Visual-Inertial-Aided Online MAV System Identification

Chuchu Chen*, Yulin Yang*, Patrick Geneva, Woosik Lee, and Guoquan Huang

***Abstract*— System modeling and parameter identification of micro aerial vehicles (MAV) are crucial for robust autonomy, especially under highly dynamic motions. Visual-inertial-aided online parameter identification has recently seen research attention due to the demanding of adaptation to platform configuration changes with minimal onboard sensor requirements. To this end, we design an online MAV system identification algorithm to tightly fuse visual, inertial and MAV aerodynamic information within a lightweight multi-state constraint Kalman filter (MSCKF) framework. In particular, while one could blindly fuse the MAV dynamic-induced relative motion constraints in EKF, we numerically show that due to the (quadrotor) MAV system modeling inaccuracy, they often become overconfident and negatively impact the state estimates. As such, we leverage the Schmidt-Kalman filter (SKF) for MAV system parameter identification to prevent corruption of state estimates. Through extensive simulations and real-world experiments, we validate the proposed SKF-based scheme and demonstrate its ability to perform robust system identification even in the presence of an inconsistent MAV dynamic model under different motions.**

I. Introduction

Micro aerial vehicles (MAV) have continued to become miniaturized with increased maneuverability due to improved thrust-to-weight ratios [1]–[6]. Accurate state estimation and system identification of MAV platforms (e.g., geometric, inertial, and aerodynamic parameters) remain at the forefront of autonomous robotic research because of the challenges of high speed, under-actuation, and platform variability. As such, visual-inertial navigation systems (VINS), which estimate the state by fusing information from an inertial measurement unit (IMU) and aiding camera sensor(s), have seen significant research efforts due to their efficiency, complementary sensing nature, light-weight and low cost [7]. In particular, EKF-based visual-inertial odometry (VIO) methods remain promising thanks to their ability to deploy on resource-constrained MAV platforms [8]–[10].

VIMO [11] and VID-fusion [12] investigated the tightly-coupled fusion of MAV dynamic information and visual-inertial measurements by extending VINS-Mono [13], which is a sliding-window optimization framework. These methods require synchronized IMU and rotor encoder measurements to preintegrate the two into a single relative motion constraint, which may cause estimator inconsistency due to the re-use of inertial information. In addition, while they did not perform online parameter identification to handle variations, they did show that the MAV dynamics could improve VIO performance by modeling external force and pre-calibrated system parameters.

MAV system parameter identification plays an essential role in many control tasks [14], [15]. These parameters can change throughout a MAV's trajectory and be influenced by environmental effects. Therefore, traditional CAD models [16], static tests [17], or offline calibration [18], [19], cannot satisfy many autonomous MAV deployments. Burri et al. [18] showed the ability to offline estimate MAV parameters while relying only on onboard sensors. *Online* concurrent parameter identification is appealing for specific scenarios, e.g., the payload or attachment location changes [20].

Online MAV system identification methods can be categorized into two classes: (i) loosely-coupled methods which perform a secondary optimization problem separate from the navigation state estimation, and (ii) tightly-coupled methods where state estimation and parameter identification are performed within a single estimator. Wüest et al. [21] proposed an online loosely-coupled filter-based method which evolved the state forward using the MAV dynamics and updated using IMU and resulting pose from a VIO or other odometry solution. No rotations between the IMU, MAV body, and center of mass were modeled, and the thrust and moment aerodynamic coefficients were not estimated in this work. Recently, Böhm et al. [22] extended the work to estimate all MAV system parameters and performed observability analyses to determine what parameters are recoverable given the rotor speed, IMU, position, and/or pose measurements. However, these methods reuse rotor speed information which might lead to inconsistent estimation solutions and introduce an additional unmodeled source of error.

In contrast, we consistently estimate the system parameters by incorporating the MAV dynamics with visual-inertial measurements, avoiding re-use of information, relaxing additional sensor requirements, and rejecting measurements that do not follow the expected dynamics (e.g., an external gust of wind). Additionally, after carefully investigating the potential model errors of MAV dynamics, we leverage the Schmidt-KF (SKF) update for robust parameter identification, which tracks all the correlations between the visual-inertial state and system parameters in a tightly-coupled manner. Specifically, the main contributions of this paper include:

- A tightly-coupled lightweight MSCKF-based estimator is developed for state estimation and parameter identification with the MAV dynamic-induced measurements, which incorporate the platform's geometric, inertial, and aerodynamic parameters.
- A robust real-time SKF formulation of parameter identification is proposed to accurately determine system

This work was partially supported by the University of Delaware (UD) College of Engineering, the NSF (IIS-1924897, MRI-2018905, SCH-2014264), and the ARL (W911NF-19-2-0226, W911NF-20-2-0098). Geneva is also supported by the UD University Doctoral Fellowship.

The authors are with the Robot Perception and Navigation Group (RPNG), University of Delaware, Newark, DE 19716, USA. Email: {ccchu,yuyang,pgeneva,woosik,ghuang}@udel.edu

*These authors contributed equally to the work.

parameters even with a simplified MAV dynamic model while retaining consistent visual-inertial state estimates.
- The proposed method is validated with Monte-Carlo simulations based on a self-designed simulator and the real-world Blackbird UAV dataset [23], shown to be able to perform accurate and robust online MAV system identification.

II. MAV Dynamics and Its Integration

The frame of references for the MAV system is illustrated in Fig. 1. The MAV states $\mathbf{x}_M$[1] and system parameters $\mathbf{x}_\theta$ are defined as:

$$\mathbf{x}_M = \begin{bmatrix} {}^M_G\bar{q}^\top & {}^G\mathbf{p}_M^\top & {}^M\boldsymbol{\omega}^\top & {}^G\mathbf{v}_M^\top \end{bmatrix}^\top \tag{1}$$

$$\mathbf{x}_\theta = \begin{bmatrix} \mathbf{x}_D^\top & \mathbf{x}_G^\top & \mathbf{x}_{MI}^\top \end{bmatrix}^\top \tag{2}$$

where:

$$\mathbf{x}_D = \begin{bmatrix} c_t & c_m \end{bmatrix}^\top \tag{3}$$

$$\mathbf{x}_{MI} = \begin{bmatrix} {}^I_M\bar{q}^\top & {}^I\mathbf{p}_M^\top \end{bmatrix}^\top \tag{4}$$

$$\mathbf{x}_G = \begin{bmatrix} m & {}^M\mathbf{j}^\top & {}^M\mathbf{p}_B^\top \end{bmatrix}^\top \tag{5}$$

where ${}^M_G\bar{q}$ is the unit quaternion representing the rotation ${}^M_G\mathbf{R}$ from the global frame $\{G\}$ to the MAV center of mass frame $\{M\}$ [24], ${}^M\boldsymbol{\omega}$ is the angular velocity of in $\{M\}$, and ${}^G\mathbf{p}_M$ and ${}^G\mathbf{v}_M$ are the MAV position and velocity in $\{G\}$, respectively. The aerodynamic parameters, $\mathbf{x}_D$, consist of the rotor thrust coefficient c_t and moment coefficient c_m. The geometrical parameters, $\mathbf{x}_G$, include the mass of the platform m, moment of inertial ${}^M\mathbf{J}$, which is a diagonal matrix defined in frame $\{M\}$ with ${}^M\mathbf{j}$ being its diagonal terms. ${}^M\mathbf{p}_B$ represents the translation between the geometric body frame $\{B\}$ and $\{M\}$.

