This is a repository copy of *A Gaussian-generalized-inverse-Gaussian joint distribution based adaptive MSCKF for visual-inertial odometry navigation*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/191632/

Version: Accepted Version

Article:

A Gaussian-Generalized-Inverse-Gaussian Joint Distribution Based Adaptive MSCKF for Visual-Inertial Odometry Navigation

Chao Xue, Yulong Huang, *Member, IEEE,* Cheng Zhao, Xiaodong Li, Lyudmila Mihaylova, *Senior Member, IEEE*, Youfu Li, *Senior Member, IEEE*, Jonathon Chambers, *Fellow, IEEE*

Abstract—The visual-inertial odometry (VIO) navigation system plays an important role in providing accurate localization information in absolute navigation information-denied environments, such as indoors and obstruction-filled scenes. However, the working environment may be dynamic, such as due to illumination variations and texture changing in which case the measurement noise of the camera will be non-stationary, and thereby the VIO exhibits poor navigation using the fixed measurement noise covariance matrix (MNCM). This paper proposes an adaptive filter framework based on the multi-state constraint Kalman filter (MSCKF). Firstly, the MNCM is regarded as an identity matrix multiplied by a scalar MNCM coefficient which together with the state vector are jointly modeled as Gaussian-generalized-inverse-Gaussian distributed to achieve adaptive adjustment of the MNCM, from which the proposed adaptive filter framework for the VIO navigation system is derived. The proposed adaptive filter framework can theoretically employ a more accurate M-NCM during the filtering and thus is expected to outperform the traditional MSCKF. Secondly, the convergence, computational complexity and initial parameters influence analyses are given to illustrate the validity of the proposed framework. Finally, simulation and experimental studies are carried out to verify the theoretical and practical effectiveness and superiority of the proposed adaptive VIO filter framework, where the EuRoC datasets testing shows the proposed method is 22% and 29% better than the traditional MSCKF in position and orientation estimation, respectively.

Index Terms—Visual-inertial odometry navigation, adaptive filter, multi-state constraint Kalman filter, generalized-inverse-Gaussian distribution, non-stationary noises

I. INTRODUCTION

A. Background

The visual-inertial odometry (VIO) integrated navigation system can provide long-term stable and reliable navigation information, and thus is widely used in unmanned systems which

C. Xue, Y. Huang, C. Zhao, and X. Li are co-first authors. This work is partially supported by the National Natural Science Foundation of China under Grant Nos. 61903097 and 61773133. Corresponding author is Y. Huang.

C. Xue, Y. Huang, C. Zhao, X. Li are with the College of Intelligent Systems Science and Engineering, Harbin Engineering University, Harbin 150001, China (e-mail: xc_adjuster@163.com; heuedu@163.com; C_Zhao2021@163.com; LiXiaodongHEU@163.com).

L. Mihaylova is with the Department of Automatic Control and Systems Engineering, University of Sheffield, Sheffield, UK (e-mail: L.S.Mihaylova@sheffield.ac.uk).

Y. Li is with the Department of Mechanical Engineering, City University of Hong Kong, Hong Kong (e-mail: meyfli@cityu.edu.hk).

J. Chambers is with the College of Intelligent Systems Science and Engineering, Harbin Engineering University, and also with the School of Engineering, Newcastle University, Newcastle Upon Tyne, UK (e-mail: Jonathon.Chambers@ncl.ac.uk).
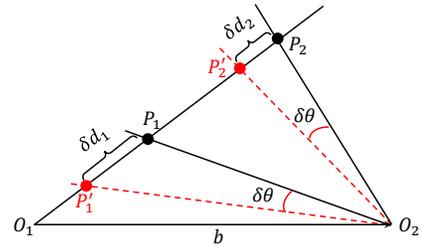

Fig. 1: Illustration for the time-varying MNCM caused by the triangulation. b denotes the baseline during two camera frames, and P_1, P_2 denote the true position of the feature points, and P_1', P_2' denote the positions estimated by the camera.

are required to work in the absolute navigation information-denied environment. The research on VIO thus has become very popular in the last ten years, and extensive literature on VIO navigation has appeared [1–10].

Despite existing VIO methods evolving towards better accuracy and less computation, insufficient attention has been paid to the VIO localization performance in dynamic scenarios, such as texture change, illumination variations and motion blur [11–13]. The above problems may lead to time-varying measurement noise statistics for the following reasons.

1) Time-varying temperature or external pressure may lead to time-varying internal parameters, which results in the inapplicability of the previously calibrated internal parameters [14].
2) The camera accuracy is related to the distance between the camera and the target. Taking the triangulation in Fig. 1 for example, the same parallax error angle $\delta\theta$ (device error) will lead to different position errors of feature points, i.e., δd_1 and δd_2, which leads to time-varying measurement noise statistics.
3) The matching errors of camera feature points also rely on the environment. For example, in an illumination changing and weakly textured environment with the camera moving very fast, different feature points are easy to identify as the same feature point, and this matching error can cause an additional measurement error, which will also result in time-varying measurement noise statistics.

In the above cases, inaccurate measurement statistics cognition may induce estimation accuracy deterioration or even divergence of VIO algorithms [25].

B. Related works

The existing VIO algorithms can be segregated into two categories, i.e., loosely-coupled VIO approaches and tightly-coupled VIO approaches. The loosely-coupled VIO approaches perform the visual odometry in the front-end and fuse the inertial measurement unit (IMU) information and visual information in the back-end. In this method, the VIO mainly focuses on the IMU information and regards the camera more as an auxiliary device, which will suffer from instability in the case of visual localization difficulty. In contrast, the tightly-coupled VIO methods accomplish the vehicle pose estimation using the IMU information while simultaneously correcting the inertial navigation system (INS) errors by employing the camera information, which can correct the drifting error of the camera and improve the overall localization accuracy. The tightly-coupled VIO methods consist of optimization-based methods and filter-based methods, where the former methods can achieve better accuracy performance while at the cost of greater computation, these include the PTAM [1], OKVIS [2], VINS-mono [3] and VINS-fusion [4]. The latter filter-based methods, which are mostly based on the Kalman filter (KF) [5–8] could be more computationally efficient but inferior in estimation accuracy compared with the optimization-based methods. Nonetheless, the filter-based methods have one unique advantage of being able to output accurate uncertainty of the estimates, which cannot be emulated by the optimization-based methods. On this account, the proposed work in this paper is a filter-based VIO algorithm.

The most typical representative method of the filter-based VIO algorithms is the traditional MSCKF (Tra-MSCKF) [5], which augments the most recent camera poses to the IMU state and utilizes multiple camera poses to constrain the positions of feature points in the three-dimensional space. After this, extensive works based on the Tra-MSCKF have been proposed. In 2013, Li and Mourikis proposed MSCKF 2.0 to address the estimation inconsistency problem in the Tra-MSCKF where the linearization error makes the unobservable azimuth information pseudo observable [6]. Sun proposed the stereo MSCKF algorithm based on a stereo camera [9], which improved the robustness of the Tra-MSCKF. Qiu et al. parameterized feature points by 1-dimensional inverse depth based on the Tra-MSCKF, and introduced zero-velocity detection into the algorithm to obtain a better performance [10].

The above nethods, however, have not considered the localization problem in dynamic scenarios, where the localization accuracy of VIO systems are severely affected, as discussed in Section I-A. To address this problem, some adaptive VIO methods have been proposed. Some works address low scene texture and motion blur [11, 13] by the direct processing of visual images at the front end to obtain robust feature tracking (removing mismatching feature points). However, it is difficult to remove all points that do not conform to the measurement model (we call them outliers). Thus, there are always missed outliers that enter the back-end and severely affect the fusion process. The above considerations motivate the development of adaptive back-end fusion algorithms to estimate unknown and time-varying MNCM.

The MNCM estimation method has been developed into four categories since first proposed in [21], i.e., the Bayesian methods [16–20], correlation methods [21, 22], maximum likelihood methods [23, 24], and covariance matching methods [27, 28]. The Sage-Husa adaptive KF belongs to the covariance matching methods, which performs recursive maximum a posterior (MAP) estimation of the noise statistics [27]. However, its estimation convergence cannot be ensured, which may cause the filter to diverge [15]. The proposed innovation-based adaptive KF in [25] belongs to the maximum likelihood methods, which estimates the noise statistics by analyzing the white innovation sequence [25]. However, it requires a long window to collect enough data to obtain accurate noise statistics estimates, which leads to a large lag in noise statistics estimation. The measurement difference method is a kind of correlation method, which computes the measurement estimate from other measurements to avoid the state estimate computation [22]. This method, however, has the same problem of large lag in noise statistics estimation as [25]. The Bayesian methods can be divided into those which adopt a Gaussian mixture model and the VB methods [15]. The Gaussian mixture Bayesian methods utilise multiple models to achieve adaption [19], which suffers from large computation burden. The proposed algorithm in [17] belongs to the VB methods, which models the MNCM as an inverse Wishart (IW) distribution and jointly estimates the MNCM and the state vector. However, this method together with the above methods are all based on an assumption that the MNCM has a constant dimension, and are thereby not suitable for the VIO navigation system with time-varying measurement dimensions. Some robust adaptive KFs can adapt to time-varying measurement dimensions such as the Huber-based KF (HKF) [29] and the robust Student's t-based KF (RSTKF) [30], because they do not rely on the historical measurements to achieve adaptive adjustment, which also makes them less effective in the case of continuous changing environments. However, there has not been any record of them being used in VIO systems. For VIO back-end fusion processing for dynamic scenarios, Yue proposed a robust adaptive filter based on fuzzy logic [31], where the trace of the calculated innovation covariance matrix serves as the input of the fuzzy inference to obtain the adjusted MNCM. Shen proposed the degree of abnormality to evaluate the ego-motion uncertainty [12], where a robust DoA-based VIO algorithm has been proposed. Both of these algorithms rely heavily on manual adjustment of parameters, which limits their accuracy and practical applicability. Recently, Zhang has proposed a VB-based adaptive stereo MSCKF (VB-MSCKF) to accommodate time-varying MNCM [13]. However, the matrix ℓ_2 norm is used to deal with measurement dimension mismatching, which will cause large approximation errors. The proposed adaptive algorithm belongs to the Bayesian methods type and has the advantages of good theoretical basis, computational efficiency, online implementation convenience, and time-varying MNCM tolerance.

C. Motivations for this paper

According to the above discussions about related works, we find that the existing adaptive methods are not suitable

for non-stationary dynamic scenarios because some of them may require an intricate parameters tuning process so that they cannot adapt well to different environments. Despite the potential effectiveness of online estimation of the time-varying MNCM, such as using the existing VB-MSCKF [13], the following drawbacks limit its estimation accuracy

- The mean-field approximation is the core technique to deal with the correlation between the state vector and auxiliary parameters but it is concurrently the main error source of the existing VB-MSCKF, which may give rise to large errors when the state vector and auxiliary parameters are severely coupled.
- In the VB-MSCKF method, the ℓ_2 norm of the IW distribution scale matrix at the previous time is extracted and taken as the prior value of its diagonal elements at the current time to realize recursive calculation of the IW distribution parameters. This operation is generally not accurate and may cause large approximation errors, as will be shown in Section V and Section VI.

The above considerations motivate us that modeling the MNCM as IW distributed may not be efficient to address the non-stationary measurement noise statistics problem.

D. Contributions and organization of this paper

In this paper, we propose an adaptive VIO framework in which we regard the MNCM as an identity matrix multiplied by a scalar parameter which is denoted as the MNCM coefficient (MNCMC), and model the state vector and the MNCMC as Gaussian-generalized-inverse-Gaussian (NGIG[1]) distributed to achieve adaptive adjustment of the MNCM. The contributions of this work are as follows.

- We propose a generalized adaptive NGIG-based VIO filter framework which can adapt to dynamic environments.
- The convergence proof and computational complexity analyses are presented to show the theoretical effectiveness of the proposed adaptive VIO filter framework.
- Various simulation and experimental tests are carried out to verify the validity and superiority of the proposed adaptive filter framework in engineering applications.

The structure of the remainder of the paper is as follows. Section II gives the state space model and introduction to the NGIG distribution in this paper. We present the proposed adaptive NGIG-based VIO framework in Section III. Section IV includes some discussions about the numerical calculation problem encountered and gives two alternative methods and corresponding properties. Simulation and experimental tests and comparisons are carried out in Section V and Section VI, respectively. This paper is concluded in Section VII.

II. PROPOSED NGIG JOINT DISTRIBUTION MODELLING

A. Linear state space model

We consider the following state space model

$$\begin{cases} \mathbf{x}_k = \boldsymbol{\Phi}_k \mathbf{x}_{k-1} + \mathbf{D}_k \mathbf{w}_k \\ \mathbf{z}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{n}_k \end{cases} \tag{1}$$

[1]The Gaussian distribution can also be called the normal distribution, so it is abbreviated as NGIG here.

TABLE I: Acronyms and Nomenclatures

Notations	Definitions
INS, IMU	Inertial Navigation System, inertial measurement unit
dof, PDF	Degree of freedom, probability density function
MAP, VIO	Maximum a posterior, visual-inertial odometry
MNCM	Measurement noise covariance matrix
MNCMC	MNCM coefficient
KF, EKF	Kalman filter, extended KF
MSCKF	Multi-state constraint KF
Tra-MSCKF	Traditional MSCKF
Inf-MSCKF	Inflated MSCKF
HKF	Huber-based KF
RSTKF	Robust Student's t-based KF
VB	Variational Bayesian
VB-MSCKF	VB-based adaptive stereo MSCKF
ANGIG-KF	Adaptive NGIG-based Kalman fitering framework
ANGIG-KF1	MAP-based ANGIG-KF
ANGIG-KF2	Approximate mean-based ANGIG-KF
RMSE	Root mean square error
PRMSE, ORMSE	Position RMSE, orientation RMSE
NEES	Normalized estimation error square
PE, OE	Position error, orientation error
APEs, AOEs	Average PEs, average OEs
IG, IW	Inverse Gamma, inverse Wishart
GIG, NGIG	Generalized inverse Gaussian, Gaussian-GIG
$x_k, \bar{x}_k, \hat{x}_k$	Discrete-time random variable at time k and
$\text{GIG}(\cdot; a, b, r)$	GIG PDF with parameters a, b, r
$\mathcal{N}(\cdot; \boldsymbol{\mu}, \mathbf{P})$	Multivariate Gaussian PDF with mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{P}$
$\text{NGIG}(\cdot, \cdot; \boldsymbol{\mu}, \mathbf{P}, a, b, r)$	NGIG joint distribution with mean $\boldsymbol{\mu}$, scale matrix $\mathbf{P}$ and dof parameters a, b, r
$\text{tr}\{\cdot\}, (\cdot)^{-1}, (\cdot)^T$	Trace, inverse and transpose operations
$\lvert \cdot \rvert, \lVert \cdot \rVert$	Determinant operation and ℓ_2 norm operations
$\text{E}\{\cdot\}, \text{Var}\{\cdot\}$	Mathematical expectation and variance operations
$\log(\cdot), \exp(\cdot)$	Logarithm and exponential operations
$\otimes$	Quaternion multiplication operation
$\mathbf{I}_n, \mathbf{0}_n$	Identity matrix and null matrix of dimension $n \times n$
:=	Defined as
$(\mathbf{x})_{1:n}$	Vector consisting of the first n elements of vector $\mathbf{x}$

where $\boldsymbol{\Phi}_k$ and $\mathbf{D}_k$ denote the state transition matrix and state noise driven matrix, respectively, and $\mathbf{n}_k$ denotes the measurement noise vector at time k. Moreover, we note that both $\mathbf{D}_k \mathbf{w}_k$ and $\mathbf{n}_k$ obey zero mean Gaussian distributions, i.e., $\mathbf{D}_k \mathbf{w}_k \sim \mathcal{N}(\mathbf{w}_k; \mathbf{0}, \mathbf{D}_k \mathbf{Q}_k \mathbf{D}_k^T)$ and $\mathbf{n}_k \sim \mathcal{N}(\mathbf{n}_k; \mathbf{0}, \mathbf{R}_k)$, where $\mathbf{R}_k = \lambda_k \mathbf{I}_{m_k}$. We assume that λ_k is slowly changing over time. However, a fixed MNCM is used in the Tra-MSCKF, which is not appropriate and even possibly leads to divergence of the Tra-MSCKF. Thus, the state and the MNCM are jointly modeled as NGIG distributed to achieve adaptive adjustment of the MNCM. Next, the NGIG distribution will be introduced and some of its properties will be specified.

B. NGIG joint distribution

Before introducing the NGIG joint distribution, we first introduce the GIG distribution. The PDF of a GIG distributed random variable λ is written as [32, 33]

$$\text{GIG}(\lambda; a, b, r) = \frac{(a/b)^{r/2}}{2K_r(\sqrt{ab})} \lambda^{(r-1)} e^{-(a\lambda + b/\lambda)/2} \tag{2}$$

where $a > 0$, $b > 0$ and r is a real parameter, and the definition of Bessel function $K_r(\cdot)$ can be found in [34].

