Supplementary Materials:
DIDO: Deep Inertial Quadrotor Dynamical Odometry

Kunyi Zhang[1,2] - kunyizhang@zju.edu.cn
Chenxing Jiang[1,2] - chenxingjiang@zju.edu.cn
Jinghang Li[2] - jhanglee.ro@gmail.com
Sheng Yang[3] - richeng.ys@alibaba-inc.com
Teng Ma[3] - damon.mt@alibaba-inc.com
Chao Xu[1,2] - chaoxu@zju.edu.cn
Fei Gao[1,2] - fgaoaa@zju.edu.cn

CONTENTS

[1]State Key Laboratory of Industrial Control Technology, Institute of Cyber-Systems and Control, Zhejiang University, Hangzhou 310027, China.
[2]Huzhou Institute, Zhejiang University, Huzhou 313000, China.
[3]Alibaba DAMO Academy Autonomous Driving Lab, Hangzhou 311121, China.

I. IMU Kinematics and Quadrotor Dynamics

A. IMU Kinematics

IMU measurements include the gyroscope $\widetilde{\boldsymbol{\omega}}$ and non-gravitational acceleration $\widetilde{\boldsymbol{a}}$, which are measured in the IMU frame (the $\mathcal{I}$ frame) and given by:

$$\begin{aligned} {}^{\mathcal{I}}\widetilde{\boldsymbol{\omega}} &= {}^{\mathcal{I}}\boldsymbol{\omega} + {}^{\mathcal{I}}\boldsymbol{b}_{\boldsymbol{\omega}} + \boldsymbol{n}_{\boldsymbol{\omega}}, \\ {}^{\mathcal{I}}\widetilde{\boldsymbol{a}} &= {}^{\mathcal{I}}\boldsymbol{a} + {}^{\mathcal{I}}\boldsymbol{b}_{\boldsymbol{a}} + {}^{\mathcal{G}}_{\mathcal{I}}\mathbf{R}^{\top\mathcal{G}}\mathbf{g} + \boldsymbol{n}_{\boldsymbol{a}}, \end{aligned} \tag{1}$$

where ${}^{\mathcal{I}}\boldsymbol{\omega}$ and ${}^{\mathcal{I}}\boldsymbol{a}$ are the true angular velocity and acceleration, ${}^{\mathcal{G}}\mathbf{g} = [0, 0, 9.8]$ is the gravity vector in the gravity-aligned frame (the $\mathcal{G}$ frame), ${}^{\mathcal{G}}_{\mathcal{I}}\mathbf{R}$ is the rotation matrix from the $\mathcal{I}$ frame to the $\mathcal{G}$ frame, $\boldsymbol{n}_{\boldsymbol{\omega}}$ and $\boldsymbol{n}_{\boldsymbol{a}}$ are the additive Gaussian white noise in gyroscope and acceleration measurements, $\boldsymbol{b}_{\boldsymbol{\omega}}$ and $\boldsymbol{b}_{\boldsymbol{a}}$ are the bias of IMU modeled as random walk:

$$\begin{aligned} \boldsymbol{n}_{\boldsymbol{\omega}} &\sim \mathcal{N}(0, \boldsymbol{\Sigma}^2_{\boldsymbol{\omega}}), \quad \dot{\boldsymbol{b}}_{\boldsymbol{\omega}} \sim \mathcal{N}(0, \boldsymbol{\Sigma}^2_{\boldsymbol{b}_{\boldsymbol{\omega}}}), \\ \boldsymbol{n}_{\boldsymbol{a}} &\sim \mathcal{N}(0, \boldsymbol{\Sigma}^2_{\boldsymbol{a}}), \quad \dot{\boldsymbol{b}}_{\boldsymbol{a}} \sim \mathcal{N}(0, \boldsymbol{\Sigma}^2_{\boldsymbol{b}_{\boldsymbol{a}}}). \end{aligned} \tag{2}$$

B. Quadrotor Dynamics

Because the propagation of kinematic states is driven by multiple propulsion units in a quadrotor system, we model the Newtonian dynamics according to [1]. The total driving force of a quadrotor in the body frame (the $\mathcal{B}$ frame) is the sum of the thrust ${}^{\mathcal{B}}\boldsymbol{F}_t$ and drag force ${}^{\mathcal{B}}\boldsymbol{F}_d$ generated by each propulsion unit as follow:

$${}^{\mathcal{B}}\boldsymbol{F} = \sum_{i=1}^{4}\left({}^{\mathcal{B}}\boldsymbol{F}_{t_i} - {}^{\mathcal{B}}\boldsymbol{F}_{d_i}\right) = \sum_{i=1}^{4}\left(\tau u_i^2 \boldsymbol{e}_3 - u_i D^{\mathcal{B}}\boldsymbol{v}_i\right), \tag{3}$$

where τ is the thrust coefficient for the propellers, $D = diag(d_x, d_y, d_z)$ is the matrix of effective linear drag coefficients, $\boldsymbol{e}_3 = [0, 0, 1]^\top$ is the z axis in any frame, and u_i and ${}^{\mathcal{B}}\boldsymbol{v}_i$ are the rotation speed and velocity of the i-th rotor, respectively. Actually, the velocity of each rotor is:

$${}^{\mathcal{B}}\boldsymbol{v}_i = {}^{\mathcal{B}}\boldsymbol{v} + {}^{\mathcal{B}}\boldsymbol{\omega} \times {}^{\mathcal{B}}\boldsymbol{r}_i^{\mathcal{B}}, \tag{4}$$

where ${}^{\mathcal{B}}\boldsymbol{v}$ and ${}^{\mathcal{B}}\boldsymbol{\omega}$ are the linear and angular velocity of the quadrotor's center of mass (CoM), ${}^{\mathcal{B}}\boldsymbol{r}_i^{\mathcal{B}}$ is the position of the i-th rotor relative to the CoM. To simplify the calculation, we ignore the velocity discrepancy of different rotors and express it as:

$${}^{\mathcal{B}}\boldsymbol{v}_i \approx {}^{\mathcal{B}}\boldsymbol{v}. \tag{5}$$

The input notations are abbreviated as:

$$U_{ss} = \sum_{i=1}^{4} u_i^2, \quad U_s = \sum_{i=1}^{4} u_i, \tag{6}$$

so we can obtain the Newtonian equation in the $\mathcal{G}$ frame:

$$m\frac{d}{dt}\left({}^{\mathcal{G}}\boldsymbol{v}\right) = {}^{\mathcal{G}}_{\mathcal{B}}\mathbf{R}\left(\tau U_{ss}\boldsymbol{e}_3 - U_s D {}^{\mathcal{G}}_{\mathcal{B}}\mathbf{R}^{\top\mathcal{G}}\boldsymbol{v}^{\mathcal{G}}_{\mathcal{B}}\right) - m^{\mathcal{G}}\mathbf{g}. \tag{7}$$

II. Two-Stage EKF for Inertial Dynamical Fusion

A. Rotation Stage

1) State: In the rotation stage, the rotation of the $\mathcal{I}$ frame in the $\mathcal{G}$ frame is taken as state:

$$\boldsymbol{x} = {}^{\mathcal{G}}_{\mathcal{I}}\boldsymbol{q}. \tag{8}$$

2) Process Model: The rotational equation is given:

$$\dot{\boldsymbol{x}} = \frac{1}{2} {}^{\mathcal{G}}_{\mathcal{I}}\boldsymbol{q} \otimes {}^{\mathcal{I}}\boldsymbol{\omega} = \frac{1}{2} {}^{\mathcal{G}}_{\mathcal{I}}\boldsymbol{q} \otimes ({}^{\mathcal{I}}\widehat{\boldsymbol{\omega}} + \boldsymbol{n}_{\boldsymbol{\omega}}), \tag{9}$$

where ${}^{\mathcal{I}}\widehat{\boldsymbol{\omega}} = {}^{\mathcal{I}}\widetilde{\boldsymbol{\omega}} - {}^{\mathcal{I}}\widehat{\boldsymbol{b}}_{\boldsymbol{\omega}}$, $\boldsymbol{n}_{\boldsymbol{\omega}} \sim \mathcal{N}(0, \boldsymbol{\Sigma}^2_{\boldsymbol{\omega}})$, and ${}^{\mathcal{I}}\widehat{\boldsymbol{b}}_{\boldsymbol{\omega}}$ is the output of the gyroscope *De-Bias Net*.

