{x})u(t) + Dv(t) \forall t; \tilde{x}(0) = \tilde{x}_0\} \quad (21)$$

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 .

⁵We 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 $\tilde{x}^t = x^t$.

⁶Sufficient 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 that will disambiguate them.

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/bounded-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/bounded-output observable.

The set of indistinguishable trajectories \mathcal{I} is an equivalence class, and when the model is observable *up to the initial condition*, it is parametrized by \tilde{x}_0 . Choosing the “true” initial condition $\tilde{x}_0 = x_0$ 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 trajectory 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).

Related work

Unknown-input observability of linear time-invariant systems has been addressed in [1, 8], for affine systems [9], and non-linear systems in [6, 17, 3]. 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 do not fit in the classes treated in the literature above, which motivates our analysis.

2 Analysis of Bearing-Augmented Navigation

2.1 Preliminary claims

Claim 1. Let $X^i(t) \in \mathbb{R}^3, i = 1, \dots, N(t); t \in \mathbb{Z}$ and $\tilde{X}^i(t) \neq X^i(t)$. Then $\pi(X^i(t)) = \pi(\tilde{X}^i(t))$ if and only if $\tilde{X}^i(t) = \sigma(t)X^i(t) \forall i, t$, where $\sigma(t) > 0$ is an arbitrary positive scalar-valued function of t (and i).

This follows directly from the definition of the projection map π .

Claim 2 (Indistinguishable Trajectories from Bearing Data Sequences). Let $X^i \in \mathbb{R}^3, i = 1, \dots, N(t) > 3$ non-coplanar points, $g(t) \in SE(3), t \in \mathbb{Z}$ and $\tilde{g}(t) \neq g(t), \tilde{X}^i \neq X^i$, then $\pi(g^{-1}(t)X^i) = \pi(\tilde{g}^{-1}(t)\tilde{X}^i) \forall i, t$ if and only if

$$\begin{cases} \tilde{X}^i = \sigma(\bar{g}X^i) \\ \tilde{g}(t) = \sigma(\bar{g}g(t)) \end{cases} \quad (22)$$

for an arbitrary constant $\sigma > 0$ and an arbitrary constant $\bar{g} \in SE(3)$.

We use the notation $\sigma(g)$ to denote a scaled rigid motion: If $g \in SE(3)$ is represented by (R, T) , then $\sigma(g)$ is represented by $(R, \sigma T)$ for a scalar $\sigma \in \mathbb{R}$. Note that if $g_1, g_2 \in SE(3), \sigma(g_1g_2) = \sigma(g_1)\sigma(g_2)$. If $X \in \mathbb{R}^3$, then $\sigma(gX) = \sigma(g)\sigma X$. To prove the claim, we write $\tilde{g}(t) \in SE(3)$ and $\tilde{X}^i \in \mathbb{R}^3$ as follows:

$$\tilde{g}(t) \doteq \bar{g}(t)g(t) \quad (23)$$

and

$$\tilde{X}^i \doteq \sigma_i(g_iX^i) \quad (24)$$

for some $\bar{g}(t), g_i \in SE(3), \sigma_i > 0$, that can be chosen so that the resulting $\tilde{g}(t), \tilde{X}^i$ are arbitrary;⁷ therefore these are *all* possible $\tilde{g}(t), \tilde{X}^i$ (and more). Among those, we select *only* those that satisfy the constraint

$$\pi(g^{-1}(t)X^i) = \pi(\tilde{g}^{-1}(t)\tilde{X}^i) \quad \forall i, t \quad (25)$$

and show that this implies that $g_i = \bar{g}(t) \quad \forall i, t$ and $\sigma_i = \sigma$, a constant, for all i .

Proof. From Claim 1 we have that

$$\tilde{g}^{-1}(t)\tilde{X}^i = \sigma(t)(g^{-1}(t)X^i) \quad (26)$$

for any positive-valued scalar function $\sigma(t)$. Substituting the expressions in (23)-(24) we have, omitting time indices,

$$g^{-1}\bar{g}^{-1}\sigma_i(g_i X^i) = \sigma(g^{-1}X^i) \quad (27)$$

or, in coordinates,

$$\underbrace{\sigma_i R^T \bar{R}^T R_i}_{\sigma_i R^T \bar{R}^T R_i} X^i + R^T (\bar{R}^T (\sigma_i T_i - \bar{T}) - T) = \underbrace{\sigma(t) R^T}_{\sigma(t) R^T} (X^i - T) \quad (28)$$

which is affine in X^i . Therefore, under general position conditions ($N(t) > 3$ non-coplanar points) the linear terms of both sides (bracketed), and the offsets on both sides, have to be equal. Equivalently, we can take the derivative of the expression above with respect to X^i (which yields the linear terms):

$$\bar{R}^T R_i \sigma_i = \sigma(t) \quad (29)$$

which is satisfied for all t, i if and only if $\bar{R}(t) = R_i$ and $\sigma_i = \sigma(t)$. Therefore, we have

$$\bar{R} = \bar{R}(t) = R_i \quad \sigma = \sigma(t) = \sigma_i \quad \forall i, t. \quad (30)$$

Now letting $X^i \neq X^j$ and taking the difference of their two corresponding equations (27) (which yields the offsets), we have that

$$\sigma \bar{R}^T T_i - \bar{R}^T \bar{T}(t) - T(t) + \sigma T(t) = 0 \quad (31)$$

which yields

$$\bar{T}(t) = \sigma T_i + (\sigma - 1) \bar{R} T(t) \quad (32)$$

and (taking the difference between two equations for $T_i \neq T_j$) also $\sigma \bar{R}^T (T_i - T_j) = 0$ which implies that $T_i = T_j \quad \forall i, j$. Substituting into the expression for \tilde{g} in (23) we obtain, in coordinates

$$\begin{bmatrix} \bar{R} & \sigma T_i + (\sigma - 1) \bar{R} T(t) \\ 0 & 1 \end{bmatrix} \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} \bar{R} & \sigma T_i \\ 0 & 1 \end{bmatrix} \begin{bmatrix} R & \sigma T \\ 0 & 1 \end{bmatrix} \quad (33)$$

or more compactly $g(t) = \sigma(\bar{g}g(t))$, which shows that

$$\bar{T}(t) = \sigma T_i, \quad \forall i, t \quad (34)$$

which proves the claim. \square

Equation (22) establishes the fact that the indistinguishable trajectories are an equivalence class parameterized by a group $\sigma(\bar{g})$, called a *gauge transformation*. We now include a constant reference frame g_a . We then have the following claim.

