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

Jianzhu Huai, Yukai Lin, Yuan Zhuang, Min Shi

{jianzhu.huai,yuan.zhuang}@whu.edu.cn, linyuk@ethz.ch, mshi2018@fau.edu

Abstract

State estimation problems that use relative observations have immanent unobservable directions. Traditional causal estimators, however, usually gain spurious information on the unobservable directions, leading to over-confident covariance inconsistent to the actual estimator errors.

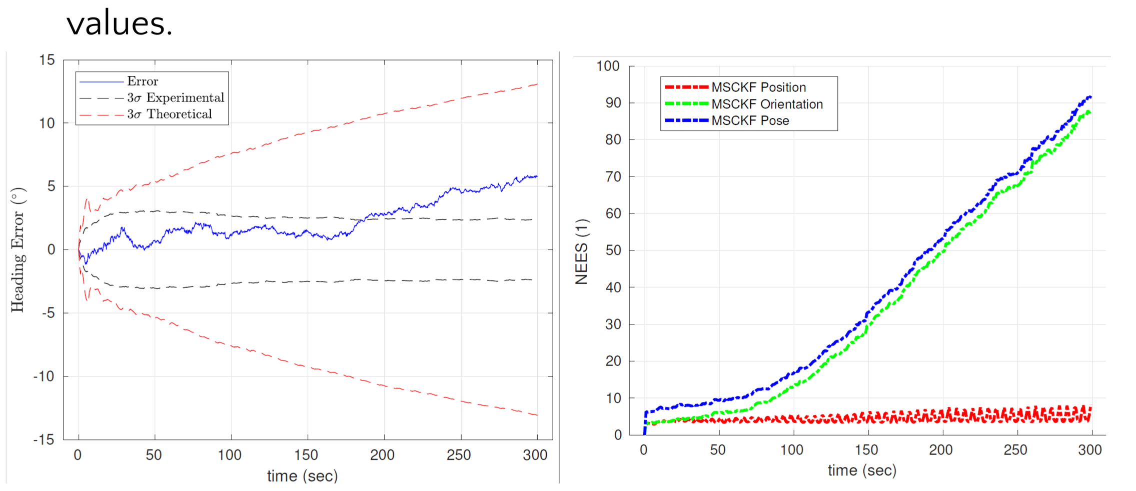
The consistency problem of fixed-lag smoothers (FLSs) has only been attacked by the first estimate Jacobian (FEJ) technique because of the complexity to analyze their observability property. But the FEJ has several drawbacks hampering its wide adoption.

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.

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.

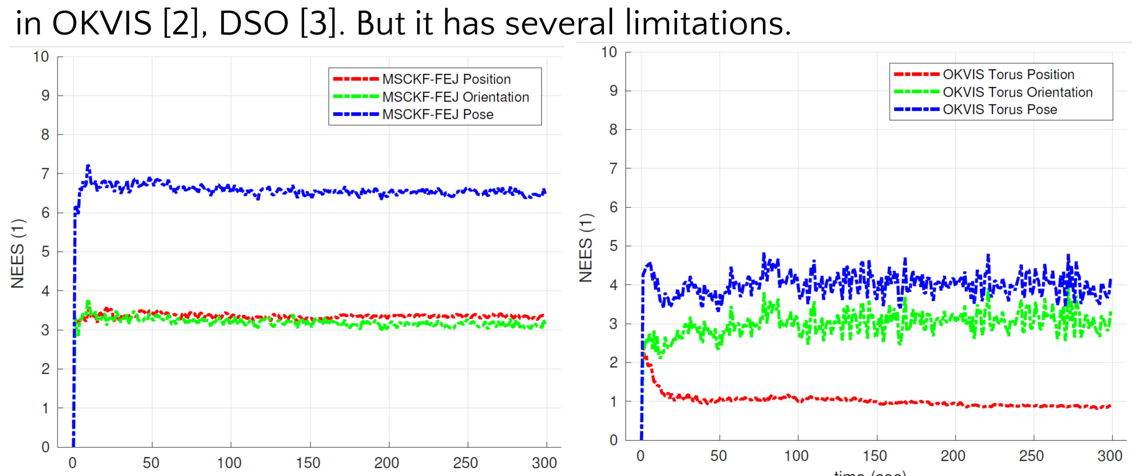
Introduction

1. Estimation with relative measurements has unobservable directions.
2. Traditional geometrical estimators, e.g., filters, and FLSs, gain spurious info along unobservable directions.
3. This leads to rank deficient observability matrices, and inconsistent covariances where standard deviations are less than theoretical values.



