Article

The International Journal of
Robotics Research
2014, Vol 33(1) 182–201

DOI: 10.1177/0278364913509675
ijr.sagepub.com

Camera-IMU-based localization: Observability analysis and consistency improvement

Joel A Hesch[1], Dimitrios G Kottas[2], Sean L Bowman[3] and Stergios I Roumeliotis[2]

Abstract

This work investigates the relationship between system observability properties and estimator inconsistency for a Vision-aided Inertial Navigation System (VINS). In particular, first we introduce a new methodology for determining the unobservable directions of nonlinear systems by factorizing the observability matrix according to the observable and unobservable modes. Subsequently, we apply this method to the VINS nonlinear model and determine its unobservable directions analytically. We leverage our analysis to improve the accuracy and consistency of linearized estimators applied to VINS. Our key findings are evaluated through extensive simulations and experimental validation on real-world data, demonstrating the superior accuracy and consistency of the proposed VINS framework compared to standard approaches.

Keywords

Vision-aided inertial navigation, visual-inertial odometry, observability analysis, estimator consistency

1. Introduction

Wide adoption of robotic technologies hinges on the ability of robots to freely navigate in our human-centric world. To do so, robots must maintain an accurate estimate of their six-degrees-of-freedom (DOF) position and orientation (pose) as they navigate in 3D. Ideally, a localization algorithm should work seamlessly outdoors and indoors. This unfortunately prohibits reliance on GPS, since coverage is not available everywhere. For this reason, researchers have focused on designing localization methods that fuse onboard sensor data to estimate the robot's ego motion.

Tracking 3D motion can be accomplished by integrating the rotational velocity and linear acceleration signals provided by an Inertial Measurement Unit (IMU). However, due to integration of sensor noise and bias, pose estimates based on IMU data alone will quickly accumulate errors. To reduce the impact of these errors, so-called aided Inertial Navigation Systems (INS) have been proposed. Laser-aided INS (LINS) methods typically rely on the existence of structural planes (Hesch et al. 2010) or height invariance in semi-structured spaces (Shen et al. 2011), and are not easily generalizable to cluttered or unstructured environments. On the other hand, Vision-aided INS (VINS) approaches, which fuse data from a camera and an IMU, can operate in both structured and unstructured areas. VINS methods have the additional benefit that both inertial and visual sensors are lightweight, inexpensive (available in most mobile devices today), and passive, hence requiring a smaller power budget compared to LINS.

Existing work on VINS has employed a number of different estimators such as the Extended Kalman Filter (EKF) (Kim and Sukkarieh 2007; Mourikis and Roumeliotis 2007; Bryson and Sukkarieh 2008), the Unscented Kalman Filter (UKF) (Ebcin and Veth 2007), and Batch-least Squares (BLS) (Strelow 2004). Non-parametric estimators, such as the Particle Filter (PF), have also been used for visual-inertial odometry (e.g. Durrie et al. (2009); Yap et al. (2011)). However, these have focused on the reduced problem of estimating a 2D robot pose, since the number of particles required is exponential in the size of the state vector. Within these works, a number of challenging issues have been addressed, such as reducing the computational cost of VINS (Mourikis and Roumeliotis 2007; Williams et al. 2011), dealing with delayed measurements (Weiss et al. 2012), increasing the accuracy of feature initialization and estimation (Jones and Soatto 2011), and improving the robustness to estimator initialization errors (Lupton and Sukkarieh 2012). Only limited attention, however,

[1]Google, Mountain View, CA, USA
[2]Department of Computer Science and Engineering, University of Minnesota, Minneapolis, USA
[3]Department of Computer and Information Science, University of Pennsylvania, Philadelphia, USA

Corresponding author:
JA Hesch, Google, Mountain View, CA 94043, USA.
Email: joel@cs.umn.edu

has been devoted to understanding how estimator inconsistency affects VINS.[1] The work described in this paper, addresses this limitation through the following three main contributions:

- We introduce a novel methodology for identifying the unobservable modes of a nonlinear system. Contrary to previous methods (Hermann and Krener 1977) that require investigating an *infinite* number of Lie derivatives, our approach employs a factorization of the observability matrix, according to the observable and unobservable modes, and only requires computing a *finite* number of Lie derivatives.
- We apply our method to VINS and determine its unobservable directions, providing their analytical form as functions of the system states.
- We leverage our results to improve the consistency and accuracy of VINS, and extensively validate the proposed estimation framework both in simulations and real-world experiments.

The rest of this paper is organized as follows: we begin with an overview of the related work (Section 2). In Section 3, we describe the system and measurement models used in VINS. Subsequently, we introduce our methodology for analyzing the observability properties of unobservable nonlinear systems (Section 4), which we leverage for determining the unobservable directions of the VINS model (Section 5). In Section 6, we present an overview of the analogous observability properties for the linearized system employed for estimation purposes, and show how linearization errors can lead to a violation of the observability properties, gain of spurious information, and estimator inconsistency. We propose an estimator modification to mitigate this issue in Section 6.1, and validate our algorithm in simulations and experimentally (Sections 7 and 8). Finally, we provide our concluding remarks and outline our future research directions in Section 9.

2. Related work

The interplay between a nonlinear system's observability properties and the consistency of the corresponding linearized estimator has become a topic of increasing interest within the robotics community in recent years. Huang et al. (Huang et al. 2008, 2010, 2011) first studied this connection for 2D Simultaneous Localization and Mapping (SLAM) and extended their work to 2D cooperative multi-robot localization. In both cases, they proved that a mismatch exists between the number of unobservable directions of the true nonlinear system and the linearized system used for estimation purposes. Specifically, the estimated (linearized) system has one-fewer unobservable direction than the true system, allowing the estimator to surreptitiously gain spurious information along the direction corresponding to global orientation.

Extending this analysis to 3D VINS is a formidable task, most notably since the VINS system state has 15 DOF instead of 3. Some authors have attempted to avoid this complexity by using abstract models (e.g. by assuming a 3D odometer that measures six-DOF pose displacements (Carlone et al. 2012)); though these approaches cannot be easily extended to the VINS. The observability properties of VINS have been examined for a variety of scenarios. For example, Mirzaei and Roumeliotis (2008) as well as Kelly and Sukhatme (2011) have studied the observability properties of IMU-camera extrinsic calibration using Lie derivatives (Hermann and Krener 1977). The former analysis, however, relies on known feature coordinates, while the latter employs an inferred measurement model (i.e. assuming the camera observes its pose in the world frame, up to scale), which requires a non-minimal set of visual measurements. This limitation is also shared by Weiss (2012), who employs symbolic/numeric software tools, rather than providing an analytical study.

Jones and Soatto (2011) investigated VINS observability by examining the indistinguishable trajectories of the system (Isidori 1989) under different sensor configurations (i.e. inertial only, vision only, vision and inertial). Their analysis, however, is restrictive due to: (a) the use of a stochastic tracking model (constant translational jerk and rotational acceleration), which cannot adequately describe arbitrary trajectories, and (b) considering the IMU biases to be known, which is a limiting assumption. They conclude that the VINS state is observable, if the six DOF of the first frame are fixed (e.g. by following the approach in Chiuso et al. (2002)). As a result, their analysis does not provide an explicit form of the unobservable directions, nor does it fully characterize the observability properties of rotation (i.e. that yaw is unobservable).

Finally, Martinelli (2012) used the concept of continuous symmetries to show that the IMU biases, 3D velocity, and absolute roll and pitch angles are observable for VINS. In this case, the unobservable directions are determined analytically for the special case of a single-point feature located at the origin, but the unobservable directions for the case of multiple points are not provided. More importantly, however, among these VINS observability studies, no one has examined the link between observability and estimator consistency, or used their observability study to bolster estimator performance.

A preliminary version of our investigation was presented at the International Workshop on the Algorithmic Foundations of Robotics (Hesch et al. 2012b). In comparison to Hesch et al. (2012b), here we present a novel method for analyzing the observability properties of an unobservable nonlinear system based on factorizing the observability matrix. Moreover, we have included a new derivation for analytically determining the observability properties of the nonlinear VINS model. Finally, we have updated our literature review to include recently published papers and

expanded the experimental validation to better demonstrate the robustness of our approach.

Li and Mourikis (2012, 2013) have also presented an investigation of estimator inconsistency utilizing linearized observability analysis of a bias-free VINS model. Based on their findings, they leveraged the First-Estimates Jacobian (FEJ) methodology of Huang et al. (2008) to reduce the impact of inconsistency in Visual-Inertial Odometry (VIO). In contrast, our observability analysis encompasses both the nonlinear and linearized systems, using the full VINS state (i.e. including IMU biases). Furthermore, the implementation of our approach is more flexible since any linearization method can be employed (e.g. computing Jacobians analytically, numerically, or using sample points) by the estimator.

In what follows, we analytically determine the observability properties of both the nonlinear VINS model and its linearized counterpart to prove that they have four unobservable DOF, corresponding to three-DOF global translations and one-DOF global rotation about the gravity vector. Then, we show that due to linearization errors, the number of unobservable directions is reduced in a standard EKF-based VINS approach, allowing the estimator to gain spurious information and leading to inconsistency. Finally, we propose a solution for reducing estimator inconsistency in VINS that is general, and can be directly applied in a variety of linearized estimation frameworks such as the EKF and UKF both for Visual SLAM (V-SLAM) and VIO.

3. VINS estimator description

We begin with an overview of the VINS propagation and measurement models, and describe the EKF employed for fusing the camera and IMU measurements. In the following analysis, we consider the general case, in which the system state contains both the sensor platform state (i.e. pose, velocity, and IMU biases) and the observed features. However, it is important to note that the same analysis applies to VINS applications that do not explicitly estimate a map of the environment, such as VIO (Mourikis and Roumeliotis 2007).

3.1. System state and propagation model

The system state comprises the IMU pose and linear velocity together with the time-varying IMU biases and a map of visual features. The $(16+3N)\times 1$ state vector is

$$\begin{aligned}\mathbf{x} &= \begin{bmatrix} {}^I\bar{q}_G^{\mathrm{T}} & \mathbf{b}_g^{\mathrm{T}} & {}^G\mathbf{v}_I^{\mathrm{T}} & \mathbf{b}_a^{\mathrm{T}} & {}^G\mathbf{p}_I^{\mathrm{T}} & | & {}^G\mathbf{p}_{f_1}^{\mathrm{T}} \cdots {}^G\mathbf{p}_{f_N}^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}} \\ &= \begin{bmatrix} \mathbf{x}_s^{\mathrm{T}} & | & \mathbf{x}_m^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}} \end{aligned} \tag{1}$$

where $\mathbf{x}_s(t)$ is the 16×1 sensor platform state, and $\mathbf{x}_m(t)$ is the $3N\times 1$ state of the map. The first component of the sensor platform state, ${}^I\bar{q}_G(t)$, is the unit quaternion representing the orientation of the *global frame* $\{G\}$ in the IMU frame $\{I\}$ at time t (see Figure 1). The frame $\{I\}$ is attached to the IMU, while $\{G\}$ is a local-vertical reference frame whose origin coincides with the initial IMU position. The sensor platform state also includes the position and velocity of $\{I\}$ in $\{G\}$, denoted by the 3×1 vectors ${}^G\mathbf{p}_I(t)$ and ${}^G\mathbf{v}_I(t)$, respectively. The remaining components are the biases, $\mathbf{b}_g(t)$ and $\mathbf{b}_a(t)$, affecting the gyroscope and accelerometer measurements, which are modeled as random-walk processes driven by the zero-mean, white Gaussian noise $\mathbf{n}_{wg}(t)$ and $\mathbf{n}_{wa}(t)$, respectively. The map, $\mathbf{x}_m$, comprises N visual features ${}^G\mathbf{p}_{f_i}$, $i=1,\ldots,N$. Note that for the case of VIO, the features are not stored in the state vector, but can be processed and marginalized on-the-fly (Mourikis and Roumeliotis 2007) (see Section 3.2). With the state of the system now defined, we turn our attention to the continuous-time kinematic model which governs the time evolution of the system state.

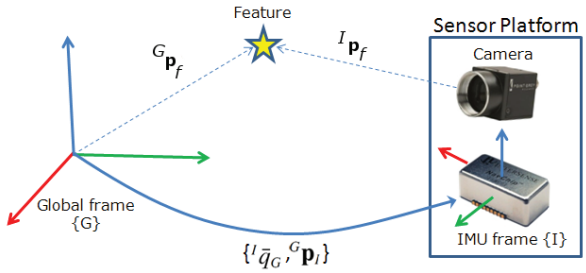


Fig. 1. The pose of the camera-IMU frame $\{I\}$ with respect to the global frame $\{G\}$ is expressed by the position vector ${}^G\mathbf{p}_I$ and the quaternion of orientation ${}^I\bar{q}_G$. The observed feature is expressed in the global frame by its 3×1 position coordinate vector ${}^G\mathbf{p}_f$, and in the sensor frame by ${}^I\mathbf{p}_f = \mathbf{C}\left({}^I\bar{q}_G\right)\left({}^G\mathbf{p}_f - {}^G\mathbf{p}_I\right)$.

3.1.1. Continuous-time model The system model describing the time evolution of the state is (see Chatfield (1997) and Trawny and Roumeliotis (2005)):

$$ {}^I\dot{\bar{q}}_G(t) = \frac{1}{2}\mathbf{\Omega}(\boldsymbol{\omega}(t))\,{}^I\bar{q}_G(t) \tag{2}$$

$$ {}^G\dot{\mathbf{p}}_I(t) = {}^G\mathbf{v}_I(t) \tag{3}$$

$$ {}^G\dot{\mathbf{v}}_I(t) = {}^G\mathbf{a}_I(t) \tag{4}$$

$$ \dot{\mathbf{b}}_g(t) = \mathbf{n}_{wg}(t) \tag{5}$$

$$ \dot{\mathbf{b}}_a(t) = \mathbf{n}_{wa}(t) \tag{6}$$

$$ {}^G\dot{\mathbf{p}}_{f_i}(t) = \mathbf{0}_{3\times 1}\,,\quad i=1,\ldots,N \tag{7}$$

In these expressions, $\boldsymbol{\omega}(t)=\begin{bmatrix}\omega_1(t) & \omega_2(t) & \omega_3(t)\end{bmatrix}^{\mathrm{T}}$ is the rotational velocity of the IMU, expressed in $\{I\}$, ${}^G\mathbf{a}_I(t)$ is the body acceleration expressed in $\{G\}$, and

$$\mathbf{\Omega}(\boldsymbol{\omega}) = \begin{bmatrix} -\lfloor\boldsymbol{\omega}\times\rfloor & \boldsymbol{\omega} \\ -\boldsymbol{\omega}^{\mathrm{T}} & 0 \end{bmatrix},\quad \lfloor\boldsymbol{\omega}\times\rfloor \triangleq \begin{bmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{bmatrix}$$

The gyroscope and accelerometer measurements, $\boldsymbol{\omega}_m$ and $\mathbf{a}_m$, are modeled as

$$\begin{aligned}\boldsymbol{\omega}_m(t) &= \boldsymbol{\omega}(t) + \mathbf{b}_g(t) + \mathbf{n}_g(t) \\ \mathbf{a}_m(t) &= \mathbf{C}({}^I\bar{q}_G(t))\,({}^G\mathbf{a}_I(t) - {}^G\mathbf{g}) + \mathbf{b}_a(t) + \mathbf{n}_a(t)\end{aligned}$$

where $\mathbf{n}_g$ and $\mathbf{n}_a$ are zero-mean, white Gaussian noise processes, and ${}^G\mathbf{g}$ is the gravitational acceleration. The matrix $\mathbf{C}(\bar{q})$ is the rotation matrix corresponding to $\bar{q}$. The observed features belong to a static scene; hence, their time derivatives are zero (see (7)). Linearizing at the current estimates and applying the expectation operator on both sides of (2)–(7), we obtain the state estimate propagation model

$$ {}^I\dot{\hat{\bar{q}}}_G(t) = \frac{1}{2}\boldsymbol{\Omega}(\hat{\boldsymbol{\omega}}(t))\,{}^I\hat{\bar{q}}_G(t) \tag{8} $$

$$ {}^G\dot{\hat{\mathbf{p}}}_I(t) = {}^G\hat{\mathbf{v}}_I(t) \tag{9} $$

$$ {}^G\dot{\hat{\mathbf{v}}}_I(t) = \mathbf{C}^{\mathrm{T}}({}^I\hat{\bar{q}}_G(t))\;\hat{\mathbf{a}}_I(t) + {}^G\mathbf{g} \tag{10} $$

$$ \dot{\hat{\mathbf{b}}}_g(t) = \mathbf{0}_{3\times 1} \tag{11} $$

$$ \dot{\hat{\mathbf{b}}}_a(t) = \mathbf{0}_{3\times 1} \tag{12} $$

$$ {}^G\dot{\hat{\mathbf{p}}}_{f_i}(t) = \mathbf{0}_{3\times 1}\,,\;\; i = 1,\ldots,N \tag{13} $$

where $\hat{\mathbf{a}}_I(t) = \mathbf{a}_m(t) - \hat{\mathbf{b}}_a(t)$, and $\hat{\boldsymbol{\omega}}(t) = \boldsymbol{\omega}_m(t) - \hat{\mathbf{b}}_g(t)$. The $(15+3N)\times 1$ error-state vector is defined as

$$ \begin{aligned} \widetilde{\mathbf{x}} &= \begin{bmatrix} {}^I\boldsymbol{\delta\theta}_G^{\mathrm{T}} & \widetilde{\mathbf{b}}_g^{\mathrm{T}} & {}^G\widetilde{\mathbf{v}}_I^{\mathrm{T}} & \widetilde{\mathbf{b}}_a^{\mathrm{T}} & {}^G\widetilde{\mathbf{p}}_I^{\mathrm{T}} & | & {}^G\widetilde{\mathbf{p}}_{f_1}^{\mathrm{T}} \cdots {}^G\widetilde{\mathbf{p}}_{f_N}^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}} \\ &= \begin{bmatrix} \widetilde{\mathbf{x}}_s^{\mathrm{T}} & | & \widetilde{\mathbf{x}}_m^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}} \end{aligned} $$

where $\widetilde{\mathbf{x}}_s(t)$ is the 15×1 error state corresponding to the sensing platform, and $\widetilde{\mathbf{x}}_m(t)$ is the $3N\times 1$ error state of the map. For the IMU position, velocity, biases, and the map, an additive error model is employed (i.e. $\widetilde{\mathbf{y}} = \mathbf{y} - \hat{\mathbf{y}}$ is the error in the estimate $\hat{\mathbf{y}}$ of a quantity $\mathbf{y}$). However, for the quaternion we employ a multiplicative error model (Trawny and Roumeliotis 2005). Specifically, the error between the quaternion $\bar{q}$ and its estimate $\hat{\bar{q}}$ is the 3×1 angle-error vector, $\boldsymbol{\delta\theta}$, implicitly defined by the error quaternion

$$ \delta\bar{q} = \bar{q}\otimes\hat{\bar{q}}^{-1} \simeq \begin{bmatrix} \frac{1}{2}\boldsymbol{\delta\theta}^{\mathrm{T}} & 1 \end{bmatrix}^{\mathrm{T}} $$

where $\delta\bar{q}$ describes the small rotation that causes the true and estimated attitude to coincide. This allows us to represent the attitude uncertainty by the 3×3 covariance matrix $\mathbb{E}[\boldsymbol{\delta\theta}\boldsymbol{\delta\theta}^{\mathrm{T}}]$, which is a minimal representation.

