Abstract—Visual-inertial odometry (VIO) is the problem of estimating a robot’s trajectory by combining information from an inertial measurement unit (IMU) and a camera and is of great interest to the robotics community. This article develops a novel Lie group symmetry for the VIO problem and applies the recently proposed equivariant filter. The proposed symmetry is compatible with the invariance of the VIO reference frame, leading to improved filter consistency. The bias-free IMU dynamics are group-affine, ensuring that filter linearization errors depend only on the bias estimation error and measurement noise. Furthermore, visual measurements are equivariant with respect to the symmetry, enabling the application of the higher order equivariant output approximation to reduce the approximation error in the filter update equation. As a result, the equivariant filter based on this Lie group is a consistent estimator for VIO with lower linearization error in the propagation of state dynamics and a higher order equivariant output approximation than standard formulations. Experimental results on the popular EuRoC and UZH FPV datasets demonstrate that the proposed system outperforms other state-of-the-art VIO algorithms in terms of both speed and accuracy.

Index Terms—Equivariant observer, simultaneous localization and mapping (SLAM), sensor fusion, visual-based navigation.

I. INTRODUCTION

VISUAL-INERTIAL odometry (VIO) is the problem of determining the trajectory of a robot from a combination of a camera and an inertial measurement unit (IMU). This problem is of enduring interest to the robotics community due to the ubiquity of systems where such sensors are available, including smart phones, virtual reality (VR) augmented reality headsets, racing drones, and more. Solutions to the “standard” variant of VIO, where only a single monocular camera is available, are of particular importance due to their wide range of applications. In addition, the IMU complements the visual data from a monocular camera by providing high-rate motion predictions and making the scale of the system observable, thereby overcoming a key weakness of the camera-only visual odometry (VO) problem.

State-of-the-art solutions for VIO are based on either the extended Kalman filter (EKF) or sliding-window optimization. EKF-based solutions, such as ROVIO [1], OpenVINS [2], and multistate constrained Kalman filter (MSCKF) [3], are generally less accurate than optimization-based methods but have lower compute and memory requirements and tend to be used in highly dynamic embedded systems applications such as VR headsets, smart phone applications, aerial vehicles, etc. On the other hand, optimization-based methods, such as VINS-mono [4] and OKVIS [5], tend to be more accurate than EKF-based methods but require significant compute and memory resources making them appropriate in applications such as automotive, larger robotic systems, etc. The main cause of loss of accuracy for EKF methods relative to optimization-based methods is associated with the accumulation of linearization errors. Recent advances in the theory of equivariant systems [6], [7] have shown that exploiting the Lie group symmetries of a system can lead to improved filter designs such as the invariant EKF (IEKF) and the equivariant filter (EqF) [8] that minimize the linearization error.

In this article, we develop a novel Lie group for the VIO problem and exploit this symmetry in the implementation of an equivariance-based VIO algorithm we term EqVIO. Unlike EKF designs, the EqF back end of our proposed system has out-of-the-box consistency properties, exact linearization of the bias-free IMU error dynamics, and a better (higher-order) linearization of the visual measurement function. The advantages of these properties are made clear in the experimental results, where EqVIO outperforms state-of-the-art EKF- and optimization-based algorithms in terms of both the accuracy of the estimated trajectory and the speed of processing each frame.

The key contributions of this article are as follows.

1) A novel Lie group, the VI-SLAM group, is developed for the VIO problem. This Lie group symmetry is compatible with the reference frame invariance of VIO. In addition, in contrast to the symmetries explored in prior literature [9], the visual measurement function of VIO is equivariant with respect to the VI-SLAM group.

2) The advantages of the VI-SLAM group are clearly demonstrated. Its $\text{SE}_2(3)$ component is shown to eliminate the linearization error in the bias-free IMU error dynamics; the only error in the propagation of the full IMU error dynamics is due to the measurement noise and bias estimation error. The novel landmark symmetry, based on $\text{SOT}(3)$ components, eliminates the second-order approximation error of the visual measurement function by exploiting...
equivariance. It also improves on and explains the well-known advantages of the inverse-depth parameterization of landmarks that is core to modern filter performance in VIO algorithms.

3) A novel VIO algorithm, EqVIO, is proposed that combines a simple feature-tracking front end and basic outlier rejection with an EqF implementation. EqVIO is shown to outperform state-of-the-art VIO algorithms in both speed and accuracy on both the popular EuRoC [10] and the challenging UZH FPV [11] datasets. Our implementation of EqVIO is open source and publicly available under a GNU GPLv3 license.¹

The proposed filter is in the class of invariant and equivariant observer designs. Analogous to previous IEKF solutions, it exploits the same $\mathbb{SE}(3)$ symmetry for the estimation of the IMU position, attitude, and velocity, and the provided estimate is statistically consistent. This is unsurprising since the EqF specializes to the IEKF when the system state can be identified with the symmetry Lie group and specific local coordinates are chosen [8, Appendix B]. However, the EqF framework used in this article enables the application of $\text{SOT}^+(3)$ as a symmetry to landmark estimation, which provides a powerful third-order approximation of the visual measurements, and cannot be applied in an IEKF setting.

This work is an extension of [12] and improves over the previous version by: providing online calibration of IMU–camera extrinsics; including the estimation of robot pose directly in the EqF rather than as a separate bundle adjustment step; detailing the effect of symmetry on the linearization of IMU error dynamics and visual measurements; and greatly expanding the experimental results to include more thorough comparisons with other state-of-the-art algorithms and demonstrate filter consistency.

II. RELATED WORK

A. Visual-Inertial Odometry

Although most VIO solutions rely on constructing a map of the robot’s local environment, the accuracy of this map is not considered important in evaluating system performance, in contrast to traditional simultaneous localization and mapping (SLAM). Some of the first systems to focus on the problem of trajectory estimation from stereo or monocular vision data, as distinct from general SLAM, were proposed in [13], [14], [15], and [16]. An important milestone in the development of VIO systems is the MSCKF [3], which approached the problem by applying a fixed-lag EKF and, notably, eliminating the estimation of landmark positions from the filter process. This modification resulted in an efficient algorithm for VIO with only linear complexity in the number of landmarks considered. Konolige et al. [17] considered a system equipped with a stereo camera and an IMU and employed bundle adjustment to solve the VIO problem. They improved their results by using specialized image features and discussed the challenge of using traditional image features in self-similar outdoor environments.

Since 2015, the monocular VIO problem specifically has seen substantial interest in the robotics community. Bloesch et al. [1] developed ROVIO: a VIO algorithm that mixes an arbitrary number of cameras with IMU measurements in an iterated EKF framework. In contrast to the majority of EKF-based VIO systems, ROVIO used a “direct” error formulation, that is, rather than obtaining feature coordinates from an image, the image pixel values were considered directly in the system model of the EKF. In parallel, Leutenegger et al. [5] developed OKVIS, which solves monocular and stereo VIO by using nonlinear optimization on a sliding window of “keyframes.” Semidirect visual odometry (SVO) [18] used a sparse set of image patches in frame-to-frame optimization to greatly reduce the presence of outliers and to solve VIO at very high speeds. While SVO is not strictly a VIO solution, as it does not use an IMU, Delmerico and Scaramuzza [19] proposed two methods to combine the output from SVO with an IMU in a filtering or smoothing framework. Recently, Qin et al. [4] have combined a range of modern SLAM techniques to develop VINS-MONO, which performs tightly coupled keyframe optimization-based VIO with efficient loop closure and achieves competitive accuracy on popular datasets. Delmerico and Scaramuzza [19] benchmarked a range of state-of-the-art VIO systems. They showed that, generally, the VIO systems optimized for speed and CPU usage suffered from relatively low accuracy, and that many of the popular systems tend to fail on challenging datasets or on limited hardware.

OpenVINS [2] is another recent VIO system, which mixes the MSCKF [3] with a traditional EKF and achieves state-of-the-art performance. The primary contribution of OpenVINS was to provide a well-documented open platform for EKF-based VIO research.

In summary, the literature on VIO is split between EKF- and optimization-based algorithms. EKF-based algorithms are preferred for their efficiency when computational resources are constrained, but suffer from linearization error that accumulates and degrades performance over time.

B. Equivariant Observers for VIO

Equivariant observers are state estimators that exploit available Lie group symmetries of a given problem. Two key examples include the IEKF [6] and the EqF [8]. The success of equivariant observers in other robotics problems has led several authors to investigate their application to inertial navigation, SLAM, and VIO. Barrau and Bonnabel [20] proposed the extended special Euclidean group $\mathbb{SE}(3)$ and show that it can be used to obtain an exact linearization of IMU error dynamics when the biases are known. This represents a clear improvement over the common representation of IMU states in the SLAM and VIO literature, which uses an on-manifold EKF [21] to obtain a minimal representation of rotation error between quaternions that is analogous to the well-known multiplicative EKF (MEKF) [22].

In the earliest work examining symmetry properties of the SLAM problem, Barrau and Bonnabel [23] proposed a novel class of Lie groups, $\mathbb{SE}_m$, and showed that this is a symmetry suitable for the classical SLAM problem. They further showed that this symmetry is compatible with the reference frame invariance of SLAM, and that the resulting IEKF consequently overcomes the well-known consistency issues of EKF-based solutions.

¹[Online]. Available: https://github.com/pvangoor/eqvio
SLAM [24], Zhang et al. [25] performed an observability analysis of the IEKF for SLAM and compared its performance to a range of other EKFs for SLAM in simulation. Wu et al. [26] then combined $\text{SE}_n(\hat{n})$ with the MSCKF concept of [3] to propose an invariant MSCKF for VIO, which they showed to be a consistent filter. They contrasted this to the original MSCKF, which exhibited growing inconsistency in a series of Monte Carlo simulation trials. Brossard et al. [27] derived an invariant unscented Kalman filter for monocular SLAM using the Lie group proposed in [23] and outperformed other invariant filters for VIO. A number of other works [9], [28], [29] have also explored applying variants of the IEKF to VIO in an MSCKF framework. Recently, Yang et al. [30] coupled the $\text{SE}_2(\hat{3})$ symmetry with the “first-estimates Jacobian” (FEJ) technique of [31]. They demonstrated improved accuracy and consistency over other filter-based algorithms implemented on the OpenVINS platform [2], evaluated in simulation and on the TUM-VI dataset [32].

Recently, van Goor et al. [33] have developed a novel Lie group for visual SLAM, under which the visual measurements of landmarks are equivariant, unlike in the previously explored $\text{SE}_n(\hat{n})$ symmetry. The IEKF cannot be directly applied using this symmetry as the Lie group is of a higher dimension than the underlying state space. This issue is overcome by the recent EqF [8], which additionally provides a framework for exploiting the equivariance of a system output function to reduce linearization error. To the best of our knowledge, there has been no equivariant observer applied to VIO with a symmetry that is compatible with visual measurements prior to this article and its previous version [12].

