

On State Estimation in Multi-Sensor Fusion Navigation: Optimization and Filtering

Feng Zhu, Zhuo Xu, Xveqing Zhang, Yuntai Zhang, Weijie Chen, Xiaohong Zhang

Abstract—The essential of navigation, perception, and decision-making which are basic tasks for intelligent robots, is to estimate necessary system states. Among them, navigation is fundamental for other upper applications, providing precise position and orientation, by integrating measurements from multiple sensors. With observations of each sensor appropriately modelled, multi-sensor fusion tasks for navigation are reduced to the state estimation problem which can be solved by two approaches: optimization and filtering. Recent research has shown that optimization-based frameworks outperform filtering-based ones in terms of accuracy. However, both methods are based on maximum likelihood estimation (MLE) and should be theoretically equivalent with the same linearization points, observation model, measurements, and Gaussian noise assumption. In this paper, we deeply dig into the theories and existing strategies utilized in both optimization-based and filtering-based approaches. It is demonstrated that the two methods are equal theoretically, but this equivalence corrupts due to different strategies applied in real-time operation. By adjusting existing strategies of the filtering-based approaches, the Monte-Carlo simulation and vehicular ablation experiments based on visual odometry (VO) indicate that the strategy adjusted filtering strictly equals to optimization. Therefore, future research on sensor-fusion problems should concentrate on their own algorithms and strategies rather than state estimation approaches.

Index Terms—Multi-sensor fusion, state estimation, factor graph optimization, Kalman filter, visual odometry.

1 INTRODUCTION

Multi-sensor fusion for navigation, perception, and decision-making is a crucial prerequisite for autonomous systems, such as unmanned aerial vehicles (UAVs), autonomous cars, mobile robots, and other emerging applications [1–3], where standalone sensor is insufficient. All these fusion tasks are to extract relevant information from substantial volume of noisy measurements, using them to estimate necessary parameters of the system. Among these, positioning and navigation is a fundamental and indispensable module for other upper-layer applications [4–6]. Wrongly estimated position and orientation of autonomous robots will lead to life and property safety incidents. When measurements from each sensor are extracted and modelled appropriately, the positioning task is boiled down to state estimation problem which can be solved by optimization or filtering.

Essentially, the nature of optimization is least squares (LSQ), which has been widely utilized in solving precise positioning issues using global navigation satellite system (GNSS). The task of GNSS is to estimate position of antennas by signals from satellites, where the orbits and satellite clocks, are firstly generated by LSQ [7–9]. Then, using LSQ, user can achieve precise positioning results with observations appropriately

modelled. Absolute position of stationary GNSS sites, or relative baseline between two arbitrary ones can also be precisely estimated by LSQ. For example, the world-famous scientific GNSS data processing software Bernese, developed by University of Bern [10], and GAMIT, developed by MIT [11], are based on LSQ, which are often used in geodesy.

However, due to vulnerable feature in urban viaducts and other signal blocked areas, the performance of GNSS degrades rapidly, leading to unreliable positioning results [12]. Fortunately, inertial measurement units (IMUs), which are independent of external infrastructure, are integrated with GNSS by extended Kalman filter (EKF), where IMU provides rigorous dynamic model by observing motion of carriers with raw measurements or absolute position of GNSS updating system states [13–16]. For example, Inertial Explorer (IE) and POSPac, which are excellent commercial software used to provide ground truth in navigation research, are developed based on EKF.

Meanwhile, EKF is also applied in simultaneous localization and mapping (SLAM) at an early stage in computer vision research. The first real-time SLAM system, MonoSLAM, is developed based on EKF [17]. Similarly, IMU is incorporated to recover scale and enable large-scale operation. One direct way to integrate measurements from cameras is to jointly estimate current IMU pose as well as visual landmarks, which results in high computational complexity [18,19]. The other way, a more time efficient and profoundly influential framework, adopts a sliding window filtering (SWF) to constrain multiple cameras and project visual bearing measurements onto the null space of the feature Jacobian matrix, rather than estimate them directly, namely MSCKF [20,21]. Based on MSCKF, researchers further utilize mapped landmarks as well as tracked ones in spacecraft landing and descent [22], apply observability constraints to avoid wrong observability property for fast UAVs [23,24]. Besides, other filtering methods, such as unscented Kalman filter (UKF), particle filter (PF) et al., are employed to handle the high degree of nonlinearity and non-Gaussian noise [25–29]. In

• Feng Zhu, Zhuo Xu, Xveqing Zhang, Yuntai Zhang, Weijie Chen, Xiaohong Zhang are with the School of Geodesy and Geomatics, Hubei Luoigia Laboratory, Wuhan University, 129 Luoyu Road, Wuhan 430079, Hubei, China. E-mail: {fzhu, zhuoxu, xveqing, officialtai, wjchen}@whu.edu.cn, xhzhang@sgg.whu.edu.cn.

• Feng Zhu is also with the Key Laboratory of Geospace Environment and Geodesy, Ministry of Education, Wuhan University, 129 Luoyu Road, Wuhan 430079, China.

• Xiaohong Zhang is also with Chinese Antarctic Center of Surveying and Mapping, Wuhan University, 129 Luoyu Road, Wuhan 430079, Hubei, China.

This study was supported by the National Natural Science Foundation of China (Grant No. 42374031), the National Natural Science Foundation of China (Grant No. 42104021), and the Special Fund of Hubei Luoigia Laboratory (Grant No. 2201000038).

(Corresponding author: Feng Zhu)

[30,31], invariant EKF (InEKF) has been introduced to preserve the observable structure of the navigation system. Nevertheless, most matured and successful approaches in filtering have assumed Gaussian noise [32], making MSCKF still the mainstream of filtering-based approaches.

On the other hand, optimization-based approaches have flourished in recent research of computer vision, as there are open-source libraries, such as Ceres, g²o, GTSAM et al., offering abilities of plugin-and-use, outlier removal, and acceleration, where the estimating states and measurements are respectively abstracted into vertices and edges by a graph to intuitively express the state estimation problem [33–36]. Even numerous advantages are featured, the foundation of optimization-based methods still fall under the LSQ. Additionally, although incremental approaches, such as incremental smoothing [34], are applied to reuse previous calculations to improve efficiency, they are still suffered from increasing size of the graph, leading to computational burdens. On that account, sliding window optimization (SWO) is widely adopted to achieve constant processing time as well as to make full use of measurements in SLAM area [5,37–40]. In order to incorporate historic information in the sliding window, SWO utilizes marginalization when window slides, i.e. Schur complement, where information of removed nodes is transferred to its Markov blanket as prior information [41]. Similar to filtering-based approaches, graph of incorporating absolute information from GNSS is constructed, including both state space representation [42,43], and observation space representation [44–46], to mitigate drifts of visual and inertial navigation systems based on SWO.

Comparing research in GNSS positioning with computer vision, it is surprising to note that optimization is preferred in computer vision while filtering is popular in positioning. [32] proposes a combined accuracy indicator to compare EKF with optimization and its experimental results show that optimization outperforms filtering based on SLAM. Other frameworks based on optimization also shows consistent conclusions that optimization performs better compared with filtering, but detailed analysis is not presented [38,43,44,47,48].

However, both filtering and optimization can be derived from maximum likelihood estimation (MLE) by assuming Gaussian noise [49,50], which indicates that optimization-based and filtering-based approaches should be mathematically equivalent under the assumption of the same measurements and observation model. Even though real-time frameworks utilize sliding window to bound computational complexity, the state estimators are still LSQ or EKF. Besides, null-space marginalization is applied in MSCKF, while Schur complement is used in SWO, the equivalence between them has been analytically demonstrated in [51]. Obviously, it is contradictory that both theories and tricks utilized in the two approaches are equivalent, yet not the estimates.

In order to find out reasons causing these discrepancies and realize potential estimating accuracy in the filtering, the two state estimation methods are theoretically and comprehensively analyzed in the paper, including state estimation theory and strategies applied in sliding window estimation, such as marginalization and first estimate Jacobian (FEJ). Then, based on the detailed analysis, existing strategies utilized in filtering-based methods are adjusted by a 2-step update strategy, which shows the same performance with optimization-based ones. To

evaluate the performance of strategy adjusted filtering, a series of Monte-Carlo simulation and vehicular ablation experiments are conducted where data collected by KITTI [52]. Specifically, the main contributions of this work include:

- 1) Discrepancies between optimization and filtering are theoretically analyzed in depth, including prior constraints, measurement utilization, linearization points, and common strategies utilized in real-time estimation, such as marginalization and FEJ. This analysis mathematically demonstrates that the differences are caused by strategies, which further leads to different constraints and measurement utilization, rather than state estimation methods.
- 2) Based on theoretical analysis, existing strategies in the filtering are adjusted by a 2-step update strategy, which fully incorporates constraints from previous window, and FEJ compensation.
- 3) Extensive experiments are performed on both Monte-Carlo simulation and ablation experiments using KITTI dataset based on visual odometry. The results show that the 2-step update strategy makes filtering equal to optimization under the mild assumption of independent and identically distributed (i.i.d.) Gaussian noise and the same measurements utilization when solving the same state estimation problem.

The rest of this paper is structured as follows. In section 2, notations and two alternative representations of the Kalman filter are introduced. In section 3, equivalence and potential discrepancies are discussed between batch estimation and sequential update. Section 4 further analyzes prior constraints, measurement utilization and FEJ compensation in sliding window cases. In section 5, a 2-step modification is proposed to adjust existing strategies in sliding window filtering based on previous analysis. In section 6, both simulation and ablation experiments are performed to validate the preceding analysis and the performance of strategy adjusted filtering-based method. Finally, this paper is concluded in section 7.

2 PRELIMINARIES

In this section, notations used throughout this paper is briefly introduced. And then two representations of the Kalman filter (KF) are mathematically derived, to simplify the subsequent analysis.

In order to analyze state estimation of the optimization-based and filtering-based approaches, assume $\mathbf{X} = [\mathbf{x}_1^T \cdots \mathbf{x}_M^T]^T$ is going to be estimated with M timestamps involved. Specifically, $\hat{\mathbf{x}}_t$ denotes the estimate of a variable \mathbf{x}_t at timestamp t and $\tilde{\mathbf{x}}_t$ denotes an observation of \mathbf{x}_t . The error state $\delta\mathbf{x}_t$ of \mathbf{x}_t is always estimated throughout this paper, which is defined by

$$\delta\mathbf{x}_t = \tilde{\mathbf{x}}_t - \mathbf{x}_t. \quad (1)$$

\mathbf{X} can be estimated only when measurements are all available, where $\mathbf{Z} = \{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_M\}$ denotes measurement set with M timestamps included. In the sliding window cases, $\mathbf{X}_t = [\mathbf{x}_t^T \cdots \mathbf{x}_{t+N}^T]^T$ and $\mathbf{Z}_t = \{\mathbf{z}_t, \mathbf{z}_{t+1}, \dots, \mathbf{z}_{t+N}\}$ denotes N involved states and measurements at timestamp t respectively.

The size of \mathbf{Z} and \mathbf{Z}_t can be expressed by s and s_t . For each measurement \mathbf{z}_t , it can be modelled as

$$\mathbf{z}_t = \mathbf{f}_t(\mathbf{x}_t) + \mathbf{n}_t. \quad (2)$$

Where, $\mathbf{f}_t(\cdot)$ is the measurement function, mapping from system states to measurements. And \mathbf{n}_t denotes the noise vector, modelled as zero-mean Gaussian noise with $\mathbf{n}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{\Lambda}_t^{-1})$.

$\mathbf{\Lambda}_t$ denotes the measurement weight matrix. By employing Taylor-series approximation, this observation model can be approximated as (3) with error state expression.

$$\delta \mathbf{z}_t = \mathbf{h}_t \delta \mathbf{x}_t + \mathbf{o}(\mathbf{x}_t) + \mathbf{n}_t. \quad (3)$$

Where, $\mathbf{o}(\mathbf{x}_t)$ denotes the higher order terms, and \mathbf{h}_t denotes the Jacobian matrix. In the linear case, there only exists first order terms, and $\mathbf{o}(\mathbf{x}_t)$ always equals to $\mathbf{0}$. While, for the nonlinear system, $\mathbf{o}(\mathbf{x}_t)$ could be ignored during state estimation.