The linearized continuous-time error-state equation is

$$ \begin{aligned} \dot{\widetilde{\mathbf{x}}} &= \begin{bmatrix} \mathbf{F}_s & \mathbf{0}_{15\times 3N} \\ \mathbf{0}_{3N\times 15} & \mathbf{0}_{3N} \end{bmatrix}\widetilde{\mathbf{x}} + \begin{bmatrix} \mathbf{G}_s \\ \mathbf{0}_{3N\times 12} \end{bmatrix}\mathbf{n} \\ &= \mathbf{F}_c\,\widetilde{\mathbf{x}} + \mathbf{G}_c\,\mathbf{n} \end{aligned} $$

where $\mathbf{0}_{3N}$ denotes the $3N\times 3N$ matrix of zeros, $\mathbf{n} = \begin{bmatrix} \mathbf{n}_g^{\mathrm{T}} & \mathbf{n}_{wg}^{\mathrm{T}} & \mathbf{n}_a^{\mathrm{T}} & \mathbf{n}_{wa}^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}$ is the system noise, $\mathbf{F}_s$ is the continuous-time error-state transition matrix corresponding to the sensor platform state, and $\mathbf{G}_s$ is the continuous-time input noise matrix, i.e.

$$ \mathbf{F}_s = \begin{bmatrix} -\lfloor\hat{\boldsymbol{\omega}}\times\rfloor & -\mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ -\mathbf{C}^{\mathrm{T}}({}^I\hat{\bar{q}}_G)\lfloor\hat{\mathbf{a}}_I\times\rfloor & \mathbf{0}_3 & \mathbf{0}_3 & -\mathbf{C}^{\mathrm{T}}({}^I\hat{\bar{q}}_G) & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 \end{bmatrix} $$

$$ \mathbf{G}_s = \begin{bmatrix} -\mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & -\mathbf{C}^{\mathrm{T}}({}^I\hat{\bar{q}}_G) & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{I}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \end{bmatrix} $$

where $\mathbf{0}_3$ is the 3×3 matrix of zeros. The system noise is modelled as a zero-mean white Gaussian process with autocorrelation $\mathbb{E}[\mathbf{n}(t)\,\mathbf{n}^{\mathrm{T}}(\tau)] = \mathbf{Q}_c\delta(t-\tau)$ which depends on the IMU noise characteristics and is computed offline (Trawny and Roumeliotis 2005).

3.1.2. Discrete-time implementation The IMU signals $\boldsymbol{\omega}_m$ and $\mathbf{a}_m$ are sampled at a constant rate $1/\delta t$, where $\delta t \triangleq t_{k+1} - t_k$. Every time a new IMU measurement is received, the state estimate is propagated using integration of (8)–(13). In order to derive the covariance propagation equation, we compute the discrete-time state transition matrix, $\boldsymbol{\Phi}_{k+1,k}$, from time-step t_k to t_{k+1}, as the solution to the following matrix differential equation:

$$ \begin{aligned} \dot{\boldsymbol{\Phi}}_{k+1,k} &= \mathbf{F}_c\boldsymbol{\Phi}_{k+1,k} \\ \text{initial condition} \quad \boldsymbol{\Phi}_{k,k} &= \mathbf{I}_{15+3N} \end{aligned} \tag{14} $$

which can be calculated analytically as we show in Hesch et al. (2012a) or numerically. We also compute the discrete-time system noise covariance matrix, $\mathbf{Q}_{d,k}$,

$$ \mathbf{Q}_{d,k} = \int_{t_k}^{t_{k+1}} \boldsymbol{\Phi}(t_{k+1},\tau)\,\mathbf{G}_c\mathbf{Q}_c\mathbf{G}_c^{\mathrm{T}}\boldsymbol{\Phi}^{\mathrm{T}}(t_{k+1},\tau)\,\mathrm{d}\tau. $$

The propagated covariance is then computed as

$$ \mathbf{P}_{k+1|k} = \boldsymbol{\Phi}_{k+1,k}\mathbf{P}_{k|k}\boldsymbol{\Phi}_{k+1,k}^{\mathrm{T}} + \mathbf{Q}_{d,k}. \tag{15} $$

3.2. Measurement update model

As the camera-IMU platform moves, the camera observes visual features which are tracked over multiple image frames. These measurements are exploited to estimate the motion of the sensing platform and (optionally) the map of the environment.

To simplify the discussion, we consider the observation of a single point $\mathbf{p}_{f_i}$. The camera measures $\mathbf{z}_i$, which is the perspective projection of the 3D point ${}^I\mathbf{p}_{f_i}$ expressed in the current IMU frame $\{I\}$, onto the image plane, i.e.

$$ \mathbf{z}_i = \frac{1}{p_z}\begin{bmatrix} p_x \\ p_y \end{bmatrix} + \eta_i \tag{16} $$

$$ \text{where} \quad \begin{bmatrix} p_x \\ p_y \\ p_z \end{bmatrix} = {}^I\mathbf{p}_{f_i} = \mathbf{C}\left({}^I\bar{q}_G\right)\left({}^G\mathbf{p}_{f_i} - {}^G\mathbf{p}_I\right) \tag{17} $$

where the measurement noise, η_i, is modeled as zero mean, white Gaussian with covariance $\mathbf{R}_i$. We note that, without loss of generality, we consider the image measurement in normalized pixel coordinates, and define the camera frame to be coincident with the IMU. In practice, we perform both intrinsic and extrinsic camera-IMU calibration off-line (Bouguet 2006; Mirzaei and Roumeliotis 2008).

The linearized error model is

$$\tilde{\mathbf{z}}_i = \mathbf{z}_i - \hat{\mathbf{z}}_i \simeq \mathbf{H}_i \tilde{\mathbf{x}} + \eta_i, \tag{18}$$

where $\hat{\mathbf{z}} = \mathbf{h}\left(\hat{\mathbf{x}}\right)$ is the expected measurement computed by evaluating (16)–(17) at the current state estimate, and the measurement Jacobian, $\mathbf{H}_i$, is

$$\mathbf{H}_i = \mathbf{H}_c \begin{bmatrix} \mathbf{H}_\theta & \mathbf{0}_{3\times 9} & \mathbf{H}_p & | & \mathbf{0}_3 \cdots \mathbf{H}_{f_i} \cdots \mathbf{0}_3 \end{bmatrix} \tag{19}$$

where the partial derivatives are

$$\begin{aligned}
\mathbf{H}_c &= \frac{\partial \mathbf{h}}{\partial {}^I\mathbf{p}_{f_i}} = \frac{1}{p_z^2}\begin{bmatrix} p_z & 0 & -p_x \\ 0 & p_z & -p_y \end{bmatrix} \\
\mathbf{H}_\theta &= \frac{\partial {}^I\mathbf{p}_{f_i}}{\partial \theta} = \lfloor \mathbf{C}\left({}^I\bar{q}_G\right)\left({}^G\mathbf{p}_{f_i} - {}^G\mathbf{p}_I\right) \times \rfloor \\
\mathbf{H}_p &= \frac{\partial {}^I\mathbf{p}_{f_i}}{\partial {}^G\mathbf{p}_I} = -\mathbf{C}\left({}^I\bar{q}_G\right) \\
\mathbf{H}_{f_i} &= \frac{\partial {}^I\mathbf{p}_{f_i}}{\partial {}^G\mathbf{p}_{f_i}} = \mathbf{C}\left({}^I\bar{q}_G\right)
\end{aligned}$$

i.e. $\mathbf{H}_c$, is the Jacobian of the perspective projection with respect to ${}^I\mathbf{p}_{f_i}$, while $\mathbf{H}_\theta$, $\mathbf{H}_p$, and $\mathbf{H}_{f_i}$, are the Jacobians of ${}^I\mathbf{p}_{f_i}$ with respect to ${}^I\bar{q}_G$, ${}^G\mathbf{p}_I$, and ${}^G\mathbf{p}_{f_i}$, respectively.

This measurement model is used, independently of whether the map of the environment $\mathbf{x}_m$ is part of the state vector (V-SLAM) or not (VIO). Specifically, for the case of V-SLAM, when features that are already mapped are observed, the measurement model (16)–(19) can be directly applied to update the filter. In particular, we compute the measurement residual,

$$\mathbf{r}_i = \mathbf{z}_i - \hat{\mathbf{z}}_i$$

the covariance of the residual,

$$\mathbf{S}_i = \mathbf{H}_i \mathbf{P}_{k+1|k} \mathbf{H}_i{}^{\mathrm{T}} + \mathbf{R}_i$$

and the Kalman gain,

$$\mathbf{K}_i = \mathbf{P}_{k+1|k} \mathbf{H}_i{}^{\mathrm{T}} \mathbf{S}_i^{-1}$$

Employing these quantities, we compute the EKF state and covariance update as

$$\begin{aligned}
\hat{\mathbf{x}}_{k+1|k+1} &= \hat{\mathbf{x}}_{k+1|k} + \mathbf{K}_i \mathbf{r}_i \\
\mathbf{P}_{k+1|k+1} &= \mathbf{P}_{k+1|k} - \mathbf{K}_i \mathbf{S}_i \mathbf{K}_i^{\mathrm{T}}.
\end{aligned}$$

When features are first observed in V-SLAM, we initialize them into the feature map. To accomplish this, we compute an initial estimate, along with covariance and cross-correlations, by solving a bundle-adjustment over a short time window (Hesch et al. 2012a). Finally, for the case of VIO, the map is not estimated explicitly; instead we use the Multi-State Constraint Kalman Filter (MSC-KF) approach (Mourikis and Roumeliotis 2007) to impose a filter update constraining all the views from which a feature was seen. To accomplish this, we employ stochastic cloning (Roumeliotis and Burdick 2002) over a window of M camera poses.

4. Nonlinear system observability analysis

In this section, we provide a brief overview of the method in Hermann and Krener (1977) for studying the observability of nonlinear systems and then introduce a new methodology for determining its unobservable directions.

4.1. Observability analysis with Lie derivatives

Consider a nonlinear, continuous-time system:

$$\begin{cases} \dot{\mathbf{x}} = \mathbf{f}_0(\mathbf{x}) + \sum_{i=1}^{\ell} \mathbf{f}_i(\mathbf{x})\, u_i \\ \mathbf{z} = \mathbf{h}(\mathbf{x}) \end{cases} \tag{20}$$

where $\mathbf{u} = \begin{bmatrix} u_1 & \dots & u_\ell \end{bmatrix}^{\mathrm{T}}$ is the control input, $\mathbf{x} = \begin{bmatrix} x_1 & \dots & x_m \end{bmatrix}^{\mathrm{T}}$ is the state vector, $\mathbf{z}$ is the output, and the vector functions $\mathbf{f}_i$, $i = 0, \dots, \ell$, comprise the process model.

Our objective is to study the observability properties of the system and to determine the directions in state-space that the measurements provide information. To this end, we compute the Lie derivatives of the system. The zeroth-order Lie derivative of the measurement function $\mathbf{h}$ is defined as the function itself (Hermann and Krener 1977):

$$\mathfrak{L}^0\mathbf{h} = \mathbf{h}(\mathbf{x})$$

Each subsequent Lie derivative is formed recursively from the definition of $\mathfrak{L}^0\mathbf{h}$. Specifically, for any ith-order Lie derivative, $\mathfrak{L}^i\mathbf{h}$, the $(i+1)$th-order Lie derivative $\mathfrak{L}_{\mathbf{f}_j}^{i+1}\mathbf{h}$ with respect to a process function $\mathbf{f}_j$ is computed as:

$$\mathfrak{L}_{\mathbf{f}_j}^{i+1}\mathbf{h} = \nabla\mathfrak{L}^i\mathbf{h} \cdot \mathbf{f}_j$$

where $\nabla\mathfrak{L}^i\mathbf{h}$ denotes the span of the ith-order Lie derivative, i.e.

$$\nabla\mathfrak{L}^i\mathbf{h} = \begin{bmatrix} \frac{\partial \mathfrak{L}^i\mathbf{h}}{\partial x_1} & \frac{\partial \mathfrak{L}^i\mathbf{h}}{\partial x_2} & \dots & \frac{\partial \mathfrak{L}^i\mathbf{h}}{\partial x_m} \end{bmatrix}$$

In order to determine the directions along which information can be acquired, we examine the span of the Lie derivatives. We do this by forming the observability matrix, $\mathcal{O}$, whose block-rows comprise of the spans of the Lie derivatives of the system, i.e.

$$\mathcal{O} = \begin{bmatrix} \nabla\mathfrak{L}^0\mathbf{h} \\ \nabla\mathfrak{L}_{\mathbf{f}_i}^1\mathbf{h} \\ \nabla\mathfrak{L}_{\mathbf{f}_i\mathbf{f}_j}^2\mathbf{h} \\ \nabla\mathfrak{L}_{\mathbf{f}_i\mathbf{f}_j\mathbf{f}_k}^3\mathbf{h} \\ \vdots \end{bmatrix}$$

where $i,j,k = 1,\ldots,\ell$. Based on Hermann and Krener (1977), to prove that a system is observable, it suffices to show that a submatrix of $\mathcal{O}$ comprising a subset of its rows is of full column rank. In contrast, to prove that a system is unobservable and find its unobservable directions, we need to: (a) Show that the infinitely many block rows of $\mathcal{O}$ can be written as a linear combination of a subset of its block rows, which form a submatrix $\mathcal{O}'$; (b) find the nullspace of $\mathcal{O}'$ in order to determine the system's unobservable directions. Although accomplishing (b) is fairly straightforward, achieving (a) is extremely challenging especially for high-dimensional systems, such as the one describing VINS.

To address this issue, in the following section, we present a new methodology that relies on a change of variables for proving that a system is unobservable and finding its unobservable directions.

4.2. Observability analysis with basis functions

In order to gain intuition for the following derivations, we provide a brief overview of the motivation for this methodology. As stated in the previous section, following the approach of Hermann and Krener (1977) for analyzing the observability properties of a nonlinear system is quite challenging. The main issue is that we must analytically compute the nullspace of a matrix with an infinite number of rows (since there are infinitely many Lie derivatives). However, our analysis can be significantly simplified if we can find a process for decomposing the observability matrix into a product of two matrices: (a) a full-rank matrix with infinitely many rows, and (b) a rank-deficient matrix with only a limited number of rows. In what follows, we show how to achieve such a factorization of the observability matrix, by computing a set of basis functions of the state, which comprise its observable modes.

We start by proving the following:

Theorem 4.1: Assume that there exists a nonlinear transformation $\boldsymbol{\beta}(\mathbf{x}) = \begin{bmatrix}\boldsymbol{\beta}_1(\mathbf{x})^\mathrm{T} & \ldots & \boldsymbol{\beta}_t(\mathbf{x})^\mathrm{T}\end{bmatrix}^\mathrm{T}$. These bases are functions of the variable $\mathbf{x}$ in (20), and the number of basis elements, t, is defined so as to fulfill:

(C1) $\boldsymbol{\beta}_1(\mathbf{x}) = \mathbf{h}(\mathbf{x})$.
(C2) $\frac{\partial \boldsymbol{\beta}}{\partial \mathbf{x}} \cdot \mathbf{f}_i, i = 0,\ldots,\ell$ is a function of $\boldsymbol{\beta}$.
(C3) The system:

$$\begin{cases} \dot{\boldsymbol{\beta}} = \mathbf{g}_0(\boldsymbol{\beta}) + \sum_{i=1}^{\ell} \mathbf{g}_i(\boldsymbol{\beta})\, u_i \\ \mathbf{z} = \mathbf{h} = \boldsymbol{\beta}_1 \end{cases} \tag{21}$$

where $\mathbf{g}_i(\boldsymbol{\beta}) = \frac{\partial \boldsymbol{\beta}}{\partial \mathbf{x}} \mathbf{f}_i(\mathbf{x}), i = 0,\ldots,\ell$, is observable.

Then:

(i) The observability matrix of (20) can be factorized as:

$$\mathcal{O} = \Xi \cdot \mathbf{B}$$

where Ξ is the observability matrix of system (21) and $\mathbf{B} \triangleq \frac{\partial \boldsymbol{\beta}}{\partial \mathbf{x}}$.

(ii) $null(\mathcal{O}) = null(\mathbf{B})$.

Proof:

(i) Based on the chain rule, the span of any Lie derivative $\nabla \mathfrak{L}^i \mathbf{h}$ can be written as:

$$\nabla \mathfrak{L}^i \mathbf{h} = \frac{\partial \mathfrak{L}^i \mathbf{h}}{\partial \mathbf{x}} = \frac{\partial \mathfrak{L}^i \mathbf{h}}{\partial \boldsymbol{\beta}} \frac{\partial \boldsymbol{\beta}}{\partial \mathbf{x}}$$

Thus, the observability matrix $\mathcal{O}$ of (20) can be factorized as:

$$\mathcal{O} = \begin{bmatrix} \nabla \mathfrak{L}^0 \mathbf{h} \\ \nabla \mathfrak{L}^1_{\mathbf{f}_i} \mathbf{h} \\ \nabla \mathfrak{L}^2_{\mathbf{f}_i \mathbf{f}_j} \mathbf{h} \\ \nabla \mathfrak{L}^3_{\mathbf{f}_i \mathbf{f}_j \mathbf{f}_k} \mathbf{h} \\ \vdots \end{bmatrix} = \begin{bmatrix} \frac{\partial \mathfrak{L}^0 \mathbf{h}}{\partial \boldsymbol{\beta}} \\ \frac{\partial \mathfrak{L}^1_{\mathbf{f}_i} \mathbf{h}}{\partial \boldsymbol{\beta}} \\ \frac{\partial \mathfrak{L}^2_{\mathbf{f}_i \mathbf{f}_j} \mathbf{h}}{\partial \boldsymbol{\beta}} \\ \frac{\partial \mathfrak{L}^3_{\mathbf{f}_i \mathbf{f}_j \mathbf{f}_k} \mathbf{h}}{\partial \boldsymbol{\beta}} \\ \vdots \end{bmatrix} \frac{\partial \boldsymbol{\beta}}{\partial \mathbf{x}} = \Xi \cdot \mathbf{B} \tag{22}$$

Next we prove that Ξ is the observability matrix of the system (21) by induction.

To distinguish the Lie derivatives of system (20), let $\mathfrak{I}$ denote the Lie derivatives of system (21). Then, the span of its zeroth-order Lie derivative is:

$$\nabla \mathfrak{I}^0 \mathbf{h} = \frac{\partial \mathbf{h}}{\partial \boldsymbol{\beta}} = \frac{\partial \mathfrak{L}^0 \mathbf{h}}{\partial \boldsymbol{\beta}}$$

which corresponds to the first block row of Ξ.