As a family of generalized distributions, the GIG distribution subsumes the inverse Gamma (IG), Gamma, exponential

and many other practical distributions as its special cases. Let's temporarily reserve the reasons why we use the GIG distribution rather than its special cases in the joint inference and focus on the NGIG distribution introduced next. With (2), the NGIG distributed joint random variables $(\mathbf{x}, \lambda)$ can be formulated as follows

$$p(\mathbf{x}, \lambda) = \underbrace{\mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \lambda\boldsymbol{\Sigma})}_{p(\mathbf{x}|\lambda)} \underbrace{\text{GIG}(\lambda; a, b, r)}_{p(\lambda)} \quad (3)$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ denote the conditional mean and conditional scale matrix of the state vector, respectively, and a, b, r are parameters of the GIG distributed MNCMC. Next, the mean and variance of the marginalized state vector $\mathbf{x}$ from the NGIG joint distribution will be given for subsequent derivations.

Remark 1. *To the best of the author's knowledge, this joint modeling of the state vector and parameters can be tracked back to [16], in which the state and parameters are jointly modeled as Gaussian-multiple-IG distributed. However, the state and the parameters are assumed as independent, which may give rise to large errors when the state vector and the parameters are severely coupled. Hence, one of the motivations of this paper is to retain the potential correlation between the state vector and the MNCMC. However, how to retain such correlation becomes another problem. In this paper, we adopt a rather simple method to establish the correlation between the state vector and the MNCMC, i.e., we multiply the same scalar λ by the covariance (which is actually called the scale matrix rather than the covariance) of the state vector. The benefit of this modeling is that a closed-form posterior solution can be derived because if we multiply the covariance of a Gaussian distribution by a coefficient, the GIG distribution is one of the conjugate distributions of this coefficient. Some properties of this modeling will be specified next.*

Proposition 1. *The mean and variance of the marginalized $\mathbf{x}$ from the NGIG distribution are calculated as*

$$\begin{cases} \mathrm{E}\{\mathbf{x}\} = \boldsymbol{\mu}, \ \mathrm{Var}\{\mathbf{x}\} = \hat{\lambda}\boldsymbol{\Sigma} \\ \mathrm{E}\{\lambda\} = \dfrac{\sqrt{b}K_{r+1}(\sqrt{ab})}{\sqrt{a}K_r(\sqrt{ab})} \end{cases} \quad (4)$$

Proof. See Appendix A. $\qquad\square$

It is notable that when the state vector and MNCMC are modeled as NGIG distributed, the VIO navigation system can use the time-varying MNCM rather than a constant MNCM, which is more suitable for a non-stationary environment. The accuracy of this MNCM is determined by the accuracy of the distribution of the MNCMC which will be corrected in the time update and measurement update. Furthermore, we want to elucidate that the marginalized state vector from this joint distribution obeys a more heavy-tailed distribution than Gaussian, which will make the proposed adaptive framework robust to measurement outliers to some extent. We will support this point by means of the following proposition.

Proposition 2. *The tail corresponding to the distribution obeyed by the marginalized state vector decays slower than Gaussian. The logarithm PDF of the marginalized state vector*
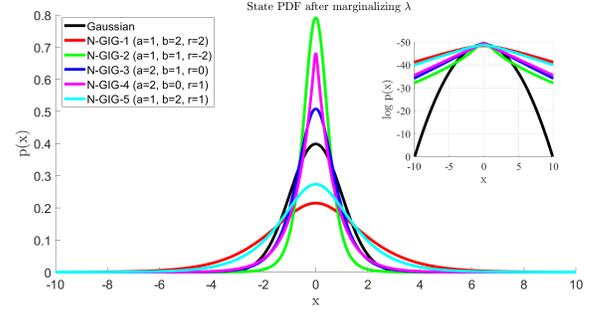

Fig. 2: The marginalized state PDF under different GIG distribution parameters.

and the Gaussian distributed state vector when $\mathbf{x}$ tends to infinity are, respectively, given as follows

$$\begin{cases} \log p(y) = -\log y - \sqrt{a}y^2 + c_y \\ \log p_{\mathcal{N}}(y) = -0.5y^4 + c_{y_N} \end{cases}$$

where $y = \sqrt{\mathbf{x}^T\mathbf{x}}/4$, and $p_{\mathcal{N}}(y)$ denotes the Gaussian PDF.

Proof. See Appendix B. $\qquad\square$

Proposition 2 indicates that the state vector obeys a heavier tail than Gaussian. To be more intuitive, the marginalized state PDF is plotted in Fig. 2, where the subfigure on the upper right corner is the logarithm marginalized state PDF. It can be observed from Fig. 2 that the tails of the marginalized state PDFs under all selected NGIG distribution parameters are much heavier than Gaussian. Besides, different degrees of robustness can be achieved through different GIG distribution parameters. Overall, as opposed to assuming the outliers are infrequent, heavy-tailed noises suppose that they occur with a relatively high probability, which is more in line with practical experience. As a result, the NGIG distribution has the advantage of resisting measurement outliers, which will make the proposed adaptive KF robust to some extent.

Remark 2. *Robustness is one of the most important properties that filtering algorithms for VIO systems should possess. Notoriously, mismatching of feature points in adjacent frames often occurs when VIO systems are operating, which gives rise to localization accuracy degradation. Thus, detecting and eliminating these outliers are necessary. On the one hand, eliminating outliers can improve the quality of obtained measurements and render the estimation results more accurate. On the other hand, abundant computation can be saved by reducing the measurement dimension. Generally, the chi-square test is utilized to perform this task. Nonetheless, selecting an appropriate chi-square threshold can be challenging. With a large threshold, the chi-square test may miss outliers which will deteriorate the estimation accuracy. While the chi-square test with a small threshold will be conservative so that extensive measurement data will be eliminated. Therefore, selecting an appropriate chi-square threshold requires much engineering experience and generally, can not be adapted to different scenarios. In this setting, the robustness of the filtering algorithm can allow the chi-square test to be appropriately conservative because the algorithm itself can provide additional protection for the system by reducing the impact of*

stray outliers. As a result, not only can the estimation accuracy be improved, but also much parameter tuning time is saved.

To this end, we express the reasons for using the GIG distribution rather than its special cases, such as the IG distribution or the Gamma distribution. On the one hand, we intend to derive a generalized framework that considers not only IG or Gamma distribution, but also all special cases of the GIG distribution. The IG or Gamma distribution are only special cases when the parameters of the GIG distribution satisfy a certain relationship. Hence, the parameters of the GIG distribution are more liberal than its special cases, which is theoretically more valuable. On the other hand, in practical applications, different interferences may occur in the case of different environments, which results in different types of heavy-tailed distributions. The generalized NGIG distribution can include more non-Gaussian cases than its special cases, and thereby accommodate more complex heavy-tailed cases.

To sum up, the NGIG distribution has an attractive property of being able to resist outliers. Hence, in the ensuing part, the state vector and MNCMC are jointly modeled as NGIG distributed to achieve an adaptive adjusting of the MNCMC, and simultaneously, achieve an accurate estimation of the state vector, from which the proposed adaptive NGIG-based Kalman filtering framework (ANGIG-KF) will be derived for VIO systems to accommodate non-stationary environments. For brevity, we simplify $\mathbf{D}_k \mathbf{Q}_k \mathbf{D}_k^T$ as $\mathbf{Q}_k$ in the remainder to denote the state noise covariance matrix.

III. AN ADAPTIVE KF BASED ON NGIG MODELING

A. The proposed ANGIG-KF

In this part, the posterior state vector and the MNCMC at time $k-1$ are modeled as NGIG distributed as follows.

$$
\begin{aligned}
&p(\mathbf{x}_{k-1}, \lambda_{k-1}|\mathbf{z}_{1:k-1}) \qquad\qquad\qquad (5)\\
&= \underbrace{\mathcal{N}(\mathbf{x}_{k-1}; \hat{\mathbf{x}}'_{k-1}, \lambda_{k-1}\hat{\mathbf{P}}'_{k-1})}_{p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1}, \lambda_{k-1})} \underbrace{\mathrm{GIG}(\lambda_{k-1}; \hat{a}_{k-1}, \hat{b}_{k-1}, \hat{r}_{k-1})}_{p(\lambda_{k-1}|\mathbf{z}_{1:k-1})}
\end{aligned}
$$

where $p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1}, \lambda_{k-1})$ denotes the conditional posterior PDF of the state vector, and $\hat{\mathbf{x}}'_{k-1}$ and $\hat{\mathbf{P}}'_{k-1}$ denote the conditional posterior mean and scale matrix of the state vector at time $k-1$, respectively, whose superscripts are used to distinguish themselves from the mean and covariance of the marginalized state vector, and $\hat{a}_{k-1}, \hat{b}_{k-1}, \hat{r}_{k-1}$ denote the posterior parameters of the GIG distribution at time $k-1$, respectively.

Theorem 1. *The joint prior PDF can be approximated as*

$$
p(\mathbf{x}_k, \lambda_k|\mathbf{z}_{1:k-1}) \approx \mathcal{N}(\mathbf{x}_k; \bar{\mathbf{x}}_k, \bar{\mathbf{P}}_k(\lambda_k))\mathrm{GIG}(\lambda_k; \bar{a}_k, \bar{b}_k, \bar{r}_k)
$$
$$
(6)
$$

where

$$
\begin{cases}
\bar{\mathbf{x}}_k = \mathbf{\Phi}_k\hat{\mathbf{x}}'_{k-1}, \ \ \bar{\mathbf{P}}_k(\lambda_k) = \mathbf{\Phi}_k\lambda_k\hat{\mathbf{P}}'_{k-1}\mathbf{\Phi}_k^T + \mathbf{Q}_k \\
\bar{a}_k = \rho\hat{a}_{k-1}, \ \ \bar{b}_k = \rho\hat{b}_{k-1}, \ \ \bar{r}_k = \rho\hat{r}_{k-1}
\end{cases}
\quad (7)
$$

and $\bar{\mathbf{x}}_k$ and $\bar{\mathbf{P}}_k$ denote the prior mean and scale matrix at time k, respectively, and $\bar{a}_k, \bar{b}_k, \bar{r}_k$ denote the posterior parameters of the GIG distribution at time k, respectively.

Proof. See Appendix C. $\qquad\qquad\square$

Unfortunately, the form of the joint distribution in (6) is not exactly the same as that in (5), which renders the posterior joint density not in an analytical form. It is known that to make the joint posterior PDF at time k analytically solvable, $\bar{\mathbf{P}}_k$ must be fully scaled by λ_k. However, it is seen from (7) that $\mathbf{Q}_k$ is not exactly scaled by λ_k, which makes $\bar{\mathbf{P}}_k$ not fully scaled. To this end, we propose to approximate $\bar{\mathbf{P}}_k$ as follows.

$$
\begin{cases}
\bar{\mathbf{P}}_k \approx \mathbf{\Phi}_k\lambda_k\hat{\mathbf{P}}'_{k-1}\mathbf{\Phi}_k^T + \dfrac{\lambda_k}{\theta_k}\mathbf{Q}_k := \lambda_k\breve{\mathbf{P}}_k(\theta_k) \\
s.t. \qquad \theta_k = \mathrm{E}\{\lambda_k\}
\end{cases}
\quad (8)
$$

where θ_k is an auxiliary parameter which takes the value of the mean of λ_k. Employing (8), we can reformulate (6) as

$$
p_{\theta_k}(\mathbf{x}_k, \lambda_k|\mathbf{z}_{1:k-1}) \approx \mathcal{N}(\mathbf{x}_k; \bar{\mathbf{x}}_k, \lambda_k\breve{\mathbf{P}}_k(\theta_k))\mathrm{GIG}(\lambda_k; \bar{a}_k, \bar{b}_k, \bar{r}_k)
$$
$$
(9)
$$

where $p_{\theta_k}(\cdot)$ represents the PDF depending on θ_k.

Remark 3. *It can be seen from (8) that we make an approximation by assuming θ_k as an unknown parameter rather than a random variable. From another point of view, we can still consider θ_k as a random variable but approximately simplify its distribution into a Dirac delta function, and then match the first moment, i.e, the mean of the approximated distribution and the distribution obeyed by λ_k, i.e., the GIG distribution in (9). The mean is a minimum variance estimate of λ_k, which utilizes the whole density information of λ_k and thus is a good approximation. Moreover, this is also the simplest approximation method, which achieves a balance between computational burden and estimation accuracy.*

Based on the assumption that the MNCMC is GIG distributed, the likelihood PDF can be formulated as the following hierarchical Gaussian form

$$
\begin{cases}
p(\mathbf{z}_k|\mathbf{x}_k, \lambda_k) = \mathcal{N}(\mathbf{z}_k; \mathbf{H}_k\mathbf{x}_k, \lambda_k\mathbf{I}_{m_k}) \\
p(\lambda_k|\mathbf{z}_{1:k-1}) = \mathrm{GIG}(\lambda_k; \bar{a}_k, \bar{b}_k, \bar{r}_k)
\end{cases}
\quad (10)
$$

where $p(\mathbf{z}_k|\mathbf{x}_k, \lambda_k)$ denotes the conditional likelihood PDF.

Theorem 2. *Exploiting (9), (10) and Bayes' rule, the conditional posterior PDF of $\mathbf{x}_k$ can be calculated as follows*

$$
p_{\theta_k}(\mathbf{x}_k|\mathbf{z}_{1:k}, \lambda_k) = \mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}'_k(\theta_k), \lambda_k\hat{\mathbf{P}}'_k(\theta_k)) \qquad (11)
$$

where the mean vector $\hat{\mathbf{x}}'_k(\theta_k)$ and the covariance matrix $\hat{\mathbf{P}}'_k(\theta_k)$ are, respectively, calculated as

$$
\begin{cases}
\hat{\mathbf{x}}'_k(\theta_k) = \bar{\mathbf{x}}_k + \mathbf{K}_k(\theta_k)\tilde{\mathbf{z}}_k \\
\hat{\mathbf{P}}'_k(\theta_k) = (\mathbf{I}_n - \mathbf{K}_k(\theta_k)\mathbf{H}_k)\hat{\mathbf{P}}_{k-1} \\
\tilde{\mathbf{z}}_k = \mathbf{z}_k - \mathbf{H}_k\bar{\mathbf{x}}_k \\
\breve{\mathbf{P}}_k^{\mathbf{zz}}(\theta_k) = \mathbf{H}_k\breve{\mathbf{P}}_k(\theta_k)\mathbf{H}_k^T + \mathbf{I}_{m_k} \\
\mathbf{K}_k(\theta_k) = \tilde{\mathbf{P}}_k\mathbf{H}_k^T(\mathbf{P}_k^{\mathbf{zz}}(\theta_k))^{-1}
\end{cases}
\quad (12)
$$

where $\mathbf{K}_k(\theta_k)$ denotes the Kalman gain, and $\tilde{\mathbf{z}}_k$ denotes the innovation vector, and $\breve{\mathbf{P}}_k^{\mathbf{zz}}(\theta_k)$ denotes the modified innovation covariance matrix, and $\mathbf{I}_{m_k}$ denotes the identity matrix with time-correlated dimension m_k.

Proof. See Appendix D. $\qquad\qquad\square$

Remark 4. *It can be seen from (9) and (11) that the state vector $\mathbf{x}_k$ is always correlated with the MNCMC λ_k in both*

TABLE II: Time update of the proposed ANGIG-KF

Function 1: $\left(\bar{\mathbf{x}}_k, \check{\mathbf{P}}_k, \check{\mathbf{P}}'_k\right) = \text{TU}\left(\hat{\mathbf{x}}'_{k-1}, \hat{\mathbf{P}}'_{k-1}, \mathbf{Q}_k, \mathbf{\Phi}_k, \hat{\theta}_{k-1}\right)$
Calculate intermediate prior parameters of $\mathbf{x}_k$:
1. $\check{\mathbf{P}}'_k = \mathbf{\Phi}_k \hat{\mathbf{P}}'_{k-1} \mathbf{\Phi}_k$
2. $\bar{\mathbf{x}}_k = \mathbf{\Phi}_k \hat{\mathbf{x}}_{k-1}, \quad \check{\mathbf{P}}_k = \check{\mathbf{P}}'_k + \mathbf{Q}_k / \hat{\theta}_{k-1}$

the time update and the measurement update. Therefore, the essential difference between the VB-based adaptive state estimation methods and the proposed method is whether to ignore the correlation between the state vector and the parameter. In the VB-based methods, the correlation between the state vector and the parameters is forcibly ignored. Although the later fixed-point iteration can compensate for some approximate errors, it may still lead to estimation accuracy degradation. In contrast, the proposed adaptive method in this paper reasonably utilizes the correlation between the state vector and MNCMC, although new errors are introduced in the time update. Discussing the advantages and disadvantages of these two kinds of adaptive methods is beyond the scope of this paper. Nevertheless, in terms of the VIO navigation system studied in this paper, the proposed adaptive algorithm can achieve better estimation accuracy than VB-based methods, as will be shown in Section V-C, VI-B, and VI-C.