A. MAV Force and Moment

The total force ${}^M\mathbf{F}$ and moment ${}^M\mathbf{M}$ of a MAV with N_r rotors are defined as [18]:

$${}^M\mathbf{F} = \sum_{i=1}^{N_r} {}^M_{A_i}\mathbf{R}\,{}^{A_i}\mathbf{F}_i \tag{6}$$

$${}^M\mathbf{M} = \sum_{i=1}^{N_r} \left({}^M_{A_i}\mathbf{R}\,{}^{A_i}\mathbf{M} + \lfloor {}^M\mathbf{p}_{A_i} \rfloor {}^M\mathbf{F}_i \right) \tag{7}$$

where ${}^M\mathbf{p}_{A_i} = {}^M_B\mathbf{R}\,{}^B\mathbf{p}_{A_i} + {}^M\mathbf{p}_B$ denotes the translation between rotor $\{A_i\}$ and center of mass frame $\{M\}$, which is typically known from the CAD model. Assuming the rotor encoder measurements $r_{m,i}$ from the i-th rotor is defined as $r_{m,i} = r_i + n_{r,i}$ with $n_{r,i}$ represents the white Gaussian noise, the individual force ${}^{A_i}\mathbf{F}_i$ and moment ${}^{A_i}\mathbf{M}_i$ from rotor $\{A_i\}$ are approximated as [21]:

$${}^{A_i}\mathbf{F}_i = c_t(r_{m,i} - n_{r,i})^2\mathbf{e}_z + \mathbf{n}_{f,i} \tag{8}$$

$${}^{A_i}\mathbf{M}_i = c_m(r_{m,i} - n_{r,i})^2\lambda_i\mathbf{e}_z + \mathbf{n}_{m,i} \tag{9}$$

[1]Throughout this paper $\hat{\mathbf{x}}$ is used to denote the estimate of a random variable $\mathbf{x}$, while $\tilde{\mathbf{x}} = \mathbf{x} \boxminus \hat{\mathbf{x}}$ is the error of this estimate. We define the orientation error quaternion, $\delta\boldsymbol{\theta}$, as $\delta\bar{q} = \bar{q} \otimes \hat{\bar{q}}^{-1} \simeq [\frac{1}{2}\delta\theta^\top \; 1]^\top$ [24]. The updated estimate from a correction $\delta\mathbf{x}$ is $\hat{\mathbf{x}}^\oplus = \hat{\mathbf{x}} \boxplus \delta\mathbf{x}$.

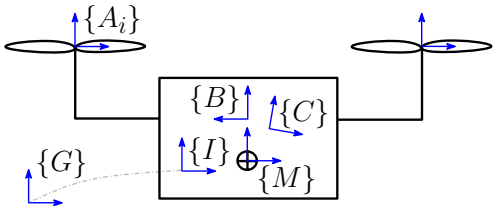

Fig. 1: Frame of references of a typical MAV system: i-th rotor frame $\{A_i\}$, geometric body frame $\{B\}$, IMU frame $\{I\}$, camera frame $\{C\}$, MAV center of mass frame $\{M\}$, and global frame $\{G\}$.

where $\lambda_i \in \{-1, 1\}$ corresponds to the rotation direction of the i-th rotor and $\mathbf{e}_z = [0, 0, 1]^\top$ is a unit vector along local z direction. As discussed in detail in Sec. IV-A, rotor speed can only measure along the local z-axis, thus, the additional noises, $\mathbf{n}_{f,i}$ and $\mathbf{n}_{m,i}$, are introduced to compensate for the inadequacy of the MAV measurements.

B. MAV Dynamic Model

The MAV dynamics are defined with the force and moment models as:

$$\begin{bmatrix} {}^M_G\dot{\bar{q}} \\ {}^G\dot{\mathbf{p}}_M \\ {}^M\dot{\boldsymbol{\omega}} \\ {}^G\dot{\mathbf{v}}_M \end{bmatrix} = \begin{bmatrix} \frac{1}{2}\boldsymbol{\Omega}\left({}^M\boldsymbol{\omega}\right) {}^M_G\bar{q} \\ {}^G\mathbf{v}_M \\ {}^M\mathbf{J}^{-1}\left({}^M\mathbf{M} - \lfloor {}^M\boldsymbol{\omega} \rfloor {}^M\mathbf{J}\,{}^M\boldsymbol{\omega}\right) \\ \frac{1}{m}{}^M_G\mathbf{R}^\top {}^M\mathbf{F} - {}^G\mathbf{g} \end{bmatrix} \tag{10}$$

where $\boldsymbol{\Omega}(\boldsymbol{\omega}) = \begin{bmatrix} -\lfloor \boldsymbol{\omega} \rfloor & \boldsymbol{\omega} \\ -\boldsymbol{\omega}^\top & 0 \end{bmatrix}$ and $\lfloor \cdot \rfloor$ is the skew-symmetric matrix; ${}^G\mathbf{g} = [0 \; 0 \; 9.81]^\top$ denotes the gravity. This can be summarized as:

$$\dot{\mathbf{x}}_M = \mathbf{f}_M\left(\mathbf{x}_M, \mathbf{x}_\theta, \mathbf{n}_M\right) \tag{11}$$

where $\mathbf{n}_M = [n_{r,1} \cdots n_{r,N_r} \; \mathbf{n}_{f,1}^\top \cdots \mathbf{n}_{f,N_r}^\top \; \mathbf{n}_{m,1}^\top \cdots \mathbf{n}_{m,N_r}^\top]^\top$ contains all noises [see Eq. (8) and (9)] with covariance $\mathbf{Q}_M$. We can integrate this model from time t_k to t_{k+1} based on first-order approximation:

$$\begin{bmatrix} {}^{M_{k+1}}_G\mathbf{R} \\ {}^G\mathbf{p}_{M_{k+1}} \\ {}^{M_{k+1}}\boldsymbol{\omega} \\ {}^G\mathbf{v}_{M_{k+1}} \end{bmatrix} = \begin{bmatrix} \Delta\mathbf{R}^\top {}^{M_k}_G\mathbf{R} \\ {}^G\mathbf{p}_{M_k} + \Delta\mathbf{p} \\ \Delta\mathbf{R}^\top\left({}^{M_k}\boldsymbol{\omega} + \Delta\boldsymbol{\omega}\right) \\ {}^G\mathbf{v}_{M_k} + \Delta\mathbf{v} \end{bmatrix} \tag{12}$$

with integration components derived as:

$$\begin{bmatrix} \Delta\mathbf{R} \\ \Delta\mathbf{p} \\ \Delta\boldsymbol{\omega} \\ \Delta\mathbf{v} \end{bmatrix} = \begin{bmatrix} \exp\left(\left({}^{M_k}\boldsymbol{\omega} + \frac{1}{2}\Delta\boldsymbol{\omega}\right)\Delta t\right) \\ {}^G\mathbf{v}_{M_k}\Delta t + \frac{1}{2}\left(\frac{1}{m}{}^{M_k}_G\mathbf{R}^\top {}^{M_k}\mathbf{F} - {}^G\mathbf{g}\right)\Delta t^2 \\ {}^{M_k}\mathbf{J}^{-1}\left({}^{M_k}\mathbf{M} - \lfloor {}^{M_k}\boldsymbol{\omega} \rfloor {}^{M_k}\mathbf{J}\,{}^{M_k}\boldsymbol{\omega}\right)\Delta t \\ \left(\frac{1}{m}{}^{M_k}_G\mathbf{R}^\top {}^{M_k}\mathbf{F} - {}^G\mathbf{g}\right)\Delta t \end{bmatrix} \tag{13}$$

where $\Delta t = t_{k+1} - t_k$ and $\exp(\cdot)$ denotes the matrix exponential [25]. We consider different sub-sets as:

- *Full model:* $\Delta\mathbf{R}$, $\Delta\mathbf{p}$, $\Delta\boldsymbol{\omega}$ and $\Delta\mathbf{v}$
- *Pose model:* $\Delta\mathbf{R}$ and $\Delta\mathbf{p}$
- *Orientation model:* $\Delta\mathbf{R}$ and $\Delta\boldsymbol{\omega}$

Therefore, the integrated discrete-time MAV dynamic model and its linearization are derived as:

$$\mathbf{x}_{M_{k+1}} = \mathbf{g}_M(\mathbf{x}_{M_k}, \mathbf{x}_\theta, \mathbf{n}_M) \tag{14}$$

$$\tilde{\mathbf{x}}_{M_{k+1}} \simeq \mathbf{\Phi}_M \tilde{\mathbf{x}}_{M_k} + \mathbf{\Phi}_\theta \tilde{\mathbf{x}}_\theta + \mathbf{G}_n \mathbf{n}_M \tag{15}$$

where $\mathbf{\Phi}_M$ is the linearized state transition matrix. $\mathbf{\Phi}_\theta$ and $\mathbf{G}_n$ represent the Jacobians for $\mathbf{x}_\theta$ and $\mathbf{n}_M$, respectively. Detailed derivations can be found in our technical report [26].

