Consistent Right-Invariant Fixed-Lag Smoother with Application to Visual Inertial SLAM

Jianzhu Huai [1], Yukai Lin [2], Yuan Zhuang [*, 1], Min Shi[3]
[1]Wuhan University
[2]ETH Zurich
[3]Washington University, St. Louis
{jianzhu.huai,yuan.zhuang}@whu.edu.cn, linyuk@ethz.ch, mshi2018@fau.edu

Abstract

State estimation problems without absolute position measurements routinely arise in navigation of unmanned aerial vehicles, autonomous ground vehicles, *etc*. whose proper operation relies on accurate state estimates and reliable covariances. Unaware of absolute positions, these problems have immanent unobservable directions. Traditional causal estimators, however, usually gain spurious information on the unobservable directions, leading to over-confident covariance inconsistent with actual estimator errors. The consistency problem of fixed-lag smoothers (FLSs) has only been attacked by the first estimate Jacobian (FEJ) technique because of the complexity to analyze their observability property. But the FEJ has several drawbacks hampering its wide adoption. To ensure the consistency of a FLS, this paper introduces the right invariant error formulation into the FLS framework. To our knowledge, we are the first to analyze the observability of a FLS with the right invariant error. Our main contributions are twofold. As the first novelty, to bypass the complexity of analysis with the classic observability matrix, we show that observability analysis of FLSs can be done equivalently on the linearized system. Second, we prove that the inconsistency issue in the traditional FLS can be elegantly solved by the right invariant error formulation without artificially correcting Jacobians. By applying the proposed FLS to the monocular visual inertial simultaneous localization and mapping (SLAM) problem, we confirm that the method consistently estimates covariance similarly to a batch smoother in simulation and that our method achieved comparable accuracy as traditional FLSs on real data.

Introduction

Positioning and navigation of a variety of vehicles, *e.g*., unmanned aerial vehicles (UAVs), autonomous ground vehicles (AGVs), depends on real-time state estimation. Accurate system state and reasonable covariance output by state estimators in real time are necessary for the proper operation of these systems. For state estimation, these systems usually fuse measurements captured by sensors that do not provide absolute positions, like cameras, lidars, inertial measurement units (IMUs), *etc*. It is well known that estimators which fuse such measurements have unobservable directions (Jones and Soatto 2011).

As reported in the literature, traditional real-time estimators, *e.g*., filters, fixed-lag smoothers (FLSs), tend to gain fictitious information on unobservable directions (Huang, Mourikis, and Roumeliotis 2010; Dong-Si and Mourikis 2011), and to output falsely optimistic covariance inconsistent to the actual state error. This inconsistency is caused by the marginalization step of real-time estimators which removes old state variables and measurements (*i.e*., factors) from an estimator and approximates those measurements by a linear prior factor. A deeper cause is that for a variable in the prior factor, its linearization point used by the prior factor differs from that used by the remaining factors. Obviously, the batch estimator and its incremental variants, *e.g*., iSAM2 (Kaess et al. 2012), do not have this issue as they do not marginalize variables.

To fix the estimator inconsistency, techniques that modify the measurement Jacobians to fit certain criteria have been proposed. For instance, the "first estimate Jacobian (FEJ)" technique (Huang, Mourikis, and Roumeliotis 2010) evaluates Jacobians relative to variables in the linear prior factor at their estimates upon marginalization. Because the Jacobian computation depends on specifics, such as an earlier estimate of a variable, it is usually difficult to apply such techniques to an existing estimator framework. A new trend is to use right invariant error formulation (Barrau and Bonnabel 2016a) where a navigation state variable (consisting of orientation, position, and velocity) is associated to a Lie group $SE_2(3)$ and the error vector is invariant to transforming the trajectory by a right multiplication. Besides mathematically elegant, it is easy to implement as it fits the conventional filtering framework. However, this formulation has not been used in FLSs, mainly because of the challenge to analyze their consistency property.

Previous work has shown that the estimator inconsistency comes along with the observability issue where the unobservable directions become spuriously observable (Hesch et al. 2014a). Thus, consistency has been predominantly studied by examining rank deficiency of the linearized observability matrix, *e.g*., (Huang, Mourikis, and Roumeliotis 2010; Dong-Si and Mourikis 2012; Brossard, Barrau, and Bonnabel 2018). The local observability matrix is acceptable in complexity for analyzing filters, but becomes very involved for dealing with FLSs, *e.g*., (Dong-Si and Mourikis 2012). Because the observability matrix is a derivative of the

*Corresponding author, yuan.zhuang@whu.edu.cn

linearized original system, we think that directly working with the linearized system can greatly simplify the observability analysis.

Based on this analysis, we prove that the right invariant error formulation leads to a consistent FLS. The claims made in the proof are validated with simulation. Furthermore, the practicality of the proposed right invariant FLS is verified with the EuRoC benchmark (Burri et al. 2016).

In summary, our contributions include

- To avoid the complexity of observability matrices, we prove that observability analysis of FLSs can be done equivalently on the linearized system.
- To clarify effects of variables on observability, we show that using different linearization points for a state variable expressed in a local coordinate frame and for a sensor parameter do not impact unobservable directions and hence consistency.
- To our knowledge, we are the first to prove and validate that FLS with the right invariant error formulation maintains consistent covariance without artificially modifying Jacobians.

The following text presents the formulation and observability analysis of the FLS to solve the visual inertial SLAM problem and the application of right invariant errors in the FLS. Then, results of simulation and real data tests are supplied. Lastly, we draw conclusions and indicate future work.

Related Work

There are several approaches to ensure consistency of traditional real-time estimators. Most of them are designed for Extended Kalman Filters (EKFs) and few are proposed for optimization-based approaches, *i.e.*, FLSs. Costante and Mancini 2020 developed a deep neural network to output state estimates and uncertainty measures, but their consistency is very challenging to analyze. The optimization-based iSAM2 (Kaess et al. 2012) method updates only affected variables as new observations arrive, keeping constant computation cost. As it keeps the entire history of variables and observations for inference, its consistency naturally follows. But it will drain the memory in a long-term operation. For EKFs, the consistency remedies include robocentric coordination (Castellanos et al. 2007), FEJ (Li and Mourikis 2013), observability constraints (Hesch et al. 2014b), and the recently developed right invariant error formulation (Barrau and Bonnabel 2016a; Zhang et al. 2017; Heo and Park 2018). The invariant error formulation for filters defines the error state in an extended Lie group such that the error state is independent of the state variable's linearization point. As a result, the inconsistency caused by using different linearization points for the same state variable is prevented.

For FLSs, to our knowledge, their consistency has only been improved with the FEJ technique (Dong-Si and Mourikis 2011). But the FEJ technique for FLSs has several downsides. The obvious one is that Jacobian matrices required by the estimators are evaluated at less accurate earlier estimates of state variables which may adversely affect state estimation accuracy. Second, it is often confusing to tell which state variable should lock its linearization point, and which Jacobian should be computed with these preset linearization points. For instance, Li and Mourikis 2013 locked linearization points for only position and velocity, and Usenko et al. 2020 locked linearization points for position, velocity, and biases once they are in the prior factor. Third, assigning and tracking linearization points requested by FEJ is often impossible for generic nonlinear solvers without hacking. For example, a solver may encapsulate state variables such that they are not tampered by external assignment. Recently, the left invariant error formulation has been used in a FLS (Brossard et al. 2020) but for the purpose of uncertainty propagation on the extended Lie group $SE_2(3)$.

