MSCEqF: A Multi State Constraint Equivariant Filter for Vision-aided Inertial Navigation

Alessandro Fornasier[1], Pieter van Goor[2], Eren Allak[1], Robert Mahony[2] and Stephan Weiss[1]

***Abstract*—This letter re-visits the problem of visual-inertial navigation system (VINS) and presents a novel filter design we dub the multi state constraint *equivariant* filter (MSC*EqF*, in analogy to the well known MSCKF). We define a symmetry group and corresponding group action that allow specifically the design of an equivariant filter for the problem of visual-inertial odometry (VIO) including IMU bias, and camera intrinsic and extrinsic calibration states. In contrast to state-of-the-art invariant extended Kalman filter (IEKF) approaches that simply tack IMU bias and other states onto the $\mathrm{SE}_2(3)$ group, our filter builds upon a symmetry that properly includes all the states in the group structure. Thus, we achieve improved behavior, particularly when linearization points largely deviate from the truth (i.e., on transients upon state disturbances). Our approach is *inherently consistent* even during convergence phases from significant errors without the need for error uncertainty adaptation, observability constraint, or other consistency enforcing techniques. This leads to greatly improved estimator behavior for significant error and unexpected state changes during, e.g., long-duration missions. We evaluate our approach with a multitude of different experiments using three different prominent real-world datasets.**

***Index Terms*—Vision-Based Navigation, Visual-Inertial SLAM**

I. Introduction and Related Work

In the past years, VINS have shown remarkable success in estimating the position and orientation of robots by relying only on low-cost and lightweight IMUs and cameras.

Popular algorithms for VINS include visual-inertial odometry (VIO) and visual-inertial simultaneous localization and mapping (VI-SLAM). VIO focuses only on the local surroundings and is, therefore, computationally simpler, less accurate, and it suffers from accumulated drift. VINS algorithms can also suffer from inconsistencies [1]. The classical extended Kalman filter (EKF)-SLAM algorithm suffers from overconfidence due to spurious information gain along the unobservable directions [2]. Different solutions have been proposed in literature to overcome the problems caused by inconsistencies. By manipulating the linearization point and enforcing the correct number of unobservable directions for the linearized system, Huang *et al.* introduced the first estimate jacobian (FEJ) [3], whereas Hesch *et al.* the observability constraint (OC) [1] as techniques aiming at solving the inconsistency issue at the cost of sub-optimal linearization points. More recently, in [4], Barrau and Bonnabel introduced the IEKF and showed that exploiting the natural symmetry of group affine systems leads to algorithms that are inherently consistent [5]. Although the IEKF theory does not apply to inertial navigation systems (INS) when IMU bias are explicitly considered, many authors [6, 7, 8, 9, 10, 11, 12] have exploited the Imperfect-IEKF framework [13] to design VINS algorithms.

In very recent research, van Goor *et al.* introduced the EqF [14, 15] as a general filter design for systems on homogeneous spaces, and proposed a symmetry for fixed landmark measurements in the context of VI-SLAM [16, 17, 18, 19, 20]. Later, Fornasier *et al.* proposed a novel symmetry for INS that couples navigation states and IMU bias and developed an EqF design for INS [21, 22] that proved superior to state-of-the-art in terms of robustness to wrong initialization, transient behavior, and consistency properties. In a very recent research study [23], the same authors analyzed the theoretical properties of different symmetry groups when employed in designing filters for inertial navigation systems, and provided a discussion of the relative strengths and weaknesses of different filter algorithms.

For *vision* aided INS systems, however, the lack of robustness against unexpected disturbances and the requirement for sophisticated tuning for a given environment and setup remain important limitations. Real-world deployments are typically constrained to precise tuning and highly engineered codebases, where the core VIO algorithm is encompassed by numerous modules responsible for tasks such as initialization, failure detection, algorithm reset, and more. A *people's visual-inertial odometry*, that is, an algorithm whose operation requires minimal knowledge, little to no tuning, and yet still functions in many different real-world scenarios, would enable a whole new tranch of real-world applications without the requirement of having highly trained engineers available. The present letter builds upon the recent results in [21, 22, 23] and is a step towards enabling this goal.

This perspective shifts the evaluation of algorithm performance from measures such as root mean square error (RMSE), accuracy, and precision, to measures such as the likelihood of failure for poor initial conditions or poor calibration. We

Manuscript received: July, 2, 2023; Revised October, 13, 2023; Accepted November, 13, 2023.

This paper was recommended for publication by Editor Pascal Vasseur upon evaluation of the Associate Editor and Reviewers' comments. This work was supported by the European Union's Horizon 2020 research and innovation program under grant agreement 871260 (BugWright2), and by the Army Research Office under Cooperative Agreement Number W911NF-21-2-0245. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Office or the U.S. Government. The U.S. Government is authorized to reproduce and distribute preprints for Government purposes notwithstanding any copyright notation herein.

[1] Alessandro Fornasier, Eren Allak and Stephan Weiss are with the Control of Networked Systems Group, University of Klagenfurt, Austria. {name.surname}@ieee.org

[2] Pieter van Goor and Robert Mahony are with the System Theory and Robotics Lab, Australian National University, Australia. {name.surname}@anu.edu.au

acknowledge that state-of-the-art VINS approaches reached a plateau in the former metrics, but there is still a large room for improvement in the latter metrics. Furthermore, this letter does not claim completeness in comparative evaluations, rather, we present here our novel findings enabling a multi state constraint equivariant filter (MSCEqF) as a step towards the *people's VIO*; compare it against OpenVINS [24], the best open-source available MSCKF [25], and see an extensive comparison covering all suitable approaches as a work that goes beyond the scope of this letter.

Apart from the different metric evaluation, this work differentiates itself from state-of-the-art by extending insights on symmetries and EqF design for fixed landmark VINS [19, 20] and INS including IMU bias into the symmetry [21, 22, 23] to the idea of a multi state constraint but *equivariant* VINS. To the best of our knowledge, the resulting algorithm is the first ever, equivariant multi state constraint filter for VIO. Our approach, dubbed MSCEqF, leverages a semi-direct product symmetry group, yielding improved linearized error dynamics when compared to other filter types [23]. Hence, the MSCEqF demonstrates consistency naturally without artificial changes of linearization points and very high robustness to poor extrinsic calibration. It not only handles significant absolute (calibration) errors but also addresses the concept of *dealing with "you don't know what you don't know"*, such as errors exceeding the prior covariance (e.g., sudden changes of calibrations states due to a disturbance during the operational phase of the robotic platform, where the state has converged already and the covariance has shrunk).

To summarize, with this work, we make the following contributions:

(i): We introduce the MSCEqF; a novel multi state constraint visual-inertial navigation system based on the equivariant filter framework, with camera and IMU self-calibration capabilities.

(ii): We demonstrate that the proposed MSCEqF achieves state-of-the-art accuracy, with superior robustness to significant absolute errors, as well as errors exceeding the prior covariance.

Our experiments show that the MSCEqF can be directly deployed in real-world scenarios with little tuning and no additional health-check modules. Furthermore, we show that the proposed MSCEqF is a naturally consistent filter without the need for FEJ, OC, or other heuristic techniques. We implemented our framework as a stand-alone C++ library, and we made it source-available to the community[1]. Wrappers for the standard middle-ware (e.g., ROS1, ROS2, etc.) will be provided such that code is available for direct use and comparison against other approaches. We derived the filter matrices in analytical form without resorting to numerical differentiation, leading to code with higher portability and lower computational complexity, appropriate for compute-limited hardware, such as nano-drones, augmented reality devices, etc.

II. Mathematical Preliminaries and Notation

A. Vector and matrix notation

Vectors describing physical quantities expressed in frame of reference $\{A\}$ are denoted by ${}^{A}\boldsymbol{v}$. Rotation matrices encoding the orientation of a frame of reference $\{B\}$ with respect to a reference $\{A\}$ are denoted by ${}^{A}\mathbf{R}_{B}$; in particular, ${}^{A}\boldsymbol{v} = {}^{A}\mathbf{R}_{B}\,{}^{B}\boldsymbol{v}$. $\mathbf{I}_n \in \mathbb{R}^{n \times n}$ denotes the n-dim identity matrix, and $\mathbf{0}_{n \times m} \in \mathbb{R}^{n \times m}$ denotes the zero matrix with n rows and m columns.

B. Lie theory

A Lie group $\mathbf{G}$ is a smooth manifold endowed with a smooth group structure. For any $X, Y \in \mathbf{G}$, the group multiplication is denoted XY, the group inverse X^{-1} and the identity element I.

Given a Lie group $\mathbf{G}$, $\mathcal{G}$ denotes the $\mathbf{G}$-Torsor [26].