Additionally, system states can be predicted by dynamic model, which can be expressed by

$$\mathbf{x}_{t|t-1} = \mathbf{g}_t(\mathbf{x}_{t-1}) + \mathbf{w}_t. \quad (4)$$

Similar with (3), by employing Taylor-series approximation and discretization, the linearized and time-discrete dynamic model can be approximated as (5) with error state expression.

$$\delta \mathbf{x}_{t|t-1} = \Phi_t \delta \mathbf{x}_{t-1} + \mathbf{w}_t. \quad (5)$$

Where, \mathbf{w}_t denotes a noise vector with covariance matrix \mathbf{Q}_t , which is also modelled as Gaussian noise $\mathbf{w}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_t)$. Suppose \mathbf{P}_{t-1} describes the covariance of $\delta \mathbf{x}_{t-1}$, and its inverse \mathbf{N}_{t-1} denotes the information matrix. The uncertainty of $\delta \mathbf{x}_{t|t-1}$ can be derived by:

$$\mathbf{P}_{t|t-1} = \Phi_t \mathbf{P}_{t-1} \Phi_t^T + \mathbf{Q}_t. \quad (6)$$

By regarding $\delta \mathbf{x}_{t|t-1}$ and $\mathbf{P}_{t|t-1}$ as a virtual measurement, $\delta \mathbf{x}_t$ can be estimated through MLE:

$$\delta \hat{\mathbf{x}}_t = \arg \max p(\delta \mathbf{z}_{t-1,p}, \delta \mathbf{z}_t | \delta \mathbf{x}_t). \quad (7)$$

Where, the virtual measurement $\delta \mathbf{z}_{t-1,p}$ has the following form with $\mathbf{w}_{t-1,p} \sim \mathcal{N}(\mathbf{0}, \mathbf{P}_{t|t-1})$.

$$\delta \mathbf{z}_{t-1,p} = \delta \mathbf{x}_{t|t-1} + \mathbf{w}_{t-1,p}. \quad (8)$$

Under Gaussian noise assumption, this MLE can be solved by LSQ, where the principal is to minimize

$$\begin{bmatrix} \delta \mathbf{z}_{t-1,p} \\ \delta \mathbf{z}_t \end{bmatrix}^T \begin{bmatrix} \mathbf{P}_{t|t-1}^{-1} \\ \mathbf{\Lambda}_t \end{bmatrix} \begin{bmatrix} \delta \mathbf{z}_{t-1,p} \\ \delta \mathbf{z}_t \end{bmatrix} = \min. \quad (9)$$

Considering (3) and (8), LSQ (10) can be further derived from (9), which also denotes the information filter.

$$\delta \hat{\mathbf{x}}_t = (\mathbf{P}_{t|t-1}^{-1} + \mathbf{h}_t^T \mathbf{\Lambda}_t \mathbf{h}_t)^{-1} (\mathbf{P}_{t|t-1}^{-1} \delta \mathbf{x}_{t|t-1} + \mathbf{h}_t^T \mathbf{\Lambda}_t \mathbf{l}_t). \quad (10)$$

Where, \mathbf{l}_t denotes the residual related part, which has the following form

$$\mathbf{l}_t = \mathbf{z}_t - \mathbf{f}(\hat{\mathbf{x}}_{t|t-1}), \quad (11)$$

$$\mathbf{b}_t = \mathbf{P}_{t|t-1}^{-1} \delta \mathbf{x}_{t|t-1} + \mathbf{h}_t^T \mathbf{\Lambda}_t \mathbf{l}_t.$$

And, $\hat{\mathbf{x}}_{t|t-1} = \mathbf{g}_t(\hat{\mathbf{x}}_{t-1})$ describes the propagated system states.

Additionally, the posteriori of the LSQ can be derived by

$$\mathbf{N}_t = \mathbf{P}_{t|t-1}^{-1} + \mathbf{h}_t^T \mathbf{\Lambda}_t \mathbf{h}_t = \mathbf{P}_t^{-1}. \quad (12)$$

Then, (10) can be rewritten as

$$\delta \mathbf{x}_t = \mathbf{P}_t \left((\mathbf{P}_{t|t-1}^{-1} - \mathbf{h}_t^T \mathbf{\Lambda}_t \mathbf{h}_t) \delta \mathbf{x}_{t|t-1} + \mathbf{h}_t^T \mathbf{\Lambda}_t \mathbf{l}_t \right). \quad (13)$$

Thus, we have

$$\delta \hat{\mathbf{x}}_t = \delta \mathbf{x}_{t|t-1} + \mathbf{P}_t \mathbf{h}_t^T \mathbf{\Lambda}_t (\mathbf{l}_t - \mathbf{h}_t \delta \mathbf{x}_{t|t-1}). \quad (14)$$

Considering

$$(\mathbf{D}^{-1} + \mathbf{C}^T \mathbf{A} \mathbf{C})^{-1} \mathbf{C}^T \mathbf{A} = \mathbf{D} \mathbf{C}^T (\mathbf{A}^{-1} + \mathbf{C} \mathbf{D}^{-1} \mathbf{C}^T)^{-1}, \quad (15)$$

coefficient in (14) can be further denoted as

$$\begin{aligned} \mathbf{P}_t \mathbf{h}_t^T \mathbf{\Lambda}_t &= (\mathbf{P}_{t|t-1}^{-1} + \mathbf{h}_t^T \mathbf{\Lambda}_t \mathbf{h}_t)^{-1} \mathbf{h}_t^T \mathbf{\Lambda}_t \\ &= \mathbf{P}_{t|t-1} \mathbf{h}_t^T (\mathbf{\Lambda}_t^{-1} + \mathbf{h}_t \mathbf{P}_{t|t-1} \mathbf{h}_t^T)^{-1} \triangleq \mathbf{K}_t. \end{aligned} \quad (16)$$

And the posteriori covariance matrix can be rewritten as:

$$\begin{aligned} \mathbf{P}_t &= (\mathbf{P}_{t|t-1}^{-1} + \mathbf{h}_t^T \mathbf{\Lambda}_t \mathbf{h}_t)^{-1} \\ &= \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1} \mathbf{h}_t^T (\mathbf{\Lambda}_t^{-1} + \mathbf{h}_t \mathbf{P}_{t|t-1} \mathbf{h}_t^T)^{-1} \mathbf{h}_t \mathbf{P}_{t|t-1} \\ &= (\mathbf{I} - \mathbf{K}_t \mathbf{h}_t) \mathbf{P}_{t|t-1}. \end{aligned} \quad (17)$$

(14), (16), and (17) form the formula of the KF, which is derived from (10). Thus, (10) and (14) are two alternative expressions of KF, namely information filtering representation (IFR) and Kalman filtering representation (KFR). Note that even the two expressions are equivalent, their time complexity is dependent on matrix inverse operation, which is approximately $\mathcal{O}(n^3)$ in IFR, and $\mathcal{O}(m^3)$ in KFR, where n and m are number of system states and measurements respectively.

Besides, KF only estimates states at a specific timestamp, but multiple variables from different timestamps can be jointly estimated in order to incorporate more information from measurements. Thus, batch estimation and sliding window strategy will be analyzed in the subsequent sections, which can be expressed with the two equivalent expressions as well. Then, (18) denotes the stacked Jacobi and residual vector in batch estimation and (19) denotes the counterparts in sliding window cases.

$$\mathbf{H} = [\mathbf{h}_1^T \quad \cdots \quad \mathbf{h}_M^T]^T, \mathbf{L} = [\mathbf{l}_1^T \quad \cdots \quad \mathbf{l}_M^T]^T, \quad (18)$$

$$\mathbf{H}_t = [\mathbf{h}_t^T \quad \cdots \quad \mathbf{h}_{t+N}^T]^T, \mathbf{L}_t = [\mathbf{l}_t^T \quad \cdots \quad \mathbf{l}_{t+N}^T]^T. \quad (19)$$

When a specific state estimation problem is expressed with either IFR or KFR, it is equivalent to the other representation. Thus, the two representations establish a bridge to analyze the equivalence between the optimization and filtering, which will be utilized to dig into the reasons that cause discrepancies between the optimization-based and filtering-based approaches in the subsequent analysis.

3 THEORETICAL ANALYSES ON OPTIMIZATION-BASED AND FILTERING-BASED APPROACHES

In this section, equivalence between batch estimation and se-

quential update is firstly analyzed with a linear system, which ensures the same Jacobian matrix. Nonlinear system is then analyzed with iteration strategy applied.

3.1 Equivalence Between Optimization and Filtering in Linear System

Toward the state estimation problem with measurement set \mathbf{Z} in a linear system, the optimal solution is to solve the full-state MLE problem, which jointly estimates all variables

$$\delta \hat{\mathbf{X}} = \arg \max p(\delta \mathbf{Z}_p, \delta \mathbf{Z} | \delta \mathbf{X}). \quad (20)$$

Similarly, regarding prior knowledge $\delta \mathbf{X}_0$ as a virtual measurement with noise vector $\mathbf{w}_p \sim \mathcal{N}(\mathbf{0}, \mathbf{P}_0)$, this MLE can be solved by full-state LSQ (21).

$$\delta \hat{\mathbf{X}} = (\mathbf{P}_0^{-1} + \mathbf{H}^T \mathbf{\Lambda} \mathbf{H})^{-1} (\mathbf{P}_0^{-1} \delta \mathbf{X}_0 + \mathbf{H}^T \mathbf{\Lambda} \mathbf{L}). \quad (21)$$

Since all measurements are stacked to estimate states, it is also known as batch estimation. In practice, measurements are often time-irrelevant, such as GNSS pseudorange or phase measurements, visual feature points, and LiDAR scans et al. Thus, weight matrix $\mathbf{\Lambda}$ satisfies:

$$\mathbf{\Lambda} = \text{diag} \{ \mathbf{\Lambda}_1 \quad \cdots \quad \mathbf{\Lambda}_M \}. \quad (22)$$

Then, considering (18), batch estimation can be rewritten as

$$\left(\mathbf{P}_0^{-1} + \sum_{i=1}^M \mathbf{h}_i^T \mathbf{\Lambda}_i \mathbf{h}_i \right) \delta \hat{\mathbf{X}} = \mathbf{P}_0^{-1} \delta \mathbf{X}_0 + \sum_{i=1}^M \mathbf{h}_i^T \mathbf{\Lambda}_i \mathbf{l}_i. \quad (23)$$

Besides, for each measurement in \mathbf{Z} , system states can be sequentially estimated by:

$$\begin{aligned} \delta \hat{\mathbf{X}}_1 &= (\mathbf{P}_0^{-1} + \mathbf{h}_1^T \mathbf{\Lambda}_1 \mathbf{h}_1)^{-1} (\mathbf{P}_0^{-1} \delta \mathbf{X}_0 + \mathbf{h}_1^T \mathbf{\Lambda}_1 \mathbf{l}_1) \\ \delta \hat{\mathbf{X}}_2 &= (\mathbf{P}_1^{-1} + \mathbf{h}_2^T \mathbf{\Lambda}_2 \mathbf{h}_2)^{-1} (\mathbf{P}_1^{-1} \delta \hat{\mathbf{X}}_1 + \mathbf{h}_2^T \mathbf{\Lambda}_2 \mathbf{l}_2) \\ &\vdots \\ \delta \hat{\mathbf{X}}_M &= (\mathbf{P}_{M-1}^{-1} + \mathbf{h}_M^T \mathbf{\Lambda}_M \mathbf{h}_M)^{-1} (\mathbf{P}_{M-1}^{-1} \delta \hat{\mathbf{X}}_{M-1} + \mathbf{h}_M^T \mathbf{\Lambda}_M \mathbf{l}_M), \end{aligned} \quad (24)$$

which is equivalent to the full-state linear KF (LKF) expressed by IFR. Considering the posteriori of the final estimates $\delta \hat{\mathbf{X}}_M$, we have

$$\begin{aligned} \mathbf{P}_M^{-1} &= \mathbf{P}_{M-1}^{-1} + \mathbf{h}_M^T \mathbf{\Lambda}_M \mathbf{h}_M \\ &= \mathbf{P}_{M-2}^{-1} + \mathbf{h}_{M-1}^T \mathbf{\Lambda}_{M-1} \mathbf{h}_{M-1} + \mathbf{h}_M^T \mathbf{\Lambda}_M \mathbf{h}_M \\ &= \mathbf{P}_0^{-1} + \sum_{i=1}^M \mathbf{h}_i^T \mathbf{\Lambda}_i \mathbf{h}_i, \end{aligned} \quad (25)$$

and the residual related part can be rewritten as

$$\begin{aligned} \mathbf{b}_M &= \mathbf{P}_{M-1}^{-1} \delta \hat{\mathbf{X}}_{M-1} + \mathbf{h}_M^T \mathbf{\Lambda}_M \mathbf{l}_M \\ &= \mathbf{P}_{M-2}^{-1} \delta \hat{\mathbf{X}}_{M-2} + \mathbf{h}_{M-1}^T \mathbf{\Lambda}_{M-1} \mathbf{l}_{M-1} + \mathbf{h}_M^T \mathbf{\Lambda}_M \mathbf{l}_M \\ &= \mathbf{P}_0^{-1} \delta \mathbf{X}_0 + \sum_{i=1}^M \mathbf{h}_i^T \mathbf{\Lambda}_i \mathbf{l}_i. \end{aligned} \quad (26)$$

Thus, final estimates in sequential update can be rewritten as (27), which shares the same mathematical form with the batch estimation.

$$\delta \hat{\mathbf{X}}_M = \left(\mathbf{P}_0^{-1} + \sum_{i=1}^M \mathbf{h}_i^T \mathbf{\Lambda}_i \mathbf{h}_i \right)^{-1} \left(\mathbf{P}_0^{-1} \delta \mathbf{X}_0 + \sum_{i=1}^M \mathbf{h}_i^T \mathbf{\Lambda}_i \mathbf{l}_i \right). \quad (27)$$

Proposition 1: The full-state MLE problem can be solved by batch estimation, or full-state LSQ (21), under Gaussian assumption. Supposing measurements are time-irrelevant, final estimates of sequential update, or full-state LKF (27), further equals to batch estimation (21) as well.

3.2 Jacobi Inconsistency in Nonlinear System

Due to linearization error introduced by Taylor-series approximation, an iterative approach should be applied. For i -th iteration at timestamp t , system states can be estimated by

$$\delta \hat{\mathbf{x}}_t^{(i)} = \left(\mathbf{P}_{t|t-1}^{-1} + \mathbf{h}_t^{(i)T} \mathbf{\Lambda}_t \mathbf{h}_t^{(i)} \right)^{-1} \left(\mathbf{P}_{t|t-1}^{-1} \delta \hat{\mathbf{x}}_t^{(i-1)} + \mathbf{h}_t^{(i)T} \mathbf{\Lambda}_t \mathbf{l}_t^{(i)} \right), \quad (28)$$

which denotes the nonlinear squares (NLSQ) and where $\delta \hat{\mathbf{x}}_t^{(0)} = \delta \mathbf{x}_{t|t-1}$ for the first iteration.