The above differential equation is discretized as follows:

$$\begin{aligned} \boldsymbol{x}_{k+1} &= \mathbf{F}(\boldsymbol{x}_k, \boldsymbol{u}_k, \boldsymbol{n}_k) \\ &= \mathbf{F}_{\boldsymbol{x}}\boldsymbol{x}_k + \mathbf{F}_{\boldsymbol{n}}\boldsymbol{n}_k, \end{aligned} \tag{10}$$

where,

$$\begin{aligned} \mathbf{F}_{\boldsymbol{x}} &= \frac{\partial \mathbf{F}}{\partial \boldsymbol{x}_k} = \mathbf{I}_4 + \frac{\Delta t}{2} \begin{bmatrix} 0 & -{}^{\mathcal{I}}\widehat{\boldsymbol{\omega}}^\top \\ {}^{\mathcal{I}}\widehat{\boldsymbol{\omega}} & \lfloor {}^{\mathcal{I}}\widehat{\boldsymbol{\omega}} \rfloor_\times \end{bmatrix}, \\ \mathbf{F}_{\boldsymbol{n}} &= \frac{\partial \mathbf{F}}{\partial \boldsymbol{n}_{\boldsymbol{\omega}}} = \frac{\Delta t}{2} \begin{bmatrix} -q_x & -q_y & -q_z \\ q_w & -q_z & q_y \\ q_z & q_w & -q_x \\ -q_y & q_x & q_w \end{bmatrix} = \frac{\Delta t}{2} \begin{bmatrix} -\boldsymbol{q}_v \\ q_w\mathbf{I}_3 + \lfloor \boldsymbol{q}_v \rfloor_\times \end{bmatrix}. \end{aligned} \tag{11}$$

Specifically, $\boldsymbol{q} = [q_w, q_x, q_y, q_z]^\top = [q_w, \boldsymbol{q}_v^\top]^\top$.

3) Measurement Model: The attitude controllers of most flying robots rely on the complementary filters [2, 3] to obtain attitude observations. Similarly, we consider the gravity alignment constraint to obtain the tilt observation:

$${}^{\mathcal{I}}\widehat{\boldsymbol{a}} \approx {}^{\mathcal{G}}_{\mathcal{I}}\mathbf{R}^{\top\mathcal{G}}\mathbf{g} + \boldsymbol{n}_{\boldsymbol{a}}. \tag{12}$$

Similarly, ${}^{\mathcal{I}}\widehat{\boldsymbol{a}} = {}^{\mathcal{I}}\widetilde{\boldsymbol{a}} - {}^{\mathcal{I}}\widehat{\boldsymbol{b}}_{\boldsymbol{a}}$, $\boldsymbol{n}_{\boldsymbol{a}} \sim \mathcal{N}(0, \boldsymbol{\Sigma}^2_{\boldsymbol{a}})$, and ${}^{\mathcal{I}}\widehat{\boldsymbol{b}}_{\boldsymbol{a}}$ is the output of the accelerator *De-Bias Net*.

The above observation equation is approximated as:

$$\boldsymbol{z}_k = \mathbf{H}(\boldsymbol{x}_k) = \mathbf{H}_{\boldsymbol{x}}\boldsymbol{x}_k, \tag{13}$$

where,

$$\mathbf{H}_{\boldsymbol{x}} = 2\begin{bmatrix} \boldsymbol{e}_3 \times \boldsymbol{q}_v & \lfloor \boldsymbol{e}_3 \times \boldsymbol{q}_v + q_w\boldsymbol{e}_3 \rfloor_\times + (\boldsymbol{q}_v \cdot \boldsymbol{e}_3)\mathbf{I}_3 - \boldsymbol{e}_3\boldsymbol{q}_v^\top \end{bmatrix}. \tag{14}$$

4) Extended Kalman Filter:

$$\begin{aligned} \boldsymbol{x}_{k+1|k} &= \mathbf{F}(\boldsymbol{x}_{k|k}, \boldsymbol{u}_k), \\ \mathbf{P}_{k+1|k} &= \mathbf{F}_{\boldsymbol{x}}\mathbf{P}_{k|k}\mathbf{F}_{\boldsymbol{x}}^T + \mathbf{F}_{\boldsymbol{u}}\mathbf{Q}_k\mathbf{F}_{\boldsymbol{u}}^T, \\ \mathbf{K}_{k+1} &= \mathbf{P}_{k+1|k}\mathbf{H}_{\boldsymbol{x}}^T\left(\mathbf{H}_{\boldsymbol{x}}\mathbf{P}_{k+1|k}\mathbf{H}_{\boldsymbol{x}}^T + \mathbf{R}_{k+1}\right)^{-1}, \\ \boldsymbol{x}_{k+1|k+1} &= \boldsymbol{x}_{k+1|k} \oplus \left(\mathbf{K}_{k+1}(\boldsymbol{z}_{k+1} - \mathbf{H}(\boldsymbol{x}_{k+1|k}))\right), \\ \mathbf{P}_{k+1|k+1} &= \left(\mathbf{I} - \mathbf{K}_{k+1}\mathbf{H}_{\boldsymbol{x}}\right)\mathbf{P}_{k+1|k}, \end{aligned} \tag{15}$$

where,

$$\mathbf{Q} = \boldsymbol{\Sigma}^2_{\boldsymbol{\omega}}, \quad \mathbf{R} = \boldsymbol{\Sigma}^2_{\boldsymbol{a}}. \tag{16}$$

B. Translation Stage

1) State: The state of the second stage is defined as:

$$\boldsymbol{x} = ({}^{\mathcal{G}}\boldsymbol{p}^{\mathcal{G}}_{\mathcal{B}},\ {}^{\mathcal{G}}\boldsymbol{v}^{\mathcal{G}}_{\mathcal{B}},\ \tau,\ \boldsymbol{d},\ {}^{\mathcal{I}}_{\mathcal{B}}\boldsymbol{q},\ {}^{\mathcal{I}}\boldsymbol{t}^{\mathcal{I}}_{\mathcal{B}}), \tag{17}$$

where ${}^{\mathcal{G}}\boldsymbol{p}^{\mathcal{G}}_{\mathcal{B}}$ and ${}^{\mathcal{G}}\boldsymbol{v}^{\mathcal{G}}_{\mathcal{B}}$ are respectively the velocity and position of the quadrotor body $\mathcal{B}$ frame expressed in the $\mathcal{G}$ frame, τ is the thrust coefficient, $\boldsymbol{d}$ is the drag vector of (d_x, d_y, d_z), and $({}^{\mathcal{I}}_{\mathcal{B}}\boldsymbol{q}, {}^{\mathcal{I}}\boldsymbol{t}^{\mathcal{I}}_{\mathcal{B}})$ is the extrinsic parameter between the $\mathcal{B}$ and the $\mathcal{I}$ frame.

2) Process Model: We regard the quadrotor dynamics as the input, and express the complete process model as follows:

$$
\begin{aligned}
{}^{\mathcal{G}}\dot{\boldsymbol{p}}_{\mathcal{B}}^{\mathcal{G}} &= {}^{\mathcal{G}}\boldsymbol{v}_{\mathcal{B}}^{\mathcal{G}}, \\
{}^{\mathcal{G}}\dot{\boldsymbol{v}}_{\mathcal{B}}^{\mathcal{G}} &= \frac{1}{m}{}_{\mathcal{B}}^{\mathcal{G}}\mathbf{R}\left(\tau U_{ss}\boldsymbol{e}_3 - U_s D_{\mathcal{B}}^{\mathcal{G}}\mathbf{R}^{\top\mathcal{G}}\boldsymbol{v}_{\mathcal{B}}^{\mathcal{G}} + \widehat{\boldsymbol{f}}_{res} + \boldsymbol{n}_{\boldsymbol{f}}\right) - {}^{\mathcal{G}}\mathbf{g}, \\
\dot{\tau} &= 0, \\
\dot{\boldsymbol{d}} &= \mathbf{0}, \\
{}_{\mathcal{B}}^{\mathcal{I}}\dot{\boldsymbol{q}} &= \mathbf{0}, \\
{}^{\mathcal{I}}\dot{\boldsymbol{t}}_{\mathcal{B}}^{\mathcal{I}} &= \mathbf{0},
\end{aligned}
\tag{18}
$$

where ${}_{\mathcal{B}}^{\mathcal{G}}\mathbf{R} = {}_{\mathcal{I}}^{\mathcal{G}}\mathbf{R}{}_{\mathcal{B}}^{\mathcal{I}}\mathbf{R}$ is the rotation from the quadrotor body $\mathcal{B}$ frame to the $\mathcal{G}$ frame, and $\widehat{\boldsymbol{f}}_{res}$ and $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{f}}^2$ ($\boldsymbol{n}_{\boldsymbol{f}} \sim \mathcal{N}(0, \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{f}}^2)$) are the *Res-Dynamics Net* outputs.