⁷For an arbitrary $\tilde{g}(t)$ and given $g(t)$, $\bar{g}(t)$ is given by $\bar{g}(t) = \tilde{g}(t)g^{-1}(t)$. Similarly, for an arbitrary \tilde{X}^i and a given X^i , $g_i = (R_i, T_i)$ can be chosen so that $T_i = R_i^T(\tilde{X}^i/\sigma_i - X^i)$ for an arbitrary R_i, σ_i .

Claim 3 (Indistinguishable Alignments). *For a number $i = 1, \dots, N(t) > 3$ points in general position (non-coplanar), and sufficiently exciting (non-constant) motion,*

$$\pi(g_a g^{-1}(t) X^i) = \pi(\tilde{g}_a \tilde{g}^{-1}(t) \tilde{X}^i) \text{ if and only if} \quad (35)$$

$$\begin{cases} \tilde{X}^i = \sigma(g_B X^i) \\ \tilde{g}(t) = \sigma(g_B g(t) g_A) \\ \tilde{g}_a = \sigma(g_a g_A) \end{cases} \quad (36)$$

for arbitrary constants $\sigma > 0, g_A, g_B \in SE(3)$

Proof. Letting $g'(t) \doteq g(t)g_a^{-1} = g(t)g_Ag_A^{-1}g_a^{-1}$, from Claim 2 we obtain $\tilde{X}^i = \sigma(g_B X^i)$ and

$$\tilde{g}'(t) = \underbrace{\sigma(g_B g(t) g_A)}_{\tilde{g}} \underbrace{g_A^{-1} g_A^{-1}}_{\tilde{g}_a} \quad (37)$$

Breaking down the ambiguity in g' into the ambiguity in its components gg_a we obtain

$$\tilde{g}(t)\tilde{g}_a^{-1} = \sigma(g_B g(t) g_A) \sigma(g_A^{-1} g_a^{-1}) \quad (38)$$

from which the result follows by the general-position and sufficient excitation conditions, equating the time-varying component of the left-hand side with the time-varying component of the right-hand side, and so for the constant components. \square

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

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

$$\pi(g_a g^{-1}(t) g_i g_a^{-1} X^j) = \pi(\tilde{g}_a \tilde{g}^{-1}(t) \tilde{g}_i \tilde{g}_a^{-1} \tilde{X}^j) \text{ if and only if} \quad (39)$$

$$\boxed{\begin{cases} \tilde{X}^j = \sigma(g_a \bar{g}_i^{-1} g_i g_a^{-1} X^j) \\ \tilde{g}(t) = \sigma(g_B g(t) g_A) \\ \tilde{g}_i = \sigma(g_B \bar{g}_i g_A) \\ \tilde{g}_a = \sigma(g_a g_A) \end{cases}} \quad (40)$$

for arbitrary constants $\sigma > 0, g_A, g_B, \bar{g}_i \in SE(3)$

Proof. Letting $X' \doteq g_i g_a^{-1} X$, from Claim 3 we obtain $\tilde{X}' = \sigma(g_B X')$, from which

$$\tilde{X}' = \sigma(g_B X') = \sigma(g_B g_i g_a^{-1} X) = \sigma(g_B \bar{g}_i g_A g_A^{-1} g_a^{-1} g_a \bar{g}_i^{-1} g_i g_a^{-1} X) \quad (41)$$

$$\tilde{g}_i \tilde{g}_a^{-1} \tilde{X} = \sigma(g_B \bar{g}_i g_A) \sigma(g_A^{-1} g_a^{-1}) \sigma(g_a \bar{g}_i^{-1} g_i g_a^{-1} X) \quad (42)$$

where we have omitted the point indices j for simplicity. Note that \tilde{g}_i is arbitrary (because so is \bar{g}_i) and \tilde{g}_a is arbitrary (because so is g_A). Therefore, equating the constant terms (independent of i, j) of both sides, the offsets (the components independent of X), and the components dependent on j , we obtain the desired result following the general-position and sufficient excitation conditions. \square

Eq. (40) 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 $g_a = Id$ and so for \tilde{g}_a , which in turn implies $g_A = Id$. The resulting ambiguity is well-known [22] and shows that scale σ 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 \bar{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 σ . 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 σ, g_A, g_B .

2.2 Indistinguishable trajectories in bearing augmentation

Claim 5 (Indistinguishable Trajectories from IMU Data). *Let $g(t) = (R(t), T(t)) \in SE(3)$ be such that*

$$\begin{cases} \dot{R} = R(\widehat{\omega}_{imu} - \widehat{\omega}_b) \\ \dot{T} = V \\ \dot{V} = R(\alpha_{imu} - \alpha_b) + \gamma \end{cases} \quad (43)$$

for some known constant γ and sufficiently exciting functions $\alpha_{imu}(t), \omega_{imu}(t)$ and for some unknown functions $\alpha_b(t), \omega_b(t)$ that are constrained to have $\|\dot{\alpha}_b(t)\| \leq \epsilon$ and $\|\dot{\omega}_b(t)\| \leq \epsilon$ at all t . If $\tilde{g}(t) \doteq \sigma(g_B g(t) g_A)$ for some constant $g_A = (R_A, T_A), g_B = (R_B, T_B), \sigma > 0$, then $\tilde{g}(t)$ satisfies (43) if and only if

$$\|T_A\| \leq \frac{2k \min_t \|\dot{\omega}_b\|}{\max_t \|\dot{\omega}_{imu}\|} \quad (44)$$

$$\|R_A - I\| \leq \frac{2 \min_t \|\dot{\omega}_b\|}{\max_t \|\dot{\omega}_{imu}\|} \quad (45)$$

$$R_B = \exp(\widehat{\omega}_B) \exp(\widehat{\gamma}\theta) \quad \text{with} \quad \|\omega_B\| \leq \left(\frac{3k \max(\min_t \|\dot{\omega}_b\|, \min_t \|\dot{\alpha}_b\|)}{\min(\max_t \|\dot{\omega}_{imu}\|, \max_t \|\dot{\alpha}_{imu}\|, \|\gamma\|)} \right) \quad (46)$$

$$|\sigma - 1| \leq \left(\frac{2k \min_t \|\dot{\alpha}_b\|}{\min(\max_t \|\dot{\omega}_{imu}\|, \max_t \|\dot{\alpha}_{imu}\|)} \right) \quad (47)$$

for a small integer constant k , with θ, T_B, \bar{g}_i arbitrary.