Assume that the span of the ith-order Lie derivative of (21) along any direction can be written as $\nabla \mathfrak{I}^i \mathbf{h} = \frac{\partial \mathfrak{L}^i \mathbf{h}}{\partial \boldsymbol{\beta}}$, which corresponds to a block row of Ξ. Then the span of the $(i+1)$th-order Lie derivative $\nabla \mathfrak{I}^{i+1}_{\mathbf{g}_j} \mathbf{h}$ along the process function $\mathbf{g}_j$ can be computed as:

$$\nabla \mathfrak{I}^{i+1}_{\mathbf{g}_j} \mathbf{h} = \frac{\partial \mathfrak{I}^{i+1}_{\mathbf{g}_j} \mathbf{h}}{\partial \boldsymbol{\beta}} = \frac{\partial (\nabla \mathfrak{I}^i \mathbf{h} \cdot \mathbf{g}_j)}{\partial \boldsymbol{\beta}} = \frac{\partial (\frac{\partial \mathfrak{L}^i \mathbf{h}}{\partial \boldsymbol{\beta}} \cdot \frac{\partial \boldsymbol{\beta}}{\partial \mathbf{x}} \mathbf{f}_j(\mathbf{x}))}{\partial \boldsymbol{\beta}}$$
$$= \frac{\partial (\frac{\partial \mathfrak{L}^i \mathbf{h}}{\partial \mathbf{x}} \cdot \mathbf{f}_j(\mathbf{x}))}{\partial \boldsymbol{\beta}} = \frac{\partial \mathfrak{L}^{i+1}_{\mathbf{f}_j} \mathbf{h}}{\partial \boldsymbol{\beta}}$$

which is also a block row of Ξ. Therefore, we conclude that Ξ is a matrix whose rows are the span of all the Lie derivatives of system (21), and thus it is the observability matrix of system (21). ■

(ii) From $\mathcal{O} = \Xi \mathbf{B}$, we have $null(\mathcal{O}) = null(\mathbf{B}) + null(\Xi) \cap range(\mathbf{B})$ (see Meyer (2000, Section 4.5.1)). Moreover, from condition C3 system (21) is observable, and Ξ is of full column rank. Therefore $null(\mathcal{O}) = null(\mathbf{B})$. ■

Based on Theorem 4.1, the unobservable directions can be determined with significantly less effort. Specifically, to find a system's unobservable directions, we first need to define the basis functions that satisfy conditions C1 and C2, and prove that matrix Ξ is of full column rank, which is condition C3. Once all the conditions are satisfied, the unobservable directions of (20) correspond to the nullspace

of matrix $\mathbf{B}$, which has finite dimensions and thus is fairly easy to find.

In the following sections, we will leverage Theorem 4.1 to prove that the VINS model is unobservable and find its unobservable directions. To do this, in Section 5.1 we first review the VINS model. In Section 5.2, we find the set of basis functions that satisfy conditions C1 and C2 of Theorem 4.1, and construct the basis functions' system as in (21) for this particular problem. In Section 5.3, we prove that the observability matrix Ξ for the basis functions' system is of full column rank, which is condition C3 of Theorem 4.1. Lastly, we determine the unobservable directions of the VINS model by finding the nullspace of matrix $\mathbf{B}$.

5. Observability analysis of the VINS model

In this section, we present the observability analysis for the VINS model using basis functions.

5.1. Revisiting the system model

For the purpose of simplifying the observability analysis, we express the IMU orientation using the Cayley–Gibbs–Rodriguez (CGR) parameterization (Shuster 1993), which is a minimal representation. Specifically, the orientation of $\{G\}$ with respect to $\{I\}$ is the 3×1 vector of CGR parameters, ${}^I\mathbf{s}_G$. Hence, we rewrite (1) as

$$\mathbf{x} = \begin{bmatrix} {}^I\mathbf{s}_G^{\mathrm{T}} & \mathbf{b}_g^{\mathrm{T}} & {}^G\mathbf{v}_I^{\mathrm{T}} & \mathbf{b}_a^{\mathrm{T}} & {}^G\mathbf{p}_I^{\mathrm{T}} & {}^G\mathbf{p}_f^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}$$

The time evolution of ${}^I\mathbf{s}_G$ is

$${}^I\dot{\mathbf{s}}_G(t) = \mathbf{D}\left({}^I\boldsymbol{\omega}(t) - \mathbf{b}_g(t)\right) \tag{23}$$

$$\text{where } \mathbf{D} \triangleq \frac{\partial \mathbf{s}}{\partial \boldsymbol{\theta}} = \frac{1}{2}\left(\mathbf{I}_3 + \lfloor \mathbf{s} \times \rfloor + \mathbf{s}\mathbf{s}^{\mathrm{T}}\right) \tag{24}$$

5.2. Determining the system's basis functions

In this section, we define the basis functions for the VINS model that satisfy conditions C1 and C2 of Theorem 4.1. We achieve this by applying C1 to obtain $\boldsymbol{\beta}_1$ and recursively employing C2 to define the additional elements $\boldsymbol{\beta}_j$, $j = 2, \ldots, 6$. We note that at each step of this process there may be multiple options for selecting $\boldsymbol{\beta}_j$, and we mitigate this by favoring bases that have a meaningful physical interpretation. After determining the bases, we present the model of the corresponding system (43), and show that it is observable in the next section.

To preserve the clarity of presentation, we retain only a few of the subscripts and superscripts in the state elements and write the system state vector as:

$$\mathbf{x} = \begin{bmatrix} \mathbf{s}^{\mathrm{T}} & \mathbf{b}_g^{\mathrm{T}} & \mathbf{v}^{\mathrm{T}} & \mathbf{b}_a^{\mathrm{T}} & \mathbf{p}^{\mathrm{T}} & \mathbf{p}_f^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}$$

The VINS model (see (2)–(7), (16)–(17), and (23)) is expressed in input-affine form as:

$$\begin{bmatrix} \dot{\mathbf{s}} \\ \dot{\mathbf{b}}_g \\ \dot{\mathbf{v}} \\ \dot{\mathbf{b}}_a \\ \dot{\mathbf{p}} \\ \dot{\mathbf{p}}_f \end{bmatrix} = \underbrace{\begin{bmatrix} -\mathbf{D}\,\mathbf{b}_g \\ \mathbf{0}_{3\times 1} \\ \mathbf{g} - \mathbf{C}^T\mathbf{b}_a \\ \mathbf{0}_{3\times 1} \\ \mathbf{v} \\ \mathbf{0}_{3\times 1} \end{bmatrix}}_{f_0} + \underbrace{\begin{bmatrix} \mathbf{D} \\ \mathbf{0}_3 \\ \mathbf{0}_3 \\ \mathbf{0}_3 \\ \mathbf{0}_3 \\ \mathbf{0}_3 \end{bmatrix}}_{\mathbf{f}_1} \boldsymbol{\omega} + \underbrace{\begin{bmatrix} \mathbf{0}_3 \\ \mathbf{0}_3 \\ \mathbf{C}^T \\ \mathbf{0}_3 \\ \mathbf{0}_3 \\ \mathbf{0}_3 \end{bmatrix}}_{\mathbf{f}_2} \mathbf{a} \tag{25}$$

$$\mathbf{z} = \frac{1}{p_z}\begin{bmatrix} p_x \\ p_y \end{bmatrix}, \quad \text{where } \begin{bmatrix} p_x \\ p_y \\ p_z \end{bmatrix} = {}^I\mathbf{p}_f = \mathbf{C}\left(\mathbf{p}_f - \mathbf{p}\right) \tag{26}$$

and $\mathbf{C} \triangleq \mathbf{C}(\mathbf{s})$. Note that f_0 is an 18×1 vector, while $\mathbf{f}_1$ and $\mathbf{f}_2$ are both 18×3 matrices which is a compact way of representing three process functions:

$$\mathbf{f}_1\boldsymbol{\omega} = f_{11} \cdot \omega_1 + f_{12} \cdot \omega_2 + f_{13} \cdot \omega_3$$
$$\mathbf{f}_2\mathbf{a} = f_{21} \cdot a_1 + f_{22} \cdot a_2 + f_{23} \cdot a_3$$

Using this model, we define the bases for this system by applying the conditions of Theorem 4.1. Specifically, we (a) select $\boldsymbol{\beta}_1$ as the measurement function $\mathbf{z}$, and (b) recursively determine the remaining bases so that $\frac{\partial \boldsymbol{\beta}_j}{\partial \mathbf{x}} \cdot \mathbf{f}_i$ can be expressed in terms of $\boldsymbol{\beta}$ for all the process functions. Note also that the definition of the bases is not unique, any basis functions that satisfy the conditions of Theorem 4.1 span the same space.

The first basis is defined as the measurement function:

$$\boldsymbol{\beta}_1 \triangleq \mathbf{h}(\mathbf{x}) = \frac{1}{p_z}\begin{bmatrix} p_x \\ p_y \end{bmatrix}$$

In order to compute the remaining basis elements, we must ensure that the properties of Theorem 4.1 are satisfied. We do so by applying C2 to $\boldsymbol{\beta}_1$.

5.2.1. Satisfying condition C2 of Theorem 4.1 for $\boldsymbol{\beta}_1$ We start by computing the span of $\boldsymbol{\beta}_1$ with respect to $\mathbf{x}$, i.e.

$$\begin{aligned} \frac{\partial \boldsymbol{\beta}_1}{\partial \mathbf{x}} &= \begin{bmatrix} \frac{\partial \boldsymbol{\beta}_1}{\partial \boldsymbol{\theta}} & \frac{\partial \boldsymbol{\beta}_1}{\partial \mathbf{s}} & \frac{\partial \boldsymbol{\beta}_1}{\partial \mathbf{b}_g} & \frac{\partial \boldsymbol{\beta}_1}{\partial \mathbf{v}} & \frac{\partial \boldsymbol{\beta}_1}{\partial \mathbf{b}_a} & \frac{\partial \boldsymbol{\beta}_1}{\partial \mathbf{p}} & \frac{\partial \boldsymbol{\beta}_1}{\partial \mathbf{p}_f} \end{bmatrix} \\ &= \underbrace{\begin{bmatrix} \frac{1}{p_z} & 0 & -\frac{p_x}{p_z^2} \\ 0 & \frac{1}{p_z} & -\frac{p_y}{p_z^2} \end{bmatrix}}_{\frac{\partial \mathbf{h}}{\partial {}^I\mathbf{p}_f}} \underbrace{\begin{bmatrix} \lfloor {}^I\mathbf{p}_f \times \rfloor \frac{\partial \boldsymbol{\theta}}{\partial \mathbf{s}} & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & -\mathbf{C} & \mathbf{C} \end{bmatrix}}_{\frac{\partial {}^I\mathbf{p}_f}{\partial \mathbf{x}}} \end{aligned} \tag{27}$$

where $\frac{\partial \boldsymbol{\theta}}{\partial \mathbf{s}} = \mathbf{D}^{-1}$ (see (24)). Once the span of the first basis function $\boldsymbol{\beta}_1$ is obtained, we project it onto *all* the process functions, f_0, $\mathbf{f}_1$, and $\mathbf{f}_2$ (see (25)), in order to determine the other basis functions that satisfy condition C2 of Theorem 4.1. During this procedure, our aim is to ensure that every term in the resulting product is a function of the existing basis elements. Whenever a term cannot be expressed

by the previously defined basis functions, we incorporate it as a new basis function.

Specifically, beginning with the projection of $\frac{\partial \boldsymbol{\beta}_1}{\partial \mathbf{x}}$ along f_0 we obtain

$$\begin{aligned}\frac{\partial \boldsymbol{\beta}_1}{\partial \mathbf{x}} \cdot f_0 &= \begin{bmatrix} \frac{1}{p_z} & 0 & -\frac{p_x}{p_z^2} \\ 0 & \frac{1}{p_z} & -\frac{p_y}{p_z^2} \end{bmatrix} \left(-\lfloor {}^I\mathbf{p}_f \times \rfloor \mathbf{b}_g - \mathbf{C}\,\mathbf{v} \right) \\ &= \begin{bmatrix} \mathbf{I}_2 & -\boldsymbol{\beta}_1 \end{bmatrix} \left(-\lfloor \begin{bmatrix} \boldsymbol{\beta}_1 \\ 1 \end{bmatrix} \times \rfloor \mathbf{b}_g - \frac{1}{p_z} \mathbf{C}\,\mathbf{v} \right) \end{aligned} \quad (28)$$

This is a function of β_1 and of other elements of the state $\mathbf{x}$, namely $\mathbf{b}_g$ and $\mathbf{v}$, as well as of functions of $\mathbf{x}$, which are $1/pz$ and $\mathbf{C}$. Hence, in order to satisfy C2, we must define new basis elements, which we select as physically interpretable quantities:

$$\begin{aligned} \boldsymbol{\beta}_2 &\triangleq \frac{1}{p_z} \\ \boldsymbol{\beta}_3 &\triangleq \mathbf{C}\,\mathbf{v} \\ \boldsymbol{\beta}_4 &\triangleq \mathbf{b}_g \end{aligned} \quad (29)$$

where $\boldsymbol{\beta}_2$ is the inverse depth to the point, $\boldsymbol{\beta}_3$ is the velocity expressed in the local frame, and $\boldsymbol{\beta}_4$ is the gyroscope bias. Rewriting (28) using these definitions we have:

$$\frac{\partial \boldsymbol{\beta}_1}{\partial \mathbf{x}} \cdot f_0 \triangleq \begin{bmatrix} \mathbf{I}_2 & -\boldsymbol{\beta}_1 \end{bmatrix} \left(-\lfloor \begin{bmatrix} \boldsymbol{\beta}_1 \\ 1 \end{bmatrix} \times \rfloor \boldsymbol{\beta}_4 - \boldsymbol{\beta}_2\, \boldsymbol{\beta}_3 \right)$$

Note that later on we will need to ensure that the properties of Theorem 4.1 are also satisfied for these new elements, $\boldsymbol{\beta}_2$, $\boldsymbol{\beta}_3$, and $\boldsymbol{\beta}_4$, but first we examine the projections of the span of $\boldsymbol{\beta}_1$ along $\mathbf{f}_1$ and $\mathbf{f}_2$.

The projections of $\frac{\partial \boldsymbol{\beta}_1}{\partial \mathbf{x}}$ along the three directions of $\mathbf{f}_1$ (i.e. $\mathbf{f}_1 \mathbf{e}_i$, $i = 1, 2, 3$, where $\begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \end{bmatrix} = \mathbf{I}_3$) are

$$\begin{aligned}\frac{\partial \boldsymbol{\beta}_1}{\partial \mathbf{x}} \cdot \mathbf{f}_1 \mathbf{e}_i &= \begin{bmatrix} \frac{1}{p_z} & 0 & -\frac{p_x}{p_z^2} \\ 0 & \frac{1}{p_z} & -\frac{p_y}{p_z^2} \end{bmatrix} \lfloor {}^I\mathbf{p}_f \times \rfloor \mathbf{e}_i \\ &= \begin{bmatrix} \mathbf{I}_2 & -\boldsymbol{\beta}_1 \end{bmatrix} \lfloor \begin{bmatrix} \boldsymbol{\beta}_1 \\ 1 \end{bmatrix} \times \rfloor \mathbf{e}_i, \quad i = 1, 2, 3 \end{aligned} \quad (30)$$

Note that in this case no new basis functions need to be defined since (30) already satisfies condition C2 of Theorem 4.1. Lastly, the projections of $\frac{\partial \boldsymbol{\beta}_1}{\partial \mathbf{x}}$ along the $\mathbf{f}_2$ directions are

$$\frac{\partial \boldsymbol{\beta}_1}{\partial \mathbf{x}} \cdot \mathbf{f}_2 \mathbf{e}_i = \mathbf{0}_{2\times 1}, \quad i = 1, 2, 3$$

Hence, by adding the new basis elements $\boldsymbol{\beta}_2$, $\boldsymbol{\beta}_3$, and $\boldsymbol{\beta}_4$, we ensure that the properties of Theorem 4.1 are fulfilled for $\boldsymbol{\beta}_1$. To make the newly defined basis functions, $\boldsymbol{\beta}_2$, $\boldsymbol{\beta}_3$, and $\boldsymbol{\beta}_4$, satisfy condition C2, we proceed by projecting their spans on the process functions.

5.2.2. Satisfying condition C2 of Theorem 4.1 for $\boldsymbol{\beta}_2$ The derivative of $\boldsymbol{\beta}_2$ (see (29)) with respect to the state is:

$$\frac{\partial \boldsymbol{\beta}_2}{\partial \mathbf{x}} = -\frac{1}{p_z^2} \mathbf{e}_3^{\mathrm{T}} \begin{bmatrix} \lfloor {}^I\mathbf{p}_f \times \rfloor \frac{\partial \boldsymbol{\theta}}{\partial \mathbf{s}} & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & -\mathbf{C} & \mathbf{C} \end{bmatrix} \quad (31)$$

Projecting (31) along f_0 we obtain

$$\begin{aligned}\frac{\partial \boldsymbol{\beta}_2}{\partial \mathbf{x}} \cdot f_0 &= -\frac{1}{p_z^2} \mathbf{e}_3^{\mathrm{T}} \left(-\lfloor {}^G\mathbf{p}_f \times \rfloor \mathbf{b}_g - \mathbf{C}\,\mathbf{v} \right) \\ &= -\boldsymbol{\beta}_2 \mathbf{e}_3^{\mathrm{T}} \left(-\lfloor \begin{bmatrix} \boldsymbol{\beta}_1 \\ 1 \end{bmatrix} \times \rfloor \boldsymbol{\beta}_4 - \boldsymbol{\beta}_2\, \boldsymbol{\beta}_3 \right) \end{aligned}$$

which is a function of only the currently enumerated basis elements.

We also project $\frac{\partial \boldsymbol{\beta}_2}{\partial \mathbf{x}}$ along the remaining input directions, i.e. $\mathbf{f}_j \mathbf{e}_i$, $j = 1, 2$, $i = 1, 2, 3$.

$$\begin{aligned}\frac{\partial \boldsymbol{\beta}_2}{\partial \mathbf{x}} \cdot \mathbf{f}_1 \mathbf{e}_i &= -\frac{1}{p_z^2} \mathbf{e}_3^{\mathrm{T}} \lfloor {}^G\mathbf{p}_f \times \rfloor \mathbf{e}_i \\ &= -\boldsymbol{\beta}_2 \mathbf{e}_3^{\mathrm{T}} \lfloor \begin{bmatrix} \boldsymbol{\beta}_1 \\ 1 \end{bmatrix} \times \rfloor \mathbf{e}_i, \quad i = 1, 2, 3 \end{aligned} \quad (32)$$

$$\frac{\partial \boldsymbol{\beta}_2}{\partial \mathbf{x}} \cdot \mathbf{f}_2 \mathbf{e}_i = 0, \quad i = 1, 2, 3$$

which does not admit any new basis elements. Thus, we see that $\boldsymbol{\beta}_2$ fulfills the properties of Theorem 4.1 without requiring us to define any new basis elements.

5.2.3. Satisfying condition C2 of Theorem 4.1 for $\boldsymbol{\beta}_3$ Following the same procedure again, we compute the span of $\boldsymbol{\beta}_3$ with respect to $\mathbf{x}$:

$$\frac{\partial \boldsymbol{\beta}_3}{\partial \mathbf{x}} = \begin{bmatrix} \lfloor \mathbf{C}\,\mathbf{v} \times \rfloor \frac{\partial \boldsymbol{\theta}}{\partial \mathbf{s}} & \mathbf{0}_3 & \mathbf{C} & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \end{bmatrix} \quad (33)$$

and then the projection of (33) along the input direction f_0

$$\begin{aligned}\frac{\partial \boldsymbol{\beta}_3}{\partial \mathbf{x}} \cdot f_0 &= -\lfloor \mathbf{C}\,\mathbf{v} \times \rfloor \mathbf{b}_g + \mathbf{C}\,\mathbf{g} - \mathbf{b}_a \\ &\triangleq -\lfloor \boldsymbol{\beta}_3 \times \rfloor \boldsymbol{\beta}_4 + \boldsymbol{\beta}_5 - \boldsymbol{\beta}_6 \end{aligned}$$

where we assign two new basis elements, i.e.

$$\begin{aligned} \boldsymbol{\beta}_5 &\triangleq \mathbf{C}\,\mathbf{g} \\ \boldsymbol{\beta}_6 &\triangleq \mathbf{b}_a \end{aligned}$$

Note again that we selected physically interpretable functions: (a) $\boldsymbol{\beta}_5$ is the gravity vector expressed in the local frame, and (b) $\boldsymbol{\beta}_6$ is the accelerometer bias. The projections of (33) along $\mathbf{f}_j \mathbf{e}_i$, $j = 1, 2$, $i = 1, 2, 3$, are

$$\begin{aligned}\frac{\partial \boldsymbol{\beta}_3}{\partial \mathbf{x}} \cdot \mathbf{f}_1 \mathbf{e}_i &= \lfloor \mathbf{C}\,\mathbf{v} \times \rfloor \mathbf{e}_i = \lfloor \boldsymbol{\beta}_3 \times \rfloor \mathbf{e}_i, \quad i = 1, 2, 3 \\ \frac{\partial \boldsymbol{\beta}_3}{\partial \mathbf{x}} \cdot \mathbf{f}_2 \mathbf{e}_i &= \mathbf{I}_3 \mathbf{e}_i = \mathbf{e}_i, \quad i = 1, 2, 3 \end{aligned}$$

which do not produce additional bases.