The above differential equation is discretized as follows:

$$
\begin{aligned}
\boldsymbol{x}_{k+1} &= \mathbf{F}(\boldsymbol{x}_k, \boldsymbol{u}_k, \boldsymbol{n}_k) \\
&= \mathbf{F}_{\boldsymbol{x}}\boldsymbol{x}_k + \mathbf{F}_{\boldsymbol{n}}\boldsymbol{n}_k,
\end{aligned}
\tag{19}
$$

where,

$$
\begin{aligned}
\mathbf{F}_{\boldsymbol{x}} &= \begin{bmatrix}
\mathbf{I}_3 & \frac{\partial \boldsymbol{p}_{k+1}}{\partial \boldsymbol{v}_k} & \mathbf{0}_{31} & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\
\mathbf{0}_3 & \frac{\partial \boldsymbol{v}_{k+1}}{\partial \boldsymbol{v}_k} & \frac{\partial \boldsymbol{v}_{k+1}}{\partial \tau} & \frac{\partial \boldsymbol{v}_{k+1}}{\partial \boldsymbol{d}} & \frac{\partial \boldsymbol{v}_{k+1}}{\partial {}_{\mathcal{B}}^{\mathcal{I}}\boldsymbol{\theta}} & \mathbf{0}_3 \\
\mathbf{0}_{13} & \mathbf{0}_{13} & 1 & \mathbf{0}_{13} & \mathbf{0}_{13} & \mathbf{0}_{13} \\
\mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_{31} & \mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\
\mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_{31} & \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_3 \\
\mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_{31} & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{I}_3
\end{bmatrix}, \\
\mathbf{F}_{\boldsymbol{n}} &= \begin{bmatrix} \mathbf{0}_3 & \frac{\partial \boldsymbol{v}_{k+1}}{\partial \boldsymbol{n}_{\boldsymbol{f}}} & \mathbf{0}_{31} & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \end{bmatrix}^{\top}.
\end{aligned}
\tag{20}
$$

Specifically,

$$
\begin{aligned}
\frac{\partial \boldsymbol{p}_{k+1}}{\partial \boldsymbol{v}_k} =& \mathbf{I}_3 \Delta t, \\
\frac{\partial \boldsymbol{v}_{k+1}}{\partial \boldsymbol{v}_k} =& \mathbf{I} - \frac{U_s \Delta t}{m}{}_{\mathcal{B}}^{\mathcal{G}}\mathbf{R} D_{\mathcal{B}}^{\mathcal{G}}\mathbf{R}^{\top}, \\
\frac{\partial \boldsymbol{v}_{k+1}}{\partial \tau} =& \frac{U_{ss}\Delta t}{m}{}_{\mathcal{B}}^{\mathcal{G}}\boldsymbol{r}_3, \\
\frac{\partial \boldsymbol{v}_{k+1}}{\partial \boldsymbol{d}} =& -\frac{\Delta t}{m}\left[{}_{\mathcal{B}}^{\mathcal{G}}\boldsymbol{r}_1{}_{\mathcal{B}}^{\mathcal{G}}\boldsymbol{r}_1^{\top\mathcal{G}}\boldsymbol{v}_{\mathcal{B}}^{\mathcal{G}} \quad {}_{\mathcal{B}}^{\mathcal{G}}\boldsymbol{r}_2{}_{\mathcal{B}}^{\mathcal{G}}\boldsymbol{r}_2^{\top\mathcal{G}}\boldsymbol{v}_{\mathcal{B}}^{\mathcal{G}} \quad {}_{\mathcal{B}}^{\mathcal{G}}\boldsymbol{r}_3{}_{\mathcal{B}}^{\mathcal{G}}\boldsymbol{r}_3^{\top\mathcal{G}}\boldsymbol{v}_{\mathcal{B}}^{\mathcal{G}}\right], \\
\frac{\partial \boldsymbol{v}_{k+1}}{\partial {}_{\mathcal{B}}^{\mathcal{I}}\boldsymbol{\theta}} =& \frac{\Delta t}{m}{}_{\mathcal{B}}^{\mathcal{G}}\mathbf{R}\Big(-\tau U_{ss}\left\lfloor \boldsymbol{e}_3 \right\rfloor_{\times} + U_s\left(\left\lfloor D_{\mathcal{B}}^{\mathcal{G}}\mathbf{R}^{\top\mathcal{G}}\boldsymbol{v}_{\mathcal{B}}^{\mathcal{G}}\right\rfloor_{\times} - D\left\lfloor {}_{\mathcal{B}}^{\mathcal{G}}\mathbf{R}^{\top\mathcal{G}}\boldsymbol{v}_{\mathcal{B}}^{\mathcal{G}}\right\rfloor_{\times}\right) - \left\lfloor \widehat{\boldsymbol{f}}_{res}\right\rfloor_{\times}\Big), \\
\frac{\partial \boldsymbol{v}_{k+1}}{\partial \boldsymbol{n}_{\boldsymbol{f}}} =& \frac{1}{m}{}_{\mathcal{B}}^{\mathcal{G}}\mathbf{R}\Delta t,
\end{aligned}
\tag{21}
$$

and $\boldsymbol{r}_i = \mathbf{R}\boldsymbol{e}_i$.

3) Measurement Models: There are three measurements corresponding to the translation stage: the IMU acceleration, the network velocity, and the network displacement measurements.

a) dynamics constraint measurements: The dynamics constraint measurements are expressed as:

$$\begin{aligned}{}^{\mathcal{I}}\widehat{\boldsymbol{a}} &= \mathbf{H}_d(\boldsymbol{x}, \boldsymbol{n}_{\boldsymbol{a}}) \\ &= \frac{1}{m}{}^{\mathcal{I}}_{\mathcal{B}}\mathbf{R}\left(\tau U_{ss}\boldsymbol{e}_3 - U_s D^{\mathcal{G}}_{\mathcal{B}}\mathbf{R}^{\top\mathcal{G}}\boldsymbol{v}^{\mathcal{G}}_{\mathcal{B}} + \widehat{\boldsymbol{f}}_{res}\right) - {}^{\mathcal{I}}\widehat{\boldsymbol{\omega}} \times ({}^{\mathcal{I}}\widehat{\boldsymbol{\omega}} \times {}^{\mathcal{I}}\boldsymbol{t}^{\mathcal{I}}_{\mathcal{B}}) - {}^{\mathcal{I}}\widehat{\boldsymbol{\alpha}} \times {}^{\mathcal{I}}\boldsymbol{t}^{\mathcal{I}}_{\mathcal{B}} + \boldsymbol{n}_{\boldsymbol{a}},\end{aligned} \tag{22}$$

whose jacobian can be given as:

$$\frac{\partial \mathbf{H}_{\boldsymbol{d}}}{\partial \boldsymbol{x}_k} = \begin{bmatrix}\mathbf{0}_3 & \frac{\partial \mathbf{H}_{\boldsymbol{d}}}{\partial \boldsymbol{v}_k} & \frac{\partial \mathbf{H}_{\boldsymbol{d}}}{\partial \tau} & \frac{\partial \mathbf{H}_{\boldsymbol{d}}}{\partial \boldsymbol{d}} & \frac{\partial \mathbf{H}_{\boldsymbol{d}}}{\partial^{\mathcal{I}}_{\mathcal{B}}\boldsymbol{\theta}} & \frac{\partial \mathbf{H}_{\boldsymbol{d}}}{\partial^{\mathcal{I}}\boldsymbol{t}^{\mathcal{I}}_{\mathcal{B}}}\end{bmatrix}, \tag{23}$$