Proof. The ambiguous rotation \tilde{R} must satisfy $\dot{\tilde{R}} = \tilde{R}(\widehat{\omega}_{imu} + \tilde{\omega}_b)$ for some $\tilde{\omega}_b$:

$$\dot{\tilde{R}} = R_B R(\widehat{\omega}_{imu} - \widehat{\omega}_b) R_A = \tilde{R} R_A^T (\widehat{\omega}_{imu} - \widehat{\omega}_b) R_A = \tilde{R} (\widehat{R_A^T \omega}_{imu} - \widehat{R_A^T \omega}_b) = \tilde{R} (\widehat{\omega}_{imu} - \underbrace{\widehat{\omega}_{imu} + \widehat{R_A^T \omega}_{imu} - \widehat{R_A^T \omega}_b}_{-\tilde{\omega}_b})$$

where the quantity in bracket is $-\tilde{\omega}_b$, which defines

$$\tilde{\omega}_b \doteq R_A^T \omega_b + (I - R_A^T) \omega_{imu}. \quad (48)$$

If $\dot{\omega}_b = \tilde{\omega}_b = 0$, under general orientations ω_{imu} , we obtain that $R_A = I$. If, however, $\dot{\omega}_b$ is not zero, but small, for instance $\|\dot{\omega}_b\| \leq \epsilon$, we obtain, using the triangular inequality, that

$$\|R_A - I\| \leq \frac{2\epsilon}{\max_t \|\dot{\omega}_{imu}(t)\|} \doteq k(\dot{\omega})\epsilon \quad (49)$$

where $k(\dot{\omega})$ is a constant that is small when $\max_t \|\dot{\omega}\|$ is large. Since R_A is close to the identity, it can be written as $R_A \simeq I + \widehat{\omega}_A$ up to higher-order terms, so the above can be written as

$$\|\omega_A\| \leq k(\dot{\omega})\epsilon \quad (50)$$

up to higher-order terms.

The ambiguous translation \tilde{T} must satisfy $\dot{\tilde{T}} = \tilde{V}$, which defines \tilde{V} , and thence \tilde{V} must satisfy $\dot{\tilde{V}} = \tilde{R}(\alpha_{imu} - \tilde{\alpha}_b) + \gamma$, which defines $\tilde{\alpha}_b$:

$$\tilde{\alpha}_b = \alpha_{imu} - \tilde{R}^T(\ddot{\tilde{T}} - \gamma) \quad (51)$$

substituting \tilde{T} , we obtain constraints on R_A, T_A, R_B . Note that we do not obtain any constraints on T_B that is *constant* but otherwise *arbitrary*.

$$\tilde{T} = \sigma(R_B RT_A + R_B T + T_B) \quad (52)$$

$$\dot{\tilde{T}} = \sigma R_B(R(\hat{\omega}_{imu} - \hat{\omega}_b)T_A + V) \quad (53)$$

$$\ddot{\tilde{T}} = \sigma R_B R([(\hat{\omega}_{imu} - \hat{\omega}_b)^2 + \dot{\hat{\omega}}_{imu} - \dot{\hat{\omega}}_b]T_A + (\alpha_{imu} - \alpha_b) + R^T \gamma) \quad (54)$$

and therefore

$$\tilde{\alpha}_b = \alpha_{imu} - \sigma R_A^T([(\hat{\omega}_{imu} - \hat{\omega}_b)^2 + \dot{\hat{\omega}}_{imu} - \dot{\hat{\omega}}_b]T_A + (\alpha_{imu} - \alpha_b) + R^T \gamma) + R_A^T R^T R_B^T \gamma \quad (55)$$

$$\tilde{\alpha}_b = \sigma R_A^T \alpha_b + (I - \sigma R_A^T) \alpha_{imu} - R_A^T R^T (\sigma I - R_B^T) \gamma - \sigma R_A^T [(\hat{\omega}_{imu} - \hat{\omega}_b)^2 + \dot{\hat{\omega}}_{imu} - \dot{\hat{\omega}}_b] T_A \quad (56)$$

The derivative is given by

$$\dot{\tilde{\alpha}}_b = \sigma R_A^T \dot{\alpha}_b + (I - \sigma R_A^T) \dot{\alpha}_{imu} - R_A^T (\hat{\omega}_b - \hat{\omega}_{imu}) R^T (\sigma I - R_B^T) \gamma - \sigma R_A^T [2(\hat{\omega}_{imu} - \hat{\omega}_b)(\dot{\hat{\omega}}_{imu} - \dot{\hat{\omega}}_b) + \ddot{\hat{\omega}}_{imu} - \ddot{\hat{\omega}}_b] T_A \quad (57)$$

or

$$R_A \dot{\alpha}_b - \sigma \dot{\alpha}_b = -2\sigma[\hat{\omega}_b \dot{\hat{\omega}}_b - \ddot{\hat{\omega}}_b] T_A + \quad (58)$$

$$+ (R_A - \sigma I) \dot{\alpha}_{imu} + \quad (59)$$

$$- \hat{\omega}_b R^T (\sigma I - R_B^T) \gamma + \quad (60)$$

$$+ \hat{\omega}_{imu} [R^T (\sigma I - R_B^T) \gamma + 2\sigma \dot{\hat{\omega}}_b T_A] \quad (61)$$

$$+ 2\sigma(\hat{\omega}_{imu} - \hat{\omega}_b) \hat{T}_A \dot{\omega}_{imu} \quad (62)$$

$$+ \sigma \hat{T}_A \ddot{\omega}_{imu} \quad (63)$$

The norm of the left-hand side is bounded above by $(1 + \sigma)\epsilon$ and below by $(1 - \sigma)\epsilon$. Because of the general-position conditions, α_{imu} can be arbitrary and so can $R, \omega_{imu}, \dot{\omega}_{imu}$ and $\ddot{\omega}_{imu}$. But if $\hat{\omega}_{imu}$ is arbitrary and independent of $\hat{\omega}_{imu}$ (at any given instant of time, due to the general-position conditions, for any $\omega_{imu}(t)$, it is possible to choose $\dot{\omega}_{imu}(t)$ arbitrarily), then $(\hat{\omega}_{imu} - \hat{\omega}_b)\dot{\hat{\omega}}_{imu}$ is also arbitrary (as time goes by, regardless of $\omega_{imu}(t)$ and $\omega_b(t)$, one can choose $\dot{\omega}_{imu}(t)$ in such a way that $(\hat{\omega}_{imu} - \hat{\omega}_b)\dot{\hat{\omega}}_{imu}$ spans \mathbb{R}^3). Therefore, the norm of each row above has to be bounded by $k\epsilon$ for a suitable constant k . In other words, there are no free parameters on the right-hand side (there are no “tilde” terms), and therefore each term must be bounded independently in order to yield a left-hand side that is bounded between $(1 - \sigma)\epsilon$ and $(1 + \sigma)\epsilon$.