III. MSCKF-based Parameter Identification

We first extend the standard MSCKF-based estimator [27], [28] to additionally estimate MAV system parameters as:

$$\mathbf{x}_k = \begin{bmatrix}\mathbf{x}_A^\top & \mathbf{x}_\theta^\top\end{bmatrix}^\top, \quad \mathbf{x}_A^\top = \begin{bmatrix}\mathbf{x}_{I_k}^\top & \mathbf{x}_f^\top & \mathbf{x}_C^\top\end{bmatrix}^\top \tag{16}$$

where:

$$\mathbf{x}_{I_k} = \begin{bmatrix}{}^{I_k}_{G}\bar{q}^\top & {}^G\mathbf{p}_{I_k}^\top & {}^G\mathbf{v}_{I_k}^\top & \mathbf{b}_{g_k}^\top & \mathbf{b}_{a_k}^\top\end{bmatrix}^\top \tag{17}$$

$$\mathbf{x}_f = \begin{bmatrix}{}^G\mathbf{p}_{f1}^\top & \cdots & {}^G\mathbf{p}_{fn}^\top\end{bmatrix}^\top \tag{18}$$

$$\mathbf{x}_C = \begin{bmatrix}\mathbf{x}_{T_{k-1}}^\top & \cdots & \mathbf{x}_{T_{k-c}}^\top\end{bmatrix}^\top \tag{19}$$

$$\mathbf{x}_{T_i} = \begin{bmatrix}{}^{I_i}_{G}\bar{q}^\top & {}^G\mathbf{p}_{I_i}^\top & {}^{I_i}\boldsymbol{\omega}^\top & {}^G\mathbf{v}_{I_i}^\top\end{bmatrix}^\top \tag{20}$$

We define the "active" state $\mathbf{x}_A$, and parameter state $\mathbf{x}_\theta$ [see Eq. (2)]. The active state contains the current IMU state $\mathbf{x}_{I_k}$, n temporal SLAM features $\mathbf{x}_f$, and c historical clone states $\mathbf{x}_C$. Each historical clone contains the IMU pose, angular ${}^I\boldsymbol{\omega}$ and linear ${}^G\mathbf{v}_I$ velocities [29]. $\mathbf{b}_g$ and $\mathbf{b}_a$ are the gyroscope and accelerometer biases, respectively.

A. IMU Kinematic Model

The state evolves with the nonlinear IMU kinematics [30]. Integrating with incoming IMU linear accelerations, ${}^I\mathbf{a}_k$, and angular velocities, ${}^I\boldsymbol{\omega}_k$, we get:

$$\mathbf{x}_{I_{k+1}} = \mathbf{g}_I\left(\mathbf{x}_{I_k}, {}^I\mathbf{a}_k, {}^I\boldsymbol{\omega}_k, \mathbf{n}_I\right) \tag{21}$$

where $\mathbf{n}_I = [\mathbf{n}_g^\top \ \mathbf{n}_a^\top \ \mathbf{n}_{wg}^\top \ \mathbf{n}_{wa}^\top]^\top$ contains zero-mean white Gaussian noises and random walks of the IMU. We can then linearize the nonlinear model and propagate the state and covariance forward [27].

Remarks: We have made the conscious decision to propagate the state forward with IMU measurements as compared to use the MAV dynamics since IMU readings can fully measure the platform's trajectory evolution. Additionally, we can leverage robust outlier gating tests to reject invalid integrated MAV measurements caused by un-modeled forces.

B. Visual Feature Measurements

When exploring the environment, bearing observations of static landmarks are tracked from images. A bearing measurement seen at timestep k can be related to the state by (simplified for presentation, model in [28] is used):

$$\mathbf{z}_{C,k} = \mathbf{h}(\mathbf{x}_{T_k}, {}^G\mathbf{p}_f) + \mathbf{n}_{C,k} =: \mathbf{\Lambda}({}^{C_k}\mathbf{p}_f) + \mathbf{n}_{C,k} \tag{22}$$

$$\mathbf{\Lambda}\left([x \ y \ z]^\top\right) = \begin{bmatrix}x/z & y/z\end{bmatrix}^\top \tag{23}$$

$${}^{C_k}\mathbf{p}_f = {}^C_I\mathbf{R}\ {}^{I_k}_G\mathbf{R}({}^G\mathbf{p}_f - {}^G\mathbf{p}_{I_k}) + {}^C\mathbf{p}_I \tag{24}$$

where $\mathbf{n}_{C,k}$ is the white Gaussian measurement noise. We can obtain the visual residual based on Eq. (22):

$$\mathbf{r}_{C,k} = \mathbf{z}_{C,k} - \mathbf{h}(\hat{\mathbf{x}}_{T_k}, {}^G\hat{\mathbf{p}}_f) - \mathbf{n}_{C,k} \tag{25}$$

$$\simeq \mathbf{H}_{T_k}\tilde{\mathbf{x}}_{T_k} + \mathbf{H}_{f_k}{}^G\tilde{\mathbf{p}}_f - \mathbf{n}_{C,k} \tag{26}$$

where $\mathbf{H}_{T_k}$ and $\mathbf{H}_{f_k}$ are the measurement Jacobians, $\tilde{\mathbf{x}}_{T_k}$ and ${}^G\tilde{\mathbf{p}}_f$ are the error states for the pose and feature, respectively.

After collecting enough measurements and "stacking" the above linearized model, the feature can be used to update the state. Features are either inserted into the state vector and updated until lost, or used to directly update the state by marginalizing the feature position through the MSCKF nullspace projection [27], [31]. Consistency is ensured through first-estimate Jacobian techniques [32]–[34].

C. MAV Dynamic-Induced Measurements

We now present how to relate the integrated MAV dynamic model, Eq. (14), to our estimation state, Eq. (16). Through rigid body constraints we have:

$$\mathbf{x}_{M_k} = \mathbf{h}_{t,k}(\mathbf{x}_{T_k}, \mathbf{x}_\theta) \tag{27}$$

$$\begin{bmatrix}{}^{M_k}_G\mathbf{R} \\ {}^G\mathbf{p}_{M_k} \\ {}^{M_k}\boldsymbol{\omega} \\ {}^G\mathbf{v}_{M_k}\end{bmatrix} = \begin{bmatrix}{}^I_M\mathbf{R}^\top {}^{I_k}_G\mathbf{R} \\ {}^G\mathbf{p}_{I_k} + {}^{I_k}_G\mathbf{R}^\top {}^I\mathbf{p}_M \\ {}^M_I\mathbf{R}\,{}^{I_k}\boldsymbol{\omega} \\ {}^G\mathbf{v}_{I_k} + {}^{I_k}_G\mathbf{R}^\top \lfloor {}^{I_k}\boldsymbol{\omega} \rfloor {}^I\mathbf{p}_M\end{bmatrix} \tag{28}$$

The MAV measurement residual and error-state Jacobians are defined accordingly:

$$\mathbf{r}_{M,k} = \hat{\mathbf{x}}_{M_{k+1}} - \mathbf{g}_M(\hat{\mathbf{x}}_{M_k}, \hat{\mathbf{x}}_\theta, \mathbf{0}) \tag{29}$$

$$= \mathbf{h}_{t,k+1}(\hat{\mathbf{x}}_{T_{k+1}}, \hat{\mathbf{x}}_\theta) - \mathbf{g}_M(\mathbf{h}_{t,k}(\hat{\mathbf{x}}_{T_k}, \hat{\mathbf{x}}_\theta), \hat{\mathbf{x}}_\theta, \mathbf{0})$$

$$\simeq \begin{bmatrix}\mathbf{H}_{T_{k+1}} & \mathbf{H}_{T_k} & \mathbf{H}_\theta\end{bmatrix}\tilde{\mathbf{x}}_k - \mathbf{G}_n\mathbf{n}_M \tag{30}$$

$$\triangleq \begin{bmatrix}\mathbf{H}_A & \mathbf{H}_\theta\end{bmatrix}\tilde{\mathbf{x}}_k - \mathbf{G}_n\mathbf{n}_M \tag{31}$$

where $\tilde{\mathbf{x}}_k = [\tilde{\mathbf{x}}_{T_{k+1}}^\top \ \tilde{\mathbf{x}}_{T_k}^\top \ \tilde{\mathbf{x}}_\theta^\top]^\top \triangleq [\tilde{\mathbf{x}}_A^\top \ \tilde{\mathbf{x}}_\theta^\top]^\top$ is the error state. The Jacobians can be derived through the chainrule [see Eq.(15)]:

$$\mathbf{H}_{T_{k+1}} = \frac{\partial \mathbf{h}_{t,k+1}}{\partial\, \mathbf{x}_{T_{k+1}}}, \quad \mathbf{H}_{T_k} = -\mathbf{\Phi}_M \frac{\partial \mathbf{h}_{t,k}}{\partial\, \mathbf{x}_{T_k}} \tag{32}$$

$$\mathbf{H}_\theta = \frac{\partial \mathbf{h}_{t,k+1}}{\partial\, \mathbf{x}_\theta} - \mathbf{\Phi}_M \frac{\partial \mathbf{h}_{t,k}}{\partial\, \mathbf{x}_\theta} - \mathbf{\Phi}_\theta \tag{33}$$

The detailed derivations are in our companion technical report [26].

