OpenVINS State Initialization: Details and Derivations

Patrick Geneva - pgeneva@udel.edu
Guoquan Huang - ghuang@udel.edu

Department of Mechanical Engineering
University of Delaware, Delaware, USA

RPNG

Robot Perception and Navigation Group (RPNG)
Tech Report - RPNG-2022-INIT
Last Updated - March 10, 2022

Contents

1 Introduction **1**

2 Related Works **1**

3 Least-squares Problem **2**
3.1 Inertial Model 2
3.2 Feature Bearing Observations 3
3.3 Discussion on Frame Transformation 4
3.4 The Linear $\mathbf{Ax} = \mathbf{b}$ Problem 4
3.5 Quadratic Constrained Least-squares 5
3.6 Recovering Inertial States 6

4 Maximum Likelihood Estimation **7**
4.1 Inertial Measurement Model 7
4.2 Camera Measurement Model 9
4.3 Prior Cost 9

References **10**

1 Introduction

A fundamental problem for visual-inertial navigation systems (VINS) is their state initialization. That is, to recover the initial observable states given inertial and visual-bearing measurements. It is well known that VINS have 4 degree-of-freedom unobservable directions corresponding to the global yaw and position [1]. This means that the other parameters are observable and thus they are recoverable given inertial and visual measurement readings. The challenge of state initialization is to recover these observable parameters without *any* prior knowledge through the construction of a linear system. This is not straightforward as orientation is non-linear, and thus without approximations and tricks, it is difficult to define such a linear problem. In what follows, we base our work off of that by Dong-Si and Mourikis which present a closed-form solution to the problem and define a quadratic constrained least-squares problem [2, 3]. An additional focus is that we wish to recover the initial state covariance for use in an filter-based visual-inertial extended Kalman filter OpenVINS [4].

2 Related Works

This is a non-exhaustive list of related work which provide a good background on the state initialization problem. Please refer to their own citations for more relevant works.

- Dong-Si and Mourikis [2, 3] – Presented methods for recovering a visual-inertial platform's attitude, velocity, feature positions, and camera-IMU extrinsic calibration. They first discussed two methods for recovering the relative rotation between the camera and IMU under ≥ 5 and a minimal solution for the case when ≥ 2 environmental features. They then formulated a orientation independent linear system which recovered the remaining quantities through a quadratically constrained least-squares problem since the magnitude of gravity is assumed known. This was then followed with a non-linear optimization to refine the state. In their technical report they additionally showed closed-form solutions for recovering IMU biases and the initialization problem sensitives to number of features and acceleration profiles.

- Agostino Martinelli [5] – Investigated the closed-form solutions to the visual-inertial problem. Additionally the minimal cases and unique solutions of each were determined in an extensive analysis. They also investigated the solution under biased inertial data and showed that very restrictive trajectories are required to fully recover the system. This closed-form solution was applied in a subsequent work by Jacques et al. [6] which focused on evaluating sensitivities to inertial biases and using a non-linear optimization to recover the gyroscope bias.

- Yang and Shen [7] – Leveraged preintegration to reduce the computation of the closed-form initialization. They additionally focused on initializing the camera-IMU transformation and modeling of measurement noise in their initialization problems.

- Qin et al. [8, 9] – Presented a robust system which leveraged a structure-from-motion result and aligned this trajectory to the inertial preintegration measurements to recover the velocity at each pose, scale, and gravity. This approach additionally recovered the gyroscope biases by optimizing the relative camera and IMU rotation changes using the initial guess of $\mathbf{b}_g$ being zero. This implementation was open-sourced in their VINS-Mono repository.

- Mur-Artal et al. [10] and Campos et al. [11, 12] – Presented an initialization method which removed the need to recover the velocity of the platform through the using two preintegrated

measurements between three keyframes to cancel all velocity terms. Once the gravity and scale of the structure-from-motion is recovered, the velocities were directly recovered through the measurement equations. They additionally introduced a "observability test" which checks the singular values of the resulting information to ensure that all parameters are recoverable, which was more computationally efficient compared to checking the invertibility of the problem.

- Evangelidis and Micusik [13] – Focused on reducing the computational demands of the linear system. They show that due to the special structure of the problem, cheap elimination can be performed by using a measurement model where both the depth of the feature in each frame and the 3D global position of the feature are considered unknown. The proposed method is investigated in terms of efficiency and accuracy.

- Zuñiga-Noël et al. [14] – More recently performed extensive validation of a similar constrained least-squares problem to Dong-Si and Mourikis [2] which instead relied on up to scale structure-from-motion poses as compared to raw feature bearing measurements. They additionally explicitly modeled the sensor noises and first recovered the gyroscope bias by optimizing the relative camera and IMU rotation changes using the initial guess of $\mathbf{b}_g$ being zero. They then recovered the scale, accelerometer bias, and gravity in a constrained least-squares system. They open sourced their code implementation.

- Scheiber et al. [15] and Delaune et al. [16] – Investigated the use of a single light-weight laser-range finder (LRF) to the initialization and scale observability problem with visual-inertial systems. They showed, as expected, this extra source of information ensure that under constant velocity motion. This is of particular importance for the state initialization problem since it is typically short and thus typically has constant acceleration on most robotic platforms. They showed that through incorporation of the LRF decreased initialization errors significantly. This is a promising direction if LRF are available.