To analyze the observability of an estimator, there are in general two categories of approaches: those based on the linearized observability matrix of the discrete system, and those based on the observability matrix built from Lie derivatives of the continuous-time system. The discrete analysis is suitable to identify unobservable directions under a degenerate motion, and the Lie differentiation analysis is suitable to identify the requirements to make all state variables observable. Other methods exist but are typically unsuitable to examine the interplay between observability and consistency (Hesch et al. 2014a). The first category includes (Huang, Mourikis, and Roumeliotis 2010; Li and Mourikis 2013; Hesch et al. 2014b; Dong-Si and Mourikis 2012; Zhang et al. 2017; Yang et al. 2020). The second category includes (Mirzaei and Roumeliotis 2008; Kelly and Sukhatme 2011; Hesch et al. 2014a; Jung, Heo, and Park 2020). The conclusions by methods from the two categories are congruent. Interestingly, in examining observability, all cited methods parameterize landmarks in the world frame rather than in a local camera frame, possibly to reduce complexity. However, we find that expressing landmarks in a local frame is actually advantageous to the consistency analysis.

Methodology

This section presents the proposed right invariant FLS applied to the visual inertial SLAM problem with the analysis of its consistency. Though many state estimation problems without absolute position measurements may exhibit the inconsistent issue in an estimator, *e.g.*, stereo visual odometry (Dong-Si and Mourikis 2012), we choose to analyze the visual inertial SLAM problem regarding consistency because its observability property has been well studied and well-known. Though some variables in the following discussion are specific to the visual inertial SLAM, the proposed method for ensuring consistency is generic enough to translate to other state estimation problems solvable by a FLS.

We first formulate the visual inertial SLAM problem from the perspective of a FLS. Second, we present the right invariant error formulation, and prove that the consistency property is guaranteed.

Visual Inertial SLAM Formulation

In a typical visual inertial SLAM problem, we try to estimate the platform state, sensor parameters, and the unknown positions of landmarks in the environment, by fusing data cap-

tured by at least one camera and an IMU rigidly mounted on the platform.

State Variables The state of the system at time t_i consists of the navigation state of the platform $\boldsymbol{\pi}_i$ and the IMU biases $\mathbf{b}_i$, *i.e.*, $\mathbf{x}_i = (\boldsymbol{\pi}_i, \mathbf{b}_i)$. In turn, a navigation state $\boldsymbol{\pi}_i$ includes orientation $\mathbf{R}_i$, velocity $\mathbf{v}_i$, and position $\mathbf{p}_i$ of the body frame $\{B\}$ (affixed to the platform) expressed in a world frame $\{W\}$ (z-axis along gravity), *i.e.*, $\boldsymbol{\pi}_i = (\mathbf{R}_i, \mathbf{v}_i, \mathbf{p}_i)$. For clarity, the considered sensor parameters are only the IMU biases $\mathbf{b}$ which includes the gyro bias $\mathbf{b}_g$ and the accelerometer bias $\mathbf{b}_a$, *i.e.*, $\mathbf{b} = (\mathbf{b}_g, \mathbf{b}_a)$. We denote by $\mathbf{x}_{0:k}$ the entire history of system states up to time t_k, *i.e.*, $\mathbf{x}_{0:k} = \{\mathbf{x}_i | i = 0, 1, \ldots, k\}$.

The SLAM problem also estimates landmark feature positions $\mathbf{f}_l$, each of which is represented by an inverse depth parameterization in an anchor camera frame $\{C_a\}$ (Civera, Davison, and Montiel 2008), *i.e.*,

$$\mathbf{f}_l = [\alpha, \beta, 1, \rho]^\intercal = [x/z, y/z, 1, 1/z]^\intercal \tag{1}$$

where $[x, y, z]^\intercal$ is the Cartesian coordinates of the landmark in $\{C_a\}$. The inverse depth parameterization is chosen for two reasons. First, it has been shown to outperform the traditional Euclidean parameterization (Solà et al. 2012; Polok et al. 2015). Second, it decouples the landmark parameters from the platform pose in the world frame, thus they remain invariant under Euclidean transform of the original problem and have no bearing on the observability analysis.

We denote by $\mathcal{X}_k$ the history of state variables up to t_k,

$$\mathcal{X}_k = \{\mathbf{x}_i | i = 0, 1, \ldots, k\} \cup \{\mathbf{f}_l | l = 1, 2, \ldots, L\}. \tag{2}$$

Measurements Measurements in the visual inertial SLAM problem include camera observations and IMU readings. The observation $\mathbf{z}_{il}$ of a landmark $\mathbf{f}_l$ in camera frame $\{C_i\}$ at t_i is represented by a projection model $\mathbf{h}$ which encodes the camera intrinsic parameters, *i.e.*,

$$\mathbf{z}_{il} = \mathbf{h}(\mathbf{T}_{BC}^{-1}\mathbf{T}_{WBi}^{-1}\mathbf{T}_{WBa}\mathbf{T}_{BC}\mathbf{f}_l) + \mathbf{n}_c \tag{3}$$

where $\mathbf{n}_c \sim N(0, \boldsymbol{\Sigma}_c)$ is 2D Gaussian noise of covariance $\boldsymbol{\Sigma}_c$, $\mathbf{T}_{BC} \in SE(3)$ is the camera extrinsic parameters, and $\mathbf{T}_{WBi} = (\mathbf{R}_{WBi}, \mathbf{p}_{WBi}) = (\mathbf{R}_i, \mathbf{p}_i)$ and $\mathbf{T}_{WBa}$ are the platform poses at the observing epoch t_i and the anchor epoch t_a. Without loss of generality, we assume $\mathbf{T}_{BC}$ is well calibrated and known. Considering that $\mathbf{T}_{WBi}$ and $\mathbf{T}_{WBa}$ are subsumed by $\mathbf{x}_i$ and $\mathbf{x}_a$, the projection model can also be written as $\mathbf{z}_{il} = \mathbf{h}(\mathbf{x}_i, \mathbf{x}_a, \mathbf{f}_l) + \mathbf{n}_c$.

In a simplified IMU model, the IMU measurements $\mathbf{a}_m$ and $\boldsymbol{\omega}_m$ are assumed to be affected by accelerometer and gyroscope biases, $\mathbf{b}_a$ and $\mathbf{b}_g$, and Gaussian white noise processes, $\boldsymbol{\nu}_a$ and $\boldsymbol{\nu}_g$, of power spectral densities, $\sigma_a^2\mathbf{I}_3$ and $\sigma_g^2\mathbf{I}_3$, respectively, *i.e.*,

$$\mathbf{a}_m = {}_B\mathbf{a}_s + \mathbf{b}_a + \boldsymbol{\nu}_a \tag{4}$$

$$\dot{\mathbf{b}}_a = \boldsymbol{\nu}_{ba} \tag{5}$$

$$\boldsymbol{\omega}_m = {}_B\boldsymbol{\omega}_{WB} + \mathbf{b}_g + \boldsymbol{\nu}_g \tag{6}$$

$$\dot{\mathbf{b}}_g = \boldsymbol{\nu}_{bg}, \tag{7}$$

where the biases are assumed to be driven by Gaussian white noise processes, $\boldsymbol{\nu}_{ba}$ and $\boldsymbol{\nu}_{bg}$, of power spectral densities, $\sigma_{ba}^2\mathbf{I}_3$ and $\sigma_{bg}^2\mathbf{I}_3$, respectively. For brevity, we denote the IMU readings from t_i to t_j by $\mathbf{u}_{i:j} = \{(\boldsymbol{\omega}_m, \mathbf{a}_m)_k | k = i, i+1, \ldots, j\}$