D. EKF Update

The standard EKF update corrects the state given either camera measurements (Sec. III-B) or MAV dynamic-induced measurements (Sec. III-C) as follows:

$$\hat{\mathbf{x}}_{k+1}^\oplus = \hat{\mathbf{x}}_{k+1} \boxplus \mathbf{K}\mathbf{r}, \quad \mathbf{P}_{k+1}^\oplus = \mathbf{P}_{k+1} - \mathbf{K}\mathbf{S}\mathbf{K}^\top \tag{34}$$

where $\mathbf{P}_{k+1}$ is the state covariance at time t_{k+1}, and:

$$\mathbf{S} = [\mathbf{H}_A \ \mathbf{H}_\theta]\mathbf{P}_{k+1}[\mathbf{H}_A \ \mathbf{H}_\theta]^\top + \mathbf{R} \tag{35}$$

$$\mathbf{K} = \mathbf{P}_{k+1}\begin{bmatrix}\mathbf{H}_A \\ \mathbf{H}_\theta\end{bmatrix}\mathbf{S}^{-1} \triangleq \begin{bmatrix}\mathbf{K}_A \\ \mathbf{K}_\theta\end{bmatrix} \tag{36}$$

where $\mathbf{K}_A$ and $\mathbf{K}_\theta$ denote the Kalman gain for the $\mathbf{x}_A$ and $\mathbf{x}_\theta$, respectively, and $\mathbf{R}$ is the measurement noise.

TABLE I: Average RMSE and NEES over 50 Monte-Carlo runs for a set of σs for the $\mathbf{n}_f$ and $\mathbf{n}_m$. VIO without MAV dynamics had an RMSE of 0.782 degrees and 0.195 meters with a NEES of 2.230 and 2.744, respectively. Ideally, the NEES should be 3. Bold-red values indicate estimate errors larger than the baseline VIO system.

σ	RMSE Ori. (deg.)			RMSE Pos. (m)			NEES Ori.			NEES Pos.		
	full	pose	ori	full	pose	ori	full	pose	ori	full	pose	ori
0.03	**2.806**	**1.271**	**3.468**	**52.589**	**0.202**	**0.309**	**1543.207**	**33.334**	**27.718**	**65.681**	**4.847**	**7.627**
0.05	**1.144**	**1.037**	**1.812**	0.191	0.188	**0.224**	**31.490**	**12.551**	**5.135**	**4.369**	3.951	3.777
0.20	**0.865**	0.777	0.718	0.190	0.188	0.188	**5.431**	3.381	2.170	3.745	3.515	2.867
0.50	0.737	0.704	0.751	0.189	0.188	0.189	2.960	2.429	2.183	3.430	3.180	2.863
1.00	0.704	0.725	0.758	0.187	0.189	0.189	2.404	2.251	2.183	3.151	2.982	2.866
1.50	0.708	0.739	0.758	0.188	0.189	0.189	2.280	2.213	2.182	3.304	2.952	2.867

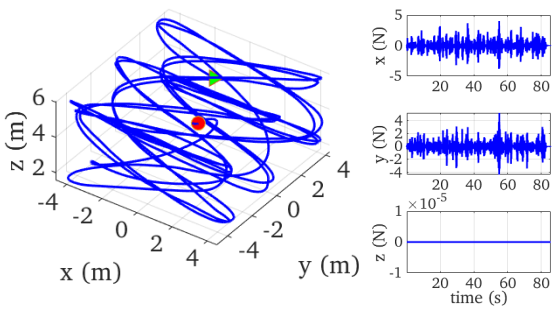

Fig. 2: Simulated trajectory for a quadrotor with total length of 323 meters (left). Differences between the recovered force ${}^{M}\mathbf{F}$ from MAV dynamics and groundtruth force (right).

IV. Numerical Analysis of MAV Dynamics

We have empirically discovered that blindly fusing the dynamic-induced measurements within the estimator often degrades the state estimation and thus parameter identification performance. To understand this phenomena, in this section, we numerically examine the model inaccuracy of the MAV dynamic model (6)-(9) and how the tuning of noise parameters $\mathbf{n}_{f,i}$ and $\mathbf{n}_{m,i}$ affects the system performance. This analysis motivates the proposed SKF that will be presented in Sec. V.

A. Demonstration of Model Inaccuracy

MAV rotor speeds can only be used to compute forces along the rotor rotation (local z) axis [see Eq. (8)], the lateral forces (on the local xy-plane of the rotor frame) cannot be captured. Thus, the dynamic model cannot accurately recover the *complete 3D* force for a MAV platform. The lateral forces (on the local xy-plane of the rotor frame) cannot be captured with rotor speed readings. This can be verified by a "realistic" simulation using a quadrotor MAV equipped with an IMU and a camera rig. The simulated MAV travels along a trajectory (left of Fig. 2) generated by RotorS [35]. The detailed simulation configurations are introduced in Sec. VI. The differences between the recovered force from MAV dynamics ${}^{M}\mathbf{F}$ [see Eq. (6)] and true force computed based on the trajectory differentiation [see Eq. (44)] are shown in the right of Fig. 2. While rotor speed can be used to accurately model the force along the z-direction, there are large non-zero differences along the xy-direction which confirms the inaccuracy of the force model.

B. Noise Parameter Tuning for MAV Model

To compensate for the model inaccuracy, two additive noise terms, $\mathbf{n}_{f,i}$ and $\mathbf{n}_{m,i}$, are introduced in Eq. (8) and (9). Since the total force and moment of the MAV system can be easily affected by the motion and environmental conditions, it is not trivial to decide what noise parameters best fit the estimator.

We herein implement a set of Monte-Carlo experiments to evaluate the system performances with different noise sigmas. We tune the noise sigmas by designing a single noise standard deviation σ related to our additive noise distribution by: $\mathbf{n}_{f,i} = [\sigma \ \ \sigma \ \ \frac{1}{10}\sigma]^\top$ and $\mathbf{n}_{m,i} = [\frac{1}{10}\sigma \ \ \frac{1}{10}\sigma \ \ \frac{1}{10}\sigma]^\top$. Note that we consider smaller noise sigmas for the moment and force z-axis, due to the fact that MAV dynamics can relatively accurately model them. The average Root Mean Squared Error (RMSE) [36] and Normalized Estimation Error Squared (NEES) [37] are used to evaluate the accuracy and consistency for the estimator, with all the results shown in Table I. The standard VIO without MAV updates achieves an average RMSE of 0.782 degrees and 0.195 meters, while the average NEES are 2.230 and 2.744 for orientation and position, respectively. All methods leverage the groundtruth calibration, and thus the errors are from the MAV dynamic model and measurements.

From Table I we note that the *full model* measurements are able to improve the pose estimation accuracy by properly tuning the noise parameters (as the case of $\sigma = 1.00$). Incorporating MAV measurements with $\sigma < 1.00$ leads to inconsistent estimation (with larger NEES), and substantially hurts the system accuracy (red values indicate worse accuracy than the benchmark VIO), and even causes the estimated trajectory to diverge. In contrast, for $\sigma > 1.00$ the system over inflates the MAV measurements and although still consistent, the estimation accuracy slightly drops (see full model with $\sigma = 1.50$).

Additionally, *full*, *pose* and *orientation* based updates are evaluated (see Sec. II-B for definition). For smaller σ (e.g., $\sigma = 0.20$), the *full* and *pose* updates easily degrade performance. The *orientation* based update generally remains consistent likely due to the platform moment being properly modeled. Its downside is that not all parameters can be calibrated (e.g., ${}^{I}\mathbf{p}_M$, and the mass m). This can be seen through inspection of Eq. (28) and (13). From these results, we notice that the *pose* update achieves comparable accuracy to the *full* update and both can identify all MAV parameters.

TABLE II: Configuration parameters and prior distributions for simulation and the estimator.