where,

$$\begin{aligned}\frac{\partial \mathbf{H}_{\boldsymbol{d}}}{\partial \boldsymbol{v}_k} &= -\frac{U_s}{m}{}^{\mathcal{I}}_{\mathcal{B}}\mathbf{R}D^{\mathcal{G}}_{\mathcal{B}}\mathbf{R}^{\top}, \\ \frac{\partial \mathbf{H}_{\boldsymbol{d}}}{\partial \tau} &= \frac{U_{ss}}{m}{}^{\mathcal{I}}_{\mathcal{B}}\boldsymbol{r}_3, \\ \frac{\partial \mathbf{H}_{\boldsymbol{d}}}{\partial \boldsymbol{d}} &= -\frac{U_s}{m}\begin{bmatrix}{}^{\mathcal{I}}_{\mathcal{B}}\boldsymbol{r}_1{}^{\mathcal{G}}_{\mathcal{B}}\boldsymbol{r}_1^{\top\mathcal{G}}\boldsymbol{v}^{\mathcal{G}}_{\mathcal{B}} & {}^{\mathcal{I}}_{\mathcal{B}}\boldsymbol{r}_2{}^{\mathcal{G}}_{\mathcal{B}}\boldsymbol{r}_2^{\top\mathcal{G}}\boldsymbol{v}^{\mathcal{G}}_{\mathcal{B}} & {}^{\mathcal{I}}_{\mathcal{B}}\boldsymbol{r}_3{}^{\mathcal{G}}_{\mathcal{B}}\boldsymbol{r}_3^{\top\mathcal{G}}\boldsymbol{v}^{\mathcal{G}}_{\mathcal{B}}\end{bmatrix}, \\ \frac{\partial \mathbf{H}_{\boldsymbol{d}}}{\partial^{\mathcal{I}}_{\mathcal{B}}\boldsymbol{\theta}} &= \frac{1}{m}{}^{\mathcal{I}}_{\mathcal{B}}\mathbf{R}\Big(-\tau U_{ss}\lfloor\boldsymbol{e}_3\rfloor_{\times} + U_s\left(\lfloor D^{\mathcal{G}}_{\mathcal{B}}\mathbf{R}^{\top\mathcal{G}}\boldsymbol{v}^{\mathcal{G}}_{\mathcal{B}}\rfloor_{\times} - D\lfloor{}^{\mathcal{G}}_{\mathcal{B}}\mathbf{R}^{\top\mathcal{G}}\boldsymbol{v}^{\mathcal{G}}_{\mathcal{B}}\rfloor_{\times}\right) - \left\lfloor\widehat{\boldsymbol{f}}_{res}\right\rfloor_{\times}\Big), \\ \frac{\partial \mathbf{H}_{\boldsymbol{d}}}{\partial^{\mathcal{I}}\boldsymbol{t}^{\mathcal{I}}_{\mathcal{B}}} &= -\lfloor{}^{\mathcal{I}}\widehat{\boldsymbol{\omega}}\rfloor_{\times}\lfloor{}^{\mathcal{I}}\widehat{\boldsymbol{\omega}}\rfloor_{\times} - \lfloor{}^{\mathcal{I}}\widehat{\boldsymbol{\alpha}}\rfloor_{\times}.\end{aligned} \tag{24}$$

Specifically, ${}^{\mathcal{I}}\widehat{\boldsymbol{\alpha}}$ is the angular acceleration in the $\mathcal{I}$ frame, which is obtained by differentiating the angular velocity, ${}^{\mathcal{I}}\widehat{\boldsymbol{\alpha}} = \frac{d}{dt}{}^{\mathcal{I}}\widehat{\boldsymbol{\omega}}$. In practice, we low-pass filter the $\widehat{\boldsymbol{\omega}}, \widehat{\boldsymbol{\alpha}}$ to reduce noise. And, the noise of IMU acceleration measurements is $\boldsymbol{n}_{\boldsymbol{a}} \sim \mathcal{N}(0, \boldsymbol{\Sigma}^2_{\boldsymbol{a}})$.

b) network velocity measurements: The velocity measurements from *V-P Net* are expressed as:

$$\begin{aligned}{}^{\mathcal{G}}\widehat{\boldsymbol{v}}^{\mathcal{G}}_{\mathcal{I}} &= \mathbf{H}_v(\boldsymbol{x}, \boldsymbol{n}_{\boldsymbol{v}}) \\ &= {}^{\mathcal{G}}\boldsymbol{v}^{\mathcal{G}}_{\mathcal{B}} - {}^{\mathcal{G}}_{\mathcal{I}}\mathbf{R}\left({}^{\mathcal{I}}\widehat{\boldsymbol{\omega}} \times {}^{\mathcal{I}}\boldsymbol{t}^{\mathcal{I}}_{\mathcal{B}}\right) + \boldsymbol{n}_{\boldsymbol{v}},\end{aligned} \tag{25}$$

whose jacobian can be given as:

$$\frac{\partial \mathbf{H}_{\boldsymbol{v}}}{\partial \boldsymbol{x}_k} = \begin{bmatrix}\mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_{31} & \mathbf{0}_3 & \mathbf{0}_3 & \frac{\partial \boldsymbol{r}_{\boldsymbol{v}}}{\partial^{\mathcal{I}}\boldsymbol{t}^{\mathcal{I}}_{\mathcal{B}}}\end{bmatrix}, \tag{26}$$

where,

$$\frac{\partial \boldsymbol{r}_{\boldsymbol{v}}}{\partial^{\mathcal{I}}\boldsymbol{t}^{\mathcal{I}}_{\mathcal{B}}} = -{}^{\mathcal{G}}_{\mathcal{I}}\mathbf{R}\lfloor{}^{\mathcal{I}}\widehat{\boldsymbol{\omega}}\rfloor_{\times}. \tag{27}$$

Specifically, the noise of network velocity measurements is $\boldsymbol{n}_{\boldsymbol{v}} \sim \mathcal{N}(0, \widehat{\boldsymbol{\Sigma}}^2_{\boldsymbol{v}})$.

c) network displacement measurements: The displacement measurements from *V-P Net* are expressed as:

$$\begin{aligned}{}^{\mathcal{G}}\widehat{\boldsymbol{p}}^{\mathcal{G}}_{\mathcal{I}} &= \mathbf{H}_p(\boldsymbol{x}, \boldsymbol{n}_{\boldsymbol{p}}) \\ &= {}^{\mathcal{G}}\boldsymbol{p}^{\mathcal{G}}_{\mathcal{B}} - {}^{\mathcal{G}}_{\mathcal{I}}\mathbf{R}^{\mathcal{I}}\boldsymbol{t}^{\mathcal{I}}_{\mathcal{B}} + \boldsymbol{n}_p,\end{aligned} \tag{28}$$

whose jacobian can be given as:

$$\frac{\partial \mathbf{H}_{\boldsymbol{p}}}{\partial \boldsymbol{x}_k} = \begin{bmatrix}\mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_{31} & \mathbf{0}_3 & \mathbf{0}_3 & \frac{\partial \mathbf{H}_{\boldsymbol{p}}}{\partial^{\mathcal{I}}\boldsymbol{t}^{\mathcal{I}}_{\mathcal{B}}}\end{bmatrix}, \tag{29}$$

where,

$$\frac{\partial \mathbf{H}_{\boldsymbol{p}}}{\partial^{\mathcal{I}}\boldsymbol{t}^{\mathcal{I}}_{\mathcal{B}}} = -{}^{\mathcal{G}}_{\mathcal{I}}\mathbf{R}. \tag{30}$$

Specifically, the noise of network displacement measurements is $\boldsymbol{n}_{\boldsymbol{p}} \sim \mathcal{N}(0, \widehat{\boldsymbol{\Sigma}}^2_{\boldsymbol{p}})$.

4) Discrete Extended Kalman Filter: The complete EKF procedures for both rotation and translation can be written in a discretized form as follows:

$$\begin{aligned}
\boldsymbol{x}_{k+1|k} &= \mathbf{F}(\boldsymbol{x}_{k|k}, \boldsymbol{u}_k), \\
\mathbf{P}_{k+1|k} &= \mathbf{F}_{\boldsymbol{x}}\mathbf{P}_{k|k}\mathbf{F}_{\boldsymbol{x}}^T + \mathbf{F}_{\boldsymbol{u}}\mathbf{Q}_k\mathbf{F}_{\boldsymbol{u}}^T, \\
\mathbf{K}_{k+1} &= \mathbf{P}_{k+1|k}\mathbf{H}_{\boldsymbol{x}}^T\left(\mathbf{H}_{\boldsymbol{x}}\mathbf{P}_{k+1|k}\mathbf{H}_{\boldsymbol{x}}^T + \mathbf{R}_{k+1}\right)^{-1}, \\
\boldsymbol{x}_{k+1|k+1} &= \boldsymbol{x}_{k+1|k} \oplus \left(\mathbf{K}_{k+1}(\boldsymbol{z}_{k+1} - h(\boldsymbol{x}_{k+1|k}))\right), \\
\mathbf{P}_{k+1|k+1} &= (\mathbf{I} - \mathbf{K}_{k+1}\mathbf{H}_{\boldsymbol{x}})\mathbf{P}_{k+1|k},
\end{aligned} \tag{31}$$

where,

$$\mathbf{Q} = \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{f}}^2, \quad \mathbf{R} = diag\{\boldsymbol{\Sigma}_{\boldsymbol{a}}^2, \quad \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{v}}^2, \quad \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{p}}^2\}. \tag{32}$$