With a sequence of IMU readings $\mathbf{u}_{i:j}$, the navigation state variable $\mathbf{x}(t_j)$ can be propagated from $\mathbf{x}(t_i)$ as expressed by $\mathbf{f}(\cdot)$,

$$\mathbf{x}(t_j|t_i) = \mathbf{f}(\mathbf{x}(t_i), \mathbf{u}_{i:j}, \mathbf{w}_{imu}), \tag{8}$$

where the continuous noises of IMU readings are stacked in $\mathbf{w}_{imu} = [\boldsymbol{\nu}_g^\intercal, \boldsymbol{\nu}_a^\intercal, \boldsymbol{\nu}_{bg}^\intercal, \boldsymbol{\nu}_{ba}^\intercal]^\intercal$. For brevity, we will drop the time symbol and keep only its index for variables in (8), *e.g.*, $\mathbf{x}_{j|i} = \mathbf{x}(t_j|t_i)$. The propagated navigation state $\boldsymbol{\pi}_{j|i}$ can be solved with the Runge-Kutta method (Jekeli 2001).

Global Bundle Adjustment and FLS

Before looking at the FLS, we first presents the basics of global bundle adjustment (BA) (Triggs et al. 2000) which is the base of the FLS. For the visual inertial SLAM problem, the objective function to be minimized in the global BA up to t_k is

$$E = \sum_{i=1}^{k} \|\mathbf{r}_x(\mathbf{x}_i, \mathbf{x}_{i|i-1})\|^2_{\boldsymbol{\Sigma}_{x,i-1:i}} + \sum_{(i,l)\in\mathcal{C}_k} \|\mathbf{r}_{il}(\mathbf{x}_i, \mathbf{x}_a, \mathbf{f}_l)\|^2_{\boldsymbol{\Sigma}_c}, \tag{9}$$

where $\mathbf{r}_x$ and $\mathbf{r}_{il}$ are residual errors associated with IMU and camera measurements, $\boldsymbol{\Sigma}_{x,k-1:k}$ and $\boldsymbol{\Sigma}_c$ are their corresponding covariance matrices, and $\mathcal{C}_k$ denotes all image measurements up to t_k. Note that the objective function does not include a gauge-fixing prior which will shadow unobservable directions.

The reprojection error $\mathbf{r}_{il}$ is usually defined to be the mismatch between predicted image coordinates of a landmark $\mathbf{f}_l$ and its measurement $\mathbf{z}_{il}$, *i.e.*, $\mathbf{r}_{il} = \mathbf{h}(\mathbf{x}_i, \mathbf{x}_a, \mathbf{f}_l) - \mathbf{z}_{il}$.

The IMU residual error $\mathbf{r}_x$ and its covariance depends on the error definitions and will be discussed later on.

Solving the objective function (9) is equivalent to finding a solution to fit the below nonlinear system,

$$\mathbf{W}\underbrace{\begin{bmatrix} \mathbf{r}_x(\mathbf{x}_1, \mathbf{x}_{1|0}) \\ \vdots \\ \mathbf{r}_x(\mathbf{x}_k, \mathbf{x}_{k|k-1}) \\ \vdots \\ \mathbf{r}_{il}(\mathbf{x}_i, \mathbf{x}_a, \mathbf{f}_l) \\ \vdots \end{bmatrix}}_{\mathbf{r}} = \mathbf{0}$$

$$\mathbf{W} = \left[\begin{array}{ccc|ccc} \boldsymbol{\Sigma}_{x,0:1}^{-1/2} & & & & & \\ & \ddots & & & & \\ & & \boldsymbol{\Sigma}_{x,k-1:k}^{-1/2} & & & \\ \hline & & & \boldsymbol{\Sigma}_{c,1}^{-1/2} & & \\ & & & & \ddots & \\ & & & & & \boldsymbol{\Sigma}_{c,m}^{-1/2} \end{array}\right] \tag{10}$$

where $m = |\mathcal{C}_k|$ is the total number of image observations.

FLS minimizes (9) by repeatedly going through two steps, linearization and marginalization, as described next.

Factor Linearization Before linearizing the measurement factors, the error state (*i.e.*, the 'small' perturbation) must be defined. Without loss of generality, we define the error state $\delta\mathbf{x}$ as a function $\boldsymbol{\eta}$ of the random variable $\mathbf{x}$ and its noise free estimate $\bar{\mathbf{x}}$, *i.e.*, $\delta\mathbf{x} = \boldsymbol{\eta}(\mathbf{x}, \bar{\mathbf{x}})$. For a variable in a real vector space, the error state is simply $\delta\mathbf{x} = \mathbf{x} - \bar{\mathbf{x}}$. Also, we define the inverse of $\boldsymbol{\eta}$ such that $\mathbf{x} = \boldsymbol{\eta}^{-1}(\bar{\mathbf{x}}, \delta\mathbf{x})$.

With the error state, the residual errors can be linearized at estimates of state variables with the first order approximation. The reprojection error is linearized as

$$\begin{aligned}\mathbf{r}_{il}(\mathbf{x}_i, \mathbf{x}_a, \mathbf{f}_l) \approx \mathbf{r}_{il}(\bar{\mathbf{x}}_i, \bar{\mathbf{x}}_a, \bar{\mathbf{f}}_l) + \mathbf{J}_{x_i,l}\delta\mathbf{x}_i + \\ \mathbf{J}_{x_a,l}\delta\mathbf{x}_a + \mathbf{J}_{f_l}\delta\mathbf{f}_l\end{aligned} \tag{11}$$

where $\mathbf{J}_{x_i,l}$, $\mathbf{J}_{x_a,l}$, and $\mathbf{J}_{f_l}$ are Jacobians of $\mathbf{r}_{il}$ relative to $\mathbf{x}_i$, $\mathbf{x}_a$, and $\mathbf{f}_l$.

The above-mentioned IMU residual error $\mathbf{r}_x$ is usually defined to be $\mathbf{r}_x(\mathbf{x}_i, \mathbf{x}_{i|i-1}) = \boldsymbol{\eta}(\mathbf{x}_i, \mathbf{x}_{i|i-1})$. It is linearized as

$$\begin{aligned}\mathbf{r}_x(\mathbf{x}_i, \mathbf{x}_{i|i-1}) &\approx \mathbf{r}_x(\bar{\mathbf{x}}_i, \bar{\mathbf{x}}_{i|i-1}) + \mathbf{A}_i\delta\mathbf{x}_i + \\ &\quad \mathbf{A}_{i|i-1}\delta\mathbf{x}_{i|i-1} \\ &= \mathbf{r}_x(\bar{\mathbf{x}}_i, \bar{\mathbf{x}}_{i|i-1}) + \mathbf{A}_i\delta\mathbf{x}_i + \\ &\quad \mathbf{A}_{i|i-1}\boldsymbol{\Phi}(t_i, t_{i-1})\delta\mathbf{x}_{i-1}\end{aligned} \tag{12}$$

where $\mathbf{A}_i$ and $\mathbf{A}_{i|i-1}$ are the Jacobians of $\mathbf{r}_x$ relative to $\delta\mathbf{x}_i$ and $\delta\mathbf{x}_{i|i-1}$, and $\boldsymbol{\Phi}(t_i, t_{i-1})$ is the discrete IMU transition matrix obtained by linearizing (8). To obtain the weight covariance $\boldsymbol{\Sigma}_{x,i-1:i}$, we note that the covariance of $\delta\mathbf{x}_{i|i-1}$, $\boldsymbol{\Sigma}_{\mathbf{x}_{i|i-1}}$, can be propagated from a zero matrix by the propagation function (8) given the defined error state $\delta\mathbf{x}$, then $\boldsymbol{\Sigma}_{x,i-1:i} = \mathbf{A}_{i|i-1}\boldsymbol{\Sigma}_{\mathbf{x}_{i|i-1}}\mathbf{A}_{i|i-1}^{\intercal}$.