C. Parameterizations of VIO Landmarks

The representation of the robot pose and environment map is known to have a significant impact on the accuracy of EKF-based SLAM and VIO approaches. Castellanos et al. [34] identified the inconsistency of the EKF for SLAM when using a straightforward inertial-frame representation of landmarks. They proposed to use a Euclidean body-fixed representation of landmarks instead and showed that this improved the consistency of the EKF. Another key work in understanding the impact of landmark representations in EKF SLAM is by Civera et al. [35], who proposed the earliest version of the inverse-depth parameterization of landmarks. The key advantage in this representation is that it is able to represent a large uncertainty in the distance of a landmark from the robot’s initial position using a Gaussian distribution. Solà [36] investigated a variety of landmark parameterizations and proposed an anchored homogeneous point representation similar to the inverse-depth parameterization. He showed that this representation provided more consistent estimation in a SLAM system compared to inverse depth in a series of simulation experiments. However, further comparisons by Solà et al. [37] showed similar performance between the inverse-depth and anchored homogeneous point representations and concluded that the inverse-depth parameterization is preferred for its lower computational cost. Li and Mourikis [38] also investigated the anchored homogeneous point and inverse-depth representations, showing that the first yielded better filter consistency, while the second resulted in better absolute filter performance. A more recent version of the inverse-depth parameterization is presented by Bloesch et al. [1], who adapted the on-manifold EKF approach developed in [21] to obtain a minimal representation of unit vectors on the sphere. While the existing inverse-depth parameterizations have been empirically shown to improve performance over the Euclidean parameterization, the cause of this improvement is not well characterized. The discovery of symmetries for visual measurements [33] motivates the development of a new compatible parameterization.

D. Consistency of Filter-Based VIO

Stochastic filters for SLAM and VIO, such as the EKF, are said to be consistent if the probability distribution they report matches the true system statistics. The straightforward EKF solution for SLAM was shown to be inconsistent in experiments carried out in [34]. A number of authors have developed advanced modifications to the standard EKF to reduce inconsistency, and these techniques continue to be used in state-of-the-art VIO systems such as OpenVINS [2]. However, significant progress has recently been made by new VIO solutions exploiting Lie group symmetries in their designs to circumvent the consistency problem entirely. Huang et al. [31] demonstrated that the inconsistency of standard EKF SLAM is associated with a mismatch between the observability of the linearized error-state system and the observability of the true system; the true system has a 6-D unobservable subspace corresponding to transformations of the reference frame, while the linearized system does not. They proposed an FEJ, which overcomes this observability issue with only a minor loss of accuracy in the EKF. Li and Mourikis [38] applied the FEJ concept to an MSCKF design and showed that this improved consistency over a standard MSCKF. Hesch et al. [39] studied the observability of VIO specifically and identified the same consistency issues in standard EKF solutions. They developed an observability constrained (OC) EKF for VIO that directly enforces the unobservable directions of the system in the update step of the EKF and showed that this significantly reduces filter inconsistency. Barrau and Bonnabel [23] proposed a novel Lie group $\text{SE}_{n+1}(3)$ and showed that an IEKF design for SLAM with this symmetry provides a consistent estimator. They showed that the linearized system admits the same unobservable directions as the true system, due to the compatibility of the proposed Lie group and the reference frame invariance of SLAM. This same symmetry was used by Wu et al. [26] to develop an IEKF for VIO. This was also shown to provide a consistent filter for VIO, as the symmetry respects the invariance of the VIO problem to changes in the reference frame yaw and position. Recently, Huai and Huang [40] have formulated the VIO problem with respect to a nonglobal moving frame and showed that an algorithm based on this approach does not suffer from the observability mismatch between the true and linearized systems. Finally, Yang et al. [30] applied an IEKF to VIO using the FEJ technique and showed that it outperforms a standard FEJ-EKF design. Recent developments have studied the VIO problem from the perspective of Lie group symmetries;
they show that designing a filter that exploits these symmetries can overcome the observability mismatch of standard EKF designs and yield a consistent solution for VIO. It is clear that a key advantage of algorithms based on invariant and equivariant principles is strong consistency properties.

### III. MATHEMATICAL PRELIMINARIES

For a comprehensive introduction to smooth manifolds and Lie groups, the authors recommend [41].

#### A. Smooth Manifolds

Given a smooth manifold $\mathcal{M}$, denote the tangent space at $\xi \in \mathcal{M}$ by $T_\xi \mathcal{M}$. The tangent bundle of $\mathcal{M}$ is written $T\mathcal{M}$. If $f : \mathcal{M} \to \mathcal{N}$ is a differentiable function between smooth manifolds, the differential of $f$ with respect to $\zeta$ at a point $\xi \in \mathcal{M}$ is

$$D_\zeta f(\zeta) : T_\xi \mathcal{M} \to T_{f(\zeta)} \mathcal{N}$$

$$u \mapsto D_\zeta f(\zeta)[u].$$

When the base point is left unspecified, the differential of $f$ is a map between tangent bundles $Df : T\mathcal{M} \to T\mathcal{N}$.

**Lie Group Theory**

A right group action of a Lie group $G$ is defined to be $G \times \mathcal{N} \to \mathcal{N}$ and forms a Lie group under composition.

**B. Lie Group Theory**

For a Lie group $G$, we write the Lie algebra as $\mathfrak{g}$. The identity is denoted $id \in G$, and left and right translation is written

$$L_X(Y) := XY, \quad R_X(Y) := YX$$

respectively. The exponential map is written $\exp$, and its inverse (when defined) is the logarithmic map $\log$. The Adjoint maps $Ad : G \times \mathfrak{g} \to \mathfrak{g}$ and $ad : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$ are defined by

$$Ad_X U = D_LX DR_X^{-1} U, \quad ad_U V = [U, V]$$

where $[\cdot, \cdot]$ is the Lie bracket on $\mathfrak{g}$. The wedge and vee operators are linear isomorphisms

$$\wedge : \mathbb{R}^{\dim \mathfrak{g}} \to \mathfrak{g}, \quad \vee : \mathfrak{g} \to \mathbb{R}^{\dim \mathfrak{g}}$$

satisfying $(u^\vee)^\wedge = u$ for all $u \in \mathfrak{g}$. When it is not clear from context, we will use a subscript to indicate which Lie group or algebra a particular operation is associated with.

A right group action of a Lie group $G$ on smooth manifold $\mathcal{M}$ is a smooth map $\phi : G \times \mathcal{M} \to \mathcal{M}$ satisfying

$$\phi(XY, \xi) = \phi(Y, \phi(X, \xi))$$

$$\phi(id, \xi) = \xi$$

for all $X, Y \in G$ and $\xi \in \mathcal{M}$. A left group action is defined similarly, except that the compatibility condition (2) is reversed.

A product Lie group is formed from the combination of multiple existing Lie groups. If $G_1, \ldots, G_n$ are Lie groups, then the product Lie group is

$$G_1 \times \cdots \times G_n := \{(X_1, \ldots, X_n) \mid X_i \in G_i\}$$

with multiplication, identity, and inverse given by

$$(X_1, \ldots, X_n)(Y_1, \ldots, Y_n) := (X_1 Y_1, \ldots, X_n Y_n)$$

$$id_{G_1 \times \cdots \times G_n} := (id_{G_1}, \ldots, id_{G_n})$$

$$(X_1, \ldots, X_n)^{-1} := (X_1^{-1}, \ldots, X_n^{-1}).$$

#### C. Important Lie Groups

The special orthogonal group is the set of 3-D rotations

$$\text{SO}(3) := \{R \in \mathbb{R}^{3 \times 3} \mid R^T R = I_3, \det(R) = 1\}$$

$$\text{so}(3) := \{\omega^\times \mid \omega \in \mathbb{R}^3\}$$

$$\omega^\times := \begin{pmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{pmatrix}.$$  

Note that $\omega^\times$ is the unique $3 \times 3$ matrix satisfying $\omega^\times v = \omega \times v$ for all $v \in \mathbb{R}^3$. For any $v \in \mathbb{R}^3$, define $(\omega)^\wedge_{\text{so}(3)} := \omega^\times$.

The scaled orthogonal transforms are

$$\text{SOT}(3) := \left\{ \begin{pmatrix} R & 0 \\ 0 & c \end{pmatrix} \mid R \in \text{SO}(3), \ c > 0 \right\}$$

$$\text{SOT}(3) := \left\{ \begin{pmatrix} \Omega & 0 \\ 0 & s \end{pmatrix} \mid \Omega \in \mathbb{R}^3, \ s \in \mathbb{R} \right\}$$

Elements of SOT(3) may be written as $Q = (R_Q, c_Q)$, where $R_Q \in \text{SO}(3)$ and $c_Q > 0$. Given $Q \in \text{SOT}(3)$ and $p \in \mathbb{R}^3$, we use the shorthand

$$Qp := c_Q R_Q p.$$  

The special Euclidean group is the set of 3-D rigid body poses

$$\text{SE}(3) := \left\{ \begin{pmatrix} R & x \\ 0 & 1 \end{pmatrix} \mid R \in \text{SO}(3), \ x \in \mathbb{R}^3 \right\}$$

$$\text{se}(3) := \left\{ \begin{pmatrix} \Omega & 0 \\ 0 & v \end{pmatrix} \mid \Omega \in \mathbb{R}^3 \right\}$$

Elements of SE(3) may be written as $P = (R_P, x_P)$, where $R_P \in \text{SO}(3)$ and $x_P \in \mathbb{R}^3$. Given $P \in \text{SE}(3)$ and $p \in \mathbb{R}^3$, we frequently use the shorthand $Pp := R_P p + x_P$ to denote the standard left group action of SE(3) on $\mathbb{R}^3$. 

---

2This shorthand notation corresponds to the normal homogeneous coordinate notation.
The extended special Euclidean group \([20]\) is

\[
SE_2(3) := \left\{ \begin{pmatrix} R & x & v \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \mid R \in SO(3), \; x, v \in \mathbb{R}^3 \right\}
\]

\[
se_2(3) := \left\{ (\Omega, v, a)_{se_2(3)} \mid \Omega, v, a \in \mathbb{R}^3 \right\}
\]

\[
(\Omega, v, a)_{se_2(3)} := \begin{pmatrix} \Omega & v & a \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.
\]

Elements of \(SE_2(3)\) may be written as \(A = (R_A, x_A, v_A)\), or as \(A = (P_A, v_A)\), where \(R_A \in SO(3), x_A, v_A \in \mathbb{R}^3\), and \(P_A = (R_A, x_A) \in SE(3)\).

### IV. PROBLEM DESCRIPTION

Choose an arbitrary inertial reference frame \(\{0\}\) and consider a robot equipped with an IMU and a camera, both of which are rigidly attached to the robot’s body. Label the camera frame \(\{C\}\), and identify the IMU frame \(\{T\}\) with the body-fixed frame of the robot \(\{B\}\). The states of the visual-inertial SLAM (VI-SLAM) problem are modeled as follows.

1. \(P_B := (R_B, x_B) \in SE(3)\) is the pose of the IMU \(\{B\}\) with respect to the inertial frame \(\{0\}\).
2. \(v_B \in \mathbb{R}^3\) is the linear velocity of the IMU expressed in the inertial frame \(\{0\}\).
3. \(b_B = (b^T_B, b^g_B) \in \mathbb{R}^6\) is the IMU bias (where \(b^T_B\) and \(b^g_B\) are the gyroscope and accelerometer biases, respectively).
4. \(T = (R_T, x_T) \in SE(3)\) is the pose of the camera \(\{C\}\) with respect to the body-fixed frame \(\{B\}\).
5. \(p_i \in \mathbb{R}^3\) are the coordinates of landmark \(i\) in the inertial frame \(\{0\}\).