III. One-Stage EKF for Inertial Dynamical Fusion

A. State

The state of the system is defined as:

$$\boldsymbol{x} = ({}^{\mathcal{G}}_{\mathcal{I}}\boldsymbol{q}, \ {}^{\mathcal{G}}\boldsymbol{p}^{\mathcal{G}}_{\mathcal{B}}, \ {}^{\mathcal{G}}\boldsymbol{v}^{\mathcal{G}}_{\mathcal{B}}, \ \tau, \ \boldsymbol{d}, \ {}^{\mathcal{I}}_{\mathcal{B}}\boldsymbol{q}, \ {}^{\mathcal{I}}\boldsymbol{t}^{\mathcal{I}}_{\mathcal{B}}). \tag{33}$$

B. Process Model

We regard the quadrotor dynamics and gyroscope outputs of *De-Bias Net* as the input, and express the complete process model as follows:

$$\begin{aligned}
{}^{\mathcal{G}}_{\mathcal{I}}\dot{\boldsymbol{q}} &= \frac{1}{2}{}^{\mathcal{G}}_{\mathcal{I}}\boldsymbol{q} \otimes ({}^{\mathcal{I}}\widehat{\boldsymbol{\omega}} + \boldsymbol{n}_{\boldsymbol{\omega}}), \\
{}^{\mathcal{G}}\dot{\boldsymbol{p}}^{\mathcal{G}}_{\mathcal{B}} &= {}^{\mathcal{G}}\boldsymbol{v}^{\mathcal{G}}_{\mathcal{B}}, \\
{}^{\mathcal{G}}\dot{\boldsymbol{v}}^{\mathcal{G}}_{\mathcal{B}} &= \frac{1}{m}{}^{\mathcal{G}}_{\mathcal{B}}\mathbf{R}\left(\tau U_{ss}\boldsymbol{e}_3 - U_s D^{\mathcal{G}}_{\mathcal{B}}\mathbf{R}^{\top\mathcal{G}}\boldsymbol{v}^{\mathcal{G}}_{\mathcal{B}} + \widehat{\boldsymbol{f}}_{res} + \boldsymbol{n}_{\boldsymbol{f}}\right) - {}^{\mathcal{G}}\mathbf{g}, \\
\dot{\tau} &= 0, \\
\dot{\boldsymbol{d}} &= \mathbf{0}, \\
{}^{\mathcal{I}}_{\mathcal{B}}\dot{\boldsymbol{q}} &= \mathbf{0}, \\
{}^{\mathcal{I}}\dot{\boldsymbol{t}}^{\mathcal{I}}_{\mathcal{B}} &= \mathbf{0},
\end{aligned} \tag{34}$$

The above differential equation is discretized as follows:

$$\begin{aligned}
\boldsymbol{x}_{k+1} &= \mathbf{F}(\boldsymbol{x}_k, \boldsymbol{u}_k) \\
&= \mathbf{F}_{\boldsymbol{x}}\boldsymbol{x}_k + \mathbf{F}_{\boldsymbol{n}}\boldsymbol{n}_k,
\end{aligned} \tag{35}$$

where,

$$\begin{aligned}
\mathbf{F}_{\boldsymbol{x}} &= \begin{bmatrix}
\frac{\partial^{\mathcal{G}}_{\mathcal{I}}\boldsymbol{\theta}_{k+1}}{\partial^{\mathcal{G}}_{\mathcal{I}}\boldsymbol{\theta}_k} & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_{31} & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\
\mathbf{0}_3 & \mathbf{I}_3 & \mathbf{I}_3\Delta t & \mathbf{0}_{31} & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\
\frac{\partial \boldsymbol{v}_{k+1}}{\partial^{\mathcal{G}}_{\mathcal{I}}\boldsymbol{\theta}_k} & \mathbf{0}_3 & \frac{\partial \boldsymbol{v}_{k+1}}{\partial \boldsymbol{v}_k} & \frac{\partial \boldsymbol{v}_{k+1}}{\partial \tau} & \frac{\partial \boldsymbol{v}_{k+1}}{\partial \boldsymbol{d}} & \frac{\partial \boldsymbol{v}_{k+1}}{\partial^{\mathcal{I}}_{\mathcal{B}}\boldsymbol{\theta}} & \mathbf{0}_3 \\
\mathbf{0}_3 & \mathbf{0}_{13} & \mathbf{0}_{13} & 1 & \mathbf{0}_{13} & \mathbf{0}_{13} & \mathbf{0}_{13} \\
\mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_{31} & \mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\
\mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_{31} & \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_3 \\
\mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_{31} & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{I}_3
\end{bmatrix}, \\
\mathbf{F}_{\boldsymbol{n}} &= \begin{bmatrix}
\frac{\partial^{\mathcal{G}}_{\mathcal{I}}\boldsymbol{\theta}_{k+1}}{\partial \boldsymbol{n}_{\boldsymbol{\omega}}} & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_{31} & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\
\mathbf{0}_3 & \mathbf{0}_3 & \frac{\partial \boldsymbol{v}_{k+1}}{\partial \boldsymbol{n}_{\boldsymbol{f}}} & \mathbf{0}_{31} & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3
\end{bmatrix}^{\top}.
\end{aligned} \tag{36}$$

Specifically,

$$
\begin{aligned}
\frac{\partial {}^{\mathcal{G}}_{\mathcal{I}}\boldsymbol{\theta}_{k+1}}{\partial {}^{\mathcal{G}}_{\mathcal{I}}\boldsymbol{\theta}_{k}} =& \mathbf{Exp}(-{}^{\mathcal{I}}\widehat{\boldsymbol{\omega}}\Delta t), \\
\frac{\partial \boldsymbol{v}_{k+1}}{\partial {}^{\mathcal{G}}_{\mathcal{I}}\boldsymbol{\theta}_{k}} =& \frac{\Delta t}{m} {}^{\mathcal{G}}_{\mathcal{I}}\mathbf{R}\Big(-\tau U_{ss} \left\lfloor {}^{\mathcal{I}}_{\mathcal{B}}\mathbf{r}_3 \right\rfloor_\times + U_s \left({}^{\mathcal{G}}_{\mathcal{I}}\mathbf{R} \left\lfloor {}^{\mathcal{I}}_{\mathcal{B}}\mathbf{R} D {}^{\mathcal{G}}_{\mathcal{B}}\mathbf{R}^{\top \mathcal{G}} \boldsymbol{v}^{\mathcal{G}}_{\mathcal{B}} \right\rfloor_\times - {}^{\mathcal{G}}_{\mathcal{B}}\mathbf{R} D {}^{\mathcal{I}}_{\mathcal{B}}\mathbf{R}^{\top} \left\lfloor {}^{\mathcal{G}}_{\mathcal{I}}\mathbf{R}^{\top \mathcal{G}} \boldsymbol{v}^{\mathcal{G}}_{\mathcal{B}} \right\rfloor_\times \right) - \left\lfloor {}^{\mathcal{I}}_{\mathcal{B}}\mathbf{R} \widehat{\boldsymbol{f}}_{res} \right\rfloor_\times \Big), \\
\frac{\partial \boldsymbol{v}_{k+1}}{\partial \boldsymbol{v}_{k}} =& \mathbf{I} - \frac{U_s \Delta t}{m} {}^{\mathcal{G}}_{\mathcal{B}}\mathbf{R} D {}^{\mathcal{G}}_{\mathcal{B}}\mathbf{R}^{\top}, \\
\frac{\partial \boldsymbol{v}_{k+1}}{\partial \tau} =& \frac{U_{ss}\Delta t}{m} {}^{\mathcal{G}}_{\mathcal{B}}\boldsymbol{r}_3, \\
\frac{\partial \boldsymbol{v}_{k+1}}{\partial \boldsymbol{d}} =& -\frac{\Delta t}{m} \left[{}^{\mathcal{G}}_{\mathcal{B}}\boldsymbol{r}_1 {}^{\mathcal{G}}_{\mathcal{B}}\boldsymbol{r}_1^{\top \mathcal{G}} \boldsymbol{v}^{\mathcal{G}}_{\mathcal{B}} \quad {}^{\mathcal{G}}_{\mathcal{B}}\boldsymbol{r}_2 {}^{\mathcal{G}}_{\mathcal{B}}\boldsymbol{r}_2^{\top \mathcal{G}} \boldsymbol{v}^{\mathcal{G}}_{\mathcal{B}} \quad {}^{\mathcal{G}}_{\mathcal{B}}\boldsymbol{r}_3 {}^{\mathcal{G}}_{\mathcal{B}}\boldsymbol{r}_3^{\top \mathcal{G}} \boldsymbol{v}^{\mathcal{G}}_{\mathcal{B}} \right], \\
\frac{\partial \boldsymbol{v}_{k+1}}{\partial {}^{\mathcal{I}}_{\mathcal{B}}\boldsymbol{\theta}} =& \frac{\Delta t}{m} {}^{\mathcal{G}}_{\mathcal{B}}\mathbf{R}\Big(-\tau U_{ss} \left\lfloor \boldsymbol{e}_3 \right\rfloor_\times + U_s \left(\left\lfloor D {}^{\mathcal{G}}_{\mathcal{B}}\mathbf{R}^{\top \mathcal{G}} \boldsymbol{v}^{\mathcal{G}}_{\mathcal{B}} \right\rfloor_\times - D \left\lfloor {}^{\mathcal{G}}_{\mathcal{B}}\mathbf{R}^{\top \mathcal{G}} \boldsymbol{v}^{\mathcal{G}}_{\mathcal{B}} \right\rfloor_\times \right) - \left\lfloor \widehat{\boldsymbol{f}}_{res} \right\rfloor_\times \Big), \\
\frac{\partial {}^{\mathcal{G}}_{\mathcal{I}}\boldsymbol{\theta}_{k+1}}{\partial \boldsymbol{n}_{\boldsymbol{\omega}}} =& \mathbf{J}_r \left({}^{\mathcal{I}}\widehat{\boldsymbol{\omega}}\Delta t \right) \Delta t, \\
\frac{\partial \boldsymbol{v}_{k+1}}{\partial \boldsymbol{n}_{\boldsymbol{f}}} =& \frac{1}{m} {}^{\mathcal{G}}_{\mathcal{B}}\mathbf{R}\Delta t,
\end{aligned}
\tag{37}
$$