(Left) Inconsistent σ and actual errors of heading by a visual inertial odometry method, MSCKF [1], and (right) the NEES (normalized estimation error squared) values for position, orientation, and pose, grow larger than theoretical values, 3 for orientation, and 6 for pose.

Existing approaches to ensure consistency of fixed lag smoothers stem from the first estimate Jacobian (FEJ) technique. For instance, it is used in OKVIS [2], DSO [3]. But it has several limitations.



(Left) MSCKF [1] with the FEJ technique achieves NEES (normalized estimation error squared) close to theoretical values, 3 for position, 3 for orientation, and 6 for pose. (Right) OKVIS [2] also achieves NEES values close to theoretical values with the FEJ technique.

This work proposes to use right invariant formulation to keep consistency of FLSs.

Methodology

We prove that right invariant formulation ensures consistency of FLSs.

1. To minimize the nonlinear cost function E of an estimation problem, it is converted into a linearized system.

$$E = \sum_{>t_0} Nonlinear\ Factors \xrightarrow[\text{at } \mathcal{X}]{\text{Linearize}} E = \mathbf{r}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{r} = (\bar{\mathbf{r}} + \mathbf{J}\delta\mathcal{X})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\bar{\mathbf{r}} + \mathbf{J}\delta\mathcal{X})$$

$$\mathbf{J}\mathbf{N} = \mathbf{0}$$

For visual inertial SLAM problem, rotation about gravity ϕ and global tranalation $\mathbf{t}$ is unobservable

$\mathbf{x}_0, \cdots, \mathbf{x}_k$: navigation variables
$\mathbf{f}_1, \cdots, \mathbf{f}_L$: landmark variables
ϕ: rotation about gravity
$\mathbf{t}$: 3D translation

$$\mathbf{N} = \begin{matrix} \mathbf{x}_0 \\ \vdots \\ \mathbf{x}_k \\ \mathbf{f}_1 \\ \vdots \\ \mathbf{f}_L \\ \mathcal{X} \end{matrix} \begin{matrix} \begin{bmatrix} \mathbf{N}_{\mathbf{x}_0\phi} & \mathbf{N}_{\mathbf{x}_0\mathbf{t}} \\ \vdots & \vdots \\ \mathbf{N}_{\mathbf{x}_k\phi} & \mathbf{N}_{\mathbf{x}_k\mathbf{t}} \\ \mathbf{N}_{\mathbf{f}_1\phi} & \mathbf{N}_{\mathbf{f}_1\mathbf{t}} \\ \vdots & \vdots \\ \mathbf{N}_{\mathbf{f}_L\phi} & \mathbf{N}_{\mathbf{f}_L\mathbf{t}} \end{bmatrix} \\ \phi \quad \mathbf{t} \end{matrix} = [\mathbf{N}_{\phi}(\mathcal{X}) \quad \mathbf{N}_{\mathbf{t}}(\mathcal{X})]$$

2. The fixed-lag smoother marginalizes variables and measurements to bound the problem size. These marginalized measurements are turned into linear factors.

After one marginalization at t_m

$$E = \sum_{[t_0, t_m]} Linear\ Factors + \sum_{>t_m} Nonlinear\ Factors$$

$$\bar{\mathbf{x}}_m(t_m)$$

After another marginalization at t_n

$$E = \sum_{[t_0, t_m]} Linear\ Factors + \sum_{(t_m, t_n]} Linear\ Factors + \sum_{>t_n} Nonlinear\ Factors$$

$$\bar{\mathbf{x}}_m(t_m) \neq \bar{\mathbf{x}}_m(t_n) \qquad \mathbf{J}_2 = \begin{bmatrix} \mathbf{J}_m(\bar{\mathbf{x}}_m(t_m)) \\ \mathbf{J}_n(\bar{\mathbf{x}}_m(t_n)) \\ \mathbf{J}_r \end{bmatrix}$$