For a given Lie group $\mathbf{G}$, the Lie algebra $\mathfrak{g}$ is a vector space corresponding to the tangent space at the identity of the group, together with a bilinear non-associative map $[\cdot,\cdot] : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$ called the *Lie bracket*. The Lie algebra $\mathfrak{g}$ is isomorphic to a vector space $\mathbb{R}^n$ of dimension $n = \dim(\mathfrak{g})$.

Define the *wedge* map and its inverse, the *vee* map as linear isomorphisms between the vector space and the Lie algebra

$$(\cdot)^{\wedge} : \mathbb{R}^n \to \mathfrak{g}, \qquad (\cdot)^{\vee} : \mathfrak{g} \to \mathbb{R}^n,$$

such that $(\boldsymbol{u}^{\wedge})^{\vee} = \boldsymbol{u}$, for all $\boldsymbol{u} \in \mathbb{R}^n$.

For any $X, Y \in \mathbf{G}$, define the left and right translations

$$\begin{aligned} \mathrm{L}_X &: \mathbf{G} \to \mathbf{G}, & \mathrm{L}_X(Y) &= XY, \\ \mathrm{R}_X &: \mathbf{G} \to \mathbf{G}, & \mathrm{R}_X(Y) &= YX. \end{aligned}$$

The Lie group ('big') Adjoint matrix is defined by

$$\mathbf{Ad}^{\vee}_{X} : \mathbb{R}^n \to \mathbb{R}^n, \qquad \mathbf{Ad}^{\vee}_{X}\,\boldsymbol{u} = \left(\mathrm{dL}_X \mathrm{dR}_{X^{-1}}\left[\boldsymbol{u}^{\wedge}\right]\right)^{\vee},$$

for every $X \in \mathbf{G}$ and $\boldsymbol{u}^{\wedge} \in \mathfrak{g}$, where dL_X, and dR_X denote the differentials of the left, and right translation, respectively.

The Lie algebra ('little') adjoint matrix is defined by

$$\mathbf{ad}^{\vee}_{\boldsymbol{u}} : \mathbb{R}^n \to \mathbb{R}^n, \qquad \mathbf{ad}^{\vee}_{\boldsymbol{u}}\boldsymbol{v} = \left[\boldsymbol{u}^{\wedge}, \boldsymbol{v}^{\wedge}\right]^{\vee},$$

for every $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$.

C. Important matrix Lie groups

The special orthogonal group $\mathbf{SO}(3)$, special Euclidean group $\mathbf{SE}(3)$, extended special Euclidean group $\mathbf{SE}_2(3)$, and their respective Lie algebras are defined, in matrix form, by

$$\begin{aligned}
\mathbf{SO}(3) &= \left\{\mathbf{A} \in \mathbb{R}^{3\times 3} \,\middle|\, \mathbf{A}\mathbf{A}^{\top} = \mathbf{I}_3,\ \det(\mathbf{A}) = 1\right\}, \\
\mathfrak{so}(3) &= \left\{\boldsymbol{\omega}^{\wedge} \in \mathbb{R}^{3\times 3} \,\middle|\, \boldsymbol{\omega}^{\wedge} = -\boldsymbol{\omega}^{\wedge\top}\right\}, \\
\mathbf{SE}(3) &= \left\{\begin{bmatrix} \mathbf{A} & \boldsymbol{a} \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix} \in \mathbb{R}^{4\times 4} \,\middle|\, \mathbf{A} \in \mathbf{SO}(3),\ \boldsymbol{a} \in \mathbb{R}^3\right\}, \\
\mathfrak{se}(3) &= \left\{\begin{bmatrix} \boldsymbol{\omega}^{\wedge} & \boldsymbol{v} \\ \mathbf{0}_{1\times 3} & 0 \end{bmatrix} \in \mathbb{R}^{4\times 4} \,\middle|\, \boldsymbol{\omega}^{\wedge} \in \mathfrak{so}(3),\ \boldsymbol{v} \in \mathbb{R}^3\right\}, \\
\mathbf{SE}_2(3) &= \left\{\begin{bmatrix} \mathbf{A} & \boldsymbol{a} & \boldsymbol{b} \\ \mathbf{0}_{2\times 3} & \multicolumn{2}{c}{\mathbf{I}_2} \end{bmatrix} \in \mathbb{R}^{5\times 5} \,\middle|\, \mathbf{A} \in \mathbf{SO}(3),\ \boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^3\right\}, \\
\mathfrak{se}_2(3) &= \left\{\begin{bmatrix} \boldsymbol{\omega}^{\wedge} & \boldsymbol{v} & \boldsymbol{w} \\ \mathbf{0}_{2\times 3} & \multicolumn{2}{c}{\mathbf{0}_{2\times 2}} \end{bmatrix} \in \mathbb{R}^{5\times 5} \,\middle|\, \boldsymbol{\omega}^{\wedge} \in \mathfrak{so}(3),\ \boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^3\right\}.
\end{aligned}$$

[1] https://github.com/aau-cns/MSCEqF

D. Semi-direct Bias group $\mathbf{G_{SD}} \coloneqq \mathbf{SE}_2(3) \ltimes \mathfrak{se}(3)$

The *Semi-direct Bias group* $\mathbf{G_{SD}} \coloneqq \mathbf{SE}_2(3) \ltimes \mathfrak{se}(3)$ introduced in [23], is a group structure on the tangent bundle $\mathbf{G}^{\ltimes}_{\mathfrak{g}} \coloneqq \mathbf{G} \ltimes \mathfrak{g}$ given by the semi-direct product of a group $\mathbf{G}$ with a Lie subalgebra $\mathfrak{g}$.

For a detailed introduction to equivariant filters for inertial navigation systems, semi-direct product groups and theoretical properties this work is built upon, we refer the reader to our previous works [21, 22, 23]. Moreover, [23] discuss the advantages of semi-direct product symmetries for filter design and compares it to classical solutions such as the MEKF and the IEKF.

E. Intrinsics group IN

In this work, we recognized that elements of the camera intrinsics matrix [27] form a Lie group. Thus, we introduce the intrisincs group $\mathbf{IN}$, as the matrix Lie group defined by

$$\mathbf{IN} = \left\{ \mathbf{K} = \begin{bmatrix} a & 0 & x \\ 0 & b & y \\ 0 & 0 & 1 \end{bmatrix} \in \mathbb{R}^{3\times 3} \;\middle|\; a, b > 0,\ x, y \in \mathbb{R} \right\}.$$

This matrix representation is associated with the standard camera intrinsics matrix, well-known in computer vision. A typical element of $\mathbf{IN}$ may be written as $\mathbf{K} = (a, b, x, y)$. Let $\mathbf{K}_1, \mathbf{K}_2 \in \mathbf{IN}$, then

$$\begin{aligned} \mathbf{K}_1\mathbf{K}_2 &= (a_1a_2, b_1b_2, x_1 + a_1x_2, y_1 + b_1y_2), \\ \mathbf{K}_1^{-1} &= (a_1^{-1}, b_1^{-1}, -a_1^{-1}x_1, -b_1^{-1}y_1). \end{aligned}$$

To the authors' understanding, exploiting the group structure of the $\mathbf{IN}$ group in equivariant or invariant VINS design represents a novel approach to this work.

F. Useful maps

For all $\boldsymbol{v} = (x, y, z) \in \mathbb{R}^3$, define the maps

$$\begin{aligned} &\pi_{Z_1}(\cdot) : \mathbb{R}^3 \to \mathbb{R}^3, \quad \pi_{Z_1}(\boldsymbol{v}) \coloneqq \frac{\boldsymbol{v}}{z}, \\ &\Xi(\cdot) : \mathbb{R}^3 \to \mathbb{R}^{3\times 4}, \quad \Xi(\boldsymbol{v}) = \begin{bmatrix} x & 0 & z & 0 \\ 0 & y & 0 & z \\ 0 & 0 & 0 & 0 \end{bmatrix} \in \mathbb{R}^{3\times 4}. \end{aligned}$$

For all $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c} \in \mathbb{R}^3 \mid (\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) \in \mathbb{R}^9$, define the maps

$$\begin{aligned} &\Pi(\cdot) : \mathfrak{se}_2(3) \to \mathfrak{se}(3), \quad \Pi\left((\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})^\wedge\right) = (\boldsymbol{a}, \boldsymbol{b})^\wedge \in \mathfrak{se}(3), \\ &\Upsilon(\cdot) : \mathfrak{se}_2(3) \to \mathfrak{se}(3), \quad \Upsilon\left((\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})^\wedge\right) = (\boldsymbol{a}, \boldsymbol{c})^\wedge \in \mathfrak{se}(3), \end{aligned}$$

For all $X = (A, a) \in \mathbf{SE}(3) \mid A \in \mathbf{SO}(3), a \in \mathbb{R}^3$, define

$$\Gamma(\cdot) : \mathbf{SE}(3) \to \mathbf{SO}(3), \quad \Gamma(X) = A \in \mathbf{SO}(3).$$

