On the consistency of Vision-aided Inertial Navigation

Dimitrios G. Kottas, Joel A. Hesch, Sean L. Bowman, and Stergios I. Roumeliotis

Abstract In this paper, we study estimator inconsistency in Vision-aided Inertial Navigation Systems (VINS). We show that standard (linearized) estimation approaches, such as the Extended Kalman Filter (EKF), can fundamentally alter the system observability properties, in terms of the number and structure of the unobservable directions. This in turn allows the influx of spurious information, leading to inconsistency. To address this issue, we propose an Observability-Constrained VINS (OC-VINS) methodology that explicitly adheres to the observability properties of the true system. We apply our approach to the Multi-State Constraint Kalman Filter (MSC-KF), and provide both simulation and experimental validation of the effectiveness of our method for improving estimator consistency.

1 Introduction

Many estimation problems in robotics, and in particular localization, involve nonlinear process and measurement models. Existing estimators, such as the Extended Kalman Filter (EKF), often suffer from inconsistency when applied to such tasks. As defined in [2], a state estimator is consistent if the estimation errors are zero-mean and have covariance smaller than or equal to the covariance calculated by the filter. In other words, an inconsistent estimator is overconfident in the accuracy of its estimates and its errors grow over time, possibly even causing divergence.

Julier and Uhlmann [14] first reported EKF inconsistency for 2D Simultaneous Localization and Mapping (SLAM). Since then, others have developed methods that seek to mitigate inconsistency in 2D SLAM (e.g., [1, 5]). However, little was known about the causes of inconsistency until recently. Specifically, in [9, 11] we have shown that a main cause of inconsistency is the mismatch between the observabil-

Dimitrios G. Kottas, Joel A. Hesch, Sean L. Bowman, and Stergios I. Roumeliotis
Dept. of Computer Science and Engineering, Univ. of Minnesota, Minneapolis, MN 55455, USA
e-mail: {dkottas|joel|bowman|stergios}@cs.umn.edu

ity properties of the linearized system used by the EKF [or the Unscented Kalman Filter (UKF)] and the underlying (true) nonlinear system. As a remedy, after identifying the unobservable directions either analytically [9] or numerically [10], we enforced them, either indirectly, by appropriately selecting the linearization points where the Jacobians are evaluated [11], or directly, by projecting the Jacobians onto the observable portion of the state space [10].

While most of the emphasis has been on 2D, very little is known about the inconsistency of 3D localization. This is primarily due to the complexity of the motion and measurement models involved in estimating a 15 (instead of 3) dimensional state. In this paper, we focus on Vision-aided Inertial Navigation Systems (VINS) that fuse data from a camera and an Inertial Measurement Unit (IMU) to track the six degrees-of-freedom (d.o.f.) pose of a sensing platform. Numerous VINS approaches have been presented in the literature, including methods based on the EKF [4, 22], UKF [6], and Batch-least Squares (BLS) [26]; however, these have not investigated the issue of VINS inconsistency.

The main contributions of this paper are the following:

- We study the observability properties of VINS and analytically determine the four unobservable directions (i.e., rotation about the gravity vector and global translation).
- We identify and solve the conditions that the VINS propagation and measurement Jacobians need to satisfy in order to ensure that the observability properties of the estimator match those of the true linearized system.
- We validate the proposed approach and demonstrate its capability to improve consistency through Monte-Carlo simulations and real-world experiments.

Although the proposed methodology is general enough to be applicable to any linearized estimator (e.g., EKF, UKF), in regular or inverse filter form, and regardless of the number of robot poses considered (smoother vs. filter), due to space limitations, we hereafter focus on Visual-Inertial Odometry (VIO) using the Multi-State Constraint Kalman Filter (MSC-KF) [22].

The remainder of this paper is organized as follows: We first review the related work (Sect. 2), followed by a description of the VINS model and the observability properties of VINS (Sect. 3). In Sect. 4, we introduce our Observability Constrained (OC)-VINS methodology for mitigating inconsistency, which we apply to the MSC-KF (Sect. 5). We validate our approach with simulation trials (Sect. 6) and real-world experiments (Sect. 7). Lastly, we provide our concluding remarks and discuss our future research directions (Sect. 8).

2 Related Work

For the task of IMU-camera extrinsic calibration, Mirzaei and Roumeliotis [19], as well as Kelly and Sukhatme [15], have analyzed the system observability using Lie derivatives [7] to determine when the IMU-camera transformation is observable.

Jones and Soatto [13] studied VINS observability by examining the indistinguishable trajectories of the system [12] under different sensor configurations (i.e., inertial only, vision only, vision and inertial). Martinelli [18] utilized the concept of continuous symmetries to show that the IMU biases, 3D velocity, and absolute roll and pitch angles are observable for VINS.

VINS inconsistency was recently investigated by Li and Mourikis [16]. Specifically, they studied the link between the VINS observability properties and estimator inconsistency for the bias-free case, and leveraged the First-Estimates Jacobian (FEJ) methodology of [9] to mitigate inconsistency in VIO. In contrast to their work, our approach has the advantage that any linearization method can be employed (e.g., computing Jacobians analytically, numerically, or using sample points) by the estimator. Additionally, we show that our approach is flexible enough to be applied in a variety of VINS problems such as VIO or SLAM.

In this work, we study the observability properties of the ideal linearized VINS model (i.e., the one whose Jacobians are evaluated at the true states), and show it has four unobservable d.o.f., corresponding to three-d.o.f. global translations and one-d.o.f. global rotation about the gravity vector. 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. To address this problem, we introduce a modification of the EKF-based VINS where its estimated Jacobians are updated so as to ensure that the number of unobservable directions is the same as when using the true Jacobians. In this manner, the global rotation about the gravity vector remains unobservable (as it should) and the consistency of the VINS EKF is significantly improved.

3 VINS Model

In what follows, we present an overview of the propagation and measurement models that describe the general case of a VINS, along with the main results of our observability analysis. For simplicity, we consider the case of a single physical feature being observed over multiple time steps; however, the analysis is straightforward to extend to the multiple feature case.