Theorem 3. *Employing (9)-(11) and (33), the posterior PDF of auxiliary random variable λ_k can be calculated as follows.*

$$p_{\theta_k}(\lambda_k|\mathbf{z}_{1:k}) = \text{GIG}(\lambda_k; \hat{a}_k, \hat{b}_k, \hat{p}_k) \quad (13)$$

where

$$\begin{cases} \hat{a}_k = \bar{a}_k, \ \hat{b}_k = \bar{b}_k + \Delta_k, \ \hat{r}_k = \bar{r}_k - 0.5m_k \\ \Delta_k = tr\{\tilde{\mathbf{P}}_k^{\mathbf{zz}}(\check{\mathbf{P}}_k^{\mathbf{zz}}(\theta_k))^{-1}\}, \ \tilde{\mathbf{P}}_k^{\mathbf{zz}} = \tilde{\mathbf{z}}_k\tilde{\mathbf{z}}_k^T \end{cases} \quad (14)$$

Proof. See Appendix E. □

Exploiting (11) and (13), the posterior joint PDF of the state vector and auxiliary random variable at time k can be formulated as follows.

$$p(\mathbf{x}_k, \lambda_k|\mathbf{z}_{1:k}) = \mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}'_k(\theta_k), \lambda_k\hat{\mathbf{P}}'_k(\theta_k))\text{GIG}(\lambda_k; \hat{a}_k, \hat{b}_k, \hat{r}_k) \quad (15)$$

Using (12)–(15), we can obtain the mean and covariance of the marginalized state vector as follows

$$\begin{cases} \hat{\mathbf{x}}_k = \hat{\mathbf{x}}'_k(\theta_k), \ \hat{\mathbf{P}}_k = \hat{\lambda}_k\hat{\mathbf{P}}'_k(\theta_k) \\ \hat{\lambda}_k = \dfrac{\sqrt{\hat{b}_k}K_{(\hat{r}_k+1)}\left(\sqrt{\hat{a}_k\hat{b}_k}\right)}{\sqrt{\hat{a}_k}K_{(\hat{r}_k)}\left(\sqrt{\hat{a}_k\hat{b}_k}\right)} \end{cases} \quad (16)$$

The determination of θ_k is the last basic work of the proposed adaptive filter framework. We have previously identified θ_k as the mean of λ_k, and we will discuss this in more detail. In the time update, only the prior information of λ_k is available. Hence, θ_k can be approximated as the prior mean of λ_k, i.e, $\bar{\theta}_k = \text{E}\{\lambda_k|\mathbf{z}_{1:k-1}\}$. Whereas in the measurement update, we provide two methods to determine the value of θ_k as follows

- *Non-iterative:* θ_k takes the same value as that in the time update, i.e., $\bar{\theta}_k$.

TABLE III: Measurement update of the proposed ANGIG-KF

Function 2: $\left(\hat{\mathbf{x}}_k, \hat{\mathbf{P}}_k, \hat{\mathbf{x}}'_k, \hat{\mathbf{P}}'_k, \hat{a}_k, \hat{b}_k, \hat{r}_k, \hat{\theta}_k\right) = \text{MU}\left(\bar{\mathbf{x}}_k, \check{\mathbf{P}}_k,\right.$
$\left. \check{\mathbf{P}}'_k, \mathbf{H}_k, M, \rho, \epsilon, \mathbf{Q}_k, \hat{\theta}_{k-1}, \hat{a}_{k-1}, \hat{b}_{k-1}, \hat{r}_{k-1}, \tilde{\mathbf{z}}_k, flag\right)$
If $flag = 1$:
Prior parameters propagation of the GIG distribution:
1. $\bar{a}_k = \hat{a}_{k-1}, \bar{b}_k = \rho\hat{b}_{k-1}, \bar{r}_k = \rho\hat{r}_{k-1}$
2. $\bar{\theta}_k = \text{E}(\lambda_k|\mathbf{z}_{1:k-1})$
3. $\tilde{\mathbf{P}}_k^{\mathbf{zz}} = \tilde{\mathbf{z}}_k\tilde{\mathbf{z}}_k^T$
Iteration initialization:
4. $\hat{\theta}_k^{(0)} = \bar{\theta}_k, \hat{a}_k^{(0)} = \bar{a}_k, \hat{b}_k^{(0)} = \bar{b}_k, \hat{r}_k^{(0)} = \bar{r}_k$
Fixed-point iteration of the unknown parameter θ_k:
for $i = 0 : M - 1$ **do**
5. $\check{\mathbf{P}}_k^{(i+1)} = \check{\mathbf{P}}'_k + \mathbf{Q}_k/\theta_k^{(i)}$
6. $\check{\mathbf{P}}_k^{\mathbf{zz},(i+1)} = \mathbf{H}_k\check{\mathbf{P}}_k^{(i+1)}\mathbf{H}_k^T + \mathbf{I}_{m_k}$
7. $\Delta_k^{(i+1)} = \text{tr}\{\tilde{\mathbf{P}}_k^{\mathbf{zz}}(\check{\mathbf{P}}_k^{\mathbf{zz},(i+1)})^{-1}\}$
8. $\hat{a}_k^{(i+1)} = \bar{a}_k, \ \hat{b}_k^{(i+1)} = \bar{b}_k + \Delta_k^{(i+1)}, \ \hat{r}_k^{(i+1)} = \bar{r}_k - 0.5m_k$
9. $\hat{\theta}_k^{(i+1)} = \text{E}^{(i)}(\lambda_k|\mathbf{z}_{1:k})$
10. If $\left|\theta_k^{(i+1)} - \theta_k^{(i)}\right|/\theta_k^{(i)} \leq \epsilon$, terminate the iteration
end for
Analytical measurement update of the state vector:
11. $\hat{\theta}_k = \theta_k^{(N_{min})}$
12. $\mathbf{K}_k = \tilde{\mathbf{P}}_k\mathbf{H}_k^T(\check{\mathbf{P}}_k^{\mathbf{zz},(N_{min})})^{-1}$
13. $\hat{\mathbf{x}}'_k = \bar{\mathbf{x}}_k + \mathbf{K}_k\tilde{\mathbf{z}}_k$
14. $\hat{\mathbf{P}}'_k = (\mathbf{I}_n - \mathbf{K}_k\mathbf{H}_k)\check{\mathbf{P}}_k^{(N_{min})}$
Actual posterior state estimate and covariance calculation:
15. $\hat{\mathbf{x}}_k = \hat{\mathbf{x}}'_k, \ \hat{\mathbf{P}}_k = \hat{\theta}_k\hat{\mathbf{P}}'_k$
Else: $\hat{\theta}_k = \hat{\theta}_{k-1}, \hat{\mathbf{x}}'_k = \bar{\mathbf{x}}_k, \hat{\mathbf{P}}'_k = \check{\mathbf{P}}_k, \hat{a}_k = \bar{a}_k, \hat{b}_k = \bar{b}_k, \hat{r}_k = \bar{r}_k,$
$\quad \hat{\mathbf{x}}_k = \hat{\mathbf{x}}'_k, \hat{\mathbf{P}}_k = \hat{\theta}_k\hat{\mathbf{P}}'_k$

- *Iterative:* θ_k is calculated as the posterior mean of the auxiliary random variable λ_k.

The non-iterative method does not absorb the measurement information, the advantage of which lies in its ability to reduce the computational complexity while retaining a satisfactory estimation accuracy. Furthermore, the iterative method attempts to use the measurement information to improve the estimation accuracy. However, we can observe from (14) that parameter θ_k and the posterior PDF of λ_k are coupled, which results in the difficulty in determining θ_k. Therefore, the fixed-point iteration method is utilized to obtain an approximate solution of θ_k. That is to say, in the (i)-th iteration, $\hat{\theta}_k^{(i-1)}$ at the $(i-1)$-th iteration participates in the posterior parameters updating in (14). Then $\hat{\theta}_k^{(i)}$ at the (i)-th iteration is calculated as the posterior mean of λ_k using its updated PDF at the i-th iteration, i.e., $\text{E}^{(i)}\{\lambda_k|\mathbf{z}_{1:k}\}$. Obviously, the iterative method requires the knowledge of the posterior information of λ_k. Therefore, at least one iteration is needed to ensure that $\hat{\theta}_k$ absorbs the measurement information to refine its accuracy. The iteration will however consume many computing resources. Thus, the iterative method can provide better estimation accuracy but at the cost of a greater computational burden. Nevertheless, later we will show that the iterative sequence of the proposed algorithm can achieve convergence with few iterations, which would prevent the iterative-based method from being too computationally expensive. Further comparisons of these two methods will be illustrated in Section V-A.

The proposed robust Kalman filtering framework has now been presented, which is summarized in Table II and Table III, where $flag$ is employed to determine whether to perform an extended KF (EKF) update, as will be detailed in Table IV. With (4) and (15), we can calculate the joint posterior PDF of the state vector and MNCMC recursively.

B. Implementation details of the whole improved MSCKF based on the proposed ANGIG-KF

This part provides the detailed implementation steps of the improved MSCKF based on the proposed ANGIG-KF to obtain a clearer view of the whole improved MSCKF algorithm. We highlight that the proposed method can be easily extended to the EKF framework, and the only difference between the Tra-MSCKF and the proposed method lies in the filter algorithm, where the former uses the traditional KF and the latter employs the proposed ANGIG-KF. The detailed implementations of the whole improved MSCKF using the proposed ANGIG-KF are shown in Table IV.

Remark 5. *There are several implementation details to note:*
*(1) Notice from step 8 in **Function 2** that the parameter a_k does not change in the measurement update. Multiplying $\hat{a}_{k-1}$ by ρ which is less than 1, is not appropriate because it will approach zero quickly. Thus, a_k will not propagate like b_k and r_k but remain constant in the whole ANGIG-KF. The influence of the initial parameters selection on the proposed algorithm will be discussed in detail in Section IV-C.*
*(2) The chi-square test will be performed at every iteration when calling **Function 2**. Moreover, $\mathbf{H}_k$ and $\tilde{\mathbf{z}}_k$ will be also reconstructed at each iteration.*
(3) Speaking precisely, the frequencies of the time update and measurement update are not exactly equal because the calculation frequency of the IMU is different from the camera frequency. In this paper, we assume the parameters of the GIG distribution do not propagate when there is no measurement update. In other words, parameters a_k, b_k and r_k are propagated and updated only when the measurement update occurs. It is reasonable because these parameters are related to the camera accuracy which is uncorrelated with the IMU.
(4) Considering that $\hat{\theta}_k$ is a reasonable estimate of λ_k, we set the posterior estimated $\hat{\theta}_k$ as the estimated MNCMC for subsequent analyses.

Note that the mean operation in step 2 and step 9 of **Function 2** is significant during the implementation of the whole improved MSCKF. Unfortunately, the calculation of this mean requires calculating the Bessel function in (4), which suffers from a numerical calculation problem under certain conditions. In this case, the mean of λ_k is unavailable. In the next section, two alternative methods are given to substitute the mean, and some properties of the proposed ANGIG-KF based on these two methods are specified.

IV. DISCUSSIONS ABOUT THE ANGIG-KF BASED ON ALTERNATIVE METHODS

In this section, the discussions will be divided into two parts. In the first part, the numerical calculation problem caused by
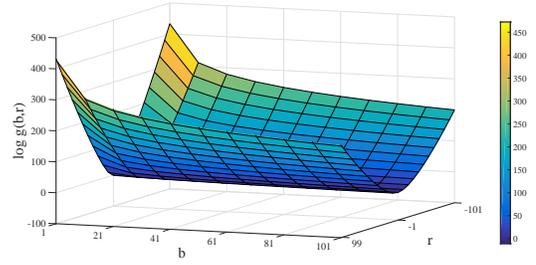

Fig. 3: The value of the logarithm of $g(b,r)$ with respect to parameter b and r.

TABLE V: Parameter settings of the GIG distribution.

Parameters	a	b	r
Value intervals	1	[1, 100]	[-100, 100]

the Bessel function calculation is illustrated and two alternative methods are given to solve this problem. In the second part, we conduct some discussions about the properties of the ANGIG-KF based on the previously proposed alternative methods to verify the validity of the proposed algorithm. Finally, the application scenarios of the proposed algorithm are presented.

A. Two alternative methods for solving the numerical calculation problem

Before giving the alternative methods, we make some explanations about how the numerical calculation problem arises. We first plot the surface of the Bessel function $g(a,b,r) = K_r(\sqrt{ab})$ with respect to its parameters in Fig. 3, where $a = 1$, i.e., $g(b,r) = K_r(\sqrt{b})$ and the logarithm of $g(b,r)$ is selected to make the graph clearer. The detailed parameter settings of the GIG distribution are listed in Table V. Fig. 3 shows that the value of $g(b,r)$ increases extremely fast with parameter r decreasing, which is exactly how this parameter will change as shown in step 8 in **Function 2**. Although parameter b will increase with time, leading to the increase of $g(b,r)$ as Fig. 3 shows, this trend is far less significant than the impact when parameter r decreases, which makes the Bessel functions in the numerator and the denominator quickly go to infinity. Next, we will give two alternative methods to avoid the calculation of Bessel functions.

1) Alternative method 1: approximate mean-based ANGIG-KF: In this method, we employ an approximate mean which is based on the upper and lower bounds of the Bessel function ratio to implement the proposed adaptive KF. It has been proven from [43] that the Bessel function ratio $K_{r+1}(\sqrt{ab})/K_r(\sqrt{ab})$ can be approximated by its upper bound (UB) and lower bound (LB) as follows [43]

$$\underbrace{\frac{r + \sqrt{r^2 + ab}}{\sqrt{ab}}}_{LB} \leq \underbrace{\frac{K_{r+1}(\sqrt{ab})}{K_r(\sqrt{ab})}}_{True\ ratio} \leq \underbrace{\frac{r + 1.5 + \sqrt{(r + 1.5)^2 + ab}}{\sqrt{ab}}}_{UB} \quad (17)$$

Employing (4) and (17), the approximate mean is defined as

$$\mathrm{E}\{\lambda\} \approx \omega \frac{\sqrt{b}}{\sqrt{a}} LB + (1 - \omega) \frac{\sqrt{b}}{\sqrt{a}} UB \quad (18)$$

where $\omega \in [0, 1]$ is a tuning weight regulating the distance from LB and UB to the approximate mean. For ease of

TABLE IV: Detailed implementations of the whole MSCKF based on the proposed ANGIG-KF.

Inputs: Nominal state and measurement noises $\mathbf{Q}_0$ and $\mathbf{R}_0$, iteration number M, forgetting factor ρ and iterative threshold ϵ.

Initialize: $\hat{\mathbf{x}}_0 = \mathbf{0}_{15\times 1}, \mathbf{P}_0, a_0, b_0, r_0, \hat{\theta}_0, \mathbf{P}_0' = \mathbf{P}_0/\hat{\theta}_0, flag = 0$

for $k = 1 : endsteps$

Step 1: Run IMU calculation to obtain the uncorrected navigation information and calculate $\mathbf{\Phi}_k$ and $\mathbf{Q}_k$

Step 2: Run the TU function: $\left(\bar{\mathbf{x}}_k, \check{\mathbf{P}}_k, \check{\mathbf{P}}_k' \right) = \text{TU} \left(\hat{\mathbf{x}}_{k-1}', \hat{\mathbf{P}}_{k-1}', \mathbf{Q}_k, \mathbf{\Phi}_k, \hat{\theta}_{k-1} \right)$

Step 3: **if**, the camera records an image, **do** update $\bar{\mathbf{x}}_k, \check{\mathbf{P}}_k, \check{\mathbf{P}}_k', \mathbf{Q}_k$; **else do** $\bar{\mathbf{x}}_k, \check{\mathbf{P}}_k, \check{\mathbf{P}}_k', \mathbf{Q}_k$ retain unchanged. **end if**

Step 4: **if**, any of the EKF update conditions are satisfied, **do** { (1) obtain the feature information
(2) perform chi-square test and eliminate outliers (3) construct $\mathbf{H}_k, \mathbf{r}_k$ and set $flag = 1$ }; **else do** set $flag = 0$. **end if**

Step 5: Run the MU function: $\left(\hat{\mathbf{x}}_k, \hat{\mathbf{P}}_k, \hat{\mathbf{x}}_k', \hat{\mathbf{P}}_k', \hat{a}_k, \hat{b}_k, \hat{r}_k, \hat{\theta}_k \right) = \text{MU} \left(\bar{\mathbf{x}}_k, \check{\mathbf{P}}_k, \check{\mathbf{P}}_k', \mathbf{H}_k, M, \rho, \epsilon, \mathbf{Q}_k, \hat{\theta}_{k-1}, \hat{a}_{k-1}, \hat{b}_{k-1}, \hat{r}_{k-1}, \tilde{\mathbf{z}}_k, flag \right)$

Step 6: (1) Add the bias correction $\hat{\mathbf{x}}_k$ to the IMU navigation information; (2) Set the bias correction to zero, i.e., $\hat{\mathbf{x}}_k = \mathbf{0}$. **end for**
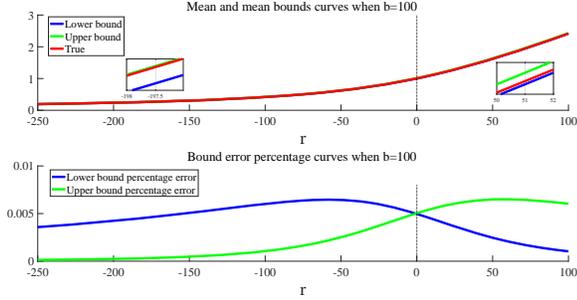

Fig. 4: Mean and mean bounds curves when $b = 100$.

understanding, we have plotted the errors between the true ratio and its upper and lower bounds respectively in a finite range when parameter b is set as 100 in Fig. 4.

