Equivariant Filters for Visual Spatial Awareness

Pieter van Goor

A thesis submitted for the degree of
Doctor of Philosophy at
The Australian National University

February 2023

Except where otherwise indicated, this thesis is my own original work.

Pieter van Goor
1 February 2023

I conducted the research in this thesis under the supervision of Prof. Robert Mahony, Prof. Tarek Hamel, and Prof. Jochen Trumpf. Much of the research conducted during my PhD and presented in this thesis has been published in highly regarded international conferences and journals, listed below. The papers that include results presented in this thesis are marked with a $\star$.

Journal Papers

- $\star$ P. van Goor, R. Mahony, T. Hamel, and J. Trumpf, "Constructive observer design for Visual Simultaneous Localisation and Mapping," *Automatica*, vol. 132, p. 109803, 2021
- $\star$ P. van Goor, T. Hamel, and R. Mahony, "Equivariant Filter (EqF)," *IEEE Transactions on Automatic Control*, pp. 1–13, 2022
- $\star$ P. van Goor and R. Mahony, "EqVIO: Equivariant Filter for Visual Inertial Odometry," *Submitted to IEEE Transactions on Robotics*, 2022
- R. Mahony, P. van Goor, and T. Hamel, "Observer Design for Nonlinear Systems with Equivariance," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 5, no. 1, pp. 221–252, May 2022

Conference Papers

- $\star$ P. van 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)*, Dec. 2019, pp. 2543–2549
- $\star$ ——, "An Observer Design for Visual Simultaneous Localisation and Mapping with Output Equivariance," *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 9560–9565, Jan. 2020
- $\star$ P. van Goor, T. Hamel, and R. Mahony, "Equivariant Filter (EqF): A General Filter Design for Systems on Homogeneous Spaces," in *2020 59th IEEE Conference on Decision and Control (CDC)*, Dec. 2020, pp. 5401–5408
- $\star$ P. van Goor and R. Mahony, "An Equivariant Filter for Visual Inertial Odometry," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, May 2021, pp. 14 432–14 438
- Y. Ng, P. van Goor, R. Mahony, and T. Hamel, "Attitude Observation for Second Order Attitude Kinematics," in *2019 IEEE 58th Conference on Decision and Control (CDC)*, Dec. 2019, pp. 2536–2542
- Z. Wang, Y. Ng, P. van Goor, and R. Mahony, "Event Camera Calibration of Per-pixel Biased Contrast Threshold," in *Australasian Conference on Robotics and Automation*, 2019

- Y. Ng, P. van Goor, and R. Mahony, "Pose observation for second order pose kinematics," *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 2317–2323, 2020
- Y. Ng, P. van Goor, T. Hamel, and R. Mahony, "Equivariant Systems Theory and Observer Design for Second Order Kinematic Systems on Matrix Lie Groups," in *2020 59th IEEE Conference on Decision and Control (CDC)*, Dec. 2020, pp. 4194–4199
- R. Mahony, P. van Goor, M. Henein, R. Pike, J. Zhang, and Y. Ng, "Equivariant Visual Odometry in the Wild," in *2020 59th IEEE Conference on Decision and Control (CDC)*, Dec. 2020, pp. 1314–1319
- Y. Ge, P. van Goor, and R. Mahony, "An Equivariant Filter for Stereo-Camera SLAM," in *2020 Australasian Conference on Robotics and Automation*. ARAA, 2020
- P. van Goor and R. Mahony, "Autonomous Error and Constructive Observer Design for Group Affine Systems," in *2021 60th IEEE Conference on Decision and Control (CDC)*, Dec. 2021, pp. 4730–4737

Acknowledgments

None of this research would have happened without the support of my primary supervisor, Robert Mahony. Rob has inspired me with his guidance, wisdom, unrelenting passion for research, and openness to new ideas. He taught me what it means to do good research, and I would not be where I am if not for his careful mentorship. I am proud to have had Rob's support throughout my PhD journey, and am forever grateful for the opportunities he has given me.

I also want to acknowledge the fantastic support from the other members of my supervisory panel, Tarek Hamel and Jochen Trumpf, who always made themselves available for me when I needed help. Jochen, thank you for answering my most abstract mathematical questions. Every time we speak I end up learning far more than I asked for. Tarek, thank you for your flexibility in communicating so frequently between Australia and France, and for all the time we spent together deriving observer equations.

The systems theory and robotics (STR) group has also been an incredible source of support and comradery during my PhD. In no particular order, I would like to thank Mina Henein, Jean-Luc Stevens, Zheyu Zhuang, Sean O'Brien, Gerard Kennedy, Jack Henderson, Yonhon Ng, Ziwei Wang, Shawn Ge, and Matthew Hampsey for their friendship, and for always being open to discussing new ideas. I would not have made it half as far in these last three years if it weren't for all the inspiring conversations we have had. I also want to thank Alex Martin and Andrew Tridgell for their engineering support throughout my project. I have learned a lot from you both, and appreciate the time and effort you both put in to turning my research from theory into reality.

Last, but far from least, I would like to thank my friends and family. To my parents Clara, Bill, Peter, and Nicoline, and to my brothers Karel and Jip, thank you for your faith in me and everything you have done to lead me here. Finally, to Steph, thank you for everything.

Abstract

Algorithms that estimate the navigation states and local environment of a robotic vehicle are core enabling technologies for autonomous systems. The first work that exploited symmetry and equivariance for state estimation in robotics was published in 2000-2005 and had immediate impact as an enabling technology for attitude estimation in the growth of the commercial aerial robotics industry. Recently, the discovery of a novel symmetry structure has allowed equivariant observer theory to be applied to Simultaneous Localisation and Mapping (SLAM), the archetypical spatial awareness problem in robotics. Beyond classical SLAM, however, it has remained unclear how to extend the equivariant observer theory to Visual Spatial Awareness (VSA); the general problem of using the rich visual information obtained from a camera, potentially in combination with other sensors, to model the physical relationship between a mobile robot and its environment. VSA will be one of the enabling technologies across a wide range of emerging autonomous systems applications including autonomous vehicles, aerial vehicles, virtual reality, and augmented reality. Essentially, any future autonomous system that operates in an unstructured dynamic environment alongside humans will need high-performance, robust VSA solutions. This thesis brings together the new field of equivariant systems theory with the application domain of visual spatial awareness, and contributes equally to the development of general theory motivated by the application domain and the development of real world algorithms based on the theory.

Existing equivariant observer designs suffer from key limitations. Constructive equivariant observers are specific to their application and difficult to generalise outside of the specific class of left- or right-invariant systems, while the invariant extended Kalman filter can only be applied to the subclass of group-affine systems defined with Lie group state space. To overcome these issues, this thesis proposes the Equivariant Filter (EqF), a general linearisation-based observer design that can be applied to any system exhibiting a Lie group symmetry. I show that it is possible to obtain a second-order (rather than first-order) approximation of the system measurement function when output is compatible with the chosen system symmetry. Observers based on this framework have demonstrably lower linearisation error, better performance, and better robustness than state-of-the-art stochastic filters.

To address the Visual SLAM (VSLAM) and Visual Inertial Odometry (VIO) problems, I develop new Lie groups, which are the first symmetries for which the measurements of VSLAM and VIO are equivariant. I use the new VSLAM symmetry to develop a novel constructive observer design for VSLAM with almost-globally asymptotically stabilisable error dynamics. For the VIO problem, I combine the newly proposed EqF with the new VIO symmetry and generate open-source code that exhibits better than state-of-the-

art performance on standard robotics datasets and at a speed between two and six of comparable algorithms.

Through the introduction of the EqF and new symmetries for VSLAM and VIO, this thesis presents a clear advance in the theory of equivariant observers and a powerful new approach to filter design for VSA problems.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

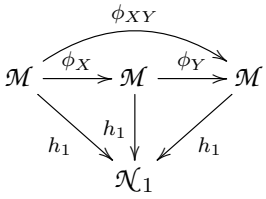
1.1 Motivation

State estimation is the problem of combining knowledge of the dynamics and measurements of a system to derive information about the states of that system. The design of estimators or *observers* for state estimation is a fundamental problem in control theory. The successful and safe operation of a system often depends on accurate estimates of that system's states. Many systems encountered in robotics, avionics, mechatronics, and computer vision exhibit symmetries, and observers for such systems that leverage these symmetries frequently outperform those that do not [22; 23; 24; 25; 26; 27; 28; 29; 30; 31; 32; 33; 34; 35; 36; 37; 38; 39]. Some examples include the attitude dynamics of quadrotors [40; 41; 24; 25; 28; 29; 31; 42; 35], the kinematics of wheeled robotic vehicles [27; 43; 32; 44; 34; 11], and the estimation of homographies between planar surfaces observed by a camera [30; 33; 36]. 15-20 years ago, the early success of equivariant observers for attitude estimation of flying robots equipped with inertial measurements units (IMUs) [24; 26] resulted in a wider interest in understanding the symmetries of a diverse range of robotic systems [26; 29; 30; 32]. Many of these powerful observer designs can be viewed as products of two core guiding philosophies: respect geometry and exploit symmetry. Working in the language of differentiable manifolds instead of local coordinate representations avoids singularities and ambiguities, and often allows the convergence of an observer's error dynamics to be characterised globally. Taking advantage of the symmetries of a system can lead to desirable outcomes in observer design including autonomous or synchronous error dynamics and improved linearisations of the system. These principles have guided the research presented here, and they are core to the development and application of equivariant observer theory in this thesis.

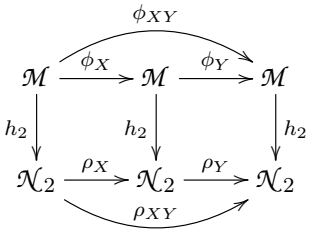
Equivariant Observers

Objective 1. Develop a new approach to equivariant observer theory based on understanding system symmetries as actions of Lie groups on homogeneous spaces, and create a novel observer design for general equivariant systems.

Formally, a symmetry of a system is a Lie group with a group action on the state space manifold of that system. For many systems encountered in robotics, there is a particular symmetry with which the dynamics or measurements of the system are compatible; that is, they are either *equivariant* or *invariant* with respect to the action of the symmetry. Equivariance, derived from 'equi-' and 'variance' meaning 'same' and 'change', occurs when applying a function after transformation by the system symmetry changes the result in such a way that the symmetry could also be applied directly to the result of the function instead. Invariance, derived from 'in-' and 'variance' meaning 'no' and 'change', occurs instead when the result of a function is unaffected by applying the system symmetry before the function. Concretely, consider a Lie group action $\phi : \mathbf{G} \times \mathcal{M} \to \mathcal{M}$, and two functions $h_1 : \mathcal{M} \to \mathcal{N}_1, h_2 : \mathcal{M} \to \mathcal{N}_2$. Then h_1 is invariant and h_2 is equivariant if there exists a Lie group action $\rho : \mathbf{G} \times \mathcal{N}_2 \to \mathcal{N}_2$ such that the following diagrams



Invariance



Equivariance

commute for all $X, Y \in \mathbf{G}$.

I refer to any system with a Lie group symmetry that is compatible with its dynamics as an *equivariant system* [45], and any state estimation algorithm for such a system that uses the symmetry as an *equivariant observer*. Current state-of-the-art equivariant observer design methodologies depend on strong assumptions around the state and dynamics of an equivariant system. Constructive observer designs require the system dynamics to be left- or right-invariant [46], or at most invariant with a bias term [47]. Invariant Extended Kalman Filter (IEKF) designs instead require the system dynamics to be 'group affine', a weaker condition than invariance, but restrict the state space of the system to be a Lie group [48]. However, there are many equivariant systems for which the state space is not a Lie group, and many for which the dynamics are neither invariant nor group affine. The approach taken in this thesis to analysing equivariant systems is distinct from much of the prior literature: rather than modelling the state space of a system with a particular Lie group structure, the system symmetry is developed as a Lie group that acts transitively on the system state space. This distinction between system and symmetry is central to the development of a new theory of equivariant observers that applies not just to systems that can be modelled as Lie group systems with group affine dynamics, but to any system for which a symmetry can be identified.

Visual Spatial Awareness

Objective 2. Extend the recent advances in equivariant observers for classical SLAM and create new symmetries compatible with the geometry and measurements of visual spatial awareness problems.

Visual Spatial Awareness (VSA) is the general problem of using the rich visual information obtained from a camera, often in combination with data from complementary sensors, to model a robot's local environment and estimate a robot's navigation states (e.g. position, attitude, velocity, etc.) with respect to that model. Among problems in spatial awareness, VSA is of particular interest as cameras are ubiquitous, low-cost, lightweight, and small, and each frame provides a large amount of information. A reliable solution to VSA is crucial to many robotic systems and tasks including GPS-denied navigation, virtual and augmented reality (VR/AR), autonomous cars, robotic exploration, and safe human-robot interaction. In the wealth of literature dedicated to this problem, two broad categories of solutions have emerged. *Filter-based* solutions to VSA require less computational resources than *optimisation-based* solutions, but optimisation-based solutions tend to be more accurate [49]. The reduced accuracy of filter-based algorithms is unavoidable, as they rely on linearisations of the system that continuously introduce error into the filter. Optimisation-based algorithms instead repeatedly re-linearise the system to improve accuracy, but do so at the expense of increased computational cost.

Recently, equivariant observers have been applied to Simultaneous Localisation and Mapping (SLAM): one of the classic problems in VSA. This was made possible by the discovery of a novel Lie group by Barrau and Bonnabel [50], who exploited this symmetry to apply an IEKF to the classical SLAM problem. In classical SLAM, the robot's environment is modelled as a collection of landmark points, and the positions of landmarks with respect to the robot can be measured directly. This is in contrast to the more modern visual SLAM (VSLAM), where a monocular camera measures only the bearings of landmarks with respect to the robot. The equivariant observers applied to VSA using this SLAM symmetry [51; 52; 53; 54; 55] have been shown to be consistent filters; a theoretical advantage over filter-based VSA solutions. However, the impact on practical performance of this symmetry for VSA problems is not clear. One issue with these approaches to VSA is that only the full position measurements of classical SLAM are equivariant under the action of the SLAM symmetry proposed in [50], while the bearing measurements characteristic of VSA do not exhibit any such compatibility. The emphasis on compatibility of the output function rather than the system dynamics is a novel feature of the work presented here, and leads to clear improvements in performance over other state-of-the-art VSA solutions. This thesis seeks to extend the ideas of equivariant classical SLAM and find new symmetries that are compatible with the visual outputs of VSA.

1.2 Thesis Contributions

The key contributions of this thesis are as follows.

- **VSLAM Group:** A new Lie group is proposed and it is shown to have compatible symmetry actions on the state and output spaces of visual SLAM. This is the first Lie group symmetry for VSLAM for which the visual measurement function is equivariant (Section 4.3).
- **Constructive VSLAM Observer:** Using the equivariance of the visual measurements under the new VSLAM symmetry, a symmetric output error is defined. The dynamics of this error are then studied to develop a novel constructive observer design with almost globally asymptotically stabilisable error dynamics (Section 4.4).
- **Equivariant Filter:** The Equivariant Filter (EqF) is proposed as a filter design for general equivariant systems on homogeneous spaces. The EqF uses symmetry of a system to define a global equivariant error, and does not suffer from the limitations of local coordinates or embeddings common to other non-linear observer designs for systems on manifolds (Section 5.5). In an example problem of bearing estimation on the sphere, the EqF is shown to outperform an extended Kalman Filter (EKF) (Section 5.6).
- **Equivariant Output Approximation:** It is shown that, in the presence of an equivariant output function, the output approximation used by the EqF can be improved to eliminate second-order errors (Section 5.4). This has a clear effect on the performance of the EqF in the example bearing estimation problem (Section 5.6) and in the real-world application to visual inertial odometry (VIO) (Section 6.6).
- **VI-SLAM Symmetry:** Combining the insights from the VSLAM group with a symmetry for inertial measurements units (IMUs), a new symmetry is developed for visual inertial SLAM (VI-SLAM). This symmetry is compatible with the reference frame invariance of VI-SLAM, leads to invariance of IMU error dynamics, and provides equivariance of the visual measurements (Section 6.4).
- **EqVIO:** The EqF is applied with the VI-SLAM symmetry to develop a full system for VIO called EqVIO. EqVIO is shown to achieve state-of-the-art accuracy on popular VIO datasets at far higher speeds than comparable algorithms (Section 6.6). The code for EqVIO is open-source and available online[1].

1.3 Thesis Outline

Beyond this introductory chapter, this thesis contains the following chapters.

[1] https://github.com/pvangoor/eqvio

- **Chapter 2** reviews the literature on the topics of equivariant observer theory and visual spatial awareness, and the recent works at their intersection.
- **Chapter 3** provides preliminary notations and mathematics that are used throughout the remainder of the thesis. This includes definitions and detailed descriptions of important Lie groups and their Lie algebras including $\mathbf{SO}(3)$, $\mathbf{SE}(3)$, $\mathbf{SOT}(3)$, $\mathbf{SE}_2(3)$.
- **Chapter 4** develops the VSLAM Lie group and shows that the visual measurements of landmarks are equivariant under its action. It also details a constructive observer design for the VSLAM problem based on this symmetry with almost-globally asymptotically stabilisable error dynamics. The performance of this observer is verified both in simulation and in practice using real-world experimental data. This chapter is largely drawn from work published in [5; 6; 1].
- **Chapter 5** introduces the Equivariant Filter (EqF). A clear description of relevant equivariant systems theory provides the framework in which the EqF observer dynamics and error system are defined. It is shown that equivariant measurements can be leveraged to eliminate second-order errors from the output approximation used in the EqF, and the consequences for performance of the EqF are shown by simulations of an example problem. This chapter is largely drawn from work published in [7; 2].
- **Chapter 6** proposes EqVIO: a system for visual inertial odometry (VIO) that applies an EqF with a novel symmetry for visual inertial SLAM (VI-SLAM). EqVIO exhibits far higher speeds than comparable algorithms while achieving similar or better accuracy on popular datasets. This chapter is largely drawn from work published in [8; 3].
- **Chapter 7** concludes the thesis. It provides remarks on the key outcomes of the thesis as a whole, and suggests some broad directions for future research.

Chapter 2

Literature Review

The literature review that follows is divided into three key sections. The first section discusses the literature of equivariant observer theory. Early equivariant observers emerged for a number of key applications, including attitude estimation, where exploiting Lie group symmetries led to performance gains over the existing state-of-the-art solutions [56; 57; 28; 29; 58; 42; 26; 32; 33]. This motivated substantial research effort into developing generic observer architectures for systems with Lie group symmetries, with many solutions demonstrating high performance both in simulation and real-world experiments [59; 37; 46; 60; 47; 39; 61; 4]. The second section describes some of the key literature in the development of simultaneous localisation and mapping (SLAM), and, in particular, in the development of the related problem of visual inertial odometry (VIO). Early solutions to SLAM tended to be based on the extended Kalman filter (EKF) or variants thereof, and substantial research efforts were dedicated to solving some of the issues associated with the linearisations used therein [62; 63; 64; 65; 66]. Around 2010, the SLAM community developed novel optimisation-based solutions that exploited the sparsity of the problem to achieve real-time performance, and a number of powerful practical systems were developed [17; 67; 68; 69; 70; 18; 71]. However, for the problem of VIO specifically, filter-based solutions have remained competitive due to their relatively low processing time and computational resource requirements [72; 73; 74; 75; 76; 77]. In the third and final section, the previous sections are unified and the recent literature on applications of equivariant observers to visual spatial awareness problems is discussed. The novel $\mathbf{SE}_n(3)$ Lie group has been used by a number of authors to develop exciting novel solutions to SLAM and VIO using equivariant observers, but these incur significant linearisation errors due to the incompatibility of the symmetry with visual measurements [50; 78; 51; 52; 55; 79; 80]. This thesis proposes a novel symmetry of visual measurements, develops a new equivariant observer architecture, and combines these to yield a new state-of-the-art solution for visual spatial awareness.

2.1 Equivariant Observer Theory

2.1.1 Applications of Equivariant Observers

An *equivariant observer* is a state estimation algorithms that exploits Lie group symmetries of a given system. The importance of Lie groups in control theory was first recognised in the 1970s [56; 81; 82], where the study of systems with symmetry was concerned particularly with their control. The problem of state estimation for systems with Lie group symmetries was first considered in the 1990s, when Salcudean created a non-linear observer for the angular velocity of a rigid body by using the quaternion representation of attitude in a mechanical energy approach to observer design [83]. Later work by Baerveldt and Klang [40] applied a complementary filter framework to estimating the attitude of a helicopter. They Formulated the estimation problem in terms of the incline angle of the helicopter as a linear space, and achieved highly accurate results between -30 and 30 degrees. In a similar vein, [22] used a high gain observer for attitude estimation of walking robots when angles are assumed to be kept between -30 and 30 degrees by a controller.

In the 2000s, there was an increase in the number of works examining Lie group symmetries for observer design. Aghannan and Rouchon [57] studied non-linear observers for a wide class of non-linear dynamic systems by exploiting the invariance properties of Lagrangian systems. Driven by the emergence of MEMS[1] gyroscopes and accelerometers, much of the research from this period was dedicated to finding new and more robust observer designs for attitude estimation. In [84], an approach based on mechanical energy is used to design a non-linear observer that estimates the attitude of a rigid body in the presence of constant gyroscope bias. Markley [41] examined the effect of attitude error representation in the popular multiplicative extended Kalman filter (MEKF). Mahony *et al.* [24] constructed a complementary filter directly on the special orthogonal group $\mathbf{SO}(3)$ for attitude estimation for the situation where attitude can be measured at low frequency. Hamel and Mahony then extended this work in [25], where they formulated a different SO(3) complementary filter by considering error in the inertial measurements directly. Additionally, they showed that gyroscope bias can be fully estimated from a single bearing measurement, and that the attitude converges to a stabiliser set consistent with that single measurement. In parallel, Bonnabel *et al.* [85] proposed the left invariant extended Kalman filter (L-IEKF) for attitude estimation, and developed a general theory for extended Kalman filters for invariant systems on Lie groups. Following on from [25], Mahony *et al.* [28] designed a complementary filter on $\mathbf{SO}(3)$ using output measurement error, and show the effectiveness of the filter in real-world experiments. Similarly, Madgwick *et al.* [29] designed an attitude filter for 9-DoF IMU (MARG) systems, provided practical implementation using unit quaternions, and showed convincing experimental results. Grip *et al.* [31] extended [28] to consider attitude estimation with time-varying reference vectors and biased gyroscope and vector measurements. To unify many of the previous complementary filter designs for attitude, Trumpf *et al.* [58] also extended the

[1] Microelectromechanical Systems

complementary filter from [28] to consider time-varying reference vectors in a generic framework, and classified the filter's convergence properties for a wide range of sensing modes. More recently, Izadi and Sanyal [42] proposed a novel approach to the attitude estimation problem based on the principles of Lie group variational integrators, which have highly stable energy properties [86]. Another novel approach is presented in [35], where a hybrid observer design approach is used to develop an observer for biased attitude kinematics on SO(3). Finally, in recent work by Ng *et al.* [9], second order attitude kinematics are analysed through the introduction of a tangent group structure for $\mathrm{T}\mathbf{SO}(3)$.

Encouraged by the success of equivariant observers for attitude estimation, many authors looked to extend these advances to other systems with Lie group symmetries. As early as 2003, Rehbinder and Ghosh [23] developed a non-linear pose observer from visual and inertial sensors by separating the attitude and position estimation problems and analysing the system's symmetry and observability. Later, Bonnabel *et al.* [26] proposed an observer for velocity aided inertial navigation that preserved and respected the system symmetry. They additionally provided a brief development on the general theory for invariant observers. In [27], Cheviron *et al.* formulated an observer for robotic vehicle kinematics using the special Euclidean group $\mathbf{SE}(3)$ in the complementary filter framework of [24]. Similarly, Vasconcelos *et al.* [43] proposed an observer on $\mathbf{SE}(3)$ for rigid body pose from landmark measurements, proved almost-global asymptotic and local exponential stability, and considered the problem of biased velocity measurements. Another approach to observer design on $\mathbf{SE}(3)$ is provided in [32], where the error dynamics of the observer were analysed in exponential coordinates and shown to be convergent on the entire domain where the Lie group logarithm is well-defined. Long *et al.* [44] showed that the exponential coordinates of $\mathbf{SE}(2)$ provide an accurate representation the 'banana' distributions often encountered in the propagation of uncertain kinematics of wheeled robotic vehicles. Hua *et al.* [34] improved on the $\mathbf{SE}(3)$ observer proposed in [27] by considering measurements with homogeneous coordinates in the real-projective space $\mathbb{RP}(3)$. Very recently, Ng *et al.* [11] proposed a tangent group structure for $\mathrm{T}\mathbf{SE}(3)$ and developed an observer for second-order pose kinematics using this symmetry. In parallel to the work on $\mathbf{SE}(3)$ for estimation of robotic vehicles, some authors also considered the problem of estimating homographies between images of planar scenes using the special linear group $\mathbf{SL}(3)$. Hamel *et al.* [30] approached homography estimation in the complementary framework of [37]. Grabe *et al.* [33] examined the problem of camera motion estimation when a majority of image features are coplanar. Finally Hua *et al.* [36] built on [30] to develop an observer for homography estimation and demonstrated its robustness to fast motion and image occlusions in real-world experiments. Inspired by the success of equivariant observers in attitude estimation, the understanding and exploitation of Lie group symmetries in other systems has led to high-performance observer designs for a diverse range of robotic systems, including those defined on $\mathbf{SE}(n)$ and $\mathbf{SL}(3)$ in particular.

2.1.2 Equivariant Observer Architectures

The success of observers with Lie group symmetries in a range of robotics problems [41; 28; 85; 27; 30] inspired the development of a general theory of equivariant observers. Bonnabel [85] proposed the left-invariant extended Kalman filter ((L-)IEKF) for systems on Lie groups with left-invariant dynamics. He showed that the error dynamics are independent from the observer state and demonstrated an application of the architecture to attitude estimation. In [59], Bonnabel *et al.* develop a general architecture for designing symmetry-preserving observers and provided three examples of its application. Separately, Lageman *et al.* [37] proposed a general design method for gradient-like observers for systems on Lie groups with invariant dynamics. They extended the concept of synchrony to systems on Lie groups, and showed the proposed observers exhibit almost-global convergence properties. This was followed by Mahony *et al.* [46], who introduced the concept of a *lift* of the system kinematics to a symmetry Lie group, and developed an observer architecture with autonomous error dynamics and powerful convergence guarantees for left- and right-invariant systems. In parallel, Bourmaud *et al.* [87] extended the discrete EKF to systems on Lie groups by modelling probabilities as concentrated Gaussian distributions (CGDs) on Lie groups. Chirikjian and Kobilarov [88], also using CGDs on Lie groups, described a way to model the probabilities associated with nonlinear measurements on Lie groups to the second order by avoiding the assumption of a linearised measurement function. Following their earlier work [87], Bourmaud *et al.* [60] used a similar approach to develop a continuous-discrete EKF for systems on Lie groups with continuous-time dynamics and discrete-time measurements. This continuous-discrete EKF on Lie groups generalises the earlier work on the IEKF [85] by removing the requirement that the Lie group system be invariant to left or right translation. Bourmaud *et al.* [38] then extended [87; 60] to propose an iterated EKF on Lie groups with outlier rejection based on a statistical test on the Lie group, and demonstrated its effectiveness in examples involving the Lie groups $\mathbf{SE}(3), \mathbf{SL}(3), \mathbf{SIM}(3)$. In [47], Koshravian *et al.* extended the developments of [37; 46] to consider systems on Lie groups with biased inputs, and provided convergence guarantees for systems on Lie groups with faithful matrix representations. Saccon *et al.* [89] derived an abstract second-order minimum energy filter for systems on Lie groups, and showed how to obtain explicit formulas for the filter equations in the case of a system on a matrix Lie group.

Recently, Barrau and Bonnabel [48; 39] built on the earlier works [85; 59] to generalise the IEKF from invariant systems on Lie groups to the broader class of *group affine* systems. They showed that for group affine systems, as was the case for invariant systems, the error dynamics of the IEKF are independent of the observer state and linear in the logarithmic coordinates of the Lie group. They also discussed the relationship between discrete-time group affine systems and the automorphism group of a Lie group in [90], and proposed a theory of invariant Luenberger observers by drawing analogues to linear systems theory. Zhang *et al.* [61] proposed an invariant particle filter for systems on Riemannian manifolds and systems on matrix Lie groups using a Stratonovich stochas-

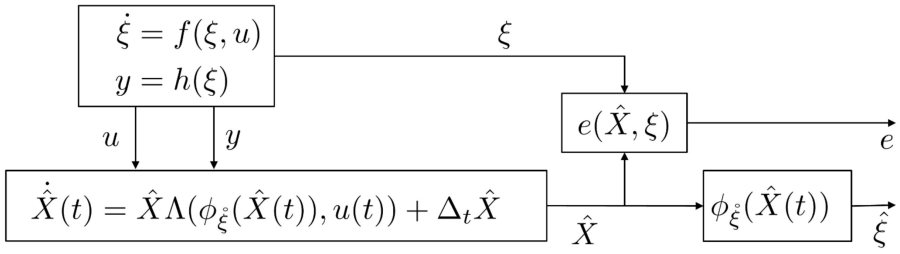


Figure 2.1: The general architecture used in equivariant observer design [4].

tic differential equation. Adapting the IEKF framework, Lavoie *et al.* [91] proposed the invariant $\mathcal{H}_\infty$ filter and demonstrated its performance in an example on the Lie group $\mathbf{SE}_2(3)$. Joshi *et al.* [92] took a different approach to equivariant observer design by examining general manifolds with Lie group actions. They separated the observer design problem into one of estimation on the orbit of the group action and on the remaining quotient manifold, and showed two example of the approach for practical problems. In [12], Ng *et al.* examined the symmetry of the tangent group $\mathrm{T}\mathbf{G}$ to design observers for second order kinematics on general Lie groups. Mahony and Trumpf [93] proposed an equivariant filter design for systems on Lie groups and described the relationships between invariant, group affine, and equivariant systems. The block diagram shown in Figure 2.1 shows the basic structure common to equivariant observer architectures; a lifted system on a Lie group is used to define an global equivariant error on the system state that needs only to be driven to a known constant origin configuration [4]. Collectively, the work in general observer design for equivariant systems has seen significant advances in the last decade, and has had impact in both the nonlinear observer and robotics communities.

2.2 Visual Spatial Awareness

2.2.1 Filter-based Simultaneous Localisation and Mapping

The Simultaneous Localisation and Mapping (SLAM) problem, as pictured in Figure 2.2, has been considered an important technology for mobile robotics since the 90s [16; 94]. Early work by Thrun *et al.* [95] provided a description of SLAM as a maximum likelihood problem, but recognised that the particular form was impractical for implementation of real-time solutions at the time. A key step in the SLAM literature was the work by Dissanayake *et al.* [62], who used an Extended Kalman Filter (EKF) to reduce SLAM to a local probability maximisation problem, and showed some of the filter's convergence properties. In [96; 64; 97], Montemerlo *et al.* developed an approach to SLAM based on Rao-Blackwellized particle filtering called *FAST SLAM*. By reducing the number of particles in the filter, FAST SLAM was able to achieve faster speeds than EKF-based systems. Additionally, particle filters proved to be more robust to non-Gaussian noise. Thrun *et al.* formulated their solution to SLAM in [98] by applying an Extended Information Fil-

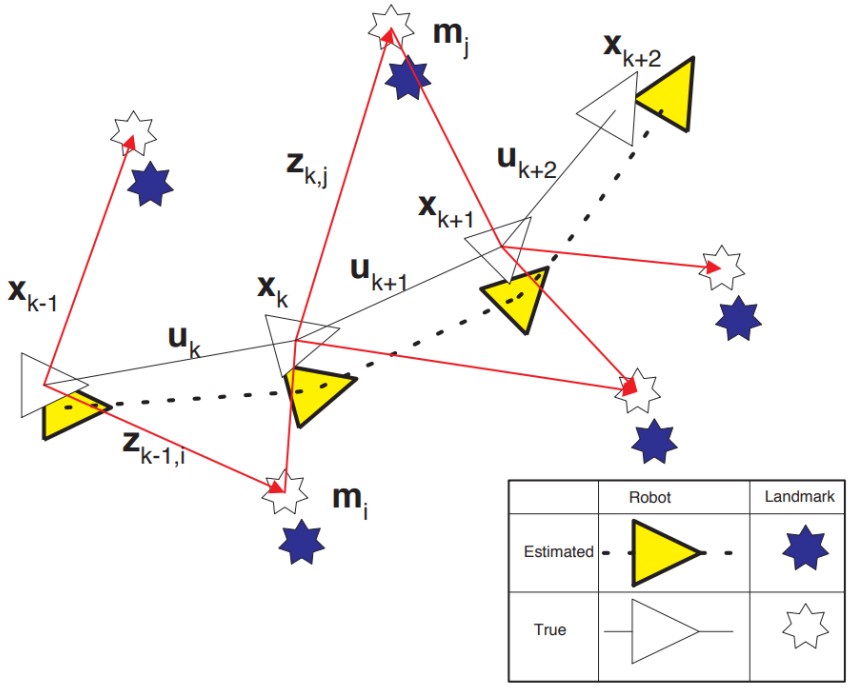


Figure 2.2: A diagramatic representation of the fundamental problem of SLAM [16].

ter (EIF). The EIF is the dual of the EKF; whereas the EKF stores a matrix representing uncertainty in the state, the EIF stores the inverse of this matrix representing confidence in the state. In this reformulation, the update step of the EIF has a linear computational complexity, but computational complexity of the prediction step becomes quadratic. Finally, Eustice *et al.* [99] developed an exactly sparse delayed-state filter for the SLAM problem where all historical poses were included in a filter with an exactly sparse covariance matrix. This formulation relied on a view-based representation of SLAM, where measurements from the robot's sensors do not model the environment explicitly, instead providing relative pose information. As a result, the filter's storage requirements are linear in the number of measurements, although its accuracy is demonstrated to be similar to a comparable EKF on a real-world dataset.

In addition to the question of how to solve the SLAM problem from a mathematical perspective, many authors looked at practical problems concerning the implementation of SLAM in real-world system. The data association problem is the problem of correctly identifying correspondences between measurements, and Neira *et al.* [63] provided an early solution to this by using the predicted locations and estimated uncertainties of landmarks from an EKF to set up and solve a maximum likelihood problem. The computational cost of SLAM was another barrier to real-world implementation, and was addressed in work by Guivant *et al.* [100], who developed approximations that allowed the EKF to perform only local updates most of the time. These methods resulted in a sub-optimal but nonetheless competitive solution for SLAM, and were popular due their significant reductions to the computational cost of EKF systems. Tardos *et al.* [101] fur-

ther reduced the computational cost for real-world SLAM systems by using an EKF for local *submaps* of the global environment, and then joining these maps in a graph optimisation scheme. Later work by the authors [102] showed through simulation experiments that the traditional formulation of EKF was inconsistent; that is, the EKF grew overconfident of its state estimate over time. They proposed a reformulation of the EKF state into robo-centric coordinates, and showed that by combining this with the submaps concept introduced in [101] greatly reduced the inconsistency issues. Estrada *et al.* [103] then showed that the submap joining problem could be solved in linear time to achieve a near-optimal solution. However, Bailey *et al.* [65] continued experiments around the consistency of filter-based SLAM solutions, and showed that even advanced variants of the EKF such as the iterated EKF and the Unscented Kalman Filter did not completely solve the problem. Combining many of the previous ideas developed by the community, a paper by Davison *et al.* [104] introduced *MonoSLAM*, a full SLAM system that worked in real-time from only images taken by a camera without any prior information about the camera motion. A key idea that enabled the performance of the system was the use of an active prediction step: stored image features were only compared to new image features in the area where the new features were expected to be. Finally, Paz *et al.* [66] developed a divide and conquer approach to EKF SLAM, resulting in local quadratic complexity, but global linear complexity, without needing to make approximations other than the linearisations common to EKF.

2.2.2 Optimisation-based Simultaneous Localisation and Mapping

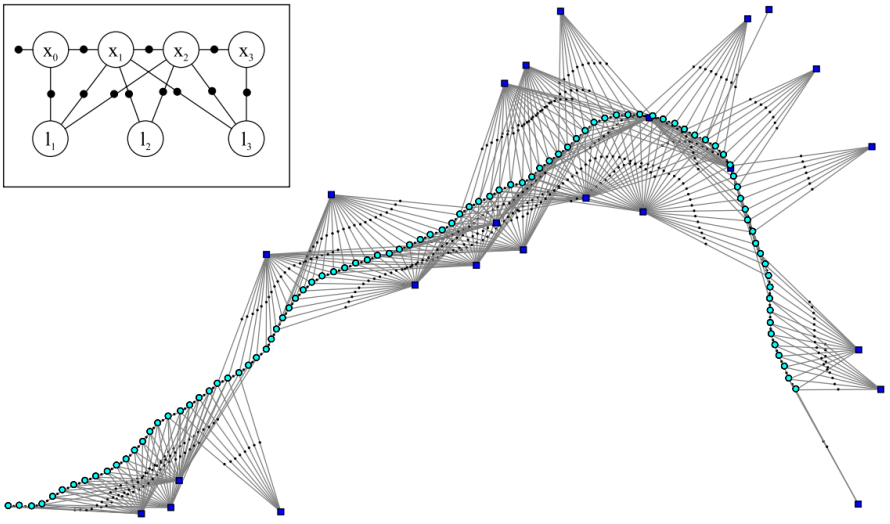


Figure 2.3: The graph optimisation formulation for SLAM is shown with lines between landmarks and robot poses representing measurements that impose a cost on the solution [17].

While very early papers, such as [105], proposed solving SLAM as a graph optimisation problem, filter-based SLAM was the main solution for SLAM until the mid 2000s. Thrun *et al.* [106] and Dellaert and Kaess [17] created some of the first computationally

feasible solutions to SLAM as a graph optimisation problem, by exploiting the sparse nature of the SLAM problem. In [106], the graph SLAM problem was formalised, and a solution based on non-linear optimisation using a sparse information matrix was proposed. In [17], the relationship between graph optimisation and matrix optimisation for SLAM was also explored, and through clever re-ordering of matrix rows and columns, the computational cost of the optimisation process was greatly reduced. In this same paper, Dellart *et al.* also showed that the optimisation-based SLAM solution can be computed faster than an EKF solution. Subsequent work by the same authors showed how QR matrix factorisation and periodic variable reordering can be used to further reduce computational time [107]. In [68], Sibley *et al.* demonstrated that the speed of convergence and the accuracy of the solution to optimisation-based SLAM can be increased by dynamically choosing the frame of reference in which the optimisation is carried out. A general framework for sparse graph optimisation problems was developed in [108] by Kummerle *et al.* , with direct application to the SLAM problem. A paper by Kaess *et al.* [69] built on the previous results from [107] by developing a *Bayes tree* data structure for SLAM, which allowed computation of the optimisation solution to be sped up even further. Finally, in a key paper by Strasdat *et al.* [109], filter-based and optimisation-based SLAM were compared in a comprehensive set of experiments for the case of monocular and stereo camera sensor systems, and it was shown that optimisation-based SLAM outperforms filter-based SLAM in speed and accuracy in both of these scenarios.

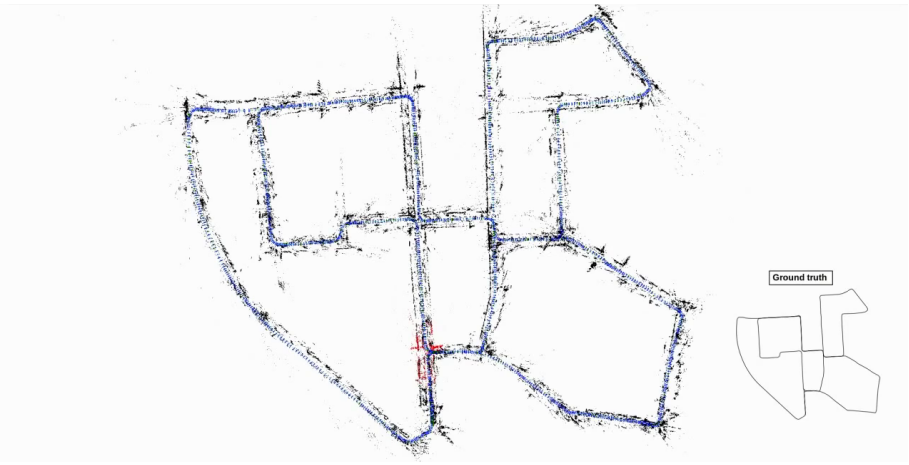


Figure 2.4: An example of the results obtained from ORB-SLAM when run on a dataset of a car driving through some interconnected streets [18].

The advances made in the theory of optimisation-based SLAM lead to new implementations of real-world SLAM systems with improved performance from the traditional filter-based SLAM. Klein *et al.* [67] developed *PTAM*: an implementation of the optimisation-based SLAM that separates global mapping and local tracking into separate problem threads, resulting in high speed processing. In [70], Engel *et al.* developed *LSD-SLAM*: an approach to monocular SLAM that does not use image features but instead works by performing optimisations directly over image data. This allowed for a

more detailed estimation of the environment than traditional SLAM methods; however, it was unable to perform the loop closures that typically ensured consistency of the map produced by SLAM systems. A key work by Mur-Artal *et al.* [18] developed *ORB-SLAM*, a full system for SLAM in the case of a monocular camera sensor system. ORB-SLAM relies on the optimisation-based SLAM system developed in [108], and employs a wide range of heuristics to ensure the Gaussian noise assumptions of the optimisation remain valid. The authors also provide a large set of benchmarking experiments to compare the performance of ORB-SLAM to LSD-SLAM and PTAM. A second version called *ORB-SLAM2* was developed in [71], and incorporates solutions for stereo and RGBD camera systems. Engel *et al.* [110] developed *Direct Sparse Odometry* (DSO), a system that extends the ideas of LSD-SLAM and includes loop closures. A key feature of DSO is that, due to the high accuracy of the local trajectory from LSD-SLAM, the loop closure system requires only a sparse set of features to achieve high accuracy, and can thus achieve low computational time by reducing the number of landmarks that need to be stored. In [111], Usenko *et al.* developed a novel visual-inertial SLAM system by using a local VIO system to yield non-linear factors that marginalise the relevant information for inclusion in an optimisation-SLAM framework. Rosinol *et al.* [112] developed Kimera: an open-source c++ framework for visual-inertial SLAM that combines modules for VIO, global mapping, dense mesh reconstruction, and metric-semantic reconstruction. ORB-SLAM3 [113] adds a capability for visual-inertial SLAM system and a novel multiple map system to ORB-SLAM2. Finally, Zhang *et al.* [114] builds on the non-parametric sensor registration framework developed by Clark *et al.* [115] to propose a continuous visual odometry solution that takes advantages of both geometric and semantic information in a unified manner.