3.1 State vector and propagation model

The 19×1 system state includes the IMU pose and linear velocity together with the time-varying IMU biases and the 3D coordinates of the feature (see Fig. 1), i.e.,

$$\mathbf{x} = \begin{bmatrix} {}^I\bar{q}_G^T & \mathbf{b}_g^T & {}^G\mathbf{v}_I^T & \mathbf{b}_a^T & {}^G\mathbf{p}_I^T \mid {}^G\mathbf{f}^T \end{bmatrix}^T \tag{1}$$
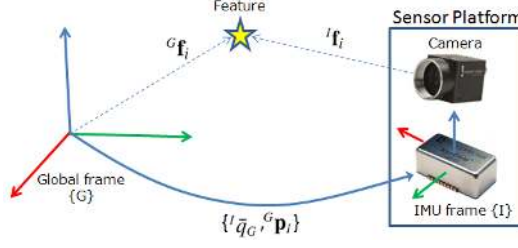
Fig. 1: Sensor platform comprising an IMU and a camera. $\{^I\bar{q}_G, {}^G\mathbf{p}_I\}$ are the quaternion of orientation and position vector describing the pose of the sensing frame $\{I\}$ with respect to the global frame $\{G\}$. The feature's 3D coordinates in $\{G\}$ and $\{I\}$ are denoted as $^G\mathbf{f}$, and $^I\mathbf{f}$, respectively.

where $^I\bar{q}_G(t)$, $^G\mathbf{p}_I(t)$, and $^G\mathbf{v}_I(t)$ are the orientation, position, and velocity of the IMU frame $\{I\}$ with respect to the global frame $\{G\}$, $\mathbf{b}_g(t)$ and $\mathbf{b}_a(t)$ are the gyroscope and accelerometer biases, and $^G\mathbf{f}$ is the feature's position expressed in $\{G\}$.

The system model describing the time evolution of the state is (see [27]):

$$^I\dot{\bar{q}}_G(t) = \tfrac{1}{2}\Omega(^I\omega(t))^I\bar{q}_G(t), \quad {}^G\dot{\mathbf{p}}_I(t) = {}^G\mathbf{v}_I(t), \quad {}^G\dot{\mathbf{v}}_I(t) = {}^G\mathbf{a}(t) \tag{2}$$

$$\dot{\mathbf{b}}_g(t) = \mathbf{n}_{wg}(t), \quad \dot{\mathbf{b}}_a(t) = \mathbf{n}_{wa}(t), \quad {}^G\dot{\mathbf{f}}(t) = 0 \ , \tag{3}$$

where $^I\omega$ and $^G\mathbf{a}$ are the rotational velocity and linear acceleration, and

$$\Omega(\omega) \triangleq \begin{bmatrix} -\lfloor \omega \times \rfloor & \omega \\ -\omega^T & 0 \end{bmatrix}, \quad \lfloor \omega \times \rfloor \triangleq \begin{bmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{bmatrix}. \tag{4}$$

Here, $\lfloor \omega \times \rfloor$ denotes the skew-symmetric matrix parameterized by ω. The time-varying biases are modeled as random-walk processes driven by white zero-mean Gaussian noise $\mathbf{n}_{wg}(t)$ and $\mathbf{n}_{wa}(t)$, respectively, with autocorrelations $\mathbb{E}[\mathbf{n}_{\mathbf{wg}}(t)\mathbf{n}_{\mathbf{wg}}{}^T(\tau)] = \mathbf{Q}_{wg}\delta(t-\tau)$ and $\mathbb{E}[\mathbf{n}_{\mathbf{wa}}(t)\mathbf{n}_{\mathbf{wa}}{}^T(\tau)] = \mathbf{Q}_{wa}\delta(t-\tau)$. The gyroscope and accelerometer measurements, ω_m and $\mathbf{a}_m$, are:

$$\omega_m(t) = {}^I\omega(t) + \mathbf{b}_g(t) + \mathbf{n}_g(t) \tag{5}$$

$$\mathbf{a}_m(t) = \mathbf{C}(^I\bar{q}_G(t))\,(^G\mathbf{a}(t) - {}^G\mathbf{g}) + \mathbf{b}_a(t) + \mathbf{n}_a(t). \tag{6}$$

The noise terms, $\mathbf{n}_g$ and $\mathbf{n}_a$, are modeled as zero-mean white Gaussian random processes. The gravitational acceleration $^G\mathbf{g}$ is known with respect to the global frame $\{G\}$. The matrix $\mathbf{C}(\bar{q})$ is the rotation matrix corresponding to the quaternion $\bar{q}$. Lastly, the time derivative of $^G\mathbf{f}$ is zero, since the camera observes a static scene.

Linearizing at the current estimates and applying the expectation operator on both sides of (2)-(3), we obtain the state estimate propagation model

$$^I\dot{\hat{\bar{q}}}_G(t) = \frac{1}{2}\Omega(^I\hat{\omega}(t))^I\hat{\bar{q}}_G(t), \quad {}^G\dot{\hat{\mathbf{p}}}_I(t) = {}^G\hat{\mathbf{v}}_I(t) \ , \quad {}^G\dot{\hat{\mathbf{v}}}_I(t) = \mathbf{C}^T(^I\hat{\bar{q}}_G(t))\hat{\mathbf{a}}(t) + {}^G\mathbf{g} \quad (7)$$

$$\dot{\hat{\mathbf{b}}}_g(t) = \mathbf{0}_{3\times 1} \ , \quad \dot{\hat{\mathbf{b}}}_a(t) = \mathbf{0}_{3\times 1}, \quad {}^G\dot{\hat{\mathbf{f}}}(t) = 0 \quad (8)$$

where $\hat{\mathbf{a}}(t) = \mathbf{a}_m(t) - \hat{\mathbf{b}}_a(t)$, and $^I\hat{\omega}(t) = \omega_m(t) - \hat{\mathbf{b}}_g(t)$.

The 18×1 error-state vector is defined as

$$\widetilde{\mathbf{x}} = \left[\ ^I\delta\theta_G^T \ \widetilde{\mathbf{b}}_g^T \ {}^G\widetilde{\mathbf{v}}_I^T \ \widetilde{\mathbf{b}}_a^T \ {}^G\widetilde{\mathbf{p}}_I^T \ {}^G\widetilde{\mathbf{f}}^T \right]^T. \quad (9)$$

For the IMU position, velocity, biases, and the observed feature, an additive error model is utilized (i.e., $\tilde{\mathbf{y}} = \mathbf{y} - \hat{\mathbf{y}}$ is the error in the estimate $\hat{\mathbf{y}}$ of a quantity $\mathbf{y}$), while for the quaternion we employ a multiplicative error model [27]. The main advantage of such an error model is that it allows us to represent the attitude uncertainty by a 3×3 covariance matrix, which is a minimal representation.