Linearization turns the nonlinear system (10) to a set of linear equations that we try to satisfy at once,

$$\mathbf{W}\left(\underbrace{\begin{bmatrix}\mathbf{r}_x(\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_{1|0}) \\ \vdots \\ \mathbf{r}_x(\bar{\mathbf{x}}_k, \bar{\mathbf{x}}_{k|k-1}) \\ \vdots \\ \mathbf{r}_{il}(\bar{\mathbf{x}}_i, \bar{\mathbf{x}}_a, \bar{\mathbf{f}}_l) \\ \vdots\end{bmatrix}}_{\bar{\mathbf{r}}} + \mathbf{J}\underbrace{\begin{bmatrix}\delta\mathbf{x}_0 \\ \vdots \\ \delta\mathbf{x}_k \\ \vdots \\ \delta\mathbf{f}_l \\ \vdots\end{bmatrix}}_{\delta\mathcal{X}_k}\right) = \mathbf{0}$$

$$\mathbf{J} = \left[\begin{array}{cccc|c} \mathbf{A}_{1|0}\boldsymbol{\Phi}_{1|0} & \mathbf{A}_1 & & & \\ & \cdots & & & \\ & & \mathbf{A}_{k|k-1}\boldsymbol{\Phi}_{k|k-1} & \mathbf{A}_k & \\ \hline & & \cdots & & \cdots \\ & \mathbf{J}_{x_i,l} & & \mathbf{J}_{x_a,l} & \mathbf{J}_{f_l} \\ & & \cdots & & \cdots \end{array}\right]. \tag{13}$$

Unobservable Directions and Nullspace To analyze the observability of the visual inertial SLAM problem, we need to specify the unobservable directions and relate them to the objective function (9) and the linearized system (13).

First, let's define a transformation $\mathcal{T}_\xi$ (minimally parameterized by $\boldsymbol{\xi}$) of the considered problem which transforms all state variables from the present world frame $\{W\}$ to another one, say $\{W_y\}$. The transformation applies to all variables relevant to $\{W\}$, *i.e.*, $\mathbf{x}_{0:k}$, which become $\mathbf{y}_{0:k}$ after the transformation.

With measurements from a camera and a consumer-grade IMU, it is impossible to determine the absolute position and heading of the platform (Jones and Soatto 2011) which are the unobservable directions for the visual inertial SLAM problem. When a transformation involves only an translation $\delta\mathbf{t}$ and a rotation about gravity $\delta\phi$, *i.e.*, $\boldsymbol{\xi} = [\delta\phi \;\; \delta\mathbf{t}]$, the value of the objective function (9) is invariant to the transformation because the residual errors do not change w.r.t the new variables $\mathbf{y}_{0:k}$, *i.e.*,

$$\begin{aligned}\mathbf{r}_x(\mathbf{x}_i, \mathbf{x}_{i|i-1}) &= \mathbf{r}_x(\mathcal{T}_\xi(\mathbf{x}_i), \mathcal{T}_\xi(\mathbf{x}_{i|i-1})) \\ \mathbf{r}_{il}(\mathbf{x}_i, \mathbf{x}_a, \mathbf{f}_l) &= \mathbf{r}_{il}(\mathcal{T}_\xi(\mathbf{x}_i), \mathcal{T}_\xi(\mathbf{x}_a), \mathbf{f}_l)\end{aligned} \tag{14}$$

Thus, the linearized system (13) still holds but with $\mathbf{y}_{0:k}$.

Next, we reveal that the unobservable directions correspond to the nullspace of $\mathbf{J}$ in (13). When $\boldsymbol{\xi}$ is close to the zero vector, the objective function after the transformation $\mathcal{T}_\xi$ can be linearized at the estimates for $\mathcal{X}_k$, and the linearized system becomes

$$\mathbf{W}(\bar{\mathbf{r}} + \mathbf{J}\delta\mathcal{Y}_k) = \mathbf{0} \tag{15}$$

where $\delta\mathcal{Y}_k$ is the error between the transformed state variables $\mathcal{Y}_k = \mathcal{T}_\xi(\mathcal{X}_k)$ and their linearization points $\bar{\mathcal{X}}_k$. By comparing (13) and (15), we observe that

$$\begin{aligned}\mathbf{0} &= \mathbf{J}(\delta\mathcal{Y}_k - \delta\mathcal{X}_k) \\ &= \mathbf{J}(\boldsymbol{\eta}(\mathcal{T}_\xi(\mathcal{X}_k), \bar{\mathcal{X}}_k) - \boldsymbol{\eta}(\mathcal{X}_k, \bar{\mathcal{X}}_k)) \\ &= \mathbf{J}\underbrace{\frac{\partial\boldsymbol{\eta}(\mathcal{T}_\xi(\mathcal{X}_k), \bar{\mathcal{X}}_k)}{\partial\boldsymbol{\xi}}}_{\mathbf{N}_J}\boldsymbol{\xi}.\end{aligned} \tag{16}$$

As the expression holds for arbitrary small $\boldsymbol{\xi}$, we have

$$\mathbf{J}\mathbf{N}_J = \mathbf{0}, \tag{17}$$

which means that changes on the column space of $\mathbf{N}_J$ to variables do not affect the linearized system. Thus, the nullspace of $\mathbf{J}$, $\mathbf{N}_J$ corresponds to the unobservable directions of the problem. In this sense, the Jacobian matrix of the system $\mathbf{J}$ is equivalent to the classic observability matrix in revealing the unobservable directions. Indeed, the observability matrix can be obtained from $\mathbf{J}$ by basic row operations as shown in (Dong-Si and Mourikis 2011).

Factor Marginalization The FLS has been a popular approach to the visual inertial SLAM problem, *e.g.*, (Rosinol et al. 2020). Essentially, it solves the problem by repeated linearization of factors, and gradually marginalizes old variables from the global BA problem to bound problem size. Every marginalization step creates a linear prior factor for variables connected to those removed variables.