For all $X = (A, a, b) \in \mathbf{SE}_2(3) \mid A \in \mathbf{SO}(3), a, b \in \mathbb{R}^3$, define

$$\begin{aligned} &\chi(\cdot) : \mathbf{SE}_2(3) \to \mathbf{SE}(3), \quad \chi(X) = (A, a) \in \mathbf{SE}(3), \\ &\Theta(\cdot) : \mathbf{SE}_2(3) \to \mathbf{SE}(3), \quad \Theta(X) = (A, b) \in \mathbf{SE}(3). \end{aligned}$$

III. Visual Inertial Navigation System

A. System definition

Consider a mobile platform equipped with a camera observing global visual features ${}^G\boldsymbol{p}_f$, and an IMU providing biased acceleration and angular velocity measurements, denoted by ${}^I\boldsymbol{w} = \left({}^I\boldsymbol{\omega}, {}^I\boldsymbol{a}\right)$. Define ${}^G\mathbf{T}_I = \left({}^G\mathbf{R}_I, {}^G\boldsymbol{v}_I, {}^G\boldsymbol{p}_I\right)$ to be the extended pose of the system, where ${}^G\mathbf{R}_I$ corresponds to the rigid body orientation, whereas ${}^G\boldsymbol{p}_I$ and ${}^G\boldsymbol{v}_I$ denote the IMU position and velocity with respect to the global frame, respectively. Define ${}^G\mathbf{P}_I = \left({}^G\mathbf{R}_I, {}^G\boldsymbol{p}_I\right)$. Define ${}^I\boldsymbol{b} = \left({}^I\boldsymbol{b}_{\boldsymbol{\omega}}, {}^I\boldsymbol{b}_{\boldsymbol{a}}\right)$ to be the gyroscope and accelerometer biases, respectively. Let g denote the magnitude of the acceleration due to gravity, and let ${}^G\boldsymbol{e}_3$ denote the direction of gravity in the global frame. Finally, define ${}^I\mathbf{S}_C$ to be the camera extrinsic calibration, and $\mathbf{K}$ be the camera intrinsic calibration.

For the sake of readability, from now on, we suppress all the subscripts and superscripts that are not strictly required.

Define the matrices $\mathbf{W}, \mathbf{B}, \mathbf{D}, \mathbf{G}$ to be

$$\mathbf{W} = \begin{bmatrix} \boldsymbol{\omega}^\wedge & \boldsymbol{a} & \mathbf{0}_{3\times 1} \\ \mathbf{0}_{1\times 3} & 0 & 0 \\ \mathbf{0}_{1\times 3} & 0 & 0 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} \boldsymbol{b}_{\boldsymbol{\omega}}^\wedge & \boldsymbol{b}_a & \mathbf{0}_{3\times 1} \\ \mathbf{0}_{1\times 3} & 0 & 0 \\ \mathbf{0}_{1\times 3} & 0 & 0 \end{bmatrix},$$

$$\mathbf{D} = \begin{bmatrix} \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 1} & \mathbf{0}_{3\times 1} \\ \mathbf{0}_{1\times 3} & 0 & 1 \\ \mathbf{0}_{1\times 3} & 0 & 0 \end{bmatrix}, \quad \mathbf{G} = \begin{bmatrix} \mathbf{0}_{3\times 3} & g\,\boldsymbol{e}_3 & \mathbf{0}_{3\times 1} \\ \mathbf{0}_{1\times 3} & 0 & 0 \\ \mathbf{0}_{1\times 3} & 0 & 0 \end{bmatrix}.$$

Finally, the visual-inertial navigation system is written

$$\dot{\mathbf{T}} = \mathbf{T}\left(\mathbf{W} - \mathbf{B} + \mathbf{D}\right) + \left(\mathbf{G} - \mathbf{D}\right)\mathbf{T}, \tag{1a}$$

$$\dot{\boldsymbol{b}} = \boldsymbol{\tau}, \tag{1b}$$

$$\dot{\mathbf{S}} = \mathbf{S}\,\boldsymbol{\mu}^\wedge, \tag{1c}$$

$$\dot{\mathbf{K}} = \mathbf{K}\,\boldsymbol{\zeta}^\wedge, \tag{1d}$$

where $\boldsymbol{\tau}, \boldsymbol{\mu}, \boldsymbol{\zeta}$ are used to model the deterministic dynamics of the bias and calibration states and are zero when these states are modeled as constants, as they are in our formulation.

Define $\xi_I = (\mathbf{T}, \boldsymbol{b}) \in \mathcal{SE}_2(3) \times \mathbb{R}^6$ to be the inertial navigation state. Define $\xi_S = (\mathbf{S}, \mathbf{K}) \in \mathcal{SE}(3) \times \mathcal{IN}(3)$ to be the camera calibration state. Then the full system state is defined as $\xi = (\xi_I, \xi_S) \in \mathcal{M} \coloneqq \mathcal{SE}_2(3) \times \mathbb{R}^6 \times \mathcal{SE}(3) \times \mathcal{IN}(3)$. Define $u = (\boldsymbol{w}, \boldsymbol{\tau}, \boldsymbol{\mu}, \boldsymbol{\zeta}) \in \mathbb{L} \subset \mathbb{R}^{18}$ to be the system's input. Note that in this work, visual features are not considered as part of the state since the dependency of measurement on features is removed through nullspace projection.

Without loss of generality, let us consider the case of a single feature $\boldsymbol{p}_f$. The camera measurement is modeled as the measurement of the bearing of the feature $\boldsymbol{p}_f$ seen from the camera.

$$h(\xi, \boldsymbol{p}_f) = \mathbf{K}\pi_{Z_1}\left((\mathbf{PS})^{-1} * \boldsymbol{p}_f\right), \tag{2}$$

where the operation $* : \mathcal{SE}(3) \times \mathbb{R}^3 \to \mathbb{R}^3$ is defined by $\mathbf{P} * \boldsymbol{v} = \mathbf{R}\,\boldsymbol{v} + \boldsymbol{p}$ for all $\mathbf{P} = (\mathbf{R}, \boldsymbol{p}) \in \mathcal{SE}(3),\ \boldsymbol{v} \in \mathbb{R}^3$.

B. Symmetry of the visual-inertial navigation system

The symmetry for the inertial navigation state ξ_I is given by the Semi-Direct symmetry group $\mathbf{G_{SD}} \coloneqq (\mathbf{SE}_2(3) \ltimes \mathfrak{se}(3))$, the symmetry for the extrinsic calibration state is given by the special Euclidean group $\mathbf{SE}(3)$, and the symmetry for the intrinsic calibration state is given by the intrinsics group $\mathbf{IN}$. The complete symmetry for the visual-inertial navigation system is thus defined to be the product group $\mathbf{G} \coloneqq \mathbf{G_{SD}} \times \mathbf{SE}(3) \times \mathbf{IN}$.

Let $X = \left(\left(D, \delta\right), E, L\right) \in \mathbf{G}$, with $D = \left(A, a, b\right) \in \mathbf{SE}_2(3)$ such that $A \in \mathbf{SO}(3)$, $a, b \in \mathbb{R}^3$. Define the subgroups $B = \chi\left(D\right) \in \mathbf{SE}(3)$, and $C = \Theta\left(D\right) \in \mathbf{SE}(3)$. Finally, define $E \in \mathbf{SE}(3)$, and $L \in \mathbf{IN}$.

Lemma 3.1. *Define $\phi : \mathbf{G} \times \mathcal{M} \to \mathcal{M}$ as*

$$\phi\left(X, \xi\right) \coloneqq \left(\mathbf{T}D, \mathbf{Ad}^\vee_{B^{-1}}\left(\boldsymbol{b} - \delta^\vee\right), C^{-1}\mathbf{S}E, \mathbf{K}L\right) \in \mathcal{M}. \tag{3}$$

Then, ϕ is a transitive right group action of $\mathbf{G}$ on $\mathcal{M}$.

C. Lifted system

The implementation of the equivariant filter (EqF) requires a lift $\Lambda : \mathcal{M} \times \mathbb{L} \to \mathfrak{g}$ to define a lifted system on the symmetry group $\mathbf{G}$ that projects down to the original system dynamics via the proposed group action ϕ. The transitivity of ϕ guarantees the existence of such a lift [28], and the following theorem provides an explicit form for a lift of the system studied in this paper.

