An Observer Design for Visual Simultaneous Localisation and Mapping with Output Equivariance

Pieter van Goor * **Robert Mahony** * **Tarek Hamel** ** **Jochen Trumpf** *

* *Australian Centre for Robotic Vision, Research School of Electrical, Energy and Materials Engineering, Australian National University, Australia (email:* `{first name}.{last name}@anu.edu.au`*)*
** *I3S (University Côte d'Azur, CNRS, Sophia Antipolis) and Insitut Universitaire de France, (email:* `thamel@i3s.unice.fr`*)*

Abstract: Visual Simultaneous Localisation and Mapping (VSLAM) is a key enabling technology for small embedded robotic systems such as aerial vehicles. Recent advances in equivariant filter and observer design offer the potential of a new generation of highly robust algorithms with low memory and computation requirements for embedded system applications. This paper studies observer design on the symmetry group proposed in (van Goor et al., 2019), in the case where inverse depth measurements are available. Exploiting this symmetry leads to a simple fully non-linear gradient based observer with almost global asymptotic and local exponential stability properties. Simulation experiments verify the observer design, and demonstrate that the proposed observer achieves similar accuracy to the widely used Extended Kalman Filter with significant gains in processing time (linear verses quadratic bounds with respect to number of landmarks) and qualitative improvements in robustness.

Keywords: Equivariant Observer, Visual Odometry (VO), Simultaneous Localisation and Mapping (SLAM).

1. INTRODUCTION

Visual Simultaneous Localisation and Mapping (VSLAM) and the closely related Visual Odometry (VO) are established topics in the robotics community (Kaess et al., 2012; Leutenegger et al., 2015; Faessler et al., 2016; Forster et al., 2017b,a). They are key components of almost all aerial robotic systems (Delmerico and Scaramuzza, 2018) and are used in a host of other robotic applications (Bonin-Font et al., 2008) including autonomous driving and underwater robotics. VSLAM is used to refer to the case of the general SLAM problem where the available measurements are the bearings of landmarks such as provided by image features obtained using a monocular camera. Visual Odometry is a variant of the VSLAM problem where the solution is optimised for local consistency of the localisation of the system and updates landmark states only for currently visible landmarks. While the VO community has focused on embedded systems applications, and places a premium on algorithms with low computational and memory requirements (Delmerico and Scaramuzza, 2018), the VSLAM community has placed a premium on large scale map building, loop closure and accuracy (Stachniss et al., 2016; Cadena et al., 2016). As a consequence, many state-of-the-art VO systems use filter based formulations (Mourikis and Roumeliotis, 2007; Bloesch et al., 2015; Forster et al., 2017b; Lynen et al., 2013) in contrast to the full trajectory smoothing and graph based optimization formulation accepted as the community standard for SLAM problems (Cadena et al., 2016). Well engineered trajectory smoothing algorithms using short sliding-windows are still highly competitive algorithms for VO (Kaess et al., 2012; Leutenegger et al., 2015; Qin et al., 2017; Forster et al., 2017a).

The non-linear observer community has become interested in the visual SLAM and VO problem in the last few years. Work by Guerrerio et al.(Guerreiro et al., 2013) and Lourenço et al.(Lourenço et al., 2016) propose a non-linear observer for the "robo-centric" SLAM problem. Recent work by Barrau et al.(Barrau and Bonnabel, 2016, 2017) introduce a symmetry group $\mathbf{SE}_{n+1}(3)$ for the SLAM problem and use this to derive an Invariant Kalman Filter algorithm that overcomes consistency issues that have plagued the EKF algorithms from the classical SLAM era (Dissanayake et al., 2011; Zhang et al., 2017). Parallel work by Mahony et al.(Mahony and Hamel, 2017) show that this symmetry acts transitively on a principle fibre bundle $\mathcal{M}_n(3)$ which forms a natural geometric state-space for the SLAM problem, overcoming the gauge uncertainty present in the usual pose-map state representation. However, the symmetry induced by the group $\mathbf{SE}_{n+1}(3)$ applies only to the SLAM configuration state and is not compatible with bearing measurements. As a consequence, applying the $\mathbf{SE}_{n+1}(3)$ symmetry to the visual SLAM problem still requires linearisation of the output map. A new symmetry for the VSLAM problem was proposed in (van Goor et al., 2019) along with a non-linear observer. However, in this prior work the observer is closely based on (Hamel and Samson, 2016) and is derived in local coordinates and then lifted onto the symmetry group. It is of interest to consider

the case where the observer is designed explicitly using the symmetry structure.

In this paper, we present a highly robust, simple, and computationally cheap non-linear observer for the visual SLAM problem based on the new symmetry of the SLAM configuration space, first presented in (van Goor et al., 2019). The symmetry is associated with the novel $\mathbf{VSLAM}_n(3)$ Lie-group, which acts on the raw pose-map configuration coordinates and is compatible with the SLAM configuration manifold $\mathcal{M}_n(3)$ (Mahony and Hamel, 2017). The symmetry group structure introduced allows direct application of previous work by Mahony et al.(Mahony et al., 2013) in development of non-linear observers to yield a novel observer for continuous-time VSLAM. In the design of this observer, it is assumed measurements of the inverse depths of landmarks are available in addition to the bearing measurements. In practice, the inverse depth may be measured by using optical flow, triangulation, or depth cameras. The resulting algorithm is fully non-linear; no linearisation is required of the system or output maps. The approach has the advantage that constant gains can be used in the filter (no Riccati gains need be computed on-line) leading to lower computation and memory requirements. This additionally leads to a reduction in the number of parameters that need to be tuned in comparison with a standard EKF, making the proposed filter simpler to use in practice. The inherent symmetry of the approach ensures high levels of robustness and Theorem 5.1 proves almost global asymptotic and local exponential stability of the error coordinates. The convergence properties of the filter are demonstrated through a simulation experiment. Additional simulation experiments compare an EKF with our observer. These show that our observer achieves comparable mean RMSE to the EKF and has fewer outliers, and operates with a computational complexity that is only linear in the number of landmarks compared to quadratic complexity for the EKF.