3 Least-squares Problem

3.1 Inertial Model

The inertial measurement unit (IMU) provides angular velocities $\boldsymbol{\omega}$ and linear accelerations $\mathbf{a}$ in the inertial frame. These can be used to recover how the state evolves from one timestep to the next with the following state dynamics:

$$
{}^{k+1}_{G}\mathbf{R} = {}^{k+1}_{k}\Delta\mathbf{R}\ {}^{k}_{G}\mathbf{R} \tag{1}
$$

$$
{}^{G}\mathbf{p}_{k+1} = {}^{G}\mathbf{p}_{k} + {}^{G}\mathbf{v}_{k}\Delta T - \frac{1}{2}{}^{G}\mathbf{g}\Delta T^2 + {}^{k}_{G}\mathbf{R}^{\top}\int_{t_k}^{t_{k+1}}\int_{t_k}^{s}{}^{k}_{u}\Delta\mathbf{R}\left(\mathbf{a}_m - \mathbf{b}_a - \mathbf{n}_a\right)duds \tag{2}
$$

$$
{}^{G}\mathbf{v}_{k+1} = {}^{G}\mathbf{v}_{k} - {}^{G}\mathbf{g}\Delta T + {}^{k}_{G}\mathbf{R}^{\top}\int_{t_k}^{t_{k+1}}{}^{k}_{u}\Delta\mathbf{R}\left(\mathbf{a}_m - \mathbf{b}_a - \mathbf{n}_a\right)du \tag{3}
$$

$$
\mathbf{b}_{\omega_{k+1}} = \mathbf{b}_{\omega_k} + \int_{t_k}^{t_{k+1}}\mathbf{n}_{\omega b}\ du \tag{4}
$$

$$
\mathbf{b}_{a_{k+1}} = \mathbf{b}_{a_k} + \int_{t_k}^{t_{k+1}}\mathbf{n}_{ab}\ du \tag{5}
$$

From the above, we define the following preintegrated IMU measurements [17]:

$$ {}^{k}\boldsymbol{\alpha}_{k+1} = \int_{t_k}^{t_{k+1}} \int_{t_k}^{s} {}^{k}_{u}\Delta\mathbf{R}\left(\mathbf{a}_m - \mathbf{b}_a - \mathbf{n}_a\right) du ds \tag{6} $$

$$ {}^{k}\boldsymbol{\beta}_{k+1} = \int_{t_k}^{t_{k+1}} {}^{k}_{u}\Delta\mathbf{R}\left(\mathbf{a}_m - \mathbf{b}_a - \mathbf{n}_a\right) du \tag{7} $$

We can then wish to remove the global frame by integrating relative to the first inertial frame. This can be derived for the position as:

$$ {}^{I_0}\mathbf{p}_{I_{k+1}} = {}^{I_0}_{G}\mathbf{R}({}^{G}\mathbf{p}_{I_{k+1}} - {}^{G}\mathbf{p}_{I_0}) \tag{8} $$

$$ = {}^{I_0}_{G}\mathbf{R}\left({}^{G}\mathbf{p}_{I_k} + {}^{G}\mathbf{v}_{I_k}\Delta T - \frac{1}{2}{}^{G}\mathbf{g}\Delta T^2 + {}^{I_k}_{G}\mathbf{R}^{\top k}\boldsymbol{\alpha}_{k+1} - {}^{G}\mathbf{p}_{I_0}\right) \tag{9} $$

$$ = {}^{I_0}_{G}\mathbf{R}({}^{G}\mathbf{p}_{I_k} - {}^{G}\mathbf{p}_{I_0}) + {}^{I_0}_{G}\mathbf{R}{}^{G}\mathbf{v}_{I_k}\Delta T - \frac{1}{2}{}^{I_0}_{G}\mathbf{R}{}^{G}\mathbf{g}\Delta T^2 + {}^{I_0}_{G}\mathbf{R}{}^{I_k}_{G}\mathbf{R}^{\top k}\boldsymbol{\alpha}_{k+1} \tag{10} $$

$$ = {}^{I_0}\mathbf{p}_{I_k} + {}^{I_0}\mathbf{v}_{I_k}\Delta T - \frac{1}{2}{}^{I_0}\mathbf{g}\Delta T^2 + {}^{I_k}_{I_0}\mathbf{R}^{\top k}\boldsymbol{\alpha}_{k+1} \tag{11} $$

We can thus have the following relative preintegration equations:

$$ {}^{I_{k+1}}_{I_0}\mathbf{R} = {}^{I_{k+1}}_{I_k}\Delta\mathbf{R}\ {}^{I_k}_{I_0}\mathbf{R} \tag{12} $$

$$ {}^{I_0}\mathbf{p}_{I_{k+1}} = {}^{I_0}\mathbf{p}_{I_k} + {}^{I_0}\mathbf{v}_{I_k}\Delta T - \frac{1}{2}{}^{I_0}\mathbf{g}\Delta T^2 + {}^{I_k}_{I_0}\mathbf{R}^{\top k}\boldsymbol{\alpha}_{k+1} \tag{13} $$

$$ {}^{I_0}\mathbf{v}_{I_{k+1}} = {}^{I_0}\mathbf{v}_{I_k} - {}^{I_0}\mathbf{g}\Delta T + {}^{I_k}_{I_0}\mathbf{R}^{\top k}\boldsymbol{\beta}_{k+1} \tag{14} $$