Theorem 3.2. *Define the map $\Lambda : \mathcal{M} \times \mathbb{L} \to \mathfrak{g}$ by*

$$\Lambda\left(\xi, u\right) \coloneqq \left(\left(\Lambda_1\left(\xi, u\right), \Lambda_2\left(\xi, u\right)\right), \Lambda_3\left(\xi, u\right), \Lambda_4\left(\xi, u\right)\right),$$

where $\Lambda_1 : \mathcal{M} \times \mathbb{L} \to \mathfrak{se}_2(3)$, $\Lambda_2 : \mathcal{M} \times \mathbb{L} \to \mathfrak{se}(3)$, $\Lambda_3 : \mathcal{M} \times \mathbb{L} \to \mathfrak{se}(3)$, and $\Lambda_4 : \mathcal{M} \times \mathbb{L} \to \mathfrak{in}$ are given by

$$\Lambda_1\left(\xi, u\right) \coloneqq \left(\mathbf{W} - \mathbf{B} + \mathbf{D}\right) + \mathbf{T}^{-1}\left(\mathbf{G} - \mathbf{D}\right)\mathbf{T}, \tag{4a}$$

$$\Lambda_2\left(\xi, u\right) \coloneqq \left(\mathbf{ad}^\vee_{\boldsymbol{b}^\wedge}\left(\Pi\left(\Lambda_1\left(\xi, u\right)\right)^\vee\right) - \boldsymbol{\tau}\right)^\wedge, \tag{4b}$$

$$\Lambda_3\left(\xi, u\right) \coloneqq \left(\mathbf{Ad}^\vee_{\mathbf{S}^{-1}}\left(\Upsilon\left(\Lambda_1\left(\xi, u\right)\right)^\vee\right) + \boldsymbol{\mu}\right)^\wedge, \tag{4c}$$

$$\Lambda_4\left(\xi, u\right) \coloneqq \boldsymbol{\zeta}^\wedge, \tag{4d}$$

Then Λ is a lift for the system in Equ. (1) with respect to the symmetry group $\mathbf{G}$.

The existence of the lift allows the construction of a lifted system on the symmetry group [28]. Let $X \in \mathbf{G}$ be the state of the lifted system, and let $\mathring{\xi} = \left(\mathring{\mathbf{T}}, \mathring{\boldsymbol{b}}, \mathring{\mathbf{S}}, \mathring{\mathbf{K}}\right) \in \mathcal{M}$ be an arbitrarily chosen element of the original state in Equ. (1), called the origin. Then the lifted system is defined

$$\dot{X} = \mathrm{d}L_X \Lambda\left(\phi_{\mathring{\xi}}\left(X\right), u\right). \tag{5}$$

IV. Multi State Constraint Equivariant Filter

A. Filter state definition

Define $\hat{X} = \left(\left(\left(\hat{D}, \hat{\delta}\right), \hat{E}, \hat{L}\right), \hat{E}_1, \cdots, \hat{E}_k\right) \in \mathbf{G} \times \mathbf{SE}(3)^k$ to be the filter's state evolving on the symmetry group. Similarly to the original formulation [25] we maintain a sliding window of k past $\hat{E}$ elements in the state of the filter, corresponding to the different times a camera measurement was collected.

B. Error dynamics and state transition matrix

Let $e = \phi_{\hat{X}^{-1}}\left(\xi\right)$ denote the equivariant error. Normal coordinates [15] of the state space $\mathcal{M}$ in a neighborhood of the origin $\mathring{\xi}$ are $\boldsymbol{\varepsilon} = \vartheta\left(e\right) \coloneqq \log\left(\phi^{-1}_{\mathring{\xi}}\left(e\right)\right)^\vee \in \mathbb{R}^{25}$, where $\log : \mathbf{G} \to \mathfrak{g}$ is the logarithm of the symmetry group.

Recall the derivation of the linearized error dynamics in [15]

$$\begin{aligned}
&\dot{\varepsilon} \approx \mathbf{A}^0_t \varepsilon,\\
&\mathbf{A}^0_t = \mathrm{D}_e|_{\mathring{\xi}}\, \vartheta\left(e\right) \mathrm{D}_\xi|_{\hat{\xi}}\, \phi_{\hat{X}^{-1}}\left(\xi\right) \mathrm{D}_E|_I\, \phi_{\hat{\xi}}\left(E\right) \cdot\\
&\qquad \cdot \mathrm{D}_\xi|_{\phi_{\hat{X}}\left(\mathring{\xi}\right)} \Lambda\left(\xi, u\right) \mathrm{D}_e|_{\mathring{\xi}}\, \phi_{\hat{X}}\left(e\right) \mathrm{D}_\varepsilon|_{\mathbf{0}}\, \vartheta^{-1}\left(\varepsilon\right).
\end{aligned}$$

The state matrix $\mathbf{A}^0_t$ is given by

$$\mathbf{A}^0_t = \begin{bmatrix} {}_1\mathbf{A} & {}_2\mathbf{A} & \mathbf{0}_{9\times6} & \mathbf{0}_{9\times4}\\ {}_3\mathbf{A} & {}_4\mathbf{A} & \mathbf{0}_{6\times6} & \mathbf{0}_{6\times4}\\ {}_5\mathbf{A} & {}_6\mathbf{A} & {}_7\mathbf{A} & \mathbf{0}_{6\times4}\\ \mathbf{0}_{4\times9} & \mathbf{0}_{4\times6} & \mathbf{0}_{4\times6} & \mathbf{0}_{4\times4} \end{bmatrix} \in \mathbb{R}^{25\times25}, \tag{6}$$

where

$$\begin{aligned}
{}_1\mathbf{A} &= \begin{bmatrix} \multicolumn{1}{c}{\mathbf{\Psi} - \mathbf{ad}^\vee_{\mathring{\boldsymbol{b}}}} & \multicolumn{2}{c}{\mathbf{0}_{6\times3}}\\ \left(\mathring{\mathbf{R}}^T \mathring{\boldsymbol{v}}\right)^\wedge - \hat{b}^\wedge \mathring{\boldsymbol{b}}_\omega{}^\wedge & \mathbf{I}_3 & \mathbf{0}_{3\times3} \end{bmatrix} \in \mathbb{R}^{9\times9},\\
{}_2\mathbf{A} &= \begin{bmatrix} \mathbf{I}_3 & \mathbf{0}_{3\times3}\\ \mathbf{0}_{3\times3} & \mathbf{I}_3\\ \hat{b}^\wedge & \mathbf{0}_{3\times3} \end{bmatrix} \in \mathbb{R}^{9\times6},\\
{}_3\mathbf{A} &= \begin{bmatrix} \mathbf{ad}^\vee_{\mathring{\boldsymbol{b}}} \mathbf{\Psi} - \mathbf{ad}^\vee_{\left(\mathbf{Ad}^\vee_{\hat{B}}\ \boldsymbol{w} + \hat{\delta}^\vee + \boldsymbol{\theta}\right)} \mathbf{ad}^\vee_{\mathring{\boldsymbol{b}}} & \mathbf{0}_{6\times3} \end{bmatrix} \in \mathbb{R}^{6\times9},\\
{}_4\mathbf{A} &= \mathbf{ad}^\vee_{\left(\mathbf{Ad}^\vee_{\hat{B}}\ \boldsymbol{w} + \hat{\delta}^\vee + \boldsymbol{\theta}\right)} \in \mathbb{R}^{6\times6},\\
{}_5\mathbf{A} &= \mathbf{Ad}^\vee_{\mathring{\mathbf{S}}^{-1}} \begin{bmatrix} -\boldsymbol{\psi}_1^\wedge & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3}\\ -\boldsymbol{\psi}_3^\wedge - \mathring{\boldsymbol{b}}_\omega{}^\wedge \hat{b}^\wedge & \mathbf{I}_3 & -\boldsymbol{\psi}_2^\wedge \end{bmatrix} \in \mathbb{R}^{6\times9},\\
{}_6\mathbf{A} &= \mathbf{Ad}^\vee_{\mathring{\mathbf{S}}^{-1}} \begin{bmatrix} \mathbf{I}_3 & \mathbf{0}_{3\times3}\\ \hat{b}^\wedge & \mathbf{0}_{3\times3} \end{bmatrix} \in \mathbb{R}^{6\times6},\\
{}_7\mathbf{A} &= \mathbf{ad}^\vee_{\left(\mathbf{Ad}^\vee_{\mathring{\mathbf{S}}^{-1}}\ \boldsymbol{\varrho}\right)} \in \mathbb{R}^{6\times6},
\end{aligned}$$

with

$$\begin{aligned}
&\boldsymbol{\psi}_1 = \hat{A}\,\boldsymbol{\omega} + \delta^\vee_\omega \in \mathbb{R}^3, && \boldsymbol{\theta} = \left(\mathbf{0}_{3\times1}, g\left(\mathring{\mathbf{R}}^T \boldsymbol{e}_3\right)\right) \in \mathbb{R}^6,\\
&\boldsymbol{\psi}_2 = \boldsymbol{\psi}_1 - \mathring{\boldsymbol{b}}_\omega \in \mathbb{R}^3, && \mathbf{\Psi} = \begin{bmatrix} \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3}\\ g\left(\mathring{\mathbf{R}}^T \boldsymbol{e}_3\right)^\wedge & \mathbf{0}_{3\times3} \end{bmatrix} \in \mathbb{R}^{6\times6},\\
&\boldsymbol{\psi}_3 = \hat{a} - \boldsymbol{\psi}_1^\wedge \hat{b} \in \mathbb{R}^3, && \boldsymbol{\varrho} = \left(\boldsymbol{\psi}_2, \boldsymbol{\psi}_4\right) \in \mathbb{R}^3.\\
&\boldsymbol{\psi}_4 = \hat{a} + \mathring{\mathbf{R}}^T \mathring{\boldsymbol{v}} - \boldsymbol{\psi}_2^\wedge \hat{b} \in \mathbb{R}^3.
\end{aligned}$$

The discrete-time state transition matrix is defined by $\mathbf{\Phi} = \exp\left(\mathbf{A}^0_t \Delta T\right)$ for time steps ΔT.