5.2.4. Satisfying condition C2 of Theorem 4.1 for $\boldsymbol{\beta}_4$ We proceed by examining the span of $\boldsymbol{\beta}_4$ with respect to $\mathbf{x}$, i.e.

$$\frac{\partial \boldsymbol{\beta}_4}{\partial \mathbf{x}} = \begin{bmatrix} \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \end{bmatrix} \quad (34)$$

with corresponding projections

$$\frac{\partial \boldsymbol{\beta}_4}{\partial \mathbf{x}} \cdot f_0 = \mathbf{0}_{3\times 1}$$
$$\frac{\partial \boldsymbol{\beta}_4}{\partial \mathbf{x}} \cdot \mathbf{f}_j \mathbf{e}_i = \mathbf{0}_{3\times 1}, \quad j = 1,2,\ i = 1,2,3$$

We note here that no additional basis elements are produced.

5.2.5. Satisfying condition C2 of Theorem 4.1 for $\boldsymbol{\beta}_5$ The derivative of $\boldsymbol{\beta}_5$ with respect to $\mathbf{x}$ is:

$$\frac{\partial \boldsymbol{\beta}_5}{\partial \mathbf{x}} = \begin{bmatrix} \lfloor \mathbf{C}\,\mathbf{g} \times \rfloor \frac{\partial \boldsymbol{\theta}}{\partial \mathbf{s}} & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \end{bmatrix} \tag{35}$$

Projecting (35) along the input directions, we obtain

$$\frac{\partial \boldsymbol{\beta}_5}{\partial \mathbf{x}} \cdot f_0 = -\lfloor \mathbf{C}\,\mathbf{g} \times \rfloor \mathbf{b}_g = -\lfloor \boldsymbol{\beta}_5 \times \rfloor \boldsymbol{\beta}_4$$
$$\frac{\partial \boldsymbol{\beta}_5}{\partial \mathbf{x}} \cdot \mathbf{f}_1 \mathbf{e}_i = \lfloor \mathbf{C}\,\mathbf{g} \times \rfloor \mathbf{e}_i = \lfloor \boldsymbol{\beta}_5 \times \rfloor \mathbf{e}_i, \quad i = 1,2,3$$
$$\frac{\partial \boldsymbol{\beta}_5}{\partial \mathbf{x}} \cdot \mathbf{f}_2 \mathbf{e}_i = \mathbf{0}_{3\times 1}, \quad i = 1,2,3$$

All of these are either a function of the existing basis elements, or are equal to zero, and thus we do not need to define any additional bases.

5.2.6. Satisfying condition C2 of Theorem 4.1 for $\boldsymbol{\beta}_6$ Lastly, we examine the span of the remaining basis element $\boldsymbol{\beta}_6$, i.e.

$$\frac{\partial \boldsymbol{\beta}_6}{\partial \mathbf{x}} = \begin{bmatrix} \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 \end{bmatrix} \tag{36}$$

The projections of (36) along the input directions are

$$\frac{\partial \boldsymbol{\beta}_6}{\partial \mathbf{x}} \cdot f_0 = \mathbf{0}_{3\times 1}$$
$$\frac{\partial \boldsymbol{\beta}_6}{\partial \mathbf{x}} \cdot \mathbf{f}_j \mathbf{e}_i = \mathbf{0}_{3\times 1}, \quad j = 1,2,\ i = 1,2,3$$

which do not produce any additional basis elements.

At this point, we have proved that the conditions C1 and C2 of Theorem 4.1 are satisfied for all of the basis elements; hence, we have defined a complete basis set for the VINS model:

$$\boldsymbol{\beta}_1 = \mathbf{h}(\mathbf{x}) \tag{37}$$
$$\boldsymbol{\beta}_2 = \frac{1}{p_z} \tag{38}$$
$$\boldsymbol{\beta}_3 = \mathbf{C}\,\mathbf{v} \tag{39}$$
$$\boldsymbol{\beta}_4 = \mathbf{b}_g \tag{40}$$
$$\boldsymbol{\beta}_5 = \mathbf{C}\,\mathbf{g} \tag{41}$$
$$\boldsymbol{\beta}_6 = \mathbf{b}_a \tag{42}$$

These correspond to the landmark projection on the image plane (37), the inverse depth to the landmark (38), the velocity expressed in the local frame (39), the gyro bias (40), the gravity vector expressed in the local frame (41), and the accelerometer bias (42). Based on Theorem 4.1, the resulting system in the basis functions (see (21)) is:

$$\begin{bmatrix} \dot{\boldsymbol{\beta}}_1 \\ \dot{\boldsymbol{\beta}}_2 \\ \dot{\boldsymbol{\beta}}_3 \\ \dot{\boldsymbol{\beta}}_4 \\ \dot{\boldsymbol{\beta}}_5 \\ \dot{\boldsymbol{\beta}}_6 \end{bmatrix} = \underbrace{\begin{bmatrix} \bar{\bar{\boldsymbol{\beta}}}_1 \left(-\lfloor \bar{\boldsymbol{\beta}}_1 \times \rfloor \boldsymbol{\beta}_4 - \boldsymbol{\beta}_2 \boldsymbol{\beta}_3 \right) \\ \boldsymbol{\beta}_2 \mathbf{e}_3^{\mathrm{T}} \left(\lfloor \bar{\boldsymbol{\beta}}_1 \times \rfloor \boldsymbol{\beta}_4 + \boldsymbol{\beta}_2 \boldsymbol{\beta}_3 \right) \\ -\lfloor \boldsymbol{\beta}_3 \times \rfloor \boldsymbol{\beta}_4 + \boldsymbol{\beta}_5 - \boldsymbol{\beta}_6 \\ \mathbf{0}_{3\times 1} \\ -\lfloor \boldsymbol{\beta}_5 \times \rfloor \boldsymbol{\beta}_4 \\ \mathbf{0}_{3\times 1} \end{bmatrix}}_{g_0} + \underbrace{\begin{bmatrix} \bar{\bar{\boldsymbol{\beta}}}_1 \lfloor \bar{\boldsymbol{\beta}}_1 \times \rfloor \\ -\boldsymbol{\beta}_2 \mathbf{e}_3^{\mathrm{T}} \lfloor \bar{\boldsymbol{\beta}}_1 \times \rfloor \\ \lfloor \boldsymbol{\beta}_3 \times \rfloor \\ \mathbf{0}_3 \\ \lfloor \boldsymbol{\beta}_5 \times \rfloor \\ \mathbf{0}_3 \end{bmatrix}}_{g_1} \boldsymbol{\omega} + \underbrace{\begin{bmatrix} \mathbf{0}_{2\times 3} \\ \mathbf{0}_{1\times 3} \\ \mathbf{I}_3 \\ \mathbf{0}_3 \\ \mathbf{0}_3 \\ \mathbf{0}_3 \end{bmatrix}}_{g_2} \mathbf{a}$$
$$\mathbf{y} = \boldsymbol{\beta}_1, \tag{43}$$

where $\bar{\boldsymbol{\beta}}_1 = \begin{bmatrix} \boldsymbol{\beta}_1^{\mathrm{T}} & 1 \end{bmatrix}^{\mathrm{T}}$ denotes $\boldsymbol{\beta}_1$ expressed as a 3×1 homogeneous vector, and $\bar{\bar{\boldsymbol{\beta}}}_1 = \begin{bmatrix} \mathbf{I}_2 & -\boldsymbol{\beta}_1 \end{bmatrix}$. In the next section, we will show that system (43) is observable by proving its observability matrix Ξ is of full column rank. Therefore, the basis functions $\boldsymbol{\beta}_1$ to $\boldsymbol{\beta}_6$ correspond to the observable modes of system (25)–(26), and the system model (43) governs the time evolution of the observable state.

5.3. Determining the system's observability matrix and its unobservable directions

Based on Theorem 4.1, the observability matrix $\mathcal{O}$ of the VINS model (see (25)) is the product of the observability matrix Ξ of system (43) with the matrix $\mathbf{B}$ comprising the derivatives of the basis functions. In what follows, we first prove that matrix Ξ is of full column rank. Then, we find the nullspace of matrix $\mathbf{B}$, which according to Theorem 4.1 corresponds to the unobservable directions of the VINS model.

Lemma 5.1: System (43) is observable.

Proof: See Appendix B.

Since system (43) is observable, based on Theorem 4.1, we can find the unobservable directions of system (25) from the nullspace of matrix $\mathbf{B}$.

Theorem 5.2: The VINS model (25) is unobservable, and its unobservable sub-space is spanned by four directions (see (45)) corresponding to the IMU-camera global position and its rotation around the gravity vector in the global frame.

Proof: System (43) satisfies the conditions of Theorem 4.1. Therefore, $null(\mathcal{O}) = null(\mathbf{B})$, which spans the unobservable subspace of the original system (25). Stacking the derivatives of the basis functions with respect to the variable $\mathbf{x}$, the matrix $\mathbf{B}$ can be written as (see (27), (31), (33), (34), (35), and (36)):

$$\mathbf{B} = \underbrace{\begin{bmatrix} \boldsymbol{\zeta} & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{I}_3 \end{bmatrix}}_{\mathbf{B}_1} \underbrace{\begin{bmatrix} \lfloor {}^I\mathbf{p}_f \times \rfloor \frac{\partial \boldsymbol{\theta}}{\partial \mathbf{s}} & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & -\mathbf{C} & \mathbf{C} \\ \lfloor \mathbf{C}\,\mathbf{v} \times \rfloor \frac{\partial \boldsymbol{\theta}}{\partial \mathbf{s}} & \mathbf{0}_3 & \mathbf{C} & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \lfloor \mathbf{C}\,\mathbf{g} \times \rfloor \frac{\partial \boldsymbol{\theta}}{\partial \mathbf{s}} & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 \end{bmatrix}}_{\mathbf{B}_2} \tag{44}$$

where we have factorized $\mathbf{B} = \mathbf{B}_1\mathbf{B}_2$ to further simplify the proof, and, for conciseness, we have denoted the first subblock of $\mathbf{B}_1$ as

$$\boldsymbol{\zeta} = \begin{bmatrix} \frac{1}{p_z} & 0 & -\frac{p_x}{p_z^2} \\ 0 & \frac{1}{p_z} & -\frac{p_y}{p_z^2} \\ 0 & 0 & -\frac{f}{p_z^2} \end{bmatrix}$$

It is easy to verify that $\mathbf{B}_1$ is full rank, since it is comprised of block-diagonal identity matrices, as well as the 3×3 upper-triangular matrix $\boldsymbol{\zeta}$, which is itself full rank (since $\frac{1}{p_z} \neq 0$). Hence, we can study the unobservable modes of VINS by examining the right nullspace of $\mathbf{B}_2$.

The 15×18 matrix $\mathbf{B}_2$ is rank deficient by exactly four, and these four unobservable modes are spanned by the columns of the following matrix

$$\mathbf{N} = \begin{bmatrix} \mathbf{0}_3 & \frac{\partial \mathbf{s}}{\partial \boldsymbol{\theta}}\mathbf{C}\,\mathbf{g} \\ \mathbf{0}_3 & \mathbf{0}_{3\times 1} \\ \mathbf{0}_3 & -\lfloor \mathbf{v} \times \rfloor \mathbf{g} \\ \mathbf{0}_3 & \mathbf{0}_{3\times 1} \\ \mathbf{I}_3 & -\lfloor \mathbf{p} \times \rfloor \mathbf{g} \\ \mathbf{I}_3 & -\lfloor \mathbf{p}_f \times \rfloor \mathbf{g} \end{bmatrix} \tag{45}$$

By multiplying $\mathbf{B}_2$ from the right with $\mathbf{N}$, it is straightforward to verify that $\mathbf{N}$ is indeed the right nullspace of $\mathbf{B}_2$ (see (44) and (45)). We note that the first three columns of $\mathbf{N}$ correspond to globally translating the feature and the IMU-camera sensor pair together, while the fourth column corresponds to global rotations about the gravity vector.

We further prove that there are no additional right nullspace directions by showing that the 15×18 matrix $\mathbf{B}_2$ has rank 14 (note that if $\mathbf{B}_2$ had five or more right nullspace directions, then it would be of rank 13 or less). To do so, we examine the left nullspace of $\mathbf{B}_2$. Specifically, we postulate that $\mathbf{B}_2$ has a left nullspace comprising the block elements $\mathbf{M}_1, \ldots, \mathbf{M}_5$, i.e.

$$\mathbf{0} = [\mathbf{M}_1 \quad \mathbf{M}_2 \quad \mathbf{M}_3 \quad \mathbf{M}_4 \quad \mathbf{M}_5] \begin{bmatrix} \lfloor {}^I\mathbf{p}_f \times \rfloor \frac{\partial \boldsymbol{\theta}}{\partial \mathbf{s}} & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & -\mathbf{C} & \mathbf{C} \\ \lfloor \mathbf{C}\,\mathbf{v} \times \rfloor \frac{\partial \boldsymbol{\theta}}{\partial \mathbf{s}} & \mathbf{0}_3 & \mathbf{C} & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \lfloor \mathbf{C}\,\mathbf{g} \times \rfloor \frac{\partial \boldsymbol{\theta}}{\partial \mathbf{s}} & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 \end{bmatrix}$$

Based on the relationships involving the second and fourth block columns of $\mathbf{B}_2$, we see that $\mathbf{M}_3\mathbf{I}_3 = \mathbf{0}$ and $\mathbf{M}_5\mathbf{I}_3 = \mathbf{0}$, which can only hold if both $\mathbf{M}_3$ and $\mathbf{M}_5$ are zero. From the third and sixth columns of $\mathbf{B}_2$ we see that $\mathbf{M}_2\mathbf{C} = \mathbf{0}$ and $\mathbf{M}_1\mathbf{C} = \mathbf{0}$, which again can only hold if $\mathbf{M}_1$ and $\mathbf{M}_2$ are zero, since the rotation matrix $\mathbf{C}$ is full rank. Thus far, the only potentially nonzero element in the left nullspace of $\mathbf{B}_2$ is $\mathbf{M}_4$. By writing the relationship involving the first block column of $\mathbf{B}_2$, we obtain

$$\mathbf{M}_4 \lfloor \mathbf{C}\,\mathbf{g} \times \rfloor \frac{\partial \boldsymbol{\theta}}{\partial \mathbf{s}} = \mathbf{0}$$

The matrix $\frac{\partial \boldsymbol{\theta}}{\partial \mathbf{s}}$ is full rank, hence, the only nonzero $\mathbf{M}_4$ which can satisfy this relationship is

$$\mathbf{M}_4 = \pm (\mathbf{C}\,\mathbf{g})^{\mathrm{T}}$$

Therefore, we conclude that $\mathbf{B}_2$ has a one dimensional left nullspace, i.e.

$$\mathbf{M} = \begin{bmatrix} \mathbf{0}_{1\times 3} & \mathbf{0}_{1\times 3} & \mathbf{0}_{1\times 3} & (\mathbf{C}\,\mathbf{g})^{\mathrm{T}} & \mathbf{0}_{1\times 3} \end{bmatrix} \tag{46}$$

Since $\mathbf{B}_2$ is a matrix of dimensions 15×18 with exactly one left null vector (see (46)), it is of rank 14. Applying this fact to determine the dimension of the right nullspace, we see that the right nullspace comprises $18 - 14 = 4$ directions, which are spanned by $\mathbf{N}$ [see (45)]. ■

6. Observability-constrained VINS

Whereas in the previous section we concerned ourselves with the properties of the underlying nonlinear system, here we analyze the observability properties of the linearized system employed for estimation purposes. When using a linearized estimator, such as the EKF, errors in linearization while evaluating the system and measurement Jacobians change the directions in which information is acquired by the estimator.[2] We postulate that if this information mistakingly lies along unobservable directions, it can lead to larger errors, erroneously smaller uncertainties, and inconsistency. We first analyze this issue, and subsequently, present an Observability-Constrained VINS (OC-VINS) that explicitly adheres to the observability properties of VINS.

The observability matrix (Maybeck 1979) is defined as a function of the linearized measurement model, $\mathbf{H}$, and the discrete-time state transition matrix, Φ, which are in turn functions of the linearization point, $\mathbf{x}$, i.e.

$$\mathbf{M}(\mathbf{x}) = \begin{bmatrix} \mathbf{H}_1 \\ \mathbf{H}_2\Phi_{2,1} \\ \vdots \\ \mathbf{H}_k\Phi_{k,1} \end{bmatrix} \tag{47}$$

where $\Phi_{k,1} \triangleq \Phi_{k-1,k-2}\cdots\Phi_{2,1}$ is the state transition matrix from time step 1 to k. In Hesch et al. (2012a), we show that $\Phi_{k+1,k}$ (see (14)) has the following block structure:

$$\Phi_{k+1,k} = \begin{bmatrix} \Phi^R_{k+1,k} & \mathbf{0}_{15\times 3N} \\ \mathbf{0}_{3N\times 15} & \mathbf{I}_{3N} \end{bmatrix} \tag{48}$$

$$\text{with } \Phi^R_{k+1,k} = \begin{bmatrix} \Phi_{11} & \Phi_{12} & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \Phi_{31} & \Phi_{32} & \mathbf{I}_3 & \Phi_{34} & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_3 \\ \Phi_{51} & \Phi_{52} & \delta t\mathbf{I}_3 & \Phi_{54} & \mathbf{I}_3 \end{bmatrix}$$

To simplify the discussion, we consider a single landmark in the state vector (i.e. $N = 1$), and write the first block row of $\mathbf{M}(\mathbf{x})$ as (see (19))

$$\mathbf{H}_1 = \mathbf{H}_{c_1}\mathbf{C}\left({}^{I_1}\bar{q}_G\right) \cdot \begin{bmatrix} \lfloor {}^G\mathbf{p}_f - {}^G\mathbf{p}_{I_1} \times \rfloor \mathbf{C}\left({}^{I_1}\bar{q}_G\right)^{\mathrm{T}} & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & -\mathbf{I}_3 & \mathbf{I}_3 \end{bmatrix}$$

where ${}^{I_1}\bar{q}_G$, denotes the rotation of $\{G\}$ with respect to frame $\{I\}$ at time step 1, and for the purposes of the observability analysis, all the quantities appearing in the previous expression are the true ones. As shown in Hesch et al. (2012a), the kth block row, for $k > 1$, is of the form:

$$\mathbf{H}_k\Phi_{k,1} = \mathbf{H}_{c_k}\mathbf{C}\left({}^{I}\bar{q}_{G_k}\right)\begin{bmatrix}\boldsymbol{\Upsilon}_k & \mathbf{D}_k & -\mathbf{I}_3\delta t_{k-1} & \mathbf{E}_k & -\mathbf{I}_3 & \mathbf{I}_3\end{bmatrix} \quad (49)$$

where

$$\boldsymbol{\Upsilon}_k = \lfloor {}^{G}\mathbf{p}_f - {}^{G}\mathbf{p}_{I_1} - {}^{G}\mathbf{v}_{I_1}\delta t_{k-1} + \frac{1}{2}{}^{G}\mathbf{g}\delta t_{k-1}^2 \times \rfloor \mathbf{C}\left({}^{I_1}\bar{q}_G\right)^{\mathrm{T}}$$

$$\delta t_{k-1} = (k-1)\,\delta t$$

We note that in (49) $\mathbf{D}_k$ and $\mathbf{E}_k$ are both time-varying matrices, which do not affect the observability properties.