From the second line

$$\|R_A - \sigma I\| \leq \frac{k\epsilon}{\|\dot{\alpha}_{imu}\|} \quad (64)$$

Since $R_A \simeq I + \hat{\omega}_A$, up to higher-order terms, we have

$$\|(1 - \sigma)I + \hat{\omega}_A\| \leq \frac{k\epsilon}{\|\dot{\alpha}_{imu}\|} \quad (65)$$

Using the fact that $\|\omega_A\| \leq \frac{2\epsilon}{\|\dot{\omega}_{imu}\|}$ and the reverse triangular inequality, we obtain

$$|1 - \sigma| - \|\omega_A\| \leq |1 - \sigma| - \frac{2\epsilon}{\|\dot{\omega}_{imu}\|} \leq \frac{k\epsilon}{\|\dot{\alpha}_{imu}\|} \quad (66)$$

and therefore

$$|1 - \sigma| \leq \frac{2\epsilon}{\|\dot{\omega}_{imu}\|} + \frac{k\epsilon}{\|\dot{\alpha}_{imu}\|} \quad (67)$$

where the constant $k(\dot{\omega}_{imu}, \dot{\alpha}_{imu})$ is small when *both* $\dot{\omega}_{imu}$ and $\dot{\alpha}_{imu}$ are small. The last row yields

$$\|T_A\| \leq \frac{k\epsilon}{\sigma\|\ddot{\omega}_{imu}\|} \leq \frac{2k\epsilon}{\|\ddot{\omega}_{imu}\|} \quad (68)$$

The third row yields

$$\|\widehat{\omega}_b R^T (\sigma I - R_B^T) \gamma\| \leq k\epsilon \quad (69)$$

and since R can be arbitrary, we obtain

$$\|(\sigma I - R_B^T) \gamma\| \leq \frac{\epsilon}{\|\omega_b\|} \quad (70)$$

However, the right-hand side can be arbitrarily large, as ω_b may be small. If we decompose R_B into a rotation about gravity and one orthogonal to it, we have that

$$R_B = \exp(\widehat{\omega}_B) \exp(\widehat{\gamma}\theta) \quad (71)$$

where ω_B is a vector that is orthogonal to γ , then $(\sigma I - R_B)\gamma \simeq (\sigma I - (I + \widehat{\omega}_B) \exp(\widehat{\gamma}\theta))\gamma = (\sigma - 1)\gamma + \widehat{\omega}_B\gamma$. From the fourth equation, using the reverse triangular inequality, we obtain

$$\|(\sigma I - R_B^T) \gamma\| - \|2\sigma\dot{\omega}_b T_A\| \leq \|R^T (\sigma I - R_B^T) \gamma + 2\sigma\dot{\omega}_b T_A\| \leq k\epsilon \quad (72)$$

and therefore

$$\|(\sigma I - R_B^T) \gamma\| \leq k\epsilon + \|2\sigma\dot{\omega}_b T_A\| \leq k\epsilon + 2|\sigma|\|\dot{\omega}_b\|\|T_A\| \leq k\epsilon + 2|\sigma|k(\ddot{\omega}_{imu})\epsilon^2 \quad (73)$$

where the latter term only comprises higher-order terms, that we neglect. Finally, we have that

$$\|(\sigma I - R_B^T) \gamma\| \simeq \|(\sigma I - (I + \widehat{\omega}_B) \exp(\widehat{\gamma}\theta)) \gamma\| \leq k\epsilon + \text{H.O.T.} \quad (74)$$

(H.O.T. stands for “higher-order terms”) and since $\exp(\widehat{\gamma}\theta)\gamma = \gamma$ for any θ , we have

$$\|(\sigma - 1)\gamma + \widehat{\omega}_B\gamma\| \leq k\epsilon \quad (75)$$

and, again using the reverse triangular inequality,

$$\|\widehat{\gamma}\omega_B\| \leq |\sigma - 1|\|\gamma\| + k\epsilon \quad (76)$$

from which

$$\|\omega_B\| \leq |\sigma - 1| + \frac{k}{\|\gamma\|}\epsilon \quad (77)$$

Note, however, that θ can be arbitrary, and therefore R_B is unconstrained in its rotation about gravity, but has to have a small off-gravity component. To summarize, compliance with the dynamical model imposed by a rigid motion driven by IMU measurements, we have that the set of $g_A = (R_A, T_A), g_B = (R_B, T_B), \sigma$ has to satisfy:

$$T_A = 0 + \tilde{T}_A \quad \text{with} \quad \|\tilde{T}_A\| \leq \frac{2k \min_t \|\dot{\omega}_b\|}{\max_t \|\ddot{\omega}_{imu}\|} \quad (78)$$

$$R_A = I + \widehat{\omega}_A \quad \text{with} \quad \|\omega_A\| \leq \frac{2 \min_t \|\dot{\omega}_b\|}{\max_t \|\ddot{\omega}_{imu}\|} \quad (79)$$

$$T_B = \text{const.} \quad \text{arbitrary} \quad (80)$$

$$R_B = \exp(\widehat{\omega}_B) \exp(\widehat{\gamma}\theta) \quad \text{with} \quad \|\omega_B\| \leq \left(\frac{3k \max(\min_t \|\dot{\omega}_b\|, \min_t \|\dot{\alpha}_b\|)}{\min(\max_t \|\ddot{\omega}_{imu}\|, \max_t \|\ddot{\alpha}_{imu}\|, \|\gamma\|)} \right) \quad (81)$$