C. Multi state constraint

Consider the measurement model in Equ. (2), applying the action of the symmetry group to the state space in Equ. (3) yields

$$h\left(\phi_X\left(\xi\right)\right) = \mathbf{K}L\pi_{Z_1}\left(E^{-1}\left(\mathbf{PS}\right)^{-1} * \boldsymbol{p}_f\right) \tag{7}$$

Recall the equivariant error $e = \phi_{\hat{X}^{-1}}\left(\xi\right) = \vartheta^{-1}\left(\varepsilon\right)$. Define $\tilde{y} = \varsigma\left(\boldsymbol{p}_f\right) - \varsigma\left(\hat{\boldsymbol{p}}_f\right)$, where $\varsigma\left(\cdot\right)$ represents the chosen feature parametrization. The true feature can then be written as $\boldsymbol{p}_f = \varsigma^{-1}\left(\varsigma\left(\hat{\boldsymbol{p}}_f\right) + \tilde{y}\right)$. Therefore, the measurement model in Equ. (2) can be linearized at $\varepsilon = \mathbf{0}$, and $\tilde{y} = \mathbf{0}$ as follows:

$$\begin{aligned} h\left(\xi, \boldsymbol{p}_f\right) &= h\left(\phi_{\hat{X}}\left(\vartheta^{-1}\left(\varepsilon\right)\right), \varsigma^{-1}\left(\varsigma\left(\hat{\boldsymbol{p}}_f\right) + \tilde{y}\right)\right) \\ &= h\left(\hat{\xi}, \hat{\boldsymbol{p}}_f\right) + \mathbf{C}_t\varepsilon + \mathbf{C}_t^f\tilde{y} + \cdots. \end{aligned} \tag{8}$$

Let us derive the $\mathbf{C}_t$, and $\mathbf{C}_t^f$ for the *anchored inverse depth* parametrization [29, 25] of the feature. Note that the matrix $\mathbf{C}_t^f$ can be computed for any desired parametrization.

Let ${}^{A}\mathbf{P}^{A}\mathbf{S}$ be the pose of the anchor, defined as the pose of the camera where the feature $\boldsymbol{p_f}$ has been first seen. Define the feature in the anchor frame as $\boldsymbol{a}_f = \left({}^{A}\mathbf{P}^{A}\mathbf{S}\right)^{-1} * \boldsymbol{p}_f$, with $\boldsymbol{a}_f = \left(a_{f_x}, a_{f_y}, a_{f_z}\right) \in \mathbb{R}^3$. The anchored inverse depth parametrization is written

$$\boldsymbol{z} = \varsigma\left(\boldsymbol{p}_f\right) = \left(\boldsymbol{z}_1, z_2\right) = \left(\left(\frac{a_{f_x}}{a_{f_z}}, \frac{a_{f_y}}{a_{f_z}}\right), \frac{1}{a_{f_z}}\right), \tag{9}$$

$$\boldsymbol{p}_f = \varsigma^{-1}\left(\boldsymbol{z}\right) = \left({}^{A}\mathbf{P}^{A}\mathbf{S}\right) * \begin{bmatrix} \frac{\boldsymbol{z}_1}{z_2} \\ \frac{1}{z_2} \end{bmatrix}. \tag{10}$$

Then the matrix $\mathbf{C}_t^f$ is written

$$\begin{aligned} \mathbf{C}_t^f\tilde{y} &= \mathring{\mathbf{K}}\hat{L}d_{\pi_{Z_1}}\Gamma\left(\left(\hat{\mathbf{P}}\hat{\mathbf{S}}\right)^{-1}{}^{A}\hat{\mathbf{P}}^{A}\hat{\mathbf{S}}\right)\frac{1}{\hat{z}_2}\begin{bmatrix} \mathbf{I}_2 & -\frac{\hat{\boldsymbol{z}}_1}{\hat{z}_2} \\ \mathbf{0}_{1\times 2} & -\frac{1}{\hat{z}_2} \end{bmatrix}\tilde{y} \\ &= \mathring{\mathbf{K}}\hat{L}d_{\pi_{Z_1}}\Gamma\left(\hat{E}^{-1A}\hat{E}\right)\frac{1}{\hat{z}_2}\begin{bmatrix} \mathbf{I}_2 & -\frac{\hat{\boldsymbol{z}}_1}{\hat{z}_2} \\ \mathbf{0}_{1\times 2} & -\frac{1}{\hat{z}_2} \end{bmatrix}\tilde{y}, \end{aligned} \tag{11}$$

where we have used $\hat{\xi} \coloneqq \phi_{\hat{X}}\left(\mathring{\xi}\right)$ to map between the estimated state in the homogeneous space $\hat{\xi}$, and the estimated state in the symmetry group $\hat{X}$. Therefore

$$\left(\hat{\mathbf{P}}\hat{\mathbf{S}}\right)^{-1}{}^{A}\hat{\mathbf{P}}^{A}\hat{\mathbf{S}} = \hat{E}^{-1}\mathring{\mathbf{S}}^{-1}\hat{C}\hat{C}^{-1}\mathring{\mathbf{P}}^{-1}\mathring{\mathbf{P}}^{A}\hat{C}^{A}\hat{C}^{-1}\mathring{\mathbf{S}}^{A}\hat{E} = \hat{E}^{-1A}\hat{E}.$$

According to [15], the $\mathbf{C}_t$ matrix is defined by

$$\begin{aligned} \mathbf{C}_t\varepsilon &= \mathrm{D}_\xi|_{\hat{\xi}}\, h\left(\xi\right)\,\mathrm{D}_e|_{\mathring{\xi}}\,\phi_{\hat{X}}\,\mathrm{D}_\varepsilon|_{\mathbf{0}}\,\vartheta^{-1}\left(\varepsilon\right)\left[\varepsilon\right] \\ &= \mathring{\mathbf{K}}\hat{L}d_{\pi_{Z_1}}\Gamma\left(\hat{E}^{-1}\right)\left[\left({}^{A}E\,\hat{\boldsymbol{a}}_f\right)^{\wedge} \quad -\mathbf{I}_3\right]\varepsilon_E - \\ &- \mathring{\mathbf{K}}\hat{L}d_{\pi_{Z_1}}\Gamma\left(\hat{E}^{-1}\right)\left[\left({}^{A}E\,\hat{\boldsymbol{a}}_f\right)^{\wedge} \quad -\mathbf{I}_3\right]\varepsilon_{{}^{A}E} + \\ &+ \mathring{\mathbf{K}}\Xi\left(\hat{L}\pi_{Z_1}\left(\hat{E}^{-1A}E\,\hat{\boldsymbol{a}}_f\right)\right)\varepsilon_L, \end{aligned} \tag{12}$$

where ε_E, and $\varepsilon_{{}^{A}E}$ represent respectively the error in normal coordinates for the element E of the symmetry group corresponding to the most recent pose and to the anchor pose, whereas ε_L represent the error in normal coordinates that is related to the camera intrinsics.

To compute the matrix $\mathbf{C}_t$ in Equ. (12), an estimate of the feature position in the anchor frame is required. To this end, when a feature has been seen from multiple views a linear-nonlinear least square problem can be solved [25, 24].

Finally, to remove the dependency of the features, and hence perform a filter update, we employ nullspace marginalization of the matrix $\mathbf{C}_t^f$ in Equ. (8), according to the original formulation [25].

V. Experiments

In this letter, we perform a series of experiments to evaluate the accuracy, consistency, and, more importantly, robustness of the proposed MSCEqF. We perform many experiments on real-world data to evaluate robustness to expected and unexpected errors in the camera extrinsic calibration. In all these experiments, we limit our comparison to filter-based MSCKF algorithms for VIO, and in particular, to the best available one we believe represents the state-of-the-art, that is OpenVINS [24]. For a fair comparison, we turned off OpenVINS's persistent features (SLAM features), and only compare against its pure MSCKF part. Furthermore, in all the experiments, OpenVINS's MSCKF parameters were specifically tuned, for each dataset, according to the authors' suggested parameters. In contrast, the proposed MSCEqF shares the same tuning parameters across all the experiments and datasets.