Parameter	Value	Parameter	Value	Parameter	Value	Prior
IMU Freq. (hz)	200	${}^{B}\mathbf{p}_{A_1}$ (m)	0.21, 0.00, 0.05	${}^{I}_{M}\bar{q}$	0, 0, 0, 1	2.86°
Cam Freq. (hz)	10	${}^{B}\mathbf{p}_{A_2}$ (m)	0.00, 0.21, 0.05	${}^{I}\mathbf{p}_{M}$ (m)	0, 0, 0	0.15
Rotor Freq. (hz)	300	${}^{B}\mathbf{p}_{A_3}$ (m)	-0.21, 0.00, 0.05	${}^{M}\mathbf{p}_{B}$ (m)	0, 0, 0	0.05
Pixel Noise (pix)	1	${}^{B}\mathbf{p}_{A_4}$ (m)	0.00, -0.21, 0.05	c_t (N s^2/rad^2)	9.9865e-06	5.0e-6
Rotor White Noise (rad/s)	0.043	λ_1,λ_2,λ_3,λ_4	1, -1, 1, -1	c_m (N s^2/rad^2)	1.455784e-7	1.0e-6
Gyro. White Noise	1.6968e-4	Accel. Rand. Walk	3.0000e-2	${}^{M}\mathbf{j}$	0.01, 0.01, 0.02	0.005
Accel. White Noise	2.0000e-2	Gyro. Rand. Walk	1.9393e-4	Mass (kg)	1	0.15

The *pose* model thus will be used in our later experiments.

Remarks: Based on the above discussions, we conclude that the force and moment noise parameters are crucial for the compensation of potential MAV dynamic inaccuracies when modeling the forces and moments. Proper noise parameter tuning is necessary for a consistent estimator but limits the application of the proposed system. To address this issue, we propose the ensuing SKF estimator design.

V. SKF-based Parameter Identification

As evident from the previous section, tightly-coupled EKF-based estimation strongly assumes accurate dynamic and measurement models. However, the commonly-used MAV dynamic model inherently possesses inaccuracy in practice, which, if not properly addressed, can be detrimental to parameter identification and state estimation. Therefore, inspired by [38], we propose to leverage the Schmidt-Kalman filter (SKF) formulation, which relaxes the requirements for noise parameter tuning and protects the visual-inertial state estimation performance, making the estimator suitable for practical applications.

Specifically, we partition the state vector into the active state $\mathbf{x}_A$ and parameter state $\mathbf{x}_\theta$ during the MAV measurement update process [see Eq. (16)]. The state estimate $\hat{\mathbf{x}}_{k+1}$ and the corresponding covariance $\mathbf{P}_{k+1}$ at timestep k+1 are represented as:

$$\hat{\mathbf{x}}_{k+1} = \begin{bmatrix} \mathbf{x}_{A_{k+1}} \\ \mathbf{x}_{\theta_{k+1}} \end{bmatrix}, \quad \mathbf{P}_{k+1} = \begin{bmatrix} \mathbf{P}_{AA} & \mathbf{P}_{A\theta} \\ \mathbf{P}_{A\theta}^\top & \mathbf{P}_{\theta\theta} \end{bmatrix} \tag{37}$$

By setting the Kalman gain related to the active state to zero $\mathbf{K}_A = \mathbf{0}$ from Eq. (36) , the Schmidt update equations are:

$$\hat{\mathbf{x}}^{\oplus}_{k+1} = \hat{\mathbf{x}}_{k+1} \boxplus \begin{bmatrix} \mathbf{0} \\ \mathbf{K}_\theta \end{bmatrix} \mathbf{r}_{M,k+1} \tag{38}$$

$$\mathbf{P}^{\oplus}_{k+1} = \mathbf{P}_{k+1} - \begin{bmatrix} \mathbf{0} & \Delta\mathbf{P}_{A\theta} \\ \Delta\mathbf{P}_{A\theta}^\top & \Delta\mathbf{P}_{\theta\theta} \end{bmatrix} \tag{39}$$

$$\Delta\mathbf{P}_{A\theta} = \mathbf{K}_A\mathbf{H}_x \begin{bmatrix} \mathbf{P}_{A\theta}^\top & \mathbf{P}_{\theta\theta}^\top \end{bmatrix}^\top, \ \Delta\mathbf{P}_{\theta\theta} = \mathbf{K}_\theta\mathbf{S}\mathbf{K}_\theta^\top \tag{40}$$

It is clear that the active state and its covariance do not change during the update, preventing the corruption of VIO if the MAV dynamic-induced measurements are inconsistent. However, the system parameters $\mathbf{x}_\theta$ will still be updated and SKF still tracks correlations in the covariance properly.

Remarks: We stress that the MAV dynamic-induced measurements are processed with the SKF update which will only correct $\mathbf{x}_\theta$ and their correlations. The visual measurements are processed with standard EKF update (see Sec. III-D) allowing state corrections to update $\mathbf{x}_A$ and indirectly refine $\mathbf{x}_\theta$ through the tracked correlations.

Alternatively, we can simply employ a decoupled Schmidt-Kalman filter (DSKF) update, which has the same mean update as SKF [see Eq. (38)] but different covariance update [see Eq. (39)] with:

$$\Delta\mathbf{P}_{A\theta} = \mathbf{0}\ , \quad \Delta\mathbf{P}_{\theta\theta} = \mathbf{K}_\theta\mathbf{S}\mathbf{K}_\theta^\top \tag{41}$$

This updated DSKF covariance drops all the correlations between $\mathbf{x}_A$ and $\mathbf{x}_\theta$ to prevent indirect updates.

VI. Monte-Carlo Simulations

The RotorS [35] MAV simulator was used to generate a *realistic* and *feasible* fully excited motion trajectory for a commonly-used quadrotor model (see left of Fig. 2). We leverage the OpenVINS [28] simulation framework, which fits a B-spline to the trajectory to allow for the calculation of groundtruth accelerations and velocities. From these, noisy IMU readings and corresponding biases, visual feature bearing measurements, and rotor speed measurements are generated. The key configurations and parameters are listed in Table II.

A. Rotor Speed Generation

With the B-spline formulation of the simulated trajectory, we can generate IMU angular velocity ${}^{I}\boldsymbol{\omega}$, angular acceleration ${}^{I}\boldsymbol{\alpha}$, and linear acceleration ${}^{G}\mathbf{a}_I$ at any desired time. The MAV motion states are computed with the rigid body constraints accordingly:

$${}^{M}\boldsymbol{\omega} = {}^{M}_{I}\mathbf{R}\,{}^{I}\boldsymbol{\omega} \tag{42}$$

$${}^{G}\mathbf{a}_M = {}^{G}\mathbf{a}_I + {}^{I}_{G}\mathbf{R}^\top \left(\lfloor {}^{I}\boldsymbol{\omega} \rfloor \lfloor {}^{I}\boldsymbol{\omega} \rfloor + \lfloor {}^{I}\boldsymbol{\alpha} \rfloor \right) {}^{I}\mathbf{p}_M \tag{43}$$

The total force and moment at a specific time are:

$$\begin{bmatrix} {}^{M}\mathbf{F} \\ {}^{M}\mathbf{M} \end{bmatrix} = \begin{bmatrix} m \cdot {}^{M}_{G}\mathbf{R} \left({}^{G}\mathbf{a}_M + {}^{G}\mathbf{g} \right) \\ {}^{M}\mathbf{J}\,{}^{M}\boldsymbol{\alpha}_M + \lfloor {}^{M}\boldsymbol{\omega}_M \rfloor {}^{M}\mathbf{J}\,{}^{M}\boldsymbol{\omega}_M \end{bmatrix} \tag{44}$$

Then, with Eq. (6) and (7), we formulate the following linear system to solve for the rotor speeds:

$$\begin{bmatrix} \mathbf{B}_1 & \mathbf{B}_2 & \mathbf{B}_3 & \mathbf{B}_4 \end{bmatrix} \cdot \mathbf{r} = \begin{bmatrix} {}^{M}\mathbf{F} \\ {}^{M}\mathbf{M} \end{bmatrix} \tag{45}$$

where $\mathbf{r} = [r_1^2 \ \cdots \ r_4^2]^\top$ and $\mathbf{B}_i$ is defined as:

$$\mathbf{B}_i = \begin{bmatrix} c_t\,{}^{M}_{A_i}\mathbf{R}\mathbf{e}_z \\ c_m\lambda_i\,{}^{M}_{A_i}\mathbf{R}\mathbf{e}_z + c_t \lfloor {}^{M}\mathbf{p}_{A_i}\mathbf{e}_z \rfloor \end{bmatrix} \tag{46}$$

Finally, white Gaussian noise $n_{r,i}$ is added to r_i for realistic simulation of rotor encoder measurement $r_{m,i}$ (e.g., $r_{m,i} =$

$r_i + n_{r,i}$). We note that rotor speeds generated from our simulator are similar to the one from RotorS [see [26]].