For traditional errors, the nullspace column corresponding to rotation about gravity disappears, i.e.,

$$\begin{bmatrix} \mathbf{J}_m(\bar{\mathbf{x}}_m(t_m)) \\ \mathbf{J}_n(\bar{\mathbf{x}}_m(t_n)) \\ \mathbf{J}_r(\mathcal{X}_n) \end{bmatrix} [\mathbf{N}_{\phi}(\mathcal{X}_n) \quad \mathbf{N}_{\mathbf{t}}(\mathcal{X}_n)] = \begin{bmatrix} \mathbf{d}_m & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

3. The local parameters, e.g., landmarks expressed in a host camera frame, and IMU biases, do not affect nullspace matrix.

$$\mathbf{N} = \begin{matrix} \mathbf{x}_0 \\ \vdots \\ \mathbf{x}_k \\ \mathbf{f}_1 \\ \vdots \\ \mathbf{f}_L \end{matrix} \begin{matrix} \begin{bmatrix} \mathbf{N}_{\mathbf{x}_0\phi} & \mathbf{N}_{\mathbf{x}_0\mathbf{t}} \\ \vdots & \vdots \\ \mathbf{N}_{\mathbf{x}_k\phi} & \mathbf{N}_{\mathbf{x}_k\mathbf{t}} \\ \mathbf{N}_{\mathbf{f}_1\phi} & \mathbf{N}_{\mathbf{f}_1\mathbf{t}} \\ \vdots & \vdots \\ \mathbf{N}_{\mathbf{f}_L\phi} & \mathbf{N}_{\mathbf{f}_L\mathbf{t}} \end{bmatrix} \\ \phi \quad \mathbf{t} \end{matrix} = \begin{bmatrix} \mathbf{N}_{\mathbf{x}_0\phi} & \mathbf{N}_{\mathbf{x}_0\mathbf{t}} \\ \vdots & \vdots \\ \mathbf{N}_{\mathbf{x}_k\phi} & \mathbf{N}_{\mathbf{x}_k\mathbf{t}} \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{bmatrix}$$

4. For right invariant errors, the nullspace matrix is independent of the pose and velocity of the agent. It depends on only gravity vector $\mathbf{g}$.

$$[\mathbf{N}_{\mathbf{x}_0\phi} \quad \mathbf{N}_{\mathbf{x}_0\mathbf{t}}] = \begin{bmatrix} \mathbf{g} & 0 \\ 0 & 0 \\ 0 & \mathbf{I}_3 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

5. The RI-FLS is consistent under a mild assumption that the left Jacobians of the pose and velocity residual errors are roughly identity.

$$\begin{bmatrix} \mathbf{J}_m \\ \mathbf{J}_n \\ \mathbf{J}_r(\mathcal{X}_n) \end{bmatrix} [\mathbf{N}_{\phi} \quad \mathbf{N}_{\mathbf{t}}] = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