where,

$$
\mathbf{J}_r(\boldsymbol{\theta}) = \mathbf{I}_3 - \frac{1 - cos||\boldsymbol{\theta}||}{||\boldsymbol{\theta}||^2} \left\lfloor \boldsymbol{\theta} \right\rfloor_\times + \frac{||\boldsymbol{\theta}|| - sin||\boldsymbol{\theta}||}{||\boldsymbol{\theta}||^3} \left\lfloor \boldsymbol{\theta} \right\rfloor_\times^2 \tag{38}
$$

C. Measurement Model

1) gravity alignment measurements: The attitude controllers of most flying robots rely on the complementary filters [2, 3] to obtain attitude observations. Similarly, we consider the gravity alignment constraint to obtain the tilt observation:

$$
\begin{aligned}
{}^{\mathcal{I}}\widehat{\boldsymbol{a}} &= \mathbf{H}_g(\boldsymbol{x}, \boldsymbol{n}_{\boldsymbol{g}}) \\
&\approx {}^{\mathcal{G}}_{\mathcal{I}}\mathbf{R}^{\top \mathcal{G}}\mathbf{g} + \boldsymbol{n}_{\boldsymbol{g}},
\end{aligned}
\tag{39}
$$

whose jacobian can be given as:

$$
\frac{\partial \mathbf{H}_{\boldsymbol{g}}}{\partial \boldsymbol{x}_k} = \left[\frac{\partial \mathbf{H}_{\boldsymbol{g}}}{\partial {}^{\mathcal{G}}_{\mathcal{I}}\boldsymbol{\theta}_k} \quad \mathbf{0}_3 \quad \mathbf{0}_3 \quad \mathbf{0}_{31} \quad \mathbf{0}_3 \quad \mathbf{0}_3 \quad \mathbf{0}_3 \right], \tag{40}
$$

where,

$$
\frac{\partial \mathbf{H}_{\boldsymbol{g}}}{\partial {}^{\mathcal{G}}_{\mathcal{I}}\boldsymbol{\theta}_k} = \left\lfloor {}^{\mathcal{G}}_{\mathcal{I}}\mathbf{R}^{\top \mathcal{G}}\mathbf{g} \right\rfloor_\times . \tag{41}
$$

2) dynamics constraint measurements: The dynamics constraint measurements are expressed as:

$$
\begin{aligned}
{}^{\mathcal{I}}\widehat{\boldsymbol{a}} &= \mathbf{H}_d(\boldsymbol{x}, \boldsymbol{n}_{\boldsymbol{a}}) \\
&= \frac{1}{m} {}^{\mathcal{I}}_{\mathcal{B}}\mathbf{R} \left(\tau U_{ss}\boldsymbol{e}_3 - U_s D {}^{\mathcal{G}}_{\mathcal{B}}\mathbf{R}^{\top \mathcal{G}} \boldsymbol{v}^{\mathcal{G}}_{\mathcal{B}} + \widehat{\boldsymbol{f}}_{res} \right) - {}^{\mathcal{I}}\widehat{\boldsymbol{\omega}} \times ({}^{\mathcal{I}}\widehat{\boldsymbol{\omega}} \times {}^{\mathcal{I}}\boldsymbol{t}^{\mathcal{I}}_{\mathcal{B}}) - {}^{\mathcal{I}}\widehat{\boldsymbol{\alpha}} \times {}^{\mathcal{I}}\boldsymbol{t}^{\mathcal{I}}_{\mathcal{B}} + \boldsymbol{n}_{\boldsymbol{a}},
\end{aligned}
\tag{42}
$$

whose jacobian can be given as:

$$
\frac{\partial \mathbf{H}_{\boldsymbol{d}}}{\partial \boldsymbol{x}_k} = \left[\frac{\partial \mathbf{H}_{\boldsymbol{d}}}{\partial {}^{\mathcal{G}}_{\mathcal{I}}\boldsymbol{\theta}_k} \quad \mathbf{0}_3 \quad \frac{\partial \mathbf{H}_{\boldsymbol{d}}}{\partial \boldsymbol{v}_k} \quad \frac{\partial \mathbf{H}_{\boldsymbol{d}}}{\partial \tau} \quad \frac{\partial \mathbf{H}_{\boldsymbol{d}}}{\partial \boldsymbol{d}} \quad \frac{\partial \mathbf{H}_{\boldsymbol{d}}}{\partial {}^{\mathcal{I}}_{\mathcal{B}}\boldsymbol{\theta}} \quad \frac{\partial \mathbf{H}_{\boldsymbol{d}}}{\partial {}^{\mathcal{I}}\boldsymbol{t}^{\mathcal{I}}_{\mathcal{B}}} \right], \tag{43}
$$

where,

$$
\begin{aligned}
\frac{\partial \mathbf{H}_{\boldsymbol{d}}}{\partial {}_{\mathcal{I}}^{\mathcal{G}}\boldsymbol{\theta}_k} &= -\frac{U_s}{m}{}_{\mathcal{B}}^{\mathcal{I}}\mathbf{R}D{}_{\mathcal{B}}^{\mathcal{I}}\mathbf{R}^\top \left\lfloor {}_{\mathcal{I}}^{\mathcal{G}}\mathbf{R}^{\top\mathcal{G}}\boldsymbol{v}_{\mathcal{B}}^{\mathcal{G}} \right\rfloor_\times, \\
\frac{\partial \mathbf{H}_{\boldsymbol{d}}}{\partial \boldsymbol{v}_k} &= -\frac{U_s}{m}{}_{\mathcal{B}}^{\mathcal{I}}\mathbf{R}D{}_{\mathcal{B}}^{\mathcal{G}}\mathbf{R}^\top, \\
\frac{\partial \mathbf{H}_{\boldsymbol{d}}}{\partial \tau} &= \frac{U_{ss}}{m}{}_{\mathcal{B}}^{\mathcal{I}}\boldsymbol{r}_3, \\
\frac{\partial \mathbf{H}_{\boldsymbol{d}}}{\partial \boldsymbol{d}} &= -\frac{U_s}{m}\left[{}_{\mathcal{B}}^{\mathcal{I}}\boldsymbol{r}_1{}_{\mathcal{B}}^{\mathcal{G}}\boldsymbol{r}_1^{\top\mathcal{G}}\boldsymbol{v}_{\mathcal{B}}^{\mathcal{G}} \quad {}_{\mathcal{B}}^{\mathcal{I}}\boldsymbol{r}_2{}_{\mathcal{B}}^{\mathcal{G}}\boldsymbol{r}_2^{\top\mathcal{G}}\boldsymbol{v}_{\mathcal{B}}^{\mathcal{G}} \quad {}_{\mathcal{B}}^{\mathcal{I}}\boldsymbol{r}_3{}_{\mathcal{B}}^{\mathcal{G}}\boldsymbol{r}_3^{\top\mathcal{G}}\boldsymbol{v}_{\mathcal{B}}^{\mathcal{G}}\right], \\
\frac{\partial \mathbf{H}_{\boldsymbol{d}}}{\partial {}_{\mathcal{B}}^{\mathcal{I}}\boldsymbol{\theta}} &= \frac{1}{m}{}_{\mathcal{B}}^{\mathcal{I}}\mathbf{R}\Big(- kU_{ss}\left\lfloor \boldsymbol{e}_3 \right\rfloor_\times + U_s\left(\left\lfloor D{}_{\mathcal{B}}^{\mathcal{G}}\mathbf{R}^{\top\mathcal{G}}\boldsymbol{v}_{\mathcal{B}}^{\mathcal{G}}\right\rfloor_\times - D\left\lfloor {}_{\mathcal{B}}^{\mathcal{G}}\mathbf{R}^{\top\mathcal{G}}\boldsymbol{v}_{\mathcal{B}}^{\mathcal{G}}\right\rfloor_\times\right) - \left\lfloor \widehat{\boldsymbol{f}}_{res}\right\rfloor_\times\Big), \\
\frac{\partial \mathbf{H}_{\boldsymbol{d}}}{\partial {}^{\mathcal{I}}\boldsymbol{t}_{\mathcal{B}}^{\mathcal{I}}} &= -\left\lfloor {}^{\mathcal{I}}\widehat{\boldsymbol{\omega}}\right\rfloor_\times \left\lfloor {}^{\mathcal{I}}\widehat{\boldsymbol{\omega}}\right\rfloor_\times - \left\lfloor {}^{\mathcal{I}}\widehat{\boldsymbol{\alpha}}\right\rfloor_\times.
\end{aligned}
\tag{44}
$$