2. NOTATION

The special orthogonal group is the set of rotation matrices and is denoted $\mathbf{SO}(3)$ with Lie algebra $\mathfrak{so}(3)$. The special Euclidean group is the set of rigid body transformations and is denoted $\mathbf{SE}(3)$ with Lie algebra $\mathfrak{se}(3)$. The group of positive real numbers equipped with multiplication is denoted $\mathbf{MR}$ with Lie algebra $\mathfrak{mr}$. We use the notation $R_P \in \mathbf{SO}(3)$ and $x_P \in \mathbb{R}^3$ to denote the rotation and translation components of a rigid-body transformation $P \in \mathbf{SE}(3)$ and write

$$P = \begin{pmatrix} R_P & x_P \\ 0 & 1 \end{pmatrix}. \qquad (1)$$

The pose of a vehicle moving in Euclidean space is written $P \in \mathbf{SE}(3)$. The kinematics of such a pose frame are written as

$$\dot{P} = PU, \quad \dot{R}_P = R_P \Omega_U^\times, \quad \dot{x}_P = R_P V_U \qquad (2)$$

where $\Omega_U = (\Omega_1, \Omega_2, \Omega_3)^\top$ and V_U are the body-fixed rotational and translational velocity vectors, respectively, and

$$U = (\Omega_U^\times, V_u)^\wedge := \begin{pmatrix} \Omega_U^\times & V_U \\ 0 & 0 \end{pmatrix}, \quad \Omega_U^\times = \begin{pmatrix} 0 & -\Omega_3 & \Omega_2 \\ \Omega_3 & 0 & -\Omega_1 \\ -\Omega_2 & \Omega_1 & 0 \end{pmatrix}. \qquad (3)$$

One has that $\Omega_U^\times w = \Omega_U \times w$ for any $w \in \mathbb{R}^3$ where $\times$ refers to the vector (cross) product.

For a unit vector $y \in \mathrm{S}^2 \subset \mathbb{R}^3$, the projector is given by
$$\Pi_y := I_3 - yy^\top, \qquad (4)$$
and has the property $\Pi_y = -y^\times y^\times$.

3. PROBLEM FORMULATION

The *total space* coordinates for the SLAM problem are defined with respect to an unknown fixed but arbitrary reference $\{0\}$. Let $P \in \mathbf{SE}(3)$ represent the body-fixed frame coordinates of the robot with respect to this reference frame. Let

$$p_i \in \mathbb{R}^3, \qquad i = 1, \ldots, n,$$

be sparse points in the environment expressed with respect to the reference frame $\{0\}$. The *total space* of the SLAM problem is the product space

$$\mathcal{T}_n(3) = \mathbf{SE}(3) \times \mathbb{R}^3 \times \cdots \times \mathbb{R}^3 \qquad (5)$$

made up of these raw coordinates $\Xi := (P, p_1, \ldots, p_n)$. The bearing of a point p_i co-located with the robot pose centre x_P is undefined, so the VSLAM problem can only be considered on the *reduced total space*

$$\mathcal{T}_n^\circ(3) = \{(P, p_i) \in \mathcal{T}_n(3) \mid (\forall i)\ p_i \neq x_P\}. \qquad (6)$$

Moreover, since all the measurements of the VSLAM problem considered are made in the body-fixed frame the solution is only well defined up to an $\mathbf{SE}(3)$ gauge transformation (Kanatani and Morris, 2001). This property can be expressed as an invariance of the problem formulation and leads to the quotient structure of the SLAM manifold proposed in (Mahony and Hamel, 2017). To keep the derivation simple and more accessible, in the present paper we will define the group actions and derive the observer on the reduced total space.

The measurements considered are spherical coordinates y_i of body-fixed frame observations of points in the environment, which in practice may be obtained from a calibrated monocular camera. Additionally, in this analysis, we assume inverse depth estimates z_i are also available. That is, for a given robot pose P and environment point p_i,

$$y_i := \frac{R_P^\top(p_i - x_P)}{|p_i - x_P|}, \qquad (7a)$$

$$z_i := |p_i - x_P|^{-1}. \qquad (7b)$$

The combined output space is $\mathcal{N}_n(3) = (S^2 \times \mathbb{R}^+) \times \cdots \times (S^2 \times \mathbb{R}^+)$ and we write $(y, z) = h(\Xi) = (h_1(\Xi), \ldots, h_n(\Xi)) = ((y_1, z_1), \ldots, (y_n, z_n))$, where appropriate.