2.2.3 Visual (Inertial) Odometry

Visual Odometry (VO) is a variant of VSA where the only available sensor is a (typically monocular) camera, and the core objective is the estimation of a robot's trajectory, with less importance placed on accurately mapping the local environment. Visual Inertial Odometry (VIO) is the related problem where gyroscope and accelerometer measurements from an Inertial Measurement Unit (IMU) are available in addition to the camera images. One of the first systems to focus on the problem of trajectory estimation from stereo or monocular vision data as distinct from general SLAM was proposed in [116; 117], and used a RANSAC scheme to optimise the estimated pose between image frames. PTAM [67] is another early work that emphasised the distinction between the tasks of local tracking and global mapping. MonoSLAM [104] solved the single camera SLAM problem using an Extended Kalman Filter. The system also removed landmarks from the state when they are heuristically determined to be unreliable, resulting in reduced computational cost and improved reliability of the retained feature set. The Multi-State Constrained Kalman Filter (MSCKF) [72] approached the VIO problem by applying an fixed-lag EKF to the problem, and, notably, eliminating the estimation of landmark

positions from the filter process. This modification resulted in a high speed algorithm for VIO with only linear complexity in the number of landmarks considered. Howard [118] developed one of the first systems to focus primarily on visual odometry over the mapping problem for a stereo camera. The proposed system achieved real-time performance by changing the focus from rejecting outliers to instead detecting inliers. In [119], Konolige *et al.* considered a system equipped with a stereo camera and an IMU, and employed bundle adjustment in a to solve the VIO problem. They improved their results by using specialised image features, and discussed the challenge of using traditional image features in self-similar outdoor environments. Since 2015, the VIO problem specifically has seen substantial interest in the robotics community. Bloesch *et al.* [73; 19] developed *ROVIO*: a VIO algorithm that mixes monocular cameras and IMU measurements in an iterated EKF framework. In contrast to the majority of filter-based VSA systems, ROVIO used a 'direct' error formulation; rather than obtaining feature coordinates from an image, the image pixel values were considered directly in the system model of the EKF. An example screenshot taken from ROVIO is shown in Figure 2.5 and shows the deformed patches being tracked by the algorithm. In parallel, Leutenegger *et al.* [74] developed *OKVIS*, which solves monocular or stereo VIO by using non-linear optimisation on a sliding window of 'keyframes'. Semi-direct Visual Odometry (SVO) [75] used a sparse set of image patches in a frame-to-frame optimisation framework to greatly reduce the presence of outliers and to solve VO at high speeds that remain state-of-the-art. Recently, Qin *et al.* [76] combined a range of modern SLAM techniques to develop VINS-MONO, which performs tightly coupled keyframe optimisation-based VIO with efficient loop closure, and achieves competitive accuracy on popular datasets. Delmerico and Scaramuzza [49] benchmarked a range of state-of-the-art VIO systems. They showed that, generally, the VIO systems optimised 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 [77] is another recent VIO system, which mixes the MSCKF [72] with a traditional EKF and achieves state-of-the-art performance. The key contribution of OpenVINS was to provide a well-documented open platform for filter-based VIO research. The wide range of VO and VIO systems developed in the last decade shows the enduring interest of the robotics community in VSA.

2.3 Equivariant Filters for Visual Spatial Awareness

Equivariant observers have seen great success in state estimation for robotic systems, and visual spatial awareness remains a challenging problem. This has led several authors to investigate the application of equivariant observers to VSA problems. The earliest work in this area was by Barrau and Bonnabel [50], who proposed a novel class of Lie groups, $\mathbf{SE}_n(m)$, and showed that this is a symmetry compatible with 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

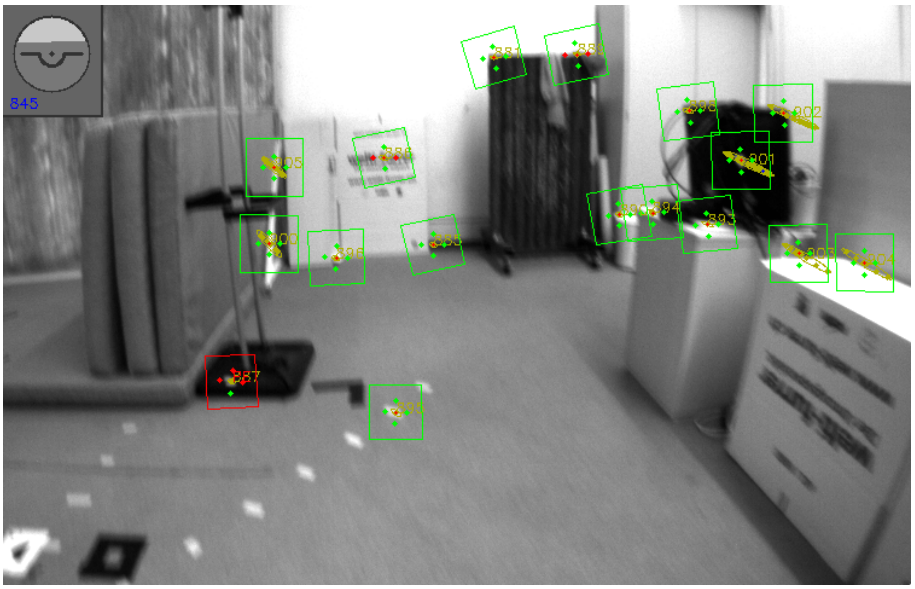


Figure 2.5: A screenshot taken from ROVIO showing information about the patch features being tracked on the current image frame [19].

consistency issues of EKF SLAM [102; 65]. In parallel, Mahony *et al.* [78] developed the same Lie group under a different name, $\mathbf{SLAM}_n(3)$, and proposed a complementary filter for SLAM with almost-global convergence properties. Another contribution of this work was to develop the SLAM manifold, where each element of the space is an equivalence class containing every possible choice of reference frame for the SLAM state. Wu *et al.* [51] combined the novel geometry proposed in [50] with the MSCKF concept of [72] to propose an invariant MSCKF for VIO, which is shown to be a consistent filter unlike the original MSCKF. Zlotnik and Forbes [120] extended [78] to consider biased velocity inputs using the framework proposed in [47]. Two other approaches to visual SLAM by the nonlinear observer community appear in [121] and [122]. Bjorne and Brekke [121] determined the attitude of the robot using an attitude and heading reference system, and then solved visual SLAM as a linear systems problem. Hamel and Samson [122] assume the attitude of the robot is fixed or known, and then use a Riccati observer to solve visual SLAM after providing a rigorous analysis of observability. In [54; 55], Heo *et al.* applied an invariant MSCKF to the problem of VIO with both point and line landmarks, achieving results comparable to other MSCKF algorithms on simulated and real-world datasets. Brossard *et al.* [53] derived an invariant unscented Kalman filter (IUKF) for monocular SLAM using the Lie group proposed in [50], and outperformed other invariant fitlers for VIO. The same authors [52] then proposed an IUKF for the stereo VIO problem using the MSCKF design method [72], and provide some results on popular robotics datasets. Recently, Hashim [79] proposed a nonlinear observer for classical SLAM based on the $\mathbf{SE}_n(3)$ symmetry, and was able to guarantee the error to respect a chosen prescribed performance function. Song *et al.* [80] adapted the concept of the $\mathbf{SE}_n(3)$ symmetry to a symmetry for object SLAM, and showed the consistency of the resulting IEKF.

The majority of equivariant observers for VSA are designed around the $\mathbf{SE}_n(3)$ symmetry first developed for classical SLAM [50; 78]. Recent works have successfully applied invariant filters to VIO and stereo VIO in this symmetry, but have required linearisation of the visual measurements, which are not equivariant with respect to $\mathbf{SE}_n(3)$. Therefore, while this approach has been shown to yield advantages in terms of filter consistency, the impact on filter performance is not yet definitive.

Chapter 3

Preliminaries and Notation

3.1 Smooth Manifolds

For a smooth manifold $\mathcal{M}$, let $\mathrm{T}_\xi\mathcal{M}$ denote the tangent space of $\mathcal{M}$ at ξ, let $\mathrm{T}\mathcal{M}$ denote the tangent bundle, and let $\mathfrak{X}(\mathcal{M})$ denote the space of vector fields over $\mathcal{M}$. Given a vector field $f \in \mathfrak{X}(\mathcal{M})$, $f(\xi) \in \mathrm{T}_\xi\mathcal{M}$ denotes the value of f at $\xi \in \mathcal{M}$.

Given a differentiable function between smooth manifolds $h : \mathcal{M} \to \mathcal{N}$, the linear map

$$\begin{aligned} \mathrm{D}_\xi|_{\xi'} h(\xi) : \mathrm{T}_{\xi'}\mathcal{M} &\to \mathrm{T}_{h(\xi')}\mathcal{N}, \\ v &\mapsto \mathrm{D}_\xi|_{\xi'} h(\xi)[v], \end{aligned}$$

denotes the differential of h with respect to the argument ξ evaluated at ξ'. The map

$$\begin{aligned} \mathrm{D}h : \mathrm{T}\mathcal{M} &\to \mathrm{T}\mathcal{N}, \\ (\xi', v) &\mapsto (h(\xi'), \mathrm{D}_\xi|_{\xi'} h(\xi)[v]), \end{aligned}$$

denotes the differential of h where the base point is implicit in the argument. That is, given $v \in \mathrm{T}_{\xi'}\mathcal{M}$ for some $\xi' \in \mathcal{M}$, and a function $h : \mathcal{M} \to \mathcal{N}$, we write

$$\mathrm{D}h[v] := \mathrm{D}_\xi|_{\xi'} h(\xi)[v] \in \mathrm{T}_{h(\xi')}\mathcal{N}.$$

In general, we denote the composition of two maps h_1 and h_2 by $h_1 \circ h_2$, with $h_1 \circ h_2(\xi) := h_1(h_2(\xi))$. For linear maps H_1, H_2 we choose to denote the composition by $H_1 \cdot H_2$ or $H_1 H_2$ to emphasise the linearity and the link to matrix multiplication. We apply this notation frequently in the context of the chain rule,

$$\begin{aligned} \mathrm{D}_\xi|_{\xi'}(h_1 \circ h_2)(\xi)[v] &= \mathrm{D}_\eta|_{h_2(\xi')} h_1(\eta) \cdot \mathrm{D}_\xi|_{\xi'} h_2(\xi)[v], \\ \mathrm{D}(h_1 \circ h_2)[v] &= \mathrm{D}h_1 \mathrm{D}h_2[v]. \end{aligned}$$

Denote the n-sphere as

$$\mathrm{S}^n := \{x \in \mathbb{R}^n \mid |x| = 1\}.$$

The 1-sphere S^1 is simply the circle, and forms a Lie group under addition of angles. For any n, the projection onto the sphere $\pi_{\mathrm{S}^n} : \mathbb{R}^n \setminus \{0\} \to \mathrm{S}^n$ is defined to be

$$\pi_{\mathrm{S}^n}(x) := \frac{x}{|x|}. \tag{3.1}$$

3.2 Lie Groups

A general Lie group is denoted $\mathbf{G}$ and has Lie algebra $\mathfrak{g}$. A Lie group $\mathbf{G}$ is a smooth manifold equipped with a group structure so that the product and inverse operations are smooth. The identity element is written $\mathrm{id} \in \mathbf{G}$. For any $X \in \mathbf{G}$, the left and right translations by X are denoted L_X and R_X respectively, and are defined by

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

where XY denotes the group product between X and Y. The adjoint map $\mathrm{Ad} : \mathbf{G} \times \mathfrak{g} \to \mathfrak{g}$ is defined by

$$\mathrm{Ad}_X U = \mathrm{D}L_X \cdot \mathrm{D}R_{X^{-1}}[U]$$

for every $X \in \mathbf{G}$ and $U \in \mathfrak{g}$. For any $X \in \mathbf{G}$, Ad_X is linear and invertible, and $(\mathrm{Ad}_X)^{-1} = \mathrm{Ad}_{X^{-1}}$. Moreover, for any $X, Y \in \mathbf{G}$, one has that

$$\mathrm{Ad}_X \, \mathrm{Ad}_Y = \mathrm{Ad}_{XY} \, .$$

The Lie algebra $\mathfrak{g}$ is a vector space equipped with an antisymmetric bilinear operator called the Lie bracket $[\cdot, \cdot] : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$. The adjoint operator $\mathrm{ad} : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$ is defined by

$$\mathrm{ad}_U V = \mathrm{D}_{X|\mathrm{id}}(\mathrm{Ad}_X V)[U],$$

and satisfies $\mathrm{ad}_U V = [U, V]$. Let $\mathfrak{g}$ be G-dimensional. Then one can choose linear isomorphisms 'wedge' $\cdot^\wedge : \mathbb{R}^G \to \mathfrak{g}$ and 'vee' $\cdot^\vee : \mathfrak{g} \to \mathbb{R}^G$ so that $(x^\wedge)^\vee = x$ and $(U^\vee)^\wedge = U$ for all $x \in \mathbb{R}^G$ and $U \in \mathfrak{g}$. Using such maps, any linear map $A : \mathfrak{g} \to \mathfrak{g}$ may be written as a matrix $A^\vee \in \mathbb{R}^{G \times G}$ defined by $A^\vee x = (Ax^\wedge)^\vee$. In particular, the adjoint maps Ad_X and ad_U have $G \times G$ matrix representations written $\mathrm{Ad}_X^\vee$ and $\mathrm{ad}_U^\vee$, respectively.

A product Lie group is formed from the combination of multiple existing Lie groups. If $\mathbf{G}_1, ..., \mathbf{G}_n$ are Lie groups, then the product Lie group is

$$\mathbf{G}_1 \times \cdots \times \mathbf{G}_n := \{(X_1, ..., X_n) \mid X_i \in \mathbf{G}_i\}, \tag{3.2}$$

with multiplication, identity, and inverse given by

$$\begin{aligned}(X_1, ..., X_n)(Y_1, ..., Y_n) &:= (X_1Y_1, ..., X_nY_n),\\ \mathrm{id}_{\mathbf{G}_1\times\cdots\times\mathbf{G}_n} &:= (\mathrm{id}_{\mathbf{G}_1}, ..., \mathrm{id}_{\mathbf{G}_n}),\\ (X_1, ..., X_n)^{-1} &:= (X_1^{-1}, ..., X_n^{-1}).\end{aligned}$$

A right action of a Lie-group $\mathbf{G}$ on a manifold $\mathcal{M}$ is a smooth map $\phi : \mathbf{G} \times \mathcal{M} \to \mathcal{M}$ that satisfies,

$$\phi(Y, \phi(X, \xi)) = \phi(XY, \xi), \tag{3.3}$$

$$\phi(\mathrm{id}, \xi) = \xi, \tag{3.4}$$

for any $X, Y \in \mathbf{G}$ and any $\xi \in \mathcal{M}$. For a fixed $X \in \mathbf{G}$, the partial map $\phi_X : \mathcal{M} \to \mathcal{M}$ is defined by $\phi_X(\xi) := \phi(X, \xi)$. Likewise, for a fixed $\xi \in \mathcal{M}$, the partial map $\phi_\xi : \mathbf{G} \to \mathcal{M}$ is defined by $\phi_\xi(X) := \phi(X, \xi)$. The stabiliser subgroup of ϕ at a point $\xi \in \mathcal{M}$ is denoted

$$\mathrm{stab}_\phi(\xi) := \{S \in \mathbf{G} \,|\, \phi(S, \xi) = \xi\}.$$

An action ϕ is called *transitive* if, for any $\xi, \xi' \in \mathcal{M}$, there exists $X \in \mathbf{G}$ such that $\phi(X, \xi) = \xi'$. We will choose all symmetries in this thesis to be right actions, noting that any left action can be transformed to a right action by considering the inverse parameterization of the group [45]. This choice reflects many widely accepted system models in the robotics community, where right-handed symmetries are naturally associated with the body-fixed sensor suites typical for mobile robot applications.

A homogeneous space $\mathcal{M}$ is a manifold with a smooth and *transitive* symmetry action $\phi : \mathbf{G} \times \mathcal{M} \to \mathcal{M}$, where $\mathbf{G}$ is a Lie group. Let m denote the dimension of $\mathcal{M}$. For any element $\mathring{\xi} \in \mathcal{M}$ one may always choose an m-dimensional subspace $\mathfrak{m} \subset \mathfrak{g}$ such that $\mathrm{D}_{E|\mathrm{id}}\phi(E, \mathring{\xi})$ is a linear isomorphism $\mathfrak{m} \to \mathrm{T}_{\mathring{\xi}}\mathcal{M}$. Let $\cdot^\wedge : \mathbb{R}^m \to \mathfrak{m} \subset \mathfrak{g}$ be a linear isomorphism that identifies $\mathfrak{m}$ with $\mathbb{R}^m$, and let its inverse be $\cdot^\vee : \mathfrak{m} \to \mathbb{R}^m$. Then, at least in a local neighbourhood $\mathcal{U}_{\mathring{\xi}}$ of $\mathring{\xi}$, the map $\vartheta : \mathcal{U}_{\mathring{\xi}} \subset \mathcal{M} \to \mathbb{R}^m$, defined by the unique element $\vartheta(\xi) \in \mathbb{R}^m$ such that $\phi(\exp(\vartheta(\xi)^\wedge), \mathring{\xi}) = \xi$, is well defined and smooth. The map $\vartheta : \mathcal{U}_{\mathring{\xi}} \to \mathbb{R}^m$ is called a normal coordinate chart of the homogeneous space (about $\mathring{\xi}$) [123]. Note that normal coordinates are not unique since the choice of subspace $\mathfrak{m}$ and its identification with $\mathbb{R}^m$ are arbitrary.

Proposition 3.2.1. *Any right action $\phi : \mathbf{G} \times \mathcal{M} \to \mathcal{M}$ induces a right action on vector fields over $\mathcal{M}$, denoted $\Phi : \mathbf{G} \times \mathfrak{X}(\mathcal{M}) \to \mathfrak{X}(\mathcal{M})$, and defined by*

$$\Phi(X, f) := \mathrm{D}\phi_X \cdot f \circ \phi_X^{-1},$$

for any $f \in \mathfrak{X}(\mathcal{M})$ and $X \in \mathbf{G}$. For a fixed $X \in \mathbf{G}$, Φ_X is a linear map on $\mathfrak{X}(\mathcal{M})$.

Proof. To see that Φ is a group action, let $X, Y \in \mathbf{G}$ and $f \in \mathfrak{X}(\mathcal{M})$. Then,

$$\begin{aligned}\Phi(Y, \Phi(X, f)) &= \mathrm{D}\phi_Y \mathrm{D}\phi_X f \circ \phi_X^{-1} \circ \phi_Y^{-1}, \\ &= \mathrm{D}\phi_{XY} f \circ \phi_{Y^{-1}X^{-1}}, \\ &= \Phi(XY, f).\end{aligned}$$

To prove linearity let $c_1, c_2 \in \mathbb{R}$ and let $f_1, f_2 \in \mathfrak{X}(\mathcal{M})$, then

$$\begin{aligned}&\Phi(X, c_1 f_1 + c_2 f_2)(\xi) \\ &= \mathrm{D}\phi_X (c_1 f_1 + c_2 f_2)(\phi_X^{-1}(\xi)), \\ &= c_1 \mathrm{D}\phi_X (f_1)(\phi_X^{-1}(\xi)) + c_2 \mathrm{D}\phi_X (f_2)(\phi_X^{-1}(\xi)), \\ &= c_1 \Phi(X, f_1)(\xi) + c_2 \Phi(X, f_2)(\xi),\end{aligned}$$

where the second-last line follows from the linearity of the differential $\mathrm{D}\phi_X$. □

3.2.1 Special Orthogonal Group $\mathbf{SO}(3)$

The special orthogonal group is a matrix Lie group denoted by $\mathbf{SO}(3)$ with matrix Lie algebra $\mathfrak{so}(3)$. The group and algebra are defined by

$$\begin{aligned}\mathbf{SO}(3) &:= \{R \in \mathbb{R}^{3\times 3} \mid R^\top R = I_3,\ \det(R) = 1\}, \\ \mathfrak{so}(3) &:= \{W \in \mathbb{R}^{3\times 3} \mid W^\top = -W\}.\end{aligned}$$

For any column vector $\Omega = (\Omega_1, \Omega_2, \Omega_3) \in \mathbb{R}^3$, the corresponding skew-symmetric matrix is denoted

$$\Omega^\times := \begin{pmatrix} 0 & -\Omega_3 & \Omega_2 \\ \Omega_3 & 0 & -\Omega_1 \\ -\Omega_2 & \Omega_1 & 0 \end{pmatrix} \in \mathfrak{so}(3).$$

This matrix has the property that, for any $v \in \mathbb{R}^3$, $\Omega^\times v = \Omega \times v$ where $\Omega \times v$ is the vector (cross) product between Ω and v. For any unit vector $y \in \mathrm{S}^2 \subset \mathbb{R}^3$ and any vector $v \in \mathbb{R}^3$,

$$y^\times y^\times v = y y^\top v - v. \tag{3.5}$$

Let the Lie algebra wedge operator be given by $\Omega^\wedge := \Omega^\times$. Then the adjoint maps have matrix forms

$$\mathrm{Ad}_R^\vee = R, \qquad\qquad \mathrm{ad}_{\Omega^\times}^\vee = \Omega^\times,$$

for any $R \in \mathbf{SO}(3)$ and $\Omega^\times \in \mathfrak{so}(3)$.

3.2.2 Scaled Orthogonal Transformations $\mathbf{SOT}(3)$

The scaled orthogonal transformations is a matrix Lie group denoted by $\mathbf{SOT}(3)$ with matrix Lie algebra $\mathfrak{sot}(3)$. It is the direct product of the special orthogonal group $\mathbf{SO}(3)$ and the multiplicative positive real numbers $\mathbf{MR} = \{a \in \mathbb{R} \mid a > 0\}$. The group and algebra are defined by

$$\begin{aligned}\mathbf{SOT}(3) &:= \{aR \mid R \in \mathbf{SO}(3),\ a > 0\},\\ \mathfrak{sot}(3) &:= \{sI_3 + W \mid W \in \mathfrak{so}(3),\ s \in \mathbb{R}\}.\end{aligned}$$

For a matrix $Q \in \mathbf{SOT}(3)$, the notation $a_Q \in \mathbf{MR}$ and $R_Q \in \mathbf{SO}(3)$ represent the scaling and rotation components of Q, respectively. Likewise, for a matrix $U \in \mathfrak{sot}(3)$, the notation $\Omega_W^\times \in \mathfrak{so}(3)$, with $\Omega_W \in \mathbb{R}^3$, and $s_W \in \mathbb{R}$ represent the angular and scale velocity components of W, respectively.

Let the Lie algebra wedge operator be given by

$$\begin{pmatrix}\Omega\\ s\end{pmatrix}^\wedge := \Omega^\times + sI_3,$$

where $\Omega \in \mathbb{R}^3$ and $s \in \mathbb{R}$. Then the adjoint maps have matrix forms

$$\mathrm{Ad}_Q^\vee = \begin{pmatrix}R_Q & 0_{3\times 1}\\ 0_{1\times 3} & 1\end{pmatrix}, \qquad \mathrm{ad}_W^\vee = \begin{pmatrix}\Omega_W^\times & 0_{3\times 1}\\ 0_{1\times 3} & 0\end{pmatrix},$$

for any $Q \in \mathbf{SOT}(3)$ and $W \in \mathfrak{sot}(3)$.

3.2.3 Special Euclidean Group $\mathbf{SE}(3)$

The special Euclidean matrix Lie group is denoted $\mathbf{SE}(3)$ with Lie algebra $\mathfrak{se}(3)$. For a matrix $P \in \mathbf{SE}(3)$, the notation $R_P \in \mathbf{SO}(3)$ and $x_P \in \mathbb{R}^3$ represent the rotation and translation components of P, respectively; that is

$$P = \begin{pmatrix}R_P & x_P\\ 0 & 1\end{pmatrix} \in \mathbf{SE}(3).$$

Likewise, for a matrix $U \in \mathfrak{se}(3)$, the notation $\Omega_U^\times \in \mathfrak{so}(3)$, with $\Omega_U \in \mathbb{R}^3$, and $V_U \in \mathbb{R}^3$ represent the angular and linear velocity components of U, respectively; i.e.

$$U = \begin{pmatrix}\Omega_U^\times & V_U\\ 0 & 0\end{pmatrix} \in \mathfrak{se}(3).$$

Let the Lie algebra wedge operator be given by

$$\begin{pmatrix} \Omega \\ V \end{pmatrix}^{\wedge} := \begin{pmatrix} \Omega^{\times} & V \\ 0_{1\times 3} & 0 \end{pmatrix}$$

where $\Omega, V \in \mathbb{R}^3$. Then the adjoint maps have matrix forms

$$\mathrm{Ad}_P^{\vee} = \begin{pmatrix} R_P & 0_{3\times 3} \\ x_P^{\times} R_P & R_P \end{pmatrix}, \qquad \mathrm{ad}_U^{\vee} = \begin{pmatrix} \Omega_U^{\times} & 0_{3\times 3} \\ V_U^{\times} & \Omega_U^{\times} \end{pmatrix},$$

for any $P \in \mathbf{SE}(3)$ and $U \in \mathfrak{se}(3)$.

3.2.4 Extended Special Euclidean Group $\mathbf{SE}_2(3)$

The Extended Special Euclidean group is a matrix Lie group is denoted $\mathbf{SE}_2(3)$ with Lie algebra $\mathfrak{se}_2(3)$. For a matrix $P \in \mathbf{SE}_2(3)$, the notation $R_P \in \mathbf{SO}(3)$ and $x_P, v_P \in \mathbb{R}^3$ represent the rotation and two translation components of X, respectively; that is

$$P = \begin{pmatrix} R_P & x_P & v_P \\ 0_{2\times 3} & & I_2 \end{pmatrix} \in \mathbf{SE}_2(3).$$

Likewise, for a matrix $U \in \mathfrak{se}_2(3)$, the notation $\Omega_U^{\times} \in \mathfrak{so}(3)$, with $\Omega_U \in \mathbb{R}^3$, and $V_U, A_U \in \mathbb{R}^3$ represent the angular and two linear velocity components of U, respectively; i.e.

$$U = \begin{pmatrix} \Omega_U^{\times} & V_U & A_U \\ 0_{2\times 3} & & 0_{2\times 2} \end{pmatrix} \in \mathfrak{se}_2(3).$$

Let the Lie algebra wedge operator be given by

$$\begin{pmatrix} \Omega \\ V \\ A \end{pmatrix}^{\wedge} := \begin{pmatrix} \Omega^{\times} & V & A \\ 0_{2\times 3} & & 0_{2\times 2} \end{pmatrix}$$

where $\Omega, V, A \in \mathbb{R}^3$. Then the adjoint maps have matrix forms

$$\mathrm{Ad}_P^{\vee} = \begin{pmatrix} R_P & 0_{3\times 3} & 0_{3\times 3} \\ x_P^{\times} R_P & R_P & 0_{3\times 3} \\ v_P^{\times} R_P & 0_{3\times 3} & R_P \end{pmatrix}, \qquad \mathrm{ad}_U^{\vee} = \begin{pmatrix} \Omega_U^{\times} & 0_{3\times 3} & 0_{3\times 3} \\ V_U^{\times} & \Omega_U^{\times} & 0_{3\times 3} \\ A_U^{\times} & 0_{3\times 3} & \Omega_U^{\times} \end{pmatrix},$$

for any $P \in \mathbf{SE}_2(3)$ and $U \in \mathfrak{se}_2(3)$.

3.3 Almost semi-globally asymptotically stabilising controls and corrections

The concept of semi-global asymptotic stabilisability was introduced in [124] to model the dependence of gain on the basin of attraction in feedback stabilisation of a dynamical system. In the context of an observer analysis, this definition can be transferred to stabilisability of the error dynamics by correction. However, the classical concept introduced by Teel and Praly does not capture topological constraints associated with stability analysis on manifolds. For a large class of manifolds, including Lie-groups with $\mathbf{SO}(3)$ as a subgroup, these topological constraints prevent the existence of globally smooth asymptotically stable error dynamics [125]. On such spaces, smooth error dynamics will always admit an exception set χ of unstable or hyperbolic critical points that cannot be part of the basin of attraction of the desired equilibrium.

We consider a system-observer pair, where the observer has the internal model principle, coupled through a correction function $\Delta(\hat{x}, y)$ depending on the observer state and system output. We assume that there is a well defined error function $e : \mathbf{G} \times \mathcal{M} \to \mathcal{M}$ where $\mathbf{G}$ is the observer state space and $\mathcal{M}$ is the system state space. The error dynamics evolve on the manifold $\mathcal{M}$ depending on the system and observer state evolution as well as any exogenous inputs such as velocities.

Definition 3.3.1. *An equilibrium $e_\star$ of the error dynamics of a system-observer pair, on a manifold $\mathcal{M}$, is* almost globally stable *if its basin of attraction is the complement of an exception set $\chi \subset \mathcal{M}$ of measure zero.*

An equilibrium $e_\star$ of the error dynamics of a system-observer pair, on a manifold $\mathcal{M}$, is said to be almost semi-globally stabilisable *if, for each compact set $K \subset \mathcal{M}$ in the complement of an exception set $\chi \subset \mathcal{M}$ of measure zero, there exists a choice of correction $\Delta(\hat{x}, y)$ such that $e_\star$ is an asymptotically stable equilibrium of the error dynamics with basin of attraction containing K.*

Chapter 4

Symmetry of Visual Simultaneous Localisation and Mapping

4.1 Chapter Introduction

Figure 4.1: The Disco Parrot UAV used to gather video data to test the proposed observer.

Simultaneous Localisation and Mapping (SLAM) has been an established problem in mobile robotics for at least the last 30 years [126]. Visual SLAM (VSLAM) refers to the special case where the only exteroceptive sensors available are cameras, and is frequently used to refer to the challenging situation where only a single monocular camera is available. The inherent non-linearity of the VSLAM problem remains challenging [127] and state-of-the-art solutions suffer from high computational complexity and poor scalability [126; 109]. Due to the low cost and low weight, as well as the ubiquity of single camera systems, the VSLAM problem remains an active research topic [126; 49].

In this chapter we present a novel non-linear equivariant observer for the VSLAM problem. The approach uses the SLAM manifold state-space proposed in [78] along with a novel symmetry Lie-group, $\mathbf{SLAM}_n^{\mathrm{V}}(3)$, introduced by van Goor *et al.* [5]. We provide equivariant group actions on the state and output spaces leading to the definition of the

lifted system, a lifted observer and more importantly an intrinsic error that is globally defined. We propose a Lyapunov function expressed in the intrinsic error coordinates and use this to construct an observer for the visual SLAM problem posed on the symmetry group $\mathbf{SLAM}_n^{\mathrm{v}}(3)$. This is in contrast to the majority of state-of-the-art algorithms which depend on local error coordinates and local linearisation [127]. The recent IEKF results [52] exploit a global symmetry of the state-space, however, the symmetry used is not compatible with visual bearings and the resulting algorithm still depends on local linearisation of the measurement function. In our proposed algorithm, separate constant gains for landmark bearing and depth estimates are used, making the design algebraically simple and leading to low computational cost. We show that the error dynamics are *almost semi-globally asymptotically stabilisable* (Def. 3.3). The resulting algorithm has low computation and memory requirements, making it ideally suited to embedded systems applications in consumer electronics.

This chapter consists of five sections alongside the introduction and conclusion. In Section 4.2, we formulate the kinematics, state-space and output of the VSLAM system, and in Section 4.3 we introduce the Lie group $\mathbf{SLAM}_n^{\mathrm{v}}(3)$ and its actions on the state and output spaces. In Section 4.4 we derive a non-linear observer on the Lie group, and in Sections 4.5 and 4.6 we provide the results of a simulation and a real-world experiment carried out using a Disco Parrot UAV (Figure 4.1). The principal contribution of the chapter is theoretical and the experimental sections support this by illustrating the properties of the algorithm and demonstrating that it functions on real-world data. We do not aspire to provide a comprehensive benchmark of performance against state-of-the-art SLAM systems in the present chapter.

4.2 Problem Formulation

Fix an arbitrary reference frame $\{0\}$. Let $P \in \mathbf{SE}(3)$ and $p_i \in \mathbb{R}^3, i = 1, ..., n$ represent the robot pose and landmark coordinates, respectively, defined with respect to $\{0\}$. The raw coordinates of the SLAM problem are written $(P, p_1, ..., p_n) \in \mathbf{SE}(3) \times \mathbb{R}^3 \times \cdots \times \mathbb{R}^3$. The notation $(P, p_i) \equiv (P, p_1, ..., p_n)$ is used for simplicity in the sequel.

The physical measurements in a monocular VSLAM system are the bearings (3D directions) of landmarks perceived by the robot. We assume from now on that the observed landmarks and the robot are not collocated to ensure that the bearing measurements are well defined. Interestingly, this assumption has a substantive impact on the nature of the global symmetries that can be admitted. We make this assumption explicit, defining the *VSLAM total space* $\mathcal{T}_n^{\mathrm{v}}(3)$ to be

$$\mathcal{T}_n^{\mathrm{v}}(3) = \left\{ (P, p_i) \in \mathbf{SE}(3) \times \mathbb{R}^3 \times \cdots \times \mathbb{R}^3 \mid p_i \neq x_P, i = 1, ..., n \right\}. \tag{4.1}$$

We assume that the trajectory of the robot remains in $\mathcal{T}_n^{\mathrm{v}}(3)$ for all time.

Given two configurations $(P, p_i), (P', p_i') \in \mathcal{T}_n^{\mathrm{v}}(3)$, then $(P, p_i) \simeq (P', p_i')$ if there exists $S \in \mathbf{SE}(3)$ such that $(P, p_i) = (S^{-1}P', R_S^\top(p_i' - x_S))$. That is, two sets of coordinates in

the total space $\mathcal{T}_n^{\mathrm{V}}(3)$ are considered equivalent when they are related by a rigid body transformation of the reference frame $\{0\}$. It is straightforward to show the relation $\simeq$ is an equivalence relation on $\mathcal{T}_n^{\mathrm{V}}(3)$ and

$$\lfloor P, p_i \rfloor := \left\{ (S^{-1}P, R_S^\top (p_i - x_S)) \,\middle|\, S \in \mathbf{SE}(3) \right\}.$$

is the associated equivalence class. The VSLAM manifold is the set

$$\mathring{\mathcal{M}}_n(3) = \{ \lfloor P, p_i \rfloor \mid (P, p_i) \in \mathcal{T}_n^{\mathrm{V}}(3) \},$$

with quotient manifold structure. This is the open subset of the SLAM manifold $\mathcal{M}_n(3)$ considered in [78] without those equivalence classes where a landmark is co-located with the robot. Two configurations are equivalent on the SLAM manifold, $(P^1, p_i^1) \simeq (P^2, p_i^2)$, if and only if the ego-centric coordinates of the landmarks are equal, $R_{P^1}^\top (p_i^1 - x_{P^1}) = R_{P^2}^\top (p_i^2 - x_{P^2})$ for all i.

4.2.1 System Dynamics

Assume that the robot is moving in a static environment. Define a velocity input vector space $\mathbb{V} = \mathfrak{se}(3)$ to contain the rigid-body velocity $U \in \mathfrak{se}(3)$ of the robot. The kinematics of the VSLAM system are given by the system function $f : \mathcal{T}_n^{\mathrm{V}}(3) \times \mathbb{V} \to \mathrm{T}\mathcal{T}_n^{\mathrm{V}}(3)$,

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}(P, p_i) &= f((P, p_i), U), \\ &:= (PU, 0). \end{aligned} \tag{4.2}$$

4.2.2 Visual Measurements

The camera measurements are modelled as elements of the sphere S^2. Each individual bearing is given by an output function $h^i : \mathcal{T}_n^{\mathrm{V}}(3) \to \mathrm{S}^2$,

$$h^i(P, p_i) := \frac{R_P^\top (p_i - x_P)}{\|p_i - x_P\|}, \tag{4.3}$$

The output functions are well defined on $\mathring{\mathcal{M}}_n(3)$ since

$$\begin{aligned} h^i(S^{-1}P, R_S^\top (p_i - x_S)) &= \frac{(R_S^\top R_P)^\top (R_S^\top (p_i - x_S) - R_S^\top (x_P - x_S))}{\|R_S^\top (p_i - x_S) - R_S^\top (x_P - x_S)\|}, \\ &= \frac{R_P^\top (p_i - x_P)}{\|p_i - x_P\|}. \end{aligned} \tag{4.4}$$

The full output space of the VSLAM system is defined as a product of n spheres,

$$\mathcal{N}_n^{\mathrm{V}}(3) := \mathrm{S}^2 \times \cdots \times \mathrm{S}^2,$$

with a combined output function $h : \mathcal{T}_n^{\mathrm{V}}(3) \to \mathcal{N}_n^{\mathrm{V}}(3)$

$$h(P, p_i) := \left(\frac{R_P^\top (p_1 - x_P)}{\|p_1 - x_P\|}, \ldots, \frac{R_P^\top (p_n - x_P)}{\|p_n - x_P\|} \right). \tag{4.5}$$

4.3 Symmetry of Visual SLAM

Define a Lie group

$$\mathbf{SLAM}_n^{\mathrm{V}}(3) = \mathbf{SE}(3) \times \mathbf{SOT}(3)^n,$$

with product Lie group structure. The superscript 'V' stands for 'visual'. This Lie group was first proposed in van Goor *et al.* [5]. The associated Lie algebra is denoted $\mathfrak{vslam}_n(3)$. We write elements of $\mathbf{SLAM}_n^{\mathrm{V}}(3)$ as $(A, Q_i) \equiv (A, Q_1, ..., Q_n) \in \mathbf{SLAM}_n^{\mathrm{V}}(3)$. The group product, identity and inverse are given by

$$\begin{aligned} &(A_1, (Q_1)_i) \cdot (A_2, (Q_2)_i) = (A_1 A_2, (Q_1 Q_2)_i), \\ &\mathrm{id} = (I_4, (I_4)_i), \qquad (A, Q_i)^{-1} = (A^{-1}, Q_i^{-1}). \end{aligned}$$

4.3.1 State symmetry

Lemma 4.3.1. *The mapping* $\Upsilon : \mathbf{SLAM}_n^V(3) \times \mathcal{T}_n^V(3) \to \mathcal{T}_n^V(3)$ *defined by*

$$\Upsilon((A, Q_i), (P, p_i)) := (PA, a_{Q_i}^{-1} R_{PA} R_{Q_i}^\top R_P^\top (p_i - x_P) + x_{PA}), \tag{4.6}$$

is a transitive right group action of $\mathbf{SLAM}_n^{\mathrm{V}}(3)$ *on* $\mathcal{T}_n^V(3)$.

Proof. Trivially, $\Upsilon((I_4, I4_i), (P, p_i)) = (P, p_i)$ for any $(P, p_i) \in \mathcal{T}_n^{\mathrm{V}}(3)$. Let $(A_1, (Q_1)_i), (A_2, (Q_2)_i) \in \mathbf{SLAM}_n^{\mathrm{V}}(3)$ and $(P, p_i) \in \mathcal{T}_n^{\mathrm{V}}(3)$ be arbitrary. Then

$$\begin{aligned} &\Upsilon((A_1, (Q_1)_i), \Upsilon((A_2, (Q_2)_i), (P, p_i))) \\ &\quad = (P A_2 A_1, (a_1^{-1} a_2^{-1} R_{PA_2A_1} R_{Q_1}^\top R_{Q_2}^\top R_P^\top (p - x_P)) + x_{PA_2A_1})_i), \\ &\quad = \Upsilon((A_2 A_1, (Q_2 Q_1)_i), (P, p_i)), \\ &\quad = \Upsilon((A_2, (Q_2)_i) \cdot (A_1, (Q_1)_i), (P, p_i)). \end{aligned}$$

Thus Υ is a group action. To see that Υ is transitive, let $(P, p_i), (P', p_i') \in \mathcal{T}_n^{\mathrm{V}}(3)$ be arbitrary. Choose $(A, Q_i) \in \mathbf{SLAM}_n^{\mathrm{V}}(3)$ to satisfy

$$A = P^{-1} P', \qquad a_{Q_i} = \frac{\|p_i - x_P\|}{\|p_i' - x_{P'}\|}, \qquad R_{Q_i} \frac{R_{P'}^\top (p_i' - x_{P'})}{\|p_i' - x_{P'}\|} = \frac{R_P^\top (p_i - x_P)}{\|p_i - x_P\|}.$$

Then $\Upsilon((A, Q_i), (P, p_i)) = (P', p_i')$. □

The action Υ of $\mathbf{SLAM}_n^{\mathrm{V}}(3)$ on $\mathcal{T}_n^{\mathrm{V}}(3)$ is shown in Figure 4.2. Given $(A, Q_i) \in \mathbf{SLAM}_n^{\mathrm{V}}(3)$ and $(P, p_i) \in \mathcal{T}_n^{\mathrm{V}}(3)$, the action transforms the robot pose P by right-translation by A. The

landmark points are transformed by considering their body-fixed coordinates; applying a rotation $R_{Q_i}^\top$ and scaling $a_{Q_i}^{-1}$; transforming them along with the robot pose; and finally writing the result in inertial coordinates.