The linearized continuous-time error-state equation is

$$\dot{\widetilde{\mathbf{x}}}(t) = \begin{bmatrix} \mathbf{F}_s(t) & \mathbf{0}_{15\times 3} \\ \mathbf{0}_{3\times 15} & \mathbf{0}_3 \end{bmatrix} \widetilde{\mathbf{x}}(t) + \begin{bmatrix} \mathbf{G}_s(t) \\ \mathbf{0}_{3\times 12} \end{bmatrix} \mathbf{n} = \mathbf{F}_c(t)\widetilde{\mathbf{x}}(t) + \mathbf{G}_c(t)\mathbf{n} \quad (10)$$

where $\mathbf{n} = \left[\mathbf{n}_g^T \ \mathbf{n}_{wg}^T \ \mathbf{n}_a^T \ \mathbf{n}_{wa}^T \right]^T$, $\mathbf{F}_c$ is the error-state transition matrix, and $\mathbf{G}_c$ is the input noise matrix, with

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

The discrete-time state transition matrix from time t_1 to t, $\Phi(t,t_1)$, is computed in analytical form [8] as the solution to the matrix differential equation $\dot{\Phi}(t,t_1) = \mathbf{F}_c(t)\Phi(t,t_1)$, with initial condition $\Phi(t_1,t_1) = \mathbf{I}_{18}$. As we show in [8], the structure of $\Phi(t_{k+1},t_k) = \Phi(t_{k+1},t_1)\Phi(t_k,t_1)^{-1}$ when the state contains the IMU pose, velocity, biases, and a single landmark is given by

$$\Phi_{k+1} = \Phi(t_{k+1},t_k) = \begin{bmatrix} \Phi_{11} & \Phi_{12} & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \Phi_{31} & \Phi_{32} & \mathbf{I}_3 & \Phi_{34} & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \Phi_{51} & \Phi_{52} & \delta t\mathbf{I}_3 & \Phi_{54} & \mathbf{I}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{I}_3 \end{bmatrix}. \quad (11)$$

In the ensuing analysis, we consider $k \geq 1$ and define $\Phi_1 := \Phi(t_1,t_1) = \mathbf{I}_{18}$.

3.2 Measurement model

As the sensor platform moves in the environment, the camera observes point features, which are tracked across images. Generally, in a VINS [21], these measurements are exploited to concurrently estimate the motion of the sensing platform and, optionally, the structure of the environment.

We employ the pinhole camera model to describe the perspective projection of the 3D point $\mathbf{f}$ on the image plane and model the measurement $\mathbf{z}_k$ at time step t_k, i.e.,

$$\mathbf{z}_k = 1/z \left[x \; y\right]^T + \eta_k, \quad \left[x \; y \; z\right]^T = {}^I\mathbf{f} = \mathbf{C}\left({}^I\overline{q}_G\right)\left({}^G\mathbf{f} - {}^G\mathbf{p}_I\right), \tag{12}$$

where $\mathbf{C}\left({}^I\overline{q}_G\right)$ is the rotation matrix from $\{G\}$ to $\{I\}$ and η_k follows a Gaussian distribution with $\mathbb{E}[\eta_k] = \mathbf{0}_{2\times 1}$ and $\mathbb{E}[\eta_k \eta_k^T] = \sigma_\eta^2 \mathbf{I}_2$. Note also that, without loss of generality, we express the image measurement in normalized pixel coordinates, and consider the camera frame to be coincident with the IMU[1]. By differentiating the nonlinear measurement model (12), we obtain the measurement Jacobian:

$$\mathbf{H}_k = \mathbf{H}_{cam}\left[\mathbf{H}_{\theta_G} \; \mathbf{0}_{3\times 9} \; \mathbf{H}_{\mathbf{p}_I} \mid \mathbf{H}_{\mathbf{f}}\right] \tag{13}$$

$$\mathbf{H}_{cam} = \frac{1}{z}\begin{bmatrix} 1 & 0 & \frac{-x}{z} \\ 0 & 1 & \frac{-y}{z} \end{bmatrix}, \quad \mathbf{H}_{\theta_G} = \lfloor {}^I\mathbf{f}\times\rfloor, \quad \mathbf{H}_{\mathbf{p}_I} = -\mathbf{C}\left({}^I\bar{q}_G\right), \quad \mathbf{H}_{\mathbf{f}} = \mathbf{C}\left({}^I\bar{q}_G\right).$$

3.3 System observability analysis

In order to compute the analytical expressions of the four unobservable directions, we form the observability matrix $\mathbf{M}(\mathbf{x}^*) = \mathbf{M}$, as a function of the linearization point $\mathbf{x}^*$, for a system observing a feature over time steps $t_1 \ldots t_N$, i.e.,

$$\mathbf{M} = \begin{bmatrix} \mathbf{H}_1 \Phi_1 \\ \mathbf{H}_2 \Phi_2 \Phi_1 \\ \vdots \\ \mathbf{H}_N \Phi_N \cdots \Phi_1 \end{bmatrix}. \tag{14}$$

Lemma: In VINS, when using the true state to evaluate $\mathbf{H}$ and Φ, each block row of the observability matrix has the following form

$$\mathbf{M}_k = \mathbf{H}_{cam,k}\mathbf{C}\left({}^I\bar{q}_{G,k}\right)\left[\lfloor {}^G\mathbf{f} - {}^G\mathbf{p}_{I,1} - {}^G\mathbf{v}_{I,1}\delta t_{k-1} + \tfrac{1}{2}{}^G\mathbf{g}\delta t_{k-1}^2 \times\rfloor \mathbf{C}\left({}^I\bar{q}_{G,1}\right)^T \; \mathbf{D}_k \; -\mathbf{I}\delta t_{k-1} \; \mathbf{E}_k \; -\mathbf{I}_3 \; \mathbf{I}_3\right],$$

where $\delta t_{k-1} = (k-1)\delta t$, and $\mathbf{D}_k$ and $\mathbf{E}_k$ are time-varying matrices.
   Proof: See [8].
Theorem: The right nullspace of the observability matrix of a VINS spans four directions, i.e.,

---

[1] We perform both intrinsic and extrinsic camera/IMU calibration off-line [3, 19].