Remark 6. *Several caveats to note,*

(1) Results are similar when b is set as other values.

(2) The bounds in [43] are not tight, as proved in [44]. However, the form of the bounds in (19) is the simplest and fits our algorithm.

(3) The bounds in [43] only apply when r is a positive integer. However, they still work when we remove the integer assumption. In addition, the Bessel function ratio can be well bounded by LB and UB when $r > 0$ but turns out to be too large to be bounded when $r < 0$. Hence, we expand the upper bound in [43] to bound the Bessel function ratio and set a parameter ω to adjust the weights of UB and LB.

(4) Fig. 4 highlights that, in the finite range, the approximate errors are bounded and tend to decrease as parameter r decreases, which is exactly how parameter r_k tends to change with time. Thus, the bounds work well in the finite range, although we cannot guarantee the quality of these bounds when parameter r gets smaller. Nonetheless, simulation and experimental results show that the approximate mean works very well, indicating that the bounds in (19) are reasonable (at least in all our simulation and experimental scenarios).

Define the prior LB, prior UB, posterior LB and posterior UB at time k as

$$\begin{cases} \overline{\text{LB}}_k = \frac{\bar{r}_k + \sqrt{\bar{r}_k^2 + \bar{a}_k \bar{b}_k}}{\sqrt{\bar{a}_k \bar{b}_k}}, & \overline{\text{UB}}_k = \frac{\bar{r}_k + 1.5 + \sqrt{(\bar{r}_k + 1.5)^2 + \bar{a}_k \bar{b}_k}}{\sqrt{\bar{a}_k \bar{b}_k}} \\ \widehat{\text{LB}}_k = \frac{\hat{r}_k + \sqrt{\hat{r}_k^2 + \hat{a}_k \hat{b}_k}}{\sqrt{\hat{a}_k \hat{b}_k}}, & \widehat{\text{LB}}_k = \frac{\hat{r}_k + 1.5 + \sqrt{(\hat{r}_k + 1.5)^2 + \hat{a}_k \hat{b}_k}}{\sqrt{\hat{a}_k \hat{b}_k}} \end{cases}$$
(19)

where the time indexes are omitted for brevity. With (18) and

(19), the prior and posterior approximate means of λ_k can be established as follows

$$\begin{cases} \bar{\theta}_k \approx \omega \frac{\sqrt{\bar{b}_k}}{\sqrt{\bar{a}_k}} \overline{\text{LB}}_k + (1 - \omega) \frac{\sqrt{\bar{b}_k}}{\sqrt{\bar{a}_k}} \overline{\text{UB}}_k \\ \hat{\theta}_k^{(i)} \approx \omega \frac{\sqrt{\hat{b}_k^{(i)}}}{\sqrt{\hat{a}_k^{(i)}}} \widehat{\text{LB}}_k^{(i)} + (1 - \omega) \frac{\sqrt{\hat{b}_k^{(i)}}}{\sqrt{\hat{a}_k^{(i)}}} \widehat{\text{UB}}_k^{(i)} \end{cases}$$
(20)

Substituting the prior and posterior means in step 2 and step 9 of **Function 2** with the prior and posterior approximate means in (20), respectively, we can perform the approximate mean-based ANGIG-KF.

2) Alternative method 2: maximum a posterior-based ANGIG-KF: In this method, we utilize the MAP of the GIG distribution for the following two reasons

- The MAP of a distribution represents the point with the highest probability, which is also a good term for approximating the density.
- The MAP of the GIG distribution can be obtained without causing any numerical calculation problems.

The MAP of the GIG distribution is given as follows

$$\text{MAP}\{\lambda_k\} = \frac{r_k - 1 + \sqrt{(r_k - 1)^2 + a_k b_k}}{a_k}$$
(21)

which can be easily obtained without any approximations by taking the derivative of the PDF in (2) and setting the derivative to zero. Likewise, the prior and posterior MAP of λ_k can also be substituted into **Function 2** as follows

$$\begin{cases} \bar{\theta}_k = \text{MAP}\{\lambda_k | \mathbf{z}_{1:k-1}\} := \frac{\bar{r}_k - 1 + \sqrt{(\bar{r}_k - 1)^2 + \bar{a}_k \bar{b}_k}}{\bar{a}_k} \\ \hat{\theta}_k^{(i)} = \text{MAP}^{(i)}\{\lambda_k | \mathbf{z}_{1:k}\} := \frac{\hat{r}_k^{(i)} - 1 + \sqrt{(\hat{r}_k^{(i)} - 1)^2 + \hat{a}_k^{(i)} \hat{b}_k^{(i)}}}{\hat{a}_k^{(i)}} \end{cases}$$
(22)

Substituting the prior and posterior means in step 2 and step 9 of **Function 2** with the prior and posterior MAPs in (22), respectively, we can perform the MAP-based ANGIG-KF.

Remark 7. *We summarize these two methods as follows. Comparing (20) and (22), we find that the MAP is also a lower bound of the mean of λ_k, which establishes the relationship between the approximate mean and the MAP. However, although the mean theoretically outperforms the MAP as discussed in [45], it is hard to tell whether the approximate mean-based ANGIG-KF or MAP-based ANGIG-KF is better. The reason is that although we can guarantee the mean is wrapped in its upper and lower bounds given in (19), whether the true mean is nearer to the upper bound or the lower*

bound is uncertain. Accordingly, the approximate mean no longer possesses the attractive property of minimum variance. Nevertheless, with r_k decreasing, the difference between the MAP and the approximate mean, and the difference between the lower bound and the upper bound will also decrease because the only distinctions between these two alternative methods are different constants modifying from r_k, i.e., $r_k - 1$, r_k and $r_k + 1.5$, which will be less obvious in this setting. To sum up, all these results indicate that the performance of the approximate mean-based ANGIG-KF and the MAP-based ANGIG-KF may be similar, as will be discussed in Section V-C. In this paper, the tuning parameter is suggested to lie in the range $[0.4, \ 0.8]$ and Section V-C will show the proposed ANGIG-KF2 with the tuning weight $\omega \in [0.4, \ 0.8]$ has similar estimation accuracy.

B. Properties analyses of the two alternative method-based ANGIG-KFs

In this part, the convergence analyses and the computational complexity analyses of the two alternative method-based ANGIG-KFs are presented first, then we give the effects of initial parameters selections of the proposed ANGIG-KF.

1) Convergence analyses: We first give the convergence analysis of the MAP-based ANGIG-KF, and the same result applies to the approximate mean-based ANGIG-KF by some simple mathematical derivations. The convergence of the fixed-point iteration process step 5–step 10 in **Function 2** will be proved through confirming the boundedness and monotonicity of the iterative sequence $\{\theta_k^{(i)}\}_{i=0}^{+\infty}$.

Theorem 4. *The iterative sequence $\{\theta_k^{(i)}\}_{i=0}^{+\infty}$ is bounded and monotonic in the MAP-based ANGIG-KF.*

Proof. See Appendix F. $\qquad\square$

Once the iterative sequence is bounded and monotonic, it can at least achieve local convergence. Because when $\theta_k^{(1)}$ at the first iteration is greater than the initial value $\theta_k^{(0)}$, the sequence will tend to its upper bound, and vice versa. Thus, the fixed-point iteration process step 5–step 10 in **Function 2** can be locally convergent in the MAP-based ANGIG-KF.

For the approximate mean-based ANGIG-KF, the proof would be quite similar. Combining (18) and (19), we write down the approximate mean at time k as follows.

$$\mathrm{E}\{\lambda_k\} \approx \omega \frac{r_k + \sqrt{r_k^2 + a_k b_k}}{a_k} \qquad (23)$$
$$+ (1 - \omega) \frac{r_k + 1.5 + \sqrt{(r_k + 1.5)^2 + a_k b_k}}{a_k}$$

Comparing (21) and (23), we can claim that the approximate mean is essentially the convex combination of two translated MAPs because both terms that comprise the approximate mean can be obtained from the MAP by simply translating parameter r_k. Hence, the iterative sequence in the approximate mean-based ANGIG-KF has the same monotonicity as that in the MAP-based ANGIG-KF. Likewise, the former sequence is also bounded because both two terms in the approximate mean are bounded. To this end, we can conclude that the iterative

TABLE VI: Computational complexities of the ANGIG-KF.

ANGIG-KF (Table 3)	
step 3	m^2
step 5-step 10	$M(n^2 + mn^2 + m^2n + m^3 + \mathcal{O}(m^3))$
step 12-step 15	$2mn^2 + m^2n + \mathcal{O}(m^3) + mn + n^3 + n^2$

sequence in the approximate mean-based ANGIG-KF can be locally convergent as well. However, it should be noted that the convergence here does not indicate that the estimated MNCMC can converge to the true MNCMC, but that step 5-step 10 in **Function 2** can locally converge. The convergence accuracy will be verified in the simulation part.

2) Computational complexity analysis: In this part, the computational complexity of the proposed ANGIG-KF is calculated by counting the floating point operations involved with multiplications of matrices and inverse of matrices. Several notes are given first.

- To simplify the analyses, some operations like the chi-square test are not considered.
- The analysis is based on the stable state dimension.
- For brevity, we only take the floating point operations in the measurement update into account.

The computational complexities of some basic steps of the proposed ANGIG-KF are listed in Table VI. According to Table VI, one can conclude that the total computational complexities of the proposed ANGIG-KF in a single measurement update can be calculated as follows.

$$fl_{PR} = m^2 + M[n^2 + mn^2 + m^2n + m^3 + \mathcal{O}(m^3)] + 2mn^2$$
$$+ m^2n + mn + n^3 + n^2 \qquad (24)$$

where fl_{PR} denotes the floating point operations of the proposed ANGIG-KF, respectively.

It can be seen from (24) the computational complexity of the proposed ANGIG-KF increases linearly with the iteration number M. Intuitively, our algorithm has greater computational complexity when iteration number increases. Nevertheless, we can observe from (24) that if we do not iterate, i.e., when $M = 1$, our algorithm will be more computationally efficient. Generally, this case is not reasonable because the iterative process is necessary to make the posterior solutions more accurate. However, later in Section V-A, we will show that the proposed ANGIG-KF still performs well without the iterative process. That is to say, the non-iterative method-based ANGIG-KF can achieve a satisfactory result with a small computational burden.

C. Influences of initial parameters selections

This part firstly provides theoretical convergent value derivations of the proposed method under a stationary scenario to verify the effectiveness of the proposed method, and then conducts a numerical simulation to verify the performance of the proposed algorithm under different initial parameters.

Assumption 1. *If the noise statistics are constant, the following approximation holds*

$$\mathrm{tr}\left\{\tilde{\mathbf{P}}_k^{\mathbf{zz}}(\mathbf{P}_k^{\mathbf{zz}})^{-1}\right\} = \mathrm{tr}\left\{\tilde{\mathbf{P}}_k^{\mathbf{zz}}\left(\mathbf{H}_k\bar{\mathbf{P}}_k\mathbf{H}_k^T + \mathbf{R}\right)^{-1}\right\} \approx m$$
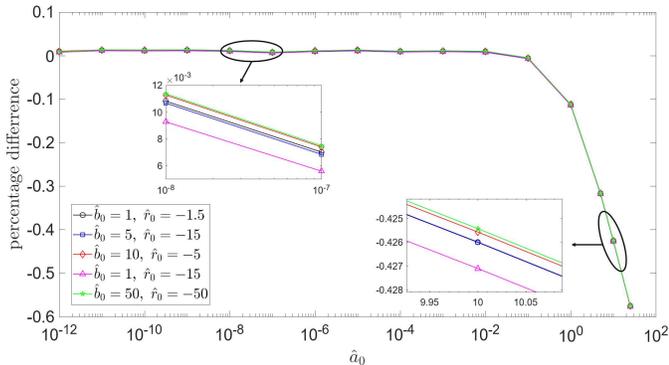$$(25)$$

Fig. 5: Percentage differences between the estimated MNCMC and the true MNCMC.

where $\mathbf{P}_k^{zz}$ denotes the innovation covariance matrix and satisfies $\mathbf{P}_k^{zz} = \lambda \check{\mathbf{P}}_k^{zz}$, and except for $\bar{\mathbf{P}}_k$ denotes the prior state covariance matrix here, other quantities have the same definitions as those in Section III.

Next, we provide analyses of the effects of different initial parameters on the proposed algorithm under a stationary condition. Note that the lower bound in (20) is selected due to similar forms of the lower bound, upper bound, and MAP.

Proposition 3. *Under Assumption 1, the influences of initial parameters $\hat{b}_0$ and $\hat{r}_0$ on the proposed algorithm decay exponentially over time, and the influence of initial parameter $\hat{a}_0$ reduces as $\hat{a}_0$ decreases.*

Proof. See Appendix G. □

Utilizing **Proposition 3**, it can be concluded that due to propagation step (7), the influences of initial parameters $\hat{b}_0$ and $\hat{r}_0$ are trivial over time. Although the parameter $\hat{a}_0$ does not propagate, its influence on the proposed algorithm is also trivial if a small $\hat{a}_0$ is selected. In order to verify this conclusion, a concise numerical simulation will be carried out to verify the aforementioned conclusion.

In this simulation, the parameters $\mathbf{F}_k = 1$, $\mathbf{H}_k = 1$, $\mathbf{Q}_k = 1$, $\rho = 1 - e^{-4}$, and the number of Monte Carlo runs is 10000. Besides, we select parameter $\hat{a}_0$ in the interval $[10^{-12}, 25]$, and set λ from 0.1 to 25, and randomly select some initial parameters for $\hat{b}_0$ and $\hat{r}_0$. Note that it is not recommended to select a smaller $\hat{a}_0$ because numerical calculation problems will be induced. Fig. 5 plots the percentage differences between the estimated MNCMC and the true MNCMC using the proposed algorithm. We average the percentage differences over different λ.

Fig. 5 shows that the influences of different $\hat{b}_0$ and $\hat{r}_0$ are small on the convergent estimated MNCMC. In addition, small $\hat{a}_0$ ranging from 10^{-12} to 10^{-1} contributes to better estimates of true MNCMC, which corresponds to the aforementioned conclusion. Hence, it is recommended that $\hat{a}_0$ is selected in the interval of $[10^{-12}, 10^{-1}]$.

D. Application scenarios of the proposed algorithm

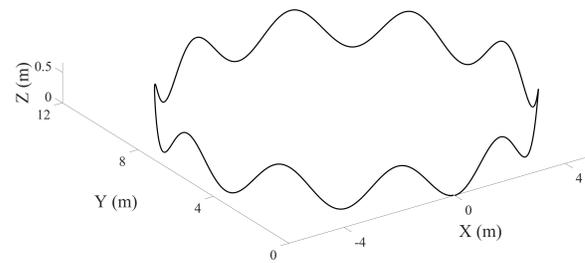The proposed algorithm is based on an assumption that the MNCM is a time-varying scalar multiplied by an identity matrix in this paper. Therefore, stringently speaking, the measurement noise of the sensor needs to satisfy the same degree of variation in each dimension. This condition can be approximately satisfied on sensors that measure in all directions based on the same principle, such as the camera, Global Position System, Doppler velocity log, ultra wide band and odometry. However, the proposed algorithm may perform poorly when each dimension of the measurement is based on different principles, such as the liDAR and the ultra-short baseline, whose range measurement is based on the time of flight but the bearing measurement is generally based on the phase difference. In this case, using a scalar to adjust the MNCM is not appropriate. Nevertheless, the VIO navigation system with time-varying and high dimensional measurements is still one of the best application scenarios for verifying the effectiveness and superiority of the proposed algorithm.

Fig. 6: Simulated groundtruth.

V. SIMULATION TESTS

In this section, we carry out some simulation tests to verify the validity and superiority of the proposed ANGIG-KF using the simulated data. A robot is assumed to move in a circle in the $x - y$ plane while undertaking sinusoidal movement along the z axis, which is shown Fig. 6, and the feature points are randomly generated on a cylinder surrounding the trajectory. Stationary feature points are randomly generated on the cylinder surrounding the trajectory, and every recorded image contains 3 feature points. The update frequency of the IMU and the camera are 100Hz and 10Hz, respectively. Note that we do not simulate absolute localization information such as the GNSS signal. Therefore, the localization accuracy of all algorithms will deteriorate over time.

100 Monte Carlo runs are executed and the simulation time is set as 250s (equivalent of 25000 INS updates) for all scenarios. The root mean square error (RMSE) and normalized estimation error square (NEES) are chosen as the performance metrics of the estimation accuracy and consistency for all compared algorithms [30]. In addition, for clearer comparison, the RMSEs and NEESs from all algorithms are smoothed using a moving average method with span of 100 steps.