A. Robustness

Robustness is an important property of a modern filter-based visual-inertial odometry algorithm. It is the ability to function with significant yet known errors, as well as the ability to deal with unknown unknowns. In simpler terms, it refers to how well an algorithm performs under non-ideal conditions, such as imperfect tuning parameters, poor calibration, or unexpected changes in the sensor's extrinsic parameters during field operations.

To assess the robustness of the proposed MSCEqF and the MSCKF, we ran a series of experiments using widely-known dataset for evaluating VIO algorithms. Specifically, the Euroc dataset [30], the TUM-VI dataset [31], and the UZH-FPV dataset [32]. For each dataset, we selected two sequences and ran each estimator $6 \times 6 \times 6 = 216$ times (for a total number of runs of 2592). In these experiments, we intentionally initialized the filters with incorrect camera extrinsic parameters, introducing errors in six steps ranging from (15°, 0.05m), to (90°, 0.3m). For each error step, we ran the estimators with six different priors (initial covariance) accounting for initial calibration errors in the range of the six error steps. For each pair (prior, error) we run each estimator six times. Finally, for each individual run, we classified an estimator as converged or diverged based on a position error threshold.

Based on the results of the experiment in Fig. 1, we derive the following noteworthy observations. In absolute terms, there seems to be an upper limit of absolute error that, no matter the prior, makes the estimators diverge. Although this limit highly depends on the dataset, for each of the tested sequences, the proposed MSCEqF possesses a higher error limit, and hence improved robustness to known absolute error. In relative terms, the proposed MSCEqF seems to deal better with unknown errors since the line at which the estimator fails is straight and does not bend towards the left side as it appears to happen for the MSCKF. Encouraged by these results, we ran an additional experiment on the *V1_01_easy* sequence of the Euroc dataset, introducing new, smaller priors and errors to effectively evaluate whether the estimators are able to manage errors that are smaller in absolute terms but outside the prior covariance. Fig. 2 clearly shows that the MSCEqF is indeed a more robust filter, able to deal with unexpected errors. Finally, Fig. 3 shows the convergence of the camera extrinsic parameters for both filters evaluated on the Euroc *V1_01_easy*

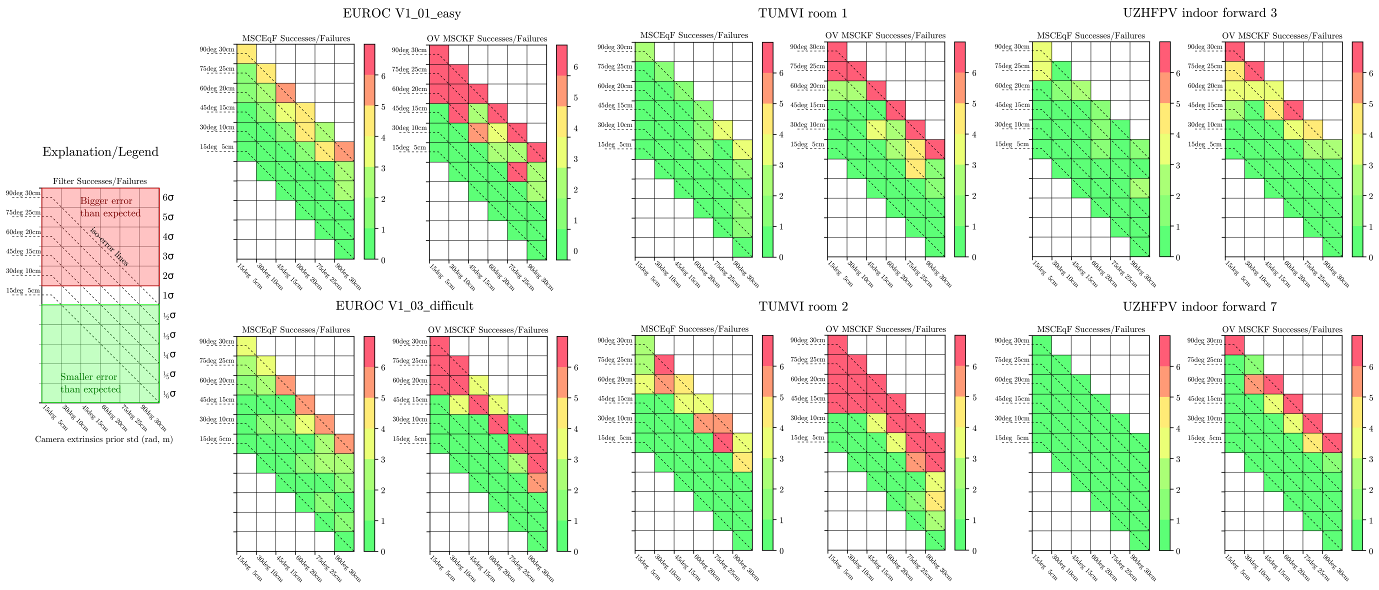

Figure 1. Results of the experiment evaluating the robustness of the proposed MSCEqF and OpenVINS's MSCKF. In these grid plots, the x-axis is the prior standard deviation the estimators are set with. The y-axis is how many σ-levels that error corresponds to. Labeled diagonal dashed lines represent iso-error lines (lines along with the error is constant). The bottom part of each grid represents expected errors, thus errors falling within $1/6\sigma$-$1/2\sigma$, whereas the top part of each grid represents unexpected errors, thus errors falling within 2σ-6σ. According to the colorbar, the color of each cell shows the number of failures.

sequence, with an initial error of (30°, $0.1\mathrm{m}$) and an initial covariance to match the error. The error plots clearly show that the proposed MSCEqF not only is a more robust filter, but it also converges faster.

Quantifying robustness in robotics, however, remains an ongoing challenge. In the presented evaluation, we have chosen the camera extrinsic calibration as a state subjected to error. Even though static and dynamic initialization approach exists [33, 34] for such a problem, in our formulation, extrinsic parameters are treated as regular state variables, and our proposed algorithm showcases inherent robustness by successfully attaining reliable estimation, for both expected and unexpected errors, eliminating the need of any auxiliary module. This characteristic sets our algorithm apart from conventional VIO algorithms, emphasizing its superior robustness.

B. Accuracy

Our next experiment focuses on the classical and widely-used metric for evaluating the performance of visual-inertial odometry algorithms [35], namely the RMSE of the absolute trajectory error (ATE). For this experiment, we ran the proposed MSCEqF and OpenVINS's MSCKF on all Euroc sequences [30]. The results presented in Tab. I demonstrate that the proposed MSCEqF achieves state-of-the-art accuracy comparable to the MSCKF. It should be noted that in our evaluation, we aligned each estimate with the groundtruth using the initial state rather than finding the optimal alignment that minimizes the error throughout the entire trajectory.

C. Consistency

An estimator is said to be consistent if the estimated covariance of the error reflects its real distribution; in other words, an estimator is consistent if the error is unbiased and within the sigma bounds of the estimated covariance. Consistency of the proposed MSCEqF is proven by compatibility of the group action ϕ in Equ. (3), and invariance of the lift Λ in Equ. (4), to reference frame transormations [20, 10]. This ensures that the filter does not gain spurious information along the unobservable directions.

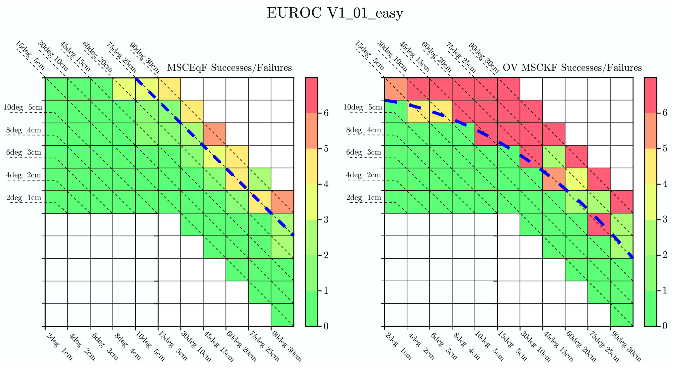

Figure 2. Grid plot showing the robustness of the proposed MSCEqF compared to OpenVINS's MSCKF for unexpected errors, thus the ability to deal with *you don't know what you don't know*. The x-axis is the prior standard deviation the estimators are set with. The y-axis is how many σ-levels that error corresponds to. Diagonal dashed lines represent iso-error lines. The blue bold dashed line is the limit at which each estimator fails. According to the colorbar, the color of each cell represents the number of failures.