Fig. 1 shows a diagram of the full VI-SLAM configuration. Let \(M^3 := SE(3) \times \mathbb{R}^3\) denote the navigation state space and define the navigation state \(\xi_B := (P_B, v_B) \in M^3\). We frequently use the notation \(\xi := (\xi_B, b_B, T, p_i)\) as shorthand for the full VI-SLAM state. To ensure that the visual measurements are always well defined, we assume that the system trajectory considered never passes through an exception set

\[
\mathcal{E} := \{(\xi_B, b_B, T, p_i) \in M^3 \times SE(3) \times (\mathbb{R}^3)^n \mid (P_B T)^{-1} p_i = 0 \text{ for any } i\}
\]

corresponding to all situations where the camera center coincides with a landmark point. To formalize this, we define the VI-SLAM total space

\[
\mathcal{T}^{VI}_n := M^3 \times SE(3) \times (\mathbb{R}^3)^n - \mathcal{E}
\]

and consider the VI-SLAM problem on \(\mathcal{T}^{VI}_n\). Note that \(\mathcal{T}^{VI}_n\) is an open subset of a smooth manifold and as such is itself a smooth manifold.

### A. VI-SLAM Dynamics

Let the acceleration due to gravity in the inertial frame \(\{0\}\) be \(g e_3\), where \(g \approx 9.81 \text{ m/s}^2\) and \(e_3 \in S^2\) is the standard direction of gravity in the inertial frame. Let the measured angular velocity and linear acceleration obtained from the IMU be \((\Omega, a) \in L := \mathbb{R}^3 \times \mathbb{R}^3\), where \(L\) is the input space. Then, the VI-SLAM dynamics \(f : L \to \mathcal{T}^{VI}_n\) are

\[
\dot{\xi} = f(\Omega, a)(\xi); \quad \dot{\eta}_B = R_B(\Omega - b_B^\Omega) \\
\dot{x}_B = v_B \\
\dot{v}_B = R_B(a - b_B^a) + g e_3 \\
\dot{b}_B^\Omega = 0 \\
\dot{b}_B^a = 0 \\
\dot{T} = 0 \\
\dot{p}_i = 0.
\]

### B. VI-SLAM Measurements

The camera measurements are modeled as \(n\) bearing measurements of the landmarks \(p_i\) in the camera frame \(\{C\}\) on the manifold \(\mathcal{N}^n_3 := (S^2)^n\), where the superscript “V” stands for visual measurements. The measurement function \(h : \mathcal{T}^{VI}_n \to \mathcal{N}^n_3\) is given by

\[
h(\xi) := (h^1(\xi), \ldots, h^n(\xi))
\]

\[
h^k(\xi_B, b_B, T, p_i) := \pi_{S^2}((P_B T)^{-1}(p_i))
\]

where \(\pi_{S^2}\) is defined as in (1). Modeling the bearing measurements directly on the sphere rather than the image plane enables the proposed system to model a wide variety of monocular cameras. The consideration of different camera models is omitted from this article but is included in the EqVIO software package.

### C. Invariance of VI-SLAM

Let \(e_3\) be the standard gravity direction and define the semidirect product group

\[
S^1 \ltimes_{e_3} \mathbb{R}^3 := \{(\theta, x) \mid \theta \in S^1, x \in \mathbb{R}^3\}
\]

with group product, identity, and inverse

\[
(\theta^1, x^1) \cdot (\theta^2, x^2) = (\theta^1 + \theta^2, x^1 + R_{e_3}(\theta^1)x^2)
\]

\[
\text{id}_{S^1 \ltimes_{e_3} \mathbb{R}^3} = (0, 0_{3 \times 1})
\]

\[
(\theta, x)^{-1} = (-\theta, -R_{e_3}(-\theta)x)
\]
where \( R_{e_3}(\theta) \in \text{SO}(3) \) is the anticlockwise rotation of an angle \( \theta \) about the axis \( e_3 \). Then, \( S^1 \times_{e_3} \mathbb{R}^3 \) is isomorphic to the subgroup

\[
\text{SE}_{e_3}(3) := \{(R, x) \in \text{SE}(3) \mid R e_3 = e_3\} \subset \text{SE}(3).
\]

Define \( \alpha : \text{SE}_{e_3} \times \mathcal{T}^n \to \mathcal{T}^n \) by

\[
\alpha(S, (\xi_B, b_B, T, p_i)) := (S^{-1} P_B, R_S^T v_B, b_B, T, S^{-1}(p_i)).
\]

(7)

Then, \( \alpha \) is a right group action of \( \text{SE}_{e_3} \) on \( \mathcal{T}^n \). For a given \( S \in \text{SE}_{e_3} \), the action \( \alpha(S, \cdot) \) represents a change of inertial reference frame from \{0\} to \{1\}, where \( S \) is the pose of \{1\} with respect to \{0\}. Moreover, any change of reference \( S \in \text{SE}_{e_3} \) leaves the direction of gravity \( e_3 \) unchanged.

**Proposition IV.1:** The dynamics (5) and measurements (6) of VI-SLAM are invariant with respect to \( \alpha \), i.e.,

\[
f_{(\Omega_a, \delta)}(\alpha(S, (\xi_B, b_B, T, p_i))) = d\alpha_S f_{(\Omega_a, \delta)}((\xi_B, b_B, T, p_i))
\]

\[h(\alpha(S, (\xi_B, b_B, T, p_i))) = h((\xi_B, b_B, T, p_i))
\]

for any \( S \in \text{SE}_{e_3} \).

V. SYMMETRY OF VI-SLAM

The VI-SLAM symmetry action proposed in this article combines the symmetry for IMU dynamics developed in [20] with the VSLAM symmetry developed in [33]. Before discussing the full symmetry group, we discuss the separate symmetries and how they lead to lower linearization error in the filter development given in the following. This section of this article is written in a more tutorial style to provide the reader with intuition underlying the proposed algorithm.

The advantage of the extended Euclidean symmetry (\( \text{SE}_e(3) \)) used for the navigation states lies in providing a locally linear coordinate representation of the ideal IMU dynamics. That is, for ideal IMU dynamics, then, using this representation leads to zero linearization error during the propagation step of the filter, assuming appropriate Gaussian noise models in local coordinates of course. This property is lost once the bias states and calibration states are added; however, the resulting update still has considerably lower linearization error in these coordinates than classical formulations. Section V-A is based on prior work by Barrau and Bonnabel [20], and the results presented therein can equally be applied in an IEKF.

The advantage of the scaled orthogonal transform (\( \text{SOT}(3) \)) symmetry used for the landmark states lies in providing a framework in which the measurement linearization error can be minimized. To make this point clear, we analyze the common landmark parameterizations used in the literature and study the linearization error. This provides a clear theoretical justification for the inverse-depth parameterization that is state of the art in VIO algorithms and goes on to demonstrate that the \( \text{SOT}(3) \) symmetry leads to lower measurement linearization error again. The material in Section V-B is novel to this article.

A. Symmetry of IMU Dynamics

Let \( f^i : \mathbb{L} \to \mathcal{X}(M^i) \) denote the IMU dynamics considered in (5) without bias, i.e.,

\[
\frac{d}{dt}(R_B, x_B, v_B) = f_{(\Omega_a)}(R_B, x_B, v_B)
\]

(8)

Filter designs, such as the EKF, on-manifold EKF [21, MEKF [3], and EqF [8], model the evolution of the probability distribution of the system state, given an initial distribution. In each of these filters, the filter’s error coordinates are taken to be normally distributed, and the dynamics of these error coordinates are linearized to propagate the estimated distribution.

Naive EKF designs model the robot’s attitude through embedded coordinates as a unit quaternion \( q \in \mathbb{H} \cong \mathbb{R}^4 \) [42, 43, 44, 45]. This leads to the (overparameterized) error coordinates

\[
\epsilon_{\text{EKF}}((\xi_B, \xi_B)) := \begin{pmatrix} q_B - \hat{q}_B \\ x_B - x_B \xi_B \\ v_B - v_B \xi_B \end{pmatrix}
\]

In the MEKF and the typical on-manifold EKF, the error coordinates are instead defined using the logarithm of \( \text{SO}(3) \)

\[
\epsilon_{\text{MEKF}}((\xi_B, \xi_B)) := \begin{pmatrix} \log_{\text{SO}(3)}(R_B R_B^\top)^\top \\ x_B - x_B \xi_B \\ v_B - v_B \xi_B \end{pmatrix}
\]

Note that, unlike the EKF, the MEKF has a minimal (9-D) representation of filter error.

In an EqF design, the filter state is part of a Lie group rather than the state space \( M^i \) [8]. Consider the Lie group \( \text{SE}_2(3) \) with right group action \( \varphi : \text{SE}_2(3) \times M^i \to M^i \) given by

\[
\varphi(A, \xi_B) := (R_B R_A, x_B + R_B x_A, v_B + R_B v_A)
\]

(10)

where \( A = (R_A, x_A, v_A) \in \text{SE}_2(3) \). In this case, \( \text{SE}_2(3) \) and \( M^i \) are isomorphic as smooth manifolds, but the distinction is important for systems on general homogeneous spaces, such as the full VI-SLAM system, where the Lie group may be of a higher dimension than the state space. Choose the origin configuration \( \xi_B = (I_3, 0, 0) \in M^i \), and define a local coordinate chart \( \vartheta_B((\xi_B)) := \log_{\text{SE}_2(3)}(R_B, x_B, v_B) \in \mathbb{R}^9 \). Then, the EqF error coordinates are

\[
\epsilon_{\text{EqF}}(\hat{\xi}_B, \xi_B) := \vartheta(\varphi(\hat{\xi}_B^{-1}, \xi_B))
\]

\[
= \begin{pmatrix} \log_{\text{SE}_2(3)}(R_B R_A^\top) x_B - R_B R_A^\top x_A \\ v_B - R_B R_A^\top v_A \\ 0 \\ 0 \end{pmatrix}
\]

(11)

where \( \hat{A} = (R_A, x_A, v_A) \in \text{SE}_2(3) \) is the filter state.

Each of these filters model their error coordinates as being drawn from a normal distribution with zero mean and covariance \( \Sigma, \varepsilon \sim N(0, \Sigma) \), where \( \Sigma \) is the filter’s Riccati matrix. The reported probability distribution of these filters is not guaranteed...
to match the true distribution of the state in general, as the propagation of the covariance depends on the linearization of the error dynamics. For the EqF, however, this linearization is dependent on the chosen symmetry group $G$, rather than on a chosen set of coordinates as in the EKF. In some special cases, the system dynamics are group affine with respect to the symmetry group $G$, and this results in an exactly linear propagation of the error coordinates [6].

Barrau and Bonnabel [20] showed that the bias-free IMU dynamics are group affine with respect to the action (10) of $\mathrm{SE}_2(3)$. As a result, using the Lie group $\mathrm{SE}_2(3)$, the propagation of bias-free IMU dynamics in the EqF has no linearization error in normal coordinates. Hence, as long as the initial distribution of the error coordinates is Gaussian, the probability distribution estimated by the EqF will match the true distribution of the state exactly. This is demonstrated in the following example.