$$\theta \text{ arbitrary} \quad (82)$$

$$\sigma = 1 + \tilde{\sigma} \quad \text{with} \quad |\tilde{\sigma}| \leq \left(\frac{2k \min_t \|\dot{\alpha}_b\|}{\min(\max_t \|\ddot{\omega}_{imu}\|, \max_t \|\ddot{\alpha}_{imu}\|)} \right) \quad (83)$$

where $\|\dot{\omega}_b\| \leq \epsilon$ and k is a small constant. \square

2.3 Gauge transformations

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 $\epsilon \rightarrow 0$ is parametrized by an arbitrary $T_B \in \mathbb{R}^3$ and $\theta \in \mathbb{R}$,

$$\begin{cases} \tilde{T} = \exp(\hat{\gamma}\theta)T + T_B \\ \tilde{R} = \exp(\hat{\gamma}\theta)R \\ \tilde{T}_{t_i} = \exp(\hat{\gamma}\theta)\bar{T}_{t_i} + T_B \\ \tilde{R}_{t_i} = \exp(\hat{\gamma}\theta)\bar{R}_{t_i} \\ \tilde{T}_{cb} = T_{cb} \\ \tilde{R}_{cb} = R_{cb} \end{cases} \quad \text{up to } \mathcal{O}\left(\frac{\|\dot{\omega}_b\|}{\|\dot{\omega}_{imu}\|}, \frac{\|\dot{\alpha}_b\|}{\|\dot{\alpha}_{imu}\|}, \frac{1}{\|\gamma\|}\right) \quad (84)$$

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

$$\tilde{X}^j = g_a \bar{g}_i^{-1} g_i g_a^{-1} X^j. \quad (85)$$

If each group i contains at least 3 non-coplanar points, it is possible to fix \bar{g}_i by parametrizing $X^j \doteq \bar{y}_{t_i}^j Z^j$ and imposing three directions $y_{t_i}^j = \tilde{y}_{t_i}^j = y^j(t_i), j = 1, \dots, 3$, 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 i we impose

$$R_{t_i} = \tilde{R}_{t_i} = \hat{R}(t_i) \text{ and } T_{t_i} = \tilde{T}_{t_i} = \hat{T}(t_i) \quad (86)$$

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

$$\hat{R}(t_i) = \exp(\hat{\gamma}\theta)\hat{R}(t_i) \quad (87)$$

and therefore $\theta = 0$; similarly,

$$T_B = (I - \exp(\hat{\gamma}\theta))T(t_i) = 0 \quad (88)$$

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 $[I_{2 \times 20}]R_0\alpha_{imu} = 0$, 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 ([12] 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 \tilde{x} that is related to the true one x as follows:

$$\tilde{X}^{\text{ref}} = (1 + \tilde{\sigma})\tilde{R}_{cb}e^{\hat{\omega}_B}e^{\hat{\gamma}\theta}e^{\hat{\omega}_A}\tilde{R}_{cb}^T(X^{\text{ref}} - T_A) + (1 + \tilde{\sigma})(\tilde{R}_{cb}e^{\hat{\omega}_A}T_B + \tilde{R}_{cb}T_A + \tilde{T}_{cb}) \quad (89)$$

$$\tilde{X}^j = (1 + \tilde{\sigma})\tilde{R}_{cb}\bar{R}_i\tilde{R}_t_i\tilde{R}_{cb}^T(X^j - T_A) + (1 + \tilde{\sigma})(\tilde{R}_{cb}\bar{R}_i\tilde{T}_t_i + \tilde{R}_{cb}\bar{T}_i + \tilde{T}_{cb}) \quad (90)$$

$$\tilde{T} = e^{\hat{\gamma}\theta}T + T_B(1 + \tilde{\sigma}) + \hat{\omega}_B e^{\hat{\gamma}\theta}T + e^{\hat{\omega}_B}e^{\hat{\gamma}\theta}RT_A(1 + \tilde{\sigma}) \quad (91)$$

$$\tilde{R} = e^{\hat{\omega}_B}e^{\hat{\gamma}\theta}Re^{\hat{\omega}_A} \quad (92)$$

$$\tilde{T}_{t_i} = e^{\hat{\gamma}\theta}\bar{T}_i + T_B(1 + \tilde{\sigma}) + \hat{\omega}_B e^{\hat{\gamma}\theta}\bar{T}_i + e^{\hat{\omega}_B}e^{\hat{\gamma}\theta}\bar{R}_iT_A(1 + \tilde{\sigma}) \quad (93)$$

$$\tilde{R}_{t_i} = e^{\hat{\omega}_B}e^{\hat{\gamma}\theta}\bar{R}_i e^{\hat{\omega}_A} \quad (94)$$

$$\tilde{T}_{cb} = T_{cb} + \tilde{\sigma}T_{cb} + R_{cb}T_A(1 + \tilde{\sigma}) \quad (95)$$

$$\tilde{R}_{cb} = R_{cb} \exp(\hat{\omega}_A) \quad (96)$$

$$\begin{aligned} \tilde{\alpha}_b &= (57) \\ \tilde{\omega}_b &= (48) \end{aligned}$$

where

$$\|T_A\| \leq \frac{2k \min_t \|\dot{\omega}_b\|}{\max_t \|\ddot{\omega}_{imu}\|}$$

$$\|\omega_A\| \leq \frac{2 \min_t \|\dot{\omega}_b\|}{\max_t \|\ddot{\omega}_{imu}\|}$$

$$\|\omega_B\| \leq \left(\frac{3k \max(\min_t \|\dot{\omega}_b\|, \min_t \|\dot{\alpha}_b\|)}{\min(\max_t \|\ddot{\omega}_{imu}\|, \max_t \|\ddot{\alpha}_{imu}\|, \|\gamma\|)} \right)$$

$$|\tilde{\sigma}| \leq \left(\frac{2k \min_t \|\dot{\alpha}_b\|}{\min(\max_t \|\ddot{\omega}_{imu}\|, \max_t \|\ddot{\alpha}_{imu}\|)} \right)$$

and arbitrary θ , T_B and suitable constant κ . 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 [21], 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. (84).