It is straightforward to verify that the right nullspace of $\mathbf{M}(\mathbf{x})$ spans four directions, i.e.

$$\mathbf{M}(\mathbf{x})\,\mathbf{N}_1 = \mathbf{0} \quad (50)$$

$$\mathbf{N}_1 = \begin{bmatrix}\mathbf{0}_3 & \mathbf{C}\left({}^{I_1}\bar{q}_G\right){}^{G}\mathbf{g} \\ \mathbf{0}_3 & \mathbf{0}_{3\times 1} \\ \mathbf{0}_3 & -\lfloor {}^{G}\mathbf{v}_{I_1} \times \rfloor {}^{G}\mathbf{g} \\ \mathbf{0}_3 & \mathbf{0}_{3\times 1} \\ \mathbf{I}_3 & -\lfloor {}^{G}\mathbf{p}_{I_1} \times \rfloor {}^{G}\mathbf{g} \\ \mathbf{I}_3 & -\lfloor {}^{G}\mathbf{p}_{f} \times \rfloor {}^{G}\mathbf{g}\end{bmatrix} = \begin{bmatrix}\mathbf{N}_{t_1} & | & \mathbf{N}_{r_1}\end{bmatrix} \quad (51)$$

where $\mathbf{N}_{t_1}$ corresponds to global translations and $\mathbf{N}_{r_1}$ corresponds to global rotations about the gravity vector, which are the same as those of the nonlinear system.[3]

Ideally, any estimator we employ should correspond to a system with an unobservable subspace that matches these directions, both in number and structure. However, when linearizing about the estimated state $\hat{\mathbf{x}}$, $\mathbf{M}(\hat{\mathbf{x}})$ gains rank due to errors in the state estimates across time (Hesch et al. 2012a). Hence, the linearized system used by the EKF has different observability properties than the nonlinear system it approximates, which leads to the acquisition of nonexistent information about global rotations around the gravity vector (yaw). To address this problem and ensure that (51) is satisfied for every block row of $\mathbf{M}$ when the state estimates are used for computing $\mathbf{H}_\ell$, and $\Phi_{\ell,1}$, $\ell = 1,\ldots,k$, we must ensure that $\mathbf{H}_\ell\Phi_{\ell,1}\mathbf{N}_1 = \mathbf{0}$, $\ell = 1,\ldots,k$ (see (47) and (50)).

One way to enforce this is by requiring that at each time step, $\Phi_{\ell+1,\ell}$ and $\mathbf{H}_\ell$ satisfy the following conditions:

$$\mathbf{N}_{\ell+1} = \Phi_{\ell+1,\ell}\mathbf{N}_\ell \quad (52)$$

$$\mathbf{H}_\ell\mathbf{N}_\ell = \mathbf{0}, \quad \ell = 1,\ldots,k \quad (53)$$

where $\mathbf{N}_\ell$, $\ell \geq 1$ is computed analytically (see (54), (55), and Hesch et al. (2012a)). This can be accomplished by appropriately modifying $\Phi_{\ell+1,\ell}$ and $\mathbf{H}_\ell$ following the process described in the next section.

6.1. OC-VINS: Algorithm description

Hereafter, we present our OC-VINS algorithm which enforces the observability constraints dictated by the VINS system structure. Rather than changing the linearization points explicitly (e.g. as in Huang et al. (2008)), we maintain the nullspace, $\mathbf{N}_k$, at each time step, and use it to enforce the unobservable directions. We refer to the first set of block rows of $\mathbf{N}_k$ as the nullspace corresponding to the robot state, which we term $\mathbf{N}_k^R$, whereas the last block row of $\mathbf{N}_k$ is the nullspace corresponding to the feature state, i.e. $\mathbf{N}_k^f$. Specifically, the 15×4 nullspace sub-block, $\mathbf{N}_k^R$, corresponding to the robot state is analytically defined as (see (51) and Hesch et al. (2012a)):

$$\mathbf{N}_1^R = \begin{bmatrix}\mathbf{0}_3 & \mathbf{C}\left({}^{I}\hat{\bar{q}}_{G,1|1}\right){}^{G}\mathbf{g} \\ \mathbf{0}_3 & \mathbf{0}_{3\times 1} \\ \mathbf{0}_3 & -\lfloor {}^{G}\hat{\mathbf{v}}_{I,1|1} \times \rfloor {}^{G}\mathbf{g} \\ \mathbf{0}_3 & \mathbf{0}_{3\times 1} \\ \mathbf{I}_3 & -\lfloor {}^{G}\hat{\mathbf{p}}_{I,1|1} \times \rfloor {}^{G}\mathbf{g}\end{bmatrix}$$

$$\mathbf{N}_k^R = \begin{bmatrix}\mathbf{0}_3 & \mathbf{C}\left({}^{I}\hat{\bar{q}}_{G,k|k-1}\right){}^{G}\mathbf{g} \\ \mathbf{0}_3 & \mathbf{0}_{3\times 1} \\ \mathbf{0}_3 & -\lfloor {}^{G}\hat{\mathbf{v}}_{I,k|k-1} \times \rfloor {}^{G}\mathbf{g} \\ \mathbf{0}_3 & \mathbf{0}_{3\times 1} \\ \mathbf{I}_3 & -\lfloor {}^{G}\hat{\mathbf{p}}_{I,k|k-1} \times \rfloor {}^{G}\mathbf{g}\end{bmatrix} = \begin{bmatrix}\mathbf{N}_{t,k}^R & | & \mathbf{N}_{r,k}^R\end{bmatrix} \quad (54)$$

The 3×4 nullspace sub-block, $\mathbf{N}_k^f$, corresponding to the feature state, is a function of the feature estimate at time t_ℓ when it was initialized, i.e.

$$\mathbf{N}_k^f = \begin{bmatrix}\mathbf{I}_3 & -\lfloor {}^{G}\hat{\mathbf{p}}_{f_{\ell|\ell}} \times \rfloor {}^{G}\mathbf{g}\end{bmatrix}. \quad (55)$$

6.1.1. Modification of the state transition matrix Φ During the propagation step, we must ensure that (52) is satisfied. We expand (52) by substituting the definitions of the state transition matrix (48) and the nullspace for both the robot state (54) and the feature (55), i.e.

$$\mathbf{N}_{k+1} = \Phi_{k+1,k}\mathbf{N}_k$$

$$\Leftrightarrow \quad \begin{bmatrix}\mathbf{N}_{k+1}^R \\ \mathbf{N}_{k+1}^f\end{bmatrix} = \begin{bmatrix}\Phi_{k+1,k}^R & \mathbf{0}_{15\times 3} \\ \mathbf{0}_{3\times 15} & \mathbf{I}_3\end{bmatrix}\begin{bmatrix}\mathbf{N}_k^R \\ \mathbf{N}_k^f\end{bmatrix}$$

which, after multiplying out, provides two relationships that should be satisfied:

$$\mathbf{N}_{k+1}^R = \Phi_{k+1,k}^R\,\mathbf{N}_k^R \quad (56)$$

$$\mathbf{N}_{k+1}^f = \mathbf{N}_k^f \quad (57)$$

From the definition of $\mathbf{N}_k^f$ (see (55)), it is clear that (57) holds automatically, and does not require any modification of $\Phi_{k+1,k}$. However, (56) will in general not hold, and hence it requires changing $\Phi_{k+1,k}^R$ such that $\mathbf{N}_{k+1}^R = \Phi_{k+1,k}^R\,\mathbf{N}_k^R$.

In order to determine which elements of $\Phi_{k+1,k}^R$ should be modified to satisfy (56), we further analyze the structure of this constraint. To do so, we partition $\mathbf{N}_k^R$ into two

components: (a) the first three columns corresponding to the unobservable translation, $\mathbf{N}^R_{t,k}$, and (b) the fourth column corresponding to the unobservable rotation about the gravity vector, $\mathbf{N}^R_{r,k}$ (see (51)). We rewrite (56) based on this partitioning to obtain:

$$\begin{aligned}&\mathbf{N}^R_{k+1} = \Phi^R_{k+1,k}\,\mathbf{N}^R_k \\ \Leftrightarrow\quad &\begin{bmatrix}\mathbf{N}^R_{t,k+1} & \mathbf{N}^R_{r,k+1}\end{bmatrix} = \Phi^R_{k+1,k}\begin{bmatrix}\mathbf{N}^R_{t,k} & \mathbf{N}^R_{r,k}\end{bmatrix}\end{aligned}$$

which is equivalent to satisfying the following two relationships simultaneously, i.e.

$$\mathbf{N}^R_{t,k+1} = \Phi^R_{k+1,k}\,\mathbf{N}^R_{t,k} \tag{58}$$

$$\mathbf{N}^R_{r,k+1} = \Phi^R_{k+1,k}\,\mathbf{N}^R_{r,k} \tag{59}$$

Treating these in order, we see that (58) is automatically satisfied, since every block row results in $\mathbf{0}_3 = \mathbf{0}_3$ or $\mathbf{I}_3 = \mathbf{I}_3$, i.e.

$$\begin{aligned}&\mathbf{N}^R_{t,k+1} = \Phi^R_{k+1,k}\,\mathbf{N}^R_{t,k} \\ \Leftrightarrow\quad &\begin{bmatrix}\mathbf{0}_3\\ \mathbf{0}_3\\ \mathbf{0}_3\\ \mathbf{0}_3\\ \mathbf{I}_3\end{bmatrix} = \begin{bmatrix}\Phi_{11} & \Phi_{12} & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3\\ \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3\\ \Phi_{31} & \Phi_{32} & \mathbf{I}_3 & \Phi_{34} & \mathbf{0}_3\\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_3\\ \Phi_{51} & \Phi_{52} & \delta t\mathbf{I}_3 & \Phi_{54} & \mathbf{I}_3\end{bmatrix}\begin{bmatrix}\mathbf{0}_3\\ \mathbf{0}_3\\ \mathbf{0}_3\\ \mathbf{0}_3\\ \mathbf{I}_3\end{bmatrix}\end{aligned}$$

We proceed by expanding the second relationship elementwise (see (59)) and we obtain

$$\begin{aligned}&\mathbf{N}^R_{r,k+1} = \Phi^R_{k+1,k}\,\mathbf{N}^R_{r,k} \Leftrightarrow \\ &\begin{bmatrix}\mathbf{C}\left({}^I\hat{\bar{q}}_{G,k+1|k}\right){}^G\mathbf{g}\\ \mathbf{0}_{3\times1}\\ -\lfloor{}^G\hat{\mathbf{v}}_{I,k+1|k}\times\rfloor{}^G\mathbf{g}\\ \mathbf{0}_{3\times1}\\ -\lfloor{}^G\hat{\mathbf{p}}_{I,k+1|k}\times\rfloor{}^G\mathbf{g}\end{bmatrix} = \begin{bmatrix}\Phi_{11} & \Phi_{12} & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3\\ \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3\\ \Phi_{31} & \Phi_{32} & \mathbf{I}_3 & \Phi_{34} & \mathbf{0}_3\\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_3\\ \Phi_{51} & \Phi_{52} & \delta t\mathbf{I}_3 & \Phi_{54} & \mathbf{I}_3\end{bmatrix}\\ &\qquad\begin{bmatrix}\mathbf{C}\left({}^I\hat{\bar{q}}_{G,k|k-1}\right){}^G\mathbf{g}\\ \mathbf{0}_{3\times1}\\ -\lfloor{}^G\hat{\mathbf{v}}_{I,k|k-1}\times\rfloor{}^G\mathbf{g}\\ \mathbf{0}_{3\times1}\\ -\lfloor{}^G\hat{\mathbf{p}}_{I,k|k-1}\times\rfloor{}^G\mathbf{g}\end{bmatrix}\end{aligned}$$

From the first block row we have that

$$\begin{aligned}\mathbf{C}\left({}^I\hat{\bar{q}}_{G,k+1|k}\right){}^G\mathbf{g} &= \Phi_{11}\mathbf{C}\left({}^I\hat{\bar{q}}_{G,k|k-1}\right){}^G\mathbf{g}\\ \Rightarrow\ \Phi_{11} &= \mathbf{C}\left({}^{I,k+1|k}\hat{\bar{q}}_{I,k|k-1}\right)\end{aligned} \tag{60}$$

The requirements for the third and fifth block rows are:

$$\Phi_{31}\mathbf{C}\left({}^I\hat{\bar{q}}_{G,k|k-1}\right){}^G\mathbf{g} = \lfloor{}^G\hat{\mathbf{v}}_{I,k|k-1}\times\rfloor{}^G\mathbf{g} - \lfloor{}^G\hat{\mathbf{v}}_{I,k+1|k}\times\rfloor{}^G\mathbf{g} \tag{61}$$

$$\begin{aligned}\Phi_{51}\mathbf{C}\left({}^I\hat{\bar{q}}_{G,k|k-1}\right){}^G\mathbf{g} = \delta t\lfloor{}^G\hat{\mathbf{v}}_{I,k|k-1}\times\rfloor{}^G\mathbf{g} + \lfloor{}^G\hat{\mathbf{p}}_{I,k|k-1}\times\rfloor{}^G\mathbf{g}\\ - \lfloor{}^G\hat{\mathbf{p}}_{I,k+1|k}\times\rfloor{}^G\mathbf{g}\end{aligned} \tag{62}$$

both of which are of the form $\mathbf{Au} = \mathbf{w}$, where $\mathbf{u}$ and $\mathbf{w}$ comprise nullspace elements that are fixed (see (54)), and we seek to find a perturbed $\mathbf{A}^*$, for $\mathbf{A} = \Phi_{31}$ and $\mathbf{A} = \Phi_{51}$, that fulfills the constraint. To compute the minimum perturbation, $\mathbf{A}^*$, of $\mathbf{A}$, we formulate the following minimization problem

$$\min_{\mathbf{A}^*} ||\mathbf{A}^* - \mathbf{A}||^2_{\mathscr{F}}\ ,\quad \text{s.t. } \mathbf{A}^*\mathbf{u} = \mathbf{w} \tag{63}$$

where $||\cdot||_{\mathscr{F}}$ denotes the Frobenius matrix norm. After employing the method of Lagrange multipliers, and solving the corresponding KKT optimality conditions (Boyd and Vandenberghe 2004), the optimal $\mathbf{A}^*$ that fulfills (63) is:

$$\mathbf{A}^* = \mathbf{A} - (\mathbf{Au} - \mathbf{w})(\mathbf{u}^\mathrm{T}\mathbf{u})^{-1}\mathbf{u}^\mathrm{T} \tag{64}$$

In summary, satisfying (52) only requires modifying three block elements of Φ_k during each propagation step. Specifically, we compute the modified Φ_{11} from (60), and Φ_{31} and Φ_{51} from (63)–(64) and construct the observability-constrained discrete-time state transition matrix. We then proceed with covariance propagation (see (15)).

6.1.2. Modification of the measurement Jacobian $\mathbf{H}$ During each update step, we seek to satisfy (53), i.e. $\mathbf{H}_k\mathbf{N}_k = \mathbf{0}$. Based on (19), (54), and (55) we can write this relationship *per feature* as

$$\mathbf{H}_c\begin{bmatrix}\mathbf{H}_\theta & \mathbf{0}_{3\times9} & \mathbf{H}_p & | & \mathbf{H}_f\end{bmatrix}\begin{bmatrix}\mathbf{0}_3 & \mathbf{C}\left({}^I\hat{\bar{q}}_{G,k|k-1}\right){}^G\mathbf{g}\\ \mathbf{0}_3 & \mathbf{0}_{3\times1}\\ \mathbf{0}_3 & -\lfloor{}^G\hat{\mathbf{v}}_{I,k|k-1}\times\rfloor{}^G\mathbf{g}\\ \mathbf{0}_3 & \mathbf{0}_{3\times1}\\ \mathbf{I}_3 & -\lfloor{}^G\hat{\mathbf{p}}_{I,k|k-1}\times\rfloor{}^G\mathbf{g}\\ \mathbf{I}_3 & -\lfloor{}^G\hat{\mathbf{p}}_{f_{\ell|\ell}}\times\rfloor{}^G\mathbf{g}\end{bmatrix} = \mathbf{0} \tag{65}$$

The first block column of (65) requires that $\mathbf{H}_f = -\mathbf{H}_p$. Hence, we rewrite the second block column of (65) as

$$\mathbf{H}_c\begin{bmatrix}\mathbf{H}_\theta & \mathbf{H}_p\end{bmatrix}\begin{bmatrix}\mathbf{C}\left({}^I\hat{\bar{q}}_{G,k|k-1}\right){}^G\mathbf{g}\\ \left(\lfloor{}^G\hat{\mathbf{p}}_{f_{\ell|\ell}}\times\rfloor - \lfloor{}^G\hat{\mathbf{p}}_{I,k|k-1}\times\rfloor\right){}^G\mathbf{g}\end{bmatrix} = \mathbf{0}$$

This is a constraint of the form $\mathbf{Au} = \mathbf{0}$, where $\mathbf{u}$ is a fixed quantity determined by elements in the nullspace, and $\mathbf{A}$ comprises elements of the measurement Jacobian $\mathbf{H}_k$. We compute the optimal $\mathbf{A}^*$ that satisfies this relationship using (63)–(64), which is a special case of this optimization problem when $\mathbf{w} = \mathbf{0}$. After computing the optimal $\mathbf{A}^*$, we recover the Jacobian as

$$\mathbf{H}_c\mathbf{H}_\theta = \mathbf{A}^*_{1:2,1:3} \tag{66}$$

$$\mathbf{H}_c\mathbf{H}_p = \mathbf{A}^*_{1:2,4:6} \tag{67}$$

$$\mathbf{H}_c\mathbf{H}_f = -\mathbf{A}^*_{1:2,4:6} \tag{68}$$

where the subscripts (i:j, m:n) denote the matrix sub-block spanning rows i to j, and columns m to n. After computing the modified measurement Jacobian, we proceed with the filter update as described in Section 3.2.

6.2. Application to the MSC-KF

The MSC-KF (Mourikis and Roumeliotis 2007) is a VINS that performs tightly coupled visual-inertial odometry over a sliding window of M poses, while maintaining linear complexity in the number of observed features. The key advantage of the MSC-KF is that it exploits all the constraints for each feature observed by the camera over M poses, without requiring to build a map or estimate the features as part of the state vector. We hereafter describe how to apply our OC-VINS methodology to the MSC-KF.