Let the initial value of the true state $q_0(0) = \exp_{\mathrm{SE}_2(3)}(\eta_0)$, where $\eta_0 \in \mathbb{R}^9$ is drawn from the distribution $N(0, \Sigma_0)$ with

$$\Sigma_0 = \begin{pmatrix} 0.2^2 I_3 & 0 & 0 \\ 0 & 0.01^2 I_3 & 0 \\ 0 & 0 & 0.01^2 I_3 \end{pmatrix}.$$  

Initialize each of the filters with this data, and let the angular velocity be $\Omega = (0, 0, 0.1)$ rad/s and the linear acceleration be $a = (0.1, 0.0, 0.0)$ m/s².  

Fig. 2 shows the true and estimated distributions of the IMU position after integrating in 5 s increments. At each time, the true distribution of the system is shown by sampling 2000 particles according to the initial distribution of the state and integrating them independently. The estimated distributions are obtained by integrating the filter equations in 5 s increments, before sampling 2000 points in the error coordinates and mapping them to the estimated state using the filter’s observer state. The figure clearly shows the advantages of applying an appropriate symmetry to the propagation of IMU uncertainty, as the distribution reported by the EqF matches the true distribution far more closely than that of the EKF or MEKF.

**Remark V.1:** The inclusion of biases in the gyroscope and accelerometer measurements means that the full IMU dynamics (5) are not group affine with respect to the action (10) [27]. However, the linearization error introduced by the bias states is proportional to the error in the bias estimates, which is often small and can be reduced quickly by using an initialization maneuver in practice. Recent work by Fornasier et al. [46], [47] shows that alternative symmetries exist for including input biases in an EqF, but their analysis is beyond the scope of this article and left for future work.

An EKF is not an optimal estimator, unlike the linear Kalman filter, due to the accumulation of linearization errors over time. By exploiting symmetry properties as above, the linearization error in each step of the EKF can be reduced or even eliminated completely. This leads to improved filter designs that closely reflect the stochastics of the underlying system and provide more accurate state estimates.

### B. Symmetry of Visual Landmarks

Consider the simplified system of a single landmark $q \in \mathbb{R}^3 \setminus \{0\}$ in the camera-fixed frame. If the camera-fixed angular and linear velocity are $\Omega_C, v_C \in \mathbb{R}^3$, respectively, then the dynamics of the landmark are

$$\dot{q} = f^\mathcal{N}_{(\Omega_C, v_C)}(q) := -\Omega_C \times q - v_C. \quad (12)$$

The visual measurement of the landmark is

$$h^\mathcal{N}(q) = \frac{q}{|q|}. \quad (13)$$

A parameterization of the landmark is a diffeomorphism $\zeta : U \subset \mathbb{R}^3 \rightarrow \mathcal{M}^N \subset \mathbb{R}^k$, where $U$ is an open subset of $\mathbb{R}^3$ and $\mathcal{M}^N$ is a smooth 3-D submanifold of $\mathbb{R}^k$ for some $k \geq 3$.

The Euclidean parameterization was commonly used in early works on visual SLAM [24] and is defined by

$$\zeta_{\text{Euc}}(q) := q, \quad \zeta_{\text{Euc}}^{-1}(z) := z. \quad (14)$$

The inverse-depth parameterization and variants thereof are used more frequently in recent literature [1], [2], [35]. Here, we consider the archetypical example given by

$$\zeta_{\text{ID}}(q) := \begin{pmatrix} \arccos\left(\frac{z_1}{|q|}\right) \\ \frac{\tan(\zeta_2, q_3)}{|q|} \end{pmatrix}, \quad \zeta_{\text{ID}}^{-1}(z) := \frac{1}{z_3} \begin{pmatrix} \cos(z_1) \\ \sin(z_2) \sin(z_1) \\ \cos(z_2) \sin(z_1) \end{pmatrix}. \quad (15)$$

We introduce a new parameterization for landmarks with visual measurements by exploiting the $\text{SOT}(3)$ symmetry actions.
developed in [33]. The polar parameterization is a novel parameterization for visual landmarks, defined by
\[
\varsigma_{\text{SOT}(3)}(q) := \left( \begin{array}{c} \arccos\left( \frac{\mathbf{e}_3 \cdot q}{|q|} \right) \\
\arccos\left( \frac{\mathbf{e}_3 \cdot \mathbf{e}_1}{|\mathbf{e}_3 \times \mathbf{e}_1|} \right) \\
-\log(|q|) \end{array} \right)
\]
and about \( q \in \mathcal{Q}_1 \) [33]
\[
\varsigma_{-1}\varsigma_{\text{SOT}(3)}(z) := \exp_{\text{SOT}(3)} \left( \left( \begin{array}{ccc} 1 & z_2 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1 \end{array} \right) \right)^\vee \varsigma_{\text{SOT}(3)}^{-1} \mathbf{e}_3.
\]
(16)

This parameterization provides normal coordinates for \( \mathbb{R}^3 \) about \( \mathbf{e}_3 \) with respect to the right action of \( \text{SOT}(3) \) defined in Lemma V.2.

**Lemma V.2.** Let \( \varphi^V : \text{SOT}(3) \times \mathbb{R}^3 \setminus \{0\} \to \mathbb{R}^3 \setminus \{0\} \) and \( \rho^V : \text{SOT}(3) \times S^2 \to S^2 \) be defined by
\[
\varphi^V(Q, q) := c_Q^{-1} R_Q^\top q \\
\rho^V(Q, y) := R_Q^\top y.
\]
(17)

Then, \( \varphi^V \) and \( \rho^V \) are transitive right group actions, and the visual measurement function (13) is equivariant with respect to these actions, i.e.,
\[
h^V(\varphi^V(Q, q), q) = \varphi^V(Q, h^V(q))
\]
for all \( Q \in \text{SOT}(3) \) and \( q \in \mathbb{R}^3 \setminus \{0\} \).

Let \( \hat{q} \in \mathbb{R}^3 \setminus \{0\} \) be the estimated landmark position and \( q \in \mathbb{R}^3 \setminus \{0\} \) be the true landmark position. Then, the parameterized state error is
\[
\varepsilon := \varsigma(q) - \varsigma(\hat{q})
\]
for a particular parameterization \( \varsigma \). The true state can be identified in terms of the estimated state and the parameterized state error
\[
q = \varsigma^{-1}(\varepsilon + \varsigma(\hat{q})).
\]

Linearizing the dynamics (12) and measurement (13) in terms of \( \varepsilon \) as \( \varepsilon = 0 \) yields
\[
\begin{align*}
\dot{f}^V_{(\Omega, v)}(q) &= f^V_{(\Omega, v)} \circ \varsigma^{-1}(\varepsilon + \varsigma(\hat{q})) \\
&= f^V_{(\Omega, v)}(\hat{q}) + D_{q} f^V_{(\Omega, v)}(\hat{q}) \cdot D_s |_{\varsigma(\hat{q})} \varsigma(s) |_{\varepsilon} + \mu^f(q) \\
h^V(q) &= h^V \circ \varsigma^{-1}(\varepsilon + \varsigma(\hat{q})) \\
&= h^V(\hat{q}) + D_{p} h^V(\hat{q}) \cdot D_s |_{\varsigma(\hat{q})} \varsigma(s) |_{\varepsilon} + \mu^h(q)
\end{align*}
\]
(19)
where \( \mu^f(q) \) and \( \mu^h(q) \) are the dynamics and measurement linearization errors, respectively, and capture all higher order terms. In general, both \( \mu^f(q) \) and \( \mu^h(q) \) are \( O(|\varepsilon|^2) \), but there are important exceptions. First, the dynamics (12) are exactly linear in the Euclidean parameterization (14), and hence, \( \mu^f_{\text{base}}(q) \equiv 0 \). Second, the equivariant output approximation proposed in [8] is available when using the polar parameterization as these are the normal coordinates associated with the action \( \varphi^V \) (17).

Specifically, applying [8, Lemma V.3] to this example yields
\[
\begin{align*}
C^*_\varepsilon &= \frac{1}{2} \left( D_{E |_{\text{id}}}^V \rho^V(E, y) + D_{E |_{\text{id}}}^V \rho^V(E, \mathbf{e}_3) \right) |_{\varsigma_{\text{SOT}(3)}}^\vee
\end{align*}
\]
and therefore
\[
C^*_\varepsilon = \frac{1}{2} \left( y + \mathbf{e}_3 \right)^\times \Omega_e.
\]
(20)

This provides an alternative measurement approximation that reduces the output linearization error to \( \mu^h_{\text{SOT}(3)}(q) = O(|\varepsilon|^3) \) by exploiting equivariance and using the available system measurement value \( y = h(q) \).

To see the effects of different parameterizations on the linearization errors, let \( \hat{q} = (0, 0, 5) \) and consider the domain
\[
U = \left\{ (z \tan(\theta), 0, z) \in \mathbb{R}^3 \mid z \in (0.1, 10), \theta \in \left(-\frac{\pi}{6}, \frac{\pi}{6}\right) \right\}.
\]
(20)

Let \( \Omega_C = (0.0, 0.2, 0.0) \) and \( v_C = (0.0, 0.0, 0.1) \). This linear and angular velocity is typical of a camera moving forward and turning about the vertical axis of the image.

Figs. 3 and 4 show the dynamics and output linearization errors \( \mu^f(q) \) and \( \mu^h(q) \), respectively, for each of the parameterization (14)-(16). Fig. 4 also shows the equivariant output approximation [8] in the polar parameterization. As expected, the Euclidean parameterization yields no dynamics linearization error over any part of the domain. Meanwhile, the inverse-depth and polar parameterizations have nonzero dynamics linearization errors that increase toward the edges of the domain, with the polar parameterization having a lower maximum linearization error. On the other hand, for the output linearization, there is a clear advantage to using the inverse depth or polar over the Euclidean parameterization. Finally, while the ordinary performance of the inverse-depth and polar parameterizations is similar, the polar parameterization is able to use the equivariant output approximation, which greatly reduces its output linearization error everywhere in \( U \).

**C. Composite Symmetry for VI-SLAM**

Define the VI-SLAM group to be
\[
\text{SLAM}^m(3) := \mathbf{S}E_2(3) \times \text{SOT}(3)^m.
\]
(21)

This is a product Lie group in the sense of (4). This group captures the symmetries that are fundamental to the VIO system: \( \mathbf{S}E_2(3) \) for the IMU dynamics and \( \text{SOT}(3)^m \) for \( n \) landmarks with visual measurements. In a practical implementation, however, it is also necessary to consider the extrinsic calibration parameters and IMU biases, for which an \( \mathbf{S}E(3) \) and an \( \mathbb{R}^6 \) symmetry can be used, respectively. To this end, define the
Fig. 3. Norm of linearization error (18) of the landmark dynamics (12) for the Euclidean (14), inverse-depth (15), and polar (16) parameterizations over the domain defined in (20). The Euclidean parameterization has zero dynamics linearization error since the dynamics are exactly linear in these coordinates.

Fig. 4. Norm of linearization error (19) of the visual landmark measurement (13) for the Euclidean (14), inverse-depth (15), and polar (16) parameterizations over the domain defined in (20). The bottom-right subplot shows the linearization error obtained when applying the equivariant output approximation [33] in the polar parameterization.

extended direct product group

\[ G := \text{SE}_2(3) \times \mathbb{R}^6 \times \text{SE}(3) \times \text{SOT}(3)^n \]

\[ \simeq \text{SLAM}_n^\text{VI}(3) \times \mathbb{R}^6 \times \text{SE}(3). \]  

We denote a typical element as \( X = (A, \beta, B, Q_1, \ldots, Q_n) \in G \) and frequently use the shorthand \((A, \beta, B, Q_1)\).