3 Measurement model reduction

The measurement model $y = \pi(gX) + n$ involves navigation states $g \in SE(3)$ as well as constant unknown parameters $X \in \mathbb{R}^{3 \times N}$, that are also modeled as states. To distinguish them, we indicate the former with x and the latter with p . In that case, the measurement equation is of the form $y = h(x, p) + n$. In some cases,⁸ one may be interested in reducing the model by *eliminating* the unknown states p . This can be done in an approximate manner via linearization, or in an exact manner using the geometry of the space of unknowns. For the case of bearing measurements, this has been first done in [24] for the continuous-case, and [23, 25, 26] for the discrete-time case. The linearization version has been done in [21] and successfully demonstrated on a cellphone for the case of vision measurements.

⁸After having augmented the model by *adding* the same states p .

3.1 Approximate model reduction via linearization (bearing)

We consider the linear approximation of the measurement equation:

$$y = h(x, p) = h(\hat{x}, \hat{p}) + \underbrace{\frac{\partial h}{\partial p}(\hat{x}, \hat{p}) \tilde{p}}_{H_p} + \underbrace{\frac{\partial h}{\partial x}(\hat{x}, \hat{p}) \tilde{x}}_{H_x} + \underbrace{\mathcal{O}(\|\tilde{x}\|^2, \|\tilde{p}\|^2)}_{\tilde{n}} + n \quad (97)$$

where the matrices H_p and H_x are the Jacobians of the measurements with respect to the state and the unknown parameters p , and the new residual \tilde{n} includes measurement noise as well as linearization error.

If the Jacobian H_p has full column rank, it is possible to eliminate the dependency of the measurement model from \tilde{p} by multiplying both sides of the equation above by its orthogonal complement H_p^\perp , obtaining

$$r \doteq H_p^\perp(y - h(\hat{x}, \hat{p})) = \underbrace{H_p^\perp(\hat{x}, \hat{p}) H_p \tilde{p}}_{=0} + H_p^\perp(\hat{x}, \hat{p}) H_x(\hat{x}, \hat{p}) \tilde{x} + \bar{n} \quad (98)$$

where $\bar{n} = H_p^\perp \tilde{n}$. Unfortunately, in neither the bearing-only nor the range-only case is the matrix H_p full column rank. In order to enable eliminating states \tilde{p} , [21] stack a number of temporal samples $y(t)$ on top of each other, so the stacked matrix H_p becomes full column rank, and reduction is made possible. Writing the contribution from individual points, we have

$$\begin{bmatrix} r^i(t_1) \\ r^i(t_2) \\ \vdots \\ r^i(t_i) \end{bmatrix} = \underbrace{\begin{bmatrix} H_p(\hat{x}(t_1), \hat{p}^i) \\ H_p(\hat{x}(t_2), \hat{p}^i) \\ \vdots \\ H_p(\hat{x}(t_i), \hat{p}^i) \end{bmatrix}}_{\mathbf{H}_p(\hat{x}, \hat{p}^i)}^\perp \begin{bmatrix} \ddots & & & \\ & H_x(\hat{x}(t), \hat{p}^i) & & \\ & & \ddots & \\ & & & H_x(\hat{x}(t_i), \hat{p}^i) \end{bmatrix} \begin{bmatrix} \tilde{x}(t_1) \\ \tilde{x}(t_2) \\ \vdots \\ \tilde{x}(t_i) \end{bmatrix} + \begin{bmatrix} \bar{n}(t_1) \\ \bar{n}(t_1) \\ \vdots \\ \bar{n}(t_i) \end{bmatrix} \quad (99)$$

or in compact form as

$$\mathbf{r}(t, t_i) = \mathbf{H}(\hat{x}, \hat{p}) \begin{bmatrix} \tilde{x}(t_1) \\ \tilde{x}(t_2) \\ \vdots \\ \tilde{x}(t_i) \end{bmatrix} + \bar{\mathbf{n}}(t, t_i) \quad (100)$$

Note that the noise term \bar{n} depends on the unknown structure \hat{p} and the nominal motion states \hat{x} , in addition to the linearization error. So even if the measurement noise was reasonably uncorrelated, certainly \bar{n} is heavily correlated, and even more so once the delay-line (sequence of temporally adjacent measurements) are considered as a batch. One could introduce additional states to model the correlation terms, but this would defeat the purpose of model reduction, which is to eliminate states.

In addition, even if we have eliminated the dependency on the error-state \tilde{p} (the ‘‘structure correction’’ term), the Jacobians still depend on the (unknown) nominal structure \hat{p} . In [21], the current estimates of the motion states $\hat{x} + \tilde{x}$ are used to triangulate points in \mathbb{R}^3 , which is possible for the case of bearing measurements, to obtain a coarse approximation of the position of each point which is taken to be \hat{p}^i , and used to compute the Jacobians above. Specifically, given at least two time instants, for instance $\tau_i = 0$ and $t_i = t$, from (10) we get

$$\widehat{y^i(t)} \widehat{y^i(0)} Z^i = \tilde{n}^i(t) \quad (101)$$

that can be solved in a least-squares fashion to estimate \hat{Z}^i and therefore $p^i = X^i$.

In the next section, we show how model reduction is possible in an *exact* fashion, in a manner that does not require linearization, and does not require an estimate of the structure states \hat{p} .

3.2 Model reduction via epipolar geometry (bearing): The Essential filter

Reducing the model for the case of bearing measurements can be done by eliminating the structure states via Epipolar geometry. Assuming the feature appears at $t = 0$, for any time $t > 0$ we have

$$y^T(t) \underbrace{R^T(t)\hat{T}(t)}_Q y^i(0) = n^i(t) \quad (102)$$

Here, there is no linearization, the independence of structure is exact, there is no need to triangulate the structure, and no need to assemble multiple data in a batch. Since the equation above is linear in Q , we write it as

$$\chi(t)Q(x) = n(t) \quad (103)$$

where $\chi(t) = (\bar{y}(t) \otimes \bar{y}(0))^T$ is the Kronecker product of $y(t)$ and $y(0)$.

4 Delay Line State Augmentation

Representing groups of points via their pose results in augmenting the state with a collection of “key-poses,” corresponding to time instants t_i when groups of features appear. Rather than picking key poses one could augment the state with a *delay-line*, or sliding window, a collection of adjacent time samples of the state, as suggested in [21]. For instance, if $x(t)$ is the state, the delay-line is the augmented state $x^m(t) = \{x(t), x(t-1), \dots, x(t-m)\}$.