Let $U \in \mathfrak{se}(3)$ denote the velocity of the robot. Assume that the environment points p_i being observed are static, and thus do not have a velocity. The tangent space of $\mathcal{T}_n(3)$ at a point $\Xi = (P, p_1, \ldots, p_n)$ can be identified with the matrix subspace

$$(PU, u_1, ..., u_n), \quad U \in \mathfrak{se}(3), \quad u_1, ..., u_n \in \mathbb{R}^3.$$

The system kinematics can then be written as

$$\frac{\mathrm{d}}{\mathrm{d}t}(P, p_1, \ldots, p_n) = (PU, 0, \ldots, 0). \qquad (8)$$

We assume that the robot velocity $U \in \mathfrak{se}(3)$ is measured. We will also measure the optical flow of each landmark

$$\phi_i \in T_{y_i} \mathrm{S}^2 \subset \mathbb{R}^3$$

by numerically differentiating the coordinates of y_i between consecutive measurements. Here, we express $\phi_i \in T_{y_i}S^2$ using the coordinates obtained by embedding $S^2 \subset \mathbb{R}^3$. Define a measurement velocity space

$$\mathbb{V} = \mathfrak{se}(3) \times \mathbb{R}^3 \times \cdots \times \mathbb{R}^3 \qquad (9)$$

with elements $(U, \phi_1, \ldots, \phi_n)$.

4. SYMMETRY OF THE VSLAM PROBLEM

The symmetry group of the VSLAM for n landmarks in Euclidean 3-space with separate bearing and range measurements problem is the *visual SLAM group* $\mathbf{VSLAM}_n(3)$ first described in (van Goor et al., 2019). However, in this paper the VSLAM group and its actions are presented in a different form to (van Goor et al., 2019).

In the following, we will write $(Q_A, a)_1$ instead of the more formal $((Q_A)_1, a_1)$ and sometimes write $(Q_A, a)_i$ to represent the tuple $(Q_A, a)_1, \ldots, (Q_A, a)_n$. Similarly, we will sometimes write (P, p_i) instead of $(P, p_1, \ldots, p_n)$.

The VSLAM group (van Goor et al., 2019) may be written

$$\mathbf{VSLAM}_n(3) = \{(A, (Q_A, a)_1, \ldots, (Q_A, a)_n) \mid$$
$$A \in \mathbf{SE}(3), \ (Q_A)_i \in \mathbf{SO}(3), \ a_i \in \mathbf{MR}, \ i = 1, \ldots, n\}.$$

Lemma 4.1. The set $\mathbf{VSLAM}_n(3)$ is a Lie group, defined as

$$\mathbf{VSLAM}_n(3) := \mathbf{SE}(3) \times (\mathbf{SO}(3) \times \mathbf{MR})^n.$$

The visual SLAM group acts as a symmetry group on the reduced total space $\mathcal{T}_n^\circ(3)$.

Lemma 4.2. The mapping $\Upsilon : \mathbf{VSLAM}_n(3) \times \mathcal{T}_n^\circ(3) \to \mathcal{T}_n^\circ(3)$ defined by

$$\Upsilon((A, (Q_A, a)_i), (P, p_i))$$
$$= (PA, (a^{-1}R_{PA}Q_A^\top R_P^\top(p - x_P) + x_{PA})_i), \qquad (10)$$

is a right group action of $\mathbf{VSLAM}_n(3)$ on $\mathcal{T}_n^\circ(3)$.

The group action for the robot pose is rigid-body transformation. The group action for environment points is considerably more subtle and can be understood conceptually as a sequence of operations: firstly, the reference frame coordinates of an environment point are written in the body-fixed frame, this point is then rotated by $Q_A^\top$ and then scaled by a^{-1}, before these body-fixed frame coordinates are rewritten in the inertial frame using the *new body-fixed frame* reference.

A key property of the proposed structure is that there is a compatible group operation on the output $\mathcal{N}_n(3)$ of the system.

Lemma 4.3. The action $\rho : \mathbf{VSLAM}_n(3) \times \mathcal{N}_n(3) \to \mathcal{N}_n(3)$ defined by

$$\rho((A, (Q_A, a)_i), (y, z)_i) := ((Q_A^\top y, az)_i) \qquad (11)$$

is a transitive right action on $\mathcal{N}_n(3)$. Furthermore, one has

$$\rho((A, (Q_A, a)_i), h(\Xi)) = h(\Upsilon((A, (Q_A, a)_i), \Xi))$$

where h is given by (7). That is, h is equivariant with respect to the actions Υ and ρ.

4.1 Lift of the SLAM kinematics

A key aspect of the proposed approach is that the symmetry group $\mathbf{VSLAM}_n(3)$ and the reduced total space $\mathcal{T}_n^\circ(3)$ are quite different spaces. The difference is particularly
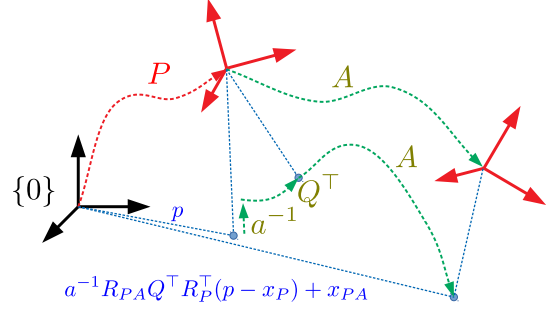

Fig. 1. Group action $\Upsilon((A, (Q_A, a)_i), (P, p_1, \ldots, p_n))$. The pose $P \mapsto PA$, that is the tip point of the pose is updated by the correction $A \in \mathbf{SE}(3)$. The body fixed-frame environment points $R_A^\top(p_i - x_P)$ are rotated by $Q_A^\top$ and scaled by a^{-1} in the body-fixed frame before *transforming with the robot pose* to a new point p_i' which is rewritten in the inertial frame.

clear in studying the structure of the lifted kinematics on the $\mathbf{VSLAM}_n(3)$ group.