$$\mathbf{M}\mathbf{N}_1 = \mathbf{0}, \quad \mathbf{N}_1 = \begin{bmatrix} \mathbf{0}_3 & \mathbf{C}\left({}^I\bar{q}_{G,1}\right){}^G\mathbf{g} \\ \mathbf{0}_3 & \mathbf{0}_{3\times 1} \\ \mathbf{0}_3 & -\lfloor {}^G\mathbf{v}_{I,1} \times \rfloor {}^G\mathbf{g} \\ \mathbf{0}_3 & \mathbf{0}_{3\times 1} \\ \mathbf{I}_3 & -\lfloor {}^G\mathbf{p}_{I,1} \times \rfloor {}^G\mathbf{g} \\ \mathbf{I}_3 & -\lfloor {}^G\mathbf{f} \times \rfloor {}^G\mathbf{g} \end{bmatrix} = \begin{bmatrix} \mathbf{N}_{R,1} \\ \mathbf{N}_{f,1} \end{bmatrix} \tag{15}$$

where the 15×4 matrix $\mathbf{N}_{R,1}$ comprises the nullspace elements corresponding to the sensor platform (robot) state, and the 3×4 matrix $\mathbf{N}_{f,1}$ corresponds to the feature. We note that, the first three columns of the matrix $\mathbf{N}_1$ correspond to global translations, while its fourth column corresponds to global rotations about ${}^G\mathbf{g}$.

Proof: See [8].

In the preceding analysis, the system is linearized at the true state, hence, the aforementioned unobservable directions correspond to the true linearized system model. In practice, we do not have access to the true state and hence we typically linearize at the current state estimate. However, this causes the observability matrix of the estimated system $\widehat{\mathbf{M}} = \mathbf{M}(\hat{\mathbf{x}})$ to have higher rank (a fact that can be easily verified by numerically evaluating $\widehat{\mathbf{M}}$ during any experiment).

4 Observability-Constrained VINS (OC-VINS)

Ideally, we would like to design a filter that adheres to the true unobservable directions of the system. However, this would require knowledge of the *true* sensor pose and landmark position, which is clearly unrealizable in practice. Alternatively, we require that the estimator adheres to the *number* and *structure* of nullspace directions by ensuring that $\widehat{\mathbf{M}}\widehat{\mathbf{N}}_1 = \mathbf{0}$ is satisfied for every block row of $\widehat{\mathbf{M}}$, i.e.,

$$\widehat{\mathbf{H}}_k\widehat{\boldsymbol{\Phi}}_k\ldots\widehat{\boldsymbol{\Phi}}_1\widehat{\mathbf{N}}_1 = \mathbf{0}, \quad k \geq 1. \tag{16}$$

We do so by appropriately modifying $\widehat{\boldsymbol{\Phi}}_k$ and $\widehat{\mathbf{H}}_k$ at each time step so that[2]

$$\widehat{\mathbf{N}}_{k+1} = \widehat{\boldsymbol{\Phi}}_{k+1}\widehat{\mathbf{N}}_k \ , \quad \widehat{\mathbf{H}}_k\widehat{\mathbf{N}}_k = \mathbf{0}, \quad k \geq 1 \tag{17}$$

where $\widehat{\mathbf{N}}_k$, $k \geq 1$ is computed analytically based on (15). We hereafter present our method for initializing the nullspace (Sect. 4.1), and employing the nullspace to preserve the system observability properties during the propagation (Sect. 4.2) and update (Sect. 4.3) steps of the filter.

---

[2] Although this could also be accomplished by appropriate selection of the linearization points (as in [9]), we instead choose to employ information projections of the Jacobians $\widehat{\mathbf{H}}_k$ and $\widehat{\boldsymbol{\Phi}}_k$ since this allows more freedom in the way that we compute these matrices (e.g., analytically, using sample points as in the UKF, or through numerical integration for $\widehat{\boldsymbol{\Phi}}_k$).

4.1 Nullspace definition at time step k

At each time step, we compute the nullspace $\widehat{\mathbf{N}}_k$ as a function of the state estimate, and use it to enforce the unobservable directions. For the robot state, the initial nullspace as well as the nullspace at all subsequent times are [8]

$$
\widehat{\mathbf{N}}_{R,1} = \begin{bmatrix} \mathbf{0}_3 & \mathbf{C}\left(^I\hat{\bar{q}}_{G,1|1}\right){}^G\mathbf{g} \\ \mathbf{0}_3 & \mathbf{0}_{3\times 1} \\ \mathbf{0}_3 & -\lfloor{}^G\hat{\mathbf{v}}_{I,1|1}\times\rfloor{}^G\mathbf{g} \\ \mathbf{0}_3 & \mathbf{0}_{3\times 1} \\ \mathbf{I}_3 & -\lfloor{}^G\hat{\mathbf{p}}_{I,1|1}\times\rfloor{}^G\mathbf{g} \end{bmatrix} \quad,\quad \widehat{\mathbf{N}}_{R,k} = \begin{bmatrix} \mathbf{0}_3 & \mathbf{C}\left(^I\hat{\bar{q}}_{G,k|k-1}\right){}^G\mathbf{g} \\ \mathbf{0}_3 & \mathbf{0}_{3\times 1} \\ \mathbf{0}_3 & -\lfloor{}^G\hat{\mathbf{v}}_{I,k|k-1}\times\rfloor{}^G\mathbf{g} \\ \mathbf{0}_3 & \mathbf{0}_{3\times 1} \\ \mathbf{I}_3 & -\lfloor{}^G\hat{\mathbf{p}}_{I,k|k-1}\times\rfloor{}^G\mathbf{g} \end{bmatrix}, \quad (18)
$$

where the notation $\hat{\mathbf{x}}_{i|j}$ denotes the estimate of quantity $\mathbf{x}$ at time-step i computed using measurements up to time-step j. For each feature, the corresponding nullspace block element is

$$
\widehat{\mathbf{N}}_{f,k} = \begin{bmatrix} \mathbf{I}_3 & -\lfloor{}^G\hat{\mathbf{f}}_{\ell|\ell}\times\rfloor{}^G\mathbf{g} \end{bmatrix}, \quad (19)
$$

where $^G\hat{\mathbf{f}}_{\ell|\ell}$ is the first estimate of the feature's position, initialized at time step ℓ.