Meanwhile, the optimal solution of the nonlinear system can be obtained by jointly estimating all variables as well. The batch estimation solution of i -th iteration can be obtained by

$$\delta \hat{\mathbf{X}}^{(i)} = \left(\mathbf{P}_0^{-1} + \mathbf{H}^{(i)T} \mathbf{\Lambda} \mathbf{H}^{(i)} \right)^{-1} \left(\mathbf{P}_0^{-1} \delta \mathbf{X}^{(i-1)} + \mathbf{H}^{(i)T} \mathbf{\Lambda} \mathbf{L}^{(i)} \right). \quad (29)$$

Different from linear system, the Jacobian matrix is affected by system states, which is evaluated at the estimates of previous iteration:

$$\mathbf{H}^{(i)} = \begin{bmatrix} \mathbf{h}_1^{(i)T} \big|_{\mathbf{X}=\hat{\mathbf{X}}^{(i-1)}} & \cdots & \mathbf{h}_M^{(i)T} \big|_{\mathbf{X}=\hat{\mathbf{X}}^{(i-1)}} \end{bmatrix}^T. \quad (30)$$

With i.i.d. Gaussian noise assumption, where (22) is satisfied, batch estimation in nonlinear system (29) can be rewritten as

$$\left(\mathbf{P}_0^{-1} + \sum_{i=1}^M \mathbf{h}_i^{(i)T} \mathbf{\Lambda}_i \mathbf{h}_i^{(i)} \right) \delta \hat{\mathbf{X}}^{(i)} = \mathbf{P}_0^{-1} \delta \mathbf{X}^{(i-1)} + \sum_{i=1}^M \mathbf{h}_i^{(i)T} \mathbf{\Lambda}_i \mathbf{l}_i^{(i)}. \quad (31)$$

Then by analogy with (24), the final estimates after M -th measurement using sequential update can be derived by

$$\delta \mathbf{X}_M^{(i)} = \left(\mathbf{P}_{M-1}^{-1} + \mathbf{h}_M^{(i)T} \mathbf{\Lambda}_M \mathbf{h}_M^{(i)} \right)^{-1} \left(\mathbf{P}_{M-1}^{-1} \delta \hat{\mathbf{X}}_{M-1}^{(i-1)} + \mathbf{h}_M^{(i)T} \mathbf{\Lambda}_M \mathbf{l}_M^{(i)} \right), \quad (32)$$

which also denotes the full-state iterative extended Kalman filter (IEKF) expressed by IFR and theoretically equals to the full-state NLSQ, i.e. batch estimation in nonlinear system (29). However, the sequential update iterates (32) until it converges at each measurement, and then compensates the estimates. Thus, Jacobi of next coming measurement is evaluated at previous estimates instead of initial value. For analysis, by stacking Jacobian matrices of each measurement update at i -th iteration in the sequential update, we have:

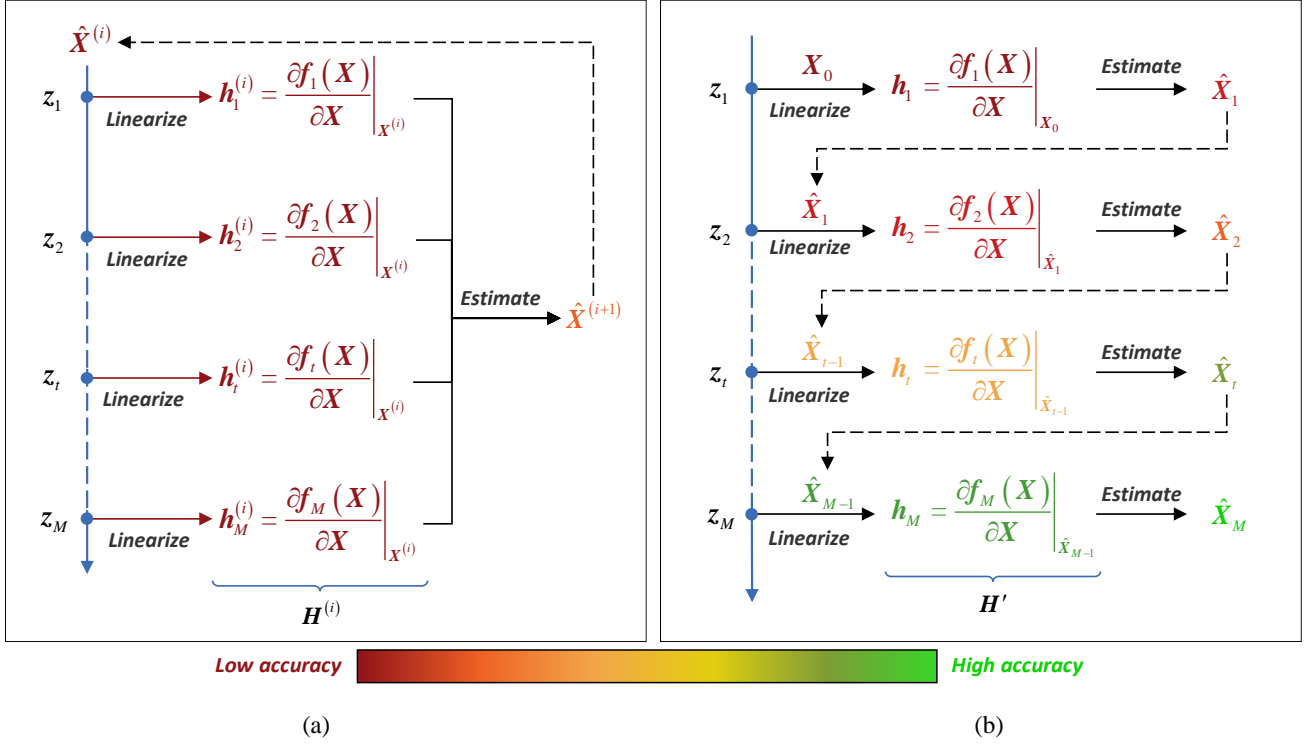


Fig. 1. Illustration of the Jacobian inconsistency using batch estimation and sequential update in a nonlinear system. Batch estimation evaluates Jacobian matrix at the same linearization points, which is consistent (a). On the contrary, the sequential update iterates and compensates at each measurement, leading to different linearization points (b). Thus, the Jacobi of sequential update is inconsistent. Color represents the accuracy of linearization points and the estimates.

$$\mathbf{H}^{(i)} = \begin{bmatrix} \mathbf{h}_1^{(i)T} /_{\mathbf{X}=\hat{\mathbf{X}}_1^{(i-1)}} & \cdots & \mathbf{h}_M^{(i)T} /_{\mathbf{X}=\hat{\mathbf{X}}_M^{(i-1)}} \end{bmatrix}^T. \quad (33)$$

Where, $\hat{\mathbf{X}}_t^{(i-1)}$ denotes the estimates using t -th measurement after iteration $i-1$. Compared (30) with (33), sequential update corrupts the consistency of the Jacobian matrix at the view of linearization points, as illustrated in Fig. 1. This inconsistency leads to wrong observations $\Delta \mathbf{H}$ in the Jacobi and estimates $\Delta \mathbf{X}$, as in (34).

$$\begin{aligned} \mathbf{Z} &= \mathbf{H}^{(i)} \hat{\mathbf{X}}^{(i)} = \mathbf{H}^{(i)} \hat{\mathbf{X}}^{(i)} \\ &= (\mathbf{H}^{(i)} + \Delta \mathbf{H}) (\hat{\mathbf{X}}^{(i)} + \Delta \mathbf{X}) \end{aligned} \quad (34)$$

Thus, $\Delta \mathbf{X}$ can be expressed by the inconsistency of the Jacobi, which implies that wrongly estimated states are severely affected by linearization points.

$$\Delta \mathbf{X} = -(\mathbf{H}^{(i)} + \Delta \mathbf{H})^{-g} \Delta \mathbf{H} \hat{\mathbf{X}}^{(i)} \quad (35)$$

Where, $-g$ is generalized inverse operation.

In order to achieve optimal estimates and avoid wrong observations introduced by inconsistency of the Jacobi. The feasible strategy is to update states measurement by measurement, which is regarded as once iteration (see Algorithm 1). Then, the Jacobi and residual vector are evaluated at compensated estimates in the next iteration, which ensures the same Jacobian matrix and residual vector with (33).

Algorithm 1 Full-state Sequential Update Using KFR

set $i \leftarrow 0, t \leftarrow 1, \text{MaxIter}, \delta \mathbf{X}_0^{(0)} \leftarrow \delta \mathbf{X}^{(0)}, \mathbf{P}_0$

while $i < \text{MaxIter}$

while $t \leq M$

// evaluate Jacobi and residual at $\mathbf{X}^{(i)}$

$$\mathbf{h}_t^{(i)} \leftarrow \left. \frac{\partial f_t(\mathbf{X})}{\partial \mathbf{X}} \right|_{\mathbf{X}^{(i)}}, \mathbf{l}_t^{(i)} \leftarrow \mathbf{z}_t - f_t(\mathbf{X}^{(i)})$$

// full-state sequential update

$$\mathbf{K}_t^{(i)} \leftarrow \mathbf{P}_{t-1} \mathbf{h}_t^{(i)T} (\mathbf{\Lambda}_t^{-1} + \mathbf{h}_t^{(i)} \mathbf{P}_{t-1} \mathbf{h}_t^{(i)T})^{-1} \quad (36)$$

$$\delta \mathbf{X}_t^{(i)} \leftarrow \delta \mathbf{X}_{t-1}^{(i)} + \mathbf{K}_t^{(i)} (\mathbf{l}_t^{(i)} - \mathbf{h}_t^{(i)} \delta \mathbf{X}_{t-1}^{(i)}) \quad (37)$$

$$\mathbf{P}_t \leftarrow (\mathbf{I} - \mathbf{K}_t^{(i)} \mathbf{h}_t^{(i)}) \mathbf{P}_{t-1} \quad (38)$$

end while

$$\hat{\mathbf{X}}_0^{(i+1)} \leftarrow \mathbf{X}_0^{(i)} - \delta \mathbf{X}_M^{(i)}$$

if converge

$$\text{return } \hat{\mathbf{X}}_0^{(i+1)}, \mathbf{P}_M$$

end if

$$i \leftarrow i + 1$$

end while

Proposition 2: For nonlinear system, NLSQ (29) and IEKF (32) are theoretically equivalent as well, although the Jacobian matrix is dependent on linearization points. Due to iteration strategies, estimating states using all measurements, compensating the estimates and then doing next iteration in sequential update (37) should be applied, which remains the consistency of the Jacobi compared with batch estimation. Instead, sequential

update iterates and compensates at each measurement (32), which leads to Jacobian inconsistency (35).

Based on previous analysis, some conceptions can be clarified here. Batch estimation, i.e. full-state LSQ or NLSQ, falls under optimization, where they mathematically share the same formula but iteration strategy is applied in nonlinear cases. Meanwhile, sequential update, i.e. full-state LKF or IEKF, is a category of filtering. But note that the EKF is different from IEKF, where IEKF adopts iteration strategy while EKF not. Thus, IEKF equals to NLSQ but EKF not.

4 SLIDING WINDOW STRATEGY APPLIED IN OPTIMIZATION-BASED AND FILTERING-BASED APPROACHES

In this section, discrepancies caused by sliding window strategy, which is used to balance computational burdens with the accuracy, are discussed. Since sliding window optimization (SWO) utilizes information matrix while sliding window filtering (SWF) operates on covariance matrix in practice, IFR and KFR are utilized respectively in SWO and SWF for analysis.

First, estimates in the current window t are analyzed. The MLE problem of the SWO or SWF at current window can be described by:

$$\delta \hat{\mathbf{X}}_t = \arg \max p(\delta \mathbf{Z}_{t,p}, \delta \mathbf{Z}_t | \delta \hat{\mathbf{X}}_t). \quad (39)$$

Compared with (20), SWO or SWF is a case of batch estimation, where states in the window are jointly estimated. Therefore, parameters in the current window can be estimated equivalently, i.e. $\hat{\mathbf{X}}_t$, provided the same initial values of parameters, measurements, observation model, and stochastic model in both estimation methods according to proposition 2.

After attaining the estimates, window will slide to maintain the same computational burdens. Since state estimation is equivalent in both SWO and SWF, sliding window strategies are concentrated, where we mathematically show how states are estimated at the next window $t+1$.

In the following analysis, suppose that \mathbf{x}_m and its measurements are to be removed from window in both SWO and SWF. And \mathbf{x}_b denotes all direct neighboring parameters of \mathbf{x}_m , also known as Markov blanket of \mathbf{x}_m (see Fig. 2).

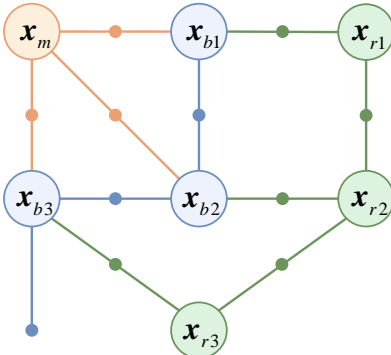


Fig. 2. Illustration of the Markov blanket of \mathbf{x}_m which will be removed from window. Edges denote observations between nodes and nodes denote the system states. \mathbf{x}_m (orange) is directly connected to $\mathbf{x}_b = [\mathbf{x}_{b1}^T \ \mathbf{x}_{b2}^T \ \mathbf{x}_{b3}^T]^T$ by observations. Thus, \mathbf{x}_b (blue) is the Markov

blanket of \mathbf{x}_m . Other states are denoted by \mathbf{x}_r (green).