The Lie-algebra of $\mathbf{VSLAM}_n(3)$ is

$$\mathfrak{vslam}_n(3) = \mathfrak{se}(3) \times (\mathfrak{so}(3) \times \mathfrak{mr}) \times \cdots \times (\mathfrak{so}(3) \times \mathfrak{mr}).$$

We write $(U, (W, w)_i) \in \mathfrak{vslam}_n(3)$ with $U \in \mathfrak{se}(3)$, $W_i \in \mathfrak{so}(3)$ and $w_i \in \mathfrak{r}$, where

$$U = \begin{pmatrix} \Omega_U^\times & V_U \\ 0 & 0 \end{pmatrix}.$$

In order to implement an observer on the VSLAM group, it is necessary to lift the velocity measurements $(U, \phi_i) \in \mathbb{V}$ (9) to elements $(U, (W, u)_i) \in \mathfrak{vslam}_n(3)$ such that the resulting group velocity is compatible with the system kinematics. That is, an algebraic map $\lambda : \mathcal{T}_n^\circ(3) \times \mathbb{V} \to \mathfrak{vslam}_n(3)$ is required such that

$$\mathrm{d}\Upsilon_{(P, p_i)}\lambda((P, p_i), (U, \phi_i)) = (PU, 0, \ldots, 0). \qquad (12)$$

Proposition 4.4. The map $\lambda : \mathcal{T}_n^\circ(3) \times \mathbb{V} \to \mathfrak{vslam}_n(3)$ defined by

$$\lambda((P, p_i), (U, \dot{y}_i)) := \left(U, \left((\phi \times y)^\times, zy^\top V_U\right)_i\right), \qquad (13)$$

where $(y, z)_i = h((P, p_i))$ satisfies the lift condition (12).

Proof. Under the static landmark assumption $\dot{p}_i = 0$, the optic flow $\phi_i = \dot{y}_i$ is given by

$$\phi_i = -\Omega_U^\times y_i - z_i(I - y_i y_i^\top)V_U. \qquad (14)$$

Let $(\lambda_A, (\lambda_Q^\times, \lambda_a)_i) := \lambda((P, p_i), (U, \phi_i))$. Evaluating the left-hand side of (12), one has

$$\mathrm{d}\Upsilon_{(P, p_i)}\lambda((P, p_i), (U, \phi_i))$$
$$= \mathrm{D}\Upsilon_{(P, p_i)}(\mathrm{id})[(\lambda_A, (\lambda_Q^\times, \lambda_a)_i)],$$
$$= (P\lambda_A, (-\lambda_a(p - x_P) + R_P\Omega_{\lambda_A}^\times R_P^\top(p - x_P)$$
$$\quad - R_P\lambda_Q^\times R_P^\top(p - x_P) + R_P V_U)_i).$$

This expression may be written in terms of $(y, z)_i$ as follows.

$$\mathrm{d}\Upsilon_{(P, p_i)}\lambda((P, p_i), (U, \phi_i))$$
$$= (P\lambda_P, (-\lambda_a R_P z^{-1}y + R_P(\Omega_{\lambda_A}^\times - \lambda_Q^\times)z^{-1}y + R_P V_U)_i).$$

Multiply the landmark velocity terms by $R_P^\top$ and substitute in the values for λ to obtain

$$-\lambda_a z^{-1} y + (\Omega_{\lambda_A}^\times - \lambda_Q^\times) z^{-1} y + V_U$$
$$= -y y^\top V + z^{-1} (\Omega_U - (\phi \times y)^\times) y + V_U,$$
$$= z^{-1} \Omega_U^\times y - z^{-1} y^\times y^\times \phi + (I - y y^\top) V,$$
$$= z^{-1} \Omega_U^\times y - z^{-1} (\Omega_U^\times y_i + z_i (I - y_i y_i^\top) V_U) + (I - y y^\top) V,$$
$$= 0.$$

Hence $\mathrm{d}\Upsilon_{(P,p_i)} \lambda((P,p_i),(U,\phi_i)) = (P\lambda_A, R_P 0) = (PU, 0)$ as required.

The lifted velocity $(U,(W,u)_i) = \lambda((P,p_i),(U,\phi_i))$ induces kinematics on the symmetry group that project down to the state space trajectory. Since the group action is not free, the stabiliser of Υ is non-trivial, and there are directions in $\mathfrak{vslam}_n(3)$, in particular $y_i^\times \in \mathfrak{so}(3)$, that are not constrained by the lift requirement (12). The lift in direction $y_i^\times \in \mathfrak{so}(3)$ is chosen to be zero without loss of generality.

The lift λ will enable us to go on and apply the observer design methodology developed in (Mahony et al., 2013).

5. OBSERVER DESIGN

We approach the observer design by considering the lifted kinematics of the system on the symmetry group and designing the observer on $\mathbf{VSLAM}_n(3)$. Let $\Xi(t) = (P(t), p_i(t)) \in \mathcal{T}_n(3)$ be the 'true' configuration of the SLAM problem, noting that $\Xi(t)$ is defined relative to some arbitrary reference $\{0\}$. Let $X(t) = (A(t), (Q(t), a(t))_i) \in \mathbf{VSLAM}_n(3)$ and define the *lifted kinematics* (Mahony et al., 2013)

$$\frac{\mathrm{d}}{\mathrm{d}t} X(t) = X\lambda(\Upsilon(X,\Xi(0)),(U,\phi_i))$$
$$X(0) = \mathrm{id}. \tag{15}$$

Equation (15) evolves on the VSLAM group where $(U,\phi_i) \in \mathbb{V}$ are the measured velocities and λ is the lift function (13).