We now define the integration from the first $\{I_0\}$ frame:

$$ {}^{I_{k+1}}_{I_0}\mathbf{R} = {}^{I_{k+1}}_{I_0}\Delta\mathbf{R}\ {}^{I_0}_{I_0}\mathbf{R} \tag{15} $$

$$ \triangleq {}^{I_{k+1}}_{I_0}\Delta\mathbf{R} \tag{16} $$

$$ {}^{I_0}\mathbf{p}_{I_{k+1}} = {}^{I_0}\mathbf{p}_{I_0} + {}^{I_0}\mathbf{v}_{I_0}\Delta T - \frac{1}{2}{}^{I_0}\mathbf{g}\Delta T^2 + {}^{I_0}_{I_0}\mathbf{R}^{\top 0}\boldsymbol{\alpha}_{k+1} \tag{17} $$

$$ \triangleq {}^{I_0}\mathbf{v}_{I_0}\Delta T - \frac{1}{2}{}^{I_0}\mathbf{g}\Delta T^2 + {}^{0}\boldsymbol{\alpha}_{k+1} \tag{18} $$

$$ {}^{I_0}\mathbf{v}_{I_{k+1}} = {}^{I_0}\mathbf{v}_{I_0} - {}^{I_0}\mathbf{g}\Delta T + {}^{I_0}_{I_0}\mathbf{R}^{\top 0}\boldsymbol{\beta}_{k+1} \tag{19} $$

$$ \triangleq {}^{I_0}\mathbf{v}_{I_0} - {}^{I_0}\mathbf{g}\Delta T + {}^{0}\boldsymbol{\beta}_{k+1} \tag{20} $$

Note that the time offset ΔT is now from time t_0 to t_{k+1}.

3.2 Feature Bearing Observations

Our camera observes environmental features as it moves along its trajectory. For a feature we consider the following relation to our state:

$$ \mathbf{z}_{n,k} = \begin{bmatrix} u_n \\ v_n \end{bmatrix} + \mathbf{n}_{n,k} \tag{21} $$

$$ = \mathbf{h}\left({}^{I_k}_{I_0}\bar{q}, {}^{I_0}\mathbf{p}_{I_k}, {}^{I_0}\mathbf{p}_f\right) + \mathbf{n}_{n,k} \tag{22} $$

$$ = \boldsymbol{\Lambda}({}^{C_k}\mathbf{p}_f) + \mathbf{n}_{n,k} \tag{23} $$

$$ {}^{C_k}\mathbf{p}_f = {}^{C}_{I}\mathbf{R}\,{}^{I_k}_{I_0}\mathbf{R}({}^{I_0}\mathbf{p}_f - {}^{I_0}\mathbf{p}_{I_k}) + {}^{C}\mathbf{p}_I \tag{24} $$

where $\boldsymbol{\Lambda}([x\ y\ z]^\top) = [x/z\ y/z]^\top$, and $\mathbf{z}_{n,k}$ is the normalized feature bearing. Here we can assume we know the camera distortion parameters to recover the normalized pixel coordinates, $\mathbf{z}_{n,k}$, and that the extrinsic transform between the camera and IMU, $\{{}^{C}_{I}\mathbf{R}, {}^{C}\mathbf{p}_I\}$, is known with reasonable accuracy. The extrinsic calibration can be recovered, see [2, 3], but under short initialization periods they are likely not fully recoverable and thus we assume we have a "good enough" guess. It is also important to note that the feature, ${}^{I_0}\mathbf{p}_f$, is represented in the $\{I_0\}$ frame.

We can now define the following linear measurement observation which removes the need for the division in $\boldsymbol{\Lambda}(\cdot)$. As presented in [2, 3], we consider the following:

$$ \begin{bmatrix} 1 & 0 & -u_n \\ 0 & 1 & -v_n \end{bmatrix} {}^{C_k}\mathbf{p}_f = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{25} $$

One can check that the left side of the above equation when multiplied out, should equate to zero. This shows that the difference between the normalized feature observation and the projected feature should be zero.

3.3 Discussion on Frame Transformation

Here we note the impact of transforming into the first inertial frame. Typically gravity in the global frame is known, e.g, ${}^{G}\mathbf{g} = [0\ 0\ 9.81]^\top$, and thus we would need to recover this non-linear rotation ${}^{I_0}_{G}\mathbf{R}$ to relate it to our inertial states. If we instead preintegrate in the local $\{I_0\}$ frame, then are required to recover the gravity in the local frame ${}^{I_0}\mathbf{g}$ which has unknown direction, but known magnitude. The orientation change is known through the process of the preintegration, i.e., ${}^{I_0}_{I_0}\mathbf{R} = \mathbf{I}$ and ${}^{I_{k+1}}_{I_0}\mathbf{R}$ can be found through gyroscope integration, we now have a *linear* problem in respect to ${}^{I_0}\mathbf{g}$.

3.4 The Linear Ax = b Problem

From a high level, we wish the recover a linear system of unknown variables from our measurements. We can combine our inertial integration from Section 3.1 and the feature observation at time t_k from Section 3.2 to get the following:

$$ \underbrace{\begin{bmatrix} 1 & 0 & -u_n \\ 0 & 1 & -v_n \end{bmatrix}}_{\boldsymbol{\Gamma}} {}^{C_k}\mathbf{p}_f = \underbrace{\begin{bmatrix} 0 \\ 0 \end{bmatrix}}_{\mathbf{0}_2} \tag{26} $$

$$ \boldsymbol{\Gamma}\left({}^{C}_{I}\mathbf{R}\,{}^{I_k}_{I_0}\mathbf{R}\left({}^{I_0}\mathbf{p}_f - {}^{I_0}\mathbf{p}_{I_k}\right) + {}^{C}\mathbf{p}_I\right) = \mathbf{0}_2 \quad \text{via Eq. (24)} \tag{27} $$