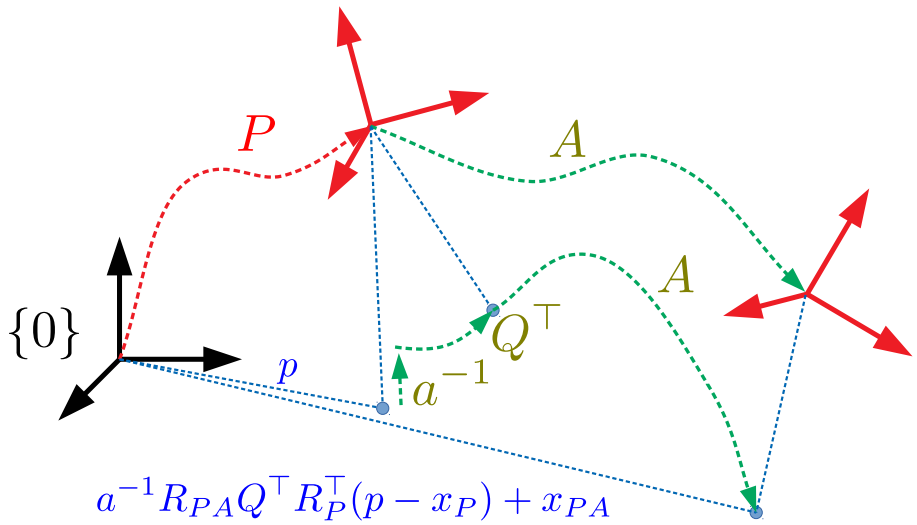


Figure 4.2: The action Υ of the VSLAM group on the state space for some given $(A, Q_i) \in \mathbf{SLAM}_n^{\mathrm{V}}(3)$ and $(P, p_i) \in \mathcal{T}_n^{\mathrm{V}}(3)$. The pose P is mapped to PA. The body fixed frame landmark points $R_P^\top(p_i - x_P)$ are rotated by $R_{Q_i}^\top$ and scaled by $a_{Q_i}^{-1}$ in the body-fixed frame before *transforming with the robot pose* to a new point p_i' which is then rewritten in the inertial frame.

4.3.2 Output symmetry

There is an action ρ of the $\mathbf{SLAM}_n^{\mathrm{V}}(3)$ group on the output space such that the measurement function h defined in (4.5) is equivariant with respect to Υ and ρ. The following lemmas define this action ρ and show the equivariance of h for the proposed $\mathbf{SLAM}_n^{\mathrm{V}}(3)$ geometry. The authors know of no output action that makes bearing outputs equivariant with respect to prior geometries proposed for SLAM [50; 78]. The equivariance structure enables the development of a globally-defined intrinsic error.

Proposition 4.3.2. *The mapping* $\rho : \mathbf{SLAM}_n^V(3) \times \mathcal{N}_n^V(3) \to \mathcal{N}_n^V(3)$ *defined by*

$$\rho((A, Q_i), y_i) = R_{Q_i}^\top y_i, \tag{4.7}$$

is a right group action of $\mathbf{SLAM}_n^V(3)$ *on* $\mathcal{N}_n^V(3)$.

Proof. The proof is straightforward. □

Lemma 4.3.3. *The output* $h : \mathcal{T}_n^V(3) \to \mathcal{N}_n^V(3)$ (4.5) *is equivariant with respect to actions* Υ (4.6) *and* ρ (4.7). *That is, for any* $X \in \mathbf{SLAM}_n^V(3)$ *and any* $\xi \in \mathcal{T}_n^V(3)$,

$$h(\Upsilon(X, \xi)) = \rho(X, h(\xi)).$$

Proof. Let $X = (A, Q_i)$ and $\xi = (P, p_i)$ be arbitrary. Then

$$\begin{aligned} h(\Upsilon(X,\xi)) &= h\left(PA, a_{Q_i}^{-1} R_{PA} R_{Q_i}^\top R_P^\top (p_i - x_P) + x_{PA}\right), \\ &= \frac{R_{PA}^\top (a_{Q_i}^{-1} R_{PA} R_{Q_i}^\top R_P^\top (p_i - x_P) + x_{PA} - x_{PA})}{\|a_{Q_i}^{-1} R_{PA} R_{Q_i}^\top R_P^\top (p_i - x_P) + x_{PA} - x_{PA}\|}, \\ &= \frac{R_{Q_i}^\top R_P^\top (p_i - x_P)}{\|p_i - x_P\|}, \\ &= \rho(X, h(\xi)). \end{aligned}$$

□

4.3.3 Lift of the dynamics

In order to consider the system on the $\mathbf{SLAM}_n^{\mathrm{V}}(3)$ group, the kinematics from the total space must be lifted onto the group. A lift is a map $\Lambda : \mathcal{T}_n^{\mathrm{V}}(3) \times \mathbb{V} \to \mathfrak{vslam}_n(3)$ such that

$$\mathrm{D}\Upsilon_{(P,p_i)}(\mathrm{id})\left[\Lambda((P,p_i),U)\right] = f((P,p_i),U) \tag{4.8}$$

where $f((P,p_i),U)$ is given by (4.2).

Lemma 4.3.4. *The function $\Lambda : \mathcal{T}_n^{V}(3) \times \mathbb{V} \to \mathfrak{vslam}_n(3)$, defined by*

$$\Lambda((P,p_i),U) := (U, (\Lambda_R(U, R_P^\top(p_i - x_P)), \Lambda_a(U, R_P^\top(p_i - x_P)))), \tag{4.9}$$

where $\Lambda_R : \mathfrak{se}(3) \times (\mathbb{R}^3 \setminus \{0\}) \to \mathfrak{so}(3)$ and $\Lambda_a : \mathfrak{se}(3) \times (\mathbb{R}^3 \setminus \{0\}) \to \mathbb{R}_+$ are given by

$$\Lambda_R\left(U, q\right) := \left(\Omega_U + \frac{q \times V_U}{|q|^2}\right)^\times, \qquad \Lambda_a\left(U, q\right) := \frac{q^\top V_U}{|q|^2},$$

is a lift in the sense of (4.8) *of the kinematics* (4.2) *onto $\mathfrak{vslam}_n(3)$ with respect to the group action* (4.6)*.*

Proof. Given $(P,p_i) \in \mathcal{T}_n^{\mathrm{V}}(3)$, let $q_i := R_P^\top(p_i - x_P)$, $W_i^\times := \Lambda_R(U, q_i)$, and $w_i := \Lambda_a(U, q_i)$. Then

$$\mathrm{D}\Upsilon_{(P,p_i)}(\mathrm{id})\left[\Lambda((P,p_i),U)\right] = \mathrm{D}\Upsilon_{(P,p_i)}(\mathrm{id})\left[(U,(W_i^\times, w_i))\right] = (PU, v_i) \tag{4.10}$$

where v_i is

$$v_i = \left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0} \left[(1 + s w_i)^{-1} R_{P(I_4 + sU)} (I_3 + s W_i^\times)^\top q_i + x_{P(I_4+sU)}\right].$$

Computing this derivative and evaluating at $s = 0$ one obtains

$$\begin{aligned} v_i &= -w_i R_P q_i + R_P \Omega_U^\times q_i + R_P (W_i^\times)^\top q_i + R_P V_U, \\ &= -\frac{q_i^\top V_U}{|q_i|^2} R_P q_i + R_P \Omega_U^\times q_i + R_P V_U \\ &\qquad - R_P \left(\Omega_U + \frac{q_i \times V_U}{|q_i|^2} \right)^\times q_i, \qquad (4.11) \\ &= -R_P \frac{q_i q_i^\top}{|q_i|^2} V_U + R_P V_U + R_P \frac{q_i q_i^\top}{|q_i|^2} V_U - R_P V_U, \qquad (4.12) \\ &= 0, \end{aligned}$$

where (4.11) follows from substituting for w_i and W_i with the full expressions for Λ_a and Λ_R, and (4.12) follows from cancelling the Ω_U term and using the relationship (3.5) as well as rearranging the first term. The result follows from substituting directly into (4.10) and recalling (4.2). □

4.4 Observer design

Figure 4.3 shows a schematic overview of the proposed observer. The key features of equivariant observer design are the distinction between the observer state $\hat{X} \in \mathbf{SLAM}_n^{\mathrm{V}}(3)$ and the estimated state $(\hat{P}, \hat{p}_i) \in \mathcal{T}_n^{\mathrm{V}}(3)$, and the design of the correction term around the output error $(d_i) = \rho(\hat{X}^{-1}, y_i)$ rather than the raw measurements $(y_i) \in \mathcal{N}_n^{\mathrm{V}}(3)$.

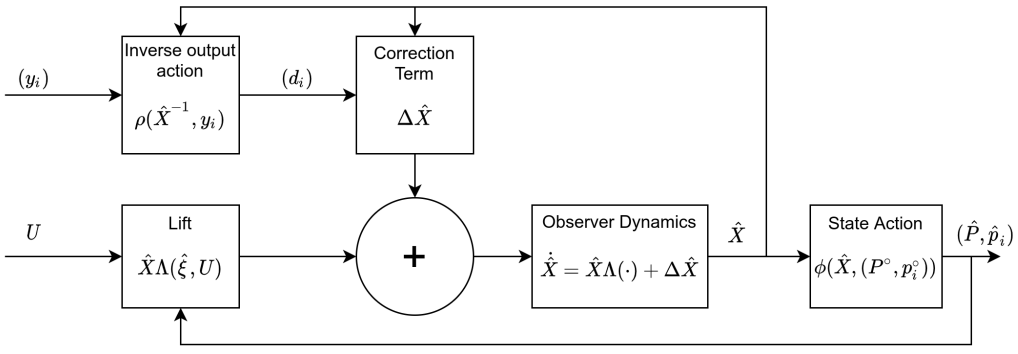


Figure 4.3: An overview of the observer system. Note the distinction between the observer state $\hat{X}$ and the estimated state $(\hat{P}, \hat{p}_i)$, and that the correction term Δ is based on the output error (d_i) rather than the raw measurement (y_i).

4.4.1 Lifted System Dynamics

Let $\xi = (P, p_i) \in \mathcal{T}_n^{\mathrm{V}}(3)$ denote the true state of the VSLAM system. The kinematics of ξ are given by (4.2). Choose an arbitrary origin configuration $\mathring{\xi} = (\mathring{P}, \mathring{p}_i) \in \mathcal{T}_n^{\mathrm{V}}(3)$. The

lifted system is (Lemma 4.3.4)

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}X &= X\Lambda(\Upsilon(X,\mathring{\xi}),U),\\ &= (AU,(R_Q\Lambda_R,a_Q\Lambda_a)_i),\end{aligned} \tag{4.13}$$

for $X=(A,Q_i)\in\mathbf{SLAM}_n^{\mathrm{V}}(3)$. If $\Upsilon(X(0),\mathring{\xi})=\xi(0)$ then trajectories of the lifted system kinematics project to trajectories of the VSLAM kinematics (4.2) [46]. That is, recalling (4.6)

$$\begin{aligned}(P(t),p_i(t)) &= \Upsilon(X(t),\mathring{\xi}),\\ &= (\mathring{P}A, a_{Q_i}^{-1}R_{\mathring{P}A}R_{Q_i}^{\top}R_{\mathring{P}}^{\top}(\mathring{p}_i-x_{\mathring{P}})+x_{\mathring{P}A}),\end{aligned}$$

for all $t\geq 0$, where we have dropped the time dependence of A and Q_i from the notation to improve readability.

4.4.2 Landmark Observer

Let $X=(A,Q_i)$ be a trajectory of the lifted system (4.13) associated with a trajectory $\xi=(P,p_i)$ of the true system satisfying (4.2) for measured input signal $U=(\Omega_U^{\times},V_U)$. Let $y_i:=h^i(\xi)$ (4.3) denote the output.

Fix an arbitrary origin configuration $\mathring{\xi}=(\mathring{P},\mathring{p}_i)\in\mathcal{T}_n^{\mathrm{V}}(3)$. The observer state is $\hat{X}=(\hat{A},\hat{Q}_i)\in\mathbf{SLAM}_n^{\mathrm{V}}(3)$, with kinematics given by

$$\frac{\mathrm{d}}{\mathrm{d}t}\hat{X} := \hat{X}\Lambda(\Upsilon(\hat{X},\mathring{\xi}),U)-\Delta_{\hat{X}}\hat{X}, \qquad \hat{X}(0)=\mathrm{id}, \tag{4.14}$$

where Λ is the lift defined by (4.9) and $\Delta_{\hat{X}}=(\Delta,(\Gamma,\gamma)_i)\in\mathfrak{vslam}_n(3)$ is a correction term that is chosen later. The state estimate is given by $\hat{\xi}=(\hat{P},\hat{p}_i)=\Upsilon(\hat{X},\mathring{\xi})$ (4.6). Let $\mathring{y}_i:=h^i(\mathring{\xi})$ denote the origin output, and let d_i denote the output error [46], defined as

$$d_i=\rho(\hat{X}^{-1},y_i)=\rho(X\hat{X}^{-1},\mathring{y}_i). \tag{4.15}$$

Define the true range $r_i=\|p_i-x_P\|$ and estimated range $\hat{r}_i=\|\hat{p}_i-x_{\hat{P}}\|$ for each i. The range error is

$$\tilde{r}_i=\frac{\hat{r}_i}{r_i}, \tag{4.16}$$

for each i.

Define the twice differentiable barrier function $\beta_{\underline{c}}^{\epsilon}:(\epsilon,\infty)\to[0,\infty)$ to be

$$\beta_{\underline{c}}^{\epsilon}(c):=\begin{cases}\frac{(c-\underline{c})^2}{(\underline{c}-\epsilon)^2(c-\epsilon)}, & \epsilon<c<\underline{c}\\ 0, & c\geq\underline{c}\end{cases}, \tag{4.17}$$

for parameters $0<\epsilon<\underline{c}$.

The landmark correction terms Γ_i and γ_i for $\Delta_{\hat{X}}$ are defined to be

$$\begin{aligned}\Gamma_i &:= \left(\frac{d_i^\top R_{\hat{Q}_i} V_U}{\hat{r}_i(1+d_i^\top \mathring{y}_i)} - \frac{k_i}{(1+d_i^\top \mathring{y}_i)^2}\right)(d_i^\times \mathring{y}_i)^\times + \frac{1}{\hat{r}}\left((\mathring{y}_i - d_i)^\times R_{\hat{Q}_i} V_U\right)^\times, \\ \gamma_i &:= \frac{\alpha_i}{\hat{r}_i^2}\left((1-d_i^\top \mathring{y}_i) d_i^\top R_{\hat{Q}_i} V_U - \mathring{y}_i^\top (d_i \times R_{\hat{Q}_i} V_U)^\times d_i\right) \\ &\quad + \frac{1}{\hat{r}_i}(\mathring{y}_i - d_i)^\top R_{\hat{Q}_i} V_U + \frac{\alpha_i}{\hat{r}_i}\beta^\epsilon_{\underline{c}}(\hat{r}_i),\end{aligned} \tag{4.18}$$

where k_i and α_i are constant positive scalars. The robot pose correction term $\Delta \in \mathbf{SE}(3)$ can be chosen to be any continuous function of the observer state and measurements (*cf.* §4.4.3).

Theorem 4.4.1. *Consider the observer $\hat{X} \in \mathbf{SLAM}_n^V(3)$ with kinematics* (4.14)*. Assume that U is bounded, and that $y_i^\times y_i^\times V_U$ is persistently exciting, in the sense that there exist $T > 0$ and $\mu > 0$ such that*

$$\frac{1}{T}\int_t^{t+T} \|y_i^\times y_i^\times V_U\| \mathrm{d}\tau \geq \mu, \tag{4.19}$$

for each i and all $t > 0$. Assume that there exist bounds $0 < \underline{r} < \overline{r} \in \mathbb{R}$ such that

$$\underline{r} \leq ||p_i - x_P|| \leq \overline{r}$$

for all time.

Then the landmark correction terms (4.18) *define an almost semi-globally asymptotically stabilising (Def. 3.3) correction term for the error dynamics of $d_i, \tilde{r}_i$* (4.15,4.16) *around the equilibrium $(\mathring{y}_i, 1)$ with exception set*

$$\chi = \{(d, \tilde{r})_i \in (\mathrm{S}^2 \times \mathbb{R}_+)^n \mid d_i = -\mathring{y}_i \text{ for some } i \in [1, \ldots, n]\}. \tag{4.20}$$

Moreover, as $(d_i, \tilde{r}_i) \to (\mathring{y}_i, 1)$, the estimated state $\hat{\xi} \to \xi$ converges to the true state up to the SLAM manifold equivalence.

Proof. The outline of the proof is as follows. We begin by choosing the parameters of the correction terms to depend on the initial conditions. Next, it is shown that the observer equations are well-defined for all time. We introduce storage functions in (4.24) and proceed to show that they are non-increasing over time in (4.25). Then we apply Barbalat's lemma along with persistence of excitation to show that the storage functions converge to zero over time, leading to (4.26) and (4.27). Finally, we show that convergence of the error dynamics is equivalent to convergence of the true and estimated states on the SLAM manifold.

Let K be a compact set in the complement of χ (4.20). Then there exists $\tilde{r}^m > 0$ such that $(d_i, \tilde{r}_i) \in K$ implies that $\tilde{r}_i \geq \tilde{r}^m$. Choose $k_0 = \frac{1}{2}\tilde{r}^m$ and assume that $(d_i(0), \tilde{r}_i(0)) \in$

K. Then

$$\tilde{r}_i(0) \geq k_0, \qquad \hat{r}_i(0) \geq k_0 r_i(0) \geq k_0 \underline{r} > 0,$$

for each i. Choose $\underline{c} = \underline{r}$ and $\epsilon = \min(k_0, \frac{1}{2})\underline{r}$. Then $\hat{r}_i(0) > \epsilon > 0$ for every i. By continuity of the solutions there exist $T_i > 0$ such that $\hat{r}_i(t) > \epsilon$ and $1 + d_i^\top \mathring{y}_i > 0$, and hence $\beta^\epsilon_{\underline{r}}(\hat{r}_i)$ and Γ_i are well-defined, for $t \in [0, T_i)$. (We will show later that T_i can be chosen arbitrarily large, and that $\hat{r}_i(t) > \epsilon$ and $1 + d_i^\top \mathring{y}_i > 0$ hold for all time.)

By definition, $r_i = a_{Q_i}^{-1}\mathring{r}_i$, and $\hat{r}_i = a_{\hat{Q}_i}^{-1}\mathring{r}_i$. Differentiating these with respect to time yields

$$\begin{aligned}
\dot{r}_i &= -V_U^\top y_i, \\
&= -d_i^\top R_{\hat{Q}_i} V_U, \\
\dot{\hat{r}}_i &= -V_U^\top \hat{y}_i + \hat{r}_i \gamma_i, \\
&= -V_U^\top y_i + \alpha_i \beta^\epsilon_{\underline{r}}(\hat{r}_i) + \frac{\alpha_i}{\hat{r}_i}\left((1 - d_i^\top \mathring{y}_i) d_i^\top R_{\hat{Q}_i} V_U - \mathring{y}_i^\top (d_i \times R_{\hat{Q}_i} V_U)^\times d_i\right),
\end{aligned} \tag{4.21}$$

where $\hat{y}_i = h^i(\hat{\xi})$ is the estimated measurement. Since $\beta^\epsilon_{\underline{r}}$ (4.17) is a barrier function ensuring that $\hat{r}_i(t) > \epsilon$ ($\beta^\epsilon_{\underline{r}}(\hat{r}_i) \to \infty$ as $\hat{r}_i \searrow \epsilon$) and the remaining terms in (4.21) are bounded, it follows that $\hat{r}_i$ is well defined $\forall t \in [0, T_i)$.

Differentiating the output error d_i yields

$$\begin{aligned}
\dot{d}_i &= \frac{\mathrm{d}}{\mathrm{d}t}\rho(X\hat{X}^{-1}, \mathring{y})_i = \frac{\mathrm{d}}{\mathrm{d}t} R_{\hat{Q}_i} R_{Q_i}^\top \mathring{y}_i, \\
&= \left(R_{\hat{Q}_i}\Lambda_R(U, R_{\hat{P}}^\top(\hat{p}_i - x_{\hat{P}})) - \Gamma_i R_{\hat{Q}_i}\right) R_{Q_i}^\top \mathring{y}_i - R_{\hat{Q}_i}\Lambda_R(U, R_P^\top(p_i - x_P)) R_{Q_i}^\top \mathring{y}_i.
\end{aligned}$$

Observe that $R_{\hat{P}}^\top(\hat{p}_i - x_{\hat{P}}) = \hat{r}_i \hat{y}_i$ and $R_P^\top(p_i - x_P) = r_i y_i$. Using this, the derivative of d_i is simplified to

$$\begin{aligned}
\dot{d}_i &= \mathrm{Ad}_{R_{\hat{Q}_i}}\left(\Lambda_R(U, \hat{r}_i \hat{y}_i) - \Lambda_R(U, r_i y_i)\right) d_i - \Gamma_i d_i, \\
&= \mathrm{Ad}_{R_{\hat{Q}_i}}\left(\frac{\hat{y}_i \times V_U}{\hat{r}_i} - \frac{y_i \times V_U}{r_i}\right)^\times d_i - \Gamma_i d_i, \\
&= \left(\frac{\mathring{y}_i^\times R_{\hat{Q}_i} V_U}{\hat{r}_i} - \frac{d_i^\times R_{\hat{Q}_i} V_U}{r_i}\right)^\times d_i - \Gamma_i d_i,
\end{aligned} \tag{4.22}$$

where the last step is obtained by using various identities involving the skew symmetric operator.

Recalling (4.22) and using identities involving the skew symmetric operator and pro-

jector, the derivative of $1 - \mathring{y}_i^\top d_i$ is

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\left(1 - \mathring{y}_i^\top d_i\right) = -\mathring{y}_i^\top \dot{d}_i, \\
&= -\mathring{y}_i^\top \left(\frac{\mathring{y}_i{}^\times R_{\hat{Q}_i} V_U}{\hat{r}_i} - \frac{d_i^\times R_{\hat{Q}_i} V_U}{r_i}\right)^\times d_i + \mathring{y}_i^\top \Gamma_i d_i, \\
&= -\mathring{y}_i^\top \left(\frac{d^\times R_{\hat{Q}_i} V_U}{\hat{r}_i} - \frac{d_i^\times R_{\hat{Q}_i} V_U}{r_i}\right)^\times d_i + \left(\frac{d_i^\top R_{\hat{Q}_i} V_U}{\hat{r}_i(1 + d_i^\top \mathring{y}_i)} - \frac{k_i}{(1 + d_i^\top \mathring{y}_i)^2}\right) \mathring{y}_i^\top (d_i^\times \mathring{y}_i)^\times d_i, \\
&= (r_i^{-1} - \hat{r}_i{}^{-1}) \mathring{y}_i^\top \left(d^\times R_{\hat{Q}_i} V_U\right)^\times d_i + \left(\frac{d_i^\top R_{\hat{Q}_i} V_U}{\hat{r}_i(1 + d_i^\top \mathring{y}_i)} - \frac{k_i}{(1 + d_i^\top \mathring{y}_i)^2}\right) \mathring{y}_i{}^\top \Pi_{d_i} \mathring{y}_i, \\
&= (r_i^{-1} - \hat{r}_i{}^{-1}) \mathring{y}_i^\top \left(d^\times R_{\hat{Q}_i} V_U\right)^\times d_i + \left(\frac{d_i^\top R_{\hat{Q}_i} V_U}{\hat{r}_i} - \frac{k_i}{1 + d_i^\top \mathring{y}_i}\right)(1 - d_i^\top \mathring{y}_i). \qquad (4.23)
\end{aligned}$$

Observe that $k_i \frac{1 - d_i^\top \mathring{y}_i}{1 + d_i^\top \mathring{y}_i} \to \infty$ as $1 - d_i^\top \mathring{y}_i \nearrow 2$. Since all other terms in (4.23) are bounded, it follows that $1 - d_i^\top \mathring{y}_i < 2 - \nu$ for some small $\nu > 0$, and hence $1 + d_i^\top \mathring{y}_i > \nu > 0$ and Γ_i is well-defined and bounded for all $t \in [0, T_i)$.

We show by contradiction that the domain of definition $[0, T_i)$ can be extended arbitrarily. Suppose, for some i, that T_i' is the largest value such that Γ_i and $\beta_{\underline{r}}^\epsilon(\hat{r}_i)$ are well-defined. Then Γ_i and $\beta_{\underline{r}}^\epsilon(\hat{r}_i)$ are both bounded on $[0, T_i')$ by the arguments above, and continuous. It follows that their limits as $t \to T_i'$ exist and are finite. But then, by continuity of solutions, Γ_i and $\beta_{\underline{r}}^\epsilon(\hat{r}_i)$ can be extended to $[0, T_i' + t_\Delta)$ for some sufficiently small $t_\Delta > 0$. This contradicts the assumption that T_i' is the maximum value for which Γ_i and $\beta_{\underline{r}}^\epsilon(\hat{r}_i)$ are well-defined on $[0, T_i')$, and therefore no such maximum value can exist. Hence Γ_i and $\beta_{\underline{r}}^\epsilon(\hat{r}_i)$ are well-defined on $[0, \infty)$, and so are the observer dynamics.

For each i, define the storage function

$$\begin{aligned}
l_i(d_i, \tilde{r}_i; r_i) &:= \frac{r_i}{2}\|\mathring{y}_i - d_i\|^2 + \frac{r_i^2}{2\alpha_i}(1 - \tilde{r}_i)^2 \\
&= r_i(1 - \mathring{y}_i^\top d_i) + \frac{1}{2\alpha_i}(r_i - \hat{r}_i)^2, \qquad (4.24)
\end{aligned}$$

in the variables $(d_i, \tilde{r}_i)$ where the second line follows from $\|d_i\| = \|\mathring{y}_i\| = 1$ and substituting for $\tilde{r}_i$. Note that l_i depends on the time varying range r_i as a parameter. By assumption, $r_i \geq \underline{r} > 0$ and the storage functions l_i are positive definite.

The derivative of l_i may now be computed as follows:

$$\begin{aligned}
\dot{l}_i &= \dot{r}_i(1 - \mathring{y}_i^\top d_i) + r_i \frac{\mathrm{d}}{\mathrm{d}t}(1 - \mathring{y}_i^\top d_i) + \frac{1}{\alpha_i}(r_i - \hat{r}_i)(\dot{r}_i - \dot{\hat{r}}_i), \\
&= -(1 - \mathring{y}_i^\top d_i) d_i^\top R_{\hat{Q}_i} V_U + r_i(r_i^{-1} - \hat{r}_i^{-1}) \mathring{y}_i^\top \left(d^\times R_{\hat{Q}_i} V_U \right)^\times d_i \\
&\quad + r_i \left(\frac{d_i^\top R_{\hat{Q}_i} V_U}{\hat{r}_i} - \frac{k_i}{1 + d_i^\top \mathring{y}_i} \right) (1 - d_i^\top \mathring{y}_i) + (\hat{r}_i - r_i) \beta_{\underline{r}}^{\epsilon}(\hat{r}_i) \\
&\quad + (\hat{r}_i - r_i) \hat{r}_i^{-1} (1 - d_i^\top \mathring{y}_i) d_i^\top R_{\hat{Q}_i} V_U - (\hat{r}_i - r_i) \hat{r}_i^{-1} \mathring{y}_i^\top (d_i \times R_{\hat{Q}_i} V_U)^\times d_i, \\
&= r_i(r_i^{-1} - \hat{r}_i^{-1}) \mathring{y}_i^\top \left(d^\times R_{\hat{Q}_i} V_U \right)^\times d_i - k_i r_i \frac{1 - d_i^\top \mathring{y}_i}{1 + d_i^\top \mathring{y}_i} + (\hat{r}_i - r_i) \beta_{\underline{r}}^{\epsilon}(\hat{r}_i) \\
&\quad - (\hat{r}_i - r_i) \hat{r}_i^{-1} \mathring{y}_i^\top (d_i \times R_{\hat{Q}_i} V_U)^\times d_i, \\
&= -k_i r_i \frac{1 - d_i^\top \mathring{y}_i}{1 + d_i^\top \mathring{y}_i} + (\hat{r}_i - r_i) \beta_{\underline{r}}^{\epsilon}(\hat{r}_i).
\end{aligned} \tag{4.25}$$

The second term is negative semi-definite, since it is zero when $\hat{r}_i \geq \underline{r}$ (4.17), and otherwise $\hat{r}_i < \underline{r} \leq r_i$ and $\hat{r}_i - r_i \leq 0$. It follows that $\dot{l}_i(t) \leq 0$, and $l_i(t) \leq l_i(0)$ for all t.

Barbalat's Lemma [128, Lemma 4.2] is used to prove $\dot{l}_i \to 0$. To show that $\dot{l}_i$ is uniformly continuous, it is sufficient to show that $\ddot{l}_i$ is bounded. Recall that r_i is bounded above and below by assumption, and that $\hat{r}_i(t) > \epsilon$ for all time. Computing the second derivative of l_i, one has

$$\ddot{l}_i = -k_i \dot{r}_i \frac{1 - \mathring{y}_i{}^\top d_i}{1 + \mathring{y}_i{}^\top d_i} - \frac{2 r_i \mathring{y}_i{}^\top \dot{d}_i}{(1 + \mathring{y}_i{}^\top d_i)^2} + (\hat{r}_i - r_i) \frac{\partial \beta_{\underline{r}}^{\epsilon}}{\partial \hat{r}_i} \dot{\hat{r}}_i.$$

It suffices to show that the component terms are all bounded. In the first and second terms, $1/(1 + \mathring{y}_i{}^\top d_i)$ is bounded from the above discussion. The components $r_i \mathring{y}_i{}^\top \dot{d}_i$ and $\dot{r}_i(1 - \mathring{y}_i{}^\top d_i)$ are also bounded due to the boundedness of $\hat{r}_i$, r_i, and the assumption that U is bounded. As for the third term, the component $\dot{\hat{r}}_i$ is bounded since the velocity input V_U is bounded and $\hat{r}_i$ is lower bounded. This means that both $\beta_{\underline{r}}^{\epsilon}(\hat{r}_i)$ and its derivative with respect to $\hat{r}_i$ are bounded. Therefore, $\dot{l}_i$ is uniformly continuous, and by Barbalat's lemma, $\dot{l}_i \to 0$. This implies that $d_i \to \mathring{y}_i$, since both terms that appear in $\dot{l}_i$ (4.25) are non-positive .

It remains to show $\tilde{r}_i \to 1$ (or equivalently $\hat{r}_i \to r_i$). Applying Barbalat's Lemma to $d_i(t)$ and exploiting the same bounding arguments as before, it is straightforward to verify that $\dot{d}_i \to 0$. It is also easily verified that $\Gamma_i \to 0$ as $d_i \to \mathring{y}_i$.

From (4.22), $\dot{d}_i$ may be written as

$$\begin{aligned}
\dot{d}_i &= \left(\frac{\mathring{y}_i{}^{\times} R_{\hat{Q}_i} V_U}{\hat{r}_i} - \frac{d_i^{\times} R_{\hat{Q}_i} V_U}{r_i} \right)^{\times} d_i - \Gamma_i d_i, \\
&= \frac{d_i^{\times} d_i^{\times} R_{\hat{Q}_i} V_U}{r_i} - \frac{d_i^{\times} \mathring{y}_i{}^{\times} R_{\hat{Q}_i} V_U}{\hat{r}_i} - \Gamma_i d_i, \\
&= (r_i^{-1} - \hat{r}_i^{-1}) d_i^{\times} d_i^{\times} R_{\hat{Q}_i} V_U + \frac{d_i^{\times} (d_i - \mathring{y}_i)^{\times} R_{\hat{Q}_i} V_U}{\hat{r}_i} - \Gamma_i d_i, \\
&= (r_i^{-1} - \hat{r}_i^{-1}) R_{\hat{Q}_i} y_i^{\times} y_i^{\times} V_U + \frac{d_i^{\times} (d_i - \mathring{y}_i)^{\times} R_{\hat{Q}_i} V_U}{\hat{r}_i} - \Gamma_i d_i.
\end{aligned}$$

Clearly, $\hat{r}_i^{-1} d_i^{\times} (d_i - \mathring{y}_i)^{\times} R_{\hat{Q}_i} V_U \to 0$ as $d_i \to \mathring{y}_i$. Hence

$$\dot{d} \to (r_i^{-1} - \hat{r}_i^{-1}) \left(R_{\hat{Q}_i} y_i{}^{\times} y_i{}^{\times} V_U \right), \tag{4.26}$$

as $d_i \to \mathring{y}_i$. Therefore, $(\hat{r}_i - r_i) \frac{1}{\hat{r}_i r_i} \| y_i{}^{\times} y_i{}^{\times} V_U \| \to 0$. Recall that $r_i \leq \overline{r}$. The estimated range $\hat{r}_i$ is also bounded above by a constant $\overline{\hat{r}}_i$ depending on the initial value of $l_i(0)$. Since $|r_i - \hat{r}_i| < (\overline{r}\overline{\hat{r}}_i) |r_i^{-1} - \hat{r}_i^{-1}|$, it follows from (4.26) that

$$(r_i - \hat{r}_i) \| y_i{}^{\times} y_i{}^{\times} V_U \| \to 0. \tag{4.27}$$

Each l_i must converge to a positive constant $c_i^0 \leq l_i(0)$ as $\dot{l}_i \to 0$. Therefore, (4.24) ensures that $(\hat{r}_i - r_i) \to \pm \sqrt{2 \alpha_i c_i^0}$. Integrating (4.27) over a period of time T, and using the fact that $(r_i - \hat{r}_i)$ is converging to a constant, it follows that

$$(r_i - \hat{r}_i) \int_t^{t+T} \| y_i{}^{\times} y_i{}^{\times} V_U \| d\tau \to 0.$$

Using the persistence of excitation assumption (4.19), it must be that $\hat{r}_i \to r_i$.

Recall the exception set χ (4.20). It is straightforward to verify that this set has measure zero. Since the initial choice of compact set K in the complement of χ was arbitrary, then the equilibrium $(\mathring{y}_i, 1)$ of $(d_i, \tilde{r}_i)$ is almost semi-globally asymptotically stabilisable by the proposed correction terms (Def. 3.3).

Observe that, at the equilibrium,

$$\begin{aligned}
\hat{y}_i = \rho(\hat{X}, \mathring{y}_i) = \rho(\hat{X}, d_i) = y_i, \\
\hat{r}_i = \tilde{r}_i r_i = r_i.
\end{aligned}$$

Using this, the ego-centric coordinates of the SLAM configuration and their estimates satisfy

$$R_P^{\top} (p_i - x_P) = r_i y_i = \hat{r}_i \hat{y}_i = R_{\hat{P}}^{\top} (\hat{p}_i - x_{\hat{P}}),$$

and thus

$$p_i = (R_{\hat{P}P^{-1}})^\top(\hat{p}_i - x_{\hat{P}P^{-1}}).$$

Therefore, at the equilibrium point, each p_i is related to each $\hat{p}_i$ by the same rigid body transformation $S = \hat{P}P^{-1}$. Moreover, it is clear that $P = S^{-1}\hat{P}$. Hence, the two configurations on $\mathcal{T}_n^{\mathrm{V}}(3)$ are equivalent on the SLAM manifold, $\xi \simeq \hat{\xi}$. This completes the proof.

4.4.3 Total Space Representative

In Theorem 4.4.1 the convergence result is independent of the choice of correction term Δ. This is due to the ego-centric nature of the group action considered and the invariance properties of the SLAM manifold, which cause the inertial frame of a SLAM system to be unobservable. In essence, choosing Δ will influence the element $S \in \mathbf{SE}(3)$ that relates the reference and estimated states, but does not influence the convergence of the SLAM error. Nevertheless, it is clear that as the error converges, it is desirable that the $S(t)$ that relates the reference and established states converges to a constant, essentially capturing the "inertial map" property that is desired in visual odometry. It is a key contribution of this chapter to observe that imposing this constraint is a separate requirement from the underlying SLAM solution, that is, we must introduce an additional criterion that captures this property and then use this to design the correction term Δ.

The criterion that we propose to minimize is the weighted mean velocity of the landmark points

$$\sum_{i=1}^{n} \kappa_i \|\dot{\hat{p}}_i\|^2$$

For a static environment, the true landmark points are not moving. For the observer estimate these points may be moving, due to residue velocity associated with the landmark error correction, but also importantly, due to the correction term Δ that is moving the entire SLAM configuration. The motion due to Δ will be strongly correlated, while it is expected that the residue velocity due to the correction terms will be uncorrelated and for large constellations of points average to zero. Choosing Δ to minimize this additional criteria can be thought of as minimizing instantaneous map drift.

Proposition 4.4.2. *Let the origin configuration $\mathring{\xi} = (\mathring{P}, \mathring{p}) \in \mathcal{T}_n^V(3)$, the observer state $\hat{X} = (\hat{A}, \hat{Q}_i) \in \mathbf{SLAM}_n^V(3)$, and the correction term $\Delta_{\hat{X}} = (\Delta, (\Gamma, \gamma)_i)$ be defined as in the statement of Theorem 4.4.1. Let $\hat{\xi} = (\hat{P}, \hat{p}_i) = \Upsilon(\hat{X}, \mathring{\xi}) \in \mathcal{T}_n^V(3)$ be the estimated state on the total space, and let $\hat{q}_i = R_{\hat{P}}^\top(\hat{p}_i - x_{\hat{P}})$ for each i. Then the solution to*

$$\Delta = \underset{\Delta \in \mathfrak{se}(3)}{\operatorname{argmin}} \left\{ \sum_{i=1}^{n} \kappa_i \|\dot{\hat{p}}_i\|^2 \right\}, \tag{4.28}$$

where κ_i are positive scalars, is given by

$$\Delta = \mathrm{Ad}_{\hat{A}}(\Omega_\Delta^\times, V_\Delta), \tag{4.29}$$

where Ω_Δ and V_Δ are determined by

$$\begin{pmatrix}\Omega_\Delta \\ V_\Delta\end{pmatrix} = -\left(\sum_{i=1}^{n}\kappa_i \begin{pmatrix}\hat{q}_i^\times \hat{q}_i^\times & \hat{q}_i^\times \\ -\hat{q}_i^\times & I_3\end{pmatrix}\right)^{-1} \tag{4.30}$$

$$\left(\sum_{i=1}^{n}\kappa_i \begin{pmatrix}\hat{q}_i^\times \,\mathrm{Ad}_{R_{\hat{Q}_i}^\top}(\Gamma_i)\hat{q}_i \\ \gamma_i \hat{q}_i + \mathrm{Ad}_{R_{\hat{Q}_i}^\top}(\Gamma_i)\hat{q}_i\end{pmatrix}\right), \tag{4.31}$$

so long as the inverse in (4.30) *remains well-defined.*

Proof. First, observe that

$$\begin{aligned}\hat{q}_i &= R_{\hat{P}}^\top(\hat{p}_i - x_{\hat{P}}),\\ &= R_{\hat{P}}^\top(a_{\hat{Q}_i}^{-1} R_{\hat{P}} R_{\hat{Q}_i}^\top R_{\mathring{P}}^\top(\mathring{p}_i - x_{\mathring{P}}) + x_{\hat{P}} - x_{\hat{P}}),\\ &= a_{\hat{Q}_i}^{-1} R_{\hat{Q}_i}^\top R_{\mathring{P}}^\top(\mathring{p}_i - x_{\mathring{P}}).\end{aligned}$$

Equation (4.28) presents a weighted least squares problem, and to solve it we analyse the component expressions $\kappa_i \dot{\hat{p}}_i$. The time derivative of each $\hat{p}_i$ needs to be computed. Recall that the velocity lift Λ is defined precisely so that $\dot{\hat{p}}_i = 0$ when the correction terms are set to zero. Since differentiation is a linear operation, this means that

$$\begin{aligned}\dot{\hat{p}}_i &= \frac{\mathrm{d}}{\mathrm{d}t}\left(R_{\mathring{P}\hat{A}} a_{\hat{Q}_i}^{-1} R_{\hat{Q}_i}^\top R_{\mathring{P}}^\top(\mathring{p}_i - x_{\mathring{P}}) + x_{\mathring{P}\hat{A}}\right)\\ &= -R_{\mathring{P}\hat{A}}\Omega_\Delta^\times a_{\hat{Q}_i}^{-1} R_{\hat{Q}_i}^\top R_{\mathring{P}}^\top(\mathring{p}_i - x_{\mathring{P}}) - R_{\mathring{P}\hat{A}} V_\Delta\\ &\qquad + \gamma_i R_{\mathring{P}\hat{A}} a_{\hat{Q}_i}^{-1} R_{\hat{Q}_i}^\top R_{\mathring{P}}^\top(\mathring{p}_i - x_{\mathring{P}}) + R_{\mathring{P}\hat{A}} a_{\hat{Q}_i}^{-1} R_{\hat{Q}_i}^\top \Gamma_i R_{\mathring{P}}^\top(\mathring{p}_i - x_{\mathring{P}}),\\ &= -R_{\mathring{P}\hat{A}}\Omega_\Delta^\times \hat{q}_i - R_{\mathring{P}\hat{A}} V_\Delta + \gamma_i R_{\mathring{P}\hat{A}}\hat{q}_i + R_{\mathring{P}\hat{A}}\,\mathrm{Ad}_{R_{\hat{Q}_i}^\top}(\Gamma_i)\hat{q}_i,\\ \|\dot{\hat{p}}_i\| &= \| - \Omega_\Delta^\times \hat{q}_i - V_\Delta + \gamma_i \hat{q}_i + \mathrm{Ad}_{R_{\hat{Q}_i}^\top}(\Gamma_i)\hat{q}_i\|,\\ &= \left\|\begin{pmatrix}-\hat{q}_i^\times & I_3\end{pmatrix}\begin{pmatrix}\Omega_\Delta \\ V_\Delta\end{pmatrix} - \left(\gamma_i \hat{q}_i + \mathrm{Ad}_{R_{\hat{Q}_i}^\top}(\Gamma_i)\hat{q}_i\right)\right\|.\end{aligned}$$

Therefore, by the theory of Weighted Least Squares, (4.30) is exactly the solution to (4.28), as required.

Proposition 4.4.2 provides a clear way to choose a correction term based on the static landmark assumption, and allows for scalars κ_i to be chosen to weight the optimisation. The computational and memory costs of the correction terms scale linearly with the number of landmarks as opposed to alternative observer designs [126; 109] which scale quadratically. The correction term Δ may be chosen in another way, such as by finding the solution of an optimisation problem involving previous measurements and

inputs. However, regardless of the choice of Δ the stability of the observer on the SLAM manifold is guaranteed according to Theorem 4.4.1.

4.5 Simulation Results

To verify the landmark observer design in Theorem 4.4.1 and the robot correction term in Proposition 4.4.2, we conducted a simulation of a flying vehicle equipped with a monocular camera, observing 5 stationary landmarks as it moves in a circular trajectory with a constant body-fixed velocity $U = (\Omega_U^\times, V_U)$, where $\Omega_U = (0, 0, 0.5)$rad/s and $V_U = (1.5, 0, 0)$m/s. For simplicity, it is assumed that the camera frame coincides with the body-fixed frame of the vehicle. The initial position of the vehicle was set to $(3, 3, 5)$m with its rotational axes aligned with the inertial frame $\{0\}$. The positions of the landmarks were initialised to random positions $(p_i^1, p_i^2, 0)$ on the ground plane with $p_i^1, p_i^2 \sim N(0, 5^2)$m.