Choose an arbitrary origin configuration
$$\Xi^\circ = (P^\circ, p_i^\circ) \in \mathcal{T}_n(3).$$
If the initial condition $X(0) \in \mathbf{VSLAM}_n(3)$ of the lifted kinematics satisfies $\Upsilon(X(0), \Xi^\circ) = \Xi(0)$ then (15) induces a trajectory that satisfies
$$\Xi(t) = \Upsilon(X(t), \Xi^\circ) \in \mathcal{T}_n(3)$$
for all time (Mahony et al., 2013).

Let the observer be defined as
$$\hat{X} = (\hat{A}, (\hat{Q}, \hat{a})_i) \in \mathbf{VSLAM}_n(3).$$
The lifted kinematics (15) provide the internal model for the observer design. That is, the kinematics of the observer are given by

$$\frac{\mathrm{d}}{\mathrm{d}t} \hat{X}(t) = \hat{X}\lambda(\Upsilon(X,\Xi^\circ),(U,\phi_i)) - \Delta\hat{X},$$
$$\hat{X}(0) = \mathrm{id}, \tag{16}$$

where $\Delta \in \mathfrak{vslam}_n(3)$ is an innovation term to be assigned. Note that $\lambda(\Upsilon(X,\Xi^\circ),(U,\phi_i))$ is shown in (13) to depend only on the measured quantities y_i, z_i, U, ϕ_i, and therefore can be implemented in the observer kinematics (16). The configuration estimate generated by the observer is given by

$$\hat{\Xi} = (\hat{P}, \hat{p}_i) = \Upsilon(\hat{X}, \Xi^\circ) \in \mathcal{T}_n^\circ(3)$$
given the reference $\Xi^\circ \in \mathcal{T}_n(3)$.

Theorem 5.1. Consider the kinematics (15) evolving on $\mathbf{VSLAM}_n(3)$ along with bounded outputs $\mathbf{y} = \{(y,z)_i\} = h(\Xi(t)) \in \mathcal{N}_n(3)$ given by (7). Fix an arbitrary origin configuration $\Xi^\circ = (P^\circ, p_i^\circ)$ and define the output error $\mathbf{e} = \{(e_y, e_z)_i\}$ as

$$\mathbf{e} = \rho(\hat{X}^{-1}, \{(y,z)_i\}) = \rho(E^{-1}, \{(y^\circ, z^\circ)_i\}) \in \mathcal{N}_n(3). \tag{17}$$

where $\{(y^\circ, z^\circ)_i\} = h(\Xi^\circ)$ and $E = \hat{X} X^{-1}$. Consider the observer defined in (16) and choose the innovation term $\Delta := (\Delta_A, (\Delta_Q, \Delta_a)^i)$ as follows:

$$\Delta_Q^i := -k_{Q_i} (e_{y_i} \times y_i^\circ)^\times \tag{18}$$

$$\Delta_a^i := -k_{a_i} \frac{e_{z_i} - z_i^\circ}{e_{z_i}} \tag{19}$$

$$\Delta_A := -k_A \mathrm{Ad}_{\hat{A}} \begin{pmatrix} \Omega_\Delta^\times & V_\Delta \\ 0 & 0 \end{pmatrix}, \tag{20}$$

$$\begin{pmatrix} \Omega_\Delta \\ V_\Delta \end{pmatrix} := \left(\sum_{i=1}^n \begin{pmatrix} \Pi_{\hat{y}_i} & \hat{z}_i \hat{y}_i^\times \\ -\hat{z}_i \hat{y}_i^\times & \hat{z}_i^2 \Pi_{\hat{y}_i} \end{pmatrix} \right)^{-1} \left(\sum_{i=1}^n \begin{pmatrix} -\hat{y}_i^\times \phi_i \\ -\hat{z}_i \phi_i \end{pmatrix} \right) - \begin{pmatrix} \Omega_U \\ V_U \end{pmatrix},$$

where the gains k_{a_i}, k_{Q_i} and k_A are positive scalars (for $i = 1, \ldots, n$), and the matrix inverse in the definition of Δ_A is assumed to be well-defined. Then the configuration estimate $\hat{\Xi}(t) = \Upsilon(\hat{X}(t), \Xi^\circ)$ converges almost globally asymptotically and locally exponentially to the true state $\Xi(t) = \Upsilon(X(t), \Xi^\circ)$ up to a possibly time-varying element in $\mathbf{SE}(3)$.