$$ \boldsymbol{\Gamma}\left({}^{C}_{I}\mathbf{R}\,{}^{I_k}_{I_0}\Delta\mathbf{R}\left({}^{I_0}\mathbf{p}_f - {}^{I_0}\mathbf{v}_{I_0}\Delta T - \frac{1}{2}{}^{I_0}\mathbf{g}\Delta T^2 + {}^{0}\boldsymbol{\alpha}_k\right) + {}^{C}\mathbf{p}_I\right) = \mathbf{0}_2 \quad \text{via Eq. (18)} \tag{28} $$

$$ \underbrace{\boldsymbol{\Gamma}\,{}^{C}_{I}\mathbf{R}\,{}^{I_k}_{I_0}\Delta\mathbf{R}}_{\boldsymbol{\Upsilon}}\left({}^{I_0}\mathbf{p}_f - {}^{I_0}\mathbf{v}_{I_0}\Delta T - \frac{1}{2}{}^{I_0}\mathbf{g}\Delta T^2\right) = -\left(\boldsymbol{\Gamma}\,{}^{C}\mathbf{p}_I + \boldsymbol{\Gamma}\,{}^{C}_{I}\mathbf{R}\,{}^{I_k}_{I_0}\Delta\mathbf{R}\,{}^{0}\boldsymbol{\alpha}_k\right) \tag{29} $$

If we stack multiple observations of the *same* feature, we can construct the following linear system.

$$\begin{bmatrix} \vdots & \vdots & \vdots \\ -\boldsymbol{\Upsilon}\Delta T & \boldsymbol{\Upsilon} & -\boldsymbol{\Upsilon}\Delta T^2 \\ \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} {}^{I_0}\mathbf{v}_{I_0} \\ {}^{I_0}\mathbf{p}_f \\ {}^{I_0}\mathbf{g} \end{bmatrix} = \begin{bmatrix} \vdots \\ -\left(\boldsymbol{\Gamma}\,{}^{C}\mathbf{p}_I + \boldsymbol{\Upsilon}^0\boldsymbol{\alpha}_k\right) \\ \vdots \end{bmatrix} \tag{30}$$

If we have more than five frames, we can recover the above linear system (i.e., $2N \geq 9$ is satisfied for $N = 5$). If gravity is constrained to have a known magnitude, then this reduces the degrees-of-freedom, thus we can recover the state with only four observations (i.e., $2N \geq 8$ is satisfied for $N = 4$). This is different from the results in [2, 3] since we do not try to recover the extrinsic camera to IMU transform. This above system can then be stacked for *all* feature observations, increasing its robustness at the same time.

3.5 Quadratic Constrained Least-squares

We now look to leverage our prior knowledge of the magnitude of gravity. We can first define a constrained linear least-squares optimization problem as:

$$\text{minimize } \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2 = \left\| \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ {}^{I_0}\mathbf{g} \end{bmatrix} - \mathbf{b} \right\|_2 \tag{31}$$

$$\text{subject to } \left\| {}^{I_0}\mathbf{g} \right\|_2 = g \tag{32}$$

where the magnitude of gravity, ${}^{I_0}\mathbf{g}$, is known to be g, and the state $\mathbf{x}_1$ contains the velocity and all feature positions. We can then define the following Lagrange multiplier [2, 3]:

$$\min\, L(\mathbf{x}, \lambda) = \left\| \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ {}^{I_0}\mathbf{g} \end{bmatrix} - \mathbf{b} \right\|_2^2 - \lambda({}^{I_0}\mathbf{g}^\top {}^{I_0}\mathbf{g} - g^2) \tag{33}$$

which is equivalent to the optimization problem as follows:

$$\text{minimize } {}^{I_0}\mathbf{g}^\top \mathbf{D}\, {}^{I_0}\mathbf{g} - 2\mathbf{d}^\top {}^{I_0}\mathbf{g} \tag{34}$$

$$\text{subject to } {}^{I_0}\mathbf{g}^\top {}^{I_0}\mathbf{g} = g^2 \tag{35}$$

with the following matrices:

$$\mathbf{D} = \mathbf{A}_2^\top (\mathbf{I} - \mathbf{A}_1(\mathbf{A}_1^\top \mathbf{A}_1)^{-1}\mathbf{A}_1^\top)\mathbf{A}_2 \tag{36}$$

$$\mathbf{d} = \mathbf{A}_2^\top (\mathbf{I} - \mathbf{A}_1(\mathbf{A}_1^\top \mathbf{A}_1)^{-1}\mathbf{A}_1^\top)\mathbf{b} \tag{37}$$

The optimal value of ${}^{I_0}\mathbf{g}$ has the same solution as the following [18, 19]:

$$\text{minimize } \lambda \tag{38}$$

$$\text{subject to } \mathbf{D}\, {}^{I_0}\mathbf{g} = \lambda\, {}^{I_0}\mathbf{g} + \mathbf{d} \tag{39}$$

$${}^{I_0}\mathbf{g}^\top {}^{I_0}\mathbf{g} = g^2 \tag{40}$$

The solution for $\{\lambda, {}^{I_0}\mathbf{g}\}$ is one which satisfies the following (given vector $\mathbf{e} \neq \mathbf{0}$) [18, Thm. 5.1]:

$$\left(\mathbf{D} - \lambda \mathbf{I}_2\right)^2 \mathbf{e} = \frac{1}{g^2}\mathbf{d}\mathbf{d}^\top \mathbf{e} \tag{41}$$

The above equation describes a quadratic eigenvalue problem, which has the solution as:

Listing 1 Matlab code used to symbolically derive solution to quadratic problem.

```
% symbolic variables
D = sym('D',[3,3],'real');
d = sym('d',[3,1],'real');
lambda = sym('lambda','real');
g = sym('g','positive');

% evaluate determinate
expression = det((D - lambda*eye(3,3))^2 - (1/g^2)*(d*d'));
collected = collect(expression, lambda)
```

$$\det\left(\left(\mathbf{D}-\lambda\mathbf{I}_2\right)^2-\frac{1}{g^2}\mathbf{d}\mathbf{d}^\top\right)=0 \tag{42}$$