The origin position $\mathring{P}$ of the robot is set to the identity I_4, and the origin landmark positions $\mathring{p}_i$ are set to $10y_i(0)$, where $y_i(0)$ are the measured bearings to the true landmark positions at time 0. That is, the estimated points are initialised with correct bearings and an arbitrary depth of 10 m. The observer is defined on $\mathbf{VSLAM}_5(3)$ with kinematics given by (4.14), landmark correction terms (Γ_i, γ_i) given by (4.18), where $k_i = 5$ and $\alpha_i = 500$, and robot correction term Δ given as in Proposition 4.4.2 with $\kappa_i = 1$. At the end of the simulation, the estimated system state is aligned with the true system state by matching the true and estimated robot poses. Figure 4.4 shows the trajectories of estimated landmark positions and robot position over time, as well as the true landmark positions and the true robot trajectory, and Figure 4.5 shows the evolution of each of the landmarks' associated storage functions, as defined in (4.24).

This simulation provides a simple demonstration of performance of the proposed observer and illustrates typical trajectories of the landmark estimates during a repeating motion such as the circle. The estimated landmark positions can be seen to converge to the true landmark positions in a natural manner. The choice to initialise landmarks as having a bearing matching the initial measurement is a natural one for practical implementation of the algorithm, although Theorem 4.4.1 provides that almost any initial conditions will converge. This almost semi-global convergence is a key property of the observer presented here that is not available in many of the state-of-the-art solutions.

4.6 Experimental Results

To demonstrate the observer described in Theorem 4.4.1 in a real-world scenario, we gathered video, GPS, and IMU data from a Disco Parrot fixed-wing UAV flying outdoors. Image features were identified using OpenCV's goodFeaturesToTrack, and subsequently tracked using OpenCV's calcOpticalFlowPyrLK. These image features were then corrected for camera intrinsics and converted to spherical bearing coordinates before be-

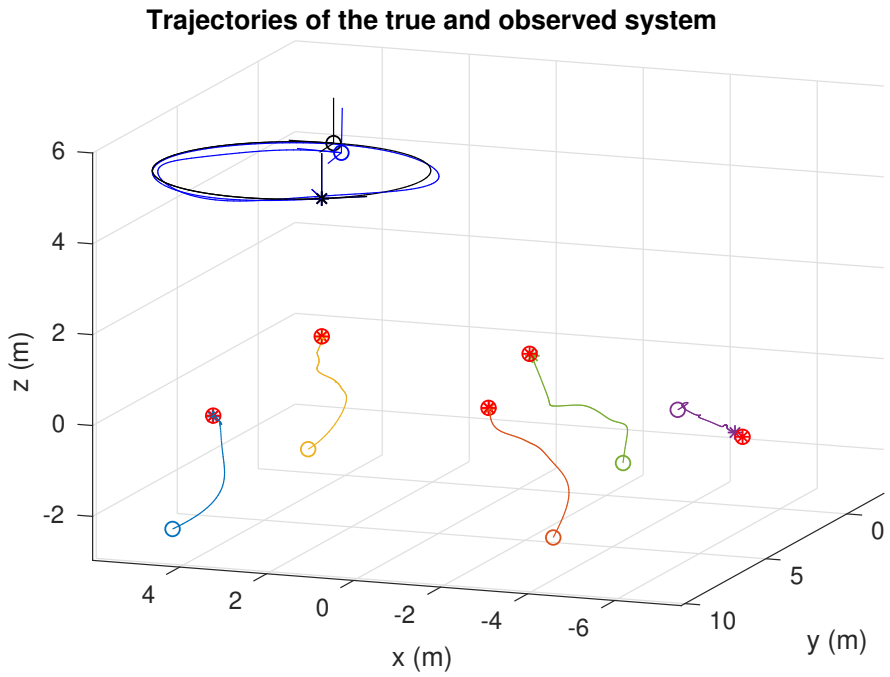


Figure 4.4: The simulated trajectories of observer landmark estimates from initial positions with correct bearings but fixed depth of 10m. The observer landmark trajectories are shown in a range of colours matching those used in Figure 4.5, and the true landmark positions are shown in red. The observer robot trajectory is shown in blue, and the true robot trajectory is shown in black. The ($\circ$) and ($\star$) markers, respectively, denote the start and end of the trajectories of all the objects shown.

ing used as landmark inputs to the observer. Landmarks are added to the system state after being observed for two frames so that their depths can be initialised from optical flow. When a landmark is no longer visible, it is removed from the observer state.

Initially, the input velocities $U = (\Omega_U, V_U)$ to the system were estimated by combining the GPS signal (to obtain scale information) with egomotion estimated using the IMU and optical flow from the video stream, as outlined in [129]. Once sufficiently many landmarks are initialised, the optical flow vectors of each of the landmarks were combined with the existing landmark estimates to compute the input velocity $U = (\Omega_U, V_U)$. The observer was implemented using Euler integration with gain parameters set to $k_i = 5.0$ and $\alpha_i = 0.5$ for each i. The video recorded had a frame rate of 30 fps, leading to the Euler integration step being set to $\mathrm{d}t = 0.033$ s. GPS data was recorded at 25 Hz in order to compare with the observer's estimated trajectory. The observer trajectory was aligned to the GPS trajectory using the Umeyama method [130]. Figure 4.6 shows the aligned trajectories according to the observer and according to the GPS in the x and y directions, where the z direction refers to the plane's altitude. Figure 4.7 shows the final positions of all landmark points in addition to the observer- and GPS-estimated trajectories. Figure 4.8 shows a frame taken from the video stream used in the experiment, with lines to rep-

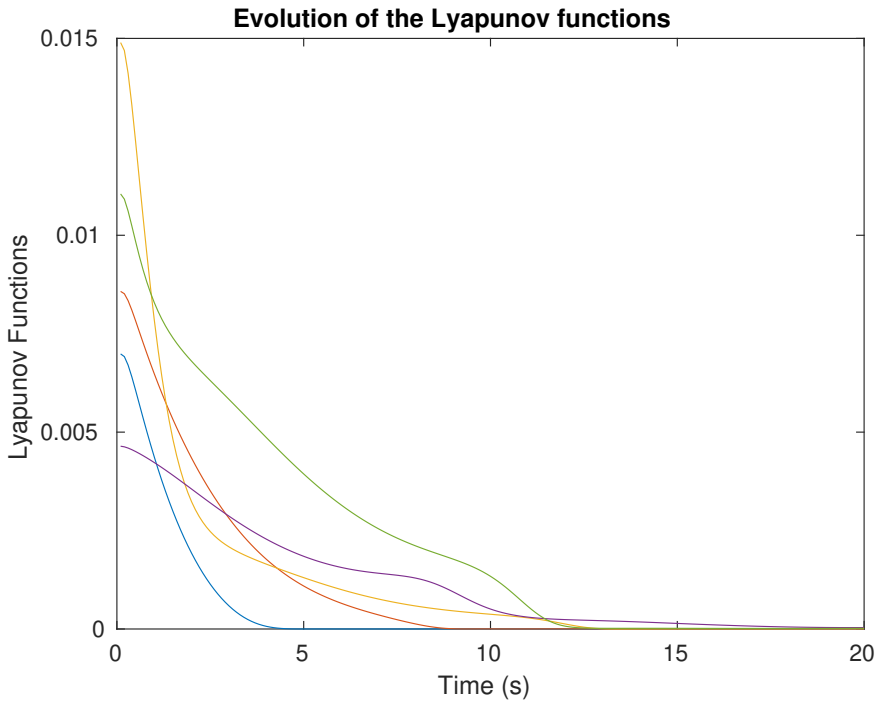


Figure 4.5: The evolution of the storage functions of each of the landmarks in the system shown in Figure 4.4. The colours match the colours of the landmark trajectories in Figure 4.4. The initial convergence of the landmarks is quick as the bearings converge, and then slows as the depths gradually converge.

resent the optical flow tracking overlaid. A video showcasing the feature tracking system is available online[1]. The quality of the trajectory and map provided in Figure 4.7 show the robustness of the observer to noisy bearing measurements in practice.

4.7 Chapter Conclusion

This chapter presents an observer design posed on a symmetry group for the VSLAM problem. The SLAM manifold introduced in [78] and the symmetry group discussed in [5] are reintroduced and exploited in the observer design. The observer is formulated on output errors, and provides a clear way to change the gains for bearing and depth of landmarks separately. The almost semi-global convergence of the proposed observer improves on the properties of state-of-the-art Extended Kalman Filter systems, which suffer from linearisation errors. While research into the development of non-linear observers for the SLAM problem is only recent, the observer for VSLAM presented in this chapter demonstrates some of the key advantages the approach can offer.

[1] https://www.youtube.com/watch?v=QzIxh2eM1_s

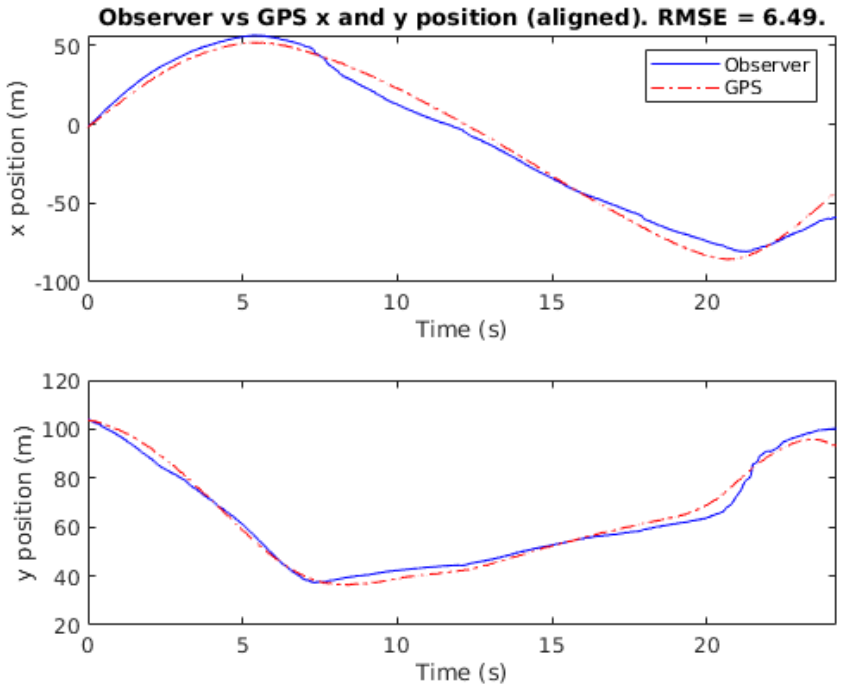


Figure 4.6: The x and y positions of the UAV according to the aligned Observer (blue) and GPS (red).

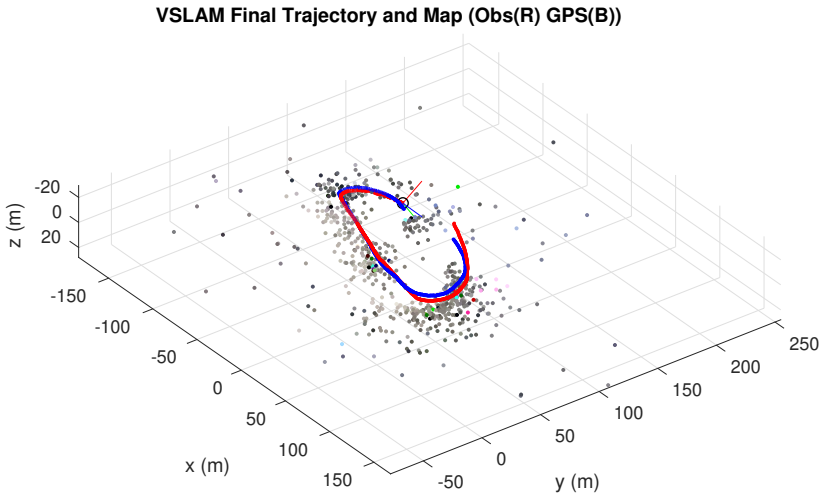


Figure 4.7: The full trajectory of the UAV according to the aligned Observer (blue) and GPS (red), and the final positions of all of the landmarks, coloured with the colour of the pixel where they were first observed.

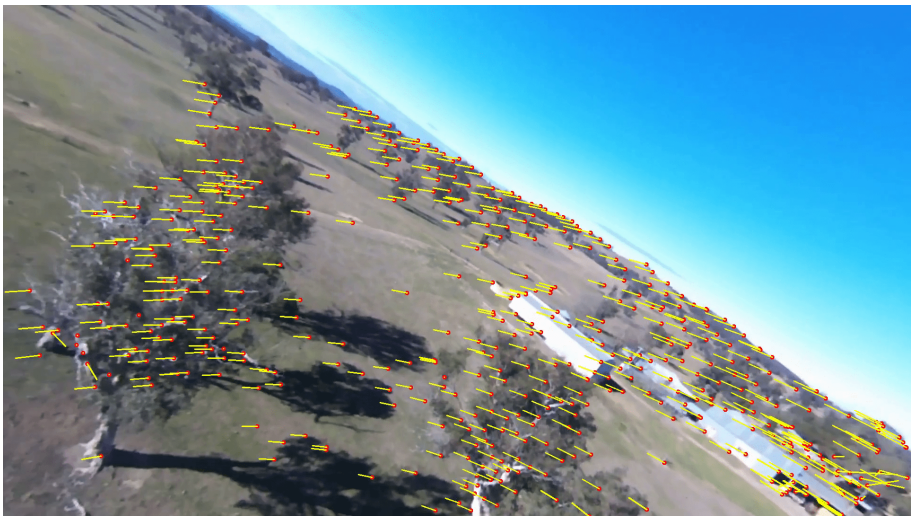


Figure 4.8: A single frame of the video stream used in the experiment. Red circles represent the image features being tracked, and yellow lines represent the vector of motion of the image features between the current frame and the previous frame.

Chapter 5

Equivariant Filter

5.1 Chapter Introduction

The importance of Lie group symmetries in analysing nonlinear control systems has been recognised since the 1970s [56; 81; 82]. In this chapter, we propose the Equivariant Filter (EqF): a novel filter design for equivariant kinematic systems posed on homogeneous spaces. The filter is derived by exploiting the Lie group symmetry of the kinematic system to derive global error coordinates. The EqF observer dynamics are defined on the symmetry Lie group; however, the correction is computed using a Riccati equation associated with linearised error kinematics about a fixed origin on the homogeneous state space. The proposed architecture fully exploits the symmetry properties of the system without requiring that the system model is posed explicitly on the Lie group and applies to any system with just the basic equivariance property. In contrast, the existing state-of-the-art observer/filter design methodologies for systems with symmetry depend on assuming properties of the system; invariance for the constructive designs [46] or the more general group affine structure that is only defined for systems with Lie group state space for the IEKF designs [48]. Interestingly, the proposed EqF specialises to the IEKF [48] when the system considered is posed directly on a Lie group and displays the specific group affine structure required for the IEKF derivation. Barrau *et al.* [48] showed that the IEKF admits an exact linearisation of the deterministic part of the state equation of the error dynamics, a property that the EqF shares on compatible systems, leading to significant performance gains versus an EKF derived without regard for the symmetry. In addition, the EqF is designed to accommodate symmetries that are compatible with the configuration output. When such a symmetry is used, we propose a novel approximation of the output equation that eliminates second-order error in the output linearisation. Implementing the equivariant filter design methodology with this approximation leads to improved performance and we term the resulting observer the EqF*. The EqF methodology extends existing filter design methodologies for invariant systems by both applying to a broad class of systems on homogeneous spaces (rather than only systems with Lie group state space) and exploiting equivariance of the configuration output to further enhance filter performance compared to state-of-the-art.

We demonstrate the potential of the EqF and EqF* using a simple example of single-

bearing estimation; that is, the problem of determining the bearing of a fixed direction (with state space the sphere S^2) in an inertial frame with respect to a rotating frame with known angular velocity. There are no global coordinates on the sphere and application of the EKF requires consideration of local or embedded coordinates. Moreover, the state space is not a Lie group, precluding the direct application of filter design methods that require a Lie group state space. We simulate a classical EKF, the EqF and the EqF* on this system. The simulation results demonstrate the known advantage [48] of exploiting symmetry versus (even a careful) EKF design in local coordinates and goes on to demonstrate a significant performance advantage for the EqF* over the standard EqF. The interested reader can also find a more tutorial exposition of equivariant observer design with a discussion of the equivariant filter in the preprint [4].

In §5.2, the design of an EKF for the example of single-bearing estimation is detailed to motivate the developments in the sequel. General systems on homogeneous spaces are defined and discussed in §5.3. In §5.4, the notion of a lifted system is used to develop the dynamics of a global state error by exploiting the Lie symmetry. We also discuss the linearisation of the error dynamics and show that the existence of an equivariant output leads to a better (lower error) linearisation of the output map. In §5.5, the EqF equations are presented and we provide some insight into tuning the filter in practice. The example problem of single-bearing estimation is revisited in §5.6. We show simulation results to demonstrate the performance of the EqF and EqF* compared to an EKF. In §5.7, we provide concluding remarks. In Appendix A, we provide a step-by-step design methodology for implementing an EqF, and in Appendix B we show how the EqF specialises to the invariant extended Kalman filter for a specific subclass of systems on Lie groups. Open source code[1] is publicly available for the implementation of the EqF for general systems based on numerical differentiation.

5.2 Motivating Example

In this section we present the problem of single bearing estimation. This example has been chosen to be as simple as possible algebraically while presenting a problem where the equivariant filter approach is of interest. Consider a robot equipped with a gyroscope that measures its angular velocity $\Omega \in \mathbb{R}^3$ and a magnetometer that measures the magnetic field in the robot's body-fixed frame $\eta \in \mathrm{S}^2$. The noise free dynamics of η are

$$\dot{\eta} = f_\Omega(\eta) := -\Omega^\times \eta. \tag{5.1}$$

[1] https://github.com/STR-ANU/auto_eqf

where $\Omega^\times \in \mathbb{R}^{3\times 3}$ is the matrix

$$\Omega^\times := \begin{pmatrix} 0 & -\Omega_3 & \Omega_2 \\ \Omega_3 & 0 & -\Omega_1 \\ -\Omega_2 & \Omega_1 & 0 \end{pmatrix}.$$

The noise free measurement model considered is a 3-axis magnetometer

$$y = h(\eta) = c_m \eta \in \mathbb{R}^3, \tag{5.2}$$

where c_m is the (known) magnetic field strength. The system is nonlinear due to the manifold structure of the sphere S^2 comprising both its state and its measurement.

Since the system is nonlinear, a nonlinear filter design is required and the extended Kalman filter is the industry standard choice. To provide context for the main contribution of the chapter we provide a sketch of the derivation of a classical EKF on the sphere for direction estimation. In Section 5.6 we will provide simulation results that compare the classical EKF for this problem to the proposed EqF that is the main contribution of the chapter.

EKF for direction estimation: The EKF requires a system to be written in Euclidean coordinates and one must either choose local coordinates for the sphere or embed the sphere in $\mathbb{R}^3$ and introduce a state constraint in the filter design [131]. The sphere S^2 has many possible choices of local coordinates, some of which cover the space almost globally. However, all choices introduce significant nonlinearities in the system dynamics leading to loss of filter performance. To avoid this, and provide a state-of-the-art EKF implementation, we instead choose to implement the EKF by embedding the system dynamics (5.1) and measurement (5.2) from the sphere into $\mathbb{R}^3$ and incorporating a nonlinear state constraint as in [131]. That is, the extended noise free system in $\mathbb{R}^3$ is

$$\begin{aligned} \dot{\eta} &= -\Omega^\times \eta, \\ h(\eta) &= c_m \frac{\eta}{|\eta|}, \\ g(\eta) &= \|\eta\|^2 = 1, \end{aligned} \tag{5.3}$$

where g is the nonlinear state constraint, and can be treated as an additional measurement. This approach is particularly attractive for direction estimation on the sphere since the embedded dynamics (5.1) on $\mathbb{R}^3$ are linear time-varying. Let $\hat{\eta} \in \mathbb{R}^3$ be the EKF state estimate and define the state error and innovation $\tilde{\eta} := \eta - \hat{\eta} \in \mathbb{R}^3$ and $\tilde{y} := h(\eta) - h(\hat{\eta})$, respectively. The output linearisation is given by

$$\tilde{y} = c_m \left(I_3 - \frac{\hat{\eta}\hat{\eta}^\top}{\hat{\eta}^\top \hat{\eta}} \right) \tilde{\eta} + O(|\tilde{\eta}|^2), \tag{5.4}$$

noting that $\hat{\eta} \in \mathbb{R}^3$. The constraint innovation $\tilde{z} = g(\eta) - g(\hat{\eta})$ has linearisation

$$\tilde{z} = -2\hat{\eta}^\top \tilde{\eta} + O(|\tilde{\eta}|^2). \tag{5.5}$$

The linearised constraint must be assigned a nonzero virtual covariance to avoid the filter from becoming singular, and this choice is a design parameter without stochastic justification.

The proposed EKF described above delivers the best performance for a classical EKF design for the direction estimation problem that the authors are aware of. However, it is clear that applying a classical EKF to this problem is not straightforward. The practitioner must choose between local and embedded coordinates. The former are not globally well-defined and introduce linearisation errors. The latter require the introduction and linearisation of state constraints. The resulting algorithms are not intrinsic in any sense, depending on a whole sequence of design decisions, and although the performance of the filters is accepted, a question always remains about whether a different choice of coordinates, or an embedding with a different state constraint could have improved the results.

5.3 Problem Description

Let $\mathcal{M}$ be a smooth m-dimensional manifold termed the *state space*. An affine (kinematic) system on $\mathcal{M}$ may be written

$$\dot{\xi} = f_0(\xi) + \sum_i f_i(\xi) u_i,$$

for some vector fields $f_0, f_1, ..., f_l \in \mathfrak{X}(\mathcal{M})$ and scalar input signals $u_1, ..., u_l \in \mathbb{R}$. Such a system is represented by an affine system function [4]

$$\begin{aligned} f &: \mathbb{L} \to \mathfrak{X}(\mathcal{M}), \\ u &\mapsto f_u \in \mathfrak{X}(\mathcal{M}), \end{aligned} \tag{5.6}$$

where $f_u(\xi) := f_0(\xi) + \sum_i f_i(\xi) u_i$ and $\mathbb{L}$ is a real vector space termed the *input space* with $u = (u_1, ..., u_l) \in \mathbb{L}$ the components of u. We refer to f_0 as the drift term of the system and f_i as the input vector fields. Trajectories $\xi(t) \in \mathcal{M}$ on a time interval $[0, \infty)$ of the system considered are solutions of the ordinary differential equation

$$\dot{\xi} = f_{u(t)}(\xi), \qquad \xi(0) \in \mathcal{M}, \tag{5.7}$$

with initial condition $\xi(0)$ and measured input signal $u(t) \in \mathbb{L}$. We will assume $u(t)$ is sufficiently smooth to ensure unique well defined solutions for all time. The configuration

output [45] for a kinematic system is a function

$$h : \mathcal{M} \to \mathcal{N} \subset \mathbb{R}^n, \tag{5.8}$$

where $\mathcal{N}$ is a smooth manifold termed the *output space*, embedded in $\mathbb{R}^n$.

5.3.1 Systems on Homogeneous Spaces

Let $\mathbf{G}$ be a Lie group with Lie algebra $\mathfrak{g}$, and suppose that $\mathcal{M}$ is a homogeneous space of $\mathbf{G}$; that is, there exists a smooth, transitive, right group action of $\mathbf{G}$ on $\mathcal{M}$,

$$\phi : \mathbf{G} \times \mathcal{M} \to \mathcal{M}. \tag{5.9}$$

A *lift* [45] for the system function (5.7) is a map $\Lambda : \mathcal{M} \times \mathbb{L} \to \mathfrak{g}$ satisfying

$$\mathrm{D}_X|_{\mathrm{id}}\phi_\xi(X)\,[\Lambda(\xi, u)] = f_u(\xi), \tag{5.10}$$

for every $\xi \in \mathcal{M}$ and $u \in \mathbb{L}$. Any kinematic system (5.7) defined on a homogeneous space admits a lift $\Lambda : \mathcal{M} \times \mathbb{L} \to \mathfrak{g}$ satisfying (5.10) [46; 45]. In the particular case where $\mathrm{D}_X|_{\mathrm{id}}\phi_\xi(X)$ is invertible (ϕ is a free group action [132]), the lift is unique [45].

5.3.2 Equivariant Systems

Equivariance of a system is a powerful structural property that can be formulated for any system on a homogeneous space. There are many established examples of equivariant systems [28; 59; 133; 42; 47; 48; 134; 1] where exploiting the equivariant structure has led to high performance observers and filters.

A kinematic system (5.7) is termed *equivariant* if there exists a smooth right group action $\psi : \mathbf{G} \times \mathbb{L} \to \mathbb{L}$ such that

$$\mathrm{D}\phi_X f_u(\xi) = f_{\psi_X(u)}(\phi_X(\xi)),$$

for all $u \in \mathbb{L}$, $\xi \in \mathcal{M}$ and $X \in \mathbf{G}$. In this case, one has

$$\begin{aligned} f_{\psi_X(u)}(\xi) &= f_{\psi_X(u)}(\phi_X(\phi_X^{-1}(\xi))), \\ &= \mathrm{D}\phi_X f_u(\phi_X^{-1}(\xi)), \\ &= \Phi_X f_u(\xi). \end{aligned} \tag{5.11}$$

That is, $f_{\psi_X(u)} = \Phi_X f_u$ as vector fields on $\mathcal{M}$, or equivalently, the diagram

$$\begin{array}{ccc} \mathbb{L} & \xrightarrow[\psi_X]{} & \mathbb{L} \\ {\scriptstyle f}\downarrow & & \downarrow{\scriptstyle f} \\ \mathfrak{X}(\mathcal{M}) & \xrightarrow[\Phi_X]{} & \mathfrak{X}(\mathcal{M}) \end{array}$$

commutes for every $X \in \mathbf{G}$.

Suppose the system (5.7) is equivariant with state symmetry $\phi : \mathbf{G} \times \mathcal{M} \to \mathcal{M}$ and input symmetry $\psi : \mathbf{G} \times \mathbb{L} \to \mathbb{L}$. A lift Λ for the system is equivariant if

$$\Lambda(\phi(X, \xi), \psi(X, u)) = \mathrm{Ad}_{X^{-1}} \Lambda(\xi, u), \tag{5.12}$$

for all $X \in \mathbf{G}$, $\xi \in \mathcal{M}$, and $u \in \mathbb{L}$, or equivalently, the diagram

$$\begin{array}{ccc} \mathfrak{g} & \xrightarrow[\mathrm{Ad}_{X^{-1}}]{} & \mathfrak{g} \\ \uparrow{\scriptstyle \Lambda} & & \uparrow{\scriptstyle \Lambda} \\ \mathcal{M} \times \mathbb{L} & \xrightarrow[\phi_X \times \psi_X]{} & \mathcal{M} \times \mathbb{L} \end{array}$$

commutes for every $X \in \mathbf{G}$.

Remark 5.3.1. *On a matrix Lie group where the group action is right translation, the lift is just* $\Lambda(X, u) = X^{-1}\dot{X} = X^{-1} f_u(X) \in \mathfrak{g}$. *However, for a general system on a homogeneous space there may be many choices of equivariant lift. In such a case, an equivariant lift can be found by expanding the conditions* (5.10) *and* (5.12) *for the particular system, and searching for a solution for* Λ *[45].*

The system (5.6) has *equivariant output* if there exists an action $\rho : \mathbf{G} \times \mathcal{N} \to \mathcal{N}$ satisfying

$$\rho(X, h(\xi)) = h(\phi(X, \xi)), \tag{5.13}$$

for all $X \in \mathbf{G}$ and ξ, or equivalently, the diagram

$$\begin{array}{ccc} \mathcal{M} & \xrightarrow[\phi_X]{} & \mathcal{M} \\ {\scriptstyle h}\downarrow & & \downarrow{\scriptstyle h} \\ \mathcal{N} & \xrightarrow[\rho_X]{} & \mathcal{N} \end{array}$$

commutes for every $X \in \mathbf{G}$.

5.4 Equivariant Error System

The filter design proposed in this chapter can be applied to any equivariant system where an equivariant lift can be found. In [45] the authors showed that any system on a homogeneous space can be extended to an equivariant system, and for any equivariant system an equivariant lift can always be constructed although the resulting construction may be infinite dimensional in the input space. Thus, in principle, the proposed EqF design applies to all systems on homogeneous spaces, although the authors note that the primary systems of interest are those which are equivariant directly, or for which the extension terminates in a finite dimensional input space.

5.4.1 Observer Architecture

Consider an equivariant kinematic system (5.7) with state symmetry $\phi : \mathbf{G} \times \mathcal{M} \to \mathcal{M}$ and input symmetry $\psi : \mathbf{G} \times \mathbb{L} \to \mathbb{L}$. Let $\Lambda : \mathcal{M} \times \mathbb{L} \to \mathfrak{g}$ be an equivariant lift for this system. Given a fixed but arbitrary $\mathring{\xi} \in \mathcal{M}$ and a known input signal $u(t) \in \mathbb{L}$, the *lifted system* [45] is defined by the ODE

$$\dot{X} = \mathrm{D}L_X \Lambda(\phi(X, \mathring{\xi}), u), \quad \phi(X(0), \mathring{\xi}) = \xi(0), \tag{5.14}$$

where $X(t) \in \mathbf{G}$. The trajectory of the lifted system projects down to the original system trajectory [45] by

$$\phi(X(t), \mathring{\xi}) \equiv \xi(t). \tag{5.15}$$

Define the observer state to be an element of the group $\hat{X} \in \mathbf{G}$, and use the lifted system as the internal model for the observer dynamics

$$\dot{\hat{X}} = \mathrm{D}L_{\hat{X}} \Lambda(\phi(\hat{X}, \mathring{\xi}), u) + \mathrm{D}R_{\hat{X}} \Delta, \quad \hat{X}(0) = \mathrm{id}, \tag{5.16}$$

where the correction term Δ remains to be chosen [45; 4].

The state estimate of the observer is given by the projection

$$\hat{\xi} = \phi(\hat{X}(t), \mathring{\xi}).$$

Thus, the state of the observer is posed on the symmetry group rather than the state space of the system kinematics. The correction term Δ will be chosen by applying a Riccati observer to global error dynamics linearised about the fixed origin $\mathring{\xi}$.

5.4.2 Global Error Dynamics

Let $\xi \in \mathcal{M}$ be the true state of the system. Choose an arbitrary fixed origin $\mathring{\xi} \in \mathcal{M}$ and let $\hat{X} \in \mathbf{G}$ be a state observer with dynamics given by (5.16). Define the *global state error*

$$e := \phi(\hat{X}^{-1}, \xi). \tag{5.17}$$

Note that $\phi(\hat{X}, \mathring{\xi}) = \xi$ if and only if $e = \mathring{\xi}$. Therefore, the goal of the filter design will be to drive $e \to \mathring{\xi}$. Define the *origin velocity*

$$\mathring{u} := \psi(\hat{X}^{-1}, u). \tag{5.18}$$

Note that the origin velocity $\mathring{u}(t)$ can always be constructed since both $\hat{X}(t)$ and $u(t)$ are available to the observer. The action $\psi(\hat{X}^{-1}, u)$ is the equivariant system generalisation of the well known adjoint action $\mathrm{Ad}_{\hat{X}}\, U$ that transforms between left and right invariant algebra elements for invariant vector fields on a Lie group. In the error dynamics (5.19) derived below, the action $\psi_{\hat{X}^{-1}}$ transforms the measured system input into the correct

representation for the error dynamics around the chosen origin $\mathring{\xi}$.

Lemma 5.4.1. *Let the global state error e be defined as in* (5.17) *and the origin velocity $\mathring{u}$ be defined as in* (5.18)*. The dynamics of e are given by*

$$\dot{e} = \mathrm{D}\phi_e \left(\Lambda(e, \mathring{u}) - \Lambda(\mathring{\xi}, \mathring{u}) - \Delta \right), \tag{5.19}$$

and depend only on e, $\mathring{u}$ and the correction Δ.

Proof. Let $X(t)$ be a solution to the lifted system (5.14) satisfying $\phi(X(0), \mathring{\xi}) = \xi(0)$. Then $\xi \equiv \phi(X, \mathring{\xi})$, and

$$e = \phi(\hat{X}^{-1}, \xi) = \phi(X\hat{X}^{-1}, \mathring{\xi}) = \phi(E, \mathring{\xi}), \tag{5.20}$$

where $E := X\hat{X}^{-1}$. Computing the dynamics of E, one has

$$\begin{aligned}
\dot{E} &= \mathrm{D}R_{\hat{X}^{-1}} \mathrm{D}L_X \Lambda(\phi(X, \mathring{\xi}), u) - \mathrm{D}L_X \mathrm{D}L_{\hat{X}^{-1}} \mathrm{D}R_{\hat{X}^{-1}} \left(\mathrm{D}L_{\hat{X}} \Lambda(\phi(\hat{X}, \mathring{\xi}), u) + \mathrm{D}R_{\hat{X}} \Delta \right), \\
&= \mathrm{D}L_E \, \mathrm{Ad}_{\hat{X}} \left(\Lambda(\phi(X, \mathring{\xi}), u) - \Lambda(\phi(\hat{X}, \mathring{\xi}), u) \right) - \mathrm{D}L_E \Delta, \\
&= \mathrm{D}L_E \left(\Lambda(\phi(E, \mathring{\xi}), \psi(\hat{X}^{-1}, u)) - \Lambda(\mathring{\xi}, \psi(\hat{X}^{-1}, u)) \right) - \mathrm{D}L_E \Delta,
\end{aligned} \tag{5.21}$$

where the last line follows from the equivariance of the lift. Recalling (5.18), the dynamics of e follow from (5.20) and (5.21),

$$\begin{aligned}
\dot{e} &= \mathrm{D}\phi_{\mathring{\xi}} \dot{E}, \\
&= \mathrm{D}\phi_{\phi(E, \mathring{\xi})} \mathrm{D}L_{E^{-1}} \dot{E}, \\
&= \mathrm{D}\phi_e \left(\Lambda(\phi(E, \mathring{\xi}), \mathring{u}) - \Lambda(\mathring{\xi}, \mathring{u}) - \Delta \right), \\
&= \mathrm{D}\phi_e \left(\Lambda(e, \mathring{u}) - \Lambda(\mathring{\xi}, \mathring{u}) - \Delta \right),
\end{aligned}$$

as required. □

5.4.3 Linearisation

Error Dynamics

Let $e \in \mathcal{M}$ denote the global state error (5.17), and let $\mathring{u} \in \mathbb{L}$ denote the origin velocity (5.18). Fix a local coordinate chart $\vartheta : \mathcal{U}_{\mathring{\xi}} \to \mathbb{R}^m$ where $\mathcal{U}_{\mathring{\xi}} \subset \mathcal{M}$ is a neighbourhood of the fixed origin $\mathring{\xi}$, and $\vartheta(\mathring{\xi}) = 0$. Let ε be the local coordinates of the state error,

$$\varepsilon = \vartheta(e). \tag{5.22}$$

The EqF correction is designed by linearising the pre-observer error dynamics of ε at zero, that is the dynamics (5.19) with the correction term $\Delta \equiv 0$ set to zero.

Lemma 5.4.2. *Let ϑ be local coordinates on $\mathcal{M}$ in an open neighbourhood $\mathcal{U}_{\mathring{\xi}} \subset \mathcal{M}$ around $\mathring{\xi}$. Assume $e(t) \in \mathcal{U}_{\mathring{\xi}}$ for all time. The linearised pre-observer ($\Delta \equiv 0$) dynamics of $e(t)$ about $\varepsilon = 0$ are*

$$\dot{\varepsilon} = \mathring{A}_t \varepsilon + O(|\varepsilon|^2), \tag{5.23}$$

$$\mathring{A}_t := \mathrm{D}_e|_{\mathring{\xi}} \vartheta(e) \cdot \mathrm{D}_E|_{\mathrm{id}} \phi_{\mathring{\xi}}(E) \cdot \mathrm{D}_e|_{\mathring{\xi}} \Lambda(e, \mathring{u}) \cdot \mathrm{D}_\varepsilon|_0 \vartheta^{-1}(\varepsilon). \tag{5.24}$$

Proof. The nonlinear pre-observer dynamics of the global state error (5.19) in local coordinates $\varepsilon = \vartheta(e)$ are

$$\dot{\varepsilon} = \mathrm{D}\vartheta \cdot \mathrm{D}\phi_{\vartheta^{-1}(\varepsilon)}(\Lambda(\vartheta^{-1}(\varepsilon), \mathring{u}) - \Lambda(\mathring{\xi}, \mathring{u})). \tag{5.25}$$

Clearly, $\Lambda(\vartheta^{-1}(0), \mathring{u}) = \Lambda(\mathring{\xi}, \mathring{u})$ since $\vartheta^{-1}(0) = \mathring{\xi}$. Hence, linearising $\dot{\varepsilon}$ about $\varepsilon = 0$ yields

$$\begin{aligned}
&\mathrm{D}\vartheta \cdot \mathrm{D}\phi_{\vartheta^{-1}(\varepsilon)}(\Lambda(\vartheta^{-1}(\varepsilon), \mathring{u}) - \Lambda(\mathring{\xi}, \mathring{u})) \\
&= \mathrm{D}\vartheta \cdot \mathrm{D}\phi_{\vartheta^{-1}(0)}(\Lambda(\vartheta^{-1}(0), \mathring{u}) - \Lambda(\mathring{\xi}, \mathring{u})) \\
&\quad + \mathrm{D}_\varepsilon|_0 \left(\mathrm{D}\vartheta \cdot \mathrm{D}\phi_{\vartheta^{-1}(\varepsilon)}(\Lambda(\vartheta^{-1}(\varepsilon), \mathring{u}) - \Lambda(\mathring{\xi}, \mathring{u})\right)[\varepsilon] + O(|\varepsilon|^2), && (5.26) \\
&= \mathrm{D}_\varepsilon|_0 \left(\mathrm{D}\vartheta \cdot \mathrm{D}\phi_{\vartheta^{-1}(\varepsilon)}\right)[\varepsilon] \cdot \left(\Lambda(\vartheta^{-1}(0), \mathring{u}) - \Lambda(\mathring{\xi}, \mathring{u})\right) \\
&\quad + \mathrm{D}\vartheta \cdot \mathrm{D}\phi_{\vartheta^{-1}(0)} \cdot \mathrm{D}_\varepsilon|_0 \left(\Lambda(\vartheta^{-1}(\varepsilon), \mathring{u}) - \Lambda(\mathring{\xi}, \mathring{u})\right)[\varepsilon] + O(|\varepsilon|^2), && (5.27) \\
&= \mathrm{D}\vartheta \cdot \mathrm{D}\phi_{\mathring{\xi}} \cdot \mathrm{D}_e|_{\mathring{\xi}} \Lambda(e, \mathring{u}) \cdot \mathrm{D}\vartheta^{-1}[\varepsilon] + O(|\varepsilon|^2), && (5.28) \\
&= \mathring{A}_t \varepsilon + O(|\varepsilon|^2).
\end{aligned}$$

Here (5.26) is the first order Taylor expansion in local coordinates about $\varepsilon = 0$. Equation (5.27) follows since $\Lambda(\vartheta^{-1}(0), \mathring{u}) - \Lambda(\mathring{\xi}, \mathring{u}) = 0$ and by expanding the first-order term using the product rule. Equation (5.28) is the result of applying $\Lambda(\vartheta^{-1}(0), \mathring{u}) - \Lambda(\mathring{\xi}, \mathring{u}) = 0$ to eliminate the first term, while noting that $\mathrm{D}_\varepsilon|_0 \Lambda(\mathring{\xi}, \mathring{u}) \equiv 0$ and applying the chain rule to simplify the second term in (5.27). The final line follows from (5.24). □

The primary role of the error dynamics linearisation is in the covariance propagation instantiated in the Ricatti equation (5.37) where the covariance is propagated along the pre-observer trajectories [4]. Barrau and Bonnabel [48] showed that, if the manifold $\mathcal{M} = \mathbf{G}$, the origin is chosen $\mathring{\xi} = \mathrm{id}$, the local coordinates are chosen $\vartheta(e) := \log(e)$, and the system dynamics are 'group affine' (cf. Appendix B), then the pre-observer dynamics are exact, $\dot{\varepsilon} = \mathring{A}_t \varepsilon$. This *exact* linearisation of the pre-observer error dynamics removes the $O(|\varepsilon|^2)$ linearisation error and significantly improves the performance of the filter by reducing linearisation error in the Ricatti equation. Even without the group affine property, the equivariant structure of the error dynamics provides significant advantages over the standard EKF. The error dynamics are linearised at a single point $\mathring{\xi}$ and using a single coordinate chart ϑ, which can be designed intentionally to minimise the $O(|\varepsilon|^2)$ linearisation error. As we show in Lemma 5.4.3, it is also possible to exploit the equivariant system structure to reduce linearisation error in the output approximation, further

improving the filter performance.

System Output

Consider the state error e (5.17), and let $\xi \in \mathcal{M}$ and $\hat{X} \in \mathbf{G}$ denote the true system state and observer state respectively. Let $\varepsilon \in \mathbb{R}^m$ represent local coordinates for e as in (5.22). The output $y = h(\xi)$ can be written

$$h(\xi) = h(\phi(\hat{X}, e)) = h(\phi_{\hat{X}}(\vartheta^{-1}(\varepsilon))). \tag{5.29}$$

Note that substituting $\varepsilon = 0$ gives

$$h(\phi_{\hat{X}}(\vartheta^{-1}(0))) = h(\phi_{\hat{X}}(\mathring{\xi})) = h(\hat{\xi}).$$

In common with error state Kalman filters, the output $y = y(\varepsilon)$ is considered as a function of the error (5.29) while the predicted output $\hat{y} = h(\hat{\xi})$ is considered an independent signal. Define the output residual

$$\tilde{y} = y(\varepsilon) - \hat{y}. \tag{5.30}$$

Linearising $\tilde{y}$ as a function of $\varepsilon \in \mathbb{R}^m$ around $\varepsilon = 0$ yields

$$\tilde{y} = C_t \varepsilon + O(|\varepsilon|^2), \tag{5.31}$$

$$C_t := \mathrm{D}_{\xi}|_{\hat{\xi}} h(\xi) \cdot \mathrm{D}_{e}|_{\mathring{\xi}} \phi_{\hat{X}}(e) \cdot \mathrm{D}_{\varepsilon}|_{0} \vartheta^{-1}(\varepsilon). \tag{5.32}$$

The matrix $C_t \in \mathbb{R}^{n \times m}$, the Jacobian of (5.29), is termed the *standard output matrix*.