Proof. Let $X(t) \in \mathbf{VSLAM}_n(3)$ satisfy the lifted kinematics (15) with $\Upsilon(X(0), \Xi^\circ) = \Xi(0)$. It follows that $\Xi(t) = \Upsilon(X(t), \Xi^\circ)$ (Mahony et al., 2013). Define

$$E = \hat{X} X^{-1} =: (\tilde{A}, (\tilde{Q}, \tilde{a})_i) \in \mathbf{VSLAM}_n(3), \tag{21}$$

with $\tilde{A} := \hat{A} A^{-1}$, $(\tilde{Q}, \tilde{a})_i = (\hat{Q} Q^\top, \hat{a} a^{-1})_i$. Using (15) and (16), it is straightforward to verify that

$$\dot{E} = (-\Delta_A \tilde{A}, (-\Delta_Q \tilde{Q}, -\Delta_a \tilde{a})_i). \tag{22}$$

Using the fact that $E^{-1} = (\tilde{A}^{-1}, (\tilde{Q}^\top, \tilde{a}^{-1})_i)$ then each element of equation (17) becomes

$$(e_{y_i}, e_{z_i}) = (\tilde{Q} y_i^\circ, \tilde{a}^{-1} z_i^\circ). \tag{23}$$

Based on (22), the error kinematics satisfy

$$(\dot{e}_{y_i}, \dot{e}_{z_i}) = (-\Delta_Q^i e_{y_i}, \Delta_a^i e_{z_i}). \tag{24}$$

We first prove almost-global asymptotic and local exponential stability of the equilibrium $(e_{y_i}, e_{z_i}) = (y_i^\circ, z_i^\circ)$ for the error kinematics (24). Consider the following candidate (positive definite) Lyapunov function $\mathcal{L} : \mathcal{N}_n(3) \to \mathbb{R}^+$,

$$\mathcal{L} = \frac{1}{2} \sum_{i=1}^n \left(|e_{y_i} - y_i^\circ|^2 + (e_{z_i} - z_i^\circ)^2 \right). \tag{25}$$

Differentiating $\mathcal{L}$ and using (18) and (19), one gets:

$$\dot{\mathcal{L}} = \sum_{i=1}^n \left((e_{y_i} - y_i^\circ)^\top \dot{e}_{y_i} + (e_{z_i} - z_i^\circ) \dot{e}_{z_i} \right),$$

$$= \sum_{i=1}^n \left(-(e_{y_i} - y_i^\circ)^\top \Delta_Q^i e_{y_i} + (e_{z_i} - z_i^\circ) \Delta_a^i e_{z_i} \right),$$

$$= -\sum_{i=1}^n \left(k_{y_i} |e_{y_i} \times y_i^\circ|^2 + k_{z_i} (e_{z_i} - z_i^\circ)^2 \right).$$

The time derivative of the Lyapunov function is negative definite and equal to zero when $e_{y_i} = \pm y_i^\circ$ and $e_{z_i} = z_i^\circ$. Direct application of Lyapunov's theorem ensures that

the equilibrium $(e_{y_i}, e_{z_i}) = (y_i^\circ, z_i^\circ)$ is almost-globally *asymptotically* stable [1].

To prove local exponential stability of the observer it suffices to split the Lyapunov function into two parts

$$\mathcal{L} = \mathcal{L}_y + \mathcal{L}_z$$

$\mathcal{L}_z = \frac{1}{2} \sum_{i=1}^n (e_{z_i} - z_i^\circ)^2$ and verify that $\dot{\mathcal{L}}_z \leq -2 \min(k_{z_i})\mathcal{L}_z$, with $\mathcal{L}_z$ converging exponentially to zero. Consider $\mathcal{L}_y = \sum_{i=1}^n \mathcal{L}_{y_i}$ with $\mathcal{L}_{y_i} = \frac{1}{2} |e_{y_i} - y_i^\circ|^2$. If there exists a positive number ϵ such that $\mathcal{L}_{y_i} \leq 2 - \epsilon$, that is $e_{y_i}(0)$ is not in the opposite direction of $y^\circ(0)$, for all $i = \{1, \ldots, n\}$, then

$$\dot{\mathcal{L}}_{y_i} \leq -2 \min(k_{y_i})\epsilon \mathcal{L}_{y_i}$$

This demonstrates local exponential stability (in a large domain) of the equilibrium $(e_{y_i}, e_{z_i}) = (y_i^\circ, z_i^\circ)$.

In the limit, at the stable equilibrium point $(e_{y_i}, e_{z_i}) = (y_i^\circ, z_i^\circ))$, (23) implies that

$$(\hat{y}, \hat{z})_i = \left(Q_{\hat{A}}^\top y^\circ, \hat{a} z^\circ \right)_i = \left(Q_{\hat{A}}^\top e_y, \hat{a} e_z \right)_i = (y, z)_i,$$

for all $i = 1, \ldots, n$. This in turn implies

$$\rho(\hat{X}, (y^\circ, z^\circ)_i) = h(\Upsilon(\hat{X}, \Xi^\circ)) = h(\Upsilon(X, \Xi^\circ))$$
$$= \rho(X, (y^\circ, z^\circ)_i)$$

Regarding just the central equality, and noting that h only preserves the relative pose on $\mathcal{T}_n^\circ(3)$, it follows that

$$\Xi = \Upsilon(X, \Xi^\circ) = \Upsilon(\hat{X}, \Xi^\circ) \cong \hat{\Xi}$$

The symbol $\cong$ indicates that $\Xi = \hat{\Xi}$ up to the possibly time-varying gauge transformation $\tilde{A}$. That is

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

This concludes the proof of the almost-global asymptotic and local exponential stability.

Remark 5.2. Observe that the output error **e** is independent of the **SE**(3) innovation Δ_A and the primary stability analysis in Theorem 5.1 is undertaken on the output space, not the state-space. This is a key property of the **VSLAM**$_n$(3) symmetry and is intrinsic to the invariance of the underlying SLAM problem discussed in Section §3. The particular choice of innovation Δ_A in (20) minimizes the least squares drift in the visual odometry error as observed from direct measurements of landmark coordinates and optical flow. This is only one of a family of possible choices (for example, Mahony and Hamel (2017)), however, further analysis of this question is beyond the scope of the present paper.