According to the engineering experience, the nominal M-NCMC λ_0 in the normalized coordinate system is set as $\lambda_0 = 1$ which corresponds to a pixel error in the pixel coordinate system[2]. The actual λ_k jumps between $4\lambda_0$ and $2\lambda_0$ over time. This simulation will be segregated into three parts.

---

[2]We highlight that the MNCMC actually does not have any units. We add the units of the MNCMC just to make the meaning of the MNCMC more obvious, i.e., $\lambda_0 = 1$ (pix^2), where pix is the abbreviation of pixel.

TABLE VIII: PRMSE and simulation time comparisons under different iteration numbers.

Iteration numbers	ANGIG-KF1		ANGIG-KF2	
	PRMSE1	Time1 (ms)	PRMSE2	Time2 (ms)
1	0.130	7.053	0.130	7.049
2	0.130	7.699	0.129	7.836
3	0.130	8.816	0.129	8.812
4	0.130	9.750	0.129	9.969
5	0.130	10.558	0.129	11.099
6	0.130	11.533	0.129	11.695
7	0.130	12.197	0.129	12.679

In the first part, we verify that both the MAP-based ANGIG-KF (ANGIG-KF1) and the approximate mean-based ANGIG-KF (ANGIG-KF2) under pre-set scenario can converge as the number of iterations increases, which provides theoretical basis for the choice of iteration numbers. In the second part, the estimation accuracy of all algorithms is compared in this case to verify the superiority of the proposed ANGIG-KF. Finally, we have some discussions about the relationship between the MAP-based ANGIG-KF and the approximate mean-based ANGIG-KF.

A. Convergence verification of the proposed ANGIG-KF

In this part, the convergence verifications of the proposed MAP-based ANGIG-KF (abbreviated as ANGIG-KF1) and approximate mean-based ANGIG-KF (abbreviated as ANGIG-KF2) are conducted. Parameter settings are listed in Table VII except for the iteration number N setting from 1 to 8.

TABLE VII: Parameter settings for simulation.

Algorithms	Parameter settings
VB-MSCKF [13]	$a = 0.4, \nu_0 = 10, \mathbf{V}_0 = 0.1 * \mathbf{I}_6$
RSTKF [30]	dof $= 5, M = 50, \epsilon = 1e - 8$
HKF [29]	tuning parameter $\gamma = 1.345, M = 50, \epsilon = 1e - 8$
ANGIG-KF1	$\hat{a}_0 = 10, \hat{b}_0 = 10, \hat{r}_0 = 1, \rho = 0.99$
ANGIG-KF2	$\hat{a}_0 = 1, \hat{b}_0 = 0, \hat{r}_0 = -0.5, \rho = 0.99, \omega = 0.5$

We list the position RMSEs (PRMSEs) and execution time[3] under different iteration numbers in Table VIII to determine which iteration number is appropriate. It can be seen from Table VIII that the accuracy improvement stopped when the iteration number is larger than 2. Furthermore, we can observe from the execution time column that a mere 0.8% accuracy improvement increases the average execution time by 10%, which indicates that the benefits attained from iterations are not obvious and performing too many iterations is not necessary. We attribute this result to retaining the correlation between the state vector and the MNCMC, which renders the initial estimated $\hat{\theta}_k$ close to the optimal solution and the iterative sequence converges very fast. Thus, the proposed algorithm will not execute iterations in the remaining simulation tests and experimental tests. That is to say, we adopt the non-iterative method which is mentioned in Section III-B.

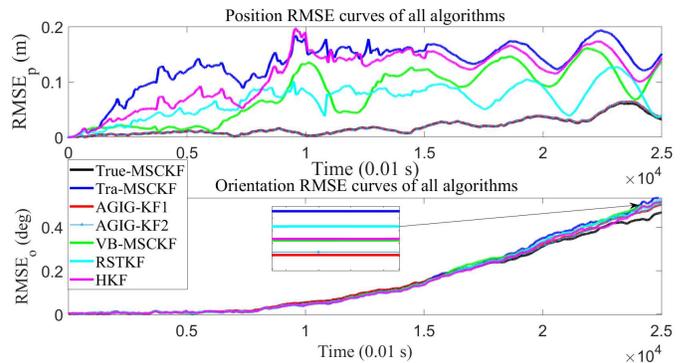[3]The execution time here denotes the average time running a single **Function 2**.

Fig. 7: PRMSE and ORMSE curves of all algorithms.

B. Superiority verification of the proposed ANGIG-KF

In this part, the RMSEs of the existing Tra-MSCKF with nominal MNCM [5], True-MSCKF with true MNCM [5], RSTKF [30], HKF [29], VB-MSCKF [13] and the proposed ANGIG-KF1 and ANGIG-KF2 are compared under the pre-selected case. Parameter settings for all compared algorithms are shown in Table VII. Note that the proposed ANGIG-KF is divided into the ANGIG-KF1 and ANGIG-KF2. However, we do not distinguish these two algorithms but will discuss them in Section V-C. In the following comparison, the ANGIG-KF1 and ANGIG-KF2 are collectively called the ANGIG-KF.

In this case, we simulate a scenario where the MNCM changes suddenly in steps at fixed intervals. Fig. 7 depict the PRMSE and orientation RMSE (ORMSE) curves of the proposed ANGIG-KF and other compared algorithms. It can be seen from (41) that the proposed ANGIG-KF propagates a_{k-1}, b_{k-1} and r_{k-1} with forgetting factor ρ, which makes the current estimated MNCM related to the historical MNCM estimates. Nevertheless, the RSTKF and HKF only use the current measurement to estimate the MNCM, which will be less accurate in this MNCM setting. The estimated MNCMC curve is plotted in Fig. 8. We can observe from Fig. 8 that even if the MNCM changes suddenly, the proposed ANGIG-KF still retains excellent performance, whose position and orientation estimation accuracies are 85.48% and 3.1% better than the Tra-MSCKF. Note that the transient performance of the estimated MNCMC is not attractive. A sudden MNCMC change at the current time is difficult for the proposed ANGIG-KF to capture rapidly since the MNCMC estimate is based on historical information. Nevertheless, we can see from Fig. 8 that the steady performance of the ANGIG-KF is fairly good. To be more convincing, another scenario in which the MNCMC experiences four sudden changes is considered. Note that only the estimated and true MNCMCs in this supplementary scenario are shown in Fig. 9. The results are similar to Fig. 8.

Furthermore, we should explain that the estimated trajectory of the MNCMC slightly lags behind the true trajectory, which is because the estimated MNCMC is obtained through past measurement data learning. As stated in [40], adaptive KF-based learning is a reactive scheme, where the change of the estimated noise parameters will always be delayed. Nevertheless, the overall tracking performance is satisfying.
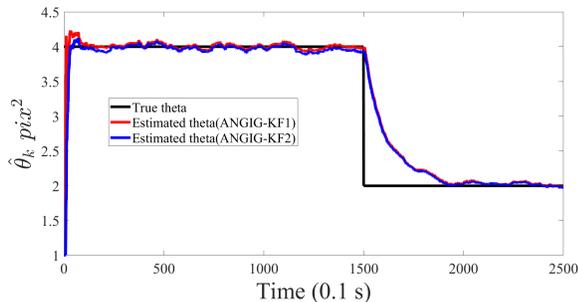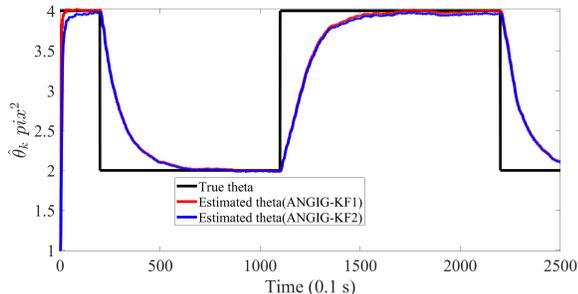
Fig. 8: Estimated MNCMC curves.

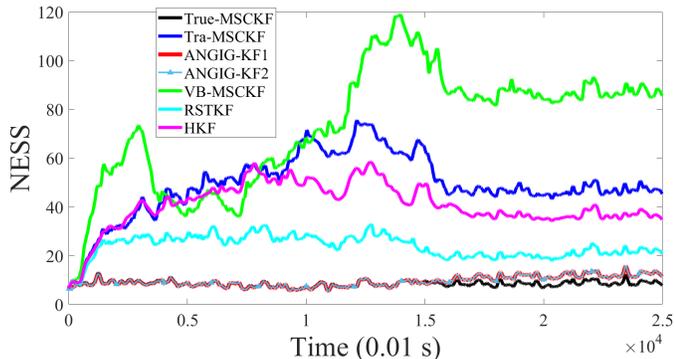Fig. 9: Estimated MNCMC curves in the supplementary scenario.

Fig. 10: NEES curves of all algorithms.

In addition to the estimation accuracy, the consistency[4] is also a significant metric to evaluate the algorithm. Therefore, we plot the NEES curves of position, velocity and altitude estimation of all algorithms in Fig. 10. It can be seen from Fig. 10 that the Tra-MSCKF has the worst consistency, which is due to an adventurous MNCM setting and the inability to adjust the MNCM. Moreover, both RSTKF and HKF are not consistent. The reason is that the existing RSTKF and HKF are essentially robust algorithms and cannot accurately estimate the time-varying MNCM. Without an accurate MNCM estimate, the consistencies of the existing RSTKF and HKF are not guaranteed. Furthermore, notice that the NEES of the VB-MSCKF is not particularly conservative as we expected. The reason is that the VB-MSCKF gives up the fixed-point iteration

[4]There are two important factors that can induce inconsistency of a VIO system, i.e., parameter inaccuracy and erroneous observability. The latter one can be solved by the FEJ technique [35] or the OC-EKF [36]. Note that this simulation focuses on the inconsistency caused by parameter inaccuracy.

TABLE IX: ARMSEs of the VB-MSCKF under different initial parameters.

Algorithms		Orientation ARMSEs (deg)	Position ARMSEs (m)
Tra-MSCKF		0.1584	0.1317
True-MSCKF		0.1484	0.0185
ANGIG-KF1		0.1513	0.0195
VB-MSCKF	$\hat{\nu}_0=10$, $\hat{\mathbf{V}}_0=0.01*\mathbf{I}_4$	0.15614	0.078781
VB-MSCKF	$\hat{\nu}_0=10$, $\hat{\mathbf{V}}_0=0.01*\mathbf{I}_6$	0.15614	0.078781
VB-MSCKF	$\hat{\nu}_0=10$, $\hat{\mathbf{V}}_0=0.05*\mathbf{I}_6$	0.15632	0.078780
VB-MSCKF	$\hat{\nu}_0=10$, $\hat{\mathbf{V}}_0=0.1*\mathbf{I}_6$	0.15599	0.078698
VB-MSCKF	$\hat{\nu}_0=10$, $\hat{\mathbf{V}}_0=0.5*\mathbf{I}_6$	0.15586	0.078242
VB-MSCKF	$\hat{\nu}_0=10$, $\hat{\mathbf{V}}_0=1*\mathbf{I}_6$	0.15626	0.078598
VB-MSCKF	$\hat{\nu}_0=5$, $\hat{\mathbf{V}}_0=0.5*\mathbf{I}_6$	0.15625	0.07776
VB-MSCKF	$\hat{\nu}_0=15$, $\hat{\mathbf{V}}_0=0.5*\mathbf{I}_6$	0.15634	0.078428
VB-MSCKF	$\hat{\nu}_0=20$, $\hat{\mathbf{V}}_0=0.5*\mathbf{I}_6$	0.15632	0.078498

to pursue computational efficiency, which makes it difficult for the VB learning process to estimate accurate MNCM parameters. In contrast, due to retaining the correlation of the state vector and parameters, the proposed ANGIG-KF not only ensures consistency (closer to 9) but also takes consideration of the estimation accuracy, which contributes to its best performance overall. Moreover, the results of the proposed ANGIG-KF also indicate that an accurate MNCM not only improves the estimation accuracy, but also improves the estimation consistency.

C. Discussions

In this part, we carry out some discussions for the remaining issues about the performance of the existing VB-MSCKF and the proposed algorithm, and the weight selection problem.

We have verified that the initial parameters have trivial influences on the proposed algorithm in Section IV-C. In this part, we show that this conclusion also applies to the VB-MSCKF. Firstly, the influences of the initial parameters on the existing adaptive VB-MSCKF algorithm are similar to those on the proposed ANGIG-KF because a similar propagation operation is selected to propagate the parameter model, i.e., propagate parameters with a forgetting factor ρ. As a result, the influences of initial parameters on the VB-MSCKF is also exponentially decaying. To substantiate this point, we have carried out a simulation, in which different initial parameters are selected for the VB-MSCKF. The results are shown in Table IX. It can be seen from Table IX that the accuracy improvement of selecting different initial parameters $\hat{u}_0$ and $\hat{U}_0$ on the existing VB-MSCKF is trivial, which is in concert with the aforementioned corollary.

In addition, we show the RMSE curves of the Tra-MSCKF, ANGIG-KF1, and ANGIG-KF2 with different tuning weight $\omega = 0.4, 0.5, 0.6, 0.7, 0.8$ in Fig. 11. It can be seen from Fig. 11 that the proposed ANGIG-KF2 with different tuning weight $\omega = 0.4, 0.5, 0.6, 0.7, 0.8$ has almost the same estimation accuracy. Nonetheless, it still outperforms the Tra-MSCKF.

Besides, we have discussed earlier in **Remark 7** that the performance of the MAP and approximate mean may be
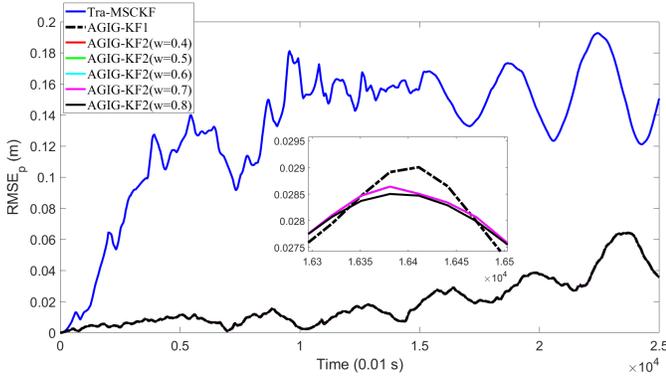
Fig. 11: PRMSEs of the proposed ANGIG-KF under different tuning weight $\omega = 0.4, 0.5, 0.6, 0.7, 0.8$.

similar due to their similar form, and the same results can be obtained through all foregoing RMSE curves. As a matter of fact, if the true mean approaches its upper bound, the approximate mean with the weight $\omega = 0.5$ may perform better than the MAP, while it may be inferior to the MAP if the true mean is nearer to its lower bound. Nevertheless, the tuning weight parameter ω is set as a constant in all simulation tests because we do not have any prior information about the true mean. The constant weight in this simulation is obtained by experience.

VI. EXPERIMENTAL TESTS

In order to be more convincing, the effectiveness and superiority of the proposed ANGIG-KF will be verified in various environments through real-world datasets. In this experiment, we carry out several tests with the aircraft EuRoC datasets [37] provided by ETH Zurich, and the in-vehicle KITTI datasets [38] co-founded by Toyota American Institute of Technology, under various stationary and complex environments. Here are some introductions to these datasets. The EuRoC datasets use a 200Hz IMU and a 20Hz monochrome binocular camera and the groundtruth of which is obtained by the Vicon motion capture system. The EuRoC datasets include two scenarios: a machine hall and an ordinary room. The KITTI datasets contain abundant outdoor scenarios and use multiple sensors in which a 100Hz IMU, a 10Hz stereo camera and a GPS are utilized in our experimental tests. The groundtruth can be provided by a high-accuracy INS/GNSS integrated navigation system. The experimental tests are divided into three parts. In Part I, the experimental environment and parameter settings are presented. Experimental tests are carried out in Part II and Part III, where Part II performs the experimental tests on the indoor stationary EuRoC datasets and Part III focuses on the outdoor complex KITTI datasets. Details will be given in respective parts.

A. Experimental environment and parameter settings

Before the tests, we want to note that the experiments require a neat and generic VIO framework which should at least satisfy the following principles: 1) The framework only runs the MSCKF algorithm without using any other complex

optimizations to improve the localization accuracy, such as the online alignment. 2) All compared algorithms can be well embedded into the framework.