Consider a marginalization step where variables prior to t_m are marginalized. The objective function (9) becomes

$$\begin{aligned} E_m = & \sum_{i=1}^{m} \|\mathbf{r}_x(\bar{\mathbf{x}}_i, \bar{\mathbf{x}}_{i|i-1}) + \mathbf{A}_i \delta\mathbf{x}_i + \\ & \mathbf{A}_{i|i-1}\mathbf{\Phi}(t_i, t_{i-1})\delta\mathbf{x}_{i-1}\|^2_{\mathbf{\Sigma}_{x,i-1:i}} + \\ & \sum_{(i,l)\in\mathcal{M}} \|\mathbf{r}_{il}(\bar{\mathbf{x}}_i, \bar{\mathbf{x}}_a, \bar{\mathbf{f}}_l) + \mathbf{J}_{x_i,l}\delta\mathbf{x}_i + \\ & \mathbf{J}_{x_a,l}\delta\mathbf{x}_a + \mathbf{J}_{f_l}\delta\mathbf{f}_l\|^2_{\mathbf{\Sigma}_c} + \\ & \sum_{i=m+1}^{k} \|\mathbf{r}_x(\mathbf{x}_i, \mathbf{x}_{i|i-1})\|^2_{\mathbf{\Sigma}_{x,i-1:i}} + \\ & \sum_{(i,l)\in\mathcal{C}_k\setminus\mathcal{M}} \|\mathbf{r}_{il}(\mathbf{x}_i, \mathbf{x}_a, \mathbf{f}_l)\|^2_{\mathbf{\Sigma}_c}, \end{aligned} \tag{18}$$

where $\mathcal{M}$ is the set of marginalized camera observations. The first two linear terms of E_m are obtained by fixing linearization points for the marginalized nonlinear factors. For convenience of analysis, none of the marginalized terms is discarded. In implementation, the first two linear terms of E_m are equivalently represented by a marginalization factor which is obtained by the Schur complement method.

As the optimizer iterates, E_m will be relinearized. For a variable in the marginalization factor (*e.g.*, $\mathbf{x}_m$), a nonlinear term of E_m usually will be linearized at a different estimate ($\bar{\mathbf{x}}'_m$) than the one ($\bar{\mathbf{x}}_m$) used in the linear terms of E_m. Thus, the Jacobian matrix in the linearized system (13) will have blocks evaluated at different points for the same variables in the marginalization factor. For the traditional error definition, this causes shrunk nullspace of $\mathbf{J}$ and inconsistent covariances as shown in (Dong-Si and Mourikis 2011).

Right Invariant Fixed-Lag Smoother

In contrast to traditional error definitions, the right invariant error formulation does not suffer from this inconsistency in observable dimensions and covariances.

The Right Invariant Error The right invariant error is defined relative to the navigation state $\boldsymbol{\pi}_i$, viewed as an element X_i of $SE_2(3)$ (Barrau and Bonnabel 2016b), *i.e.*,

$$X_i = \begin{bmatrix} \mathbf{R}_i & \mathbf{v}_i & \mathbf{p}_i \\ \mathbf{0}_{1\times3} & 1 & 0 \\ \mathbf{0}_{1\times3} & 0 & 1 \end{bmatrix} \in SE_2(3). \tag{19}$$

The right invariant error $\boldsymbol{\xi}_{\pi,i}$ consisting of rotational error $\delta\boldsymbol{\theta}_i$, velocity error $\delta\mathbf{v}_i$, and positional error $\delta\mathbf{p}_i$, is given by

$$\boldsymbol{\xi}_{\pi,i} = (\delta\boldsymbol{\theta}_i, \delta\mathbf{v}_i, \delta\mathbf{p}_i), \tag{20}$$

$$X_i = \exp(\mathcal{L}(\boldsymbol{\xi}_{\pi,i}))\bar{X}_i, , \tag{21}$$

$$\mathcal{L}(\boldsymbol{\xi}_\pi) = \begin{bmatrix} \delta\boldsymbol{\theta}_\times & \delta\mathbf{v} & \delta\mathbf{p} \\ \mathbf{0}_{1\times3} & 1 & 0 \\ \mathbf{0}_{1\times3} & 0 & 1 \end{bmatrix} \tag{22}$$

where $\exp(\cdot)$ is the matrix exponent, and $\mathcal{L}(\boldsymbol{\xi}_\pi)$ is the Lie operator for $SE_2(3)$, computed with the skew operator $(\cdot)_\times$. The closed form expression for the exponential map of $\boldsymbol{\xi}_\pi$ is,

$$\exp(\mathcal{L}(\boldsymbol{\xi}_\pi)) = \begin{bmatrix} \exp(\delta\boldsymbol{\theta}_\times) & \mathbf{J}_l(\delta\boldsymbol{\theta})\delta\mathbf{v} & \mathbf{J}_l(\delta\boldsymbol{\theta})\delta\mathbf{p} \\ \mathbf{0}_{1\times3} & 1 & 0 \\ \mathbf{0}_{1\times3} & 0 & 1 \end{bmatrix} \tag{23}$$

where $\mathbf{J}_l(\cdot)$ is the left Jacobian for $SO(3)$ (Barfoot and Furgale 2014).

"Right invariance" is on the grounds that the error for X is the same as that for its transformed variable, XY, obtained by right multiplication with an element $Y \in SE_2(3)$, as shown by $XY = \exp(\mathcal{L}(\boldsymbol{\xi}_\pi))(\bar{X}Y)$. That is, the right invariant error is independent of the system state.

Consistency Property For the right invariant errors, assuming $\Delta t = t_i - t_{i-1}$ is small (*e.g.*, 0.1s), the discrete transition matrix $\mathbf{\Phi}(t_i, t_{i-1})$ is found to be

$$\begin{aligned} \mathbf{\Phi}_{i|i-1} &= \begin{bmatrix} \mathbf{\Phi}_\pi & \mathbf{\Phi}_{\pi,b} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \\ \mathbf{\Phi}_\pi &= \begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{g}_\times\Delta t & \mathbf{I} & \mathbf{0} \\ \mathbf{g}_\times\Delta t^2/2 & \mathbf{I}\Delta t & \mathbf{I} \end{bmatrix} \\ \mathbf{\Phi}_{\pi,b} &= \begin{bmatrix} -\mathbf{R}\Delta t & \mathbf{0} \\ -\mathbf{v}_\times\mathbf{R}\Delta t - \mathbf{g}_\times\mathbf{R}\frac{\Delta t^2}{2} & -\mathbf{R}\Delta t \\ -\mathbf{p}_\times\mathbf{R}\Delta t - \mathbf{v}_\times\mathbf{R}\frac{\Delta t^2}{2} - \mathbf{g}_\times\mathbf{R}\frac{\Delta t^3}{6} & -\mathbf{R}\frac{\Delta t^2}{2} \end{bmatrix} \end{aligned} \tag{24}$$

where $\mathbf{g}$ is the gravity vector in $\{W\}$, and we drop the subscript 'i' of $(\mathbf{R}, \mathbf{v}, \mathbf{p})$ for brevity. Thanks to right invariance, $\mathbf{\Phi}_\pi$ is independent of the state variable $\boldsymbol{\pi}_i$.

Another useful finding is that the parameters of landmarks anchored at a camera frame and sensor parameters (*e.g.*, biases) do not interfere with nullspace of the coefficient matrix $\mathbf{J}$ of the linearized system (13). Thus, their Jacobians can be safely ignored in analyzing consistency.

The right invariance property together with the above finding lead to the proof that the right invariant error formulation can ensure consistency of the FLS as detailed in the supplementary material.

One point worth noting is that the proof approximates two component Jacobians for the IMU residual error (12), $\mathbf{A}_i$ and $\mathbf{A}_{i|i-1}$, by identities,

$$\mathbf{A}_i \approx \mathbf{I}_{15} \quad \mathbf{A}_{i|i-1} \approx -\mathbf{I}_{15}. \tag{25}$$

This approximation is also used in (Dong-Si and Mourikis 2011) for proving consistency of the FEJ technique. It is reasonable when the IMU residual error $\mathbf{r}_x$ is small, and we found that using the exact $\mathbf{A}_i$ and $\mathbf{A}_{i|i-1}$ led to slight inconsistency (see Fig. 4).