Specifically, ${}^{\mathcal{I}}\widehat{\boldsymbol{\alpha}}$ is the angular acceleration in the $\mathcal{I}$ frame, which is obtained by differentiating the angular velocity, ${}^{\mathcal{I}}\widehat{\boldsymbol{\alpha}} = \frac{d}{dt}{}^{\mathcal{I}}\widehat{\boldsymbol{\omega}}$. In practice, we low-pass filter the $\widehat{\boldsymbol{\omega}}, \widehat{\boldsymbol{\alpha}}$ to reduce noise. And, the noise of IMU acceleration measurements is $\boldsymbol{n}_{\boldsymbol{a}} \sim \mathcal{N}(0, \boldsymbol{\Sigma}_{\boldsymbol{a}}^2)$.

3) network velocity measurements: The velocity measurements from *V-P Net* are expressed as:

$$
\begin{aligned}
{}^{\mathcal{G}}\widehat{\boldsymbol{v}}_{\mathcal{I}}^{\mathcal{G}} &= \mathbf{H}_v(\boldsymbol{x}, \boldsymbol{n}_{\boldsymbol{v}}) \\
&= {}^{\mathcal{G}}\boldsymbol{v}_{\mathcal{B}}^{\mathcal{G}} - {}_{\mathcal{I}}^{\mathcal{G}}\mathbf{R}\left({}^{\mathcal{I}}\widehat{\boldsymbol{\omega}} \times {}^{\mathcal{I}}\boldsymbol{t}_{\mathcal{B}}^{\mathcal{I}}\right) + \boldsymbol{n}_{\boldsymbol{v}},
\end{aligned}
\tag{45}
$$

whose jacobian can be given as:

$$
\frac{\partial \mathbf{H}_{\boldsymbol{v}}}{\partial \boldsymbol{x}_k} = \begin{bmatrix} \frac{\partial \mathbf{H}_{\boldsymbol{v}}}{\partial {}_{\mathcal{I}}^{\mathcal{G}}\boldsymbol{\theta}_k} & \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_{31} & \mathbf{0}_3 & \mathbf{0}_3 & \frac{\partial \mathbf{H}_{\boldsymbol{v}}}{\partial {}^{\mathcal{I}}\boldsymbol{t}_{\mathcal{B}}^{\mathcal{I}}} \end{bmatrix},
\tag{46}
$$

where,

$$
\begin{aligned}
\frac{\partial \mathbf{H}_{\boldsymbol{v}}}{\partial {}_{\mathcal{I}}^{\mathcal{G}}\boldsymbol{\theta}_k} &= {}_{\mathcal{I}}^{\mathcal{G}}\mathbf{R}\left\lfloor {}^{\mathcal{I}}\widehat{\boldsymbol{\omega}} \times {}^{\mathcal{I}}\boldsymbol{t}_{\mathcal{B}}^{\mathcal{I}}\right\rfloor_\times, \\
\frac{\partial \mathbf{H}_{\boldsymbol{v}}}{\partial {}^{\mathcal{I}}\boldsymbol{t}_{\mathcal{B}}^{\mathcal{I}}} &= -{}_{\mathcal{I}}^{\mathcal{G}}\mathbf{R}\left\lfloor {}^{\mathcal{I}}\widehat{\boldsymbol{\omega}}\right\rfloor_\times.
\end{aligned}
\tag{47}
$$

Specifically, the noise of network velocity measurements is $\boldsymbol{n}_{\boldsymbol{v}} \sim \mathcal{N}(0, \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{v}}^2)$.

4) network displacement measurements: The displacement measurements from *V-P Net* are expressed as:

$$
\begin{aligned}
{}^{\mathcal{G}}\widehat{\boldsymbol{p}}_{\mathcal{I}}^{\mathcal{G}} &= \mathbf{H}_p(\boldsymbol{x}, \boldsymbol{n}_{\boldsymbol{p}}) \\
&= {}^{\mathcal{G}}\boldsymbol{p}_{\mathcal{B}}^{\mathcal{G}} - {}_{\mathcal{I}}^{\mathcal{G}}\mathbf{R}{}^{\mathcal{I}}\boldsymbol{t}_{\mathcal{B}}^{\mathcal{I}} + \boldsymbol{n}_p,
\end{aligned}
\tag{48}
$$

whose jacobian can be given as:

$$
\frac{\partial \mathbf{H}_{\boldsymbol{p}}}{\partial \boldsymbol{x}_k} = \begin{bmatrix} \frac{\partial \mathbf{H}_{\boldsymbol{p}}}{\partial {}_{\mathcal{I}}^{\mathcal{G}}\boldsymbol{\theta}_k} & \mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_{31} & \mathbf{0}_3 & \mathbf{0}_3 & \frac{\partial \mathbf{H}_{\boldsymbol{p}}}{\partial {}^{\mathcal{I}}\boldsymbol{t}_{\mathcal{B}}^{\mathcal{I}}} \end{bmatrix},
\tag{49}
$$

where,

$$
\begin{aligned}
\frac{\partial \mathbf{H}_{\boldsymbol{p}}}{\partial {}_{\mathcal{I}}^{\mathcal{G}}\boldsymbol{\theta}_k} &= {}_{\mathcal{I}}^{\mathcal{G}}\mathbf{R}\left\lfloor {}^{\mathcal{I}}\boldsymbol{t}_{\mathcal{B}}^{\mathcal{I}}\right\rfloor_\times, \\
\frac{\partial \mathbf{H}_{\boldsymbol{p}}}{\partial {}^{\mathcal{I}}\boldsymbol{t}_{\mathcal{B}}^{\mathcal{I}}} &= -{}_{\mathcal{I}}^{\mathcal{G}}\mathbf{R}.
\end{aligned}
\tag{50}
$$

Specifically, the noise of network displacement measurements is $\boldsymbol{n}_{\boldsymbol{p}} \sim \mathcal{N}(0, \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{p}}^2)$.

D. Discrete Extended Kalman Filter

The complete EKF procedures for both rotation and translation can be written in a discretized form as follows:

$$
\begin{aligned}
\boldsymbol{x}_{k+1|k} &= \mathbf{F}(\boldsymbol{x}_{k|k}, \boldsymbol{u}_k), \\
\mathbf{P}_{k+1|k} &= \mathbf{F}_{\boldsymbol{x}} \mathbf{P}_{k|k} \mathbf{F}_{\boldsymbol{x}}^T + \mathbf{F}_{\boldsymbol{u}} \mathbf{Q}_k \mathbf{F}_{\boldsymbol{u}}^T, \\
\mathbf{K}_{k+1} &= \mathbf{P}_{k+1|k} \mathbf{H}_{\boldsymbol{x}}^T \left(\mathbf{H}_{\boldsymbol{x}} \mathbf{P}_{k+1|k} \mathbf{H}_{\boldsymbol{x}}^T + \mathbf{R}_{k+1} \right)^{-1}, \\
\boldsymbol{x}_{k+1|k+1} &= \boldsymbol{x}_{k+1|k} \oplus \left(\mathbf{K}_{k+1} (\boldsymbol{z}_{k+1} - h(\boldsymbol{x}_{k+1|k})) \right), \\
\mathbf{P}_{k+1|k+1} &= (\mathbf{I} - \mathbf{K}_{k+1} \mathbf{H}_{\boldsymbol{x}}) \mathbf{P}_{k+1|k},
\end{aligned}
\tag{51}
$$

where,

$$
\mathbf{Q} = diag\{\boldsymbol{\Sigma}_{\boldsymbol{\omega}}^2, \quad \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{f}}^2\}, \quad \mathbf{R} = diag\{\boldsymbol{\Sigma}_{\boldsymbol{a}}^2, \quad \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{v}}^2, \quad \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{p}}^2\}. \tag{52}
$$

IV. Observability Analysis

According to the observability analysis method developed in [4], we could analyze the observability of a control affine system by checking the observability rank criterion.