4.2 OC propagation: modification of the state transition matrix Φ

During each propagation step, we must ensure that $\widehat{\mathbf{N}}_{k+1} = \widehat{\Phi}_{k+1}\widehat{\mathbf{N}}_k$. We note that the first block-column of this constraint is automatically satisfied by the structure of $\widehat{\Phi}_{k+1}$ [see (20)], so we focus on the fourth column of $\widehat{\mathbf{N}}_{k+1} = \widehat{\Phi}_{k+1}\widehat{\mathbf{N}}_k$, which we write element-wise as:

$$
\begin{bmatrix} \mathbf{C}\left(^I\hat{\bar{q}}_{G,k+1|k}\right){}^G\mathbf{g} \\ \mathbf{0}_{3\times 1} \\ -\lfloor{}^G\hat{\mathbf{v}}_{I,k+1|k}\times\rfloor{}^G\mathbf{g} \\ \mathbf{0}_{3\times 1} \\ -\lfloor{}^G\hat{\mathbf{p}}_{I,k+1|k}\times\rfloor{}^G\mathbf{g} \\ -\lfloor{}^G\hat{\mathbf{f}}_{\ell|\ell}\times\rfloor{}^G\mathbf{g} \end{bmatrix} = \begin{bmatrix} \widehat{\Phi}_{11} & \widehat{\Phi}_{12} & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \widehat{\Phi}_{31} & \widehat{\Phi}_{32} & \mathbf{I}_3 & \widehat{\Phi}_{34} & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \widehat{\Phi}_{51} & \widehat{\Phi}_{52} & \delta t\mathbf{I}_3 & \widehat{\Phi}_{54} & \mathbf{I}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{I}_3 \end{bmatrix} \begin{bmatrix} \mathbf{C}\left(^I\hat{\bar{q}}_{G,k|k-1}\right){}^G\mathbf{g} \\ \mathbf{0}_{3\times 1} \\ -\lfloor{}^G\hat{\mathbf{v}}_{I,k|k-1}\times\rfloor{}^G\mathbf{g} \\ \mathbf{0}_{3\times 1} \\ -\lfloor{}^G\hat{\mathbf{p}}_{I,k|k-1}\times\rfloor{}^G\mathbf{g} \\ -\lfloor{}^G\hat{\mathbf{f}}_{\ell|\ell}\times\rfloor{}^G\mathbf{g} \end{bmatrix}. \quad (20)
$$

This results in constraints on the block-elements $\widehat{\Phi}_{11}$, $\widehat{\Phi}_{31}$, and $\widehat{\Phi}_{51}$ (see [8]). Specifically, for $\widehat{\Phi}_{11}$ we require that:

$$
\mathbf{C}\left(^I\hat{\bar{q}}_{G,k+1|k}\right){}^G\mathbf{g} = \widehat{\Phi}_{11}\mathbf{C}\left(^I\hat{\bar{q}}_{G,k|k-1}\right){}^G\mathbf{g} \quad \Rightarrow \widehat{\Phi}_{11} = \mathbf{C}\left(^{I,k+1|k}\hat{\bar{q}}_{I,k|k-1}\right). \quad (21)
$$

The constraints for $\widehat{\Phi}_{31}$ and $\widehat{\Phi}_{51}$ are

$$\widehat{\boldsymbol{\Phi}}_{31}\mathbf{C}\left({}^{I}\hat{\bar{q}}_{G,k|k-1}\right){}^{G}\mathbf{g} = \lfloor {}^{G}\hat{\mathbf{v}}_{I,k|k-1} - {}^{G}\hat{\mathbf{v}}_{I,k+1|k} \times \rfloor {}^{G}\mathbf{g} \tag{22}$$

$$\widehat{\boldsymbol{\Phi}}_{51}\mathbf{C}\left({}^{I}\hat{\bar{q}}_{G,k|k-1}\right){}^{G}\mathbf{g} = \lfloor \delta t {}^{G}\hat{\mathbf{v}}_{I,k|k-1} + {}^{G}\hat{\mathbf{p}}_{I,k|k-1} - {}^{G}\hat{\mathbf{p}}_{I,k+1|k} \times \rfloor {}^{G}\mathbf{g} \tag{23}$$

both of which are in the form $\mathbf{Au} = \mathbf{w}$, where $\mathbf{u}$ and $\mathbf{w}$ are nullspace elements that are known. We seek to find a perturbed $\mathbf{A}^*$, for $\mathbf{A} = \widehat{\boldsymbol{\Phi}}_{31}$ and $\mathbf{A} = \widehat{\boldsymbol{\Phi}}_{51}$ that fulfills the constraints (22) and (23). We formulate this as a minimization problem:

$$\min_{\mathbf{A}^*} ||\mathbf{A}^* - \mathbf{A}||_{\mathscr{F}}^2, \quad \text{subject to } \mathbf{A}^*\mathbf{u} = \mathbf{w} \tag{24}$$

where $||\cdot||_{\mathscr{F}}$ denotes the Frobenius matrix norm. Applying the method of Lagrange multipliers, we solve (24) in closed form as $\mathbf{A}^* = \mathbf{A} - (\mathbf{Au} - \mathbf{w})(\mathbf{u}^T\mathbf{u})^{-1}\mathbf{u}^T$.

We compute $\widehat{\boldsymbol{\Phi}}_{11}$ from (21), and $\widehat{\boldsymbol{\Phi}}_{31}$ and $\widehat{\boldsymbol{\Phi}}_{51}$ from (24) and construct the observability constrained discrete-time propagation Jacobian matrix.

4.3 OC update: modification of the measurement matrix H

During each update step, the measurement Jacobian must satisfy $\widehat{\mathbf{H}}_k \widehat{\mathbf{N}}_k = \mathbf{0}$, i.e.,

$$\widehat{\mathbf{H}}_{cam}\left[\widehat{\mathbf{H}}_{\theta_G}\ \mathbf{0}_{3\times 9}\ \widehat{\mathbf{H}}_{\mathbf{p}_I} \mid \widehat{\mathbf{H}}_{\mathbf{f}}\right]\begin{bmatrix} \mathbf{0}_3 & \mathbf{C}\left({}^{I}\hat{\bar{q}}_{G,k|k-1}\right){}^{G}\mathbf{g} \\ \mathbf{0}_3 & \mathbf{0}_{3\times 1} \\ \mathbf{0}_3 & -\lfloor {}^{G}\hat{\mathbf{v}}_{I,k|k-1}\times\rfloor{}^{G}\mathbf{g} \\ \mathbf{0}_3 & \mathbf{0}_{3\times 1} \\ \mathbf{I}_3 & -\lfloor {}^{G}\hat{\mathbf{p}}_{I,k|k-1}\times\rfloor{}^{G}\mathbf{g} \\ \mathbf{I}_3 & -\lfloor {}^{G}\hat{\mathbf{f}}_{\ell|\ell}\times\rfloor{}^{G}\mathbf{g} \end{bmatrix} = \mathbf{0}. \tag{25}$$