Simulation Results

This section presents the simulation results, validating that the FLS formulated with right invariant errors has consistent covariances.

Σ	Gyroscope	Accelerometer
Bias White Noise	$\sigma_{bg}^2/f\mathbf{I}_3$ with $\sigma_{bg} = 2 \cdot 10^{-5}$ $rad/s^2/\sqrt{Hz}$	$\sigma_{ba}^2/f\mathbf{I}_3$ with $\sigma_{ba} = 5.5 \cdot 10^{-5}$ $m/s^3/\sqrt{Hz}$
White Noise	$\sigma_g^2 f\mathbf{I}_3$ with $\sigma_g = 1.2 \cdot 10^{-3}$ $rad/s/\sqrt{Hz}$	$\sigma_a^2 f\mathbf{I}_3$ with $\sigma_a = 8 \cdot 10^{-3}$ $m/s^2/\sqrt{Hz}$

Table 1: Covariances of the zero-mean Gaussian distributions from which discrete noise samples are drawn. f is the IMU sampling rate.

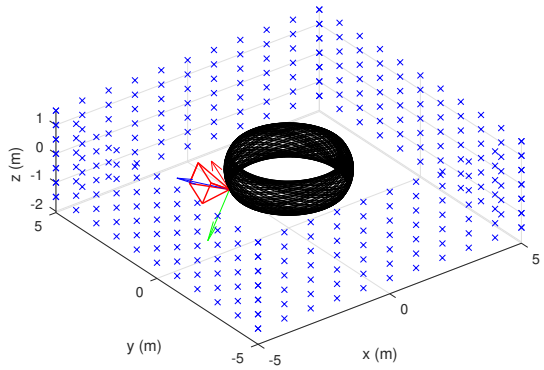

Figure 1: Simulated scene with general torus motion lasting for five minutes. A sample camera view frustum is shown by the red wireframe on the black trajectory.

Error Metrics The consistency of a FLS is measured by the Normalized Estimation Error Squared (NEES) of components of the navigation state variable. The expected value of NEES for a variable is its degrees of freedom, hence 3 for positional error $\delta\mathbf{p}_{WB}$, 3 for orientation error $\delta\boldsymbol{\theta}_{WB}$, and 6 for pose error $\delta\mathbf{T}_{WB} = (\delta\mathbf{p}_{WB}, \delta\boldsymbol{\theta}_{WB})$. An inconsistent estimator will optimistically estimate the covariance, thus the computed NEES is greater than its expected value. Following (Bar-Shalom, Li, and Kirubarajan 2004, (3.7.6-1)), with n_s successful runs of an estimator, the NEES $\boldsymbol{\epsilon}$ for position, orientation, and pose at epoch t is given by

$$\boldsymbol{\epsilon}_X(t) = \frac{1}{n_s}\Sigma_{i=1}^{n_s}\delta\mathbf{X}(t)^\intercal\boldsymbol{\Sigma}_X^{-1}(t)\delta\mathbf{X}(t) \tag{26}$$

where $\mathbf{X} = \mathbf{p}_{WB}, \boldsymbol{\theta}_{WB}, \mathbf{T}_{WB}$, and $\boldsymbol{\Sigma}_X$ is its covariance.

The accuracy of the estimated state is measured by Root Mean Square Error (RMSE) for components of the state vector. A component $\mathbf{X}$'s RMSE r_X at t is given by

$$r_X(t) = \sqrt{\frac{1}{n_s}\Sigma_{i=1}^{n_s}\delta\mathbf{X}(t)^\intercal\delta\mathbf{X}(t)} \tag{27}$$

Simulation Setup A scene with point landmarks distributed on four walls was simulated. A monocular camera-IMU platform traversed the scene for five minutes with a torus trajectory (Fig. 1). The platform moved at an average velocity 2.30 m/s.

The camera captured images of size 752×480 at 10Hz. The image observations were corrupted by white Gaussian noise of 1 pixel standard deviation at each direction. The simulated inertial measurements were sampled at f=100 Hz, corrupted by random walk biases and additive white noise. Discrete noise samples were drawn from Gaussian distributions tabulated in Table 1. These noise parameters were chosen to be realistic for a consumer-grade IMU.

Estimator Setup The proposed FLS was implemented with the `IncrementalFixedLagSmoother` in GTSAM (Dellaert 2012) which wraps the iSAM2 (Kaess et al. 2012) method. By setting the time horizon to a large value, it turns into the iSAM2 which gives results very close to a batch solution (Forster et al. 2017). Also, GTSAM provides a `BatchFixedLagSmoother` wrapping a Levenberg-Marquardt solver which ensures consistency by locking variables in the marginalization factor.

We compared several estimators, the incremental FLS (Inc. FLS), the batch FLS, iSAM2, and the proposed FLS with the right invariant error (RI-FLS). The first three estimators used the error state defined in (Forster et al. 2017). Except for iSAM2, the other estimators adopted a time horizon of 1 second.

A simulation frontend was created to provide feature tracks to an estimator. It associated observations of a landmark between consecutive frames and between current frame and a selected earlier reference frame. For the torus motion, the average feature track length was 5.8, and the average number of observed landmarks in an image was 40.5.

All estimators were initialized with the true pose but a noisy velocity estimate affected by noise of Gaussian distribution $N(\mathbf{0}, 0.05^2\mathbf{I}_3\, m^2/s^4)$. Each estimator ran 100 times, and only successful runs (with the error in position ≤ 100 m at the end), were used to compute the error metrics.

Estimator Consistency For the above estimators, the evolution of NEES is visualized in Fig. 2. The NEES values averaged over the last 10 seconds to smooth out jitters are tabulated in Table 2. From the NEES curves and their final values, we see that both incremental FLS and batch FLS did not output consistent covariances, and incremental FLS performed better than batch FLS in terms of orientation NEES. On the other hand, the proposed RI-FLS and iSAM2 achieved NEES values very close to the reference. It is expected that iSAM2 exhibits consistency as it does not drop out variables. It is a bit surprising that RI-FLS achieved even better NEES than iSAM2, indicating that the right invariant formulation is effective for ensuring consistency.

To assess the state estimation accuracy, the RMSE values for each dimension of position, orientation, and IMU biases, are drawn in Fig. 3. Unsurprisingly, iSAM2 achieved best accuracy for all these variables. Incremental FLS and batch FLS had an issue in constraining errors on one horizontal direction of the gyro bias. All estimators estimated well the accelerometer bias. RI-FLS outperformed other FLSs in position accuracy, and achieved good orientation accuracy.

RI-FLS variants We also examine the effect of approximating the IMU residual Jacobians, and evaluate a RI-FLS variant with smart factors (Forster et al. 2017).

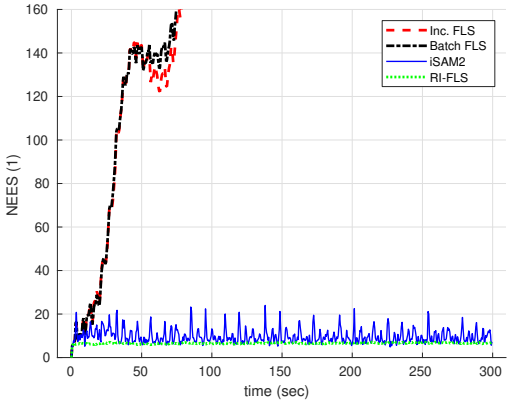

Figure 2: The history of NEES for pose of estimators including incremental FLS, batch FLS, iSAM2, and right invariant FLS. The expected value of NEES for pose is 6.