**Lemma V.3:** The map \( \phi : G \times T_n^\text{VI}(3) \to T_n^\text{VI}(3) \) defined by

\[
\phi((A, \beta, B, Q_1), (\xi_B, b_B, T, p_i)) := (\varphi^B(A, \xi_B), b_B + \beta, P_A^{-1}TB, P_A^{-1}TBQ_i^{-1}T^{-1}P_B^{-1}(p_i))
\]

\[ \text{(23)} \]

where \( \varphi^B \) is defined as in (10), is a transitive right group action.

**Lemma V.4:** The group action \( \phi \) (23) is compatible with the VI-SLAM invariance \( \alpha \) (7), i.e.,

\[ \phi(X, \alpha(S, \xi)) = \alpha(S, \phi(X, \xi)) \]

for all \( \xi \in T_n^\text{VI}(3), S \in \text{SE}_{n_0}(3), \) and \( X \in G \).

**Lemma V.5:** The map \( \rho : G \times \mathcal{N}_n^\text{VI}(3) \to \mathcal{N}_n^\text{VI}(3) \) defined by

\[ \rho((A, \beta, B, Q_1), (y_i)) := (R_Q^T y_i) \]

\[ \text{(24)} \]

is a right group action. In addition, the measurement function (6) is equivariant with respect to the actions \( \phi \) (23) and \( \rho \), i.e.,

\[ h(\phi(X, \xi)) = \rho(X, h(\xi)) \]

for all \( X \in G \) and \( \xi \in T_n^\text{VI}(3) \).

Overall, the structure presented provides a complete symmetry of the VIO dynamics and measurement. This is summarized by the diagram

\[ T_n^\text{VI}(3) \xrightarrow{\alpha_s} T_n^\text{VI}(3) \xrightarrow{h} \mathcal{N}_n^\text{VI}(3) \]

\[ \phi \]

\[ T_n^\text{VI}(3) \xrightarrow{\alpha_s} T_n^\text{VI}(3) \xrightarrow{h} \mathcal{N}_n^\text{VI}(3) \]

\[ \phi \]

\[ T_n^\text{VI}(3) \xrightarrow{\alpha_s} T_n^\text{VI}(3) \xrightarrow{h} \mathcal{N}_n^\text{VI}(3) \]

\[ h \]

which commutes for any \( X \in G \) and \( S \in \text{SE}_{n_0}(3) \).

The proposed EqF formulation has two key differences to our previous work [12]. First, while both the proposed EqF and the EqF presented in [12] feature the same \( \text{SE}_2(3) \) symmetry associated with the inertial states, the EqF presented in [12] did not include the translation and yaw components directly. Instead, the translation and yaw components (corresponding to the invariance \( \alpha_S \)) were excluded from the filter, and a secondary bundle lift was used to compute an update to the these states separately from the EqF equations. The advantage of the new formulation is that the EqF now provides a covariance estimate
for the translation and yaw. The second key difference is the inclusion of the camera extrinsics in the VI-SLAM state with an associated \( \text{SE}(3) \) symmetry. In [12], the camera extrinsics were simply assumed to be fixed constant.

VI. EqF for VIO

While the performance of an EKF for VIO depends largely on what coordinates are chosen, the performance of an EqF depends instead on the symmetry used. The VI-SLAM symmetry yields a few key advantages.

The EqF equipped with the VI-SLAM symmetry is a consistent estimator for VIO. In Lemma VI.1, Lemma V.3 is used to lift the system dynamics from the state-space manifold to the chosen symmetry group \( \mathcal{G} \), and the resulting lift \( \Lambda \) is shown to be invariant to the reference frame transformation \( \alpha \). Then, since both the group action \( \phi \) (23) and the lift \( \Lambda \) (25) are compatible with \( \alpha \) (7) (cf. Lemmas V.4 and VI.1), the proposed EqF naturally also respects the invariance. Specifically, if the Riccati matrix \( \Sigma \) of the EqF is treated as a covariance, then the Fisher information \( \Sigma^{-1} \) is nonincreasing along the unobservable directions spanned by the differential of \( \alpha \), and therefore, the EqF is consistent by [48, Th. 2].

The EqF uses a higher order (more precise) output approximation than that available to the EKF. The system output function is shown to be equivariant in Lemma V.5, and the local coordinates \( \vartheta \) chosen in Section VI-B are normal with respect to the group action \( \phi \) (23). Therefore, the equivariant output approximation \( C^*_\vartheta \) is available [8, Lemma V.3], and, unlike the usual first-order approximation of the output function provided by an EKF, it has no second-order error terms, i.e.,

\[
y - h(\hat{\xi}) = C^*_\vartheta \varepsilon + O(|\varepsilon|^3)
\]

where \( y \) is the system measurement, \( \hat{\xi} \) is the EqF state estimate, and \( \varepsilon \) is the local error coordinates of the EqF.

A. Lifted Dynamics and Consistency

The existence of a transitive action by the VI-SLAM group on the VI-SLAM manifold guarantees the existence of a lift for the system dynamics (5) in the sense of [7]. Although [7] provides a constructive algorithm to build lifts, in practice, it is easiest to guess the lift and then prove that it satisfies the required conditions.

**Lemma VI.1:** The map \( \Lambda : \mathcal{T}^\text{VI}_n(3) \times (\mathbb{R}^3 \times \mathbb{R}^3) \to \mathfrak{g} \), given by

\[
\begin{align*}
\Lambda((\xi_b, b, T, p_i), (\Omega, \alpha)) &:= \\
&= \left((\Omega, R^T_{\xi} v, a)^{\alpha}_{\mathcal{T}(3)}, 0, (\Omega_C, v_C)^{\alpha}_{\mathcal{T}(3)}\right) \quad (25)
\end{align*}
\]

is a lift [7] of the system dynamics (5), i.e.,

\[
D_\mathcal{T}\mathfrak{g}\mathcal{L}((\xi_b, b, T, p_i), (\Omega, \alpha)) = f((\xi_b, b, T, p_i))
\]

Moreover, \( \Lambda \) is invariant with respect to the action \( \alpha \) (7), i.e.,

\[
\Lambda((\alpha(S, (\xi_b, b, T, p_i)), (\Omega, \alpha)) = \Lambda((\xi_b, b, T, p_i), (\Omega, \alpha))
\]

for all \( S \in \text{SE}(3) \).

**Proof:** The proof that \( \Lambda \) satisfies the lift condition (26) closely follows the proof of [33, Lemma 4.4], and the invariance of \( \Lambda \) to \( \alpha \) is straightforward. Both have been omitted to save space. \( \square \)

B. Origin Choice and Local Coordinates

Let \( \hat{\xi} = (\xi_b, \hat{b}_n, \hat{T}, \hat{p}_i) \in \mathcal{T}^\text{VI}_n(3) \) denote the chosen fixed origin configuration, with \( \hat{p}_i := P_b T e_3 \) for every \( i \). For generality, we leave the remaining terms \( \xi_b, \hat{b}_n, \) and \( T \) arbitrary.

Define the map \( \vartheta : \mathcal{U}_c \subseteq \mathcal{T}^\text{VI}_n(3) \to \mathbb{R}^{21 + 3n} \) by

\[
\begin{align*}
\vartheta((\xi_b, b, T, p_i)) &:= \\
&= \left(\log_{\text{SE}(3)}(\tilde{R}_B R^{-1}_B, \tilde{R}_B^T (x_B - \hat{x}_B), \tilde{R}_B^T (y_B - \hat{y}_B))^\nu \quad b_b - \hat{b}_b \quad \log_{\text{SE}(3)}((P_b T)^{-1}(P_b T))^\nu \quad \varsigma_{\text{SOT}(3)}((P_b T)^{-1}(p_1)) \quad \vdots \quad \varsigma_{\text{SOT}(3)}((P_b T)^{-1}(p_n)) \right)
\end{align*}
\]

(27)

to be the coordinate chart for \( \mathcal{T}^\text{VI}_n(3) \) about \( \hat{\xi} \), where \( \mathcal{U}_c \) is a large neighborhood of \( \hat{\xi} \) and \( \varsigma_{\text{SOT}(3)}(\cdot) \) is the polar parameterization (16). Then, \( \vartheta \) provides normal coordinates for \( \mathcal{T}^\text{VI}_n(3) \) about \( \hat{\xi} \) with respect to the action \( \phi \).

C. EqF Dynamics

Denote the true state of the system \( \xi = (\xi_b, b, T, p_i) \in \mathcal{T}^\text{VI}_n(3) \), let the input signals be \( (\Omega, \alpha) \in \mathbb{R}^3 \times \mathbb{R}^3 \), and denote the measurements as \( y = h(\xi) \).

Let \( \hat{X} = (A, \beta, B, Q_c) \in \mathcal{G} \) be the observer state and \( \Sigma \in \mathbb{S}_+(21 + 3n) \) be the Riccati matrix, where \( \mathbb{S}_+(k) \) denotes the set of positive-definite \( k \times k \) matrices. Define \( A_t, B_t, \) and \( C^*_t \) to be the state, input, and equivariant output matrices of the EqF, as defined in [8], respectively, and let \( D_\mathcal{T}\mathfrak{g}\mathcal{L}\phi(\xi)(E) \) be a fixed right inverse of \( D_\mathcal{T}\mathfrak{g}\mathcal{L}\phi(\xi) \). Then, the EqF dynamics [8] are defined to be

\[
\begin{align*}
\dot{\hat{X}} &= \hat{X} \Lambda(\phi(\hat{X}, \hat{\xi}), (\Omega, \alpha)) - \Delta \hat{X} \\
\Delta &= D_\mathcal{T}\mathfrak{g}\mathcal{L}\phi(\xi)(E) \cdot \varsigma_{\text{SOT}(3)}(\dot{\hat{\xi}}) - \Sigma C^*_t \hat{N}^{-1} \hat{N}^{-1} - (y - h(\hat{\xi})) \\
\Sigma &= A_t \Sigma + \Sigma A^*_t + M_t + B_t M^m_t B^*_t - \Sigma C^*_t \hat{N}^{-1} C^*_t \Sigma
\end{align*}
\]

(28)

where \( \Delta \in \mathfrak{g} \) is the EqF correction term, and:

1. \( \Sigma(0) = \Sigma_0 \in \mathbb{S}_+(21 + 3n) \) is the initial Riccati gain;
2. \( M_t \in \mathbb{S}_+(21 + 3n) \) is the state gain;
3. \( M^m_t \in \mathbb{S}_+(6) \) is the input gain;
4. \( \hat{N}_t \in \mathbb{S}_+(2n) \) is the output gain;
\[ T^\text{VI}_n(3) \]

\[ \xi(t) = \phi(\hat{X}(t), \hat{\xi}) \]

\[ \psi^{-1} \]

\[ \psi \]

\[ e(t) = \phi(\hat{X}(t)^{-1}, \xi(t)) \]

\[ \mathbb{R}^{21+3n} \]

\[ \varepsilon(t) = \vartheta(e(t)) \]

Fig. 5. Illustration of the relationships between the true state \( \xi \), the estimated state \( \hat{\xi} \), the observer state \( \hat{X} \), and the global error \( e \). The true and estimated states are related to the error and the origin by the transformation \( \varphi \). The error \( e \) is linearized through local coordinates to yield \( \varepsilon = \vartheta(e) \).

following their descriptions in [8].