This is a six-order polynomial in terms of λ and can be found via Listing 1. The roots (and thus solution) to this polynomial can be found through eigen decomposition of its polynomial Companion matrix. The smallest λ solution which satisfies the constraint exactly will give the optimal. We can recover the state as follows:

$$\mathbf{x}=\begin{bmatrix}\mathbf{x}_1\\ {}^{I_0}\mathbf{g}\end{bmatrix}=\begin{bmatrix}-(\mathbf{A}_1^\top\mathbf{A}_1)^{-1}\mathbf{A}_1^\top\mathbf{A}_2(\mathbf{D}-\lambda\mathbf{I})^{-1}\mathbf{d}+(\mathbf{A}_1^\top\mathbf{A}_1)^{-1}\mathbf{A}_1^\top\mathbf{b}\\ (\mathbf{D}-\lambda\mathbf{I})^{-1}\mathbf{d}\end{bmatrix} \tag{43}$$

3.6 Recovering Inertial States

The complete state of the linear system with N environmental features is as follows:

$$\mathbf{x}=\begin{bmatrix}{}^{I_0}\mathbf{v}_{I_0}^\top & {}^{I_0}\mathbf{p}_{f,1}^\top & \cdots & {}^{I_0}\mathbf{p}_{f,N}^\top & {}^{I_0}\mathbf{g}^\top\end{bmatrix}^\top \tag{44}$$

We now wish to recover the inertial state at each camera timestep which will be used as the initial guess for our non-linear optimization in the following section. Using our inertial propagation we can recover the inertial state at each time as:

$$\begin{bmatrix}{}^{I_k}_{I_0}\mathbf{R}\\ {}^{I_0}\mathbf{p}_{I_k}\\ {}^{I_0}\mathbf{v}_{I_k}\end{bmatrix}=\begin{bmatrix}{}^{I_{k+1}}_{I_0}\Delta\mathbf{R}\\ {}^{I_0}\mathbf{v}_{I_0}\Delta T-\frac{1}{2}{}^{I_0}\mathbf{g}\Delta T^2+{}^0\boldsymbol{\alpha}_{k+1}\\ {}^{I_0}\mathbf{v}_{I_0}-{}^{I_0}\mathbf{g}\Delta T+{}^0\boldsymbol{\beta}_{k+1}\end{bmatrix} \tag{45}$$

We have highlighted in red the components which we recovered from our constrained least-squares and ΔT is from time t_0 to t_k. Now we wish to align the first frame of reference with that of gravity. This means that we will compute a frame $\{G\}$ such that gravity is ${}^G\mathbf{g}=[0\ 0\ 9.81]^\top$ with its position and yaw being at the $\{I_0\}$ origin. We can find the rotation with its z-axis along the gravity direction as:

$$\mathbf{r}_z=\text{normalize}({}^{I_0}\mathbf{g}) \tag{46}$$

$$\mathbf{r}_x = \text{normalize}(\mathbf{e}_1 - \mathbf{r}_z \mathbf{r}_z^\top \mathbf{e}_1) \tag{47}$$

$$\mathbf{r}_y = \text{normalize}(\lfloor \mathbf{r}_z \rfloor \mathbf{r}_x) \tag{48}$$

$${}^{I_0}_{G}\mathbf{R} = \begin{bmatrix} \mathbf{r}_x & \mathbf{r}_y & \mathbf{r}_z \end{bmatrix} \tag{49}$$

where we note that Eq. (47) is the Gram-Schmidt process and $\mathbf{r}_y$ is found through the cross-product of the first two axes. We can the recover the following states in the global:

$$\begin{bmatrix} {}^{I_k}_{G}\mathbf{R} \\ {}^{G}\mathbf{p}_{I_k} \\ {}^{G}\mathbf{v}_{I_k} \end{bmatrix} = \begin{bmatrix} {}^{I_k}_{I_0}\mathbf{R} \, {}^{I_0}_{G}\mathbf{R} \\ {}^{I_0}_{G}\mathbf{R}^\top {}^{I_0}\mathbf{p}_{I_k} \\ {}^{I_0}_{G}\mathbf{R}^\top {}^{I_0}\mathbf{v}_{I_k} \end{bmatrix} \tag{50}$$

$${}^{G}\mathbf{p}_f = {}^{I_0}_{G}\mathbf{R}^\top {}^{I_0}\mathbf{p}_f \tag{51}$$

Note again that the frame $\{G\}$ and $\{I_0\}$ have the same origin thus only a rotation of position is required.

4 Maximum Likelihood Estimation

Using the linear system we have recovered an initial guess of the states. From here, we wish to refine the estimate such that they are closer to their true and also recover the covariance of the initial state. This process will also take into account the noise properties of the system such that each measurement is weighted relative to the accuracy of its sensor. We can define the following state which we wish to optimize:

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_{I,1}^\top & \cdots & \mathbf{x}_{I,C}^\top & {}^{G}\mathbf{p}_{f,1}^\top & \cdots & {}^{G}\mathbf{p}_{f,N}^\top \end{bmatrix}^\top \tag{52}$$

$$\mathbf{x}_{I,k} = \begin{bmatrix} {}^{I_k}_{G}\bar{q}^\top & {}^{G}\mathbf{p}_{I_k}^\top & {}^{G}\mathbf{v}_{I_k}^\top & \mathbf{b}_{g,k}^\top & \mathbf{b}_{a,k}^\top \end{bmatrix}^\top \tag{53}$$

where we have C inertial states at each imaging timestep, and N environmental features. We can define the following optimization problem with inertial $\mathbb{C}_I$, camera $\mathbb{C}_C$, and prior factors $\mathbb{C}_P$:

$$\min_{\mathbf{x}} \sum \mathbb{C}_I + \sum \mathbb{C}_C + \sum \mathbb{C}_P \tag{54}$$

This is optimized using ceres-solver [20].