For the rotation stage which is driven by Eq. (9) and observed by Eq. (12), the rotation ${}^{\mathcal{G}}_{\mathcal{I}}\boldsymbol{q}$ is composed of two observable angles (roll and pitch) and an unobservable yaw angle.

For the translation stage, we first write the process model of the system in control affine form:

$$
\dot{\boldsymbol{x}} = \boldsymbol{f}_0(\boldsymbol{x}) + \sum_{i=1}^{n} \boldsymbol{f}_i(\boldsymbol{x}) \boldsymbol{u}_i. \tag{53}
$$

Since the four motor speeds of the quadrotor are integrated into the two inputs U_{ss} and U_s in Eq. (18), our system could be presented:

$$
\begin{aligned}
\dot{\boldsymbol{x}} &= \boldsymbol{f}_0(\boldsymbol{x}) + \boldsymbol{f}_1(\boldsymbol{x}) U_{ss} + \boldsymbol{f}_2(\boldsymbol{x}) U_s \\
&= \begin{bmatrix} {}^{\mathcal{G}}\boldsymbol{v}_{\mathcal{B}}^{\mathcal{G}} \\ -{}^{\mathcal{G}}\mathbf{g} \\ \mathbf{0}_{10\times1} \end{bmatrix} + \begin{bmatrix} \mathbf{0}_{3\times1} \\ \frac{1}{m} {}^{\mathcal{G}}_{\mathcal{B}}\mathbf{R} \tau \boldsymbol{e}_3 \\ \mathbf{0}_{10\times1} \end{bmatrix} U_{ss} + \begin{bmatrix} \mathbf{0}_{3\times1} \\ -\frac{1}{m} {}^{\mathcal{G}}_{\mathcal{B}}\mathbf{R} D_{\mathcal{B}}^{\mathcal{G}} \mathbf{R}^{\top} {}^{\mathcal{G}}\boldsymbol{v}_{\mathcal{B}}^{\mathcal{G}} \\ \mathbf{0}_{10\times1} \end{bmatrix} U_s.
\end{aligned}
\tag{54}
$$

And then, we simplify the three measurement models Eq. (22, 25, 28) as:

$$
\begin{aligned}
\boldsymbol{h}_d &= \frac{1}{m} {}^{\mathcal{I}}_{\mathcal{B}}\mathbf{R} \left(\tau U_{ss} \boldsymbol{e}_3 - U_s D_{\mathcal{B}}^{\mathcal{G}} \mathbf{R}^{\top} {}^{\mathcal{G}}\boldsymbol{v}_{\mathcal{B}}^{\mathcal{G}} \right) - {}^{\mathcal{I}}\widehat{\boldsymbol{\omega}} \times ({}^{\mathcal{I}}\widehat{\boldsymbol{\omega}} \times {}^{\mathcal{I}}\boldsymbol{t}_{\mathcal{B}}^{\mathcal{I}}) - {}^{\mathcal{I}}\widehat{\boldsymbol{\alpha}} \times {}^{\mathcal{I}}\boldsymbol{t}_{\mathcal{B}}^{\mathcal{I}}, \\
\boldsymbol{h}_v &= {}^{\mathcal{G}}\boldsymbol{v}_{\mathcal{B}}^{\mathcal{G}} - {}^{\mathcal{G}}_{\mathcal{I}}\mathbf{R} \left({}^{\mathcal{I}}\widehat{\boldsymbol{\omega}} \times {}^{\mathcal{I}}\boldsymbol{t}_{\mathcal{B}}^{\mathcal{I}} \right), \\
\boldsymbol{h}_p &= {}^{\mathcal{G}}\boldsymbol{p}_{\mathcal{B}}^{\mathcal{G}} - {}^{\mathcal{G}}_{\mathcal{I}}\mathbf{R} {}^{\mathcal{I}}\boldsymbol{t}_{\mathcal{B}}^{\mathcal{I}}.
\end{aligned}
\tag{55}
$$

We use Lie derivatives to quantify the impact of changes in the control input U_{ss} and U_s on the output functions $\boldsymbol{h}_d$, $\boldsymbol{h}_v$ and $\boldsymbol{h}_p$.

$$
\begin{aligned}
L^0 \boldsymbol{h} &= \boldsymbol{h}, \\
L^{k+1}_{\boldsymbol{f}_{i_1}, \boldsymbol{f}_{i_2}, \ldots, \boldsymbol{f}_{i_{k+1}}} \boldsymbol{h} &= \nabla_{\boldsymbol{x}} \left(L^k_{\boldsymbol{f}_{i_1}, \boldsymbol{f}_{i_2}, \ldots, \boldsymbol{f}_{i_k}} \boldsymbol{h} \right) \boldsymbol{f}_{i_{k+1}}
\end{aligned}
\tag{56}
$$

We stack vertically several Lie derivatives of the unforced vector field $\boldsymbol{f}_0$ and the control input vector fields $\boldsymbol{f}_1$ and $\boldsymbol{f}_2$ into a vector $\mathcal{O}$:

$$
\mathcal{O} = \begin{bmatrix} \boldsymbol{h}_d \\ \boldsymbol{h}_v \\ \boldsymbol{h}_p \\ L_{\boldsymbol{f}_0} \boldsymbol{h}_d \\ L_{\boldsymbol{f}_0} \boldsymbol{h}_p \\ L_{\boldsymbol{f}_1} \boldsymbol{h}_d \\ L_{\boldsymbol{f}_1} \boldsymbol{h}_v \\ L_{\boldsymbol{f}_2} \boldsymbol{h}_v \end{bmatrix}, \tag{57}
$$

and calculate its gradients as observability matrix $\nabla_{\boldsymbol{x}}\mathcal{O}$. Finally, we can evaluate whether the system is locally observable by checking the observability rank criterion. In general, the rank is:

$$rank\{\nabla_{\boldsymbol{x}}\mathcal{O}\} = 11. \tag{58}$$

However, under some certain conditions the matrix has a rank deficiency. Some of these cases are when:

$$\begin{cases} {}^{\mathcal{I}}\widehat{\boldsymbol{\omega}} = \mathbf{0}_{3\times1}, {}^{\mathcal{I}}\boldsymbol{t}_{\mathcal{B}}^{\mathcal{I}} \ is\ unobservable; \\ {}^{\mathcal{G}}\boldsymbol{v}_{\mathcal{B}}^{\mathcal{G}} = \mathbf{0}_{3\times1}, D\ and\ {}_{\mathcal{B}}^{\mathcal{I}}\mathbf{R}\ are\ unobservable; \\ {}^{\mathcal{G}}\boldsymbol{v}_{\mathcal{B}}^{\mathcal{G}} = [0;0;*], yaw({}_{\mathcal{B}}^{\mathcal{I}}\mathbf{R})\ is\ unobservable, \end{cases} \tag{59}$$

which means the corresponding parameters to be estimated may not converge under the above conditions.

REFERENCES

[1] J. Svacha, J. Paulos, G. Loianno, and V. Kumar, "Imu-based inertia estimation for a quadrotor using newton-euler dynamics," *IEEE Robotics and Automation Letters*, vol. 5, no. 3, pp. 3861–3867, 2020.

[2] R. Mahony, T. Hamel, and J.-M. Pflimlin, "Nonlinear complementary filters on the special orthogonal group," *IEEE Transactions on automatic control*, vol. 53, no. 5, pp. 1203–1218, 2008.

[3] S. O. Madgwick, A. J. Harrison, and R. Vaidyanathan, "Estimation of imu and marg orientation using a gradient descent algorithm," in *2011 IEEE international conference on rehabilitation robotics*. IEEE, 2011, pp. 1–7.

[4] R. Hermann and A. Krener, "Nonlinear controllability and observability," *IEEE Transactions on automatic control*, vol. 22, no. 5, pp. 728–740, 1977.