To build a delay-line of length $m > 0$ starting from an initial time t_i and define the states $x_k(t) \doteq x(t - kdt)$ for $k = 1, \dots, m$. Assuming a constant inter-frame sampling $dt > 0$, for a linear model of the form $\dot{x} = f(x) + c(x)u = Ax + Bu$, starting at $t \geq t_i + m$ we have:

$$\left\{ \begin{array}{ll} x_m(t+dt) = x_{m-1}(t) & x_m(t) = x(t_i) \\ x_{m-1}(t+dt) = x_{m-2}(t) & x_{m-1}(t) = Ax(t_i) + Bu(t_i) \\ \vdots & \\ x_{m-k}(t+dt) = x_{m-k-1}(t) & x_{m-k}(t) = A^k x(t_i) + \sum_{j=1}^{k-1} A^{k-j} Bu(t_i + jdt) \\ \vdots & \\ x_2(t+dt) = x_1(t) & x_2(t) = A^{m-2} x(t_i) + \sum_{j=1}^{m-3} A^{m-1-j} Bu(t_i + jdt) \\ x_1(t+dt) = x(t) & x_1(t) = A^{m-1} x(t_i) + \sum_{j=1}^{m-2} A^{m-1-j} Bu(t_i + jdt) \\ \dot{x} = f(x) + c(x)u & x(t) = A^m x(t_i) + \sum_{j=1}^{m-1} A^{m-1-j} Bu(t_i + jdt) \end{array} \right. \quad (104)$$

The initial conditions on the right can be computed recursively. In the non-linear case, this is implemented as m repeated steps of the integral $x(t+dt) = x(t) + \int_t^{t+dt} f(x)d\tau + \int_t^{t+dt} c(x)du(\tau) \doteq F(x, t) + B(x, u, t)$, each step defining the initialization of one delay block: Assuming an initial time t_i , then at time $t = t_i + m$, we have

$$x_m(t) = x(t_i) \quad (105)$$

$$x_{m-1}(t) = F(x_m, t) + B(x_m, u, t) \quad (106)$$

$$x_{m-2}(t) = F(x_{m-1}, t) + B(x_{m-1}, u, t) \quad (107)$$

$$\vdots \quad (108)$$

$$x_1(t) = F(x, t) + B(x, u, t) \quad (109)$$

Once initialized, the model evolves according to (104) for an augmented state and linearization

$$x^m(t) = \begin{bmatrix} x_m(t) \\ x_{m-1}(t) \\ \vdots \\ x_1(t) \\ x(t) \end{bmatrix}, \quad F = \begin{bmatrix} 0 & I & & & \\ & 0 & I & & \\ & & \ddots & & \\ & & & 0 & I \\ 0 & \dots & 0 & \frac{\partial f}{\partial x}(x(t)) & \end{bmatrix}, \quad C = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \frac{\partial c}{\partial x}(x(t)) \end{bmatrix} \quad (110)$$

Starting from $t > t_i + m$, we have measurements available, of the form

$$y(t-m) = y(t_i) = h(x_m(t)) + n(t_i) \quad (111)$$

$$y(t-m+1) = y(t_i+1) = h(x_{m-1}(t)) + n(t_i+1) \quad (112)$$

$$\vdots \quad (113)$$

$$y(t-1) = y(t_i+m-1) = h(x_1(t)) + n(t_i+m-1) \quad (114)$$

$$y(t) = y(t_i+m) = h(x(t)) + n(t_i+m) \quad (115)$$

which we could collate into an augmented measurement equation $y^m(t)$. However, this measurement cannot be used at every time instant t , as the sliding window would result in highly (temporally) correlated noise n^m . Instead, it must be used at block intervals of km time instants, at which point it provides constraints on each of the states $x^m(t)$. Alternatively, a batch of measurements $y^m(t)$ can be used only once, at time $t = t_i + m$, to provide measurement constraints on the augmented states, as done in [21].

4.1 Multi-State Constraints with Epipolar Geometry

The epipolar constraint can be interpreted as an implicit measurement equation involving inter-frame pose, regardless of the position of points. If $X(t) = R^T(t)(X_0 - T(t))$, the Essential matrix at time t is of the form $Q(t) = R^T(t)\widehat{T}(t)$.⁹ The inter-frame Essential matrix, $Q(t_2, t_1)$ is then the Essential matrix determined by the motion between t_1 and t_2 ,¹⁰ that has a rotational component $R(t_2, t_1) \doteq R(t_2)^T R(t_1)$ and a translational component $T(t_2, t_1) = R^T(t_2)(T(t_1) - T(t_2))$. Therefore, $Q(t_2, t_1) = R^T(t_1)R(t_2)[R^T(t_2)(\widehat{T(t_1)} - T(t_2))]$, or equivalently

$$Q(t_2, t_1) = R^T(t_1)[\widehat{T}(t_1) - \widehat{T}(t_2)]R(t_2) \quad (116)$$

independent of the choice of spatial frame, since the transformation from the camera frame at t_1 to the spatial frame is annihilated by the transformation from the spatial frame to the camera frame at t_2 . In the presence of an unknown alignment transformation $g_{cb} = (R_{cb}, T_{cb})$ one can easily verify that

$$Q(t_2, t_1; R_{cb}, T_{cb}) = R^T(t_1)[\widehat{T}(t_1) - \widehat{T}(t_2) + R(t_2)R_{cb}^T T_{cb}]R(t_2)R_{cb}^T \quad (117)$$

or, equivalently,

$$Q(t_2, t_1; R_{cb}, T_{cb}) = Q(t_2, t_1)R_{cb}^T + R^T(t_1)R(t_2)R_{cb}^T \widehat{T}_{cb} \quad (118)$$

⁹Note that if the motion model was instead $X(t) = R(t)X_0 + T(t)$, then the Essential matrix would have the form $Q(t) = \widehat{T}(t)R(t)$. The equivalence between the two can be arrived at by noticing that $\widehat{R^T T} = R^T \widehat{T} R$.

¹⁰From $X(t_1) = R^T(t_1)(X_0 - T(t_1))$ and $X(t_2) = R^T(t_2)(X_0 - T(t_2))$.