4.1 Inertial Measurement Model

Inertial readings are related to two bounding states through continuous preintegration [17] (the IMU frame $\{I_k\}$ has been shorten to $\{k\}$):

$${}^{G}\mathbf{p}_{k+1} = {}^{G}\mathbf{p}_k + {}^{G}\mathbf{v}_k \Delta T - \frac{1}{2} {}^{G}\mathbf{g} \Delta T^2 + {}^{k}_{G}\mathbf{R}^\top {}^{k}\boldsymbol{\alpha}_{k+1} \tag{55}$$

$${}^{G}\mathbf{v}_{k+1} = {}^{G}\mathbf{v}_k - {}^{G}\mathbf{g} \Delta T + {}^{k}_{G}\mathbf{R}^\top {}^{k}\boldsymbol{\beta}_{k+1} \tag{56}$$

$${}^{k+1}_{G}\bar{q} = {}^{k+1}_{k}\bar{q} \otimes {}^{k}_{G}\bar{q} \tag{57}$$

where ΔT is the difference between the bounding pose timestamps (t_k, t_{k+1}) and ${}^{k}\boldsymbol{\alpha}_{k+1}, {}^{k}\boldsymbol{\beta}_{k+1}$ are defined by the following integrations of the IMU measurements:

$${}^{k}\boldsymbol{\alpha}_{k+1} = \int_{t_k}^{t_{k+1}} \int_{t_k}^{s} {}^{k}_{u}\mathbf{R} \left(\mathbf{a}_m - \mathbf{b}_a - \mathbf{n}_a \right) duds \tag{58}$$

$$ {}^{k}\boldsymbol{\beta}_{k+1} = \int_{t_k}^{t_{k+1}} {}^{k}_{u}\mathbf{R}\left(\mathbf{a}_m - \mathbf{b}_a - \mathbf{n}_a\right) du \tag{59} $$

We note that the preintegrated measurements, ${}^{k}\boldsymbol{\alpha}_{k+1}$, ${}^{k}\boldsymbol{\beta}_{k+1}$, ${}^{k+1}_{k}\bar{q}$ are dependent on the *true* biases. This dependency is addressed through a first order Taylor series expansion about the current bias estimates $\bar{\mathbf{b}}_w$ and $\bar{\mathbf{b}}_a$ at time t_k (assumed to be known before linear initialization):

$$ {}^{k}\boldsymbol{\alpha}_{k+1} \simeq {}^{k}\breve{\boldsymbol{\alpha}}_{k+1} + \frac{\partial \boldsymbol{\alpha}}{\partial \mathbf{b}_a}\Big|_{\bar{\mathbf{b}}_a} \Delta\mathbf{b}_a + \frac{\partial \boldsymbol{\alpha}}{\partial \mathbf{b}_w}\Big|_{\bar{\mathbf{b}}_w} \Delta\mathbf{b}_w \tag{60} $$

$$ {}^{k}\boldsymbol{\beta}_{k+1} \simeq {}^{k}\breve{\boldsymbol{\beta}}_{k+1} + \frac{\partial \boldsymbol{\beta}}{\partial \mathbf{b}_a}\Big|_{\bar{\mathbf{b}}_a} \Delta\mathbf{b}_a + \frac{\partial \boldsymbol{\beta}}{\partial \mathbf{b}_w}\Big|_{\bar{\mathbf{b}}_w} \Delta\mathbf{b}_w \tag{61} $$

$$ {}^{k+1}_{k}\bar{q} \simeq \bar{q}(\Delta\mathbf{b}_w)^{-1} \otimes {}^{k+1}_{k}\breve{q} \tag{62} $$

where ${}^{k}\breve{\boldsymbol{\alpha}}_{k+1}$, ${}^{k}\breve{\boldsymbol{\beta}}_{k+1}$, ${}^{k+1}_{k}\breve{q}$ are the preintegrated measurements evaluated at the current bias estimates. In particular, ${}^{k+1}_{k}\breve{q}$ can be found using the zeroth order quaternion integrator [21]. The quaternion which models multiplicative orientation corrections due to linearized bias change is:

$$ \bar{q}(\Delta\mathbf{b}_w) = \begin{bmatrix} \frac{\boldsymbol{\theta}}{||\boldsymbol{\theta}||}\sin\frac{||\boldsymbol{\theta}||}{2} \\ \cos\frac{||\boldsymbol{\theta}||}{2} \end{bmatrix} \tag{63} $$

$$ \boldsymbol{\theta} = \frac{\partial \bar{q}}{\partial \mathbf{b}_w}\Big|_{\bar{\mathbf{b}}_w} \left(\mathbf{b}_{w(k)} - \bar{\mathbf{b}}_w\right) \tag{64} $$

where $\Delta\mathbf{b}_w := \mathbf{b}_{w(k)} - \bar{\mathbf{b}}_w$ and $\Delta\mathbf{b}_a := \mathbf{b}_{a(k)} - \bar{\mathbf{b}}_a$ are the differences between the true biases and the current bias estimate used as the linearization point. The new preintegration measurements can now be computed *once* and changes in the bias estimates can be taken into account through the above Taylor series. The final measurement residual is as follows:

$$ r_I(\mathbf{x}) = \begin{bmatrix} 2\text{vec}\left({}^{k+1}_{V}\bar{q} \otimes {}^{k}_{V}\bar{q}^{-1} \otimes {}^{k+1}_{k}\breve{q}^{-1} \otimes \bar{q}(\Delta\mathbf{b}_w)\right) \\ \mathbf{b}_{w,k+1} - \mathbf{b}_{w,k} \\ \begin{pmatrix} {}^{k}_{V}\mathbf{R}\left({}^{V}\mathbf{v}_{k+1} - {}^{V}\mathbf{v}_k + {}^{V}_{G}\mathbf{R}\,{}^{G}\mathbf{g}\Delta T\right) \\ -{}^{k}\breve{\boldsymbol{\beta}}_{k+1} - \frac{\partial \boldsymbol{\beta}}{\partial \mathbf{b}_a}\Big|_{\bar{\mathbf{b}}_a}\Delta\mathbf{b}_a - \frac{\partial \boldsymbol{\beta}}{\partial \mathbf{b}_w}\Big|_{\bar{\mathbf{b}}_w}\Delta\mathbf{b}_w \end{pmatrix} \\ \mathbf{b}_{a,k+1} - \mathbf{b}_{a,k} \\ \begin{pmatrix} {}^{k}_{V}\mathbf{R}\left({}^{V}\mathbf{p}_{k+1} - {}^{V}\mathbf{p}_k - {}^{V}\mathbf{v}_k\Delta T + \frac{1}{2}{}^{V}_{G}\mathbf{R}\,{}^{G}\mathbf{g}\Delta T^2\right) \\ -{}^{k}\breve{\boldsymbol{\alpha}}_{k+1} - \frac{\partial \boldsymbol{\alpha}}{\partial \mathbf{b}_a}\Big|_{\bar{\mathbf{b}}_a}\Delta\mathbf{b}_a - \frac{\partial \boldsymbol{\alpha}}{\partial \mathbf{b}_w}\Big|_{\bar{\mathbf{b}}_w}\Delta\mathbf{b}_w \end{pmatrix} \end{bmatrix} $$

where $\text{vec}(\cdot)$ returns the vector portion of the quaternion (i.e., the top three elements) and the bias errors are the difference between biases in the bounding states.

We use combined continuous preintegration factors that included both the inertial and bias errors together and relate to the full 15 degree-of-freedom state (see Equation 53). This combined continuous preintegration factor better models the measurement error state dynamics due to bias drift over the integration interval. Thus we have the following:

$$ \mathbb{C}_I \triangleq \|\mathbf{0} - r_I(\mathbf{x})\|^2_{\mathbf{P}^{-1}} \tag{65} $$

where $\mathbf{P}$ is the measurement covariance and we refer the reader to the original paper for it and the state Jacobians [17, 22].

4.2 Camera Measurement Model

We follow the camera model used in OpenVINS [4]. Each observation of a feature can be written as a function of the state by:

$$\begin{aligned}
\mathbf{z}_{c,k} &= h_c(\mathbf{x}) + \mathbf{n}_{c,k} && (66)\\
&= h_d(\mathbf{z}_{n,k},\ \boldsymbol{\zeta}) + \mathbf{n}_{c,k} && (67)\\
&= h_d(h_p({}^{C_k}\mathbf{p}_f),\ \boldsymbol{\zeta}) + \mathbf{n}_{c,k} && (68)\\
&= h_d(h_p(h_t({}^{G}\mathbf{p}_f,\ {}^{C_k}_{G}\mathbf{R},\ {}^{G}\mathbf{p}_{C_k})),\ \boldsymbol{\zeta}) + \mathbf{n}_{c,k} && (69)
\end{aligned}$$

where $\mathbf{z}_{c,k}$ is the raw uv pixel coordinate; $\mathbf{n}_{c,k}$ the raw pixel noise and typically assumed to be zero-mean white Gaussian $\mathbf{R}_c$; $\mathbf{z}_{n,k}$ is the normalized undistorted uv measurement; ${}^{C_k}\mathbf{p}_f$ is the landmark position in the current camera frame; ${}^{G}\mathbf{p}_f$ is the landmark position in the global frame and depending on its representation may also be a function of state elements; and $\{{}^{C_k}_{G}\mathbf{R}, {}^{G}\mathbf{p}_{C_k}\}$ denotes the current camera pose (position and orientation) in the global frame which are related to $\{{}^{I_k}_{G}\mathbf{R}, {}^{G}\mathbf{p}_{I_k}\}$ through the extrinsic calibration.

The measurement functions h_d, h_p, and h_t correspond to the intrinsic distortion, projection, and transformation functions and the corresponding measurement Jacobians can be computed through a simple chain rule. We can then define the following cost for feature observations:

$$\mathbb{C}_C \triangleq \|\mathbf{z}_c - h_c(\mathbf{x})\|^2_{\mathbf{R}_c^{-1}} \tag{70}$$

4.3 Prior Cost

It is important that we provide priors to the states which are unobservable in nature. For a visual-inertial system the global yaw and position are unobservable [1]. Additionally, under short windows with limited rotation, the gyroscope and especially accelerometer biases can nearly be unobservable.