4.1 Sliding Window Operation in SWO

SWO utilizes all measurements to estimate states in the current window, but it only converts measurements of \mathbf{x}_m into the posteriori (see Fig. 3). Thus, SWO only considers measurements connected to \mathbf{x}_m , denoted by $\mathbf{Z}_{t,m}$, when window slides, which can be expressed by

$$\underbrace{\left(\mathbf{P}_{t|t-1}^{-1} + \sum_{i=t}^{t+S_{t,m}} \mathbf{h}_i^T \mathbf{\Lambda}_i \mathbf{h}_i \right)}_{\mathbf{N}_{t,m}} \delta \hat{\mathbf{X}}_{t,m} = \underbrace{\mathbf{P}_{t|t-1}^{-1} \delta \mathbf{X}_{t|t-1}}_{\mathbf{b}_{t,m}} + \sum_{i=t}^{t+S_{t,m}} \mathbf{h}_i^T \mathbf{\Lambda}_i \mathbf{l}_i. \quad (40)$$

The superscript of iteration number i is omitted for simplicity. Note that measurements in $\mathbf{Z}_{t,m}$ connect \mathbf{x}_m and \mathbf{x}_b , but are irrelevant to \mathbf{x}_r , so that $\mathbf{N}_{t,m}$ has the following form:

$$\mathbf{N}_{t,m} = \begin{bmatrix} \mathbf{N}_{mm} & \mathbf{N}_{mb} \\ \mathbf{N}_{mb}^T & \mathbf{N}_{bb} \end{bmatrix}. \quad (41)$$

In order to remove \mathbf{x}_m from window without information loss, marginalization is carried out by Schur complement over the Markov blanket [53]:

$$\begin{aligned} \begin{bmatrix} \mathbf{N}_{mm} & \mathbf{N}_{mb} \\ \mathbf{N}_{mb}^T & \mathbf{N}_{bb} \end{bmatrix} \begin{bmatrix} \delta \mathbf{x}_{t,m} \\ \delta \mathbf{x}_{t,b} \end{bmatrix} &= \begin{bmatrix} \mathbf{b}_m \\ \mathbf{b}_b \end{bmatrix} \\ \Rightarrow \underbrace{(\mathbf{N}_{bb} - \mathbf{N}_{mb}^T \mathbf{N}_{mm}^{-1} \mathbf{N}_{mb})}_{\mathbf{N}_{t,b}} \delta \mathbf{x}_{t,b} &= \underbrace{\mathbf{b}_b - \mathbf{N}_{mb}^T \mathbf{N}_{mm}^{-1} \mathbf{b}_m}_{\mathbf{b}_{t,b}}. \end{aligned} \quad (42)$$

Considering

$$\delta \hat{\mathbf{x}}_{t,b} = \mathbf{N}_{t,b}^{-1} \mathbf{b}_{t,b}, \quad (43)$$

variable $\mathbf{x}_{t,b}$ is constrained by $\delta \hat{\mathbf{x}}_{t,b}$ and $\mathbf{N}_{t,b}^{-1}$ in the window $t+1$.

Additionally, remind that $\mathbf{N}_{t,b}^{-1}$ is dependent on linearization points of \mathbf{x}_m and \mathbf{x}_b due to nonlinearity, where it is evaluated at the estimates $\hat{\mathbf{X}}_t$. Furthermore, $\mathbf{N}_{t,b}^{-1}$ consists of measurements of \mathbf{x}_m , which have been removed. Thus, $\mathbf{N}_{t,b}^{-1}$ is unable to be updated with the newest estimates at window $t+1$, i.e. linearization points are fixed. However, \mathbf{x}_b is still estimated in the next window, which leads to linearization points of the measurement Jacobian matrix and residual vector changed. Thereafter, linearization points of the same variables are different compared the priori with the Jacobi and residuals, resulting in wrong observability of the system [54]. Thus, linearization points should remain the fixed instead of the updated, known as first estimate Jacobian (FEJ). Obviously, FEJ may lead to large linearization error, which can be represented by the difference between current estimates and fixed linearization points of the variable \mathbf{x}_b :

$$\Delta \mathbf{x}_{t+1,b} = \hat{\mathbf{x}}_{t+1,b} - \hat{\mathbf{x}}_{t,b}. \quad (44)$$

Fortunately, the first order of this linearization error can be compensated by regarding $\mathbf{b}_{t,b}$ as a function of \mathbf{x}_b [38], which has the following form

$$\Delta \mathbf{b}_{t+1,b} = \mathbf{N}_{t,b} \Delta \mathbf{x}_{t+1,b}. \quad (45)$$

It shows that $\Delta \mathbf{x}_{t+1,b}$ is also constrained by $\mathbf{N}_{t,b}^{-1}$. Then, combining (43), $\delta \hat{\mathbf{x}}_{t,b} + \Delta \mathbf{x}_{t+1,b}$ and $\mathbf{N}_{t,b}^{-1}$ can be regarded as a virtual measurement (46) with $\mathbf{w}_{t,p} \sim \mathcal{N}(\mathbf{0}, \mathbf{N}_{t,b}^{-1})$.

$$\delta \mathbf{z}_{t,p} = \delta \hat{\mathbf{x}}_{t,b} + \Delta \mathbf{x}_{t+1,b} + \mathbf{w}_{t,p}. \quad (46)$$

Note that only \mathbf{x}_b , the Markov blanket of \mathbf{x}_m , is constrained in the window $k+1$ and other unconstrained variables are theoretically equivalent to being constrained by infinite covariance. Therefore, these constraints can be expressed by

$$\mathbf{P}_{t+1|t} = \begin{bmatrix} \mathbf{N}_{t,b}^{-1} & & \\ & \infty \mathbf{I} & \\ & & \infty \mathbf{I} \end{bmatrix}, \delta \mathbf{X}_{t+1|t} = \begin{bmatrix} \delta \hat{\mathbf{x}}_{t,b} + \Delta \mathbf{x}_{t+1,b} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}. \quad (47)$$

Where, variables in the next window are defined by

$$\mathbf{X}_{t+1} = [\mathbf{x}_b^T \quad \mathbf{x}_r^T \quad \mathbf{x}_n^T]^T, \quad (48)$$

and \mathbf{x}_n denotes newly coming variables after window slides.

By utilizing measurements in \mathbf{Z}_{t+1} and considering the virtual measurements, variables in the window $k+1$ can be optimized by (49) expressed with IFR.

$$\left(\mathbf{P}_{t+1|t}^{-1} + \sum_{i=t+1}^{t+S_{t+1}} \mathbf{h}_i^T \mathbf{\Lambda}_i \mathbf{h}_i \right) \delta \hat{\mathbf{X}}_{t+1} = \mathbf{P}_{t+1|t}^{-1} \delta \mathbf{X}_{t+1|t} + \sum_{i=t+1}^{t+S_{t+1}} \mathbf{h}_i^T \mathbf{\Lambda}_i \mathbf{l}_i. \quad (49)$$

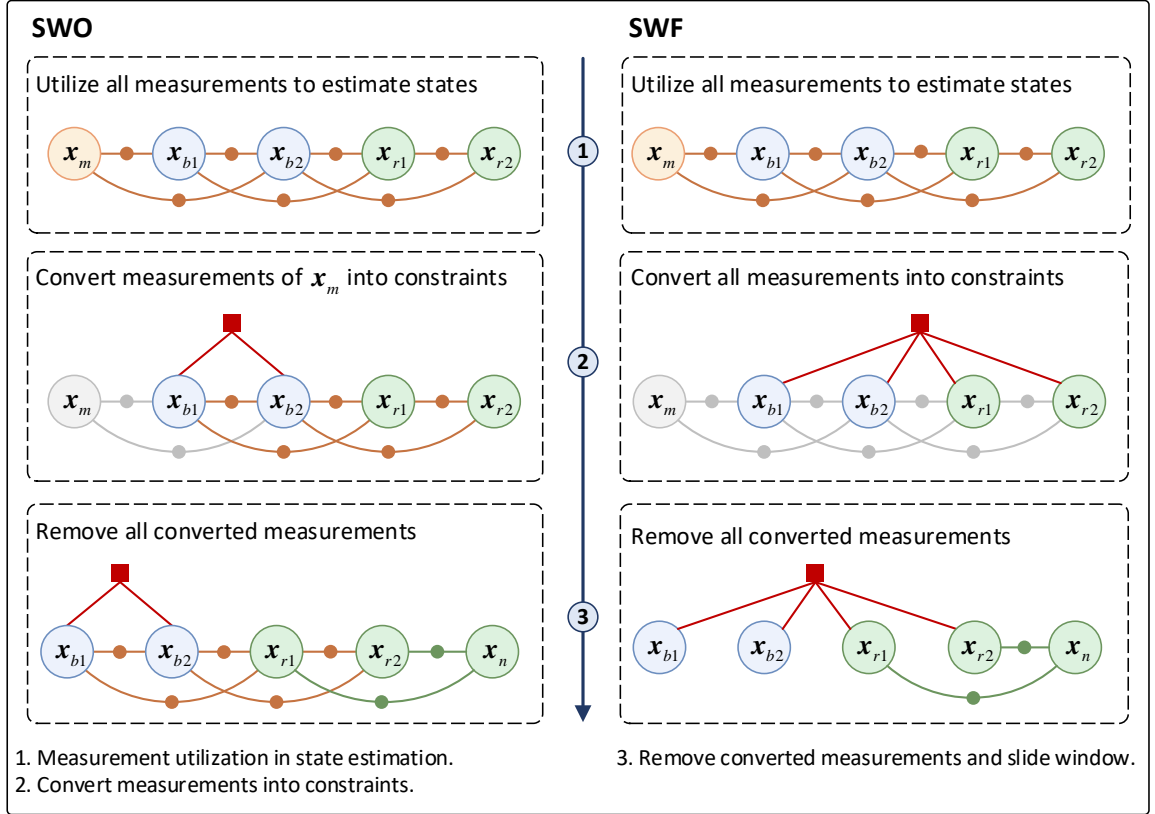


Fig. 3. Illustration of sliding window strategy in SWO and SWF. All measurements in the window are utilized (brown) to optimize states in both SWO and SWF. Differently, SWO only converts measurements of \mathbf{x}_m into constraints and then removes them (grey), which constrains \mathbf{x}_b (red). And the rest measurements can be utilized in the next window. SWF converts all utilized measurements into the priori, which constrains \mathbf{x}_b and \mathbf{x}_r in the next window. And thus, all these converted measurements are removed. With FEJ strategy taken into consideration, linearization points of \mathbf{x}_r are early fixed in SWF compared with SWO. \mathbf{x}_n denotes newly coming states after window slides.

4.2 Sliding window operation in SWF

In SWF, all measurements in the window are used to update states and then are converted into the posteriori (50), which is different from SWO, as shown in Fig. 3.

$$\delta \hat{\mathbf{X}}_t = \delta \mathbf{X}_{t|t-1} + \mathbf{K}_t (\mathbf{L}_t - \mathbf{H}_t \delta \mathbf{X}_{t|t-1}), \quad (50)$$

$$\mathbf{P}_t = (\mathbf{I} - \mathbf{K}_t \mathbf{H}_t) \mathbf{P}_{t|t-1}.$$

Considering $\mathbf{P}_t = \mathbf{N}_t^{-1}$, \mathbf{P}_t has the following form by mathematically inverting \mathbf{N}_t (see Appendix):

$$\mathbf{P}_t = \mathbf{N}_t^{-1} = \begin{bmatrix} \mathbf{N}_{mm} & \mathbf{N}_{mb'} \\ \mathbf{N}_{mb'}^T & \mathbf{N}_{b'b'} \end{bmatrix}^{-1} = \begin{bmatrix} (\mathbf{N}_{mm} - \mathbf{N}_{mb'} \mathbf{N}_{b'b'}^{-1} \mathbf{N}_{mb'}^T) & \mathbf{C} \\ \mathbf{C}^T & (\mathbf{N}_{b'b'} - \mathbf{N}_{mb'}^T \mathbf{N}_{mm}^{-1} \mathbf{N}_{mb'})^{-1} \end{bmatrix}, \quad (51)$$

where,

$$\mathbf{N}_{b'b'} = \begin{bmatrix} \mathbf{N}_{bb} & \mathbf{N}_{br} \\ \mathbf{N}_{br}^T & \mathbf{N}_{rr} \end{bmatrix}, \mathbf{N}_{mb'} = [\mathbf{N}_{mb} \quad \mathbf{N}_{mr}]. \quad (52)$$

When window slides, \mathbf{x}_m with measurements in \mathbf{Z}_t is removed from window, and covariance of \mathbf{x}_m is directly deleted. Note that estimated error state $\delta \hat{\mathbf{X}}_t$ should be compensated back to the propagated value via

$$\hat{\mathbf{X}}_t = \mathbf{X}_{t|t-1} - \delta \hat{\mathbf{X}}_t. \quad (53)$$

Thus, current estimates $\hat{\mathbf{X}}_t$ are regarded as zero error and then $\delta \hat{\mathbf{X}}_t$ is set to zero vector after compensation, which is the closed-loop system. Thereafter, \mathbf{x}_b and \mathbf{x}_r are constrained by

$$\mathbf{N}_{t,b'}^{-1} = (\mathbf{N}_{b'b'} - \mathbf{N}_{mb'}^T \mathbf{N}_{mm}^{-1} \mathbf{N}_{mb'})^{-1}, \delta \hat{\mathbf{x}}_{t,b'} = \begin{bmatrix} \delta \hat{\mathbf{x}}_b \\ \delta \hat{\mathbf{x}}_r \end{bmatrix} = \mathbf{0}. \quad (54)$$

Similarly, linearization points of \mathbf{x}_b and \mathbf{x}_r should remain the fixed instead of the updated according to FEJ strategy. However, the leading increasing linearization error is not compensated in SWF. Thus, the priori is denoted by (55) with $\mathbf{w}'_{t,p} \sim \mathcal{N}(\mathbf{0}, \mathbf{N}_{t,b'}^{-1})$

$$\delta \mathbf{z}'_{t,p} = \delta \hat{\mathbf{x}}_{t,b'} + \mathbf{w}'_{t,p}. \quad (55)$$

Then, constraints in the window $k+1$ can be expressed by

$$\mathbf{P}'_{t+1|t} = \begin{bmatrix} (\mathbf{N}_{b'b'} - \mathbf{N}_{mb'}^T \mathbf{N}_{mm}^{-1} \mathbf{N}_{mb'})^{-1} \\ \infty \mathbf{I} \end{bmatrix}, \delta \mathbf{X}'_{t+1|t} = \begin{bmatrix} \delta \hat{\mathbf{x}}_{t,b'} \\ \mathbf{0} \end{bmatrix}. \quad (56)$$

Consequently, incorporating virtual measurements (55), variables in the window $k+1$ can be estimated by

$$\begin{aligned} \delta \hat{\mathbf{X}}_{t+1} &= \mathbf{K}'_{t+1} \mathbf{L}'_{t+1}, \\ \mathbf{K}'_{t+1} &= \mathbf{P}'_{t+1|t} \mathbf{H}'_{t+1}^T (\mathbf{\Lambda}_{t+1}^{-1} + \mathbf{H}'_{t+1} \mathbf{P}_{t|t-1} \mathbf{H}_{t+1}^T)^{-1} \end{aligned} \quad (57)$$

Where, \mathbf{H}'_{t+1} and \mathbf{L}'_{t+1} are corresponding Jacobian matrix as well as residual vector using measurements in \mathbf{Z}'_{t+1} of SWF, where the two measurement sets of SWF and SWO are different from each other and will be discussed in the following.