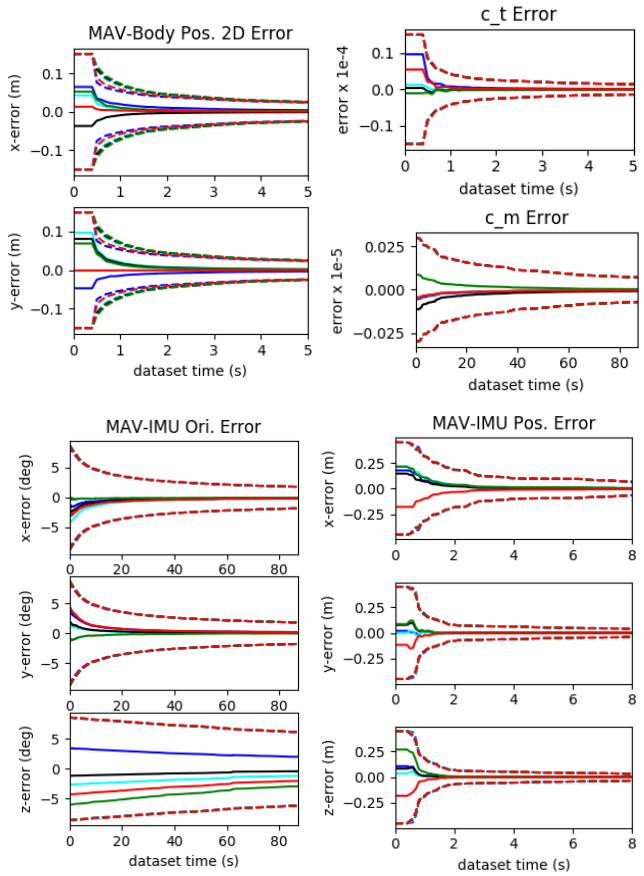

Fig. 3: Parameter estimation errors (solid) and 3σ bounds (dotted) for five runs (colors) with different realizations of the noise and initial perturbations.

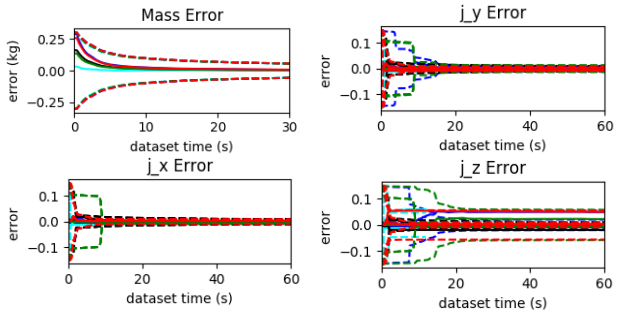

Fig. 4: Estimation errors (solid) and 3σ bounds (dotted) for m and ${}^{M}\mathbf{j}$ over 5 runs (colors) with different initial guesses. Note that m and ${}^{M}\mathbf{j}$ are calibrated separately.

B. Online Parameter Identification

We first investigate the MAV system parameter identification performance of the proposed system. In order to show the robustness, the proposed SKF-based estimator is run on the simulation trajectory five times, with the MAV *pose model* and the system noise $\sigma = 0.20$. For each run, we perturb the MAV system parameters using a Gaussian distribution with standard deviation shown in Table II.

As demonstrated by [18], [21], [22], [26], MAV parameters have the following observability proprieties:

- The z component of ${}^{B}\mathbf{p}_{M}$ is unobservable
- c_t, c_m, m and ${}^{M}\mathbf{j}$ are jointly unobservable

We consequently perform three sets of simulations to verify the capability of the proposed estimator: (i) estimate c_t, c_m, $\mathbf{x}_{MI}$ and the xy-components of ${}^{M}\mathbf{p}_{B}$, (ii) the mass m, and (iii) moment of inertia ${}^{M}\mathbf{j}$. The first sets of results are illustrated in Fig. 3 and show that parameters converge quickly to the ground truth even with significant and different initial perturbations. The m and moment of inertia ${}^{M}\mathbf{j}$ estimation results are presented in Fig. 4. We note that both the z-components of MAV-IMU orientation ${}^{I}_{M}\mathbf{R}$ and moment of inertia ${}^{M}\mathbf{j}$ converge much slower than other parameters, which reconfirms the results from [21], [22]. We hypothesize this is due to the small (or constant) yaw change during the MAV integration interval, which induces the z-component of ${}^{I}_{M}\mathbf{R}$ and ${}^{M}\mathbf{j}$ to be less observable.

We further evaluate the parameter identification accuracy of the SKF-based estimator over different noise σs for rotor force and moment. As shown in Table III, it is as expected that the inaccurate modeling (i.e., $\sigma = 0.05$) of the MAV dynamics will cause the standard EKF based system to become inconsistent, which additionally degrades the identification accuracy. In contrast, the SKF and DSKF-based estimators still offer robust visual-inertial pose estimation since there is no erroneous information from incorrect MAV measurements corrupting $\mathbf{x}_A$. The SKF-based estimators also robustly achieve the accurate estimation performances for $\mathbf{x}_\theta$ even with different sigmas. Hence, the parameter tuning of σ can be greatly relaxed for the SKF based estimators, making the proposed system more robust for different applications.

Comparing the SKF and DSKF, the DSKF drops correlations between $\mathbf{x}_A$ and $\mathbf{x}_\theta$, which theoretically is a loss of information. Interestingly, this has little impact on the converged parameter accuracy from our simulation results. Our conjecture is that this might be due to the inaccurate prior and/or small correlations, which will be further investigated.

VII. Real-World Experiments

Blackbird UAV dataset [23] provides *real-world* inertial readings, rotor encoder measurements, and ground truth poses for a MAV with photo-realistic images rendered in a series of different environments and is suitable for the proposed estimator evaluation which fuses visual-inertial and MAV dynamic information. As general 3D motion is more appropriate for parameter identification, we picked the `Ampersand` sequence (shown in Fig. 5) in the large apartment environment with fixed yaw and different linear velocities to evaluate the proposed estimator. In the experiment, we utilize the *pose model* with hand-tuned sigmas $\mathbf{n}_f$ and $\mathbf{n}_m$ of 1.0 and 0.15 as they produce reasonable results for the EKF. OpenVINS [28] and VINS-Mono [13] were also evaluated as state-of-the-art visual-inertial baselines. We were unable to get VIMO [11] to run with sufficient accuracy with the

TABLE III: Average RMSE and NEES for IMU pose and MAV parameters (mean error $\pm$ standard deviation) over 50 Monte-Carlo runs using pose update. Mass and moment of inertia were not calibrated as they are jointly unobservable.