Therefore, we use the OpenVINS [39] as the framework of all compared algorithms because different algorithms can be implemented on the VIO navigation system by simply modifying the codes of the state prediction part and measurement update part of OpenVINS. Additionally, in our experiment, the number of the sliding windows N is set as 11 and the maximum number of feature points participating in the measurement update at each time step is set as 40. We conduct comparisons among the previous mentioned algorithms in the simulation, where we only employ the proposed ANGIG-KF2 to implement the comparison due to the similar performance of the ANGIG-KF1 and ANGIG-KF2. To compare the performance of all algorithms in experimental tests, we define following position error (PE) and orientation error (OE) as the performance metrics

$$\begin{cases} \text{PE}_k(m) & = \sqrt{(x_k - \hat{x}_k)^2 + (y_k - \hat{y}_k)^2 + (z_k - \hat{z}_k)^2} \\ \text{OE}_k(deg) & = \left\| \left({}_G^I\mathbf{q} \bigotimes {}_G^I\hat{\mathbf{q}}^{-1} \right)_{1:3} \right\| * \frac{180}{\pi} \end{cases}$$
(26)

where x_k, y_k, z_k and ${}_G^I\mathbf{q}$ denote the reference values of the global three dimensional positions and unit quaternion, respectively, and $\hat{x}_k, \hat{y}_k, \hat{z}_k$ and ${}_G^I\hat{\mathbf{q}}$ are their corresponding estimates after bias correction, respectively. Besides the algorithms in the simulation tests, the MSCKF with inflated MNCM is also added, which is thereafter called the Inf-MSCKF. The nominal MNCMCs are selected as 2.25 and 4 in the EuRoC datasets and KITTI datasets, respectively.

B. EuRoC datasets

In this part, the EuRoC datasets will be tested to verify the performance of the proposed algorithm in the case of an indoor and stable environment. The nominal value for λ_0 under the EuRoC datasets is set as $\lambda_0 = 1 \ (pix^2)$ which corresponds to a pixel error in the pixel coordinate system. We should note that although the environment is stable, the Tra-MSCKF may still suffer from estimation degradation due to the wrong cognition of the MNCM. Datasets V101-V203 and MH01-MH05 are selected to conduct comparisons of all algorithms, where dataset V203 is analyzed in detail as representative. Parameter settings for all compared algorithms in the EuRoC datasets are listed in Table X, where the weight ω is selected by engineering experience and the selection principle of $\hat{a}_0, \hat{b}_0, \hat{r}_0$ is to set $\hat{\theta}_0 = \lambda_0$. The average PEs (APEs) and average OEs (AOEs) of all tested EuRoC datasets are listed in Table XI.

TABLE X: Parameter settings of algorithms for experiments.

Algorithms	Parameter settings
ANGIG-KF2	$a_0 = 10^{-8}, b_0 = 3.5, r_0 = -0.25, \omega = 0.7$
VB-MSCKF [13]	$\nu_0 = 10, a = 0.4, \mathbf{V}_0 = 0.1 * \mathbf{I}_6$
RSTKF [30]	$dof = 5, N = 50, \epsilon = 1e-6$
HKF [29]	tuning parameter $\gamma = 1.345, M = 50, \epsilon = 1e-6$

1) Dataset V203: In the dataset V203, the camera moves rapidly indoors and suffers from sudden shakes sometimes, which leads to blurring of some images. The dataset V203 runs
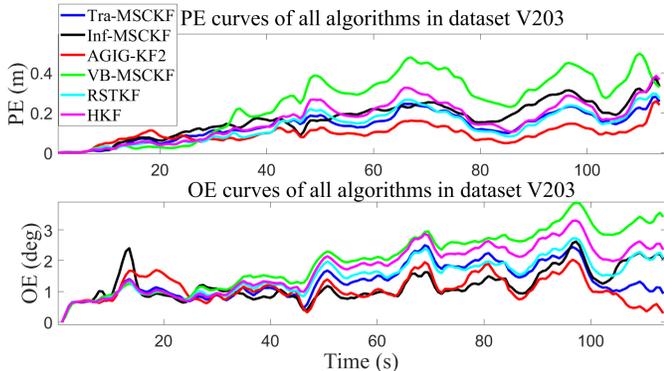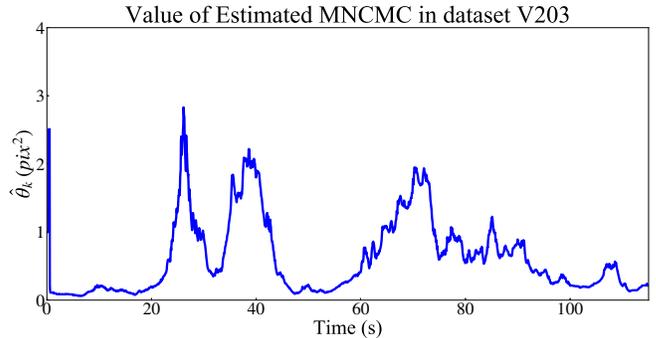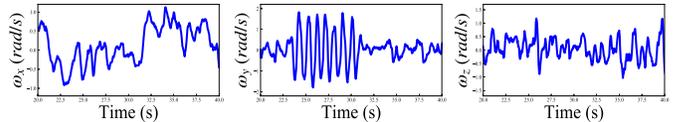
Fig. 12: PEs and OEs for dataset V203.

(a) Estimated MNCMC for dataset V203.

(b) Gyroscope outputs.

Fig. 13: Estimated MNCMC and gyroscope outputs in dataset V203.

for 114.845s and 86.128m with an average speed of 2.7km/h. The PEs and OEs of all algorithms for dataset V203 are shown in Fig. 12. It can be observed from Fig. 12 that due to the lack of absolute localization information, the whole PEs and OEs are increasing moderately overall, which is unavoidable in VIO systems. Note that the existing VB-MSCKF performs poorly compared with other algorithms, even with the Tra-MSCKF, which is inconsistent with the performance in the simulation. We make the following explanations for this result. Firstly, as illustrated in Section I-C, the inaccurate dimension matching method based on the ℓ_2 norm limits the estimation accuracy of the VB-MSCKF. Secondly, VB-based methods require certain iterations to compensate for the inaccuracy of the nominal MNCM. The VB-MSCKF, however, does not perform iterations to save computation [13], which further limits its accuracy. In the simulation, the true state noise covariance matrix is employed so that the overall estimation accuracy of all algorithms is relatively high. However, authentic scenarios in the experiment are more harsh, which are usually reflected in unknown interference and inaccurate parameters such as $\mathbf{Q}_k$. These further amplify the drawback of the VB-MSCKF. In contrast, the proposed ANGIG-KF2 retains a satisfactory estimation performance, which benefits from a more accurate modeling of MNCMC and a more efficient joint inference than VB inference, as discussed in **Remark 4**. It can be seen from Fig. 12 that both the position and orientation estimation accuracy of the proposed ANGIG-KF2 is superior than the Tra-MSCKF and other compared algorithms. To better explain this result, we plot the estimated MNCMC of this dataset in Fig. 13 to understand the underlying reason for the advantage of the proposed algorithm.

The most volatile period at around 25s is singled out for analysis. The output curves of the gyroscope during this period are plotted in Fig. 13(b). It is obvious that the orientation of the camera experienced large interferences between 25s-30s, which resulted in the severe camera shake. As a consequence, the image the camera recorded during that time was blurred, leading to the increase of the inaccuracy of the camera. The proposed ANGIG-KF2 accurately captures the vehicle maneuvering during this time and adaptively increases the estimated MNCMC to cope with the camera accuracy degradation, which gives rise to its good estimation performance.

2) Other datasets: For other datasets in the EuRoC datasets, we directly list the APEs, AOEs and execution times during single implementation of all algorithms in Table XI. It can be seen from Table XI that our algorithm outperforms others in most cases. The Inf-MSCKF is inferior to the Tra-MSCKF in most cases, which indicates that the experience MNCM setting is more reasonable in EuRoC datasets. On average, the orientation and position estimation accuracies of proposed ANGIG-KF2 are improved by 22.20% and 29.41% with respect to the Tra-MSCKF, respectively, and 24.33% and 37.5% with respect to the Inf-MSCKF, and 11.60% and 36.51% with respect to the VB-MSCKF, and 10.66% and 22.58% with respect to the RSTKF, and 19.69% and 26.38% with respect to the HKF. Thus, we can basically deduce that the essential solution of the accuracy degradation of the Tra-MSCKF is the ability to adaptively adjust the MNCM rather than inflate the MNCM. Moreover, note that the RSTKF has a rather good performance in datasets V102-V202, MH03 and MH05. Our explanation is that the experimental environments are worse than the simulation, which highlights the advantage of the robust algorithms. The robust HKF, however, does not perform well due to its inherent shortcoming of not considering the measurement noise statistics. Despite that the RSTKF performs better than the proposed ANGIG-KF2 occasionally, it consumes much larger computation to achieve this advantage.

Next, to show the consistency of all algorithms, we draw the 3-sigma bound[5] curves of all algorithms for dataset V101 in Fig. 14, where the first three rows are position 3-sigma bounds and the last three rows are orientation 3-sigma bounds. Before we analyze the consistency of all algorithms, some caveats are given as follows.

---

[5]The principle of evaluation consistency of the 3-sigma bound is to judge whether the true error is within the theoretical allowable error range (i.e., plus or minus three times of estimated standard deviation for Gaussian distributed state vector). If the bound curves do not wrap at least 99.74% of the error curve, we consider this estimate to be inconsistent.

TABLE XI: APEs and AOEs of all tested EuRoC datasets.

NO. Dataset	Performance metrics	Tra-MSCKF [5]	Inf-MSCKF [5]	VB-MSCKF [13]	ANGIG-KF2	RSTKF [30]	HKF [29]
V101	AOE (deg)	0.851	**0.607**	1.169	0.749	0.829	0.802
	APE (m)	0.071	0.636	0.085	**0.050**	0.074	0.086
	Time (ms)	12.252	12.608	12.257	12.737	24.120	39.481
V102	AOE (deg)	1.284	1.481	0.517	0.693	**0.483**	1.185
	APE (m)	0.135	0.159	0.126	**0.105**	0.111	0.139
	Time (ms)	12.950	13.612	13.330	13.987	25.922	38.752
V103	AOE (deg)	0.813	0.968	1.004	**0.793**	1.159	1.044
	APE (m)	0.092	0.095	0.119	0.102	**0.078**	0.008
	Time (ms)	13.431	13.881	13.749	14.543	24.023	39.860
V201	AOE (deg)	1.481	1.715	**0.894**	1.198	1.323	1.461
	APE (m)	0.094	0.110	0.150	**0.084**	0.120	0.098
	Time (ms)	12.081	12.168	12.359	12.302	23.138	33.644
V202	AOE (deg)	0.903	0.908	1.171	0.852	**0.849**	0.880
	APE (m)	0.116	0.122	0.152	**0.111**	0.135	0.117
	Time (ms)	12.631	12.852	12.809	13.736	23.057	36.683
V203	AOE (deg)	1.293	1.228	2.091	**1.056**	1.510	1.778
	APE (m)	0.139	0.172	0.250	**0.099**	0.138	0.168
	Time (ms)	13.570	14.418	13.824	14.647	21.172	37.034
MH01	AOE (deg)	1.440	1.281	1.114	**1.077**	1.439	1.375
	APE (m)	0.179	0.204	0.230	**0.091**	0.160	0.172
	Time (ms)	11.357	12.098	11.736	11.950	19.289	14.423
MH02	AOE (deg)	0.879	1.259	0.776	**0.596**	0.679	0.852
	APE (m)	0.131	0.161	0.144	**0.096**	0.119	0.127
	Time (ms)	11.412	11.462	12.296	12.034	19.358	14.277
MH03	AOE (deg))	2.364	2.112	**1.388**	1.781	1.795	1.565
	APE (m)	0.216	0.227	0.164	0.173	**0.158**	0.163
	Time (ms)	11.666	12.350	11.914	12.323	20.650	16.407
MH04	AOE (deg)	0.847	1.125	0.595	**0.369**	0.594	0.829
	APE (m))	0.358	0.416	0.330	**0.197**	0.305	0.368
	Time (ms)	11.118	11.220	11.295	11.795	18.759	14.001
MH05	AOE (deg)	1.017	0.829	0.853	1.078	**0.785**	0.971
	APE (m)	0.335	0.379	0.324	**0.212**	0.307	0.351
	Time (ms)	11.494	11.829	11.529	11.601	19.621	16.952
Mean	Mean_o (deg)	1.198	1.229	1.052	**0.930**	1.041	1.158
	Mean_p (m)	0.170	0.192	0.189	**0.120**	0.155	0.163

- Since the true state vector in this paper is heavy-tailed distributed, its error bounds should theoretically be a little larger than that in the Gaussian distribution to cover at least 99.74% of the error curve. Nevertheless, we still draw the 3-sigma curves for convenience.
- Since the state vector defined in this paper is the inertial navigation error, the obtained filter standard deviation is essentially the standard deviation of the inertial navigation error state rather than the navigation estimate[6]. Therefore, stringently speaking, the bound curves in Fig. 14 do not actually reflect the consistency of navigation estimate. Fortunately, the inertial navigation error is corrected to zero at every posterior moment, which renders the two standard deviations mathematically equivalent.

It can be seen from Fig. 14 that the consistency of the Inf-MSCKF is slightly better than the Tra-MSCKF due to an inflated MNCM setting. Nevertheless, all except for the proposed ANGIG-KF2 have severe inconsistent position estimates while the orientation estimates of all algorithms are mostly consistent, which verifies the good quality of the estimates

[6]The navigation estimate denotes the corrected INS updating estimate using the filter posterior state estimate

of the proposed algorithm. The reason for this superiority stems from an outstanding MNCMC estimate, which benefits the right cognition of the true MNCM of the system and concurrently improves the consistency of the estimate.

C. KITTI datasets

In this part, we carry out tests in outdoor environments which are divided into two scenarios, i.e., urban and highway. The nominal MNCMC is set as 1.5 pixel standard deviation, and $\hat{a}_0 = 10^{-8}, \hat{b}_0 = 5.75, \hat{r}_0 = -0.5$. Other parameter settings retain the same as Table X. Detailed APEs and AOEs of all algorithms in the two scenarios are listed in Table XII.

1) Urban scenario: For the urban scenario, we use dataset 2011_10_03_*drive*_0027 which is collected in the urban area (we will thereafter call this dataset Urban). In this scenario, the vehicle is frequently diverted on a road, which may lead to the feature points tracking missing due to the big perspective changing. Additionally, although the road is highly textured, abundant shadows exist along the road, for which the measurement accuracy of the camera is extremely susceptible. Furthermore, it is worth noting that the data between 37.5s and 195s are selected in this test due to the IMU measurements
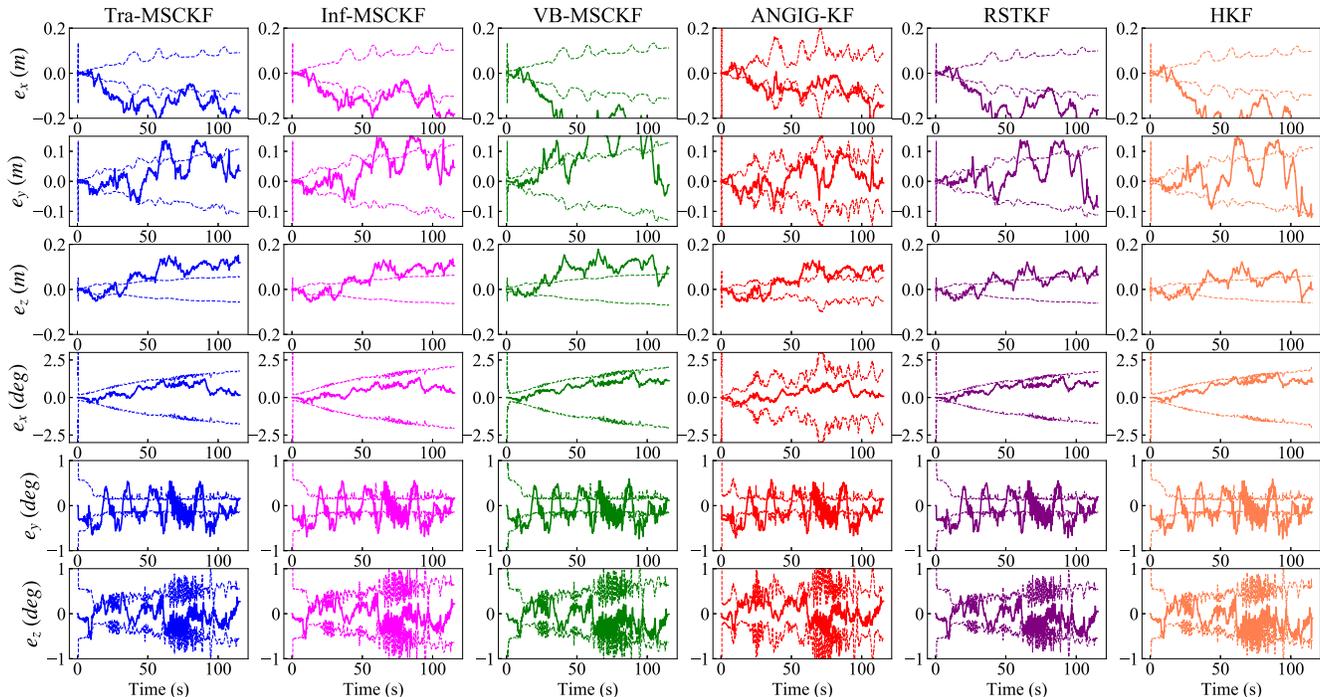
Fig. 14: 3-sigma bound curves of all algorithms for dataset V203.

missing at around 35s and 196s according to the timestamp of the dataset. In this setting, the trajectory runs for 157.5s and 1133.5m at an average speed of 26km/h. Fig. 15 gives the PEs and OEs of all algorithms. We can see from Fig. 15 that in the urban area, all algorithms suffer from certain estimation accuracy degradation compared with the indoor results tested in the EuRoC datasets, especially the existing VB-MSCKF, which means the outdoor localization is more challenging than indoor localization. Note that the performance of the VB-MSCKF degrades more in the dataset Urban, which is due to the much worse environment in this dataset (frequent diversion, shadows, shakes). Note that the estimation accuracy of the Inf-MSCKF is improved compared with the Tra-MSCKF, which means an inflated MNCM is more reasonable in this scenario. Nonetheless, both the orientation and position estimation performance of the proposed ANGIG-KF2 are superior to other algorithms, which is 19.40% and 19.87% better than the Tra-MSCKF, and 4.57% and 3.69% than the Inf-MSCKF, and 79.16% and 80.70% better than the VB-MSCKF, and 37.79% and 38.06% better than the RSTKF, and 40.09% and 40.12% better than the HKF. Moreover, the execution time of the proposed ANGIG-KF2 is not significantly increased compared with the VB-MSCKF, but its estimation accuracy is much better than the VB-MSCKF.