Theorem 5.1. *Define* $H \coloneqq (R_H, 0, p_H) \in \mathbf{SE}_2(3)$, *where* $R_H \in \mathbf{SE}_{\boldsymbol{e}_3}(3)$ *represent a anti-clockwise rotation about the vertical axis* $\boldsymbol{e}_3$, *and* p_H *represent the a translation. Define the right group action* $\alpha : \mathbf{SE}_2(3) \times \mathcal{M} \to \mathcal{M}$ *such that* $\alpha(H, \xi) \coloneqq (H^{-1}\,\mathbf{T}, \boldsymbol{b}, \mathbf{S}, \mathbf{K})$ *represents a change of reference, from* $\{G\}$ *to* $\{H\}$ *that leaves the direction of gravity unchanged.*

Then the action of the symmetry group on the state space ϕ

Table I
ATTITUDE (A), AND POSITION (P) ABSOLUTE TRAJECTORY ERROR (ATE) RMSE ON EUROC DATASET

SEQUENCE	MSCEQF		OV MSCKF		SEQUENCE	MSCEQF		OV MSCKF		SEQUENCE	MSCEQF		OV MSCKF	
	A [rad]	P [m]	A [rad]	P [m]		A [rad]	P [m]	A [rad]	P [m]		A [rad]	P [m]	A [rad]	P [m]
V1_01_easy	0.07	**0.24**	**0.05**	0.36	V2_02_medium	0.08	0.55	**0.03**	**0.17**	MH_03_medium	0.02	**0.34**	**0.01**	0.41
V1_02_medium	0.03	**0.20**	**0.02**	0.22	V2_03_difficult[2]	**0.03**	0.39	**0.03**	**0.28**	MH_04_difficult	**0.03**	**0.53**	0.04	0.61
V1_03_difficult	0.05	0.30	**0.02**	**0.18**	MH_01_easy	**0.05**	**0.29**	**0.05**	0.42	MH_05_difficult	0.03	**0.70**	**0.02**	0.78
V2_01_easy	**0.02**	**0.13**	0.05	0.18	MH_02_easy	**0.01**	**0.38**	0.03	0.54					

and the lift Λ are respectively compatible and invariant with respect to change of reference, that is

$$\alpha(H, \phi(X, \xi)) = \phi(X, \alpha(H, \xi)),$$
$$\Lambda(\alpha(H, \xi), u) = \Lambda(\xi, u).$$

Proof.

$$\begin{aligned}\phi(X, \alpha(H, \xi)) &= \left((H^{-1}\,\mathbf{T})D, \mathbf{Ad}^{\vee}_{B^{-1}}\left(\boldsymbol{b} - \delta^{\vee}\right), C^{-1}\mathbf{S}E, \mathbf{K}L\right)\\ &= \left(H^{-1}\,\mathbf{T}D, \mathbf{Ad}^{\vee}_{B^{-1}}\left(\boldsymbol{b} - \delta^{\vee}\right), C^{-1}\mathbf{S}E, \mathbf{K}L\right)\\ &= \alpha(H, \phi(X, \xi)),\end{aligned}$$

as required.

To prove the invariance of Λ to the action α, it is sufficient to show that $\Lambda_1(\alpha(H, \xi), u) = \Lambda_1(\xi, u)$.

$$\begin{aligned}\Lambda_1(\alpha(H, \xi), u) &= (\mathbf{W} - \mathbf{B} + \mathbf{D}) + (\mathbf{T}^{-1}H)(\mathbf{G} - \mathbf{D})(H^{-1}\,\mathbf{T})\\ &= (\mathbf{W} - \mathbf{B} + \mathbf{D}) + \mathbf{T}^{-1}\left(H(\mathbf{G} - \mathbf{D})H^{-1}\right)\mathbf{T}\\ &= (\mathbf{W} - \mathbf{B} + \mathbf{D}) + \mathbf{T}^{-1}\left(H(\mathbf{G} - \mathbf{D})H^{-1}\right)\mathbf{T}\\ &= (\mathbf{W} - \mathbf{B} + \mathbf{D}) + \mathbf{T}^{-1}(\mathbf{G} - \mathbf{D})\,\mathbf{T}\\ &= \Lambda_1(\xi, u),\end{aligned}$$

where we have used the fact that $H(\mathbf{G} - \mathbf{D})H^{-1} = \mathbf{G} - \mathbf{D}$. Specifically

$$\begin{aligned}H(\mathbf{G} - \mathbf{D})H^{-1} &= \begin{bmatrix}\mathbf{0}_{3\times3} & R_H g\,\boldsymbol{e}_3 & \mathbf{0}_{3\times1}\\ \mathbf{0}_{1\times3} & 0 & -1\\ \mathbf{0}_{1\times3} & 0 & 0\end{bmatrix}\\ &= \begin{bmatrix}\mathbf{0}_{3\times3} & g\,\boldsymbol{e}_3 & \mathbf{0}_{3\times1}\\ \mathbf{0}_{1\times3} & 0 & -1\\ \mathbf{0}_{1\times3} & 0 & 0\end{bmatrix}\\ &= \mathbf{G} - \mathbf{D}.\end{aligned}$$

It is straightforward to see that $R_H g\,\boldsymbol{e}_3 = g\,\boldsymbol{e}_3$ since R_H is a rotation about the $\boldsymbol{e}_3$ axis. This completes the proof. □

In this final experiment, we employed the pose (orientation and position) average normalized estimation error squared (ANEES) as a metric to analyze the consistency of the proposed MSCEqF. In particular, we used the VINSEval framework [36] to generate a photorealistic synthetic dataset of 25 runs of the same trajectory, with the same noise statistics but different noise realizations.

The ANEES for the MSCEqF was computed according to the following formula

$$\text{ANEES} = \frac{1}{Mn}\sum_{i=1}^{M}\boldsymbol{\varepsilon}_i^T\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\varepsilon}_i,$$

where M is the number of runs, $n = dim\left(\boldsymbol{\varepsilon}\right)$ is the dimension of the error $\boldsymbol{\varepsilon}$, and $\boldsymbol{\Sigma}$ is the covariance of the error. The error $\boldsymbol{\varepsilon} = \log_{\mathbf{SE}(3)}\left(\mathring{\mathbf{P}}^{-1}\,\mathbf{P}\,\hat{\mathbf{P}}^{-1}\mathring{\mathbf{P}}\right)^{\vee}$ is the pose components of the equivariant error defined in Sec. IV-B.

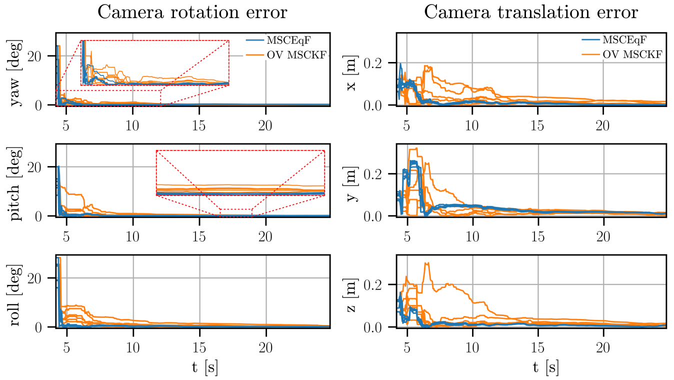

Figure 3. Absolute errors of camera extrinsic parameters for the proposed MSCEqF, and OpenVINS's MSCKF. The plots show the convergence performance of the filters evaluated on the Euroc *V1_01_easy* sequence, for 6 runs, with an initial error of (30°, 0.1m).

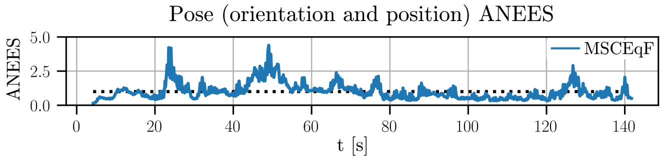

Figure 4. Pose (orientation and position) ANEES of the proposed MSCEqF for 25 runs on a custom dataset generated with the VINSEval framework.

The resulting ANEES shown in Fig. 4 fluctuates around a computed average of 1.0 and is not increasing or decreasing over time. This is a very similar average than FEJ esti-amtors [24, 37], but without requiring artificial modification of the linearization points to achieve consistency.

VI. CONCLUSION

This letter presented the *multi state constraint equivariant filter (MSCEqF)*. A novel equivariant filter formulation for the VIO problem, capable of camera intrinsic and extrinsic self-calibration. With our approach, we address the need for an VIO algorithm that achieves state-of-the-art accuracy and consistency while minimizing the need for sophisticated tuning and remaining robust against expected *and unexpected* errors. Through the presented experiments, we have demonstrated that the proposed MSCEqF successfully tackles these requirements. It exhibits robustness against both high absolute

[2] Due to non-deterministic results with varying in accuracy, we reported the best result out of 5 runs

errors and unexpected errors that exceed the prior covariance. Furthermore, the MSCEqF has been proven to be a naturally consistent estimator, achieving accuracy comparable to a state-of-the-art MSCKF algorithm but without the need for additional health-check nor consistency enforcing modules and heuristics. Future work includes the extension of the proposed MSCEqF with a polar symmetry for explicit SLAM features [20]

References

[1] J. A. Hesch, D. G. Kottas, S. L. Bowman, and S. I. Roumeliotis, "Consistency analysis and improvement of vision-aided inertial navigation," *IEEE Transactions on Robotics*, vol. 30, no. 1, pp. 158–176, 2014.