Equivariant Output Linearisation

When the system has output equivariance, the linearisation of the output function can be improved to obtain $O(|\varepsilon|^3)$ error. A filter's ability to incorporate information from measurements correctly is fundamental to its robustness and transient performance. As shown in the simulation results in §5.6, the equivariant output linearisation presented below greatly improves the EqF performance both with and without the presence of noise in the measurement signals.

Lemma 5.4.3. *Suppose the local coordinates $\vartheta : \mathcal{U}_{\mathring{\xi}} \subset \mathcal{M} \to \mathbb{R}^m$ are normal coordinates of the Lie group about $\mathring{\xi}$. Then*

$$\tilde{y} = C_t^\star \varepsilon + O(|\varepsilon|^3), \tag{5.33}$$

$$C_t^\star \varepsilon = \frac{1}{2} \left(\mathrm{D}_{E|\mathrm{id}} \rho(E, y) + \mathrm{D}_{E|\mathrm{id}} \rho(E, \hat{y}) \right) \mathrm{Ad}_{\hat{X}^{-1}} \varepsilon^\wedge, \tag{5.34}$$

where $\cdot^\wedge : \mathbb{R}^m \to \mathfrak{m} \subset \mathfrak{g}$ is the identification of $\mathfrak{m}$ with $\mathbb{R}^m$ used in defining ϑ.

Proof. By construction $\vartheta^{-1}(\varepsilon) = \phi_{\mathring{\xi}}(\exp(\varepsilon^\wedge)) = e$. Recalling (5.29) and (5.30) one has

$$\tilde{y}(\varepsilon) = h(\phi(\hat{X}, \phi(\exp(\varepsilon^\wedge), \mathring{\xi}))) - h(\phi(\hat{X}, \mathring{\xi})).$$

Clearly, $\tilde{y}(0) = 0$. Using the equivariance of h, one has

$$\begin{aligned} h(\phi(\hat{X}, \phi(\exp(\varepsilon^\wedge), \mathring{\xi}))) &= h(\phi(\exp(\varepsilon^\wedge)\hat{X}, \mathring{\xi})), \\ &= h(\phi(\hat{X}\hat{X}^{-1}\exp(\varepsilon^\wedge)\hat{X}, \mathring{\xi})), \\ &= h(\phi(\exp(\mathrm{Ad}_{\hat{X}^{-1}}(\varepsilon^\wedge)), \phi_{\hat{X}}(\mathring{\xi}))), \\ &= \rho(\exp(\mathrm{Ad}_{\hat{X}^{-1}}(\varepsilon^\wedge)), h(\hat{\xi})). \end{aligned}$$

Setting $\hat{y} = h(\hat{\xi})$ and differentiating $\tilde{y}$ at $\varepsilon = 0$ in a direction $\gamma \in \mathbb{R}^m$ yields

$$\mathrm{D}_{x|0}\tilde{y}(x)[\gamma] = \mathrm{D}_{E|\mathrm{id}}\rho_{\hat{y}}(E)\,\mathrm{Ad}_{\hat{X}^{-1}}\,\gamma^\wedge.$$

Although this formula allows for an arbitrary $\gamma \in \mathbb{R}^m$, the linearisation is computed for $\gamma = \varepsilon$.

The fact that the linearisation is computed in the same direction ε as the coordinates of the error can be exploited along with equivariance to obtain a second linearisation point. In particular, we will compute the differential of y at $\varepsilon = \vartheta(e)$ *in direction* ε. Note that $\phi_{\mathring{\xi}}(\exp(\varepsilon^\wedge)) = \phi(\hat{X}^{-1}, \xi)$, and therefore $\phi_{\mathring{\xi}}(\exp(\varepsilon^\wedge)\hat{X}) = \xi$. Then

$$\begin{aligned} \mathrm{D}_{x|\varepsilon}\tilde{y}(x)[\varepsilon] &= \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} h(\phi(\exp((1+t)\varepsilon^\wedge)\hat{X}, \mathring{\xi})), \\ &= \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} h(\phi(\exp(\varepsilon^\wedge)\exp(t\varepsilon^\wedge)\hat{X}, \mathring{\xi})), \\ &= \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} h(\phi(\exp(\varepsilon^\wedge)\hat{X}\hat{X}^{-1}\exp(t\varepsilon^\wedge)\hat{X}, \mathring{\xi})), \\ &= \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} h(\phi(\hat{X}^{-1}\exp(t\varepsilon^\wedge)\hat{X}, \xi)), \\ &= \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \rho(\exp(t\,\mathrm{Ad}_{\hat{X}^{-1}}\,\varepsilon^\wedge), h(\xi)), \\ &= \mathrm{D}_{E|\mathrm{id}}\rho(E, y)\,\mathrm{Ad}_{\hat{X}^{-1}}\,\varepsilon^\wedge. \end{aligned}$$

Although the differential is posed at the unknown error state ε, it is evaluated using only the known measurement data y.

Consider the Taylor expansion of the differential $\mathrm{D}_{x|\varepsilon}\tilde{y}(x)$ with respect to ε around $\varepsilon = 0$:

$$\mathrm{D}_{x|\varepsilon}\tilde{y}(x)[\cdot] = \mathrm{D}_{x|0}\tilde{y}(x)[\cdot] + \mathrm{D}^2_{x|0}\tilde{y}(x)[\varepsilon, \cdot] + O(|\varepsilon|^2),$$

and hence

$$\mathrm{D}^2_{x|0}\tilde{y}(x)[\varepsilon,\varepsilon] = \mathrm{D}_{x|\varepsilon}\tilde{y}(x)[\varepsilon] - \mathrm{D}_{x|0}\tilde{y}(x)[\varepsilon] + O(|\varepsilon|^3).$$

The result (5.34) follows from taking the Taylor expansion of $\tilde{y}$ with respect to ε and substituting

$$\begin{aligned}\tilde{y}(\varepsilon) &= \tilde{y}(0) + \mathrm{D}_{x|0}\tilde{y}(x)[\varepsilon] + \frac{1}{2}\mathrm{D}^2_{x|0}\tilde{y}(x)[\varepsilon,\varepsilon] + O(|\varepsilon|^3),\\ &= \frac{1}{2}(\mathrm{D}_{x|\varepsilon}\tilde{y}(x)[\varepsilon] + \mathrm{D}_{x|0}\tilde{y}(x)[\varepsilon]) + O(|\varepsilon|^3),\\ &= \frac{1}{2}(\mathrm{D}_{E|\mathrm{id}}\rho(E,y) + \mathrm{D}_{E|\mathrm{id}}\rho(E,\hat{y}))\,\mathrm{Ad}_{\hat{X}^{-1}}\,\varepsilon^\wedge + O(|\varepsilon|^3).\end{aligned}$$

□

5.5 Equivariant Filter (EqF)

Consider a kinematic system (5.6) with a state symmetry $\phi : \mathbf{G}\times\mathcal{M}\to\mathcal{M}$. Assume the system is equivariant with input symmetry $\psi : \mathbf{G}\times\mathbb{L}\to\mathbb{L}$ and has an equivariant lift $\Lambda : \mathcal{M}\times\mathbb{L}\to\mathfrak{g}$. Let $\xi\in\mathcal{M}$ denote the true state of the system, with trajectory determined by the measured input $u\in\mathbb{L}$. Denote the configuration output $y = h(\xi)$.

We construct the Equivariant Filter (EqF) as follows. Let $\hat{X}\in\mathbf{G}$ denote the observer state. Pick an arbitrary fixed origin $\mathring{\xi}\in\mathcal{M}$. For a general output map, set C_t to be the standard output matrix defined in (5.31). If the system has output equivariance then set $C_t = C_t^\star$ to be the equivariant output matrix as defined in (5.34). In this case the resulting algorithm is termed the EqF*. Let $\mathring{A}_t$ denote the state matrix as defined in (5.24). Choose an initial value for the Riccati term $\Sigma_0\in\mathbb{S}_+(m)$, where $\mathbb{S}_+(m)$ is the set of positive-definite symmetric $m\times m$ matrices, and pick a state gain matrix $M_t\in\mathbb{S}_+(m)$ and an output gain matrix $N_t\in\mathbb{S}_+(n)$. Choose a right inverse $\mathrm{D}_E|_{\mathrm{id}}\phi_{\mathring{\xi}}(E)^\dagger$ of $\mathrm{D}_E|_{\mathrm{id}}\phi_{\mathring{\xi}}(E)$; that is, $\mathrm{D}_E|_{\mathrm{id}}\phi_{\mathring{\xi}}(E)\cdot\mathrm{D}_E|_{\mathrm{id}}\phi_{\mathring{\xi}}(E)^\dagger = \mathrm{id}$.

The proposed equivariant filter is given by the solution of

$$\dot{\hat{X}} = \mathrm{D}L_{\hat{X}}\Lambda(\phi(\hat{X},\mathring{\xi}),u) + \mathrm{D}R_{\hat{X}}\Delta,\quad \hat{X}(0) = \mathrm{id}, \tag{5.35}$$

$$\Delta = \mathrm{D}_E|_{\mathrm{id}}\phi_{\mathring{\xi}}(E)^\dagger\mathrm{D}\vartheta^{-1}\Sigma C_t^\top N_t^{-1}(y - h(\phi(\hat{X},\mathring{\xi}))), \tag{5.36}$$

$$\dot{\Sigma} = \mathring{A}_t\Sigma + \Sigma\mathring{A}_t^\top + M_t - \Sigma C_t^\top N_t^{-1}C_t\Sigma,\quad \Sigma(0) = \Sigma_0. \tag{5.37}$$

If the pair $(\mathring{A}_t, C_t)$ is uniformly observable in the sense of Proposition 1 of [135], then $\Sigma(t)$ is bounded above and below, and the Riccati equation (5.37) is well-defined for all time [136].

Provided that the error trajectory $\varepsilon(t)$ remains well-defined for all time $t\geq 0$, [137, Theorem 1.1.1] provides sufficient conditions for the convergence of $\varepsilon\to 0$. In particular, if the error system is uniformly observable, the second derivative of the error dynamics is bounded, and the initial error is sufficiently small, then the error ε and the Lyapunov

function

$$\mathcal{L}(t) := \varepsilon^\top \Sigma^{-1} \varepsilon, \tag{5.38}$$

converge exponentially to zero as $t \to \infty$ [137].

Remark 5.5.1. *Recent work by the authors [133; 89; 4] showed that the Ricatti equation* (5.37) *can be augmented by a curvature modification that compensates for the reset process, where the linearisation point is continually translated to track the observer state, in the extended Kalman filter derivation [4]. Several works have shown that, during the transient at least, an appropriate curvature modification term can improve filter performance [138; 4]. Curvature is directly connected to parallel transport on a manifold and is an additional structure that can be chosen independently from the homogeneous space structure. One possible choice is to define the normal coordinates to be flat, that is, parallel transport on the manifold is just translation in local coordinates. The Ricatti equation* (5.37) *corresponds to this choice. Such a choice has the advantage of simplicity; however, the associated affine connection will usually have non-zero torsion. The relative benefit or consequence of choosing different geometries, with different curvatures, along with symmetry or torsion of the associated connections remains an open question in equivariant systems theory.*

5.5.1 EqF Gain Tuning

The choice of gain matrices Σ_0, M_t and N_t can greatly influence the performance of the EqF. In the context of a Kalman-Bucy filter, Σ_0 reflects the uncertainty in the initial state estimate, and M_t and N_t are optimally chosen to be the intensities (covariances) of zero-mean Gaussian noise terms added to the filter dynamics and output, respectively. Similar choices can be made for the EqF. The initial value of the Riccati term Σ_0 can be chosen to reflect the uncertainty in the initial state estimate as expressed in the chosen local coordinates. Suppose the measured velocity $\mathrm{d}u_m = \mathrm{d}u + \mathrm{d}\mu_u$ and the measured output $\mathrm{d}y_m = \mathrm{d}y + \mathrm{d}\nu_y$, where $\mathrm{d}\mu_u \sim \mathbf{W}(0, M_t^m)$ and $\mathrm{d}\nu_y \sim \mathbf{W}(0, N_t^m)$ are Wiener processes. Then, re-linearising the pre-observer error dynamics (5.19) with $\Delta \equiv 0$ and output (5.29) yields

$$\mathrm{d}\varepsilon = \mathring{A}_t \varepsilon \mathrm{d}t + B_t \mathrm{d}\mu_u, \tag{5.39}$$

$$\mathrm{d}\tilde{y} = C_t \varepsilon \mathrm{d}t + \mathrm{d}\nu_y, \tag{5.40}$$

where the *input matrix* B_t is obtained by linearising the pre-observer error dynamics with respect to a perturbation of the measured input,

$$\begin{aligned} B_t &:= \mathrm{D}_e|_{\mathring{\xi}} \vartheta(e) \cdot \mathrm{D}_E|_{\mathrm{id}} \phi_{\mathring{\xi}}(E) \cdot \mathrm{Ad}_{\hat{X}} \cdot \mathrm{D}_u|_{u_m} \Lambda(\hat{\xi}, u), \\ &= \mathrm{D}_e|_{\mathring{\xi}} \vartheta(e) \cdot \mathrm{D}_E|_{\mathrm{id}} \phi_{\mathring{\xi}}(E) \cdot \mathrm{D}_w|_{\mathring{u}_m} \Lambda(\mathring{\xi}, w) \cdot \psi_{\hat{X}^{-1}}. \end{aligned} \tag{5.41}$$

Based on this formulation, the EqF gain matrices can be chosen by

$$M_t = M_\varepsilon + B_t M_t^m B_t^\top, \tag{5.42}$$
$$N_t = N_\varepsilon + N_t^m. \tag{5.43}$$

The matrices $M_\varepsilon \in \mathbb{S}_+(m)$ and $N_\varepsilon \in \mathbb{S}_+(n)$ are optimally set to zero in the case of a linear system, but can otherwise be used by the practitioner to model the error introduced to the dynamics and output by linearisation.

5.6 Example Revisited: Single Bearing Estimation

5.6.1 Equivariant System

Here we describe the design preliminaries for the EqF and EqF* for the single bearing estimation problem following Algorithm 1 as described in Appendix A.

State Symmetry

Consider the Lie group of 3D rotations

$$\mathbf{SO}(3) = \{R \in \mathbb{R}^{3\times 3} \mid R^\top R = I_3,\ \det(R) = 1\}.$$

This group has a right action on the sphere $\phi : \mathbf{SO}(3) \times \mathrm{S}^2 \to \mathrm{S}^2$ given by

$$\phi(R, \eta) := R^\top \eta. \tag{5.44}$$

System Equivariance

Define the map $\psi : \mathbf{SO}(3) \times \mathbb{R}^3 \to \mathbb{R}^3$ to be

$$\psi(R, \Omega) := R^\top \Omega. \tag{5.45}$$

Then the system is equivariant with respect to the state action ϕ and input action ψ. To see this, let $\Omega \in \mathbb{R}^3$, $\eta \in \mathrm{S}^2$, and $R \in \mathbf{SO}(3)$ be arbitrary, and compute

$$\begin{aligned}
\Phi(R, f_\Omega)(\eta) &= \mathrm{D}\phi_R(-\Omega^\times R\eta), \\
&= -R^\top \Omega^\times R\eta, \\
&= -(R^\top \Omega)^\times \eta, \\
&= f_{\psi(R,\Omega)}(\eta).
\end{aligned}$$

Equivariant Lift

The Lie algebra $\mathfrak{so}(3)$ of $\mathbf{SO}(3)$ can be written as the subspace of skew-symmetric matrices,

$$\mathfrak{so}(3) := \{\omega \in R^{3\times 3} \mid \omega^\top = -\omega\}.$$

Define the candidate lift function $\Lambda : \mathrm{S}^2 \times \mathbb{R}^3 \to \mathfrak{so}(3)$ by

$$\Lambda(\eta, \Omega) := \Omega^\times \tag{5.46}$$

To check that it is indeed a lift, evaluate the lift condition (5.10).

$$\begin{aligned} \mathrm{D}_R|_{\mathrm{id}}\phi_\eta(R)\left[\Lambda(\eta,\Omega)\right] &= \mathrm{D}_R|_{\mathrm{id}}\phi_\eta(R)\left[\Omega^\times\right], \\ &= (\Omega^\times)^\top \eta, \\ &= -\Omega^\times \eta, \\ &= f_\Omega(\eta), \end{aligned}$$

as required. Next, check the equivariance of Λ as in (5.12).

$$\begin{aligned} \Lambda(\phi(R,\eta),\psi(R,\Omega)) &= \Lambda(R^\top\eta, R^\top\Omega), \\ &= (R^\top\Omega)^\times, \\ &= R^\top\Omega^\times R, \\ &= \mathrm{Ad}_{R^{-1}}\Lambda(\eta,\Omega), \end{aligned}$$

as required.

Output Equivariance

The action $\rho : \mathbf{SO}(3) \times \mathbb{R}^3 \to \mathbb{R}^3$ given by

$$\rho(R, y) = R^\top y, \tag{5.47}$$

ensures the system (5.2) has output equivariance since

$$\rho(R, h(\eta)) = R^\top(c_m\eta) = c_m R^\top \eta = h(\phi(R,\eta)).$$

Origin and State Error

Fix the origin element $\mathring{\eta} = \mathbf{e}_1 \in \mathrm{S}^2$, and let $\hat{R} \in \mathbf{SO}(3)$ denote the observer state. The global state error is given by

$$e = \phi(\hat{R}^{-1}, \eta) = \hat{R}\eta$$

where $\eta \in \mathrm{S}^2$ is the true state of the system.

Since the system exhibits output equivariance, we choose to use normal coordinates. Explicitly, define

$$\begin{aligned} \mathfrak{m} &:= \{v^\wedge \in \mathbb{R}^{3\times 3} \mid v \in \mathbb{R}^2\} \subset \mathfrak{so}(3), \\ (v_2, v_3)^\wedge &:= (0, v_2, v_3)^\times, \end{aligned}$$

where the indices $(v_2, v_3) \in \mathbb{R}^2$ are chosen to correspond to the associated indices for the embedding into $\mathfrak{so}(3)$. Then the normal coordinates for S^2 about $\mathbf{e}_1$ are given by

$$\begin{aligned} \vartheta(e) &:= -\mathrm{atan2}(|\mathbf{e}_1 \times e|, \mathbf{e}_1^\top e) \begin{pmatrix} 0_{2\times 1} & I_2 \end{pmatrix} \frac{\mathbf{e}_1 \times e}{|\mathbf{e}_1 \times e|}, \\ \vartheta^{-1}(\varepsilon) &:= \phi(\exp(\varepsilon^\wedge), \mathbf{e}_1). \end{aligned} \tag{5.48}$$

In this system, $\mathrm{D}_E|_{\mathrm{id}}\phi_{\mathbf{e}_1}(E)$ is not invertible,

$$\begin{aligned} \mathrm{D}_{E|_{\mathrm{id}}}\phi_{\mathbf{e}_1}(E)[\omega^\times] &= \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \phi(\exp(t\omega^\times), \mathbf{e}_1), \\ &= -\omega^\times \mathbf{e}_1, \\ &= \mathbf{e}_1^\times \omega. \end{aligned}$$

We propose the following right inverse (required in (5.36)),

$$\mathrm{D}_E|_{\mathrm{id}}\phi_{\mathbf{e}_1}(E)^\dagger[u] := (u_3, -u_2)^\wedge \in \mathfrak{m},$$

where $u = (0, u_2, u_3)^\top \in \mathrm{T}_{\mathbf{e}_1}\mathrm{S}^2$ is an arbitrary tangent vector in the embedded coordinates for S^2 at $\mathbf{e}_1$. To see that this indeed defines a right-inverse, compute

$$\begin{aligned} \mathrm{D}_{E|_{\mathrm{id}}}\phi_{\mathbf{e}_1}(E)\mathrm{D}_E|_{\mathrm{id}}\phi_{\mathbf{e}_1}(E)^\dagger[u] &= \mathbf{e}_1^\times \begin{pmatrix} 0 & u_3 & -u_2 \end{pmatrix}^\top, \\ &= \begin{pmatrix} 0 & u_2 & u_3 \end{pmatrix}^\top, \\ &= u. \end{aligned}$$

EqF Matrices

The EqF matrices are obtained by specialising the general matrix formulas to the specific example. The state matrix $\mathring{A}_t$ (5.23), the input matrix B_t (5.41), the standard output matrix C_t (5.32), and the equivariant output matrix $C_t = C_t^\star$ (5.34) are given by

$$\begin{aligned} \mathring{A}_t &= 0_{2\times 2}, & B_t &= \begin{pmatrix} 0_{2\times 1} & I_2 \end{pmatrix} \hat{R}, \\ C_t &= \hat{y}^\times \hat{R}^\top \begin{pmatrix} 0_{1\times 2} \\ I_2 \end{pmatrix}, & C_t^\star &= \frac{1}{2}\left(y^\times + \hat{y}^\times\right) \hat{R}^\top \begin{pmatrix} 0_{1\times 2} \\ I_2 \end{pmatrix}. \end{aligned}$$

These expressions can be compared to the EKF matrices (5.1) and (5.4) derived in §5.2.

5.6.2 EqF Implementation

We implement the EqF equations (5.35-(5.37)) as detailed in Algorithm 2, §A.

The observer dynamics are given by specialising (5.35),

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\hat{R} &= \mathrm{D}L_{\hat{R}}\Lambda(\phi(\hat{R},\mathbf{e}_1),\Omega) + \mathrm{D}R_{\hat{R}}\Delta,\\ \dot{\hat{R}} &= \hat{R}\Omega^{\times} + \Delta\hat{R},\end{aligned}$$

where the correction term Δ is computed according to (5.36).

5.6.3 Simulation Results

To verify the observer design for this example, we performed a simulation of a robot rotating with an angular velocity $\Omega(t) = (0.1\cos(2t), 0.2\sin(t), 0)$ rad/s, where t is the simulation time in seconds. The initial state, the measured angular velocity, and the measured output were chosen by

$$\begin{aligned}\eta(0) &= \frac{\mathbf{e}_1+\mu_0}{|\mathbf{e}_1+\mu_0|}, & \mu_0 &\sim N(0, 0.5^2 I_3),\\ \Omega_m &= \Omega + \mu_u, & \mu_u &\sim N(0, 0.01^2 I_3),\\ y_m &= \eta + \nu_y, & \nu_y &\sim N(0, 0.05^2 I_3),\end{aligned} \tag{5.49}$$

respectively. The state $\eta(t)$ was then computed by integrating

$$\frac{\mathrm{d}}{\mathrm{d}t}\eta(t) = f_{\Omega(t)}(\eta(t)) = -\Omega(t)^{\times}\eta(t).$$

The EqF gain matrices were chosen according to the procedure outlined in §5.5.1.

We also implemented an extended Kalman filter (EKF) as described in §5.2 to compare its performance to that of the EqF. The system (5.1,5.2), the EqF equations (5.35-5.37), and the EKF were all implemented in python3 and integrated for 5.0 s using Euler integration with a time step of 0.01 s. Both the EqF and EKF were given the initial estimate $\hat{\eta}(0) = \mathbf{e}_1$.

In order to verify the local exponential convergence of the proposed filters in the absence of noise, we performed a simulation with the gyroscope and magnetometer noise set to zero, that is, $\mu_u = 0, \nu_y = 0$. Figure 5.1 shows the absolute angle between the estimated direction and true direction and the Lyapunov value (5.38) for each filter. The results demonstrate the performance of each filter under ideal conditions where the only error is due to the difference between the initial state $\eta(0)$ and the initial estimate $\mathbf{e}_1$. It is clear to see that the EqF is locally exponentially convergent, and that the EqF* exhibits faster initial convergence.

We also performed 500 monte carlo simulations with noise added to the velocity, measurement, and initial conditions, generated according to (5.49). Figure 5.2 shows the distribution of the absolute angle between the estimated direction and true direction for each filter, as well as the Lyapunov value (5.38) for each of the filters.

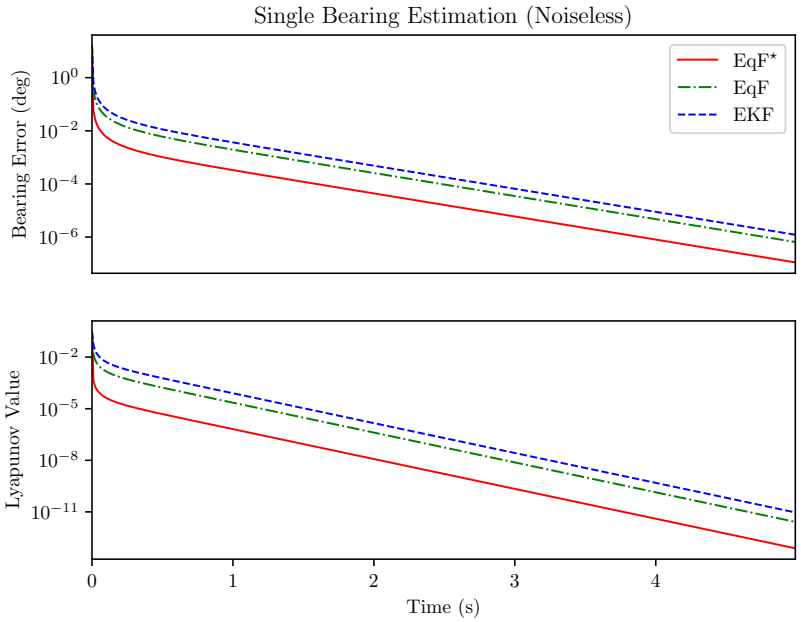


Figure 5.1: The angle error and Lyapunov value (5.38) for each of the filters without noise added to any of the gyroscope or magnetometer measurements. The EqF⋆ (red solid line) shows faster initial convergence than both the EqF (green dot-dashed line) and the EKF (blue dashed line).

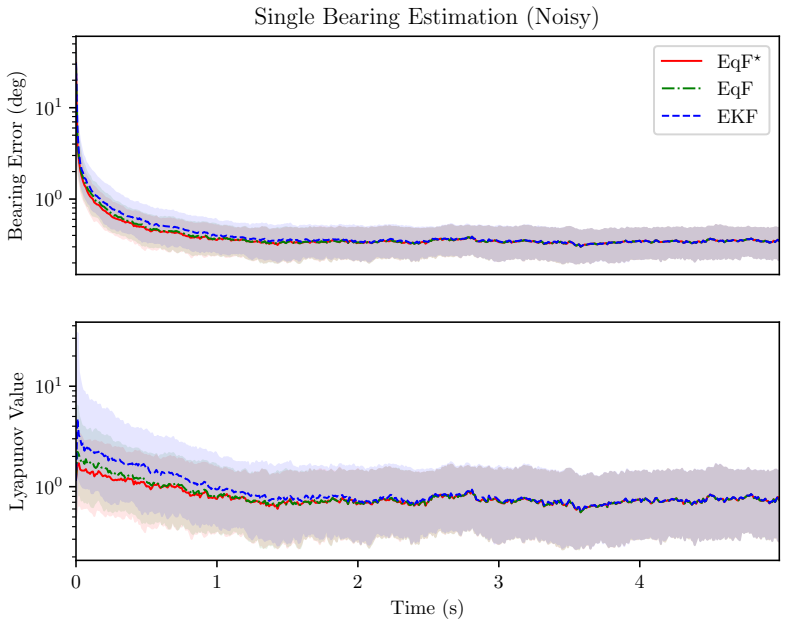


Figure 5.2: The median angle error and Lyapunov value (5.38) for each of the filters over 500 trials with noise generated for each trial independently. The EqF⋆ (red solid line) outperforms both the EqF (green dot-dashed line) and the EKF (blue dashed line) in terms of both angle error and Lyapunov value. The coloured areas show the 25th and 75th percentile for each filter's angle error and Lyapunov value.

Figure 5.3 shows the error introduced by different linearisations of the output. Each point $\eta \in \mathrm{S}^2$ other than $-\mathbf{e}_1$ has been mapped to $\mathbb{R}^2$ using spherical coordinates, and each heatmap shows, for a given filter, the absolute difference between the true measurement

residual $\tilde{y} = \eta - h(\mathbf{e}_1)$ and the linearised measurement residual $C_t\vartheta(\eta)$ for each filter. Note, in case of the EKF, $\vartheta(\eta) := \eta - \mathbf{e}_1$. The EqF and EKF show comparable performance, with the EKF having slightly higher error near $\mathbf{e}_1$ and the EqF having slightly higher error further away from $\mathbf{e}_1$. The output linearisation used by the EqF* is superior to the other filters, both near $\mathbf{e}_1$ as well as far from it, as expected from Lemma 5.4.3.

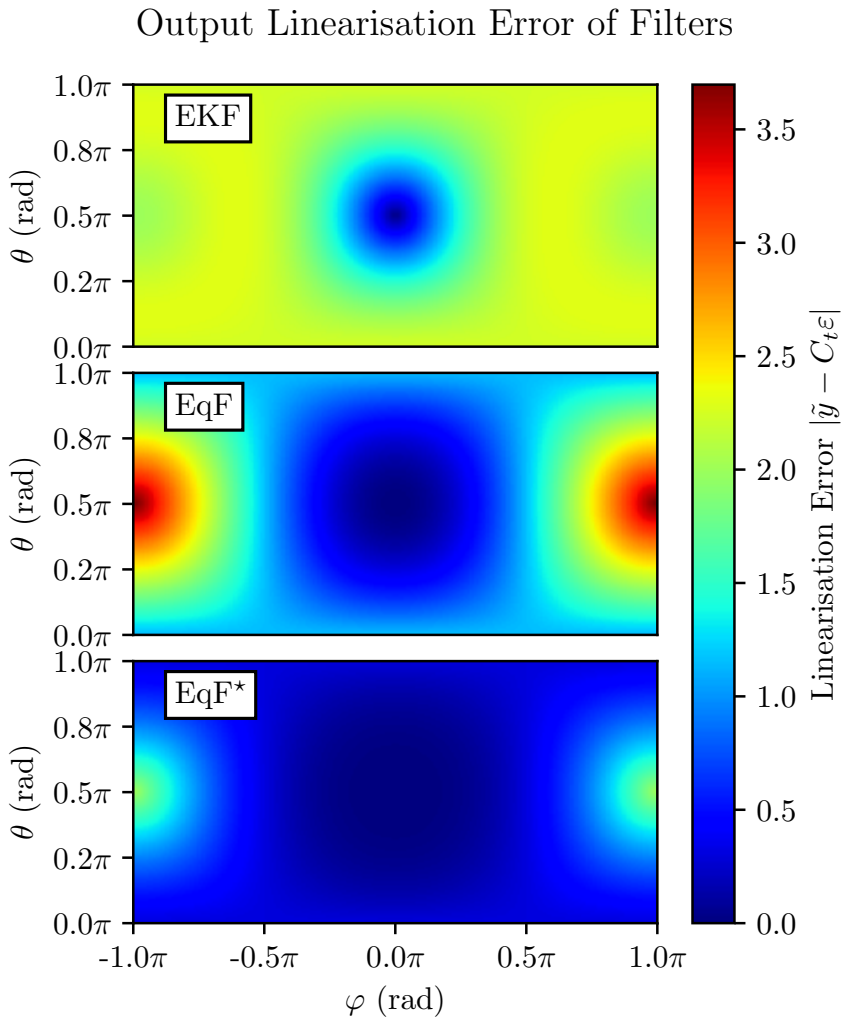


Figure 5.3: The error of linearising the output residual $\tilde{y}$ for each the EKF, EqF, and EqF*. The EKF and EqF are similar in terms of overall quality, with the EqF performing better near the linearisation point, and the EKF performing better further away. The EqF* shows clearly superior performance to both the EKF and EqF as expected from Lemma 5.4.3.

The example system shown is of interest for several reasons. First, the system is defined on a homogeneous space of a Lie group rather than on a Lie group itself, necessitating the development of a lifted system to apply equivariant observer design methods. This also precludes the application of the popular IEKF [48] as it is exclusively defined for group affine systems on Lie groups and not on the more general class of equivariant systems on homogeneous spaces. Second, the system has a symmetry compatible with the output function, enabling the use of Lemma 5.4.3 to improve the output linearisation. Figure 5.3 shows that the improved output linearisation results in a significant reduc-

tion in linearisation error over the whole space, and Figures 5.1 and 5.2 show clearly the positive effect of this improvement on filter performance.

Overall, the simulations demonstrate clearly that the EqF* outperforms the EqF, which in turn outperforms the EKF. The relative margin of improvement may appear small; however, it must be understood in the context of a problem chosen for simplicity and the implementation of an EKF that was designed carefully to exploit the specific structure available; specifically using embedding coordinates in order to obtain linear time varying state dynamics (5.1). Conversely, the EqF and EqF* were derived following the standard methodology outlined in Appendix A. For more complex problems, where there is no longer structure that can be exploited to design a clever EKF, the relative performance gap is expected to increase.

5.7 Chapter Conclusion

The Equivariant Filter (EqF) proposed in this chapter is a nonlinear observer that exploits symmetry properties of equivariant kinematic systems posed on homogeneous spaces. The key contributions of this chapter are

- Proposing a general filter design for equivariant systems with linearised dynamics about a fixed origin point rather than the time-varying state estimate.
- Demonstrating how output equivariance leads to an approximation of the output map for the EqF that has third order linearisation error (rather than the usual quadratic linearisation error), improving filter robustness and transient performance.
- Providing a simple example of the EqF application and simulation results where the EqF clearly outperforms the traditional EKF.

It is important to note that the example in the present chapter is chosen to be as simple as possible. It may appear that the mathematical overhead of the EqF is not warranted. Recent works [13; 1; 8] provide other examples where the EqF is the only symmetry based filter that can be applied.

The appendices provide details on how to derive and implement the EqF, and the conditions under which the EqF specialises to the invariant extended Kalman filter.

Chapter 6

EqVIO: An Equivariant Filter for Visual Inertial Odometry

6.1 Chapter 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), and this problem is of enduring interest to the robotics community due to the ubiquity of systems where such sensors are available, including smartphones, VR/AR headsets, racing drones, and more. State-of-the-art solutions for VIO are based on either the extended Kalman filter (EKF) or sliding-window optimisation. EKF-based solutions, such as ROVIO [19], OpenVINS [77], and MSCKF [72], are generally less accurate than optimisation-based methods but have lower compute and memory requirements and tend to be used in highly dynamic embedded systems applications such as VR headsets, smartphone applications, aerial vehicles, etc. On the other hand, optimisation-based methods, such as VINS-mono [76] and OKVIS [74], 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 optimisation-based methods is associated with the accumulation of linearisation errors. Recent advances in the theory of equivariant systems [39; 45] 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) [2] that minimize linearisation error.

In this chapter, 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 linearisation of the bias-free IMU error dynamics, and a better (higher-order) linearisation 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 optimisation-based algorithms in terms of both the accuracy of the estimated trajectory and the speed of processing each frame.

The key contributions of this chapter are as follows.

- 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. Additionally, in contrast to the symmetries explored in prior literature [52], the visual measurement function of VIO is equivariant with respect to the VI-SLAM group.
- The advantages of the VI-SLAM group are clearly demonstrated. Its $\mathbf{SE}_2(3)$ component is shown to eliminate linearisation error in the bias-free IMU error dynamics. The novel landmark symmetry, based on $\mathbf{SOT}(3)$ components, eliminates 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 parametrisation of landmarks that is core to modern filter performance in VIO algorithms.
- A novel VIO algorithm, *EqVIO*, is proposed that combines a simple feature-tracking front-end and basic outlier rejection with an equivariant filter implementation. EqVIO is shown to outperform state-of-the-art VIO algorithms in both speed and accuracy on both the popular EuRoC [20] and the challenging UZH FPV [21] datasets. Our implementation of EqVIO is open source and publicly available under a GNU GPLv3 licence[1].

6.2 Related Work

This section discusses some literature that is specifically relevant to this chapter.

6.2.1 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 Localisation 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 [116; 117; 67; 104]. An important milestone in the development of VIO systems is the Multi-State Constrained Kalman Filter (MSCKF) [72], 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 a high speed algorithm for VIO with only linear complexity in the number of landmarks considered. In [119], Konolige *et al.* 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 specialised 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.* [19] developed *ROVIO*: a VIO algorithm that mixes an

[1] https://github.com/pvangoor/eqvio

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.* [74] developed *OKVIS*, which solves monocular and stereo VIO by using non-linear optimisation on a sliding window of 'keyframes'. Semi-direct Visual Odometry (SVO) [75] used a sparse set of image patches in frame-to-frame optimisation to greatly reduce the presence of outliers and to solve VO at very high speeds. While SVO is not strictly a VIO solution, as it does not use an IMU, Delmerico and Scaramuzza [49] propose two methods to combine the output from SVO with an IMU in a filtering or smoothing framework. Recently, Qin *et al.* [76] combined a range of modern SLAM techniques to develop VINS-MONO, which performs tightly coupled keyframe optimisation-based VIO with efficient loop closure, and achieves competitive accuracy on popular datasets. Delmerico and Scaramuzza [49] benchmarked a range of state-of-the-art VIO systems. They showed that, generally, the VIO systems optimised 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 [77] is another recent VIO system, which mixes the MSCKF [72] 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-based and optimisation-based algorithms. EKF-based algorithms are preferred for their efficiency when computational resources are constrained, but suffer from linearisation error that accumulates and degrades performance over time.

6.2.2 Equivariant Observers for VIO

Equivariant observers are state estimators that exploit available Lie group symmetries of a given problem. Two key examples include the invariant EKF (IEKF) [39] and the Equivariant Filter (EqF) [2]. The success of equivariant observers in other robotics problems has led several authors to investigate their application to inertial navigation, SLAM, and VIO. In [139], Barrau and Bonnabel proposed the extended Special Euclidean group $\mathbf{SE}_2(3)$, and show that it can be used to obtain an exact linearisation 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 [140] to obtain a minimal representation of rotation error between quaternions that is analogous to the well-known multiplicative EKF (MEKF) [41].

In the earliest work examining symmetry properties of the SLAM problem, Barrau and Bonnabel [50] proposed a novel class of Lie groups, $\mathbf{SE}_n(m)$, 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 SLAM

[65]. In [141], Zhang *et al.* 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.* [51] then combined $\mathbf{SE}_n(m)$ with the MSCKF concept of [72] to propose an invariant MSCKF for VIO, which they showed to be a consistent filter unlike the original MSCKF. Brossard *et al.* [53] derived an invariant unscented Kalman filter (IUKF) for monocular SLAM using the Lie group proposed in [50], and outperformed other invariant filters for VIO. A number of other works [54; 55; 52] have also explored applying variants of the IEKF to VIO in an MSCKF framework.

Recently, van Goor *et al.* [1] developed a novel Lie group for visual SLAM under which the visual measurements of landmarks are equivariant, unlike in the previously explored $\mathbf{SE}_m(n)$ 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 [2], which additionally provides a framework for exploiting the equivariance of a system output function to reduce linearisation 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 chapter and our recent work [8].

6.2.3 Parametrisations of VIO Landmarks

The representation of the robot pose and environment map are known to have a significant impact on the accuracy of EKF-based SLAM and VIO approaches. In [102], Castellanos *et al.* 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.* [142], who proposed the earliest version of the *inverse-depth* parametrisation 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. A more recent version of the inverse-depth parametrisation is presented by Bloesch *et al.* [19], who adapted the on-manifold EKF approach developed in [140] to obtain a minimal representation of unit vectors on the sphere. While the existing inverse-depth parametrisations have been empirically shown to improve performance over the Euclidean parametrisation, the cause of this improvement is not well characterised. The discovery of symmetries for visual measurements [1] motivates the development of a new compatible parametrisation.

6.3 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 $\{I\}$ with the body-fixed frame of the

robot $\{B\}$. The states of the Visual-Inertial SLAM (VI-SLAM) problem are modelled as follows.

- $P_{\text{B}} := (R_{\text{B}}, x_{\text{B}}) \in \mathbf{SE}(3)$ is the pose of the IMU $\{B\}$ with respect to the inertial frame $\{0\}$,
- $v_{\text{B}} \in \mathbb{R}^3$ is the linear velocity of the IMU expressed in the inertial frame $\{0\}$,
- $b_{\text{B}} = (b^{\Omega}_{\text{B}}, b^{a}_{\text{B}}) \in \mathbb{R}^6$ is the IMU bias (where b^{Ω}_{B} and b^{a}_{B} are the gyroscope and accelerometer biases, respectively),
- $T = (R_T, x_T) \in \mathbf{SE}(3)$ is the pose of the camera $\{C\}$ with respect to the body-fixed frame $\{B\}$
- $p_i \in \mathbb{R}^3$ are the coordinates of landmark i in the inertial frame $\{0\}$.

Figure 6.1 shows a diagram of the full VI-SLAM configuration.

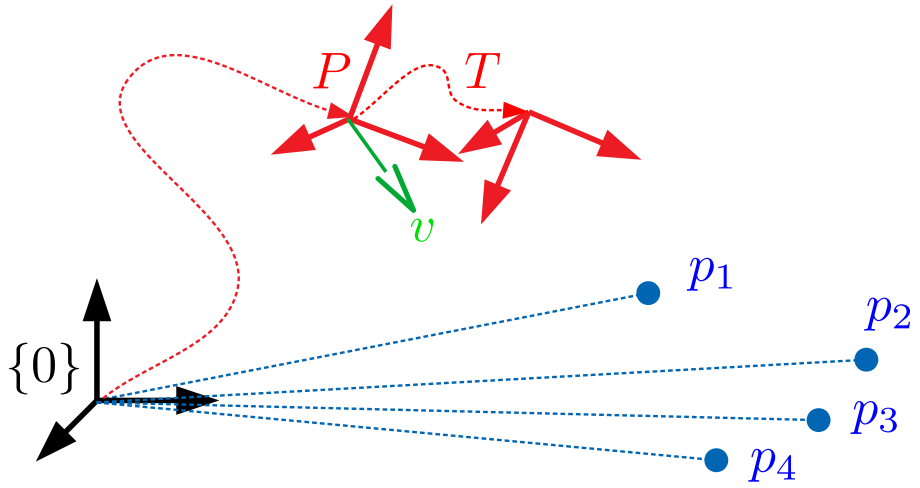


Figure 6.1: A diagram showing the states of visual-inertial SLAM. Note that the IMU biases are excluded here.