4.3 Discrepancy Analysis Between SWO and SWF

In order to analyze discrepancies between SWO and SWF, expression of state estimation using SWF in window $t+1$ (57) is rewritten as

$$\left(\mathbf{P}'_{t+1|t} + \sum_{i=t+1}^{t+s'_{t+1}} \mathbf{h}_i^T \mathbf{\Lambda}_i \mathbf{h}_i \right) \delta \hat{\mathbf{X}}'_{t+1} = \sum_{i=t+1}^{t+s'_{t+1}} \mathbf{h}_i^T \mathbf{\Lambda}_i \mathbf{I}'_i. \quad (58)$$

Then, comparing (49) with (58), it is obviously shown that variables estimated by SWO and SWF in the window $t+1$ exist discrepancies. Since they are equivalent at timestamp t , the reasons of these discrepancies can be attributed to several aspects caused by sliding window strategy.

1) **Different Prior constraints.** From (42), SWO only converts measurements of \mathbf{x}_m using marginalization, which constrains \mathbf{x}_b only. Moreover, it considers full prior constraints, including both $\mathbf{P}_{t+1|t}^{-1}$ and $\delta \mathbf{X}_{t+1|t}$ in the window $t+1$. On the contrary, SWF constrains \mathbf{x}_b and \mathbf{x}_r in the window $t+1$ (54), since it converts all measurements in \mathbf{Z}_t into the priori. Besides, note that $\delta \hat{\mathbf{x}}_{t,b'}$ is set to zero vector (54), and the priori satisfies

$$\mathbf{N}_{t,b'} \delta \hat{\mathbf{x}}_{t,b'} = \mathbf{b}_{t,b'}. \quad (59)$$

Thus, $\mathbf{b}_{t,b'}$ is a zero vector as well. However, $\mathbf{b}_{t,b}$ is obtained by marginalization in SWO (43), which is set to a nonzero vector and constrains variables in the window $t+1$. As a result, zero

vector $\delta \hat{\mathbf{x}}_{t,b'}$ is equivalent to ignoring part of constraints compared with SWO.

2) **Different Measurement Utilization.** All measurements converted to the constraints should be removed in both SWO or SWF. Otherwise, measurements will be repeatedly used. Theoretically, prior constraints $\mathbf{P}_{t+1|t}^{-1}$ in SWO can be rewritten as:

$$\mathbf{P}_{t+1|t}^{-1} \Leftrightarrow \mathbf{P}_{t|t-1}^{-1} + \sum_{i=t}^{t+s_t} \mathbf{h}_i^T \mathbf{\Lambda}_i \mathbf{h}_i. \quad (60)$$

Then the information matrix in the window $t+1$ can be expressed by

$$\mathbf{N}_{t+1} \Leftrightarrow \mathbf{P}_{t|t-1}^{-1} + \sum_{i=t}^{t+s_t} \mathbf{h}_i^T \mathbf{\Lambda}_i \mathbf{h}_i + \sum_{i=t+1}^{t+s_{t+1}} \mathbf{h}_i^T \mathbf{\Lambda}_i \mathbf{h}_i. \quad (61)$$

Suppose $\mathbf{Z}_t \cap \mathbf{Z}_{t+1} \neq \emptyset$, some measurements will be repeatedly contributed to the information matrix, leading to the wrong posteriori. Then, this further results in wrong constraints and suboptimal solution in the subsequent estimation, especially fusing measurements from multiple sources.

Consequently, measurements converted into the constraints are no longer utilized in window $t+1$. SWO only converts measurements in $\mathbf{Z}_{t,m}$ to the priori, leading to $\mathbf{Z}_t \cap \mathbf{Z}_{t+1} = \mathbf{Z}_{t,r}$ without repeatedly use. Conversely, SWF converts and removes all measurements in \mathbf{Z}_t , and thus \mathbf{Z}'_{t+1} in SWF is different from \mathbf{Z}_{t+1} in SWO, which satisfies $\mathbf{Z}_t \cap \mathbf{Z}'_{t+1} = \emptyset$. In other words, measurements in $\mathbf{Z}_{t,r}$ are *re-linearized* compared with SWF, which is the reason that SWO outperforms SWF.

3) **Different FEJ Compensation.** Linearization points of constrained states should be fixed according to FEJ strategy, where the linearization error may be increasingly larger and the first order of this linearization error can be compensated in SWO (47). However, it is ignored in SWF.

4.4 Applications of SWO and SWF in Visual SLAM: Two Representative Frameworks

Strategies in two representative and state-of-the-art schemes are further discussed based on visual inertial odometry: VINS-mono and MSCKF, which utilizes SWO and SWF respectively. Since the two schemes apply different model of inertial measurement unit (IMU), visual measurements processing is focused in state estimation.

1) **VINS-mono:** VINS-mono utilizes landmarks, which have been tracked by at least two frames, to estimate states in the window. After estimation, the oldest frame is to be removed. Next, the Markov blanket of the oldest frame is built, which contains landmarks observed by it, i.e. $\mathbf{x}_{t,1}$ and $\mathbf{x}_{t,2}$ in this case. Marginalization is then carried out according to (42), resulting in constraints of the Markov blanket in the next window. Since only measurements of the removed frame are converted to the posteriori, others remain available in the next window due to marginalization.

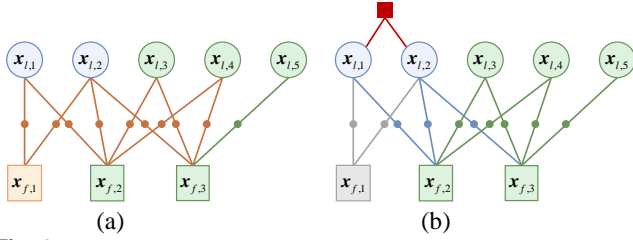


Fig. 4. An example illustrates measurement utilization and sliding window strategy in VINS-mono. All tracked landmarks are jointly estimated with frames by all measurements (a). Only measurements of the oldest frame are removed, and thus the Markov blanket of $x_{f,1}$ which consists of $x_{l,1}$ and $x_{l,2}$ is constrained (b).

2) *MSCKF*: Landmarks will be used if one of the following conditions are satisfied in MSCKF, where strategies utilized in it are simplified for discussion, as shown in Fig. 5.

- i. Landmarks are no longer observed by latest frame;
- ii. Landmarks are observed by at least three frames.

In this case, $x_{l,1}$ and $x_{l,2}$ are utilized, which satisfies the two conditions respectively. With measurements connected to the two landmarks utilized to estimate states, these measurements are all converted into constraints. Then, these measurements as well as the oldest frame, $x_{f,1}$, are removed from window. Since the two landmarks are no longer observed, they are also removed. Thus, $x_{f,2}$ and $x_{f,3}$ are constrained in the next estimation. And measurements are utilized only once.

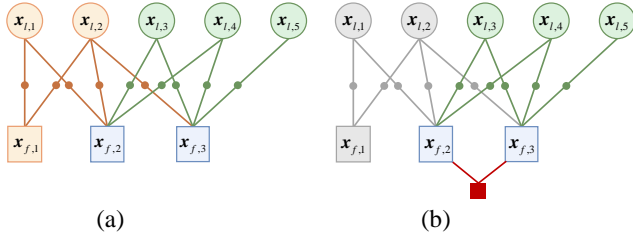


Fig. 5. An example illustrates measurement utilization and sliding window strategy in MSCKF. Only satisfied landmarks (brown) are used to update states (a). And all these used measurements are removed and converted to the priori after estimation, which constrains $x_{f,2}$ and $x_{f,3}$ (b).

Comparing the two applications of SWO with SWF in visual odometry, discrepancies between SWO and SWF will show up even from the first window, since different measurement utilization strategies are applied in SWO and SWF, and which further leads to different constraints.

Proposition 3: Theoretically, discrepancies are caused by sliding strategy instead of state estimation algorithms, where the estimates are equal at the beginning but SWO outperforms SWF in the subsequent estimation. When window slides, SWO utilizes marginalization over the Markov blanket (42) converting part of measurements, which enables re-linearization. SWF converts all used measurements into the priori and ignores part of constraints (54). Due to different measurements converted to the priori, measurement utilization is further different from each other. Besides, linearization error in SWO caused by FEJ can be compensated (45), while it is increasingly larger in SWF (55).

5 STRATEGY MODIFICATION ON SLIDING WINDOW

FILTERING

In this section, strategies of existing SWF are adjusted in order to achieve the same performance in accuracy with SWO. Measurement set Z_t is firstly divided into two parts, where they contain measurements of the removing states and the others respectively. A 2-step update strategy, ensuring the same measurement utilization and constraints, is utilized which makes SWF equivalent to SWO in accuracy.

5.1 Observation Preparation

Comparing (42) with (54), it is shown that marginalization using Schur complement in SWO shares the same mathematical form with directly deleting covariance in SWF. In other words, the prior constraints will be equivalent once the same measurements are converted. Thereafter, the measurement utilization in the next window is equal as well.

So, given the estimating states $X_t = [x_t \cdots x_{t+N}]$ with measurements contained in set Z_t , the to-be-removed states x_m should be firstly selected. Then, divide Z_t into $Z_{t,m}$ and $Z_{t,r}$, which denotes set of measurements observed by x_m and the rest respectively, where the two sets satisfy

$$Z_{t,m} \cup Z_{t,r} = Z_t, Z_{t,m} \cap Z_{t,r} = \emptyset. \quad (62)$$

5.2 2-step Update

Since the final estimates of sequential update equal to the solution of batch estimation according to proposition 1, measurements in $Z_{t,m}$ and $Z_{t,r}$ are updated sequentially in order to ensure the equivalent estimates. Then, only measurements in $Z_{t,m}$ are converted into constraints in the next window, ensuring the same priori. Additionally, with the IFR expression in (49), it is demonstrated that the ignored constraints $\delta X_{t|t-1}$ and FEJ compensation ΔX_t can also be taken into consideration in SWF equivalently. Thus, after attaining the prior constraints $P_{t|t-1}$ and $\hat{X}_{t|t-1}^{(0)}$, the 2-step update strategy in i -th iteration is adjusted as follows (see Fig. 6):

Step 1: Update states and covariance using measurements in $Z_{t,m}$ with full information incorporated and FEJ compensation.

Evaluate Jacobian matrix at $\hat{X}_{t|t-1}^{(i)}$, which is the estimates of previous iteration. Note that the estimates, $\delta \hat{X}_{t,m}^{(i)}$, should not be compensated to states. Otherwise, the consistency of Jacobi will be broken, and wrong observability will be introduced due to different linearization points applied in the same states.

Step 2: Sequentially update states by measurements belonging to Z_r regarding the estimates of step 1 as a virtual measurement. Similarly, evaluate Jacobian matrix at $\hat{X}_{t|t-1}^{(i)}$ which ensures the Jacobian consistency. Note that $\delta \hat{X}_{t,m}^{(i)}$ and $P_{t,m}^{(i)}$, should be used as virtual measurements, in order to utilize measurement information in $Z_{t,m}$, as the stated sequential update in proposition 2. The covariance matrix in this step is not updated to ensure the same measurements converted to posteriori.