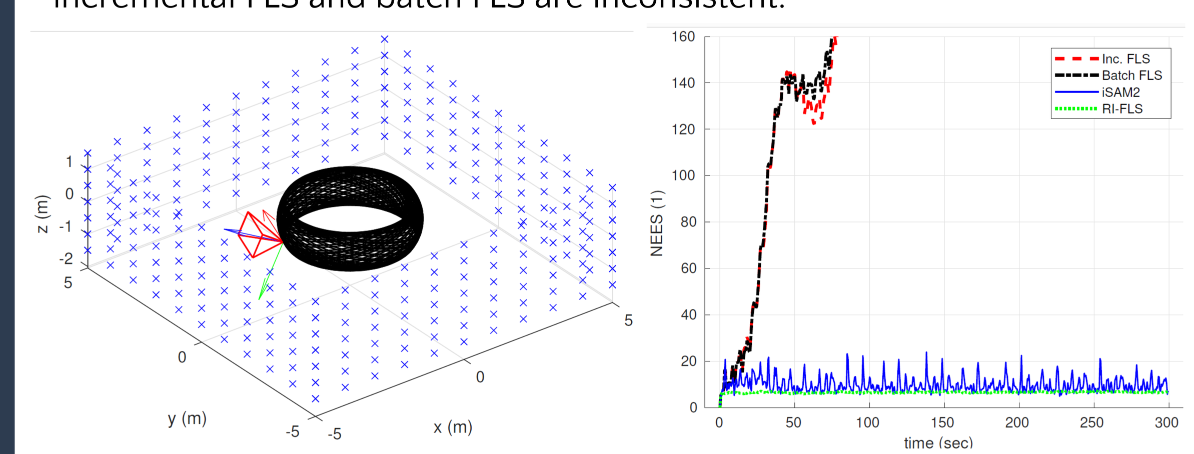
Results

Simulation

The camera and IMU system traverses a virtual room with landmarks on four walls.

For visual inertial SLAM, we compared our RI-FLS, the incremental FLS in GTSAM [4], the batch FLS in GTSAM, and iSAM2 in GTSAM.

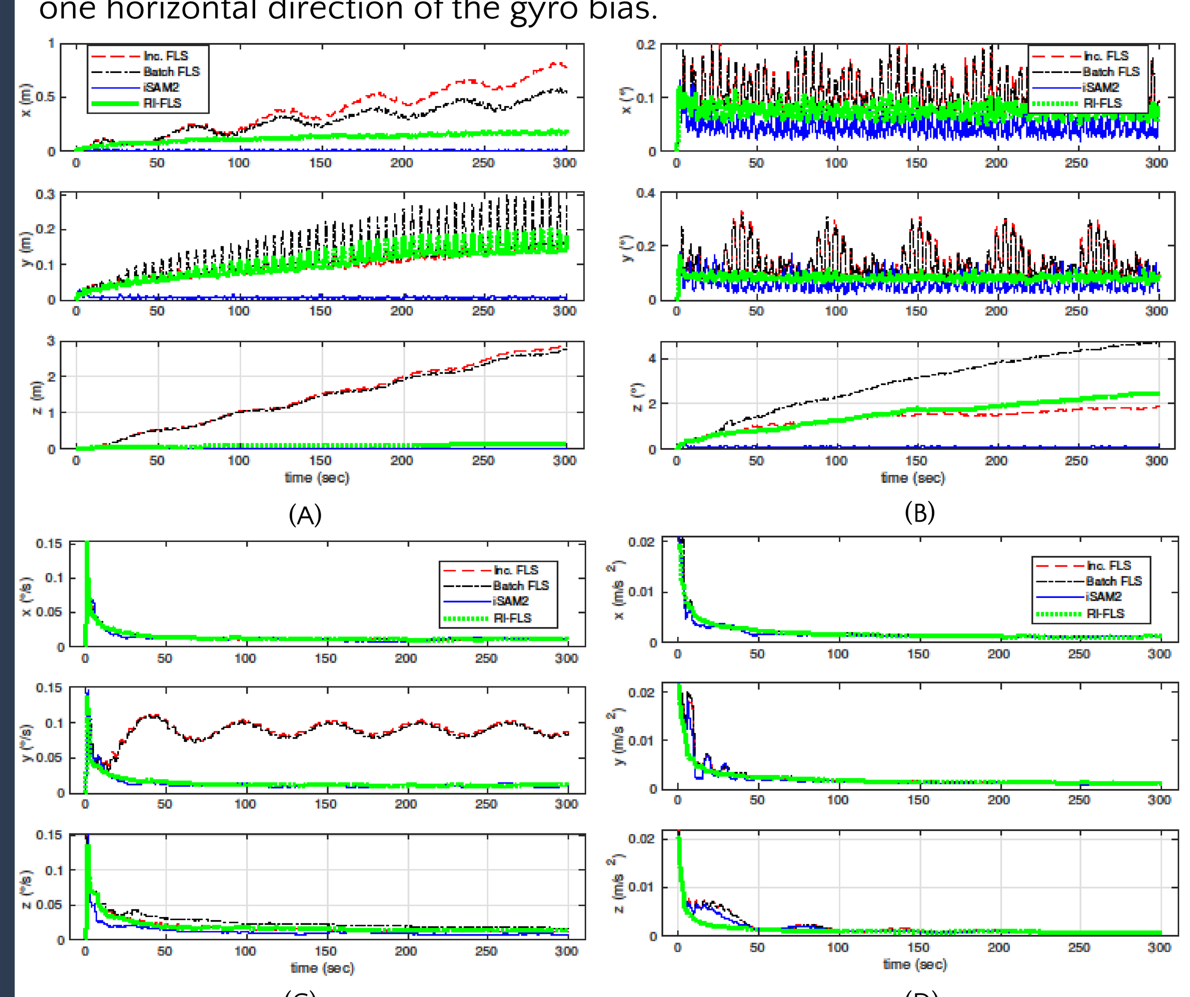
RI-FLS is consistent in terms of NEES similar to iSAM2, while incremental FLS and batch FLS are inconsistent.