To better analyze the impact on the estimated MNCMC of the outdoor complex environment, the true trajectory and the estimated MNCMC curve are drawn in Fig. 16. We list several typical spots where the MNCMC is most likely to present time-varying characteristics, including the diversion, shadows and moving objects (unexpected pedestrian appearance). In Fig. 16(b), frequent fluctuations of the estimated MNCMC from 90s to 140s may stem from continuous diversions during the movement, as circled by the red ellipse in Fig. 16(a). Due
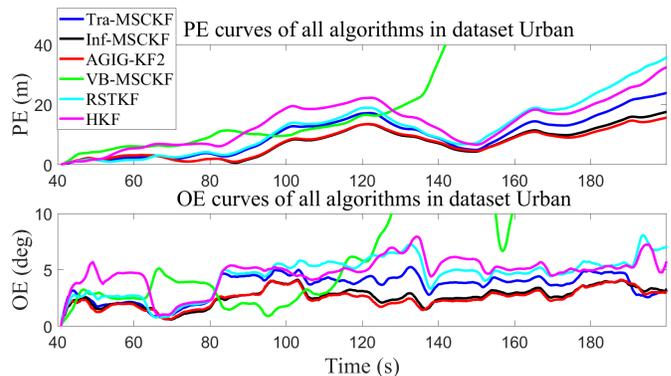

Fig. 15: PEs and OEs for dataset Urban.

to four turns and occasional shakes, the accuracy of the camera is very unstable. Moreover, the sharp rise at about 180s is due to a series of shadows appearing at the end of the trajectory. Nonetheless, even if the measurement accuracy of the camera is very unstable, the proposed ANGIG-KF2 can reasonably adjust the MNCM estimate according to the obtained camera measurement to achieve an approximate tracking of the true MNCM and obtain good position and orientation estimation accuracy.

2) Highway dataset: Finally, we use the dataset 2011_10_03_*drive*_0042 to test the performance of all algorithms under high-speed scenarios, where the vehicle moves on a straight highway with a speed of about 78km/h (denoted as dataset Highway). It is clear from Table XII that the performances of all algorithms are much worse than those under the urban case, and the VB-MSCKF even diverges in this dataset. We should first explain the reason that the execution time of the VB-MSCKF in this dataset is less than

TABLE XII: APEs and AOEs of all tested KITTI datasets.

NO. Sequence	Performance metrics	Tra-MSCKF [5]	Inf-MSCKF [5]	VB-MSCKF [13]	ANGIG-KF2	RSTKF [30]	HKF [29]
	AOE (deg)	3.717	2.759	14.379	**2.633**	4.816	5.001
Urban	APE (m)	11.736	8.748	65.435	**8.425**	15.182	15.706
	Time (ms)	15.911	16.740	16.829	17.140	34.953	49.879
	AOE (deg)	5.406	5.427	24.034	**5.361**	5.520	5.512
Highway	APE (m)	23.494	30.274	1098.726	**21.899**	39.050	26.089
	Time (ms)	11.588	11.803	11.315	12.149	16.168	16.443
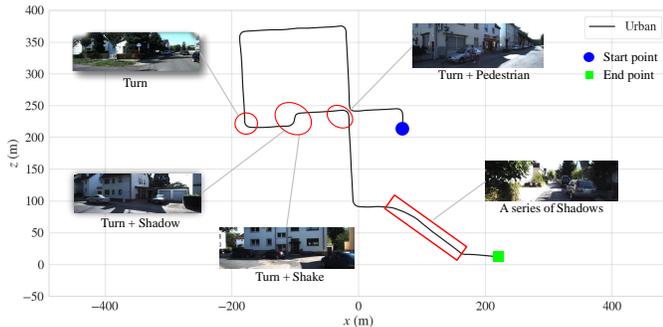
(a) Trajectory of the datasets Urban.
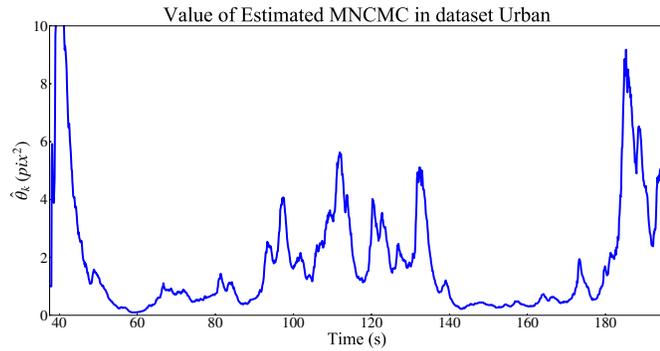
(b) Estimated MNCMC in the urban area.

Fig. 16: True trajectory and estimated MNCMC.

the Tra-MSCKF. Due to the divergence of the VB-MSCKF, its estimated trajectory has been completely separated from the true trajectory. Thus, almost all feature points are identified as outliers and eliminated by the chi-square test, which saves much execution time for the measurement update. Hence, the VB-MSCKF consumes less execution time than the Tra-MSCKF, which also indicates that too little execution time may not be good, as it means less information (both useful information and outliers) can be utilized and processed. In this case, the algorithm is easy to diverge. Furthermore, the orientation and position estimation accuracy improvement of the proposed algorithm compared to other algorithms in this case is also limited, which is 0.8% and 11.05% better than the Tra-MSCKF, and 1.21% and 27.66% than the Inf-MSCKF, and 2.88% and 43.92% better than the RSTKF, and 2.74% and 16.06% better than the HKF.

We make the following explanations for this result. Images recorded in high-speed scenarios can easily become blurred, which weakens the ability of the system to extract feature points. Besides, the captured feature points can be lost quickly due to the high speed of the system. Furthermore, the system

moves most of the time in a straight and stationary line, where the gyroscope and the accelerometer lack motivation. Hence, little information can be utilized to estimate the biases of the gyroscope and the accelerometer. Therefore, all these reasons explain that localization of the VIO navigation system in a high-speed scenario can be troublesome and difficult.

Overall, through all the above simulation and experimental results, it can be concluded that the proposed algorithm can well accommodate non-stationary environments and provide accurate and consistent localization information for VIO navigation systems.

VII. CONCLUSION

In this paper, a novel ANGIG-KF for the VIO navigation system was proposed by jointly estimating the state vector and the MNCMC which are modeled as NGIG distributed, then the implementations of the improved MSCKF based on the proposed ANGIG-KF are detailed. A numerical calculation problem is introduced and two alternative methods based on the MAP and approximate mean are, respectively, proposed to address it, where the convergence, computational complexity, and initial parameters influence analyses are given to illustrate the validity of the proposed framework. Simulation and experimental dataset tests demonstrated that the proposed algorithm can achieve a good tracking performance of the true MNCM, which contributed to the better estimation accuracy and consistency over other algorithms.

VIII. APPENDIX

A. Proof of Proposition 1

The mean and covariance matrix of the state vector can be obtained with the PDF of the state vector which is calculated by integrating λ of (3), i.e.,

$$
\begin{aligned}
\mathrm{E}\{\mathbf{x}\} &= \int \mathbf{x} \int p(\mathbf{x}, \lambda) d\mathbf{x} d\lambda \\
&= \iint (\mathbf{x}\mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \lambda\boldsymbol{\Sigma})d\mathbf{x}) \times \mathrm{GIG}(\lambda; a, b, r)d\lambda \\
&= \int \boldsymbol{\mu} \times \mathrm{GIG}(\lambda; a, b, r)d\lambda = \boldsymbol{\mu}
\end{aligned}
\tag{27}
$$

$$
\begin{aligned}
\mathrm{Var}(\mathbf{x}) &= \mathrm{E}\{(\mathbf{x} - \hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}})^T\} = \int \tilde{\mathbf{x}}\tilde{\mathbf{x}}^T \int p(\mathbf{x}, \lambda)d\mathbf{x}d\lambda \\
&= \iint \left(\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \lambda\boldsymbol{\Sigma})\right) d\mathbf{x} \times \mathrm{GIG}(\lambda; a, b, r)d\lambda \\
&= \int \lambda\boldsymbol{\Sigma} \times \mathrm{GIG}(\lambda; a, b, r)d\lambda = \hat{\lambda}\boldsymbol{\Sigma}
\end{aligned}
$$

where $\tilde{\mathbf{x}} = \mathbf{x} - \hat{\mathbf{x}}$, and $\mathrm{E}\{\lambda\}$ can be found in [32].

B. Proof of Proposition 2

Without loss of generality, we let $\mathbf{\Sigma} = \mathbf{I}_n$. First, we marginalize λ in the joint NGIG distribution as

$$p(\mathbf{x}) = \int p(\mathbf{x}|\lambda)p(\lambda)d\lambda = \frac{\exp\{-0.5\lambda\mathbf{x}^T\mathbf{x}\}}{\sqrt{|2\pi\lambda\mathbf{I}_n|}}\frac{(a/b)^{p/2}}{2K_r(\sqrt{ab})}$$
$$\times \lambda^{(r-1)}\exp\{-0.5(a\lambda+b/\lambda)\}d\lambda \qquad (28)$$
$$= (2\pi)^{-0.5n}\frac{(a/b)^{p/2}}{2K_r(\sqrt{ab})}\int \lambda^{(r-0.5n-1)}$$
$$\times \exp\{-0.5(a\lambda+(b+\mathbf{x}^T\mathbf{x})/\lambda)\}d\lambda$$

By introducing the term $\frac{2K_{r'}(\sqrt{ab'})}{(a/b')^{r'/2}}$, equation (28) can be reformulated as

$$p(\mathbf{x}) = (2\pi)^{-0.5n}\frac{(a/b)^{r/2}}{K_r(\sqrt{ab})}\frac{K_{r'}(\sqrt{ab'})}{(a/b')^{r'/2}}\int \mathrm{GIG}(\lambda;a,b',r')d\lambda$$
$$= (2\pi)^{-0.5n}\frac{(a/b)^{r/2}}{K_r(\sqrt{ab})}\frac{K_{r'}(\sqrt{ab'})}{(a/b')^{r'/2}} \qquad (29)$$

where $b' = b+\mathbf{x}^T\mathbf{x}$ and $r' = r-0.5n$. Here we focus on the tail characteristic of the PDF, thus $\mathbf{x}^T\mathbf{x}$ and b' can be assumed to approach $+\infty$. Then, using the approximation $K_r(x) \approx \sqrt{\frac{\pi}{2x}}\exp\{-x\}$, $|x| \to +\infty$ of equation (1.5) in [44], we can obtain the following approximation,

$$K_{r'}(\sqrt{ab'}) \approx \sqrt{\frac{\pi}{2\sqrt{ab'}}}\exp\{-\sqrt{ab'}\} \qquad (30)$$

With (30), the logarithm PDF of the state vector is written as

$$\log p(\mathbf{x}) = -\frac{1}{4}\log(b+\mathbf{x}^T\mathbf{x})) - \sqrt{a(b+\mathbf{x}^T\mathbf{x})} + c_{\mathbf{x}}$$
$$\approx -\frac{1}{4}\log(\mathbf{x}^T\mathbf{x}) - \sqrt{a\mathbf{x}^T\mathbf{x}} + c_{\mathbf{x}}$$
$$(31)$$

We can obtain $\log p(y) = -\log y - \sqrt{a}y^2 + c_y$, where $4\log y = \log(\mathbf{x}^T\mathbf{x})$. The logarithm PDF of the Gaussian distributed state vector can be formulated as $\log p_{\mathcal{N}}(y) = -0.5y^4 + c_{y_{\mathcal{N}}}$, where $p_{\mathcal{N}}(y)$ denotes the Gaussian PDF.

It is evident that the Gaussian PDF $p_{\mathcal{N}}(y)$ decays faster than $p(y)$ when y, i.e., $\mathbf{x}^T\mathbf{x}$ tends to infinity because parameter a retains constant and bounded.

C. Proof of Theorem 1

Employing Bayes' rule, the joint prior PDF $p(\mathbf{x}_k, \lambda_k|\mathbf{z}_{1:k-1})$ can be written as

$$p(\mathbf{x}_k, \lambda_k|\mathbf{z}_{1:k-1}) = p(\mathbf{x}_k|\lambda_k, \mathbf{z}_{1:k-1})p(\lambda_k|\mathbf{z}_{1:k-1})$$
$$= \int p(\mathbf{x}_k|\mathbf{x}_{k-1})p(\mathbf{x}_{k-1}|\lambda_k, \mathbf{z}_{1:k-1})p(\lambda_k|\mathbf{z}_{1:k-1})d\mathbf{x}_{k-1}$$
$$\approx \mathcal{N}\left(\mathbf{x}_k; \mathbf{\Phi}_k\hat{\mathbf{x}}'_{k-1}, \mathbf{\Phi}_k\lambda_{k-1}\hat{\mathbf{P}}'_{k-1}\mathbf{\Phi}_k^T + \mathbf{Q}_k\right)p(\lambda_k|\mathbf{z}_{1:k-1})$$

where the approximation is based on the assumption that λ_k is very close to λ_{k-1}. Note that according to Bayes' rule, the prior PDF of the MNCMC can be calculated as $p(\lambda_k|\mathbf{z}_{1:k-1}) = \int p(\lambda_k|\lambda_{k-1})p(\lambda_{k-1}|\mathbf{z}_{1:k-1})d\lambda_{k-1}$, which is unavailable because the dynamic model $p(\lambda_k|\lambda_{k-1})$ is unknown. In this paper, we utilize a similar method in [41] to propagate the GIG distribution parameters with the forgetting factor ρ, which is shown in (7).

D. Proof of Theorem 2

According to Bayes' rule and (9) and (10), we can obtain

$$p_{\theta_k}(\mathbf{x}_k, \lambda_k|\mathbf{z}_{1:k}) = \frac{p(\mathbf{z}_k|\mathbf{x}_k, \lambda_k)p_{\theta_k}(\mathbf{x}_k, \lambda_k|\mathbf{z}_{1:k-1})}{p_{\theta_k}(\mathbf{z}_k|\mathbf{z}_{1:k-1})} \qquad (32)$$
$$\propto \mathcal{N}(\mathbf{z}_k; \mathbf{H}_k\mathbf{x}_k, \lambda_k\mathbf{I}_{m_k})\underbrace{\mathcal{N}(\mathbf{x}_k; \bar{\mathbf{x}}_k, \lambda_k\breve{\mathbf{P}}_k(\theta_k))}_{p_{\theta_k}(\mathbf{x}_k|\mathbf{z}_{1:k-1}, \lambda_k)}p(\lambda_k|\mathbf{z}_{1:k-1})$$

where $p_{\theta_k}(\mathbf{x}_k|\mathbf{z}_{1:k-1}, \lambda_k)$ denotes the conditional prior PDF of the state vector. Using the Gaussian integral formula [42], we can obtain

$$p_{\theta_k}(\mathbf{z}_k|\mathbf{z}_{1:k-1}, \lambda_k)$$
$$= \int \mathcal{N}(\mathbf{z}_k; \mathbf{H}_k\bar{\mathbf{x}}_k, \lambda_k\mathbf{I}_{m_k})\mathcal{N}(\mathbf{x}_k; \bar{\mathbf{x}}_k, \lambda_k\breve{\mathbf{P}}_k(\theta_k))d\mathbf{x}_k$$
$$= \mathcal{N}(\mathbf{z}_k; \mathbf{H}_k\bar{\mathbf{x}}_k, \lambda_k\mathbf{P}_k^{\mathbf{zz}}(\theta_k)) \qquad (33)$$

where $\mathbf{P}_k^{\mathbf{zz}}(\theta_k)$ is given in (12). Hence, with (32) and (33), the conditional posterior PDF of the state vector can be analytically calculated as (12).