Let $\mathcal{M}^{\text{I}} := \mathbf{SE}(3) \times \mathbb{R}^3$ denote the *navigation state space* and define the *navigation state* $\xi_{\text{B}} := (P_{\text{B}}, v_{\text{B}}) \in \mathcal{M}^{\text{I}}$. We frequently use the notation $\xi := (\xi_{\text{B}}, b_{\text{B}}, T, p_i)$ 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,

$$\mathcal{E} := \{(\xi_{\text{B}}, b_{\text{B}}, T, p_i) \in \mathcal{M}^{\text{I}} \times \mathbf{SE}(3) \times (\mathbb{R}^3)^n \mid (P_{\text{B}}T)^{-1}p_i = 0 \text{ for any } i\},$$

corresponding to all situations where the camera centre coincides with a landmark point. To formalise this, we define the visual inertial SLAM (VI-SLAM) total space

$$\mathcal{T}^{\text{VI}}_n(3) := \mathcal{M}^{\text{I}} \times \mathbf{SE}(3) \times (\mathbb{R}^3)^n - \mathcal{E}$$

and consider the visual inertial SLAM problem on $\mathcal{T}^{\text{VI}}_n(3)$. Note that $\mathcal{T}^{\text{VI}}_n(3)$ is an open subset of a smooth manifold and as such is itself a smooth manifold.

Remark 6.3.1. *It is possible to identify elements of the navigation state space $\mathcal{M}^{\mathrm{I}}$ with the Lie group $\mathbf{SE}_2(3)$. However, this prescribes a group structure to $\mathcal{M}^{\mathrm{I}}$ that is not intrinsic. This distinction between the manifold state space and the Lie group symmetry is an important concept in equivariant systems theory [93]. In Section 6.4.1, the MEKF and an EqF for IMU dynamics are compared, and these filters may be viewed as applying different Lie groups to the navigation state space; $\mathbf{SO}(3) \times \mathbb{R}^3 \times \mathbb{R}^3$ for the MEKF and $\mathbf{SE}_2(3)$ for the EqF. The difference in performance shows the importance of distinguishing between state space and symmetry.*

6.3.1 VI-SLAM Dynamics

Let the acceleration due to gravity in the inertial frame $\{0\}$ be $g\mathbf{e}_3$, where $g \approx 9.81\ \mathrm{m/s^2}$ and $\mathbf{e}_3 \in \mathrm{S}^2$ is the standard direction of gravity in the inertial frame. Let the measured angular velocity and linear acceleration obtained from the IMU be $(\Omega, a) \in \mathbb{L} := \mathbb{R}^3 \times \mathbb{R}^3$, where $\mathbb{L}$ is the *input space*. Then the VI-SLAM dynamics $f : \mathbb{L} \to \mathfrak{X}(\mathcal{T}_n^{\mathrm{VI}}(3))$ are

$$\dot{\xi} = f_{(\Omega,a)}(\xi); \qquad \begin{aligned} \dot{R}_{\mathrm{B}} &= R_{\mathrm{B}}(\Omega - b_{\mathrm{B}}^{\Omega})^{\times}, \\ \dot{x}_{\mathrm{B}} &= v_{\mathrm{B}}, \\ \dot{v}_{\mathrm{B}} &= R_{\mathrm{B}}(a - b_{\mathrm{B}}^{a}) + g\mathbf{e}_3, \\ \dot{b}_{\mathrm{B}}^{\Omega} &= 0, \\ \dot{b}_{\mathrm{B}}^{a} &= 0, \\ \dot{T} &= 0, \\ \dot{p}_i &= 0. \end{aligned} \tag{6.1}$$

6.3.2 VI-SLAM Measurements

The camera measurements are modelled as n bearing measurements of the landmarks p_i in the camera frame $\{C\}$ on the manifold $\mathcal{N}_n^{\mathrm{V}}(3) := (\mathrm{S}^2)^n$ where the superscript "V" stands for visual measurements. The measurement function $h : \mathcal{T}_n^{\mathrm{VI}}(3) \to \mathcal{N}_n^{\mathrm{V}}(3)$ is given by

$$\begin{aligned} h(\xi) &:= \left(h^1(\xi), \ldots h^n(\xi)\right), \\ h^k(\xi_{\mathrm{B}}, b_{\mathrm{B}}, T, p_i) &:= \pi_{\mathrm{S}^2}\left((P_{\mathrm{B}}T)^{-1}(p_k)\right), \end{aligned} \tag{6.2}$$

where π_{S^2} is defined as in (3.1). Modelling the bearing measurements directly on the sphere rather than the image plane enables the proposed system to model a wide variety of monocular cameras. Consideration of different camera models is omitted from this chapter but is included in the EqVIO software package.

6.3.3 Invariance of Visual Inertial SLAM

Let $\mathbf{e}_3$ be the standard gravity direction and define the semi-direct product group

$$\mathrm{S}^1 \ltimes_{\mathbf{e}_3} \mathbb{R}^3 := \{(\theta, x) \,|\, \theta \in \mathrm{S}^1, x \in \mathbb{R}^3\},$$

with group product, identity and inverse

$$\begin{aligned}(\theta^1, x^1) \cdot (\theta^2, x^2) &= (\theta^1 + \theta^2, x^1 + R_{\mathbf{e}_3}(\theta^1)x^2),\\ \mathrm{id}_{\mathrm{S}^1 \ltimes_{\mathbf{e}_3}\mathbb{R}^3} &= (0, 0_{3\times 1}),\\ (\theta, x)^{-1} &= (-\theta, -R_{\mathbf{e}_3}(-\theta)x),\end{aligned}$$

where $R_{\mathbf{e}_3}(\theta) \in \mathbf{SO}(3)$ is the anti-clockwise rotation of an angle θ about the axis $\mathbf{e}_3$. Then $\mathrm{S}^1 \ltimes_{\mathbf{e}_3} \mathbb{R}^3$ is isomorphic to the subgroup

$$\mathbf{SE}_{\mathbf{e}_3}(3) := \{(R, x) \in \mathbf{SE}(3) \,|\, R\mathbf{e}_3 = \mathbf{e}_3\} \subset \mathbf{SE}(3).$$

Define $\alpha : \mathbf{SE}_{\mathbf{e}_3} \times \mathcal{T}_n^{\mathrm{VI}}(3) \to \mathcal{T}_n^{\mathrm{VI}}(3)$ by

$$\alpha(S, (\xi_{\mathrm{B}}, b_{\mathrm{B}}, T, p_i)) := (S^{-1}P_{\mathrm{B}}, R_S^\top v_{\mathrm{B}}, b_{\mathrm{B}}, T, S^{-1}(p_i)). \tag{6.3}$$

Then α is a right group action of $\mathbf{SE}_{\mathbf{e}_3}(3)$ on $\mathcal{T}_n^{\mathrm{VI}}(3)$. For a given $S \in \mathbf{SE}_{\mathbf{e}_3}(3)$, 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 $S \in \mathbf{SE}_{\mathbf{e}_3}(3)$ leaves the direction of gravity $\mathbf{e}_3$ unchanged.

Proposition 6.3.2. *The dynamics* (6.1) *and measurements* (6.2) *of VI-SLAM are invariant with respect to α; that is,*

$$\begin{aligned}f_{(\Omega,a)}(\alpha(S, (\xi_B, b_B, T, p_i))) &= \mathrm{d}\alpha_S f_{(\Omega,a)}(\xi_B, b_B, T, p_i),\\ h(\alpha(S, (\xi_B, b_B, T, p_i))) &= h(\xi_B, b_B, T, p_i),\end{aligned}$$

for any $S \in \mathbf{SE}_{\mathbf{e}_3}(3)$.

6.4 Symmetry of VI-SLAM

The VI-SLAM symmetry action proposed in this chapter combines the symmetry for IMU dynamics developed in [139] with the VSLAM symmetry developed in [1]. Before discussing the full symmetry group, we discuss the separate symmetries and how they lead to lower linearisation error in the filter development given in the sequel. This section of the chapter is written in a more tutorial style to provide the reader with intuition underlying the proposed algorithm.

The advantage of the extended Euclidean symmetry ($\mathbf{SE}_2(3)$) 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 linearisation 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 linearisation error in these coordinates than classical formulations. Section 6.4.1 is based on prior work by Barrau *et al.* [139].

The advantage of the scaled orthogonal transform ($\mathbf{SOT}(3)$) symmetry used for the landmark states lies in providing a framework in which the measurement linearisation error can be minimized. To make this point clear we analyse the common landmark parametrisations used in the literature and study the linearisation error. This provides a clear theoretical justification for the inverse depth parametrisation that is state-of-the-art in VIO algorithms and goes on to demonstrate that the $\mathbf{SOT}(3)$ symmetry leads to lower measurement linearisation error again. The material in Section 6.4.2 is novel to this chapter.

6.4.1 Symmetry of IMU Dynamics

Let $f^{\mathrm{I}} : \mathbb{L} \to \mathfrak{X}(\mathcal{M}^{\mathrm{I}})$ denote the IMU dynamics considered in (6.1) without bias; that is,

$$\frac{\mathrm{d}}{\mathrm{d}t}(R, x, v) = f_{(\Omega,a)}(R, x, v), \tag{6.4}$$

$$= (R\Omega^{\times}, v_{\mathrm{B}}, R_{\mathrm{B}}a + g\mathbf{e}_3,). \tag{6.5}$$

Filter designs such as the EKF, on-manifold EKF [140], MEKF [72] and EqF [2] 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 linearised to propagate the estimated distribution.

Traditional EKF designs model the robot's attitude through embedded coordinates as a unit quaternion $q \in \mathbb{H} \simeq \mathbb{R}^4$. This leads to the (over-parametrised) error coordinates,

$$\varepsilon_{\mathrm{EKF}}(\hat{\xi}_{\mathrm{B}}, \xi_{\mathrm{B}}) := \begin{pmatrix} q_{\mathrm{B}} - \hat{q}_{\mathrm{B}} \\ x_{\mathrm{B}} - x_{\hat{\xi}_{\mathrm{B}}} \\ v_{\mathrm{B}} - x_{\hat{\xi}_{\mathrm{B}}} \end{pmatrix}.$$

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

$$\varepsilon_{\mathrm{MEKF}}(\hat{\xi}_{\mathrm{B}}, \xi_{\mathrm{B}}) := \begin{pmatrix} \log_{\mathbf{SO}(3)}(R_{\mathrm{B}} R_{\hat{\xi}_{\mathrm{B}}}^{\top})^{\vee} \\ x_{\mathrm{B}} - x_{\hat{\xi}_{\mathrm{B}}} \\ v_{\mathrm{B}} - x_{\hat{\xi}_{\mathrm{B}}} \end{pmatrix}.$$

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

In an EqF design, the filter state is part of a Lie group rather than the state space $\mathcal{M}^{\text{I}}$ [2]. Consider the Lie group $\mathbf{SE}_2(3)$ with right group action $\varphi : \mathbf{SE}_2(3) \times \mathcal{M}^{\text{I}} \to \mathcal{M}^{\text{I}}$ given by

$$\varphi(A, \xi_{\text{B}}) := (R_{\text{B}} R_A, x_{\text{B}} + R_{\text{B}} x_A, v_{\text{B}} + R_{\text{B}} v_A), \tag{6.6}$$

where $A = (R_A, x_A, v_A) \in \mathbf{SE}_2(3)$. In this case, $\mathbf{SE}_2(3)$ and $\mathcal{M}^{\text{I}}$ 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 $\mathring{\xi}_{\text{B}} = (I_3, 0, 0) \in \mathcal{M}^{\text{I}}$, and define a local coordinate chart $\vartheta^{\text{B}}(\xi_{\text{B}}) := \log^{\vee}_{\mathbf{SE}_2(3)}(R_{\text{B}}, x_{\text{B}}, v_{\text{B}}) \in \mathbb{R}^9$. Then the EqF error coordinates are

$$\begin{aligned} \varepsilon_{\text{EqF}}(\hat{A}, \xi_{\text{B}}) &:= \vartheta^{\text{I}}(\varphi(\hat{A}^{-1}, \xi_{\text{B}})), \\ &= \log_{\mathbf{SE}_2(3)} \begin{pmatrix} R_{\text{B}} R_{\hat{A}}^{\top} & x_{\text{B}} - R R_{\hat{A}}^{\top} x_{\hat{A}} & v_{\text{B}} - R R_{\hat{A}}^{\top} v_{\hat{A}} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}^{\vee}, \end{aligned} \tag{6.7}$$

where $\hat{A} = (R_{\hat{A}}, x_{\hat{A}}, v_{\hat{A}}) \in \mathbf{SE}_2(3)$ is the filter state.

Each of these filters model their error coordinates as being drawn from a normal distribution with zero mean and covariance Σ, $\varepsilon \sim N(0, \Sigma)$, where Σ 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 linearisation of the error dynamics. For the EqF, however, this linearisation is dependent on the chosen symmetry group $\mathbf{G}$, 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 $\mathbf{G}$, and this results in an exactly linear propagation of the error coordinates [39].

Barrau and Bonnabel [139] showed that the bias-free IMU dynamics are group affine with respect to the action (6.6) of $\mathbf{SE}_2(3)$. As a result, using the Lie group $\mathbf{SE}_2(3)$, the propagation of bias-free IMU dynamics in the EqF has no linearisation 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 $\xi_{\text{B}}(0) = \exp_{\mathbf{SE}_2(3)}(\eta_0^{\wedge})$, where $\eta_0 \in \mathbb{R}^9$ is drawn from the distribution $N(0, \Sigma_0)$ with

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

Initialise 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/s^2.

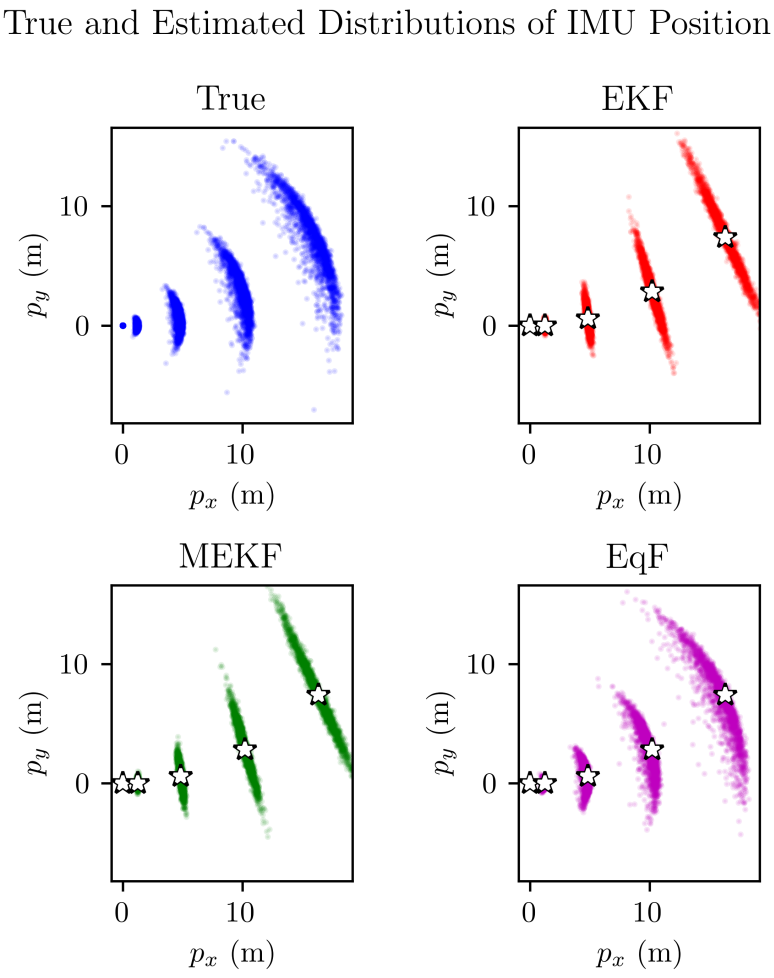


Figure 6.2: The true distribution of IMU positions at increments of 5 s obtained from integrating the dynamics (6.5) compared with the estimated distributions from the EKF, MEKF, and EqF.

Figure 6.2 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.

An EKF is not an optimal estimator, unlike the linear Kalman filter, due to the accumulation of linearisation errors over time. By exploiting symmetry properties as above, the linearisation 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.

6.4.2 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

$$\dot{q} = f^{\mathrm{V}}_{(\Omega_C, v_C)}(q) := -\Omega_C \times q - v_C. \tag{6.8}$$

The visual measurement of the landmark is

$$h^{\mathrm{V}}(q) = \frac{q}{|q|} \tag{6.9}$$

A parametrisation of the landmark is a diffeomorphism $\varsigma : U \subset \mathbb{R}^3 \to \mathcal{M}^{\mathrm{V}} \subset \mathbb{R}^k$, where U is an open subset of $\mathbb{R}^3$ and $\mathcal{M}^{\mathrm{V}}$ is a smooth 3-dimensional submanifold of $\mathbb{R}^k$ for some $k \geq 3$.

The *Euclidean* parametrisation was commonly used in early works on visual SLAM [65] and is defined by

$$\varsigma_{\mathrm{Euc}}(q) := q, \qquad \varsigma^{-1}_{\mathrm{Euc}}(z) := z. \tag{6.10}$$

The *inverse-depth* parametrisation, and variants thereof, are used more frequently in recent literature [142; 19; 77]. An archetypical example is given by

$$\varsigma_{\mathrm{ID}}(q) := \begin{pmatrix} q/|q| \\ 1/|q| \end{pmatrix}, \qquad \varsigma^{-1}_{\mathrm{ID}}(z) := \begin{pmatrix} I_3 & 0_{3\times 1} \end{pmatrix} \frac{z}{z_4}. \tag{6.11}$$

We introduce a new parametrisation for landmarks with visual measurements by exploiting the $\mathbf{SOT}(3)$ symmetry actions developed in [1]. The *polar* parametrisation is a novel parametrisation for visual landmarks, defined by

$$\begin{aligned} \varsigma_{\mathbf{SOT}(3)}(q) &:= \begin{pmatrix} -\arccos(\frac{\mathbf{e}_3^\top q}{|q|}) \frac{\mathbf{e}_3 \times q}{|\mathbf{e}_3 \times q|} \\ -\log(\frac{q}{|q|}) \end{pmatrix}, \\ \varsigma^{-1}_{\mathbf{SOT}(3)}(z) &:= \exp_{\mathbf{SOT}(3)}(z^\wedge_{\mathfrak{sot}(3)})^{-1} \mathbf{e}_3. \end{aligned} \tag{6.12}$$

This parametrisation provides normal coordinates for $\mathbb{R}^3$ about $\mathbf{e}_3$ with respect to the right action of $\mathbf{SOT}(3)$ defined in Lemma 6.4.1.

Lemma 6.4.1. *Let* $\varphi^V : \mathbf{SOT}(3) \times \mathbb{R}^3 \setminus \{0\} \to \mathbb{R}^3 \setminus \{0\}$ *and* $\rho^V : \mathbf{SOT}(3) \times \mathrm{S}^2 \to \mathrm{S}^2$ *be defined by*

$$\begin{aligned} \varphi^V(Q, q) &:= c_Q^{-1} R_Q^\top q, \\ \rho^V(Q, y) &:= R_Q^\top y. \end{aligned} \tag{6.13}$$

Then φ^V *and* ρ^V *are transitive right group actions, and the visual measurement function* (6.9) *is*

equivariant with respect to these actions; that is,

$$h^V(\varphi^V(Q,q)) = \rho^V(Q, h^V(q)),$$

for all $Q \in \mathbf{SOT}(3)$ *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 *parametrised state error* is

$$\varepsilon := \varsigma(q) - \varsigma(\hat{q}),$$

for a particular parametrisation ς. The true state can be identified in terms of the estimated state and the parametrised state error,

$$q = \varsigma^{-1}(\varepsilon + \varsigma(\hat{q})).$$

Linearising the dynamics (6.8) and measurement (6.9) in terms of ε about $\varepsilon = 0$ yields

$$\begin{aligned} f^{\mathrm{V}}_{(\Omega,v)}(q) &= f^{\mathrm{V}}_{(\Omega,v)} \circ \varsigma^{-1}(\varepsilon + \varsigma(\hat{q})) && (6.14)\\ &= f^{\mathrm{V}}_{(\Omega,v)}(\hat{q}) + \mathrm{D}_q|_{\hat{q}} f^{\mathrm{V}}_{(\Omega,v)}(q) \cdot \mathrm{D}_s|_{\varsigma(\hat{q})}\varsigma(s)[\varepsilon] + \mu^f(q), \\ h^{\mathrm{V}}(q) &= h^{\mathrm{V}} \circ \varsigma^{-1}(\varepsilon + \varsigma(\hat{q})) && (6.15)\\ &= h^{\mathrm{V}}(\hat{q}) + \mathrm{D}_p|_{\hat{q}} h^{\mathrm{V}}(q) \cdot \mathrm{D}_s|_{\varsigma(\hat{q})}\varsigma(s)[\varepsilon] + \mu^h(q), \end{aligned}$$

where $\mu^f(q)$ and $\mu^h(q)$ are the dynamics and measurement linearisation 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 (6.8) are exactly linear in the Euclidean parametrisation (6.10), and hence $\mu^f_{\mathrm{Euc}}(q) \equiv 0$. Second, the equivariant output approximation proposed in [2] is available when using the polar parametrisation as these are the normal coordinates associated with the action φ^{V} (6.13). Specifically, applying [2, Lemma V.3] to this example yields

$$\begin{aligned} C^\star_t \varepsilon &= \frac{1}{2}(\mathrm{D}_E|_{\mathrm{id}}\rho^{\mathrm{V}}(E,y) + \mathrm{D}_E|_{\mathrm{id}}\rho^{\mathrm{V}}(E,\mathbf{e}_3))\varepsilon^\wedge_{\mathfrak{sot}(3)}, \\ &= \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}|_{t=0}(\rho^{\mathrm{V}}(e^{t\varepsilon^\wedge_{\mathfrak{sot}(3)}}, y) + \rho^{\mathrm{V}}(e^{t\varepsilon^\wedge_{\mathfrak{sot}(3)}}, \mathbf{e}_3)), \\ &= \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}|_{t=0}(-R_{e^{t\varepsilon^\wedge_{\mathfrak{sot}(3)}}} y - R_{e^{t\varepsilon^\wedge_{\mathfrak{sot}(3)}}} \mathbf{e}_3), \\ &= \frac{1}{2}(-\Omega^\times_\varepsilon y - \Omega^\times_\varepsilon \mathbf{e}_3), \\ &= \frac{1}{2}(y + \mathbf{e}_3)^\times \Omega_\varepsilon, \end{aligned}$$

and therefore,

$$C^\star_t = \begin{pmatrix} \frac{1}{2}(y+\mathbf{e}_3)^\times & 0_{3\times 1}. \end{pmatrix}$$

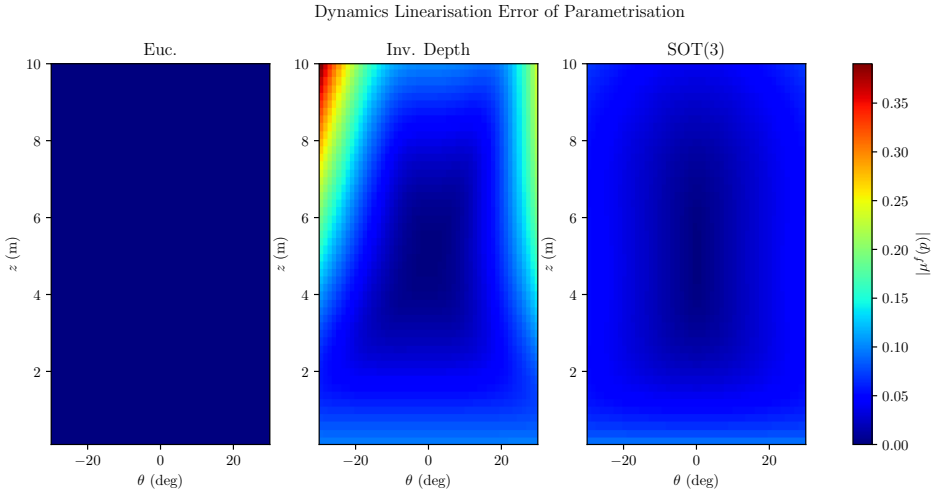


Figure 6.3: The norm of linearisation error (6.14) of the landmark dynamics (6.8) for the Euclidean (6.10), inverse-depth (6.11), and polar (6.12) parametrisations over the domain defined in (6.16). The Euclidean parametrisation has zero dynamics linearisation error since the dynamics are exactly linear in these coordinates.

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

To see the effects of different parametrisations on the linearisation errors, let $\hat{q} = (0, 0, 5)$ and consider the domain

$$U = \{(z\tan(\theta), 0, z) \in \mathbb{R}^3 \mid z \in (0.1, 10),\ \theta \in (-\frac{\pi}{6}, \frac{\pi}{6})\}. \tag{6.16}$$

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.

Figures 6.3 and 6.4 show the dynamics and output linearisation errors $\mu^f(q)$ and $\mu^h(q)$, respectively, for each of the parametrisation (6.10,6.11,6.12) Figure 6.4 also shows the equivariant output approximation [2] in the polar parametrisation. As expected, the Euclidean parametrisation yields no dynamics linearisation error over any part of the domain. Meanwhile, the inverse-depth and polar parametrisations have non-zero dynamics linearisation errors that increase toward the edges of the domain, with the polar parametrisation having a lower maximum linearisation error. On the other hand, for the output linearisation there is a clear advantage to using the inverse-depth or polar over the Euclidean parametrisation. Finally, while the ordinary performance of the inverse-depth and polar parametrisations is similar, the polar parametrisation is able to use the equivariant output approximation which greatly reduces its output linearisation error everywhere in U.

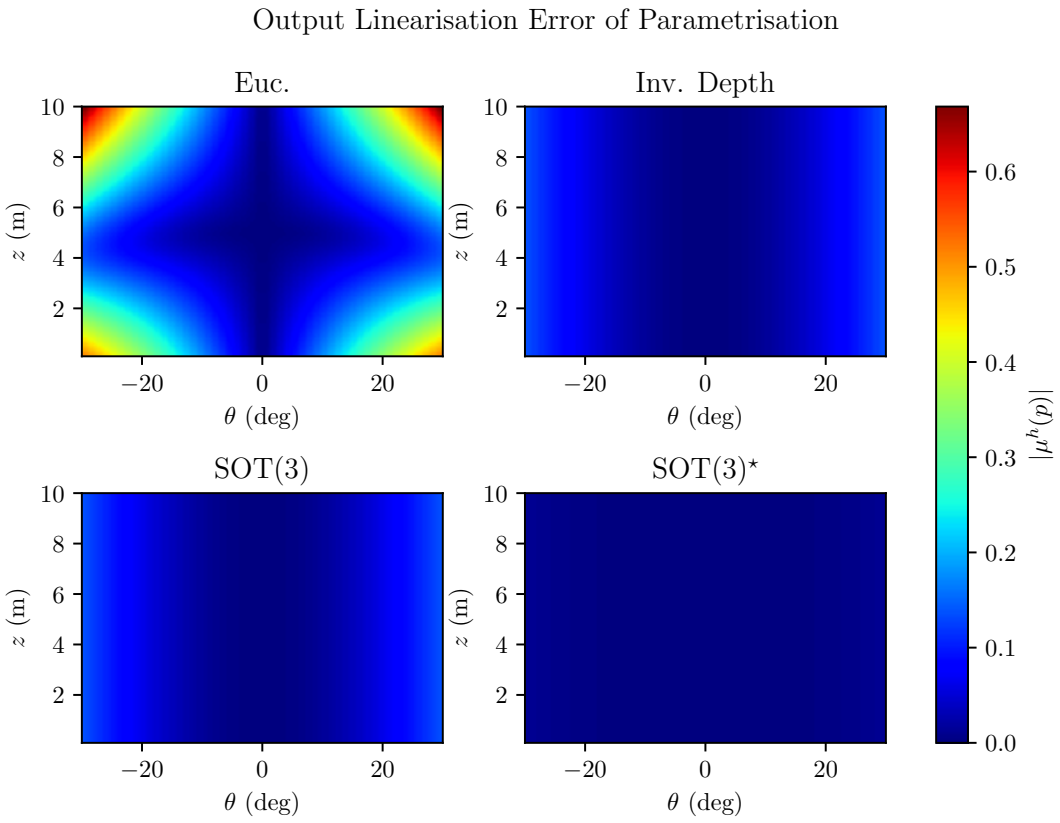


Figure 6.4: The norm of linearisation error (6.15) of the visual landmark measurement (6.9) for the Euclidean (6.10), inverse-depth (6.11), and polar (6.12) parametrisations over the domain defined in (6.16). The bottom-right subplot shows the linearisation error obtained when applying the equivariant output approximation [1] in the polar parametrisation.

6.4.3 Composite Symmetry for VI-SLAM

Define the *VI-SLAM Group* to be

$$\mathbf{SLAM}_n^{\text{VI}}(3) := \mathbf{SE}_2(3) \times \mathbf{SOT}(3)^n. \tag{6.17}$$

This is a product Lie group in the sense of (3.2). This group captures the symmetries that are fundamental to the VIO system: $\mathbf{SE}_2(3)$ for the IMU dynamics and $\mathbf{SOT}(3)^n$ 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 $\mathbf{SE}(3)$ and an $\mathbb{R}^6$ symmetry can be used, respectively. To this end, define the extended direct product group

$$\begin{aligned} \mathbf{G} &:= \mathbf{SE}_2(3) \times \mathbb{R}^6 \times \mathbf{SE}(3) \times \mathbf{SOT}(3)^n, \\ &\simeq \mathbf{SLAM}_n^{\text{VI}}(3) \times \mathbb{R}^6 \times \mathbf{SE}(3). \end{aligned} \tag{6.18}$$

We denote a typical element as $X = (A, \beta, B, Q_1, ..., Q_n) \in \mathbf{G}$, and frequently use the shorthand (A, β, B, Q_i).

Lemma 6.4.2. *The map* $\phi : \mathbf{G} \times \mathcal{T}_n^{VI}(3) \to \mathcal{T}_n^{VI}(3)$ *defined by*

$$\begin{aligned}&\phi((A, \beta, B, Q_i), (\xi_B, b_B, T, p_i))\\ &\quad := (\varphi^B(A, \xi_B), b_B + \beta, P_A^{-1} T B, P_B T B Q_i^{-1} T^{-1} P_B^{-1}(p_i)),\end{aligned} \tag{6.19}$$

where φ^B *is defined as in* (6.6)*, is a transitive right group action.*

Lemma 6.4.3. *The group action* ϕ (6.19) *is compatible with the VI-SLAM invariance* α (6.3)*; that is,*

$$\phi(X, \alpha(S, \xi)) = \alpha(S, \phi(X, \xi)),$$

for all $\xi \in \mathcal{T}_n^{VI}(3)$*,* $S \in \mathbf{SE}_{\mathbf{e}_3}(3)$ *,and* $X \in \mathbf{G}$*.*

Lemma 6.4.4. *The map* $\rho : \mathbf{G} \times \mathcal{N}_n^{V}(3) \to \mathcal{N}_n^{V}(3)$ *defined by*

$$\rho((A, \beta, B, Q_i), (y_i)) := (R_{Q_i}^{\top} y_i), \tag{6.20}$$

is a right group action. Additionally, the measurement function (6.2) *is equivariant with respect to the actions* ϕ (6.19) *and* ρ*; that is,*

$$h(\phi(X, \xi)) = \rho(X, h(\xi)),$$

for all $X \in \mathbf{G}$ *and* $\xi \in \mathcal{T}_n^{VI}(3)$*.*

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

$$\begin{array}{ccccc}
\mathcal{T}_n^{\mathrm{VI}}(3) & \xrightarrow{\alpha_S} & \mathcal{T}_n^{\mathrm{VI}}(3) & \xrightarrow{h} & \mathcal{N}_n^{\mathrm{V}}(3) \\
\Big\downarrow{\scriptstyle \phi_X} & & \Big\downarrow{\scriptstyle \phi_X} & & \Big\downarrow{\scriptstyle \rho_X} \\
\mathcal{T}_n^{\mathrm{VI}}(3) & \xrightarrow{\alpha_S} & \mathcal{T}_n^{\mathrm{VI}}(3) & \xrightarrow{h} & \mathcal{N}_n^{\mathrm{V}}(3)
\end{array}$$

which commutes for any $X \in \mathbf{G}$ and $S \in \mathbf{SE}_{\mathbf{e}_3}(3)$.

6.5 Equivariant Filter 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 6.5.1, Lemma 6.4.2 is used to lift the system dynamics from the state space manifold to the chosen symmetry group $\mathbf{G}$, and the resulting lift Λ is shown to be invariant to

the reference frame transformation α (6.3). Then, since both the group action ϕ (6.19) and the lift Λ (6.21) are compatible with α (6.3) (c.f. Lemmas 6.4.3, 6.5.1), the proposed EqF naturally also respects the invariance. Specifically, if the Riccati matrix Σ of the EqF is treated as a covariance, then the Fisher information Σ^{-1} is non-increasing along the unobservable directions spanned by the differential of α, and therefore the EqF is consistent by [143, Theorem 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 6.4.4, and the local coordinates ϑ chosen in Section 6.5.2 are normal with respect to the group action ϕ (6.19). Therefore, the equivariant output approximation $C_t^\star$ is available [2, 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. That is,

$$y - h(\hat{\xi}) = C_t^\star \varepsilon + O(|\varepsilon|^3),$$

where y is the true measurement, $\hat{\xi}$ is the EqF state estimate, and ε is the local error coordinates of the EqF.

6.5.1 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 (6.1) in the sense of [45]. Although [45] provides a constructive algorithm to build lifts, in practice it is easiest to guess the lift and then prove it satisfies the required conditions.

Lemma 6.5.1. *The map* $\Lambda : \mathcal{T}_n^{VI}(3) \times (\mathbb{R}^3 \times \mathbb{R}^3) \to \mathfrak{g}$, *given by*

$$\begin{aligned}
&\Lambda((\xi_B, b, T, p_i), (\Omega, a)) \\
&\quad := \left((\Omega, R_B^\top v, a)^\wedge_{\mathfrak{se}_2(3)},\ 0,\ (\Omega_C, v_C)^\wedge_{\mathfrak{se}(3)}, \left(\Omega_C + \frac{q_i^\times v_C}{\|q_i\|^2}, \frac{q_i^\top v_C}{\|q_i\|^2} \right)_i \right), \\
&q_i := (PT)^{-1}(p_i), \quad (\Omega_C, v_C)^\wedge_{\mathfrak{se}(3)} := \mathrm{Ad}_{T^{-1}}(\Omega, R_B^\top v)^\wedge_{\mathfrak{se}(3)},
\end{aligned} \tag{6.21}$$

is a lift [45] of the system dynamics (6.1)*. That is,*

$$\mathrm{D}_E|_{\mathrm{id}} \phi_{(\xi_B, b, T, p_i)}(E) \Lambda((\xi_B, b, T, p_i), (\Omega, a)) = f_{(\Omega, a)}(\xi_B, b, T, p_i) \tag{6.22}$$

Moreover, Λ *is invariant with respect to the action* α (6.3)*; i.e.*

$$\Lambda(\alpha(S, (\xi_B, b, T, p_i)), (\Omega, a)) = \Lambda((\xi_B, b, T, p_i), (\Omega, a)),$$

for all $S \in \mathbf{SE}_{\mathbf{e}_3}(3)$.

Proof. The proof that Λ satisfies the lift condition (6.22) closely follows the proof of [1, Lemma 4.4], and the invariance of Λ to α is straightforward. Both have been omitted to save space. □

6.5.2 Origin Choice and Local Coordinates

Let $\mathring{\xi} = (\mathring{\xi}_{\text{B}}, \mathring{b}_{\text{B}}, \mathring{T}, \mathring{p}_i) \in \mathcal{T}_n^{\text{VI}}(3)$ denote the chosen, fixed *origin configuration*, with $\mathring{p}_i := \mathring{P}_{\text{B}}\mathring{T}\mathbf{e}_3$ for every i. For generality, we leave the remaining terms $\mathring{\xi}, \mathring{b}_{\text{B}}, \mathring{T}$ arbitrary.

Define the map $\vartheta : \mathcal{U}_{\mathring{\xi}} \subset \mathcal{T}_n^{\text{VI}}(3) \to \mathbb{R}^{21+3n}$ by

$$\vartheta(\xi_{\text{B}}, b, T, p_i) := \begin{pmatrix} \log_{\mathbf{SE}_2(3)}(\mathring{R}_{\text{B}}^\top R_{\text{B}}, \mathring{R}_{\text{B}}^\top(x_{\text{B}} - \mathring{x}_{\text{B}}), \mathring{R}_{\text{B}}^\top(v_{\text{B}} - \mathring{v}_{\text{B}}))^\vee \\ b_{\text{B}} - \mathring{b}_{\text{B}} \\ \log_{\mathbf{SE}(3)}((\mathring{P}_{\text{B}}\mathring{T})^{-1}(P_{\text{B}}T))^\vee \\ \varsigma_{\mathbf{SOT}(3)}((P_{\text{B}}T_{\text{B}})^{-1}(p_1)) \\ \vdots \\ \varsigma_{\mathbf{SOT}(3)}((P_{\text{B}}T_{\text{B}})^{-1}(p_n)) \end{pmatrix}, \tag{6.23}$$

to be the coordinate chart for $\mathcal{T}_n^{\text{VI}}(3)$ about $\mathring{\xi}$, where $\mathcal{U}_{\mathring{\xi}}$ is a large neighbourhood of $\mathring{\xi}$ and $\varsigma_{\mathbf{SOT}(3)}$ is the polar parametrisation (6.12). Then ϑ provides normal coordinates for $\mathcal{T}_n^{\text{VI}}(3)$ about $\mathring{\xi}$ with respect to the action ϕ.

6.5.3 EqF Dynamics

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

Let $\hat{X} = (\hat{A}, \hat{\beta}, \hat{B}, \hat{Q}_i) \in \mathbf{G}$ 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 $\mathring{A}_t, B_t, C_t^\star$ to be the state, input, and equivariant output matrices of the EqF as defined in [2], respectively, and let $\mathrm{D}_E|_{\mathrm{id}}\phi_{\mathring{\xi}}(E)^\dagger$ be a fixed right-inverse of $\mathrm{D}_E|_{\mathrm{id}}\phi_{\mathring{\xi}}(E)$. Then the EqF dynamics [2] are defined to be

$$\begin{aligned} \dot{\hat{X}} &= \hat{X}\Lambda(\phi(\hat{X}, \mathring{\xi}), (\Omega, a)) - \Delta\hat{X}, \\ \Delta &= \mathrm{D}_E|_{\mathrm{id}}\phi_{\mathring{\xi}}(E)^\dagger \cdot \mathrm{D}_\xi|_{\mathring{\xi}}\vartheta(\xi)^{-1}\Sigma C_t^{\star\top} N_t^{-1}(y - h(\hat{\xi})), \\ \dot{\Sigma} &= \mathring{A}_t\Sigma + \Sigma\mathring{A}_t^\top + M_\varepsilon + B_t M_t^m B_t^\top - \Sigma C_t^{\star\top} N_t^{-1} C_t^\star \Sigma, \end{aligned} \tag{6.24}$$

where $\Delta \in \mathfrak{g}$ is the EqF correction term, and

- $\Sigma(0) = \Sigma_0 \in \mathbb{S}_+(21+3n)$ is the initial Riccati gain,
- $M_\varepsilon \in \mathbb{S}_+(21+3n)$ is the state gain,
- $M_t^m \in \mathbb{S}_+(6)$ is the input gain,
- $N_t \in \mathbb{S}_+(2n)$ is the output gain,

following their descriptions in [2].

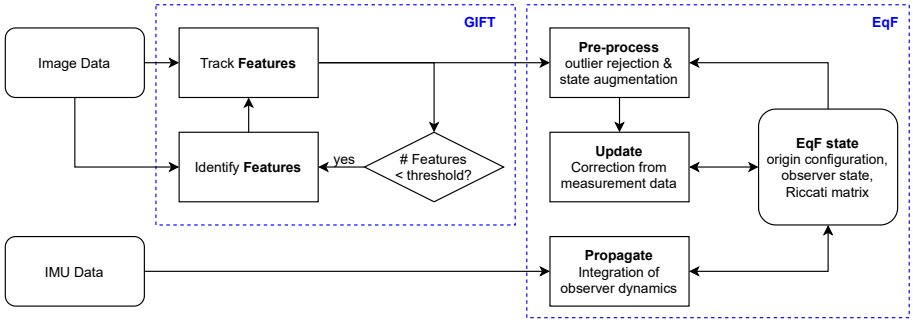


Figure 6.5: An overview of EqVIO as a system. The key components can be split into the front-end (GIFT) and the back-end (EqF).

At any time, the EqF state estimate is given by

$$\hat{\xi} = (\hat{\xi}_{\mathrm{B}}, \hat{b}, \hat{T}, \hat{p}_i) := \phi(\hat{X}, \mathring{\xi}). \tag{6.25}$$

6.6 Experimental Results

The EqF equations (6.24) are discretised (cf. § 6.6.1) and implemented in c++17 using the Eigen matrix library [144]. Visual measurements of landmarks are obtained using GIFT [1; 145] to identify and track image features. We refer to the resulting system as *EqVIO*. The four key steps in EqVIO are

- **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.
- **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.
- **Propagation:** Upon receiving an IMU signal, the EqF state and Ricatti equation are integrated without the correction terms.
- **Update:** After augmenting the state and removing outliers, the EqF state and Ricatti equation are integrated exclusively with the correction terms.

Figure 6.5 provides an overview of the system components and how each of the steps are linked.

6.6.1 Discretisation of EqF Equations

Implementation of the continuous-time EqF equations (6.24) requires them to be discretised in time. Additionally, the implementation has to respect the fact that input and output measurements are not necessarily synchronous, leading to two separate steps similar

to a discrete-time extended Kalman filter. A principled formulation of a discrete-time EqF is developed in [146], but this topic is beyond the scope of this thesis.