At any time, the EqF state estimate is given by

\[ \hat{\xi} = (\hat{\xi}_b, \hat{b}, \hat{T}, \hat{p}_i) := \phi(\hat{X}, \hat{\xi}). \quad (29) \]

The EqF is designed around the definition of a global error state, \( e = \phi(X^{-1}, \xi) \). The filter is developed by expressing the dynamics and measurements in terms of \( e \) and linearizing through the local coordinates \( \vartheta \). Fig. 5 shows how the true state \( \xi \), the estimated state \( \hat{\xi} \), the observer state \( \hat{X} \), and the global error \( e \) are related. From the probabilistic point of view, the Riccati matrix \( \Sigma \) may then be thought of as the covariance of the linearized error \( e = \vartheta(e) \), that is, the EqF estimates the distribution of the true state \( \xi \) to satisfy

\[ \vartheta(\phi(X^{-1}, \xi)) \sim N(0, \Sigma). \]

For a formal discussion of the derivation and probabilistic interpretation of the EqF, we refer the reader to [8], [49].

VII. EXPERIMENTAL RESULTS

The EqF equations (28) are discretized and implemented in C++ using the Eigen matrix library [50]. Visual measurements of landmarks are obtained using GIFT [33, 51] to identify and track image features. We refer to the resulting system as EqVIO. The four key steps in EqVIO are as follows.

1) **Features:** The number of features tracked is kept to a fixed limit that can be set by the user. When the number of features that are successfully being tracked falls below a chosen threshold, new image features are identified.

2) **Preprocessing:** If a feature is unable to be tracked or is identified as an outlier, the corresponding landmark is removed from the state. If any new features have been identified, they are added to the state with a constant initial covariance.

3) **Propagation:** Upon receiving an IMU signal, the EqF state and the Riccati equation are integrated without the correction terms.

4) **Update:** After augmenting the state and removing outliers, the EqF state and the Riccati equation are integrated exclusively with the correction terms.

Fig. 6 provides an overview of the system components and how each of the steps is linked.

A. Performance Comparisons

We compared the performance of EqVIO to other VIO systems on two popular public datasets: EuRoC [10] and UZH FPV [11]. Table I lists the algorithms considered along with information about the back-end and front-end systems used, including GFFT [52], LKT [53], FAST [54], ORB [55], and BRISK [56].

| Algorithm   | Ref. | back end       | front end       |
|-------------|------|----------------|-----------------|
| EqVIO       | *    | EqF            | GFFT + LKT      |
| ROVIO       | [1]  | iterated EKF   | image patches   |
| OpenVINS    | [2]  | MSCKF + EKF    | FAST + LKT      |
| VINS-mono   | [4]  | sliding-window | GFFT + LKT + ORB|
| MSCKF       | [3]  | MSCKF          | FAST + LKT      |
| OKVIS       | [5]  | sliding-window | GFFT + BRISK    |

Where stated, the algorithm performance results have been obtained from the recent benchmark study by Delmerico and Scaramuzza [19]. Otherwise, the algorithms were compiled using the suggested default configuration and run on an Ubuntu 20.04 desktop computer equipped with an AMD Ryzen 7 3700X eight-core processor and 16-GB memory. The tuning parameters of the algorithms were changed between the EuRoC and UZH FPV datasets, but kept constant across all sequences in those datasets. Additional system capabilities, including loop closure and map reuse, were disabled where relevant to ensure a fair comparison between algorithms. The key sensor characteristics and tuning parameters used for EqVIO are listed in Table II. For more detailed tuning parameters, refer to the configuration files provided in the open-source repository.

Table III lists the root mean square error (RMSE) of the position estimates of each of the systems in Table I on the EuRoC dataset. EqVIO achieves the best performance in five of the
sequences and also achieves the best mean performance alongside OpenVINS. In addition, Table IV lists the processing time taken per frame for each of the algorithms tested on the authors’ desktop. Table V lists the state length parameters used to achieve the results for each of the algorithms. On average, EqVIO is faster than the next fastest algorithm by a factor of 2.14 on the EuRoC dataset.

Table VI lists the RMSE of the position estimates for a subset of the systems in Table I on some sequences from the UZH FPV dataset. To ensure a fair comparison, the sequences considered here are those for which OpenVINS has publicly listed tuning parameters at the time of writing. The default tuning parameters were used for ROVIO, and its performance could perhaps be improved with tuning specific to the challenging dataset. The tuning parameters for VINS-Mono were taken from those used in VINS-Stereo (https://github.com/rising-turtle/VINS-Stereo) but with the second camera disabled. Over the selected sequences, EqVIO achieves the best performance in four of the ten sequences and achieves the best mean performance overall. In addition, from Table VII, it is clear that EqVIO is also

---

**Fig. 6.** Overview of EqVIO as a system. The key components can be split into the front end (GIFT) and the back end (EqF).

---

**TABLE II**

| Sensor Characteristic          | EuRoC | UZH FPV | Simulation |
|--------------------------------|-------|---------|------------|
| gyroscope random walk [rad/s/√Hz] | 2.43 × 10^4 | 1.19 × 10^3 | 1.0 × 10^4 |
| accelerometer random walk [m/s/√Hz] | 1.24 × 10^2 | 3.26 × 10^5 | 1.0 × 10^3 |
| gyroscope bias diffusion [rad/s^2/√Hz] | 1.34 × 10^4 | 2.00 × 10^4 | 1.0 × 10^5 |
| accelerometer bias diffusion [m/s^2/√Hz] | 4.46 × 10^3 | 6.30 × 10^3 | 1.0 × 10^4 |
| IMU rate [Hz] | 200 | 500 | 500 |
| image resolution [px] | 752 × 480 | 640 × 480 | 752 × 480 |
| camera focal length [px] | (458.654, 457.296) | (275.460, 274.995) | (458.654, 457.296) |
| image center [px] | (367.215, 248.375) | (315.958, 242.712) | (367.215, 248.375) |
| feature noise std. dev. [px] | 1.9 | 3.8 | 0.5 |
| camera frame rate [Hz] | 20 | 30 | 30 |

The bold values are the smallest values in their respective rows.

**TABLE III**

| Algorithm | BUVIO | ROVIO | OpenVINS | VINS-Mono | MGC-CPF | ORVIO |
|-----------|-------|-------|----------|-----------|----------|-------|
| Source | * | * | [19] | [19] | |
| MH01 | 0.14 | 0.21 | 0.09 | 0.24 | 0.42 | 0.16 |
| MH02 | 0.20 | 0.25 | 0.24 | 0.21 | 0.45 | 0.22 |
| MH03 | 0.09 | 0.25 | 0.16 | 0.15 | 0.23 | 0.24 |
| MH04 | 0.22 | 0.49 | 0.31 | 0.23 | 0.37 | 0.34 |
| MH05 | 0.28 | 0.52 | 0.39 | 0.28 | 0.48 | 0.47 |
| V1_01 | 0.06 | 0.10 | 0.08 | 0.07 | 0.34 | 0.09 |
| V1_02 | 0.14 | 0.10 | 0.09 | 0.28 | 0.20 | 0.20 |
| V1_03 | 0.19 | 0.14 | 0.06 | 0.16 | 0.67 | 0.24 |
| V2_01 | 0.11 | 0.20 | 0.08 | 0.07 | 0.10 | 0.13 |
| V2_02 | 0.17 | 0.14 | 0.07 | 0.16 | 0.16 | 0.16 |
| V2_03 | 0.20 | 0.14 | 0.15 | 0.22 | 1.13 | 0.29 |

Mean | 0.16 | 0.23 | 0.16 | 0.19 | 0.41 | 0.23 |

The bold values are the smallest values in their respective rows.

**TABLE IV**

| Algorithm | BUVIO | ROVIO | OpenVINS | VINS-Mono |
|-----------|-------|-------|----------|-----------|
| Source | * | * | * | * |
| MH01 | 5.13 | 13.42 | 10.27 | 36.61 |
| MH02 | 4.87 | 13.51 | 10.84 | 34.86 |
| MH03 | 5.99 | 16.10 | 11.92 | 34.35 |
| MH04 | 5.88 | 19.27 | 11.40 | 34.18 |
| MH05 | 5.36 | 19.38 | 11.46 | 35.87 |
| V1_01 | 5.87 | 19.91 | 13.45 | 33.92 |
| V1_02 | 4.78 | 21.74 | 12.70 | 26.31 |
| V1_03 | 5.20 | 18.02 | 11.07 | 24.84 |
| V2_01 | 5.74 | 35.99 | 12.01 | 42.78 |
| V2_02 | 5.33 | 19.94 | 11.59 | 28.10 |
| V2_03 | 5.59 | 20.70 | 10.52 | 22.99 |

Mean | 5.43 | 20.00 | 11.57 | 32.25 |

The bold values are the smallest values in their respective rows.
Table V

| Algorithm     | Ref. | Number of features | Number of states |
|---------------|------|--------------------|------------------|
| EQVIO         |      | 40                 | 1                |
| ROVIO         | [1]  | 25                 | 1                |
| OpenVINS      | [2]  | 50 (SLAM) + 200 (MSC) | 11               |
| VINS-mono     | [4]  | 150                | variable         |
| MSCPF         | [3]  | variable           | 30               |
| OKVIS         | [5]  | 400                | 5                |

Table VI

RMSE of Position Estimates in Meters on the UZH FPV [11] Dataset for Some of the Systems in Table I

| Algorithm          | EqVIO | ROVIO | OpenVINS | VINS-mono |
|--------------------|-------|-------|----------|-----------|
| indoor_45_12       | 0.26  | x     | 0.24     | 0.39      |
| indoor_45_13       | 0.28  | 1.53  | 0.29     | 0.74      |
| indoor_45_14       | 0.27  | 2.63  | 0.22     | x         |
| indoor_45_2        | 0.31  | x     | 0.15     | 0.62      |
| indoor_45_4        | 0.33  | x     | 0.28     | 1.11      |
| indoor_forward_10  | 0.15  | 1.23  | 0.34     | 0.53      |
| indoor_forward_5   | 0.29  | 2.43  | 0.16     | 0.12      |
| indoor_forward_6   | 0.23  | 0.65  | 0.17     | 0.28      |
| indoor_forward_7   | 0.40  | 1.09  | 0.42     | 0.54      |
| indoor_forward_9   | 0.17  | x     | 0.50     | 0.56      |
| **Mean**           | 0.27  | 1.59  | 0.28     | 0.54      |

Table VII

Average Time Taken to Process Each Frame in Milliseconds on the UZH FPV [11] Dataset for Some of the Systems in Table I

| Algorithm          | EqVIO | ROVIO | OpenVINS | VINS-mono |
|--------------------|-------|-------|----------|-----------|
| indoor_45_12       | 3.06  | 5.60  | 17.47    | 79.37     |
| indoor_45_13       | 3.24  | 28.18 | 18.29    | 54.61     |
| indoor_45_14       | 3.53  | 23.85 | 18.53    | 43.16     |
| indoor_45_2        | 3.66  | 32.21 | 18.06    | 53.02     |
| indoor_45_4        | 3.39  | 24.48 | 18.06    | 58.34     |
| indoor_forward_10  | 3.38  | 41.86 | 16.27    | 66.43     |
| indoor_forward_5   | 3.33  | 32.77 | 17.35    | 91.71     |
| indoor_forward_6   | 3.09  | 43.71 | 17.00    | 45.64     |
| indoor_forward_7   | 3.21  | 30.36 | 17.42    | 59.24     |
| indoor_forward_9   | 3.02  | 24.34 | 16.20    | 74.21     |
| **Mean**           | 3.28  | 30.30 | 17.43    | 64.02     |

Table VIII

State Length Parameters of the VIO Algorithms (UZH FPV)

| Algorithm     | Ref. | Number of features | Number of states |
|---------------|------|--------------------|------------------|
| EQVIO         | [1]  | 40                 | 1                |
| ROVIO         |      | 25                 | 1                |
| OpenVINS      | [2]  | 50 (SLAM) + 200 (MSC) | 11               |
| VINS-mono     | [4]  | 300                | variable         |

significantly faster than any of the other algorithms considered. Table VIII lists the state length parameters used to achieve the results for each of the algorithms. On average, EqVIO is faster than the next fastest algorithm by a factor of 5.3 on the UZH FPV dataset.