We can define a prior as computing the difference between a linearization point $\mathbf{x}_{lin}$ and the current estimate of the state $\mathbf{x}$. This can have some uncertainty, which for the the unobservable directions, we pick to be very small (on order $1e-5$) such that they are well constrained in the problem. This prior also enables the inversion of the information matrix to recover the covariance, which is a crucial last step in the optimization process.

$$\mathbb{C}_C \triangleq \|\mathbf{x} \boxminus \mathbf{x}_{lin}\|^2_{\boldsymbol{\Lambda}} \tag{71}$$

where $\boxminus$ is subtraction for vectors, and ${}^{G}\delta\boldsymbol{\theta} = \log({}^{I}_{G}\mathbf{R}^{\top}\,{}^{I}_{G}\mathbf{R}_{lin})$ for the orientation.

References

[1] J.A. Hesch, D.G. Kottas, S.L. Bowman, and S.I. Roumeliotis. "Consistency Analysis and Improvement of Vision-aided Inertial Navigation". In: 30.1 (2013), pp. 158–176.

[2] Tue-Cuong Dong-Si and Anastasios I Mourikis. "Estimator initialization in vision-aided inertial navigation with unknown camera-IMU calibration". In: *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE. 2012, pp. 1064–1071.

[3] Tue-Cuong Dong-Si and Anastasios I Mourikis. *Closed-form Solutions for Vision-aided Inertial Navigation*. Tech. rep. Dept. of Electrical Engineering, University of California, Riverside, 2011. URL: http://tdongsi.github.io/download/pubs/2011_VIO_Init_TR.pdf.

[4] Patrick Geneva, Kevin Eckenhoff, Woosik Lee, Yulin Yang, and Guoquan Huang. "OpenVINS: A Research Platform for Visual-Inertial Estimation". In: *Proc. of the IEEE International Conference on Robotics and Automation*. Paris, France, 2020. URL: https://github.com/rpng/open_vins.

[5] Agostino Martinelli. "Closed-form solution of visual-inertial structure from motion". In: *International journal of computer vision* 106.2 (2014), pp. 138–152.

[6] Jacques Kaiser, Agostino Martinelli, Flavio Fontana, and Davide Scaramuzza. "Simultaneous state initialization and gyroscope bias calibration in visual inertial aided navigation". In: *IEEE Robotics and Automation Letters* 2.1 (2016), pp. 18–25.

[7] Zhenfei Yang and Shaojie Shen. "Monocular visual–inertial state estimation with online initialization and camera–IMU extrinsic calibration". In: *IEEE Transactions on Automation Science and Engineering* 14.1 (2016), pp. 39–51.

[8] Tong Qin and Shaojie Shen. "Robust initialization of monocular visual-inertial estimation on aerial robots". In: *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. 2017, pp. 4225–4232.

[9] Tong Qin, Peiliang Li, and Shaojie Shen. "VINS-Mono: A robust and versatile monocular visual-inertial state estimator". In: *IEEE Transactions on Robotics* 34.4 (2018), pp. 1004–1020.

[10] Raúl Mur-Artal and Juan D Tardós. "Visual-inertial monocular SLAM with map reuse". In: *IEEE Robotics and Automation Letters* 2.2 (2017), pp. 796–803.

[11] Carlos Campos, José MM Montiel, and Juan D Tardós. "Fast and robust initialization for visual-inertial SLAM". In: *2019 International Conference on Robotics and Automation (ICRA)*. IEEE. 2019, pp. 1288–1294.

[12] Carlos Campos, José MM Montiel, and Juan D Tardós. "Inertial-only optimization for visual-inertial initialization". In: *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE. 2020, pp. 51–57.

[13] Georgios Evangelidis and Branislav Micusik. "Revisiting visual-inertial structure-from-motion for odometry and SLAM initialization". In: *IEEE Robotics and Automation Letters* 6.2 (2021), pp. 1415–1422.

[14] David Zuñiga-Noël, Francisco-Angel Moreno, and Javier Gonzalez-Jimenez. "An Analytical Solution to the IMU Initialization Problem for Visual-Inertial Systems". In: *IEEE Robotics and Automation Letters* 6.3 (2021), pp. 6116–6122.

[15] Martin Scheiber, Jeff Delaune, Stephan Weiss, and Roland Brockers. "Mid-Air Range-Visual-Inertial Estimator Initialization for Micro Air Vehicles". In: *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE. 2021, pp. 7613–7619.

[16] Jeff Delaune, David S Bayard, and Roland Brockers. "Range-visual-inertial odometry: Scale observability without excitation". In: *IEEE Robotics and Automation Letters* 6.2 (2021), pp. 2421–2428.

[17] Kevin Eckenhoff, Patrick Geneva, and Guoquan Huang. "Closed-form Preintegration Methods for Graph-based Visual-Inertial Navigation". In: *International Journal of Robotics Research* 38.5 (2019), pp. 563–586.

[18] Walter Gander, Gene H Golub, and Urs Von Matt. "A constrained eigenvalue problem". In: *Linear Algebra and its applications* 114 (1989), pp. 815–839.

[19] Emil Spjøtvoll. "A Note on a Theorem of Forsythe and Golub". In: *SIAM Journal on Applied Mathematics* 23.3 (1972), pp. 307–311.

[20] Sameer Agarwal, Keir Mierle, and Others. *Ceres Solver*. http://ceres-solver.org.

[21] Nikolas Trawny and Stergios I. Roumeliotis. *Indirect Kalman Filter for 3D Attitude Estimation*. Tech. rep. University of Minnesota, Dept. of Comp. Sci. & Eng., Mar. 2005.

[22] Kevin Eckenhoff, Patrick Geneva, and Guoquan Huang. *Continuous Preintegration Theory for Visual-Inertial Navigation*. Tech. rep. RPNG-2018-CPI. Available: http://udel.edu/~ghuang/papers/tr_cpi.pdf. University of Delaware, 2018.