The first block column of $\widehat{\mathbf{N}}_k$ dictates that $\widehat{\mathbf{H}}_{\mathbf{f}} = -\widehat{\mathbf{H}}_{\mathbf{p}_I}$. We substitute this equality, and rewrite the fourth column of (25) in a compact form as

$$\widehat{\mathbf{H}}_{cam}\left[\widehat{\mathbf{H}}_{\theta_G}\ \widehat{\mathbf{H}}_{\mathbf{p}_I}\right]\begin{bmatrix} \mathbf{C}\left({}^{I}\hat{\bar{q}}_{G,k|k-1}\right){}^{G}\mathbf{g} \\ \lfloor {}^{G}\hat{\mathbf{f}}_{\ell|\ell} - {}^{G}\hat{\mathbf{p}}_{I,k|k-1}\times\rfloor{}^{G}\mathbf{g} \end{bmatrix} = \mathbf{0}. \tag{26}$$

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. We compute the optimal perturbed matrix $\mathbf{A}^*$ that fulfills (26), by solving a problem of the same form as (24) to obtain the modified elements of the measurement Jacobian. Specifically, after computing $\mathbf{A}^* = \mathbf{A} - \mathbf{Au}\left(\mathbf{u}^T\mathbf{u}\right)^{-1}\mathbf{u}^T$, we recover the measurement Jacobian elements as

$$\widehat{\mathbf{H}}_{cam}\widehat{\mathbf{H}}_{\theta_G} = \mathbf{A}^*_{1:2,1:3}\ ,\quad \widehat{\mathbf{H}}_{cam}\widehat{\mathbf{H}}_{\mathbf{p}_I} = \mathbf{A}^*_{1:2,4:6}\ ,\quad \widehat{\mathbf{H}}_{cam}\widehat{\mathbf{H}}_{\mathbf{f}} = -\mathbf{A}^*_{1:2,4:6} \tag{27}$$

where the subscripts (i:j, m:n) denote the submatrix spanning rows i to j, and columns m to n.

5 Application: Observability-Constrained MSC-KF (OC-MSC-KF)

The MSC-KF [20] 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 utilizes 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 [24] of the sensor pose. This enables the MSC-KF to utilize delayed image measurements; in particular, it allows all of the observations of a given feature $\mathbf{f}_i$ to be processed during a single update step (when the first pose that observed the feature is about to be marginalized). Whenever the current pose is cloned, we also clone the corresponding nullspace elements to obtain an augmented nullspace, i.e.,

$$\widehat{\mathbf{N}}_k^{aug} = \begin{bmatrix} \widehat{\mathbf{N}}_k \\ \widehat{\mathbf{N}}_{k,clone} \end{bmatrix}, \quad \text{where } \widehat{\mathbf{N}}_{k,clone} = \begin{bmatrix} \mathbf{0}_3 & \mathbf{C}\left({}^I\hat{\bar{q}}_{G,k|k-1}\right){}^G\mathbf{g} \\ \mathbf{I}_3 & -\lfloor {}^G\hat{\mathbf{p}}_{I,k|k-1} \times \rfloor {}^G\mathbf{g} \end{bmatrix}. \tag{28}$$

During propagation, the current state estimate evolves forward in time by integrating (7)-(8), while the current clone poses are static. We employ (21)-(23) to compute the observability-constrained discrete-time state transition matrix $\widehat{\Phi}_k$, and propagate the covariance as

$$\mathbf{P}_{k+1|k}^{aug} = \begin{bmatrix} \widehat{\Phi}_k & \mathbf{0}_{15\times 6m} \\ \mathbf{0}_{6m\times 15} & \mathbf{I}_{6m} \end{bmatrix} \mathbf{P}_{k|k}^{aug} \begin{bmatrix} \widehat{\Phi}_k^T & \mathbf{0}_{15\times 6m} \\ \mathbf{0}_{6m\times 15} & \mathbf{I}_{6m} \end{bmatrix} + \begin{bmatrix} \mathbf{Q}_k & \mathbf{0}_{15\times 6m} \\ \mathbf{0}_{6m\times 15} & \mathbf{0}_{6m} \end{bmatrix} \tag{29}$$

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 m-th clone (i.e., the one about to be marginalized from the sliding window of poses). We utilize (26) to compute the observability-constrained measurement Jacobian, $\widehat{\mathbf{H}}_k$, for each measurement and stack all observations of the i-th feature across m time steps into a large measurement vector

$$\begin{bmatrix} \tilde{\mathbf{z}}_k \\ \vdots \\ \tilde{\mathbf{z}}_{k-m} \end{bmatrix} = \begin{bmatrix} \widehat{\mathbf{H}}_k \\ \vdots \\ \widehat{\mathbf{H}}_{k-m} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{x}}^{aug} \\ \tilde{\mathbf{f}} \end{bmatrix} + \begin{bmatrix} \eta_k \\ \vdots \\ \eta_{k-m} \end{bmatrix} = \widehat{\mathbf{H}}_x \tilde{\mathbf{x}}^{aug} + \widehat{\mathbf{H}}_f \tilde{\mathbf{f}} + \eta \tag{30}$$

where $\widehat{\mathbf{H}}_x$ and $\widehat{\mathbf{H}}_f$ are the Jacobians corresponding to the augmented state vector $\tilde{\mathbf{x}}^{aug}$, and to the feature, respectively. To avoid including $\mathbf{f}$ into the state, we marginalize it *on-the-fly* by projecting (30) onto the left nullspace of $\widehat{\mathbf{H}}_f$, $\mathbf{W}$. This yields

$$\mathbf{W}^T\tilde{\mathbf{z}} = \mathbf{W}^T\widehat{\mathbf{H}}_x\tilde{\mathbf{x}}^{aug} + \mathbf{W}^T\eta \quad \Leftrightarrow \quad \tilde{\mathbf{z}}' = \widehat{\mathbf{H}}_x'\tilde{\mathbf{x}}^{aug} + \eta', \tag{31}$$

which we employ to update the state estimate and covariance using the standard EKF update equations.[3]
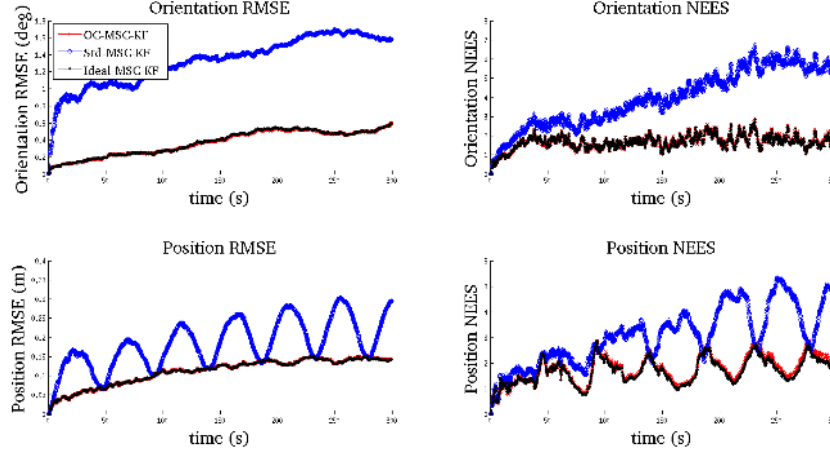

Fig. 2: The average RMSE and NEES over 30 Monte-Carlo simulation trials for orientation (above) and position (below). Note that the OC-MSC-KF attains performance indistinguishable from the Ideal-MSC-KF.