One reason for the faster processing time of EqVIO is the use of fewer features than other algorithms. Only ROVIO uses fewer features than EqVIO in these experiments, but this is because it uses a direct image difference as its measurement error, which yields more information from each feature at the cost of increased processing time. While the other algorithms could be made faster by reducing the number of features used, this generally reduces the accuracy of the estimated trajectory, as is also seen to be the case for EqVIO in Section VII-E.

While it remains an open question how to compare different VIO algorithms when accounting for different tuning parameters, the results in this section have shown that EqVIO is able to achieve competitive accuracy while processing data at much higher rates than other state-of-the-art algorithms.

B. Verification of Consistency

In order to verify the consistency of the proposed system, we performed a series of Monte Carlo simulations and computed statistics of the normalized estimation error squared (NEES) and the filter covariance. The NEES of the EqF is calculated using the following formula:

$$\text{NEES} = \frac{1}{m} \vartheta(\phi(\hat{X}^{-1}, \xi)) \Sigma^{-1} \vartheta(\phi(\hat{X}^{-1}, \xi))$$

where $m = 21 + 3n$ is the dimension of $\mathbb{T}^N_1(3), \hat{X} \in \mathbb{G}$ is the observer state, $\xi \in \mathbb{M}$ is the true system state, and $\Sigma \in \mathbb{S}_+^{21+3n}$ is the EqF Riccati matrix.

In each simulation, the robot trajectory was defined by a square pattern of side length 2 that is flown every 20 s. The landmarks were scattered uniformly on the four vertical walls around the square pattern at a distance of 1 m, with 25 landmarks scattered on each wall. Fig. 7 shows an example of the true trajectory and the true landmark positions. Every simulation was run for a total of 200 s, meaning that the robot flew around the edges of the square exactly ten times. The characteristics of the simulated IMU and camera were chosen as in Table II. EqVIO was configured with the corresponding input and output gain.
matrices, and the state gain matrix was set to zero. In addition, outlier rejection was not used, and new landmarks were simply added and removed according to their visibility in the simulation.

Fig. 8 shows the resulting statistics of the NEES of EqVIO taken over 1000 Monte Carlo simulations. NEES statistics are shown for estimates of the full state, the pose, and the attitude. The top row of the figure shows the experimental and theoretical median values and 95% confidence bounds of NEES for each time. In all three subplots, the results of the experimental trials closely match the theoretical values. The bottom row of the figure shows the experimental and theoretical distributions of NEES at the final simulation time of 200 s. Again, in all three subplots, the histogram of experimental values matches the theoretical distribution closely.

Fig. 9 shows the evolution of pose error over time for each trial, along with lines indicating three times the standard deviation reported by the filter. The errors in roll and pitch converge quickly and remain small for the whole simulation time. The errors in position and yaw grow slowly unbounded over time due to their unobservability associated with the invariance group action, as discussed in Section IV-C. Fig. 9 demonstrates that the pose error growth is in line with the estimated associated standard deviation.

These results verify that the proposed EqF provides consistent estimates of the true system state. This is a characteristic shared with other recent works that exploit symmetry for observer design in SLAM and VIO, such as [23], [26], [27], and [28], which also achieve filter consistency by exploiting additional techniques like the FEJs [2], [31] or the OC update [57]. In summary, EqVIO maintains a consistent estimate and covariance of the true state error over time, and the observability of the system states is reflected accurately in the Riccati matrix.

C. Convergence of Bias and Camera Parameters

We conducted an additional series of Monte Carlo experiments to verify the convergence of the camera extrinsics and IMU biases to their true values. In order to ensure the observability of the camera extrinsics, in particular, a different trajectory was chosen to the square trajectory used in the consistency experiments in Section VII-B. In each simulation, the IMU trajectory was defined by

\[
R_B(t) := \exp_{SO(3)} \left( \frac{\pi}{4} \begin{pmatrix} \cos(0.25t) \\ \cos(-0.3t) \\ \cos(0.2t) \end{pmatrix} \right),
\]

\[
x_B(t) := \frac{1}{2} \begin{pmatrix} \cos(0.1\pi t) \\ \cos(0.2\pi t) \\ \cos(0.15\pi t) \end{pmatrix}.
\]

The landmarks were scattered uniformly on the four vertical walls and on the floor and ceiling around the bounds of the trajectory at a distance of 1 m from the bounds of the trajectory, with 20 landmarks scattered on each surface. Fig. 10 shows an example of the true trajectory and the true landmark positions.
Every simulation was run for a total of 90 s. Other than this, the simulation configuration was the same as in Section VII-B. The characteristics of the simulated IMU and camera were defined as in Table II. EqVIO was configured with the corresponding input and output gain matrices, the state gain matrix was set to zero, outlier rejection was not used, and new landmarks included in the state depending on their visibility in the simulation.

Fig. 11 shows the convergence of the camera extrinsic errors along with bounds indicating three times the standard deviation reported by the filter. The initial extrinsic rotation errors and translation errors were drawn from zero-mean normal distributions. The initial standard deviations of rotation error and translation error were chosen to be $\sqrt{5} \times 10^{-2}$ rad and 0.05 m, respectively. The figure shows that the camera extrinsics successfully converge from each of these initial errors, and that the errors are consistent with the covariance reported by the filter. This verifies the ability of EqVIO to perform online extrinsic calibration.

Fig. 12 shows the evolution of bias error over time for each trial, along with lines indicating three times the standard deviation reported by the filter. The initial values of the gyroscope bias and accelerometer bias were drawn from zero-mean normal distributions with standard deviations 0.3 rad/s and 0.1 m/s$^2$, respectively. The figure shows that the gyroscope and accelerometer biases all converge successfully and remain within the bounds suggested by the filter covariance.

The results in this section demonstrate that EqVIO is able to estimate camera extrinsics and IMU biases even with reasonably poor initial estimates. The camera translation and accelerometer bias are seen to take longer to converge than the camera rotation and gyroscope bias. This is explained by the inherent observability of VIO: the accurate estimation of the camera translation offset depends on the IMU undergoing rotations about at least two axes. The rates of convergence are also all reflected in the estimated covariance; the errors are consistent with the standard deviations reported by the EqF. Overall, these results show that EqVIO can be used to accurately calibrate camera offsets and IMU biases online, and that the reported uncertainty of these estimates matches their true distributions.

D. Example Performance Details

In addition to the experiments comparing EqVIO’s performance with other state-of-the-art algorithms, we collected data to evaluate and verify the system’s performance. We provide examples of these additional results on the EuRoC sequence V2_01.

Fig. 13 shows the time taken to process every frame of the sequence as a flamegraph, and Fig. 14 shows histograms of the time taken for each key step of the system. These figures show...
that the time taken by EqVIO is significantly increased when new features need to be identified rather than only tracked. Nonetheless, the peak time taken for any frame is still comparable to the average processing time of the other algorithms listed in Table IV.

The absolute position and the attitude about the direction of gravity are unobservable due to the reference frame invariance described in Proposition IV.1. However, the velocity and direction of gravity with respect to the body-fixed frame are observable and, thus, expected to remain free of drift for all time. Fig. 15 shows the estimated and true values of the body-fixed gravity and linear velocity over time. Clearly, the EqF maintains a highly accurate estimate of both the body-fixed gravity direction and velocity over the whole trajectory, and, as expected, no drift is present.

Fig. 16 shows the estimated IMU biases over time. The estimated gyroscope biases converge quickly and are very stable throughout the sequence, while the estimated accelerometer biases vary significantly over time. This is associated with observability properties of the VIO problem and is also reflected in the bias convergence experiments of Section VII-C. According to the estimated ground truth provided with the EuRoC sequence V2_01, the true gyroscope and accelerometer biases are constant values $b^\Omega_B = (-0.002295, 0.024939, 0.081667)$ rad/s and $b^a_B = (-0.023601, 0.121044, 0.074783)$ m/s$^2$, respectively. The estimates shown in Fig. 16 are all close to these values once converged, with the exception of the $z$-axis of the accelerometer bias. One factor may be that EqVIO approximates the strength of gravity as 9.80665 m/s$^2$, rather than using the exact gravity of the room where the dataset was recorded.

E. Number of Landmarks

In the experiments carried out on the EuRoC and UZH FPV datasets, EqVIO was restricted to use a maximum of 40 landmarks at any given time. This is a design choice based on trading off the desired accuracy and processing time performance. Fig. 17 shows how the position RMSE and processing time of EqVIO change for varying numbers of landmarks. EqVIO
Appendix

Proof of Proposition IV.1: Let \( S \in SE_3(3), \xi = (\xi_b, b_b, T, p_i) \in T^V(3) \), and \((\Omega, a) \in L\) be arbitrary. Then, compute

\[
\begin{align*}
f_{(\Omega, a)}(\alpha(S, (\xi_b, b_b, T, p_i))) &= f_{(\Omega, a)}((S^{-1}P_b, R_S^T v_b, b_b), T, S^{-1}(p_i)) \\
&= (R_S^T R_b)(\Omega - b_b^\Omega) - 1, (R_S^T v_b), (R_S^T R_b)(a - b_b^a) + q e_3, 0, 0, 0) \\
&= (R_S^T (R_b(\Omega - b_b^\Omega)) - 1, R_S^T v_b, R_S^T (R_b(a - b_b^a) + q e_3), 0, 0, 0) \\
&= d\alpha f_{(\xi_b, b_b, T, p_i)}
\end{align*}
\]

where the second-last line follows from \( R_S^T e_3 = e_3 \). This shows that, indeed, \( f \) is invariant with respect to the action \( \alpha \). To show the invariance of \( h \), it is sufficient to show the invariance of the component functions \( h^k \) defined in (6). One has that

\[
\begin{align*}
h^k(\alpha(S, (\xi_b, b_b, T, p_i))) &= h^k(((S^{-1}_1 P_b, R_S^T v_b, b_b^1, b_b^0), T, S^{-1}(p_i))) \\
&= \pi_{S^2}( (S^{-1}_1 P_b T)^{-1} (S^{-1}(p_i))) \\
&= \pi_{S^2}( (P_b T)^{-1} S S^{-1}(p_k)) \\
&= \pi_{S^2}( (P_b T)^{-1}(p_k)) \\
&= h^k(\xi_b, b_b, T, p_i).
\end{align*}
\]

This completes the proof.