E. Proof of Theorem 3

According to Bayes' rule, we can obtain

$$p_{\theta_k}(\mathbf{x}_k|\mathbf{z}_{1:k-1}, \lambda_k)p(\mathbf{z}_k|\mathbf{x}_k, \lambda_k)$$
$$= p_{\theta_k}(\mathbf{z}_k|\mathbf{z}_{1:k-1}, \lambda_k)p_{\theta_k}(\mathbf{x}_k|\mathbf{z}_{1:k}, \lambda_k) \qquad (34)$$

Employing (6), (11) and (33), we can attain

$$p(\mathbf{x}_k, \lambda_k|\mathbf{z}_{1:k}) \propto \mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}'_k(\theta_k), \lambda_k\hat{\mathbf{P}}'_k(\theta_k)) \qquad (35)$$
$$\times \mathcal{N}(\mathbf{z}_k; \mathbf{H}_k\bar{\mathbf{x}}_k, \lambda_k\mathbf{P}_k^{\mathbf{zz}}(\theta_k))\mathrm{GIG}(\lambda_k; \bar{a}_k, \bar{b}_k, \bar{r}_k)$$

Taking the logarithm of both sides of (35), we can obtain

$$\log p(\mathbf{x}_k, \lambda_k|\mathbf{z}_{1:k})$$
$$+ (\bar{r}_k - 1)\log\lambda_k - 0.5(\bar{a}_k\lambda_k + \bar{b}_k/\lambda_k) + c_{\lambda_k}$$
$$= \log\mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}'_k(\theta_k), \lambda_k\hat{\mathbf{P}}'_k(\theta_k)) - (\bar{r}_k - 0.5m_k - 1)\log\lambda_k$$
$$- 0.5(\bar{a}_k\lambda_k + (\bar{b}_k + \Delta_k)/\lambda_k) + c_{\lambda_k} \qquad (36)$$

where c_{λ_k} denotes a constant independent of λ_k, and Δ_k is given in (14). Through matching the coefficients, we can update the posterior parameters of the GIG distribution as shown in (14) without changing the conditional posterior density of $\mathbf{x}_k$.

F. Proof of Theorem 4

First, we prove that the iterative sequence $\{\theta_k^{(i)}\}_{i=0}^{+\infty}$ is monotonic in this case. From **Function 2**, we can obtain

$$\hat{\theta}_k^{(i)} = \frac{\hat{r}_k^{(i)} - 1 + \sqrt{(\hat{r}_k^{(i)}-1)^2 + \hat{a}_k^{(i)}\hat{b}_k^{(i)}}}{\hat{a}_k^{(i)}} \qquad (37)$$
$$= \frac{\bar{r}_k - 0.5m_k - 1 + \sqrt{(\bar{r}_k - 0.5m_k - 1)^2 + \bar{a}_k(\bar{b}_k + \Delta_k^{(i)})}}{\bar{a}_k}$$

where $\Delta_k^{(i)} = \mathrm{tr}\{\tilde{\mathbf{P}}_k^{\mathbf{zz}}(\mathbf{H}_k(\breve{\mathbf{P}}_k^{(i-1)})\mathbf{H}_k^T + \mathbf{I}_{m_k})^{-1}\}$.

If $\hat{\theta}_k^{(i-2)} < \hat{\theta}_k^{(i-1)}$, we can easily attain $\mathrm{tr}\{\tilde{\mathbf{P}}_k^{\mathbf{zz}}(\mathbf{P}_k^{\mathbf{zz},(i-2)})^{-1}\} < \mathrm{tr}\{\tilde{\mathbf{P}}_k^{\mathbf{zz}}(\mathbf{P}_k^{\mathbf{zz},(i-1)})^{-1}\}$. Employing (37), we can obtain $\hat{\theta}_k^{(i-1)} < \hat{\theta}_k^{(i)}$, which means $\{\theta_k^{(i)}\}_{i=0}^{+\infty}$ is monotone increasing. In the same way, $\hat{\theta}_k^{(i-1)} < \hat{\theta}_k^{(i)}$ will be

monotone decreasing if we set $\hat{\theta}_k^{(i-2)} > \hat{\theta}_k^{(i-1)}$. Therefore, the sequence $\{\theta_k^{(i)}\}_{i=0}^{+\infty}$ is monotonic.

Next, we will prove the sequence $\{\theta_k^{(i)}\}_{i=0}^{+\infty}$ is bounded. According to (37), $\theta_k > 0$, which makes the modified innovation covariance matrix $\mathbf{P}_k^{\mathbf{zz}}(\theta_k)$ positive-definite. Then, we can obtain $\Delta_k^{(i)} \geq 0$, which makes

$$\theta_k^{(i)} \geq \frac{\bar{r}_k - 0.5m_k - 1 + \sqrt{(\bar{r}_k - 0.5m_k - 1)^2 + \bar{a}_k \bar{b}_k}}{\bar{a}_k} \qquad (38)$$

where parameters $\bar{a}_k$, $\bar{b}_k$, $\bar{r}_k$ tend to stabilize over time.

According to (12), we can obtain $\mathbf{P}_k^{\mathbf{zz}}(\theta_k) > \mathbf{I}_{m_k}$, which can derive the following inequality

$$\theta_k^{(i)} < \frac{\bar{r}_k - 0.5m_k - 1 + \sqrt{(\bar{r}_k - 0.5m_k - 1)^2 + \bar{a}_k(\bar{b}_k + \mathrm{tr}\{\tilde{\mathbf{P}}_k^{\mathbf{zz}}\})}}{\bar{a}_k} \qquad (39)$$

Inequality (39) indicates that the sequence $\{\theta_k^{(i)}\}_{i=0}^{+\infty}$ is upper bounded if the residual vector $\tilde{\mathbf{z}}_k$ is bounded, which is easy to satisfy in engineering applications. Thus, utilizing (38) and (39), the sequence $\{\theta_k^{(i)}\}_{i=0}^{+\infty}$ is bounded.

G. Proof of Proposition 3

Following (20), we have

$$\hat{\lambda}_l = \frac{\hat{r}_l + \sqrt{\hat{r}_l^2 + \hat{a}_l \hat{b}_l}}{\hat{a}_l} \qquad (40)$$

Using (7), (14) and **Assumption 1**, we can obtain

$$\begin{cases} \hat{a}_l = \hat{a}_0 \\ \hat{b}_l = \rho^l \hat{b}_0 + \cdots + \rho^2 \Delta_{l-2} + \rho \Delta_{l-1} + \Delta_l \\ \hat{r}_l = \rho^l \hat{r}_0 - \cdots - 0.5\rho^2 m - 0.5\rho m - 0.5m \end{cases} \qquad (41)$$

To obtain the average $\hat{\lambda}_l$ in the statistical sense, we average both sides of (40), and employing Jensen's inequality [46] to deal with the square root operation in (40), and using the sum formula of a geometric sequence. Then (40) can be reformulated as follows.

$$\lim_{l \to +\infty} \mathrm{E}\{\hat{\lambda}_l\} \leq \frac{\frac{-0.5m}{1-\rho} + \sqrt{(\frac{-0.5m}{1-\rho})^2 + \frac{\lambda \hat{a}_0 m}{1-\rho}}}{\hat{a}_0} \qquad (42)$$

The difference d_λ between $\lim_{l \to +\infty} \mathrm{E}\{\hat{\lambda}_l\}$ and λ can be written as

$$d_\lambda \leq \frac{-(1-\rho)\lambda^2}{\left((\sqrt{0.25m^2 + \lambda m} + 0.5m)/\hat{a}_0 + (1-\rho)\lambda\right)} < 0 \qquad (43)$$

It can be seen from (41) and (43) that the influences of $\hat{b}_0$ and $\hat{r}_0$ on d_λ decay exponentially over time, and the absolute value of d_λ reduces when $\hat{a}_0$ decreases.

REFERENCES

[1] G. Klein and D. Murray, "Parallel tracking and mapping for small AR workspaces," In *2007 6th IEEE and ACM international symposium on mixed and augmented reality*, 2007, pp. 225–234.

[2] 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, 2015.

[3] T. Qin, P. Li and S. Shen, "VINS-mono: A robust and versatile monocular visual-inertial state estimator," *IEEE Transactions on Robotics*, vol. 34, no. 4, pp. 1004–1020, Aug. 2018.

[4] T. Qin, J. Pan, S. Cao and S. Shen, "A general optimization-based framework for local odometry estimation with multiple sensors," 2019, arXiv: 1901.03638.

[5] A. I. Mourikis and S. I. Roumeliotis, "A Multi-State Constraint Kalman Filter for Vision-aided Inertial Navigation," In *Proceedings 2007 IEEE International Conference on Robotics and Automation*, 2007, pp. 3565–3572.

[6] M. Y. 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, 2013.

[7] M. Bloesch, M. Burri, S. Omari, M. Hutter, and R. Siegwart, "Iterated extended Kalman filter based visual-inertial odometry using direct photometric feedback," *The International Journal of Robotics Research*, vol. 36, no. 10, pp. 1053–1072, 2017.

[8] M. Bloesch, S. Omari, M. Hutter and R. Siegwart, "Robust visual inertial odometry using a direct EKF-based approach," In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2015, pp. 298–304.

[9] K. Sun et al., "Robust Stereo Visual Inertial Odometry for Fast Autonomous Flight," *IEEE Robotics and Automation Letters*, vol. 3, no. 2, pp. 965–972, Apr. 2018.

[10] X. Qiu, H. Zhang, and W, Fu, "Lightweight hybrid visual-inertial odometry with closed-form zero velocity update," *Chinese Journal of Aeronautics*, vol. 33, no. 12, pp. 3344–3359, 2020.

[11] T. Zhu and H. Ma, "Challenges of Dynamic Environment for Visual-Inertial Odometry," In *2018 3rd International Conference on Robotics and Automation Engineering (ICRAE)*, 2018, pp. 82–86.

[12] C. Zhai, M. Wang, Y. Yang and K. Shen, "Robust Vision-Aided Inertial Navigation System for Protection Against Ego-Motion Uncertainty of Unmanned Ground Vehicle," *IEEE Transactions on Industrial Electronics*, vol. 68, no. 12, pp. 12462–12471, Dec. 2021.

[13] Z. Zhang, P. Dong, J. Wang and Y. Sun, "Improving S-MSCKF with Variational Bayesian Adaptive Nonlinear Filter," *IEEE Sensors Journal*, vol. 20, no. 16, pp. 9437–9448, Aug. 2020.

[14] S. Liu et al., "High precision camera calibration and distortion compensation method based on CMM," C.N. Patent CN107481290A, Dec. 2017.

[15] J. Duník , O. Straka, O. Kost, et al., "Noise covariance matrices in state-space models: A survey and comparison of estimation methods-Part I", *International Journal of Adaptive Control and Signal Processing*, vol. 31, no. 11, pp. 1505–1543, 2017.

[16] S. Sarkka and A. Nummenmaa, "Recursive Noise Adaptive Kalman Filtering by Variational Bayesian Approximations," *IEEE Transactions on Automatic Control*, vol. 54, no. 3, pp. 596–600, March 2009.

[17] Y. Huang, Y. Zhang, Z. Wu, N. Li, and J. Chambers, "A novel adaptive Kalman filter with inaccurate process and measurement noise covariance matrices," *IEEE Transactions on Automatic Control*, vol. 63, no. 2, pp. 594–601, Feb. 2018.

[18] Y. Huang, Y. Zhang, Y. Zhao, and J. Chambers, "A novel robust Gaussian-Student's t mixture distribution based Kalman filter," *IEEE Transactions on Signal Processing*, vol. 67, no. 13, pp. 3606–3620, Jul. 2019.

[19] X. R. Li and Y. Bar-Shalom, "A recursive multiple model approach to noise identification," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 30, no. 3, pp. 671–684, Jul. 1994.

[20] D. Lainiotis, "Optimal adaptive estimation: Structure and parameter adaption," *IEEE Transactions on Automatic Control*, vol. 16, no. 2, pp. 160–170, 1971.

[21] R. Mehra, "On the identification of variances and adaptive Kalman filtering," *IEEE Transactions on Automatic Control*, vol. 15, no. 2, pp. 175-184, Apr. 1970.

[22] J. Duník, O. Straka and O. Kost, "Measurement difference autocovariance method for noise covariance matrices estimation," In *2016 IEEE 55th Conference on Decision and Control (CDC)*, 2016, pp. 365–370.

[23] R. L. Kashyap, "Maximum likelihood identification of stochastic linear systems," *IEEE Transactions on Automatic Control*, vol. 15, no. 1, pp. 25–34, 1970.

[24] R. H. Shumway and D. S. Stoffer, "An approach to time series smoothing and forecasting using the EM algorithm," *Journal of Time Series Analysis*, vol. 3, no. 4, pp. 253–264, 1982.

[25] A. H. Mohamed and K. P. Schwarz, "Adaptive Kalman filtering for INS/GPS," *Journal of Geodesy*, vol. 73, no. 4, pp. 193–203, May. 1999.

[26] M. Yu, "INS/GPS integration system using adaptive filter for estimating measurement noise variance," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 2, pp. 1786–1792, Apr. 2012.

[27] A. P. Sage and G. W. Husa, "Adaptive filtering with unknown prior statistics," In *Proceedings of the Joint Automatic Control Conference,*

1969, pp. 760–769.

[28] K. A. Myers and B. D. Tapley, "Adaptive sequential estimation with unknown noise statistics," *IEEE Transactions on Automatic Control*, vol. 21, no. 4, pp. 520–523, 1976.

[29] C. D. Karlgaard and H. Schaub, "Huber-based divided difference filtering," *Journal of Guidance Control & Dynamics,* vol. 30, no. 3, pp. 885–891, 2007.

[30] Y. Huang, Y. Zhang, N. Li, Z. Wu and J. A. Chambers, "A Novel Robust Student's t-Based Kalman Filter," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 53, no. 3, pp. 1545–1554, Jun. 2017.

[31] Z. Yue, B. Lian and Y. Gao, "Robust adaptive filter using fuzzy logic for tightly-coupled visual inertial odometry navigation system," *IET Radar Sonar & Navigation*, vol. 14, no. 3, pp. 364-371, 2020.

[32] L. Perreault, B Bobée, and P. F. Rasmussen, "Halphen Distribution System. I: Mathematical and Statistical Properties," *Journal of Hydrologic Engineering*, vol. 4, no. 3, pp. 189–199, 1999.

[33] V. Seshadri. "Halphen's Laws," *John Wiley & Sons*, Inc. 2004.

[34] F. Mechel, "Calculation of the modified Bessel functions of the second kind with complex argument," *Mathematics of Computation*, vol. 20, no. 95, pp. 407–407, 1996.

[35] G. P. Huang, A. I. Mourikis and S. I. Roumeliotis. "Observability-based Rules for Designing Consistent EKF SLAM Estimators," *International Journal of Robotics Research*, vol. 29, no. 5, pp. 502–528, 2010.

[36] J. A. Hesch, D. G. Kottas, S. L. Bowman and S. I. Roumeliotis, "Observability-constrained vision-aided inertial navigation", 2012.

[37] M. Burri et al., "The EuRoC micro aerial vehicle datasets," *The International Journal of Robotics Research*, vol. 35, no. 10, pp. 1157–1163, 2016.

[38] A. Geiger, P. Lenz and R. Urtasun, "Are we ready for autonomous driving? THE KITTI vision benchmark suite," In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 3354–3361.

[39] P. Geneva, K. Eckenhoff, W. Lee, Y. Yang and G. Huang, "OpenVINS: A Research Platform for Visual-Inertial Estimation," In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, 2020, pp. 4666–4672.

[40] W. Vega, A. Bachrach, A. Bry, J. Kelly and N. Roy, "CELLO: A Fast Algorithm for Covariance Estimation," In *2013 IEEE International Conference on Robotics and Automation*, 2013, pp. 3160–3167.

[41] S. Sarkka and J. Hartikainen, "Variational Bayesian adaptation of noise covariance in nonlinear Kalman filtering," 2013, arXiv: 1302.0681.

[42] B. N. Vo and W. K. Ma, "The Gaussian Mixture Probability Hypothesis Density Filter," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4091–4104, Nov. 2006.

[43] J. Segura, "Bounds for ratios of modified Bessel functions and associated Turán-type inequalities," *Journal of Mathematical Analysis and Applications*, vol. 374, no. 2, pp. 516–528, Sep. 2010.

[44] Z. Yang and Y. Chu, "On approximating the modified Bessel function of the second kind," *Journal of Inequalities and Applications*, vol. 2017, no. 1, pp. 1–8, Feb. 2017.

[45] D. G. Tzikas, A. C. Likas and N. P. Galatsanos, "The Variational Approximation for Bayesian Inference Life After the EM Algorithm," *IEEE Signal Processing Magazine*, vol. 25, no. 6, pp. 131–146, Nov. 2008.

[46] J. L. W. V. Jensen, "Sur les fonctions convexes et les inégalités entre les valeurs moyennes," *Acta Mathematica*, vol. 30, no. 1, pp. 175–193, Jan. 1906.