σ		RMSE (deg / m)	NEES (ori / pos)	$\tilde{c}_t$ (N s^2/rad^2)	$\tilde{c}_m$ (N s^2/rad^2)	${}^M\tilde{\mathbf{p}}_B$ (m)	${}^I_M\delta\boldsymbol{\theta}$ (deg)	${}^I\tilde{\mathbf{p}}_M$ (m)
0.05	EKF	2.196 / 307.436	1560.65 / 935.05	3.991e-08 $\pm$ 6.713e-08	1.319e-08 $\pm$ 1.561e-09	3.226e-05 $\pm$ 3.838e-05	0.1182 $\pm$ 0.0962	1.339e-03 $\pm$ 1.111e-04
	DSKF	0.749 / 0.184	2.209 / 2.690	2.544e-08 $\pm$ 5.888e-09	1.270e-08 $\pm$ 1.474e-09	3.168e-05 $\pm$ 3.431e-05	0.0996 $\pm$ 0.0270	1.373e-03 $\pm$ 8.718e-05
	SKF	0.749 / 0.184	2.209 / 2.690	2.337e-08 $\pm$ 6.121e-09	1.278e-08 $\pm$ 1.483e-09	3.176e-05 $\pm$ 3.604e-05	0.0996 $\pm$ 0.0324	1.402e-03 $\pm$ 8.643e-05
0.50	EKF	0.852 / 0.187	3.081 / 2.624	2.738e-08 $\pm$ 2.824e-09	7.938e-08 $\pm$ 6.740e-08	5.163e-05 $\pm$ 5.984e-05	0.4426 $\pm$ 0.3153	1.306e-03 $\pm$ 9.793e-05
	DSKF	0.749 / 0.184	2.209 / 2.690	2.659e-08 $\pm$ 6.251e-09	7.941e-08 $\pm$ 6.741e-08	5.226e-05 $\pm$ 5.760e-05	0.4428 $\pm$ 0.3179	1.322e-03 $\pm$ 1.009e-04
	SKF	0.749 / 0.184	2.209 / 2.690	2.752e-08 $\pm$ 2.781e-09	7.939e-08 $\pm$ 6.744e-08	5.152e-05 $\pm$ 5.987e-05	0.4428 $\pm$ 0.3156	1.309e-03 $\pm$ 9.810e-05
1.00	EKF	0.742 / 0.184	2.817 / 2.285	3.189e-08 $\pm$ 5.236e-09	2.446e-07 $\pm$ 2.126e-07	1.553e-04 $\pm$ 1.317e-04	0.9481 $\pm$ 0.7116	1.340e-03 $\pm$ 1.948e-04
	DSKF	0.749 / 0.184	2.209 / 2.690	3.112e-08 $\pm$ 7.903e-09	2.446e-07 $\pm$ 2.126e-07	1.557e-04 $\pm$ 1.300e-04	0.9482 $\pm$ 0.7129	1.355e-03 $\pm$ 1.916e-04
	SKF	0.749 / 0.184	2.209 / 2.690	3.198e-08 $\pm$ 5.266e-09	2.446e-07 $\pm$ 2.126e-07	1.552e-04 $\pm$ 1.315e-04	0.9482 $\pm$ 0.7117	1.342e-03 $\pm$ 1.952e-04
1.50	EKF	0.733 / 0.183	2.756 / 2.236	3.718e-08 $\pm$ 8.938e-09	4.019e-07 $\pm$ 3.498e-07	3.192e-04 $\pm$ 2.612e-04	1.3133 $\pm$ 1.0025	1.395e-03 $\pm$ 3.709e-04
	DSKF	0.749 / 0.184	2.209 / 2.690	3.642e-08 $\pm$ 1.082e-08	4.019e-07 $\pm$ 3.498e-07	3.191e-04 $\pm$ 2.592e-04	1.3133 $\pm$ 1.0026	1.409e-03 $\pm$ 3.672e-04
	SKF	0.749 / 0.184	2.209 / 2.690	3.724e-08 $\pm$ 8.961e-09	4.019e-07 $\pm$ 3.498e-07	3.190e-04 $\pm$ 2.610e-04	1.3133 $\pm$ 1.0038	1.396e-03 $\pm$ 3.713e-04

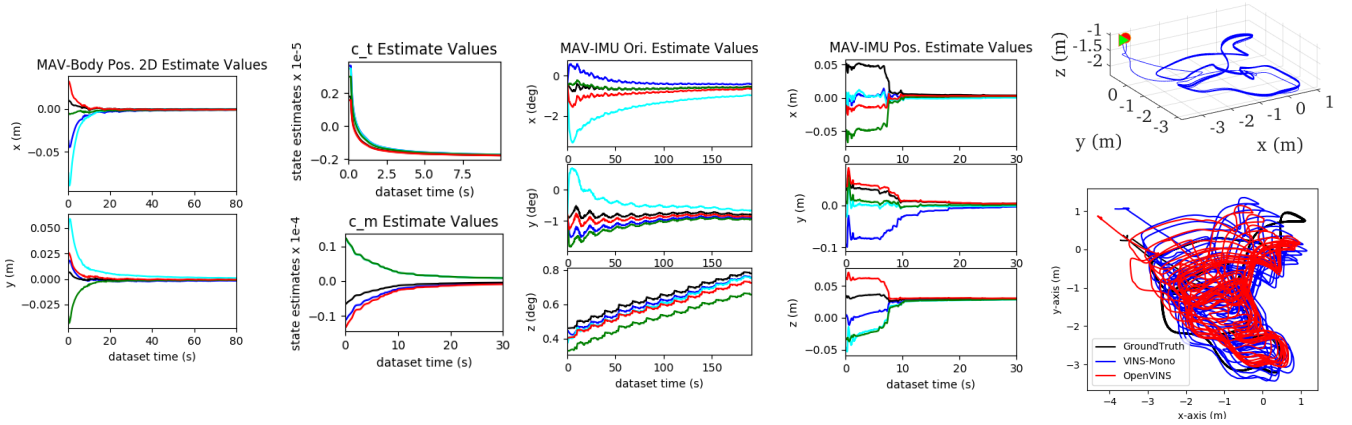

Fig. 5: Five runs of the proposed SKF estimator with pose update on the 286m Blackbird 3m/s yaw constant `Ampersand` trajectory with different initial perturbations to the MAV parameters (left). The groundtruth trajectory (top right) and estimated trajectory of VINS-Mono and OpenVINS are shown (bottom right).

TABLE IV: RMSE values for each estimator in units of degree/meters on the Blackbird `Ampersand` sequence.

Algorithm	RMSE (1 m/s)	RMSE (2 m/s)	RMSE (3 m/s)
Proposed EKF	1.463 / 0.067	1.696 / 0.119	4.195 / 0.703
Proposed SKF	1.571 / 0.069	1.703 / 0.120	3.881 / 0.720
Proposed DSKF	1.571 / 0.069	1.703 / 0.120	3.881 / 0.720
OpenVINS [28]	1.571 / 0.069	1.703 / 0.120	3.881 / 0.720
VINS-Mono [13]	1.281 / 0.075	2.851 / 0.515	4.598 / 0.965

authors' configuration file. The trajectory accuracy for all methods are reported in Table IV. The proposed system with SKF and OpenVINS have the exact same performance (as expected) and outperform VINS-Mono.

We conjecture that the EKF-based method slightly improves in accuracy for the lower velocity trajectories due to the smaller force along the local x and y directions. However, at higher speeds (3m/s), it hurts orientation accuracy. This supports our motivation to design a SKF-based estimator to protect VIO performance as the EKF-based method relies on parameter tuning of the MAV noise model. The proposed SKF-based system generates identical trajectory accuracy to the OpenVINS and demonstrates its robustness and accuracy to different motion profiles.

We next evaluate parameter identification capability by running the estimator on the same trajectory multiple times with different initial values. As shown in Fig. 5, c_t, c_m, xy-components of ${}^M\mathbf{p}_B$, and $\mathbf{x}_{MI}$ converge to the same values at the end. In lieu of groundtruth parameters, this shows both the robustness and repeatably of the proposed method. Similar to the simulation (see Fig. 3), the z-component of the MAV-IMU orientation either does not show the trend of convergence or converges slower than other parameters. This is still likely due to the small yaw change during the MAV integration interval. Moreover, the `Ampersand` trajectory is designed to be yaw constant, which could make the z direction of MAV-IMU orientation even vulnerable.

VIII. Conclusion and Future Work

Accurate system modeling and parameter identification are essential for MAV navigation and control. In this paper, we have developed a tightly-coupled real-time SKF estimator for MAV system parameter identification which avoids the requirements of an additional sensor, the synchronization and joint propagation of rotor speeds with the IMU, or the estimation of an assumed external force. Special attention was paid to the inaccuracies of the MAV dynamic model. Through numerical studies, we have shown that the naive fusion of MAV dynamics in an EKF can degrade performance due to overconfident noise parameters. The SKF-based estimator was leveraged to relax extensive parameter tuning procedures, prevent VIO performance degradation, and ensure accurate and robust parameter identification. Extensive Monte-Carlo simulations and real-world evaluations were performed to demonstrate that the proposed system achieves

accurate and robust online MAV parameter identification. In the future, we plan to extend the work with an degenerate motion analysis for system parameters along with calibration parameters (e.g., see [39]), and develop observability-aware motion planning for autonomous navigation.

ACKNOWLEDGEMENT

The authors thank Kunyi Zhang and Ziming Ding from Zhejiang University for their valuable discussion.

REFERENCES

[1] A. Gawel, M. Kamel, T. Novkovic, J. Widauer, D. Schindler, B. P. Von Altishofen, R. Siegwart, and J. Nieto, "Aerial picking and delivery of magnetic objects with mavs," in *2017 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2017, pp. 5746–5752.

[2] M. Blösch, S. Weiss, D. Scaramuzza, and R. Siegwart, "Vision based MAV navigation in unknown and unstructured environments," in *2010 International Conference on Robotics and Automation (ICRA)*, Anchorage, AK, May 3–8, 2010, pp. 21–28.