Each time the camera records an image, the MSC-KF creates a stochastic clone (Roumeliotis and Burdick 2002) of the sensor pose. This enables the MSC-KF to use delayed image measurements; in particular, it allows all of the observations of a given feature $\mathbf{p}_{f_i}$ to be processed during a single update step (when the first pose that observed the feature is about to be marginalized). Every time the current pose is cloned, we also clone the corresponding nullspace elements to obtain an augmented nullspace, i.e.

$$\mathbf{N}_k^{aug} = \begin{bmatrix} \mathbf{N}_k \\ \mathbf{N}_{k,clone} \end{bmatrix}$$

$$\text{where } \mathbf{N}_{k,clone} = \begin{bmatrix} \mathbf{0}_3 & \mathbf{C}\left({}^I\hat{\bar{q}}_{G,k|k-1}\right){}^G\mathbf{g} \\ \mathbf{I}_3 & -\lfloor {}^G\hat{\mathbf{p}}_{I,k|k-1} \times \rfloor {}^G\mathbf{g} \end{bmatrix}$$

During propagation, the current state estimate evolves forward in time by integrating (8)–(13), while the current clone poses are static. Moreover, we employ (60) and solve in closed form the optimization problem (63) for the constraints (61)–(62), using (64), so as to compute the observability-constrained discrete-time state transition matrix $\mathbf{\Phi}_{k+1,k}$, and propagate the covariance as

$$\mathbf{P}_{k+1|k}^{aug} = \mathbf{\Phi}_{k+1,k}^{aug}\mathbf{P}_{k|k}^{aug}\mathbf{\Phi}_{k+1,k}^{aug\mathrm{T}} + \begin{bmatrix} \mathbf{Q}_k & \mathbf{0}_{15\times 6M} \\ \mathbf{0}_{6M\times 15} & \mathbf{0}_{6M} \end{bmatrix}$$

$$\mathbf{\Phi}_{k+1,k}^{aug} = \begin{bmatrix} \mathbf{\Phi}_{k+1,k} & \mathbf{0}_{15\times 6M} \\ \mathbf{0}_{6M\times 15} & \mathbf{I}_{6M} \end{bmatrix}$$

where $\mathbf{P}_{i|j}^{aug}$ denotes the covariance of the augmented state corresponding to M cloned poses, along with the current state.

During the MSC-KF update step, we process all measurements of the features observed by the Mth clone (i.e. the one about to be marginalized from the sliding window of poses). We use (66)–(68) to compute the observability-constrained measurement Jacobian, $\widehat{\mathbf{H}}_k$, for each measurement and stack all observations of the ith feature across M time steps into a large measurement vector

$$\begin{bmatrix} \tilde{\mathbf{z}}_k \\ \vdots \\ \tilde{\mathbf{z}}_{k-M} \end{bmatrix} = \begin{bmatrix} \mathbf{H}_k \\ \vdots \\ \mathbf{H}_{k-M} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{x}}^{aug} \\ \tilde{\mathbf{p}}_f \end{bmatrix} + \begin{bmatrix} \boldsymbol{\eta}_k \\ \vdots \\ \boldsymbol{\eta}_{k-M} \end{bmatrix} = \mathbf{H}_x\tilde{\mathbf{x}}^{aug} + \mathbf{H}_f\tilde{\mathbf{p}}_f + \boldsymbol{\eta} \tag{69}$$

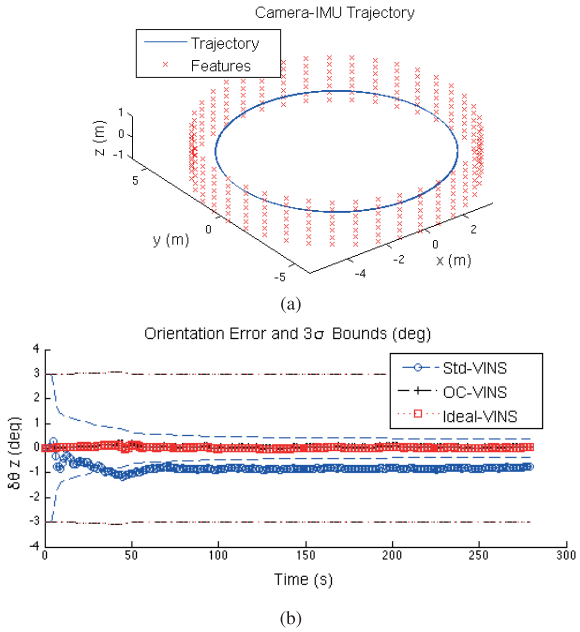


Fig. 2. (a) Camera-IMU trajectory and 3D features. (b) Errors (lines with markers) and 3σ bounds (lines without markers) for the rotation about the gravity vector, for the three filters, plotted for a single run. Note that the errors and 3σ bounds corresponding to the ideal VINS and the proposed OC-VINS are almost identical which makes it difficult to distinguish the corresponding lines in the figure.

where $\mathbf{H}_x$ and $\mathbf{H}_f$ are the Jacobians corresponding to the augmented state vector $\tilde{\mathbf{x}}^{aug}$, and to the feature, respectively. To avoid including $\mathbf{p}_f$ into the state, we marginalize it by projecting (69) onto the left nullspace of $\mathbf{H}_f$, which we term $\mathbf{W}$. This yields

$$\mathbf{W}^{\mathrm{T}}\tilde{\mathbf{z}} = \mathbf{W}^{\mathrm{T}}\mathbf{H}_x\tilde{\mathbf{x}}^{aug} + \mathbf{W}^{\mathrm{T}}\boldsymbol{\eta} \quad \Rightarrow \quad \tilde{\mathbf{z}}' = \mathbf{H}_x'\tilde{\mathbf{x}}^{aug} + \boldsymbol{\eta}',$$

which we employ to update the state estimate and covariance using the standard EKF update equations (Mourikis and Roumeliotis 2007).

7. Simulations

We conducted Monte-Carlo simulations to evaluate the impact of the proposed OC-VINS method on estimator consistency. We compared its performance to the standard VINS (Std-VINS), as well as the ideal VINS that linearizes about the true state. Since the ideal VINS has access to the true state, it is not realizable in practice, but we included it here as a baseline comparison. Specifically, we computed the Root Mean Squared Error (RMSE) and Normalized Estimation Error Squared (NEES) over 100 trials in which the camera-IMU platform traversed a circular trajectory of radius 5 m at an average velocity of 60 cm/s. The camera had 45° field of view, with $\sigma_{px} = 1px$, while the IMU was modeled with MEMS-quality sensors. The camera observed visual features distributed on the interior wall

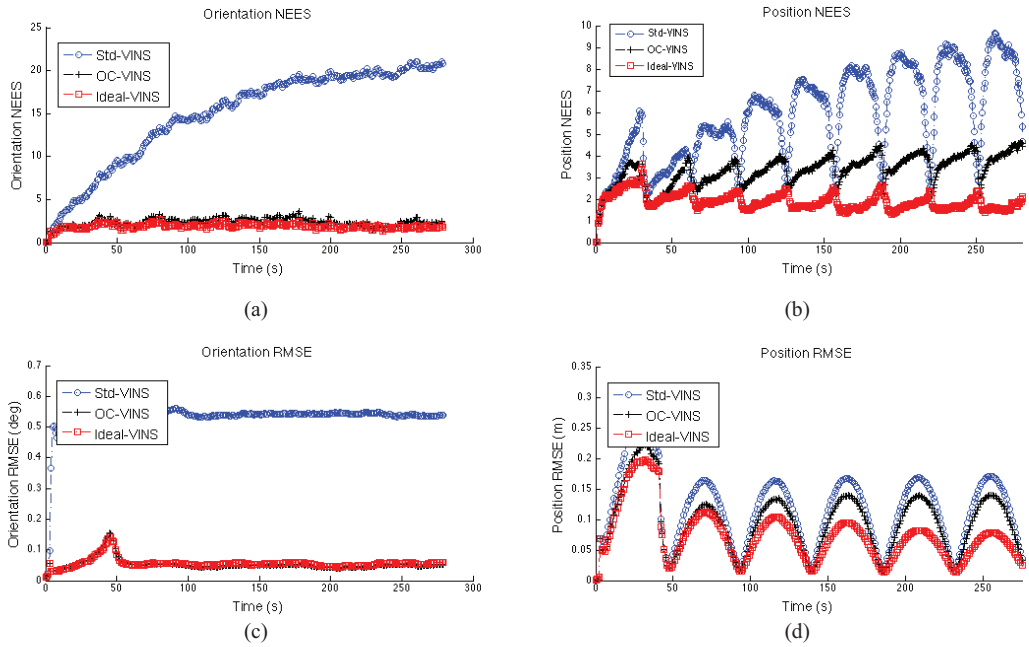


Fig. 3. The RMSE and NEES errors for position and orientation plotted for all three filters, averaged per time step over 100 Monte Carlo trials. Note that the RMSE and NEES errors, for orientation (left), corresponding to the ideal VINS and the proposed OC-VINS are almost identical which makes it difficult to distinguish the corresponding lines in the figure.

of a circumscribing cylinder with radius 6 m and height 2 m (see Figure 2(a)). The effect of inconsistency during a single run is depicted in Figure 2(b). The error and corresponding 3σ bounds of uncertainty are plotted for the rotation about the gravity vector. It is clear that the Std-VINS gains spurious information, hence reducing its 3σ bounds of uncertainty, while the Ideal-VINS and the OC-VINS do not. The Std-VINS becomes inconsistent during this run as the orientation errors fall outside of the uncertainty bounds, while both the Ideal-VINS and the OC-VINS remain consistent. Figure 3 displays the RMSE and NEES, in which we observe that the OC-VINS achieves orientation accuracy and consistency levels similar to the ideal, while significantly outperforming the Std-VINS. Similarly, the OC-VINS obtains better positioning accuracy compared to the Std-VINS.

8. Experimental results

In order to validate the proposed OC-VINS, we applied our method to the MSC-KF using real data. We hereafter provide a short overview of the system implementation, along with a discussion of the experimental results.

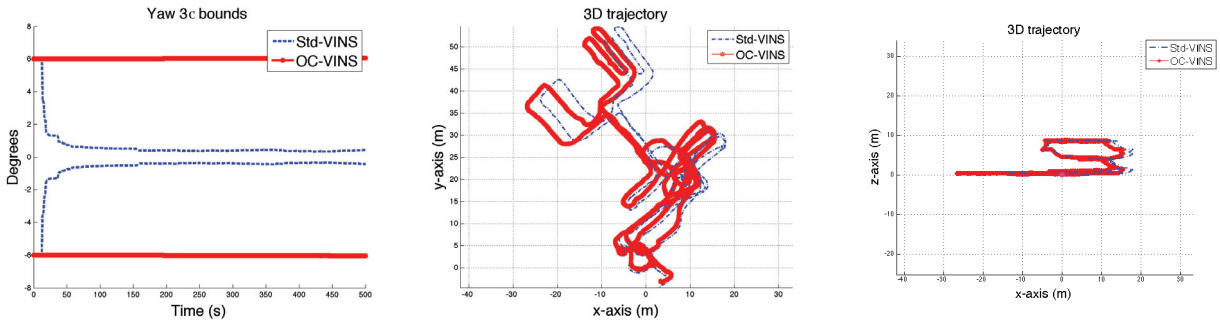
8.1. Implementation remarks

Short-term features are extracted from images using the Shi–Tomasi corner detector (Shi and Tomasi 1994). After acquiring image k, it is inserted into a sliding window buffer of M images, $\{k-M+1, k-M+2, \ldots, k\}$. We then extract features from the first image in the window and track them pairwise through the window using the KLT tracking algorithm (Lucas and Kanade 1981). To remove outliers from the resulting tracks, we use a two-point-RANSAC algorithm to find the essential matrix between successive frames (Kneip et al. 2011). Specifically, given the filter's estimated rotation between image i and j, ${}^i\hat{q}_j$ (using the gyroscope measurements), we compute the essential matrix, which now has only two DOF, from only two feature correspondences, i.e. ${}^i\xi^T\mathbf{E}^j\xi = 0$, where ${}^i\xi$ is the unit-norm vector of a feature at time step i, and $\mathbf{E} \triangleq \lfloor {}^i\rho_j \times \rfloor \mathbf{C}({}^i\hat{q}_j)$, where ${}^i\rho_j$ is the two DOF direction of motion between camera poses i and j. This approach is more robust than the five-point algorithm (Nistér 2003) because it provides 2 solutions for the essential matrix rather than up to 10, and as it requires only 2 data points, it reaches a consensus with fewer hypotheses when used in a RANSAC framework.

At every time step, the robot poses corresponding to the last M images are kept in the state vector, as described in Roumeliotis and Burdick (2002). Before marginalizing a pose, all the features that first appeared at the oldest augmented robot pose, are processed following the MSC-KF approach, as discussed in Section 3.2.

8.2. Experimental evaluation

Experiments were performed with a PointGrey Chameleon[4] camera and a Navchip IMU[5] which are rigidly attached on a light-weight sensing platform (see Figure 4). For both



(a) (b) (c)

Fig. 5. Experiment 1: (a) The estimated uncertainty in yaw computed by the Std-VINS and OC-VINS methods. (b) An x–y view of the trajectory which covered 550 m over three floors. (c) A 3D side view of the trajectory.

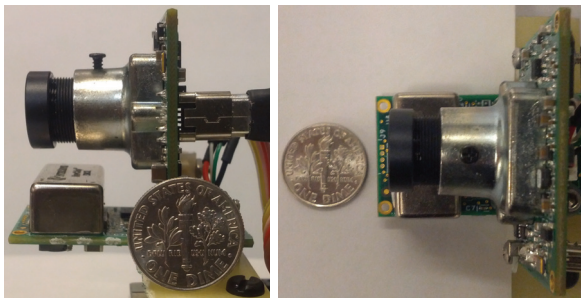


Fig. 4. The hardware setup comprises a miniature monochrome Point Grey Chameleon camera recording images at 7.5 Hz, and a rigidly attached InterSense NavChip IMU operating at 100 Hz. A coin (US dime, radius 1.8 cm) is included as a size reference.

experiments, IMU signals were sampled at a frequency of 100 Hz while camera images were acquired at 7.5 Hz. Features were tracked using a window of $M = 20$ images. In the first experiment, the platform traveled a total distance of 550 m over three floors of Walter Library at the University of Minnesota, traversing regions with a variety of lighting conditions, containing areas that were both rich and poor in distinctive features, and passing through both crowded and empty scenes. A video demonstrating the robustness of the proposed algorithm, can be found in Extension 1, accompanying the present paper. At the end of the trajectory, the sensing platform was placed back in its original configuration, so as to provide a quantitative characterization of the achieved accuracy.

For the Std-VINS the final position error was 5.33 m, while the OC-VINS achieved a final error of 4.60 m, corresponding to 0.97% and 0.83% of the total distance travelled, respectively (see Figure 5). In addition, the estimated covariances from the Std-VINS are smaller than those from the OC-VINS (see Figure 6). Furthermore, uncertainty estimates from the Std-VINS decreased in directions that are unobservable (i.e. rotations about the gravity vector); this violates the observability properties of the system and demonstrates that spurious information is injected into the filter.

Figure 5(a) highlights the difference in estimated yaw uncertainty between the OC-VINS and the Std-VINS. In contrast to the OC-VINS, the Std-VINS covariance rapidly decreases, violating the observability properties of the system. Similarly, large differences can be seen in the covariance estimates for the x- and y-position estimates (see Figure 6(a) and (b)). The Std-VINS estimates a much smaller uncertainty than the OC-VINS, supporting the claim that Std-VINS tends to be inconsistent.

The second experiment was conducted on the fifth floor of Keller Hall at the University of Minnesota, for which the floor plans are available, over a trajectory of 144 m. The IMU-camera sensor platform was initially aligned with the walls, so that the comparison of the estimated trajectory with the floor plans can provide strong qualitative evidence of the filter's improvement in accuracy and consistency. As evident from Figures 7 and 8, the Std-VINS erroneously injects information along the global yaw direction, which results in an infeasible trajectory when overlaid on the building's blueprint (i.e. the path passes through the walls). Both Std-VINS and OC-VINS achieved final accuracy smaller than 0.5% of the total distance travelled. However, as evident from the overlay of the estimated trajectories on the floor plan, the violation of the correct observability properties from the Std-VINS, leads to inconsistent yaw estimates which cause significant position errors, especially at the parts of the trajectory that are farthest from the starting point (see Figures 7(a) and (b)).

9. Conclusion and future work

In this work, we studied the observability properties of VINS, and leveraged our key results for mitigating filter inconsistency and improving the performance of linearized estimators. To do so, we first introduced a new method for determining the unobservable modes of a nonlinear system. In particular, and in contrast to previous approaches that require considering the infinitely many block rows of the observability matrix, we employ a set of auxiliary variables (basis functions) to achieve a factorization of the observability matrix, and hence a decomposition of the

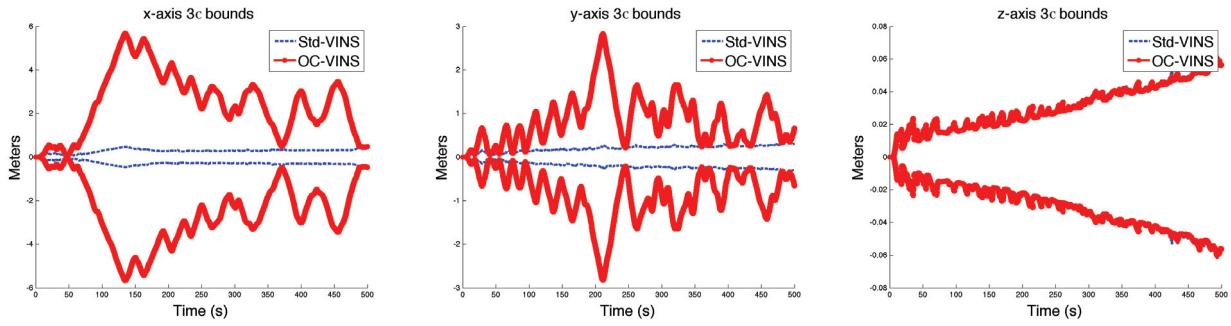


Fig. 6. Experiment 1: The estimated uncertainties for the Std-VINS and OC-VINS estimators computed for the *x*-, *y*-, and *z*-axes of position. We note that because Std-VINS is overconfident in its yaw estimate, it also becomes overconfident along the *x*- and *y*-axes since the horizontal positioning uncertainty is coupled with the global yaw uncertainty.

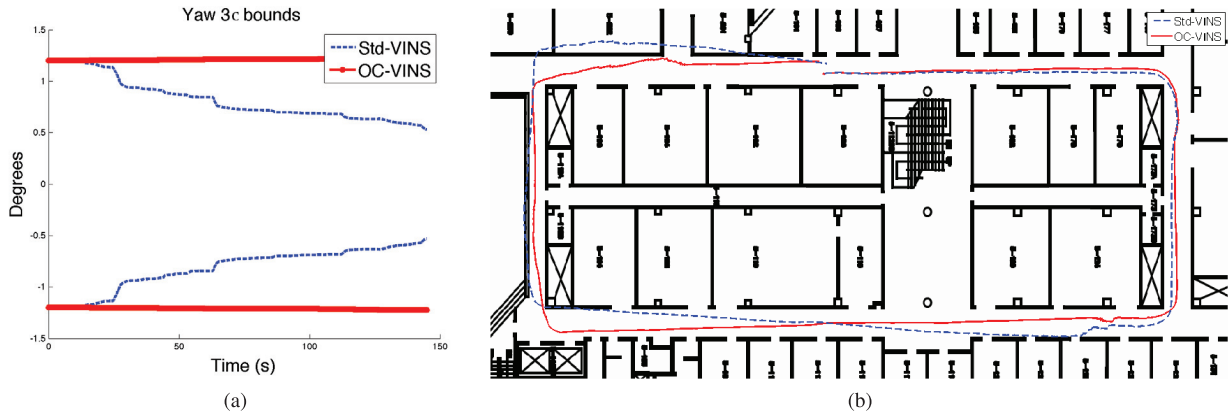


Fig. 7. Experiment 2: (a) The estimated uncertainty in yaw computed by the Std-VINS and OC-VINS methods. (b) Overhead *x*-*y* view of the trajectory, projected on the building's floor plans. The Std-VINS violation of the correct observability properties results in an infeasible trajectory (i.e. passing through walls), while the Std-VINS remains consistent with the ground truth floor drawings.