Discretisation of Propagation

Upon receiving an IMU input signal, the state of the EqF is propagated by discretising the relevant terms of (6.24). Consider the filter propagation equations

$$\begin{aligned}\dot{\hat{X}} &= \hat{X}\Lambda(\phi(\hat{X}, \mathring{\xi}), (\Omega, a)),\\ \dot{\Sigma} &= \mathring{A}_t\Sigma + \Sigma\mathring{A}_t^\top + M_\varepsilon + B_t M_t^m B_t^\top.\end{aligned}$$

The first equation may be evaluated approximately as follows. Given an initial condition $\hat{X}(t)$, an input signal (Ω, a), and a period of time $\tau > 0$ over which to integrate, the approximated solution $\hat{X}(t+\tau)$ is implemented as

$$\begin{aligned}\hat{X}(t+\tau) &= \hat{X}(t)\cdot(A^\Lambda, \beta^\Lambda, B^\Lambda, Q_1^\Lambda, ..., Q_n^\Lambda);\\ P_{A^\Lambda} &= \exp_{\mathbf{SE}(3)}\left(\tau\begin{pmatrix}(\Omega - \hat{b}^\Omega_{\text{B}})^\times & \hat{R}^\top_{\text{B}}\hat{v}_{\text{B}} + \frac{\tau}{2}((a-\hat{b}^a_{\text{B}}) - \hat{R}^\top_{\text{B}}g)\\ 0_{1\times 3} & 1\end{pmatrix}\right),\\ v_{A^\Lambda} &= \tau((a-\hat{b}^a_{\text{B}}) - \hat{R}^\top_{\text{B}}g),\\ \beta^\Lambda &= 0,\\ B^\Lambda &= \hat{T}^{-1}P_{A^\Lambda}\hat{T},\\ R_{Q_i^\Lambda} &= \text{Rot}\left(\pi_{\text{S}^2}((B^\Lambda)^{-1}\hat{q}_i), \pi_{\text{S}^2}(\hat{q}_i)\right),\\ c_{Q_i^\Lambda} &= \frac{|\hat{q}_i|}{|(B^\Lambda)^{-1}\hat{q}_i|},\\ \hat{q}_i &:= (\hat{P}\hat{T})^{-1}\hat{p}_i,\end{aligned}$$

where all state-dependent quantities on the right-hand side are as given at time t, and $\text{Rot} : \text{S}^2 \times \text{S}^2 \to \mathbf{SO}(3)$ yields a rotation matrix taking the first argument to the second; that is, $\text{Rot}(\eta_1, \eta_2)\eta_1 = \eta_2$.

The approximate Riccati matrix solution $\Sigma(t+\tau)$ is given by

$$\Sigma(t+\tau) = \exp(\tau\mathring{A}_t)\Sigma(t)\exp(\tau\mathring{A}_t)^\top + \tau(M_\varepsilon + B_t M_t^m B_t^\top),$$

where $\mathring{A}_t$ and B_t are computed based on the state and measurement at time t. An important advantage of this approximation is that it preserves positive-definiteness of the Riccati matrix.

Discretisation of Update

When an image is received by the system, features are tracked from the previous image. The tracked features are then used to update the EqF state with the new information. Let $y = (y_1, ..., y_n) \in (\text{S}^2)^n$ denote the feature measurements obtained from the image

Table 6.1: Popular VIO algorithms used to compare with the performance of EqVIO.

Algorithm	Ref.	back-end	front-end
EqVIO	*	EqF	GFTT + LKT
ROVIO	[19]	iterated EKF	image patches
OpenVINS	[77]	MSCKF + EKF	FAST + LKT
VINS-mono	[76]	sliding-window	GFTT + LKT + ORB
MSCKF	[72]	MSCKF	FAST + LKT
OKVIS	[74]	sliding-window	GFTT + BRISK

received at time t. Let $\hat{X}(t^-)$ and $\Sigma(t^-)$ denote the observer state at the time immediately prior to receiving the image. Then the updated observer states $\hat{X}(t)$ and $\Sigma(t)$ are approximated by

$$\begin{aligned}
\hat{X}(t) &= \exp_{\mathbf{G}}(\Delta)\hat{X}(t^-), \\
\Delta &= \mathrm{D}_E|_{\mathrm{id}}\phi_{\mathring{\xi}}(E)^\dagger \mathrm{D}_\xi|_{\mathring{\xi}}\vartheta(\xi)^{-1}K(y(t) - h(\hat{\xi}(t^-))), \\
\Sigma(t) &= (I - KC^\star_{t^-})\Sigma(t^-), \\
K &:= \Sigma(t^-){C^\star_{t^-}}^\top N_t^{-1}.
\end{aligned}$$

Unlike the propagation discretisation, the update step is not obtained by approximating an integral. Instead, it is treated as an instantaneous correction using the new information. The form above is similar to the equations of the standard EKF, except for the important difference in how the update step Δ is applied to the observer state $\hat{X}$.

6.6.2 Performance Comparisons

We compared the performance of EqVIO to other VIO systems on two popular public datasets: EuRoC [20] and UZH FPV [21]. Table 6.1 lists the algorithms considered along with information about the back-end and front-end systems used, including GFFT [147], LKT [148], FAST [149], ORB [150], and BRISK [151].

Where stated, the algorithm performance results have been obtained from the recent benchmark study by Delmerico and Scaramuzza [49]. 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 8-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 are including loop closure and map reuse were disabled where relevant to ensure a fair comparison between algorithms.

Table 6.2 lists the root mean square error (RMSE) of the position estimates of each of the systems in Table 6.1 on the EuRoC dataset. EqVIO achieves the best performance in many of the sequences, and also achieves the best mean performance. Additionally, Table 6.3 lists the processing time taken per frame for each of the algorithms tested on the authors' desktop. On average, EqVIO is faster than the next fastest algorithm by a factor

of 2.1 on the EuRoC dataset.

Table 6.2: RMSE of position estimates in metres on the EuRoC [20] dataset for each of the systems in Table 6.1.

Algorithm	EqVIO	ROVIO	OpenVINS	VINS-Mono	MSCKF	OKVIS
Source	*	[49]	*	*	[49]	[49]
MH_01	0.14	0.21	**0.09**	0.24	0.42	0.16
MH_02	**0.14**	0.25	0.24	0.21	0.45	0.22
MH_03	**0.10**	0.25	0.16	0.15	0.23	0.24
MH_04	**0.19**	0.49	0.31	0.23	0.37	0.34
MH_05	**0.26**	0.52	0.39	0.28	0.48	0.47
V1_01	**0.07**	0.10	0.08	**0.07**	0.34	0.09
V1_02	0.13	0.10	**0.09**	0.28	0.20	0.20
V1_03	0.22	0.14	**0.06**	0.16	0.67	0.24
V2_01	**0.06**	0.20	0.08	0.07	0.10	0.13
V2_02	0.16	0.14	**0.07**	0.16	0.16	0.16
V2_03	0.18	**0.14**	0.15	0.22	1.13	0.29
Mean	**0.15**	0.23	0.16	0.19	0.41	0.23

Table 6.4 lists the RMSE of the position estimates for a subset of the systems in Table 6.1 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. As in the EuRoC dataset, EqVIO achieves the best performance in many of the sequences and the best mean performance. Additionally, from Table 6.5 it is clear that EqVIO is also significantly faster than any of the other algorithms considered. On average, EqVIO is faster than the next fastest algorithm by a factor of 6.0 on the UZH FPV dataset. EqVIO achieved an average processing speed of 338 Hz over all the sequences, and a maximum processing speed of 400 Hz on indoor_45_13.

6.6.3 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.

Figure 6.6 shows the time taken to process every frame of the sequence as a flamegraph, and Figure 6.7 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

Table 6.3: Average time taken to process each frame in ms on the EuRoC [20] dataset for some of the systems in Table 6.1.

Algorithm	EqVIO	ROVIO	OpenVINS	VINS-Mono
Source	*	*	*	*
MH_01	**5.41**	13.42	10.27	36.61
MH_02	**5.16**	15.51	10.84	34.86
MH_03	**6.00**	16.10	11.92	34.35
MH_04	**6.08**	19.27	11.40	34.18
MH_05	**5.44**	19.38	11.46	35.87
V1_01	**6.10**	19.91	13.45	33.92
V1_02	**4.97**	21.74	12.70	26.31
V1_03	**5.75**	18.02	11.07	24.84
V2_01	**6.01**	35.99	12.01	42.78
V2_02	**5.37**	19.94	11.59	28.10
V2_03	**5.78**	20.70	10.52	22.99
Mean	**5.64**	20.00	11.57	32.26

for any frame is still comparable to the average processing time of the other algorithms listed in Table 6.3.

The absolute position and the attitude about the direction of gravity are unobservable due to the reference frame invariance described in Proposition 6.3.2. 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. Figure 6.8 shows the estimated and true values of the body-fixed gravity and linear velocity $R_{\text{B}}^{\top} v_{\text{B}}$ 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.

6.6.4 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. Figure 6.9 shows how the position RMSE and processing time of EqVIO change for varying numbers of landmarks. EqVIO was run over the entire EuRoC dataset with a single set of parameters, except that the maximum number of landmarks allowed was changed from 4 to 100 in increments of 2 for each trial. The processing time per frame and position RMSE were averaged over the whole dataset for each trial. Due to the similarities between the EqF and an EKF, the processing time indeed appears to increase quadratically with the number of landmarks considered. In contrast, the position RMSE decreases as the number of landmarks increases, although this trend is less consistent. The experiment shows that the number of landmarks used in EqVIO may be increased to yield improved

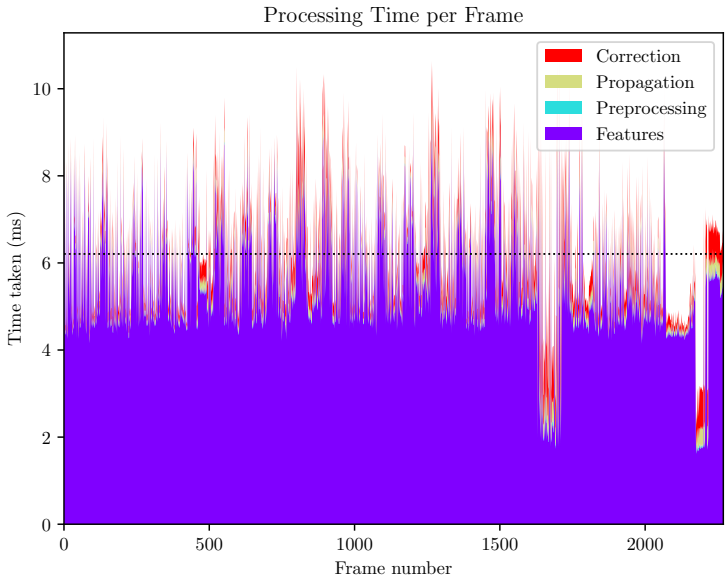


Figure 6.6: A flame graph showing the time taken to process each frame of the EuRoC sequence V2_01.

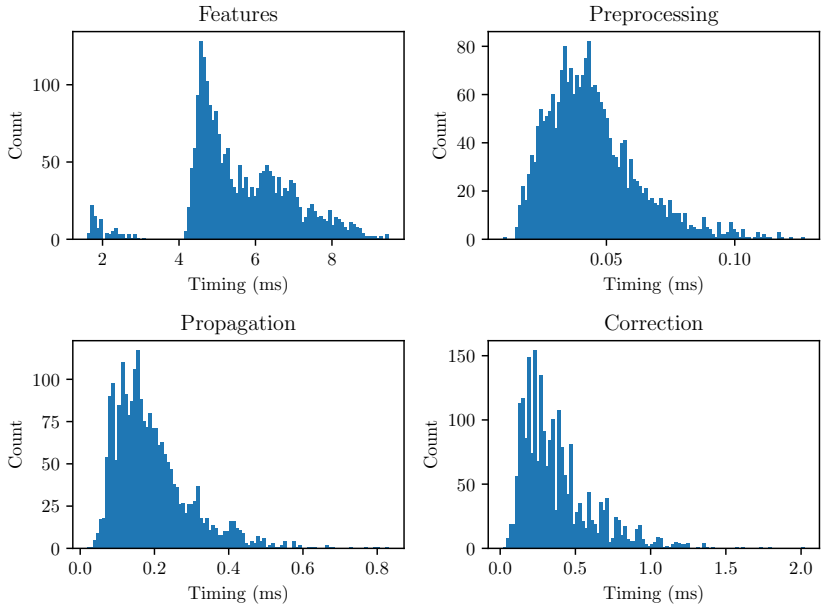


Figure 6.7: Histograms of the time taken to process each frame by each section of EqVIO in the EuRoC sequence V2_01.

Table 6.4: RMSE of position estimates in metres on the UZH FPV [21] dataset for some of the systems in Table 6.1.

Algorithm	EqVIO	ROVIO	OpenVINS	VINS-Mono
Source	*	*	*	*
indoor_45_12	0.63	x	0.63	**0.62**
indoor_45_13	**0.43**	1.66	0.50	0.82
indoor_45_14	1.79	3.36	**1.76**	1.85
indoor_45_2	0.40	x	**0.33**	0.69
indoor_45_4	**0.28**	x	0.35	1.15
indoor_forward_10	**0.32**	1.22	0.36	0.61
indoor_forward_5	0.23	2.46	**0.22**	0.32
indoor_forward_6	0.58	0.90	**0.53**	0.62
indoor_forward_7	**0.34**	1.07	0.43	0.56
indoor_forward_9	**0.64**	x	0.82	0.76
Mean	**0.56**	1.78	0.59	0.76

accuracy at the cost of increased processing time. This trade-off is useful in allowing a practitioner to tune the proposed system according to their requirements.

6.7 Chapter Conclusion

This chapter presents and develops EqVIO: a novel system for visual inertial odometry based on the recently proposed equivariant filter. A new Lie group, the VI-SLAM group, is developed for the VIO problem. It is shown that this symmetry is compatible with the well-known reference frame invariance of VIO, and that therefore the resulting EqF is a naturally consistent estimator. The VI-SLAM group incorporates the extended special Euclidean group $\mathbf{SE}_2(3)$ proposed in [139], leading to exact linearisation of the error dynamics associated with the navigation states. The VI-SLAM group also takes advantages of the $\mathbf{SOT}(3)$ symmetry of visual landmarks to enable the EqF to use the higher-order equivariant output approximation [2]. EqVIO is the system that combines an EqF based on the VI-SLAM group with a feature tracking front-end. In experimental results on two popular VIO datasets [20; 21], we show that EqVIO outperforms other state-of-the-art algorithms in terms of both speed of computation and accuracy of trajectory estimation. In summary, this chapter shows the compatibility of the novel VI-SLAM Lie group with the VIO problem, and that the resulting equivariance-based VIO system significantly outperforms alternative solutions.

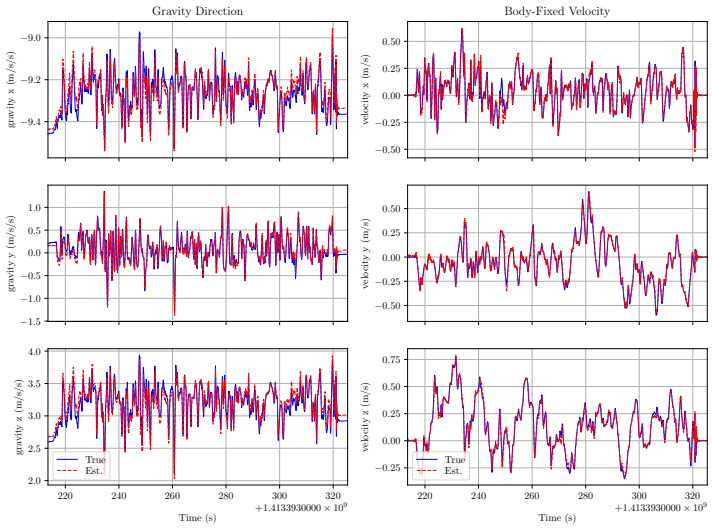


Figure 6.8: Comparison of estimated and true values of body-fixed velocity and gravity in the EuRoC sequence V2_01.

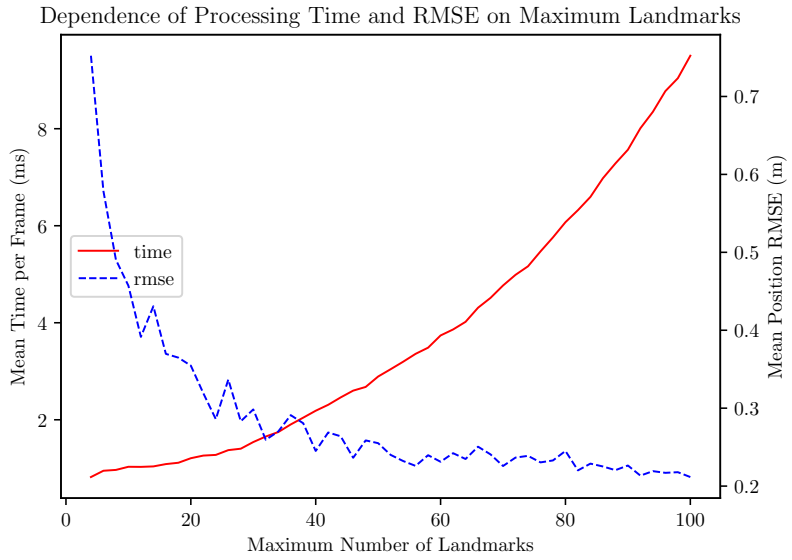


Figure 6.9: The relationship between the maximum number of landmarks used in EqVIO and the position RMSE and processing time on the EuRoC dataset.

Table 6.5: Average time taken to process each frame in ms on the UZH FPV [21] dataset for some of the systems in Table 6.1.

Algorithm	EqVIO	ROVIO	OpenVINS	VINS-Mono
Source	*	*	*	*
indoor_45_12	**2.79**	21.20	17.47	79.37
indoor_45_13	**2.50**	28.18	18.29	54.61
indoor_45_14	**3.19**	23.85	18.53	43.16
indoor_45_2	**3.30**	32.21	18.06	53.02
indoor_45_4	**2.65**	24.48	18.06	58.34
indoor_forward_10	**2.67**	41.86	16.27	66.43
indoor_forward_5	**3.10**	32.77	17.35	91.71
indoor_forward_6	**3.62**	43.71	17.00	45.64
indoor_forward_7	**2.75**	30.36	17.03	59.24
indoor_forward_9	**3.02**	24.34	16.20	74.21
Mean	**2.96**	30.30	17.43	64.02

6.8 Proofs

Proof of Proposition 6.3.2. Let $S \in \mathbf{SE}_{\mathbf{e}_3}(3)$, $\xi = (\xi_{\text{B}}, b_{\text{B}}, T, p_i) \in \mathcal{T}_n^{\text{VI}}(3)$, and $(\Omega, a) \in \mathbb{L}$ be arbitrary. Then compute

$$
\begin{aligned}
&f_{(\Omega,a)}(\alpha(S,(\xi_{\text{B}}, b_{\text{B}}, T, p_i))) \\
&= f_{(\Omega,a)}((S^{-1}P_{\text{B}}, R_S^\top v_{\text{B}}, b_{\text{B}}), T, S^{-1}(p_i)), \\
&= ((R_S^\top R_{\text{B}})(\Omega - b_{\text{B}}^\Omega)^\times, (R_S^\top v_{\text{B}}), (R_S^\top R_{\text{B}})(a - b_{\text{B}}^a) + g\mathbf{e}_3, \\
&\qquad 0, 0, 0), \\
&= (R_S^\top (R_{\text{B}}(\Omega - b_{\text{B}}^\Omega)^\times), R_S^\top v_{\text{B}}, R_S^\top (R_{\text{B}}(a - b_{\text{B}}^a) + g\mathbf{e}_3), \\
&\qquad 0, 0, 0,), \\
&= \mathrm{d}\alpha_S f_{\Omega,a}(\xi_{\text{B}}, b_{\text{B}}, T, p_i),
\end{aligned}
$$

where the second-last line follows from $R_S^\top \mathbf{e}_3 = \mathbf{e}_3$. This shows that, indeed, f is invariant with respect to the action α. To show the invariance of h, it is sufficient to show the invariance of the component functions h^k defined in (6.2). One has that

$$
\begin{aligned}
&h^k(\alpha(S,(\xi_{\text{B}}, b_{\text{B}}, T, p_i))) \\
&= h^k(((S^{-1}P_{\text{B}}, R_S^\top v_{\text{B}}, b_{\text{B}}^\Omega, b_{\text{B}}^a), T, S^{-1}(p_i))), \\
&= \pi_{\text{S}^2}\left((S^{-1}P_{\text{B}}T)^{-1}(S^{-1}(p_i))\right), \\
&= \pi_{\text{S}^2}\left((P_{\text{B}}T)^{-1}SS^{-1}(p_k)\right), \\
&= \pi_{\text{S}^2}\left((P_{\text{B}}T)^{-1}(p_k)\right), \\
&= h^k(\xi_{\text{B}}, b_{\text{B}}, T, p_i)).
\end{aligned}
$$

This completes the proof. □

Proof of Lemma 6.4.1. It is straightforward to see that φ^{V} and ρ^{V} are indeed right group actions. To see the equivariance of h^{V}, let $Q \in \mathbf{SOT}(3)$ and $q \in \mathbb{R}^3 \setminus \{0\}$ be arbitrary. Then,

$$\begin{aligned} h^{\text{V}}(\varphi^{\text{V}}(Q,q)) &= h^{\text{V}}(c_Q^{-1}R_Q^\top q), \\ &= \frac{c_Q^{-1}R_Q^\top q}{|c_Q^{-1}R_Q^\top q|}, \\ &= c_Q^{-1}R_Q^\top \frac{q}{c_Q^{-1}|R_Q^\top q|}, \\ &= R_Q^\top \frac{q}{|q|}, \\ &= \rho^{\text{V}}(Q, h^{\text{V}}(q)), \end{aligned}$$

as required. □

Proof of Lemma 6.4.2. Let $\xi = (\xi_{\text{B}}, b_{\text{B}}, T, p_i) \in \mathcal{T}_n^{\text{VI}}(3)$ and $X_1 = (A_1, \beta_1, B_1, Q_{1,i}), X_2 = (A_2, \beta_2, B_2, Q_{2,i}) \in \mathbf{G}$ be arbitrary. Then,

$$\begin{aligned} &\phi(X_2, \phi(X_1, \xi)) \\ &= \phi(X_2, (\varphi^{\text{B}}(A_1, \xi_{\text{B}}), b_{\text{B}} + \beta_1, P_{A_1}^{-1}TB_1, \\ &\qquad P_{\text{B}}TB_1Q_{1,i}^{-1}T^{-1}P_{\text{B}}^{-1}(p_i))), \\ &= (\varphi^{\text{B}}(A_2, \varphi^{\text{B}}(A_1, \xi_{\text{B}})), b_{\text{B}} + \beta_1 + \beta_2, P_{A_2}^{-1}(P_{A_1}^{-1}TB_1)B_2, \\ &\qquad P_{\text{B}}P_{A_1}P_{A_1}^{-1}TB_1B_2Q_{2,i}^{-1}(P_{A_1}^{-1}TB_1)^{-1}(P_{\text{B}}P_{A_1})^{-1} \\ &\qquad (P_{\text{B}}TB_1Q_{1,i}^{-1}T^{-1}P_{\text{B}}^{-1}(p_i))), \\ &= (\varphi^{\text{B}}(A_1A_2, \xi_{\text{B}}), b_{\text{B}} + (\beta_1 + \beta_2), (P_{A_1}P_{A_2})^{-1}T(B_1B_2), \\ &\qquad P_{\text{B}}TB_1B_2Q_{2,i}^{-1}B_1^{-1}T^{-1}P_{A_1}P_{A_1}^{-1}P_{\text{B}}^{-1} \\ &\qquad P_{\text{B}}TB_1Q_{1,i}^{-1}T^{-1}P_{\text{B}}^{-1}(p_i)), \\ &= (\varphi^{\text{B}}(A_1A_2, \xi_{\text{B}}), b_{\text{B}} + (\beta_1 + \beta_2), (P_{A_1}P_{A_2})^{-1}T(B_1B_2), \\ &\qquad P_{\text{B}}TB_1B_2Q_{2,i}^{-1}Q_{1,i}^{-1}T^{-1}P_{\text{B}}^{-1}(p_i)), \\ &= \phi(X_1X_2, \xi). \end{aligned}$$

This shows that the compatibility condition (3.3) is satisfied. For the identify condition (3.4), compute

$$\begin{aligned} &\phi(\mathrm{id}_{\mathbf{G}}, \xi) \\ &= \phi((I_5, 0, I_4, (I_4)), (\xi_{\text{B}}, b_{\text{B}}, T, p_i)), \\ &= (\varphi^{\text{B}}(I_5, \xi_{\text{B}}), b_{\text{B}} + 0, I_4TI_4, P_{\text{B}}TI_4I_4^{-1}T^{-1}P_{\text{B}}^{-1}(p_i)), \\ &= (\xi_{\text{B}}, b_{\text{B}}, p_i). \end{aligned}$$

Then, indeed, ϕ is a group action. Finally, to see that ϕ is transitive, let $\xi^1, \xi^2 \in \mathcal{T}_n^{\text{VI}}(3)$ be arbitrary, and let $X = (A, \beta, B, Q_i) \in \mathbf{G}$ such that

$$\begin{aligned} P_A &= (P_{\text{B}}^1)^{-1} P_{\text{B}}^2, \\ v_A &= (R_{\text{B}}^1)^\top (v_{\text{B}}^1 - v_{\text{B}}^2), \\ \beta &= b_{\text{B}}^2 - b_{\text{B}}^1, \\ B &= (P_{\text{B}}^1 T^1)^{-1} (P_{\text{B}}^2 T^2), \\ Q_i((P_{\text{B}}^2 T^2)^{-1} p_i^2) &= (P_{\text{B}}^1 T^1)^{-1} p_i^2. \end{aligned}$$

Then it is straightforward to see that $\phi(X, \xi^1) = \xi^2$. This completes the proof. □

Proof of Lemma 6.4.3. Let $\xi = (\xi_{\text{B}}, b_{\text{B}}, T, p_i) \in \mathcal{T}_n^{\text{VI}}(3)$, $S \in \mathbf{SE}_{\mathbf{e}_3}(3)$ and $X = (A, \beta, B, Q_i) \in \mathbf{G}$ be arbitrary. Then,

$$\begin{aligned} &\phi((A, \beta, B, Q_i), \alpha(S, (\xi_{\text{B}}, b_{\text{B}}, T, p_i)) \\ &= \phi((A, \beta, B, Q_i), (S^{-1} P_{\text{B}}, R_S^\top v_{\text{B}}, b_{\text{B}}, T, S^{-1}(p_i))), \\ &= (S^{-1} P_{\text{B}} P_A, R_S^\top v_{\text{B}} + R_S^\top R_{\text{B}} v_A, b_{\text{B}} + \beta, A^{-1} T B, \\ &\qquad S^{-1} P_{\text{B}} T B Q_i^{-1} (S^{-1} P_{\text{B}})^{-1} S^{-1}(p_i)), \\ &= (S^{-1}(P_{\text{B}} P_A), R_S^\top (v_{\text{B}} + R_{\text{B}} v_A), b_{\text{B}} + \beta, A^{-1} T B, \\ &\qquad S^{-1} P_{\text{B}} T B Q_i^{-1} P_{\text{B}}^{-1}(p_i)), \\ &= \alpha(S, (P_{\text{B}} P_A, v_{\text{B}} + R_{\text{B}} v_A, b_{\text{B}} + \beta, A^{-1} T B, \\ &\qquad P_{\text{B}} T B Q_i^{-1} P_{\text{B}}^{-1}(p_i))), \\ &= \alpha(S, \phi(X, \xi)), \end{aligned}$$

as required. □

Proof of Lemma 6.4.4. It is trivial to see that ρ is a group action. To show the equivariance of h, one examines the component measurement functions h^k. Let $X = (A, \beta, B, Q_i) \in \mathbf{G}$ and $\xi = (\xi_{\text{B}}, b_{\text{B}}, T, p_i) \in \mathcal{T}_n^{\text{VI}}(3)$. Then one has

$$\begin{aligned} &h^k(\phi((A, \beta, B, Q_i), (\xi_{\text{B}}, b_{\text{B}}, T, p_i))) \\ &= h^k(\varphi^{\text{B}}(A, \xi_{\text{B}}), b_{\text{B}} + \beta, P_A^{-1} T B, P_{\text{B}} T B Q_i^{-1} T^{-1} P_{\text{B}}^{-1}(p_i)), \\ &= \pi_{\text{S}^2}((P_{\text{B}} P_A)^{-1} (P_{\text{B}} T B Q_k^{-1} T^{-1} P_{\text{B}}^{-1}(p_k))), \\ &= \pi_{\text{S}^2}((P_{\text{B}} P_A P_A^{-1} T B)^{-1} (P_{\text{B}} T B Q_k^{-1} T^{-1} P_{\text{B}}^{-1}(p_k))), \\ &= \pi_{\text{S}^2}(Q_k^{-1} T^{-1} P_{\text{B}}^{-1}(p_k)), \\ &= \frac{c_{Q_k}^{-1} R_{Q_k}^\top T^{-1} P_{\text{B}}^{-1}(p_k)}{|c_{Q_k}^{-1} R_{Q_k}^\top T^{-1} P_{\text{B}}^{-1}(p_k)|}, \\ &= R_{Q_k}^\top \frac{T^{-1} P_{\text{B}}^{-1}(p_k)}{|T^{-1} P_{\text{B}}^{-1}(p_k)|}, \\ &= R_{Q_k}^\top h^k(\xi_{\text{B}}, b_{\text{B}}, T, p_i). \end{aligned}$$

It follows that

$$\begin{aligned}
&h(\phi((A,\beta,B,Q_i),(\xi_{\text{B}},b_{\text{B}},T,p_i)))\\
&=(h^1(\phi((A,\beta,B,Q_i),(\xi_{\text{B}},b_{\text{B}},T,p_i))),...,\\
&\qquad\qquad h^n(\phi((A,\beta,B,Q_i),(\xi_{\text{B}},b_{\text{B}},T,p_i)))),\\
&=(R_{Q_1}^{\top}h^1(\xi_{\text{B}},b_{\text{B}},T,p_1),...,R_{Q_n}^{\top}h^n(\xi_{\text{B}},b_{\text{B}},T,p_n)),\\
&=\rho((A,\beta,B,Q_i),h(\xi_{\text{B}},b_{\text{B}},T,p_i)),
\end{aligned}$$

as required. □

Chapter 7

Conclusion

To conclude this thesis, the key results are summarised and discussed below. Some directions of future work are outlined in Section 7.2.

7.1 Summary and Contributions

Recent advances in the theory of equivariant systems has led to the rise of powerful observer architectures that exploit system symmetries to outperform traditional filter designs in a range of practical problems. Current state-of-the-art equivariant observer architectures suffer from key limitations. General constructive observer designs are limited to systems with invariant dynamics, but allow for biases and delayed measurements. The linearisation-based IEKF is limited to systems on Lie groups with group-affine dynamics. The discovery of a novel symmetry group for SLAM has inspired research into the application of the IEKF to visual spatial awareness problems, but has not yielded results that compete with state-of-the-art VSA systems. This thesis extends the theory of equivariant systems by proposing new symmetries for visual spatial awareness problems and by developing a new filter design for any system for which a symmetry can be identified. The specific contributions of each technical chapter are as follows.

7.1.1 Symmetry of Visual Simultaneous Localisation and Mapping

This chapter proposed the VSLAM group and showed it to be compatible with the measurements of visual SLAM. The SLAM group [50; 78] is a symmetry of the VSLAM problem for which the dynamics are left-invariant, but the outputs are not equivariant. On the other hand, the dynamics of VSLAM are not invariant under the VSLAM group, but the outputs are equivariant. The general importance of equivariant outputs becomes clear in Chapter 5, where it is shown how this property can be exploited to obtain a higher-order approximation of the output function and improve filter performance.

A symmetric output error was defined by exploiting the equivariance of the measurement function. A novel observer for VSLAM was then defined by examining the dynamics of this error, and finding correction terms that drive this error to a constant value. As a result of this design procedure, the resulting observer could be shown to have almost globally asymptotically stabilisable error dynamics. This is a powerful re-

sult that contrasts with many of the existing observers for visual SLAM which offer only local convergence guarantees.

7.1.2 Equivariant Filter

This chapter introduced the Equivariant Filter (EqF). The EqF is a novel equivariant observer architecture that can be applied to any system on a homogeneous space, and exploits the symmetries of the system. The EqF design relies on the symmetry of a system to define a global equivariant error that can be examined using a local coordinate chart near a fixed origin configuration. This avoids the need for embedding of the system state or finding a set of changing local coordinate charts, which are required by other general observer designs for systems on manifolds. The EqF was shown to outperform an EKF for a state estimation problem on the sphere.

This chapter also showed how an equivariant output function can be exploited in an EqF design to reduce linearisation error. Specifically, it showed that the second-order error of the output approximation can be eliminated by exploiting knowledge of the measurement value and the compatibility of symmetry actions on the state and output spaces. This technique yielded a clear improvement in performance of the EqF for the example problem of bearing estimation on the sphere. The EqF differs qualitatively from previous linearisation-based equivariant observers (e.g. IEKF) in that the system state and its symmetry are strongly distinguished from one another. This separation of symmetry and system allows the EqF to be applied to a wider range of systems, and means that a user may choose the symmetry best suited to their system's characteristics when multiple options are available.

7.1.3 EqVIO: The Equivariant Filter for Visual Inertial Odometry

In this chapter, the VSLAM symmetry of Chapter 4 was extended with the $\mathbf{SE}_2(3)$ symmetry for IMU dynamics [50] to create a new visual inertial (VI-)SLAM Lie group. The VI-SLAM group was shown to exhibit compatibility with the VIO problem in a number of ways.

- The reference frame invariance of VIO is compatible with the VI-SLAM group.
- The IMU error dynamics are invariant under the VI-SLAM group.
- The visual measurements of landmarks are equivariant with respect to the VI-SLAM group.

No previous symmetries that have been applied to VIO have provided equivariance of the visual measurements of landmarks. Exploiting this output equivariance led to a significant reduction in linearisation error of landmark measurement approximation. This reduction in linearisation error of the visual measurements is key in the performance advantage of the EqF over the IEKF-based systems for VIO that depend on the classical SLAM symmetry.

EqVIO was the name given to the VIO system that combines a modern feature tracking front-end with an EqF back-end that uses the VI-SLAM symmetry. This system outperforms existing state-of-the-art VIO algorithms in terms of both speed and accuracy on both the EuRoC and UZH FPV datasets. Finally, the c++ implementation of EqVIO was made publicly available under a GNU GPLv3 license[1].

7.2 Future Work

A number of possible directions for future work are laid out in the following sections.

7.2.1 New Symmetries

The discovery of $\mathbf{SE}_2(3)$ for IMU dynamics and $\mathbf{SLAM}_n(3)$ for SLAM enabled the application of equivariant observers to previously intractable problems. One key contribution of this thesis was the development of new symmetries for VSLAM and VIO. There are many state estimation problems, however, for which there are still no compatible Lie groups known. Examples include problems involving biased IMU dynamics, multi-camera vision systems, soft robotics, and more. This means it is not currently possible to obtain the full benefit of applying equivariant observers to these systems. Discovering symmetries for such systems is thus an exciting direction for further research that would enable wider application of the techniques discussed in this thesis. While analysis of system dynamics was emphasised in prior works on equivariant systems theory, such as the study of invariant and group affine systems, the work presented in thesis instead emphasised the analysis of system measurements. This change of focus may lead to the discovery of new alternative symmetries for previously studied systems, with the potential to improve filter performance.

7.2.2 Problems in Visual Spatial Awareness

Chapter 6 addresses the VSA problem of VIO, where the visual sensor is a standard monocular camera. However, there are many other problems under the VSA umbrella, primarily distinguished by the available sensors. It is relatively straightforward to extend the EqVIO system to stereo and multiview camera systems by considering the additional measurements as non-equivariant outputs. Event cameras are bio-inspired sensors that, rather than capturing static frames at a fixed rate, capture changes in the light intensity at each pixel dynamically. Finding a way to integrate event camera measurements in a high-speed implementation like EqVIO could provide a powerful new event-based VIO system that does not lose feature tracking even in rapid motions. Another possible extension to the work in this thesis is to do away with image features and consider the full image instead as in the dense SLAM literature [70; 110]. The recent works by Clark *et al.* [115] and Zhang *et al.* [114] apply Lie group symmetries to point-cloud registration

[1] https://github.com/pvangoor/eqvio

for dense SLAM with RGBD cameras, and provide an interesting direction for potential future research. Finding a way to utilise dense measurements in an equivariant observer framework could lead to new breakthroughs in both dense SLAM research and equivariant systems theory.

7.2.3 Exploiting Symmetry Properties

In Chapter 5, it was shown that one can take advantage of equivariant output functions to obtain higher order output approximations. This adds to a growing body of literature showing the advantages of equivariant observers over generic nonlinear observer designs. An obvious question, then, is what other advantages the use of symmetry in observer design may yield. In a recent paper [4], Mahony *et al.* discussed the introduction of a curvature term in the Riccati equation of the EqF, and showed that its inclusion yielded some performance advantages in an example simulation. An interesting avenue for further research is therefore to identify the theoretical impact modifications such as the equivariant output approximation and the curvature term have on the performance of equivariant observers.

Appendix A

Equivariant Filter Implementation

The following algorithms are used to design the EqF for a given system.

Algorithm 1 EqF Design Preliminaries

1. Find a Lie group $\mathbf{G}$ and state action $\phi : \mathbf{G} \times \mathcal{M} \to \mathcal{M}$.
2. Check that the system is equivariant and compute the input symmetry $\psi : \mathbf{G} \times \mathbb{L} \to \mathbb{L}$.
3. Construct an equivariant lift $\Lambda : \mathcal{M} \times \mathbb{L} \to \mathfrak{g}$.
4. Check if there exists an action $\rho : \mathbf{G} \times \mathcal{N} \to \mathcal{N}$ such that the configuration output is equivariant.
5. Choose an origin $\mathring{\xi} \in \mathcal{M}$, choose a local coordinate chart ϑ about $\mathring{\xi}$, and fix a right-inverse $\mathrm{D}_X|_{\mathrm{id}}\phi_{\mathring{\xi}}(X)^\dagger$ of $\mathrm{D}_X|_{\mathrm{id}}\phi_{\mathring{\xi}}(X)$.
6. Initialise the observer state $\hat{X}(0) = \mathrm{id}$ and the Riccati term $\Sigma(0) = \Sigma_0 \in \mathbb{S}_+(m)$.

While the filter is presented in continuous time, in practice the equations must be implemented through numerical integration. Given input and output measurements $u \in \mathbb{L}$ and $y = h(\xi) \in \mathcal{N}$ at a given time, the steps in Algorithm 2 are executed. For the majority

Algorithm 2 EqF Design Implementation

1. Compute the origin velocity $\mathring{u} = \psi_{\hat{X}^{-1}}(u)$ and use this to obtain the state matrix $\mathring{A}_t$ (5.24) (cf. Lemma A.0.1).
2. Compute the standard output matrix C_t (5.32) or (preferably) the equivariant output matrix $C_t^\star$ (5.34).
3. Choose state and output gain matrices $M_t \in \mathbb{S}_+(m)$ and $N_t \in \mathbb{S}_+(n)$.
4. Update the observer state $\hat{X}(t)$ and Riccati state $\Sigma(t)$ by numerically approximating equations (5.35-5.37).

of applications, Euler integration or a higher order Runge-Kutta method is appropriate. Itazi and Sanyal [42] showed the effectiveness of Lie group variational integrators [86] for discretising an observer for invariant attitude dynamics, and a similar approach may

also be applied to discretising the EqF for certain systems.

In some cases it may be difficult to compute an explicit algebraic expression for the velocity action ψ. The following Lemma provides a way to implement the EqF without the need to derive ψ.

Lemma A.0.1. *The linearised state matrix $\mathring{A}_t$ defined in* (5.24) *can be written*

$$\begin{aligned}\mathring{A}_t = {} & \mathrm{D}_e|_{\mathring{\xi}}\vartheta(e)\mathrm{D}_\xi|_{\mathring{\xi}}\phi_{\hat{X}^{-1}}(\xi)\cdot\mathrm{D}_E|_{\mathrm{id}}\phi_{\mathring{\xi}}(E)\\ & \cdot\mathrm{D}_\xi|_{\phi_{\hat{X}}(\mathring{\xi})}\Lambda(\xi,u)\cdot\mathrm{D}_e|_{\mathring{\xi}}\phi_{\hat{X}}(e)\cdot\mathrm{D}_\varepsilon|_0\vartheta^{-1}(\varepsilon).\end{aligned}\tag{A.1}$$

Proof. Recall the equivariant lift condition (5.12). It follows that

$$\begin{aligned}&\mathrm{D}_E|_{\mathrm{id}}\phi_{\mathring{\xi}}(E)\cdot\mathrm{D}_e|_{\mathring{\xi}}\Lambda(e,\mathring{u})\\ &\quad=\mathrm{D}_E|_{\mathrm{id}}\phi_{\mathring{\xi}}(E)\cdot\mathrm{D}_e|_{\mathring{\xi}}\,\mathrm{Ad}_{\hat{X}}\,\Lambda(\phi_{\hat{X}}(e),u),\\ &\quad=\mathrm{D}_E|_{\mathrm{id}}\phi_{\mathring{\xi}}(E)\,\mathrm{Ad}_{\hat{X}}\cdot\mathrm{D}_e|_{\mathring{\xi}}\Lambda(\phi_{\hat{X}}(e),u),\\ &\quad=\mathrm{D}_E|_{\mathrm{id}}\phi_{\mathring{\xi}}(E)\,\mathrm{Ad}_{\hat{X}}\cdot\mathrm{D}_\xi|_{\phi_{\hat{X}}(\mathring{\xi})}\Lambda(\xi,u)\cdot\mathrm{D}_e|_{\mathring{\xi}}\phi_{\hat{X}}(e),\\ &\quad=\mathrm{D}_\xi|_{\hat{\xi}}\phi_{\hat{X}^{-1}}(\xi)\cdot\mathrm{D}_E|_{\mathrm{id}}\phi_{\hat{\xi}}(E)\cdot\mathrm{D}_\xi|_{\phi_{\hat{X}}(\mathring{\xi})}\Lambda(\xi,u)\\ &\qquad\qquad\cdot\mathrm{D}_e|_{\mathring{\xi}}\phi_{\hat{X}}(e).\end{aligned}$$

Then the expression (A.1) follows from the definition of $\mathring{A}_t$ in (5.24). Unlike (5.24), the expression (A.1) depends only the measured signal $u\in\mathbb{L}$ and not on the origin velocity $\mathring{u}=\psi_{\hat{X}^{-1}}(u)\in\mathbb{L}$. □

While the definitions of $\mathring{A}_t$ in (5.24) and (A.1) are equivalent, they present different challenges in practical implementation of the filter equations. Using the definition (5.24) requires an explicit algebraic expression for the velocity action ψ, which may be challenging to compute. On the other hand, (A.1) is independent of the algebraic expression of ψ, but requires the differentials $\mathrm{D}_\xi|_{\hat{\xi}}\phi_{\hat{X}^{-1}}(\xi)$ and $\mathrm{D}_\xi|_{\phi_{\hat{X}}(\mathring{\xi})}\Lambda(\xi,u)$ to be recomputed at different state elements $\hat{\xi}\in\mathcal{M}$ for each iteration. The expression (A.1) is particularly useful in situations where the equivariant velocity extension of a system is infinite [45].