6. SIMULATION RESULTS

The first simulation experiment was conducted to verify the observer design in Theorem 5.1. A robot is simulated to move in a circle with velocity $V_U = (0.1, 0, 0)$ m/s, $\Omega_U = (0, 0, 0.02\pi)$ rad/s on the ground, with 10 landmarks uniformly distributed in a 0.5-1 m band around the robot's path. The reference configuration ξ° of the observer is randomly set, and the observer $\hat{X}$ is initialised to the identity group element. All landmarks are assumed to be measured at all times, and no noise is added to the system. The gains of the observer are set to $k_{Q_i} = 0.05$, $k_{a_i} = 0.02$, $k_A = 0.03$ for all $i = 1, ..., 10$. The observer equations are implemented with Euler integration using a

[1] It is straightforward to verify that the equilibrium point $e_{y_i} = -y_i^\circ$ is unstable.

time step of 0.5 s. Figure 2 shows the evolution of the Lyapunov function (25) components for each landmark over 100 s. The bearing storage refers to the component $l_y^i := \frac{1}{2}|e_{y_i} - y_i^\circ|^2$ and the inverse depth storage refers to the component $l_z^i := \frac{1}{2}|e_{z_i} - z_i^\circ|^2$ for each landmark index i. The top two plots show the value of these functions for each landmark, and the bottom two plots show the log value for each landmark. The plots clearly show the almost-global asymptotic and local exponential convergence of the observer's error system.
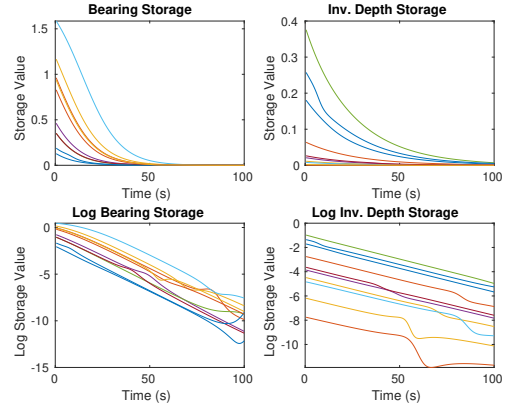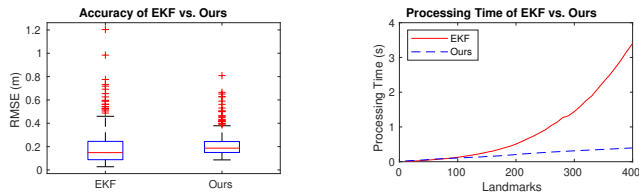

Fig. 2. The evolution of the individual components of the Lyapunov function (25) for 10 landmarks over time.

Additional simulations were carried out to compare the non-linear observer proposed in Theorem 5.1 with an Extended Kalman Filter (EKF). A robot is simulated to move in a circle with velocity $V_U = (0.1, 0, 0)$ m/s, $\Omega_U = (0, 0, 0.02\pi)$ rad/s on the ground, with n landmarks uniformly distributed in a 0.5-1 m band around the robot's path. The robot is modelled to have a sensor range of 1 m. The reference configuration ξ° is initialised without any landmarks, and the observer group element is initialised to identity. When landmarks are first seen, their inertial frame position is computed using the observer's current position estimate, and the reference configuration is augmented with this value. When landmarks are not within the sensing range, the observer equations cannot be used, and the current observer estimate of the landmark position is fixed until the landmark is next seen. All noise added to the input velocities and output measurements is drawn from zero-mean Gaussian distributions. The linear velocity noise has variance 0.2, the angular velocity noise has variance 0.1, the optical flow noise has variance 0.02, the bearing measurement noise has variance 0.01, and the inverse depth measurement noise has variance 0.4. The EKF is implemented with the system equation (8), and the measurement equations (7). The gains of the observer are set to $k_{Q_i} = 0.25$, $k_{a_i} = 0.1$, $k_A = 0.1$ for all $i = 1, ..., n$, and the observer equations were implemented using Euler integration with a time step of 0.5 s.

Figure 3a compares the statistics of the RMSE of the EKF and our observer for $n = 50$ landmarks after 100 s over 500 trials. While the EKF has a slightly lower mean RMSE, there are also more outliers due to linearisations errors. Figure 3b shows the mean computation time of the EKF and our observer for an increasing value of n between 10 and 400 landmarks over 500 trials per number of landmarks. While the processing time depends on the

(a) Boxplot of the RMSE of EKF and our observer on 50 landmarks over 500 trials.

(b) Mean computation time of EKF and our observer on 10-400 landmarks over 500 trials.

Fig. 3. Results of the simulation experiments comparing an EKF with the proposed observer.

implementation of the EKF and of our observer, the figure clearly illustrates the quadratic complexity of the EKF and the linear complexity of our observer.

7. CONCLUSION

This paper proposes a new symmetry for visual SLAM and VIO problems. This geometry is exploited to develop a visual SLAM observer and provide an almost global asymptotic and local exponential stability proof. The authors believe that the inherent simplicity and robustness of the proposed approach makes it useful as a tool for embedded robotics applications.

REFERENCES

Barrau, A. and Bonnabel, S. (2016). An ekf-slam algorithm with consistency properties. *arXiv:1510.06263*. URL `https://arxiv.org/abs/1510.06263v3`. ArXiv:1510.06263.