Proof of Lemma V.2: It is straightforward to see that \( \varphi^V \) and \( \rho^V \) are indeed right group actions. To see the equivariance of \( h^V \), let \( Q \in \text{SOT}(3) \) and \( q \in \mathbb{R}^3 \setminus \{0\} \) be arbitrary. Then

\[
\begin{align*}
h^V(\varphi^V(Q, q)) &= h^V(c^{-1} R^T q) \\
&= c^{-1} R^T q \frac{q}{|c^{-1} R^T q|} \\
&= c^{-1} R^T q \frac{q}{|c^{-1} R^T q|} \\
&= R^T q \\
&= \rho^V(Q, h^V(q))
\end{align*}
\]

as required.

Proof of Lemma V.3: Let \( \Omega = (\xi_b, b_b, T, p_i) \in T^\Omega(3) \) and \( X_1 = (A_1, \beta_1, B_1, Q_{1,i}), X_2 = (A_2, \beta_2, B_2, Q_{2,i}) \in G \) be arbitrary. Then

\[
\begin{align*}
\phi(X_2, \phi(X_1, \xi)) &= \phi(X_2, (\varphi^B(A_1, \xi_b), b_b + \beta_1, P_{A_1}^{-1} T B_1, \\
P_b T B_1 Q_{1,i}^{-1} T^{-1} P_{B_1}^{-1}(p_i))) \\
&= (\varphi^B(A_2, \varphi^B(A_1, \xi_b)), b_b + \beta_1 + \beta_2, P_{A_1}^{-1}(P_{A_2}^{-1} T B_1) B_2,
\end{align*}
\]

Fig. 17. Relationship between the maximum number of landmarks used in EqVIO and the position RMSE and processing time on the EuRoC dataset.
Let 

\[ P_B P_A I T B_1 B_2 Q_2^{-1} (P_A^{-1} T B_1)^{-1} (P_B P_A)^{-1} \]

\[ = (P_B T B_1 Q_1^{-1} T^{-1} P_B^{-1} (p_1)) \]

\[ = (\varphi^B (A_1 A_2, \xi_B, b_B + (\beta_1 + \beta_2), (P_A P_A^{-1}, T (B_1 B_2)), \)

\[ P_B T B_1 B_2 Q_2^{-1} B_1^{-1} T^{-1} P_A, P_A^{-1} P_B^{-1} \]

\[ = (P_B T B_1 Q_1^{-1} T^{-1} P_B^{-1} (p_1)) \]

\[ = (\varphi^B (A_1 A_2, \xi_B, b_B + (\beta_1 + \beta_2), (P_A P_A^{-1}, T (B_1 B_2)), \)

\[ P_B T B_1 B_2 Q_2^{-1} Q_1^{-1} T^{-1} P_B^{-1} (p_1)) \]

\[ = \phi (X_1 X_2, \xi). \]

This shows that the compatibility condition (2) is satisfied. For the identity condition (3), compute

\[ \phi (id_G, \xi) = \phi ((I_5, 0, I_4, (I_4, (\xi_B, b_B, T, p_1)) \]

\[ = (\varphi^B (I_5, \xi_B, b_B + 0, I_4 T I_4, P_B T I_4 I_4^{-1} T^{-1} P_B^{-1} (p_1)) \]

\[ = (\xi_B, b_B, p_1). \]

Then, indeed, \( \phi \) is a group action. Finally, to see that \( \phi \) is transitive, let \( \xi_1, \xi_2 \in \mathcal{T}_n^G (3) \) be arbitrary, and let \( X = (A, B, Q, i) \in G \) such that

\[ P_A = (P_B^{-1})^{-1} P_B^2 \]

\[ v_A = (P_B^{-1})^{-1} (v_B - v_B^2) \]

\[ \beta = b_B - b_B^1 \]

\[ B = (P_B^{-1})^{-1} (P_B^2 T^2) \]

\[ Q_i (P_B^{-1} T^{-1} P_B^{-1} (p_1)) \]

Then, it is straightforward to see that \( \phi (X, \xi_1) = \xi_2 \). This completes the proof. \( \square \)

Proof of Lemma V4. Let \( \xi = (\xi_B, b_B, T, p_1) \in \mathcal{T}_n^G (3), S \in \text{SE}_{3} (3), \) and \( X = (A, B, Q, i) \in G \) be arbitrary. Then

\( \phi ((A, \beta, B, Q, i) \}

\[ \phi (S, (\xi_B, b_B, T, p_1)) \]

\[ = \phi ((A, \beta, B, Q, i), (S^{-1} P_B, R_S v_B, b_B, T, S^{-1} (p_1)) \]

\[ = (S^{-1} P_B P_A, R_S v_B + R_S v_B v_B, b_B, \beta, A^{-1} T B, \]

\[ S^{-1} P_B T B Q_i^{-1} (S^{-1} P_B^{-1} S^{-1} (p_1)) \]

\[ = (S^{-1} P_B P_A, R_S T (v_B + R_B v_A), b_B, \beta, A^{-1} T B, \]

\[ S^{-1} P_B T B Q_i^{-1} (P_B^{-1} (p_1)) \]

\[ = \alpha (S, (P_B P_A, v_B + R_B v_A, b_B, \beta, A^{-1} T B, \]

\[ P_B T B Q_i^{-1} (P_B^{-1} (p_1)) \]

as required. \( \square \)

Proof of Lemma V5. It is trivial to see that \( \phi \) is a group action. To show the equivariance of \( h \), one examines the component measurement functions \( h^k \). Let \( X = (A, \beta, B, Q, i) \in G \) and

\[ \xi = (\xi_B, b_B, T, p_1) \in \mathcal{T}_n^G (3). \]

Then, one has

\[ h^k(\phi ((A, \beta, B, Q, i), (\xi_B, b_B, T, p_1))) \]

\[ = h^k(\varphi^B (A, \xi_B, b_B + \beta, A^{-1} T B, P_B T B Q_i^{-1} (P_B^{-1} (p_1))) \]

\[ = \pi_{S^2} ((P_B P_A)^{-1} (P_B T B Q_k^{-1} T^{-1} P_B^{-1} (p_k))) \]

\[ = \pi_{S^2} ((P_B P_A^{-1} T B (P_B T B Q_k^{-1} T^{-1} P_B^{-1} (p_k))) \]

\[ = \pi_{S^2} (Q_k^{-1} T^{-1} P_B^{-1} (p_k)) \]

\[ = c_{Q_k} R_k^{-1} T^{-1} P_B^{-1} (p_k) \]

\[ = R_k T^{-1} P_B^{-1} (p_k) \]

\[ = R_k h^k (\xi_B, b_B, T, p_1) \]

It follows that

\[ h(\phi ((A, \beta, B, Q, i), (\xi_B, b_B, T, p_1))) \]

\[ = h(\phi ((A, \beta, B, Q, i), (\xi_B, b_B, T, p_1))) \]

\[ = h^n (\phi ((A, \beta, B, Q, i), (\xi_B, b_B, T, p_1))) \]

\[ = (R_k h^k (\xi_B, b_B, T, p_1), \ldots, R_k h^k (\xi_B, b_B, T, p_1)) \]

\[ = \rho ((A, \beta, B, Q, i), (\xi_B, b_B, T, p_1)) \]

as required. \( \square \)
Z. Huai and G. Huang, “Robocentric visual–inertial odometry,” in Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst., 2006, pp. 3562–3568.

A. Barrasso and S. Bonnabel, “An EKF-SLAM algorithm with consistency properties,” 2016, arXiv:1510.06263.

T. Bailey, J. Nieto, J. Guivant, M. Stevens, and E. Nebot, “Consistency of the EKF-SLAM algorithm,” in Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. 2006, pp. 3562–3568.

T. Zhang, K. Wu, J. Song, S. Huang, and G. Dissanayake, “Convergence and consistency analysis for a 3-D invariant-EKF SLAM,” IEEE Robot. Autom. Lett., vol. 2, no. 2, pp. 733–740, Apr. 2017.

K. Wu, T. Zhang, D. Su, S. Huang, and G. Dissanayake, “An invariant-EKF VINS algorithm for improving consistency,” in Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. 2017, pp. 1578–1585.

M. Brossard, S. Bonnabel, and A. Barrasso, “Invariant Kalman filtering for visual inertial SLAM,” in Proc. IEEE 21st Int. Conf. Intell. Fusion, 2018, pp. 2021–2028.

S. Heo, J. H. Jung, and C. G. Park, “Consistent EKF-based visual-inertial navigation using points and lines,” IEEE Sens. J., vol. 18, no. 18, pp. 7638–7649, Sep. 2018.

S. Heo and C. G. Park, “Consistent EKF-based visual-inertial odometry on matrix lie group,” IEEE Sens. J., vol. 18, no. 9, pp. 3780–3788, May 2018.

Y. Yang, C. Chen, W. Lee, and G. Huang, “Decoupled right invariant error states for consistent visual-inertial navigation,” IEEE Robot. Autom. Lett., vol. 7, no. 2, pp. 1627–1634, Apr. 2022.

G. P. Huang, A. I. Mourikis, and S. I. Roumeliotis, “Analysis and improvement of the consistency of extended Kalman filter based SLAM,” in Proc. IEEE Int. Conf. Robot. Autom., 2008, pp. 473–479.

D. Schubert, T. Goll, N. Demmel, V. Usenko, J. Stückler, and D. Cremers, “The TUM VI benchmark for evaluating visual-inertial odometry,” in Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst., 2018, pp. 1680–1687.

P. van Goor, R. Mahony, T. Hamel, and J. Trumpf, “Constructive observer design for visual simultaneous localisation and mapping,” Automatica, vol. 132, 2021, Art. no. 109803.

J. A. Castellanos, J. Neira, and J. D. Tardós, “Limits to the consistency of EKF-based SLAM,” IFAC Proc. Vol., vol. 37, no. 8, pp. 716–721, 2004.

J. Civera, A. Davison, and J. Montiel, “Inverse depth parametrization for monocular SLAM,” IEEE Trans. Robot., vol. 24, no. 5, pp. 932–945, Oct. 2008.

J. Solà, “Consistency of the monocular EKF-SLAM algorithm for three different landmark parameterizations,” in Proc. IEEE Int. Conf. Robot. Autom., 2010, pp. 3513–3518.

J. Solà, T. Vidal-Calleja, J. Civera, and J. M. Martínez-Monti, “Impact of landmark parameterization on monocular EKF-SLAM with points and lines,” Int. J. Comput. Vis., vol. 97, no. 3, pp. 339–368, 2011.

M. Li and A. I. Mourikis, “High-precision, consistent EKF-based visual-inertial odometry,” Int. J. Robot. Res., vol. 32, no. 6, pp. 690–711, 2013.

J. A. Hesch, D. G. Kottas, S. L. Bowman, and S. I. Roumeliotis, “Consistency analysis and improvement of vision-aided inertial navigation,” IEEE Trans. Robot., vol. 30, no. 1, pp. 158–176, Feb. 2014.

Z. Hua, and G. Huang, “Robocentric visual–inertial odometry,” Int. J. Robot. Res., vol. 41, no. 7, pp. 667–689, 2022.

J. M. Lee, Introduction to Smooth Manifolds, New York, NY, USA: Springer, 2013.