6 Simulation Results

We conducted Monte-Carlo simulations to evaluate the consistency of the proposed method applied to the MSC-KF [22]. Specifically, we compared the standard MSC-KF (Std-MSC-KF) with the Observability-Constrained MSC-KF (OC-MSC-KF) (see Sect. 4 and Sect. 5). We employed the Ideal-MSC-KF, whose Jacobians are linearized at the true states, as a benchmark, since it fulfills the observability properties of the true linearized system.

   We evaluated the Root Mean Squared Error (RMSE) and Normalized Estimation Error Squared (NEES) over 30 trials (see Fig. 2) in which the camera-IMU platform traversed a circular trajectory of radius 5 m at an average speed of 60 cm/s, and observed 50 randomly distributed features per image. The camera was modeled with a 45 deg field of view, and measurement noise with $\sigma_\eta = 1$ px. The IMU was modeled with MEMS quality sensors. As evident from Fig. 2, the OC-MSC-KF outperforms the Std-MSC-KF and attains performance almost indistinguishable from the Ideal-MSC-KF in terms of RMSE and NEES. This indicates that ensuring an

---

[3] The interested reader is referred to [20] for a more complete perspective.

estimator respects the observability properties of the true system, plays a key role in improving both the accuracy and consistency of VINS.

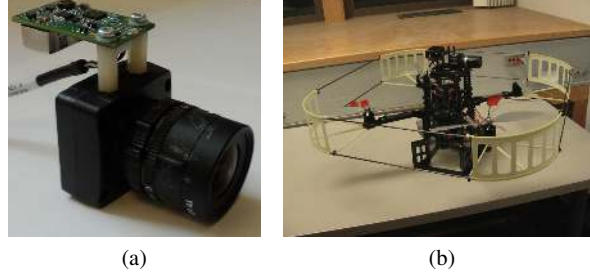<div align="center">(a)　　　　　　　　　　　　　　　　(b)</div>

Fig. 3: (a) The experimental testbed comprises a light-weight InterSense NavChip IMU and a Point Grey Chameleon Camera. The dimensions of the sensing package are approximately 6 cm tall, by 5 cm wide, by 8 cm deep. (b) An AscTech Pelican on which the camera-IMU package was mounted during the experiment.

7 Experimental Results

We further validated the proposed OC-MSC-KF on real-world data. Our hardware testbed consists of a Point Grey monochrome monocular camera with resolution 640x480 pixels and an InterSense NavChip IMU, both of which were rigidly mounted on an AscTec Pelican quadrotor (see Fig. 3). We begin with an overview of our image processing approach, followed by the experimental evaluation.

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 using the Shi-Tomasi corner detector [25] and track them pairwise through the window using the KLT tracking algorithm [17]. To remove outliers from the resulting tracks, we use a two-point algorithm to find the essential matrix between successive frames. Specifically, given the filter's estimated rotation (from the gyroscopes' measurements) between image i and j, ${}^i\hat{\bar{q}}_j$, we estimate the essential matrix from only two feature correspondences. This approach is more robust than the five-point algorithm [23] because it provides two solutions for the essential matrix rather than up to ten. Moreover, it requires only two data points, and thus 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 [24]. Upon completion of the image processing, all the features that first appeared at the oldest robot pose (corresponding to image $k-m+1$) are processed following the MSC-KF approach, as discussed in Sect. 5.

The sensor platform traversed three loops of total lenght 50 m in an indoor area and finally returned to its initial position. At the end of the trajectory, the Std-MSC-KF had a position error of 18.73 cm, while the final error for the OC-MSC-KF was 16.39 cm (approx. 0.38% and 0.33% of the distance traveled, respectively). In order to assess the impact of inconsistency on the orientation estimates of both methods, we used as ground truth the rotation between the first and last images computed independently using BLS and feature point matches. The Std-MSC-KF had final orientation error $\begin{bmatrix} 0.15 & -0.23 & -5.13 \end{bmatrix}$ deg for roll, pitch, and yaw (rpy), while the rpy errors for the OC-MSC-KF were $\begin{bmatrix} 0.19 & -0.20 & -1.32 \end{bmatrix}$ deg, respectively.

In addition to achieving higher accuracy, for yaw in particular, the OC-MSC-KF is more conservative since it strictly adheres to the unobservable directions of the system. This is evident in both the position and orientation uncertainties. We plot the y-axis position and yaw angle uncertainties in Fig. 4, as representative results. Most notably, the yaw uncertainty of the OC-MSC-KF remains approximately 1.13 deg (3σ), while for the Std-MSC-KF it reduces to 0.87 deg (3σ). This indicates that the Std-MSC-KF gains spurious orientation information, which leads to inconsistency. Lastly, we also show the 3D trajectory along with an overhead (x-y) view. It is evident that the Std-MSC-KF yaw error impacts the position accuracy, as the Std-MSC-KF trajectory exhibits a rotation with respect to the OC-MSC-KF.
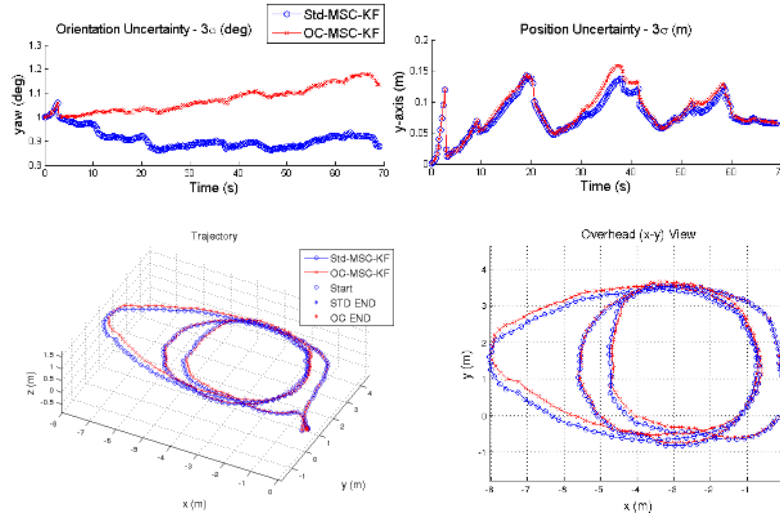

Fig. 4: (above) The position and orientation uncertainties (3σ bounds) for the yaw angle and the y-axis, which demonstrate that the Std-MSC-KF gains spurious information about its orientation. (below) The 3D trajectory and corresponding overhead (x-y) view.