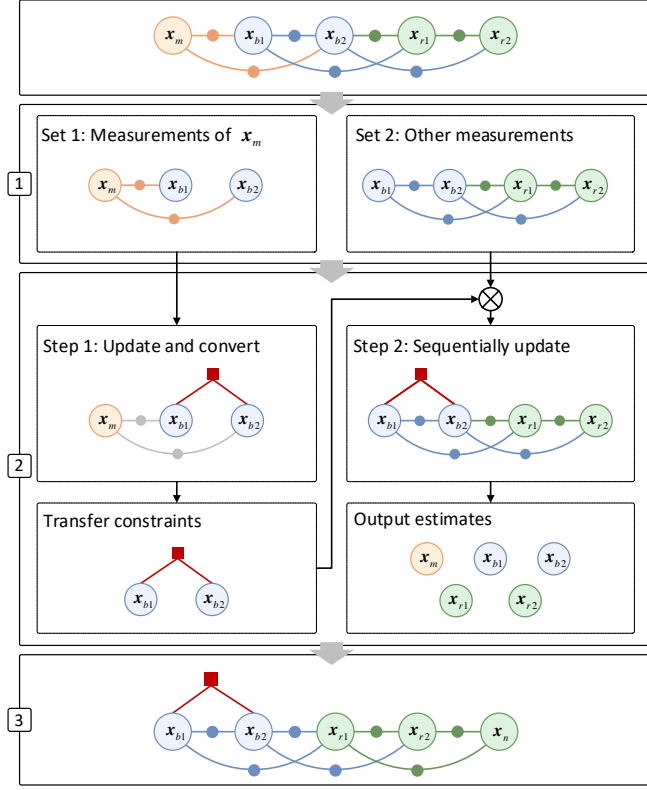


Fig. 6. Overview of the strategy adjusted SWF. Suppose variables (x_m, x_b, x_r) in the current window are to be estimated. Three procedures are executed in the strategy adjusted SWF. 1) Divide measurements into two sets, where the first set contains measurements of x_m and the other contains the rest ones. 2) Adjust SWF by a 2-step update strategy. Step 1 utilizes measurements in set 1 to update and convert them into the constraints. Step 2 sequentially updates states with constraints from step 1 incorporated, then output the estimates. 3) Slide the window and discard measurements of x_m . Thus, the 2-step update strategy enables the equivalent prior constraints, measurement utilization, and FEJ compensation compared with SWO.

Step 1 converts the same measurements compared with SWO, ensuring the equivalent constraints in the next window. On the other hand, step 2 fully utilizes all available measurements without converting them. Then, other measurements can be still used (re-linearization) in the subsequent window, which ensures the same measurement utilization and thus the same estimates as SWO in the subsequent. The whole procedure of the 2-step update can be referred to algorithm 2.

Algorithm 2 Strategy adjusted sliding window filtering

```

set  $\delta X_{t|t-1}^{(0)} \leftarrow \delta X_{t|t-1}, i \leftarrow 1, \text{MaxIter}, P_{t|t-1}$ 
divide  $Z_k$  into  $Z_{k,m}, Z_{k,r}$ , s.t.  $Z_{t,m} \cup Z_{t,r} = Z_t, Z_{t,m} \cap Z_{t,r} = \emptyset$ 
while  $i < \text{MaxIter}$ 
  // update and convert measurements in  $Z_{t,m}$ 
   $\delta \hat{X}_{t,m}^{(i)} \leftarrow \delta X_{t|t-1}^{(i-1)} + \Delta X_k$ 
   $+ P_{t|t-1}^{-1} H_{t,m}^{(i)T} \Lambda_{t,m} \left( L_{t,m}^{(i)} - H_{t,m}^{(i)} \left( \delta X_{t|t-1}^{(i-1)} + \Delta X_t \right) \right)$ 

```

$$P_{t,m}^{(i)} \leftarrow (I - K_{t,m}^{(i)} H_{t,m}^{(i)}) P_{t|t-1}$$

// sequentially update measurements in $Z_{t,r}$

$$\delta \hat{X}_{t,r}^{(i)} \leftarrow \delta \hat{X}_{t,m}^{(i)} + P_{t,m}^{(i)} H_{t,r}^{(i)T} \Lambda_{t,r} \left(L_{t,r}^{(i)} - H_{t,r}^{(i)} \delta \hat{X}_{t,m}^{(i)} \right)$$

$$\hat{X}_{t|t-1}^{(i)} \leftarrow \hat{X}_{t|t-1}^{(i-1)} - \delta \hat{X}_{t,r}^{(i)}$$

if converge

// transfer constraints converted by $Z_{t,m}$

$$P_t \leftarrow P_{t,m}^{(i)}, \delta \hat{X}_{t,m}^{(i)}$$

// return estimates updated by all measurements

$$\hat{X}_t \leftarrow \hat{X}_{t|t-1}^{(i)}$$

return $\hat{X}_t, P_t, \delta \hat{X}_{t,m}^{(i)}$

end if

$i \leftarrow i + 1$

end while

output estimates \hat{X}_t

slide window

transfer constraints $P_{t+1|t} \leftarrow \left[P_t \right]_{x_m}, \delta X_{t+1|t} \leftarrow \left[\delta X_{t,m}^{(i)} \right]_{x_m}$

Where, operator \mathbf{v}_x denotes delete elements related to x in matrix or vector \mathbf{v} . Thus, the strategy adjusted SWF has the same prior constraints, measurement utilization, and FEJ compensation.

Proposition 4: By adjusting existing strategy to the 2-step update strategy, which ensures the same measurements utilized and converted to the constraints, strategy adjusted SWF mathematically equals to SWO. It again demonstrates that the discrepancies between SWF and SWO are caused by strategies instead of state estimation algorithms.

6 EXPERIMENTAL VALIDATION BASED ON VISUAL SLAM

In this section, a series of experiments are presented based on visual SLAM to validate the theoretical analysis, in particular, proposition 2-4. Specifically, Monte-Carlo experiments are conducted first to validate the propositions. And then, the strategy adjusted SWF (SWF-SA) is examined by ablation studies based on KITTI dataset. Besides, assume that there are no external dynamic or other constraints information in all these experiments, where Φ is set to the identity. Since SLAM is a highly nonlinear system, iteration strategy has to be employed, where the threshold in algorithm 1 and algorithm 2 is also set equally. The weight of each visual measurement is equally set to 2 pixels. Other strategies, such as outlier rejection, are implemented equivalently in the two estimators as well. In our implementations, IFR in optimization-based methods, while KFR is utilized in filtering-based counterparts.

The root mean square error (RMSE) and the discrepancy are applied to evaluate experimental results. Discrepancy is defined as the difference between estimates of the filtering-based and optimization-based approaches, then the Euclidean norm is calculated.

6.1 Monte-Carlo experiments

A series of Monte-Carlo experiments are conducted, includ-

ing 1) solve camera poses as well as landmarks by batch estimation and sequential update, 2) apply real-time operation using SWO and SWF, and 3) adjust existing strategy in SWF then compare with SWO. Each experiment processes the same data of 50 Monte Carlo trials, where the camera moves along a circular path, as shown in Fig. 7. Initial values of frames and landmarks for linearization are provided for simplicity, rather than triangulate feature points and determine camera poses using structure-from-motion, which will not affect the equivalence between estimators.

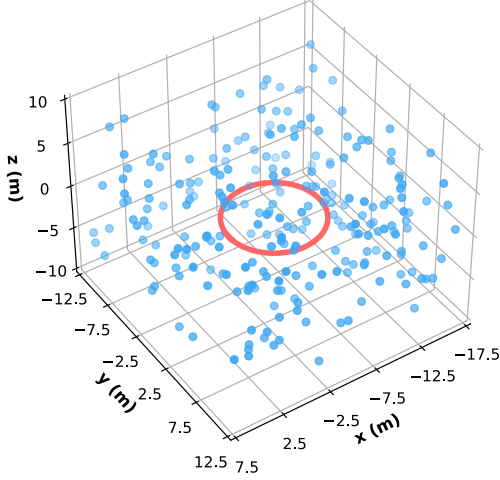


Fig. 7. A stereo camera is simulated, moving along the circular trajectory (red) with landmarks (blue) randomly generated in the space.

1) *Equivalence between batch estimation and sequential update.* The system states of each trial include all camera poses as well as all landmarks observed by at least 2 frames. Then, all tracked feature points are utilized as measurements in both approaches. Discrepancies between the two methods at arbitrary timestamps can be obtained, where Fig. 8 shows averaged discrepancies between batch estimation and the sequential update at each timestamp, demonstrating the equivalence between the two estimators if measurements, observation model, stochastic model, and linearization points are equivalently utilized. This slight difference may be introduced by computational error.

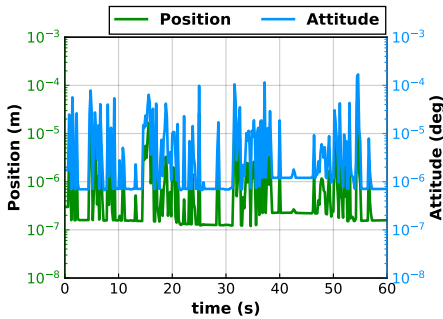


Fig. 8. Equivalence between batch estimation and sequential update based on 50 Monte Carlo trials. The maximum discrepancies of position and attitude are no more than $10^{-4}m$, 10^{-3} deg respectively, which demonstrates the equivalence between batch estimation and sequential update.

2) *Discrepancies between SWO and existing SWF.* In our implementation of SWO and SWF, the length of window is set to 20 frames as well as landmarks observed by them, and all

tracked feature points are used to update states, ensuring the same measurements utilization at the beginning. Thus, estimating results of the first frame will be equivalent (see Fig. 9). Differently, SWO adopts marginalization to convert measurements of the first frame to priori. On the contrary, SWF converts all utilized measurements into constraints and then discards them. So, this equivalence corrupts from the second window due to different measurements converted, which further leads to different measurement utilization. And thus existing SWF is inequal to SWO.

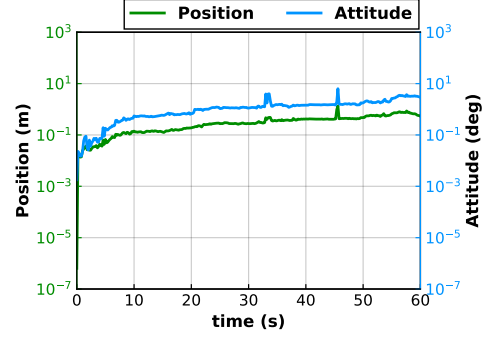


Fig. 9. Averaged discrepancies between SWO and existing SWF, where SWO and SWF are equal ($10^{-7}m$ and 10^{-7} deg respectively) at beginning due to the same measurement utilization and prior knowledge. This equivalence corrupts due to existing strategies applied in SWF, which intuitively shows the inequality between SWO and SWF.

3) *Equivalence between SWO and SWF-SA.* Strategies in SWF are adjusted, by a 2-step update strategy stated in proposition 4, which ensures the same measurement utilization and prior constraints with full constraints and FEJ compensation incorporated. Thus, discrepancies between the optimization and filtering are disappeared, remaining computational error, demonstrating the equivalence between SWO and SWF-SA, as shown in Fig. 10.

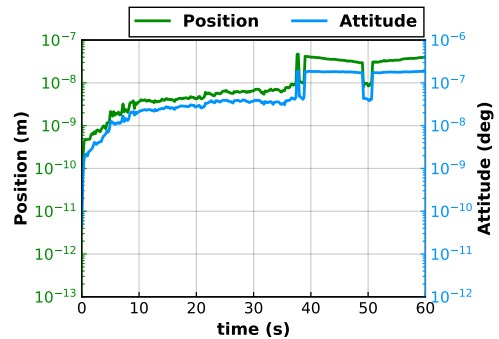


Fig. 10. Averaged discrepancies between SWO and SWF-SA based on 50 Monte Carlo trials. The maximum discrepancies of position and attitude are no more than $10^{-6}m$, 10^{-7} deg respectively, showing the equivalence between SWO and SWF-SA.

Finally, Fig. 11 shows RMSE of each trial, which demonstrates that SWO indeed significantly performs better than existing SWF due to strategies, specifically re-linearization. But with a 2-step modification, the SWF-SA achieves the same estimating accuracy with SWO. Besides, batch estimation fully utilizes measurement information, which shows the best estimating accuracy in most trials compared with sliding window estimation and is consistent with the preceding analysis.

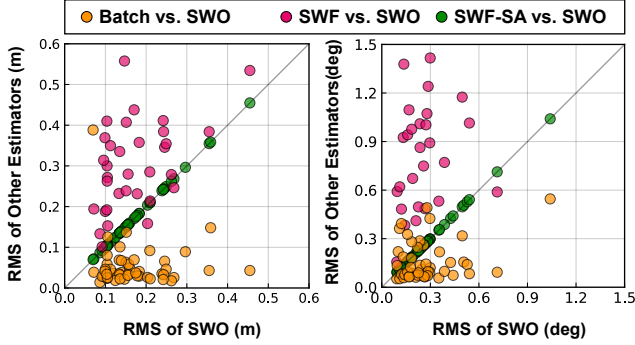


Fig. 11. Comparison of Monte-Carlo experimental results. Each circle represents RMSE of once trial and the color denotes the comparing group.

6.2 Ablation studies based on KITTI

In this section, effects of incomplete constraints and FEJ compensation in SWF are analyzed through ablation experiments based on KITTI dataset [52]. In the ablation experiments, two factors are analyzed, ignored constraints $\delta\hat{x}_{t,b}$ and FEJ compensation. Since SWO as well as SWF-SA achieves the best estimates in real-time operation from the simulation experiments, SWO is conducted as the control group and SWF-SA is firstly compared with it. Then, based on SWF-SA, the first factor is analyzed by ignoring $\delta\hat{x}_{t,b}$, while strategy of FEJ compensation remains equivalent to SWF-SA, namely SWF-FEJ. Finally, linearization error caused by FEJ is not compensated but with full priori incorporated, analyzing the second factor, namely SWF-Full. Note that other strategies keep equivalent to SWF-SA, including measurement utilization, outlier removal, et

al. And ORBSLAM 2 [55] is utilized as our frontend to provide initial values of camera frames and landmarks.