EqF Specialisation to IEKF

The EqF specialises to an invariant extended Kalman filter (IEKF) [48] for a certain subclass of equivariant systems; specifically those with group affine dynamics on a Lie group, where the origin is chosen to be the identity, and where the local coordinates are chosen to be the exponential map in the EqF implementation.

Consider a system function $f : \mathbb{L} \to \mathfrak{X}(\mathcal{G})$ where $\mathcal{G}$ is the torsor of an m-dimensional matrix Lie group $\mathbf{G} \subset \mathbf{GL}(d)$. Right translation $R : \mathbf{G} \times \mathcal{G} \to \mathcal{G}$, defined by $R_X(B) = BX$ is a smooth, transitive right action of $\mathbf{G}$ on $\mathcal{G}$. Suppose the system is equivariant with respect to R and some velocity action ψ, that is,

$$\mathrm{D}R_X f_u(P) = f_{\psi_X(u)}(PX),$$

for all $P \in \mathcal{G}$, $X \in \mathbf{G}$ and $u \in \mathbb{L}$. Define the lift $\Lambda : \mathcal{G} \times \mathbb{L} \to \mathfrak{g}$ to be

$$\Lambda(P, u) = P^{-1} f_u(P),$$

where P^{-1} is understood as a matrix inverse.

Let $P \in \mathcal{G}$ denote the true state of the system. Choose the origin element, $\mathring{P} = I_d$, to be the identity matrix and let $\hat{X} \in \mathbf{G}$ denote the state of the observer, with dynamics

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\hat{X} &:= \hat{X}\Lambda(R_{\hat{X}}(\mathring{P}), u) + \Delta\hat{X}, \\
&= \hat{X}\Lambda(\hat{X}, u) + \Delta\hat{X}, \\
&= \hat{X}\hat{X}^{-1} f_u(\hat{X}) + \Delta\hat{X}, \\
&= f_u(\hat{X}) + \Delta\hat{X}.
\end{aligned}$$

In [48], Barrau *et al.* showed (see also [45, Remark 7.1]) that if the system f is "group affine", then the dynamics of the error $E := P\hat{X}^{-1}$ depend only on u and not the origin velocity $\mathring{u}$,

$$\dot{E} = f_u(E) - E f_u(I_d) - E\Delta.$$

Let $\vartheta : \mathcal{U}_{I_d} \to \mathbb{R}^m$ be the normal coordinates for $\mathbf{G}$ about I_d so that $\varepsilon = \vartheta(E) = \log(E)^\vee$.

The pre-observer ($\Delta \equiv 0$) dynamics of ε about $\varepsilon = 0$ are exactly

$$\begin{aligned}\dot{\varepsilon} &= \mathring{A}_t \varepsilon, \\ \mathring{A}_t \varepsilon &= \left(\mathrm{D}_E|_{I_n} \Lambda(E, u)[\varepsilon^\wedge]\right)^\vee.\end{aligned}$$

Let the output space $\mathcal{N}$ be the Euclidean space $\mathbb{R}^n$, and let the configuration output $h : \mathcal{G} \to \mathbb{R}^n$ be any map, that is, h is not necessarily equivariant. Then the EqF (5.35-5.37) with standard output matrix specialises to the invariant extended Kalman filter (IEKF) proposed in [48].

In [48], the pre-observer dynamics of ε are shown to be exactly linear as a consequence of the novel 'group affine' property. In this case the IEKF is locally asymptotically stable with a constant convergence radius [48].

Bibliography

[1] P. van Goor, R. Mahony, T. Hamel, and J. Trumpf, "Constructive observer design for Visual Simultaneous Localisation and Mapping," *Automatica*, vol. 132, p. 109803, 2021. (Cited on pages xiv, 5, 51, 66, 70, 73, 77, 80, 82, and 84.)

[2] P. van Goor, T. Hamel, and R. Mahony, "Equivariant Filter (EqF)," *IEEE Transactions on Automatic Control*, pp. 1–13, 2022. (Cited on pages 5, 67, 69, 70, 74, 75, 78, 79, 82, 83, and 90.)

[3] P. van Goor and R. Mahony, "EqVIO: Equivariant Filter for Visual Inertial Odometry," *Submitted to IEEE Transactions on Robotics*, 2022. (Cited on page 5.)

[4] R. Mahony, P. van Goor, and T. Hamel, "Observer Design for Nonlinear Systems with Equivariance," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 5, no. 1, pp. 221–252, May 2022. (Cited on pages xiii, 7, 11, 48, 50, 53, 55, 59, and 100.)

[5] P. van 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)*, Dec. 2019, pp. 2543–2549. (Cited on pages 5, 27, 30, and 44.)

[6] ——, "An Observer Design for Visual Simultaneous Localisation and Mapping with Output Equivariance," *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 9560–9565, Jan. 2020. (Cited on page 5.)

[7] P. van Goor, T. Hamel, and R. Mahony, "Equivariant Filter (EqF): A General Filter Design for Systems on Homogeneous Spaces," in *2020 59th IEEE Conference on Decision and Control (CDC)*, Dec. 2020, pp. 5401–5408. (Cited on page 5.)

[8] P. van Goor and R. Mahony, "An Equivariant Filter for Visual Inertial Odometry," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, May 2021, pp. 14 432–14 438. (Cited on pages 5, 66, and 70.)

[9] Y. Ng, P. van Goor, R. Mahony, and T. Hamel, "Attitude Observation for Second Order Attitude Kinematics," in *2019 IEEE 58th Conference on Decision and Control (CDC)*, Dec. 2019, pp. 2536–2542. (Cited on page 9.)

[10] Z. Wang, Y. Ng, P. van Goor, and R. Mahony, "Event Camera Calibration of Per-pixel Biased Contrast Threshold," in *Australasian Conference on Robotics and Automation*, 2019. (Not cited.)

[11] Y. Ng, P. van Goor, and R. Mahony, "Pose observation for second order pose kinematics," *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 2317–2323, 2020. (Cited on pages 1 and 9.)

[12] Y. Ng, P. van Goor, T. Hamel, and R. Mahony, "Equivariant Systems Theory and Observer Design for Second Order Kinematic Systems on Matrix Lie Groups," in *2020 59th IEEE Conference on Decision and Control (CDC)*, Dec. 2020, pp. 4194–4199. (Cited on page 11.)

[13] R. Mahony, P. van Goor, M. Henein, R. Pike, J. Zhang, and Y. Ng, "Equivariant Visual Odometry in the Wild," in *2020 59th IEEE Conference on Decision and Control (CDC)*, Dec. 2020, pp. 1314–1319. (Cited on page 66.)

[14] Y. Ge, P. van Goor, and R. Mahony, "An Equivariant Filter for Stereo-Camera SLAM," in *2020 Australasian Conference on Robotics and Automation*. ARAA, 2020. (Not cited.)

[15] P. van Goor and R. Mahony, "Autonomous Error and Constructive Observer Design for Group Affine Systems," in *2021 60th IEEE Conference on Decision and Control (CDC)*, Dec. 2021, pp. 4730–4737. (Not cited.)

[16] H. F. Durrant-Whyte and T. Bailey, "Simultaneous Localisation and Mapping (SLAM): Part I," *IEEE Robotics and Automation Magazine (RAM)*, vol. 13, no. 2, pp. 99–110, 2006. (Cited on pages xiii, 11, and 12.)

[17] F. Dellaert and M. Kaess, "Square Root SAM: Simultaneous localization and mapping via square root information smoothing," *The International Journal of Robotics Research*, vol. 25, no. 12, pp. 1181–1203, 2006. (Cited on pages xiii, 7, 13, and 14.)

[18] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardós, "ORB-SLAM: A Versatile and Accurate Monocular SLAM System," *IEEE Transactions on Robotics*, vol. 31, no. 5, pp. 1147–1163, 2015. (Cited on pages xiii, 7, 14, and 15.)

[19] M. Bloesch, M. Burri, S. Omari, M. Hutter, and R. Siegwart, "Iterated extended Kalman filter based visual-inertial odometry using direct photometric feedback," *The International Journal of Robotics Research*, vol. 36, no. 10, pp. 1053–1072, Sep. 2017. (Cited on pages xiii, 16, 17, 67, 68, 70, 77, and 86.)

[20] M. Burri, J. Nikolic, P. Gohl, T. Schneider, J. Rehder, S. Omari, M. W. Achtelik, and R. Siegwart, "The EuRoC micro aerial vehicle datasets," *The International Journal of Robotics Research*, vol. 35, no. 10, pp. 1157–1163, Sep. 2016. (Cited on pages xvii, 68, 86, 87, 88, and 90.)

[21] 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 *2019 International Conference on Robotics and Automation (ICRA)*, May 2019, pp. 6713–6719. (Cited on pages xvii, 68, 86, 90, and 92.)

[22] H. Rehbinder and X. He, "Nonlinear pitch and roll estimation for walking robots," in *Proceedings 2000 ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation. Symposia Proceedings (Cat. No. 00CH37065)*, vol. 3. IEEE, 2000, pp. 2617–2622. (Cited on pages 1 and 8.)

[23] H. Rehbinder and B. K. Ghosh, "Pose estimation using line-based dynamic vision and inertial sensors," *IEEE Transactions on Automatic control*, vol. 48, no. 2, pp. 186–199, 2003. (Cited on pages 1 and 9.)

[24] R. Mahony, T. Hamel, and J.-M. Pflimlin, "Complementary filter design on the special orthogonal group SO (3)," in *Proceedings of the 44th IEEE Conference on Decision and Control*. IEEE, 2005, pp. 1477–1484. (Cited on pages 1, 8, and 9.)

[25] T. Hamel and R. Mahony, "Attitude estimation on SO [3] based on direct inertial measurements," in *Proceedings 2006 IEEE International Conference on Robotics and Automation, 2006. ICRA 2006.* IEEE, 2006, pp. 2170–2175. (Cited on pages 1 and 8.)

[26] S. Bonnabel, P. Martin, and P. Rouchon, "A non-linear symmetry-preserving observer for velocity-aided inertial navigation," in *American Control Conference, Proceedings of The*, Jun. 2006, pp. 2910–2914. (Cited on pages 1, 7, and 9.)

[27] T. Cheviron, T. Hamel, R. Mahony, and G. Baldwin, "Robust Nonlinear Fusion of Inertial and Visual Data for position, velocity and attitude estimation of UAV," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2007, pp. 2010–2016. (Cited on pages 1, 9, and 10.)

[28] R. Mahony, T. Hamel, and J. Pflimlin, "Nonlinear Complementary Filters on the Special Orthogonal Group," *IEEE Transactions on Automatic Control*, vol. 53, no. 5, pp. 1203–1218, Jun. 2008. (Cited on pages 1, 7, 8, 9, 10, and 51.)

[29] S. O. Madgwick, A. J. Harrison, and R. Vaidyanathan, "Estimation of IMU and MARG orientation using a gradient descent algorithm," in *2011 IEEE International Conference on Rehabilitation Robotics*. IEEE, 2011, pp. 1–7. (Cited on pages 1, 7, and 8.)

[30] T. Hamel, R. Mahony, J. Trumpf, P. Morin, and M.-D. Hua, "Homography estimation on the Special Linear group based on direct point correspondence," in *2011 50th IEEE Conference on Decision and Control and European Control Conference*, Dec. 2011, pp. 7902–7908. (Cited on pages 1, 9, and 10.)

[31] H. F. Grip, T. I. Fossen, T. A. Johansen, and A. Saberi, "Attitude Estimation Using Biased Gyro and Vector Measurements With Time-Varying Reference Vectors," *IEEE Transactions on Automatic Control*, vol. 57, no. 5, pp. 1332–1338, May 2012. (Cited on pages 1 and 8.)

[32] S. Brás, M. Izadi, C. Silvestre, A. Sanyal, and P. Oliveira, "Nonlinear observer for 3D rigid body motion," in *52nd IEEE Conference on Decision and Control*, Dec. 2013, pp. 2588–2593. (Cited on pages 1, 7, and 9.)

[33] V. Grabe, H. H. Bülthoff, D. Scaramuzza, and P. R. Giordano, "Nonlinear ego-motion estimation from optical flow for online control of a quadrotor UAV," *The International Journal of Robotics Research*, vol. 34, no. 8, pp. 1114–1135, 2015. (Cited on pages 1, 7, and 9.)

[34] M. Hua, T. Hamel, R. Mahony, and J. Trumpf, "Gradient-like observer design on the Special Euclidean group SE(3) with system outputs on the real projective space," in *2015 54th IEEE Conference on Decision and Control (CDC)*, Dec. 2015, pp. 2139–2145. (Cited on pages 1 and 9.)

[35] S. Berkane, A. Abdessameud, and A. Tayebi, "Hybrid attitude and gyro-bias observer design on $ SO (3) $," *IEEE Transactions on Automatic Control*, vol. 62, no. 11, pp. 6044–6050, 2017. (Cited on pages 1 and 9.)

[36] M.-D. Hua, J. Trumpf, T. Hamel, R. Mahony, and P. Morin, "Feature-based recursive observer design for homography estimation and its application to image stabilization," *Asian Journal of Control*, vol. 21, no. 4, pp. 1443–1458, 2019. (Cited on pages 1 and 9.)

[37] C. Lageman, J. Trumpf, and R. Mahony, "Gradient-Like Observers for Invariant Dynamics on a Lie Group," *IEEE Transactions on Automatic Control*, vol. 55, no. 2, pp. 367–377, Feb. 2010. (Cited on pages 1, 7, 9, and 10.)

[38] G. Bourmaud, R. Mégret, A. Giremus, and Y. Berthoumieu, "From Intrinsic Optimization to Iterated Extended Kalman Filtering on Lie Groups," *Journal of Mathematical Imaging and Vision*, vol. 55, no. 3, pp. 284–303, Jul. 2016. (Cited on pages 1 and 10.)

[39] A. Barrau and S. Bonnabel, "Invariant Kalman Filtering," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 1, no. 1, pp. 237–257, 2018. (Cited on pages 1, 7, 10, 67, 69, and 75.)

[40] A.-J. Baerveldt and R. Klang, "A low-cost and low-weight attitude estimation system for an autonomous helicopter," in *Proceedings of IEEE International Conference on Intelligent Engineering Systems*. IEEE, 1997, pp. 391–395. (Cited on pages 1 and 8.)

[41] F. L. Markley, "Attitude Error Representations for Kalman Filtering," *Journal of Guidance, Control, and Dynamics*, vol. 26, no. 2, pp. 311–317, 2003. (Cited on pages 1, 8, 10, and 69.)

[42] M. Izadi and A. K. Sanyal, "Rigid body attitude estimation based on the Lagrange–d'Alembert principle," *Automatica*, vol. 50, no. 10, pp. 2570–2577, Oct. 2014. (Cited on pages 1, 7, 9, 51, and 101.)

[43] J. F. Vasconcelos, R. Cunha, C. Silvestre, and P. Oliveira, "A nonlinear position and attitude observer on SE(3) using landmark measurements," *Systems & Control Letters*, vol. 59, no. 3-4, pp. 155–166, 2010. (Cited on pages 1 and 9.)

[44] A. W. Long, K. C. Wolfe, M. J. Mashner, and G. S. Chirikjian, "The banana distribution is Gaussian: A localization study with exponential coordinates," *Robotics: Science and Systems VIII*, vol. 265, 2013. (Cited on pages 1 and 9.)

[45] R. Mahony, T. Hamel, and J. Trumpf, "Equivariant Systems Theory and Observer Design," *arXiv:2006.08276 [cs, eess]*, Aug. 2020. (Cited on pages 2, 21, 51, 52, 53, 67, 82, 102, and 103.)

[46] R. Mahony, J. Trumpf, and T. Hamel, "Observers for Kinematic Systems with Symmetry," in *IFAC Proceedings Volumes*, ser. 9th IFAC Symposium on Nonlinear Control Systems, vol. 46, Jan. 2013, pp. 617–633. (Cited on pages 2, 7, 10, 34, 47, and 51.)

[47] A. Khosravian, J. Trumpf, R. Mahony, and C. Lageman, "Observers for invariant systems on Lie groups with biased input measurements and homogeneous outputs," *Automatica*, vol. 55, pp. 19–26, May 2015. (Cited on pages 2, 7, 10, 17, and 51.)

[48] 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, Apr. 2017. (Cited on pages 2, 10, 47, 48, 51, 55, 65, 103, and 104.)

[49] J. Delmerico and D. Scaramuzza, "A Benchmark Comparison of Monocular Visual-Inertial Odometry Algorithms for Flying Robots," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2018. (Cited on pages 3, 16, 27, 69, 86, and 87.)

[50] A. Barrau and S. Bonnabel, "An EKF-SLAM algorithm with consistency properties," *arXiv:1510.06263 [cs]*, Sep. 2016. (Cited on pages 3, 7, 16, 17, 18, 31, 69, 70, 97, and 98.)

[51] K. Wu, T. Zhang, D. Su, S. Huang, and G. Dissanayake, "An invariant-EKF VINS algorithm for improving consistency," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2017, pp. 1578–1585. (Cited on pages 3, 7, 17, and 70.)

[52] M. Brossard, S. Bonnabel, and A. Barrau, "Unscented Kalman filter on Lie groups for visual inertial odometry," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 649–655. (Cited on pages 3, 7, 17, 28, 68, and 70.)

[53] ——, "Invariant Kalman Filtering for Visual Inertial SLAM," in *2018 21st International Conference on Information Fusion (FUSION)*. Cambridge: IEEE, Jul. 2018, pp. 2021–2028. (Cited on pages 3, 17, and 70.)

[54] S. Heo, J. H. Jung, and C. G. Park, "Consistent EKF-Based Visual-Inertial Navigation Using Points and Lines," *IEEE Sensors Journal*, vol. 18, no. 18, pp. 7638–7649, Sep. 2018. (Cited on pages 3, 17, and 70.)

[55] 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, May 2018. (Cited on pages 3, 7, 17, and 70.)

[56] R. W. Brockett, "System theory on group manifolds and coset spaces," *SIAM Journal on control*, vol. 10, no. 2, pp. 265–284, 1972. (Cited on pages 7, 8, and 47.)

[57] N. Aghannan and P. Rouchon, "An intrinsic observer for a class of Lagrangian systems," *IEEE Transactions on Automatic Control*, vol. 48, no. 6, pp. 936–945, Jun. 2003. (Cited on pages 7 and 8.)

[58] J. Trumpf, R. Mahony, T. Hamel, and C. Lageman, "Analysis of Non-Linear Attitude Observers for Time-Varying Reference Measurements," *IEEE Transactions on Automatic Control*, vol. 57, no. 11, pp. 2789–2800, Nov. 2012. (Cited on pages 7 and 8.)

[59] S. Bonnabel, P. Martin, and P. Rouchon, "Symmetry-Preserving Observers," *IEEE Transactions on Automatic Control*, vol. 53, no. 11, pp. 2514–2526, Dec. 2008. (Cited on pages 7, 10, and 51.)

[60] G. Bourmaud, R. Mégret, M. Arnaudon, and A. Giremus, "Continuous-Discrete Extended Kalman Filter on Matrix Lie Groups Using Concentrated Gaussian Distributions," *Journal of Mathematical Imaging and Vision*, vol. 51, no. 1, pp. 209–228, Jan. 2015. (Cited on pages 7 and 10.)

[61] C. Zhang, A. Taghvaei, and P. G. Mehta, "Feedback Particle Filter on Riemannian Manifolds and Matrix Lie Groups," *IEEE Transactions on Automatic Control*, vol. 63, no. 8, pp. 2465–2480, Aug. 2018. (Cited on pages 7 and 10.)

[62] M. G. Dissanayake, P. Newman, S. Clark, H. F. Durrant-Whyte, and M. Csorba, "A solution to the simultaneous localization and map building (SLAM) problem," *IEEE Transactions on robotics and automation*, vol. 17, no. 3, pp. 229–241, 2001. (Cited on pages 7 and 11.)

[63] J. Neira and J. D. Tard/'os, "Data Association in Stochastic Mapping Using the Joint Compatibility Test," *IEEE Transactions on Robotics and Automation (TRA)*, vol. 17, no. 6, pp. 890–897, 2001. (Cited on pages 7 and 12.)

[64] M. Montemerlo and S. Thrun, "Simultaneous localization and mapping with unknown data association using FastSLAM," in *2003 IEEE International Conference on Robotics and Automation (Cat. No. 03CH37422)*, vol. 2. IEEE, 2003, pp. 1985–1991. (Cited on pages 7 and 11.)

[65] T. Bailey, J. Nieto, J. Guivant, M. Stevens, and E. Nebot, "Consistency of the EKF-SLAM algorithm," in *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2006, pp. 3562–3568. (Cited on pages 7, 13, 17, 70, and 77.)

[66] L. M. Paz, J. D. TardÓs, and J. Neira, "Divide and Conquer: EKF SLAM in $O(n)$," *IEEE Transactions on Robotics*, vol. 24, no. 5, pp. 1107–1120, Oct. 2008. (Cited on pages 7 and 13.)

[67] G. Klein and D. Murray, "Parallel tracking and mapping for small AR workspaces," in *Proceedings of the 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*. IEEE Computer Society, 2007, pp. 1–10. (Cited on pages 7, 14, 15, and 68.)

[68] D. Sibley, C. Mei, I. D. Reid, and P. Newman, "Adaptive relative bundle adjustment." in *Robotics: Science and Systems*, vol. 32, 2009, p. 33. (Cited on pages 7 and 14.)

[69] M. Kaess, H. Johannsson, R. Roberts, V. Ila, J. J. Leonard, and F. Dellaert, "iSAM2: Incremental smoothing and mapping using the Bayes tree," *The International Journal of Robotics Research*, vol. 31, no. 2, pp. 216–235, 2012. (Cited on pages 7 and 14.)

[70] C. D. Engel J., Schöps T., "LSD-SLAM: Large-Scale Direct Monocular SLAM," in *Computer Vision – ECCV 2014*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Springer, 2014, vol. 8690. (Cited on pages 7, 14, and 99.)

[71] R. Mur-Artal and J. D. Tardós, "Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras," *IEEE Transactions on Robotics*, vol. 33, no. 5, pp. 1255–1262, 2017. (Cited on pages 7 and 15.)

[72] A. I. Mourikis and S. I. Roumeliotis, "A multi-state constraint Kalman filter for vision-aided inertial navigation," in *Procedings of the International Conference on Robotics and Automation (ICRA)*, 2007. (Cited on pages 7, 15, 16, 17, 67, 68, 69, 70, 74, and 86.)

[73] M. Bloesch, S. Omari, M. Hutter, and R. Siegwart, "Robust visual inertial odometry using a direct EKF-based approach," in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2015, pp. 298–304. (Cited on pages 7 and 16.)

[74] S. Leutenegger, S. Lynen, M. Bosse, R. Siegwart, and P. Furgale, "Keyframe-based visual-inertial SLAM using nonlinear optimization," *Int. J. Robot. Research, vol. 34, no. 3, pp. 314–334, 2015.*, vol. 34, no. 3, pp. 314—334, 2015. (Cited on pages 7, 16, 67, 69, and 86.)

[75] C. Forster, Z. Zhang, M. Gassner, M. Werlberger, and D. Scaramuzza, "SVO: Semidirect Visual Odometry for Monocular and Multicamera Systems," *IEEE*

Transactions on Robotics, vol. 33, no. 2, pp. 249–265, Apr. 2017. (Cited on pages 7, 16, and 69.)

[76] T. Qin, P. Li, and S. Shen, "Vins-mono: A robust and versatile monocular visual-inertial state estimator," *IEEE Transactions on Robotics*, vol. 34, no. 4, pp. 1004–1020, 2018. (Cited on pages 7, 16, 67, 69, and 86.)

[77] P. Geneva, K. Eckenhoff, W. Lee, Y. Yang, and G. Huang, "OpenVINS: A Research Platform for Visual-Inertial Estimation," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, May 2020, pp. 4666–4672. (Cited on pages 7, 16, 67, 69, 77, and 86.)

[78] R. Mahony and T. Hamel, "A geometric nonlinear observer for simultaneous localisation and mapping," in *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, Dec. 2017, pp. 2408–2415. (Cited on pages 7, 17, 18, 27, 29, 31, 44, and 97.)

[79] H. A. Hashim, "Guaranteed Performance Nonlinear Observer for Simultaneous Localization and Mapping," *IEEE Control Systems Letters*, vol. 5, no. 1, pp. 91–96, Jan. 2021. (Cited on pages 7 and 17.)

[80] Y. Song, Z. Zhang, J. Wu, Y. Wang, L. Zhao, and S. Huang, "A Right Invariant Extended Kalman Filter for Object Based SLAM," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 1316–1323, Apr. 2022. (Cited on pages 7 and 17.)

[81] V. Jurdjevic and H. J. Sussmann, "Control systems on Lie groups," *Journal of Differential equations*, vol. 12, no. 2, pp. 313–329, 1972. (Cited on pages 8 and 47.)

[82] R. Brockett, "Lie theory and control systems defined on spheres," *SIAM Journal on Applied Mathematics*, vol. 25, no. 2, pp. 213–225, 1973. (Cited on pages 8 and 47.)

[83] S. Salcudean, "A globally convergent angular velocity observer for rigid body motion," *IEEE transactions on Automatic Control*, vol. 36, no. 12, pp. 1493–1497, 1991. (Cited on page 8.)

[84] J. Thienel and R. M. Sanner, "A coupled nonlinear spacecraft attitude controller and observer with an unknown constant gyro bias and gyro noise," *IEEE Transactions on Automatic Control*, vol. 48, no. 11, pp. 2011–2015, Nov. 2003. (Cited on page 8.)

[85] S. Bonnabel, "Left-invariant extended Kalman filter and attitude estimation," in *Procedings of the IEEE Conference on Decision and Control (CDC)*, New Orleans, LA, USA, 2007, p. 6 pages. (Cited on pages 8 and 10.)

[86] M. Leok, "Foundations of computational geometric mechanics," Ph.D. dissertation, California Institute of Technology, United States – California, 2004. (Cited on pages 9 and 101.)

[87] G. Bourmaud, R. Mégret, A. Giremus, and Y. Berthoumieu, "Discrete Extended Kalman Filter on Lie groups," in *21st European Signal Processing Conference (EUSIPCO 2013)*, Sep. 2013, pp. 1–5. (Cited on page 10.)

[88] G. Chirikjian and M. Kobilarov, "Gaussian approximation of non-linear measurement models on Lie groups," in *53rd IEEE Conference on Decision and Control*, Dec. 2014, pp. 6401–6406. (Cited on page 10.)

[89] A. Saccon, J. Trumpf, R. Mahony, and A. P. Aguiar, "Second-Order-Optimal Minimum-Energy Filters on Lie Groups," *IEEE Transactions on Automatic Control*, vol. 61, no. 10, pp. 2906–2919, Oct. 2016. (Cited on pages 10 and 59.)

[90] A. Barrau and S. Bonnabel, "Linear observed systems on groups," *Systems & Control Letters*, vol. 129, pp. 36–42, Jul. 2019. (Cited on page 10.)

[91] M.-A. Lavoie, J. Arsenault, and J. R. Forbes, "An Invariant Extended H∞ Filter," in *2019 IEEE 58th Conference on Decision and Control (CDC)*, Dec. 2019, pp. 7905–7910. (Cited on page 11.)

[92] A. A. Joshi, D. H. S. Maithripala, and R. N. Banavar, "A bundle framework for observer design on smooth manifolds with symmetry," *Journal of Geometric Mechanics*, vol. 13, no. 2, p. 247, 2021. (Cited on page 11.)

[93] R. Mahony and J. Trumpf, "Equivariant Filter Design for Kinematic Systems on Lie Groups," *IFAC-PapersOnLine*, vol. 54, no. 9, pp. 253–260, Jan. 2021. (Cited on pages 11 and 72.)

[94] T. Bailey and H. F. Durrant-Whyte, "Simultaneous Localisation and Mapping (SLAM): Part II," *IEEE Robotics and Automation Magazine (RAM)*, vol. 13, no. 3, pp. 108–117, 2006. (Cited on page 11.)

[95] S. Thrun, W. Burgard, and D. Fox, "A probabilistic approach to concurrent mapping and localization for mobile robots," *Autonomous Robots*, vol. 5, no. 3-4, pp. 253–271, 1998. (Cited on page 11.)

[96] M. Montemerlo, S. Thrun, D. Koller, B. Wegbreit *et al.*, "FastSLAM: A factored solution to the simultaneous localization and mapping problem," *Aaai/iaai*, vol. 593598, 2002. (Cited on page 11.)

[97] ——, "FastSLAM 2.0: An improved particle filtering algorithm for simultaneous localization and mapping that provably converges," in *IJCAI*, 2003, pp. 1151–1156. (Cited on page 11.)

[98] S. Thrun, Y. Liu, D. Koller, A. Y. Ng, Z. Ghahramani, and H. Durrant-Whyte, "Simultaneous localization and mapping with sparse extended information filters," *The international journal of robotics research*, vol. 23, no. 7-8, pp. 693–716, 2004. (Cited on page 11.)

[99] R. M. Eustice, H. Singh, and J. J. Leonard, "Exactly Sparse Delayed-State Filters for View-Based SLAM," *IEEE Transactions on Robotics*, vol. 22, no. 6, pp. 1100–1114, Dec. 2006. (Cited on page 12.)

[100] J. E. Guivant and E. M. Nebot, "Optimization of the simultaneous localization and map-building algorithm for real-time implementation," *IEEE transactions on robotics and automation*, vol. 17, no. 3, pp. 242–257, 2001. (Cited on page 12.)

[101] J. D. Tardós, J. Neira, P. M. Newman, and J. J. Leonard, "Robust mapping and localization in indoor environments using sonar data," *The International Journal of Robotics Research*, vol. 21, no. 4, pp. 311–330, 2002. (Cited on pages 12 and 13.)

[102] J. A. Castellanos, J. Neira, and J. D. Tardós, "Limits to the consistency of EKF-based SLAM," *IFAC Proceedings Volumes*, vol. 37, no. 8, pp. 716–721, 2004. (Cited on pages 13, 17, and 70.)

[103] C. Estrada, J. Neira, and J. D. Tardos, "Hierarchical SLAM: Real-time accurate mapping of large environments," *IEEE Transactions on Robotics*, vol. 21, no. 4, pp. 588–596, Aug. 2005. (Cited on page 13.)

[104] A. J. Davison, I. D. Reid, N. D. Molton, and O. Stasse, "MonoSLAM: Real-time single camera SLAM," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, no. 6, pp. 1052–1067, 2007. (Cited on pages 13, 15, and 68.)

[105] F. Lu and E. Milios, "Globally consistent range scan alignment for environment mapping," *Autonomous robots*, vol. 4, no. 4, pp. 333–349, 1997. (Cited on page 13.)

[106] S. Thrun and M. Montemerlo, "The graph SLAM algorithm with applications to large-scale mapping of urban structures," *The International Journal of Robotics Research*, vol. 25, no. 5-6, pp. 403–429, 2006. (Cited on pages 13 and 14.)

[107] M. Kaess, A. Ranganathan, and F. Dellaert, "iSAM: Incremental smoothing and mapping," *IEEE Transactions on Robotics*, vol. 24, no. 6, pp. 1365–1378, 2008. (Cited on page 14.)

[108] R. Kümmerle, G. Grisetti, H. Strasdat, K. Konolige, and W. Burgard, "G 2 o: A general framework for graph optimization," in *2011 IEEE International Conference on Robotics and Automation*. IEEE, 2011, pp. 3607–3613. (Cited on pages 14 and 15.)

[109] H. Strasdat, J. M. M. Montiel, and A. J. Davison, "Visual SLAM: Why filter?" *Computer Vision and Image Understanding (CVIU)*, vol. 30, no. 2, pp. 65–77, 2012. (Cited on pages 14, 27, and 41.)

[110] J. Engel, V. Koltun, and D. Cremers, "Direct sparse odometry," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 3, pp. 611–625, 2017. (Cited on pages 15 and 99.)

[111] V. Usenko, N. Demmel, D. Schubert, J. Stückler, and D. Cremers, "Visual-Inertial Mapping With Non-Linear Factor Recovery," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 422–429, Apr. 2020. (Cited on page 15.)

[112] A. Rosinol, M. Abate, Y. Chang, and L. Carlone, "Kimera: An Open-Source Library for Real-Time Metric-Semantic Localization and Mapping," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, May 2020, pp. 1689–1696. (Cited on page 15.)

[113] C. Campos, R. Elvira, J. J. G. Rodríguez, J. M. M. Montiel, and J. D. Tardós, "ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual–Inertial, and Multimap SLAM," *IEEE Transactions on Robotics*, vol. 37, no. 6, pp. 1874–1890, Dec. 2021. (Cited on page 15.)

[114] R. Zhang, T.-Y. Lin, C. E. Lin, S. A. Parkison, W. Clark, J. W. Grizzle, R. M. Eustice, and M. Ghaffari, "A New Framework for Registration of Semantic Point Clouds from Stereo and RGB-D Cameras," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, May 2021, pp. 12 214–12 221. (Cited on pages 15 and 99.)

[115] W. Clark, M. Ghaffari, and A. Bloch, "Nonparametric continuous sensor registration," *The Journal of Machine Learning Research*, vol. 22, no. 1, pp. 271:12 412–271:12 461, Jan. 2021. (Cited on pages 15 and 99.)

[116] D. Nistér, O. Naroditsky, and J. Bergen, "Visual odometry," in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, vol. 1. Ieee, 2004, pp. I–I. (Cited on pages 15 and 68.)

[117] ——, "Visual odometry for ground vehicle applications," *Journal of Field Robotics*, vol. 23, no. 1, pp. 3–20, 2006. (Cited on pages 15 and 68.)

[118] A. Howard, "Real-time stereo visual odometry for autonomous ground vehicles," in *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2008, pp. 3946–3952. (Cited on page 16.)

[119] K. Konolige, M. Agrawal, and J. Sola, "Large-scale visual odometry for rough terrain," in *Robotics Research*. Springer, 2010, pp. 201–212. (Cited on pages 16 and 68.)

[120] D. E. Zlotnik and J. R. Forbes, "Gradient-based observer for simultaneous localization and mapping," *IEEE Transactions on Automatic Control*, vol. 63, no. 12, pp. 4338–4344, 2018. (Cited on page 17.)

[121] E. Bjorne, T. A. Johansen, and E. F. Brekke, "Redesign and analysis of globally asymptotically stable bearing only SLAM," in *2017 20th International Conference on Information Fusion (Fusion)*, Jul. 2017, pp. 1–8. (Cited on page 17.)

[122] T. Hamel and C. Samson, "Riccati Observers for the Nonstationary PnP Problem," *IEEE Transactions on Automatic Control*, vol. 63, no. 3, pp. 726–741, Mar. 2018. (Cited on page 17.)

[123] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry*. New York, London, 1963, vol. 1. (Cited on page 21.)

[124] A. R. Teel and L. Praly, "Global stabilizability and observability imply semi-global stabilizability by output feedback," *Systems and Control Letters*, no. 22, pp. 313–325, 1994. (Cited on page 25.)

[125] S. P. Bhata and D. S. Bernstein, "Atopological obstruction to continuous global stabilization of rotational motion and the unwinding phenomenon," *Systems & Control Letters*, vol. 39, pp. 63–70, 2000. (Cited on page 25.)

[126] J. Fuentes-Pacheco, J. Ruiz-Ascencio, and J. M. Rendón-Mancha, "Visual simultaneous localization and mapping: A survey," *Artificial Intelligence Review*, vol. 43, no. 1, pp. 55–81, Jan. 2015. (Cited on pages 27 and 41.)

[127] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, and J. J. Leonard, "Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age," *IEEE Transactions on robotics*, vol. 32, no. 6, pp. 1309–1332, 2016. (Cited on pages 27 and 28.)

[128] J.-J. E. Slotine and W. Li, *Applied Nonlinear Control*. Englewood Cliffs, N.J: Prentice Hall, 1991. (Cited on page 38.)

[129] F. Schill, R. Mahony, and P. Corke, "Estimating ego-motion in panoramic image sequences with inertial measurements," in *Robotics Research*. Springer, 2011, pp. 87–101. (Cited on page 43.)

[130] S. Umeyama, "Least-squares estimation of transformation parameters between two point patterns," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, no. 4, pp. 376–380, 1991. (Cited on page 43.)

[131] D. Simon and T. L. Chia, "Kalman filtering with state equality constraints," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 38, no. 1, pp. 128–136, Jan. 2002. (Cited on page 49.)

[132] J. M. Lee, "Smooth Manifolds," in *Introduction to Smooth Manifolds*, ser. Graduate Texts in Mathematics, J. M. Lee, Ed. New York, NY: Springer, 2012. (Cited on page 51.)

[133] M. Zamani, J. Trumpf, and R. Mahony, "Minimum-energy filtering for attitude estimation," *IEEE Transactions on Automatic Control*, vol. 58, no. 11, pp. 2917–2921, 2013. (Cited on pages 51 and 59.)

[134] B. N. Stovner, T. A. Johansen, T. I. Fossen, and I. Schjølberg, "Attitude estimation by multiplicative exogenous Kalman filter," *Automatica*, vol. 95, pp. 347–355, Sep. 2018. (Cited on page 51.)

[135] P. Morin, A. Eudes, and G. Scandaroli, "Uniform Observability of Linear Time-Varying Systems and Application to Robotics Problems," in *Geometric Science of Information*, F. Nielsen and F. Barbaresco, Eds. Cham: Springer International Publishing, 2017, pp. 336–344. (Cited on page 58.)

[136] B. Delyon, "A note on uniform observability," *IEEE Transactions on Automatic Control*, vol. 46, no. 8, pp. 1326–1327, 2001. (Cited on page 58.)

[137] A. J. Krener, "The convergence of the extended Kalman filter," in *Directions in Mathematical Systems Theory and Optimization*. Springer, 2003, pp. 173–182. (Cited on pages 58 and 59.)

[138] M. Zamani, J. Trumpf, and R. Mahony, "Nonlinear attitude filtering: A comparison study," *arXiv preprint arXiv:1502.03990*, 2015. (Cited on page 59.)

[139] A. Barrau and S. Bonnabel, "Invariant particle filtering with application to localization," in *53rd IEEE Conference on Decision and Control*, Dec. 2014, pp. 5599–5605. (Cited on pages 69, 73, 74, 75, and 90.)

[140] C. Hertzberg, R. Wagner, U. Frese, and L. Schröder, "Integrating generic sensor fusion algorithms with sound state representations through encapsulation of manifolds," *Information Fusion*, vol. 14, no. 1, pp. 57–77, Jan. 2013. (Cited on pages 69, 70, and 74.)

[141] T. Zhang, K. Wu, J. Song, S. Huang, and G. Dissanayake, "Convergence and Consistency Analysis for a 3-D Invariant-EKF SLAM," *IEEE Robotics and Automation Letters*, vol. 2, no. 2, pp. 733–740, Apr. 2017. (Cited on page 70.)

[142] J. Civera, A. Davison, and J. Montiel, "Inverse Depth Parametrization for Monocular SLAM," *IEEE Transactions on Robotics*, vol. 24, no. 5, pp. 932–945, Oct. 2008. (Cited on pages 70 and 77.)

[143] M. Brossard, A. Barrau, and S. Bonnabel, "Exploiting Symmetries to Design EKFs With Consistency Properties for Navigation and SLAM," *IEEE Sensors Journal*, vol. 19, no. 4, pp. 1572–1579, Feb. 2019. (Cited on page 82.)

[144] G. Guennebaud, B. Jacob *et al.*, "Eigen v3," 2010. (Cited on page 84.)

[145] P. van Goor, "General Invariant Feature Tracker (GIFT)," Nov. 2021. (Cited on page 84.)

[146] Y. Ge, P. van Goor, and R. Mahony, "Equivariant Filter Design for Discrete-time Systems," in *To Appear in 2022 61st IEEE Conference on Decision and Control (CDC)*, 2022. (Cited on page 85.)

[147] J. Shi and Tomasi, "Good features to track," in *1994 Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 1994, pp. 593–600. (Cited on page 86.)

[148] S. Baker and I. Matthews, "Lucas-Kanade 20 Years On: A Unifying Framework," *International Journal of Computer Vision*, vol. 56, no. 3, pp. 221–255, Feb. 2004. (Cited on page 86.)

[149] E. Rosten and T. Drummond, "Machine Learning for High-Speed Corner Detection," in *Computer Vision – ECCV 2006*, ser. Lecture Notes in Computer Science, A. Leonardis, H. Bischof, and A. Pinz, Eds. Berlin, Heidelberg: Springer, 2006, pp. 430–443. (Cited on page 86.)

[150] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *2011 International Conference on Computer Vision*, Nov. 2011, pp. 2564–2571. (Cited on page 86.)

[151] S. Leutenegger, M. Chli, and R. Y. Siegwart, "BRISK: Binary Robust invariant scalable keypoints," in *2011 International Conference on Computer Vision*, Nov. 2011, pp. 2548–2555. (Cited on page 86.)