	NEES averaged over last 10 seconds		
	Position (1)	Orientation (1)	Pose (1)
Reference	3	3	6
Inc. FLS	644.3	5.0	657.3
Batch FLS	693.6	102.3	800.3
iSAM2	5.4	4.1	9.3
RI-FLS	3.3	3.4	6.6

Table 2: NEES computed over 100 runs for estimators including incremental FLS, batch FLS, iSAM2, and right invariant FLS.

The consistency analysis approximates IMU residual Jacobians components $\mathbf{A}_i$ and $\mathbf{A}_{i|i-1}$ with identities (25). When the exact expressions for $\mathbf{A}_i$ and $\mathbf{A}_{i|i-1}$ are used, the observability property $\mathbf{J}\mathbf{N}_J = \mathbf{0}$ may not hold.

The RI-FLS with smart factors is motivated by the fact that the GTSAM optimizer often throws the indeterminant system exception because of landmarks with low disparity that are common for real data. Smart factors fix this issue by removing landmarks from the optimizer. We think this technique will not adversely impact estimator consistency.

To confirm these thoughts, three variants of RI-FLS were tested in the above simulation setup: RI-FLS with approximated IMU Jacobians (baseline), RI-FLS with exact IMU Jacobians (RI-FLS exact), and RI-FLS with smart factors and approximated IMU Jacobians (RI-FLS smart). The history of the NEES for the three methods shown in Fig. 4 confirm that exact IMU Jacobians lead to worse NEES values, and that smart factors do not worsen NEES values.

Results on Real Data

To show practicality, we tested the incremental FLS with errors defined in (Forster et al. 2017), RI-FLS, and RI-FLS with exact Jacobians on the EuRoC benchmark. All methods were implemented with smart factors to handle degenerate landmarks and state variables were associated with consecutive camera frames in a time horizon of 1 second. The absolute translation error RMS (Zhang and Scaramuzza 2018)

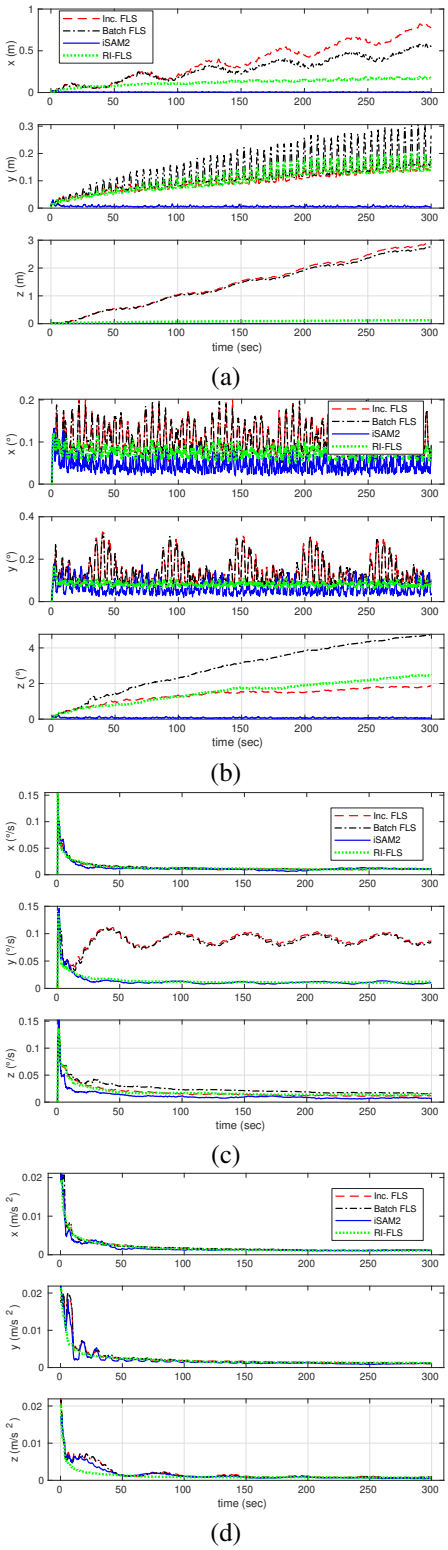

Figure 3: RMSE of position (a), orientation (b), gyro bias (c), and accelerometer bias (d), computed over 100 runs for estimators including incremental FLS, batch FLS, iSAM2, and right invariant FLS.

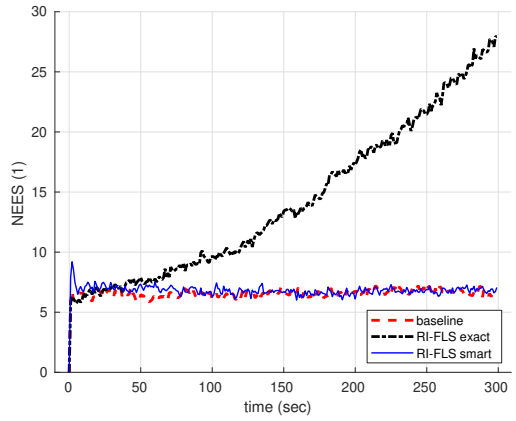

Figure 4: The history of NEES for pose of estimators including the baseline RI-FLS, RI-FLS with exact IMU Jacobians, and RI-FLS with smart factors.

Mean ATE RMS (m)	**MH_01**	**MH_05**	**V1_02**	**V2_02**
Inc. FLS	0.88	**0.68**	**0.28**	0.24
RI-FLS	**0.53**	0.89	**0.28**	0.29
RI-FLS with exact Jacobians	0.82	1.26	0.39	**0.23**

Table 3: Absolute translation error RMS averaged over 3 runs on several EuRoC sessions for incremental fixed-lag smoother (FLS), right invariant FLS (RI-FLS, the proposed), RI-FLS with exact IMU factor Jacobians. All methods use smart factors to deal with degenerate landmarks.

averaged over 3 runs on several EuRoC sequences are tabulated in Table 3 which shows that the proposed RI-FLS achieved comparable accuracy to the established method, incremental FLS with a traditional error formulation. The odometry accuracy could be improved by using the concept of keyframes as in (Forster et al. 2017).

Conclusion

To fix the inconsistent covariances output by traditional FLSs, we introduce the right invariant error formulation into the FLS framework. We analyze its observability directly with the linearized system, which has much lower analysis complexity than observability matrices. As a byproduct, we find that landmarks parameterized in a local camera frame and sensor parameters like biases do not affect the estimator consistency. In the end, we prove that the right invariant error formulation ensures the observability property of a FLS without artificially correcting Jacobians like the first estimate Jacobian method. The proposed right invariant FLS is applied to a monocular visual inertial SLAM problem. Its consistency is confirmed by simulation, and its practicality is verified with the EuRoC benchmark.

In the future, we will examine the consistency of observable parameters after marginalization, and look into the properties of the left invariant error formulation.

Acknowledgments

We thank the anonymous reviewers for stimulating comments and suggestions. Jianzhu Huai is partially funded by the National Natural Science Foundation of China (grant number 62003248).