Comparing with SWO, TABLE I gives the statistics of discrepancies of the ablation experiments. The maximum reflects the extreme discrepancies but it may be affected by wrong estimates. The average can reflect the general discrepancy over each sequence. First, SWF-SA is strictly equivalent to SWO, where both the maximum and average are no more than 10^{-6} on both translation and attitude, which is negligible during state estimation and is consistent with our previous analysis. Apparently, both $\delta\hat{x}_{t,b}$ and FEJ compensation significantly affect equivalence between SWO and SWF. Specifically, SWF-FEJ shows better equivalence compared with SWF-Full in general, which indicates that compensating linearization error caused by FEJ is necessary in a high nonlinear system. Meanwhile, incomplete constraints may lead to large extreme value on attitude, which is indispensable during state estimation as well.

It is important to note that, in implementation of MSCKF 2.0 [56], correct observable structure of the estimator using FEJ is more important than linearization error introduced by FEJ, since they utilize inertial measurement unit to provide reliable initial value. In our experiments, linearization points are obtained from a pure visual scheme, where the linearization error may be larger compared with the visual inertial system. Thus, for a high nonlinear system without precise initial value, the effect of linearization error may be greater than incomplete prior constraints.

TABLE I
STATISTICS OF DISCREPANCY COMPARED WITH SWO ON ABLATION EXPERIMENTAL RESULTS

Sequence	Translation ($\log_{10} m$)						Attitude ($\log_{10} \text{deg}$)					
	SWF-SA		SWF-FEJ		SWF-Full		SWF-SA		SWF-FEJ		SWF-Full	
	max	mean	max	mean	max	mean	max	mean	max	mean	max	mean
00	-7.98	-8.62	1.13	0.03	1.21	0.15	-6.68	-8.51	1.40	0.32	1.77	0.59
01	-8.38	-9.14	1.59	0.72	1.52	0.82	-8.28	-9.73	1.06	-0.14	0.79	0.02
02	-7.72	-8.60	0.78	0.21	0.77	0.21	-7.43	-8.52	1.47	0.34	1.17	0.45
03	-8.50	-9.08	1.37	0.27	1.37	0.16	-8.47	-9.24	1.42	0.01	1.42	0.15
04	-8.87	-9.17	0.48	-0.01	0.74	0.21	-8.85	-9.33	0.77	0.43	0.54	0.11
05	-8.12	-8.58	0.67	0.30	1.10	0.21	-7.73	-8.54	1.14	0.56	1.38	0.46
06	-8.32	-9.07	0.85	0.23	1.50	0.26	-7.80	-8.83	1.32	0.40	1.10	0.36
07	-7.35	-8.55	0.50	0.12	0.69	0.12	-6.90	-7.71	1.20	0.46	1.16	0.31
08	-7.56	-8.66	0.88	0.23	1.53	0.30	-6.74	-8.31	1.44	0.51	1.37	0.58
09	-7.88	-8.84	0.60	0.14	0.91	0.13	-7.54	-8.72	1.20	0.36	1.39	0.37
10	-8.18	-8.69	0.96	0.11	1.00	0.16	-7.85	-8.33	1.19	0.57	1.35	0.59

Note that $\log_{10} 10^x = x$, which shows that statistics in TABLE I reflect the magnitude of discrepancies in ablation experiments.

Besides, from the simulation analysis between SWO and SWF-SA (Fig. 10), discrepancies between them tend to increase. Nonetheless, KITTI dataset contains large-scale scenarios, where our experimental results intuitively show that by adjusting strategies in existing SWF, the SWF-SA strictly equals to

SWO and discrepancies between them are negligible during estimation (see Fig. 12).

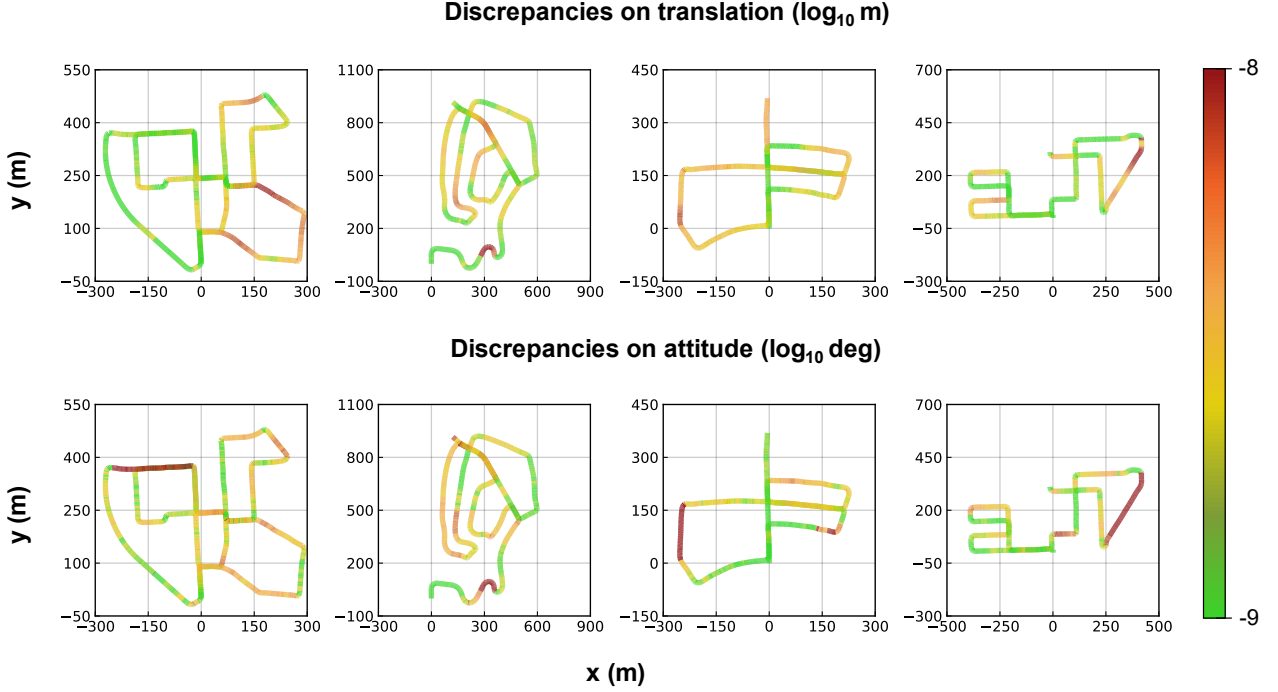


Fig. 12. Discrepancies on translations and attitudes between SWF-SA and SWO based on KITTI 00, 02, 05, and 08 sequences. Trajectories are plotted by ground truth, and color represents discrepancies between SWF-SA and SWO.

7 CONCLUDING REMARKS

In this paper, discrepancies between optimization-based and filtering-based approaches have been explored by analyzing theories and strategies between batch estimation and sequential update, between SWO and SWF. It demonstrates that the discrepancies are caused by sliding window strategy, where SWF converts different measurements into constraints compared with SWO, leading to different prior constraint and measurement utilization. Additionally, SWF ignores part of constraints and FEJ compensation. Based on previous analysis, a 2-step update strategy is applied in existing SWF, which ensures the same measurements converted, fully incorporates prior constraints and FEJ compensation. The strategy adjusted SWF is evaluated by simulation experiments and ablation experiments based on KITTI dataset, where it is shown that strategy adjusted SWF is strictly equivalent to SWO under the assumption of the same measurements, observation model, stochastic model, and strategies. The discrepancies are caused by different strategies which further leads to different measurement utilization and constraints, instead of state estimation algorithms.

Thus, future research on sensor fusion should thoroughly describe their own algorithms and strategies which essentially affect the estimating results. For example, feature correspondence is main task in computer vision, which should be deeply studied. Then, measurements from the same sensor have different quality, even including outliers, which may lead to fault estimates. For those outliers, they should be carefully detected and removed, while other healthy observations should also be properly weighted by a prior model instead of determining weights through empirical values. Additionally, it is still a challenge to estimate uncertainty of each sensor appropriately for sensor fusion problem, which may lead to wrong estimates as

stated before. As for state estimation method, it should be chosen according to dimension of system states as well as number of measurements to ensure efficiency due to different time complexity of the two state estimation methods.

8 APPENDIX A

Given a real symmetric positive definite matrix N , the inverse of it can be expressed by N^{-1} which satisfies:

$$\begin{bmatrix} N_{mm} & N_{mb} \\ N_{mb}^T & N_{bb} \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} I & \\ & I \end{bmatrix}. \quad (63)$$

Thus, we have:

$$\begin{cases} N_{mm}A + N_{mb}C = I \\ N_{mm}B + N_{mb}D = 0 \\ N_{mb}^TA + N_{bb}C = 0 \\ N_{mb}^TB + N_{bb}D = I \end{cases}. \quad (64)$$

Since N_{mm} and N_{bb} are invertible, N^{-1} can be derived by solving (64):

$$\begin{cases} A = (N_{mm} - N_{mb}N_{bb}^{-1}N_{mb}^T)^{-1} \\ C = B^T = -(N_{bb} - N_{mb}^TN_{mm}^{-1}N_{mb})^{-1}N_{mb}^TN_{mm}^{-1} \\ D = (N_{bb} - N_{mb}^TN_{mm}^{-1}N_{mb})^{-1} \end{cases}. \quad (65)$$

9 REFERENCES

- [1] A.N. U, N. Naganure, S.K. S, BEV Detection and Localisation using Semantic Segmentation in Autonomous Car Driving Systems, in: 2021 IEEE International Confer-

- ence on Electronics, Computing and Communication Technologies (CONECCT), IEEE, Bangalore, India, 2021: pp. 1–6. <https://doi.org/10.1109/CONECCT52877.2021.9622702>.
- [2] Q. Liu, X. Li, S. Yuan, Z. Li, Decision-Making Technology for Autonomous Vehicles: Learning-Based Methods, Applications and Future Outlook, in: 2021 IEEE International Intelligent Transportation Systems Conference (ITSC), IEEE, Indianapolis, IN, USA, 2021: pp. 30–37. <https://doi.org/10.1109/ITSC48978.2021.9564580>.
 - [3] S. Li, X. Li, H. Wang, Y. Zhou, Z. Shen, Multi-GNSS PPP/INS/Vision/LiDAR tightly integrated system for precise navigation in urban environments, *Information Fusion* 90 (2023) 218–232. <https://doi.org/10.1016/j.inffus.2022.09.018>.
 - [4] C. Qin, H. Ye, C.E. Pranata, J. Han, S. Zhang, M. Liu, LINS: A Lidar-Inertial State Estimator for Robust and Efficient Navigation, in: 2020 IEEE International Conference on Robotics and Automation (ICRA), IEEE, Paris, France, 2020: pp. 8899–8906. <https://doi.org/10.1109/ICRA40945.2020.9197567>.
 - [5] T. Qin, P. Li, S. Shen, VINS-Mono: A Robust and Versatile Monocular Visual-Inertial State Estimator, *IEEE Trans. Robot.* 34 (2018) 1004–1020. <https://doi.org/10.1109/TRO.2018.2853729>.
 - [6] X. Li, X. Li, S. Li, Y. Zhou, M. Sun, Q. Xu, Z. Xu, Centimeter-accurate vehicle navigation in urban environments with a tightly integrated PPP-RTK/MEMS/vision system, *GPS Solut* 26 (2022) 124. <https://doi.org/10.1007/s10291-022-01306-3>.
 - [7] X. Li, X. Chen, M. Ge, H. Schuh, Improving multi-GNSS ultra-rapid orbit determination for real-time precise point positioning, *J Geod* 93 (2019) 45–64. <https://doi.org/10.1007/s00190-018-1138-y>.
 - [8] W. Fu, G. Huang, Q. Zhang, S. Gu, M. Ge, H. Schuh, Multi-GNSS real-time clock estimation using sequential least square adjustment with online quality control, *J Geod* 93 (2019) 963–976. <https://doi.org/10.1007/s00190-018-1218-z>.
 - [9] H. Ge, T. Wu, B. Li, Characteristics analysis and prediction of Low Earth Orbit (LEO) satellite clock corrections by using least-squares harmonic estimation, *GPS Solut* 27 (2023) 38. <https://doi.org/10.1007/s10291-022-01377-2>.
 - [10] R. Dach, P. Walser, Bernese GNSS software version 5.2, University of Bern, Bern Open Publishing, 2015.
 - [11] Y. Li, Analysis of GAMIT/GLOBK in high-precision GNSS data processing for crustal deformation, *Earthquake Research Advances* 1 (2021) 100028. <https://doi.org/10.1016/j.eqrea.2021.100028>.
 - [12] X. Li, H. Li, G. Huang, Q. Zhang, S. Meng, Non-holonomic constraint (NHC)-assisted GNSS/SINS positioning using a vehicle motion state classification (VMSC)-based convolution neural network, *GPS Solut* 27 (2023) 144. <https://doi.org/10.1007/s10291-023-01483-9>.
 - [13] W. Meiling, F. Guoqiang, Y. Huachao, L. Yafeng, Y. Yi, X. Xuan, A loosely coupled MEMS-SINS/GNSS integrated system for land vehicle navigation in urban areas, in: 2017 IEEE International Conference on Vehicular Electronics and Safety (ICVES), IEEE, Vienna, Austria, 2017: pp. 103–108. <https://doi.org/10.1109/ICVES.2017.7991909>.
 - [14] X. Zhang, F. Zhu, X. Tao, R. Duan, New optimal smoothing scheme for improving relative and absolute accuracy of tightly coupled GNSS/SINS integration, *GPS Solut* 21 (2017) 861–872. <https://doi.org/10.1007/s10291-017-0601-1>.
 - [15] D. Wang, Y. Dong, Z. Li, Q. Li, J. Wu, Constrained MEMS-Based GNSS/INS Tightly Coupled System With Robust Kalman Filter for Accurate Land Vehicular Navigation, *IEEE Trans. Instrum. Meas.* 69 (2020) 5138–5148. <https://doi.org/10.1109/TIM.2019.2955798>.
 - [16] K.W. Chiang, G.J. Tsai, H.W. Chang, C. Joly, N. El-Sheimy, Seamless navigation and mapping using an INS/GNSS/grid-based SLAM semi-tightly coupled integration scheme, *Information Fusion* 50 (2019) 181–196. <https://doi.org/10.1016/j.inffus.2019.01.004>.
 - [17] A.J. Davison, I.D. Reid, N.D. Molton, O. Stasse, MonoSLAM: Real-Time Single Camera SLAM, *IEEE Trans. Pattern Anal. Mach. Intell.* 29 (2007) 1052–1067. <https://doi.org/10.1109/TPAMI.2007.1049>.
 - [18] J. Kim, S. Sukkarieh, Real-time implementation of airborne inertial-SLAM, *Robotics and Autonomous Systems* 55 (2007) 62–71. <https://doi.org/10.1016/j.robot.2006.06.006>.
 - [19] M. Bloesch, S. Omari, M. Hutter, R. Siegwart, Robust visual inertial odometry using a direct EKF-based approach, in: 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, Hamburg, Germany, 2015: pp. 298–304. <https://doi.org/10.1109/IROS.2015.7353389>.
 - [20] A.I. Mourikis, S.I. Roumeliotis, A Multi-State Constraint Kalman Filter for Vision-aided Inertial Navigation, in: *Proceedings 2007 IEEE International Conference on Robotics and Automation*, IEEE, Rome, Italy, 2007: pp. 3565–3572. <https://doi.org/10.1109/ROBOT.2007.364024>.
 - [21] M. Li, A.I. Mourikis, Improving the accuracy of EKF-based visual-inertial odometry, in: 2012 IEEE International Conference on Robotics and Automation, IEEE, St Paul, MN, USA, 2012: pp. 828–835. <https://doi.org/10.1109/ICRA.2012.6225229>.
 - [22] A.I. Mourikis, N. Trawny, S.I. Roumeliotis, A.E. Johnson, A. Ansar, L. Matthies, Vision-Aided Inertial Navigation for Spacecraft Entry, Descent, and Landing, *IEEE Trans. Robot.* 25 (2009) 264–280. <https://doi.org/10.1109/TRO.2009.2012342>.
 - [23] K. Sun, K. Mohta, B. Pfrommer, M. Watterson, S. Liu, Y. Mulgaonkar, C.J. Taylor, V. Kumar, Robust Stereo Visual Inertial Odometry for Fast Autonomous Flight, *IEEE Robot. Autom. Lett.* 3 (2018) 965–972. <https://doi.org/10.1109/LRA.2018.2793349>.
 - [24] G.P. Huang, A.I. Mourikis, S.I. Roumeliotis, Observability-based Rules for Designing Consistent EKF SLAM Estimators, *The International Journal of Robotics Re-*