Barrau, A. and Bonnabel, S. (2017). The invariant extended kalman filter as a stable observer. *IEEE Transactions on Automatic Control*, 62(4), 1797–1812. doi: DOI:10.1109/TAC.2016.2594085.

Bloesch, M., Omari, S., Hutter, M., and Siegwart, R. (2015). Robust visual inertial odometry using a direct ekf-based approach. In *International Conference on Intelligent Robotics (IROS)*.

Bonin-Font, F., Ortiz, A., and Oliver, G. (2008). Visual navigation for mobile robots: A survey. *Journal of Intelligent Robotic Systems*, 53, 263–296. doi:DOI10.1007/s10846-008-9235-4.

Cadena, C., Carlone, L., Carrillo, H., Latif, Y., Scaramuzza, D., Neira, J., Reid, I., and Leonard, J. (2016). Past, present, and future of simultaneous localization and mapping: Towards the robust-perception age. *IEEE Transactions on Robotics*, 32(6), 1309–1332.

Delmerico, J. and Scaramuzza, D. (2018). A benchmark comparison of monocular visual-inertial odometry algorithms for flying robots. In *IEEE International Conference on Robotics and Automation (ICRA)*.

Dissanayake, G., Huang, S., Wang, Z., and Ranasinghe., R. (2011). A review of recent developments in simultaneous localization and mapping. In *International Conference on Industrial and Information Systems*, 477–482.

Faessler, M., Fontana, F., Forster, C., Mueggler, E., Pizzoli, M., and Scaramuzza, D. (2016). Autonomous, vision-based flight and live dense 3d mapping with a quadrotor MAV. *Journal of Field Robotics*, 33(4), 431–450.

Forster, C., Carlone, L., Dellaert, F., and Scaramuzza, D. (2017a). On-manifold preintegration for real-time visual-inertial odometry. *IEEE Transactions on Robotics*, 33(1), 1–21.

Forster, C., Zhang, Z., Gassner, M., Werlberger, M., and Scaramuzza, D. (2017b). Svo: Semidirect visual odometry for monocular and multicamera systems. *IEEE Transactions on Robotics*, 33(2), 249–265.

Guerreiro, B., Batista, P., Silvestre, C., and Oliveira, P. (2013). Globally asymptotically stable sensor-based simultaneous localization and mapping. *IEEE Transactions on Robotics*, 29(6), 1380–1395. doi:DOI:10.1109/TRO.2013.2273838.

Hamel, T. and Samson, C. (2016). Riccati observers for position and velocity bias estimation from direction measurements. In *2016 IEEE 55th Conference on Decision and Control (CDC)*, 2047–2053. doi:10.1109/CDC.2016.7798565.

Kaess, M., Johannsson, H., Roberts, R., Ila, V., Leonard, J., and Dellaert, F. (2012). isam2: Incremental smoothing and mapping using the bayes tree. *The International Journal of Robotics Research*, 31, 216–235. doi:doi.org/10.1177/0278364911430419. URL `https://doi.org/10.1177/0278364911430419`.

Kanatani, K. and Morris, D. (2001). Gauges and gauge transformations for uncertainty description of geometric structure with indeterminacy. *IEEE Transactions on Information Theory*, 47(5), 2017–2028.

Leutenegger, S., Lynen, S., Bosse, M., Siegwart, R., and Furgale, P. (2015). Keyframe-based visual-inertial slam using nonlinear optimization. *International Jouranl on Robotic Research*, 34(3), 314—-334.

Lourenço, P., Guerreiro, B., Batista, P., Oliveira, P., and Silvestre, C. (2016). Simultaneous localization and mapping for aerial vehicles: a 3-d sensor-based gas filter. *Autonomous Robots*, 40(5), 881–902.

Lynen, S., Achtelik, M., Weiss, S., Chli, M., and Siegwart, R. (2013). A robust and modular multi-sensor fusion approach applied to mav navigation. In *Proceedings of International Conference on Intelligent Robotics (IROS)*.

Mahony, R. and Hamel, T. (2017). A geometric nonlinear observer for simultaneous localisation and mapping. In *Conference on Decision and Control*, 6 pages. Melbourne.

Mahony, R., Trumpf, J., and Hamel, T. (2013). Observers for kinematic systems with symmetry. In *Proceedings of 9th IFAC Symposium on Nonlinear Control Systems (NOLCOS)*, 17 pages. Plenary paper.

Mourikis, A. and Roumeliotis, S. (2007). A multi-state constraint Kalman filter for vision-aided inertial navigation. In *Procedings of the International Conference on Robotics and Automation (ICRA)*.

Qin, T., Li, P., and Shen, S. (2017). Vins-mono: A robust and versatile monocular visual-inertial state estimator. arXiv:1708.03852.

Stachniss, C., Thrun, S., and Leonard, J. (2016). *Simultaneous Localization and Mapping. In B. Siciliano*, chapter 46, 1153—-1176. Springer, 2nd edition.

van Goor, P., Mahony, R.E., Hamel, T., and Trumpf, J. (2019). An equivariant observer design for visual localisation and mapping. *CoRR*, abs/1904.02452. URL `http://arxiv.org/abs/1904.02452`.

Zhang, T., Wu, K., Song, J., Huang, S., and Dissanayake, G. (2017). Convergence and consistency analysis for a 3-d invariant-EKF SLAM. *IEEE Robotics and Automation Letters*, 2(2), 733–740.