[3] G. Loianno, C. Brunner, G. McGrath, and V. Kumar, "Estimation, control, and planning for aggressive flight with a small quadrotor with a single camera and imu," *IEEE Robotics and Automation Letters*, vol. 2, no. 2, pp. 404–411, 2016.

[4] M. Saska, T. Baca, J. Thomas, J. Chudoba, L. Preucil, T. Krajnik, J. Faigl, G. Loianno, and V. Kumar, "System for deployment of groups of unmanned micro aerial vehicles in gps-denied environments using onboard visual relative localization," *Autonomous Robots*, vol. 41, no. 4, pp. 919–944, 2017.

[5] T. Özaslan, G. Loianno, J. Keller, C. J. Taylor, V. Kumar, J. M. Wozencraft, and T. Hood, "Autonomous navigation and mapping for inspection of penstocks and tunnels with mavs," *IEEE Robotics and Automation Letters*, vol. 2, no. 3, pp. 1740–1747, 2017.

[6] N. Michael, S. Shen, K. Mohta, V. Kumar, K. Nagatani, Y. Okada, S. Kiribayashi, K. Otake, K. Yoshida, K. Ohno, *et al.*, "Collaborative mapping of an earthquake damaged building via ground and aerial robots," in *Field and service robotics*. Springer, 2014, pp. 33–47.

[7] G. Huang, "Visual-inertial navigation: A concise review," in *Proc. International Conference on Robotics and Automation*, Montreal, Canada, May 2019.

[8] K. Sun, K. Mohta, B. Pfrommer, M. Watterson, S. Liu, Y. Mulgaonkar, C. J. Taylor, and V. Kumar, "Robust stereo visual inertial odometry for fast autonomous flight," *IEEE Robotics and Automation Letters*, vol. 3, no. 2, pp. 965–972, April 2018.

[9] J. Delmerico and D. Scaramuzza, "A benchmark comparison of monocular visual-inertial odometry algorithms for flying robots," in *2018 International Conference on Robotics and Automation (ICRA)*, Brisbane, Australia, May 21–25, 2018.

[10] J. Delaune, D. S. Bayard, and R. Brockers, "Range-visual-inertial odometry: Scale observability without excitation," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 2421–2428, 2021.

[11] B. Nisar, P. Foehn, D. Falanga, and D. Scaramuzza, "VIMO: Simultaneous visual inertial model-based odometry and force estimation," *IEEE Robotics and Automation Letters*, vol. 4, no. 3, pp. 2785–2792, 2019.

[12] Z. Ding, T. Yang, K. Zhang, C. Xu, and F. Gao, "Vid-fusion: Robust visual-inertial-dynamics odometry for accurate external force estimation," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, 2021, pp. 14 469–14 475.

[13] T. Qin, P. Li, and S. Shen, "VINS-Mono: A robust and versatile monocular visual-inertial state estimator," *IEEE Transactions on Robotics*, vol. 34, no. 4, pp. 1004–1020, 2018.

[14] A. Ollero, M. Tognon, A. Suarez, D. Lee, and A. Franchi, "Past, present, and future of aerial robotic manipulators," *IEEE Transactions on Robotics*, 2021.

[15] S. Liu, J. Erskine, A. Chriette, and I. Fantoni, "Decentralized control and teleoperation of a multi-uav parallel robot based on intrinsic measurements," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2021, pp. 6329–6335.

[16] L. Derafa, T. Madani, and A. Benallegue, "Dynamic modelling and experimental identification of four rotors helicopter parameters," in *2006 IEEE international conference on industrial technology*. IEEE, 2006, pp. 1834–1839.

[17] N. V. Hoffer, C. Coopmans, A. M. Jensen, and Y. Chen, "A survey and categorization of small low-cost unmanned aerial vehicle system identification," *Journal of Intelligent & Robotic Systems*, vol. 74, no. 1, pp. 129–145, 2014.

[18] M. Burri, M. Bloesch, Z. Taylor, R. Siegwart, and J. Nieto, "A framework for maximum likelihood parameter identification applied on mavs," *Journal of Field Robotics*, vol. 35, no. 1, pp. 5–22, 2018.

[19] D. Six, S. Briot, J. Erskine, and A. Chriette, "Identification of the propeller coefficients and dynamic parameters of a hovering quadrotor from flight data," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 1063–1070, 2020.

[20] D. Mellinger and V. Kumar, "Minimum snap trajectory generation and control for quadrotors," in *2011 IEEE International Conference on Robotics and Automation (ICRA)*, Shanghai, China, May 9–13, 2011, pp. 2520–2525.

[21] V. Wüest, V. Kumar, and G. Loianno, "Online estimation of geometric and inertia parameters for multirotor aerial vehicles," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 1884–1890.

[22] C. Böhm, M. Scheiber, and S. Weiss, "Filter-Based Online System-Parameter Estimation for Multicopter UAVs," in *Proceedings of Robotics: Science and Systems*, Virtual, July 2021.

[23] A. Antonini, W. Guerra, V. Murali, T. Sayre-McCord, and S. Karaman, "The blackbird uav dataset," *The International Journal of Robotics Research*, vol. 39, no. 10-11, pp. 1346–1364, 2020.

[24] N. Trawny and S. I. Roumeliotis, "Indirect Kalman filter for 3D attitude estimation," University of Minnesota, Dept. of Comp. Sci. & Eng., Tech. Rep., Mar. 2005.

[25] T. D. Barfoot, *State estimation for robotics*. Cambridge University Press, 2017.

[26] C. Chen, Y. Yang, P. Geneva, W. Lee, and G. Huang, "Supplementary materials: Visual-inertial-aided online mav system identification," University of Delaware, Tech. Rep. RPNG-2022-MAV, 2022. [Online]. Available: http://udel.edu/~ghuang/papers/tr_mav.pdf

[27] A. I. Mourikis and S. I. Roumeliotis, "A multi-state constraint Kalman filter for vision-aided inertial navigation," in *Proceedings of the IEEE International Conference on Robotics and Automation*, Rome, Italy, Apr. 10–14, 2007, pp. 3565–3572.

[28] P. Geneva, K. Eckenhoff, W. Lee, Y. Yang, and G. Huang, "OpenVINS: a research platform for visual-inertial estimation," in *Proc. of the IEEE International Conference on Robotics and Automation*, Paris, France, 2020. [Online]. Available: https://github.com/rpng/open_vins

[29] M. Li, B. H. Kim, and A. I. Mourikis, "Real-time motion tracking on a cellphone using inertial sensing and a rolling-shutter camera," in *2013 IEEE International Conference on Robotics and Automation*. IEEE, 2013, pp. 4712–4719.

[30] A. B. Chatfield, *Fundamentals of High Accuracy Inertial Navigation*. AIAA, 1997.

[31] Y. Yang, J. Maley, and G. Huang, "Null-space-based marginalization: Analysis and algorithm," in *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, Vancouver, Canada, Sept. 2017, pp. 6749–6755.

[32] G. Huang, A. I. Mourikis, and S. I. Roumeliotis, "Observability-based rules for designing consistent EKF SLAM estimators," *International Journal of Robotics Research*, vol. 29, no. 5, pp. 502–528, Apr. 2010.

[33] M. Li and A. Mourikis, "High-precision, consistent EKF-based visual-inertial odometry," *International Journal of Robotics Research*, vol. 32, no. 6, pp. 690–711, 2013.

[34] C. Chen, Y. Yang, P. Geneva, and G. Huang, "FEJ2: a consistent visual-inertial state estimator design," in *Proc. International Conference on Robotics and Automation*, Philadelphia, USA, May 2022.

[35] F. Furrer, M. Burri, M. Achtelik, and R. Siegwart, "Rotors—a modular gazebo mav simulator framework," in *Robot operating system (ROS)*. Springer, 2016, pp. 595–625.

[36] Z. Zhang and D. Scaramuzza, "A tutorial on quantitative trajectory evaluation for visual (-inertial) odometry," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2018, pp. 7244–7251.

[37] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with applications to tracking and navigation: theory algorithms and software*. John Wiley & Sons, 2004.

[38] K. Eckenhoff, P. Geneva, N. Merrill, and G. Huang, "Schmidt-ekf-based visual-inertial moving object tracking," in *Proc. of the IEEE International Conference on Robotics and Automation*, Paris, France, 2020.

[39] Y. Yang, P. Geneva, X. Zuo, and G. Huang, "Online self-calibration for visual-inertial navigation systems: Models, analysis and degeneracy," *arXiv preprint arXiv:2201.09170*, 2022.