8 Conclusion

In this paper, we analyzed a root cause of inconsistency in VINS, specifically, the gain of spurious information due to incorrect system observability properties of an EKF-based VINS estimator. We introduced an observability-constrained framework for explicitly enforcing the correct number and structure of unobservable directions by modifying the system and measurement Jacobians. We applied this methodology to the MSC-KF, and showed improved consistency both in simulations and through real-world experiments. Our future research directions include investigating other sources of VINS inconsistency, such as the existence of local minima, and extending our work to systems with multiple exteroceptive sensing modalities.

Acknowledgment

References

1. T. Bailey, J. Nieto, J. Guivant, M. Stevens, and E. Nebot. Consistency of the EKF-SLAM algorithm. In *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pages 3562–3568, Beijing, China, Oct. 9–15, 2006.
2. Y. Bar-Shalom, X. R. Li, and T. Kirubarajan. *Estimation with Applications to Tracking and Navigation*. John Wiley & Sons, New York, NY, 2001.
3. J.-Y. Bouguet. Camera calibration toolbox for matlab, 2006.
4. M. Bryson and S. Sukkarieh. Observability analysis and active control for airborne SLAM. *IEEE Trans. on Aerospace and Electronic Systems*, 44(1):261–280, Jan. 2008.
5. J. A. Castellanos, R. Martivez-Cantin, and J. Neira. Robocentric map joining: Improving the consistency of EKF-SLAM. *Robotics and Autonomous Systems*, 55(1):21–29, 2007.
6. S. Ebcin and M. Veth. Tightly-coupled image-aided inertial navigation using the unscented Kalman filter. Technical report, Air Force Institute of Technology, Dayton, OH, 2007.
7. R. Hermann and A. Krener. Nonlinear controllability and observability. *IEEE Trans. on Automatic Control*, 22(5):728–740, Oct. 1977.
8. J. A. Hesch, D. G. Kottas, S. L. Bowman, and S. I. Roumeliotis. Observability-constrained vision-aided inertial navigation. Technical Report 2012-001, University of Minnesota, Dept. of Comp. Sci. & Eng., MARS Lab, Feb. 2012.
9. G. P. Huang, A. I. Mourikis, and S. I. Roumeliotis. A first-estimates Jacobian EKF for improving SLAM consistency. In *Proc. of the Int. Symposium on Experimental Robotics*, pages 373–382, Athens, Greece, July 14–17, 2008.
10. G. P. Huang, A. I. Mourikis, and S. I. Roumeliotis. On the complexity and consistency of UKF-based SLAM. In *Proc. of the IEEE Int. Conf. on Robotics and Automation*, pages 4401–4408, Kobe, Japan, May 12–17, 2009.

11. G. P. Huang, A. I. Mourikis, and S. I. Roumeliotis. Observability-based rules for designing consistent EKF SLAM estimators. *Int. Journal of Robotics Research*, 29(5):502–528, Apr. 2010.

12. A. Isidori. *Nonlinear Control Systems*. Springer-Verlag, 1989.

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

14. S. J. Julier and J. K. Uhlmann. A counter example to the theory of simultaneous localization and map building. In *Proc. of the IEEE Int. Conf. on Robotics and Automation*, pages 4238–4243, Seoul, Korea, May 21–26, 2001.

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

16. M. Li and A. I. Mourikis. Improving the accuracy of EKF-based visual-inertial odometry. In *Proc. of the IEEE Int. Conf. on Robotics and Automation*, pages 828–835, St. Paul, MN, May 14–18, 2012.

17. B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proc. of the Int. Joint Conf. on Artificial Intelligence*, pages 674–679, Vancouver, B.C., Canada, Aug. 24–28, 1981.

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

19. F. M. Mirzaei and S. I. Roumeliotis. A Kalman filter-based algorithm for IMU-camera calibration: Observability analysis and performance evaluation. *IEEE Trans. on Robotics*, 24(5):1143–1156, Oct. 2008.

20. A. I. Mourikis and S. I. Roumeliotis. A multi-state constraint Kalman filter for vision-aided inertial navigation. In *Proc. of the IEEE Int. Conf. on Robotics and Automation*, pages 3565–3572, Rome, Italy, Apr. 10–14, 2007.

21. A. I. Mourikis and S. I. Roumeliotis. A dual-layer estimator architecture for long-term localization. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition Workshops*, pages 1–8, Anchorage, AK, June 2008.

22. A. I. Mourikis, N. Trawny, S. I. Roumeliotis, A. E. Johnson, A. Ansar, and L. Matthies. Vision-aided inertial navigation for spacecraft entry, descent, and landing. *IEEE Trans. on Robotics*, 25(2):264–280, Apr. 2009.

23. D. Nistér. An efficient solution to the five-point relative pose problem. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, pages 195–202, Madison, WI, June 16–22, 2003.

24. S. I. Roumeliotis and J. W. Burdick. Stochastic cloning: A generalized framework for processing relative state measurements. In *Proc. of the IEEE Int. Conf. on Robotics and Automation*, pages 1788–1795, Washington D.C., May 11-15, 2002.

25. J. Shi and C. Tomasi. Good features to track. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, pages 593–600, Washington, DC, June 27–July 2, 1994.

26. D. W. Strelow. *Motion estimation from image and inertial measurements*. PhD thesis, Carnegie Mellon University, Pittsburgh, PA, Nov. 2004.

27. N. Trawny and S. I. Roumeliotis. Indirect Kalman filter for 3D attitude estimation. Technical Report 2005-002, University of Minnesota, Dept. of Comp. Sci. & Eng., MARS Lab, Mar. 2005.