- search 29 (2010) 502–528. <https://doi.org/10.1177/0278364909353640>.
- [25] G. Bleser, D. Strickery, Using the marginalised particle filter for real-time visual-inertial sensor fusion, in: 2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality, IEEE, Cambridge, UK, 2008: pp. 3–12. <https://doi.org/10.1109/ISMAR.2008.4637316>.
- [26] J. Georgy, A. Noureldin, C. Goodall, Vehicle navigator using a mixture particle filter for inertial sensors/odometer/map data/GPS integration, *IEEE Trans. Consumer Electron.* 58 (2012) 544–552. <https://doi.org/10.1109/TCE.2012.6227459>.
- [27] H. Zhou, Z. Yao, C. Fan, S. Wang, M. Lu, Rao-Blackwellised particle filtering for low-cost encoder/INS/GNSS integrated vehicle navigation with wheel slipping, *IET Radar, Sonar & Navigation* 13 (2019) 1890–1898. <https://doi.org/10.1049/iet-rsn.2019.0108>.
- [28] G.P. Huang, A.I. Mourikis, S.I. Roumeliotis, A Quadratic-Complexity Observability-Constrained Unscented Kalman Filter for SLAM, *IEEE Trans. Robot.* 29 (2013) 1226–1243. <https://doi.org/10.1109/TRO.2013.2267991>.
- [29] T. Cantelobre, C. Chahbazian, A. Croux, S. Bonnabel, A real-time unscented Kalman filter on manifolds for challenging AUV navigation, in: 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, Las Vegas, NV, USA, 2020: pp. 2309–2316. <https://doi.org/10.1109/IROS45743.2020.9341216>.
- [30] M. Brossard, A. Barrau, S. Bonnabel, Exploiting Symmetries to Design EKFs With Consistency Properties for Navigation and SLAM, *IEEE Sensors J.* 19 (2019) 1572–1579. <https://doi.org/10.1109/JSEN.2018.2882714>.
- [31] C. Liu, C. Jiang, H. Wang, InGVIO: A Consistent Invariant Filter for Fast and High-Accuracy GNSS-Visual-Inertial Odometry, *IEEE Robot. Autom. Lett.* 8 (2023) 1850–1857. <https://doi.org/10.1109/LRA.2023.3243520>.
- [32] H. Strasdat, J.M.M. Montiel, A.J. Davison, Real-time monocular SLAM: Why filter?, in: 2010 IEEE International Conference on Robotics and Automation, IEEE, Anchorage, AK, 2010: pp. 2657–2664. <https://doi.org/10.1109/ROBOT.2010.5509636>.
- [33] R. Kummerle, G. Grisetti, H. Strasdat, K. Konolige, W. Burgard, G²o: A general framework for graph optimization, in: 2011 IEEE International Conference on Robotics and Automation, IEEE, Shanghai, China, 2011: pp. 3607–3613. <https://doi.org/10.1109/ICRA.2011.5979949>.
- [34] M. Kaess, H. Johannsson, R. Roberts, V. Ila, J. Leonard, F. Dellaert, iSAM2: Incremental smoothing and mapping with fluid relinearization and incremental variable reordering, in: 2011 IEEE International Conference on Robotics and Automation, IEEE, Shanghai, China, 2011: pp. 3281–3288. <https://doi.org/10.1109/ICRA.2011.5979641>.
- [35] D.M. Rosen, L. Carlone, A.S. Bandeira, J.J. Leonard, SE-Sync: A certifiably correct algorithm for synchronization over the special Euclidean group, *The International Journal of Robotics Research* 38 (2019) 95–125. <https://doi.org/10.1177/0278364918784361>.
- [36] S. Agarwal, K. Mierle, T.C.S. Team, Ceres solver, (2022). <https://github.com/ceres-solver/ceres-solver>.
- [37] C. Campos, R. Elvira, J.J.G. Rodríguez, J.M.M. Montiel, J.D. Tardós, ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual-Inertial and Multi-Map SLAM, *IEEE Trans. Robot.* 37 (2021) 1874–1890. <https://doi.org/10.1109/TRO.2021.3075644>.
- [38] S. Leutenegger, S. Lynen, M. Bosse, R. Siegwart, P. Furgale, Keyframe-based visual-inertial odometry using nonlinear optimization, *The International Journal of Robotics Research* 34 (2015) 314–334. <https://doi.org/10.1177/0278364914554813>.
- [39] R. Mur-Artal, J.D. Tardos, Visual-Inertial Monocular SLAM With Map Reuse, *IEEE Robot. Autom. Lett.* 2 (2017) 796–803. <https://doi.org/10.1109/LRA.2017.2653359>.
- [40] L. von Stumberg, D. Cremers, DM-VIO: Delayed Marginalization Visual-Inertial Odometry, *IEEE Robot. Autom. Lett.* 7 (2022) 1408–1415. <https://doi.org/10.1109/LRA.2021.3140129>.
- [41] K. Eickenhoff, L. Paull, G. Huang, Decoupled, consistent node removal and edge sparsification for graph-based SLAM, in: 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, Daejeon, South Korea, 2016: pp. 3275–3282. <https://doi.org/10.1109/IROS.2016.7759505>.
- [42] T. Qin, S. Cao, J. Pan, S. Shen, A General Optimization-based Framework for Global Pose Estimation with Multiple Sensors, (2019). <http://arxiv.org/abs/1901.03642> (accessed November 1, 2022).
- [43] X. Niu, H. Tang, T. Zhang, J. Fan, J. Liu, IC-GVINS: A Robust, Real-Time, INS-Centric GNSS-Visual-Inertial Navigation System, *IEEE Robot. Autom. Lett.* 8 (2023) 216–223. <https://doi.org/10.1109/LRA.2022.3224367>.
- [44] S. Cao, X. Lu, S. Shen, GVINS: Tightly Coupled GNSS-Visual-Inertial Fusion for Smooth and Consistent State Estimation, (2021). <http://arxiv.org/abs/2103.07899> (accessed November 1, 2022).
- [45] J. Liu, W. Gao, Z. Hu, Optimization-Based Visual-Inertial SLAM Tightly Coupled with Raw GNSS Measurements, in: 2021 IEEE International Conference on Robotics and Automation (ICRA), 2021: pp. 11612–11618. <https://doi.org/10.1109/ICRA48506.2021.9562013>.
- [46] T. Li, L. Pei, Y. Xiang, W. Yu, T.-K. Truong, P3-VINS: Tightly-Coupled PPP/INS/Visual SLAM Based on Optimization Approach, *IEEE Robot. Autom. Lett.* 7 (2022) 7021–7027. <https://doi.org/10.1109/LRA.2022.3180441>.
- [47] Y. Zhang, T. Zhang, S. Huang, Comparison of EKF based SLAM and optimization based SLAM algorithms, in: 2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA), IEEE, Wuhan, 2018: pp. 1308–1313. <https://doi.org/10.1109/ICIEA.2018.8397911>.
- [48] J. Delmerico, D. Scaramuzza, A Benchmark Comparison of Monocular Visual-Inertial Odometry Algorithms for Flying Robots, in: 2018 IEEE International Conference on Robotics and Automation (ICRA), IEEE, Brisbane, QLD, 2018: pp. 2502–2509. <https://doi.org/10.1109/ICRA.2018.8460664>.
- [49] B.M. Bell, F.W. Cathey, The iterated Kalman filter up-

date as a Gauss-Newton method, *IEEE Trans. Automat. Contr.* 38 (1993) 294–297. <https://doi.org/10.1109/9.250476>.

- [50] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, J.J. Leonard, Past, Present, and Future of Simultaneous Localization and Mapping: Toward the Robust-Perception Age, *IEEE Trans. Robot.* 32 (2016) 1309–1332. <https://doi.org/10.1109/TRO.2016.2624754>.
- [51] Y. Yang, J. Maley, G. Huang, Null-space-based marginalization: Analysis and algorithm, in: 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, Vancouver, BC, 2017: pp. 6749–6755. <https://doi.org/10.1109/IROS.2017.8206592>.
- [52] A. Geiger, P. Lenz, R. Urtasun, Are we ready for autonomous driving? The KITTI vision benchmark suite, in: Conference on Computer Vision and Pattern Recognition (CVPR), 2012.
- [53] G. Sibley, L. Matthies, G. Sukhatme, Sliding window filter with application to planetary landing: Sibley et al.: Sliding Window Filter, *J. Field Robotics* 27 (2010) 587–608. <https://doi.org/10.1002/rob.20360>.
- [54] G.P. Huang, A.I. Mourikis, S.I. Roumeliotis, A first-estimates jacobian EKF for improving SLAM consistency, in: O. Khatib, V. Kumar, G.J. Pappas (Eds.), *Experimental Robotics*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2009: pp. 373–382.
- [55] R. Mur-Artal, J.D. Tardos, ORB-SLAM2: An Open-Source SLAM System for Monocular, Stereo, and RGB-D Cameras, *IEEE Trans. Robot.* 33 (2017) 1255–1262. <https://doi.org/10.1109/TRO.2017.2705103>.
- [56] M. Li, A.I. Mourikis, High-precision, consistent EKF-based visual-inertial odometry, *The International Journal of Robotics Research* 32 (2013) 690–711. <https://doi.org/10.1177/0278364913481251>.



Xueqing Zhang is currently pursuing her master's degree at the School of Geodesy and Geomatics, Wuhan University, China. She received her B.Sc. degree at the School of Geodesy and Geomatics, Wuhan University in 2022. Her research interests include multi-sensor fusion navigation technology.



Yuantai Zhang is currently pursuing his master's degree at the School of Geodesy and Geomatics, Wuhan University, China. He received his B.Sc. degree from the School of Geodesy and Geomatics, Wuhan University in 2022. His research interests include GNSS, graph optimization and its application in multi-sensor fusion.



Weijie Chen is currently pursuing his Ph.D. degree at the School of Geodesy and Geomatics, Wuhan University, China. He received his B.Sc. and M.Sc. degree at the School of Geodesy and Geomatics, Wuhan University, in 2020 and 2023 respectively. His research interests include multi-sensor fusion navigation and deep learning.



Xiaohong Zhang is currently a professor at Wuhan University. He obtained his B.Sc., M.Sc., and Ph.D. degrees with distinction in Geodesy and Engineering Surveying from Wuhan University in 1997, 1999, and 2002, respectively. His main research interests include PPP, PPP-RTK, GNSS/INS integration technology and its applications.



Feng Zhu is currently an associate professor at the School of Geodesy and Geomatics, Wuhan University, China. He obtained his B.Sc., Master, and Ph.D. degrees with distinction in Geodesy and Engineering Surveying at the School of Geodesy and Geomatics in Wuhan University in 2012, 2015 and 2019. His current research focuses on multi-sensor integration and its application in position and orientation systems and autonomous systems.



Zhuo Xu is currently pursuing his master's degree at the School of Geodesy and Geomatics, Wuhan University, China. He received his B.Sc. degree at the School of Geodesy and Geomatics, Wuhan University in 2022. His research interests include multi-sensor fusion navigation and technology.