Now, if we grow a batch of m measurements starting from an initial t_i up to time $t = t_i + m$, we have, calling $\chi(t_1, t_2) \doteq \chi(y(t_1), y(t_2))$ and $x \doteq x(t); x_k \doteq x_k(t)$,

$$\chi(t, t - m)Q(x, x_m) = \chi(t_i + m, t_i)Q(x(t_i + m), x(t_i)) = 0 \quad (119)$$

$$\chi(t, t - m + 1)Q(x, x_{m-1}) = \chi(t_i + m, t_i + 1)Q(x(t_i + m), x(t_i + 1)) = 0 \quad (120)$$

$$\vdots \quad (121)$$

$$\chi(t, t - m + k)Q(x, x_{m-k}) = \chi(t_i + m, t_i + k)Q(x(t_i + m), x(t_i + k)) = 0 \quad (122)$$

$$\vdots \quad (123)$$

$$\chi(t, t - 1)Q(x, x_1) = \chi(t_i + m, t_i + m - 1)Q(x(t_i + m), x(t_i + m - 1)) = 0 \quad (124)$$

5 Discussion

Sensor platform trajectory (position and orientation) is commonly believed to be observable from bearing and inertial sensors, along with accelerometer and gyro biases and alignment. We show that this is only the case when the biases are strictly *constant*, but as soon as the biases are allowed to vary, however, slowly, the conclusion that the state of the underlying model is observable is not valid. While the bias derivatives may be *small*, they are not *white*, and therefore they should play a role in the observability analysis.

We study the *unknown-input observability* of visual-inertial navigation by explicitly characterizing the set of indistinguishable state trajectories, and showing that they are *not* singletons, therefore contradicting prior observability analysis. We derive bounds on the volume of the indistinguishable set, that is proportional to the accelerometer and gyro bias rates (it reduces to zero when the biases are constant) and inversely proportional to derivatives of linear and angular velocities and acceleration.

Various forms of model reduction do not change the observability properties of the model, and in general are detrimental from the perspective of uncertainty management. However, as demonstrated by recent work of Roumeliotis, Mourikis and coworkers, model reduction can lead to significant advantages in the presence of intermittent measurements, for instance due to features appearing/disappearing, and in the presence of short feature tracks, that would not allow sufficient time for convergence if inserted in the state.

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] R. Beard, W. Curtis, M. Eilders, J. Evers, and J. Cloutier. Vision aided proportional navigation for micro air vehicles. In *Proceedings of the AIAA Guidance, Navigation, and Control Conference. Paper number AIAA-2007-6609. American Institute of Aeronautics and Astronautics, Hilton Head*, 2007.
- [3] 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.
- [4] F. Di Corato, M. Innocenti, and L. Pollini. Visual-inertial navigation with guaranteed convergence.
- [5] D. Diel, P. DeBitetto, and S. Teller. Epipolar constraints for vision-aided inertial navigation. In *Application of Computer Vision, 2005. WACV/MOTIONS'05 Volume 1. Seventh IEEE Workshops on*, volume 2, pages 221–228. IEEE, 2005.

- [6] 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.
- [7] Q. Fang and X. Huang. Global observability analysis imu/camera integration. *Applied Mechanics and Materials*, 380:1069–1072, 2013.
- [8] 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.
- [9] H. Hammouri and Z. Tmar. Unknown input observer for state affine systems: A necessary and sufficient condition. *Automatica*, 46(2):271–278, 2010.
- [10] J. Hesch, D. Kottas, S. Bowman, and S. Roumeliotis. Observability-constrained vision-aided inertial navigation. *University of Minnesota, Dept. of Comp. Sci. & Eng., MARS Lab, Tech. Rep*, 1, 2012.
- [11] J. Hesch, D. Kottas, S. Bowman, and S. Roumeliotis. Towards consistent vision-aided inertial navigation. In *Algorithmic Foundations of Robotics X*, pages 559–574. Springer, 2013.
- [12] 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.
- [13] E. S. Jones, A. Vedaldi, and S. Soatto. Inertial structure from motion and autocalibration. In *Workshop on Dynamical Vision*, October 2007.
- [14] J. Kelly and G. Sukhatme. Fast Relative Pose Calibration for Visual and Inertial Sensors. In *Experimental Robotics*, pages 515–524, 2009.
- [15] L. Kneip, S. Weiss, and R. Siegwart. Deterministic initialization of metric state estimation filters for loosely-coupled monocular vision-inertial systems. In *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*, pages 2235–2241. IEEE, 2011.
- [16] D. Kottas, J. Hesch, S. Bowman, and S. Roumeliotis. On the consistency of vision-aided inertial navigation. In *Proceedings of the International Symposium on Experimental Robotics*, 2012.
- [17] D. Liberzon, P. R. Kumar, A. Dominguez-Garcia, and S. Mitra. Invertibility and observability of switched systems with inputs and outputs. 2012.
- [18] T. Lupton and S. Sukkarieh. Visual-inertial-aided navigation for high-dynamic motion in built environments without initial conditions. *Robotics, IEEE Transactions on*, 28(1):61–76, 2012.
- [19] A. Martinelli. Closed-form solutions for attitude, speed, absolute scale and bias determination by fusing vision and inertial measurements. 2011.
- [20] A. Martinelli. Visual-inertial structure from motion: observability and resolvability. *IROS 2013*, 2013.
- [21] 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.
- [22] S. Soatto. 3-d structure from visual motion: modeling, representation and observability. *Automatica*, 33:1287–1312, 1997.
- [23] S. Soatto, R. Frezza, and P. Perona. Motion estimation via dynamic vision. *IEEE Transactions on Automatic Control*, 41(3):393–414, March 1996.
- [24] S. Soatto and P. Perona. Recursive 3-d visual motion estimation using subspace constraints. *Int. J. of Computer Vision*, 22(3):235–259, 1997.

- [25] S. Soatto and P. Perona. Reducing “structure from motion”: a general framework for dynamic vision. part 1: modeling. *IEEE Trans. Pattern Anal. Mach. Intell.*, 20(9):993–942, September 1998.
- [26] S. Soatto and P. Perona. Reducing “structure from motion”: a general framework for dynamic vision. part 2: Implementation and experimental assessment. *IEEE Trans. Pattern Anal. Mach. Intell.*, 20(9):943–960, September 1998.
- [27] C. Taylor. An analysis of observability-constrained kalman filtering for vision-aided navigation. In *Position Location and Navigation Symposium (PLANS), 2012 IEEE/ION*, pages 1240–1246. IEEE, 2012.