(Left) The simulation steup. (Right) RI-FLS achieves NEES values similar to iSAM2, while the incremental FLS and the batch FLS shows inconsistency with growing NEES over time.

In terms of RMSE in each dimension of position, orientation, gyro bias, and accelerometer bias, RI-FLS outperformed other FLSs in position accuracy, and achieved good orientation accuracy.

As expected, 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.

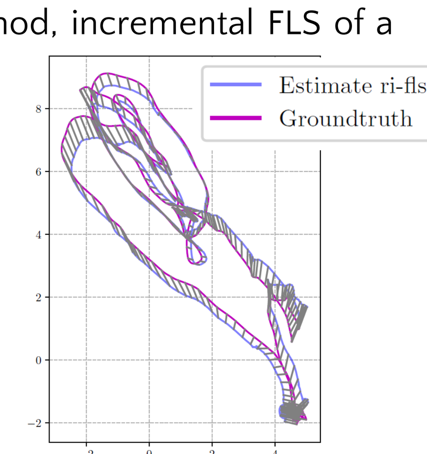


RMSE of position, orientation, gyro bias, and accelerometer bias, computed over 100 runs for estimators including incremental FLS, batch FLS, iSAM2, and right invariant FLS.

On the EuRoC benchmark [5], the proposed RI-FLS achieved comparable accuracy to the established method, incremental FLS of a traditional error formulation.

Absolute translation error (ATE) RMS averaged over 3 runs on several EuRoC sessions for incremental FLS, RI-FLS, and RI-FLS with exact IMU factor Jacobians.

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**



RI-FLS trajectory top view on MH_01

Conclusion

1. We introduce the right invariant error formulation into the FLS framework and analyze its observability directly with the linearized system, which has much lower analysis complexity than observability matrices.
2. As a byproduct, we find that landmarks parameterized in a local camera frame and sensor parameters like biases do not affect the estimator consistency.
3. 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.
4. 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 on the EuRoC benchmark.

Future work

1. Does marginalization cause spurious information to accrue in observable directions?
2. Use the keyframe scheme to improve odometry accuracy.
3. The state errors defined on the Lie group $SE_2(3)$ (which represents position, velocity, and rotation jointly) achieve much better consistency than traditional errors defined on $SO(3) \times \mathbb{R}^3$. But Kontiki [6] has argued that split interpolation on $SO(3) \times \mathbb{R}^3$ is better than joint interpolation on $SE(3)$ for reconstruction. Do the two observations conflict?

References

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

[2] OKVIS. S. Leutenegger, S. Lynen, M. Bosse, R. Siegwart, and P. Furgale, "Keyframe-based visual-inertial odometry using nonlinear optimization," *The International Journal of Robotics Research*, vol. 34, no. 3, pp. 314–334, Mar. 2015.

[3] DSO. J. Engel, V. Koltun, and D. Cremers, "Direct sparse odometry," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 3, pp. 611–625, Mar. 2018.

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

[5] EuRoC. M. Burri, J. Nikolic, P. Gohl, T. Schneider, J. Rehder, S. Omari, M. W. Achtelik, and R. Siegwart, "The EuRoC micro aerial vehicle datasets," *The International Journal of Robotics Research*, vol. 35, no. 10, pp.1157–1163, 2016.

[6] Konkiti. H. Ovrén, and F. Per-Erik. "Trajectory representation and landmark projection for continuous-time structure from motion." *The International Journal of Robotics Research*, vol. 38, no. 6, pp. 686-701, 2019.