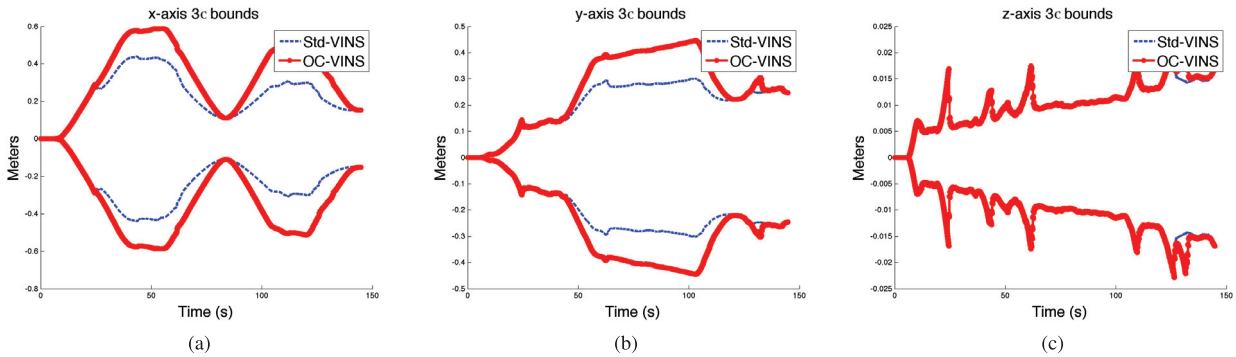


Fig. 8. Experiment 2: The estimated uncertainties for the Std-VINS and OC-VINS estimators computed for the *x*-, *y*-, and *z*-axes of position.

original system into observable and unobservable modes. Using this approach, the observability matrix of the resulting (reduced) unobservable system has a bounded number of rows which greatly simplifies the process for computing all its unobservable modes, and thus those of the original system.

Next, we applied this method to the VINS state model and derived the analytical form of the unobservable directions of the nonlinear system. Furthermore, we showed that these coincide with the unobservable directions of the linearized system, when linearization is performed around the true state. In practice, however, when the system is linearized about the state estimates, the observability properties are violated, allowing spurious information to enter into the estimator and leading to inconsistency. To address this issue, we employed our analysis for improving the

consistency and accuracy of VINS. In particular, we explicitly enforced the correct observability properties (in terms of number and structure of the unobservable directions) by performing simple modifications of the system and measurement Jacobians. Finally, we presented both simulation and experimental results that validate the superior performance and improved consistency of the proposed observability-constrained estimator.

In our future work, we are interested in analyzing additional sources of estimator inconsistency in VINS such as the existence of multiple local minima.

Funding

This work was supported by the University of Minnesota (DTC) and AFOSR (grant number FA9550-10-1-0567).

Notes

1. As defined in Bar-Shalom et al. (2001), a state estimator is consistent if the estimation errors are zero-mean and have covariance equal to the one calculated by the filter.
2. The same phenomenon occurs in other estimation frameworks such as batch least-squares, unless the entire cost function (including both IMU and vision constraints) is relinearized at every iteration.
3. Note that the unobservable directions for the linearized system when the Jacobians are evaluated at the true states (see (51)) are the same as those of the underlying nonlinear system (see (45)). The term $\frac{\partial \mathbf{s}}{\partial \boldsymbol{\theta}}$ appearing in the first element of the fourth direction of (45) is due to the different attitude representation.
4. See www.ptgrey.com.
5. See www.intersense.com.

References

Bar-Shalom Y, Li XR and Kirubarajan T (2001) *Estimation with Applications to Tracking and Navigation*. New York: John Wiley & Sons.

Bouguet JY (2006) Camera calibration toolbox for matlab. Available at: www.vision.caltech.edu/bouguetj/calibdoc/.

Boyd S and Vandenberghe L (2004) *Convex Optimization*. Cambridge; New York: Cambridge University Press.

Bryson M and Sukkarieh S (2008) Observability analysis and active control for airborne SLAM. *IEEE Transactions on Aerospace and Electronic Systems* 44(1): 261–280.

Carlone L, Macchia V, Tibaldi F and Bona B (2012) Robot localization and 3D mapping: Observability analysis and applications. In: *Proceedings of the International symposium on artificial intelligence, robotics and automation in space*, Turin, Italy, 4–6 September 2012, pp. 7–14.

Chatfield A (1997) *Fundamentals of High Accuracy Inertial Navigation*. Reston, VA: AIAA.

Chiuso A, Favaro P, Jin H and Soatto S (2002) Structure from motion causally integrated over time. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24(4): 523–535.

Durrie J, Gerritsen T, Frew EW and Pledgie S (2009) Vision-aided inertial navigation on an uncertain map using a particle filter. In: *Proceedings of the IEEE International conference on robotics and automation*, Kobe, Japan, 12–17 May 2009, pp. 4189–4194.

Ebcin S and Veth M (2007) Tightly-coupled image-aided inertial navigation using the unscented Kalman filter. Technical report, Air Force Institute of Technology, Dayton, OH.

Hermann R and Krener A (1977) Nonlinear controllability and observability. *IEEE Transactions on Automatic Control* 22(5): 728–740.

Hesch JA, Kottas DG, Bowman SL and Roumeliotis SI (2012a) *Observability-constrained vision-aided inertial navigation.* Technical report 2012-001, University of Minnesota, Department of Computer Science and Engineering, MARS Lab. Available at: www-users.cs.umn.edu/~joel/_files/Joel_Hesch_TR12.pdf.

Hesch JA, Kottas DG, Bowman SL and Roumeliotis SI (2012b) Towards consistent vision-aided inertial navigation. In: *International workshop on the algorithmic foundations of robotics*, Cambridge, MA, 13–15 June 2012, pp. 559–574.

Hesch JA, Mirzaei FM, Mariottini GL and Roumeliotis SI (2010) A Laser-aided Inertial Navigation System (L-INS) for human localization in unknown indoor environments. In: *Proceedings of the IEEE International conference on robotics and automation*, Anchorage, AK, 3–8 May 2010, pp. 5376–5382.

Huang GP, Mourikis AI and Roumeliotis SI (2008) A first-estimates Jacobian EKF for improving SLAM consistency. In: *Proceedings of the International symposium on experimental robotics*. Athens, Greece, 14–17 July 2008, pp. 373–382.

Huang GP, Mourikis AI and Roumeliotis SI (2010) Observability-based rules for designing consistent EKF SLAM estimators. *International Journal of Robotics Research* 29(5): 502–528.

Huang GP, Trawny N, Mourikis AI and Roumeliotis SI (2011) Observability-based consistent EKF estimators for multi-robot cooperative localization. *Autonomous Robots* 30(1): 99–122.

Isidori A (1989) *Nonlinear Control Systems*. Berlin, New York: Springer-Verlag.

Jones ES and Soatto S (2011) Visual-inertial navigation, mapping and localization: A scalable real-time causal approach. *International Journal of Robotics Research* 30(4): 407–430.

Kelly J and Sukhatme GS (2011) Visual-inertial sensor fusion: Localization, mapping and sensor-to-sensor self-calibration. *International Journal of Robotics Research* 30(1): 56–79.

Kim J and Sukkarieh S (2007) Real-time implementation of airborne inertial-SLAM. *Robotics and Autonomous Systems* 55(1): 62–71.

Kneip L, Chli M and Siegwart R (2011) Robust real-time visual odometry with a single camera and an IMU. In: *Proceedings of the British machine vision conference*, Dundee, UK, 29 August-2 September, 2011, pp. 16.1-16.11.

Li M and Mourikis AI (2012) Improving the accuracy of EKF-based visual-inertial odometry. In: *Proceedings of the IEEE International conference on robotics and automation*, Minneapolis, MN, 14–18 May 2012, pp. 828–835.

Li M and Mourikis AI (2013) High-precision, consistent EKF-based visualâŁ“inertial odometry. *International Journal of Robotics Research* 32(6): 690–711.

Lucas B and Kanade T (1981) An iterative image registration technique with an application to stereo vision. In: *Proceedings of the International joint conference on artificial intelligence*, Vancouver, Canada, August 1981, pp. 674–679.

Lupton T and Sukkarieh S (2012) Visual-inertial-aided navigation for high-dynamic motion in built environments without initial conditions. *IEEE Transactions on Robotics* 28(1): 61–76.

Martinelli A (2012) Vision and IMU data fusion: Closed-form solutions for attitude, speed, absolute scale, and bias determination. *IEEE Transactions on Robotics* 28(1): 44–60.

Maybeck PS (1979) *Stochastic models, estimation, and control*, Vol. 1. New York: Academic Press.

Meyer CD (2000) *Matrix Analysis and Applied Linear Algebra*. Philadelphia, PA: SIAM.

Mirzaei FM and Roumeliotis SI (2008) A Kalman filter-based algorithm for IMU-camera calibration: Observability analysis and performance evaluation. *IEEE Transactions on Robotics* 24(5): 1143–1156.

Mourikis AI and Roumeliotis SI (2007) A multi-state constraint Kalman filter for vision-aided inertial navigation. In: *Proceedings of the IEEE International conference on robotics and automation*, Rome, Italy, 10–14 April, pp. 3565–3572.

Nistér D (2003) An efficient solution to the five-point relative pose problem. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, Madison, WI, 16–22 June 2003, pp. 195–202.

Roumeliotis SI and Burdick JW (2002) Stochastic cloning: A generalized framework for processing relative state measurements. In: *Proceedings of the IEEE International conference on robotics and automation*, Washington DC, 11–15 May 2002, pp. 1788–1795.

Shen S, Michael N and Kumar V (2011) Autonomous multi-floor indoor navigation with a computationally constrained MAV. In: *Proceedings of the IEEE International conference on robotics and automation*, Shanghai, China, 9–13 May 2011, pp. 20–25.

Shi J and Tomasi C (1994) Good features to track. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, Washington DC, 21–23 June 1994, pp. 593–600.

Shuster MD (1993) A survey of attitude representations. *Journal of the Astronautical Sciences* 41(4): 439–517.

Strelow DW (2004) *Motion estimation from image and inertial measurements*. PhD Thesis, Carnegie Mellon University, Pittsburgh, PA.

Trawny N and Roumeliotis SI (2005) Indirect Kalman filter for 3D attitude estimation. Technical Report 2005-002, University of Minnesota, Department of Computer Science and Engineering, MARS Lab.

Weiss S (2012) *Vision based navigation for micro helicopters*. PhD Thesis, Swiss Federal Institute of Technology (ETH), Zurich, Switzerland.

Weiss S, Achtelik MW, Chli M and Siegwart R (2012) Versatile distributed pose estimation and sensor self-calibration for an autonomous MAV. In: *Proceedings of the IEEE International conference on robotics and automation*. Minneapolis, MN, 14–18 May 2012, pp. 957–964.

Williams B, Hudson N, Tweddle B, Brockers R and Matthies L (2011) Feature and pose constrained visual aided inertial navigation for computationally constrained aerial vehicles. In: *Proceedings of the IEEE International conference on robotics and automation*, Shanghai, China, 9–13 May 2011, pp. 431–438.

Yap T Jr., Li M, Mourikis AI and Shelton CR (2011) A particle filter for monocular vision-aided odometry. In: *Proceedings of the IEEE International conference on robotics and automation*, Shanghai, China, 9–13 May 2011, pp. 5663–5669.

Appendix A: Index to Multimedia Extensions

The multimedia extension page is found at http://www.ijrr.org

Table of Multimedia Extensions

Extension	Media Type	Description
1	Video	Experimental evaluation of OC-VINS over a 550 (m) indoor trajectory.

Appendix B

In this section, we study the observability properties of system (43), by showing that its observability matrix, $\boldsymbol{\Xi}$ [see (22)], is of full column rank; thus system (43) is observable and the basis functions $\boldsymbol{\beta}$ are the observable modes of the original system (25).

Although the observability matrix comprising the spans of all the Lie derivatives of the system (25)–(26) will, in general, have infinite number of rows, we will only use a subset of its Lie derivatives to prove that it is observable. In particular, since we aim to prove that $\boldsymbol{\Xi}$ is of full column rank, we need only to select a subset of its rows that are linearly independent. Specifically, we select the set

$$\begin{aligned}\mathscr{L} = \{&\mathscr{L}^0\mathbf{h}, \mathscr{L}^3_{\mathbf{g}_0\mathbf{g}_{13}\mathbf{g}_{21}}\mathbf{h}, \mathscr{L}^1_{\mathbf{g}_0}\mathbf{h}, \mathscr{L}^3_{\mathbf{g}_0\mathbf{g}_{13}\mathbf{g}_{13}}\mathbf{h}, \mathscr{L}^3_{\mathbf{g}_0\mathbf{g}_0\mathbf{g}_{21}}\mathbf{h},\\ &\mathcal{L}^2_{\mathbf{g}_0\mathbf{g}_0}\mathbf{h}, \mathscr{L}^3_{\mathbf{g}_0\mathbf{g}_0\mathbf{g}_{13}}\mathbf{h}, \mathscr{L}^3_{\mathbf{g}_0\mathbf{g}_0\mathbf{g}_0}\mathbf{h}\}\end{aligned}$$

where we have used the abbreviated notation $\mathbf{g}_{ij}$ to denote the jth component of the ith input, i.e. $\mathbf{g}_{ij} = \mathbf{g}_i\mathbf{e}_j$. The ordering of $\mathscr{L}$ has been selected so as to admit an advantageous structure for analyzing the observability matrix.

The observability sub-matrix corresponding to these Lie derivatives is

$$\boldsymbol{\Xi}' = \begin{bmatrix} \frac{\partial \mathscr{L}^0\mathbf{h}}{\partial \boldsymbol{\beta}} \\ \frac{\partial \mathscr{L}^3_{\mathbf{g}_0\mathbf{g}_{13}\mathbf{g}_{21}}\mathbf{h}}{\partial \boldsymbol{\beta}} \\ \frac{\partial \mathscr{L}^1_{\mathbf{g}_0}\mathbf{h}}{\partial \boldsymbol{\beta}} \\ \frac{\partial \mathscr{L}^3_{\mathbf{g}_0\mathbf{g}_{13}\mathbf{g}_{13}}\mathbf{h}}{\partial \boldsymbol{\beta}} \\ \frac{\partial \mathscr{L}^3_{\mathbf{g}_0\mathbf{g}_0\mathbf{g}_{21}}\mathbf{h}}{\partial \boldsymbol{\beta}} \\ \frac{\partial \mathscr{L}^2_{\mathbf{g}_0\mathbf{g}_0}\mathbf{h}}{\partial \boldsymbol{\beta}} \\ \frac{\partial \mathscr{L}^3_{\mathbf{g}_0\mathbf{g}_0\mathbf{g}_{13}}\mathbf{h}}{\partial \boldsymbol{\beta}} \\ \frac{\partial \mathscr{L}^3_{\mathbf{g}_0\mathbf{g}_0\mathbf{g}_0}\mathbf{h}}{\partial \boldsymbol{\beta}} \end{bmatrix} \tag{70}$$

which, after expanding all of the spans of the Lie derivatives in (70), has the following structure:

$$\boldsymbol{\Xi}' = \begin{bmatrix} \mathbf{I}_{3\times3} & \mathbf{0}_{3\times6} & \mathbf{0}_{3\times6} \\ \mathbf{0}_{1\times3} & \mathbf{0}_{1\times6} & \mathbf{0}_{1\times6} \\ \mathbf{X}_{6\times3} & \boldsymbol{\Psi}_{6\times6} & \mathbf{0}_{6\times6} \\ \mathbf{Y}_{6\times3} & \mathbf{Z}_{6\times6} & \boldsymbol{\Theta}_{6\times6} \end{bmatrix}$$

Since the second block row of $\boldsymbol{\Xi}'$ is all zero, we drop it, defining a new matrix $\boldsymbol{\Xi}''$ whose rank is the same, i.e.

$$\Xi'' = \begin{bmatrix} \mathbf{I}_{3\times3} & \mathbf{0}_{3\times6} & \mathbf{0}_{3\times6} \\ \mathbf{X}_{6\times3} & \mathbf{\Psi}_{6\times6} & \mathbf{0}_{6\times6} \\ \mathbf{Y}_{6\times3} & \mathbf{Z}_{6\times6} & \mathbf{\Theta}_{6\times6} \end{bmatrix}$$

Hence, we can prove that system (43) is observable, by showing that the matrix Ξ'' is of full column rank.

The first step in our proof consists of showing that both $\mathbf{\Psi}_{6\times6}$ and $\mathbf{\Theta}_{6\times6}$ are full-rank matrices. Specifically,

$$\mathbf{\Psi} = \begin{bmatrix} -\beta_{21} & 0 & \beta_{11}\beta_{21} & -\beta_{11}\beta_{12} & \beta_{11}^2+1 & -\beta_{12} \\ 0 & -\beta_{21} & \beta_{12}\beta_{21} & -\beta_{12}^2-1 & \beta_{11}\beta_{12} & \beta_{11} \\ \beta_{21} & 0 & -\beta_{11}\beta_{21} & 4\beta_{11}\beta_{12} & 2\beta_{12}^2-2\beta_{11}^2 & \beta_{12} \\ 0 & \beta_{21} & -\beta_{12}\beta_{21} & 2\beta_{12}^2-2\beta_{11}^2 & -4\beta_{11}\beta_{12} & -\beta_{11} \\ 0 & 0 & -2\beta_{21}^2 & 2\beta_{12}\beta_{21} & -4\beta_{11}\beta_{21} & 0 \\ 0 & 0 & 0 & 0 & -2\beta_{12}\beta_{21} & -2\beta_{21} \end{bmatrix}$$

where $\boldsymbol{\beta}_{ij}$ denotes the jth component of basis element $\boldsymbol{\beta}_i$. Examining the determinant of $\mathbf{\Psi}$, we see that

$$\begin{aligned} \det(\mathbf{\Psi}) &= -4\boldsymbol{\beta}_{21}^5\left(\boldsymbol{\beta}_{11}^2+\boldsymbol{\beta}_{12}^2-1\right)\left(2\boldsymbol{\beta}_{11}^2+2\boldsymbol{\beta}_{12}^2+1\right) \\ &= -4\frac{1}{p_z^5}\left(\frac{p_x^2}{p_z^2}+\frac{p_y^2}{p_z^2}-1\right)\left(2\frac{p_x^2}{p_z^2}+2\frac{p_y^2}{p_z^2}+1\right) \end{aligned} \quad (71)$$

where for the purpose of analyzing the determinant, we have substituted the basis element definitions (see (37) and (38)). First, we note that since the observed point cannot be coincident with the camera center (due to the physical size of the lens and optics), $p_z \neq 0$. Moreover, since we only process features whose positions can be triangulated from multiple views (i.e. features that are not at infinite distance from the camera) $\frac{1}{p_z} \neq 0$. Second, we note that all quantities in the last term are nonnegative, hence,

$$\left(2\frac{p_x^2}{p_z^2}+2\frac{p_y^2}{p_z^2}+1\right) \geq 1$$

This means that $\mathbf{\Psi}$ is only rank deficient when the relationship

$$\left(\frac{p_x^2}{p_z^2}+\frac{p_y^2}{p_z^2}-1\right) = 0$$

holds. This equation is satisfied when the observed point lies on a circle with radius 1 on the normalized image plane (i.e. at focal length 1 from the optical center). The corresponding bearing angle to a point on this circle is 45°. This corresponds to a zero-probability event, since the control inputs of the system take arbitrary values across time. Thus, we conclude that $\mathbf{\Psi}$ is generically full rank.