[2] G. P. Huang, A. I. Mourikis, and S. I. Roumeliotis, "Analysis and improvement of the consistency of extended Kalman filter based SLAM," *Proceedings - IEEE International Conference on Robotics and Automation*, pp. 473–479, 2008.

[3] ——, "A first-estimates Jacobian EKF for improving SLAM consistency," in *Experimental Robotics: The Eleventh International Symposium*. Springer, 2009, pp. 373–382.

[4] A. Barrau and S. Bonnabel, "The Invariant Extended Kalman Filter as a Stable Observer," *IEEE Transactions on Automatic Control*, vol. 62, no. 4, pp. 1797–1812, 2017.

[5] ——, "An EKF-SLAM algorithm with consistency properties," *arXiv preprint arXiv:1510.06263*, 2015.

[6] S. Heo and C. G. Park, "Consistent EKF-Based Visual-Inertial Odometry on Matrix Lie Group," *IEEE Sensors Journal*, vol. 18, no. 9, pp. 3780–3788, 5 2018.

[7] M. Brossard, S. Bonnabel, and A. Barrau, "Invariant Kalman Filtering for Visual Inertial SLAM," *2018 21st International Conference on Information Fusion, FUSION 2018*, pp. 2021–2028, 9 2018.

[8] ——, "Unscented Kalman Filter on Lie Groups for Visual Inertial Odometry," *IEEE International Conference on Intelligent Robots and Systems*, pp. 649–655, 12 2018.

[9] R. Hartley, M. Ghaffari, R. M. Eustice, and J. W. Grizzle, "Contact-aided invariant extended Kalman filtering for robot state estimation," *The International Journal of Robotics Research*, vol. 39, no. 4, pp. 402–430, 2020.

[10] K. Wu, T. Zhang, D. Su, S. Huang, and G. Dissanayake, "An invariant-EKF VINS algorithm for improving consistency," *IEEE International Conference on Intelligent Robots and Systems*, vol. 2017-September, pp. 1578–1585, 12 2017.

[11] C. Liu, C. Jiang, and H. Wang, "InGVIO: A Consistent Invariant Filter for Fast and High-Accuracy GNSS-Visual-Inertial Odometry," *IEEE Robotics and Automation Letters*, pp. 1–8, 2023.

[12] Y. Yang, C. Chen, W. Lee, and G. Huang, "Decoupled Right Invariant Error States for Consistent Visual-Inertial Navigation," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 1627–1634, 4 2022.

[13] A. Barrau and A. Barrau, "Non-linear state error based extended Kalman filters with applications to navigation," Ph.D. dissertation, Mines Paristech, 9 2015.

[14] P. Van Goor, T. Hamel, and R. Mahony, "Equivariant Filter (EqF): A General Filter Design for Systems on Homogeneous Spaces," *Proceedings of the IEEE Conference on Decision and Control*, vol. 2020-Decem, no. Cdc, pp. 5401–5408, 2020.

[15] P. van Goor, T. Hamel, and R. Mahony, "Equivariant Filter (EqF)," *IEEE Transactions on Automatic Control*, 6 2022.

[16] P. v. Goor, R. Mahony, T. Hamel, and J. Trumpf, "A Geometric Observer Design for Visual Localisation and Mapping," in *2019 IEEE 58th Conference on Decision and Control (CDC)*, 2019, pp. 2543–2549.

[17] P. van Goor, R. Mahony, T. Hamel, and J. Trumpf, "An Observer Design for Visual Simultaneous Localisation and Mapping with Output Equivariance," *arXiv preprint arXiv:2005.14347*, 2020.

[18] ——, "Constructive Observer Design for Visual Simultaneous Localisation and Mapping," *arXiv preprint arXiv:2006.05053*, 2020.

[19] P. van Goor and R. Mahony, "An Equivariant Filter for Visual Inertial Odometry," *Proceedings - IEEE International Conference on Robotics and Automation*, vol. 2021-May, pp. 1875–1881, 2021.

[20] ——, "EqVIO: An Equivariant Filter for Visual-Inertial Odometry," *IEEE Transactions on Robotics*, pp. 1–19, 2023.

[21] A. Fornasier, Y. Ng, R. Mahony, and S. Weiss, "Equivariant Filter Design for Inertial Navigation Systems with Input Measurement Biases," *2022 International Conference on Robotics and Automation (ICRA)*, pp. 4333–4339, 5 2022.

[22] A. Fornasier, Y. Ng, C. Brommer, C. Bohm, R. Mahony, and S. Weiss, "Overcoming Bias: Equivariant Filter Design for Biased Attitude Estimation With Online Calibration," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 12 118–12 125, 10 2022.

[23] A. Fornasier, Y. Ge, P. van Goor, R. Mahony, and S. Weiss, "Equivariant Symmetries for Inertial Navigation Systems," *arXiv preprint arXiv:2309.03765*, 9 2023.

[24] P. Geneva, K. Eckenhoff, W. Lee, Y. Yang, and G. Huang, "OpenVINS: A Research Platform for Visual-Inertial Estimation," *Proceedings - IEEE International Conference on Robotics and Automation*, pp. 4666–4672, 5 2020.

[25] A. I. Mourikis and S. I. Roumeliotis, "A multi-state constraint Kalman filter for vision-aided inertial navigation," *Proceedings - IEEE International Conference on Robotics and Automation*, pp. 3565–3572, 2007.

[26] R. Mahony, J. Trumpf, and T. Hamel, "Observers for kinematic systems with symmetry?" *IFAC Proceedings Volumes (IFAC-PapersOnline)*, vol. 9, no. PART 1, pp. 617–633, 2013.

[27] R. Hartley and A. Zisserman, "Multiple View Geometry in Computer Vision," *Multiple View Geometry in Computer Vision*, 3 2004.

[28] R. Mahony, T. Hamel, and J. Trumpf, "Equivariant sys-

tems theory and observer design," *arXiv*, 2020.

[29] J. Civera, A. J. Davison, and J. M. Montiel, "Inverse depth parametrization for monocular SLAM," *IEEE Transactions on Robotics*, vol. 24, no. 5, pp. 932–945, 2008.

[30] M. Burri, J. Nikolic, P. Gohl, T. Schneider, J. Rehder, S. Omari, M. W. Achtelik, and R. Siegwart, "The EuRoC micro aerial vehicle datasets," *https://doi.org/10.1177/0278364915620033*, vol. 35, no. 10, pp. 1157–1163, 1 2016.

[31] D. Schubert, T. Goll, N. Demmel, V. Usenko, J. Stuckler, and D. Cremers, "The TUM VI Benchmark for Evaluating Visual-Inertial Odometry," *IEEE International Conference on Intelligent Robots and Systems*, pp. 1680–1687, 12 2018.

[32] J. Delmerico, T. Cieslewski, H. Rebecq, M. Faessler, and D. Scaramuzza, "Are We Ready for Autonomous Drone Racing? The UZH-FPV Drone Racing Dataset," in *IEEE Int. Conf. Robot. Autom. (ICRA)*, 2019.

[33] T. C. Dong-Si and A. I. Mourikis, "Estimator initialization in vision-aided inertial navigation with unknown camera-IMU calibration," *IEEE International Conference on Intelligent Robots and Systems*, pp. 1064–1071, 2012.

[34] C. Campos, J. M. Montiel, and J. D. Tardos, "Fast and robust initialization for visual-inertial SLAM," *Proceedings - IEEE International Conference on Robotics and Automation*, vol. 2019-May, pp. 1288–1294, 5 2019.

[35] J. Delmerico and D. Scaramuzza, "A Benchmark Comparison of Monocular Visual-Inertial Odometry Algorithms for Flying Robots," *Proceedings - IEEE International Conference on Robotics and Automation*, pp. 2502–2509, 9 2018.

[36] A. Fornasier, M. Scheiber, A. Hardt-Stremayr, R. Jung, and S. Weiss, "VINSEval: Evaluation Framework for Unified Testing of Consistency and Robustness of Visual-Inertial Navigation System Algorithms," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 5 2021, pp. 13 754–13 760.

[37] C. Chen, Y. Yang, P. Geneva, and G. Huang, "FEJ2: A Consistent Visual-Inertial State Estimator Design," *Proceedings - IEEE International Conference on Robotics and Automation*, pp. 9506–9512, 2022.