References

Bar-Shalom, Y.; Li, X.; and Kirubarajan, T. 2004. *Estimation with Applications to Tracking and Navigation: Theory Algorithms and Software*. John Wiley & Sons.

Barfoot, T. D.; and Furgale, P. T. 2014. Associating uncertainty with three-dimensional poses for use in estimation problems. *IEEE Transactions on Robotics* 30(3): 679–693.

Barrau, A.; and Bonnabel, S. 2016a. An EKF-SLAM algorithm with consistency properties. Technical report. URL http://arxiv.org/abs/1510.06263.

Barrau, A.; and Bonnabel, S. 2016b. The invariant extended Kalman filter as a stable observer. *IEEE Transactions on Automatic Control* 62(4): 1797–1812.

Brossard, M.; Barrau, A.; and Bonnabel, S. 2018. Exploiting Symmetries to Design EKFs with Consistency Properties for Navigation and SLAM. *IEEE Sensors Journal* 19(4): 1572–1579.

Brossard, M.; Barrau, A.; Chauchat, P.; and Bonnabel, S. 2020. Associating uncertainty to extended poses for on Lie group IMU preintegration with rotating Earth. Technical report. URL http://arxiv.org/abs/2007.14097.

Burri, M.; Nikolic, J.; Gohl, P.; Schneider, T.; Rehder, J.; Omari, S.; Achtelik, M. W.; and Siegwart, R. 2016. The EuRoC micro aerial vehicle datasets. *The International Journal of Robotics Research* 35(10): 1157–1163.

Castellanos, J.; Martinez-Cantin, R.; Tardós, J.; and Neira, J. 2007. Robocentric map joining: Improving the consistency of EKF-SLAM. *Robotics and Autonomous Systems* 55(1): 21–29. URL https://linkinghub.elsevier.com/retrieve/pii/S0921889006001448.

Civera, J.; Davison, A.; and Montiel, J. 2008. Inverse depth parametrization for monocular SLAM. *IEEE Transactions on Robotics* 24(5): 932–945. doi:10.1109/TRO.2008.2003276.

Costante, G.; and Mancini, M. 2020. Uncertainty estimation for data-driven visual odometry. *IEEE Transactions on Robotics* 36(6): 1738–1757. doi:10.1109/TRO.2020.3001674.

Dellaert, F. 2012. Factor graphs and GTSAM: A hands-on introduction. Technical Report GT-RIM-CP&R-2012-002, Georgia Institute of Technology, Atlanta, Georgia, US.

Dong-Si, T.-C.; and Mourikis, A. I. 2011. Motion tracking with fixed-lag smoothing: Algorithm and consistency analysis. In *2011 IEEE International Conference on Robotics and Automation (ICRA)*, 5655–5662. Shanghai, China.

Dong-Si, T.-C.; and Mourikis, A. I. 2012. Consistency analysis for sliding-window visual odometry. In *2012 IEEE International Conference on Robotics and Automation (ICRA)*, 5202–5209. Saint Paul, MN, USA.

Forster, C.; Carlone, L.; Dellaert, F.; and Scaramuzza, D. 2017. On-manifold preintegration for real-time visual-inertial odometry. *IEEE Transactions on Robotics* 33(1): 1–21. URL https://ieeexplore.ieee.org/document/7557075/.

Heo, S.; and Park, C. G. 2018. Consistent EKF-based visual-inertial odometry on matrix Lie group. *IEEE Sensors Journal* 18(9): 3780–3788.

Hesch, J. A.; Kottas, D. G.; Bowman, S. L.; and Roumeliotis, S. I. 2014a. Camera-IMU-based localization: Observability analysis and consistency improvement. *The International Journal of Robotics Research* 33(1): 182–201. doi: 10.1177/0278364913509675.

Hesch, J. A.; Kottas, D. G.; Bowman, S. L.; and Roumeliotis, S. I. 2014b. Consistency analysis and improvement of vision-aided inertial navigation. *IEEE Transactions on Robotics* 30(1): 158–176. doi:10.1109/TRO.2013.2277549.

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

Jekeli, C. 2001. *Inertial Navigation Systems with Geodetic Applications*. Berlin, Germany: de Gruyter. doi:10.1515/9783110800234.

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

Jung, J. H.; Heo, S.; and Park, C. G. 2020. Observability analysis of IMU intrinsic parameters in stereo visual–inertial odometry. *IEEE Transactions on Instrumentation and Measurement* 69(10): 7530–7541. doi:10.1109/TIM.2020.2985174.

Kaess, M.; Johannsson, H.; Roberts, R.; Ila, V.; Leonard, J. J.; and Dellaert, F. 2012. iSAM2: Incremental smoothing and mapping using the Bayes tree. *The International Journal of Robotics Research* 31(2): 216–235.

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

Li, M.; and Mourikis, A. I. 2013. High-precision, consistent EKF-based visual-inertial odometry. *The International Journal of Robotics Research* 32(6): 690–711. URL https://doi.org/10.1177/0278364913481251.

Mirzaei, F. M.; and Roumeliotis, S. I. 2008. A Kalman filter-based algorithm for IMU-camera calibration: Observability analysis and performance evaluation. *IEEE Transactions on Robotics* 24(5): 1143–1156. doi:10.1109/TRO.2008.2004486.

Polok, L.; Lui, V.; Ila, V.; Drummond, T.; and Mahony, R. 2015. The effect of different parameterisations in incremental structure from motion. In *2015 Australian Conference on Robotics and Automation (ACRA)*. Canberra, Australia.

Rosinol, A.; Abate, M.; Chang, Y.; and Carlone, L. 2020. Kimera: An open-source library for real-time metric-semantic localization and mapping. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, 1689–1696. Paris, France. URL https://github.com/MIT-SPARK/Kimera.

Solà, J.; Vidal-Calleja, T.; Civera, J.; and Montiel, J. M. M. 2012. Impact of landmark parametrization on monocular EKF-SLAM with points and lines. *International Journal of Computer Vision* 97(3): 339–368. doi:10.1007/s11263-011-0492-5.

Triggs, B.; McLauchlan, P. F.; Hartley, R. I.; and Fitzgibbon, A. W. 2000. Bundle adjustment – A modern synthesis. In Triggs, B.; Zisserman, A.; and Szeliski, R., eds., *Vision Algorithms: Theory and Practice*, Lecture Notes in Computer Science, 298–372. Berlin, Heidelberg: Springer. doi:10.1007/3-540-44480-7_21.

Usenko, V.; Demmel, N.; Schubert, D.; Stückler, J.; and Cremers, D. 2020. Visual-inertial mapping with non-linear factor recovery. *IEEE Robotics and Automation Letters* 5(2): 422–429. doi:10.1109/LRA.2019.2961227.

Yang, Y.; Geneva, P.; Zuo, X.; and Huang, G. 2020. Online IMU intrinsic calibration: Is it necessary? In *Robotics: Science and Systems (RSS)*, 716–725. Corvallis, Oregon.

Zhang, T.; Wu, K.; Song, J.; Huang, S.; and Dissanayake, G. 2017. Convergence and consistency analysis for a 3-D Invariant-EKF SLAM. *IEEE Robotics and Automation Letters* 2(2): 733–740.

Zhang, Z.; and Scaramuzza, D. 2018. A tutorial on quantitative trajectory evaluation for visual(-inertial) odometry. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 7244–7251. Madrid, Spain. doi:10.1109/IROS.2018.8593941.