We now turn our attention to the 6×6 submatrix $\mathbf{\Theta}$:

$$\mathbf{\Theta} = \begin{bmatrix} -\boldsymbol{\beta}_{21} & 0 & \boldsymbol{\beta}_{11}\boldsymbol{\beta}_{21} & \boldsymbol{\beta}_{21} & 0 & -\boldsymbol{\beta}_{11}\boldsymbol{\beta}_{21} \\ 0 & -\boldsymbol{\beta}_{21} & \boldsymbol{\beta}_{12}\boldsymbol{\beta}_{21} & 0 & \boldsymbol{\beta}_{21} & -\boldsymbol{\beta}_{12}\boldsymbol{\beta}_{21} \\ 0 & -\boldsymbol{\beta}_{21} & \boldsymbol{\beta}_{12}\boldsymbol{\beta}_{21} & 0 & 0 & -\boldsymbol{\beta}_{12}\boldsymbol{\beta}_{21} \\ \boldsymbol{\beta}_{21} & 0 & -\boldsymbol{\beta}_{11}\boldsymbol{\beta}_{21} & 0 & 0 & \boldsymbol{\beta}_{11}\boldsymbol{\beta}_{21} \\ \mathbf{\Theta}_{5,1} & \mathbf{\Theta}_{5,2} & \mathbf{\Theta}_{5,3} & \mathbf{\Theta}_{5,4} & \mathbf{\Theta}_{5,5} & \mathbf{\Theta}_{5,6} \\ \mathbf{\Theta}_{6,1} & \mathbf{\Theta}_{6,2} & \mathbf{\Theta}_{6,3} & \mathbf{\Theta}_{6,4} & \mathbf{\Theta}_{6,5} & \mathbf{\Theta}_{6,6} \end{bmatrix}$$

where $\mathbf{\Theta}_{i,j}$ denotes the element in the ith row and jth column of the matrix $\mathbf{\Theta}$, with

$$\begin{aligned}
\mathbf{\Theta}_{5,1} &= -2\boldsymbol{\beta}_{21}\left(\boldsymbol{\beta}_{11}\boldsymbol{\beta}_{42}-\boldsymbol{\beta}_{12}\boldsymbol{\beta}_{41}+\boldsymbol{\beta}_{21}\boldsymbol{\beta}_{33}\right) \\
&\quad -\boldsymbol{\beta}_{21}\left(2\boldsymbol{\beta}_{11}\boldsymbol{\beta}_{42}-\boldsymbol{\beta}_{12}\boldsymbol{\beta}_{41}+\boldsymbol{\beta}_{21}\boldsymbol{\beta}_{33}\right)-2\boldsymbol{\beta}_{11}\boldsymbol{\beta}_{21}\boldsymbol{\beta}_{42} \\
\mathbf{\Theta}_{5,2} &= 2\boldsymbol{\beta}_{21}\boldsymbol{\beta}_{43}+\boldsymbol{\beta}_{21}\left(\boldsymbol{\beta}_{43}+\boldsymbol{\beta}_{11}\boldsymbol{\beta}_{41}\right)+2\boldsymbol{\beta}_{11}\boldsymbol{\beta}_{21}\boldsymbol{\beta}_{41} \\
\mathbf{\Theta}_{5,3} &= 2\boldsymbol{\beta}_{21}\left(\boldsymbol{\beta}_{42}-\boldsymbol{\beta}_{21}\boldsymbol{\beta}_{31}-\boldsymbol{\beta}_{12}\boldsymbol{\beta}_{43}\right. \\
&\quad \left.+\boldsymbol{\beta}_{11}\left(\boldsymbol{\beta}_{11}\boldsymbol{\beta}_{42}-\boldsymbol{\beta}_{12}\boldsymbol{\beta}_{41}+\boldsymbol{\beta}_{21}\boldsymbol{\beta}_{33}\right)\right)-2\boldsymbol{\beta}_{21}\boldsymbol{\beta}_{42} \\
&\quad -{\boldsymbol{\beta}_{21}}^2\left(\boldsymbol{\beta}_{31}-\boldsymbol{\beta}_{11}\boldsymbol{\beta}_{33}\right)-\boldsymbol{\beta}_{12}\boldsymbol{\beta}_{21}\left(\boldsymbol{\beta}_{43}+\boldsymbol{\beta}_{11}\boldsymbol{\beta}_{41}\right) \\
&\quad +2\boldsymbol{\beta}_{11}\boldsymbol{\beta}_{21}\left(\boldsymbol{\beta}_{11}\boldsymbol{\beta}_{42}-\boldsymbol{\beta}_{12}\boldsymbol{\beta}_{41}+\boldsymbol{\beta}_{21}\boldsymbol{\beta}_{33}\right) \\
&\quad +\boldsymbol{\beta}_{11}\boldsymbol{\beta}_{21}\left(2\boldsymbol{\beta}_{11}\boldsymbol{\beta}_{42}-\boldsymbol{\beta}_{12}\boldsymbol{\beta}_{41}+\boldsymbol{\beta}_{21}\boldsymbol{\beta}_{33}\right) \\
\mathbf{\Theta}_{5,4} &= 2\boldsymbol{\beta}_{21}\left(\boldsymbol{\beta}_{11}\boldsymbol{\beta}_{42}-\boldsymbol{\beta}_{12}\boldsymbol{\beta}_{41}+\boldsymbol{\beta}_{21}\boldsymbol{\beta}_{33}\right) \\
&\quad +\boldsymbol{\beta}_{21}\left(2\boldsymbol{\beta}_{11}\boldsymbol{\beta}_{42}-\boldsymbol{\beta}_{12}\boldsymbol{\beta}_{41}+\boldsymbol{\beta}_{21}\boldsymbol{\beta}_{33}\right)+\boldsymbol{\beta}_{11}\boldsymbol{\beta}_{21}\boldsymbol{\beta}_{42} \\
\mathbf{\Theta}_{5,5} &= -\boldsymbol{\beta}_{21}\boldsymbol{\beta}_{43}-\boldsymbol{\beta}_{21}\left(\boldsymbol{\beta}_{43}+\boldsymbol{\beta}_{11}\boldsymbol{\beta}_{41}\right)-\boldsymbol{\beta}_{11}\boldsymbol{\beta}_{21}\boldsymbol{\beta}_{41} \\
\mathbf{\Theta}_{5,6} &= {\boldsymbol{\beta}_{21}}^2\left(\boldsymbol{\beta}_{31}-\boldsymbol{\beta}_{11}\boldsymbol{\beta}_{33}\right)+\boldsymbol{\beta}_{21}\boldsymbol{\beta}_{42}-2\boldsymbol{\beta}_{21}\left(\boldsymbol{\beta}_{42}-\boldsymbol{\beta}_{21}\boldsymbol{\beta}_{31}\right. \\
&\quad \left.-\boldsymbol{\beta}_{12}\boldsymbol{\beta}_{43}+\boldsymbol{\beta}_{11}\left(\boldsymbol{\beta}_{11}\boldsymbol{\beta}_{42}-\boldsymbol{\beta}_{12}\boldsymbol{\beta}_{41}+\boldsymbol{\beta}_{21}\boldsymbol{\beta}_{33}\right)\right) \\
&\quad +\boldsymbol{\beta}_{12}\boldsymbol{\beta}_{21}\left(\boldsymbol{\beta}_{43}+\boldsymbol{\beta}_{11}\boldsymbol{\beta}_{41}\right) \\
&\quad -2\boldsymbol{\beta}_{11}\boldsymbol{\beta}_{21}\left(\boldsymbol{\beta}_{11}\boldsymbol{\beta}_{42}-\boldsymbol{\beta}_{12}\boldsymbol{\beta}_{41}+\boldsymbol{\beta}_{21}\boldsymbol{\beta}_{33}\right) \\
&\quad -\boldsymbol{\beta}_{11}\boldsymbol{\beta}_{21}\left(2\boldsymbol{\beta}_{11}\boldsymbol{\beta}_{42}-\boldsymbol{\beta}_{12}\boldsymbol{\beta}_{41}+\boldsymbol{\beta}_{21}\boldsymbol{\beta}_{33}\right) \\
\mathbf{\Theta}_{6,1} &= -2\boldsymbol{\beta}_{21}\boldsymbol{\beta}_{43}-\boldsymbol{\beta}_{21}\left(\boldsymbol{\beta}_{43}+\boldsymbol{\beta}_{12}\boldsymbol{\beta}_{42}\right)-2\boldsymbol{\beta}_{12}\boldsymbol{\beta}_{21}\boldsymbol{\beta}_{42} \\
\mathbf{\Theta}_{6,2} &= 2\boldsymbol{\beta}_{12}\boldsymbol{\beta}_{21}\boldsymbol{\beta}_{41}-\boldsymbol{\beta}_{21}\left(\boldsymbol{\beta}_{11}\boldsymbol{\beta}_{42}-2\boldsymbol{\beta}_{12}\boldsymbol{\beta}_{41}+\boldsymbol{\beta}_{21}\boldsymbol{\beta}_{33}\right) \\
&\quad -2\boldsymbol{\beta}_{21}\left(\boldsymbol{\beta}_{11}\boldsymbol{\beta}_{42}-\boldsymbol{\beta}_{12}\boldsymbol{\beta}_{41}+\boldsymbol{\beta}_{21}\boldsymbol{\beta}_{33}\right) \\
\mathbf{\Theta}_{6,3} &= 2\boldsymbol{\beta}_{21}\boldsymbol{\beta}_{41}-{\boldsymbol{\beta}_{21}}^2\left(\boldsymbol{\beta}_{32}-\boldsymbol{\beta}_{12}\boldsymbol{\beta}_{33}\right) \\
&\quad -2\boldsymbol{\beta}_{21}\left(\boldsymbol{\beta}_{41}+\boldsymbol{\beta}_{21}\boldsymbol{\beta}_{32}-\boldsymbol{\beta}_{11}\boldsymbol{\beta}_{43}\right. \\
&\quad \left.-\boldsymbol{\beta}_{12}\left(\boldsymbol{\beta}_{11}\boldsymbol{\beta}_{42}-\boldsymbol{\beta}_{12}\boldsymbol{\beta}_{41}+\boldsymbol{\beta}_{21}\boldsymbol{\beta}_{33}\right)\right) \\
&\quad +\boldsymbol{\beta}_{11}\boldsymbol{\beta}_{21}\left(\boldsymbol{\beta}_{43}+\boldsymbol{\beta}_{12}\boldsymbol{\beta}_{42}\right) \\
&\quad +2\boldsymbol{\beta}_{12}\boldsymbol{\beta}_{21}\left(\boldsymbol{\beta}_{11}\boldsymbol{\beta}_{42}-\boldsymbol{\beta}_{12}\boldsymbol{\beta}_{41}+\boldsymbol{\beta}_{21}\boldsymbol{\beta}_{33}\right) \\
&\quad +\boldsymbol{\beta}_{12}\boldsymbol{\beta}_{21}\left(\boldsymbol{\beta}_{11}\boldsymbol{\beta}_{42}-2\boldsymbol{\beta}_{12}\boldsymbol{\beta}_{41}+\boldsymbol{\beta}_{21}\boldsymbol{\beta}_{33}\right) \\
\mathbf{\Theta}_{6,4} &= \boldsymbol{\beta}_{21}\boldsymbol{\beta}_{43}+\boldsymbol{\beta}_{21}\left(\boldsymbol{\beta}_{43}+\boldsymbol{\beta}_{12}\boldsymbol{\beta}_{42}\right)+\boldsymbol{\beta}_{12}\boldsymbol{\beta}_{21}\boldsymbol{\beta}_{42} \\
\mathbf{\Theta}_{6,5} &= 2\boldsymbol{\beta}_{21}\left(\boldsymbol{\beta}_{11}\boldsymbol{\beta}_{42}-\boldsymbol{\beta}_{12}\boldsymbol{\beta}_{41}+\boldsymbol{\beta}_{21}\boldsymbol{\beta}_{33}\right) \\
&\quad +\boldsymbol{\beta}_{21}\left(\boldsymbol{\beta}_{11}\boldsymbol{\beta}_{42}-2\boldsymbol{\beta}_{12}\boldsymbol{\beta}_{41}+\boldsymbol{\beta}_{21}\boldsymbol{\beta}_{33}\right)-\boldsymbol{\beta}_{12}\boldsymbol{\beta}_{21}\boldsymbol{\beta}_{41} \\
\mathbf{\Theta}_{6,6} &= {\boldsymbol{\beta}_{21}}^2\left(\boldsymbol{\beta}_{32}-\boldsymbol{\beta}_{12}\boldsymbol{\beta}_{33}\right)-\boldsymbol{\beta}_{21}\boldsymbol{\beta}_{41} \\
&\quad +2\boldsymbol{\beta}_{21}\left(\boldsymbol{\beta}_{41}+\boldsymbol{\beta}_{21}\boldsymbol{\beta}_{32}-\boldsymbol{\beta}_{11}\boldsymbol{\beta}_{43}\right. \\
&\quad \left.-\boldsymbol{\beta}_{12}\left(\boldsymbol{\beta}_{11}\boldsymbol{\beta}_{42}-\boldsymbol{\beta}_{12}\boldsymbol{\beta}_{41}+\boldsymbol{\beta}_{21}\boldsymbol{\beta}_{33}\right)\right) \\
&\quad -\boldsymbol{\beta}_{11}\boldsymbol{\beta}_{21}\left(\boldsymbol{\beta}_{43}+\boldsymbol{\beta}_{12}\boldsymbol{\beta}_{42}\right) \\
&\quad -2\boldsymbol{\beta}_{12}\boldsymbol{\beta}_{21}\left(\boldsymbol{\beta}_{11}\boldsymbol{\beta}_{42}-\boldsymbol{\beta}_{12}\boldsymbol{\beta}_{41}+\boldsymbol{\beta}_{21}\boldsymbol{\beta}_{33}\right) \\
&\quad -\boldsymbol{\beta}_{12}\boldsymbol{\beta}_{21}\left(\boldsymbol{\beta}_{11}\boldsymbol{\beta}_{42}-2\boldsymbol{\beta}_{12}\boldsymbol{\beta}_{41}+\boldsymbol{\beta}_{21}\boldsymbol{\beta}_{33}\right)
\end{aligned}$$

Again, by examining the matrix determinant, we can show that $\mathbf{\Theta}$ is generically full rank. Specifically,

$$\begin{aligned} \det(\mathbf{\Theta}) &= 3\boldsymbol{\beta}_{21}^7\left(\boldsymbol{\beta}_{11}\boldsymbol{\beta}_{33}\boldsymbol{\beta}_{41}-\boldsymbol{\beta}_{32}\boldsymbol{\beta}_{42}-\boldsymbol{\beta}_{31}\boldsymbol{\beta}_{41}+\boldsymbol{\beta}_{12}\boldsymbol{\beta}_{33}\boldsymbol{\beta}_{42}\right) \\ &= 3\boldsymbol{\beta}_{21}^7\begin{bmatrix}\boldsymbol{\beta}_{11}\boldsymbol{\beta}_{33}-\boldsymbol{\beta}_{31} & \boldsymbol{\beta}_{12}\boldsymbol{\beta}_{33}-\boldsymbol{\beta}_{32}\end{bmatrix}\begin{bmatrix}\boldsymbol{\beta}_{41} \\ \boldsymbol{\beta}_{42}\end{bmatrix} \end{aligned} \quad (72)$$

We hereafter employ the definitions of the basis elements (see (37)–(40)) in order to analyze $\det(\mathbf{\Theta})$. As before, the first term $\boldsymbol{\beta}_{21} = \frac{1}{p_z}$ is strictly positive and finite. For the remaining two terms, it suffices to show that they and their product are generically non-zero.

Starting from the last term, we note that this is zero only when $\mathbf{b}_g = \boldsymbol{\beta}_4 = \mathbf{0}_{3\times1}$. However, this corresponds to a

different system whose system equations would need to be modified to reflect that its gyro is bias free.

The second term is a function of the feature observation, $\boldsymbol{\beta}_1 = \mathbf{h}$, and the velocity expressed in the local frame, $\boldsymbol{\beta}_3 = \mathbf{C}\,\mathbf{v}$, which can be written in a matrix vector form as

$$\begin{bmatrix}\boldsymbol{\beta}_{11}\,\boldsymbol{\beta}_{33} - \boldsymbol{\beta}_{31} & \boldsymbol{\beta}_{12}\,\boldsymbol{\beta}_{33} - \boldsymbol{\beta}_{32}\end{bmatrix}^{\mathrm{T}} = \mathbf{A}\boldsymbol{\beta}_3$$

where $\mathbf{A} = \begin{bmatrix}-\mathbf{I}_2 & \boldsymbol{\beta}_1\end{bmatrix}$. Since, generically, $\boldsymbol{\beta}_3 \neq \mathbf{0}_{3\times 1}$ (the camera is moving), and $\mathbf{A}$ is full column rank, their product cannot be zero. Thus, it suffices to examine the case for which

$$\begin{aligned}\begin{bmatrix}\boldsymbol{\beta}_{11}\,\boldsymbol{\beta}_{33} - \boldsymbol{\beta}_{31} & \boldsymbol{\beta}_{12}\,\boldsymbol{\beta}_{33} - \boldsymbol{\beta}_{32}\end{bmatrix}\begin{bmatrix}\boldsymbol{\beta}_{41}\\ \boldsymbol{\beta}_{42}\end{bmatrix} &= 0\\ \Leftrightarrow \begin{bmatrix}\boldsymbol{\beta}_{41} & \boldsymbol{\beta}_{42}\end{bmatrix}\mathbf{A}\boldsymbol{\beta}_3 &= 0\\ \Leftrightarrow \mathbf{A}\boldsymbol{\beta}_3 = \lambda\begin{bmatrix}\boldsymbol{\beta}_{42}\\ -\boldsymbol{\beta}_{41}\end{bmatrix},\ \lambda \in \mathbb{R}&\end{aligned}$$

This condition, for particular values of $\boldsymbol{\beta}_{41}$ and $\boldsymbol{\beta}_{42}$ (constant), and for time-varying values of $\boldsymbol{\beta}_1$ and hence $\mathbf{A}$, restricts $\boldsymbol{\beta}_3 = \mathbf{C}\,\mathbf{v}$ to always reside in a manifold. This condition, however, cannot hold given that arbitrary control inputs (linear acceleration and rotational velocity) are applied to the system.

We have shown that the diagonal elements of $\boldsymbol{\Xi}''$, i.e. $\boldsymbol{\Psi}$ and $\boldsymbol{\Theta}$ are both full rank (see (71) and (72)). We can now apply block-Gaussian elimination in order to show that $\boldsymbol{\Xi}''$ itself is full rank. Specifically, we begin by eliminating both $\mathbf{X}_{6\times 3}$ and $\mathbf{Y}_{6\times 3}$ using the identity matrix in the upper-left 3×3 sub block. Subsequently, we can eliminate $\mathbf{Z}_{6\times 6}$ using $\boldsymbol{\Psi}_{6\times 6}$ to obtain the following matrix whose columns span the same space:

$$\boldsymbol{\Xi}''' = \begin{bmatrix}\mathbf{I}_{3\times 3} & \mathbf{0}_{3\times 6} & \mathbf{0}_{3\times 6}\\ \mathbf{0}_{6\times 3} & \boldsymbol{\Psi}_{6\times 6} & \mathbf{0}_{6\times 6}\\ \mathbf{0}_{6\times 3} & \mathbf{0}_{6\times 6} & \boldsymbol{\Theta}_{6\times 6}\end{bmatrix}$$

Since the block-diagonal elements of $\boldsymbol{\Xi}'''$ are all full-rank, and all its off-diagonal block elements are zero, we conclude that $\boldsymbol{\Xi}'''$ is full column rank. This implies that the matrices $\boldsymbol{\Xi}''$, $\boldsymbol{\Xi}'$, and $\boldsymbol{\Xi}$ are also full column rank. Therefore system (43) is observable, and our defined basis functions comprise the observable modes of the original system (25).